

***PERSONNEL ENSEIGNANT DE LA FACULTE***  
**ANNEE UNIVERSITAIRE 2014-2015**

**DOYEN : Pr Abdelhalim TRABELSI**

***PHARMACIENS HOSPITALO-UNIVERSITAIRES***

**PROFESSEUR EMERITE**

**MICROBIOLOGIE**..... Moncef JEDDI  
**HEMATOLOGIE**..... Kamel BOUKEF  
**BIOLOGIE CELLULAIRE**..... Hassen CHAABANI

**PROFESSEURS**

**MICROBIOLOGIE**..... Chédli FENDRI  
Noureddine BOUJAARFAR  
Noura Leïla SLIM  
Mohamed KHEDER  
Ridha BEN AISSA  
Abdelhalim TRABELSI  
Ridha M'ZOUGH  
Férièle MESSAADI  
Leïla BACCAR ép. BAKIR  
Farouk BARGUELLIL  
Afef MASMUDI  
Sophia BOUHALILA-BESBES  
Maha MASTOURI  
Amel EL FRAY-RHIM  
Soumaya KETATA  
Mohamed BEN MOUSSA  
Youssef BEN SALEM  
Imène FODHA  
Meriem ZRIBI  
Emna M'HIRI

**IMMUNOLOGIE** ..... Faouzi JENHANI  
Ibtissem GHEDIRA  
Salma FEKI  
Nabil SAKLY  
Ezzeddine GHAZOUANI

**HEMATOLOGIE**..... Nour El Houda TOUMI  
Mondher KORTAS  
Slama HMIDA  
Saloua JEMNI  
Touhami MAHJOUR  
Abderraouf OUESLATI  
Nasreddine GRITLI  
Mohsen HASSINE  
Najet MOJAAT  
Brahim NCIRI  
Salem ABID  
Mohamed ZILI  
Choumous KALLEL  
Houda KAABI  
Néjia JMILI-BRAHAM  
Batoul HOUISSA  
Sondos LAMLOUM ép. HIZEM

<b>BIOCHIMIE</b> .....	Mohamed Fadhel NAJJAR Néziha KAABACHI Ali BOUSLAMA Mohamed Chiheb BEN RAYANA Mohamed Adel BEN AMOR Taïeb MESSAOUD Kacem MAHDOUANI Haïfa SENHAJI Asma OMEZZINE Selima FERCHICHI Amina BIBI Zied AOUNI Fadoua NEFFATI - HASSINE
<b>PARASITOLOGIE</b> .....	Hamouda BABBA Sâadia FOURATI-GARGOURI
<b>TOXICOLOGIE</b> .....	Abderrazak LEHDHILI Youssef BRAHAM Wahiba DOUKI Hayet GHORBAL-BRAHAM Dorra AMIRA
<b>PHARMACIE GALENIQUE</b> .....	Souad SFAR GANDOURA
<b>PHARMACOLOGIE</b> .....	Nabil CHOUCHENE Nadia FENINA
<b>CHIMIE ANALYTIQUE</b> .....	Mohamed KALLEL Féthi SAFTA Farhat FARHAT
<b>PHARMACOGNOSIE</b> .....	Rachid CHEMLI Kamel GHEDIRA
<b>PHYSIOLOGIE HUMAINE ET EXPLORATION FONCTIONNELLE</b> .....	Thabet TABKA

### **MAITRES DE CONFERENCES AGREGES**

<b>BIOCHIMIE</b> .....	Chakib MAZIGH Nabila BEN REJEB Hassen BOUZIDI Henda CHAHED
<b>HEMATOLOGIE</b> .....	Fatma ABDERRAZAK Tahar CHAKROUN Inès JDIDI Manel CHAABANE
<b>MICROBIOLOGIE</b> .....	Anis BEL HAJ KHELIFA Asma GHARIANI Mohamed Salim ASLI
<b>IMMUNOLOGIE</b> .....	Radhia KOCHKAR Yousr GALAI
<b>PARASITOLOGIE</b> .....	Boutheïna JEMLI
<b>TOXICOLOGIE</b> .....	Kaïs GARRAB

<b>CYTOGENETIQUE</b> .....	Ghaya MERDESSI
<b>PHARMACIE GALENIQUE</b> .....	Imène LIMAYEM Emna ZRIBI Senda BAHRI Mohamed Ali EL YOUSFI
<b>PHARMACIE CLINIQUE</b> .....	Zeïneb OUAHCHI
<b>CHIMIE ANALYTIQUE</b> .....	Inès EL FRADI
<b>BIOPHYSIQUE PHARMACEUTIQUE</b> .....	Najla AYACHI
<b>CHIMIE THERAPEUTIQUE</b> .....	Olfa GLOULOU
<b>PHARMACOLOGIE</b> .....	Mehdi DRIDI
 <b><u>ASSISTANTS</u></b>	
<b>BIOCHIMIE</b> .....	Ilham HELLARA Yassine CHAABOUNI Lobna BOUACIDA Faten AMDOUNI Soumaya BERRIRI
<b>HEMATOLOGIE</b> .....	Saida BEN HATIRA Mourad CHAARI Amina BOUATTAY ép. BOUZOUTA Meriem BECHEUR Mouna SASSI Houda ROMDHANE Nada DAMMAK Inès GHARIANI Sami LAYOUNI Asma AYED Sonia CHOUAIEB
<b>MICROBIOLOGIE</b> .....	Chérifa CHAOUCH Selma SMAOUI Aïda ELARGOUBI Yomna BEN LAMINE Adam BEN NASR Hager BATIKH Salma SAMMOUD Sana KAMMOUN JMAL Chema MAROUANE MAAZOUN
<b>VIROLOGIE</b> .....	Saoussen KACEM
<b>PARASITOLOGIE</b> .....	Nawel SABBAH
<b>TOXICOLOGIE</b> .....	Meriem ZOUARI Asma GHORBEL Mohamed Hachem SAADAOUI
<b>CYTOGENETIQUE</b> .....	Elyes Slim GHEDIRA
<b>PHARMACIE GALENIQUE</b> .....	Ahmed AMRI Balsam KACEM
<b>PHARMACIE CLINIQUE</b> .....	Zakia Inès HARZALLAH Khouloud BEN JEDDOU Nesrine HASNI Dorra JARRAYA Ghassen BELHABIB

<b>CHIMIE ANALYTIQUE</b> .....	Amel HASSAIRI Hanène OUESLATI Inès GANA
<b>PHARMACOGNOSIE</b> .....	Houda BOUATTOUR Safa DERBAL Amine TRABELSI
<b>PHYSIOLOGIE</b> .....	Emna CHTOUROU Rabii SELMI Amani CHERIF
<b>CHIMIE THERAPEUTIQUE</b> .....	Fayçal JENAN Mohamed Ali ESSOUSSI
<b>PHARMACOLOGIE</b> .....	Manel TURKI Meriem RAZGALLAH – KHROUF Wafa CHAOUCH (Mise en disponibilité durant 1 an à partir du 1/6/2013)
<b>BIOLOGIE CELLULAIRE</b> .....	Amira BEN GAIED
<b>CHIMIE ORGANIQUE</b> .....	Donia GHEDIRA
<b>IMMUNOLOGIE</b> .....	Meriem GHOZZI

***PERSONNEL PHARMACIEN UNIVERSITAIRE***

**PROFESSEURS**

<b>PHARMACOLOGIE</b> .....	Abderrahman BOURAOUI Letaïef REKIK
<b>BIOPHYSIQUE</b> .....	Ridha HAMDANE

**ASSISTANT**

<b>PHARMACIE GALENIQUE</b> .....	Mohamed Ali LASSOUED
----------------------------------	----------------------

## ***PERSONNEL ENSEIGNANT SCIENTIFIQUE***

### **PROFESSEURS**

BIOCHIMIE.....	Jemni BEN CHIBANI Khira MAAROUFI
MICROBIOLOGIE.....	Mahjoub AOUNI Amina BEN FEDHILA-BAKHROUF
PHYSIOLOGIE ANIMALE.....	Dalila SAIDANE-MOSBAHI
PHYSIOLOGIE .....	Ahmed Nouredine HELLAL

### **Maître de Conférences**

PHYSIOLOGIE .....	Hassen BEN ABDENNABI
BOTANIQUE, PHYSIOLOGIE VEGETALE.....	Zohra MARZOUK
CHIMIE.....	Chedly BOUDOUKHANE

### **MAITRES ASSISTANTS**

BOTANIQUE, PHYSIOLOGIE VEGETALE.....	Saïda CHAKROUN
BOTANIQUE .....	Dhabbia KHIARI
PHARMACIE GALENIQUE.....	Taoufik DOUSS
CHIMIE .....	Béchir M'RABET Jamil KRAIEM Najah CHERIAA ép. JEGUIRIM Faouzi AYARI
CHIMIE ANALYTIQUE .....	Mouna TOUATI ép. BELHAJ YOUSSEF
PHARMACOGNOSIE.....	Fatma BANNOUR
SCIENCES BIOLOGIQUES PHARMACEUTIQUES.....	Meriem BEN HAMIDA REBII
PARASITOLOGIE.....	Najoua HOUAS (Coopération 5 ans à partir de 2013) Slim MRAD
BIOLOGIE CELLULAIRE .....	Soumaya KILANI
GENETIQUE.....	Sana SFAR
BIOPHYSIQUES.....	Sonia SFAR LAATAOUI Raja GUEDOUAR
IMMUNOLOGIE.....	Wahiba SAKLY

### **ASSISTANT**

CHIMIE .....	Hager TRAD
BIOLOGIE CELLULAIRE .....	Latifa CHKIOUA
BOTANIQUE .....	Ameur EL AISSI Hanène ALMADINI
PHYSIQUE.....	Hanène OUERFELLI

***PROFESSEUR DE L'ENSEIGNEMENT SECONDAIRE***

<b>ANGLAIS</b> .....	Mounira ZOUITEN-BESBES Fredj ALAYA
<b>INFORMATIQUE</b> .....	Anis NASRI Amel JABLAOUI
<b>EDUCATION PHYSIQUE</b> .....	Hichem EL BAKKARI

---

**Liste arrêtée le 12/09/2014**

# بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ قِسْمُ جَائِزَاتِ

بمضراساندة بهذه الكلية وزملائه في الدراسة

## اقسم

أَنْ أَجْلَ كُلِّ الَّذِينَ سَاهَمُوا فِي تَلْقِيَنِي مَبَادِيءَ  
فَنَ الصَّيْدَلَةِ وَأَعْتَرَلَهُمْ عَنَّا اعْتِرَافِي بِالْجَمِيلِ  
وَذَلِكَ بَوْلَانِي الدَّائِمِ لِتَعَالِيهِمْ  
وَأَنْ أُمَارَسَ مِهْنَتِي بِضَمِيرٍ خَالِصٍ وَأَنْ اتَّبِعَ عِنْدَ مَبَاشَرَةٍ  
مِهْنَةَ الصَّيْدَلَةِ لِأَلِقَوَانِينَ الْقَائِمَةِ وَالْجَارِي  
بِهَا الْعَمَلِ فَحَسَبَ بِرَجْمِ سَنَنِ الشَّرَفِ وَالزَّاهَةِ وَالتَّعَفُّفِ  
وَأَنْ لَا أَبُوحَ إِلَى أَحَدٍ بِالْأَسْرَارِ الَّتِي تَعَهَّدَ لِي  
وَالَّتِي قَدْ أَطْلَعَ عَلَيْهَا بِمُنَاسَبَةٍ مَمَارَسَتِي لِفَنِّ الصَّيْدَلَةِ  
وَأَنْيَّ بِمَحَافِظَتِي عَلَى هَذَا الْقِسْمِ وَتَمَسُّكِي بِهِ بِقُوَّةٍ  
أَكُونَ جَدِيرًا بِاحْتِرَامِ زَمَلَانِي وَتَقْدِيرِهِمْ وَحِرْفَانِي  
لِمَافِيهِ خَيْرَ الصَّيْدَلَةِ وَالصِّحَّةِ الْعَامَّةِ





## *Serment de Galien*

*Je jure en présence des maîtres de la faculté,  
des conseillers de l'ordre des pharmaciens  
et de mes conseillers:*

*D'honorer ceux qui m'ont instruit dans les préceptes de  
mon art et de leur témoigner ma reconnaissance en restant fidèle à  
leur enseignement.*

*D'exercer dans l'intérêt de la santé publique,  
ma profession avec conscience et de respecter non seulement la  
légalisation en vigueur, mais aussi les règles  
de l'honneur, de la probité et du désintéressement*

*De ne jamais oublier ma responsabilité et mes devoirs  
envers le malade et sa dignité humaine*

*En aucun cas, je consentirai à utiliser mes connaissances et mon  
état pour corrompre les mœurs et favoriser des actes criminels.*

*Que les hommes m'accordent leur estime si je suis fidèle  
à mes promesses. Que je sois couvert d'opprobre et méprisé  
de mes confrères si j'y manque.*



A ma mère

mon frère

ma famille

mes amis

et à tous ceux qui me soutiennent ...

## Remerciements

A notre maître et présidente du jury **Pr. Asma OMEZZINE**, je vous présente mes sincères remerciements de m'honorer en présidant le jury de ma thèse en vue de l'obtention du diplôme de Docteur en Pharmacie ;

A notre maître et juge **Pr. Imene FODHA**, je vous suis reconnaissant d'avoir accepté de faire partie du jury, je vous exprime ma profonde gratitude ;

A notre maître et juge **Pr. Catherine ETCHEBEST**, je vous suis reconnaissant de m'avoir accepté au Master Bioinformatique à l'Université Paris Diderot. Ce travail ne serait pas possible sans votre aide et conseil .Votre présence nous honore tous ;

A notre maître et directeur de thèse, **Mr le Doyen Pr. Abdelhalim TRABELSI**, je vous exprime mon profond respect et je vous remercie d'avoir accepté de m'encadrer pour ce travail ;

A notre maître et co-directeur de thèse, **Dr. Meriam BEN HAMIDA-REBAÏ**, je vous suis reconnaissant d'avoir accepté de m'encadrer à distance pour ce travail, ce qui a demandé beaucoup d'efforts. Je vous remercie pour votre temps et votre conseil de qualité. Je vous souhaite du courage pour votre entreprise.

## Sommaire

<b>1. INTRODUCTION</b>	<b>1</b>
<b>2. SYNTHÈSE BIBLIOGRAPHIQUE</b>	<b>2</b>
<b>2.1. Relation structure spatiale - activité</b>	<b>2</b>
2.1.1. Méthodes de prédiction de l'activité : <i>Quantitative Structure-Activity Relationship (QSAR)</i>	3
2.1.2. Exemples de relations démontrées entre la structure et la fonction d'une protéine	3
<b>2.2. Résolution de la structure</b>	<b>9</b>
2.2.1. Les techniques de résolution de la structure	10
2.2.2. Les limites de ces techniques	13
<b>2.3. Intérêt de la bioinformatique structurale</b>	<b>16</b>
2.3.1. Prédiction de la structure à partir de la séquence protéique	16
2.3.2. Apport de la modélisation des protéines membranaires dans la découverte et la conception des molécules à intérêt biothérapeutique	24
2.3.3. L'intérêt de la présente thèse	27
<b>3. MATÉRIELS ET MÉTHODES</b>	<b>29</b>
<b>3.1. La structure des protéines</b>	<b>30</b>
<b>3.2. HeliStatistix</b>	<b>30</b>
<b>3.3. Définitions des paramètres</b>	<b>31</b>
3.3.1. Calcul de l'axe d'une hélice	32
3.3.2. Cassure d'hélice	33
3.3.3. Moment hydrophobe d' <i>Eisenberg</i>	33
3.3.4. L'hydrophobicité moyenne d'une hélice	33
3.3.5. Calcul de l'angle de <i>tilt</i>	34
3.3.6. Contact entre deux hélices en interaction	34
3.3.7. Calcul de l'angle $\Omega$ de <i>packing</i>	35
3.3.8. Calcul de l'orientation du <i>packing</i>	35

<b>4. RESULTATS ET DISCUSSION</b>	<b>36</b>
4.1. La base de données	36
4.2. L'angle de <i>tilt</i>	38
4.2.1. Distribution des angles de <i>tilt</i>	38
4.2.2. Angle de <i>tilt</i> et la longueur d'hélice	40
4.2.3. Angle de <i>tilt</i> et hydrophobicité	43
4.2.4. Angle de <i>tilt</i> et composition en acides aminés	45
4.2.5. Angle de <i>tilt</i> et cassures d'hélices	47
4.2.6. Angle de <i>tilt</i> et hélices voisines	51
4.3. Angle de <i>packing</i>	52
4.3.1. Distribution des angles de <i>packing</i>	52
4.3.2. Angle de <i>packing</i> et hélices voisines de séquence	56
4.3.3. Autres paramètres des hélices en interaction et leurs angles de <i>packing</i>	58
<b>5. CONCLUSION</b>	<b>69</b>
<b>REFERENCES</b>	<b>I</b>
<b>ANNEXES</b>	<b>I</b>

## Liste des figures

<b>Figure 1</b> – Relation structure-activité-fonction des protéines.	2
<b>Figure 2</b> – Figure schématique de l'exodomaine de l'hétérodimère de l'intégrine désactivée.	4
<b>Figure 3</b> – Différentes étapes de l'activation des intégrines.	5
<b>Figure 4</b> – A – Figure schématique d'un canal à potassium. B – et sa composition en hélices transmembranaires par chaîne.	6
<b>Figure 5</b> – A – Figure schématique d'une protéine de transport membranaire et B – sa représentation en structures secondaires hélicoïdales par chaîne.	7
<b>Figure 6</b> – Mécanisme d'action de la Lactose Perméase.	8
<b>Figure 7</b> – La structure tridimensionnelle de la première protéine membranaire avec la représentation des hélices transmembranaires par chaîne.	9
<b>Figure 8</b> – Principe de la diffraction des rayons X par les cristaux.	11
<b>Figure 9</b> – Schéma des différentes étapes de la détermination de la structure par cristallographie à rayons X.	11
<b>Figure 10</b> – Figure schématique du principe de la résonance magnétique nucléaire.	12
<b>Figure 11</b> – Résolution de la structure protéique par RMN associée à la modélisation in silico.	13
<b>Figure 12</b> – Comparaison entre l'évolution des structures des protéines solubles et les protéines membranaires depuis 1985.	14
<b>Figure 13</b> – L'évolution du nombre de structures uniques de protéines membranaires depuis 1985.	14
<b>Figure 14</b> – Image d'un séquenceur de troisième génération.	17
<b>Figure 15</b> – Evolution du nombre de séquences de protéines par rapport aux nombre de structures résolues.	17

<b>Figure 16</b> – Les différentes étapes de la détermination de la structure tridimensionnelle d'une protéine à partir de la séquence d'acides aminés en utilisant la modélisation comparative.	19
<b>Figure 17</b> – Evolution de la performance des équipes participantes à la compétition <i>CASP</i> .	21
<b>Figure 18</b> – Taux de réussite dans le développement du médicament par classe thérapeutique.	26
<b>Figure 19</b> – Schéma représentant le fonctionnement de l'outil HeliStatistiX.	31
<b>Figure 20</b> – Schéma représentant l'axe d'une hélice et la méthode des calculs des paramètres.	32
<b>Figure 21</b> – La distribution des angles de tilt.	38
<b>Figure 22</b> – Probabilités des angles de tilt.	39
<b>Figure 23</b> – Distribution corrigée de l'angle de tilt.	40
<b>Figure 24</b> – La distribution de la longueur des hélices exprimée en nombre d'acides aminés.	41
<b>Figure 25</b> – L'angle de <i>tilt</i> est en corrélation statistiquement significative avec la longueur de l'hélice.	42
<b>Figure 26</b> – Distribution de l'hydrophobicité moyenne des hélices transmembranaires.	44
<b>Figure 27</b> – Distribution du moment hydrophobe des hélices transmembranaires.	44
<b>Figure 28</b> – Fréquence des acides aminés dans le premier et dernier tour d'hélice en fonction de l'angle de <i>tilt</i> .	47
<b>Figure 29</b> – Distribution des cassures d'hélices.	48
<b>Figure 30</b> – Angle de tilt et nombre de cassures par hélice.	49
<b>Figure 31</b> – Distribution des hélices en fonction des cassures.	50
<b>Figure 32</b> – Distribution des angles de cassures. Seul le plus grand angle de cassure de chaque hélice a été représenté.	50
<b>Figure 33</b> – Distribution du nombre d'hélices voisines à chaque hélice.	52
<b>Figure 34</b> – Histogramme de distribution des angles de packing.	53

<b>Figure 35</b> – La distribution de l’angle de packing dans les protéines solubles.	54
<b>Figure 36</b> – Distribution des angles de packing d’hélice qui ne sont pas en interaction.	54
<b>Figure 37</b> – Préférence des angles de packing après correction du biais statistique.	56
<b>Figure 38</b> – A – Distribution de la longueur de boucle de connexion des hélices voisines qui sont en interaction. B – Distribution de la longueur de boucle de connexion des hélices voisines qui ne sont pas en interaction.	57
<b>Figure 39</b> – A – Distribution des angles de packing des hélices qui sont voisines de séquence. B – Distribution des angles de packing des hélices qui ne sont pas voisines de séquence.	59
<b>Figure 40</b> – Distribution de l’angle de packing des hélices en orientation antiparallèle.	60
<b>Figure 41</b> – Distribution de l’angle de packing des hélices en orientation parallèle.	61
<b>Figure 42</b> – Distribution de l’angle de packing en fonction du nombre de contacts entre hélices.	62
<b>Figure 43</b> – Représentation de la distance minimale entre deux hélices en fonction de l’angle de packing.	63
<b>Figure 44</b> – L’influence des cassures des hélices sur l’angle de packing.	65
<b>Figure 45</b> – Corrélation entre le moment hydrophobe d’ <i>Eisenberg</i> des hélices en interaction et l’angle de packing.	67
<b>Figure 46</b> – Différence du moment hydrophobe entre les couples d’hélices en interaction en fonction de l’angle de packing.	68

## Liste des tableaux

<b>Tableau I</b> – L'échelle d'hydrophobicité de <i>Pliska</i> et <i>Fauchère</i> .	34
<b>Tableau II</b> – Caractéristiques des protéines étudiées.	37

## **Liste des abréviations**

**3D:** tridimensionnelle

**BLAST:** basic local alignment search tool

**CASP:** critical assessment of protein structure prediction

**CEA:** commissariat à l'énergie atomique et aux énergies renouvelables

**CSV:** comma separated values

**MQAP:** model quality assessment prediciton

**OPM:** orientation of proteins in membranes

**PDB:** protein data bank

**QSAR:** quantitative structure activity relationship

**RMN:** résonance magnétique nucléaire

## 1. INTRODUCTION

Les protéines membranaires jouent un rôle central dans la machinerie cellulaire (récepteurs, canaux à ions) et de ce fait constituent des cibles thérapeutiques idéales. Connaître la structure tridimensionnelle des protéines est donc un prérequis au développement de ligands efficaces et spécifiques. Les hélices  $\alpha$  sont les principales structures secondaires des protéines membranaires, qui leur assurent une topologie particulière. Elles sont donc directement impliquées dans la structure finale de la protéine. Les progrès effectués dans la résolution des structures tridimensionnelles des protéines a permis de rendre un grand nombre de structures protéiques disponibles. Cependant, le nombre de structures de protéines membranaires reste faible du fait de leur complexité structurale. La modélisation *in silico*, complète actuellement les méthodes biophysiques classiques, elle permet d'améliorer le processus de la résolution des structures.

L'analyse statistique des caractéristiques structurales de protéines membranaires permettra dans le présent travail de dégager les principales relations entre les hélices, principales structures secondaires des protéines membranaires, et l'environnement lipidique de la membrane. Cette analyse s'étend pour étudier la composition des hélices en contact avec le milieu hydrophile, ainsi que les caractéristiques des hélices telle que la topologie de l'hélice en termes de cassures et de longueur. Les corrélations statistiques ainsi révélées permettront de réduire l'espace de recherche possible des paramètres lors de l'inférence de la structure à partir de la séquence d'acides aminés de la protéine, un défi qui reste au premier plan pour la communauté scientifique.

Pour ce travail, un outil de calcul des angles entre hélices, entre l'hélice et la membrane protéique ainsi qu'une vingtaine de paramètres a été développé. Le même outil pourra être utilisé pour l'amélioration de la résolution de nouvelles structures protéiques membranaires dans le futur.

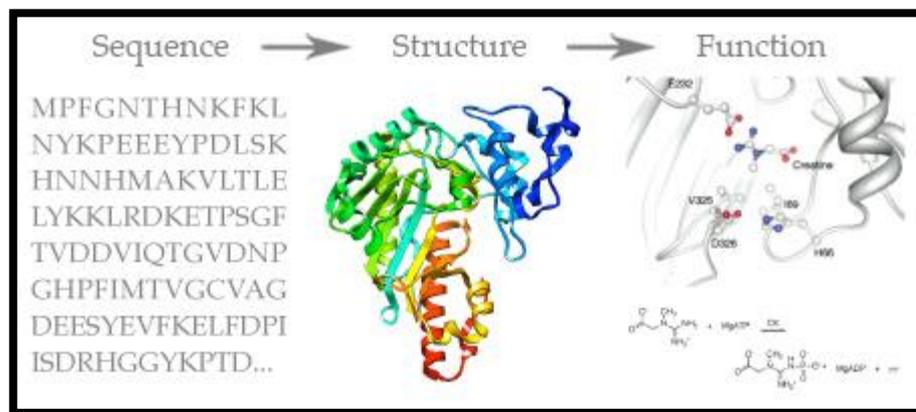
## 2. SYNTHÈSE BIBLIOGRAPHIQUE

Les récepteurs membranaires ont des structures et des fonctions diverses. La membrane biologique est formée d'une bicouche lipidique, dans laquelle sont insérées des protéines réceptrices. Selon le type de la cellule et la localisation dans la cellule, différentes protéines confèrent à la membrane ses fonctions spécifiques.

Un nombre relativement limité de structures et de conformations, se traduisant par l'arrangement spatial des atomes des acides aminés qui composent les protéines, assure une grande diversité des rôles.

### 2.1. Relation structure spatiale - activité

A l'échelle atomique, la fonction se traduit par l'interaction entre les atomes des protéines et ceux de l'effecteur, le ligand en l'occurrence. Des structures ou des séquences apparentées (Figure 1) peuvent renseigner sur des fonctions similaires (1).



**Figure 1** – Relation structure-activité-fonction des protéines. (Source : <http://persweb.wabash.edu>).

### **2.1.1. Méthodes de prédiction de l'activité : *Quantitative Structure-Activity Relationship (QSAR)***

Classiquement, la méthode QSAR s'intéresse à l'évaluation du potentiel d'une molécule candidate pour les essais cliniques, via l'étude de nombreux descripteurs de différentes natures. L'avènement des structures protéiques à haute résolution a permis l'utilisation de nouveaux descripteurs, qui prennent en compte la structure tridimensionnelle de la protéine cible, ce qui a permis d'affiner les modèles prédictifs.

Les modèles QSAR (2) sont des modèles de régression qui corrélient une variable de prédiction (structure chimique, propriétés physico-chimiques) à une variable de réponse (la fonction biologique). Un jeu de données permet d'abord établir une corrélation et de construire un modèle, qui permettra ensuite de prédire la fonction d'une nouvelle protéine dont la structure est récemment élucidée.

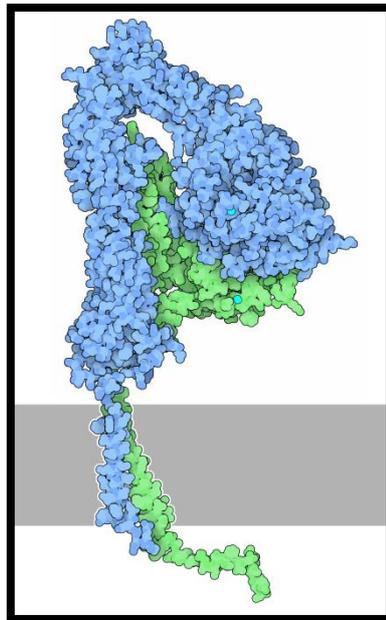
Les études de la dynamique moléculaire, science qui s'intéresse au comportement de la protéine cible au cours du temps, ont été également intégrées dans la méthode QSAR, pour former la QSAR 3D (1). Des descripteurs tels que le champ de force auquel est soumise la protéine, le champ électrostatique, les régions hydrophobes ainsi que les champs stériques ont permis d'avoir une approche qui englobe toute la protéine, plutôt que la recherche d'un fragment support de l'activité, dit pharmacophore.

### **2.1.2. Exemples de relations démontrées entre la structure et la fonction d'une protéine**

Cette section s'intéressera à trois exemples parmi les grandes classes fonctionnelles protéiques (transporteurs, canaux, protéines d'adhésion).

### 2.1.2.1 L'activation des intégrines

Les intégrines sont des protéines membranaires intégrales, polytopiques et hétérodimériques (Figure 2). Elles sont formées de deux sous unités :  $\alpha$  et  $\beta$ . La fixation d'un ligand, principalement de la famille des laminines, déclenche une cascade de signalisation intracellulaire, impliquée dans le mouvement et l'adhésion de la cellule. Seules les intégrines activées peuvent fixer le ligand. A ce jour, deux modèles d'activation des intégrines ont été décrits : un modèle *deadbolt* et un modèle *switchblade*, le plus communément accepté (3). Dans ce dernier modèle (4), l'intégrine passe d'une conformation repliée, inactive, à une conformation étendue, active, qui permet de fixer différents ligands.

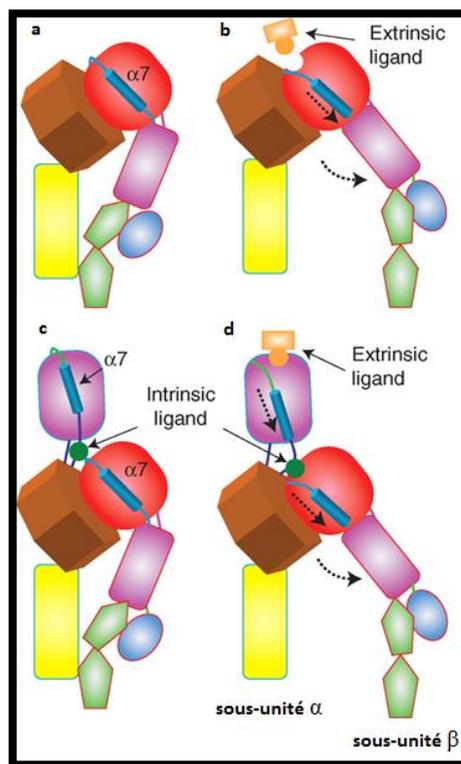


**Figure 2** – Figure schématique de l'exodomaine de l'hétérodimère de l'intégrine désactivée. Sous unité  $\alpha$  : vert, sous unité  $\beta$  : bleu. (Source : *Protein Data Bank*, code d'accès : 1JV2).

Celle-ci sont soit initiées par un facteur extrinsèque (4) et facteur intrinsèque dans deux sous types de la famille des intégrines.

Les hélices transmembranaires des sous unités  $\alpha$  et  $\beta$  jouent un rôle important dans l'activation de la protéine, donc dans son activité biologique. En effet, il a

été démontrée qu'à l'état inactif, les deux hélices transmembranaires de l'intégrine  $\alpha$ IIb $\beta$ III sont associées et en contact. L'hélice de la sous-unité  $\alpha$ IIb est perpendiculaire par rapport au plan de la membrane, et celle de la sous-unité  $\beta$ III est légèrement inclinée. La présence de glycines dans l'interface entre les deux hélices, ainsi que le rapprochement de deux paires consécutives de résidus phénylalanine de  $\alpha$ IIb vers  $\beta$ III, favorisent les interactions électrostatiques entre  $\alpha$ IIb(D723) and  $\beta$ III(R995). La disposition des hélices transmembranaires, reflète celles des segments extracellulaires ou éctodomaines, grâce à deux segments de liaison. En effet, le domaine extracellulaire (Figure 3) est composé de deux chaînes  $\alpha$ IIb et  $\beta$ III qui sont stabilisées à l'état inactif par plusieurs points de contacts. Egalement, l'hélice  $\alpha$ 7 de la sous-unité  $\beta$  masque au départ le site de liaison du ligand (5).



**Figure 3** – Différentes étapes de l'activation des intégrines (5).

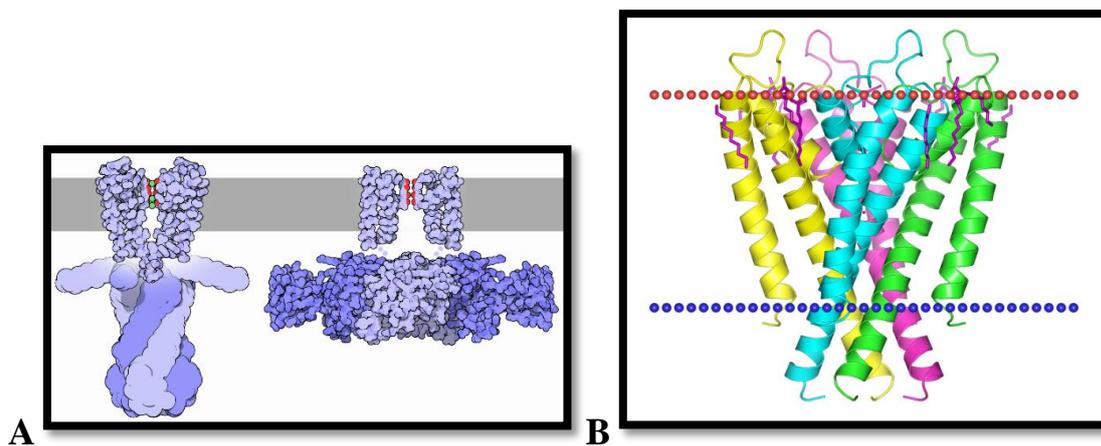
Le mouvement de l'hélice transmembranaire de  $\beta$ III sous l'influence d'un ligand extrinsèque (Figure 3a-b) ou intrinsèque (Figure 3c-d), d'un angle incliné par rapport à la membrane, à un angle qui lui y est perpendiculaire, se traduit dans

l'ectodomaine par un éloignement de III et IIb, pour abolir tous les contacts et permettre l'extension de la protéine, et son activation. Durant ce processus, ces changements impliquent également le mouvement de l'hélice  $\alpha 7$ , pour libérer le site actif et permettre la liaison du ligand.

La structure des hélices transmembranaires des intégrines, a permis donc de renseigner sur le mécanisme d'action et l'activation de ces protéines.

### 2.1.2.2. Les canaux à potassium

Les canaux à potassium sont les protéines membranaires intégrales polytopiques les plus étudiées. Ils permettent un passage passif des ions potassium et leur structure particulière explique leur fonction biologique (Figure 4 A-B).

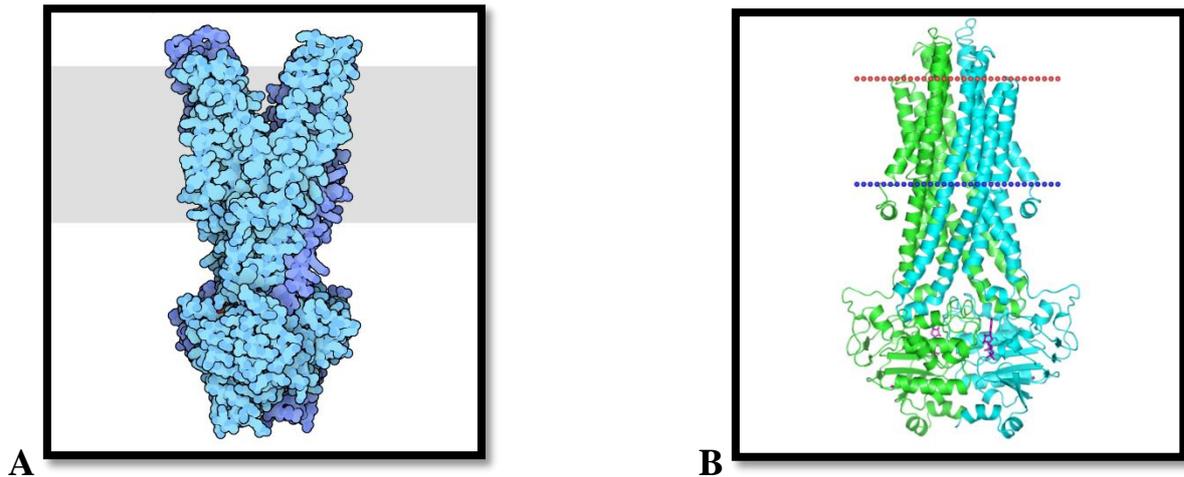


**Figure 4** – A – Figure schématique d'un canal à potassium. B – et sa composition en hélices transmembranaires par chaîne. (Source : Protein Data Bank, code d'accès : 1BL8).

En effet, la majorité des protons des amides de cette protéine sont non échangeables avec le solvant. Ceci reflète une structure rigide, neutre et passive vis-à-vis de son milieu. Cette protéine reste toujours active et sous la même conformation (6), même si elle est insérée dans une membrane à fort taux de lipides, qui exerceraient une grande pression sur sa structure interne. De plus, la structure même de la protéine montre huit hélices, symétriques deux à deux (chaines C, D, J et H), présentant plusieurs contacts entre elles (Figure 4 B).

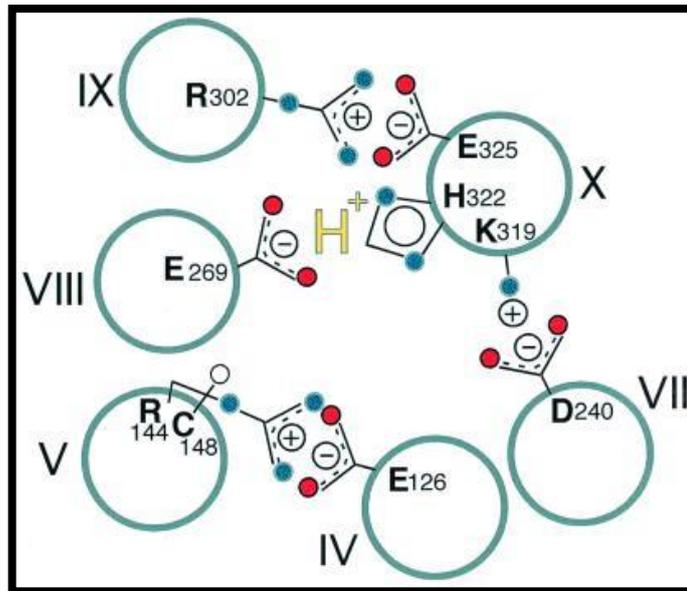
### 2.1.2.3. Les transporteurs membranaires

Les transporteurs membranaires, quant à eux, assurent un transport actif, qui demande de l'énergie pour faire basculer la protéine d'une conformation à une autre (Figure 5). Malgré le caractère hydrophobe des transporteurs, 90% des protons des amides sont échangeables avec le solvant, ce qui permet une grande flexibilité, inhérente à la fonction de la protéine (7).



**Figure 5** – A – Figure schématique d'une protéine de transport membranaire et B – sa représentation en structures secondaires hélicoïdales par chaîne. (Source : *Protein Data Bank*, code d'accès : 2ONJ).

Des études sur la Lactose Perméase (Figure 6) (8), ont permis de conclure que la composition en lipides de la membrane, influence l'activité. En effet, une membrane reconstituée d'un fort taux de lipides (800 lipides pour une molécule de perméase) diminue l'activité, exprimée en nombre de molécules de lactose transportées. Ceci serait dû à l'effet compressif qu'exerceraient les lipides sur les hélices  $\alpha$  du domaine transmembranaire. L'angle que forment ces hélices avec la membrane est augmenté, impliquant l'inactivation de la perméase. La disposition des hélices n'est pas symétrique, avec peu de points de contacts entre elles, ce qui permet une meilleure flexibilité de la protéine.



**Figure 6** – Mécanisme d'action de la Lactose Perméase. IV-X désigne l'hélice impliquée. R, E, C, D, H, K désigne respectivement arginine, glutamate, cystéine, aspartate, histidine et lysine (10).

Le mécanisme d'action des transporteurs est aussi directement lié à la structure. En effet, la Lactose Perméase, à l'état inactif, fixe un proton entre l'histidine 322 et le glutamate 269. L'activation, faisant suite à la fixation du lactose sur le site actif, implique un changement de conformation et une fixation du proton sur le glutamate 325. Le site actif est internalisé, impliquant une diminution de l'affinité de la protéine pour son substrat, suivi d'une déprotonation du glutamate 325. Le lactose est ainsi fixé à l'extérieur de la cellule et relâché à l'intérieur. L'histidine 322 et le glutamate 269 fixent alors un nouveau proton et le cycle recommence (Figure 6).

La différence de structures explique la différence de fonctions (9). Les transporteurs échangent des molécules à poids moléculaires élevés contre un proton. Cette opération est à caractère actif, c'est-à-dire qu'elle se fait contre le gradient de concentration cytoplasme-milieu extracellulaire. La flexibilité de la protéine est nécessaire pour le transport de différents substrats. Les canaux à potassium, quant à eux, assurent le transport sélectif des ions potassium, suivant le gradient de concentration, ce qui explique la rigidité de la protéine.

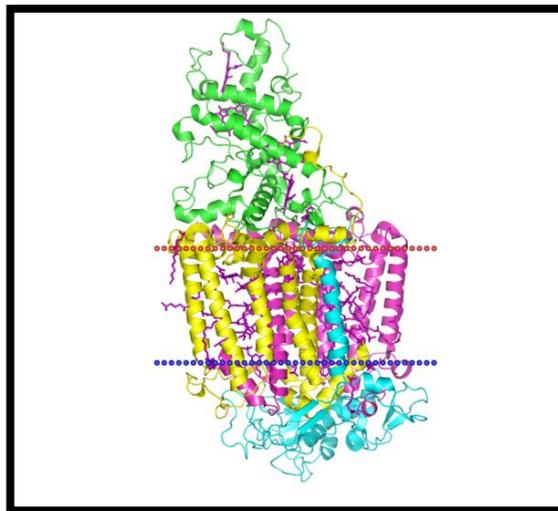
Ces exemples illustrent l'étroite relation entre la structure de la protéine membranaire et son activité biologique voire son potentiel thérapeutique.

## 2.2. Résolution de la structure

De manière générale, les protéines membranaires intégrales sont constituées d'un domaine extracellulaire, d'un ou plusieurs domaines transmembranaires et d'un domaine cytoplasmique.

La première structure protéique à haute résolution est celle de la myoglobine (11), déterminée en 1958, suivie peu après par celle de l'hémoglobine. Le travail qui a commencé en 1937, a valu à ses auteurs *Kendrew* et *Perutz* de partager le prix *Nobel* de chimie en 1962.

Ce n'est qu'en 1985, soit plus de 20 ans après, que la première protéine membranaire intégrale est résolue. Il s'agit du centre réactionnel photosynthétique de *Rhodospseudomonasviridis* (Source : *Protein Data Bank*, code d'accès : 1DXR) (Figure 7). Les auteurs *Johann Deisenhofer*, *Robert Huber*, and *Hartmut Michel* (12) ont reçu le prix *Nobel* de chimie en 1988.



**Figure 7**– La structure tridimensionnelle de la première protéine membranaire avec la représentation des hélices transmembranaires par chaîne (11).

Ces distinctions démontrent la difficulté de la tâche à laquelle sont confrontées les scientifiques, en particulier, quand il s'agit des protéines membranaires.

Depuis, plusieurs techniques ont émergé, ce qui a permis la résolution d'un nombre relativement important de structures protéiques. La *Protein Data Bank* (PDB) regroupe toutes ces structures sous forme de fichiers contenant les coordonnées atomiques constituant les acides aminés (13). A cette date plus de 50,000 structures protéiques ont été répertoriées.

### **2.2.1. Les techniques de résolution de la structure**

Les techniques les plus utilisées pour déterminer la structure protéique à une échelle atomique sont la cristallographie, la Résonance Magnétique Nucléaire (RMN) et la microscopie électronique. La digestion protéique ainsi que le dichroïsme circulaire permettent également d'affiner les structures obtenues et/ou d'en vérifier la validité.

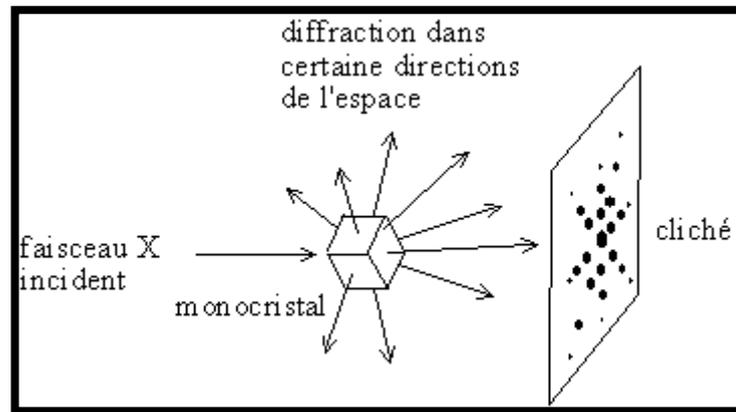
#### **2.2.1.1. Cristallographie à diffraction des rayons X**

La cristallographie à rayons X est un processus qui permet de déterminer la structure d'une protéine sous une forme cristalline (14). Il est donc nécessaire au préalable d'exprimer la protéine sous une forme cristalline pure, sous certaines conditions. Il s'agit donc d'une étape vitesse limitante surtout pour les protéines membranaires. Il faut ensuite éclairer ce cristal par un faisceau de rayons X. Les photons incidents n'ayant pas atteint l'énergie d'excitation, seront diffractés par les électrons des atomes de la protéine formant des tâches sur une plaque photographique derrière le cristal, qui seront uniques à chaque cristal (Figure 8). L'analyse de l'intensité ces tâches (ou profil de diffraction du rayon X par le cristal), permet dans un second temps de déterminer la structure de la protéine.

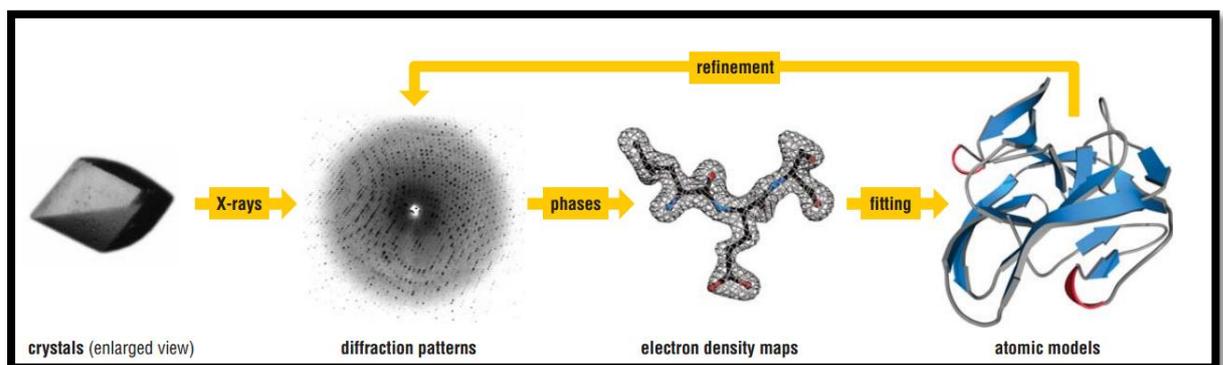
Lorsque les molécules du cristal diffractent les rayons X, un profil complexe est obtenu avec des ondes de différentes amplitudes. On obtient une première structure par addition des ondes. Puis, après un cycle itératif d'affinement où la

position des atomes dans la structure prédite est ajustée, on obtient un modèle de structure le plus concordant possible avec le profil de diffraction (Figure 9).

Cette technique classique reste aujourd'hui la plus répandue pour la résolution des structures protéiques.



**Figure 8** – Principe de la diffraction des rayons X par les cristaux. Lorsqu'il est irradié par un faisceau de rayon X, chaque atome du cristal diffuse une onde qui est projetée sur une plaque photographique. (Source : biochimej.angers.fr)

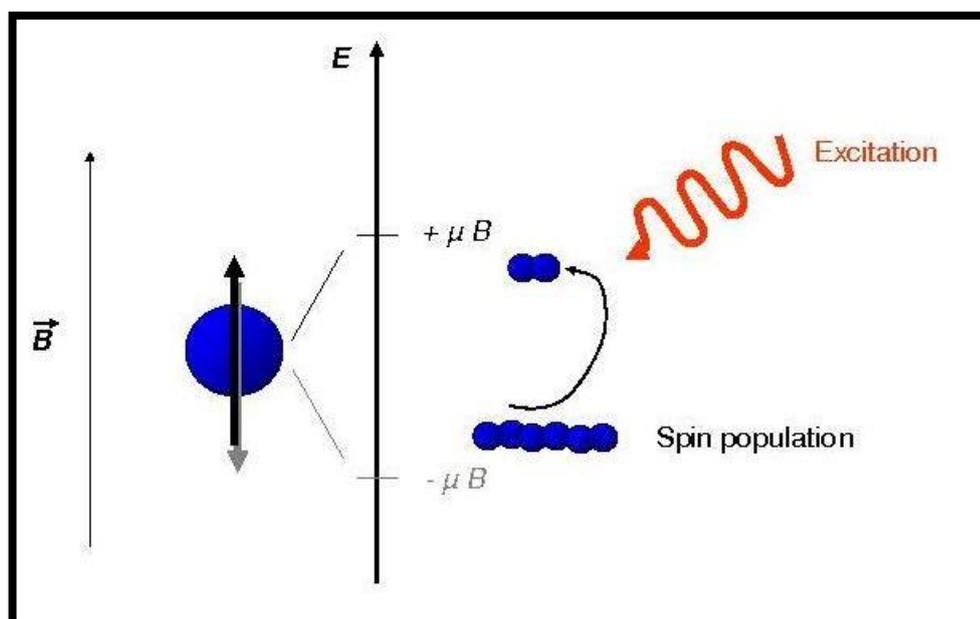


**Figure 9** – Schéma des différentes étapes de la détermination de la structure par cristallographie à rayons X. (Source : Protein structure and function, p 169)

### 2.2.1.2. Spectroscopie par Résonance Magnétique Nucléaire (RMN)

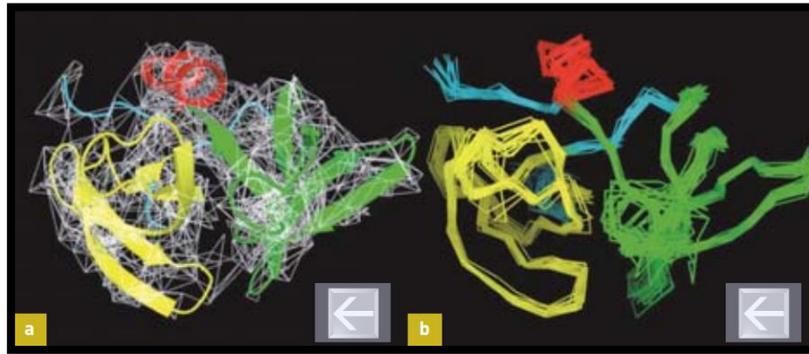
Le principe de la RMN consiste à mesurer les spins des noyaux atomiques composant la protéine, plongée en solution dans un champ magnétique intense (15). Les spins des noyaux s'orientent le long de l'axe principal du champ magnétique (Figure 10), ce qui rend possible la perturbation de l'état d'équilibre de ces spins grâce à une série d'impulsions électromagnétiques. Le courant induit lors du retour à l'équilibre est mesuré dans une bobine de réception située à proximité de l'échantillon. La combinaison de l'ensemble des contributions des différents atomes en solution est contenue dans le signal enregistré. Sa transformée de Fourier fournit les différentes fréquences de résonance des spins observés, ce qui permet la déduction de la structure atomique de la protéine.

Cette technique est facilitée par le fait qu'une grande partie des atomes composant les protéines, sont le carbone, oxygène, hydrogène et azote, atomes ayant un spin atomique non nul et donc aisément détectés par RMN.



**Figure 10** – Figure schématique du principe de la résonance magnétique nucléaire. (Source : neel.cnrs.fr)

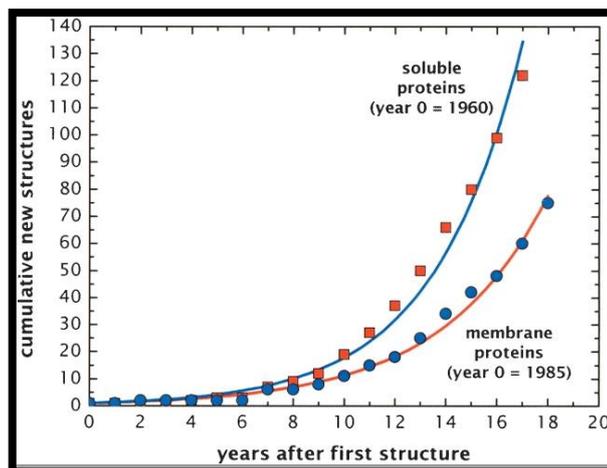
La RMN présente plusieurs avantages par rapport à la cristallographie car les mesures ne détruisent pas les échantillons et elle permet de mesurer également les fluctuations conformationnelles des protéines ce qui permet de renseigner sur la dynamique moléculaire de la protéine. La RMN peut être également associée à d'autres techniques pour affiner la structure finale (Figure 11).



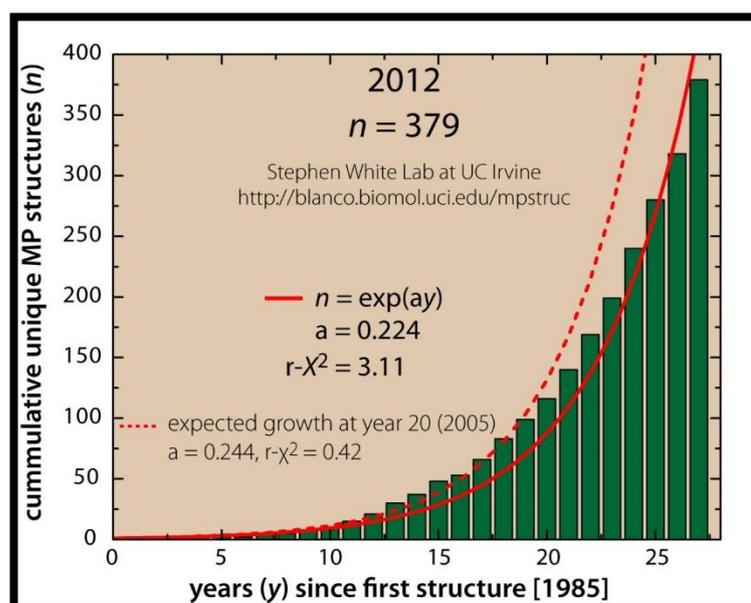
**Figure 11**– Résolution de la structure protéique par RMN associée à la modélisation in silico. Cette technique permet de poser des contraintes à la structure protéique (a) ce qui permet de déduire ensuite un ensemble de structures compatibles avec l'expérience (b) très proches les unes des autres. (Source : La RMN pour comprendre les protéines, CEA).

### 2.2.2. Les limites de ces techniques

Les protéines membranaires intégrales constituent 25% du protéome. Seulement 379 structures (moins de 1% des structures totales) ont été résolues jusqu'à ce jour (Figure 12) et sont répertoriées dans la *Membrane proteins of known 3D structure* (16). En ne tenant compte que des structures uniques, non redondantes, elles ne sont en réalité que 148 (Figure 13) structures à être disponibles aujourd'hui, dont 39 proviennent d'organismes eucaryotes (17). Ceci est dû à la complexité de l'extraction de la protéine de la membrane qui nécessite l'utilisation de détergents, à l'instabilité de ces protéines et également à l'hydrophobicité de leur surface.



**Figure 12** – Comparaison entre l'évolution des structures des protéines solubles et les protéines membranaires depuis 1985 (16).



**Figure 13** – L'évolution du nombre de structures uniques de protéines membranaires depuis 1985. (Source : <http://bianco.biomol.uci.edu/mpstruc>)

### 2.2.2.1. Expression et purification

Les protéines membranaires intégrales sont purifiées (18) et exprimées dans des bactéries, levures, cellules d'insectes ou dans des cellules de mammifères. De nombreux paramètres peuvent donc influencer cette étape. En effet, *Escherichia coli*, qui est le système d'expression le plus utilisé (19), ne peut exprimer

efficacement les protéines eucaryotes, d'où l'utilisation de systèmes eucaryotes plus complexes.

Dans un premier temps, la protéine doit s'insérer dans la membrane de l'hôte pour reconstituer sa conformation native. Ensuite, la composition en lipides de la membrane, qui varie d'un système à un autre, influence la stabilité de la protéine. Il faut également tenir compte des modifications post-traductionnelles, qui ne peuvent pas être assurées par l'hôte. Il est donc clair que le choix de l'hôte a un effet sur la structure finale retrouvée.

Les détergents utilisés pour solubiliser la protéine, sont eux aussi des facteurs déterminants dans la résolution de la structure. En effet, le choix du détergent est souvent basé sur le rendement en termes de protéines extraites, mais ceci ne garantit pas la stabilité de la protéine dans les micelles de détergents.

#### **2.2.2.2. Cristallisation**

Afin d'obtenir des cristaux de protéines qui permettent d'avoir des structures à haute résolution, plusieurs paramètres sont variés (20), tels que le PH, l'ajout de sels, de polymères, de détergents, à différentes concentrations. Une fois les conditions de cristallisation fixées, un processus d'optimisation visant à améliorer la qualité des cristaux est mis en place. Pour les protéines solubles, des systèmes entièrement automatisés, opérant à la précision du nanolitre permettent cette optimisation. Même si ces systèmes testent un nombre important de conditions initiales (21), les plupart des protéines membranaires ne parviennent pas à se cristalliser (22), d'où la nécessité de l'utilisation d'approches non conventionnelles, au cas par cas, souvent gourmandes en temps et en ressources.

#### **2.2.2.3. Résolution de la structure**

Après la diffraction des rayons X par les cristaux, les données générées sont traitées de manière automatique, pour les protéines solubles, ce qui n'est pas le

cas des protéines membranaires. La présence d'un grand nombre de micelles de détergent au contact des parties hydrophobes de la protéine, rend les cristaux fragiles à l'irradiation lors de la diffraction par les rayons X, difficiles à manipuler, et génèrent des structures à basse résolution (23). Également, l'anisotropisme des cristaux de la même protéine est plus fréquent avec les protéines membranaires.

### **2.3. Intérêt de la bioinformatique structurale**

Des améliorations notables des techniques de résolution des structures des protéines membranaires ont eu lieu. Cependant, comme le souligne le point 2.2.2, la nature complexe de ces protéines, est un facteur qui limite la résolution de ces structures. Ceci est illustré par le faible nombre de structures uniques résolues jusqu'à aujourd'hui. La structure étant directement reliée aux fonctions biologiques et au potentiel thérapeutique, de nouvelles approches bioinformatiques, dites *in silico* (24), ont permis de surmonter plusieurs difficultés et de remettre ce sujet de recherche à l'ordre du jour (25).

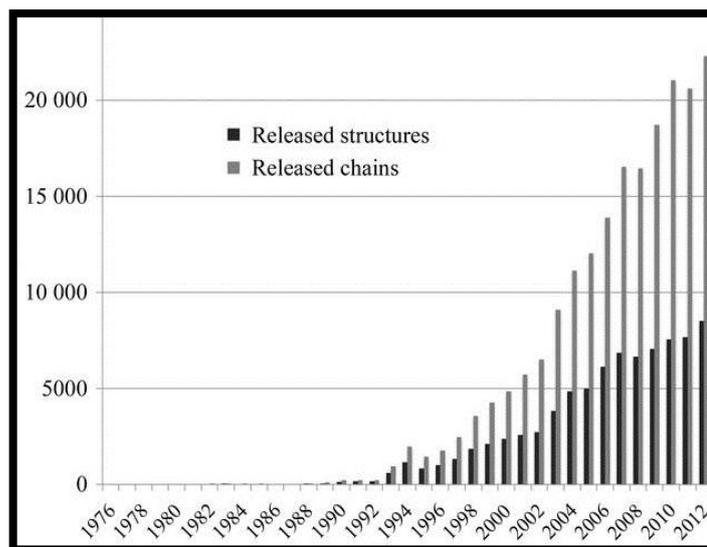
#### **2.3.1. Prédiction de la structure à partir de la séquence protéique**

Les deux dernières décennies ont, sans doute, été marquées par l'avancée des techniques de séquençage du génome. Le séquençage du premier génome humain a été achevé en avril 2003. Depuis les premières initiatives dans les années 1990, le projet *Human Genome Project* (26) a nécessité 20 ans de travail et une collaboration de six pays, une dizaine de laboratoires et une centaine de chercheurs. Le coût a été estimé à 1 milliard de dollars. En 2014, le séquençage du génome, est une opération quasi-routinière. En effet, les technologies de 3<sup>ème</sup> génération (Figure 14), permettent de réaliser un séquençage en un jour, avec un coût de mille dollars par génome (27).



**Figure 14** – Image d'un séquenceur de troisième génération. (Source : [www.pacificbiosciences.com](http://www.pacificbiosciences.com))

Cette avancée a profité à la détermination des structures protéiques. En effet, la séquence protéique de la plupart des protéines membranaires intégrales, eucaryotes, sont disponibles dans différents bases de données. Il est néanmoins difficile de déduire la structure d'une protéine à partir de sa séquence (Figure 15) (28), encore moins sa fonction. Les structures protéiques étant plus conservées que leurs séquences, donc des séquences différentes peuvent donner des structures similaires.



**Figure 15**– Evolution du nombre de séquences de protéines par rapport aux nombre de structures résolues (28).

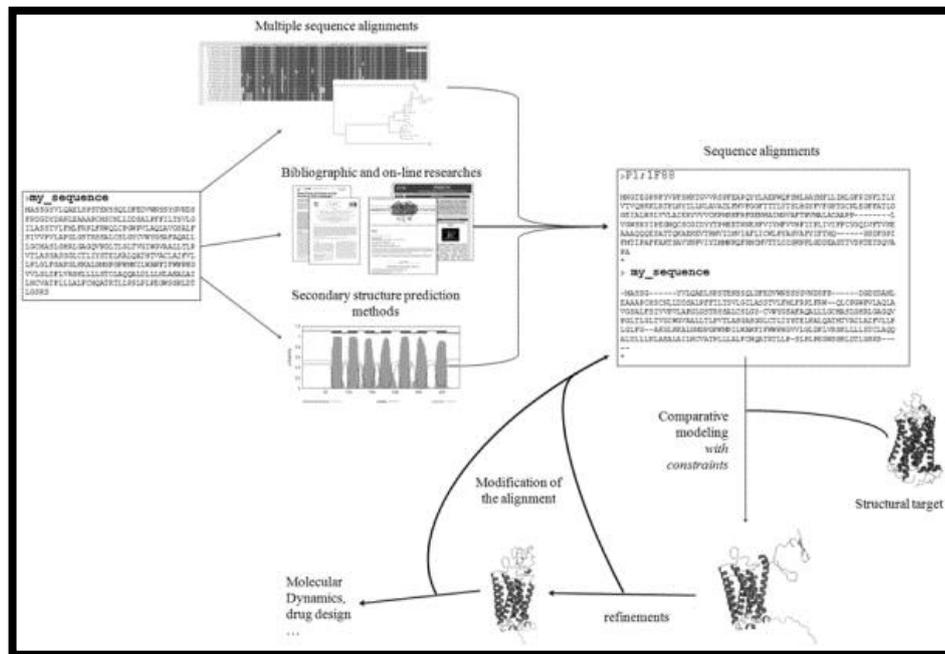
Néanmoins, la bioinformatique, en essayant de prédire la structure et donc la fonction de la protéine, a apporté plusieurs solutions et outils qui s'attaquent à cette problématique d'un angle différent. En effet, les méthodes basées sur l'apprentissage (*machine learning*), et le développement des algorithmes de prédiction, ont permis le développement d'outils fiables d'assignation des structures secondaires, des structures tridimensionnelles, de l'accessibilité au solvant entre autres propriétés des protéines.

### **2.3.1.1. Différentes méthodes bioinformatiques de prédiction de la structure**

#### **2.3.1.1.1. Modélisation comparative**

Lorsque la structure d'une protéine n'a pas été déterminée, il existe des méthodes bioinformatiques qui peuvent nous aider à déterminer la conformation 3D de la protéine d'intérêt en se basant sur la conformation tridimensionnelle d'une protéine similaire.

Quand la structure d'une protéine similaire à celle qu'on cherche à modéliser, existe dans la *Protein Data Bank*, il est possible d'utiliser cette structure comme modèle primaire qui servira à guider la modélisation de la protéine d'intérêt (29). Il suffit d'effectuer un alignement multiple de la séquence protéique avec les séquences des protéines dont la structure est connue (Figure 16) (23). Ensuite, la sélection de la structure modèle est basée sur un score de similarité. L'intégration de données expérimentales provenant de la littérature (30), l'assignation des structures secondaires, la recherche des interactions entre hélices avec l'utilisation des outils de prédiction de contact avec les lipides de la membrane, permettent d'affiner le modèle de structure finale. Le modèle final doit également refléter les propriétés biochimiques de la molécule (31).



**Figure 16** – Les différentes étapes de la détermination de la structure tridimensionnelle d'une protéine à partir de la séquence d'acides aminés en utilisant la modélisation comparative (23).

Lors de la première étape de recherche de séquences similaires, des outils d'alignement peuvent être utilisés, tel que *PSI-BLAST* (32). Une approche phylogénétique est souvent recommandée, dans le but de détecter les régions conservées dans la séquence protéique et qui sont le plus souvent, responsables du maintien de la structure et de la fonction. Il est important aussi de conduire ses propres expériences biologiques, comme la mutagenèse dirigée, afin de déterminer les résidus les plus impliqués dans le maintien de la structure. Une recherche bibliographique, suffit parfois à trouver toutes les informations nécessaires.

Dans un deuxième temps, il est nécessaire d'assigner les structures secondaires et d'extraire des informations qui serviront de base de départ pour la modélisation comparative (33). Il s'agit de retrouver le nombre d'hélices ou de feuillets, leur longueur et leur position. Des outils basés sur l'apprentissage et le recensement d'un grand nombre de structures secondaires permettent d'effectuer cette étape.

La troisième étape consiste en la réalisation de la modélisation proprement dite. *Modeller* (34) est un logiciel qui permet de générer plusieurs structures possibles, sur la base des données fournies. La modélisation peut être améliorée grandement, en imposant des contraintes au modèle final, telles que la distance entre les résidus, l'existence de ponts disulfure et les contacts entre hélices. La recherche de telles contraintes fait l'objet du présent travail et dont le but est d'améliorer les modèles des protéines membranaires intégrales polytopiques.

Finalement, un ou plusieurs modèles sont sélectionnés et une étape de validation des structures est réalisée. Différents métriques permettent d'évaluer le modèle (voir 2.3.1.2).

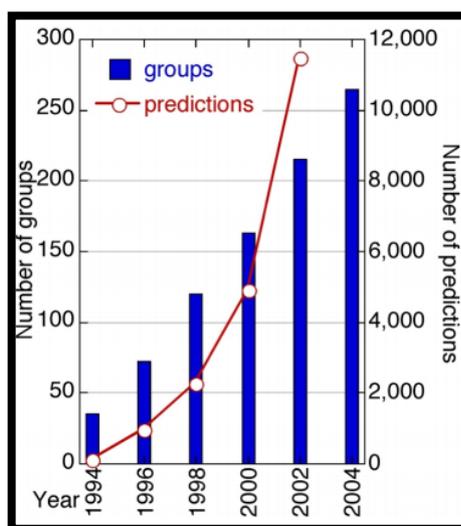
#### **2.3.1.1.2. Modélisation par *Threading***

Cette méthode concerne les protéines pour lesquelles aucune structure homologue résolue n'est disponible pour servir de modèle primaire. Elle se base sur l'identification de caractères structuraux partagés par le plus grand nombre de protéines. Dans un premier temps, il y a une étape d'apprentissage sur les structures résolues dans la *Protein Data Bank*. Ensuite, chaque acide aminé de la séquence dont on cherche à déterminer la structure, est aligné dans une structure probable (35). Finalement, il s'agit d'évaluer la conformité des caractéristiques de chaque acide aminé aux caractéristiques précédemment déterminés. Le modèle qui se conforme le plus aux caractéristiques générales partagés par les structures disponibles est retenu. Cette méthode se base sur deux constats : le premier est que, d'après la base de données *CATH* (cathdb.info), le nombre de conformations possibles dans la nature est relativement faible (autour de 1300), le deuxième étant le fait que 99% des nouvelles entrées dans la *Protein Data Bank* sont similaires aux structures précédemment résolues.

### 2.3.1.1.3. Prédiction de novo

Prédire la structure à partir de la séquence d'acides aminés a été classé par *Science* comme l'un des 100 défis scientifiques que la science doit relever. Après la synthèse de la protéine, celle-ci retrouve sa structure native, parmi une infinité de structures possibles, en 10 microsecondes. Ce processus prendrait 30 ans en temps machine, si les scientifiques tentait de le reproduire. C'est le paradoxe de *Leventhal* (36). La prédiction *de novo* essaie de déterminer la structure d'une protéine à partir de la séquence en acides aminés (37), sans aucune information préalable apportée par une structure similaire.

Actuellement, les algorithmes prédictifs s'intéressent aux petites protéines, vu les capacités énormes de calculs requis. Depuis 1994, la compétition *CASP* (*Critical Assessment for Protein Structure Prediction*) (38), réunit tous les deux ans les différents groupes de recherche pour comparer l'efficacité de leurs méthodes de détermination des structures (Figure 17). Des séquences protéiques dont les structures viennent d'être résolues et gardées secrètes, sont remises aux compétiteurs, et les différents algorithmes sont jugés sur leur efficacité à prédire au mieux la structure finale.



**Figure 17** – Evolution de la performance des équipes participantes à la compétition *CASP*. (Source : <http://predictioncenter.org>)

Les différentes méthodes détaillées dans le point précédent, essayent de prédire la structure tridimensionnelle d'une protéine à partir de sa séquence protéique. Mêmes si elles sont toutes différentes de par l'approche adoptée pour résoudre la structure, néanmoins, elles ont tous un point commun : elles se basent sur des observations biologiques et statistiques des protéines étudiées et infèrent ensuite ces connaissances sur des protéines moins connues.

Les méthodes d'apprentissage ou *machine learning*, prennent en entrée un ensemble de données relatives à des structures protéiques, et en déduisent ensuite une série de paramètres, de corrélations, qui sont partagés par ces structures et leur sont uniques (39) : c'est la phase d'apprentissage. Dans un deuxième temps, l'algorithme généralise ces corrélations à d'autres séquences dont la structure est inconnue. D'autres méthodes prennent en entrée des contraintes physico-chimiques pour modéliser la protéine. Par exemple, sachant que deux hélices  $\alpha$ , voisines de séquence, sont automatiquement en contact dans l'espace tridimensionnel, cette contrainte est appliquée dans les méthodes d'assignation des structures secondaires, ainsi que dans la modélisation moléculaire. Ceci dans le but de se rapprocher le plus possible de la réalité, de comprendre les mécanismes de formation des structures natives et enfin d'améliorer les prédictions, la modélisation et les études de dynamique des protéines membranaires.

### **2.3.1.2. La validation des modèles**

Le calcul de l'écart quadratique moyen, permet de déterminer la distance moyenne interatomique entre le modèle primaire et le modèle généré *in silico*. Le modèle ayant un écart moindre pourra être éventuellement sélectionné pour d'autres études de validation. Ces différentes méthodes sont regroupées sous le nom de *Model Quality Assessment Prediction* (MQAP) (40). Il s'agit dans un premier temps d'évaluer un modèle sur le plan global, c'est-à-dire, sa conformité par rapport aux modèles primaires et aux autres modèles générés.

Puis sur le plan local, dans le but de vérifier la qualité de l'alignement ou le positionnement des résidus.

Dans un deuxième temps, les propriétés physiques sont évaluées. Dans ce cas, des scores sont attribués aux modèles suivant leur stéréochimie, chiralité, liaisons interatomiques et propriétés électrostatiques. L'outil *PROCHECK* (41) intègre un ensemble de fonctionnalités qui permettent la mesure de ces paramètres et constitue un moyen efficace pour la validation des propriétés physiques d'un modèle. Le diagramme de *Ramachandran* (42), en l'occurrence, permet de vérifier les angles de torsions interatomiques et de les classer en zones « favorables », « acceptables » et « non acceptables ». Le calcul de l'énergie de solvation et des liaisons hydrogène sont également importantes pour la validation du modèle (43).

D'autres scores se basent sur une comparaison du modèle aux bases de données de structures de protéines existantes. Ces scores dits « statistiques » permettent de caractériser les propriétés atomiques dans une base de données de conformations résolues expérimentalement. Les paramètres les plus étudiés sont la distance interatomique, la distribution des groupes chargés, les volumes atomiques et les liaisons hydrogènes. Le service web *VERIFY 3D* (44) est l'un des premiers moyens de calcul de ces scores mis à la disposition des structuralistes. Dans un premier temps, il y a apprentissage des caractéristiques 1D-3D de chaque protéine dans une base de données déterminée expérimentalement. Les caractéristiques 1D sont celles qui s'intéressent à l'homologie entre séquences d'acides aminés (score de compatibilité séquence-séquence) et le positionnement des résidus. Les caractéristiques 3D s'intéressent quant à elles à l'environnement préféré par certains résidus, c'est le score de compatibilité séquence-structure. Les descripteurs 3D basés sur l'accessibilité au solvant, l'hydrophobicité et la polarité des atomes environnants permettent d'assigner une classe environnementale à chaque résidu. Ensuite, un score est calculé sur la base de la conformité des résidus de la séquence cible aux

propriétés des résidus déterminés dans la base de données (45). D'autres scores évaluent le potentiel de voisinage entre résidus ou potentiel de paires, et l'adéquation des structures secondaires. Les méthodes récentes combinent les scores physiques avec les scores statistiques, ce qui permet une meilleure évaluation des modèles (46).

Les différentes étapes détaillées dans les points précédents (2.3.1.1.1 /2.3.1.1.2 /2.3.1.1.3), sont aussi bien valables pour les protéines solubles que pour les protéines membranaires. Toutefois, les différents outils utilisés ont été conçus principalement pour les protéines solubles et le plus souvent ils ne sont pas valables pour les protéines membranaires, ce qui rend le processus de modélisation de ces protéines, moins évident, nécessitant parfois des approches non conventionnelles. Telle que l'intervention manuelle, pour améliorer l'alignement par exemple.

### **2.3.2. Apport de la modélisation des protéines membranaires dans la découverte et la conception des molécules à intérêt biothérapeutique**

Les protéines membranaires assurent un ensemble de rôles indispensables à la vie de la cellule. Les pompes, les convertisseurs d'énergie, les récepteurs, les transporteurs, les enzymes et les canaux constituent deux tiers des cibles thérapeutiques et sont codées par un tiers du génome humain.

Connaitre la structure des protéines membranaires a permis de mieux comprendre les mécanismes moléculaires de leur activation, et leurs mécanismes d'action (47) et de synthétiser efficacement des ligands, doués d'une spécificité remarquable, pour le traitement des maladies impliquant ces récepteurs.

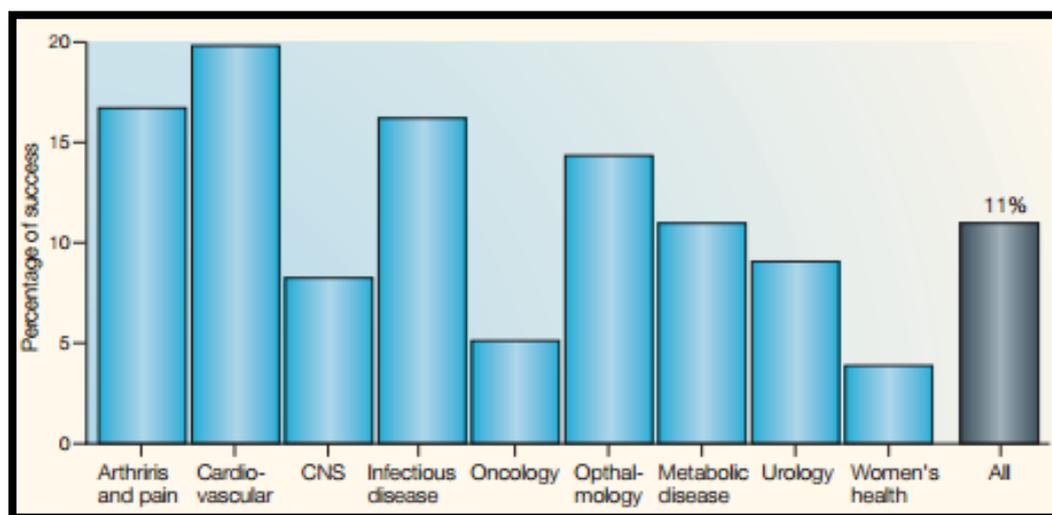
La conception du médicament, dans sa phase moderne, au début de la seconde moitié du vingtième siècle, était basée sur un large écrantage (*screening*) de différents composés chimiques de différentes natures et des essais *in vivo* sur les animaux. Cette approche basée sur la recherche de la molécule active

rallongeait le temps de la commercialisation du médicament sur le marché et augmentait les coûts. Il faut savoir que 90% des molécules candidates ne sont pas mises sur le marché (Figure 18) (48). Le début des années 1990, a connu l'émergence d'un nouveau paradigme, qui stipule que la compréhension des mécanismes moléculaires induits par les xénobiotiques doit constituer le rationnel derrière la sélection d'une molécule candidate pour le passage en phase de développement : c'est le *Rational Drug Design* (49). Le principe en est simple : si un principe actif, la molécule, devait se lier au récepteur pour produire l'effet thérapeutique escompté, comment faire en sorte que cette molécule se lie à ce récepteur de la manière la plus spécifique qui soit. En effet, les effets indésirables détectés lors des phases de développement du médicament, ainsi qu'après la commercialisation, en phase de pharmacovigilance, sont sans doute la première cause de l'arrêt du développement du médicament ou de son retrait du marché avec les pertes colossales qui s'en suivent pour l'industrie pharmaceutique et pour les patients. En admettant que le récepteur est une serrure et que le médicament est la clef, comment trouver la clef qui ouvre uniquement cette serrure. Le concept de la serrure et la clef est la base sur laquelle repose le *Rational Drug Design*. C'est donc une approche dont le point de départ est la cible impliquée dans la maladie et non le principe actif.

La modélisation moléculaire permet de mieux comprendre la cible et les sites d'interaction avec le ligand. Dans un premier temps, les bioinformaticiens détectent dans la structure, les poches, support de l'interaction avec le ligand, ce qui permet de déterminer une structure du ligand qui se lie spécifiquement et avec une grande affinité à cette cible. Une molécule virtuelle, le *lead*, est ainsi construite en prenant en compte les similarités avec les protéines déjà connues. Ensuite, ce chef de file est optimisé : Des molécules dérivées sont obtenues par manipulation de la structure afin de limiter la toxicité et d'augmenter l'affinité. Les propriétés pharmacocinétiques sont également prises en compte pour assurer

une bonne biodisponibilité et une meilleure distribution. Finalement, avant de passer à la synthèse chimique, un écrantage virtuel sur la chimiothèque permet de ressortir les composés qui partagent le plus de similarité en termes de structure chimique avec le composé à synthétiser, ceci afin de guider la chimie médicinale.

Dans les années 1990, le succès relatif de la société *AGURON*, qui a développé le *Prinomastat* (50) par une approche de bioinformatique structurale a poussé les industries pharmaceutiques à adopter ce paradigme dans la recherche de médicament. Le *Prinomastat* est un agent anticancéreux puissant agissant contre les métastases incontrôlées dans les cancers majeurs, dont le développement s'est arrêté en phase III en 2000 par manque d'efficacité dans le cancer métastatique retardé.



**Figure 18** – Taux de réussite dans le développement du médicament par classe thérapeutique. Le pourcentage moyen de réussite d'un composé de la phase une à la phase quatre est évalué à 11% (48).

### 2.3.3. L'intérêt de la présente thèse

Ce travail s'intéresse à l'analyse des protéines membranaires intégrales polytopiques. La modélisation de ces protéines est difficile, pour les raisons détaillées dans le point 2.2.2 Comprendre le mécanisme natif du repliement de la protéine membranaire et les événements impliquant la structure et la chimie de la protéine est une étape nécessaire pour prédire la structure de cette dernière. A ce jour le paradigme le plus accepté est celui de *Popot et Engelman* (51), qui stipule que le repliement natif des hélices des protéines membranaires est un processus séquentiel qui a lieu en deux étapes : l'insertion des hélices dans la membrane avec une certaine topologie et ensuite l'adoption d'une conformation, résultante de l'interaction entre les hélices et entre les lipides et les hélices pour se stabiliser dans une structure finale. Plusieurs expériences de biochimie ont permis de valider ce modèle, même s'il ne reflète pas tous les détails du processus. Il s'agit donc dans un premier temps de localiser la position des hélices dans la séquence protéique et de caractériser leur topologie. Ensuite, il faut déterminer la structure tertiaire préférée, via la caractérisation de l'entassement des hélices ou *packing*.

La détermination de la position des hélices et leurs topologies est grandement facilitée par les expériences de biochimie, et par les connaissances structurales acquises des protéines similaires. Il est donc possible de retrouver ces informations à partir de la structure et à l'aide de l'expérimentation.

La caractérisation de la structure tertiaire est donc un facteur limitant dans la prédiction de la structure. Caractériser le *packing* des hélices dont la topologie est connue, a fait l'objet de plusieurs études, dont une majorité concerne les protéines solubles (52, 53). Le but de ce travail est de caractériser la structure tertiaire des hélices transmembranaires, en se basant sur le travail pionnier de *Bowie* (54). Il s'agit de décrire les hélices transmembranaires en termes de longueur, nombre de contact avec les hélices voisines, composition en acides aminés de la première et dernière boucle, calcul de l'angle formé avec la

membrane, et de l'angle formé entre deux hélices voisines de séquences et dans l'espace. L'auteur reconnaît dans son article, où il y analyse les hélices de trois protéines membranaires, qu'une étude plus large, englobant plus de protéines étudiées, est nécessaire pour généraliser les connaissances retrouvées. Ce travail vise donc à analyser un plus grand nombre de protéines, sur les critères de *Bowie* (55) et d'autres critères d'analyses statistiques plus récentes basées sur la structure (56). Ceci permettra de mieux comprendre le mécanisme de repliement des protéines membranaires et indéniablement d'améliorer le processus de prédiction de la structure à partir de la séquence.

### 3. MATERIELS ET METHODES

Le présent travail s'intéresse donc à l'analyse topologique des hélices transmembranaires par des outils bioinformatiques et statistiques.

Pour se faire, un outil en langage *Python* a été créé afin de réaliser ces analyses sur une base de données de protéines membranaires. *HeliStatistiX* est présenté ainsi comme un outil flexible, fiable et très complet. Il englobe une grande majorité des paramètres décrivant les hélices transmembranaires tels que l'angle formé entre l'axe de l'hélice et le plan de la protéine, la longueur de l'hélice, son moment hydrophobe, la présence de cassures et l'angle de ses cassures. Mais aussi, l'interaction entre hélices et l'angle formé par elles, ainsi que la boucle de connexion et le nombre de points de contacts, entre autres paramètres.

Les différentes corrélations entre paramètres et autres tendances statistiques avérées dans ce travail permettront aux bioinformaticiens de réduire l'espace de recherche lors de la prédiction de la structure tridimensionnelle de protéines transmembranaires.

L'importance de ce travail réside dans le fait qu'une grande base de données de protéines transmembranaires a été générée et étudiée dans le but de comparer et étayer nos résultats à ceux des précédents travaux de Bowie et al. (54, 56), travaux précurseurs dans ce domaine. *HeliStatistiX* n'est donc pas le premier outil qui permet de réaliser ce type d'analyses (57, 58), cependant, il examine davantage de paramètres et permet de regrouper les résultats d'analyse en un fichier unique, ce qui permet de procéder plus aisément et rapidement aux analyses statistiques.

L'objectif sera donc de confirmer les corrélations entre les paramètres précédemment caractérisés dans la littérature mais aussi d'en déterminer d'autres. Ceci dans le but d'élucider un peu plus les mécanismes qui régissent l'interaction des hélices transmembranaires avec leur environnement lipidique,

de comprendre le repliement de la protéine dans la bicouche lipidique, la formation de structures secondaires ainsi que de déterminer les différences entre protéines membranaires et protéines globulaires.

### 3.1. La structure des protéines

Les protéines membranaires contenant un ou plusieurs segments transmembranaires ont été identifiées par leurs codes PDB. Les fichiers de structures correspondants ont été récupérés dans la base de données OPM. L'avantage qu'offre OPM par rapport à la PDB est la correction des coordonnées atomiques par rapport au plan médian de la protéine. Autrement dit, le plan de la protéine est situé à l'origine du repère.

### 3.2. HeliStatistiX

Cet outil a été codé en *Python version 2.7* (59), il prend en entrée un fichier *config* en trois lignes distinctes (Figure 19) (Annexe 1) :

- Le code PDB des protéines séparés par un point-virgule,
- Les chaînes protéiques transmembranaires de chaque protéine séparées par une virgule pour la même protéine et par un point-virgule entre protéines différentes,
- Les positions de début et de fin de chaque hélice. Ces informations sont prises de la base de données OPM (60).

Les fichiers PDB téléchargés sont mis dans le même dossier que le script de l'outil et son fichier *config*.

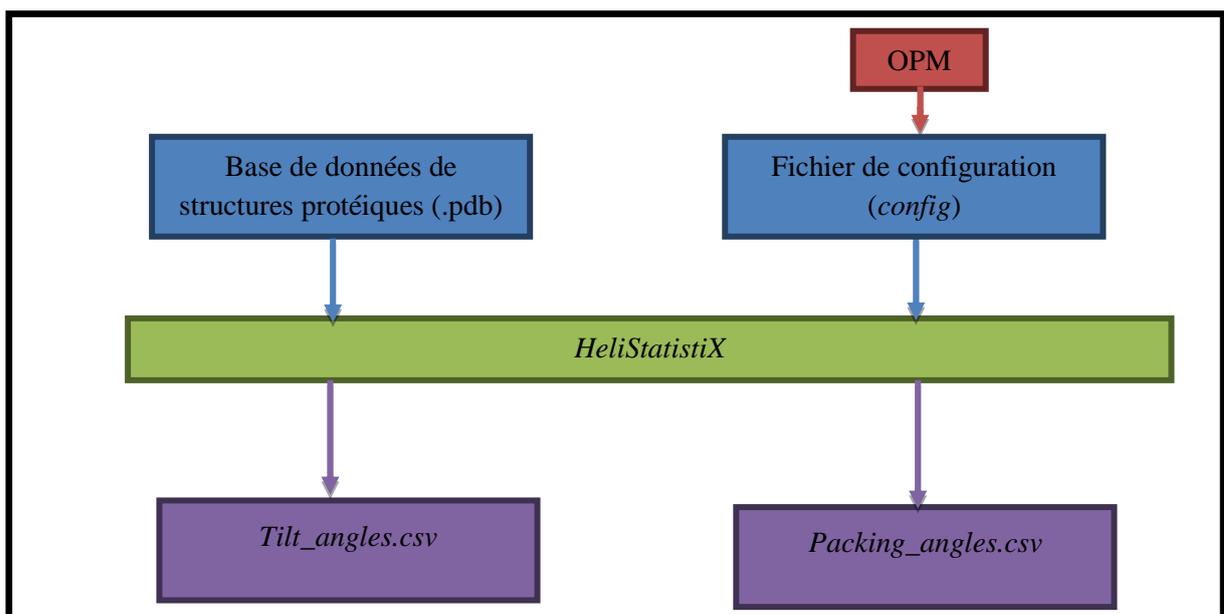
Après l'exécution du programme, deux fichiers tableau CSV sont créés en sortie :

- Le premier *Tilt\_angles* recense les angles de *tilt* des hélices ainsi que neuf autres paramètres : La longueur de l'hélice en acides aminés, la longueur en Angströms (Å), l'hydrophobicité moyenne de l'hélice, le moment hydrophobe d'*Eisenberg*, le nombre d'hélices voisines, le nombre de cassures, l'angle de la plus grande cassure par rapport à l'axe de l'hélice, la composition en acides

aminés du premier tour et du dernier tour d'hélices, chaque tour étant défini sur quatre acides aminés.

- Le second *Packing\_angles* regroupe toutes les hélices en interaction et les angles qu'elles forment ainsi que huit autres paramètres : l'orientation de l'hélice (Parallèle / Antiparallèle), nombre de contacts entre hélices en interaction, la distance minimum entre les centres d'hélices en interaction, la longueur de la boucle de connexion, le nombre de cassures dans le couple d'hélices, le moment hydrophobe d'*Eisenberg* de chaque hélice en interaction, le pourcentage des hélices voisines qui sont en interaction.

Les différentes analyses ont été effectuées à partir de ces deux fichiers grâce à un script en langage statistique *R* (61) : Le calcul du coefficient de corrélation de *Pearson*, ainsi que les figures.



**Figure 19** – Schéma représentant le fonctionnement de l'outil *HeliStatistiX*.

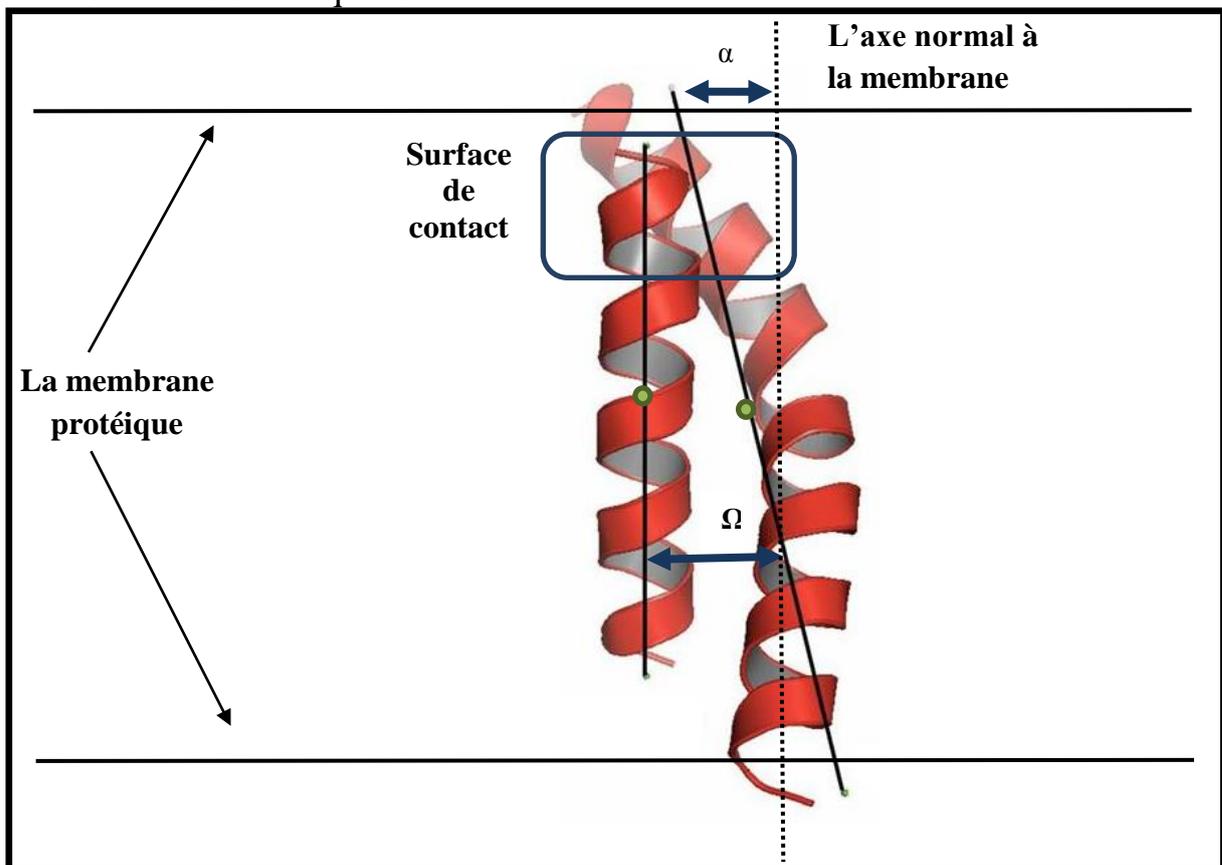
### 3.3. Définitions des paramètres

La plupart des paramètres ont été définis comme dans l'article de référence de Bowie *et al* (50), afin de pouvoir comparer les résultats et constater l'effet de l'étude d'un plus grand nombre de protéines (33 contre trois protéines dans l'article de référence). L'accent a été mis sur deux paramètres qui sont l'angle de

*tilt* et l'angle de *packing*. L'angle de *tilt* permet de comprendre l'influence qu'exerce l'environnement de l'hélice sur son orientation tandis que l'angle de *packing* évalue l'interaction de l'hélice avec son environnement.

### 3.3.1. Calcul de l'axe d'une hélice

Le calcul de l'axe de l'hélice se fait selon la méthode de *Chothia* (52). C'est la moyenne des axes locaux de chaque tour d'hélice défini sur quatre acides aminés (Figure 20). Chaque axe local est le vecteur normal au plan formé par les acides aminés de chaque tour.



**Figure 20** – Schéma représentant l'axe d'une hélice et la méthode des calculs des paramètres.  $\Omega$  est l'angle de *packing*,  $\alpha$  est l'angle de *tilt*, en pointillé l'axe normale à la membrane.

Un axe local est calculé pour chaque tour d'hélice. L'axe final est la moyenne de tous les axes locaux. Il y a cassure si l'angle entre deux axes locaux successifs est supérieur à  $20^\circ$ .

L'angle de *packing* est l'angle entre les axes de deux hélices.

L'angle de *tilt* est celui que forme l'axe de l'hélice avec le vecteur normal à la bicouche lipidique (vecteur vertical). Finalement, deux hélices sont en contact si les deux centres (points verts sur la figure 20) sont à moins de 25 Å de distance et qu'au moins trois résidus sont en contact.

### 3.3.2. Cassure d'hélice

La présence de cassure dans une hélice : est définie lorsque deux axes locaux successifs forment un angle de plus de 20°.

### 3.3.3. Moment hydrophobe d'*Eisenberg*

Le moment hydrophobe d'*Eisenberg* noté  $\langle uH \rangle$  est défini comme dans la littérature (62-64) et calculé de la façon suivante :

$$\langle uH \rangle = \frac{1}{N} \left\langle \left[ \sum_{n=1}^N H_n \sin(n\delta) \right]^2 + \left[ \sum_{n=1}^N H_n \cos(n\delta) \right]^2 \right\rangle^{1/2}$$

Avec  $H$ , l'hydrophobicité du  $n^{\text{ième}}$  acide aminé dans la séquence,  $n\delta$  l'angle séparant les chaînes latérales, avec  $\delta = 100^\circ$  pour les hélices  $\alpha$ .

Plus le moment d'une hélice  $N$  est élevé, plus son caractère amphiphile est prononcé. Sa valeur varie entre 0 et 3.26. Le moment hydrophobe est le vecteur somme de tous les moments hydrophobes de chaque acide aminé, déterminant ainsi sa longueur et sa direction (Annexe 2).

### 3.3.4. L'hydrophobicité moyenne d'une hélice

L'hydrophobicité moyenne d'une hélice varie entre -1.01 et 2.25, elle est notée  $\langle H \rangle$ . C'est la somme de l'hydrophobicité de chaque de chaque acide aminé de la séquence de l'hélice divisée par le nombre d'acides aminés. L'hydrophobicité est déterminée (Tableau 1) selon l'échelle de *Pliska et Fauchère* (65).

Acide aminé	Hydrophobicité	Acide aminé	Hydrophobicité
Ala	0.31	Leu	1.7
Asp	-0.77	Phe	1.79
Glu	-0.64	Thr	0.26
Ile	1.8	Val	1.22
Met	1.23	Asn	-0.6
Ser	-0.04	Gln	-0.22
Tyr	0.96	His	0.13
Arg	-1.01	Lys	-0.99
Cys	1.54	Pro	0.72
Gly	0	Trp	2.25

**Tableau I** – L'échelle d'hydrophobicité de *Pliska* et *Fauchère* (65).

Cette échelle a pu être déterminée grâce aux coefficients de partition Octanol/Eau de chaque résidu.

### 3.3.5. Calcul de l'angle de *tilt*

L'angle de *tilt* est l'angle calculé entre le vecteur de l'axe de la protéine et le vecteur de l'axe de l'hélice selon la formule du produit scalaire :

$$\vec{u} \cdot \vec{v} = \|\vec{u}\| \cdot \|\vec{v}\| \cdot \cos(\vec{u}, \vec{v})$$

### 3.3.6. Contact entre deux hélices en interaction

Le contact entre deux hélices est défini selon la méthode *Chothia* (52), basée sur un seuil de distance : Il y a contact si deux hélices ont plus de trois résidus en contact avec une distance entre les centres des hélices inférieure à 25 Å . Deux résidus sont en contacts si la distance entre leurs carbones  $\alpha$  respectifs est inférieure à 10 Å.

### **3.3.7. Calcul de l'angle $\Omega$ de *packing***

L'angle de *packing*, ou de torsion est l'angle entre les deux vecteurs normaux aux axes de deux hélices en interaction. L'origine de chaque vecteur est le point où la distance entre les deux hélices est la plus faible.

### **3.3.8. Calcul de l'orientation du *packing***

L'orientation de deux hélices en interaction est parallèle si les vecteurs des axes des deux hélices ont le même sens et antiparallèle s'ils ont un sens contraire.

## 4. RESULTATS ET DISCUSSION

### 4.1. La base de données

La base de données de protéines membranaires utilisée dans ce travail regroupe 33 protéines transmembranaires différentes et non redondantes ayant 48 segments transmembranaires différents. Ce qui représente un total de 231 hélices recensées dans le Tableau 2. 350 interactions entre hélices ont été caractérisées selon les critères décrits précédemment.

<b>Protéine (Code PDB)</b>	<b>Chaîne transmembranaire</b>	<b>Nombre d'hélices</b>
1dxr	L	5
	M	5
1e12	A	7
1eys	L	5
	M	5
1h2s	A	7
	B	2
1jb0	B	11
	F	1
	L	3
1kqf	C	4
1ors	C	5
1q16	C	5
1r3j	C	2
1v55	B	2
1wpg	A	10
1xio	A	7
2bhw	A	3

2bl2	A	4
2f2b	A	8
2nq2	A	10
2qts	B	2
2uuh	A	4
2vpz	C	8
2x2v	A	2
2zxe	A	10
3a7k	A	7
3cx5	C	8
	E	1
	H	1
3gd8	A	8
3gia	A	12
3hd6	A	12
3ldc	A	2
2j8c	M	5
3s8g	A	13
1yq3	C	3
	D	3
2xtv	A	6
2xqu	A	2
3m73	A	10

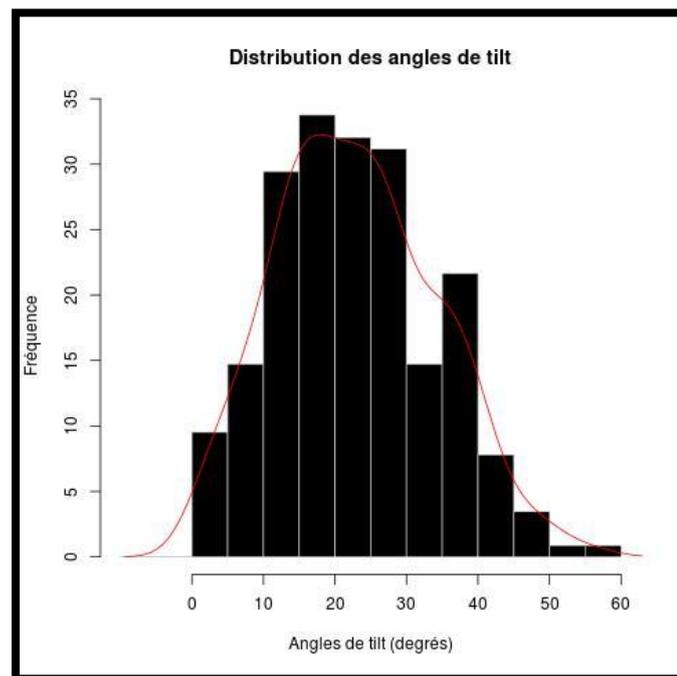
**Tableau II** – Caractéristiques des protéines étudiées.

Les positions de début et de fin de chaque hélice peuvent être retrouvées dans la base de données OPM en entrant le code PDB de chaque protéine.

## 4.2. L'angle de *tilt*

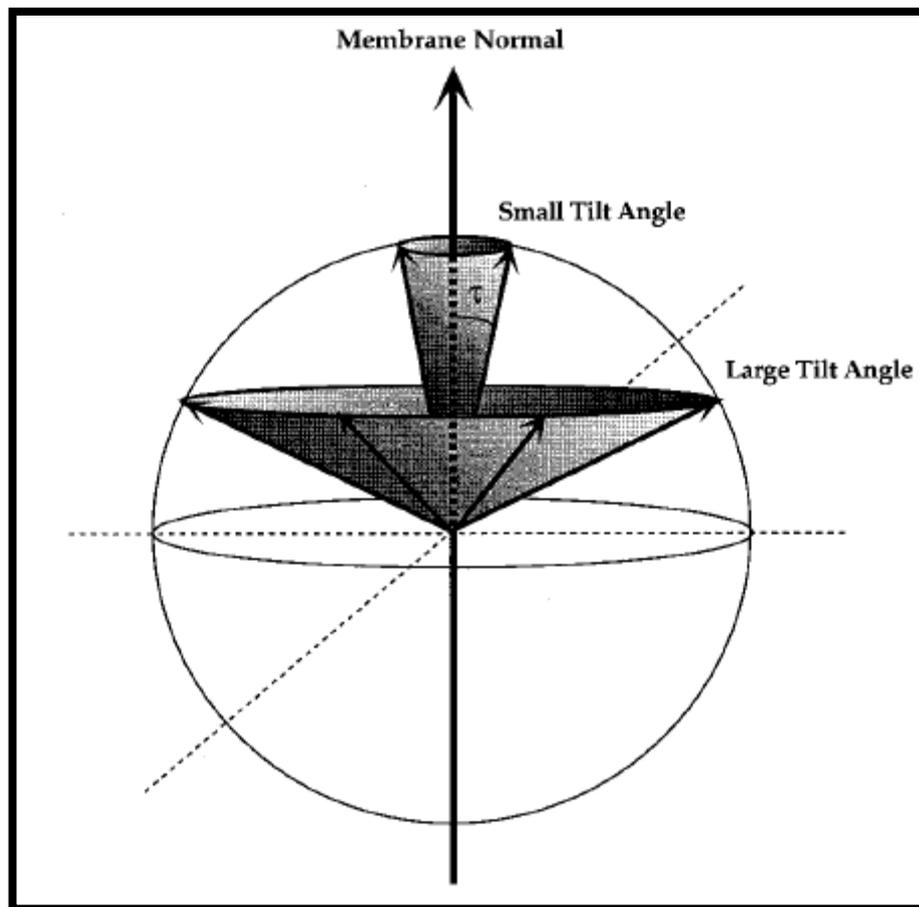
### 4.2.1. Distribution des angles de *tilt*

La figure 21 montre la distribution des angles de *tilt*. La moyenne se situe à  $30^\circ$  avec un écart type de  $9,57^\circ$ . Le graphique ne permet pas de dégager une tendance des angles de *tilt*, avec une majorité des angles se situant dans l'intervalle compris entre  $10^\circ$  et  $40^\circ$ , sans prendre en compte le biais statistique de préférence pour les grands angles.



**Figure 21** – La distribution des angles de *tilt*. Cette distribution n'est pas normale (test de Shapiro :  $W = 0.98$ ,  $p = 0.03 < 0.05$ ).

Cependant la préférence pour tous les angles n'est pas la même, car dans un repère tridimensionnel, il existe plusieurs manières pour générer un grand angle par rapport à un petit angle (54). Autrement dit, en l'absence de contraintes, en prenant l'exemple d'une sphère, un grand angle a plus de probabilité d'être sélectionné au hasard (Figure 22).



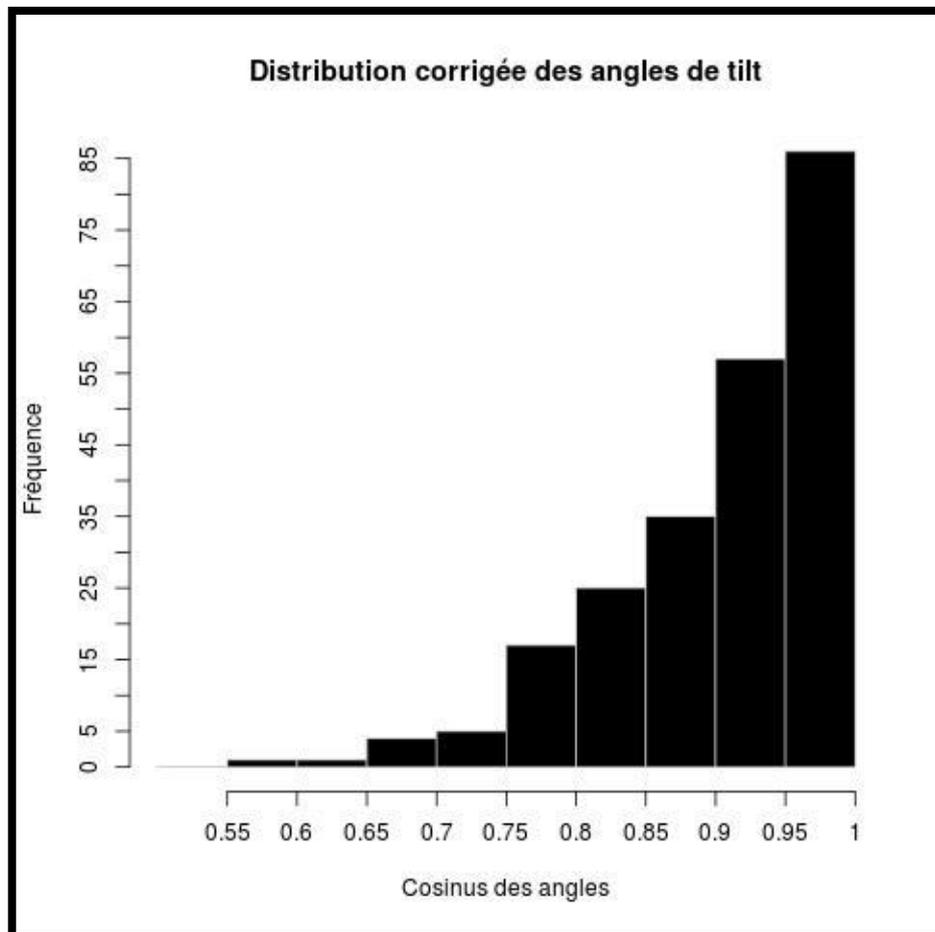
**Figure 22** – Probabilités des angles de tilt (54).

Dans un repère tridimensionnel, plus un angle est grand, plus il a de possibilités d'être sélectionné, comme démontré par la figure 22. Ceci est à l'origine d'un biais statistique qui favorise les grands angles dans la distribution des angles de *tilt*. La probabilité de chaque angle est proportionnelle à la circonférence de l'anneau formé par les différentes possibilités de sélectionner cet angle, elle-même proportionnelle au cosinus de l'angle.

Pour pallier à ce biais statistique, la distribution selon le cosinus de l'angle a été proposée (Figure 23). Cette distribution corrigée montre une préférence pour les petits angles ( $0^\circ - 18^\circ$ ) de *tilt*, correspondant aux valeurs de cosinus comprises entre 0.95 et 1.

Ceci confirme le modèle selon lequel, sous les contraintes physiques dans la membrane, les hélices des protéines membranaires s'y insèrent avec un angle quasi-nul (51), ceci a été confirmé par des modèles *in silico* (66). Les hélices

étant ancrées de part et d'autres de la membrane, la longueur idéale est celle de la membrane (30 Å) (67) et celles ayant des longueurs différentes (supérieures ou inférieures à 30 Å ) ont donc des angles de *tilt* qui dévient de la normale à la membrane.

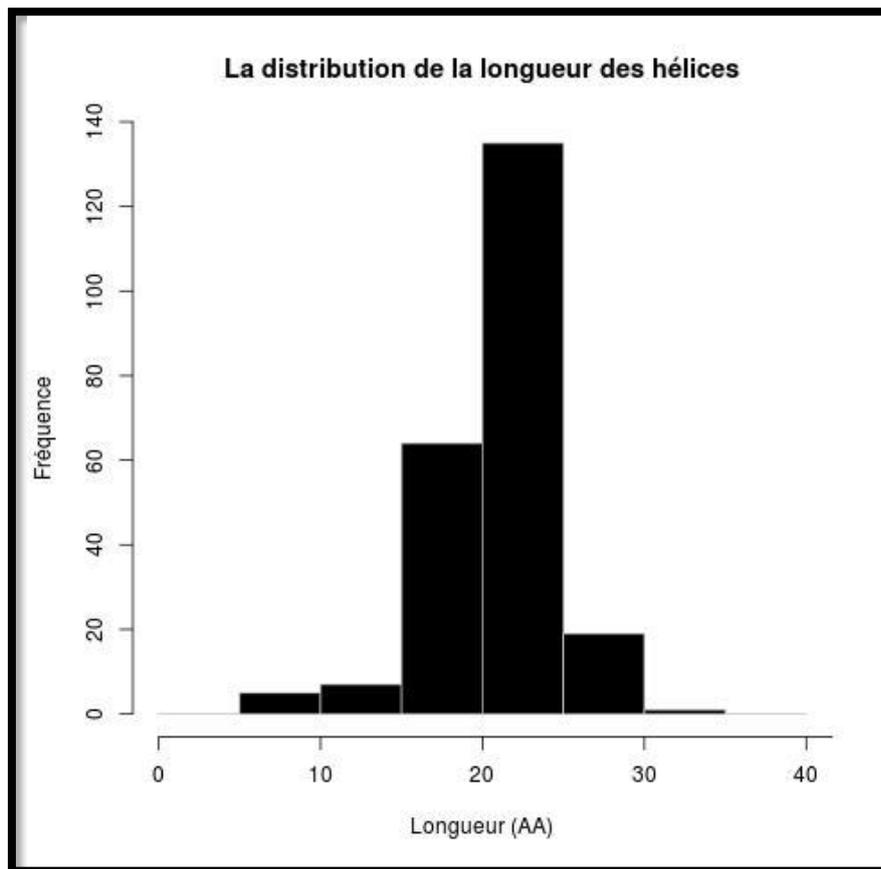


**Figure 23** – Distribution corrigée de l'angle de tilt.

#### 4.2.2. Angle de *tilt* et la longueur d'hélice

##### 4.2.2.1. La longueur d'hélice

Une hélice est définie au minimum sur un tour par quatre acides aminés. La longueur moyenne des hélices est de 22,1 acides aminés (écart-type 2,3 acides aminés) soit 32.4 Å (écart-type 2,6 Å) (Figure 24). Ces résultats sont en accord avec les résultats retrouvés dans l'article de référence (54) selon lequel la longueur moyenne des hélices serait de 26.4 acides aminés.



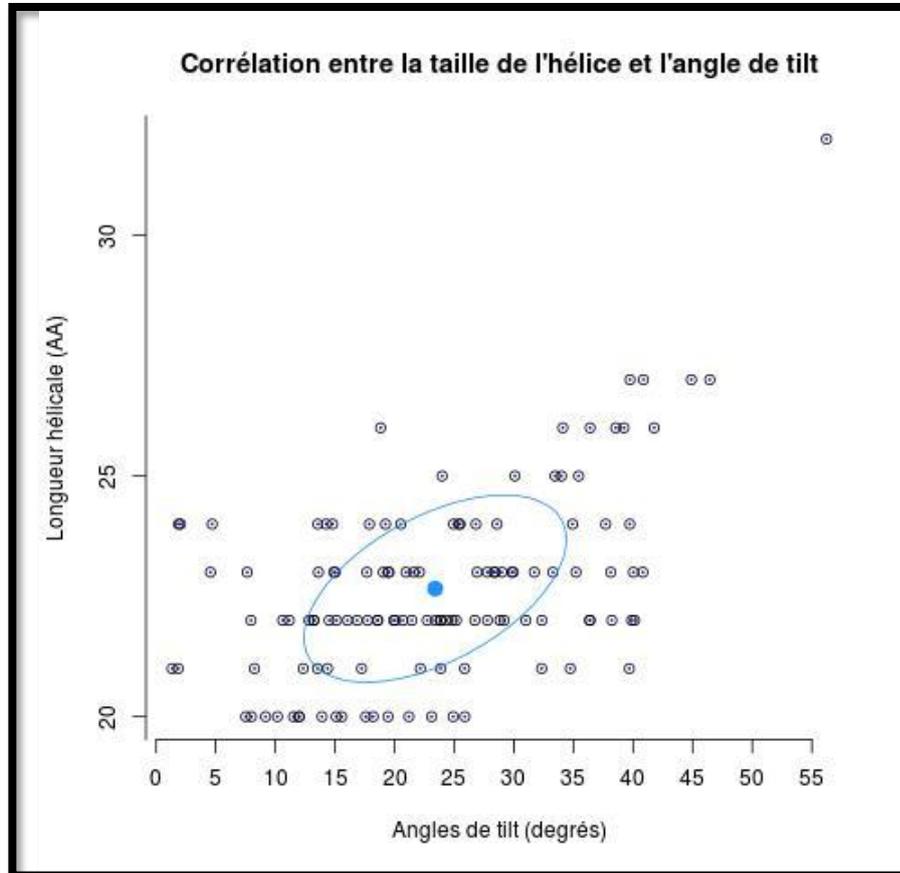
**Figure 24** – La distribution de la longueur des hélices exprimée en nombre d'acides aminés.

La majorité des hélices membranaires ont une longueur située entre 20 et 25 acides aminés avec une moyenne de 21 acides aminés soit  $32 \text{ \AA}$ , ce qui correspond à la longueur de la bicouche lipidique dans laquelle sont insérées les hélices ( $\sim 30 \text{ \AA}$ ) (55). Cette contrainte impose une certaine préférence pour la longueur des hélices, les hélices des protéines solubles, dépourvus de telles contraintes ont une distribution différente (55).

#### 4.2.2.2. Corrélation entre angle de *tilt* et la longueur d'hélice

Le précédent résultat démontre aussi que la plupart des hélices devrait s'insérer de manière parallèle à l'axe normal au plan de la protéine. L'hélice étant ancrée de part et d'autre de la membrane, un angle de *tilt* de  $0^\circ$  devrait être énergétiquement favorable au vue de telles contraintes.

Pour l'étude de cette corrélation, seules les hélices de plus de 19 acides aminés, susceptibles d'être ancrées à la membrane des deux côtés sont retenues. Egalement, les hélices présentant une cassure ont été retirées, car les cassures ou *kinks* peuvent influencer l'angle de *tilt* (voir 4.2.5).



**Figure 25** – L'angle de *tilt* est en corrélation statistiquement significative avec la longueur de l'hélice.

Le coefficient de corrélation de *Pearson* calculé s'élève à 0.53 ( $p < 0.001$ ) (test du coefficient de corrélation est significatif) entre la longueur d'hélice et l'angle de *tilt* (Figure 25) ce qui démontre l'hypothèse précédemment émise selon laquelle plus l'hélice est longue, plus elle est déviée de la normale à la membrane.

Toutefois, la corrélation reste faible, mais elle devient plus grande si les angles sont catégorisés en intervalles de  $5^\circ$ . D'autres raisons peuvent expliquer la faible corrélation, notamment, la possible présence d'hélices en position parallèle à la membrane, qui sont toutefois rares. Finalement, l'extension de l'hélice de plus

d'un tour au-delà du domaine transmembranaire est possible et pourrait fausser la corrélation. Ceci pourra être corrigé si l'on ne s'intéresse qu'aux parties transmembranaires des hélices.

### 4.2.3. Angle de *tilt* et hydrophobicité

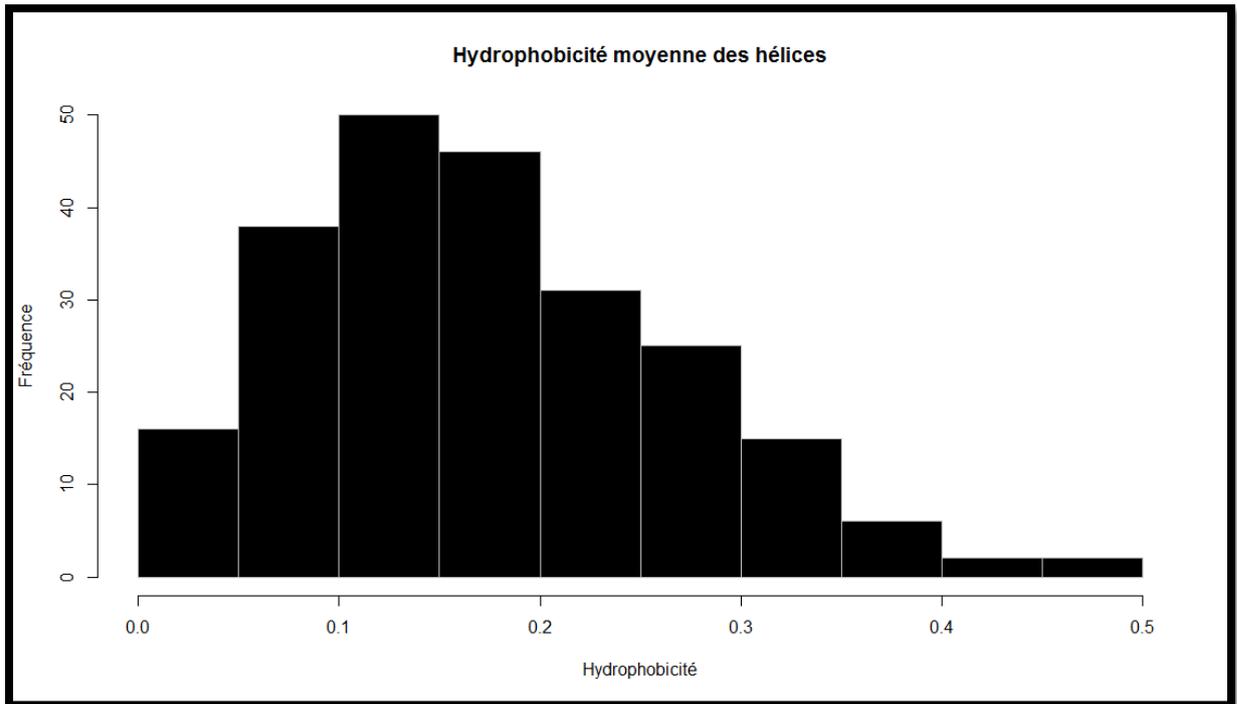
#### 4.2.3.1. L'hydrophobicité des hélices

La Figure 26 et l'annexe 3 montrent la distribution de l'hydrophobicité moyenne des hélices qui s'étend sur une échelle de 0 à 0.5. Ce profil confirme le caractère hydrophobe prononcé des hélices transmembranaires. La moyenne est égale à 0.17 avec un écart-type de 0.091.

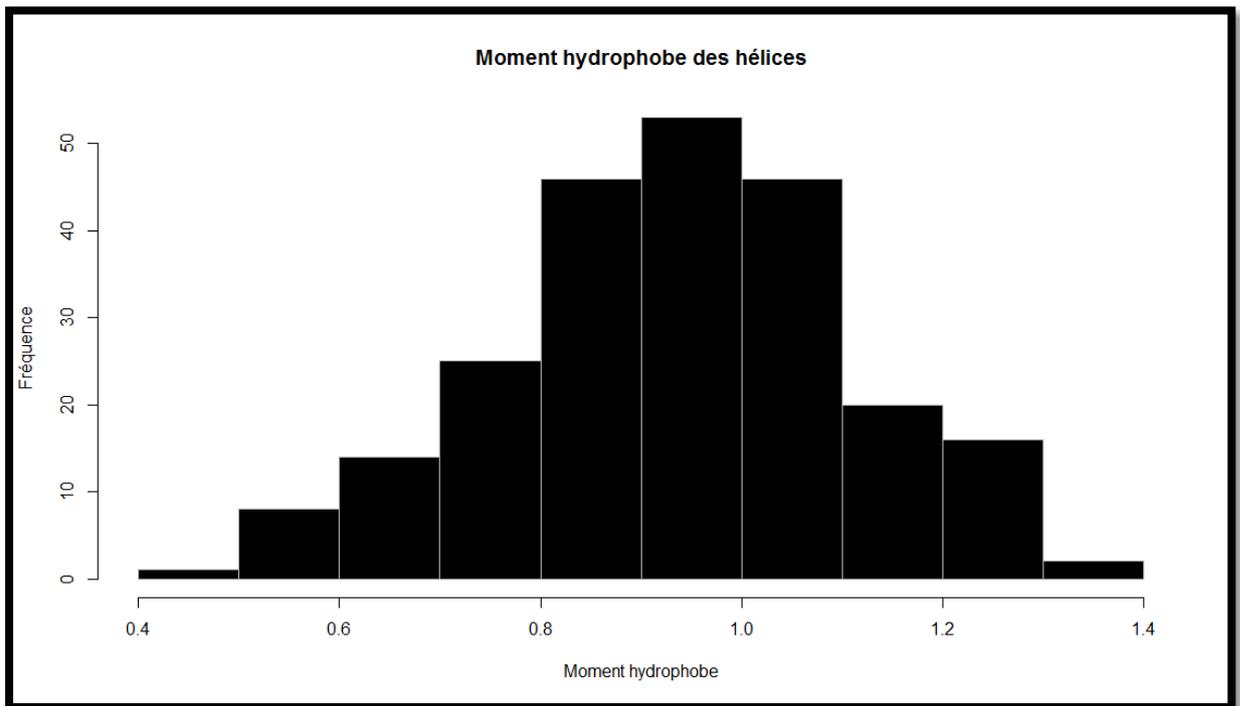
La figure 27 représentant la distribution du moment hydrophobe des hélices membranaires montre que ce paramètre varie de 0.4 à 1.4 selon une distribution normale (test de Shapiro :  $W = 0.99$ ,  $p = 0.511 > 0.05$ ) (Annexe 4). La moyenne est 0.93 et l'écart-type égale à 0.17.

#### 4.2.3.2. Corrélation entre angle de *tilt* et hydrophobicité

L'hélice enfouie dans le cœur hydrophobe de la membrane, tend à garder cette position énergétiquement favorable (68) et les contacts avec l'environnement hydrophile sont minimisés, c'est le *hydrophobic mismatch* (69). Le *tilting* de l'hélice d'un certain angle est l'un des mécanismes qui permet de garder cette position favorable (70). En calculant l'hydrophobicité moyenne et le moment hydrophobe d'*Eisenberg*, aucune corrélation n'a pu être déterminée entre l'hydrophobicité (Annexe 5) et la tendance amphiphile de l'hélice (Annexe 6) avec l'angle de *tilt*. Ceci est dû au fait que toutes les hélices sont hydrophobes et qu'elles diffèrent peu au niveau des échelles d'hydrophobicité. En fait, les différences se situent essentiellement au niveau des interfaces membrane-cytoplasme et membrane-milieu extracellulaire. Nous nous sommes donc intéressés à la composition des hélices à ces niveaux.



**Figure 26** – Distribution de l'hydrophobicité moyenne des hélices transmembranaires.



**Figure 27** – Distribution du moment hydrophobe des hélices transmembranaires.

#### 4.2.4. Angle de *tilt* et composition en acides aminés

##### 4.2.4.1. Composition du premier et dernier tour d'hélice

Dans la matrice de composition du premier et dernier tour d'hélice en fonction de l'angle de *tilt* (Figure 28), les acides aminés hydrophobes (en bas) sont séparés des acides aminés hydrophiles (en haut) par la glycine. Les résidus hydrophobes sont majoritairement présents dans le premier et dernier tour d'hélice par rapport aux protéines solubles (56) (Annexe 7).

La leucine est également l'acide aminé le plus abondant dans les extrémités, comme à l'intérieur de l'hélice (56). En comparant la fréquence des acides aminés hydrophiles présents aux extrémités par rapport à l'intérieur de l'hélice (71), ceux-ci sont plus abondants du fait que leur présence à l'intérieur de la membrane est énergétiquement défavorable. En particulier L'histidine, l'asparagine et la glutamine sont essentiels pour stabiliser l'hélice dans la membrane, en accord avec les modèles précédemment publiés, car leur présence résulte en une diminution de l'énergie libre (72).

##### 4.2.4.2. Corrélation entre angle de *tilt* et composition en acides aminés

La distribution des acides aminés au sein du premier et dernier tour d'hélice en fonction de l'angle de *tilt* ne démontre pas une influence claire d'un acide aminé en particulier sur l'angle de *tilt* (Figure 28).

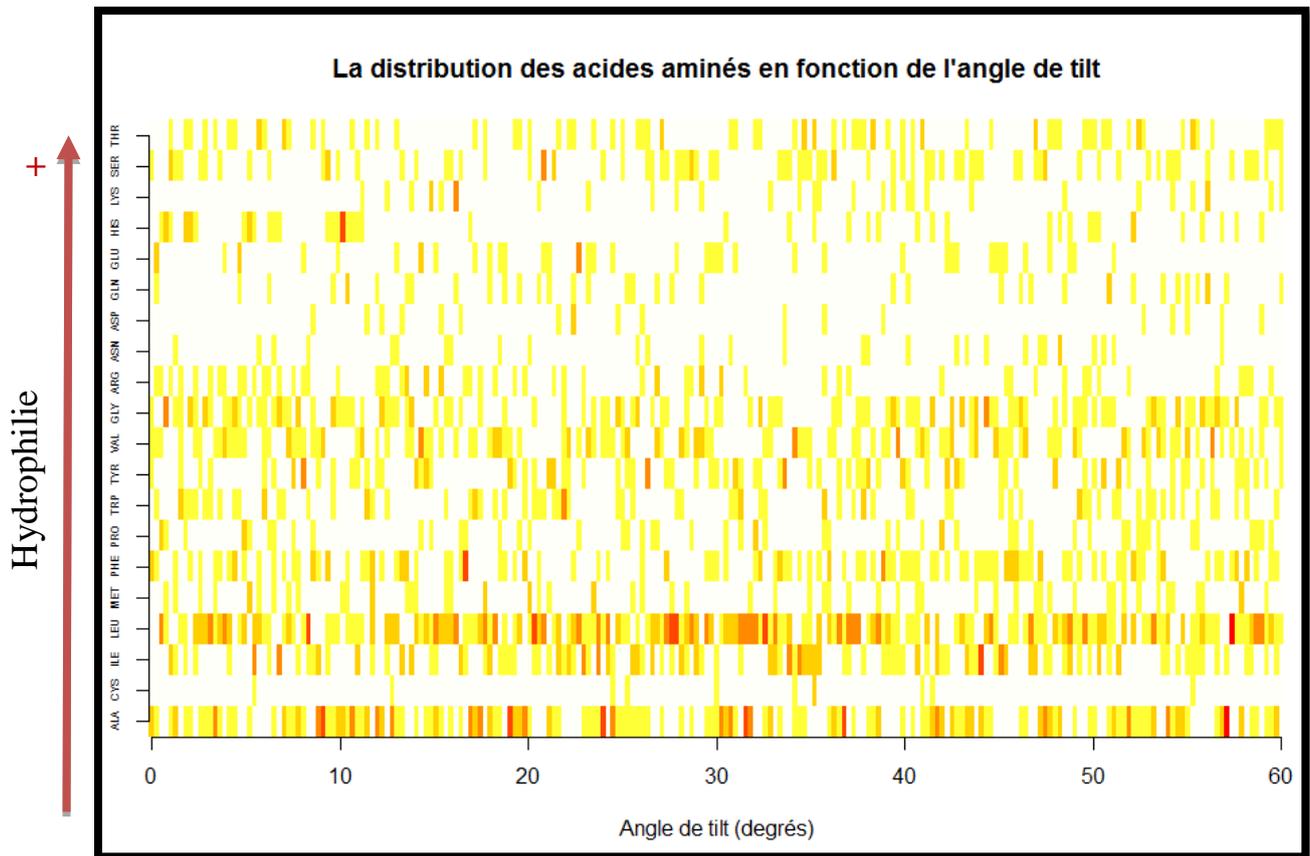
D'après cette même figure, les acides aminés aromatiques tels que la tyrosine, la proline et le tryptophane sont présents dans plusieurs intervalles d'angles. Ceci est possiblement lié au fait que la partie hydrophobe est ancrée à la membrane lipidique, du fait que leurs chaînes latérales soient à caractère hydrophile.

On remarque également (Annexe 9), que certains acides aminés hydrophiles sont absents dans des intervalles entiers d'angles, tels que l'acide aspartique et la lysine de  $0^\circ$  à  $10^\circ$ , ou très présents telle que l'histidine dans le même intervalle. Il est possible que pour deux hélices de même longueur, l'angle de *tilt* soit d'autant plus petit que certains acides aminés hydrophiles sont présents, du fait d'une moindre intensité de l'*hydrophobic mismatch* (73).

Sur la matrice, cet effet apparaît à partir du seuil d'angle ( $10$  à  $15^\circ$ ) et donc d'une la longueur d'hélice (20-25 acides aminés). Les hélices ayant une longueur compatible avec la longueur de membrane, ne sont pas *tiltées* et donc la composition du premier et dernier tour n'influence pas l'angle de *tilt*. Enfin, d'après la figure 28, on constate que les acides aminés polaires, tels que l'arginine, la glutamine et la lysine ont une fréquence qui suit l'augmentation de l'angle de *tilt*.

Ces acides aminés favorisent la conformation hélicoïdale des résidus et la continuité de l'hélice. Les hélices fortement *tiltées* sont plus enfouies dans la membrane alors que les hélices moins *tiltées* sont plus au contact de l'environnement hydrophile extérieur. De ce fait, ces acides aminés polaires favorisent la continuité de l'hélice dans l'environnement membranaire d'où leur présence dans le premier et dernier tour des hélices enfouies.

Ceci peut expliquer le fait que les acides aminés exposés, de manière générale, sont beaucoup moins conservés dans les hélices que ceux qui sont enfouis dans la membrane. Ils permettent à la protéine de s'adapter à l'environnement extérieur et assurent ainsi les fonctions des diverses protéines membranaires (interactions avec le milieu extérieur, transport, transduction du signal) (74).



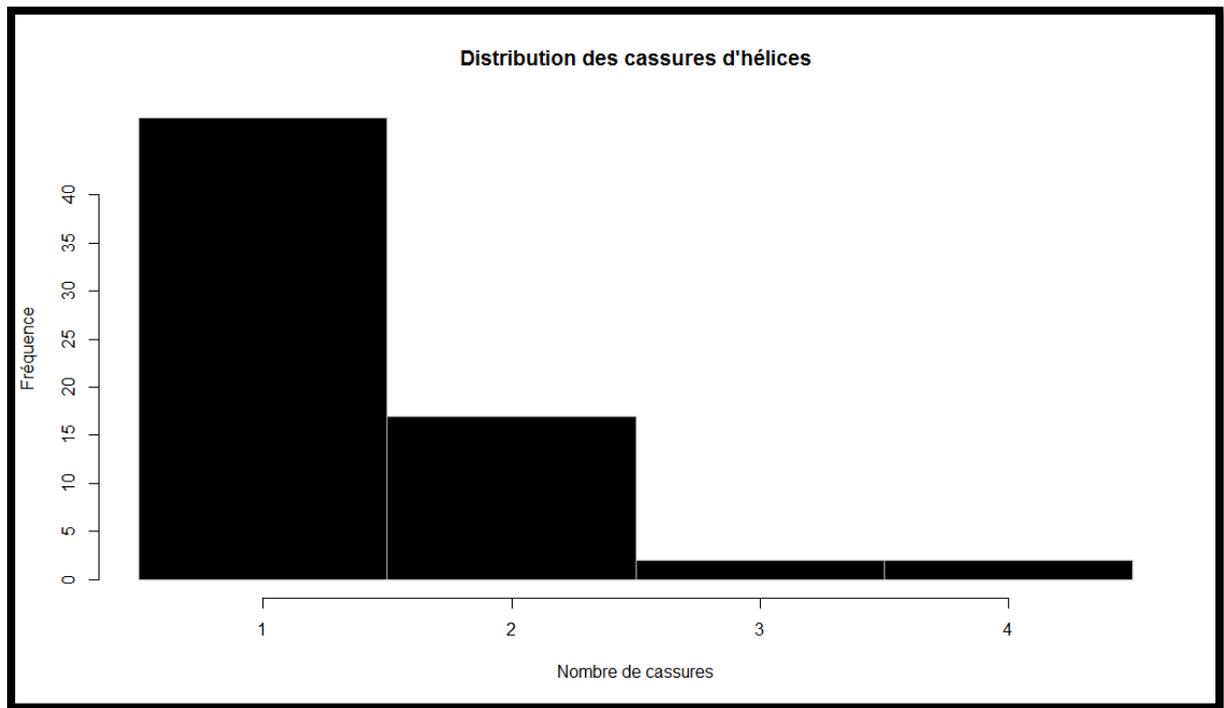
**Figure 28** – Fréquence des acides aminés dans le premier et dernier tour d’hélice en fonction de l’angle de *tilt*. Légende : spectre du blanc au rouge. Blanc : fréquence zéro, Rouge : fréquence cinq.

#### 4.2.5. Angle de *tilt* et cassures d’hélices

##### 4.2.5.1. Cassures d’hélices

Une majorité des hélices (162 – 70%) sont intactes. Les statistiques de cassures d’hélices montrent que 69 hélices sont cassées à un endroit au moins.

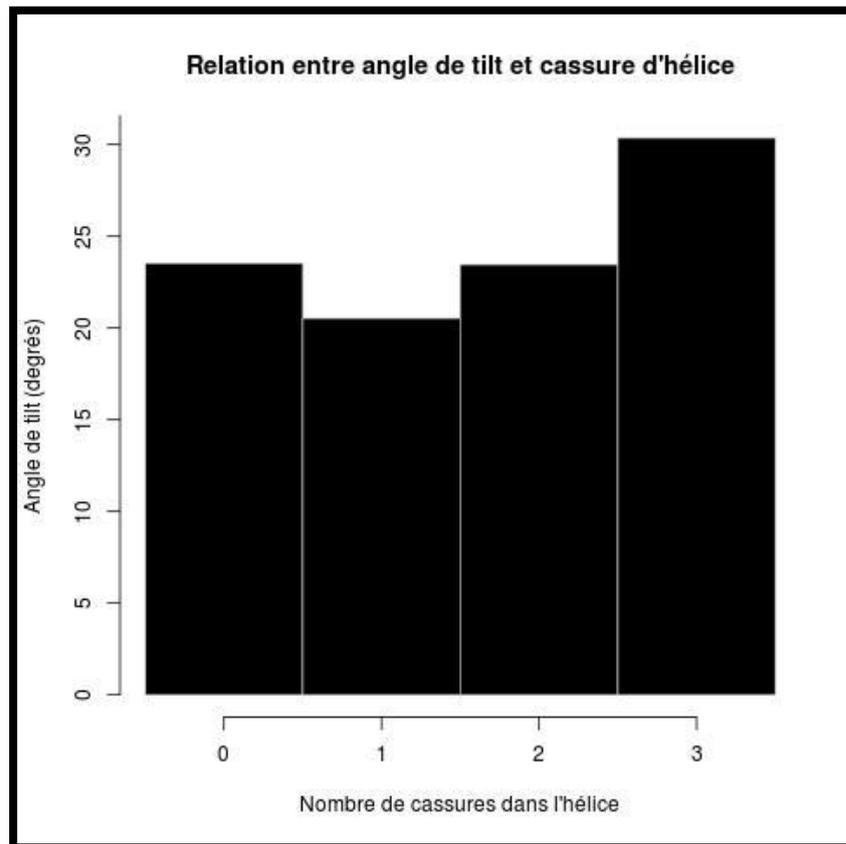
La distribution des cassures d’hélices (Figure 29) montre que la majorité des hélices cassées, le sont à un endroit. Les cassures à plus qu’un endroit sont plus rares.



**Figure 29** – Distribution des cassures d'hélices.

#### 4.2.5.2. Corrélation entre angle de *tilt* et cassures d'hélices

Sous les contraintes imposées par le *hydrophobic mismatch*, une hélice pourra soit s'incliner d'un certain angle par rapport à la membrane, soit se casser. Ce constat a été notamment confirmé par des simulations de dynamique moléculaire (73). Indépendamment de la composition de l'hélice et en supposant que l'inclinaison est préalable à la cassure, on s'attend donc à ce que la présence de cassure soit corrélée à l'angle de *tilt*. Ainsi, l'angle que forme cette cassure avec l'axe local de l'hélice, serait plus grand quand l'hélice aurait subi une contrainte plus grande, donc un angle de *tilt* plus grand. Effectivement, en prenant la moyenne des angles ayant zéro, une, deux et trois cassures, un coefficient de corrélation de *Pearson* de 0.726 (*p non significative*) a été déterminé (Figure 30).



**Figure 30** – Angle de *tilt* et nombre de cassures par hélice. Dans ce graphique les hélices ont été regroupées par nombre de cassures.

Cependant, une faible corrélation (0.344) entre l'angle de cassure et l'angle de *tilt* a été trouvée (Annexe 8). La figure montre une corrélation positive entre les deux paramètres : plus l'angle de *tilt* est grand plus l'angle de cassure l'est aussi. Il est fort probable que cette corrélation puisse être plus élevée si la base de données avait été plus importante, ceci d'autant plus que le nombre d'hélices présentant des cassures est faible dans ce jeu de données (Figure 31).

La distribution du plus grand angle de cassure de chaque hélice (Figure 32) montre que le plus grand nombre se situe entre 20° et 30°, mais la distribution est large et peut atteindre des valeurs plus élevées. Cet angle peut influencer le calcul de l'axe de l'hélice donc celui des angles de *tilt* et de *packing*.

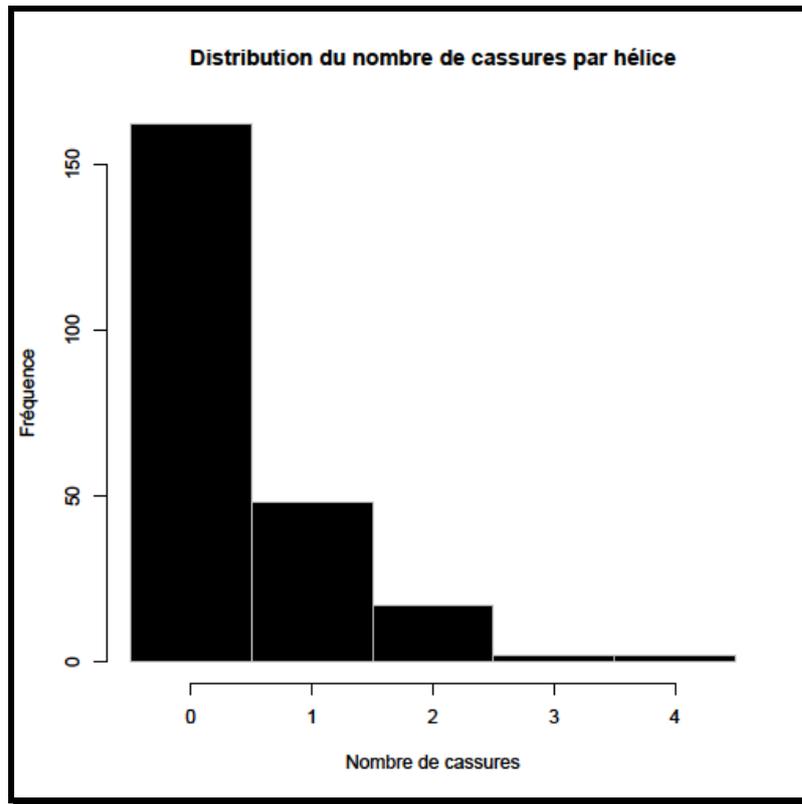


Figure 31 – Distribution des hélices en fonction des cassures.

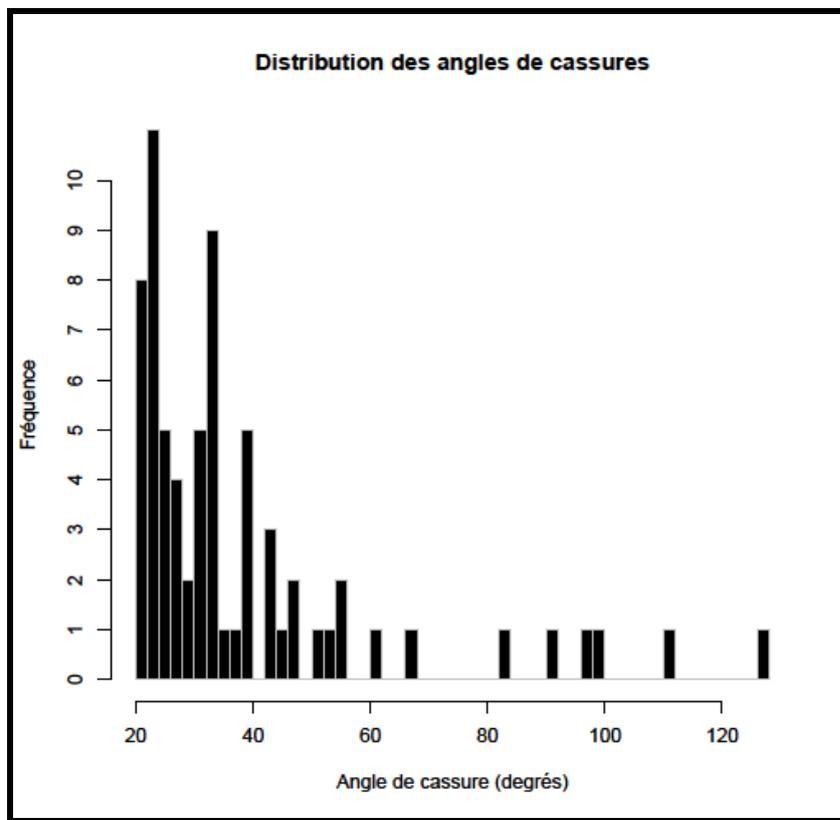


Figure 32 – Distribution des angles de cassures. Seul le plus grand angle de cassure de chaque hélice a été représenté.

## 4.2.6. Angle de *tilt* et hélices voisines

### 4.2.6.1. Hélices voisines

Une hélice est dite voisine d'une autre si elle remplit les critères de contact suivants : au moins trois résidus en contact et une distance minimale entre les centres de 25 Å). La distribution des hélices voisines (Figure 33) montre que les hélices sont rarement isolées.

En effet, 12% des hélices n'ont qu'une seule hélice voisine alors que 30% des hélices ont trois hélices voisines et 20% en ont quatre.

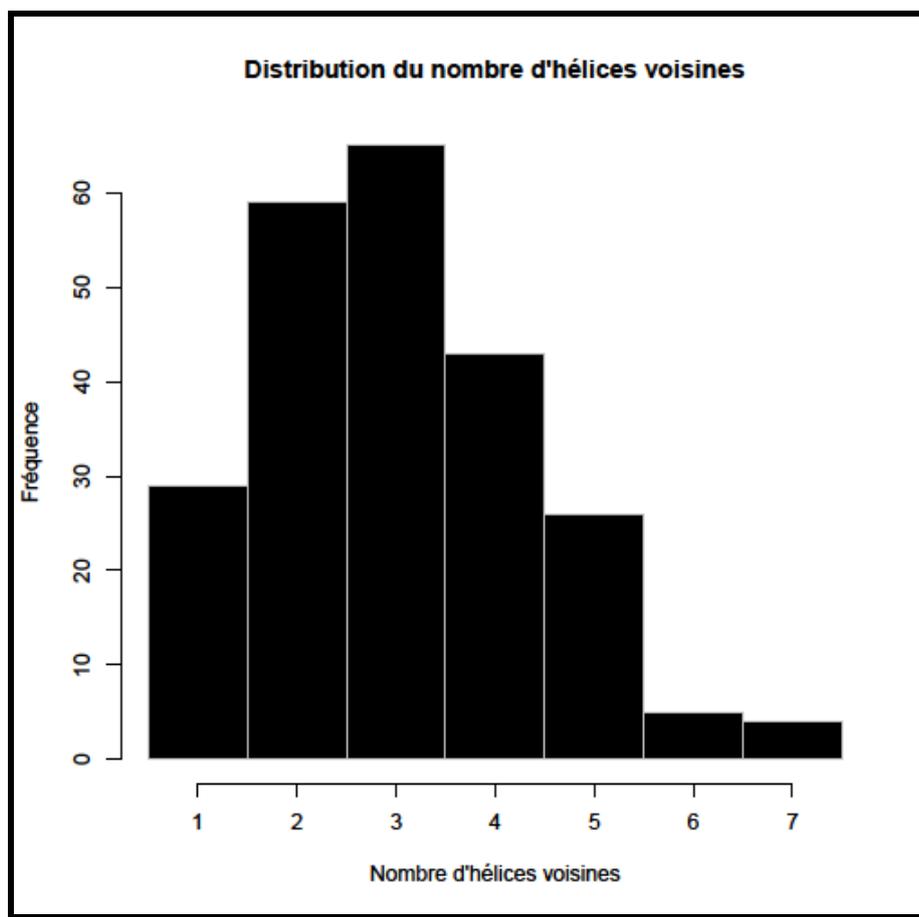
### 4.2.6.2. Corrélation entre angle de *tilt* et nombre d'hélices voisines

On peut donc penser que le nombre d'hélices avoisinant une hélice pourrait influencer son angle de *tilt*, ou que l'encombrement dans l'espace avoisinant l'hélice favorise un certain angle de *tilt*.

Néanmoins, la corrélation entre les deux paramètres est de 0.16, ce qui ne permet pas de conclure à une influence directe du nombre d'hélices voisines sur l'angle de *tilt*.

En prenant seulement les hélices ayant trois hélices voisines et plus (Figure 33) (avec un maximum à sept), la corrélation augmente à 0.24.

Il est possible que ce facteur ne soit pas à lui seul déterminant mais participe avec d'autres paramètres à l'inclinaison de l'hélice.

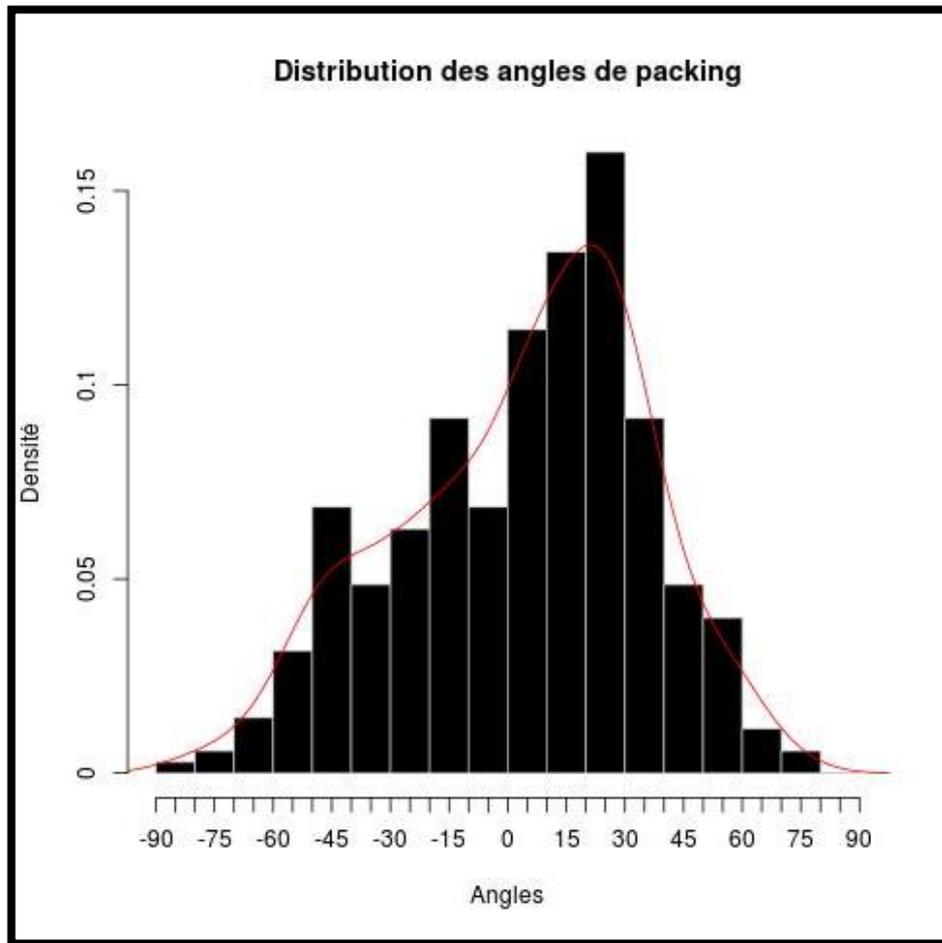


**Figure 33** – Distribution du nombre d’hélices voisines à chaque hélice.

### 4.3. Angle de *packing*

#### 4.3.1. Distribution des angles de *packing*

Sur les critères définis précédemment, l’angle de *packing* est l’angle entre deux axes d’hélices en contact. La distribution de l’angle de *packing* dans le présent travail (Figure 34) montre que l’intervalle préféré est compris entre  $[0,30^\circ]$ . Dans l’article de référence (54), l’auteur montre également que l’angle préféré de *packing* se situe dans l’intervalle de 0 à  $30^\circ$ , mais ce n’est pas la valeur d’angle la plus observée dans l’analyse effectuée sur 88 interactions entre hélices (54). Les valeurs d’angles les plus observées ont été catégorisées en classes (*a*, *b* et *c*) (Figure 35) (voir paragraphe suivant). Le fait que la base de données soit plus grande a permis d’estimer le biais statistique qui favorise les grands angles (54).

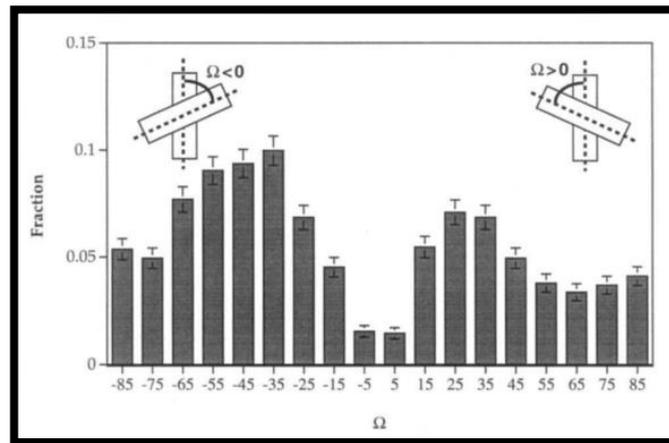


**Figure 34** – Histogramme de distribution des angles de *packing*. Il s’agit d’une distribution binomiale où les pics se situent à  $-45^\circ$  et à  $+30^\circ$ .

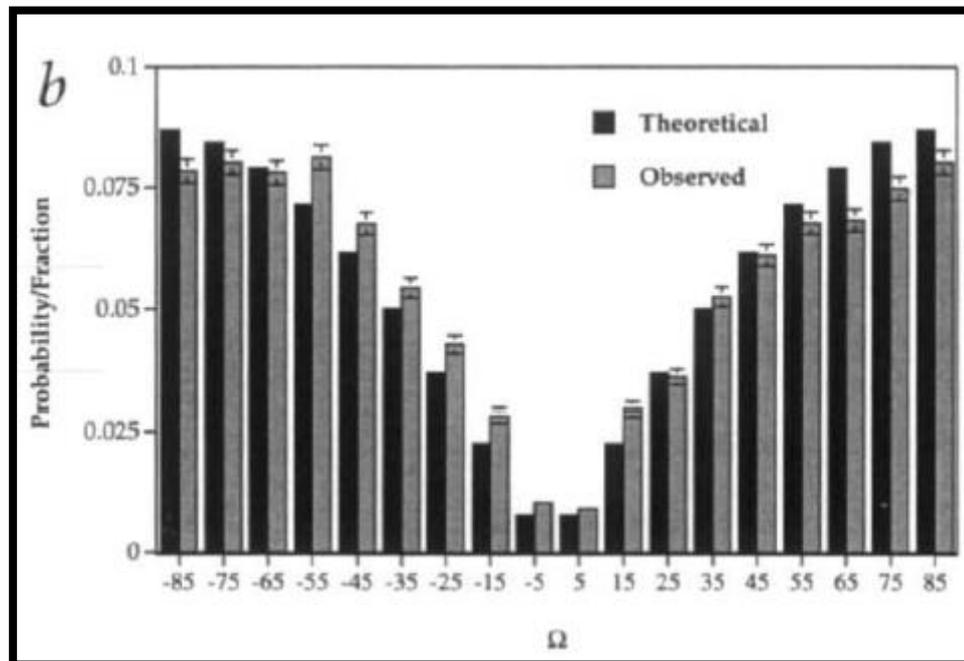
Néanmoins, une correction du biais statistique par une distribution au hasard d’angles de *packing* où les hélices ne sont soumises à aucune contraintes et sans contacts entre elles (Figure 36) (55), a été effectuée selon la formule suivante (Figure 37) :

*Densité corrigée (Fig 37)*

$$= \frac{\text{Densité observée (Fig 34)}}{\text{Densité d'hélices qui ne sont pas en contact (Fig 36)}}$$



**Figure 35** – La distribution de l’angle de *packing* dans les protéines solubles (55).



**Figure 36** – Distribution des angles de *packing* d’hélice qui ne sont pas en interaction (55). La distribution observée a été faite sur une base de données de protéines globulaires et membranaires totalisant 12,605 angles de *packing*.

Les interactions incluses dans cette distribution (Figure 36) sont celles d’hélices qui n’ont aucun résidu en contact et qui ont plus de 15 Å de distance entre elles. Cette distribution a permis de corriger la distribution observée des angles de

*packing* observée entre hélices en interactions. La correction s'est faite en divisant les densités de distribution.

Dans cette figure, les valeurs d'angles ayant une préférence supérieure à 1 sont significativement préférées par les hélices. L'intervalle de  $0^\circ$  à  $30^\circ$  ou classe *c* (52), décrit par *Chothia* (classes a, b et c respectivement  $-52^\circ$ ,  $+75^\circ$ ,  $+23^\circ$ ) (52) et *Walther* (classes a, b et c respectivement  $-37^\circ$ ,  $+83^\circ$ ,  $+22^\circ$ ) (75), est effectivement l'intervalle d'angle de *packing* préféré par l'interaction hélice-hélice. Cette contrainte est absente dans les protéines solubles où les angles de *packing* varient largement (Figure 35) (55).

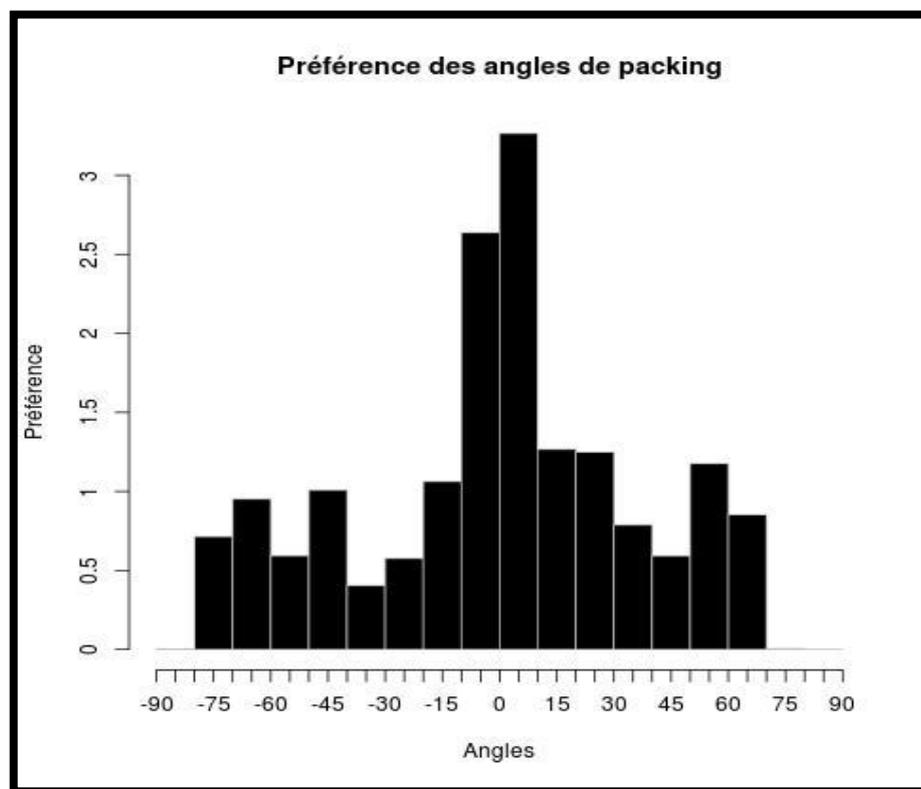
Ce constat vient du fait que les hélices transmembranaires sont soumises aux forces présentes dans la membrane lipidiques décrites auparavant pour les angles de *tilt*. De plus, la plus grande proximité entre les hélices dans cet environnement favorise une plus grande surface de contact, donc un petit angle. De ce fait les grands angles de *packing* ne sont pas favorisés dans les protéines membranaires. Ce constat a été démontré par plusieurs modèles (76, 77) prenant en compte plusieurs paramètres intervenant dans le *packing* : le rayon de *Van der Waals* et les volumes de contacts. Les raisons de cette préférence pour ces valeurs d'angles ont été amputées au milieu hydrophobe de la membrane lipidique, mais il a été démontré que la composition même des hélices en est la principale cause (78). Les petits résidus assurant un *packing* plus rapproché, sont plus fréquemment observés dans les hélices transmembranaires que dans les hélices des protéines solubles (78) où des résidus plus grands sont plus fréquemment observés dans l'interface de contact. Il en résulte une distribution où trois valeurs d'angles sont les plus observées (Figure 35), les classes *a*, *b* et *c* correspondant respectivement aux valeurs d'angles  $-37^\circ$ ,  $+83^\circ$  et  $+22^\circ$ . Ceci renforce la théorie de l'influence du milieu sur la préférence pour les angles de *packing*.

### 4.3.2. Angle de *packing* et hélices voisines de séquence

#### 4.3.2.1. Propriétés des hélices voisines de séquence

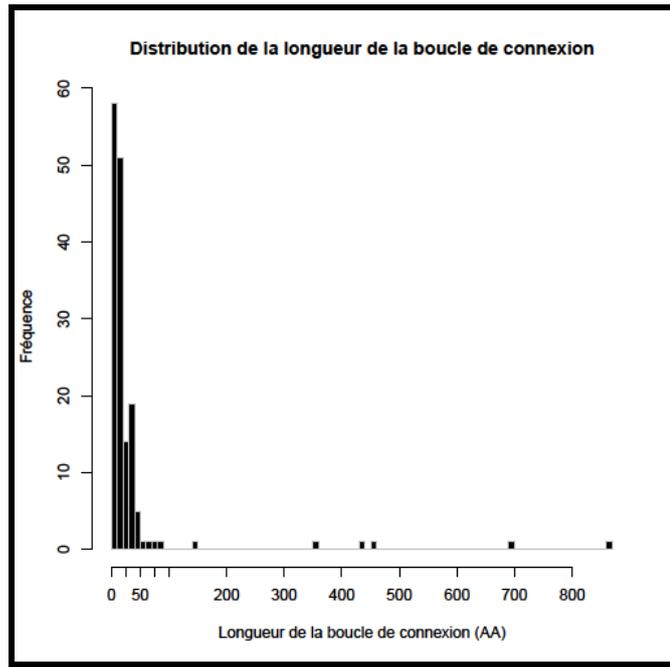
L'analyse des interactions a montré que 82.43% des hélices voisines en séquence sont en interaction. Ceci confirme plusieurs modèles où il est communément admis que les hélices voisines en séquence sont automatiquement en interaction (79).

La moyenne de la longueur de la boucle de connexion entre les 157 hélices voisines de séquences (Figure 38 A) et en interaction est de 35.68 acides aminés avec un minimum à deux et un maximum à 868. D'autre part, les 44 hélices voisines (Figure 38 B), qui ne sont pas en interaction, ont une longueur de boucle de connexion moyenne de 62.79 acides aminés, variant de trois à 465, donc deux fois supérieure à la moyenne des hélices en interaction. Ce résultat montre que ce paramètre est une valeur prédictive pour déterminer si les hélices voisines interagissent ou non.

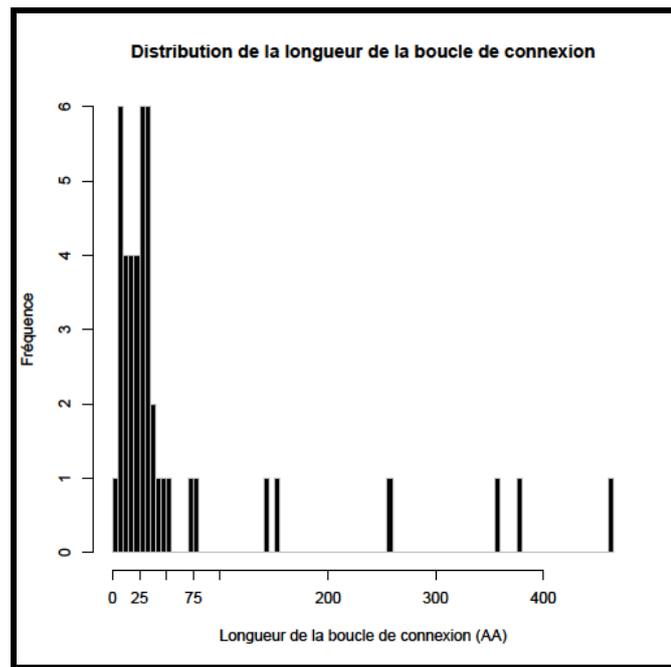


**Figure 37** – Préférence des angles de *packing* après correction du biais statistique.

A



B



**Figure 38** – A – Distribution de la longueur de boucle de connexion des hélices voisines qui sont en interaction. B – Distribution de la longueur de boucle de connexion des hélices voisines qui ne sont pas en interaction.

#### 4.3.2.2. Corrélation entre angle de *packing* et hélices voisines de séquence

Concernant l'influence de la promiscuité des hélices en interaction avec l'angle de *packing*, aucune corrélation significative n'a pu être déterminée. Les hélices voisines de séquences ont toutefois un angle de *packing* dans les classes *a* et *c* (Figure 39 B). Les hélices non voisines de séquence préfèrent un angle de *packing* moins grand, celui de la classe *c* (Figure 39 A).

#### 4.3.3. Autres paramètres des hélices en interaction et leurs angles de *packing*

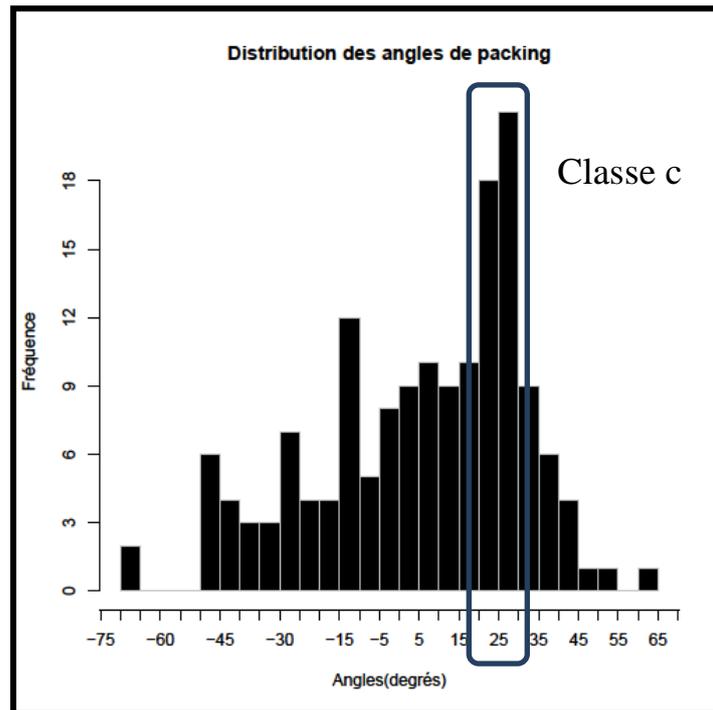
##### 4.3.3.1. Orientation des hélices en interaction

L'orientation antiparallèle est favorisée dans les interactions avec 66% des hélices qui présentent cet arrangement ; ceci vient en accord avec les résultats trouvés dans d'autres études. Les contraintes topologiques favoriseraient cette orientation pour les hélices d'une même sous-unité (79), d'autant plus que les hélices voisines de séquence sont presque automatiquement en interaction.

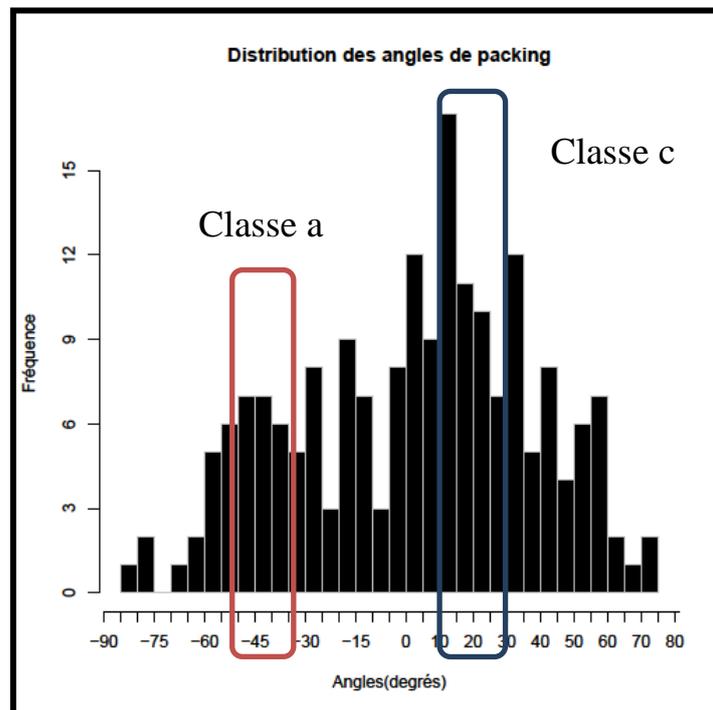
Par conséquent, nous avons choisi d'étudier les interactions entre hélices de sous-unités différentes afin de minimiser l'impact de ces contraintes. Ceci a mis en évidence le fait que 65% des hélices de sous unités différentes en interaction sont aussi antiparallèles. Il est donc clair que les contraintes topologiques n'influencent pas à elles seules l'orientation antiparallèle des hélices en interaction et que cette orientation stabiliserait mieux le *packing*.

La distance moyenne entre les hélices à orientation parallèle est de 5.68 Å contre 5.41 Å pour l'orientation antiparallèle, sachant que la moyenne pour toutes les interactions est de 5.5Å avec un écart-type de 1.2 Å.

A

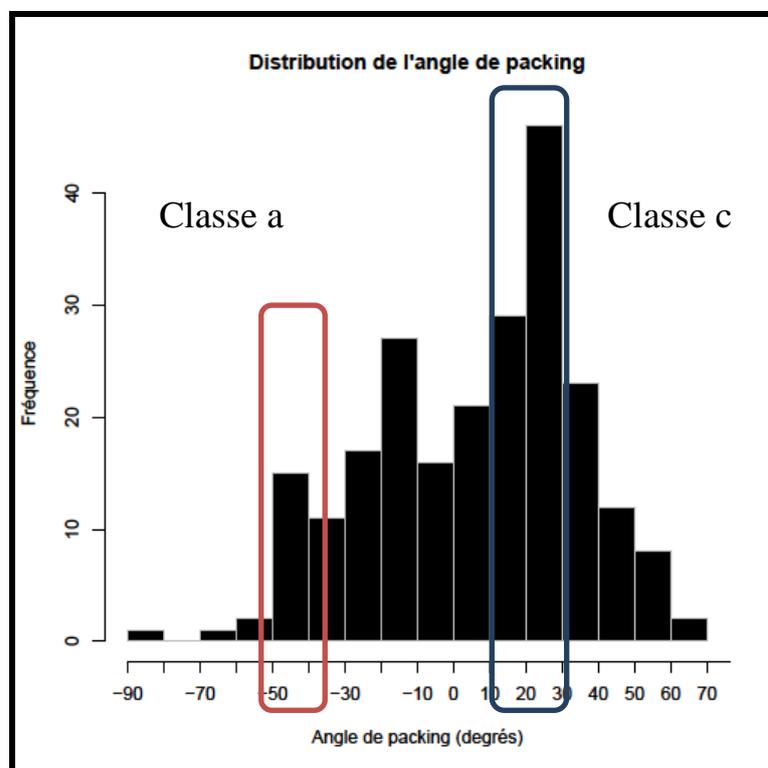


B

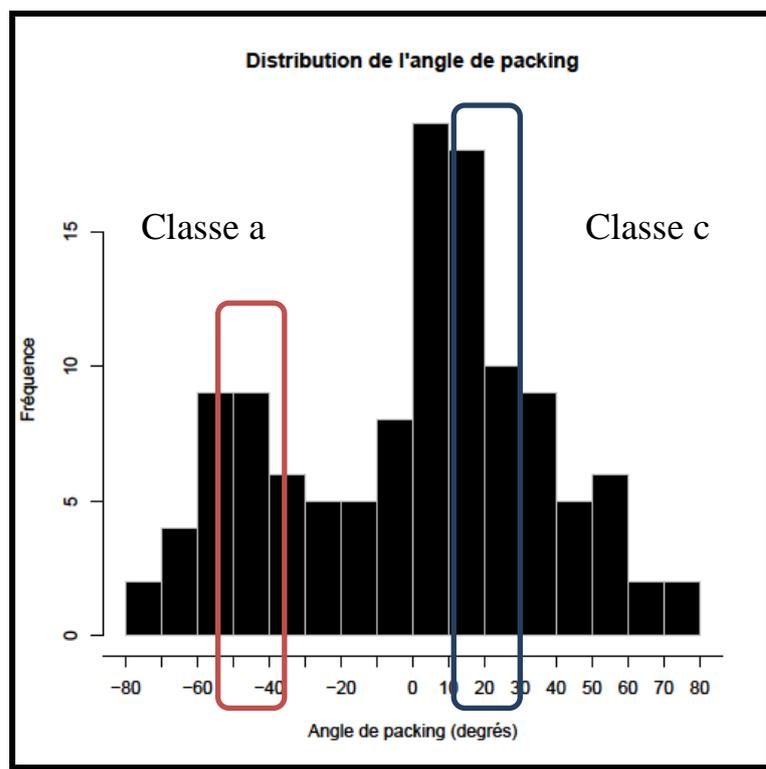


**Figure 39** – A – Distribution des angles de *packing* des hélices qui sont voisines de séquence. B – Distribution des angles de *packing* des hélices qui ne sont pas voisines de séquence.

Le nombre de contacts est lui aussi augmenté pour l'orientation antiparallèle avec une moyenne de 52.83 contacts contre 43.54 contacts pour l'orientation parallèle. Ceci démontre que l'orientation antiparallèle, avec une distance plus faible et un plus grand nombre de contacts, est favorisée dans les interactions, car elle est plus stable. Plusieurs hypothèses ont été formulées quant à la préférence pour l'orientation antiparallèle (75, 80-82). Néanmoins, la différence observée entre les orientations n'est pas décisive, surtout en comparant le nombre de contacts maximum. En effet, il s'élève à 148 pour les *packings* parallèles contre 125 pour les *packings* antiparallèles. La distance minimum de contact est de 3.73 Å pour l'orientation parallèle et de 3.78 Å pour l'orientation antiparallèle. Ceci démontre que les hélices en orientation parallèle peuvent aussi avoir un contact très rapproché et peuvent également être très stables selon certains modèles (83). Les Figures 40 et 41 montrent que la distribution des angles de *packing* des deux types d'orientations se situent préférentiellement dans les classes *a* et *c*.



**Figure 40** – Distribution de l'angle de *packing* des hélices en orientation antiparallèle.

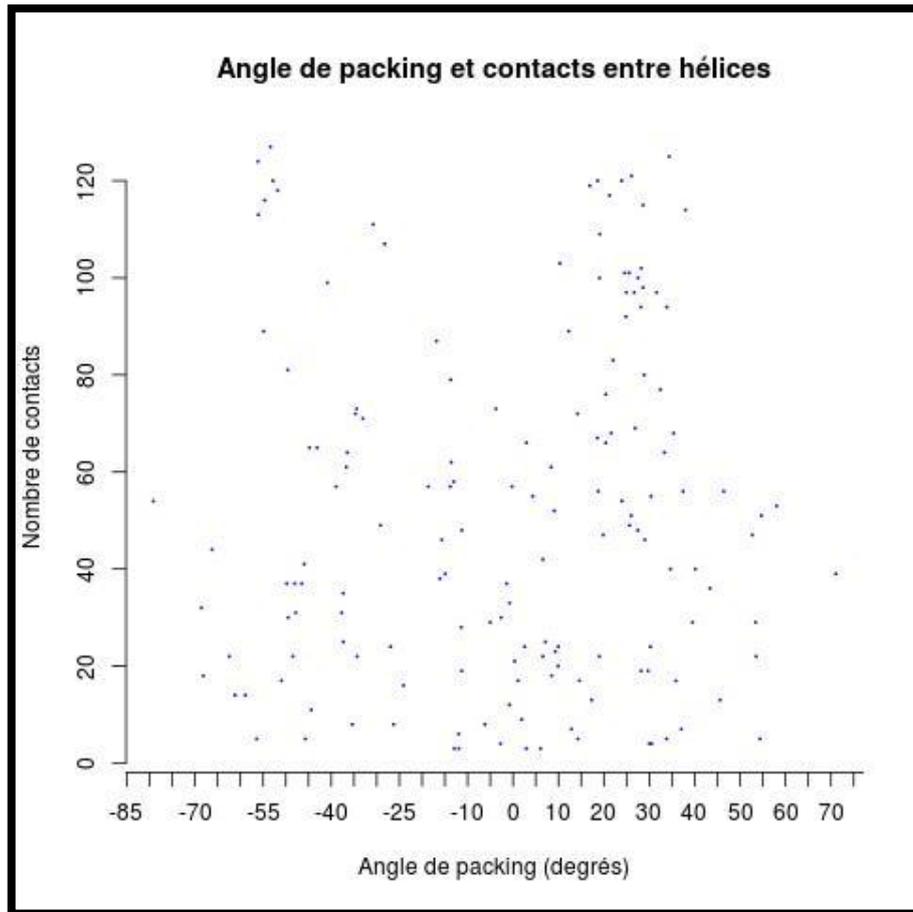


**Figure 41** – Distribution de l'angle de *packing* des hélices en orientation parallèle.

#### 4.3.3.2. Corrélation entre angle de *packing*, le nombre de contacts et la distance minimum entre les hélices en interaction

Cette démarche vise à déterminer si le nombre de contacts et la distance minimum entre les hélices exercent une influence sur l'angle de *packing* entre deux hélices en interaction.

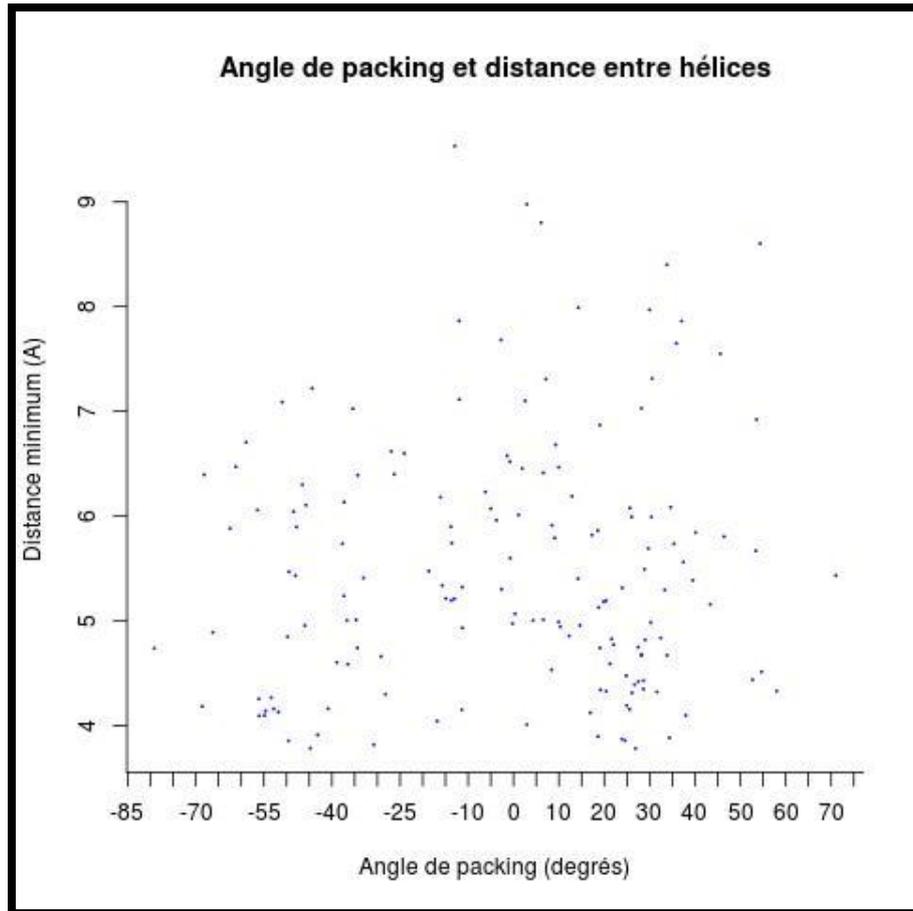
Pour ce faire, seules les interactions entre hélices sans cassures sont prises en compte. Le nombre de contacts est proportionnellement faible pour les petits angles, augmente pour un certain intervalle d'angle (autour de  $+30^\circ$  et  $-55^\circ$ ) et diminue pour les très grands angles. D'après la figure 42, il semblerait que les contacts ne soient pas en corrélation directe avec l'angle de *packing*.



**Figure 42** – Distribution de l’angle de *packing* en fonction du nombre de contacts entre hélices.

Il était attendu qu’un plus grand nombre de contacts soit observé pour les petits angles et diminue à mesure que l’angle devienne plus grand.

Concernant la distance minimum séparant deux hélices en interaction (Figure 43), on constate un phénomène semblable : une faible promiscuité pour les petits angles autour de  $0^\circ$ , puis une augmentation pour une certaine plage d’angle en particulier autour de  $25^\circ$  et de  $50^\circ$  pour diminuer encore pour les très grands angles. Cette distribution binomiale suit en effet celle de la distribution des angles de *packing* (Figure 26) où l’on remarque deux pics autour de  $-50^\circ$  et  $20^\circ$ .



**Figure 43** – Représentation de la distance minimale entre deux hélices en fonction de l’angle de packing. La distribution est inversée par rapport à celle des contacts. Deux minima sont observés autour de  $+25^\circ$  et  $-55^\circ$ . La corrélation entre les deux paramètres est égale à 0.137.

Cette observation est en accord avec la théorie de *Chothia* sur les classes d’angles les plus observés pour les *packings* (84).

Cette théorie généraliste a néanmoins été formulée sur les protéines solubles dont les structures sont plus faciles à obtenir. Les plages d’angles autour de  $-50^\circ$  et  $+30^\circ$  correspondent en effet aux classes *a* et *c* de *packing*. Mais ces contraintes physiques ne sont pas responsables à elles seules de la détermination de l’angle de *packing* préféré par les hélices.

Dans le point 4.3.1, la distribution des angles a été corrigée par une distribution d'angles d'hélices qui ne sont pas en interaction, donc sans contacts, ce qui suffit à corriger le biais statistique imposé par les contraintes d'orientations.

En résumé, le nombre de contacts et la distance minimum entre les hélices dresse un profil similaire à la distribution des angles de *packing*, mais ne constitue pas un facteur qui pourrait déterminer la préférence pour une certaine valeur d'angle.

Il serait également intéressant de calculer le contact par des méthodes de tessellation, qui offrent une perspective plus réaliste des contacts que les méthodes basées sur le seuil (85, 86).

#### **4.3.3.3. Les cassures dans les hélices en interaction**

Il s'agit de déterminer l'influence que pourrait avoir les cassures dans l'une ou l'autre des hélices en interaction sur l'angle de *packing* (Figure 44).

En effet, la cassure modifie l'axe de l'hélice et on trouve que le plus grand nombre de cassure est situé dans les grands angles ( $65^\circ$ ) mais également les petits angles autour de  $0^\circ$ .

Cette matrice démontre donc que la présence d'une cassure dans une hélice peut influencer l'angle de *packing*.



**Figure 44** – L'influence des cassures des hélices sur l'angle de *packing*.

Légende : spectre du blanc au rouge. Blanc : zéro cassure, Rouge : six cassures.

Le nombre de cassures totales a été également calculé en fonction de l'angle de *packing*.

La cassure peut orienter l'axe d'une hélice dans le sens d'un rapprochement avec la seconde hélice (petit angle) ou au contraire dans le sens contraire, ce qui résulte en un grand angle de *packing*.

Toutefois, aucune influence systématique des cassures sur les angles de *packing* n'a pu être déterminée. Plusieurs méthodes ont été proposées afin d'éviter cette influence sur le calcul de l'axe de l'hélice.

Il est possible, en effet, de calculer les angles de *packing* locaux au foyer du contact (entre les deux axes locaux en contact) (57, 58).

#### 4.3.3.4. Angle de *packing* et moment hydrophobe

Quelle est l'influence du caractère amphiphile des hélices en interaction sur l'angle qu'elles forment entre elles ? Afin de répondre à cette question, le moment hydrophobe d'*Eisenberg* de chaque hélice en interaction a été représenté en fonction de l'angle de *packing* (Figure 45). Le moment hydrophobe des hélices dans la base de données se situe entre 0.4 et 1.36 avec une hydrophobicité moyenne de 0.17.

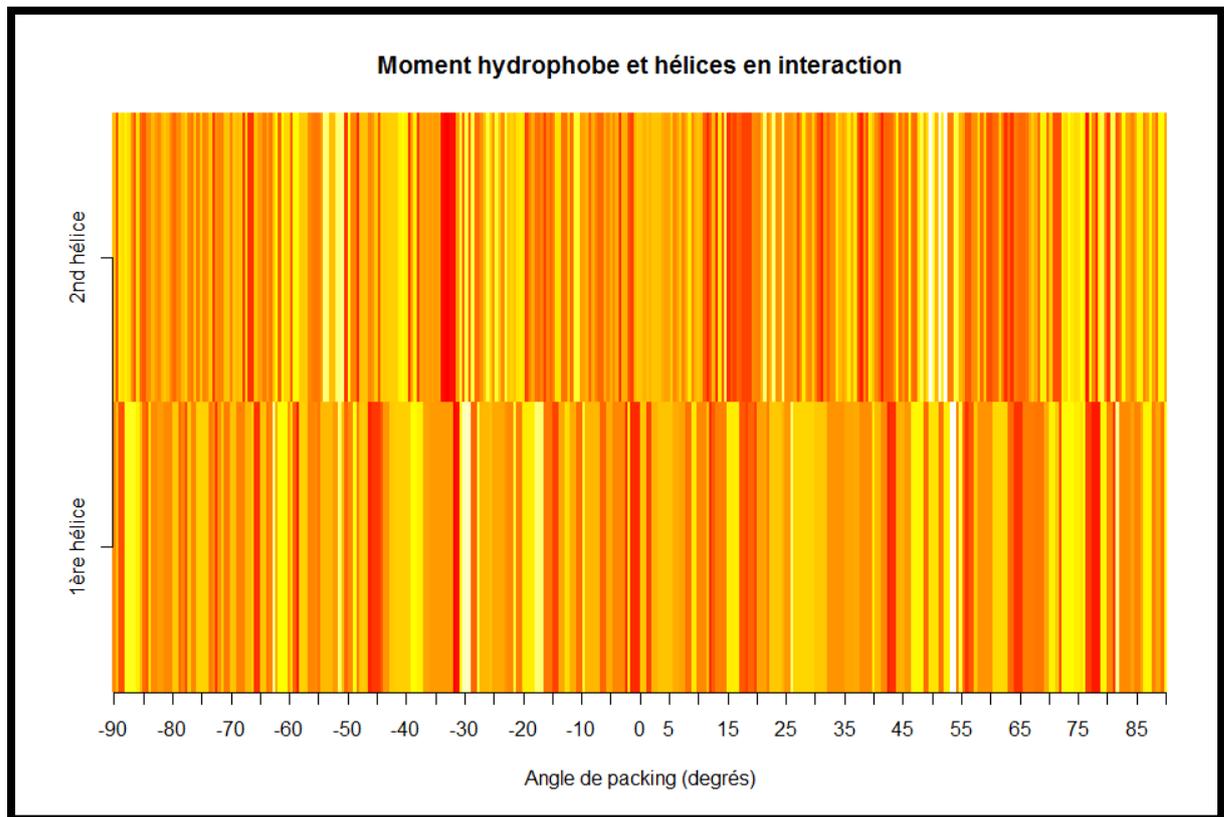
On remarque que les hélices ayant un caractère amphiphile proche tendent à former de petits angles surtout aux alentours de 0°.

Au contraire les hélices ayant un caractère amphiphile différent favorisent les grands angles surtout +50°, +70°, -50°. Cette règle n'est pas générale car pour les valeurs d'angles de -65° à -85°, les hélices ont un moment hydrophobe comparable. Il existe plusieurs raisons qui pourraient expliquer le fait qu'il n'y ait pas de corrélation directe entre les deux paramètres.

La première est que les hélices dans les segments transmembranaires sont à la fois très hydrophobes et qu'ils ont un caractère amphiphile peu prononcé, les différences entre les moments hydrophobes sont donc faibles et ne peuvent jouer un rôle dans l'interaction entre hélice.

Ensuite, le moment hydrophobe est un vecteur orienté, il est possible que deux hélices ayant un caractère amphiphile différent forment un petit angle, ceci dépend de la configuration spatiale de l'interaction entre les hélices.

La différence entre le moment hydrophobe des couples d'hélices en interaction (Figure 46) montre que dans l'intervalle d'angle de [-30°,55°], une faible différence du moment hydrophobe tend à favoriser les petits angles et vice versa.

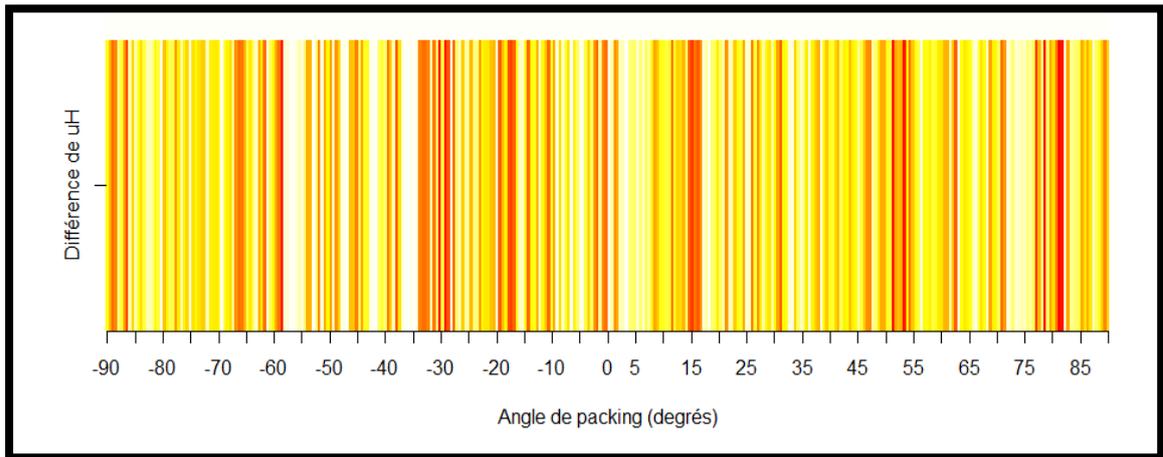


**Figure 45** – Corrélation entre le moment hydrophobe d'*Eisenberg* des hélices en interaction et l'angle de *packing*. Légende : spectre du blanc au rouge. Blanc : 0.4, Rouge : 1.36.

L'exception à la règle se situe essentiellement aux angles n'appartenant pas à cet intervalle et qui ne démontrent aucune tendance. Il est possible que l'influence du moment hydrophobe diminue au-delà d'une certaine distance entre les hélices.

En effet, le nombre d'interactions aux extrêmes est faible, comme peut l'attester la distribution des angles de *packing* (Figure 34). Certains modèles ne confèrent pas à l'hélice transmembranaire un tel degré de liberté et la contraignent à former un angle de *packing* situé entre  $-70^\circ$  et  $+70^\circ$  (54).

D'une manière plus générale, il semblerait qu'il n'existe pas de tendance systématique ou de corrélation entre l'angle de *packing* et le caractère amphiphile d'une hélice.



**Figure 46** – Différence du moment hydrophobe entre les couples d'hélices en interaction en fonction de l'angle de *packing*. Légende : spectre du blanc au rouge. Blanc : 0.002, Rouge : 3.26.

## 5. CONCLUSION

La résolution de la structure des protéines membranaires revêt un aspect important dans la conception des molécules à intérêt thérapeutique puisque ces protéines sont dans la plupart des cas, des cibles thérapeutiques. L'amélioration des techniques biophysiques de résolution des structures a permis d'avoir un nombre de plus en plus important des structures tridimensionnelles disponibles. En réalité, ce nombre reste faible par rapport aux protéines solubles, à cause des difficultés techniques liées à la topologie complexe de ces protéines. Le séquençage à haut-débit a permis d'avoir un grand nombre de séquences protéiques disponibles. La modélisation *in silico* paraît donc une alternative aux méthodes classiques de résolution de la structure. Pour ce faire, Il faut déterminer quels paramètres permettent d'inférer la structure des protéines à partir de la séquence protéique. Le présent travail s'est intéressé en particulier aux hélices  $\alpha$  transmembranaires et aux angles qu'elles forment entre elles (*packing*) et avec le plan de la protéine (*tilt*). Une base de données de protéines membranaires non redondantes a été utilisée pour l'analyse statistique en incluant plusieurs paramètres tels que la longueur de l'hélice et son hydrophobicité. L'avantage que présente les protéines membranaires sur les protéines globulaires est qu'elles sont présentes dans un environnement lipidique qui impose des contraintes stériques aux hélices, ce qui réduit l'espace possible des paramètres (87). Ces paramètres sont donc orientés vers certaines plages de valeurs qui assurent une stabilité énergétique et stérique au sein de la membrane. Ce travail s'inscrit, comme d'autres, dans l'effort d'orienter le processus de modélisation des protéines membranaires, en fournissant entre autres un outil d'analyse de structures protéiques et a permis de mettre en évidence plusieurs contraintes telles que la relation entre la longueur de l'hélice et l'angle de *tilt* ainsi que la prédiction de l'interaction de deux hélices en fonction de la longueur de leur boucle de connexion.

## REFERENCES

1. **le Coutre J, Kaback HK.** Structure-function relationships of integral membrane proteins: membrane transporters vs channels. *Biopolymers*. 2000;55(4):297-307.
2. **Allen TW, Separovic F.** Membrane protein structure and function. *Biochim biophys acta*. 2012;1818(2):125.
3. **Arnaout MA, Mahalingam B, Xiong JP.** Integrin structure, allostery, and bidirectional signaling. *Annu rev cell dev bi*. 2005;21:381-410.
4. **Campbell ID, Humphries MJ.** Integrin structure, activation, and interactions. *Cold Spring Harb perspect biol*. 2011;3(3).
5. **Hynes RO.** Integrins: bidirectional, allosteric signaling machines. *Cell*. 2002;110(6):673-87.
6. **Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, et al.** The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity. *Science*. 1998;280(5360):69-77.
7. **Giacomini KM, Huang SM, Tweedie DJ, Benet LZ, Brouwer KL, Chu X, et al.** Membrane transporters in drug development. *Nat Rev Drug Discov*. 2010;9(3):215-36.
8. **Venter H, Ashcroft AE, Keen JN, Henderson PJ, Herbert RB.** Molecular dissection of membrane-transport proteins: mass spectrometry and sequence determination of the galactose-H<sup>+</sup> symport protein, GalP, of *Escherichia coli* and quantitative assay of the incorporation of [ring-2-<sup>13</sup>C]histidine and (<sup>15</sup>NH<sub>3</sub>). *Biochem j*. 2002;363(Pt 2):243-52.
9. **de Brevern AG, Wong H, Tournamille C, Colin Y, Le Van Kim C, Etchebest C.** A structural model of a seven-transmembrane helix receptor: the Duffy antigen/receptor for chemokine (DARC). *Biochim biophys acta*. 2005;1724(3):288-306.
10. **Sahin-Toth M, Karlin A, Kaback HR.** Unraveling the mechanism of the lactose permease of *Escherichia coli*. *P Natl Acad Sci USA*. 2000;97(20):10729-32.

11. **Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC.** A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*. 1958;181(4610):662-6.
12. **Deisenhofer J, Epp O, Miki K, Huber R, Michel H.** X-ray structure analysis of a membrane protein complex. Electron density map at 3 Å resolution and a model of the chromophores of the photosynthetic reaction center from *Rhodospseudomonas viridis*. *J mol biol*. 1984;180(2):385-98.
13. **Bernstein FC, Koetzle TF, Williams GJ, Meyer EF, Jr., Brice MD, Rodgers JR, et al.** The Protein Data Bank: a computer-based archival file for macromolecular structures. *J mol biol*. 1977;112(3):535-42.
14. **Arora A, Tamm LK.** Biophysical approaches to membrane protein structure determination. *Curr opin struc biol*. 2001;11(5):540-7.
15. **Andronesi OC, Becker S, Seidel K, Heise H, Young HS, Baldus M.** Determination of membrane protein structure and dynamics by magic-angle-spinning solid-state NMR spectroscopy. *J Am Chem Soc*. 2005;127(37):12965-74.
16. **White SH.** The progress of membrane protein structure determination. *Protein sci : a publication of the Protein Society*. 2004;13(7):1948-9.
17. **Bill RM, Henderson PJ, Iwata S, Kunji ER, Michel H, Neutze R, et al.** Overcoming barriers to membrane protein structure determination. *Nat biotechnol*. 2011;29(4):335-40.
18. **Smith SM.** Strategies for the purification of membrane proteins. *Methods Mol Biol*. 2011;681:485-96. Epub 2010/10/28.
19. **Zoonens M, Miroux B.** Expression of membrane proteins at the *Escherichia coli* membrane for structural studies. *Methods Mol Biol*. 2010;601:49-66.
20. **Caffrey M.** Crystallizing membrane proteins for structure-function studies using lipidic mesophases. *Biochem Soc t*. 2011;39(3):725-32.
21. **Li D, Lee J, Caffrey M.** Crystallizing Membrane Proteins in Lipidic Mesophases. A Host Lipid Screen. *Cryst growth des*. 2011;11(2):530-7.
22. **Carpenter EP, Beis K, Cameron AD, Iwata S.** Overcoming the challenges of membrane protein crystallography. *Curr opin struc biol*. 2008;18(5):581-6.

23. **de Brevern AG.** 3D structural models of transmembrane proteins. *Methods Mol Biol.* 2010;654:387-401.
24. **Leach AR.** *Molecular modelling: principles and applications*: Pearson College Division; 2001.
25. **Luthy R, Bowie JU, Eisenberg D.** Assessment of protein models with three-dimensional profiles. *Nature.* 1992;356(6364):83-5.
26. **Collins FS, Morgan M, Patrinos A.** The Human Genome Project: lessons from large-scale biology. *Science.* 2003;300(5617):286-90.
27. **Schadt EE, Turner S, Kasarskis A.** A window into third-generation sequencing. *Hum mol genet.* 2010;19(R2):R227-40.
28. **Mistry J, Kloppmann E, Rost B, Punta M.** An estimated 5% of new protein structures solved today represent a new Pfam family. *Acta crystallogr D, Biological crystallography.* 2013;69(Pt 11):2186-93.
29. **Centeno NB, Planas-Iglesias J, Oliva B.** Comparative modelling of protein structure and its impact on microbial cell factories. *Microb cell fact.* 2005;4:20.
30. **de Fays K, Tibor A, Lambert C, Vinals C, Denoel P, De Bolle X, et al.** Structure and function prediction of the *Brucella abortus* P39 protein by comparative modeling with marginal sequence similarities. *Protein eng.* 1999;12(3):217-23.
31. **Liu T, Tang GW, Capriotti E.** Comparative modeling: the state of the art and protein drug target structure prediction. *Comb chem high t scr.* 2011;14(6):532-47.
32. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al.** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids res.* 1997;25(17):3389-402.
33. **Ginalski K.** Comparative modeling for protein structure prediction. *Curr opin struc biol.* 2006;16(2):172-7.
34. **Sali A, Blundell TL.** Comparative protein modelling by satisfaction of spatial restraints. *J mol biol.* 1993;234(3):779-815.
35. **Doerr A.** Membrane protein structures. *Nat methods.* 2009;6(1).

- 36. Dill KA, Ozkan SB, Weikl TR, Chodera JD, Voelz VA.** The protein folding problem: when will it be solved? *Curr opin struc biol.* 2007;17(3):342-6.
- 37. Bowie JU, Luthy R, Eisenberg D.** A method to identify protein sequences that fold into a known three-dimensional structure. *Science.* 1991;253(5016):164-70.
- 38. Moult J, Fidelis K, Kryshchuk A, Tramontano A.** Critical assessment of methods of protein structure prediction (CASP)--round IX. *Proteins.* 2011;79 Suppl 10:1-5.
- 39. Krogh A, Larsson B, von Heijne G, Sonnhammer EL.** Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J mol biol.* 2001;305(3):567-80.
- 40. Cozzetto D, Kryshchuk A, Ceriani M, Tramontano A.** Assessment of predictions in the model quality assessment category. *Proteins.* 2007;69 Suppl 8:175-83.
- 41. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM.** AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J biomol NMR.* 1996;8(4):477-86.
- 42. Gopalakrishnan K, Sowmiya G, Sheik SS, Sekar K.** Ramachandran plot on the web (2.0). *Protein pept lett.* 2007;14(7):669-71.
- 43. Hazai E, Bikadi Z.** Homology modeling of breast cancer resistance protein (ABCG2). *J struct biol.* 2008;162(1):63-74.
- 44. Eisenberg D, Luthy R, Bowie JU.** VERIFY3D: assessment of protein models with three-dimensional profiles. *Method enzymol.* 1997;277:396-404.
- 45. Kolinski A, Betancourt MR, Kihara D, Rotkiewicz P, Skolnick J.** Generalized comparative modeling (GENECOMP): a combination of sequence comparison, threading, and lattice modeling for protein structure prediction and refinement. *Proteins.* 2001;44(2):133-49.
- 46. Kasrayan A, Bocola M, Sandstrom AG, Laven G, Backvall JE.** Prediction of the *Candida antarctica* lipase A protein structure by comparative modeling and site-directed mutagenesis. *Chembiochem : a European journal of chemical biology.* 2007;8(12):1409-15.

- 47. Overington JP A-LB, Hopkins AL.** How many drug targets are there? *Nat Rev Drug Discov.* 2006;5:993-6.
- 48. Kola I, Landis J.** Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov.* 2004;3(8):711-5.
- 49. Lee CH, Huang HC, Juan HF.** Reviewing ligand-based rational drug design: the search for an ATP synthase inhibitor. *Int j mol sci.* 2011;12(8):5304-18.
- 50. Wang A, Savas U, Hsu MH, Stout CD, Johnson EF.** Crystal structure of human cytochrome P450 2D6 with prinomastat bound. *J biol chem.* 2012;287(14):10834-43.
- 51. Popot JL, Engelman DM.** Membrane protein folding and oligomerization: the two-stage model. *Biochemistry.* 1990;29(17):4031-7.
- 52. Chothia C, Levitt M, Richardson D.** Helix to helix packing in proteins. *J mol biol.* 1981;145(1):215-50.
- 53. Reddy BV, Blundell TL.** Packing of secondary structural elements in proteins. Analysis and prediction of inter-helix distances. *J mol biol.* 1993;233(3):464-79.
- 54. Bowie JU.** Helix packing in membrane proteins. *J mol biol.* 1997;272(5):780-9.
- 55. Bowie JU.** Helix packing angle preferences. *Nat struct biol.* 1997;4(11):915-7.
- 56. Baeza-Delgado C, Marti-Renom MA, Mingarro I.** Structure-based statistical analysis of transmembrane helices. *European biophysics journal : EBJ.* 2013;42(2-3):199-207.
- 57. Dalton JA, Michalopoulos I, Westhead DR.** Calculation of helix packing angles in protein structures. *Bioinformatics.* 2003;19(10):1298-9.
- 58. Bansal M, Kumar S, Velavan R.** HELANAL: a program to characterize helix geometry in proteins. *J biomol struct dyn.* 2000;17(5):811-9.
- 59. Rossum Gv.** Python tutorial, TechnicalReport CS-R9526. 2.7 ed. Amsterdam: Centrum voor Wiskunde en Informatica (CWI); May 1995.

- 60. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI.** OPM: orientations of proteins in membranes database. *Bioinformatics*. 2006;22(5):623-5.
- 61. Team RC.** R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
- 62. Eisenberg D, Schwarz E, Komaromy M, Wall R.** Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J mol biol*. 1984;179(1):125-42.
- 63. Eisenberg D, Weiss RM, Terwilliger TC.** The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature*. 1982;299(5881):371-4.
- 64. Eisenberg D, Weiss RM, Terwilliger TC.** The hydrophobic moment detects periodicity in protein hydrophobicity. *P Natl Acad Sci USA*. 1984;81(1):140-4.
- 65. Fauchère J, and Pliska, V.** Hydrophobic parameters ( $\pi$ ) of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. *EUR J Med Chem*. 1983(8):369-75.
- 66. Cheng Y, Liu A, Guo Y, Qi X.** Computational models for the helix tilt angle. *J chem inf comp sci*. 2004;44(6):1952-60.
- 67. Engelman DM, Steitz TA, Goldman A.** Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu rev biophys bio*. 1986;15:321-53.
- 68. White SH, Ladokhin AS, Jayasinghe S, Hristova K.** How membranes shape protein structure. *J biol chem*. 2001;276(35):32395-8.
- 69. Kim T, Im W.** Revisiting hydrophobic mismatch with free energy simulation studies of transmembrane helix tilt and rotation. *Biophys j*. 2010;99(1):175-83.
- 70. Yeagle PL, Bennett M, Lemaitre V, Watts A.** Transmembrane helices of membrane proteins may flex to satisfy hydrophobic mismatch. *Biochim biophys acta*. 2007;1768(3):530-7.
- 71. Mokrab Y, Stevens TJ, Mizuguchi K.** Lipophobicity and the residue environments of the transmembrane alpha-helical bundle. *Proteins*. 2009;74(1):32-49.

- 72. Johansson AC, Lindahl E.** Position-resolved free energy of solvation for amino acids in lipid membranes from molecular dynamics simulations. *Proteins*. 2008;70(4):1332-44.
- 73. de Jesus AJ, Allen TW.** The determinants of hydrophobic mismatch response for transmembrane helices. *Biochim biophys acta*. 2013;1828(2):851-63.
- 74. Mokrab Y, Stevens TJ, Mizuguchi K.** A structural dissection of amino acid substitutions in helical transmembrane proteins. *Proteins*. 2010;78(14):2895-907.
- 75. Walther D, Eisenhaber F, Argos P.** Principles of helix-helix packing in proteins: the helical lattice superposition model. *J mol biol*. 1996;255(3):536-53.
- 76. Gerstein M, Chothia C.** Perspectives: signal transduction. *Proteins in motion*. *Science*. 1999;285(5434):1682-3.
- 77. Tsai J, Gerstein M.** Calculations of protein volumes: sensitivity analysis and parameter database. *Bioinformatics*. 2002;18(7):985-95.
- 78. Eilers M, Shekar SC, Shieh T, Smith SO, Fleming PJ.** Internal packing of helical membrane proteins. *P Natl Acad Sci USA*. 2000;97(11):5796-801.
- 79. Baldwin JM.** The probable arrangement of the helices in G protein-coupled receptors. *The EMBO journal*. 1993;12(4):1693-703.
- 80. Deber CM, Li SC.** Peptides in membranes: helicity and hydrophobicity. *Biopolymers*. 1995;37(5):295-318.
- 81. Yeates TO, Komiya H, Rees DC, Allen JP, Feher G.** Structure of the reaction center from *Rhodobacter sphaeroides* R-26: membrane-protein interactions. *P Natl Acad Sci USA*. 1987;84(18):6438-42.
- 82. Ben-Tal N, Honig B.** Helix-helix interactions in lipid bilayers. *Biophys j*. 1996;71(6):3046-50.
- 83. Bormann BJ, Engelman DM.** Intramembrane helix-helix association in oligomerization and transmembrane signaling. *Annu rev bioph biom*. 1992;21:223-42.
- 84. Chothia C, Levitt M, Richardson D.** Structure of proteins: packing of alpha-helices and pleated sheets. *P Natl Acad Sci USA*. 1977;74(10):4130-4.

- 85. Esque J, Oguey C, de Brevern AG.** Comparative analysis of threshold and tessellation methods for determining protein contacts. *Journal of chemical information and modeling*. 2011;51(2):493-507.
- 86. Faure G, Bornot A, de Brevern AG.** Analysis of protein contacts into Protein Units. *Biochimie*. 2009;91(7):876-87.
- 87. Guo Z, Kraka E, Cremer D.** Description of local and global shape properties of protein helices. *J mol model*. 2013;19(7):2901-11.

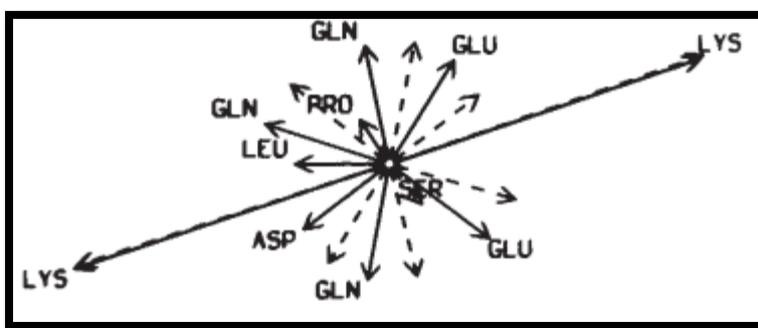
## ANNEXES

**Annexe 1 – Code de l’outil *HeliStatistiX***

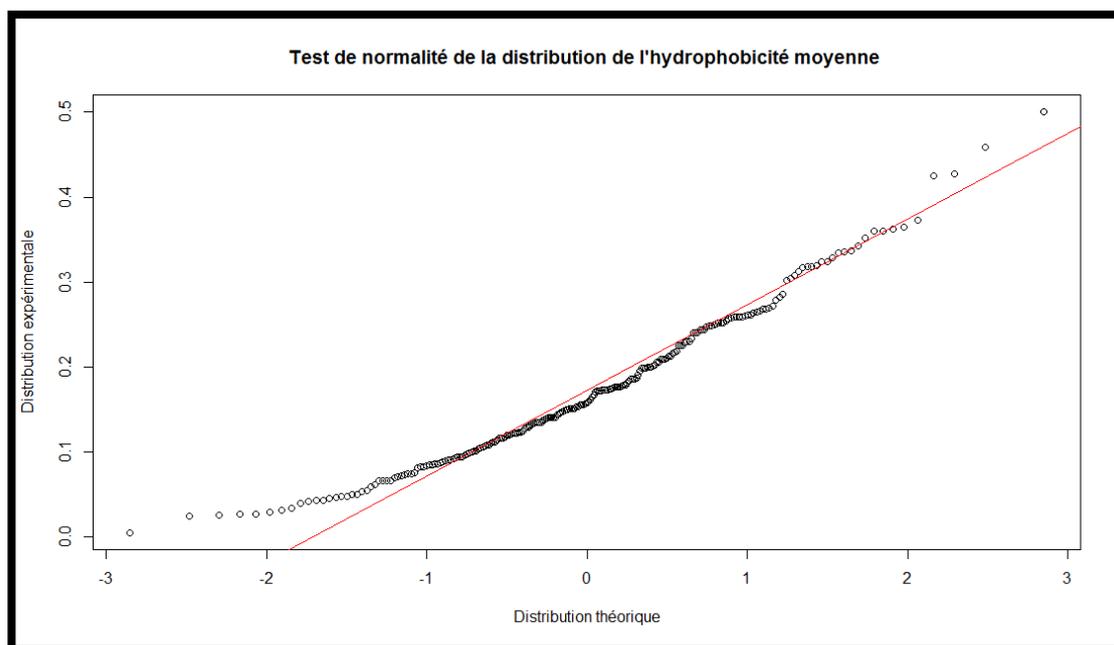
Le code python de l’outil *HeliStatistiX* est disponible en ligne sur le lien suivant :

<https://github.com/marouenbg/HeliStatistiX>

Ainsi que tous les fichiers .pdb et le fichier *config* analysés dans le présent travail.

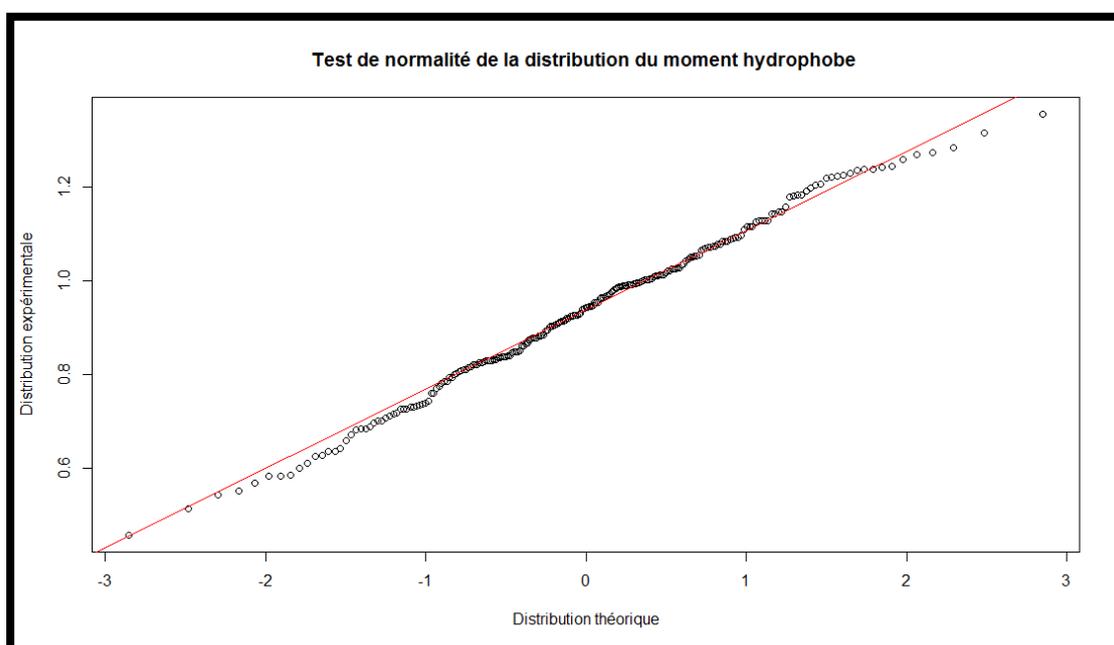
**Annexe 2 – Calcul du moment hydrophobe d’Eisenberg.**

Cette représentation graphique montre la contribution de chaque résidu à la formation du moment hydrophobe de l’hélice (62).



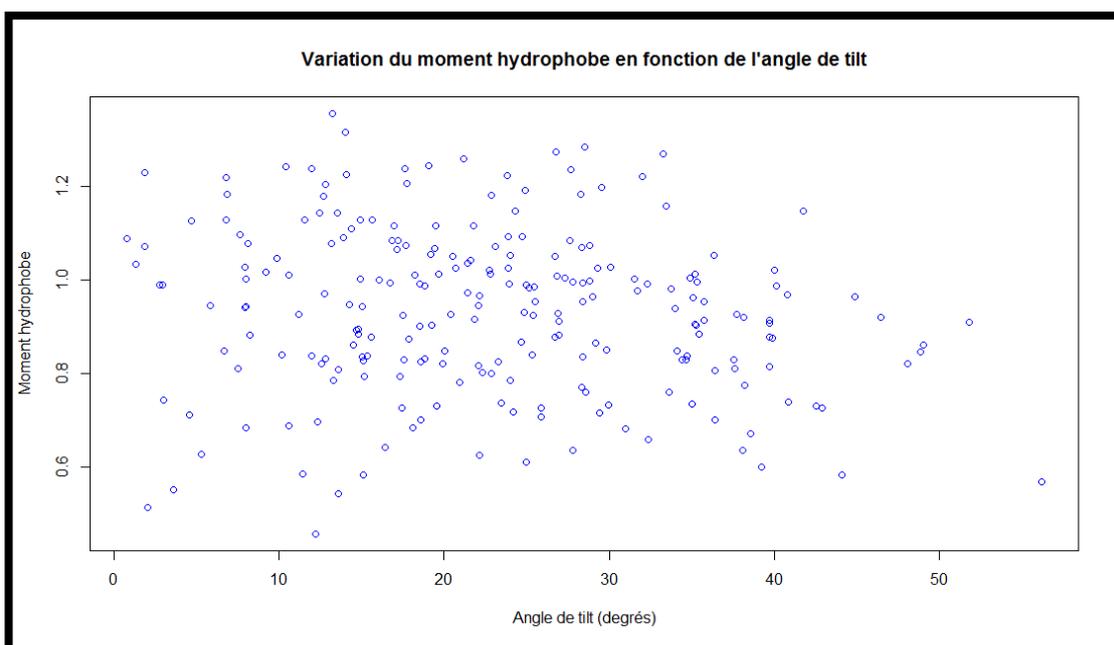
### Annexe 3 - Test de normalité de la distribution de l'hydrophobicité moyenne.

Le schéma montre les valeurs observées de l'hydrophobicité moyenne en fonction des valeurs théoriques normalement distribués. Cette distribution n'est pas normale (test de Shapiro :  $W = 0.96$ ,  $p = 0.00003 < 0.05$ ).



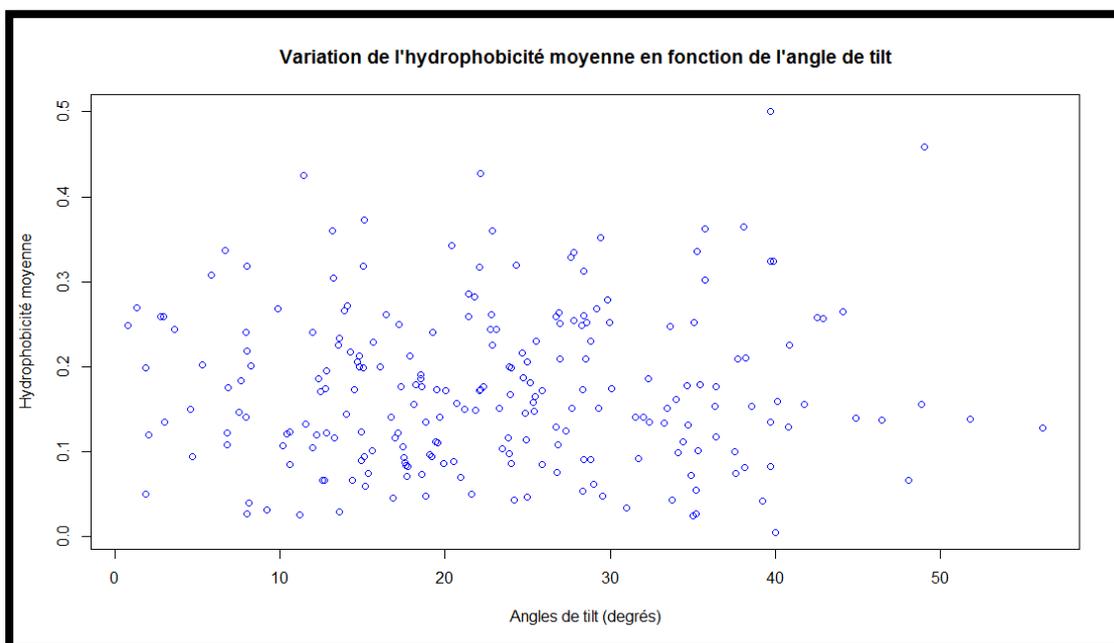
### Annexe 4 – Test de normalité de la distribution du moment hydrophobe.

Le schéma montre les valeurs observées du moment hydrophobe en fonction des valeurs théoriques normalement distribués.

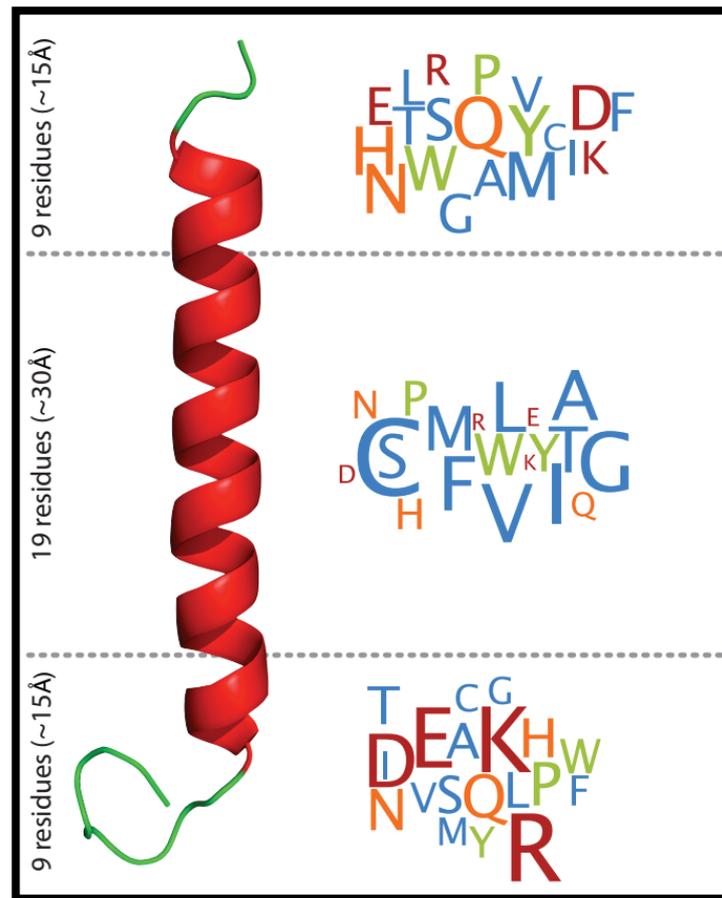


**Annexe 5** – Variation du moment hydrophobe en fonction de l'angle de *tilt*.

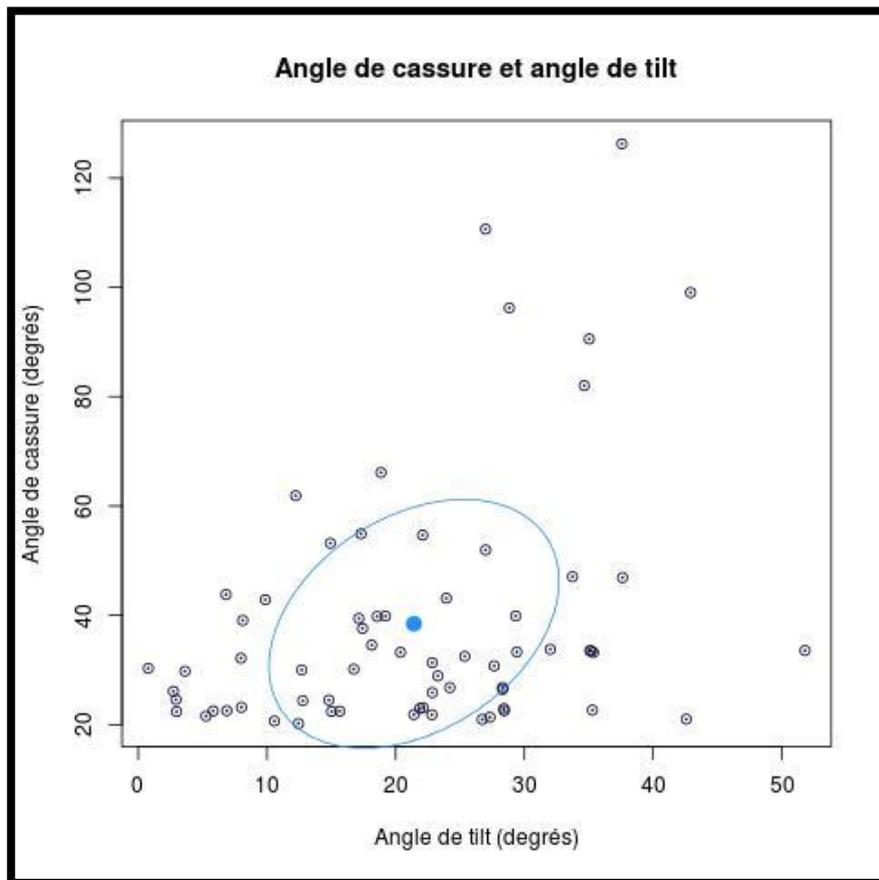
( $r = -0.15$ ,  $p = 0.682$ )



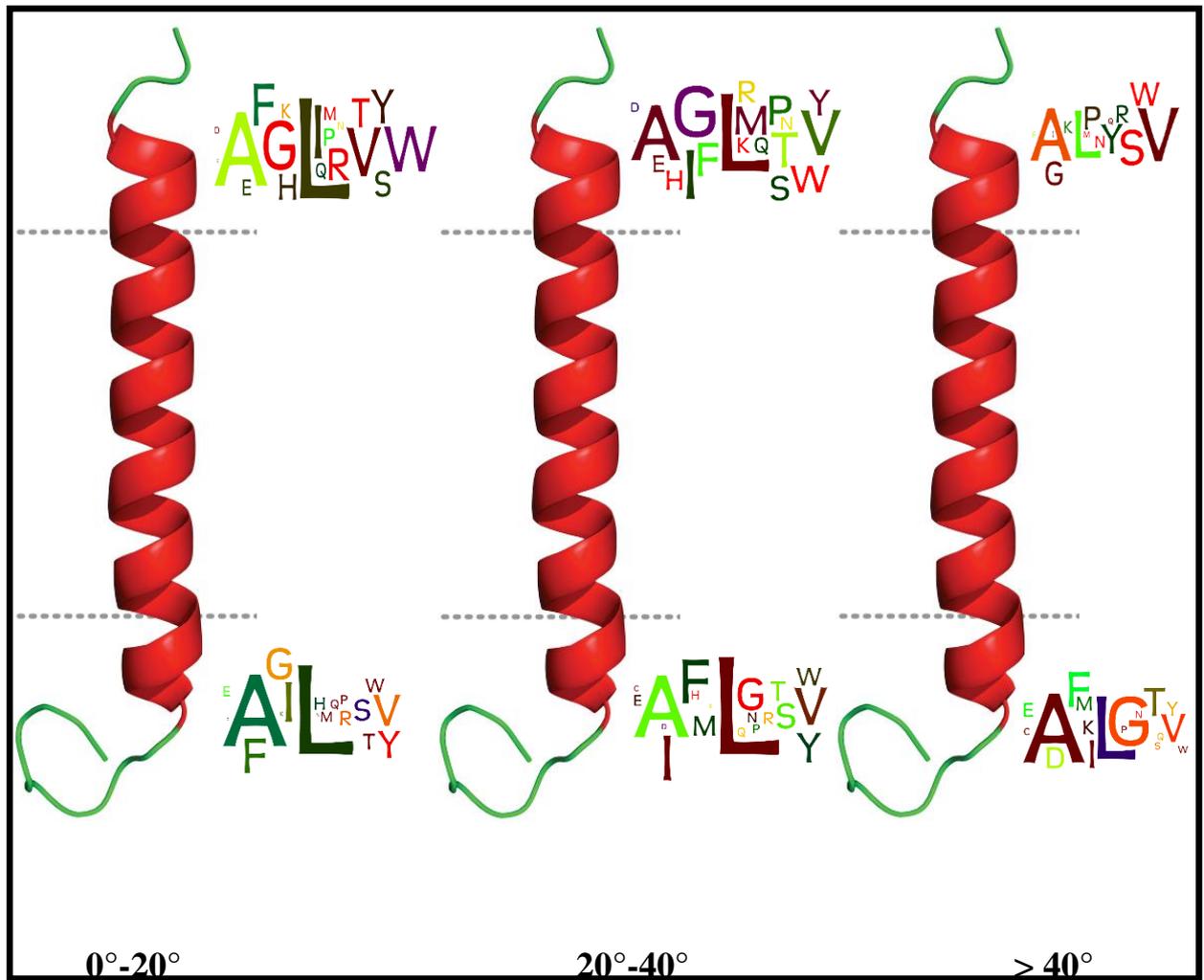
**Annexe 6** – Variation de l'hydrophobicité moyenne en fonction de l'angle de *tilt*. ( $r = 0.0271$ ,  $p = 0.019$ )



**Annexe 7** – Fréquence des acides aminés selon la position dans l'hélice des protéines solubles. Figure extraite d'une étude des protéines solubles (56).



**Annexe 8** – Angle de cassure en fonction de l'angle de *tilt*. ( $p = 0.062$ ,  $r = 0.34$ )



**Annexe 9** - Fréquence des acides aminés au premier tour et au dernier tour d'hélice en fonction de l'angle de *tilt*.