

Institut de Pharmacologie & Toxicologie
Professeur : L. Schild.
Assistant : O. Michiellin

Travail de diplôme

**Etudes structurales et fonctionnelles de la
boucle extracellulaire de la sous-unité alpha du
canal épithélial à sodium (α rENaC)**

Pierre DEVAUD

Le 10 février 2003

TABLE DES MATIERES

1. Introduction	1
1.1. Le canal à sodium épithélial (ENaC)	1
1.1.1. Famille des ENaC/dégénérines	1
1.1.2. Distribution tissulaire	2
1.1.3. Fonction physiologique	2
1.1.3.1. <i>Généralités</i>	2
1.1.3.2. <i>Les reins</i>	3
1.1.4. Régulation de son activité	6
1.1.4.1. <i>Voies hormonales</i>	6
1.1.4.2. <i>Voies non hormonales</i>	7
1.1.5. Propriétés	8
1.1.6. Structure	9
1.1.6.1. <i>Stoechiométrie des sous-unités</i>	9
1.1.6.2. <i>Structure primaire et topologie membranaire</i>	10
1.1.6.3. <i>Pore et filtre de sélectivité</i>	11
1.1.6.4. <i>Site de liaison à l'amiloride</i>	12
1.1.6.5. <i>Modèle structural du pore du canal</i>	13
1.1.6.6. <i>Gating du canal</i>	13
1.2. Structure protéique : hélices α amphipathiques	14
1.2.1. Hélices α : un important élément de structure secondaire	14
1.2.2. Amphipathicité	15
1.3. But du travail de diplôme	18
2. Matériel et méthodes	19
2.1. Bioinformatique	19
2.1.1 Langage de programmation : Perl	19
2.1.2. Programme de prédiction des hélices α amphipathiques	21
2.1.2.1. <i>Input</i>	21
2.1.2.2. <i>Algorithme de prédiction : moment hydrophobe</i>	22
2.1.2.3. <i>Output</i>	25
2.1.3. Programme de définition de substitutions optimales d'acides aminés	25
2.1.3.1. <i>Algorithme de substitutions</i>	25
2.1.3.2. <i>Algorithme d'évaluation des substitutions</i>	26
2.2. Biologie moléculaire	27

2.2.1. Mutagenèse dirigée	27
2.2.2. Sous-clonage	30
2.2.3. Séquençage	31
2.2.4. Transcription in vitro	31
2.2.5 Expression des mutants α F-rENaC dans les ovocytes de <i>Xenopus</i>	31
3. Résultats et discussion	32
3.1. Bioinformatique	32
3.1.1. Test du programme de prédiction d'hélices α amphipathiques	32
3.1.2. Hélices α amphipathiques dans α -rENaC	36
3.1.2.1. <i>Prédiction α-rENaC</i>	36
3.1.2.2. <i>Prédiction sur la famille ENaC/DEG</i>	38
3.1.2.3. <i>Définition de substitutions destructrices optimales</i>	39
3.2. Biologie moléculaire	40
3.2.1. Préparation des mutants	40
3.2.2. Electrophysiologie	41
4. Conclusion et perspectives	43
Remerciements	44
Références	45
Annexes	49

1. Introduction

1.1. Le canal à sodium épithélial (ENaC)

1.1.1. Famille des ENaC/dégénérines

Le canal à sodium épithélial (ENaC) constitue l'un des membres les plus étudiés de la famille ENaC/dégénérines. Découverte au début des années quatre-vingt dix, cette famille de gènes codant pour des canaux sodiques épithéliaux a tout d'abord été mise en évidence par le biais de ses deux premiers membres *deg-1* et *mec-4* (*degenerin (DEG) genes*) lors d'études portant sur la voie mécanosensorielle de *Caenorhabditis elegans*. Très peu de temps après l'identification de la famille de gènes DEG chez *C. elegans*, une stratégie de clonage fonctionnel dans les ovocytes de *Xenopus laevis* permet l'isolation et le séquençage du cDNA codant pour la sous-unité α du canal à sodium épithélial sensible à l'amiloride (EnaC est un complexe hétéromultimérique formé principalement de trois sous-unités: α , β et γ).

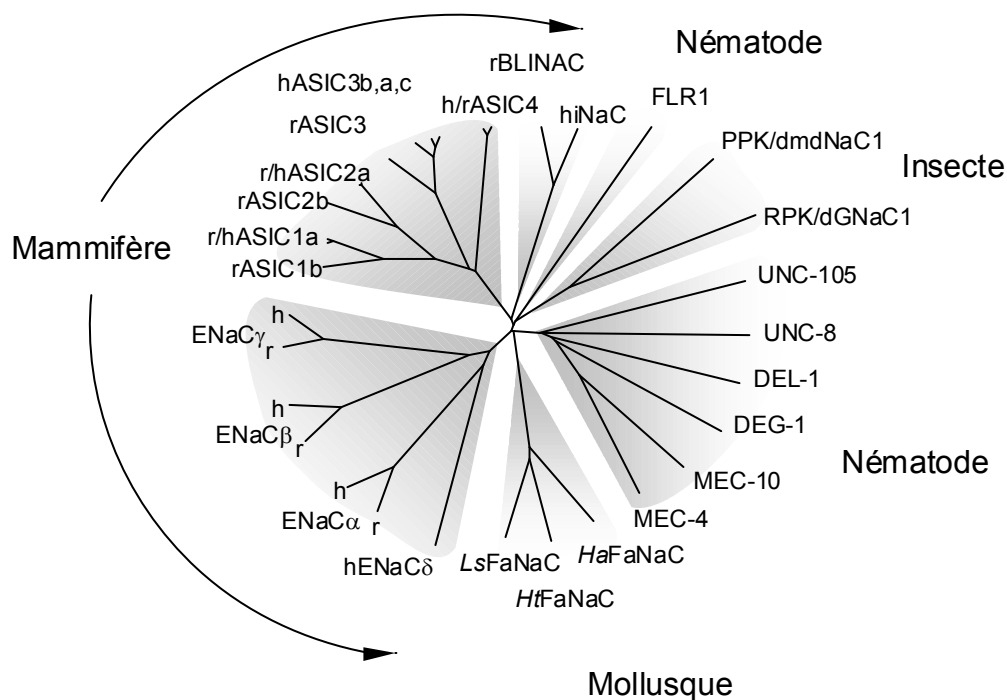


Fig. 1 : Arbre phylogénétique des canaux à sodium épithéliaux de la famille des ENaC/DEG montrant l'organisation en sous-famille de séquences en relation. Les canaux des vertébrés sont divisés en trois groupes : ENaC, *acid-sensing ion channels* (ASICs) et *brain-liver-intestine sodium channel* (BLINaC)/*human intestine sodium channel* (hINaC). Les canaux des invertébrés peuvent être divisés en quatre groupes : 1) les Dégénérines de *Caenorhabditis elegans*; 2) les canaux de la *Drosophila* RPK/dGNaC1 et PPK/dmdNaC1; 3) les *FMRFamide-gated sodium channel* (FaNaC) exprimés dans les mollusques; 4) le canal FLR-1 de *C. elegans* différent des dégénérines. L'identité des séquences entre les différentes sous-familles est de ~ 15 à 20 % alors que l'identité au sein de la sous-famille des sous-unités d'ENaC est de ~ 30 %.

Lors de son clonage, ce canal était déjà connu pour jouer un rôle crucial dans la réabsorption du sodium dans la partie distale du tubule rénal et pour être une cible importante de l'activité de l'aldostérone à ce niveau. Lors d'alignement de séquences, ENaC et les dégénéérines montrèrent une homologie substantielle et de cette relation fut née la famille de gènes que l'on nomma d'après leur nom respectif: ENaC/dégénéérines. Depuis la mise en évidence de cette famille par ces membres fondateurs, beaucoup d'autres gènes issus de vertébrés et d'invertébrés ont grossi ses rangs. A ce jour, plus de trente membres ont été identifiés. Les plus importants sont communément divisés en sept sous-familles selon leur homologie de séquence comme le montre la Figure 1.

1.1.2. Distribution tissulaire

La distribution d'ENaC peut être étudiée à l'aide de tests de *northern blot* ou de *western blot* exécutées sur diverses cellules dont les cellules épithéliales. Pour les épithélia, ces tests peuvent être complétés par différentes expériences électrophysiologiques visant à démontrer la présence d'une conductance sensible au bloqueur spécifique d'ENaC (amiloride). Une des expériences les plus courantes consiste en la mesure d'un courant de court-circuit (I_{sc}) et d'une conductance à travers un épithélium reconstitué en chambre ou un épithélium tubulaire perfusé *in vitro*. Tous ces tests ont permis de mettre en évidence une expression variée d'ENaC. A titre d'exemple, il peut être mentionné que ce canal a entre autres été mis en évidence dans la peau d'amphibiens et de mammifères, dans le colon, dans les reins et conduits urinaires, dans les conduits sudoripares et salivaires ou encore dans les poumons.

1.1.3. Fonction physiologique

1.1.3.1. Généralités

ENaC permet le transport du Na^+ à travers la membrane apicale de certaines cellules épithéliales polarisées. Sa fonction de base dans ces épithélia est de permettre un transport transcellulaire vectoriel de Na^+ . Ce transport transépithélial à travers une cellule implique d'une manière générale trois étapes représentées dans la Figure 2. L'important gradient électrochimique de Na^+ présent au niveau de la membrane apicale permet le passage du Na^+ dans la cellule. Au niveau de la membrane basolatérale, le transport actif du Na^+ par la Na^+/K^+ -ATPase permet le maintien du gradient électrochimique présent au sein de la membrane apicale ainsi que le transport du Na^+ côté basolatéral. Diverses conductances potassiques situées dans les membranes apicales ou basolatérales assurent l'évacuation du K^+ importé dans la cellule par l'activité de la Na^+/K^+ -ATPase.

Par les différents flux ioniques qu'il génère à travers certains épithélia, le transport actif du Na^+ devient capital pour le contrôle de la composition et du volume des fluides présents de chaque côté de ces épithélia. Dans les reins par exemple, l'activité d'ENaC qui représente

l'étape limitante de la réabsorption du Na^+ dans la partie distale du néphron joue un rôle important dans la maintenance de la composition sodique et potassique sanguine et par conséquent dans le contrôle du volume et de la pression sanguine.

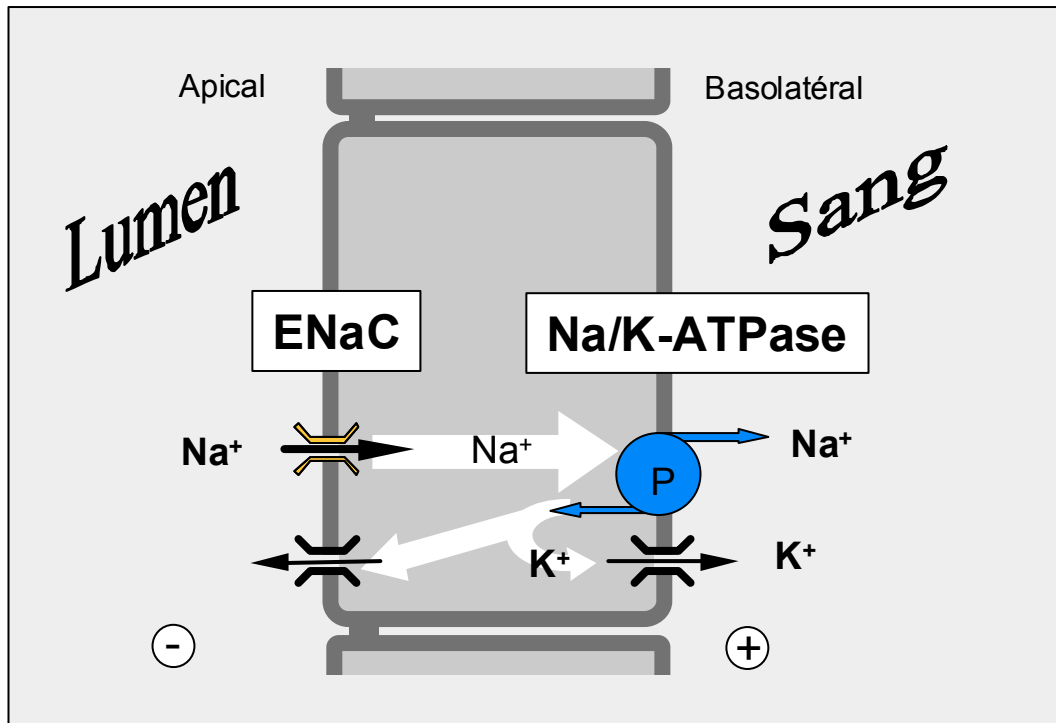


Fig. 2 : Représentation simplifiée du transport transépithélial Na^+ à travers une cellule. ENaC permet l'entrée du Na^+ dans la cellule du côté apicale et la Na^+/K^+ -ATPase le fait ressortir du côté basolatéral. Des canaux K^+ sont présents du côté apical et basolatéral. Du côté apical ils permettent la sécrétion du K^+ dans le lumen tubulaire.

1.1.3.2. Les reins

Bien qu'ENaC possède une distribution tissulaire variée, sa fonction au sein de ces divers tissus reste encore peu connue pour un grand nombre d'entre eux. Une des fonctions les mieux documentées est son rôle dans la réabsorption rénale du sodium donnée ici en exemple.

L'eau et le sodium sont des substances fondamentales pour le bon équilibre de l'organisme. L'eau représente 60% du poids corporel et le sodium est le cation quantitativement le plus important du milieu extracellulaire. Il joue un rôle essentiel dans la régulation du volume extracellulaire, de la pression artérielle et de l'osmorégulation. Le maintien des concentrations sodiques physiologiques est majoritairement assuré par les reins. Le rein peut être séparé anatomiquement en trois parties : une partie corticale (cortex), une partie médullaire (médulla interne et externe). Plus précisément, le rein est constitué de plusieurs unités fonctionnelles appelées les néphrons (un tubule rénal et son glomérule) qui traversent les régions corticale et médullaire. Le rein humain comporte environ un million de néphrons alors que celui de la souris en possède environ 100'000. Les différentes parties du néphron sont décrites dans la Figure 3. Le glomérule est formé par l'invagination de capillaires dans la capsule de Bowman

et c'est à ce niveau qu'a lieu la filtration du plasma. Vient alors le tubule contourné proximal (PCT: *Proximal convoluted tubule*; PST: *Proximal straight tubule*) très perméable à l'eau et à partir duquel 65% des électrolytes filtrés sont réabsorbés. On trouve ensuite l'anse de Henle avec sa partie descendante (DLH: *Descending limb of Henle's loop*), où l'eau est absorbée vers la médulla et où l'osmolarité intratubulaire augmente progressivement, suivie de sa partie ascendante (ALH: *Ascending limb of Henle's loop*) relativement imperméable à l'eau. La région distale du néphron est constituée de la partie médullaire ascendante épaisse de l'anse de Henle (TALH: *Thick ascending limb of Henle's loop*) et du tubule contourné distal (DCT: *Distal convoluted tubule*) eux aussi relativement imperméables à l'eau. Enfin, le dernier segment est appelé tubule collecteur cortical (CCD: *Cortical collecting duct*; OMCD: *Outer medullary collecting duct*; IMCD: *Inner medullary collecting duct*). Il est particulièrement important pour le maintien de la balance hydrique car c'est là que s'établit l'ajustement final de la réabsorption de l'eau et du sodium.

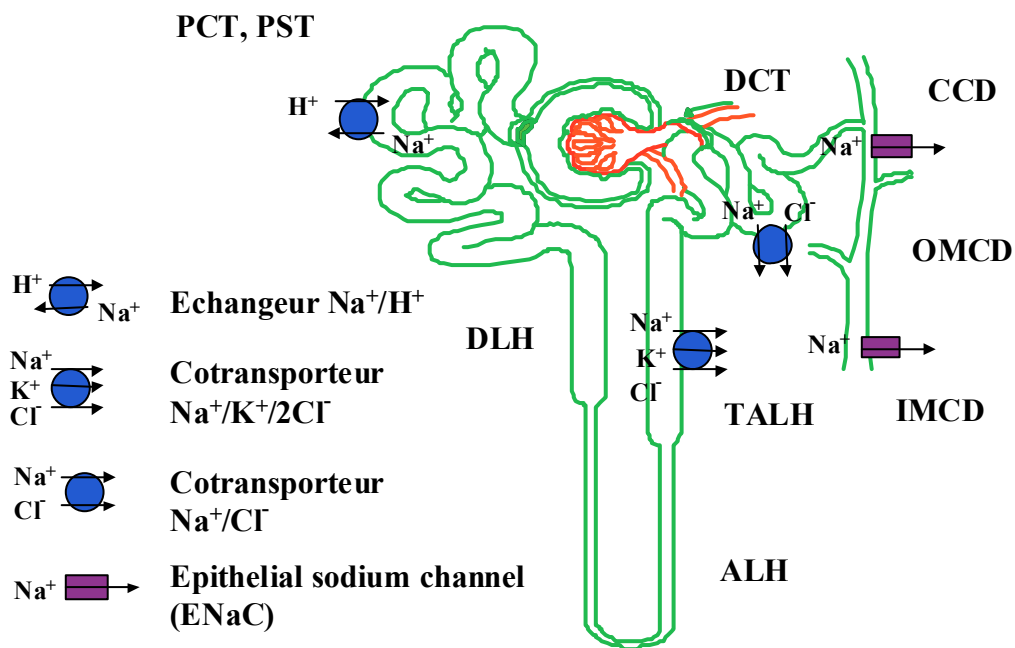


Fig. 3: Représentation schématique du néphron. Les principaux transporteurs impliqués dans la réabsorption du Na^+ le long du néphron sont mentionnés. PCT : *Proximal Convoluted Tubule*. PST : *Proximal straight tubule*. DLH : *Descending limb of Henle's loop*. ALH : *Ascending limb of Henle's loop*. TALH : *Thick ascending limb of Henle's loop*. DCT : *Distal convoluted tubule*. CCD : *Cortical collecting duct*. OMCD : *Outer medullary collecting duct*. IMCD : *Inner medullary collecting duct*.

La réabsorption du Na^+ au niveau des reins est réalisée par différents transporteurs localisés le long du néphron comme le représente la Figure 3. Ces transporteurs spécifiquement localisés

dans certains segments [1] permettent tous l'entrée du Na^+ à travers la membrane apicale des cellules épithéliales. Parmi ces transporteurs, les plus importants sont :

- L'échangeur Na^+/H^+ du tubule proximal associé à de multiples cotransporteurs (PCT, PST).
- Le cotransporteur $\text{Na}^+/\text{K}^+/2\text{Cl}^-$ de la partie ascendante de l'anse de Henle (TALH).
- Le cotransporteur de Na^+/Cl^- du tubule contourné distal (DCT).
- Le canal ENaC du tubule distal et des conduits collecteurs (CCD, OMCD, IMCD).

ENaC est exprimé dans la partie distale du néphron où la réabsorption du Na^+ est contrôlée par l'hormone minéralocorticoïde aldostérone. Le long de la partie distale du néphron, il y a clairement une expression d'ENaC hétérogène. Ainsi, son expression est plus importante dans les régions corticales superficielles comme les tubules collecteurs corticaux (CCD) que dans les régions médullaires plus profondes comme les tubules collecteurs médullaires internes (IMCD). Bien que la majorité du Na^+ réabsorbé dans le rein en conditions physiologiques le soit dans la partie précédent le tubule collecteur cortical (95% du sodium réabsorbé), la réabsorption du Na^+ dans ce tubule est essentielle pour une régulation fine de la réabsorption sodique et pour un contrôle précis de la composition sodique des fluides. L'importance de cette partie finale du néphron dans la réabsorption du Na^+ s'explique par le fait que la réabsorption sodique à ce niveau peut s'accroître de façon considérable en présence d'aldostérone ou de vasopressine dont la sécrétion est stimulée respectivement par une perte de sel et une déshydratation.

Au niveau des reins, les rôles physiologiques et pathophysiologiques d'ENaC dans l'homéostasie sodique et potassique ont clairement été définis par le biais d'études génétiques réalisées sur des patients souffrants d'hypertension (syndrome de Liddle) ou d'hypotension (pseudohypoaldostéronisme de type 1). Le syndrome de Liddle est une forme rare d'hypertension héréditaire caractérisée par des degrés divers d'hypokaliémie associés ou non à une alcalose métabolique et par un faible taux plasmatique d'aldostérone et de rénine. Ces caractéristiques reflètent une réabsorption anormalement élevée du Na^+ au niveau du néphron distal ayant pour conséquence une augmentation marquée du volume du fluide extracellulaire et provoquant le développement précoce d'une hypertension artérielle. En 1994, des études effectuées sur des patients atteints du syndrome de Liddle [2] démontrèrent une relation directe entre l'hypertension observée et un segment du chromosome 16 contenant, entre autres, les gènes codant pour les sous-unités β et γ d'ENaC. La comparaison minutieuse de ces gènes chez des personnes atteintes ou non du syndrome de Liddle permit d'isoler plusieurs mutations dans les sous-unités β et γ associées à une hyperactivation des canaux. Parmi les mutations mises en évidence, une grande partie d'entre elles est responsable de l'altération d'un motif riche en proline (PY) situé en C-terminus. Cette altération a pour

résultat d'augmenter de 3 à 5 fois le courant observé lors d'expériences *in vitro* faites dans des ovocytes de *Xenopus laevis*. Si la mise en évidence des bases génétiques du syndrome de Liddle démontra clairement le rôle crucial d'ENaC dans le maintien du juste équilibre entre la réabsorption et l'excrétion du Na^+ au niveau des reins, le rôle d'ENaC dans l'homéostasie du Na^+ devint encore plus évident lors de l'identification de mutations associées au pseudoaldostéronisme de type 1 (PHA-I) et caractérisées par une réduction ou une perte totale d'activité d'ENaC. Tout comme le syndrome de Liddle, le pseudoaldostéronisme de type 1 est un syndrome héréditaire. Cependant, il est caractérisé par des symptômes opposés au syndrome de Liddle comme par exemple l'hyperkaliémie, l'hyponatriémie, l'hypotension ou encore des taux plasmatiques élevés d'aldostérone et de rénine. Les mutations décelées parmi les patients atteints de cette maladie autosomale récessive touchent l'une ou l'autre des sous-unités d'ENaC et aboutissent à la formation de sous-unités non fonctionnelles par troncations, délétions ou encore substitutions. Bien que les mécanismes par lesquels ces mutations permettent une réduction ou une perte d'activité du canal n'ont pas encore été tous élucidés, ces mutations permettent encore une fois d'établir un lien direct entre une altération d'activité d'ENaC et des symptômes typiques d'une perte d'homéostasie du Na^+ .

1.1.4. Régulation de son activité

L'importance de la balance sodique corporelle et de l'homéostasie sodique intracellulaire chez les vertébrés et invertébrés requiert un contrôle très précis de l'entrée du sodium par les membranes apicales des cellules épithéliales du rein. De par sa localisation dans une région rénale essentielle pour une régulation fine de la réabsorption du Na^+ , ENaC se doit pour remplir sa fonction d'être une protéine dont l'activité est soumise à un contrôle minutieux par certaines voies régulatrices.

Les études de régulation effectuées à ce jour permettent de distinguer communément deux voies régulatrices majeures de l'activité de ce canal: une voie hormonale et une voie non hormonale.

1.1.4.1. Voies hormonales

Au niveau des reins, les régulations hormonales contribuent au maintien de l'homéostasie de l'organisme entier et protègent ce dernier des changements de contenu en sel et des variations de volume.

- *L'aldostérone*: l'aldostérone, un minéralocorticoïde synthétisé dans la partie corticale de la glande surénale, est le stimulus le plus puissant (augmentation d' env. 20 fois le courant non stimulé) pour la réabsorption du sodium dans la partie distale du néphron. Son action est médiée par l'activation d'ENaC et de la Na^+/K^+ -ATPase. Bien que les

mécanismes moléculaires de son action soient encore peu connus, des progrès importants ont été fait ces dernières années. Ainsi, il semblerait qu'après son entrée dans la cellule cible, l'aldostérone se lie au récepteur des minéralocorticoïdes dans le cytoplasme et active ce dernier. Une fois activé, le récepteur aux minéralocorticoïdes lié à l'aldostérone se dirige alors dans le noyau où sa liaison à certaines séquences de DNA (*Hormone Responsive Element*) inhibe ou induit la transcription de gènes (*AIPs: Aldosterone-induced transcripts; ARTs: Aldosterone-repressed transcripts*) [3]. Certains AITs sont directement les sous-unités de la Na^+/K^+ -ATPase ou d'ENaC mais beaucoup d'autres candidats AITs ou ARTs ont été identifiés par la méthode *SAGE* [4]. La présence de transcrits induits différents des sous-unités de la Na^+/K^+ -ATPase ou d'ENaC invite à penser que la notion conventionnelle qui veut que l'aldostérone agisse par l'induction unique de la synthèse de nouvelles sous-unités et par la seule augmentation de Na^+/K^+ -ATPase et d'ENaC en surface cellulaire ne soit pas tout à fait complète. La réponse à l'aldostérone passe très certainement aussi par la modification de Na^+/K^+ -ATPase ou d'ENaC préexistants par certains produits d'AITs ayant comme résultat une augmentation ou une diminution de leur activité.

- *La vasopressine*: une autre hormone capable d'augmenter la perméabilité sodique de la partie distale du néphron est l'hormone peptidique antidiurétique dénommée vasopressine. Nettement moins puissante que l'aldostérone (augmentation d'env. 5 fois le courant non stimulé), cette hormone est capable de produire un effet synergique avec l'aldostérone. Cet effet suggère que son mode d'action n'emprunte pas les mêmes voies régulatrices que l'aldostérone et des études ont en effet démontré que l'activité de la vasopressine serait médiée par une voie spécifique passant par l'activation de l'adénylate cyclase et l'augmentation du cAMP intracellulaire qui s'en suit. D'une façon générale, il est admis que cette augmentation de cAMP a pour double effet d'augmenter le nombre d'ENaC en surface cellulaire et de modifier la probabilité d'ouverture de ces derniers par le biais de phosphorylations directes et indirectes.
- *Autres hormones*: il existe encore d'autres hormones connues pour être capables d'agir sur l'activité d'ENaC. Parmi celles-ci, l'insuline et l'oxytocine.

1.1.4.2. Voies non hormonales

En plus de certaines hormones qui régulent l'activité d'ENaC par le biais de voies régulatrices diverses et encore incomplètement définies, des travaux ont montré qu'ENaC est directement régulé par un certain nombre de facteurs protéiques ou non.

- *Les facteurs non protéiques*: les facteurs non protéiques agissant sur l'activité d'ENaC sont relativement nombreux [5]. Les concentrations sodiques internes et externes, la

concentration cytoplasmique du calcium ou encore le pH intracellulaire peuvent être cités à titre d'exemple mais la liste est loin d'être exhaustive.

- *Les facteurs protéiques*: en accord avec la méthode *SAGE* qui montre qu'un certain nombre de transcrits sont produits en réponse à une stimulation par l'aldostérone, des travaux récents ont démontré que certaines protéines possèdent une action stimulatrice ou inhibitrice sur ENaC. A titre d'exemple, deux protéines connues pour leur effet modulateur sur l'activité d'ENaC peuvent être mentionnées bien que la compréhension des mécanismes d'action de ces protéines régulatrices soit encore incomplète. CAPI1, une sérine protéase, a pour effet d'augmenter le courant sensible à l'amiloride lorsqu'elle est coexprimée avec ENaC dans les ovocytes de *Xenopus*. La modulation de l'activité d'ENaC par cette protéine semble être médiée par une protéolyse directe ou indirecte du domaine extracellulaire. Nedd4, une ubiquitine ligase, agit au contraire en diminuant le courant sensible à l'amiloride lorsqu'elle est coexprimée avec ENaC dans les ovocytes de *Xenopus*. Son activité semble entre autres s'accomplir par l'ubiquitination de plusieurs résidus lysine localisés en N-terminal des sous-unités β et γ . Cette ubiquitination agit comme signal de retrait du canal de la membrane apicale afin qu'il soit dégradé ou recyclé.

1.1.5. Propriétés

Une caractéristique importante et hautement spécifique d'ENaC réside dans le fait que ce canal est bloqué par une faible concentration d'amiloride (K_i amiloride: 0.1 à 1 μ M) en comparaison des autres membres de la famille des ENaC/DEG. Cette propriété spécifique est très largement utilisée pour la mise en évidence de courants sodiques associés à ENaC. Cependant, d'autres propriétés importantes peuvent être mentionnées :

- *Haute sélectivité ionique* : ENaC ne laisse passer que le Na^+ et un petit nombre de cations monovalents de plus petites tailles comme Li^+ et H^+ . Il possède une perméabilité très restreinte pour les cations monovalents de plus grosse taille comme K^+ ($P_{\text{Na}}/P_{\text{K}} > 10$) et pour les cations divalents comme Ca^{++} .
- *Constitutivement actif* : ce canal ne nécessite pas de stimuli (ex. contraintes mécaniques ou ligands externes) pour être actif.
- *Conductance unitaire faible* : à température ambiante, la conductance unitaire d'ENaC exprimée dans les ovocytes de *Xenopus* est relativement petite avec une valeur de 4-5 pS pour une concentration Na^+ de 140 mM et de 9-10 ps en présence de plus de 100mM de Li^+ .
- *Cinétique lente* : le temps d'ouverture moyen d'ENaC est de l'ordre de la seconde (1 à 2 secondes).

- *Voltage indépendant* : aucune dépendance marquée au voltage n'a pu être mise en évidence. Une légère augmentation de la probabilité d'ouverture est toutefois à noter lors d'hyperpolarisations.

1.1.6. Structure

1.1.6.1 Stoechiométrie des sous-unités

ENaC est un complexe hétéromultimérique formé par l'association de trois sous-unités : α , β et γ (dans certains tissus comme par exemple le pancréas ou les ovaires, une sous-unité δ possédant une séquence d'acides aminés proche de celle d' α et capable de former un canal fonctionnel en présence de β et γ a été mise en évidence. Cependant, le rôle physiologique de cette sous-unité dans ces tissus reste encore à découvrir). Ces trois sous-unités d'ENaC sont des glycoprotéines possédant un poids moléculaire de 60 à 75 kDa et partageant une identité en acides aminés relativement importante (~ 30%). Les expériences initiales de clonage démontrèrent que la présence de ces trois sous-unités est requise pour l'obtention d'une activité maximale d'ENaC [3]. Bien que la présence des trois sous-unités α , β et γ soit une condition sine qua non pour un maximum d'activité, ces sous-unités sont, dans une certaine mesure et en accord avec leur homologie relativement importante, interchangeables. Par exemple, l'expression de la sous-unité α seule dans les ovocytes de *Xenopus* permet la génération d'un petit courant sensible à l'amiloride qui suggère que les canaux homomériques α -ENaC sont fonctionnels. De plus, des expériences similaires basées sur l'expression simultanée des sous-unités α - β , α - γ ou β - γ dans les ovocytes ont permis de démontrer que 1) les sous-unités β et γ peuvent partiellement se substituer l'une à l'autre dans le complexe hétéromultimérique [6, 7] et 2) qu'elles sont aptes à former des canaux fonctionnels [8] bien que nettement moins performants que les canaux issus des autres combinaisons testées.

Contrairement à la composition en sous-unités d'ENaC pour laquelle il existe une entente quasi générale de la part des groupes de recherche, le nombre de chacune de ces sous-unités présent dans le canal fonctionnel n'est pas encore un sujet ayant obtenu l'unanimité. Si certains groupes proposent un modèle constitué de quatre sous-unités (2α , 1β et 1γ) arrangés autour d'un pore central, d'autres au contraire prétendent qu'ENaC est un complexe bien plus important formé de neuf sous-unités (3α , 3β et 3γ). Parmi les expériences ayant permis de déterminer une organisation tétramérique des sous-unités d'ENaC, les plus importantes sont les suivantes: 1) détermination de l'abondance relative en surface cellulaire des trois sous-unités d'ENaC par l'analyse quantitative d'anticorps radioactifs liés dirigés contre l'épitope FLAG introduit dans le domaine extracellulaire des sous-unités α , β et γ [9 et 10]. 2) analyse quantitative de la contribution de chaque sous-unité à la fonction du canal par une analyse originellement utilisée par MacKinnon pour calculer la stoechiométrie du canal au potassium *Shaker* et consistant en l'utilisation de mutants possédant par exemple une moindre affinité à l'amiloride que le canal *wild type* [10]. 3) expression de constructions trimériques ou

tétramériques faites de sous-unités α , β et γ liées à la "queue leu leu" dans différentes combinaisons [10]. Bien que de nombreuses expériences tendent à démontrer une composition tétramérique du canal ENaC, il est important de noter que toutes ne permettent pas d'aboutir à la stœchiométrie 2α , 1β et 1γ . Certaines favorisent plus une stœchiométrie 3α , 3β et 3γ . Entre autres évidences, des images de microscopie électronique réalisées sur des cryofracture d'ovocytes exprimant ENaC ont révélé des particules de 8 nm de diamètre compatible avec le canal ENaC [11]. En se basant sur l'hypothèse que l'aire transversale des protéines membranaires dans une microscopie par cryofracture est proportionnelle au nombre de domaines transmembranaires, la microscopie électronique nous révèle la présence de 17 segments transmembranaires chez ENaC correspondant à 8 ou 9 sous-unités. Bien qu'à ce jour aucune des deux stœchiométries proposées ne puisse s'imposer par le biais d'évidences claires, il est important de mentionner que la composition tétramérique des sous-unités est cohérente avec la stœchiométrie d'autres canaux sélectifs Na^+ ou K^+ et en particulier avec la stœchiométrie des canaux composés de sous-unités possédant deux segments transmembranaires (ex. canal au potassium Kcsa).

1.1.6.2. Structure primaire et topologie membranaire

Généralement, les régions de conservation de séquences d'acides aminés au sein d'une famille de protéines décèlent des éléments structuraux importants à la fonction protéique. Dans le cas d'ENaC, certaines séquences sont entièrement conservées chez tous les membres de la famille des ENaC/DEG. Comme le montre la Figure 4, elles impliquent: 1) le motif His-Gly (HG) situé dans la partie N-terminale cytoplasmique proche du segment transmembranaire M1 et impliqué dans le *gating* du canal chez ENaC (c.f 1.1.6.4. *Gating* du canal), 2) la séquence FPxxTxC suivant M1 (post-M1), 3) certains résidus dans la région du segment transmembranaire M2 (pre-M2/M2) comme le motif G/SxS participant au filtre de sélectivité (cf. 1.1.6.3. Pore et filtre de sélectivité) ou le résidu G/S représentant le site de liaison de l'amiloride (cf. 1.1.6.5. Site de liaison à l'amiloride) et 4) des domaines riches en cystéine (CRD II et III) localisés dans la boucle extracellulaire et probablement actif dans le maintien de la structure tertiaire d'ENaC (ponts dissulfures). D'autres séquences ne sont conservées que parmi les sous-unités d'ENaC. C'est le cas du motif riche en proline PPPxYxxL (PY) localisé dans la partie C-terminale cytoplasmique et jouant un rôle dans l'interaction protéique des sous-unités avec des protéines de régulation (ex. Nedd4).

La topologie membranaire des sous-unités d'ENaC avec leurs parties N- et C-terminales cytoplasmiques, leurs deux segments transmembranaires M1 et M2 et la très large boucle extracellulaire (~ 50kDa) qui les relie a été prédite à partir de la séquence primaire d'acides aminés (profile hydrophobe) puis a été confirmée par plusieurs approches différentes (ex. digestion par protéases, sites de glycosylation) [12]. Outre la mise en évidence des deux segments transmembranaires M1 et M2, le profile hydrophobe des sous-unités d'ENaC a

permis de révéler deux courtes séquences d'acides aminés hydrophobes avant chaque segment transmembranaire. En accord avec des expériences de digestion, celle précédant M2 (pre-M2) est accessible aux protéases du côté intracellulaire. Dans tous les canaux sélectifs Na^+ ou K^+ , le court segment hydrophobe situé en amont du dernier domaine transmembranaire côté C-terminal forme le pore externe du canal. Cette analogie suggère que le segment pre-M2 des sous-unités d'ENaC pourrait participer à la formation du pore externe du canal.

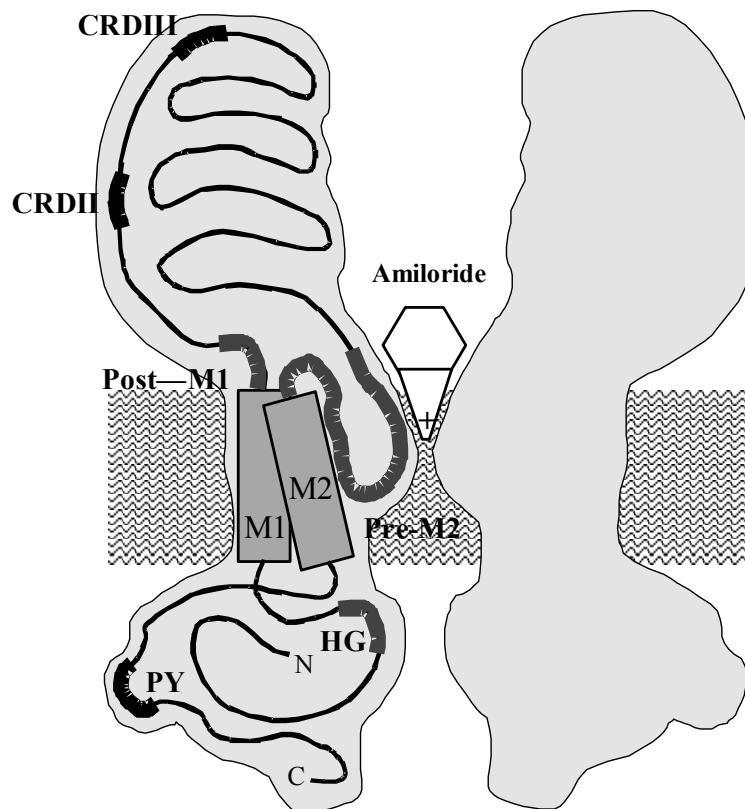


Fig. 4: Représentation schématique de la topologie membranaire d'ENaC et des domaines conservés dont un rôle fonctionnel a été proposé pour certains. Le canal ENaC est probablement formé de 4 sous-unités analogues (deux sont représentées) arrangées de façon symétrique autour d'un pore central. Chaque sous-unité possède deux segments transmembranaires M1 et M2. La boucle pre-M2 (filtre de sélectivité, site de liaison à l'amiloride) longe la partie étroite du pore et du vestibule externe du canal. Le motif PY est impliqué dans les phénomènes de régulation du canal (liaison de Nedd4 et endocytose). Le motif HG participe aux mécanismes de *gating*. Le domaine post-M1 représente la séquence FPxxTxC encore associée à aucune fonction. Les domaines CRDII et III constituent les parties riches en cystéines probablement impliquées dans l'établissement de liaisons dissulfures importantes dans le maintien de la structure tertiaire de chaque sous-unité.

1.1.6.3. Pore et filtre de sélectivité

Bien des travaux ont été réalisés dans le but de déterminer les domaines des sous-unités participant à la formation du pore et du filtre de sélectivité d'ENaC. Plusieurs expériences ont permis de conclure à la participation active du second domaine hydrophobe dans le pore du canal. Cette région est hautement conservée parmi les membres de la famille ENaC/DEG. Elle implique une hélice α typique, assez longue pour traverser la membrane plasmique et

supposée longer le pore du canal (M2), et un segment de résidus hydrophobes moins structuré précédent M2 (pre-M2). Des études récentes de mutagenèse invitent à penser que le motif conservé de trois acides aminés G/SxS du segment pre-M2 des sous-unités α , β et γ participe activement à la formation du filtre de sélectivité [13]. Des mutations du résidu α S589 de ce motif permettent le passage de cations monovalents plus larges (K^+ , Rb^+ , Cs^+ , NH_4^+) et de cations divalents à travers le canal. Ces changements de propriétés résultent d'un agrandissement du pore au niveau du filtre de sélectivité qui, selon la façon dont ENaC distingue les cations, se doit de constituer la partie la plus étroite du pore. De manière surprenante, les mutants révélèrent que plus l'acide aminé introduit en position α 589 avait une chaîne latérale encombrante plus la perméabilité du canal pour les larges cations était importante. Ces résultats suggèrent que la chaîne latérale du résidu α S589 n'est pas dirigée dans la lumière du canal plus vraisemblablement bordée par la chaîne principale des résidus G/SxS. Une telle représentation du filtre de sélectivité d'ENaC bordé par les groupes carbonyles de trois résidus amphiphyles ressemble beaucoup à la structure de base du filtre de sélectivité du KcsA et suggère grandement la présence de mécanismes de sélectivité similaires des cations traversants. La structure cristalline de ce canal montre que l'étroit filtre de sélectivité est le lieu d'interactions fines entre le ion passant et les résidus longeant le pore du canal. Dans les canaux potassiques, les oxygènes des groupes carbonyles de la chaîne principale se substituent aux molécules dH₂O entourant le ion et accélèrent ainsi sa déshydratation dans le pore pour lui permettre le passage.

1.1.6.4. Site de liaison à l'amiloride

Déjà mentionné dans un paragraphe précédent, l'activité d'ENaC est inhibée par la présence côté extracellulaire d'une très faible concentration d'amiloride. Des expériences ont démontré que cet effet inhibiteur de l'amiloride résultait d'une réduction substantielle de la probabilité d'ouverture sans altération de la conductance du canal [14]. Ces observations suggèrent par conséquent que l'amiloride agit comme un bloqueur capable d'obstruer physiquement le pore d'ENaC.

Une série d'expériences visant à muter les acides aminés précédant le segment transmembranaire M2 (pre-M2) des sous-unités d'ENaC a révélé l'existence de résidus (β G525, γ G537 et α S583) impliqués dans l'effet inhibiteur de l'amiloride [15]. Ainsi, lorsque ces acides aminés possédant des positions homologues sont mutés, une diminution de l'affinité apparente du canal pour l'amiloride est observée ($\sim 1000x$ pour β G525 et γ G537 et $\sim 20x$ pour α S583). Ces observations suggèrent que ces résidus sont directement impliqués dans l'interaction du canal avec l'amiloride et qu'ils participent au site de blocage par ce dernier. Partant du principe que l'amiloride agit comme un bloqueur au niveau du pore, ces observations démontrèrent pour la première fois que les segments pre-M2 des différentes

sous-unités d'ENaC participent au pore externe du canal et que les sous-unités s'arrangent bien autour d'un pore central.

Outre les résidus β G525, γ G537 et α S583 du pre-M2, le segment WYRFHY localisé en N-terminal du CRDII de la sous-unité α a été proposé comme jouant un rôle dans l'interaction s'établissant entre ENaC et l'amiloride [16, 17]. Cependant, cette hypothèse est le sujet de vives controverses car elle n'a pas pu être confirmée par de nouvelles expériences.

1.1.6.5 Modèle structural du pore du canal

En se basant sur la structure cristalline du canal Kcsa très proche d'ENaC par sa topologie membranaire et possédant de nombreuses similitudes avec ce dernier et en tenant compte des informations obtenues lors des diverses expériences réalisées pour la définition du site de liaison de l'amiloride, un modèle structural du pore du canal représenté dans la Figure 5 a été élaboré. Cependant, ce modèle n'est encore qu'une simple hypothèse car certains éléments comme la formation du pore par le segment transmembranaire M2 ne sont encore supportés par aucune évidence expérimentale.

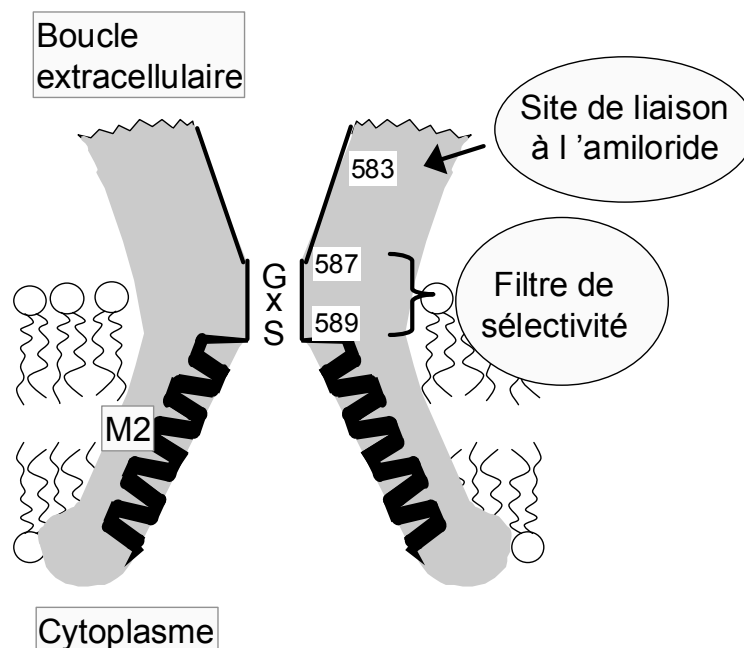


Fig. 5 : Modèle de la région du pore du canal ENaC. Ce modèle d'ENaC est basé sur des données fonctionnelles de canaux mutants et sauvages ainsi que sur une comparaison topologique entre ENaC et KcsA dont la structure cristalline est connue. Deux des quatre sous-unités sont représentées. Les secondes hélices α transmembranaires M2 ("hélices internes" chez Kcsa) et les segments pre-M2 sont indiqués. La numérotation des acides aminés représente celle observée chez la sous-unité α . Contrairement aux sous-unités du Kcsa qui ne contiennent que ~ 30 résidus entre les deux segments transmembranaires M1 et M2, la partie extracellulaire des sous-unités d'ENaC dont l'organisation structurale est inconnue représente $>50\%$ de leur masse et se déploie bien au-delà de la représentation schématique. Dans ce modèle, le vestibule externe est longé par les segments pre-M2 contenant le site de liaison à l'amiloride. Puis le vestibule se rétrécit et les segments pre-M2 forment alors le filtre de sélectivité.

1.1.6.6. *Gating du canal*

Des mutations provoquant une perte de fonction ou une diminution de l'activité d'ENaC ont été identifiées dans les trois sous-unités chez des patients souffrant de pseudohypoaldostéronisme de type 1 (PHA-I). Parmi celles-ci, une mutation dans la sous-unité α substituant une glycine (G95S) du motif HG précédent le premier segment transmembranaire M1 réduit la probabilité d'ouverture (P_o) d'ENaC. Lors d'expériences de substitutions systématiques par l'alanine [18], cet acide aminé (G95) de même que l'histidine qui le précède (H94) et une arginine en aval (R98) ont été associés à une nette diminution de courant Na^+ (I_{Na}) sans altération d'expression de surface et sans modification de la conductance. Ces expériences ont ainsi clairement indiqué que le domaine précédent le segment transmembranaire M1 (pre-M1) contenant H94, G95 (motif HG) et R98 est impliqué dans les mécanismes de modulation de la cinétique d'ouverture d'ENaC (*gating*). Ces mécanismes sont cependant toujours largement inconnus et la façon dont ces résidus contrôlent le *gating* du canal doit encore être élucidée.

1.2. Structure protéique: hélices α amphipathiques

1.2.1. Hélices α : un important élément de structure secondaire

L'hélice α est une structure périodique en bâtonnet formé par le repliement d'une chaîne polypeptidique en une structure répétitive régulière. La chaîne polypeptidique principale étroitement enroulée forme la partie interne du bâtonnet et les chaînes latérales se projettent à l'extérieur en un arrangement hélicoïdal comme le montre la Figure 6. L'hélice α est stabilisée par des liaisons hydrogène entre les groupes NH et CO de la chaîne principale. Le groupe CO de chaque acide aminé est uni par liaison hydrogène au groupe NH de l'acide aminé situé quatre résidus plus avant dans la séquence linéaire. Ainsi, tous les groupes CO et NH de la chaîne principale sont unis par liaison hydrogène. Toutes ces liaisons hydrogène sont dirigées dans la même direction de sorte que les unités peptidiques sont alignés avec une même orientation le long de l'axe de l'hélice. Puisqu'une unité peptidique possède un moment dipolaire dû à la polarité différente des groupements NH et CO, ces moments dipolaires sont également alignés le long de l'axe de l'hélice α . Il en résulte que l'hélice α est globalement un dipôle net avec une charge positive partielle à l'extrémité aminée et une charge négative partielle à l'extrémité C-terminale.

Chaque résidu d'une hélice α est disposé par rapport au suivant selon une translation de 1.5 Å le long de l'axe de l'hélice et une rotation de 100°, ce qui donne 3.6 résidus d'acides aminés par tour d'hélice. Ainsi, les acides aminés distants de trois ou quatre résidus dans la séquence linéaire sont spatialement très proches l'un de l'autre dans une hélice α . En revanche, les acides aminés distants de deux résidus dans la séquence linéaire sont situés sur des côtés

opposés de l'hélice et ne peuvent ainsi pas entrer en contact. Le pas de l'hélice α est de 5.4 Å (1.5 Å x 3.6).

En théorie, le sens d'enroulement d'une hélice peut être droit (sens des aiguilles d'une montre) ou gauche (sens inverse des aiguilles d'une montre). Cependant, les hélices α rencontrées dans les protéines sont presque toujours droites car la formation d'une hélice gauche n'est pas possible avec des acides aminés L étant donné la trop grande proximité des chaînes latérales et du groupement CO.

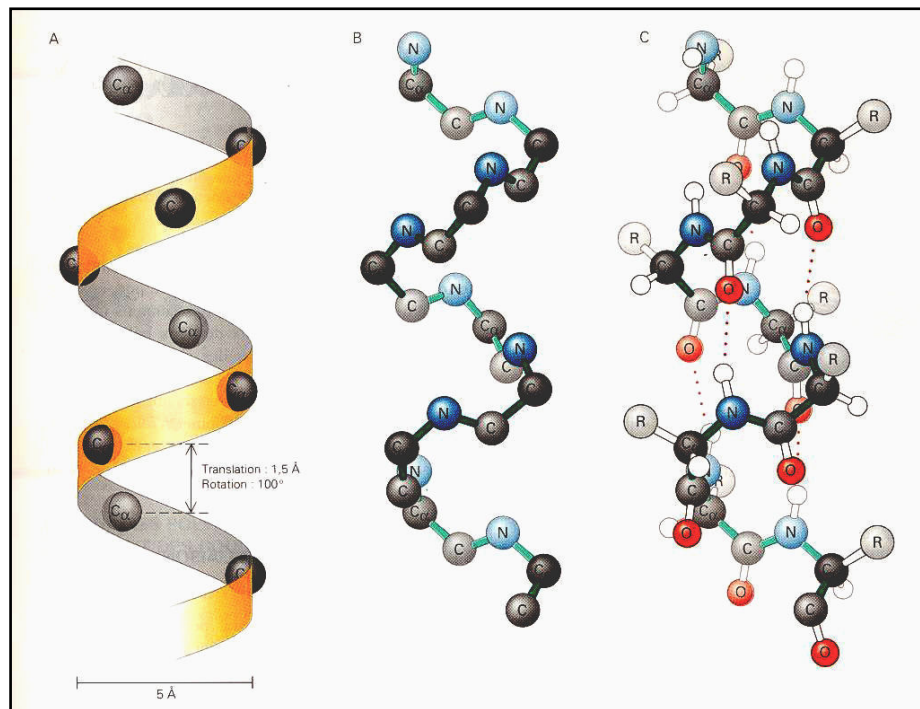


Fig. 6: Modèle d'une hélice α droite. A : seuls les atomes de carbone α sont représentés sur un ruban hélicoïdal. B : seuls les atomes d'azote N, de carbone α C $_{\alpha}$ et de carbone du carbonyle C du squelette sont représentés. C : l'hélice dans sa totalité. Les liaisons hydrogène (figurés par un pointillé) entre les groupes NH et CO stabilisent l'hélice α .

Il existe des différences faibles mais réelles entre les chaînes latérales des acides aminés qui favorisent ou défavorisent leur présence dans les hélices α . Ainsi, selon la table de prédiction d'appartenance à une hélice α de Chou & Fasman, l'alanine (A), l'acide aspartique (D), l'acide glutamique (E), la glutamine (Q), l'histidine (H), l'isoleucine (I), la leucine (L), la lysine (K), la méthionine (M), la phenylalanine (F), le tryptophane (W) et la valine (V) représentent les acides aminés les plus fréquemment observés dans une structure hélicoïdale.

1.2.2. Amphipathicité

Le terme "amphipathique", introduit par Harlley en 1936 [19] et utilisé pour décrire l'organisation particulière de structures protéiques secondaires comme l'hélice α , signifie

littéralement "passion/souffrance" (*pathos*) "des deux côtés" (*amphi*). Cet adjectif est littéralement plus correct que son synonyme "amphiphyle", un terme inventé plus récemment par Tanford [20], car ce dernier suggère que les moitiés hydrophiles et hydrophobes d'une molécule possèdent respectivement de fortes affinités pour les environnements polaires et apolaires. Or, l'effet hydrophobe dépend de façon prédominante des affinités/répulsions (*pathos*) de l'eau pour les molécules polaires et apolaires alors que l'affinité (*philos*) des molécules apolaires entre elles (ex. force de van des Waals) contribue pour peu à l'effet hydrophobe.

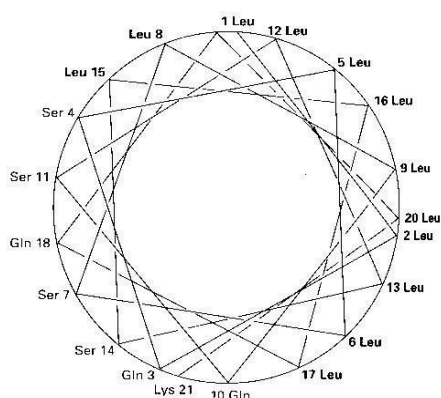
Structurellement parlant, l'amphipathicité se traduit par une ségrégation spatiale des résidus hydrophobes et hydrophiles dans les structures protéiques. Bien que cette ségrégation s'observe dans les deux structures secondaires généralement connues (ex. hélices α , feuilletts β), la conformation hélicoïdale est la structure qui présente le plus fréquemment cette ségrégation de domaines hydrophiles et hydrophobes. Par définition, une hélice α amphipathique est donc une hélice α dans laquelle une des faces longeant l'axe de l'hélice est hydrophile alors que la face opposée est hydrophobe. Cette description qualitative des hélices α amphipathiques définit les caractères essentiels de ce motif conformationnel mais ne représente en aucun cas une définition claire et précise de ce qu'est une hélice α comme pourrait le faire une définition quantitative basée par exemple sur l'étendue de la séparation spatiale des domaines hydrophobes et hydrophiles. Cette absence de définitions structurale ou expérimentale précises représente un des handicaps majeurs dans l'étude fonctionnelle des hélices α amphipathiques car elle permet le regroupement dans la littérature scientifique d'un nombre impressionnant d'hélices α "amphipathiques" aux degrés d'amphipathicité plus que variables résultant de prédiction et d'expériences hautement subjectives.

D'un point de vue fonctionnel, les hélices α correspondant à la définition qualitative de l'amphipathicité furent tout d'abord décrites comme étant un motif à structure/fonction unique impliqué dans l'interaction lipidique [21]. Déjà avant cela, Perutz et al [22] avaient remarqué lors d'études de cristallographie par rayon X que les hélices α de protéines globulaires possédaient fréquemment une face apolaire étroite longeant l'axe central et faisant face à la partie interne apolaire de la protéine. Dans les années huitante, les domaines hélicoïdaux amphipathiques ont été observés dans d'autres protéines capables de s'associer aux lipides, comme certaines hormones polypeptidiques (ex. endorphines), des venins polypeptidiques (ex. bombolitine), des antibiotiques peptidiques (ex. magainines) ou encore certains complexes protéiques transmembranaires (ex. bacteriorhodopsine). De plus, des hélices α amphipathiques impliquées dans des interactions protéines-protéines intra- et intermoléculaires ont été décrites dans un certain nombre de protéines comme les protéines globulaires (ex. hémoglobine) ou encore les protéines contenant des motifs *coiled-coil* (ex. myosin). En bref, les propriétés fonctionnelles suggérées pour les hélices α amphipathiques impliquent l'association lipidique, la perturbation membranaire (ex. fusion ou lyse), la

catalyse de certaines interactions hormone-récepteur, la transduction de certains signaux transmembranaires et le regroupement de segments transmembranaires hélicoïdaux.

Plusieurs méthodes existent pour la détection et la caractérisation de domaines protéiques possédant une structure hélicoïdale amphipathique potentielle. Deux méthodes graphiques, la méthode appelée *helical wheel* de Shiffer-Edmundson [23] et celle dénommée *helical net* de Lim [24], sont des méthodes de choix pour débiter une analyse. La première méthode consiste en la projection des chaînes latérales de chaque résidu sur un plan circulaire perpendiculaire à l'axe central de l'hélice comme le montre la Figure 7A. Chaque résidu représenté sur le cercle forme avec le résidu qui le précède un angle de 100° défini par le nombre théorique de résidus par tour observé dans la structure hélicoïdale. Si le segment ainsi représenté est une hélice amphipathique, les résidus hydrophiles tendront alors à se regrouper sur une face du cercle alors que les résidus hydrophobes se regrouperont sur la face opposée. Basée sur le même principe que la précédente, la seconde méthode est une représentation de la position des chaînes latérales projetées sur une surface cylindrique enroulée autour de l'axe central de l'hélice comme le montre la Figure 7B. Ces deux méthodes peuvent être facilement appliquées et fournissent une interprétation visuelle rapide du motif. Cependant, elles possèdent un certain nombre d'inconvénients. Tout d'abord, si ces deux méthodes sont facilement applicables à de courts segments bien déterminés, elles ne le sont plus lorsqu'il s'agit d'analyser de longues séquences protéiques dans le but d'identifier les séquences possédant un fort potentiel hélicoïdal amphipathique. D'autre part, ces méthodes font appel à un jugement subjectif pour évaluer la crédibilité du motif et délimiter la fin de l'hélice α amphipathique. Enfin, elles ne permettent en aucun cas d'établir une comparaison entre hélices α amphipathiques.

A.



B.

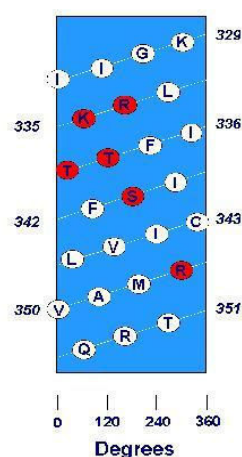


Fig. 7: Représentation des deux méthodes qualitatives d'analyse de l'amphipathicité d'une hélice α . A : *Helical wheel*. B : *Helical net*.

En réponse aux limites observées dans les méthodes graphiques, des méthodes plus quantitatives ont été cherchées. Un excellent exemple de telles méthodes est celui de la méthode du "moment hydrophobe" utilisé dans ce présent travail (cf. 2.1.2.2. Algorithme de prédiction: moment hydrophobe) et basé sur la détection d'une périodicité donnée dans une séquence de nombres représentant les résidus. Cependant, malgré l'apparition de méthodes plus quantitatives capables d'associer un degré d'amphipathicité à une hélice α , il est important de mentionner, en réponse à la remarque établie précédemment dans le présent paragraphe, qu'une définition claire et une classification des hélices α amphipathiques n'a pas encore été établie car chaque méthode nécessite un choix de paramètres qui n'aboutissent pas toujours aux mêmes prédictions.

1.3. But du travail de diplôme

Bien que bon nombre de connaissances concernant la structure/fonction d'ENaC ait été réunies ces dernières années et bien que l'étude de structures connues de canaux présentant des similitudes avec ENaC (ex. KcsA) nous renseignent sur bien des aspects de sa structure, il existe encore aujourd'hui une grande part de mystère entourant l'organisation d'ENaC. Parmi les éléments les moins documentés, la large boucle extracellulaire d'ENaC, caractère unique parmi les canaux ioniques, est certainement un sujet d'importance capitale pour une meilleure compréhension de la structure du vestibule externe du canal. A ce jour, les informations relatives à la structure de cette boucle se résument à des prédictions structurales réalisées par le biais du site Internet de M. Rost & Sander (*Server PHDsec Site: <http://cubic.bioc.columbia.edu/predictprotein/>*) [25, 26]. Ces prédictions, basées sur la séquence primaire des acides aminés (prédiction par un système de réseau de neurones possédant un taux moyen d'exactitude $> 72\%$) ont été initialement réalisées sur la séquence totale de la sous-unité α d'ENaC. Elles mettent en évidence la présence de 5 hélices α à fortes probabilités distinctes des deux segments transmembranaires déjà connus (H1: a.a 69 à 89; H2: a.a 165 à 187; H3: a.a 271 à 289; H4: a.a 440 à 455; H5: a.a 569 à 579). Conformément à la topologie membranaire d'ENaC, caractérisée par la localisation extracellulaire d'une immense majorité des acides aminés, 4 hélices α prédites (H2 à H5) sur 5 se trouvent dans la boucle extracellulaire du canal. Partant du principe que les 4 hélices α localisées dans cette boucle participent activement à la formation du pore externe, il est intéressant de penser que certaines de ces hélices α constituent peut-être l'interface hydrophobe/hydrophile du vestibule externe par le biais d'une organisation amphipathique. Cette hypothèse d'hélices α amphipathiques partiellement insérées dans la membrane plasmique et participant au pore externe d'ENaC est intéressante car elle permettrait éventuellement d'expliquer la présence des 17 segments transmembranaires mis en évidence dans les expériences de cryofracture [11] tout en respectant la stœchiométrie 2α , β et γ soutenue par bon nombre d'expériences.

Tester l'hypothèse d'une éventuelle participation d'hélices α amphipathiques à la formation du pore externe d'ENaC constitue le but de ce travail de diplôme. Pour atteindre ce but, une double approche a été suivie:

- *Bioinformatique*: dans un premier temps, il a été nécessaire de créer de petits programmes informatiques en Perl capables de 1) prédire quantitativement le degré d'amphipathicité d'une hélice α en se basant sur la séquence primaire de cette dernière et 2) déterminer les substitutions d'acides aminés permettant une destruction maximale des amphipathicités observées.
- *Biologie moléculaire et électrophysiologie*: puis, basé sur le degré d'amphipathicité observé pour les hélices α de la sous-unité α d'ENaC et sur la définitions des substitutions d'acides aminés les plus destructrices, des mutagenèses ont été réalisées. Les mutants ainsi obtenus ont alors été exprimés dans les ovocytes de *Xenopus* afin de vérifier si la suppression d'éventuelles hélices α amphipathiques possédait un impact sur le fonctionnement du canal.

2. Matériel et méthodes

2.1. Bioinformatique

2.1.1. Langage de programmation: Perl

Perl est le diminutif de *Practical Extraction and Report Language* et est un langage de programmation créé par Larry Wall dans le milieu des années huitante [27]. Il représente aujourd'hui un langage très populaire utilisé de façon intensive dans les domaines de la bioinformatique ou de la programmation web. En plus d'être un langage de programmation, Perl est aussi une application (souvent référencé comme interpréteur ou compilateur) capable d'être installée sur un ordinateur personnel. Ainsi, le terme Perl fait aussi bien référence au langage dans lequel les programmes sont écrits qu'à l'application présente sur l'ordinateur qui permet à ces programmes de fonctionner. Schématiquement, l'application Perl est capable de lire un programme écrit en Perl, de le traduire en instructions que l'ordinateur peut interpréter puis de l'exécuter.

Les raisons qui font de Perl un langage tant apprécié par les biologistes (d'autres langages de programmation tels que C ou Java sont aussi utilisés en bioinformatique) sont multiples et variées mais à titre d'exemple les raisons suivantes peuvent être mentionnées:

- *Programmation facilitée* : Perl possède un certain nombre de caractéristiques qui simplifient certaines tâches courantes en bioinformatique. Il est par exemple capable de manipuler des informations contenues dans les fichiers textes ASCII (un des modes de sauvegarde et de contrôle des données textuelles utilisés par les ordinateurs) qui représentent exactement le genre de fichiers dans lesquels la plupart des données biologiques importantes se trouvent (GenBank, base de données PDB, ...). Outre cet

avantage incontestable, certaines de ses fonctions permettent également de considérablement simplifier les manipulations de longues séquences comme celles des acides nucléiques ou encore des protéines. Il est aussi important de mentionner que si, de par son architecture, Perl est un outil de choix pour traiter les données biologiques, il n'en est pas moins un langage de programmation très bien adapté à d'autres tâches. Ainsi, Perl rend par exemple très pratique l'écriture de programmes capables de contrôler un ou plusieurs autres programmes. Il peut également être facilement utilisé pour mettre en valeur les sites web des différents laboratoires de recherche en permettant par exemple la création de pages dynamiques. Bien qu'un petit aperçu des différentes possibilités offertes par ce langage de programmation, ces quelques exemples illustrent clairement l'aide précieuse que peut fournir Perl dans un laboratoire de recherche.

- *"Prototypage" rapide* : une autre raison d'utiliser Perl en recherche biologique réside en la rapidité avec laquelle un programmeur peut écrire un programme ("prototypage"). Avec Perl, beaucoup de problèmes peuvent être résolus en un nombre de lignes de code bien inférieur au nombre de lignes de code nécessaire à C ou Java pour accomplir la même tâche. Dans le domaine de la recherche, cette caractéristique est un avantage certain car il est souvent nécessaire d'écrire de petits programmes capables d'accomplir des tâches uniques ou occasionnelles ou qui nécessitent de fréquentes modifications.
- *Portabilité*: la portabilité représente le nombre de systèmes d'ordinateur sur lequel le langage fonctionne. Dans le cas de Perl, il n'y a aucun problème car ce langage est disponible sur pratiquement tous les systèmes modernes. Ainsi, un programme créé sur Mac puis transféré sur PC Windows fonctionnera très bien avec pas ou peu de modifications. Là encore, la portabilité offerte par Perl est un atout indiscutable dans le domaine de la recherche où les échanges entre scientifiques sont monnaie courante.
- *Maintenance des programmes* : la maintenance se résume par l'activité qui assure le bon fonctionnement d'un programme. Elle implique en outre l'ajout de nouvelles fonctionnalités à un programme, son extension permettant la manipulation de nouvelles données ou encore la réparation des *bugs*. Les programmes prennent un certains temps à être écrits mais la maintenance est une partie importante dans l'édition d'un programme car il est courant de devoir le retravailler une fois fonctionnel. Pour cette raison, il est important d'utiliser un langage et un style permettant une maintenance relativement aisée et ceci est facilement réalisable avec Perl (syntaxe facilitant la lecture et la compréhension des programmes par d'autres programmeurs).
- *Apprentissage rapide* : une dernière raison importante qui peut être citée est la rapidité avec laquelle il est possible d'apprendre à écrire un programme fonctionnel en Perl. Ce langage de programmation permet au néophyte de se mettre en route très rapidement sans avoir à maîtriser préalablement une quantité énorme d'informations. Avec un

minimum de règles et de fonctions, il est déjà possible d'écrire des petits programmes très bien adaptés à l'exécution de tâches simples et très utiles en bioinformatique. Cependant, bien que Perl permette une mise en route rapide de son apprentissage, sa maîtrise n'en est pas moins une entreprise de longue haleine qui peut prendre un temps considérable pour être menée à bien.

2.1.2. Programme de prédiction des hélices α amphipathiques

Afin de déterminer l'importance d'hélices α amphipathiques dans l'organisation structurale du pore externe d'ENaC, il a été nécessaire de créer un premier programme en Perl capable de nous renseigner quantitativement sur le potentiel amphipathique des hélices α mises en évidence par le site Internet de prédictions structurales (*Server PHDsec Site*). Ce programme nous indique à partir d'une séquence protéique (*input*) le degré d'amphipathicité (moment hydrophobe) de chaque segment, de taille définie par l'utilisateur, de cette séquence. Les degrés d'amphipathicité ainsi que d'autres informations relatives à chaque segment considéré sont ensuite rendus sous forme de graphique (*output*) pour permettre une meilleure interprétation des résultats.

2.1.2.1. *Input*

Les séquences protéiques (structure primaire) utilisées par le programme pour quantifier le potentiel amphipathique des structures hélicoïdales des protéines correspondantes peuvent se présenter dans différents formats de fichiers:

- *Format Fasta*: ce format est le plus simple que l'on puisse trouver. Il représente simplement la séquence sous forme de lignes de caractères (ex: AFGTR...) codant pour les vingt acides aminés. Il suffit donc au programme créé dans le cadre de ce travail de lire les lignes du fichier pour obtenir la séquence dont il a besoin pour fonctionner. En tête de ces lignes se trouvent généralement des informations relatives à la séquence représentée. Il est important de mentionner pour la suite que ce format ne contient aucune indication concernant la structure à laquelle chaque acide aminé représenté participe.
- *Format Protein Data Bank (PDB)*: la principale source d'information concernant les structures des protéines est la base de donnée PDB (<http://www.rcsb.org/pdb/>). Le format spécifique associé aux fichiers de cette base de donnée est par conséquent un standard pour l'échange d'informations structurales car il permet, par le biais de programmes de visualisation, d'obtenir une représentation 3D des protéines. Les fichiers PDB (chaque fichier représente une structure moléculaire) sont longs et leur format relativement complexe. Sans entrer dans les détails (voir *The Protein Data Bank*

Contents Guide du site pour une description complète du format), le format est caractérisé par une suite d'informations structurées par thèmes. Si une partie du fichier est constitué des thèmes portant sur les aspects généraux de la molécule considérée (nom, source, auteur, publications, séquence primaire, ...), son immense majorité est constitué d'un thème concernant la position de chaque atome dans la molécule. Contrairement au format Fasta, ce format contient bien évidemment l'information nécessaire à la détermination de la structure secondaire à laquelle participe chaque acide aminé de la séquence. Bon nombre d'applications existantes sont ainsi capables d'extraire la structure protéique primaire et secondaire de ces fichiers. C'est par le biais de ces applications que le programme créé dans ce travail obtient la séquence et la structure secondaire de la protéine associée au fichier PDB.

- *Format Server PHDsec Site (PHD)*: faisant suite à la soumission *online* d'une séquence protéique au site Internet de prédiction structurale, un fichier HTML contenant une foule d'informations relatives à cette séquence est retourné. Tout comme les fichiers PDB, le fichier HTML est caractérisé par une organisation thématique. Parmi les thèmes présents dans le fichier figurent celui concernant la prédiction des structures secondaires (*PROF predictions*). Au sein de ce thème, la séquence protéique et la structure secondaire associée à cette séquence sont représentées sous forme de lignes de caractères superposées. Il suffit alors au programme de prédiction amphipathique de lire ces lignes pour obtenir la séquence et la structure secondaire dont il a besoin pour fonctionner. Encore une fois, ce format comporte une information structurale associée à la séquence primaire et se distingue par conséquent du format Fasta.

2.1.2.2. *Algorithme de prédiction: moment hydrophobe*

Une fois la structure protéique primaire introduite dans l'ordinateur sous forme des différents formats susmentionnés, le programme exécute à partir de cette séquence une suite de processus visant à établir le profil amphipathique de la protéine considérée:

1. Lecture de la séquence protéique et génération de segments (fenêtres de calcul) de taille définie par l'utilisateur.

ex: taille 11

YVFGTRNNLPOVFET... → séquence.

YVFGTRNNLPOVFET... → fenêtre 1.

YVFGTRNNLPOVFET... → fenêtre 2.

YVFGTRNNLPOVFET... → fenêtre 3.

etc.

2. Calcul, pour chaque fenêtre, du moment hydrophobe et de l'hydrophobicité moyenne.

Le moment hydrophobe est une interprétation quantitative de la méthode graphique *Helical wheel* (c.f 1.2.2. Amphipathicité) introduite par Eisenberg et al. [28]. Comme le montre la Figure 8, dans cette méthode un vecteur de taille proportionnelle à l'hydrophobicité de l'acide aminé considéré est représenté partant de l'axe central de l'hélice α et passant par la position du C_α de l'acide aminé. Le moment hydrophobe est ensuite défini comme étant le produit scalaire de la somme vectorielle des tous les acides aminés ainsi représentés participant à la fenêtre analysée. Il peut donc être défini de la manière suivante:

$$|\vec{\mu}_h| = \sum H_i, \quad i = 1 \text{ à } N$$

De façon à pouvoir comparer les moments hydrophobes de fenêtre de différentes longueurs (N), Eisenberg et al. ont également défini le moment hydrophobe moyen:

$$\langle \mu_h \rangle = |\vec{\mu}_h| / N$$

D'un point de vue mathématique, la somme des représentations vectorielles des acides aminés décrites précédemment se traduit par la formule suivante:

$$\mu(100^\circ) = \left\{ \left[\sum_{K=1}^N H_K \sin(K100^\circ) \right]^2 + \left[\sum_{K=1}^N H_K \cos(K100^\circ) \right]^2 \right\}^{1/2}$$

Où $\mu(100^\circ)$ représente la somme vectorielle (moment hydrophobe $|\mu_h|$), H_K l'hydrophobicité du $K^{\text{ième}}$ résidu de la fenêtre et 100° la périodicité angulaire typique observée dans une hélice α de l'orientation des chaînes latérales des résidus.

Pratiquement, le programme applique à chaque fenêtre qu'il génère la formule mathématique démontrée ci-dessus puis divise le résultat par le nombre de résidus de la fenêtre pour obtenir le moment hydrophobe moyen ($\langle \mu_h \rangle$). Pour ce faire, il nécessite l'hydrophobicité de chaque résidu participant à la fenêtre. L'échelle d'hydrophobicité utilisée par le programme est une échelle consensus normalisée réalisée par Eisenberg et al. (cf. Annexe D) [29]. Il est à noter également que la périodicité angulaire d'une hélice α est rarement exactement de 100° et qu'elle fluctue fréquemment entre 95° et 105° . Pour tenir compte de cette variation, le programme a été conçu de manière à tester toutes les périodicités angulaires situées entre 95° et 105° . N'est alors conservé, pour chaque fenêtre, que la périodicité angulaire associée au moment hydrophobe le plus élevé.

Il est important de mentionner ici que le calcul du moment hydrophobe tel qu'il vient d'être défini n'a de sens que si la séquence de la fenêtre analysée participe à une structure secondaire hélicoïdale. Bien entendu, ceci est loin d'être le cas pour la majorité des fenêtres et bon nombre d'entre elles sont par conséquent faussement modélisées en hélice α par le programme. Sans les informations structurales propres à chaque fenêtre, les moments hydrophobes qui leur sont associés sont par conséquent de peu d'intérêt. Ce problème est graphiquement résolu (cf. 2.1.2.3. *Output*) pour les formats de séquence comportant une information structurale associée à la structure primaire de la protéine (format PDB et PHD) mais demeure pour les formats ne possédant pas cette information (format Fasta).

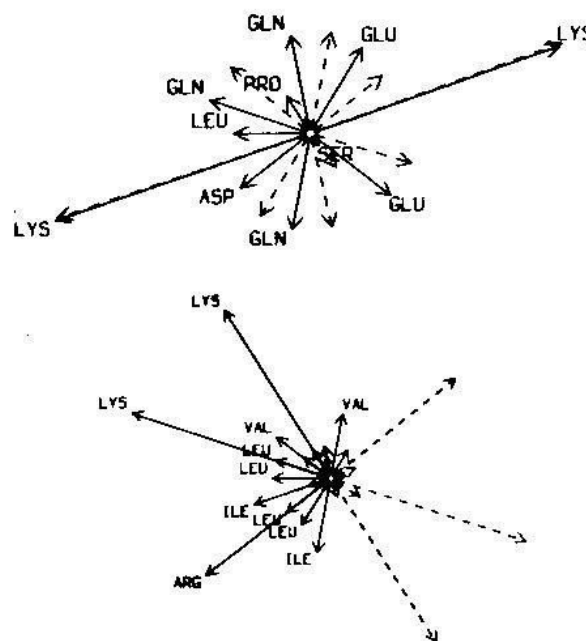


Fig. 8 : Représentation graphique de la contribution, par résidu, au moment hydrophobe ($|\mu_h|$). Les résidus hydrophobes possèdent des valeurs positives pour leur hydrophobicité et sont représentés par des vecteurs en trait plein s'éloignant de l'axe central de l'hélice α . Les résidus hydrophiles ont des valeurs négatives et leur position par rapport à l'axe central de l'hélice α est représentée par des vecteurs en pointillés s'éloignant du centre. La contribution vectorielle de chaque résidu hydrophile au moment hydrophobe est représentée par des vecteurs en trait continu tourné de 180° .

Outre le moment hydrophobe moyen, le programme calcule pour chaque fenêtre l'hydrophobicité moyenne. Cette valeur est calculée à partir de la même échelle d'hydrophobicité utilisée pour le calcul du moment hydrophobe. Elle représente simplement la somme des hydrophobicités présentes dans la fenêtre divisée par le nombre de résidus de cette dernière. Cette mesure associée à celle du moment hydrophobe permet l'identification de segments transmembranaires potentiellement amphipathiques.

3. Création d'un fichier contenant les valeurs calculées associées à l'acide aminé occupant la place centrale de la fenêtre.

ex: taille 11

YVFGTRNNLPOVFET... → séquence.

YVFGTRNNLPOVFET... → valeurs fenêtre 1 associées à R.

YVFGTRNNLPOVFET... → valeurs fenêtre 2 associées à N.

YVFGTRNNLPOVFET... → valeurs fenêtre 3 associées à N.

etc.

Selon le type de format utilisé en *input* (format de séquence PDB ou PHD), le fichier contient également pour chaque acide aminé l'information relative à sa participation ou non participation à une structure secondaire hélicoïdale. Cette information est codée sous la forme d'une valeur qui peut ainsi facilement être représentée graphiquement lors de l'*output* au côté des autres valeurs précédemment décrites.

2.1.2.3. *Output*

A partir du fichier créé contenant les informations susmentionnées, le programme contrôle l'exécution d'une application (Xmgr) capable de représenter ces informations (cf. Annexe A). Le graphique (*output*) ainsi construit représente la position de chaque acide aminé dans la séquence (abscisse) et les valeurs qui leur sont associées (ordonnée). De cette manière, les résultats sont clairement représentés et il est relativement facile par le biais de la position des acides aminés centraux de distinguer les régions possédant à la fois une structure hélicoïdale et un degré d'amphipathicité élevé.

2.1.3. Programme de définition de substitutions optimales d'acides aminés

Dans le but de définir les substitutions les plus aptes à détruire l'amphipathicité des hélices α présentes dans la sous-unité α d'ENaC et mises en évidence par le premier programme informatique, un deuxième programme informatique en Perl a été créé. Ce nouveau programme reprend dans les grandes lignes le même corps central que le programme précédemment décrit. Il possède cependant deux nouveaux modules qui s'y intègrent de façon à lui offrir de nouvelles fonctionnalités.

2.1.3.1 *Algorithme de substitutions*

Fonctionnant en amont de l'algorithme de prédiction, ce module a pour effet de créer une série de double substitutions dans une région hélicoïdale de la séquence protéique délimitée par l'utilisateur. Le choix d'une double substitution s'est imposé suite aux observations faites de l'incapacité pour une mutation simple de détruire massivement l'amphipathicité des hélices α .

L'insertion et la délétion de deux acides aminés respectivement au centre et en C-terminal de l'hélice α amphipathique était une solution également intéressante autorisant la rotation de 180° d'une moitié d'hélice α . Cependant, cette solution n'a pas été retenue car elle représentait des manipulations plus difficilement réalisables en laboratoire. De façon exhaustive, le module créé va ainsi générer toutes les combinaisons de substitutions doubles possibles.

ex: substitutions doubles exhaustives de 2 a.a (A et C)

... YVFGTRNNLPOVFET... → région "sauvage" de 15 a.a délimitée par l'utilisateur.
... AAFGTRNNLPOVFET... → région "mutante" n° 1.
... ACFGTRNNLPOVFET... → " " n° 2.
... AVAGTRNNLPOVFET... → " " n° 3.
... AVCGTRNNLPOVFET... → " " n° 4.
...
... AVFGTRNNLPOVFEA... → " " n° 27.
... AVFGTRNNLPOVFEC... → " " n° 28.
... YAAGTRNNLPOVFET... → " " n° 29.
... YACGTRNNLPOVFET... → " " n° 30.
... YAFATRNNLPOVFET... → " " n° 31.
... YAFCTRNNLPOVFET... → " " n° 32.
...
... YVFGTRNNLPOVFAA... → " " n° x.
... YVFGTRNNLPOVFAC... → " " n° x+1.

Afin de ne pas détruire la structure hélicoïdale de la région substituée, la liste des acides aminés utilisés pour effectuer les substitutions a été choisie en fonction de la capacité de chaque acide aminé à participer à la formation d'une hélice α . Cette liste élaborée à partir des données de Chou & Fasman contient les acides aminés suivants: l'alanine (A), l'acide aspartique (D), l'acide glutamique (E), la glutamine (Q), l'histidine (H), l'isoleucine (I), la leucine (L), la lysine (K), la méthionine (M), la phenylalanine (F), le tryptophane (W) et la valine (V).

Chaque séquence "mutante" ainsi générée par ce module est ensuite testée par le programme de prédiction d'hélices α amphipathiques représentant le corps central de ce deuxième programme informatique.

2.1.3.2. *Algorithme d'évaluation des substitutions*

Contrairement au module précédent, ce module fonctionne cette fois en aval de l'algorithme de prédiction. Une fois le profil amphipathique de la séquence mutée accompli par le corps central du programme, ce deuxième module détermine l'impact de la substitution générée sur le degré d'amphipathicité de la région délimitée par l'utilisateur. Pour ce faire, il mesure l'aire sous la courbe du moment hydrophobe pour la région substituée comme le montre la Figure 9. La substitution la plus dévastatrice est ensuite retenue par ce module qui sélectionne la séquence mutante possédant l'aire sous la courbe la plus faible.

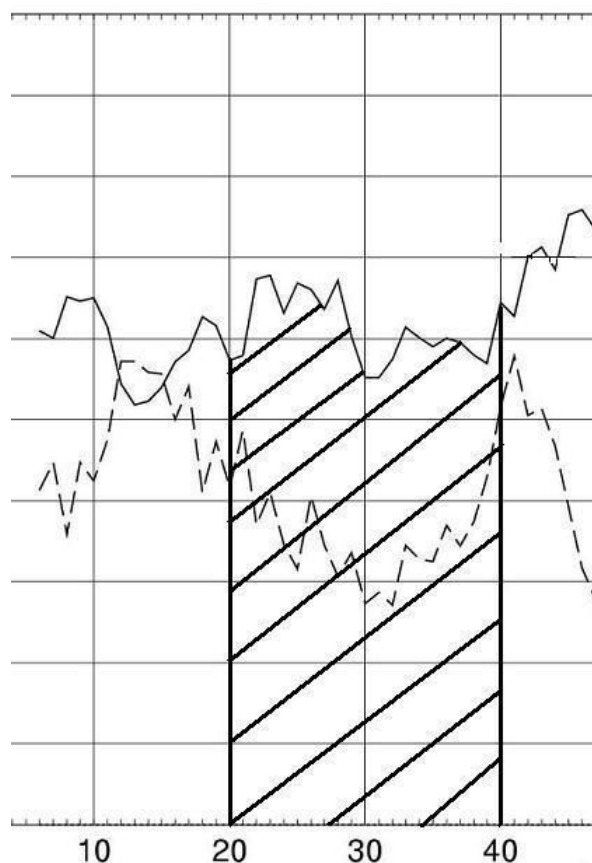


Fig.9 : Représentation du calcul de l'aire sous la courbe du moment hydrophobe effectué par le programme de définition de substitutions optimales. Profil amphipathique de la sous-unité α d'ENaC (fenêtre 11) montrant le moment hydrophobe moyen (trait continu), l'hydrophobicité moyenne (trait en interrompu), la position des résidu et l'aire sous la courbe (zone hachurée).

2.2. Biologie moléculaire

2.2.1. Mutagenèse dirigée

Pour réaliser les divers mutants portant les substitutions optimales mises en évidence par les programmes informatiques et permettant de détruire l'amphipathicité des trois hélices visées (H1, H2 et H3), la méthode de mutagenèse dirigée selon Nelson et Long a été utilisée [30]. Cette technique, dont le principe est illustré par la Figure 10, utilise la réaction de polymérisation en chaîne (PCR) [31] et nécessite par conséquent l'utilisation de *primers*. Les séquences des *primers* utilisés dans ce travail sont représentées dans le Tableau 1. Afin de faciliter les étapes visant à vérifier la présence de l'insert muté dans l'ADN plasmidique obtenu lors du sous-clonage, des mutations ayant pour effet de créer ou de détruire un site de restriction ont été incorporées dans les *primers* porteurs des mutations relatives aux substitutions optimales. Dans le cadre de ce travail, la matrice de base des deux premières PCR a toujours été le DNA linéarisé du vecteur de transcription pSD5 (cf. Annexe B) contenant la sous-unité α F-rENaC.

Hélice concernée	Nom du primer	Séquence 5' -> 3'
H1	Muté α h7087	ATT GAA TTC CAC <u>ATC</u> TCC TAC CGG GAG <u>AAG</u> TTC CA
H2	Muté α h168186	ATT AAA GAG GAG CTC <u>ATA</u> GAG CTG GAC <u>ATC</u> ATC AC
H3	Muté α h274287	GAT GCA <u>AAG</u> ATT GAG TGG TAT <u>CGC</u> TTC CAT TAC AT
H1	Hybride α 800::VSVG	G ATA CTG GGC TAT CCC CTT CCG GTC CAC TTG GGG
H2 et H3	Hybride α 1100::VSVG	GA TAC TGG GCT ATC CGG GAT GAA GTC ATT CTG CTC TGT G
H1, H2 et H3	SP6	ATT TAG GTG ACA CTA TAG
H1, H2 et H3	VSVG	G ATA CTG GGC TAT CC

Tab. 1: Séquences des *primers* utilisés lors de la mutagenèse dirigée. Les bases soulignées indiquent les mutations introduites.

Les différentes PCR nécessaires à l'introduction des mutations ont été effectuées à l'aide du tampon PCR et de la *Taq* polymérase provenant de la firme Amersham Pharmacia Biotech et en utilisant un thermocycler Perkin Elmer Cetus avec les protocoles suivants:

Etape 1: elle est effectuée dans un volume de 50 μ l (H₂O: 37.5 μ l, tampon PCR 10x: 5 μ l, dNTP 1.25mM: 4 μ l, *primer* muté: 1 μ l, *primer* hybride: 1 μ l, pSD5/ α F-rENaC linéarisé: 1 μ l, *Taq* polymérase 5 u/ μ l: 0.5 μ l).

Programme:	dénaturation initiale	94°C	1 minute
	dénaturation	94°C	1 minute
	hybridation	55°C	1 minute
	élongation	72°C	1 minute 30 seconde

Etape 2: elle est effectuée dans un volume de 97.5 μ l (produit étape 1: 2 μ l, H₂O: 76 μ l, tampon PCR 10x: 10 μ l, dNTP 1.25mM: 8 μ l, pSD5/ α F-rENaC linéarisé: 1 μ l, *Taq* polymérase 5 u/ μ l: 0.5 μ l).

Programme:	dénaturation initiale	94°C	1 minute
	dénaturation	94°C	1 minute
	hybridation	63°C	1 minute
	élongation	72°C	12 minutes

Etape 3: elle est effectuée dans un volume de 100 μ l (produit étape 2: 97.5 μ l, *primer* SP6: 1 μ l, *primer* VSVG: 1 μ l, *Taq* polymérase 5 u/ μ l: 0.5 μ l).

Programme:	dénaturation initiale	94°C	1 minute
	dénaturation	94°C	1 minute
	hybridation	55°C	1 minute
	élongation	72°C	1 minute 30 seconde

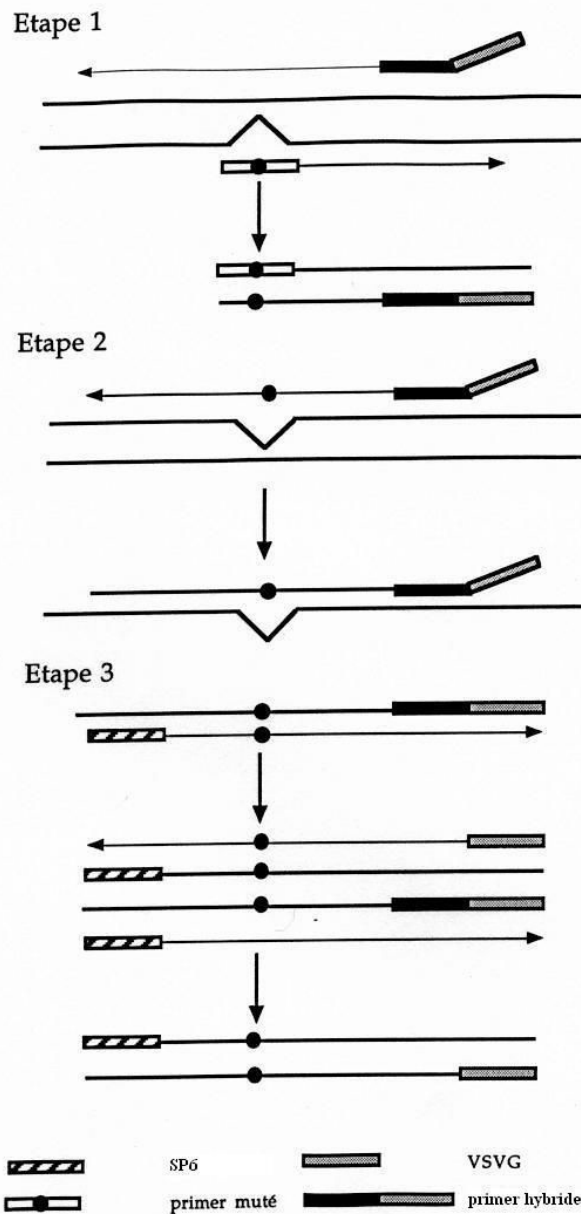


Fig. 10 : Mutagenèse dirigée selon Nelson et Long (1989). 1) première PCR réalisée avec un *primer* hybride (α 1100::VSVG) et un *primer* porteur de la mutation à introduire (α h7087, α h168186 ou α h274287). Le fragment ainsi amplifié porte la mutation désirée en son extrémité 5'. 2) élongation du brin antisens non codant. La PCR effectuée utilise le produit de la première étape comme *primer*. 3) amplification du fragment compris entre le *primer* SP6 et le *primer* VSVG (partie non complémentaire à l'ADN du *primer* hybride). L'utilisation du *primer* VSVG permet l'amplification spécifique des fragments comportant la mutation.

2.2.2. Sous-clonage

Chaque ADN muté (insert) ainsi amplifié a tout d'abord et purifié sur colonne dans 30 μ l (Genomed PCR purification kit) puis a été digéré à l'aide des enzymes de restriction *MluI* (GIBCO BRL 10u/ μ l) + *BsgI* (New England Biolabs 3u/ μ l) pour H1 et *MluI* + *XbaI* (Roche 10u/ μ l) pour H2 et H3 dans un volume final de 40 μ l (insert: 30 μ l, tampon enz. 10x: 4 μ l, enz.: 2 μ l chacune, H₂O: 4 μ l; 1h à 37°C). Les produits digérés ont été séparés par migration sur gel d'agarose de 0.8% (tampon TBE 1x + bromure d'éthidium 0.2 μ g/ml) [32]. La taille des fragments d'intérêts a pu être estimée à l'aide du marqueur de taille 1kb (Invitrogene). Les bandes à 524 pb (H1) et 956 pb (H2 et H3) représentant les fragments portant les mutations, ont été extraites à l'aide d'un scalpel stérile puis l'ADN a été récupéré en utilisant le Kit MiniElute Gel Extraction de QIAGEN.

Simultanément, le vecteur pSD5 contenant la séquence codante pour α F-rENaC de même qu'un gène de résistance à un antibiotique (ampicilline) a été préparé de la manière suivante: les fragments compris entre les sites de restriction *MluI* et *BsgI* (H1) ou *MluI* et *XbaI* (H2 et H3) ont tout d'abord été enlevés par une double digestion exécutée dans un volume final de 20 μ l (vecteur ~ 500ng/ μ l: 2 μ l, tampon enz. 10x: 2 μ l, enz.: 2 μ l chacune, H₂O: 14 μ l; 1h à 37°C). Puis, l'extrémité cohésive 5' du vecteur ainsi linéarisé a été déphosphorylée à l'aide de la phosphatase alcaline (Boehringer Mannheim) dans un volume final de 30 μ l (vecteur linéarisé: 18 μ l, tampon 10x: 3 μ l, phosphatase 1 u/ μ l: 1 μ l, H₂O: 8 μ l; 30 min. à 37°C). Ces étapes réalisées avec succès, le vecteur a été isolé des fragments digérés puis récupéré à l'aide de la même méthode précédemment décrite dans le cadre de l'isolation des fragments PCR digérés.

Une fois en possession du vecteur linéarisé et des fragments PCR digérés, il s'est ensuivi une ligation de ces éléments à l'aide de la T4 DNA ligase et du kit de ligation de Roche (*Rapid DNA ligation kit*) dans un volume final de 20 μ l (rapport de masse vecteur/insert: 1/7, tampon dilution DNA 5x: 2 μ l, tampon ligase 2x: 10 μ l, ligase: 1 μ l, H₂O: complété à 20 μ l; 30 min. à température ambiante).

Les ligations ont été utilisées pour transformer la souche DH10 β de *E.coli* selon la méthode décrite par Chung et Miller (produit de ligation dans 100 μ l DH10 β , 30 min. sur glace, 2 min. à 42°C, 2 min. sur glace, transfert dans 900 μ l LB, 1h 30 min. dans agitateur à 37°C, centri. 1 min. 13'000rpm, culot resuspendu dans 200 μ l) [33]. L'étalement sur boîtes LB + Ampicilline (100 μ g/ml) des 200 μ l obtenus puis leur incubation à 37°C O/N ont ensuite permis de sélectionner les colonies ayant acquis la résistance à l'ampicilline par intégration du plasmide ligé et par conséquent possédant le gène α F-rENaC muté.

L'ADN plasmidique de ces colonies recombinantes a été extrait en utilisant le Kit QIAprep Spin Miniprep de Qiagen puis digéré dans un volume final de 20 μ l (vecteur ~ 500ng/ μ l: 1 μ l, tampon enz. 10x: 2 μ l, enz.: 1 μ l, H₂O: 16 μ l; 1h à 37°C) à l'aide des enzymes pour lesquelles des sites de restriction ont été ajoutés (*SacI* pour l'hélice H2) ou supprimés (*SacI* pour l'hélice

H1 et *KpnI* pour l'hélice H3) afin de s'assurer de la présence du fragment muté (digestion contrôle).

2.2.3 Séquençage

Dans le but de vérifier la présence des mutations ciblées, de détecter d'autres mutations éventuelles non désirées et de contrôler que le cadre de lecture du fragment inséré est correct, des séquençages ont été réalisés sur les plasmides sélectionnés lors des digestions contrôles (cf. Annexe C). Tous les séquençages ont été effectués par la firme Synergene et les *primers* utilisés sont SP6 (tableau 1) pour le séquençage de l'hélice H1 et SP6 + $\alpha 440$ (ATG ATG TAC TGG CAG TTC GC) pour le séquençage des hélices H2 et H3.

2.2.4. Transcription *in vitro*

Les plasmides portant les mutations désirées ont été linéarisés à l'aide de l'enzyme de restriction *BglII* (Invitrogen 10u/ μ l) dans un volume final de 20 μ l (vecteur \sim 500ng/ μ l: 2 μ l, tampon enz. 10x: 2 μ l, enz.: 1 μ l, H₂O: 14 μ l; 1h à 37°C). Une fois linéarisé, ces plasmides ont tout d'abord été purifiés (Précipitation: 0.1 vol. NaAcétate 3M, 2 vol. EtOH 100%; -20°C O/N. Centrifugation: 30 min. à 13'000 rpm. Lavage: EtOH 70%; 2 min. à 13'000 rpm. Dilution: 10 μ l d'H₂O *RNase free*) puis utilisés pour synthétiser les cRNA à l'aide de la SP6 RNA polymérase (Promega 15u/ μ l) dans un volume final de 50 μ l (Tampon 5x transcr. SP6: 10 μ l, A/C/U-TP 5mM & GTP 1mM: 15 μ l, GpppG 5mM: 5 μ l, DTT 100mM: 5 μ l, RNAsin 40u/ μ l: 1.3 μ l, H₂O *RNase free*: 5.7 μ l, BSA 400ng/ μ l: 1 μ l, plasmide linéarisé & purifié: 5 μ l, SP6 RNA polymérase 15u/ μ l: 2 μ l; 1h à 40°C) [34]. Une fois la synthèse des RNA terminée, les plasmides ont été éliminés (*DNAseI RNase free*: 1 μ l; 30 min. à 37°C). Grâce au vecteur pSD5 utilisé, le RNA final porte en 5' la coiffe 7-méthylguanosine et en 3' la queue polyadénylée.

La quantification du RNA synthétisé a été effectuée par migration des échantillons dilués (2x et 5x) sur gel d'agarose 0.8% (tampon TAE 1x + bromure d'éthidium 0.2 μ g/ml) en présence de RNA polyA⁻ de concentration connue. Des dilutions ont été effectuées afin d'obtenir une concentration finale de 100ng/ μ l.

2.2.5. Expression des mutants α F-rENaC dans les ovocytes de *Xenopus*

Les ovocytes ont été extraits des ovaires de *Xenopus laevis*, traités pendant environ 1h par de la collagénase en milieu *Modified Barth's Saline* (NaCl 85mM, KCl 1mM, NaHCO₃ 2.4mM, MgSO₄ 0.82mM, Hepes 16.3mM, NaOH 4.08mM; pH 7.4) puis lavés et conservés à 19°C en milieu MBS additionné de pénicilline (10 μ g/ml) et de streptomycine (5 μ g/ml). Ils ont ensuite été sélectionnés sur des critères de taille, de sphéricité et de pigmentation correspondant au stade V de leur croissance afin d'être injectés [35]. L'injection d'une quantité égale de cRNA

de chaque sous-unité d'ENaC a été effectuée grâce à une micropipette et chaque ovocytes a ainsi reçu 10ng de cRNA de chaque sous-unité.

Les mesures électrophysiologiques ont été effectuées environ 24h après l'injection, temps nécessaire à l'expression d'ENaC par les ovocytes. Pendant ce temps, les ovocytes ont été incubés dans du *Modified Barth's Saline low sodium* (NaCl 10mM, MgSO₄ • 7H₂O 0.82mM, CaCl₂ • 2H₂O 0.41 mM, KCl 2mM, Ca(NO₃)₂ 0.33mM, Hepes 5mM, NMDG 80mM; pH 7.4) additionné de pénicilline (10µg/ml) et de streptomycine (5µg/ml).

Le courant macroscopique de sodium résultant de l'expression d'ENaC (courant amiloride sensible) a été mesuré sur l'ovocyte entier par la technique de *voltage-clamp* à deux électrodes (*TEV-clamp*). L'une des électrodes sert à mesurer le potentiel de membrane à chaque instant et la seconde permet de maintenir ce potentiel à -100mV par une injection de courant contrebalançant le courant résultant de l'entrée du sodium par les canaux ENaC. Ces mesures ont été effectuées dans du milieu *Frog Ringer* (NaCl 120mM, KCl 2.5mM, CaCl₂ • 2H₂O 1.8mM, Hepes 10mM; pH 7.2) additionné ou non d'amiloride (10µM). Toutes les mesures ont été effectuées à température ambiante.

3. Résultats et discussion

3.1. Bioinformatique

3.1.1. Test du programme de prédiction d'hélices α amphipathiques

Avant d'avoir pu utiliser le programme de prédiction d'hélices α amphipathiques (cf. 2.1.2.) sur la sous-unité α d'ENaC, il a été nécessaire de tester ce programme sur des hélices α amphipathiques connues pour vérifier l'exactitude de ses prédictions. Dans ce but, des recherches dans la littérature scientifique ont été exécutées afin de mettre à jour une éventuelle base de données de telles hélices. Malgré ces recherches approfondies, bien peu d'éléments susceptibles de permettre un test sérieux du programme ont été trouvés. Si une base de donnée réalisée par Segrest et al. dans le cadre d'une tentative de classification d'hélices α amphipathiques [36] existe bel et bien, cette dernière apparaît rapidement comme inutilisable. Elle regroupe en effet des hélices α possédant une "amphipathicité" qualitativement très variable et parfois douteuse (observations faites par la méthode *helical wheel*) en conséquence d'une définition imprécise de l'amphipathicité. Ainsi, lors du test du programme réalisé avec une septantaine d'hélices α issues de cette base de donnée, seules ~ 50% d'entre elles présentaient, pour une fenêtre minimale de 9 résidus (minimum significatif), un moment hydrophobe moyen situé entre 0.6 et 0.9 (respectivement: valeur limite établie dans ce projet en tenant compte des bruits de fond et valeur maximale observée). Etabli sur la base de donnée de Segrest et al. [36], le test montre clairement que l'algorithme de prédiction utilisé ne permet sûrement pas de détecter efficacement toutes les structures hélicoïdales "amphipathiques" de la littérature. Cependant, il est important de mentionner une nouvelle

fois que la base de donnée utilisée dans ce test comporte des structures hélicoïdales dont l'amphipathicité a certainement été définies par l'utilisation des deux méthodes quantitatives hautement subjectives que sont l'*helical wheel* et l'*helical net*. Elle comporte par conséquent des structures amphipathiques très variables. Le sentiment de l'absence d'une définition claire de l'amphipathicité capable de déterminer efficacement qu'elles devraient être les structures pour lesquelles un programme de prédiction devrait émettre un signal fort est très nettement renforcé à la lecture des premières lignes de la revue de Segrest et al. [36]. En effet, l'introduction de cette revue reste très vague sur la définition de l'amphipathicité utilisée pour établir la base de donnée. Elle justifie la présence des hélices α dans cette base de donnée en ces termes: "*In most instances, evidence for the existence of amphipathic/amphiphilic helices is indirect but convincing*" !

Dans l'impossibilité de tester le programme de prédiction à l'aide d'une base de donnée cohérente et clairement définie, il a été nécessaire de trouver un autre moyen pour valider les prédictions du programme. Pour ce faire, une cinquantaine d'hélices α de la base de donnée de Segrest et al. dont la structure est présente dans la *Protein Data Bank* ont été choisies dans le but d'associer pour chacune d'entre elles une structure protéique (programme de visualisation 3D) aux résultats de prédiction du programme. L'idée était ainsi de pouvoir déterminer visuellement si les pics des moments hydrophobes moyens associés à certaines hélices α par le programme correspondaient bien à la réalité d'une structure hélicoïdale amphipathique (observation d'une ségrégation des résidus hydrophobes/hydrophiles basée sur l'échelle d'hydrophobicité d'Eisenberg [29]). Elle permettait également d'associer des degrés visuels d'amphipathicité aux moments hydrophobes moyens généralement compris entre 0.1 et 0.9. A l'aide de cette approche, les remarques générales présentées dans le tableau 2 ont pu être faites.

D'une manière générale, le test du programme de prédiction basé sur une détermination visuelle de la signification des moments hydrophobes moyens donne des résultats satisfaisants. Par le biais de ce test, il est clairement démontré que les fenêtres de calcul possédant un moment hydrophobe moyen égal ou supérieur à 0.6 (classe1) peuvent sans trop d'erreur être considérée comme potentiellement amphipathiques. De plus, dans cette classe le degré d'amphipathicité semble être assez correctement représenté par le moment hydrophobe moyens. Ainsi, une structure hélicoïdale caractérisée par un moment hydrophobe moyen de 0.6 possédera une ségrégation de ses résidus hydrophobes/hydrophiles moins marquée qu'une structure caractérisée par un moment hydrophobe moyen de 0.9. En ce qui concerne les fenêtres de calcul présentant un moment hydrophobe moyen situé entre 0.4 et 0.5 (classe 2), la prédiction d'hélices α amphipathiques est plus difficilement réalisable. En effet, cette classe regroupant aussi bien de "bonnes" comme de "mauvaises" structures amphipathiques, un segment protéique hélicoïdal correspondant à une fenêtre de calcul caractérisée par un tel

moment hydrophobe moyen ne peut pas à coup sûr être considéré comme possédant une structure amphipathique.

Classe	Mom. hydrophobes	Remarques
1	0.6 à 0.9	Les structures qui correspondent aux fenêtres de calcul présentant de tels moments hydrophobes moyens et dont certains exemples sont montrés dans la Figure 11A et B sont, dans le cadre de ce travail, qualifiées comme bonnes à très bonnes en terme d'amphipathicité. La ségrégation des résidus hydrophobes/hydrophiles y est très nettement observée. Le passage progressif de 0.9 à 0.6 se traduit structurellement par la présence toujours plus marquée d'un résidu hydrophobe ou hydrophile dans le domaine respectivement hydrophile ou hydrophobe de la structure hélicoïdale.
2	0.4 et 0.5	Les fenêtres de calcul caractérisées par de telles moments hydrophobes moyens sont associées à des structures possédant une amphipathicité plus que variable comme le montre les exemples de la Figure 11C à F. Si quelques unes d'entre elles peuvent être qualifiées de bonnes (voir très bonnes dans de rares cas) par comparaison aux structures observées dans la classe 1, d'autres au contraire se voient qualifiées de moyennes à mauvaises. Dans ces structures moyennes ou mauvaises, la ségrégation des résidus hydrophobes/hydrophiles y est encore parfois observable mais la présence de résidus hydrophobes ou hydrophiles ne correspondant pas au domaine auquel ils participent y est de plus en plus soutenue.
3	0.1 à 0.3	Dans cette dernière classe de moments hydrophobes, se trouvent des structures correspondant aux fenêtres de calcul qualifiées de mauvaise. Comme le montre l'exemple de la Figure 11G, elles possèdent clairement une absence de structure amphipathique. Cette absence se traduit soit par la seule présence de résidus hydrophobes ou hydrophiles soit par la présence de résidus hydrophobes et hydrophiles aléatoirement disséminés le long de l'axe central de l'hélice α .

Tab. 2 : Observations structurales concernant les moments hydrophobes moyens observés dans les divers profils amphipathiques étudiés.

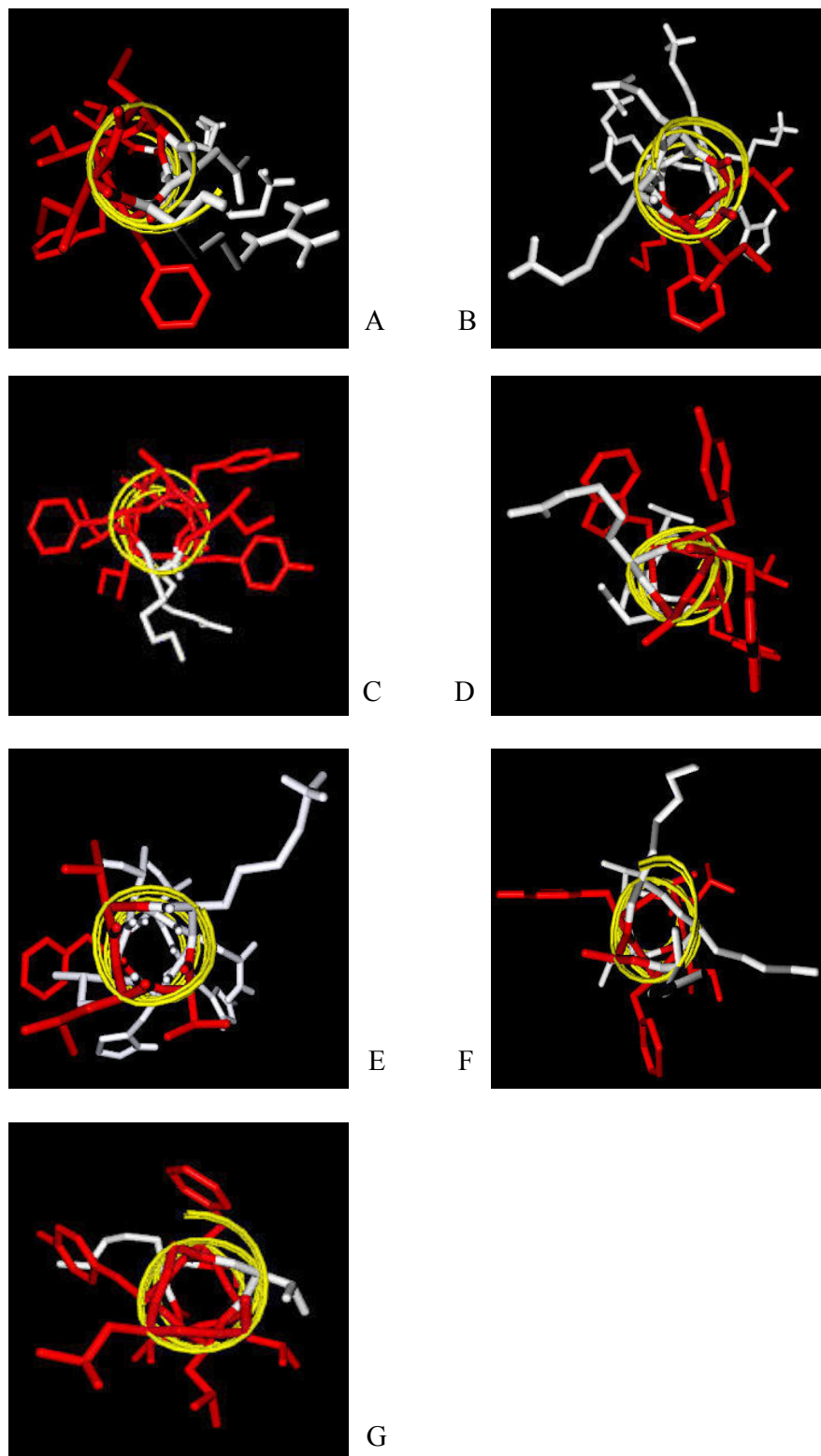


Fig. 11: Exemples structuraux des différentes classes de moments hydrophobes décrits dans le tableau 2. A et B: classe 1. C à F: classe 2. G: classe 3. Rouge: résidus hydrophobes. Blanc: résidus hydrophiles. Jaune: représentation de l'hélice α par un ruban hélicoïdal. A: Myohémérithrine a.a 27 à 37; B: Myohémérithrine a.a 74 à 84; C: Calcium ATPase a.a 289 à 299; D: Bactériorodhopsine a.a 80 à 90; E: Myohémérithrine a.a 47 à 57; F: Bactériorodhopsine a.a 38 à 48; G: Bactériorodhopsine a.a 21 à 31.

Cette classe démontre d'une certaine manière les limites de l'algorithme de prédiction choisi dans le cadre de ce travail. En effet, si la nature vectorielle du moment hydrophobe permet élégamment d'intégrer les deux composantes essentielles à la prédiction du degré d'amphipathicité d'une hélice α (ségrégation structurale des résidus hydrophobes/hydrophiles et degré d'hydrophobicité de chaque résidu), elle comporte cependant le désavantage de ne pas définir l'importance réciproque de ces deux composantes dans la détermination du moment hydrophobe. Ce manquement rend ainsi possible les observations faites dans la classe 2 qui associe aussi bien, à des moments hydrophobes moyens proches, des structures hélicoïdales possédant une ségrégation idéale mais des résidus à l'hydrophobicité faible que des structures à la ségrégation douteuse mais présentant une disposition d'hydrophobicité telle qu'il en émane un moment hydrophobe moyen non négligeable. Dans cette classe 2, le faible signal d'une éventuelle structure hélicoïdale amphipathique se perd par conséquent dans un bruit de fond important et empêche ainsi toute prédiction fiable concernant le degré d'amphipathicité des hélices α étudiées. Contrairement à la classe précédente, la classe 2 ne permet pas une bonne représentation du degré d'amphipathicité d'une hélice α par le moment hydrophobe. Passé cette zone d'ombre que représente la classe 2, le moment hydrophobe moyen peut à nouveau clairement être associé à un degré d'amphipathicité. Parmi les structures hélicoïdales caractérisées par un moment hydrophobe moyen inférieur ou égal à 0.3 (classe 3) aucune ne peut être considérée comme amphipathique et l'ambiguïté observée lors de la classe 2 disparaît bel et bien.

En résumé, le programme de prédiction d'hélices α amphipathiques élaboré dans le cadre de ce travail rend bien compte d'une réalité structurale pour les segments protéiques caractérisés par un moment hydrophobe moyen élevé (> 0.6). Passé cette limite inférieure, les prédictions ne sont plus soutenues par des évidences structurales. Il est par conséquent difficile voir impossible d'affirmer être en présence d'une structure hélicoïdale amphipathique pour un segment protéique caractérisé par un moment hydrophobe moyen < 0.6 .

3.1.2. Hélices α amphipathiques dans α -rENaC

3.1.2.1. Prédiction sur α -rENaC

Avec la conviction que les moments hydrophobes moyens > 0.6 possèdent une signification structurale en terme d'amphipathicité, la prédiction d'éventuelles hélices α amphipathiques chez α -rENaC a été effectuée. Initialement réalisée avec une fenêtre de 9 résidus, la prédiction a rapidement mis en évidence la présence d'un certain nombre de segments hélicoïdaux possédant un moment hydrophobe moyen > 0.7 (préféré à la limite 0.6 pour être sûr d'isoler des segments structurellement significatifs). Cependant, partant du principe que 9 résidus (2.5 tours d'hélices) ne représentaient pas une hélice de taille significative, une nouvelle approche a été utilisée pour isoler les segments hélicoïdaux les plus grands

possédant un moment hydrophobe moyen > 0.7 . Cette dernière consistait tout d'abord en une première prédiction réalisée avec une fenêtre de 20 résidus (statistiquement parlant plus la fenêtre est grande plus le moment hydrophobe moyen diminue) suivie d'une série de prédictions établies sur la base de fenêtres toujours plus petites. Au fur et à mesure de la diminution des tailles de fenêtre, certains fragments hélicoïdaux ont clairement été mis en évidence par une augmentation progressive de leur moment hydrophobe moyen. La taille maximale de chaque fragment a ensuite été déterminée comme étant la plus grande fenêtre possédant un moment hydrophobe moyen > 0.7 . A l'aide de cette approche, trois des cinq hélices α prédites par le *Server PHDsec Site* ont été identifiées comme possédant dans leur séquence des fragments potentiellement amphipathiques (cf. Annexe A). Les caractéristiques des hélices α et des fragments amphipathiques ayant permis leur mise en évidence sont présentées dans le tableau 3.

Hélice	Position des acides aminés	Fragment	Taille	$\langle \mu_h \rangle$	Périodicité angulaire
1	69 – 89	⁷⁴ IEFHRSYRELFQF ⁸⁶	13	~ 0.75	103°
2	165 – 187	¹⁶⁸ IKEELEELDRI ¹⁷⁸	11	~ 0.71	103°
3	271 - 289	²⁷¹ GVDAVREWYRF ²⁸¹	11	~ 0.80	105°

Tab. 3 : Caractéristiques des hélices α et des fragments hélicoïdaux décrits dans le texte.

Parmi les séquences hélicoïdales d' α -rENaC révélées par le programme de prédiction et contenant un potentiel amphipathique important, certaines représentent des domaines connus, ou sont localisées proches de domaines connus, associés à des fonctions précises dans la littérature scientifique. Leur intérêt est de ce fait non négligeable. Il en est ainsi tout d'abord de la séquence ⁷⁴IEFHRSYRELFQF⁸⁶ participant à l'hélice α intracellulaire H1. Cette dernière se trouve juste en amont d'un groupe de 10 acides aminés dont le motif HG hautement conservés dans la famille des ENaC/DEG (T92 – C101). Cette courte séquence de 10 acides aminés consécutifs a été associée, lors d'une série de mutations, à une réduction importante de la probabilité d'ouverture (P_o) d'ENaC. Elle est par conséquent fortement suspectée de participer activement aux phénomènes de *gating* du canal [18]. Bien que ce segment hélicoïdal amphipathique ne se situe en aucune manière dans la boucle extracellulaire et qu'il ne participe ainsi pas à l'organisation structurale du pore externe du canal (but initial de ce travail), son étude fonctionnelle présente cependant un intérêt tout particulier lié à sa localisation. Ainsi, son étude permettra éventuellement une meilleure compréhension des phénomènes de *gating* d'ENaC. Ces phénomènes sont peut-être liés au positionnement de certains résidus que l'ancrage à la membrane plasmique d'une hélice α amphipathique permet de maintenir.

Outre le fragment amphipathique participant à l'hélice α H1, une deuxième séquence hélicoïdale amphipathique revêt une importance toute particulière de par sa localisation. En effet, la séquence ²⁷¹GVDVREWYRF²⁸¹ constituant une partie de l'hélice α H2 représente la majorité (4 résidus sur 6) du domaine WYRFHY. Ce domaine, et notamment l'acide aminé H282, a été démontré par Kleyman et al. comme participant activement à la liaison unissant l'amiloride au canal ENaC [16, 17]. Dans le modèle proposé par Kleyman et al., ce domaine constitue un des sites de la boucle extracellulaire qui sont essentiels à la stabilisation de la liaison de l'amiloride. Par repliement protéique, ce site se trouverait à proximité du site de liaison de l'amiloride généralement reconnu [15] et situé dans le pore du canal. Dans un sens, ce modèle localise l'hélice α amphipathique dans le pore externe du canal et nous permet d'envisager un rôle actif de cette hélice α amphipathique dans la formation de ce pore et dans le positionnement éventuel de résidus impliqués dans le bloque du canal par l'amiloride. Egalement associés à ce fragment amphipathique de l'hélice α H2 par leur proximité, deux acides aminés (H282 et R280) semblent jouer un rôle important dans l'inhibition de l'activité du canal par le nickel (Ni^{2+}) [37]. Récemment publiées, les évidences d'une activité inhibitrice du Ni^{2+} par modification de la probabilité d'ouverture (P_o) du canal ne sont pas encore clairement comprises et acceptées. Cependant, si elles devaient être confirmées, elles apporteraient peut-être un soutien supplémentaire en faveur de l'importance d'une présence amphipathique dans l'organisation du pore externe du canal.

Localisé dans une région dénuée de tout domaine fonctionnel connu, le dernier fragment hélicoïdal amphipathique localisé dans l'hélice α H3 ne possède pas d'autre intérêt que celui de sa seule présence dans la boucle extracellulaire. L'étude fonctionnelle de ce fragment va par conséquent se borner à déterminer son importance d'un point de vue général.

3.1.2.2. Prédiction sur la famille ENaC/DEG

Afin d'obtenir une première indication sur l'importance éventuelle des hélices α amphipathiques mises en lumière par le programme, des prédictions ont été réalisées sur une série restreinte de membres de la famille des ENaC/DEG. Partant du principe qu'un élément relativement conservé au sein d'une famille de protéine décèle une structure importante à la fonction protéique, ces prédictions ont permis d'accorder une certaine importance fonctionnelle des trois segments hélicoïdaux amphipathiques susmentionnés. En effet, comme le montre la Figure 12, le schéma de la localisation des segments hélicoïdaux amphipathiques observés chez les différents membres de la famille présente quelques traits communs. Cette remarque est surtout valable pour les membres de la sous-famille des ENaC où la présence d'au moins deux hélices α amphipathiques sur trois est constamment observée autour du segment transmembranaire TM1. En ce qui concerne les quelques autres membres de la famille, le *pattern* caractéristique mis en évidence dans les membres de la sous-famille des ENaC est nettement moins évident. Cependant, quelques hélices α amphipathiques se

trouvent régulièrement à proximité de TM1 que ce soit côté intra ou extracellulaire. Ajoutées au fait que les séquences hélicoïdales amphipathiques des différents membres de la famille possèdent parfois très peu d'homologie, ces observations suggèrent que les structures hélicoïdales amphipathiques représentent peut-être des éléments structuraux importants à la fonction protéique.

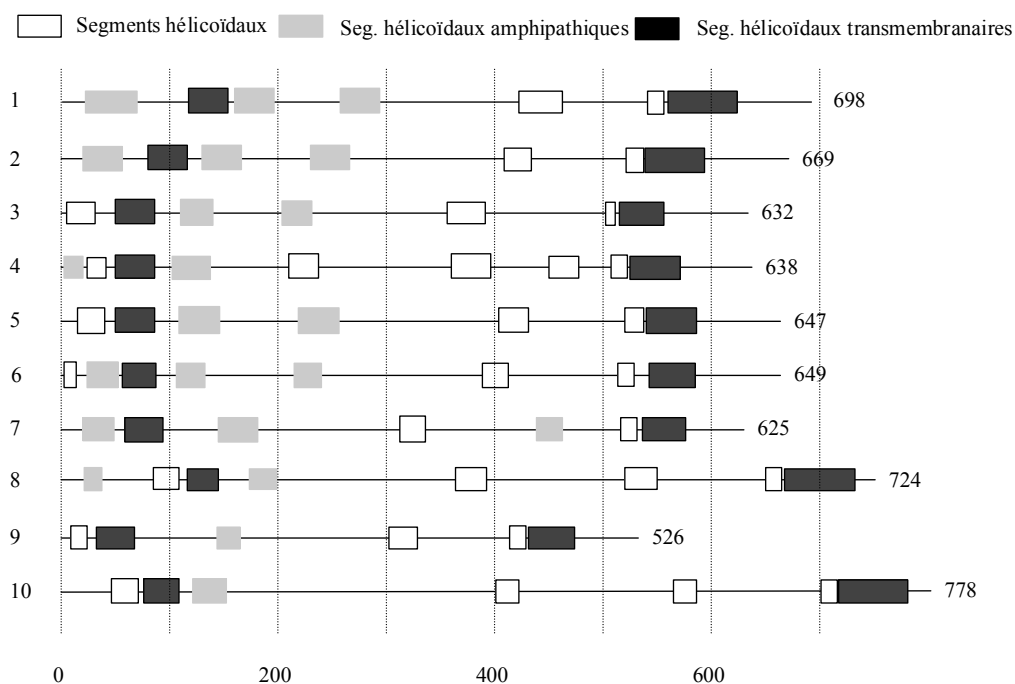


Fig. 12 : Schéma de localisation des hélices α dans la famille des ENaC/DEG. 1. α -rENaC. 2. α -hENaC. 3. α -xENaC. 4. β -rENaC. 5. β -xENaC. 6. γ -hENaC. 7. heFaNaCh. 8. cMec10. 9. rAsic1. 10. cDeg1. Définition des segments hélicoïdaux amphipathiques : moment hydrophobe moyen > 0.6.

3.1.2.3. Définition de substitutions destructrices optimales

Dans le but d'entreprendre l'étude fonctionnelle des différents fragments mis en évidence par le programme de prédiction, il a été décidé de procéder à un certain nombre de substitutions d'acides aminés visant à détruire l'amphipathicité des hélices α . Le choix des substitutions à reproduire en laboratoire a été déterminé à l'aide d'un deuxième programme informatique (cf. 2.2.3.). Ce programme a été établi afin de définir qu'elles sont les substitutions d'un fragment hélicoïdal donné les plus destructrices en terme d'amphipathicité.

Réalisée sur la sous-unité α d'ENaC et ceci pour les trois fragments isolés lors de l'exécution du premier programme, la détermination des substitutions destructrices optimales a donné les résultats représentés dans le tableau 4.

Hélice	Fragment "sauvage"	Fragment "mutant"	Substitutions
1	⁷⁴ IEFHRSYREL ⁸⁶ FQF	⁷⁴ IEFHISYRE ⁸⁶ K ⁸¹ FQF	R78I L83K
2	¹⁶⁸ IKEELEELDRI ¹⁷⁸	¹⁶⁸ IKEELI ¹⁷⁸ ELDII	E173I R177I
3	²⁷¹ GVDAVREWYRF ²⁸¹	²⁷¹ GVDA ²⁸¹ KIEWYRF	V275K R276I

Tab. 4 : Substitutions destructrices optimales pour chaque fragment mis en évidence par le programme de prédiction de structures hélicoïdales amphiphatiques.

Une fois les substitutions définies, de nouvelles prédictions de structure secondaire ont été réalisées (*Server PHDsec Site*) sur la base des séquences "mutantes". Ces nouvelles prédictions ont été effectuées afin de vérifier que la présence des acides aminés substitués n'empêchait en aucune manière la formation de la structure hélicoïdale initialement prédite. Ainsi, les résultats présentés ci-dessus représentent le fruit de plusieurs séries de substitutions et constituent les substitutions les plus destructrices n'altérant aucunement les prédictions de structures secondaires initialement observées.

3.2. Biologie moléculaire

3.2.1 Préparation des mutants

Fort des informations obtenues lors de la partie bioinformatique de ce travail de diplôme, il a ensuite été possible de s'engager dans l'étude fonctionnelle des fragments hélicoïdaux amphipathiques mis en évidence. La première étape de la partie "pratique" de ce travail de diplôme a tout d'abord été la construction de vecteurs d'expression pSD5 (cf. Annexe B) contenant des gènes mutés de la sous-unité α F-rENaC (la sous-unité "flaggée" a été choisie en prévision d'éventuelles études d'expression de surface). Comme le montre le tableau 5, les mutations introduites par mutagenèse dirigée (cf. 2.2.1.) dans le gène de la sous-unité α F correspondent aux substitutions décrites dans le paragraphe ci-dessus.

Différents problèmes ont été rencontrés lors des étapes de constructions du vecteur d'expression pSD5. Des problèmes tout d'abord liés aux mutagenèses dirigées pour lesquelles il a fallu déterminer les bonnes conditions d'amplification par PCR. Puis des problèmes avec les enzymes de restriction qu'il a été nécessaire de changer plusieurs fois avant de trouver celles ne possédant pas de coupures compatibles ou autorisant un bon taux de ligation (cf. 2.2.2.). Cette première étape pratique faussement considérée comme ne devant pas poser trop

Hél.	Frag.	Séquence protéique/nucléique												
1	"sauv."	I	E	F	H	R	S	Y	R	E	L	F	Q	F
		ATT	GAA	TTC	CAC	CGC	TCC	TAC	CGG	GAG	CTC	TTC	CAG	TTC
	"mut."	I	E	F	H	<u>I</u>	S	Y	R	E	<u>K</u>	F	Q	F
		ATT	GAA	TTC	CAC	ATC	TCC	TAC	CGG	GAG	AAG	TTC	CAG	TTC
2	"sauv."	I	K	E	E	L	E	E	L	D	R	I		
		ATT	AAA	GAG	GAG	CTG	GAA	GAG	CTG	GAC	CGC	ATC		
	"mut."	I	K	E	E	L	<u>I</u>	E	L	D	<u>I</u>	I		
		ATT	AAA	GAG	GAG	CTG	ATA	GAG	CTG	GAC	ATC	ATC		
3	"sauv."	G	V	D	A	V	R	E	W	Y	R	F		
		GGG	GTG	GAT	GCA	GTG	AGG	GAG	TGG	TAC	CGC	TTC		
	"mut."	G	V	D	A	<u>K</u>	<u>I</u>	E	W	Y	R	F		
		GGG	GTG	GAT	GCA	AAG	ATT	GAG	TGG	TAC	CGC	TTC		

Tab. 5 : Mutations introduites dans le gène de la sous-unité αF et substitutions correspondantes.

de problème a, en réalité, occupé la majorité de cette deuxième partie pratique ne laissant que peu de place aux caractérisations fonctionnelles des fragments amphipathiques. Malgré ces problèmes et avec beaucoup de patience et de travail, il a tout de même été possible d'obtenir les constructions envisagées dont la présence des mutations a été confirmée par séquençage (cf. Annexe C).

3.2.2 Electrophysiologie

Bien que le temps à disposition pour l'étude fonctionnelle proprement dite des fragments hélicoïdaux amphipathiques n'ait pas été aussi long qu'espéré, il a tout de même été possible d'entreprendre une première expérience d'électrophysiologie visant à déterminer l'impact fonctionnel de la perte d'amphipathicité de ces fragments. Ainsi, une fois les constructions validées par le séquençage, ces dernières ont été utilisées sans problèmes dans les étapes de transcriptions (cf. 2.2.4.). Ces étapes ont alors permis la production des ARN_m de la sous-unité αF d'ENaC nécessaires à l'injection des ovocytes (cf. 2.2.5.). Les ovocytes injectés, des mesures électrophysiologiques par *voltage clamp* à deux électrodes (TEV-*clamp*) ont été effectuées dont les résultats sont donnés dans la Figure 13.

Outre le résultat obtenu pour le deuxième fragment amphipathique ($\alpha FH2\beta\gamma$) participant à l'hélice $\alpha H2$ prédite par le *Server PHDsec Site* pour lequel la différence avec le contrôle ($\alpha F\beta\gamma$) n'est pas significative, les résultats suggèrent que la destruction de l'amphipathicité des deux autres fragments ($\alpha FH1\beta\gamma$ et $\alpha FH3\beta\gamma$) possède un impact fonctionnel. En effet, les courants Na⁺ macroscopiques sensibles à l'amiloride présentent une nette diminution et une

nette augmentation pour respectivement les fragments amphipathiques participant aux hélices α H1 et H3.

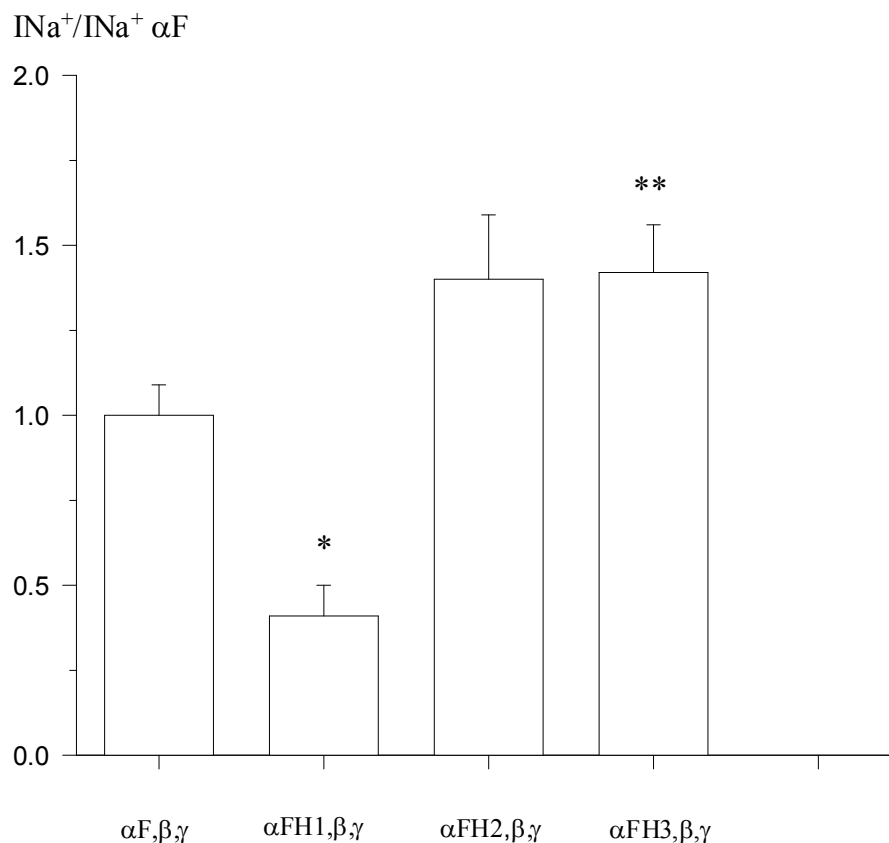


Fig. 13 : Effet des substitution dans α F-rENaC sur le courant Na^+ macroscopique sensible à l'amiloride (*voltage clamp* -100mV). Les courants ont été normalisés par rapport au courant observé dans les ovocytes exprimant ENaC type sauvage $\alpha\text{F}\beta\gamma$ (n=17). $\alpha\text{FH1}\beta\gamma$ et $\alpha\text{FH2}\beta\gamma$ (n=16), $\alpha\text{FH3}\beta\gamma$ (n=15). La barre d'erreur représente la SEM. * et ** Signification statistique respectivement $P<0.05$ et $P<0.01$.

Bien que ces études fonctionnelles préliminaires ne permettent pas vraiment de déterminer plus précisément quels sont les mécanismes qui sous-tendent cette variation des courants macroscopiques, il est cependant possible pour chaque fragment amphipathique de faire quelques observations sur la base de leur localisation. Pour commencer, il est remarquable que la destruction de l'amphipathicité du fragment participant à l'hélice α H1 ($\alpha\text{FH1}\beta\gamma$) soit associée à une diminution du courant Na^+ macroscopique. En effet, de par sa localisation proche d'une région de 10 acides aminés (T92 – C101) également associée à une diminution de courant lors de mutations les affectant [18], il est intéressant de penser que ces deux éléments participent en réalité à la même entité fonctionnelle. La diminution du courant faisant suite aux mutations de la région des 10 acides aminés ayant été attribuée à une diminution de la probabilité d'ouverture d'ENaC (P_0), il est justifiable de penser que la réduction du courant constatée lors de la destruction de l'amphipathicité du fragment considéré soit également la conséquence d'une réduction de cette probabilité. Ces deux

éléments juxtaposés représenteraient par conséquent une partie essentielle du *gating* du canal que des études fonctionnelles plus approfondies permettraient éventuellement de mieux définir.

Contrairement au premier fragment hélicoïdal amphipathique (α FH1 β , γ) possédant des résultats significatifs, le fragment amphipathique de l'hélice α H3 (α FH3 β γ) ne présente pas des résultats en relation avec les fonctions décrites pour la région dans lequel il se situe. En effet, constituant une partie importante d'un domaine associé par Kleyman et al. [16, 17] à un site participant à l'effet inhibiteur de l'amiloride (WYRFHY), une quelconque modification de l'amphipathicité de ce site devrait se traduire, suite à une modification structurale majeure, par une diminution du courant Na^+ macroscopique résultant d'une perte d'inhibition par l'amiloride. Or, les résultats obtenus pour ce fragment hélicoïdal amphipathique présente une nette augmentation du courant macroscopique incompatible avec l'altération d'un site inhibiteur. Ces observations peuvent par conséquent suggérer deux choses : 1) la perte d'amphipathicité du segment associé au site décrit par Kleyman et al. n'altère pas la fonction de ce dernier et la variation observée du courant résulte de l'altération d'une fonction indépendante (deux entités fonctionnelles distinctes) et 2) le site de Kleyman et al. associé au fragment amphipathique ne constitue pas un site de liaison à l'amiloride et la variation du courant découle de la modification de mécanismes fonctionnels encore inconnus (une seule entité fonctionnelle). Cette seconde observation constitue très probablement l'hypothèse la plus plausible. Elle est notamment soutenue par les résultats expérimentaux de Shild et al. [38] qui démontrent l'absence de toute perte d'inhibition par l'amiloride chez le canal mutant comportant la sous-unité α H282D. Dans les deux cas, des études fonctionnelles plus poussées sont nécessaires pour vérifier une éventuelle absence d'altération du site de liaison à l'amiloride et pour essayer de mieux comprendre quelle autre entité fonctionnelle peut être associée aux résultats obtenus.

Les résultats obtenus lors de la destruction de l'amphipathicité du segment associé à l'hélice α H2 (α FH2 β γ) ne présente pas une différence significative par rapport au contrôle. Les expériences réalisées ne permettent par conséquent pas d'associer un impact quelconque de la perte d'amphipathicité de ce segment. Localisé dans une région absente de tout domaine à la fonction connue, des résultats positifs auraient peut-être permis de mettre en évidence une nouvelle région importante dans la fonction protéique. Malgré l'absence de résultats probants, une étude fonctionnelle plus importante peut s'avérer intéressante notamment au regard de la localisation de ce segment dans l'organisation structurale du pore externe d'ENaC.

4. Conclusion et perspectives

Initialement axé sur la mise en évidence de structures hélicoïdales amphipathiques au sein de la boucle extracellulaire, ce travail de diplôme a non seulement répondu aux attentes des premiers jours mais c'est en plus intéressé à l'organisation structurale de l'extrémité

cytoplasmique N-terminale du canal. En décelant la présence de deux fragments amphipathiques associés à des hélices α localisées dans la partie extracellulaire d'ENaC et en démontrant l'importance fonctionnelle de l'un d'entre eux, il représente un premier pas vers une meilleure compréhension de l'organisation du pore externe de ce canal. Outre les informations relatives à la structure du pore externe d'ENaC, ce projet a également fourni des informations importantes sur la présence et l'importance fonctionnelle d'un fragment amphipathique participant à une hélice α intracellulaire en N-terminal.

Afin de compléter les maigres études fonctionnelles réalisées dans le cadre de ce travail de diplôme, il serait intéressant de réaliser, pour chaque fragment hélicoïdal amphipathique mis en évidence, un certain nombre d'expériences visant à mieux définir leur rôle dans le fonctionnement d'ENaC. Concernant le fragment amphipathique localisé au sein de l'hélice α H1 prédite par le site *Server PHDsec Site*, des expériences de *patch clamp* en *single channel* devrait être effectuées sur un canal comportant la sous-unités α mutante correspondante. Ces expériences permettraient de visualiser une modification de la probabilité d'ouverture (P_0) du canal et valideraient peut-être l'hypothèse faite dans ce travail d'une éventuelle participation de ce segments hélicoïdal amphipathique aux phénomènes de *gating*.

Suite aux observations contradictoires faites à propos de la région extracellulaire dans laquelle le fragment amphipathique associé à l'hélice α H3 se situe, il serait intéressant de tester l'affinité pour l'amiloride du canal ENaC comportant la sous-unité α mutante correspondante. Ces expériences de *voltage clamp* à deux électrodes effectuées sur des ovocytes injectés et réalisées en présence de diverses concentrations d'amiloride permettraient ainsi de clarifier un point obscur de ce travail en évaluant clairement la participation de ce segment au site de liaison de l'amiloride. Outre les questions relatives à l'amiloride, d'autres questions se posent sur la région où se trouve localisé ce fragment. En effet, cette région a également été récemment associée par Kleyman et al. [37] à un site de liaison du Ni^{2++} , inhibiteur potentiel du canal. La présence du fragment amphipathique à proximité de ce site constitue une bonne raison pour vérifier si l'inhibition d'ENaC par le Ni^{2++} représente bien une réalité. Ainsi, des expériences de *voltage clamp* à deux électrodes effectuées sur des ovocytes coexprimant la sous-unité α mutante et les sous-unités β et γ et réalisées en présence de nickel permettraient de mieux définir l'activité inhibitrice de Ni^{2++} . Elle pourrait éventuellement permettre de mettre également en évidence le rôle du fragment hélicoïdal amphipathique dans ce contexte.

Remerciements

Je voudrais par le biais de ces quelques lignes remercier tout d'abord le Prof. Laurent Schild de m'avoir donné la possibilité d'effectuer mon travail de diplôme au sein de son groupe. Je tiens à le remercier vivement de m'avoir "concocté" un projet mêlant à la fois des méthodes traditionnelles de la biologie et de la bioinformatique malgré mon inexpérience dans ce

dernier domaine. Merci aussi de ses précieux conseils et du dévouement sans limite qu'il a montré à mon égard.

Pour le domaine de la bioinformatique, je tiens ici à exprimer toute ma gratitude au Dr. Olivier Michielin de l'Institut Suisse de Bioinformatique (S.I.B) pour son aide capitale et ses excellents conseils relatifs à toutes les approches informatiques de ce travail. Bien que mes connaissances en Linux ou encore en Perl aient été limitées, il m'a permis en très peu de temps de prendre un énorme plaisir lors de l'élaboration de la partie bioinformatique.

Ma reconnaissance s'adresse également à Muriel Auberson et à Yvan Pfister pour leur encadrement et leur aide si cruellement appréciée pour la partie biologie moléculaire de ce projet. Un merci tout particulier pour leur bonne humeur qui, malgré mes questions scabreuses et mon impatience face à l'adversité, ne les a jamais quitté et m'a permis d'effectuer mon travail de diplôme dans les meilleures conditions.

Finalement, mes remerciements s'adressent également à Ivan Gautschy que j'ai côtoyé dans le cadre de mes expériences électrophysiologiques. Merci pour les injections de maître et les mesures qu'il a effectuées pour moi lors des derniers instants de mon travail de diplôme alors que le temps m'était compté.

Références

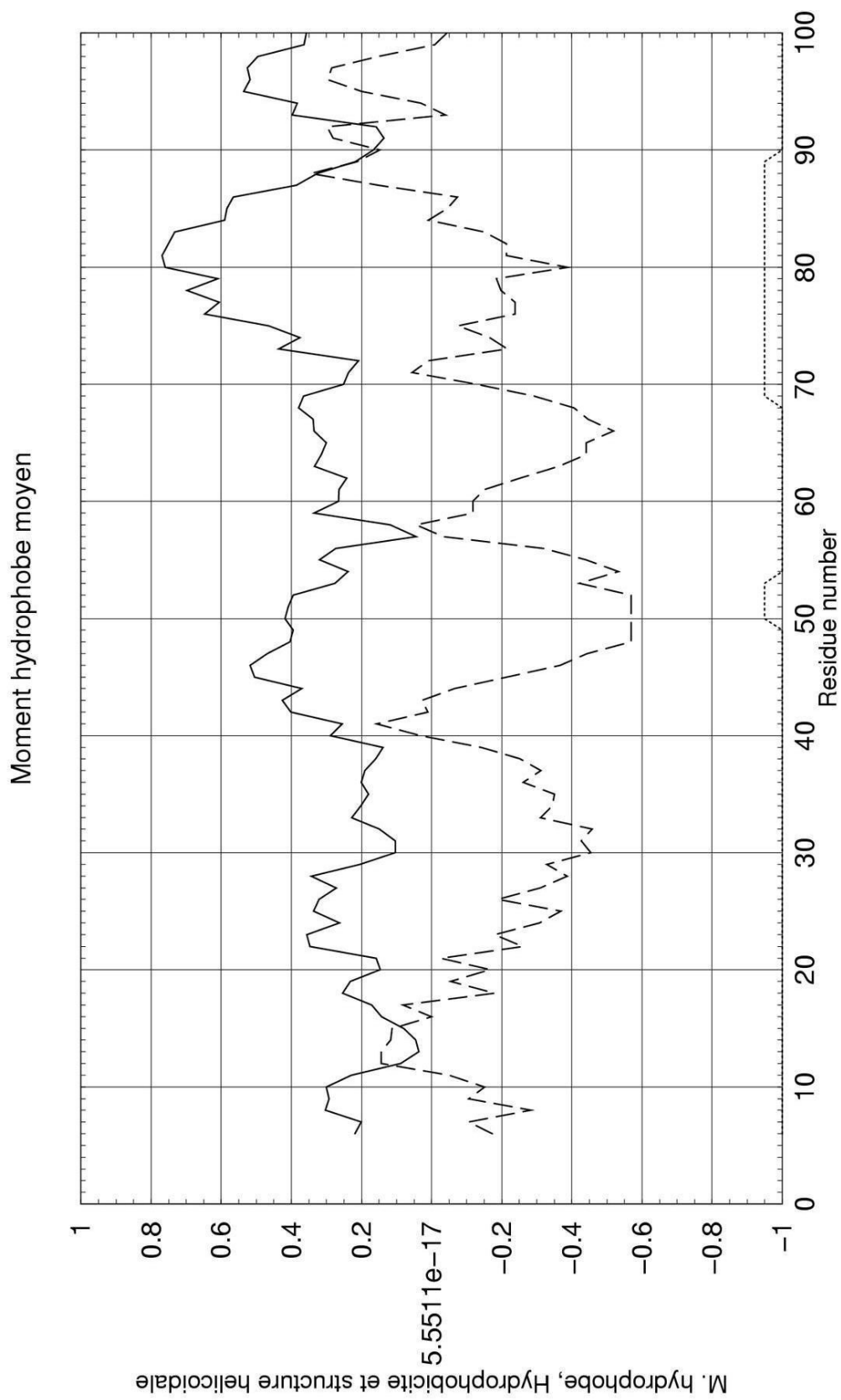
1. Roskopf D., Dusing R., Siffert W.: **Membrane sodium-proton exchange and primary hypertension.** *Hypertension* 1993, **21**: 607-617.
2. Shimkets R., Warnock D., Bositis C., Nelson-Williams C., Hansson J.: **Liddle's syndrome: heritable human hypertension caused by mutations in the β subunit of the epithelial sodium channel.** *Cell* 1994, **79**: 407-418.
3. Verrey F.: **Transcriptional control of sodium transport in tight epithelia by adrenal steroids.** *J. Membr. Biol.* 1995, **144**: 93-110.
4. Robert-Nicoud M., Flahaut M., Elalouf JM., Nicod M., Salinas M., Bens M., Horisberger JD., Rossier BC., Firsov D.: **Transcriptome of a mouse kidney cortical collecting duct cell line: effect of aldosterone and vasopressin.** *PNAS* 2001, **98**: 2712-2716.
5. Garty H., Palmer LG.: **Epithelial sodium channels: function, structure and regulation.** *Physiol. Rev.* 1997, **77**: 359-396.
6. Canessa CM., Schild L., Buell G., Thorens B., Gautschi I., Horisberger JD., Rossier BC.: **Amiloride-sensitive epithelial Na^+ channel is made of three homologous subunits.** *Nature* 1994, **367**: 463-467.
7. McNicholas CM., Canessa C.: **Diversity of channels generated by different combinations of epithelial sodium channel subunits.** *J Gen Physiol* 1997, **109**: 681-692.

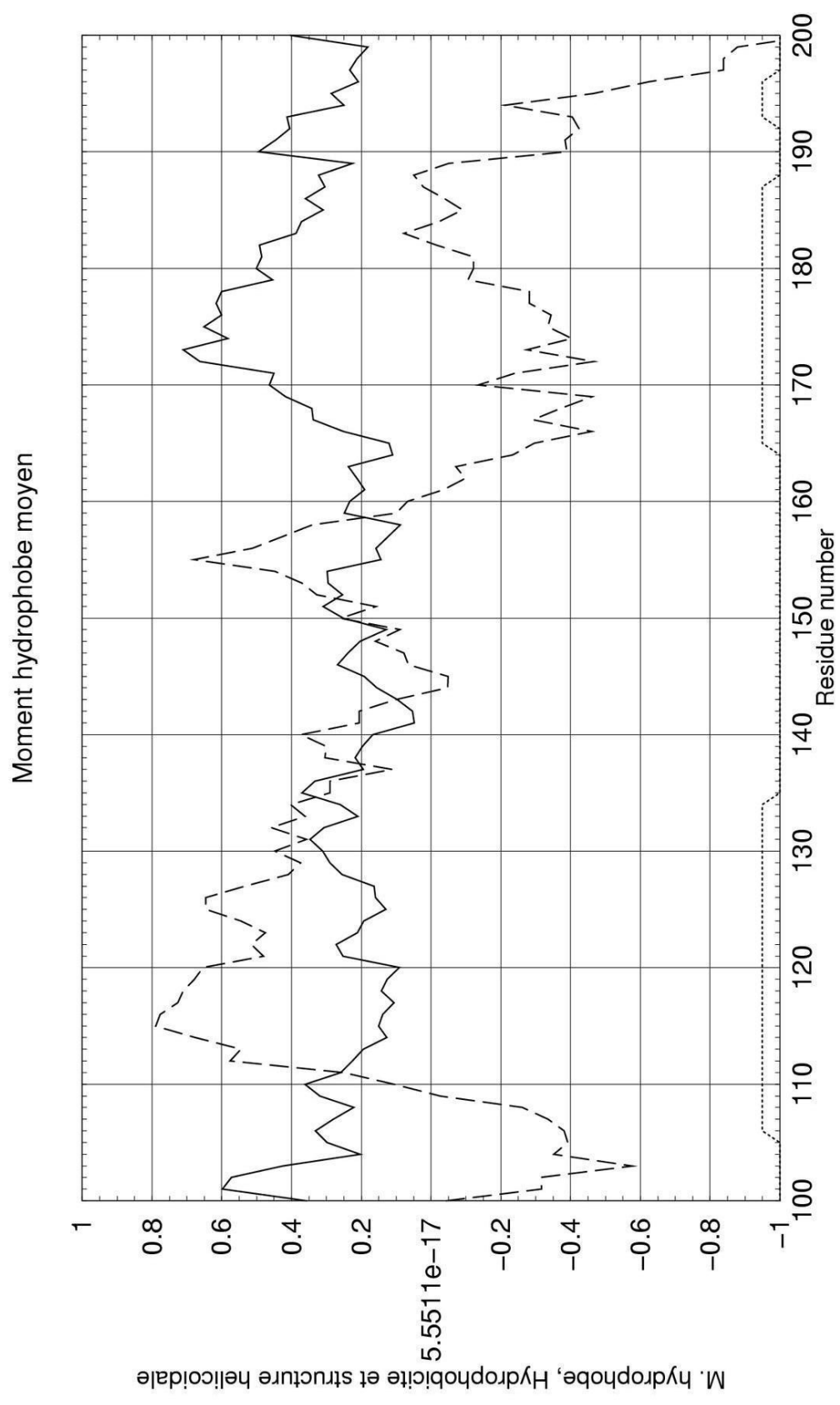
8. Bonny O., Chraïbi A., Loffing J., Jaeger NF., Grunder S., Horisberger JD., Rossier BC.: **Functional expression of a pseudohypoaldosteronism type I mutated epithelial Na⁺ channel lacking the pore-forming region of its alpha subunit.** *J. Clin. Invest.* 1999, **104**: 967-974.
9. Firsov D., Schild L., Gautschi I., Merillat AM., Schneeberger E., Rossier BC.: **Cell surface expression of the epithelial Na⁺ channel and a mutant causing Liddle's syndrome: a quantitative approach.** *Proc. Natl. Acad. Sci USA.* 1996, **93**: 15370-15375.
10. Firsov D., Gautschi I., Merillat AM., Rossier BC., Schild L.: **The heterotetrameric architecture of the epithelial sodium channel (ENaC).** *EMBO J.* 1998, **17**: 344-352.
11. Eskandari S., Snyder PM., Kreman M., Zampighi GA., Welsh MJ., Wright EM.: **Number of subunits comprising the epithelial sodium channel.** *J. Biol. Chem.* 1999, **274**: 27281-27286.
12. Garty H., Palmer LG.: **Epithelial sodium channels – function, structure and regulation.** *Physiol. Rev.* 1997, **77**: 359-396.
13. Kellenberger S., Gautschi I., Schild L.: **A single point mutation in the pore region of the epithelial Na⁺ channel changes ion selectivity by modifying molecular sieving.** *Proc. Natl. Acad. Sci. USA* 1999, **96**: 4170-4175.
14. Hamilton KL., Eaton DC.: **Single-channel recordings from amiloride-sensitive epithelial sodium channel.** *AM. J. Physiol. Cell Physiol.* 1985, **249**: C200-C207.
15. Schild L., Schneeberger E., Gautschi I., Firsov D.: **Identification of amino acid residues in the α , β , γ subunits of the epithelial sodium channel (ENaC) involved in amiloride block and ion permeation.** *J. Gen. Physiol.* 1997, **109**: 15-26.
16. Ismailov II., Kieber-Emmons T., Lin CM., Berdiev BK., Shlyonsky VG., Patton HK., Fuller CM., Worreli R., Zuckerman JB., Sun W., Eaton DC., Benos DJ., Kleyman TR.: **Identification of an amiloride binding domain within the alpha-subunit of the epithelial Na⁺ channel.** *J. Biol. Chem.* 1997, **272**: 21075-21083.
17. Kieber-Emmons T., Lin C., Foster MH., Kleyman TR.: **Antiidiotypic antibody recognizes an amiloride binding domain within the alpha-subunit of the epithelial Na⁺ channel.** *J. Biol. Chem.* 1999, **274**: 9648-9655.
18. Gründer S., Fowler Jaeger N., Gautschi I., Schild L., Rossier BC.: **Identification of a highly conserved sequence at the N-terminus of the epithelial Na⁺ channel α subunit involved in gating.** *Eur. J. Physiol.* 1999, **438**: 709-715.
19. Hartley GS.: **Aqueous solutions of paraffin-chain salts.** Paris: Hermann & Cie. 1936.
20. Tanford C.: **The hydrophobic effect, 1st Edition.** New-York: John Wiley & Sons. 1973.
21. Segrest JP., Jackson RL., Morrisett JD., Gotto AM.: **A molecular theory for protein-lipid interactions in plasma lipoproteins.** *FEBS Lett.* 1974, **38**: 247-253.

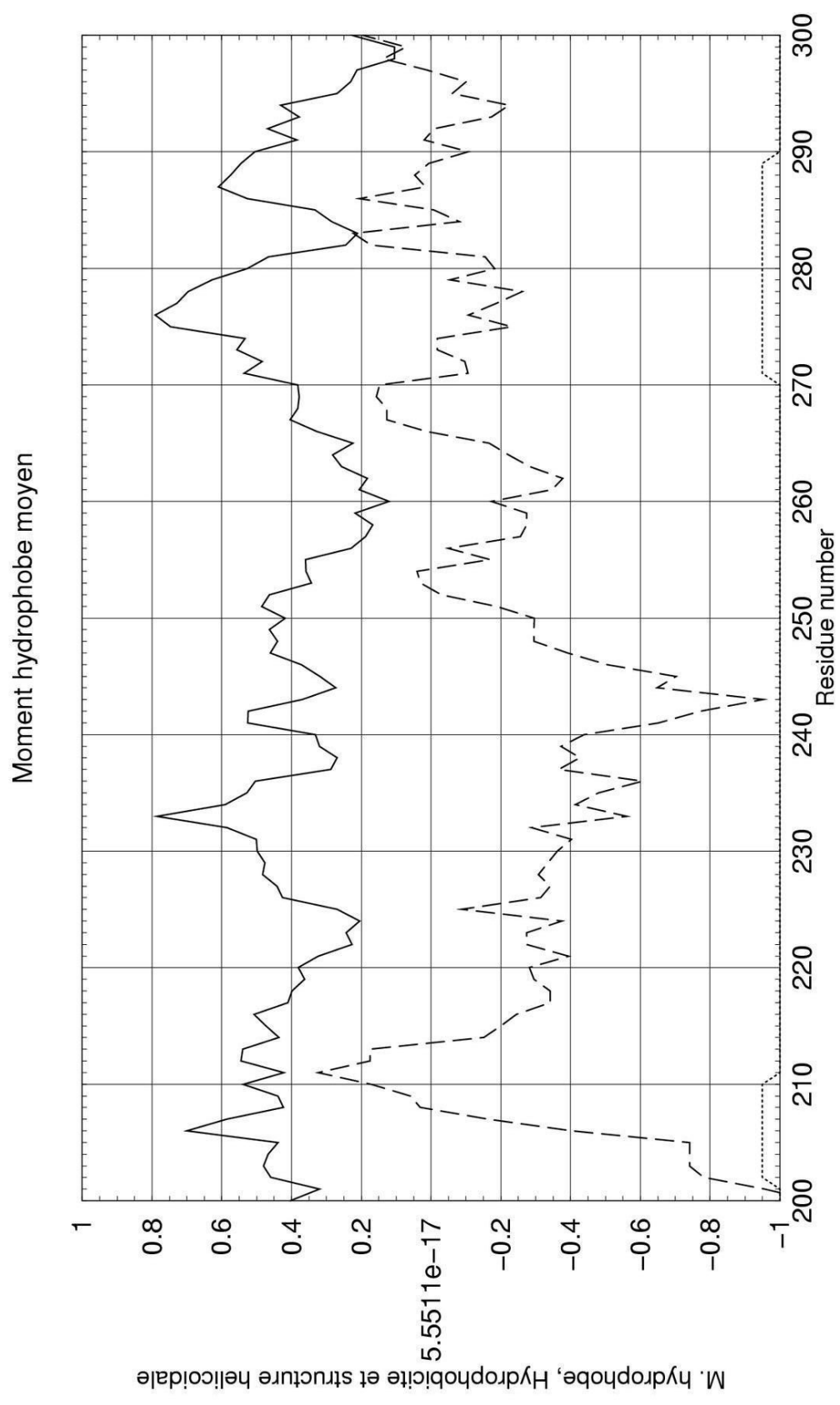
22. Perutz MF., Kendrew JC., Watson HC.: **Structure and function of haemoglobin II. Some relations between polypeptide chain configuration and amino acid sequence.** *J. Mol. Biol.* 1965, **13**: 669-678.
23. Schiffer M., Edmundson AB.: **Use of helical wheels to represent the structures of proteins and to identify segments with helical potential.** *Biophys. J.* 1967, **7**: 121-135.
24. Lim VI.: **Polypeptide chain folding through a highly helical intermediate as a general principle of globular protein structure formation.** *FEBS Lett.* 1978, **89**: 10-14.
25. Rost B.: **PHD: predicting one-dimensional protein structure by profile based neural networks.** *Methods in Enzymology.* 1996, **266**: 525-539.
26. Rost B., Fariselli P., Casadio R.: **Topology prediction for helical transmembrane proteins at 86% accuracy.** *Protein Science.* 1996, **7**: 1704-1718.
27. Tisdall J.: **Beginning Perl for Bioinformatics.** *O'Reilly Edition* 2001.
28. Eisenberg D., Weiss RM., Terwilliger TC.: **The helical hydrophobic moment: a measure of the amphiphilicity of helix.** *Nature.* 1982, **299**: 371-375.
29. Eisenberg D., Schwarz E., Komarony M., Wall R.: **Analysis of membrane and surface protein sequences with the hydrophobic moment plot.** *J. Mol. Biol.* 1984, **179**: 125-142.
30. Nelson RM., Long GL.: **A general method of site-specific mutagenesis using a modification of the *Thermus aquaticus* polymerase chain reaction.** *Anal. Biochem.* 1989, **180**: 147-151.
31. Saiki RK., Gelfand DH., Stoffler S., Scharf SJ., Higuchi R., Horn GT., Mullis KB., Erlich HA.: **Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.** *Science.* 1988, **239**: 487-491.
32. Maniatis T., Fritsch EF., Sambrook J.: **Molecular cloning: A laboratory manual.** *NY: Cold Spring Harbor Laboratory Publications.* 1989.
33. Chung CT., Miller RH.: **A rapid and convenient method for the preparation and storage of competent bacterial cells.** *Nucleic Acids Research.* 1988, **16**: 3580.
34. Melton DA., Krieg PA., Rebagliati MR., Maniatis T., Zinn K., Green MR.: **Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter.** *Nucleic Ac. Res.* 1984, **18**: 7035-7056.
35. Schild L., Canessa CM., Shimkets RA., Gautschi I., Lifton RP., Rossier BC.: **A mutation in the epithelial sodium channel causing Liddle disease increases channel activity in the *Xenopus laevis* oocytes expression system.** *Proc. Natl. Acad. Sci. USA* 1995, **92**: 5699-5703.

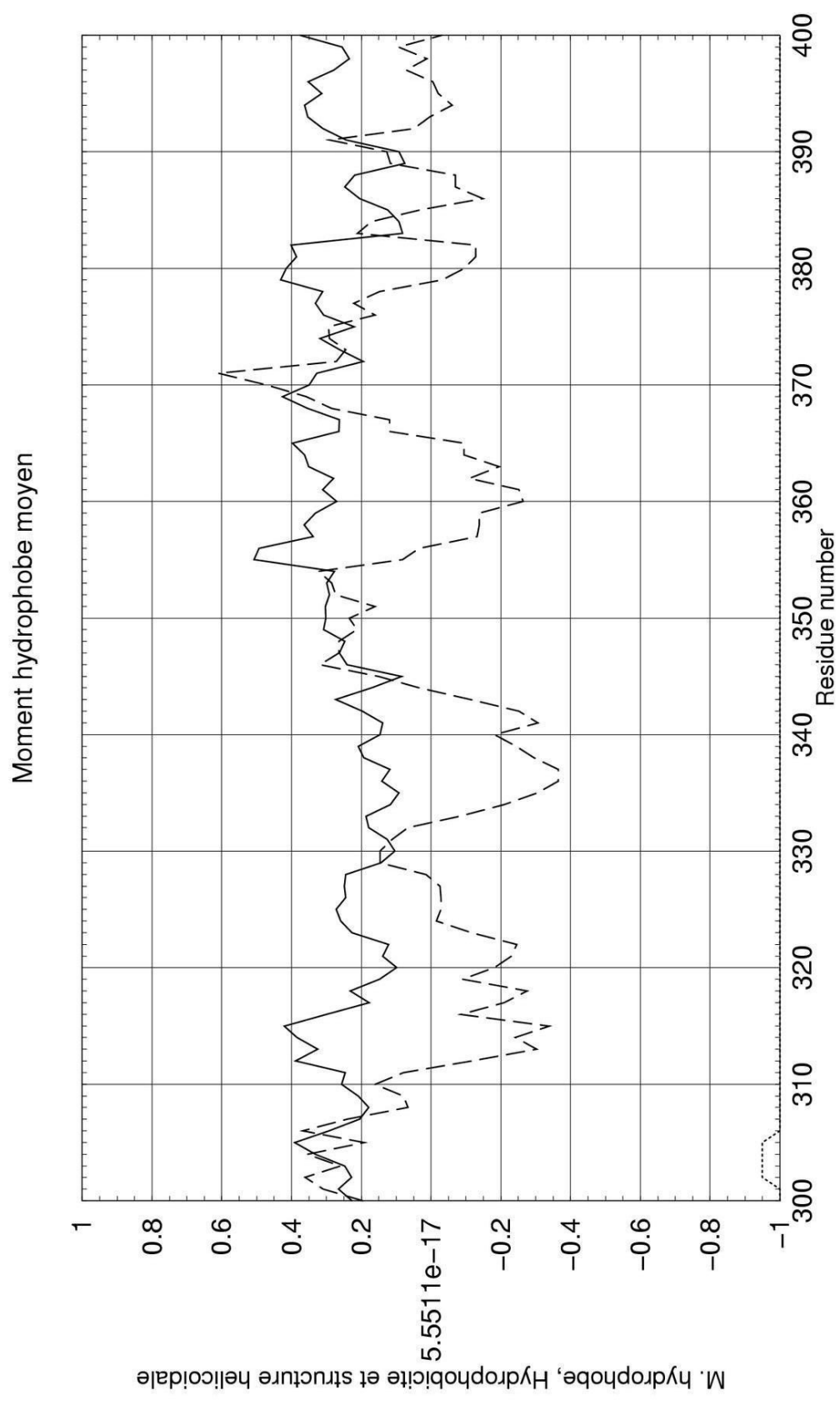
36. Segrest JP., De Loof H., Dohlman JG., Brouillette CG., Anantharamaiah GM.: **Amphipathic helix: classes and properties.** *Proteins: structure, function and genetics.* 1990, **8**: 103-117.
37. Sheng S., Perry CJ., Kleyman TR. : **External nickel inhibits epithelial sodium channel by binding to histidine residues within the extracellular domains of α and γ subunits and reducing channel open probability.** *J. Biol. Chem.* 2002, **277**: 50098-50111.
38. Kellenberger S., Hoffmann-Pochon N., Gautschi I., Schneeberger E., Schild L.: **On the molecular basis of ion permeation in the epithelial Na^+ channel.** *J. Gen. Physiol.* 1999, **114**: 13-30.

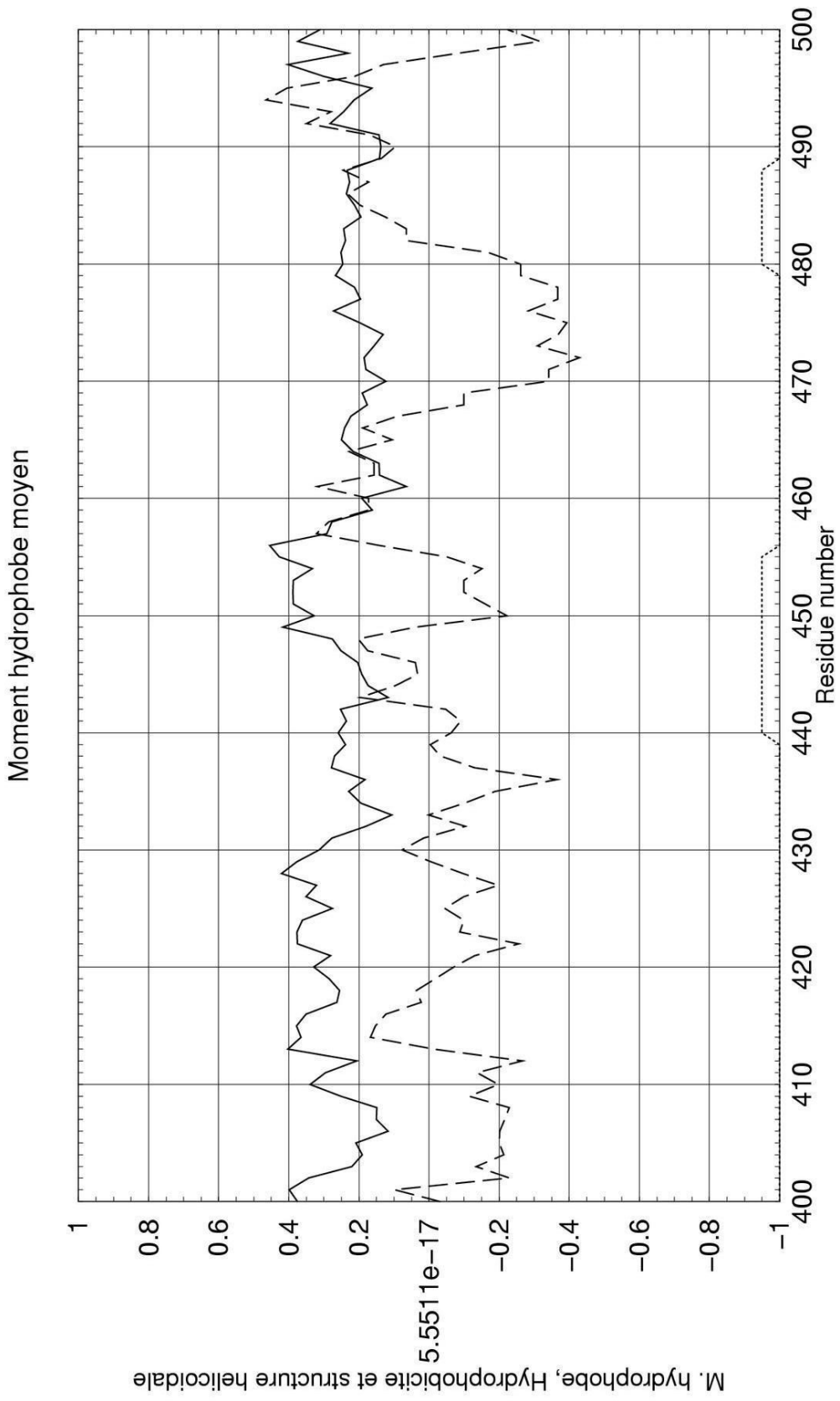
Annexe A : Exemple d'une prédiction amphipathique. α -rENaC, fenêtre de taille 11. Trait continu : moment hydrophobe moyen. Trait interrompu : hydrophobicité moyenne. Trait pointillé : structure hélicoïdale.

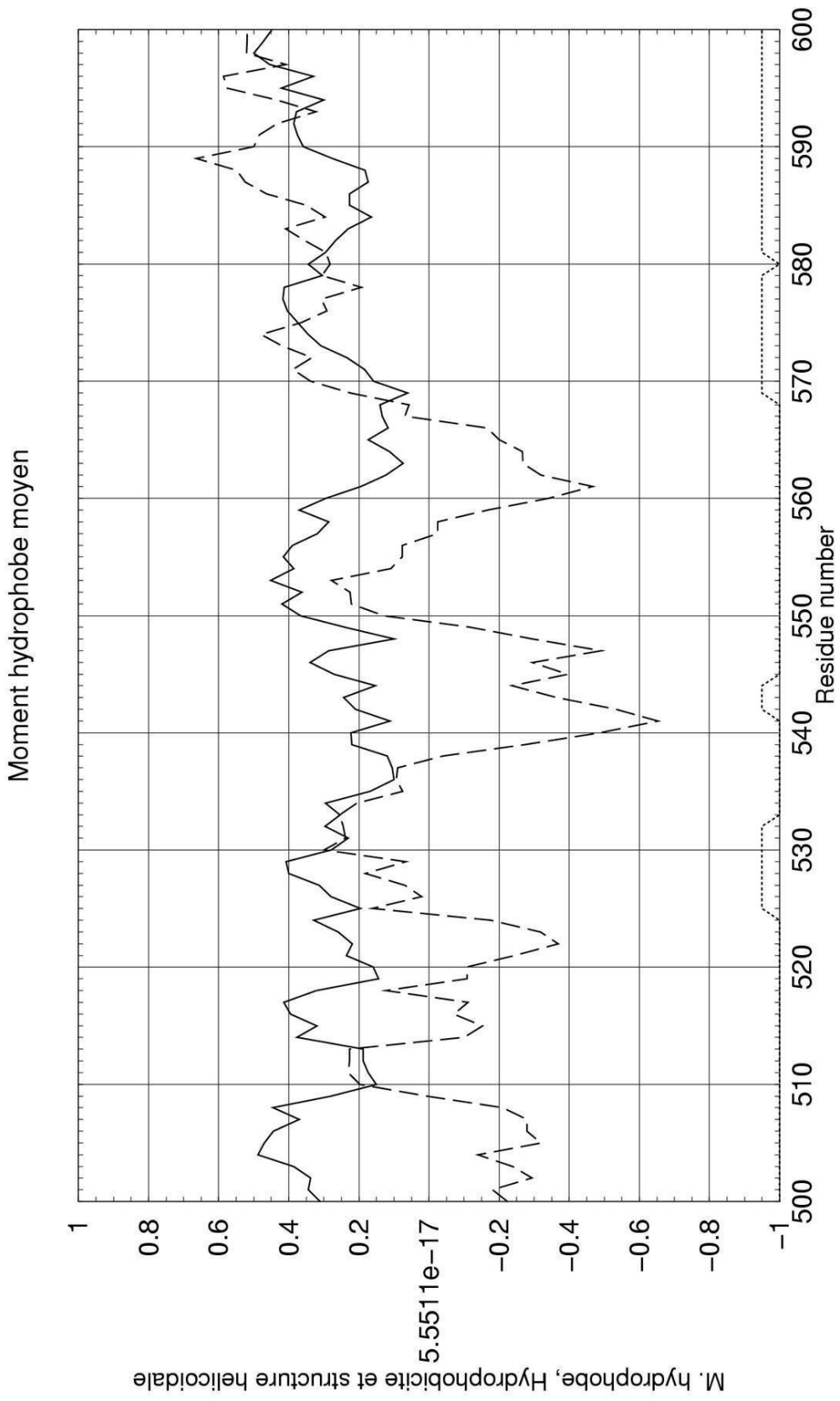


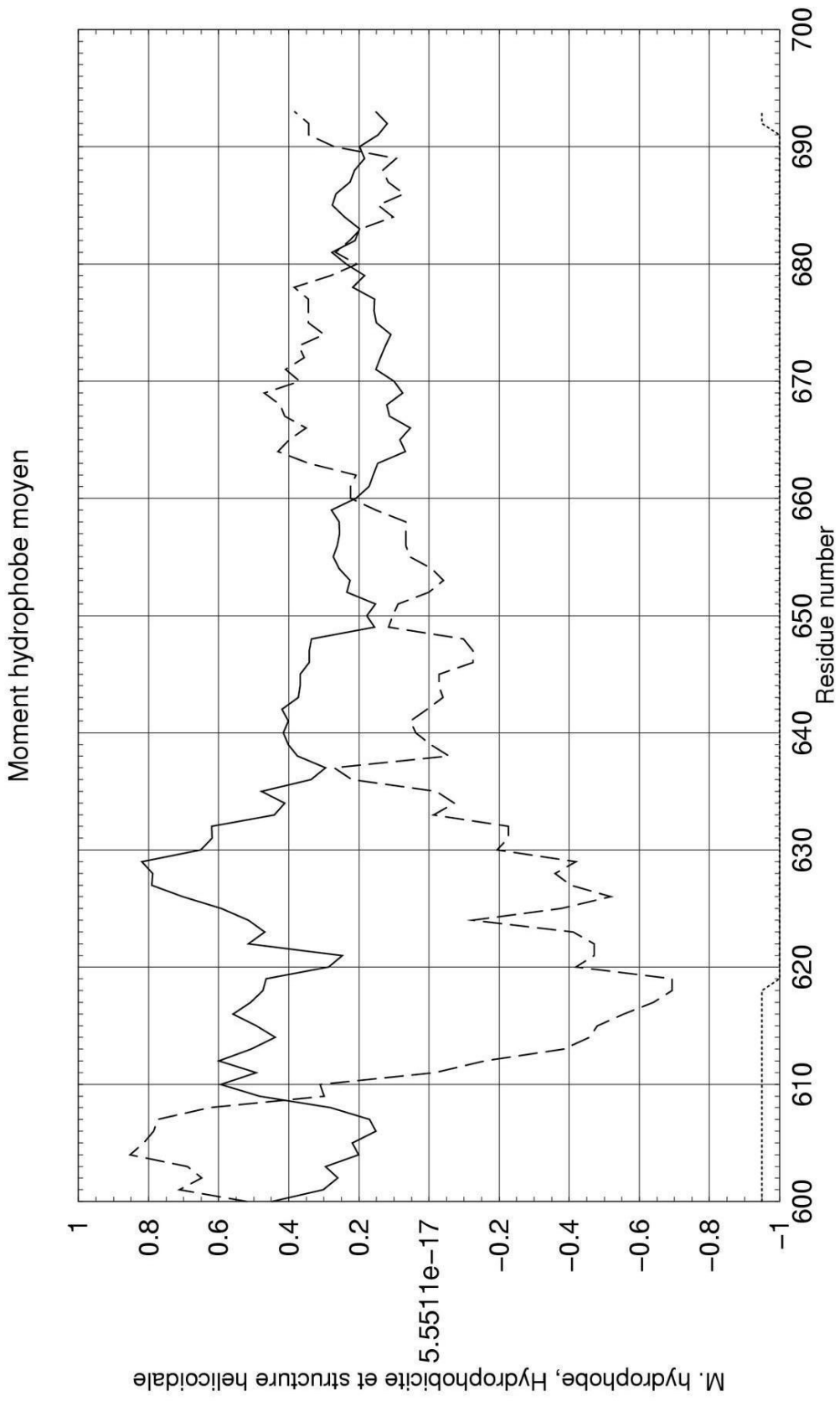




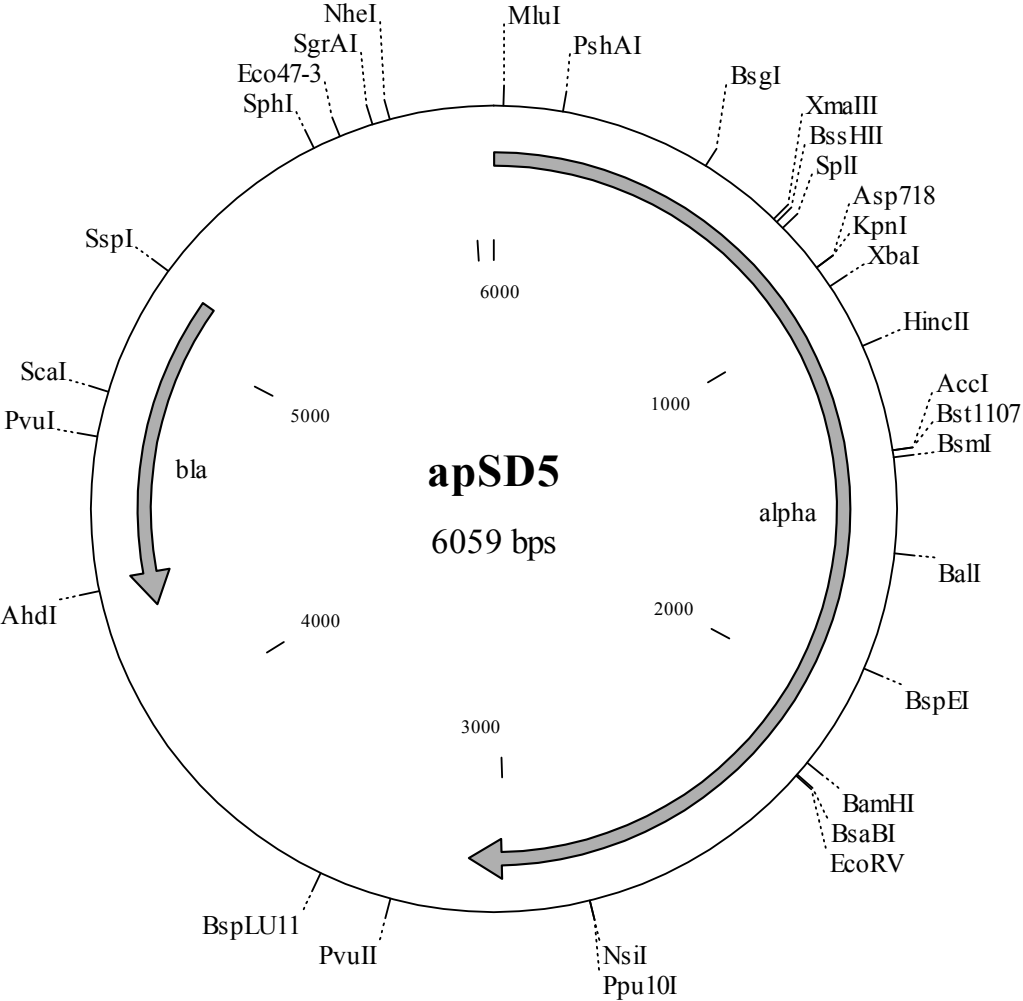








Annexe B : Carte du vecteur d'expression pSD5.



Annexe C : Alignement des séquences mutantes d' α F-rENaC obtenues par mutagenèse dirigée. Les mutations introduites sont soulignées. Les séquences commencent toutes par le site de l'enzyme de restriction *MluI* et se terminent par les sites de *BsgI* (α FH1) et *XbaI* (α FH2 et α FH3) représentés par un soulignement.

α F : ACGCGTCCGCCGGACAGCCCCATTCTGCCTTCACGCTAATGATGCTGGACCACACCAGAGCCCCCTGAGCT

α FH1 : ACGCGTCCGCCGGACAGCCCCATTCTGCCTTCACGCTAATGATGCTGGACCACACCAGAGCCCCCTGAGCT

α FH2 : ACGCGTCCGCCGGACAGCCCCATTCTGCCTTCACGCTAATGATGCTGGACCACACCAGAGCCCCCTGAGCT

α FH3 : ACGCGTCCGCCGGACAGCCCCATTCTGCCTTCACGCTAATGATGCTGGACCACACCAGAGCCCCCTGAGCT

α F : CAACATTGACCTAGACCTTTCACGCCTCCAACTCGCCTAAGGGGTCCATGAAGGGCAACCAATTCAAGGAG

α FH1 : CAACATTGACCTAGACCTTTCACGCCTCCAACTCGCCTAAGGGGTCCATGAAGGGCAACCAATTCAAGGAG

α FH2 : CAACATTGACCTAGACCTTTCACGCCTCCAACTCGCCTAAGGGGTCCATGAAGGGCAACCAATTCAAGGAG

α FH3 : CAACATTGACCTAGACCTTTCACGCCTCCAACTCGCCTAAGGGGTCCATGAAGGGCAACCAATTCAAGGAG

α F : CAAGACCCTTGTCCCTCAGCCCCATGCAAGGACTGGGGAAAGGGGGACAAACGTGAAGAGCAGGGCCTGG

α FH1 : CAAGACCCTTGTCCCTCAGCCCCATGCAAGGACTGGGGAAAGGGGGACAAACGTGAAGAGCAGGGCCTGG

α FH 2 : CAAGACCCTTGTCCCTCAGCCCCATGCAAGGACTGGGGAAAGGGGGACAAACGTGAAGAGCAGGGCCTGG

α FH3 : CAAGACCCTTGTCCCTCAGCCCCATGCAAGGACTGGGGAAAGGGGGACAAACGTGAAGAGCAGGGCCTGG

α F : GCCCGGAACCCCTCAGCACCCCGGCAGCCCCACCGAGGAGGAGGCACTGATTGAATTCCACCGCTCCTA

α FH1 : GCCCGGAACCCCTCAGCACCCCGGCAGCCCCACCGAGGAGGAGGCACTGATTGAATTCCACATCTCCTA

α FH2 : GCCCGGAACCCCTCAGCACCCCGGCAGCCCCACCGAGGAGGAGGCACTGATTGAATTCCACCGCTCCTA

α FH3 : GCCCGGAACCCCTCAGCACCCCGGCAGCCCCACCGAGGAGGAGGCACTGATTGAATTCCACCGCTCCTA

α F : CCGGGAGCTCTTCCAGTTCTTGCAACAACACCACCATCCACGGGGCCATCCGGCCTGGTGTGCTCCAAA

α FH1 : CCGGGAGAGTTCAGTTCTTGCAACAACACCACCATCCACGGGGCCATCCGGCCTGGTGTGCTCCAAA

α FH2 : CCGGGAGCTCTTCCAGTTCTTGCAACAACACCACCATCCACGGGGCCATCCGGCCTGGTGTGCTCCAAA

α FH3 : CCGGGAGCTCTTCCAGTTCTTGCAACAACACCACCATCCACGGGGCCATCCGGCCTGGTGTGCTCCAAA

α F : CACAACCGCATGAAGACGGCCTTCTGGGCGGTGCTGTGGCTGTGCACCTTCGGCATGATGTACTGGCAGT

α FH1 : CACAACCGCATGAAGACGGCCTTCTGGGCGGTGCTGTGGCTGTGCACCTTCGGCATGATGTACTGGCAGT

α FH2 : CACAACCGCATGAAGACGGCCTTCTGGGCGGTGCTGTGGCTGTGCACCTTCGGCATGATGTACTGGCAGT

α FH3 : CACAACCGCATGAAGACGGCCTTCTGGGCGGTGCTGTGGCTGTGCACCTTCGGCATGATGTACTGGCAGT

α F : TCGCCTTGCTGTTTCGAGGAGTACCTCAGCTACCCAGTGAGCCTCAACATCAACCTCAATTTCAGACAAGCT

α FH1 : TCGCCTTGCTGTTTCGAGGAGTACCTCAGCTACCCAGTGAGCCTCAACATCAACCTCAATTTCAGACAAGCT

α FH2 : TCGCCTTGCTGTTTCGAGGAGTACCTCAGCTACCCAGTGAGCCTCAACATCAACCTCAATTTCAGACAAGCT

α FH3 : TCGCCTTGCTGTTTCGAGGAGTACCTCAGCTACCCAGTGAGCCTCAACATCAACCTCAATTTCAGACAAGCT

α F : GGTCTTCCCCTGCCGTCACTGTCTGCACCCCTTAATCCTTACAGATACACTGAAATTAAGAGGAGCTGGAA

α FH1 : GGTCTTCCCCTGCCGTCACTGTCTGCAC

α FH2 : GGTCTTCCCCTGCCGTCACTGTCTGCACCCCTTAATCCTTACAGATACACTGAAATTAAGAGGAGCTCATA

α FH3 : GGTCTTCCCCTGCCGTCACTGTCTGCACCCCTTAATCCTTACAGATACACTGAAATTAAGAGGAGCTGGAA

α F : GAGCTGGACCGCATCACGGAGCAGACGCTTTTTTGACTTGTACAAAATACAACCTTCTCCCTACGACTACAAGG

α FH2 : GAGCTGGACATCATCACGGAGCAGACGCTTTTTTGACTTGTACAAAATACAACCTTCTCCCTACGACTACAAGG

α FH3 : GAGCTGGACCGCATCACGGAGCAGACGCTTTTTTGACTTGTACAAAATACAACCTTCTCCCTACGACTACAAGG

α F : ACGACGATGAGAAAGCGCAGCTCCCCGGCAGCCTCCTGGGTGCTTTCCCGCACCCCCCTGCAGCGCCTGCCGCAC

α FH2 : ACGACGATGACAAGCGCAGCTCCCCGGCAGCCTCCTGGGTGCTTTCCCGCACCCCCCTGCAGCGCCTGCCGCAC

α FH3 : ACGACGATGACAAGCGCAGCTCCCCGGCAGCCTCCTGGGTGCTTTCCCGCACCCCCCTGCAGCGCCTGCCGCAC

α F : TCCACCTCCGCCCTACTCCGGCCGCACGGCGCAGCGGGTCTTCCAGCGTACGCGACAACAATCCCCAA

α FH2 : TCCACCTCCGCCCTACTCCGGCCGCACGGCGCAGCGGGTCTTCCAGCGTACGCGACAACAATCCCCAA

α FH3 : TCCACCTCCGCCCTACTCCGGCCGCACGGCGCAGCGGGTCTTCCAGCGTACGCGACAACAATCCCCAA

α F : GTGGACCCGGAAGGACTGGAAGATCGGGCTTCCAACTGTGCAACCAGAAACAATCAGACTGTTTCTACCCAGA

α FH2 : GTGGACCCGGAAGGACTGGAAGATCGGGCTTCCAACTGTGCAACCAGAAACAATCAGACTGTTTCTACCCAGA

α FH3 : GTGGACCCGGAAGGACTGGAAGATCGGGCTTCCAACTGTGCAACCAGAAACAATCAGACTGTTTCTACCCAGA

α F : CATACTCCTCTGGGGTGGATGCAGTGAGGGAGTGGTACCGCTTCCATTACATCAACAATCTGTCCAGACT

α FH2 : CATACTCCTCTGGGGTGGATGCAGTGAGGGAGTGGTACCGCTTCCATTACATCAACAATCTGTCCAGACT

α FH3 : CATACTCCTCTGGGGTGGATGCAAAGATIGAGTGGTAICGCTTCCATTACATCAACAATCTGTCCAGACT

α F : GTCGGACACCTCGCCCCGCTCTAGA

α FH2 : GTCGGACACCTCGCCCCGCTCTAGA

α FH3 : GTCGGACACCTCGCCCCGCTCTAGA

Annexe D : Code du programme de prédiction des hélices α amphipathiques.

```
#!/usr/bin/perl
# Fasta_PDB_PHD_file_parse => clean
#
# Programs that parse protein Fasta, PDB and PHD files to output hydrophobic
# moment
# hydrophobicity, a.a hydrophobe/phile ratio, Eisenberg cut off and helix structure.
#

# Variables and constants
my$filename = $ARGV[0];
my$win_size = $ARGV[1];
my$graph_opt = $ARGV[2];
my%seq_struct_chains = ();
my@seq_chains = ();
my@struct_chains = ();
my@result_chain = ();
my$result_chain = ();
my@result_chains = ();
my$outputfile = "";
my$file_abrev = "";
my$xmlgr = '/usr/bin/X11/xmlgr';

# Eisenberg scale for hydrophobic moment and hydrophobicity calculations
my%hyscale = (
    "r" => "-2.53",
    "d" => "-0.90",
    "e" => "-0.74",
    "n" => "-0.78",
    "k" => "-1.50",
    "q" => "-0.85",
    "h" => "-0.40",
    "s" => "-0.18",
    "t" => "-0.05",
    "y" => "0.26",
    "g" => "0.48",
    "p" => "0.12",
    "c" => "0.29",
    "a" => "0.62",
    "w" => "0.81",
    "m" => "0.64",
    "f" => "1.19",
    "v" => "1.08",
    "i" => "1.38",
    "l" => "1.06",
);
```

```

# Rearrangement of filename and arguments
chomp $filename;
chomp $win_size;

# Test of command line arguments
unless($ARGV[0] and $ARGV[1] and $ARGV[2]) {
    print "\nEnter filename and window size at the command prompt:
File_parser_95_105 [filename][window size][graph (y/n)]!\n\n";
    exit;
}

# Test for file type and compute accordingly
if ($filename =~ /\.fas$/) {

# Withdrawal of fasta file sequence and structure information
    %seq_struct_chains = screen_fasta($filename);
}

elsif ($filename =~ /\.pdb$/) {

# Withdrawal of pdb file sequence and structure information
    %seq_struct_chains = screen_pdb($filename);
}

elsif ($filename =~ /\.phd$/) {

# Withdrawal of phd file sequence and structure information
    %seq_struct_chains = screen_phd($filename);
}
else {
    print "\nYour format file is incompatible with this program !! \n\n";
    exit;
}

# Extraction of sequence and structure informations for future amphiphilic and
# structure computation
@seq_chains = keys %seq_struct_chains;
@struct_chains = values %seq_struct_chains;

# Print out the arrays for control
for (my$i = 0; $i<scalar@seq_chains; $i++) {
    print "\n\n Sequence $i:\n";
    print $seq_chains[$i], "\n", $struct_chains[$i], "\n\n";
}

# Amphiphilic and structure prediction over sequences
for (my$i = 0; $i<scalar@seq_chains; $i++) {
    @result_chain = amphiphilic_pred($seq_chains[$i], $struct_chains[$i], $win_size);
}

```

```

$result_chain = join('Z', @result_chain);
$result_chains[$i] = $result_chain;
}

# Moment_score file construction and launching of graphical interface XMGR
for(my$i = 0; $i<scalar@result_chains; $i++) {
    @result_chain = split('Z', $result_chains[$i]);
    $file_abbrev = $i.$win_size.$filename;
    $outfile =
"/home/devaud/Proteins_databases/Program_output/Moment_score_$file_abbrev";
    open(MOMENTHYDROPHOB_SCORE, ">$outfile") or die "Could not open
file \"$outfile\" to write to !\n";
    print MOMENTHYDROPHOB_SCORE "@result_chain";
    close(MOMENTHYDROPHOB_SCORE);
    if($graph_opt =~ /y/) {
        system("$xmgr -param /home/devaud/Perl/Xmgr_param/Hydromom.param -nxy
/home/devaud/Proteins_databases/Program_output/Moment_score_$file_abbrev &");
    }
}
exit;

#####
#####
#                               Subroutines
#####
#####

# amphiphilic_pred
#####
# -given a protein sequence, a score option and a window size return an array with
# a.a position, mean
# hydrophobicity, Eisenberg cut off and hydrophobic moment.
#

sub amphiphilic_pred {

    use strict;
    use warnings;

    # Declare variables and constants
    my($prot_seq, $prot_struct, $win_size) = @_ ;
    my$length_space = 0;
    my$low_prot_seq = "";
    my$length_prot_seq = 0;
    my$win_seq = "";
    my$win_hydro_moment = 0;
    my$win_angle_deg_optim = 0;
    my$win_mean_hydrophob = 0;
    my$win_cut_off = 0;
    my$pos_struct = 0;

```

```

my$win_pho_phy_ratio = 0;
my$pos = 0;
my$win_result = "";
my@results = ();

# Clean off protein sequence and protein structure prediction
$prot_seq =~ s/(\s+)$/;/;
$length_space = length($1);
$prot_struct =~ s/\s{$length_space}$/;/;

# Change upperclass character in lowerclass character
$low_prot_seq = class_conv($prot_seq);

# Screening of the protein for hydrophobic moment computation of each window
$length_prot_seq = length $low_prot_seq;
for (my$p = int($win_size/2); $p<$length_prot_seq - int($win_size/2); ++$p){
    $win_seq = substr($low_prot_seq, $p - int($win_size/2), $win_size);
    ($win_hydro_moment, $win_angle_deg_optim) =
hydro_moment($win_seq);
    $win_mean_hydrophob = mean_hydrophob($win_seq);
    $pos_struct = pos_struct($prot_struct, $p);
    $pos = $p+1;
    $win_result = "$pos $win_hydro_moment $win_mean_hydrophob
$pos_struct $win_angle_deg_optim\n";
    push (@results, $win_result);
}
return @results;
}

# class_conv
#####
#
# -given a protein sequence return that sequence in lower class character
#

sub class_conv {
    my($prot_seq) = @_;
    $prot_seq =~ tr/RDENKQHSTYGPCAWMFVIL/rdenkqhstygpcawmfvil/;
    if ($prot_seq =~ /[bjouxz]/) {
        print "A letter in the protein sequence doesn't correspond to an amino acid ! \n";
        exit;
    }
    return $prot_seq;
}

# call_stride
#####
#
# -given a PDB filename, return the output from the "stride"

```

```

# secondary structure prediction program
#

sub call_stride {

    use strict;
    use warnings;

    # Declare variables and constants
    my($pdb_filename) = @_;
    my$stride = 'stride';
    my$stride_option = '-o';
    my@results = ();

    -e $pdb_filename or die "File \"$pdb_filename\" doesn't seem to exist in current
    directory !\n";

    # Start up the program, capture and return the output
    @results = `sstride $stride_option $pdb_filename`;
    return @results;
}

# hydro_moment
#####
#
# -given a protein sequence return its hydrophobic moment
#

sub hydro_moment {

    # Declare variables and constants
    my($win_seq) = @_;
    my@aa = ();
    my$sin = 0;
    my$cos = 0;
    my$hydro_moment = 0;
    my$hydro_moment_optim = 0;
    my$angle_deg_optim = 0;

    # Core
    @aa = split("",$win_seq);
    for(my$angle_deg = 95; $angle_deg<=105;$angle_deg++){
        for(my$i = 0; $i < scalar@aa; $i++){
            my$j = $i+1;
            $sin = $sin+($hyscale{$Saa[$i]}*(sin rad($j*$angle_deg)));
            $cos = $cos+($hyscale{$Saa[$i]}*(cos rad($j*$angle_deg)));
        }
        $hydro_moment = (sqrt(($sin*$sin)+($cos*$cos)))/scalar@aa;
        if($hydro_moment>$hydro_moment_optim) {

```

```

        $hydro_moment_optim = $hydro_moment;
        $angle_deg_optim = $angle_deg;
    }
    $sin = 0;
    $cos = 0;
}
return ($hydro_moment_optim, $angle_deg_optim);
}

# mean_hydrophob
#####
#
# -given a protein sequence return its mean hydrophobicity
#

sub mean_hydrophob {

    # Declare variables and constants
    my($win_seq) = @_;
    my@aa = ();
    my$hydrophob = 0;
    my$mean_hydrophob = 0;

    # Core
    @aa = split(",",$win_seq);
    for(my$i = 0; $i < scalar@aa; $i++){
        $hydrophob = $hydrophob+($hyscale{$Saa[$i]});
    }
    $mean_hydrophob = $hydrophob/scalar@aa;
    return $mean_hydrophob;
}

# parse_raw
#####
#
# -given an array raw output, extract the primary sequence and the
# secondary structure prediction returning them in a hash with:
# keys = sequence/values = structure
# prediction.
#

sub parse_raw {

    use strict;
    use warnings;

    # Declare variables and constants
    my(@raw_output) = @_;
    my$seq = "";
    my$str = "";

```

```

my%seq_struct = ();

# Extract the lines of sequence and structure and process them to discard all the
spaces
foreach(@raw_output) {
    if($_ =~ /^AA/) {
        $seq = substr($_, 11, ((length$_) -11));
        $seq =~ s/\s//g;
    }
    if($_ =~ /^PROF_sec/) {
        $str = substr($_, 11, ((length$_) -11));
    }
}
# Create sequence/structure hash
$seq_struct{$seq} = $str;
return %seq_struct;
}

# parse_stride
#####
#
# -given an array stride output, extract the primary sequence and the
# secondary structure prediction returning them in a hash with keys = sequence and
# values = structure
# prediction.
#

sub parse_stride {

    use strict;
    use warnings;

    # Declare variables and constants
    my(@stride_report) = @_;
    my$last_line_residu_start = 0;
    my$line_residu_start = 0;
    my$chain = 0;
    my%seq_chains = ();
    my$pos = 0;
    my$str = "";
    my%seq_struct_chains = ();

    # Extract the lines of interest
    my@seq = grep(/^SEQ /, @stride_report);
    my@str = grep(/^STR /, @stride_report);

    # Process SEQ lines to discard all but the sequence and create sequence hash
    # according to chains
    for (@seq) { $_ = substr($_, 5, 55) }
    for (@str) { $_ = substr($_, 10, 50);}

```

```

for (@seq) {
    $line_residu_start = substr($_, 0, 4);
    $line_residu_start =~ s/\s//g;
    if ($line_residu_start > $last_line_residu_start) {
        if (defined $seq_chains{$chain}) {
            $seq_chains{$chain} .= substr($_, 5, 50);
        }
        else {
            $seq_chains{$chain} = substr($_, 5, 50);
        }
    }
    else {
        $chain++;
        $seq_chains{$chain} = substr($_, 5, 50);
    }
}
$last_line_residu_start = $line_residu_start;
}

# Process STR lines and create structure hash matching sequence hash
$str = join(" ", @str);
for (my$i = 0; $i < scalar keys %seq_chains; $i++) {
    if (defined $seq_struct_chains{$seq_chains{$i}}) {
    }
    else {
        $seq_struct_chains{$seq_chains{$i}} = substr($str, $pos,
(length $seq_chains{$i})-1);
    }
    $pos = $pos + length $seq_chains{$i};
}
return (%seq_struct_chains);
}

# pos_struct
#####
#
# -given a position in protein sequence return its numerical helical structure
# prediction (0: non helical; 1: helical)
#

sub pos_struct {

    # Declare variables and constants
    my($prot_struct, $pos) = @_;
    my$struct = "";
    my$num_struct = 0;

    # Core
    $struct = substr($prot_struct, $pos, 1);
    if($struct =~ /H/) { $num_struct = -0.95}
    else { $num_struct = -1}
}

```



```

    return $num_struct;
}

# rad
#####
#
# -given an angle in degree return that angle in radian
#

sub rad {

    # Declare variables and constants
    my($angle_deg) = @_;
    my$angle_rad = 0;

    # Core
    $angle_rad = $angle_deg*0.017453293;
    return $angle_rad;
}

# raw_screen
#####
#
# -given a PHD filename, return the phd file data containing sequence and
# secondary structure prediction
#

sub raw_screen {

    use strict;
    use warnings;

    # Declare variables and constants
    my(@phd_file) = @_;
    my$start_line = 0;
    my$end_line = 0;
    my@results = ();

    # Parse roughly the file, capture and return the raw output
    for(my$i = 0; $i<scalar@phd_file; $i++) {
        if($phd_file[$i] =~ /^PROF results \(\normal\)|^PHD results \(\normal\)/)
        {$start_line = $i}
        if($phd_file[$i] =~ /^SUB_sec/) {$end_line = $i}
    }
    for(my$i = $start_line; $i<=$end_line; $i++) {
        push(@results, $phd_file[$i]);
    }
    return @results;
}

```

```

# screen_fasta
#####
#
# -given a fasta file return a sequence/structure prediction containing hash
# (according to empty structure information of such format file the values
# of the hash are noted as x)
#

sub screen_fasta {

    use strict;
    use warnings;

    # Declare variables and constants
    my($fasta_filename) = @_ ;
    my@fasta_file = ();
    my$seq = "";
    my$struct = "";
    my%seq_struct = ();

    # Test and opening of fasta file
    -e $fasta_filename or die "File \"$fasta_filename\" doesn't seem to exist in current
directory !\n";
    open(FASTAFILENAME, $fasta_filename) or die "Could not open
\"$fasta_filename\" !\n";
    @fasta_file = <FASTAFILENAME>;
    close(FASTAFILENAME);

    # Parse the file and extract sequence informations
    $seq = join("", @fasta_file);
    $seq =~ s/\s//g;
    $struct = "X" x(length$seq);
    $seq_struct{$seq} = $struct;
    return %seq_struct;
}

# screen_pdb
#####
#
# -given a PDB file return a sequence/structure prediction containing hash
#

sub screen_pdb {

    use strict;
    use warnings;

    # Declare variables and constants
    my($pdb_filename) = @_ ;

```

```

my@stride_output = ();
my%seq_struct_chains = ();

# Test and opening of pdb file
-e $pdb_filename or die "File \"$pdb_filename\" doesn't seem to exist in current
directory !\n";
open(PDBFILENAME, $pdb_filename) or die "Could not open \"$pdb_filename\"
!\n";
close(PDBFILENAME);

# Call "stride" on pdb file and collect the report
@stride_output = call_stride($pdb_filename);

# Parse the stride report and class primary sequence and secondary
# structure prediction in a hash
%seq_struct_chains = parse_stride(@stride_output);
return %seq_struct_chains;
}

# screen_phd
#####
#
# -given a PHD file return a sequence/structure prediction containing hash
#

sub screen_phd {

    use strict;
    use warnings;

    # Declare variables and constants
    my($phd_filename) = @_ ;
    my@phd_file = ();
    my@raw_output = ();
    my%seq_struct = ();

    # Test and opening of phd file
    -e $phd_filename or die "File \"$phd_filename\" doesn't seem to exist in current
directory !\n";
    open(PHDFILENAME, $phd_filename) or die "Could not open \"$phd_filename\"
!\n";
    @phd_file = <PHDFILENAME>;
    close(PHDFILENAME);

    # Screen phd file and collect the line of interest
    @raw_output = raw_screen(@phd_file);

    # Parse the raw output and class primary sequences and corresponding secondary
    # structures prediction in a hash
    %seq_struct = parse_raw(@raw_output);

```

```
    return %seq_struct;  
}
```