

République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة أبو بكر بلقايد – تلمسان
Université ABOUBEKR BELKAID – TLEMCEM
كلية علوم الطبيعة والحياة، وعلوم الأرض والكون
Faculté des Sciences de la Nature et de la Vie, et Sciences de la Terre et de l'Univers
Département de BIOLOGIE



MÉMOIRE

Présenté par

KOURAK Widad

En vue de l'obtention du

Diplôme de MASTER

En Biochimie Appliquée

Thème

ANALYSE COMPARATIVE DE DEUX GROUPES DE LA FAMILLE DES CONNEXINES

Soutenu le 27/06/2023, devant le jury composé de :

Président	Dr. Meriem BENMANSOUR	Maître de conférences	Université de Tlemcen
Encadrant	Dr. Amal BRIKCI NIGASSA	Maître de conférences	Université de Tlemcen
Examineur	Dr. Amine BEDRANE	Maître assistant	Université de Tlemcen

Année universitaire 2022/2023

Dédicaces

À mes chers parents, vous êtes ma source de soutien inébranlable, mes piliers dans les moments de doute et ma motivation constante. Votre amour et votre encouragement inconditionnels ont été essentiels pour mener à bien ce travail.

À mes chères sœurs, Naoual et Chahrazed, votre amour et votre précieuse présence sont mes sources constantes de réconfort. Je suis infiniment reconnaissante d'avoir des sœurs aussi extraordinaires que vous.

À mes chers frères, Diden, Mohammed-Rédha, Saddek, Mokhtar et Houssame eddine, pour leur appui et leur encouragement permanents. Merci d'avoir été les frères exceptionnels que vous êtes, et que ce travail puisse être une source de fierté partagée entre nous.

À toute la famille KOURAK.

À mes précieuses amies.

Et finalement, à la belle Linda.

Remerciements

En préambule à ce mémoire, je souhaite exprimer ma gratitude à Allah tout Puissant pour m'avoir accordé la patience et le courage nécessaires durant ces longues années d'études.

Ce travail n'aurait pas pu voir le jour sans l'aide et l'encadrement du Dr Amal BRIKCI NIGASSA, que je tiens à remercier chaleureusement pour la qualité exceptionnelle de son encadrement, sa patience et sa disponibilité tout au long de la réalisation de ce mémoire.

Je tiens également à remercier profondément Dr Meriem BENMANSOUR pour l'honneur qu'elle me fait en acceptant de présider ce jury.

Mes sincères remerciements vont également au Dr Amine BEDRANE d'avoir accepté d'examiner et de discuter ce travail.

Je tiens à adresser mes remerciements les plus sincères à tous mes professeurs pour leurs efforts exceptionnels, qui ont grandement contribué à mon parcours académique.

Enfin, mes profonds remerciements vont également à toutes les personnes qui m'ont apporté leur aide et leur soutien, que ce soit de près ou de loin.

Merci du fond du cœur à tous et à toutes pour votre contribution à la réalisation de ce mémoire.

تركزت دراستنا على الكونكسينات، وهي بروتينات غشائية تدخل في تكوين قنوات ناقلة بين الخلايا. قمنا بتحليل التوبولوجيا الغشائية ومحاذاة التسلسلات للمجموعتين الرئيسيتين، ألفا وبيتا، المعبر عنهما في الأنسجة الظهارية، باستخدام أدوات معلوماتية (بروتسكال، دلتا جي بريدكتور، توبكونز و كلاستال أوميغا). أكد التحليل الخصائص المتوقعة للكونكسينات، كالمجالات المحفوظة مثل المناطق الغشائية الكارهة للماء وعناصر السيستين في كل حلقة خارجية، والمجالات غير المحفوظة بما فيها الحلقة الداخلية والنهاية الكاربوكسيلية. حددنا أنماطاً خاصة لكل مجموعة ترتبط بالتوافق الاختلاطي و تتحكم في تكوين القنوات النصفية. بالإضافة إلى ذلك، قمنا بتنفيذ محاكاة عن طريق نيتفوس 1.3 سمحت لنا توقع مواقع الفسفرة، بما في ذلك المواقع المؤكدة علمياً (S255، S262، S279، S282، S297، S368 و S372) والجديدة غير مؤكدة (S219، S364، S365 و S369).

كلمات مفتاحية: الكونكسينات البشرية، محاذاة متعددة لتسلسلات الأحماض الأمينية، توقع مواقع الفسفرة، الكينازات، توبولوجيا الغشاء.

ANALYSE COMPARATIVE DE DEUX GROUPES DE LA FAMILLE DES CONNEXINES

Notre étude s'est concentrée sur les connexines, des protéines transmembranaires impliquées dans la formation des jonctions communicantes. Nous avons analysé la topologie membranaire et les alignements de séquences de deux principaux groupes, Alpha et Beta, exprimés dans le tissu épithélial, en utilisant des outils de bioinformatique (ProtScale, ΔG Predictor, TOPCONS et Clustal Omega). L'analyse a confirmé les caractéristiques attendues des connexines, telles que la conservation des régions transmembranaires hydrophobes et des résidus de cystéine des domaines extracellulaires et la divergence des boucles intracellulaires et domaines C-terminaux. Nous avons identifié des motifs spécifiques à chaque groupe associés à la compatibilité hétéromérique. De plus, nous avons prédit des sites de phosphorylation par NetPhos3.1, y compris des sites confirmés expérimentalement (S255, S262, S279, S282, S297, S368 et S372) et nouveaux non-confirmés (S219, S364, S365 et S369).

Mots clés : connexines humaines, alignement de séquences multiple, prédiction des sites de phosphorylation, kinases, topologie membranaire

COMPARATIVE ANALYSIS OF TWO GROUPS WITHIN THE CONNEXIN FAMILY

Our study focused on connexins, transmembrane proteins involved in the formation of gap junctions. We analyzed the membrane topology and sequence alignments of two main groups, Alpha and Beta, expressed in epithelial tissue, using bioinformatics tools (ProtScale, ΔG Predictor, TOPCONS and Clustal Omega). The analysis confirmed the expected characteristics of connexins, such as the conservation of hydrophobic transmembrane regions and cysteine residues in the extracellular domains, as well as the divergence of intracellular loops and C-terminal domains. We identified group-specific motifs associated with heteromeric compatibility. Additionally, we predicted phosphorylation sites using NetPhos3.1, including experimentally confirmed (S255, S262, S279, S282, S297, S368 and S372) and novel sites (S219, S364, S365 and S369).

Mots clés : Human connexins, multiple sequence alignment, prediction of phosphorylation sites, kinases, membrane topology.

Liste des figures

Figure 1.1 : Structure de la connexine, du connexon et de la jonction <i>GAP</i>	6
Figure 1.2 : Structures des gènes des connexines	9
Figure 1.3 : Structure topologique d'une connexine.....	10
Figure 1.4 : Différentes combinaisons possibles de connexons.....	13
Figure 1.5 : Synthèse, assemblage et dégradation des jonctions <i>gap</i>	15
Figure 1.6 : Sites de phosphorylation dans le domaine C-terminale de la Cx43 et ses kinases respectives	18
Figure 2.1 : Le format Genbank/DDBJ.....	25
Figure 2.2 : Le format FASTA.....	26
Figure 2.3 : L'alignement global de deux séquences protéiques	29
Figure 2.4 : L'alignement local de deux séquences protéiques.....	29
Figure 2.5 : Alignement multiple des 21 isoformes des connexines humaines.....	30
Figure 3.1 : Interface web de la base de données UniProtKB/SWISS-PROT	36
Figure 3.2 : Séquence de la connexine 43	38
Figure 3.3 : Interface de ProtScale d'ExpAsy.....	39
Figure 3.4 : Interface de ΔG predictor 1.0.....	40
Figure 3.5 : Interface de TOPCONS	41
Figure 3.6 : Interface web du programme Clustal Omega	42
Figure 3.7 : Interface d'outil Jalview	43
Figure 3.8 : Code couleur pour le pourcentage d'identité dans Jalview.....	44
Figure 3.9 : Interface de NetPhos 3.1	44

Figure 3.10: Interface de Protter.....	45
Figure 4.1 : Profils d'hydropathie de la Cx43 et la Cx26 par ProtScale	47
Figure 4.2 : Diagrammes d'hydrophobicité de la Cx43 et la Cx26 par ΔG predictor 1.0 ..	47
Figure 4.3 : Topologie membranaire prédite de la Cx43 et la Cx26 par TOPCONS	48
Figure 4.4 : Représentation en domaines de la Cx43 et la Cx26	49
Figure 4.5 : Alignement multiple des isoformes Cx37, Cx40, Cx43 et Cx46	51
Figure 4.6 : Matrice du pourcentage d'identité des isoformes Cx37, Cx40, Cx43 et Cx46	52
Figure 4.7 : Alignement multiple des isoformes Cx26, Cx30, Cx30.3, Cx31, Cx31.1 et Cx32	54
Figure 4.8 : Matrice de pourcentage d'identité des isoformes Cx26, Cx30, Cx30.3, Cx31, Cx31.1 et Cx32.....	54
Figure 4.9 : Visualisation de la topologie et des sites de phosphorylation de la Cx43	58
Figure 4.10 : Visualisation de la topologie et des sites de phosphorylation de la Cx26....	59

Liste des tableaux

Tableau 1.1 : Classification des connexines humaines et murines	8
Tableau 2.1 : Classification, contenu et adresses des banques de données les plus connues	23
Tableau 2.2 : Les programmes d'alignement de séquences les plus couramment utilisés .	31
Tableau 2.3 : Les outils de prédiction des modifications post-traductionnelles les plus couramment utilisés.....	32
Tableau 3.1 : Expression tissulaire des connexines humaines	35
Tableau 3.2 : Nomenclature des dix isoformes des connexines humaines selon UniProtKB	37
Tableau 3.3 : Échelle d'hydrophatie de Kyte et Doolittle (1982)	39
Tableau 4.1 : Tableau récapitulatif des prédictions des trois programmes	49
Tableau 4.2 : Prédiction des sites de phosphorylation de la Cx43 par Netphos3.1	55
Tableau 4.3 : Prédiction des sites de phosphorylation de la Cx26 par Netphos3.1.....	57

Liste des abréviations

3'UTR	Région 3' Non Traduite
5'UTR	Région 5' Non Traduite
ADN	Acide Désoxyribonucléique
AMPc	Adénosine-3',5'-Monophosphate Cyclique
ATP	Adénosine Triphosphate
CAM	Cell Adhesion Molecules
Cdc2	Cell Division Control 2
Cdk5	Cyclin -Dependent Kinase 5
CKI	Caséine Kinase 1
CKII	Caséine Kinase 2
DNA-PK	DNA-Dependent Protein Kinase
EGFR	Récepteur du Facteur de Croissance Epidermique
GMPc	Guanosine Monophosphate Cyclique
GSK3	Glycogène Synthase Kinase 3
hCx	Connexine Humaine
Inx	Innexine
IP₃	Inositol Triphosphate
MAPK	Mitogen-Activated Protein Kinase
mCx	Connexine Murine
Panx	Pannexine
PKA	Protéine Kinase A
PKB	Protéine Kinase B
PKC	Protéine Kinase C
PKG	Protéine Kinase G
RSK	Kinase S6 Ribosomique
Src	Sarcoma

Table des matières

Introduction	1
Partie 1 : Synthèse Bibliographique	4
Chapitre I : Connexines et jonctions communicantes.....	5
1. Introduction	6
2. La structure des jonctions communicantes.....	6
3. Les connexines	7
3.1 Expression génique des connexines	7
3.2 Structure des connexines.....	9
3.3 Distribution tissulaire des connexines.....	11
3.4 Cycle de vie des connexines	12
3.4.1 Synthèse et assemblage	12
3.4.1.1 Compatibilité entre connexines	13
3.4.2 Transport	14
3.4.3 Intégration à la membrane plasmique et formation de la jonction communicante	14
3.4.4 Dégradation des canaux	14
4. Fonctions des jonctions communicantes	16
4.1 Couplage électrique.....	16
4.2 Couplage métabolique.....	16
5. Régulation de la communication jonctionnelle.....	17
5.1 Concentration du calcium intracellulaire	17
5.2 pH intracellulaire.....	17
5.3 Phosphorylation des connexines	17
5.3.1 Phosphorylation de la Connexine 43.....	18
6. Jonctions gap et pathologies	19

Chapitre II - La bioinformatique.....	20
1. Introduction	21
2. La bioinformatique	21
3. Les banques de données biologiques	22
4. Les différents formats de fichiers en bioinformatique.....	24
4.1 Les formats propres aux bases de données	24
4.2 Les formats utilisés par les outils d'analyse des séquences	26
5. Programmes d'analyse des séquences protéiques	27
5.1 Méthodes de prédiction de la localisation des domaines protéiques.....	27
5.1.1 Les méthodes de prédiction des segments transmembranaires	27
5.1.2 Les méthodes de détection du peptide signal	28
5.2 Alignement des séquences	28
5.2.1 Alignement global	28
5.2.2 Alignement local	29
5.2.3 Alignement par paires	29
5.2.4 Alignement multiple.....	29
5.3 Étude des modifications post-traductionnelles	31
Partie 2 : Travaux expérimentaux et Résultats	33
Chapitre III : Matériel et Méthodes	34
1. Choix de Connexines.....	35
2. Étude de la topologie membranaire de la Cx43 et la Cx26	38
2.1 ProtScale	38
2.2 ΔG predictor 1.0.....	40
2.3 TOPCONS	40
3. Alignement des séquences multiple des 10 isoformes	41
3.1 Alignement des séquences avec Clustal Omega	41
3.2 Édition de l'alignement avec Jalview version 2.11.2.6.....	43
4. Étude des modifications post-traductionnelles : La phosphorylation avec NetPhos3.1	44
5. Visualisation de la topologie membranaire et les sites de phosphorylation des connexines	45

Chapitre IV : Résultats et Discussion	46
1. Analyse de la topologie membranaire de la Cx43 et la Cx26	47
1.1 Résultats de Protscale.....	47
1.2 Résultats de ΔG predictor 1.0	47
1.3 Résultats de TOPCONS	48
1.4 Prédiction de la topologie de la Cx43 et la Cx26 par les trois méthodes.....	48
2. Alignement des séquences multiple des 10 isoformes	49
2.1 Alignement des connexines Alpha.....	49
2.2 Alignement des connexines Bêta	52
3. Prédiction des sites potentiels de phosphorylation	55
3.1 Sites de phosphorylation prédits de la Cx43	55
3.2 Sites de phosphorylation prédits de la Cx26.....	57
4. Visualisation des domaines et sites prédits	58
Conclusion	60
Annexes	63
Références Bibliothèques	67

Introduction

Chez les organismes pluricellulaires, les cellules se différencient et se spécialisent afin d'accomplir des fonctions spécifiques. Cependant, pour que l'organisme fonctionne correctement et maintienne l'homéostasie, ces cellules spécialisées doivent communiquer et travailler ensemble. La communication directe entre elles leur permet d'échanger des informations, de coordonner leurs activités physiologiques et de répondre aux stimuli externes.

Parmi les différents mécanismes de communication cellulaire, les jonctions communicantes occupent une place prépondérante. Elles sont des structures membranaires hautement spécialisées qui permettent le passage de petites molécules entre le cytoplasme des cellules voisines via des canaux spécialisés. Ce mécanisme assure un couplage électrique et métabolique entre les cellules (Beyer et Berthoud, 2009).

Chez les vertébrés une jonction communicante consiste en deux héli-canaux appelés connexons, avec un connexon fourni par chaque cellule. Un connexon est un hexamère de connexines (Wong *et al.*, 2016). Les connexines sont une famille de protéines transmembranaires, comprenant vingt-et-une isoformes homologues chez l'Homme, classées en cinq groupes (alpha, bêta, gamma, delta, epsilon) en fonction de leur boucle intracytoplasmique et leur extrémité C-terminale (Beyer et Berthoud, 2009). Plusieurs facteurs sont responsables de la régulation fonctionnelle des connexines, comme l'augmentation de la concentration intracellulaire des ions de calcium (Ca^{2+}) et la diminution du pH. De plus, de nombreux travaux ont mis l'accent sur l'impact des modifications post-traductionnelles des connexines qui régulent leur transport à la membrane, la perméabilité de leurs canaux jonctionnels ainsi que l'internalisation de leurs canaux afin qu'ils soient dégradés par la voie lysosomale ou par le protéasome (Laird, 2006).

L'importance des connexines est mise en évidence par le grand nombre de maladies liées à leurs mutations qui conduisent au dysfonctionnement des jonctions communicantes. Ces mutations ont été associées à vingt-huit maladies génétiques. Parmi celles-ci, six sont causées par des mutations de la Cx43, notamment des troubles cardiovasculaires (Allagnat *et al.*, 2005). et huit autres sont le résultat de mutations de la Cx26, notamment des maladies cutanées liées à l'hyperkératose et à la surdité (Srinivas *et al.*, 2018 ; Qiu *et al.*, 2022).

Avec l'avancement de la technologie, l'étude *in silico* des protéines possède un vaste champ d'application en apportant des éléments de structure et même de fonction comme la prédiction des modifications post-traductionnelles. Ceci nous permet une meilleure compréhension des voies de signalisation et des interactions cellulaires, sans avoir recours à

des expériences physiques coûteuses ou chronophages. Cela est achevé grâce à des bases de données spécialisées et des outils disponibles et facilement accessibles. Dans ce travail, nous nous sommes intéressés à effectuer une analyse qui consiste de performer des prédictions de la topologie et des sites potentiels de phosphorylation pour les dix isoformes exprimées dans le tissu épithélial, à savoir les Cx37, Cx40, Cx43 et Cx46 du groupe Alpha, ainsi que les Cx26, Cx30, Cx30.3, Cx31, Cx31.1 et Cx32 du groupe Bêta.

Notre premier objectif est la prédiction de la topologie afin de discriminer les domaines transmembranaires des domaines intra- et extracellulaires. Dans ce but, nous avons utilisé trois programmes de prédiction de la topologie membranaire : ProtScale (Gasteiger *et al.*, 2005), Δ G Predictor (Hessa *et al.*, 2007) et TOPCONS (Bernsel *et al.*, 2009).

Le deuxième objectif est d'effectuer un alignement multiple des séquences de connexines Alpha et Bêta par l'outil Clustal Omega (Sievers *et al.*, 2011) afin d'identifier les régions de séquence les plus conservées et de déterminer les caractéristiques propres à chaque groupe.

Finalement, le dernier objectif concerne la prédiction des sites de phosphorylation au niveau des domaines C-terminaux de la Cx43 et de la Cx26 en utilisant l'outil NetPhos3.1 (Blom *et al.*, 2004).

Ce mémoire commence par une introduction qui présente la problématique de notre étude et nos objectifs projetés, suivie de deux parties :

La première est une synthèse bibliographique qui comporte deux chapitres.

- Le chapitre I présente une recherche approfondie sur les connexines et les jonctions communicantes, en fournissant des informations depuis leur synthèse jusqu'à la formation des plaques jonctionnelles et de leur dégradation.
- Le chapitre II présente les notions de bioinformatique en détaillant divers concepts et méthodes utilisées dans ce domaine.

La deuxième partie concerne les travaux expérimentaux réalisés dans notre étude. Elle comporte également deux chapitres :

- Le chapitre III détaille les méthodes utilisées dans ce mémoire.
- Le chapitre IV présente les résultats obtenus par ces méthodes avec leurs interprétations et une discussion.

Enfin, ce mémoire se termine par une conclusion.

Partie 1:
Synthèse Bibliographique

**Chapitre I:
Connexines et jonctions
communicantes**

1. Introduction

L'homéostasie dans un organisme pluricellulaire est maintenue grâce à différents processus, dont la communication implique (i) la **communication juxtacrine** via des molécules liées à la membrane plasmique, (ii) la **communication paracrine** et **endocrine** par l'intermédiaire de molécules libérées dans le milieu extracellulaire telles que les cytokines, les facteurs de croissance, les hormones et les neurotransmetteurs qui agissent sur des récepteurs membranaires ou nucléaires (Talbot, 2012).

Cependant, la communication paracrine permet également le transfert direct des messagers à travers des canaux spécifiques intercellulaires regroupés dans des régions spécialisées de la membrane plasmique. Ces canaux forment la «jonction communicante» (Meşe *et al.*, 2007).

2. La structure des jonctions communicantes

Les jonctions communicantes, également connues sous le nom de *nexus*, *macula communicans* et jonctions *GAP* (Saez *et al.*, 2003), sont des structures membranaires hautement spécialisées qui relient les cytoplastes de deux cellules voisines. Ces cellules sont séparées par une mince fente appelée « *gap* ». Chez les vertébrés, la jonction *GAP* est formée par l'association de deux demi-canaux cylindriques creux appelés connexons, et chaque connexon résulte de l'oligomérisation de six sous-unités de protéines transmembranaires appelées connexines (Chanson *et al.*, 2018) (figure 1.1).

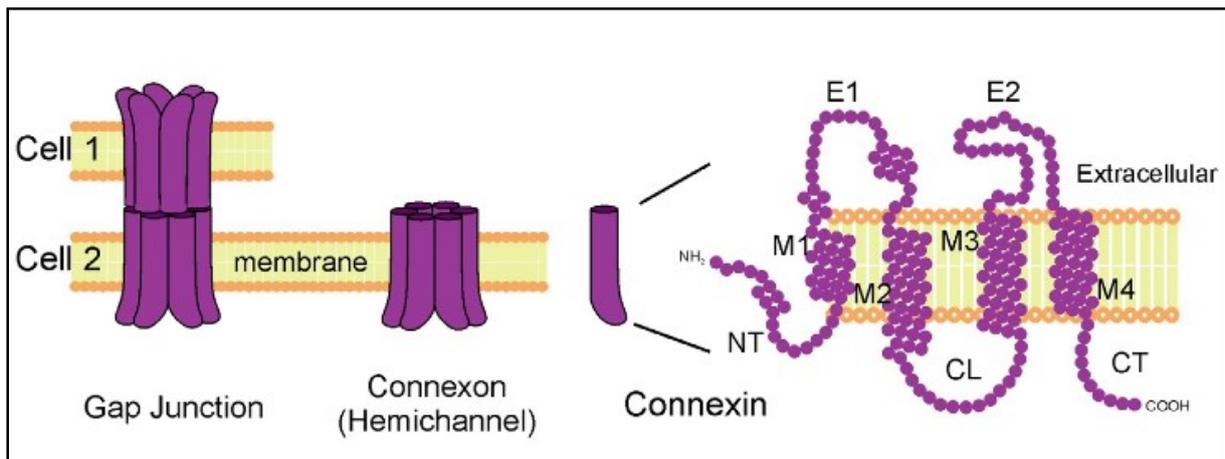


Figure 1.1 : Structure de la connexine, du connexon et de la jonction *GAP* (Bai *et al.*, 2018)

Par ailleurs, chez les invertébrés, les homologues des connexines sont appelés innexines (Inx). Ces dernières s'organisent de la même manière en hexamères (innexons), formant ainsi des héli-canaux qui, par arrimage au héli-canaux des cellules voisines, forment les canaux jonctionnels (Beyer et Berthoud, 2018).

Il est intéressant de noter que les vertébrés et les chordés inférieurs contiennent des homologues de l'innexine, appelés pannexines (Panx) (Beyer et Berthoud, 2018 ; Baranova *et al.*, 2004). Dont la topologie est similaire, les pannexons font référence à des canaux à membrane unique (Panchin *et al.*, 2000).

Les jonctions communicantes permettent l'échange direct de molécules d'un poids moléculaire inférieur à 1200 Da (Talbot, 2012), des ions (K^+ et Ca^{2+}), des seconds messagers (AMPC, GMPc, Inositol 1,4,5-trisphosphate (IP_3)) et des petits métabolites (glucose, ATP) entre les cellules adjacentes (Meşe *et al.*, 2007) à travers le pore du canal jonctionnel d'un diamètre de 2 nm. Ce transfert permet aux cellules de coordonner leurs activités physiologiques, y compris la contraction synchronisée dans le muscle cardiaque (Hervé *et al.*, 2008) et la sécrétion coordonnée dans les cellules épithéliales glandulaires (Chanson *et al.*, 2018).

3. Les connexines

3.1 Expression génique des connexines

Les connexines forment une grande famille de protéines transmembranaires comprenant 20 et 21 membres dans les génomes murins (mCx) et humains (hCx) respectivement. Parmi ces isoformes, dix-neuf (19) sont communs aux deux espèces et peuvent être regroupés en paires orthologues (Söhl et Willecke, 2004).

L'expression des connexines orthologues ne présente pas une distribution tissulaire parfaite. Par exemple, lorsqu'on examine la connexine hCx30.2, on remarque qu'elle est exprimée dans le cœur humain. Par contre, son orthologue chez la souris, mCx30.2, n'est pas exprimé (Söhl *et al.*, 2003).

Les connexines ont deux nomenclatures distinctes. La première, proposée par Beyer et collaborateurs (1987), les catégorise selon « espèce d'origine - Cx - masse moléculaire en kDa ». Par exemple, la protéine hCx43 désigne la connexine 43 humaine ayant une masse moléculaire de 43 kilodaltons.

La seconde nomenclature regroupe les connexines en cinq sous-familles (alpha, bêta, gamma, delta et epsilon) en se basant sur la similitude de leurs séquences et la longueur de leur boucle cytoplasmique. Chaque connexine est désignée par « Gj (*Gap junction*) - Sous-famille (en lettre grecque) - numéro d'ordre de découverte ». Par exemple, la Cx43 appartenant à la sous-famille alpha (α) est nommée Gja1 (tableau 1.1).

Tableau 1.1 : Classification des connexines humaines et murines (Söhl et Willecke, 2004)

Humain				Souris			
Connexine	Gène	Chromosome	Masse moléculaire (KDa)	Connexine	Gène	Chromosome	Masse moléculaire (KDa)
α							
hCx43	GJA1	6q21-q23.2	43,008	mCx43	Gja1	10 (29.0 cM)	43,004
hCx46	GJA3	13q12.11	46,655	mCx46	Gja3	14 (21.0 cM)	46,302
hCx37	GJA4	1p35.1	37,413	mCx37	Gja4	4 (57.6 cM)	37,596
hCx40	GJA5	1q21.1	40,38	mCx40	Gja5	3 (45.6 cM)	40,413
hCx50	GJA8	1q21.1	48,173	mCx50	Gja8	3 (47.1 cM)	49,597
hCx59	GJA9	1p34	58,842	-	-	-	-
hCx62	GJA10	6q15-q16	61,871	mCx57	Gja10	4 (16.1 cM)	57,114
β							
hCx32	GJB1	Xq13.1	32,024	mCx32	Gjb1	X (41.0 cM)	32,003
hCx26	GJB2	13q11-q12	26,2	mCx26	Gjb2	14 (21.0 cM)	26,411
hCx31	GJB3	1p34	30,817	mCx31	Gjb3	4 (58.0 cM)	30,901
hCx30.3	GJB4	1p34.3	30,419	mCx30.3	Gjb4	4 (57.5 cM)	30,388
hCx31.1	GJB5	1p35.1	31,088	mCx31.1	Gjb5	4 (57.5 cM)	31,394
hCx30	GJB6	13q12	30,396	mCx30	Gjb6	14 (22.5 cM)	30,366
hCx25	GJB7	6q15	25,892	-	-	-	-
γ							
hCx45	GJC1	17q21.31	45,482	mCx45	Gjc1	11	45,665
hCx47	GJC2	1q42.13	47,427	mCx47	Gjc2	11 (34.0 cM)	46,603
hCx30.2	GJC3	7q22.1	30,313	mCx29	Gjc3	5	28,981
δ							
hCx36	GJD2	15q14	36,248	mCx36	Gjd2	2 (F3)	36,085
hCx31.9	GJD3	17q21.2	31,933	mCx30.2	Gjd3	11	30,219
hCx40.1	GJD4	10p11.21	40,14	mCx39	Gjd4	18	39,996
ϵ							
hCx23	GJE1	6q24.1	23	mCx23	Gje1	10	23

Bien que les gènes distincts codant les connexines soient situés sur des chromosomes différents, ils présentent une conservation considérable et une structure similaire. Selon Sáez et collaborateurs (2003), la majorité des gènes qui codent pour les connexines sont composés de deux exons et d'un intron de longueur variable. Où l'exon1 contient la région 5'-UTR et l'exon2 comprend la totalité de la séquence codante complète et la région 3'-UTR (figure 1.2).

Ce modèle classique peut comporter des exceptions telles que des épissages au niveau de la région 5'-UTR ou de la région codante par des introns notamment pour la Cx36 (Condorelli *et al.*, 1998 ; Söhl et Willecke, 2004).

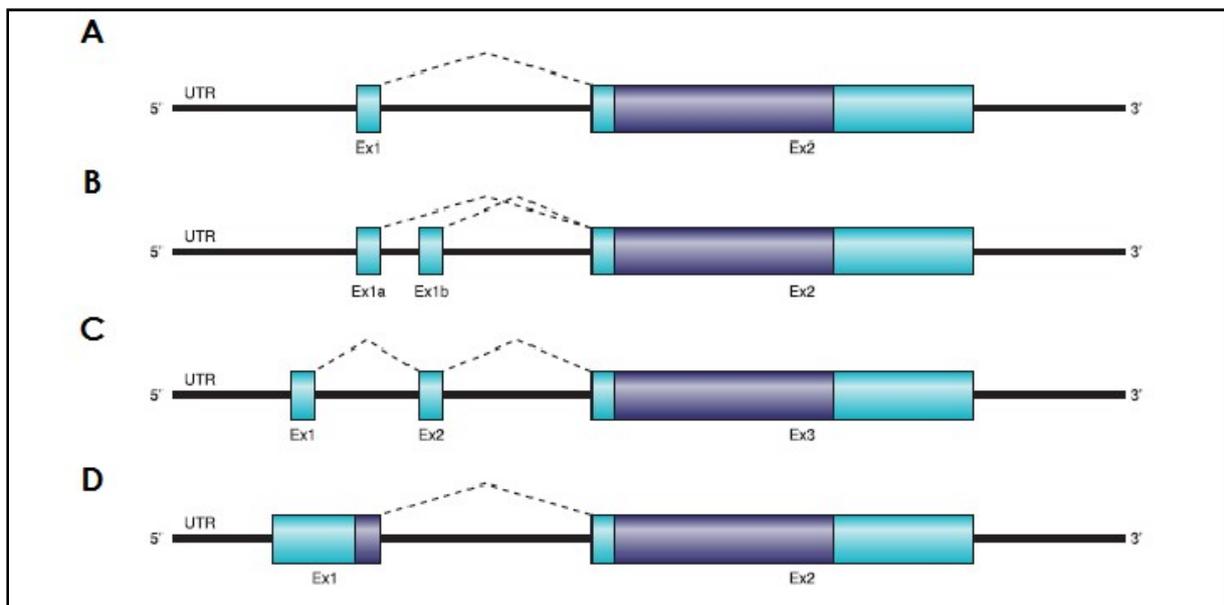


Figure 1.2 : Structures des gènes des connexines (Bosco *et al.*, 2011)

3.2 Structure des connexines

Les recherches menées par Zimmer et collaborateurs (1987) ainsi que les études ultérieures menées par Milks et collaborateurs (1988), Yeager et Gilula (1992) et Zhang et Nicholson (1994) ont révélé que toutes les connexines ont une topologie similaire, montrant la présence de quatre régions transmembranaires hydrophobes (M1, M2, M3 et M4). Chacun de ces domaines présente une conformation en hélice alpha et est composé d'environ 20 à 28 résidus d'acides aminés (Milks *et al.*, 1988).

L'analyse des séquences de ces domaines transmembranaires a montré que les régions M1, M2 et M4 sont principalement composées de résidus hydrophobes. En revanche, la région M3 présente des résidus aromatiques qui, initialement, étaient pensés pour aligner la lumière du canal intercellulaire (Milks *et al.*, 1988 ; Bennett *et al.*, 1991). Cependant, il a été

démonstré par la suite que cette région se loge effectivement dans les sillons des domaines transmembranaires voisins, formant ainsi, avec le domaine M4, l'extérieur de la paroi du canal face aux lipides environnants de la membrane (Maeda *et al.*, 2009).

Les domaines transmembranaires sont reliés par une boucle intracellulaire (I) et deux boucles extracellulaires (E1 et E2). Ces dernières semblent adopter une structure en tonneau β (feuillettes β antiparallèles), contenant chacune 3 résidus de cystéines, sauf la Cx23 qui ne présente que deux cystéines (Iovine *et al.*, 2008). Les extrémités N-terminales et C-terminales sont intracytoplasmiques.

La formation du canal jonctionnel entre les cellules adjacentes par l'arrimage de deux connexons est assurée par les interactions non covalentes entre les boucles extracellulaires E1 et E2 (Yeager *et al.*, 1996), formées respectivement de 31 et 34 acides aminés. Les trois résidus cystéines hautement conservés dans chaque boucle sont responsables de la formation des ponts disulfures intramoléculaires, qui contribuent à la stabilisation de la structure tertiaire du canal (John et Revel, 1991 ; Foote *et al.*, 1998).

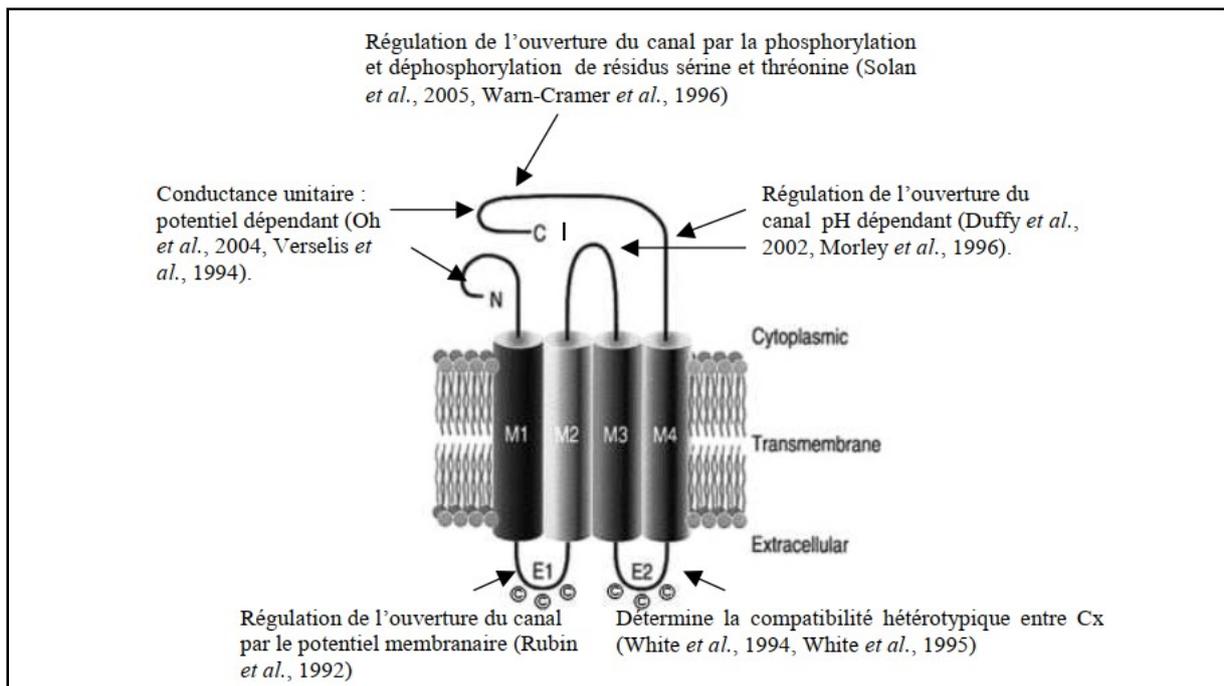


Figure 1.3 : Structure topologique d'une connexine (Abbaci, 2008)

M1, M2, M3, M4 : domaines transmembranaires ; I : boucle intracellulaire ; E1 et E2 : boucles extracellulaires ; C : résidus cystéines ; N : Extrémité N-terminale ; C : Extrémité C-terminale.

L'analyse des séquences de connexines a révélé que l'extrémité amino-terminale, les domaines transmembranaires et les boucles extracellulaires sont les domaines les plus conservés à travers les différentes espèces (Kumar *et al.*, 1986 ; Beyer *et al.*, 1987).

Les domaines les moins conservés sont la boucle intracytoplasmique et l'extrémité C-terminale (Kovacs *et al.*, 2007), qui présentent des séquences et des longueurs très variables. Ces variations expliquent les différences de poids moléculaire qui caractérisent chaque isoforme (Willecke *et al.*, 2002). Par exemple, la Cx46 possède 191 résidus au niveau de l'extrémité C-terminale, tandis que la Cx26 n'en a que 16 (Evans *et al.*, 2006).

Le domaine C-terminal possède de nombreux sites de modifications post-traductionnelles qui vont impacter sa fonction, notamment les sites de phosphorylation par différentes kinases telles que PKA, PKC, MAPK (Hervé *et al.*, 2004 ; Lampe et Lau, 2004).

3.3 Distribution tissulaire des connexines

Les connexines sont largement présentes dans presque tous les types cellulaires des vertébrés, mais certaines exceptions sont connues. Il s'agit des cellules complètement différenciées telles que les cellules musculaires squelettiques adultes, les globules rouges, certains neurones et les spermatozoïdes (Saez *et al.*, 2003).

Le taux d'expression varie pour chaque isoforme. Certaines sont ubiquitaires, comme la Cx43, alors que d'autres sont limitées à un nombre restreint de tissus, telles que la Cx26 et la Cx32. Par contre, certaines connexines sont exclusives, comme la Cx50 qui n'est exprimée que dans le cristallin (Talbot, 2012).

Il est possible que différentes isoformes de connexines soient co-exprimées dans un même type cellulaire (Laird, 2006), comme les hépatocytes qui expriment simultanément la Cx26 et la Cx32. De plus, dix isoformes de connexines ont été identifiées dans la peau.

Le profil d'expression des connexines peut également être restreint par le stade de développement de l'organisme. Par exemple, la mCx40 est fortement exprimée au niveau du cœur pendant la période fœtale, mais son expression diminue après la naissance (Van Kempen *et al.*, 1996).

Cycle de vie des connexines

La durée de vie des connexines est limitée à seulement quelques heures, soit entre 1,5 et 5h, ce qui leur permet de répondre aux besoins physiologiques de la cellule grâce à leur synthèse et dégradation continue (Fallon et Goodenough, 1981 ; Laird *et al.*, 1991).

3.3.1 Synthèse et assemblage

La synthèse et l'assemblage de la plupart des connexines commencent au niveau du réticulum endoplasmique (RE) en suivant la voie classique de sécrétion. Dans ce compartiment, elles s'oligomérisent en héli-canaux hexamériques (Falk *et al.*, 1994 ; Ahmad *et al.*, 1999). Cependant, pour la Cx43 et la Cx46, Musil et Goodenough (1993) ont découvert que des monomères de ces isoformes étaient présents dans l'appareil de Golgi, et leur association en connexons semblait se produire dans le réseau trans-Golgi.

Si le repliement des connexines est correct, la synthèse se poursuit dans le compartiment intermédiaire (ERGIC : Compartiment Intermediaire Réticulum Endoplasmique Golgi), afin d'atteindre l'appareil de Golgi (Segretain et Falk, 2004). Si les connexines sont mal repliées (et par conséquent non fonctionnelles), elles seront poly-ubiquitinylées et dégradées dans le protéasome (Brikci Nigassa, 2014).

De manière exceptionnelle, la voie empruntée par la Cx26 diffère de la voie classique en passant directement du réticulum endoplasmique à la membrane plasmique de manière indépendante de l'appareil de Golgi (Ahmad et Evans, 2002).

Une ou plusieurs protéines chaperonnes protègent les connexines pendant leur synthèse bien qu'elles ne soient pas glycosylées. Il est ainsi suggéré que des disulfures isomérases interviennent en raison des ponts disulfures formés entre les résidus cystéine des boucles extracellulaires (Laird, 2006).

Ces observations nous permettent de déduire que le compartiment d'assemblage des connexons varie en fonction de l'isotype de la connexine et de leur compatibilité sélective qui contrôle la formation des héli-canaux homo- ou hétéromériques.

3.3.1.1 Compatibilité entre connexines

L'oligomérisation de différentes isoformes de connexines produit deux types de connexons: ceux constitués d'un même type de connexines, appelés connexons homomériques, et ceux composés de deux ou plusieurs types de connexines, appelés connexons hétéromériques. De même, la formation du canal jonctionnel complet par l'association de deux connexons peut être homotypique si les deux connexons sont identiques ou hétérotypique si les deux connexons sont différents (Koval *et al.*, 2014).

Les connexons hétéromériques alors peuvent être assemblés selon de nombreuses combinaisons possibles (figure 1.4). Ces combinaisons sont contrôlées par la compatibilité hétéromérique sélective des connexines, qui semble être possible uniquement au sein de la même sous-famille (Talbot, 2012).

Par exemple, des connexons hétéromériques constitués de Cx26 et Cx32 (sous-famille β) ont été observés dans le foie (Sosinsky, 1995), tandis que d'autres constitués de Cx46 et Cx50 (sous-famille α) ont été observés dans le cristallin (Hopperstad *et al.*, 2000). Par contre, il a été démontré que Cx26 (GJ β 2) et Cx43 (GJ α 1) ne forment pas de connexon hétéromérique (Gemel *et al.*, 2004).

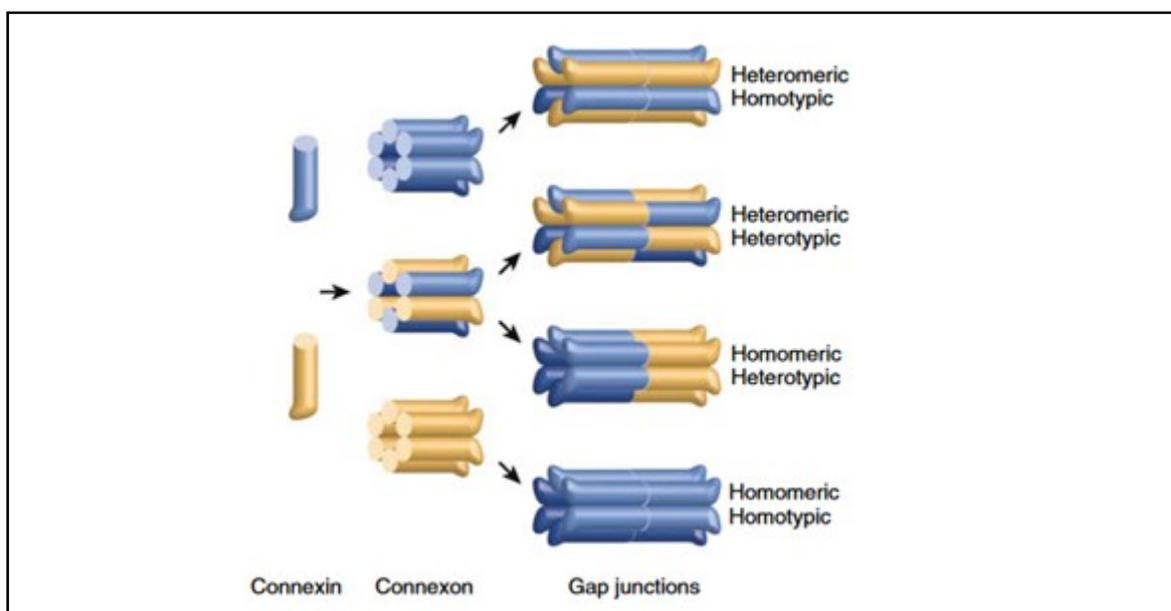


Figure 1.4 : Différentes combinaisons possibles de connexons (Meşe *et al.*, 2007)

Ces combinaisons diverses jouent un rôle crucial dans la perméabilité sélective des canaux. Bevans et collaborateurs (1998) ont montré que les hémi-canaux homomériques de Cx32 étaient perméables à l'AMPc et au GMPc, tandis que les hémi-canaux hétéromériques de Cx26/Cx32 présentaient une perméabilité réduite à l'AMPc.

Cette compatibilité sélective entre les différentes isoformes semble être liée au lieu de synthèse (le réticulum endoplasmique (RE) ou l'appareil de Golgi), empêchant ainsi les isoformes d'avoir un « contact physique » permettant l'hétéro-oligomérisation. Ou alors à la sélectivité aux molécules chaperonnes spécifiques qui se lient à certaines connexines pour contrôler leur interaction (Abbaci, 2008).

3.3.2 Transport

Les hémi-canaux nouvellement assemblés empruntent différentes voies de transport pour atteindre la membrane cellulaire. La majorité des connexons de la Cx26 sont directement insérés dans la membrane plasmique, de même que certains hémi-canaux de la Cx43. Cependant, dans la plupart des cas la Cx43 et plus rarement de la Cx26, suivent la voie sécrétoire classique et semblent être transportées le long des microtubules dans des vésicules de transport jusqu'à la membrane plasmique (Brikci-Nigassa *et al.*, 2012).

3.3.3 Intégration à la membrane plasmique et formation de la jonction communicante

Les connexons s'intègrent à la bicouche phospholipidique et se déplacent latéralement le long de la membrane sous la direction de la N et la E-cadhérine, pour rejoindre la plaque jonctionnelle où ils forment des jonctions *gap* avec les hémi-canaux des cellules adjacentes (Meda, 1996 ; Wei *et al.*, 2005 ; Laird, 2006). Les canaux jonctionnels récemment formés diffusent à la périphérie de la plaque jonctionnelle tandis que les canaux plus anciens se trouvent au centre et sont destinés à être éliminés par internalisation dans des connexosomes, où ils sont dégradés par les protéasomes ou les lysosomes (Laird, 2006).

3.3.4 Dégradation des canaux

Le mécanisme de dégradation des jonctions communicantes implique l'internalisation des canaux jonctionnels marqués pour l'élimination dans des connexosomes. Ces derniers sont des vésicules à double membrane également appelées jonctions annulaires qui se détachent de la membrane cellulaire pour être détruites par la voie lysosomale et/ou protéosomale ubiquitine-dépendante (Jordan *et al.*, 2001 ; Qin *et al.*, 2003 ; Segretain et Falk, 2004).

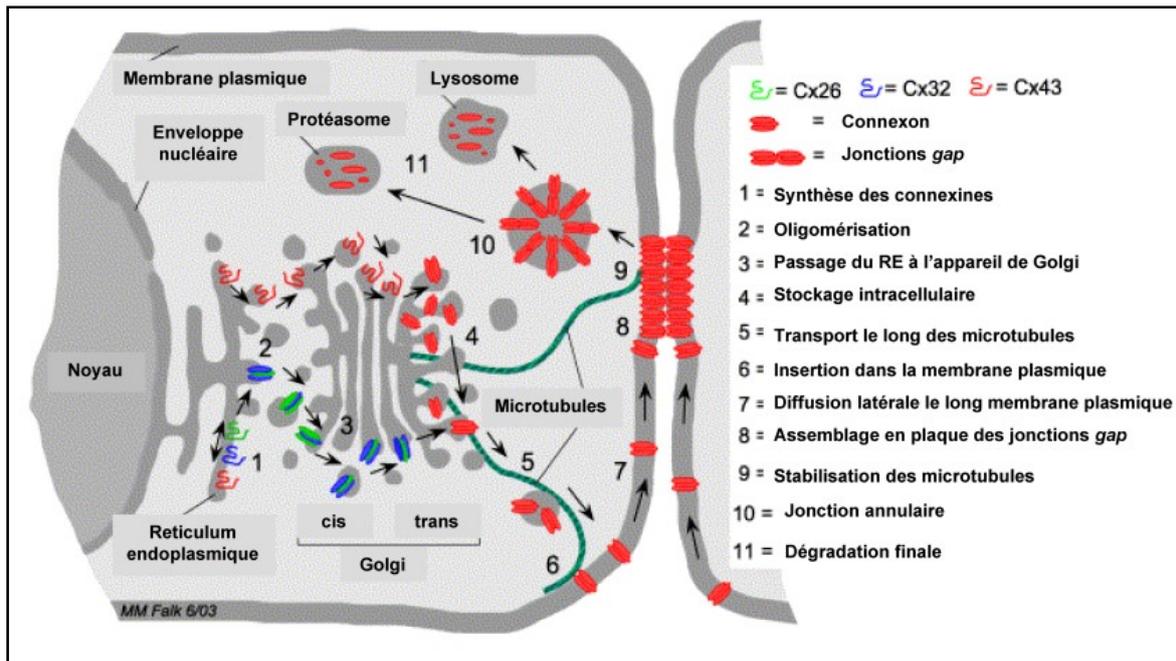


Figure 1.5 : Synthèse, assemblage et dégradation des junctons gap (Segretain et Falk, 2004)

(1) la synthèse des isoformes de Cxs au niveau des membranes du réticulum endoplasmique (2) l'oligomérisation en héli-canaux selon une conformation homo- ou hétéromérique (3) les connexons traversent dans l'appareil de Golgi (4) les connexons sont emmagasinés dans le réseau trans-golgien (5) le transport le long des microtubules (6) l'intégration des connexons dans la bicouche lipidique (7) le déplacement latéral des héli-canaux le long de la membrane cellulaire (8) le regroupement des canaux jonctionnels en plaques (9) la stabilisation de l'extrémité des microtubules par liaison aux Cx43 des junctons gap (10) l'internalisation dans les connexosomes (11) La dégradation totale par les voies lysosomales et/ou protéosomales ubiquitine-dépendante.

4. Fonctions des jonctions communicantes

Les jonctions communicantes sont responsables de la signalisation paracrine en permettant aux cellules adjacentes d'échanger directement des molécules et des ions entre leurs cytoplasmes via leurs canaux intercellulaires. Toutefois, les connexons non-appariés présents dans la surface cellulaire, permettent aussi l'échange de messagers paracrines avec le milieu extracellulaire. Ces messagers, tels que l'ATP, interviennent dans des fonctions physiologiques variées (Goodenough et Paul, 2003 ; Goodenough et Paul, 2009).

4.1 Couplage électrique

Le couplage ionique des jonctions communicantes implique le passage de courant ionique (Ca^{2+} , Na^+ , Cl^- , K^+) établissant une communication électrique entre les cellules excitables, comme les neurones, les cellules musculaires lisses et cardiaques, ainsi que les cellules sécrétrices et sensorielles, qui subissent une dépolarisation membranaire. Ce mécanisme essentiel génère une réponse synchronisée des cellules, telle que la contraction musculaire ou la libération d'insuline par les cellules β pancréatiques (Oyamada *et al.*, 1994 ; Bruzzone *et al.*, 1996 ; Cao *et al.*, 1997 ; Bennett et Zukin, 2004).

4.2 Couplage métabolique

Le couplage métabolique des jonctions communicantes établit une voie de communication directe entre les cellules adjacentes non-excitables en favorisant la diffusion rapide de molécules hydrophiles ($\text{PM} < 1200 \text{ Da}$), comme les nucléotides, les acides aminés, les cofacteurs vitaminiques et les seconds messagers (AMPc , IP_3 , Ca^{2+}) (Sáez *et al.*, 2003 ; Goodenough et Paul, 2009) ainsi que les pARNi et les μARN (Wolvetang *et al.*, 2007 ; Lim *et al.*, 2011). Ce mécanisme est indispensable pour accomplir certaines fonctions biologiques, telles que la prolifération et la différenciation, permettant ainsi des réponses cellulaires coordonnées à divers stimuli afin de maintenir l'homéostasie tissulaire (Saez *et al.*, 2003 ; Goodenough et Paul, 2009).

5. Régulation de la communication jonctionnelle

Plusieurs facteurs sont responsables de la régulation fonctionnelle des jonctions communicantes. Ces facteurs exercent leur influence sur certains domaines des connexines, induisant des changements conformationnels qui influent sur l'ouverture et la fermeture des canaux jonctionnels.

5.1 Concentration du calcium intracellulaire

La calmoduline (CaM) activée par le Ca^{2+} intracellulaire interagit avec les connexines au niveau de leur boucle intracellulaire et C-terminale, ce qui entraîne la fermeture des canaux jonctionnels. Cette interaction entre la CaM et les connexines se produit lorsque la concentration de Ca^{2+} intracellulaire est élevée, de l'ordre du micromolaires (μM) (Zhou *et al.*, 2007 ; Zou *et al.*, 2014 ; Peracchia, 2020).

5.2 pH intracellulaire

La communication jonctionnelle est susceptible d'être réduite via l'acidification intracellulaire. Les niveaux de pH acides peuvent provoquer des modifications structurales dans la boucle intracytoplasmique et le domaine C-terminal de la Cx43, conduisant à la fermeture des pores jonctionnels. La sensibilité à l'acidification intracellulaire peut différer d'une cellule à l'autre, selon le type de connexines exprimées et la nature homo- ou hétérotypique des jonctions *gap* (Peracchia, 2004 ; Wei *et al.*, 2004 ; Hirst-Jensen *et al.*, 2007).

5.3 Phosphorylation des connexines

La phosphorylation est une modification post-traductionnelle qui consiste à ajouter des groupes phosphate à des acides aminés spécifiques : la sérine, la thréonine ou la tyrosine, modifiant la charge, l'hydrophobicité et la structure de toutes les connexines, à l'exception de la Cx26 (Aasen *et al.*, 2018). Elle se déroule principalement dans l'extrémité C-terminale (Lampe et Lau, 2000) ou dans certains cas exceptionnels, dans la boucle cytoplasmique, notamment chez la Cx36 et la Cx56 (Berthoud *et al.*, 1997 ; Urschel *et al.*, 2006 ; Račkauskas *et al.*, 2010). Elle peut influencer de manière significative les propriétés des jonctions *gap*, notamment leur perméabilité, leur durée de vie et leur localisation intracellulaire, ainsi que leur trafic et leurs interactions protéine-protéine (Aasen *et al.*, 2018).

La phosphorylation des connexines entraîne une régulation différentielle de la communication jonctionnelle. Elle joue un rôle crucial dans la modulation de l'ouverture et de la fermeture des canaux en termes de durée moyenne et de probabilité (Verrecchia *et al.*, 1999 ; Duthe *et al.*, 2000), en induisant l'interaction de l'extrémité C-terminale avec la boucle cytoplasmique (Hervé *et al.*, 2011).

5.3.1 Phosphorylation de la Connexine 43

La Cx43 est la connexine la plus largement répandue, elle possède 21 sites de phosphorylation au niveau de son domaine C-terminal (Axelsen *et al.*, 2013).

Ces phosphorylations sont gérées par de multiples voies de signalisation, impliquant la protéine kinase A (PKA) et la protéine kinase C (PKC) (Reynhout *et al.*, 1992 ; Sáez *et al.*, 1997), les tyrosine kinases de la famille Src (Lin *et al.*, 2001) et la famille des MAPK (Warn-Cramer *et al.*, 1996 ; Warn-Cramer *et al.*, 1998 ; Cameron *et al.*, 2003 ; Sirnes *et al.*, 2009).

La phosphorylation de la Cx43 par la PKA augmente la communication et favorise son insertion dans la membrane plasmique (Atkinson *et al.*, 1995 ; Paulson *et al.*, 2000 ; TenBroek *et al.*, 2001 ; Matsumura *et al.*, 2006 ; Sirnes *et al.*, 2009), tandis que la PKC l'inhibe. De même, la phosphorylation de la Cx43 par les MAPK entraîne son internalisation (Leithe et Rivedal, 2004).

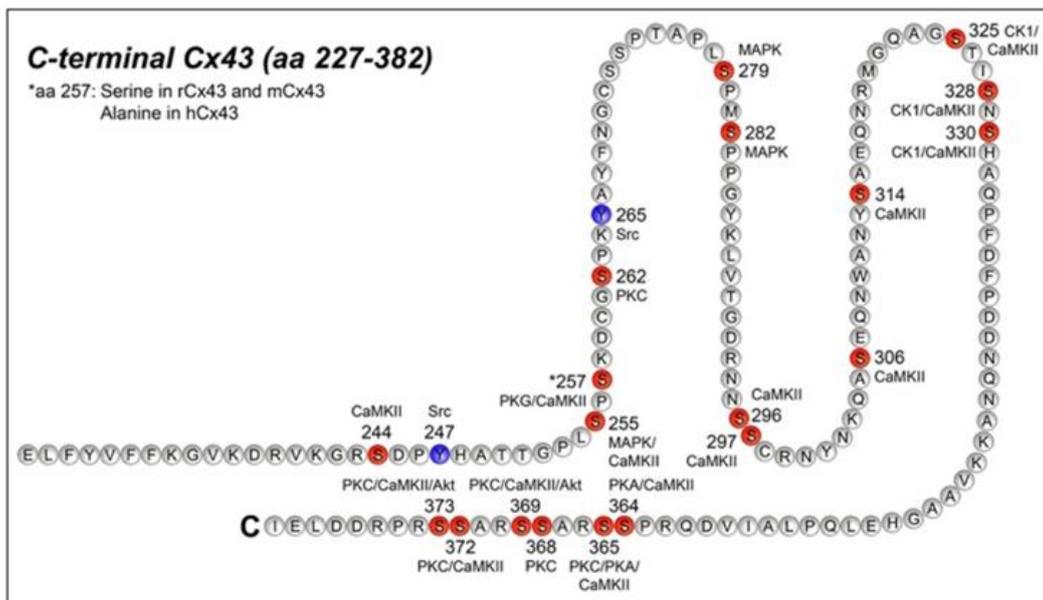


Figure 1.6 : Sites de phosphorylation dans le domaine C-terminale de la Cx43 et ses kinases respectives (Axelsen *et al.*, 2013)

6. Jonctions *gap* et pathologies

Des mutations dans les gènes des connexines ont été associées à plusieurs maladies humaines, conduisant à des troubles qui affectent un large éventail de tissus et d'organes, notamment la peau, les yeux, les oreilles, le cœur et le système nerveux.

Par exemple, des mutations du gène GJB2, qui code pour la Cx26, ont été associées à la surdité héréditaire non syndromique (tel que la surdité neurosensorielle à transmission autosomique récessive et dominante), ainsi qu'une forme de surdité associée à une hyperkératose palmo-plantaire connue sous le nom de syndrome de Vohwinkel. Tandis que des mutations du gène GJB6, qui code pour la Cx30, ont été associées à la perte d'audition (Forme dominante de surdité non syndromique) et à des troubles cutanés (Syndrome de Clouston). Les mutations du gène GJB3 et GJB4, qui code pour la Cx31 et la Cx30.3 respectivement, ont été impliquées dans des troubles cutanés notamment l'érythrokratodermie variabilis (Allagnat *et al.*, 2005),

Des études ont révélé deux catégories principales de mutations dans la famille bêta connexine. Le premier type de mutation conduit à une déficience dans le trafic et l'absence de fonction du canal, qui est appelé "mutations de perte de fonction". Le deuxième type forme des héli-canaux « fuyants » en ciblant la membrane plasmique, appelé "mutations de gain de fonction" (Martin et Van Steensel, 2015).

Des mutations dans les gènes des Alpha connexines ont également été associées à plusieurs maladies. Des mutations du gène GJA1, qui code pour la Cx43, ont été associées à divers troubles cardiovasculaires, notamment la cardiomyopathie hypertrophique, l'insuffisance cardiaque et l'ischémie (Boengler et Schulz, 2017). Par ailleurs, les mutations de la Cx43 sont également associées à la dysplasie oculo-dento-digitale, une maladie caractérisée par une dysmorphie cranio-faciale, une surdité et diverses anomalies neurologiques et ophtalmologiques (Allagnat *et al.*, 2005).

En outre, plusieurs troubles neurologiques, notamment la maladie de Charcot-Marie-Tooth liée au chromosome X et l'épilepsie, qui sont associées à des mutations de GJB1 codant pour la Cx32 (Bone *et al.*, 1997) et de GJD2 codant pour la Cx36 (Wang et Wu, 2021) respectivement. Par ailleurs, des mutations du gène GJA3, qui code pour la Cx46 responsable du maintien de la transparence du cristallin, ont été associées à l'apparition de cataractes sévères et à d'autres troubles oculaires (Shi *et al.*, 2022).

Chapitre II :

La bioinformatique

1. Introduction

La recherche scientifique s'est principalement appuyée sur les expériences réalisées *in vivo* et *in vitro*, mais ces approches présentent des limites. Les investigations réalisées *in vivo* peuvent poser des risques aux animaux de l'étude ou ne pas refléter avec précision les résultats de ces modèles animaux chez l'Homme. D'autre part, les recherches *in vitro* ne peuvent pas rendre compte de la complexité des interactions qui se produisent *in vivo*.

Par conséquent, les méthodes *in silico* ont gagné en popularité ces dernières années car elles utilisent diverses techniques, telles que l'apprentissage automatique « *machine learning* » (Fox et Kriegl, 2006), l'exploration de données « *data mining* » (Zaki *et al.*, 2007) et les outils d'analyse pour étudier les hypothèses pharmacologiques, concevoir de nouveaux médicaments, ainsi que pour l'analyse génomique et protéomique. Les méthodes *in silico* offrent plusieurs avantages, notamment une réduction des coûts et du temps, ainsi que la possibilité de fournir des résultats plus expérimental.

2. La bioinformatique

L'institut américain de recherche sur le génome humain (*NHGRI : National Human Genome Research Institute*) définit la bioinformatique comme une sous-discipline scientifique qui consiste à utiliser la technologie informatique pour collecter, stocker, analyser et diffuser des données et des informations biologiques telles que les séquences d'ADN et d'acides aminés ou les annotations sur ces séquences et ceci afin d'améliorer notre compréhension de la santé et des maladies.

La bioinformatique a des applications diverses dans de nombreux domaines. Par exemple, dans l'analyse du génome, les outils de bioinformatique sont utilisés pour comparer les génomes de différentes espèces et identifier les gènes uniques, permettant aux chercheurs de reconstruire les arbres évolutifs et d'étudier les relations entre les espèces. De même, dans l'analyse des protéines, la bioinformatique aide à identifier les domaines protéiques et leur architecture, afin de modéliser les structures protéiques et mieux comprendre leurs fonctions. Dans la conception des médicaments, la bioinformatique joue un rôle crucial dans la modélisation des interactions moléculaires entre les composés bioactifs.

Grâce aux contributions de la bioinformatique, les scientifiques ont fait des progrès significatifs dans la compréhension des systèmes biologiques complexes et dans le développement de nouvelles approches pour améliorer la santé humaine.

3. Les banques de données biologiques

Les banques de données biologiques sont des répertoires d'informations généralement accessibles en ligne, régulièrement mis à jour et révisés pour fournir des données précises et actualisées. Elles jouent un rôle central en bioinformatique en offrant aux scientifiques l'opportunité d'accéder à une grande variété de données biologiquement pertinentes (Baxevanis et Bateman, 2015), organisées de manière spécifique pour faciliter le stockage et la récupération efficace de l'information.

Il est possible de classer les banques de données biologiques de deux manières différentes :

La première classification les catégorise en banque de données « généralistes » et en banque de données « spécialisées ». Les banques de données généralistes peuvent inclure une vaste collection de données et fournir un ensemble d'informations plutôt hétérogènes. Par exemple, GenBank (Benson *et al.*, 2012), EMBL «*European Molecular Biology Laboratory*» (Stoesser *et al.*, 2002), DDBJ «*DNA Data Bank of Japan* » (Tateno *et al.*, 2002). En revanche, les banques de données spécialisées se concentrent sur un ensemble de données plus homogène établi autour d'un thème spécifique. Par exemple, PROSITE (Bairoch *et al.*, 1997).

La deuxième classification catégorise les banques de données biologiques en banque de données « primaire », « secondaire » et « composite ». Une banque de données primaire ne contient que des informations de séquence ou de structure, tandis que les banques de données secondaires contiennent des données sélectionnées et annotées, dérivées de l'analyse et du traitement des données primaires. Enfin, les banques de données composites intègre des données provenant de diverses sources telles que des bases de données des séquences, de littérature et de structure, ce qui permet d'éviter la nécessité de rechercher dans plusieurs sources séparées. Par exemple, le NCBI «*National Center for Biotechnology Information*» qui sert de hub central pour de nombreuses autres bases de données, y compris GenBank, PubMed, BioSample, et bien d'autres,

Le tableau 2.1 présente les banques de données les plus connues, ainsi que leur classification et leur contenu.

Tableau 2.1 : Classification, contenu et adresses des banques de données les plus connues

Nom	Contenu	Adresse (URL)
Banque de données généralistes / primaire		
GenBank (Benson <i>et al.</i> , 2012)	Collection de toutes les séquences d'ADN disponibles publiquement	https://www.ncbi.nlm.nih.gov/genbank/
EMBL (Stoesser <i>et al.</i> , 2002)	Séquences d'ADN / ARN collectées à partir de la littérature scientifique, des offices de brevets et est directement soumise par les chercheurs	https://www.embl.org/
DDBJ (Tateno <i>et al.</i> , 2002)	Séquences d'ADN principalement provenant de chercheurs japonais.	https://www.ddbj.nig.ac.jp
UniProt « <i>Universal Protein Resource</i> » (UniProt Consortium, 2019)	Swiss-Prot: séquences protéiques corrigées manuellement, et TrEMBL: séquences non corrigées	https://www.uniprot.org/
PDB « <i>Protein Data Bank</i> » (Burley <i>et al.</i> , 2017)	Données structurales tridimensionnelles des protéines	https://www.rcsb.org/
Banque de données spécialisées / secondaire		
PROSITE (Bairoch <i>et al.</i> , 1997)	Collections des familles, les domaines et les motifs protéiques	https://prosite.expasy.org/
CATH « <i>Class, Architecture, Topology, Homology</i> » (Orengo <i>et al.</i> , 1997)	Classe les domaines protéiques en relations évolutives et structurales.	http://www.cathdb.info/
SCOP2 « <i>Structural Classification of Proteins database</i> » (Andreeva <i>et al.</i> , 2014)	Classification des domaines structuraux des protéines	https://scop.mrc-lmb.cam.ac.uk/scop2/
eMOTIF (Huang et Brutlag, 2001)	collection de plus de 170.000 motifs de séquences protéiques hautement spécifiques et sensibles	https://motif.stanford.edu/distributions/emotif/
Banque de données composite		
NCBI (Sayers <i>et al.</i> , 2021)	Séquences de nucléotides et de protéines, données d'expression génique, projets de séquençage du génome, etc.	https://www.ncbi.nlm.nih.gov/

4. Les différents formats de fichiers en bioinformatique

La génération et l'exploitation des données biologiques ont pris diverses formes, conduisant à la création de différents formats de fichiers. Chacun de ces formats a son propre cas d'utilisation en fonction de la compatibilité avec des logiciels spécifiques, des exigences de traitement et d'analyse des données, de la lisibilité humaine et de l'efficacité de stockage. Il existe deux principaux types :

4.1 Les formats propres aux bases de données : le format GenBank, le format EMBL et le format DDBJ.

Les formats GenBank, EMBL et DDBJ sont utilisés pour stocker des séquences nucléotidiques entières ainsi que leurs annotations telles que les noms de gènes, les fonctions, les numéros d'accession (partagés entre les trois bases de données) (Benson *et al.*, 2010), les références bibliographiques et bien d'autres données supplémentaires. Ces formats facilitent la récupération et le partage des séquences ainsi que des informations qui leur sont associées.

Par exemple, le format GenBank est composé d'une section d'annotation et d'une section de séquence. Le début de la section d'annotation est marqué par une ligne commençant par le mot «*LOCUS*». Le début de la section de séquence est marqué par une ligne commençant par le mot «*ORIGIN*» et la fin de la section est marquée par une ligne contenant seulement «*//*» (Leonard et Littlejohn, 2004) (figure 2.1).

Le format DDBJ est très similaire au format GenBank et suit la même structure, tandis que le format EMBL a une structure légèrement différente. Dans ce format, l'entrée de la séquence commence par une ligne d'identification qui commence par «*ID*», suivie de lignes d'annotation supplémentaires. Le début de la séquence est marqué par une ligne qui commence par «*SQ*» et la fin de la séquence est marquée par deux barres obliques «*//*».

RecName: Full=Gap junction alpha-1 protein; AltName: Full=Connexin-43; Short=Cx43; AltName: Full=Gap junction 43 kDa heart protein

UniProtKB/Swiss-Prot: P17302.2
[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to:

LOCUS	CXA1 HUMAN	382 aa	linear	PRI 03-MAY-2023
--------------	------------	--------	--------	-----------------

DEFINITION RecName: Full=Gap junction alpha-1 protein; AltName: Full=Connexin-43; Short=Cx43; AltName: Full=Gap junction 43 kDa heart protein.

ACCESSION P17302

VERSION P17302.2

DBSOURCE UniProtKB: locus CXA1_HUMAN, accession [P17302](#);
class: standard.
extra accessions: B2R5U9, Q6FHU1, Q9Y5I8
created: Aug 1, 1990.
sequence updated: Jan 23, 2007.
annotation updated: May 3, 2023.

KEYWORDS 3D-structure; Acetylation; Cataract; Cell junction; Cell membrane; Disease variant; Disulfide bond; Endoplasmic reticulum; Gap junction; Hypotrichosis; Isopeptide bond; Membrane; Palmoplantar keratoderma; Phosphoprotein; Reference proteome; S-nitrosylation; Transmembrane; Transmembrane helix; Ubl conjugation.

SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (residues 1 to 382)
AUTHORS Fishman, G.I., Spray, D.C. and Levinwand, L.A.
TITLE Molecular characterization and functional expression of the human cardiac gap junction channel
JOURNAL J Cell Biol 111 (2), 589-598 (1990)
PUBMED [1696265](#)
REMARK NUCLEOTIDE SEQUENCE [MRNA].;
TISSUE=Heart muscle

ORIGIN

```

1 mgdwsalgkl ldkvqaysta ggkvwlsvlf ifrilllgt vesawgdeqs afrcntqqpg
61 cenvcydksf pishvrfwvl qiifsvptl lylahvfyvm rkeeklnkke eelkvaqtdg
121 vnvdmhklqi eikkfkygie ehgkvkrmrg llrtyiisil fksifevafli qwyiygfs
181 lsavytckrd pcphqvdcfl srptektifi ifmlvvslvs lalnielfy vffkgvkdrv
241 kgksdpyhat sgalspakdc gsqkyayfng cssptaplsp mspggyklvt gdrnsscrn
301 ynkqaseqnrw anysaeqnrn gqagstisns haqpfdfpdd nqnskklaag helqplaidv
361 qrpssrassr assrprpddl ei
//

```

Figure 2.1 : Le format Genbank/DBJ

4.2 Les formats utilisés par les outils d'analyse des séquences

De nombreux formats tels que FASTA, FASTQ, SAM, BAM, VCF, GFF, GTF et BED sont supportés par les programmes d'analyse. FASTA est utilisé pour les séquences de nucléotides ou de protéines. FASTQ présente à la fois la séquence de nucléotides et ses scores de qualité Phred¹ correspondants. SAM et BAM sont plutôt utilisés pour des données d'alignement de séquences, alors que VCF représente des variations génétiques, et GFF, GTF et BED fournissent des informations d'annotation génomique (Zhang, 2016).

Cependant, le format de fichier qui est le plus couramment utilisé et qui est supporté par la plupart des bases de données et des outils d'analyse de séquences est le format FASTA (figure 2.2). La première ligne de chaque nouvelle entrée commence par un chevron « > » suivi d'une description d'une seule ligne de la séquence qui suit. La ligne suivante contient la séquence en code à lettre unique. Dans ce format, un trait d'union (-) représente un écart de longueur indéterminée; La lettre U représente la sélénocystéine dans les séquences de protéines, et un astérisque (*) dans une séquence de protéines indique une fin de traduction (Leonard et Littlejohn, 2004).

```

FASTA ▾

RecName: Full=Gap junction alpha-1 protein; AltName: Full=Connexin-43; Short=Cx43; AltName: Full=Gap junction 43 kDa heart protein

UniProtKB/Swiss-Prot: P17302.2
GenPept Identical Proteins Graphics
>sp|P17302.2|CXA1_HUMAN RecName: Full=Gap junction alpha-1 protein; AltName: Full=Connexin-43; Short=Cx43; AltName: Full=Gap junction 43 kDa heart protein
MGDWSALGKLLDKVQAYSTAGGKVNLSVLFIFRIILLGTAVESAWGDEQSAFRCNTQQPGCENVCYDKSF
PISHVRFWVLQIIFVSVPTLLYLAHVYVMRKEEKLNKKEEELKVAQTDGVMVDMHLKQIEIKKFKYGIIE
EHGKVKMRGGLLRTYIISILFKSIFEVAFLLIQWYIYGFSLSAVYTCRDPDPCPHQVDCFLSRPTEKTIFI
IFMLVVSLSLALNIIELFYVFFKGVKDRVKGKSDPYHATSGALSPAKDCGSQKYAYFNGCSPTAPLSP
MSPPGYKLVTDGRNNSCRNRYNKQASEQNWANYSAEQNRMGQAGSTISNSHAQPFDFPDDNQNSKKLAAG
HELQPLAIVDQRPSSRASSRASSRPRPDDLEI
    
```

Figure 2.2 : Le format FASTA

¹ Le score de qualité Phred est une mesure de la qualité de la base dans une séquence qui donne une estimation de la probabilité qu'une base soit erronée. Il est stocké sous forme de caractère ASCII (une lettre, un chiffre ou un symbole), dont la valeur ASCII indique la précision de l'appel de base. Plus le score Phred est élevé, plus la qualité de la base est considérée comme élevée et plus la probabilité que la base soit correcte est élevée.

5. Programmes d'analyse des séquences protéiques

5.1 Méthodes de prédiction de la localisation des domaines protéiques

La prédiction de la localisation des protéines repose sur l'analyse des données issues de la séquence protéique, en se basant sur les propriétés des résidus d'acides aminés dans la structure primaire. Cette information est ensuite exploitée pour déterminer -par exemple- le nombre des domaines transmembranaires des protéines membranaires ou détecter un signal-motif spécifique (Laroum, 2011).

La prédiction peut être réalisée à l'aide de diverses méthodes. Certaines couvrent un large éventail de localisations possibles, tandis que d'autres se concentrent sur des localisations ou des motifs spécifiques. Nous allons nous concentrer sur deux catégories de méthodes de prédiction dans ce mémoire : les méthodes de prédiction des segments transmembranaires et les méthodes de prédiction du peptide signal.

5.1.1 Les méthodes de prédiction des segments transmembranaires

La prédiction de la topologie des protéines membranaires repose sur l'identification de la disposition des domaines transmembranaires, des domaines intracellulaires et des domaines extracellulaires en fonction des propriétés hydrophobes/hydrophiles des acides aminés de la séquence (structure primaire). De même, la prédiction des segments transmembranaires dépend de leur structure secondaire. Certaines méthodes sont spécifiquement conçues pour prédire les segments structurés en hélices alpha (Ahmed *et al.*, 2010), tandis que d'autres sont spécifiques pour les segments structurés en feuillets bêta (Bagos *et al.*, 2005).

Les méthodes de prédiction des segments membranaires structurés en hélices alpha sont les plus répandues dues à la prévalence de ce type de structure et de la facilité de son identification (Laroum, 2011).

Kyte et Doolittle (1982) ont élaboré une des premières méthodes de prédiction qui repose sur le principe que les acides aminés hydrophobes sont plus susceptibles de se trouver dans la membrane cellulaire, tandis que les acides aminés hydrophiles sont plus susceptibles de se trouver à la surface, soit du côté extracellulaire, soit du côté intracellulaire. Ils ont établi une échelle allant de -4,5 à +4,5.

5.1.2 Les méthodes de détection du peptide signal

Les peptides signal (PS) sont des courtes séquences d'acides aminés situées à l'extrémité N-terminale des protéines, qui régulent la sécrétion et la translocation de ces dernières chez tous les organismes vivants (Teufel *et al.*, 2022). Les PS peuvent être prédits à partir des données de séquence. L'identification du peptide signal et de son site de clivage par les outils de bioinformatique permet la prédiction des protéines secrétées.

De nombreux outils existent pour la prédiction de la localisation du peptide signal ainsi que de son site de clivage, cela dépend de l'algorithme utilisé par le programme et de l'organisme spécifique considéré.

SignalP¹ est l'un des premiers et des plus populaires programmes pour effectuer ce type de prédiction. La version initiale utilisait des réseaux de neurones (Nielsen *et al.*, 1997), puis les versions ultérieures ont utilisé des modèles de chaînes de Markov cachées (Nielsen *et al.*, 1999). La version la plus récente, SignalP-6.0, utilise l'apprentissage automatique « *machine learning* » et les modèles de langage de protéines pour prédire la présence de peptides signal et l'emplacement de leurs sites de clivage dans les protéines d'Archaea, de bactéries à Gram positif, de bactéries à Gram négatif et d'eucaryotes (Teufel *et al.*, 2022).

5.2 Alignement des séquences

L'alignement des séquences est un outil fondamental de la bioinformatique qui a de nombreuses applications dans l'analyse des séquences protéiques. Ces applications comprennent notamment la modélisation de la structure, la prédiction des sites fonctionnels, l'analyse phylogénétique et la recherche de séquences dans des bases de données (Pei, 2008).

L'objectif d'alignement des séquences est d'identifier les régions de ressemblance, déterminant ainsi les caractéristiques propres et partagées entre deux ou plusieurs séquences d'acides aminés. Ce processus peut être réalisé à l'aide de différents algorithmes et méthodes, tels que l'alignement de séquences par paires ou multiples. Ces méthodes peuvent être globales ou locales.

5.2.1 Alignement global

L'algorithme d'alignement global de Needleman et Wunsch (1970) est utilisé pour aligner deux séquences du début à la fin. Il produit un alignement indépendamment de la similitude entre les séquences. Cet algorithme convient lorsque les deux séquences ont une longueur

¹ <https://dtu.biolib.com/SignalP-6>

similaire et ont un degré significatif de similitude tout au long de leur séquence (Likic, 2008) (figure 2.3).

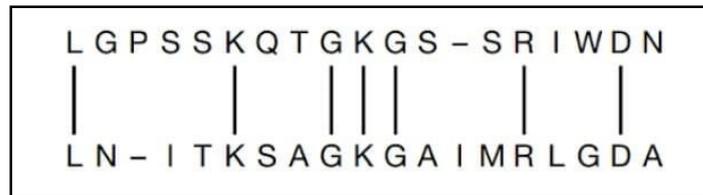


Figure 2.3 : L’alignement global de deux séquences protéiques

5.2.2 Alignement local

L'algorithme d'alignement local de Smith et Waterman (1981) convient pour comparer des séquences qui sont substantiellement différentes les unes des autres et qui peuvent différer significativement en longueur, mais qui n'ont que de courtes régions de similarité. L'algorithme recherche l'alignement optimal en identifiant de courtes régions de similarité ou de « *matches* » entre les deux séquences et en permettant l'introduction des brèches « *gaps* » dans l'alignement sans prendre en compte l'alignement des autres régions de séquences (figure 2.4).

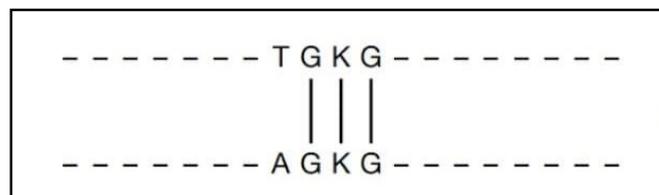


Figure 2.4 : L’alignement local de deux séquences protéiques

5.2.3 Alignement par paires

L'alignement par paire est utilisé pour comparer une séquence donnée à une séquence de référence afin d'identifier leurs similitudes et leurs différences. Le but de l'alignement par paire est de trouver le meilleur alignement possible entre les deux séquences en introduisant des gaps ou des mis-matches. L'alignement par paire peut être réalisé en utilisant des méthodes d'alignement globales ou locales.

5.2.4 Alignement multiple

L'alignement de séquences multiples (figure 2.5) est le processus d'aligner trois ou plusieurs séquences afin d'identifier les régions conservées et les insertions/suppressions

(indels) dans toutes les séquences. Les méthodes d'alignement de séquences multiples peuvent être progressives ou itératives.

Les méthodes progressives commencent par aligner les deux séquences les plus similaires, puis ajoutent une à une les séquences supplémentaires à l'alignement en fonction de leur similarité avec l'alignement existant. Ce processus est répété jusqu'à ce que toutes les séquences soient alignées (Thompson *et al.*, 1999).

Les méthodes itératives, en revanche, utilisent un premier alignement de départ (tel qu'un alignement par paire) pour générer un nouvel alignement plus précis. Le processus est répété en affinant l'alignement à chaque itération jusqu'à ce que la convergence soit atteinte (Hirosawa *et al.*, 1995 ; Meshoul et Layeb, 2017).

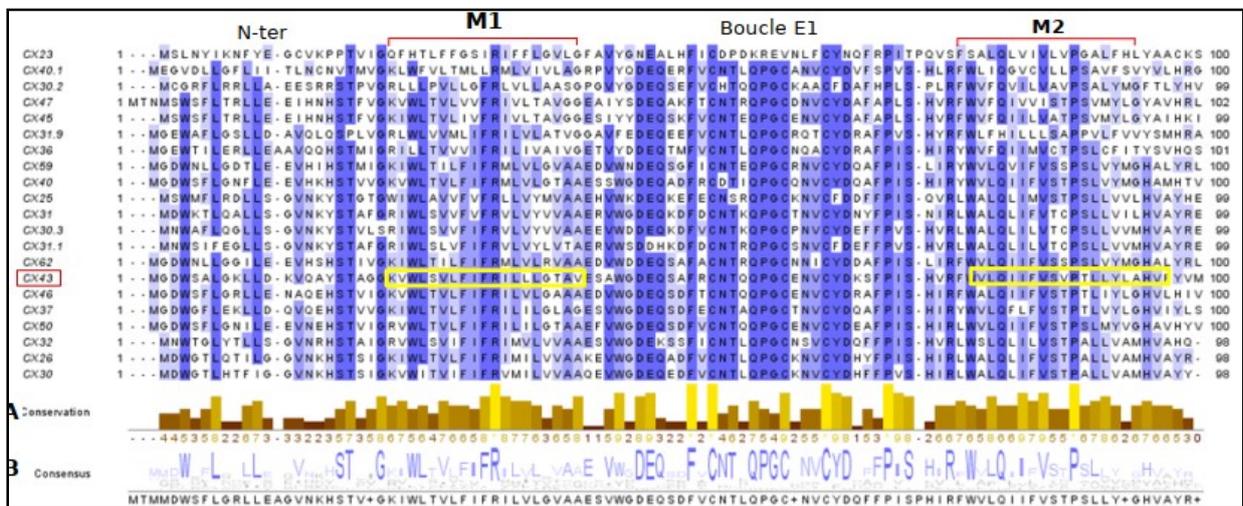


Figure 2.5 : Alignement multiple des 21 isoformes des connexines humaines (Benayad, 2018)

Le tableau 2.2 présente les outils d'alignement les plus couramment utilisés.

Tableau 2.2 : Les programmes d'alignement de séquences les plus couramment utilisés

Alignement	Méthode	Programme	Serveur Web
Par paires	Global	EMBOSS Needle	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/psa/emboss_needle
		EMBOSS Stretcher	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/psa/emboss_stretcher
		GSEARCH2SEQ	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/psa/ggsearch2seq
	Local	EMBOSS Water	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/psa/emboss_water
		EMBOSS Matcher	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/psa/emboss_matcher
		LALIGN	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/psa/lalign
BLAST		https://blast.ncbi.nlm.nih.gov/Blast.cgi	
SSEARCH2SEQ	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/psa/ssearch2seq		
Multiple	Progressive	Clustal Omega	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/msa/clustalo
		MAFFT	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/msa/mafft
		T-Coffee	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/msa/tcoffee
		EMBOSSCons	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/msa/emboss_cons
		KALIGN	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/msa/kalign
		MUSCLE	https://wwwdev.ebi.ac.uk/Tools/jdispatcher/msa/muscle
	Itérative	WebPRANK	https://www.ebi.ac.uk/goldman-srv/webprank/
		DIALIGN	https://dialign.gobics.de/
		PRRN	https://www.genome.jp/tools-bin/prrn

5.3 Étude des modifications post-traductionnelles

Les modifications post-traductionnelles, désignent les modifications apportées aux chaînes latérales d'acides aminés des protéines après leur synthèse (Ramazi et Zahiri, 2021). Elles ont un impact significatif sur la fonction des protéines en affectant la stabilité, la localisation, la régulation et les interactions des protéines entre eux et avec d'autres molécules.

On peut distinguer deux grandes catégories de modifications post-traductionnelles : l'ajout covalent des molécules et le clivage d'une liaison peptidique dans le processus de conversion d'un précurseur protéique en une protéine mature (Virág *et al.*, 2020).

Il existe plusieurs outils bioinformatiques disponibles pour permettre l'analyse de ces modifications. Ces outils utilisent différents algorithmes pour prédire les sites potentiels de phosphorylation, de glycosylation, de palmitoylation, etc. Cependant, ces logiciels ne peuvent pas prédire les conséquences fonctionnelles des modifications post-traductionnelles. Les résultats d'études expérimentales sont utilisés pour construire des bases de données. Ces bases de données fournissent des informations qui sont ensuite utilisées pour entraîner les

programmes conçus afin d'effectuer ces prédictions. Le tableau 2.3 présente certains des outils les plus populaires.

Tableau 2.3 : Les outils de prédiction des modifications post-traductionnelles les plus couramment utilisés

AA modifié	Principales fonctions	Programme	Serveur Web
Phosphorylation			
Ser Thr	Régulation de l'activité enzymatique	NetPhos3.1	https://services.healthtech.dtu.dk/services/NetPhos-3.1/
		GPS5.0	http://gps.biocuckoo.cn/download.php
		KinasePhos3.0	https://awi.cuhk.edu.cn/KinasePhos/
		PhosphoSite+	https://www.phosphosite.org/kinaseLibraryAction.action
		PhosphoSVM	http://sysbio.unl.edu/PhosphoSVM/
		ScanSite	https://scansite4.mit.edu/#home
		Musite	http://musite.sourceforge.net/
DISPHOS1.3	http://phospho.elm.eu.org/links.html		
Acétylation			
Lys Nter	Localisation subcellulaire, Stabilité de la protéine, Régulation des interactions ADN-protéine	NetAcet	https://services.healthtech.dtu.dk/services/NetAcet-1.0/
		GPS-PAIL	http://pail.biocuckoo.org/download.php
		DeepAcet	https://github.com/Sunmile/DeepAcet
Glycosylation O/N liée			
Ser Thr Asn	Repliement/ Stabilité des protéines, Adhésion cellulaire	YingOYang1.2	https://services.healthtech.dtu.dk/services/YinOYang-1.2/
		NetNGlyc1.0	https://services.healthtech.dtu.dk/services/NetNGlyc-1.0/
		NetOGlyc4.0	https://services.healthtech.dtu.dk/services/NetOGlyc-4.0/
		GPP	https://comp.chem.nottingham.ac.uk/glyco/
		DictyOGlyc1.1	https://services.healthtech.dtu.dk/services/DictyOGlyc-1.1/
		GLYCOPP	http://crdd.osdd.net/raghava/glycopp/
Méthylation			
Lys Arg	Code des histones	PRmePRed	http://bioinfo.icgeb.res.in/PRmePRed/
		NmSEER 2.0	http://www.rnanut.net/nmseer-v2/
		Met-predictor	https://sourceforge.net/projects/met-predictor/
Poly-ubiquitination			
Lys Nter	Dégradation protéosomique	UbPred	http://www.ubpred.org/
		UbiNet2.0	https://awi.cuhk.edu.cn/~ubinet/index.php
		GPS-Uber	http://gpsuber.biocuckoo.cn/online.php
S-palmitoylation			
Cys	liaison avec la membrane	PalmPred	http://proteininformatics.org/mkumar/palmpred/index.html
		CKSAAP-Palm	http://doc.aporc.org/wiki/CKSAAP-Palm
		GPS-Palm	https://gpspalm.biocuckoo.cn/download.php

AA : Acide aminé.

Partie 2 :

Travaux expérimentaux et

Résultats

Chapitre III :

Matériel et Méthodes

1. Choix de Connexines

Les connexines humaines sont une famille de protéines transmembranaires impliquées dans la formation des jonctions communicantes (*gap junctions*). Chez l'Homme, vingt-et-une isoformes des connexines homologues ont été identifiées, classées en cinq groupes majeurs (alpha, bêta, gamma, delta, epsilon) en fonction de leur boucle intracytoplasmique et de leur extrémité C-terminale (Beyer et Berthoud, 2009).

Dans ce travail, nous nous sommes intéressés à l'analyse des dix isoformes exprimées dans le tissu épithélial, à savoir les Cx37, Cx40, Cx43 et Cx46 du groupe Alpha, ainsi que les Cx26, Cx30, Cx30.3, Cx31, Cx31.1 et Cx32 du groupe Bêta. Notre choix a été motivé par une recherche bibliographique approfondie portant sur la distribution tissulaire des connexines (tableau 3.1) où nous avons observé leur forte expression dans les tissus épithéliaux.

Tableau 3.1 : Expression tissulaire des connexines humaines

Connexine	Profil d'expression tissulaire
α	
Cx43	Ubiquitaire
Cx46	Tissu épithélial (cristallin, épithélium alvéolaire)
Cx37	Tissu musculaire lisse, tissu épithélial (peau), ovaires
Cx40	Tissu musculaire cardiaque, endothélium, tissu épithélial (peau), tissu nerveux
Cx50	Tissu épithélial (cristallin)
Cx59	Génome humain
Cx62	Tissu nerveux (cellules horizontales de la rétine)
β	
Cx32	Tissu nerveux (cellules de Schwann, oligodendrocytes), tissu épithélial (foie, peau)
Cx26	Tissu nerveux (cellules gliales), tissu épithélial (épithélium respiratoire, foie, pancréas, peau), cochlée
Cx31	Tissu épithélial (épithélium respiratoire, peau), cochlée
Cx30.3	Tissu épithélial (peau)
Cx31.1	Tissu épithélial (peau)
Cx30	Tissu nerveux, tissu épithélial (épithélium respiratoire, peau), cochlée
Cx25	Génome humain
γ	
Cx45	Tissu nerveux (astrocytes, neurones, oligodendrocytes), tissu épithélial (cellules épithéliales du canal pancréatique, peau), tissu conjonctif (ostéoblastes), tissu musculaire cardiaque
Cx47	Tissu nerveux (oligodendrocytes)
Cx30.2	Tissu épithélial (cellules épithéliales du canal pancréatique, foie), tissu nerveux (oligodendrocytes)
δ	
Cx36	Tissu nerveux (neurones), tissu épithélial spécialisé (cellules β pancréatiques)
Cx31.9	Génome humain
Cx40.1	Génome humain
ϵ	
Cx23	Génome humain

Pour cela, nous avons analysé la topologie membranaire des connexines représentatives de chaque groupe (la Cx43 pour le groupe Alpha et la Cx26 pour le groupe Bêta). Ensuite, nous avons effectué un alignement de séquences multiples des isoformes Alpha et Bêta afin d'identifier les régions de séquence les plus conservées, permettant ainsi de déterminer les caractéristiques propres à chaque groupe. Enfin, nous avons procédé à la prédiction des sites potentiels de phosphorylation pour la Cx43 et la Cx26.

Les dix séquences de connexines utilisées dans ce travail ont été acquises à partir de la base de données UniProtKB/SWISS-PROT (<https://www.uniprot.org/>), une base de données de séquences de protéines de haute qualité, non redondante, corrigées et annotées manuellement (figure 3.1).

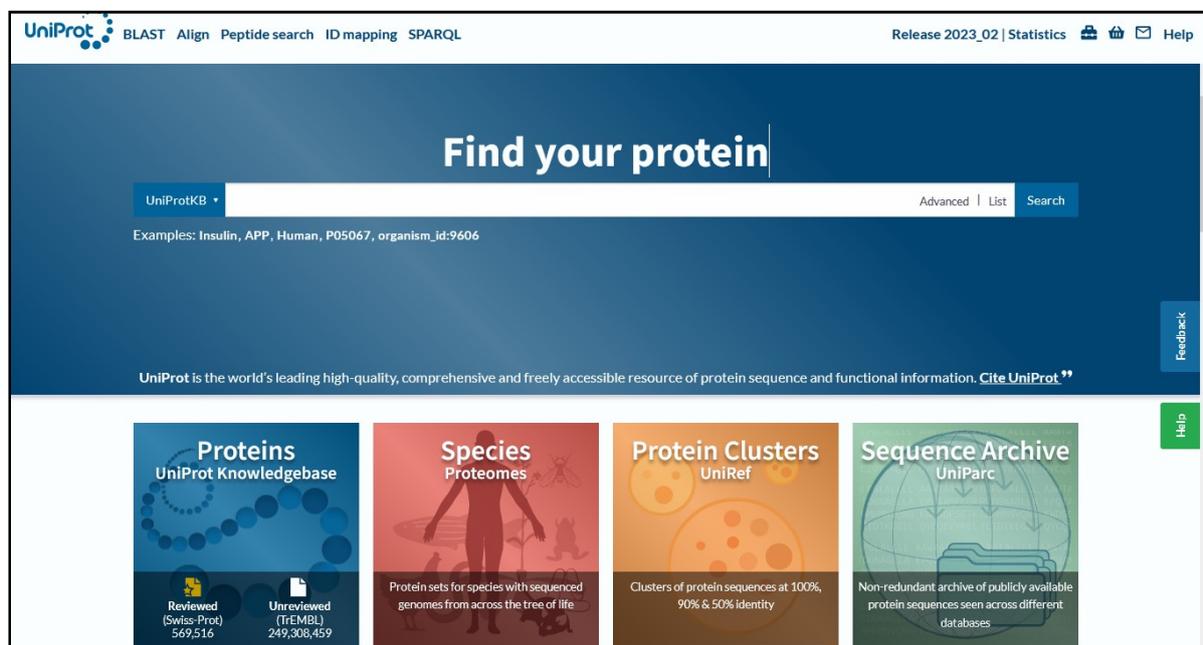


Figure 3.1 : Interface web de la base de données UniProtKB/SWISS-PROT

La base de données UniProtKB/Swiss-Prot fournit une description détaillée de la protéine recherchée, comprenant la nomenclature complète, ainsi que des informations telles que le nombre d'acides aminés, la séquence, le gène, l'organisme et le score d'annotation. Dans cette nomenclature, nous trouverons l'abréviation « GJ » pour *gap junction* suivie par une lettre grecque qui correspond au groupe dont la protéine fait partie (Alpha ou Bêta) et d'un nombre pour indiquer l'ordre de découverte au sein de ce groupe.

Les protéines contenues dans cette base de données sont toutes archivées sous un identifiant appelé "numéro d'accession". Le tableau 3.2 présente les dix isoformes impliquées dans notre étude, telles qu'elles sont répertoriées dans UniProtKB/Swiss-Prot.

Tableau 3.2: Nomenclature des dix isoformes des connexines humaines selon UniProtKB

Connexine	Numéro d'accèsion	Noms des connexines selon UniProtKB	Nombre d'acides aminés
α			
Cx43	P17302	Full=Gap junction alpha-1 protein; AltName: Full=Connexin-43; Short=Cx43; GeneName: GJA1	382
Cx46	Q9Y6H8	Full=Gap junction alpha-3 protein; AltName: Full=Connexin-46; Short=Cx46; GeneName: GJA3	435
Cx37	P35212	Full=Gap junction alpha-4 protein; AltName: Full=Connexin-37; Short=Cx37; GeneName: GJA4	333
Cx40	P36382	Full=Gap junction alpha-5 protein; AltName: Full=Connexin-40; Short=Cx40; GeneName: GJA5	358
β			
Cx32	P08034	Full=Gap junction beta-1 protein; AltName: Full=Connexin-32; Short=Cx32; GeneName: GJB1	283
Cx26	P29033	Full=Gap junction beta-2 protein; AltName: Full=Connexin-26; Short=Cx26; GeneName: GJB2	226
Cx31	O75712	Full=Gap junction beta-3 protein; AltName: Full=Connexin-31; Short=Cx31; GeneName: GJB3	270
Cx30.3	Q9NTQ9	Full=Gap junction beta-4 protein; AltName: Full=Connexin-30.3; Short=Cx30.3; GeneName: GJB4	266
Cx31.1	O95377	Full=Gap junction beta-5 protein; AltName: Full=Connexin-31.1; Short=Cx31.1; GeneName: GJB5	273
Cx30	O95452	Full=Gap junction beta-6 protein; AltName: Full=Connexin-30; Short=Cx30; GeneName: GJB6	261

La séquence peptidique est fournie sous la forme d'un fichier texte codant les acides aminés selon le format FASTA, qui utilise une seule lettre pour représenter chaque acide aminé. Ce format est supporté par tous les programmes utilisés dans ce travail (figure 3.2).

A

B

```
>sp|P17302|CXA1_HUMAN Gap junction alpha-1 protein OS=Homo sapiens OX=9606 GN=GJA1 PE=1 SV=2
MGDWSALGKLLDKVQAYSTAGGKVVLSVLFIFRILLLLGTAVESAWGDEQSAFRCNTQQPG
CENVCYDKSFPISHVRFWVLQIIFVSVPTLLYLAHVYVMRKEEKLNKKEEELKVAQTDG
VNVDMHLKQIEIKKFKYGIEEHGKVKMRGGLLRITYIISILFKSIFEVAVLLIQWYIYGFS
LSAVYTCRDPCPHQVDCFLSRPTEKTIFFIFMLVSLVSLALNIIELFYVFFKGVKDRV
KGKSDPYHATSGALSPAKDCGSQKYAYFNGCSSPTAPLSPMSPPGYKLVTDGRNNSSCRN
YNKQASEQNNANYSAEQNRMGQAGSTISNSHAQPFDPPDDNQNQSKKLAAGHELQPLAIVD
QRPSSRASSRASSRPRPDDLEI
```

Figure 3.2: Séquence de la connexine 43

A : Recherche de la séquence sur UniProtKB/Swiss-Prot **B :** La séquence de la Cx43 au format FASTA

2. Étude de la topologie membranaire de la Cx43 et la Cx26

La prédiction de la topologie des connexines implique la discrimination des domaines transmembranaires des domaines intra- et extracellulaires sur la base des propriétés des résidus d'acides aminés de leur structure primaire.

Afin d'obtenir des informations détaillées sur la topologie de la Cx43 et la Cx26 ainsi que les positions des acides aminés qui délimitent chaque région, des outils de prédiction topologique ont été utilisés, ProtScale (Gasteiger *et al.*, 2005), ΔG Predictor (Hessa *et al.*, 2007) et TOPCONS (Bernsel *et al.*, 2009).

2.1 ProtScale

Le profil d'hydropathie est un graphique qui représente la capacité hydrophobe de chaque acide aminé de la protéine. Il est utilisé pour prédire les domaines hydrophobes d'une protéine.

Il repose sur le principe que les acides aminés hydrophobes sont plus susceptibles de se retrouver dans la membrane cellulaire, tandis que les acides aminés hydrophiles sont plus susceptibles de se retrouver à la surface, que ce soit du côté intra- ou extracellulaire.

ProtScale (<https://web.expasy.org/protscale/>) est un programme qui utilise ce profil d'hydropathie pour prédire les domaines transmembranaires d'une protéine. Il glisse une fenêtre de taille fixe le long de la séquence protéique, et l'hydropathie moyenne des acides aminés dans la fenêtre est déterminée. La valeur d'hydropathie moyenne est attribuée au résidu au centre de la fenêtre, résumant efficacement l'hydropathie des résidus voisins (figure 3.3).

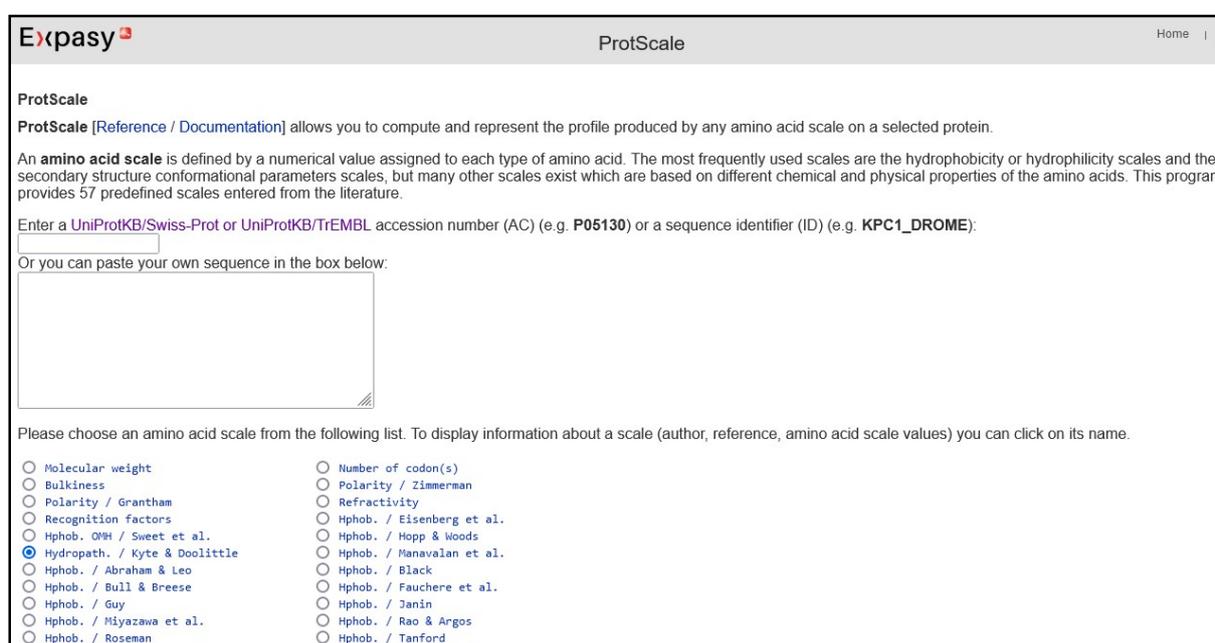


Figure 3.3: Interface de ProtScale d’ExpASY

Nous avons d'abord saisi chaque séquence comme fichier d'entrée (ou simplement saisi le numéro d'accession de la protéine fourni par UniProtKB). Ensuite, nous avons sélectionné l'échelle de Kyte et Doolittle (1982) (tableau 3.3) et ajusté la taille de la fenêtre à 19 résidus. En cliquant sur « *Submit* ». L'outil génère un graphique du profil d'hydropathie sous la forme d'un fichier image, ainsi que d'autres informations, notamment la position des acides aminés.

Tableau 3.3 : Échelle d’hydropathie de Kyte et Doolittle (1982)

Acide aminé	Acide aminé	Acide aminé	Acide aminé
Ala (A) : +1.8	Gln (Q) : -3.5	Leu (L) : +3.8	Ser (S) : -0.8
Arg (R) : -4.5	Glu (E) : -3.5	Lys (K) : -3.9	Thr (Y) : -0.7
Asn (N) : -3.5	Gly (G) : -0.4	Met (M) : +1.9	Trp (W) : -0.9
Asp (D) : -3.5	His (H) : -3.2	Phe (F) : +2.8	Tyr (T) : -1.3
Cys (C) : +2.5	Ile (I) : +4.5	Pro (P) : -1.6	Val (V) : +4.2

2.2 ΔG predictor 1.0

Le ΔG predictor 1.0 (<https://dgpred.cbr.su.se/index.php?p=fullscan>) (figure 3.4) est un programme spécialement conçu pour rechercher et localiser des hélices transmembranaires potentielles dans une séquence de protéine en utilisant son mode de *Full Protein Scan*.

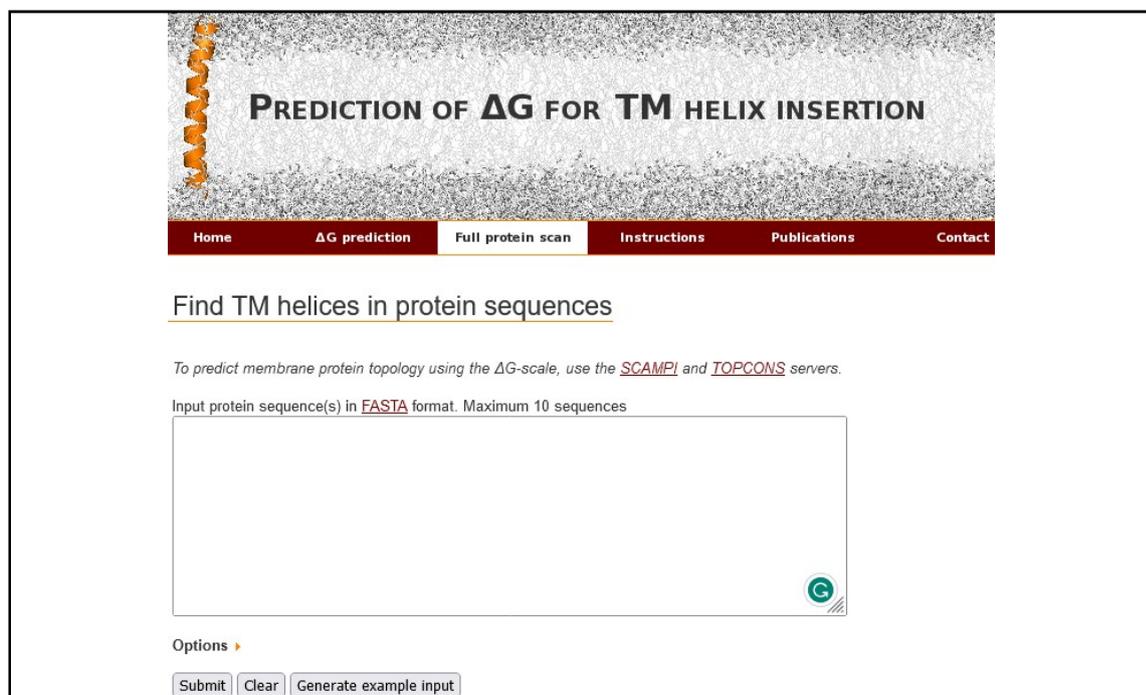


Figure 3.4: Interface de ΔG predictor 1.0

Basé sur les travaux expérimentaux de Hessa et collaborateurs (2005), le programme utilise une fenêtre glissante de 19 à 23 acides aminés pour analyser toute la longueur de la séquence peptidique en donnant une prédiction de la différence d'énergie libre apparente, ΔG_{app} de l'insertion de cette séquence d'hélice transmembranaire putative dans la membrane du réticulum endoplasmique via le translocon Sec61.

Une valeur négative de ΔG_{app} suggère que le translocon Sec61 reconnaît la séquence comme une hélice transmembranaire et l'intègre dans la membrane. Les résultats sont par la suite traduits par un graphique appelée « diagramme d'hydrophobicité ».

2.3 TOPCONS

TOPCONS (<https://topcons.net/>) (figure 3.5) est un prédicteur de la topologie des protéines plus spécifique que les deux autres programmes mentionnés précédemment. Cette spécificité est attribuée à sa capacité à différencier un segment transmembranaire de la partie hydrophobe du peptide signal, qui peut être localisée dans la région N-terminale. Auparavant,

il était difficile de distinguer cette région du segments transmembranaires (Bernsel *et al.*, 2009).

Le fonctionnement de TOPCONS repose sur l'application de cinq algorithmes différents pour la prédiction de la topologie. Parmi ces algorithmes, un utilise la séquence FASTA de la protéine, tandis que les quatre autres nécessitent un profil de séquence provenant de bases de données telles que UniProt. Ces algorithmes permettent de prédire la localisation de chaque résidu de la séquence protéique. La localisation est annotée comme suit : à l'intérieur « i », l'extérieur « o », membranaire « m » et le peptide signal est noté « s ». La combinaison des résultats de ces algorithmes permet de calculer la topologie consensus finale (Brensel *et al.*, 2009).

Figure 3.5: Interface de TOPCONS

TOPCONS produit une représentation graphique mettant en évidence les positions prédites des peptides signal et des hélices transmembranaires. Il a également fourni des prédictions pour la localisation de chaque résidu, ainsi que les séquences correspondantes.

3. Alignement des séquences multiple des 10 isoformes

3.1 Alignement des séquences avec Clustal Omega

Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) est un outil d'alignement multiple de séquences largement utilisé en bioinformatique. Bien qu'il soit principalement conçu pour l'alignement des séquences de protéines, Clustal Omega peut également être

utilisé avec des séquences de nucléotides et peut aligner plus que 4000 séquences à la fois (Sievers *et al.*, 2011). Son algorithme repose sur les étapes suivantes :

Tout d'abord, Clustal Omega effectue l'alignement de toutes les paires de séquences deux à deux en utilisant l'algorithme de Needleman et Wunsch (1970). Cela permet de déterminer le score de divergence entre chaque paire de séquences afin de créer la matrice de distance.

À partir de la matrice de distances, Clustal Omega construit un arbre guide et suivant l'ordre de branchement donné par cet arbre il aligne progressivement les deux séquences les plus proches et crée un profil, puis il aligne la séquence suivante dans l'arbre contre ce profil. Ce processus est répété jusqu'à ce que toutes les séquences soient alignées (figure 3.6).

The screenshot shows the Clustal Omega web interface. At the top, there is a teal header with the logo 'Clustal Omega' and navigation links: 'Input form', 'Web services', 'Help & Documentation', and 'Bioinformatics Tools FAQ'. Below the header, there is a breadcrumb trail: 'Tools > Multiple Sequence Alignment > Clustal Omega'. A light blue box contains a 'Service Announcement' about the Job Dispatcher Services beta website. The main heading is 'Multiple Sequence Alignment' in teal. Below this, a paragraph describes the tool: 'Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools.' An 'Important note' states: 'This tool can align up to 4000 sequences or a maximum file size of 4 MB.' The interface is divided into steps. 'STEP 1 - Enter your input sequences' is highlighted. Below this, there is a dropdown menu labeled 'Enter or paste a set of' with 'PROTEIN' selected. Below the dropdown is a large text area labeled 'sequences in any supported format:'.

Figure 3.6 : Interface web du programme Clustal Omega

Une fois l'alignement effectué, les résultats sont disponibles sous la forme d'un fichier ClustalW (*result.clustalw_aln*). Nous avons ensuite utilisé le programme Jalview pour apporter des modifications à l'alignement. Les fichiers de sortie pour chaque groupe sont donnés en annexe.

3.2 Édition de l'alignement avec Jalview version 2.11.2.6

Jalview (Waterhouse *et al.*, 2009) est un outil conçu pour la visualisation et l'analyse d'alignements de séquences multiples. Il offre une interface facile à utiliser qui permet aux chercheurs d'examiner et de manipuler les alignements afin de mieux analyser la conservation des séquences, les caractéristiques structurelles et les annotations fonctionnelles.

Jalview offre une variété de fonctionnalités, y compris le code couleur des résidus en fonction de leurs propriétés, la mise en évidence des régions conservées, la génération d'arbres phylogénétiques et l'affichage de structures 3D avec Jmol.

Nous avons utilisé le code couleur de Jalview pour mettre en évidence les résidus d'acides aminés dans notre alignement. Tout d'abord, nous avons téléchargé la dernière version du programme à partir de son site officiel (<https://www.jalview.org/>). Ensuite, nous avons importé notre alignement préexistant en naviguant vers *File* → *Input Alignment* → *From File*, et sélectionné le fichier *result.clustalw_aln* de notre alignement. Ensuite, nous avons sélectionné l'option « *Colour* » puis « *Percentage Identity* » (figure 3.7).

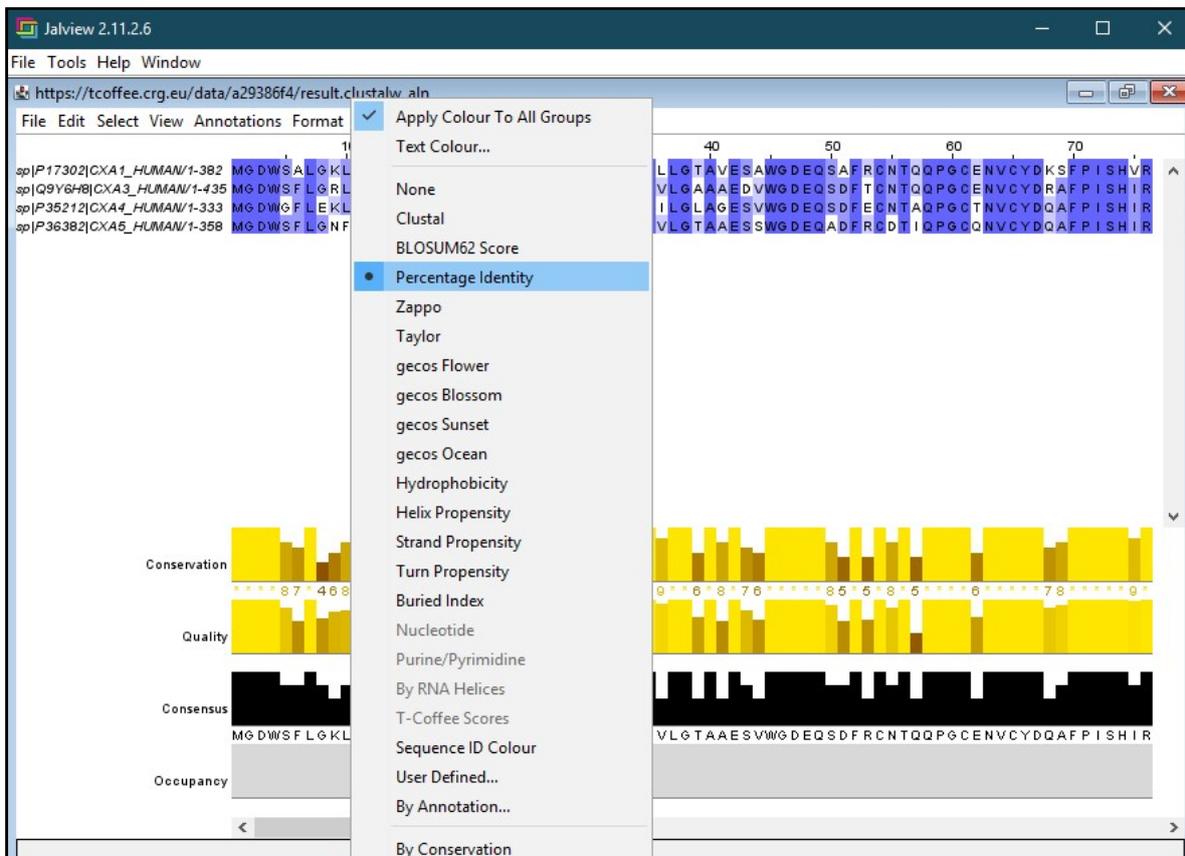


Figure 3.7: Interface d'outil Jalview

Jalview colore les résidus d'acides aminés en fonction de leur pourcentage d'identité et la séquence logo est construite à partir des identités des résidus à chaque colonne. Ce code couleur facilite l'identification des régions conservées dans l'alignement. Les résidus présentant une forte identité se voient attribuer la même couleur, ce qui met en évidence leur conservation (figure 3.8).

> 80 %
> 60 %
> 40%
< 40%

Figure 3.8: Code couleur pour le pourcentage d'identité dans Jalview

4. Étude des modifications post-traductionnelles : La phosphorylation avec NetPhos3.1

NetPhos 3.1 (Blom *et al.*, 2004) est la version la plus récente du programme NetPhos (Blom *et al.*, 1999). Bénéficiant de l'algorithme du réseau neuronal artificiel pour prédire les sites de phosphorylation de la sérine, de la thréonine ou de la tyrosine. NetPhos 3.1 (<https://services.healthtech.dtu.dk/services/NetPhos-3.1/>) (figure 3.9) peut prédire les sites de phosphorylation de 17 kinases au total (ATM, CKI, CKII, CaM-II, DNAPK, EGFR, GSK3, INSR, PKA, PKB, PKC, PKG, RSK, SRC, cdc2, cdk5 et p38MAPK) (Li *et al.*, 2018) dans les séquences de protéines eucaryotes avec une sensibilité comprise entre 69% et 96% (Blom *et al.*, 1999 ; Blom *et al.*, 2004).

NetPhos - 3.1

Generic phosphorylation sites in eukaryotic proteins

The **NetPhos 3.1** server predicts serine, threonine or tyrosine phosphorylation sites in eukaryotic proteins using ensembles of neural networks. Both generic and kinase specific predictions are performed. The **generic** predictions are identical to the predictions performed by NetPhos 2.0. The **kinase specific** predictions are identical to the predictions by NetPhosK 1.0. Predictions are made for the following 17 kinases:

ATM, CKI, CKII, CaM-II, DNAPK, EGFR, GSK3, INSR, PKA, PKB, PKC, PKG, RSK, SRC, cdc2, cdk5 and p38MAPK.

Submission	Instructions	Output format	PhosphoBase	Downloads
------------	--------------	---------------	-------------	-----------

Submission

Sequence submission: paste the sequence(s) and/or upload a local file
 Paste a single sequence or several sequences in FASTA format into the field below:

Submit a file in FASTA format directly from your local disk:
 Aucun fichier sélectionné.

Residues to predict serine threonine tyrosine all three
For each residue display only the best prediction

Figure 3.9: Interface de NetPhos 3.1

5. Visualisation de la topologie membranaire et les sites de phosphorylation des connexines

Enfin, nous avons résumé visuellement nos résultats en utilisant Protter (<http://wlab.ethz.ch/protter>) (figure 3.10), un outil open-source conçu par Omasits et collaborateurs (2014) pour la visualisation des protéoformes. À l'aide de Protter, nous avons illustré les différents domaines de nos protéines et annoté les sites de phosphorylation prédits, afin de communiquer les aspects essentiels de notre recherche.

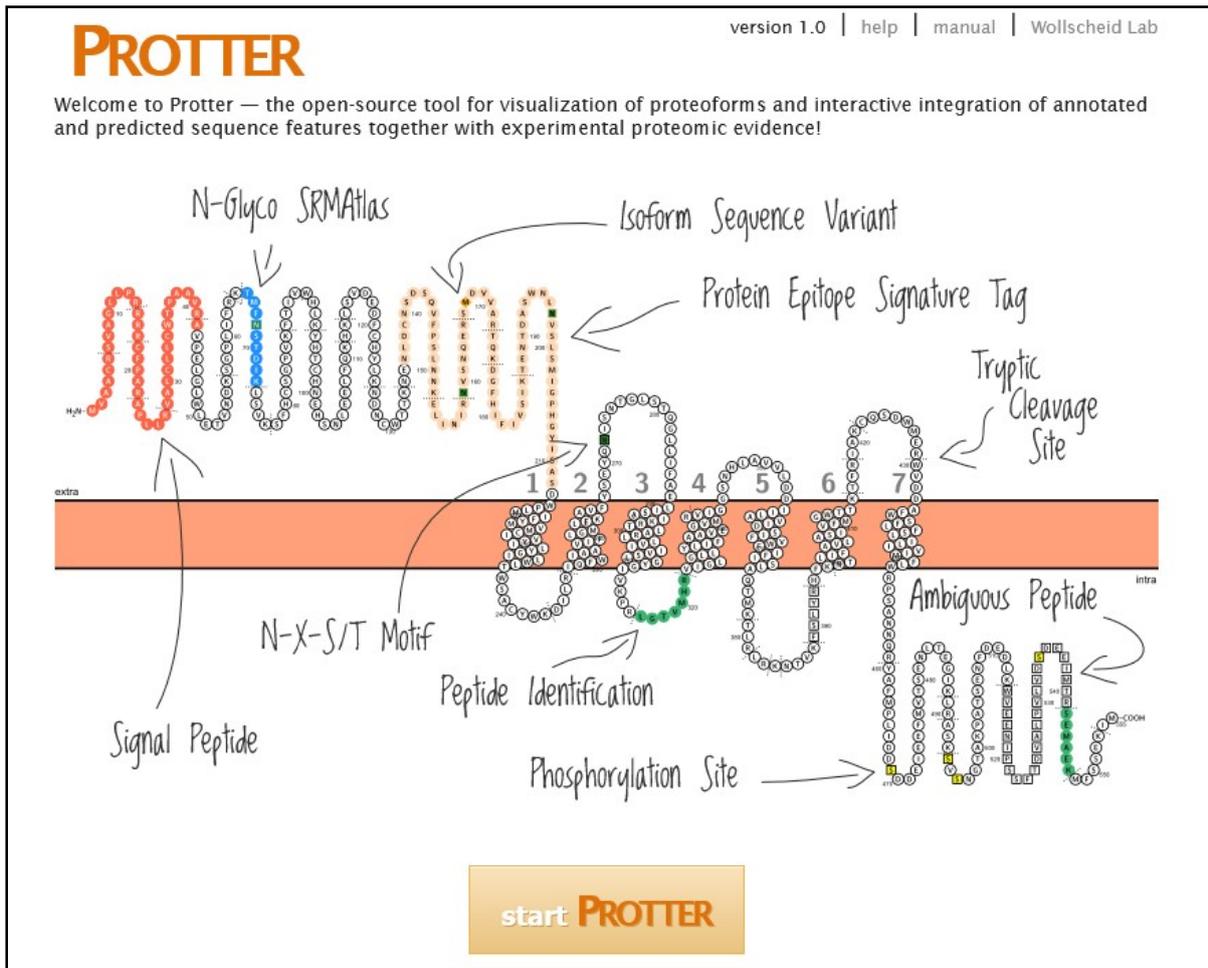


Figure 3.10: Interface de Protter

Chapitre IV :

Résultats et Discussion

1. Analyse de la topologie membranaire de la Cx43 et la Cx26

1.1 Résultats de ProtScale

L'analyse des profils d'hydrophobie pour chaque isoforme montre la présence de quatre régions hydrophobes. Pour la Cx43, ces domaines correspondent aux séquences : 24-44 ; 77-97 ; 156-176 ; 208-228. Quant à la Cx26, ils correspondent aux séquences : 21-40 ; 74-94 ; 136-156 ; 190-210.

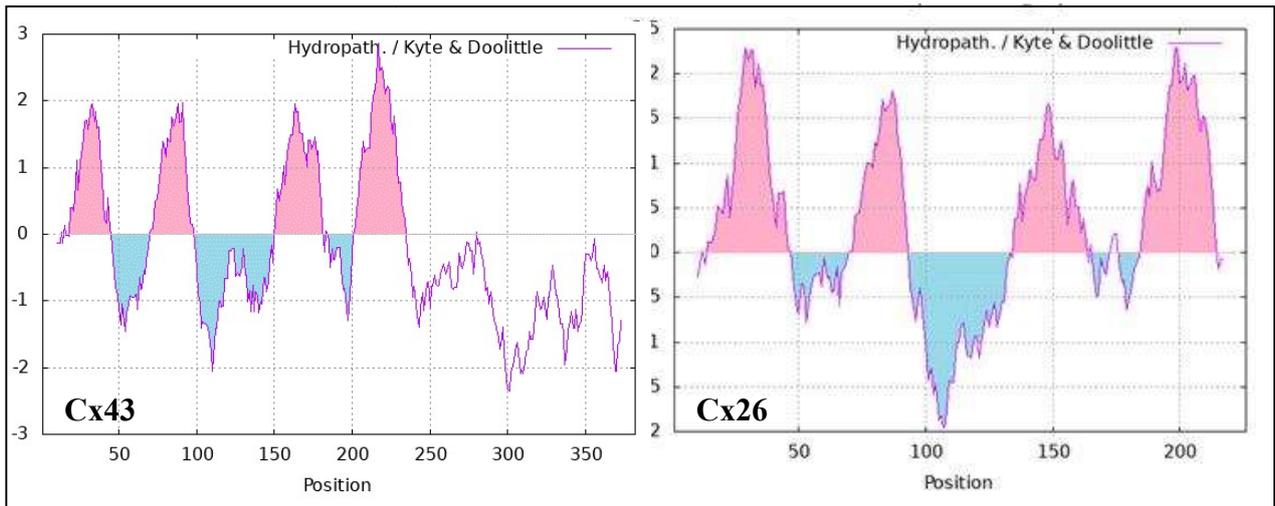


Figure 4.1 : Profils d'hydrophobie de la Cx43 et la Cx26 par ProtScale

1.2 Résultats de ΔG predictor 1.0

L'analyse des diagrammes d'hydrophobicité prédit la présence de quatre domaines transmembranaires qui adoptent un motif de repliement en hélice alpha (α). Pour la Cx43, ces domaines correspondent aux séquences : 23-41 ; 75-94 ; 151-172 ; 207-229. Quant à la Cx26, ils correspondent aux séquences : 22-44 ; 74-93 ; 132-154 ; 191-213.

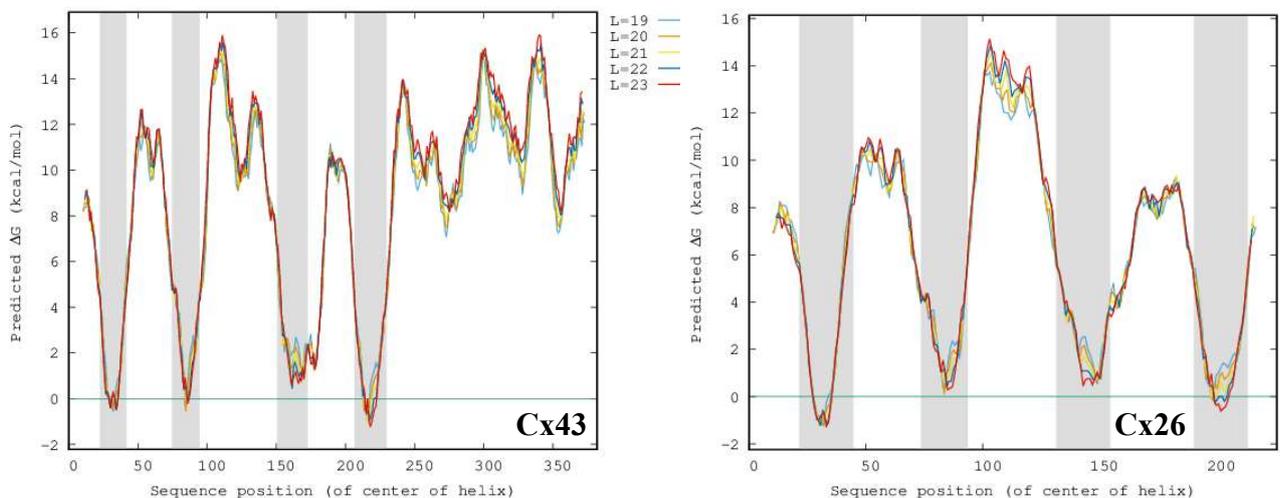


Figure 4.2 : Diagrammes d'hydrophobicité de la Cx43 et la Cx26 par ΔG predictor 1.0

1.3 Résultats de TOPCONS

Par l'analyse de la représentation schématique de la Cx43 et la Cx26, les résultats précédents sont confirmés avec des précisions sur la localisation et l'orientation de ses domaines par rapport à la membrane cellulaire. Le logiciel TOPCONS identifie quatre régions transmembranaires hydrophobes (pour la Cx43: 21-41 ; 78-98 ; 156-176 ; 208-228 et pour la Cx26: 22-42 ; 76-96 ; 132-152 ; 192-212) et indique que les extrémités N-terminale et C-terminale sont dirigées vers l'intérieur de la cellule.

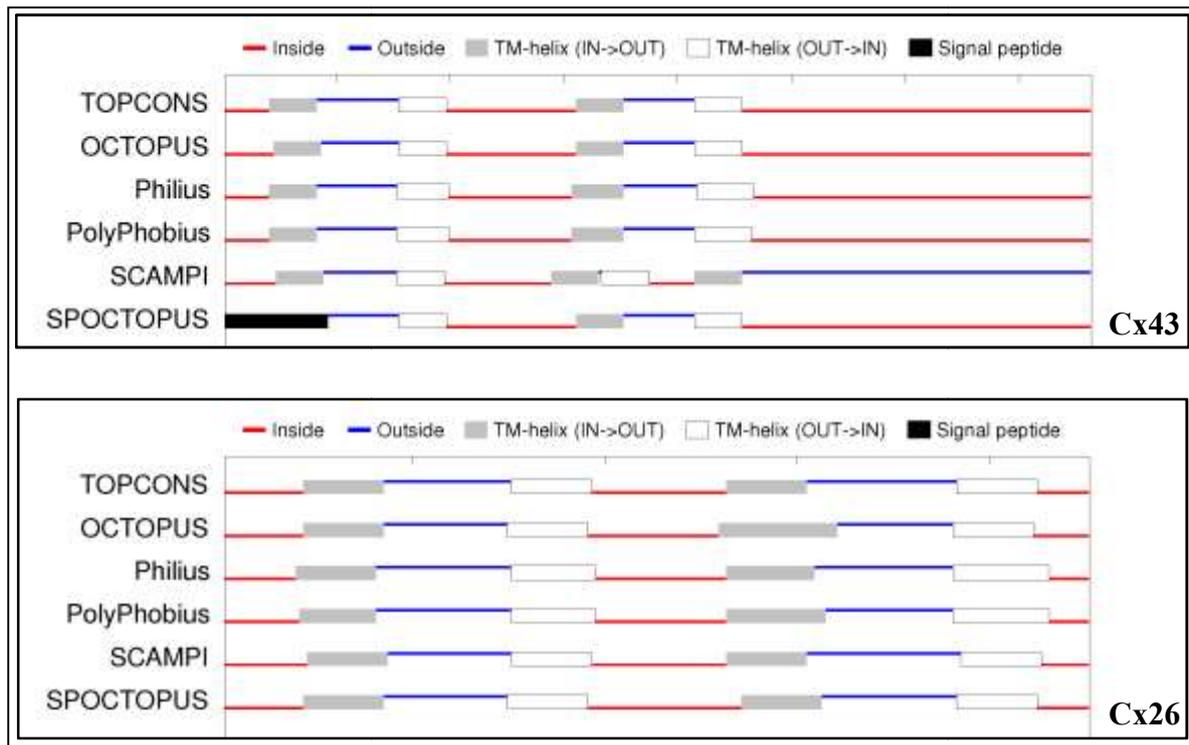


Figure 4.3 : Topologie membranaire prédite de la Cx43 et la Cx26 par TOPCONS

1.4 Prédiction de la topologie de la Cx43 et la Cx26 par les trois méthodes

Pour minimiser l'erreur dans le choix des limites de la séquence de chaque segment, nous avons sélectionné la valeur la plus élevée à la borne inférieure et la valeur la plus basse à la borne supérieure des valeurs de séquence fournies par chaque programme. Cela augmente les chances que les acides aminés prédits soient transmembranaires.

Le tableau 4.1 classe ces valeurs réduites, que nous avons ensuite utilisées pour générer une représentation visuelle (figure 4.4) de chaque isoforme à l'aide de DOG2.0.1 (Ren *et al.*, 2009) qui est un logiciel illustre des domaines protéiques. Grâce à cette approche, nous avons pu représenter les segments transmembranaires de chaque connexine. Un tableau

détaillant les prédictions pour toutes les isoformes impliquées dans notre étude est donné en annexe.

Tableau 4.1 : Tableau récapitulatif des prédictions des trois programmes

		M1	M2	M3	M4	Séquence prédite
Cx43	Protoscale	24-44	77-97	156-176	208-228	V24-V41 ; W78-A94 ; I156-I172 ; I208-L228
	Δ G predictor	23-41	75-94	151-172	207-229	
	TOPCONS	21-41	78-98	156-176	208-228	
Cx26	Protoscale	21-40	74-94	136-156	190-210	K22-A40 ; L76-M93 ; Y136-Y152 ; T192-L210
	Δ G predictor	22-44	74-93	132-154	191-213	
	TOPCONS	22-42	76-96	132-152	192-212	

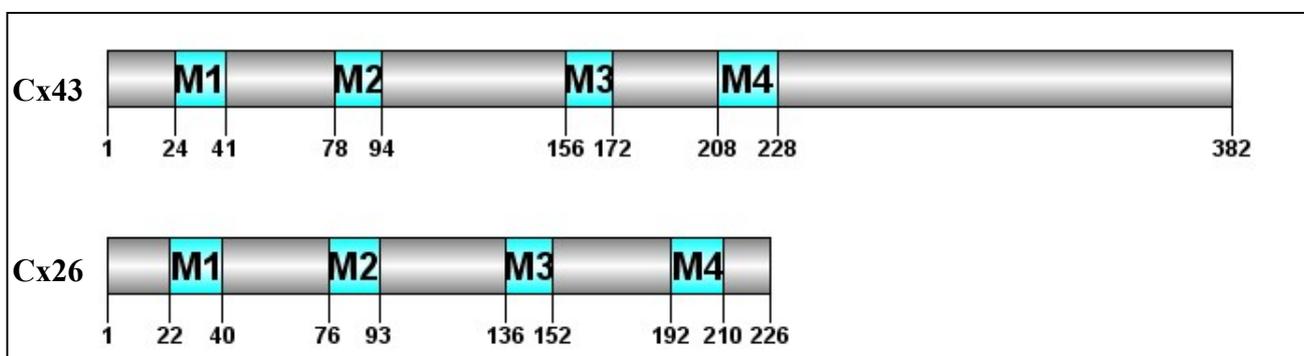


Figure 4.4 : Représentation en domaines de la Cx43 et la Cx26

2. Alignement des séquences multiple des 10 isoformes

2.1 Alignement des connexines Alpha

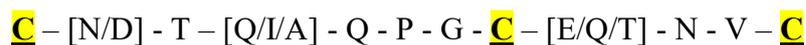
La figure 4.5 présent le résultat de l'alignement multiple des isoformes Cx37, Cx40, Cx43 et Cx46, qui font partie de la famille des Alpha connexines, effectué à l'aide de l'outil Clustal Omega et coloré en fonction du pourcentage d'identité par Jalview.

Nous observons que les domaines M1 et M2 présentent plutôt une amphiphilie, avec une haute conservation des acides aminés hydrophobes (L, V, F, I, P, W, et A), ainsi que des acides aminés hydrophiles (G, K, S, Y, Q et R). Cela est également visualisé dans l'histogramme. Ces résultats sont en accord avec les travaux de Maeda et collaborateurs (2009), qui ont précisé que ces domaines constituent la face interne du pore (côté hydrophile de leurs hélices alpha fait face à la lumière du canal).

Le domaine M3 présente plusieurs acides aminés conservés : L, T, Y, S, E, F et Q, parmi lesquels les résidus d'acides aminés aromatiques, la tyrosine (Y) et la phénylalanine (F). Ces résidus aromatiques, conformément aux résultats de Maeda et collaborateurs (2009), permettent à cette région de se loger efficacement dans les sillons des domaines transmembranaires voisins.

Le domaine M4 est majoritairement composé de résidus hydrophobes conservés (F, I, M, L et V), ainsi que de quelques résidus hydrophiles (S, E, K). Cette nature principalement hydrophobe lui permet de former, en association avec le domaine M3, l'extérieur de la paroi du canal, faisant face aux lipides de la membrane (Maeda *et al.*, 2009).

L'alignement multiple révèle également un degré très élevé de conservation dans les boucles extracellulaires E1 et E2 qui sont essentiels à l'arrimage des héli-canaux. En particulier pour les trois résidus de cystéine. Dans E1, ces résidus occupent les positions C54, C61 et C65, suivant une séquence consensus CX₆CX₃C:



Et pour la boucle E2, les trois résidus de cystéines C187, C192 et C198 sont conservés selon une séquence consensus exacte CX₄CX₅C :



Ces résultats sont soutenus par les travaux de John et Revel (1991), qui ont détaillé la séquence consensus pour chaque boucle extracellulaire et ont mis en évidence l'importance de la conservation de ces résidus cystéine dans le maintien de la stabilité de la structure tertiaire du canal (Foote *et al.*, 1998) grâce à la formation de ponts disulfures intramoléculaires.

En revanche, la boucle intracellulaire présente une faible conservation. Cela implique que la séquence codante de cette région diffère considérablement d'une isoforme à l'autre, à l'exception de quelques résidus d'acide glutamique (E), d'histidine (H), d'arginine (R) et de glycine (G) qui semblent être conservés.

Par ailleurs, l'analyse de la limite boucle « I » domaine M3 a révélé la présence de résidus d'acides aminés basiques conservés, l'arginine (R) et la lysine (K), formant un motif d'acide aminé caractéristique qui est corrélé à la compatibilité hétéromérique des connexines alpha et joue un rôle crucial dans le processus d'oligomérisation des isoformes de ce groupe en connexons. Selon les travaux de Koval et collaborateurs (2014), ces connexines sont désignées sous l'appellation connexines de type « R ».

La région C-terminale est la moins conservée et présente une divergence considérable entre les quatre isoformes, avec des séquences et des longueurs variées. Cette divergence est clairement visible grâce au grand nombre d'insertions « gap » introduites lors de l'alignement. Cependant, nous pouvons observer quelques résidus conservés, notamment plusieurs résidus de sérine (S) qui, selon les travaux de Lampe et Lau (2004), sont des sites de phosphorylation par les kinases et jouent un rôle important dans la régulation de la communication intercellulaire par jonction gap (GJC).

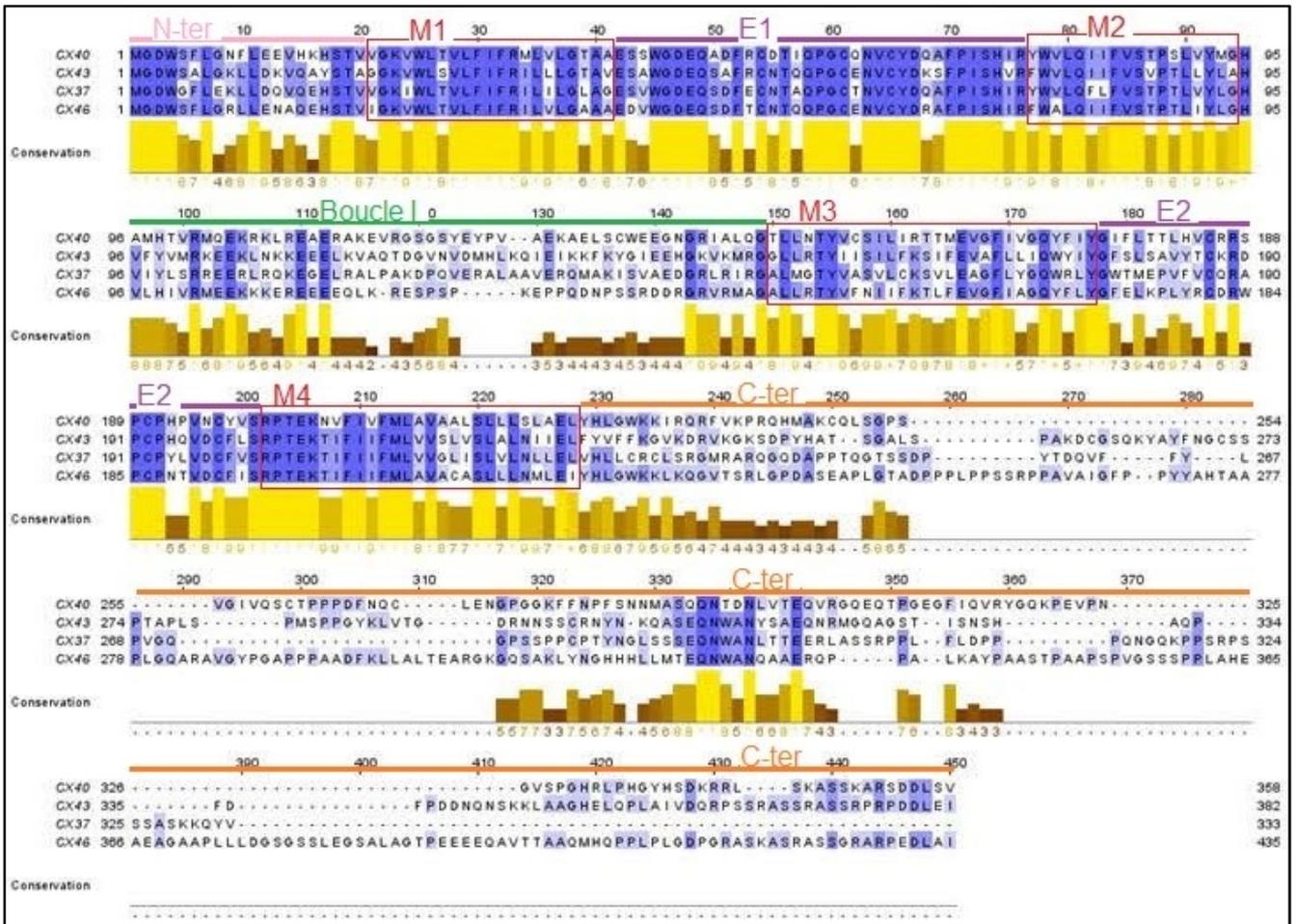


Figure 4.5 : Alignement multiple des isoformes Cx37, Cx40, Cx43 et Cx46

Le programme Clustal Omega a fourni une matrice contenant les scores d'alignement par paires de toutes les isoformes. En examinant la matrice d'identité des connexines Alpha (figure 4.6), nous observons un taux d'identité d'environ 50%.

Les séquences les plus proches sont celles de la Cx46 et de la Cx37, avec un pourcentage d'identité de 48,01%. Ensuite, les connexines Cx43 et Cx37 montrent un taux d'identité de 47,43%.

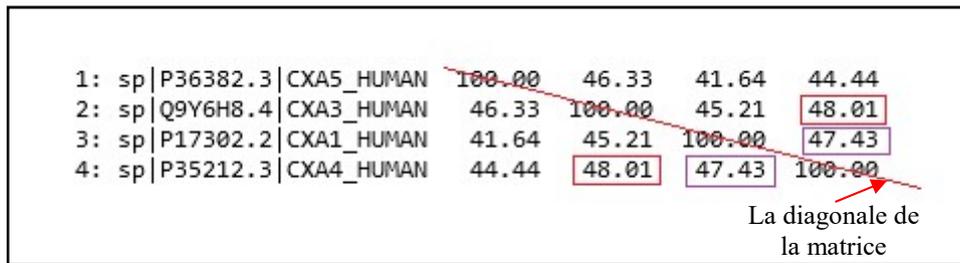


Figure 4.6 : Matrice du pourcentage d'identité des isoformes Cx37, Cx40, Cx43 et Cx46

2.2 Alignement des connexines Bêta

La figure 4.7 présente l'alignement multiple des connexines Cx26, Cx30, Cx30.3, Cx31, Cx31.1 et Cx32, qui font partie de la famille des Bêta connexines.

L'analyse des domaines transmembranaires M1 et M2 montre un haut degré de conservation, où il apparaît que ces domaines sont de nature amphiphile en raison de la présence de résidus hydrophobes (L, V, F, I, A, W, P et M), ainsi que de résidus hydrophiles (R et Q). Ces observations sont fortement soutenues par les travaux approfondis de Maeda et collaborateurs (2009), qui ont réalisé une étude détaillée de la structure du canal jonctionnel de la Cx26 et ont attribué cette amphiphilie au rôle des domaines M1 et M2 dans la formation du pore du canal.

L'analyse de l'alignement au niveau du domaine transmembranaire M3 révèle la conservation des résidus aromatiques W, Y et F. Ces résultats sont similaires à ceux observés chez les connexines Alpha et sont en accord avec les travaux de Maeda et collaborateurs (2009). Ce domaine présente également la conservation d'autres résidus, à savoir G, L, T et S.

De plus, on observe la présence d'un motif di-tryptophane (WW) très conservé au début du domaine M3, ce motif étant corrélé à la compatibilité hétéromérique sélective des connexines Bêta au sein de la même sous-famille, également appelées connexines de type « W », selon les travaux de Koval et collaborateurs (2014).

Le domaine M4 est en grande partie composé de résidus hydrophobes conservés (F, M, I et L), avec quelques résidus hydrophiles (T, E, K, N). Ces résultats sont similaires à ceux obtenus pour le domaine M4 des connexines Alpha et sont en accord avec les travaux de Maeda et collaborateurs (2009).

Les boucles extracellulaires E1 et E2 sont également hautement conservées. Elles contiennent chacune trois résidus cystéines avec un nombre constant d'acides aminés qui les

séparent, selon des séquences consensus prédéterminées (John et Revel, 1991). Pour la boucle E1, la séquence est CX₆CX₃C, qui correspond à :



Et pour la boucle E2, les trois résidus de cystéines sont conservés selon la séquence consensus CX₄CX₅C, qui correspond à :



La boucle intracytoplasmique (I), en revanche, est très peu conservée car elle présente une similarité minimale entre les séquences de chaque isoforme. Cependant, il est important de noter qu'il existe certains résidus hydrophiles conservés, à savoir l'arginine (R) et l'acide glutamique (E) ainsi que la tyrosine (Y).

La plus faible conservation chez les connexines Bêta se trouve dans leur domaine C-terminal, comme le montrent les nombreux gaps introduits par le programme d'alignement. Cela est dû aux longueurs variables ainsi qu'aux différentes séquences présentes dans chaque isoforme. Cependant, nous observons la présence d'un résidu tyrosine (Y), leucine (L) et cystéine (C) hautement conservés au début de ce segment. Les connexines Bêta présentent également un domaine C-terminal plus court par rapport aux connexines Alpha.

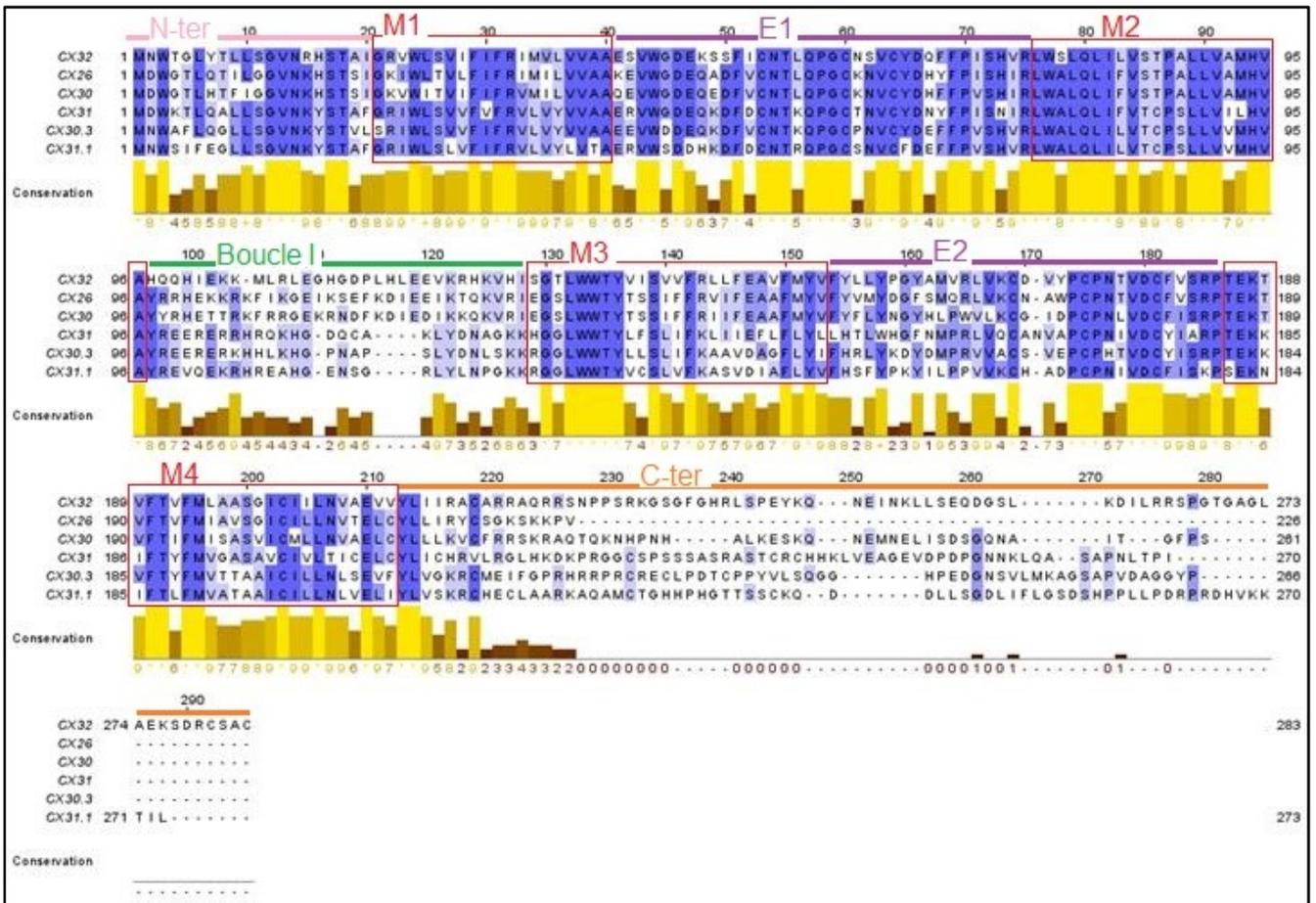


Figure 4.7 : Alignement multiple des isoformes Cx26, Cx30, Cx30.3, Cx31, Cx31.1 et Cx32

La matrice d'identité des connexines bêta (figure 4.8) révèle un taux de conservation plus élevé que celui des connexines alpha, ce qui indique une plus grande conservation entre les isoformes de cette famille. Les séquences les plus proches parmi eux sont celles de Cx26 et Cx30, qui présentent un taux d'identité de 76,55 %. Ensuite, les connexines Cx26 et Cx32 montrent un taux d'identité de 63,56 %.

1:	sp	P08034.1	CXB1_HUMAN	100.00	63.56	56.15	43.85	48.44	43.56
2:	sp	P29033.3	CXB2_HUMAN	63.56	100.00	76.55	52.94	51.13	47.96
3:	sp	O95452.2	CXB6_HUMAN	56.15	76.55	100.00	45.06	46.59	45.38
4:	sp	O75712.1	CXB3_HUMAN	43.85	52.94	45.06	100.00	58.78	55.00
5:	sp	Q9NTQ9.1	CXB4_HUMAN	48.44	51.13	46.59	58.78	100.00	59.85
6:	sp	O95377.2	CXB5_HUMAN	43.56	47.96	45.38	55.00	59.85	100.00

La diagonale de la matrice

Figure 4.8 : Matrice de pourcentage d'identité des isoformes Cx26, Cx30, Cx30.3, Cx31, Cx31.1 et Cx32

La grande conservation observée dans les connexines bêta par rapport aux connexines alpha peut potentiellement être attribuée à la longueur de ses séquences, car elles sont plus courtes que les connexines alpha, en particulier dans les boucles intracytoplasmiques (en comparant le nombre d'acides aminés (31 vs 54)) et la région C-terminale, qui sont les segments les plus divergents.

3. Prédiction des sites potentiels de phosphorylation

Netphos3.1 donne les résultats des sites de phosphorylation prédits sous la forme de neuf acides aminés centrés autour du site accepteur du phosphate (P-4 **S*** P+4) tout en associant des scores de probabilité pour chaque prédiction. Nous avons choisi une valeur seuil de prédiction de 0.5 dans notre travail. Les meilleurs scores sont présentés dans le tableau 4.2 pour la Cx43 et dans le tableau 4.3 pour la Cx26 avec leurs kinases respectives.

3.1 Sites de phosphorylation prédits de la Cx43

Tableau 4.2 : Prédictions des sites de phosphorylation de la Cx43 par Netphos3.1

Site de phosphorylation	Score	kinase	Séquence	Site déterminé expérimentalement
S255	0.531	MAPK	SGALS*PAKD	+
S262	0.838	PKC	KDCGS*QKYA	+
S279	0.698	cdk5	TAPLS*PMSP	+
	0.652	MAPK		
S282	0.630	cdk5	LSPMS*PPGY	+
	0.588	MAPK		
S297	0.533	PKC	RNNS*CRNY	+
S364	0.824	PKC	DQRPS*SRAS	-
S365	0.503	GSK3	QRPS*RASS	-
S368	0.862	PKC	SSRAS*SRAS	+
S369	0.501	RSK	SRAS*RASS	-
S372	0.887	PKC	SSRAS*SRPR	+

L'analyse du tableau 4.2 montre que tous les sites prédits pour la Cx43 se trouvent au niveau du C-terminal. Ceci est en accord avec les travaux d'Axelsen et collaborateurs (2013).

Par ailleurs, ces sites seront classés en deux groupes : sites expérimentalement prouvés et sites non prouvés. En examinant les trois sites de phosphorylation, S364, S365 et S369, dont la phosphorylation est prédite respectivement par la PKC, GSK3 et la RSK, nous observons un score élevé pour la première kinase et faible pour les deux autres.

Le premier site prédit, S364 (sérine 364), auquel a été attribué un score de 0,824, se trouve au niveau du domaine C-terminal de la Cx43. Kemp et Pearson (1990) ont étudié la phosphorylation des résidus sérine par la protéine kinase C et ont proposé un motif de reconnaissance pour cette kinase correspondant à $xRxxS^*xRx$ (où x peut être n'importe quel acide aminé). En comparant cette séquence consensus avec notre séquence DQRPS*SRAS, nous constatons qu'elle satisfait partiellement aux exigences du motif où la position P+2 est occupée par un résidu d'arginine. Par ailleurs, Le résidu arginine de la séquence consensus occupe la position P-3, alors que dans notre séquence, il est plus proche de la sérine et occupe la position P-2, suggérant que le nombre d'acides aminés séparant le site accepteur de phosphate et cet acide aminé basique n'est peut être pas important

Le deuxième site prédit, S365 (sérine 365), qui a un score de 0,503, est également situé dans le domaine C-terminal de Cx43. Shinde et collaborateurs (2017) étaient plus intéressés à la glycogène synthase kinase-3 (GSK3) et ont proposé le motif $(S^*/T^*)xxx(S/T)$ comme séquence consensus de l'enzyme. En examinant cette séquence consensus avec notre séquence QRPS*RASS, nous constatons une correspondance parfaite.

Le troisième site prédit correspond à la sérine 369 (S369) avec un score de 0.501, et il se localise dans le domaine C-terminal de la Cx43. Kemp et Pearson (1990) ont également montré le motif de reconnaissance par la kinase S6 ribosomique (RSK), qui est $xRxxS^*x$. En comparant ce motif avec notre séquence SRASS*RASS, nous constatons une exacte similitude.

Par ailleurs, aucune étude expérimentale n'a montré la phosphorylation des connexines par ces deux enzymes, GSK3 et RSK. Par conséquent, il est nécessaire de réaliser des études supplémentaires afin d'éclaircir ce point et de confirmer ou de réfuter cette prédiction.

NetPhos 3.1 a également prédit des sites de phosphorylation vérifiés expérimentalement, S262/S297/S368/S372 prédits par la PKC, S255/S279/S282 prédits par les MAPK et S279/S282 prédits par la cdk5. Ces sites de phosphorylation sont largement documentés dans la littérature.

La phosphorylation par la PKC entraîne une réduction de la perméabilité du canal, la phosphorylation de S262 a été liée également à une augmentation de la prolifération cellulaire par un mécanisme inconnu (Doble *et al.*, 2004), tandis que la phosphorylation de S297 réduit la communication intercellulaire par jonctions *gap* (GJIC) mais la signification fonctionnelle de ce site est actuellement inconnue (Procida *et al.*, 2009). La phosphorylation de la Cx43 au site S368 entraîne une réduction de la conductance unitaire du canal, il a également été démontré qu'elle est associée à l'ubiquitination de la protéine et au désassemblage des jonctions *gap* (Liao *et al.*, 2013), tandis que la prévention de la phosphorylation de S368 a augmenté la localisation de la Cx43 à la membrane cellulaire (Wang *et al.*, 2013). Pour le site S372, l'effet régulateur de la phosphorylation par la PKC n'a pas été encore démontré.

La phosphorylation par les MAPK sur les sites S255/279/282 facilite la fermeture des canaux et le désassemblage des jonctions *gap* (Lastwika *et al.*, 2019) par internalisation dans des connexosomes, où ils sont dégradés par les protéasomes (Solan et Lampe, 2020).

Par ailleurs, la kinase Cdk5 a été prédite pour phosphoryler la Cx43 au niveau de deux sérines S279 et S282. Selon Qi et collaborateurs (2016), cette phosphorylation inhibe le transport de la Cx43 jusqu'à la membrane et favorise l'internalisation et la dégradation de façon protéasome-dépendante.

3.2 Sites de phosphorylation prédits de la Cx26

Tableau 4.3 : Prédictions des sites de phosphorylation de la Cx26 par Netphos3.1

Site de phosphorylation	Score	kinase	Séquence	Site déterminé expérimentalement
S219	0.910	PKC	IRYCS*GKSK	-

Pour la Cx26, les résultats de NetPhos3.1 étaient très limités, car cette isoforme possède une région C-terminale très courte. Nous avons réussi à prédire un site de phosphorylation correspondant à la sérine S219 par la kinase prédite PKC et qui se trouve dans la séquence IRYCS*GKSK du domaine C-terminal. Le programme a attribué un score très élevé à cette prédiction, soit 0.910. La PKC reconnaît le motif xRxxS*xRx (Kemp et Pearson, 1990). Il faut noter que dans notre prédiction, l'arginine qui se trouve à la position P+2 du site accepteur est remplacée par un autre acide aminé basique, la lysine (K).

4. Visualisation des domaines et sites prédits

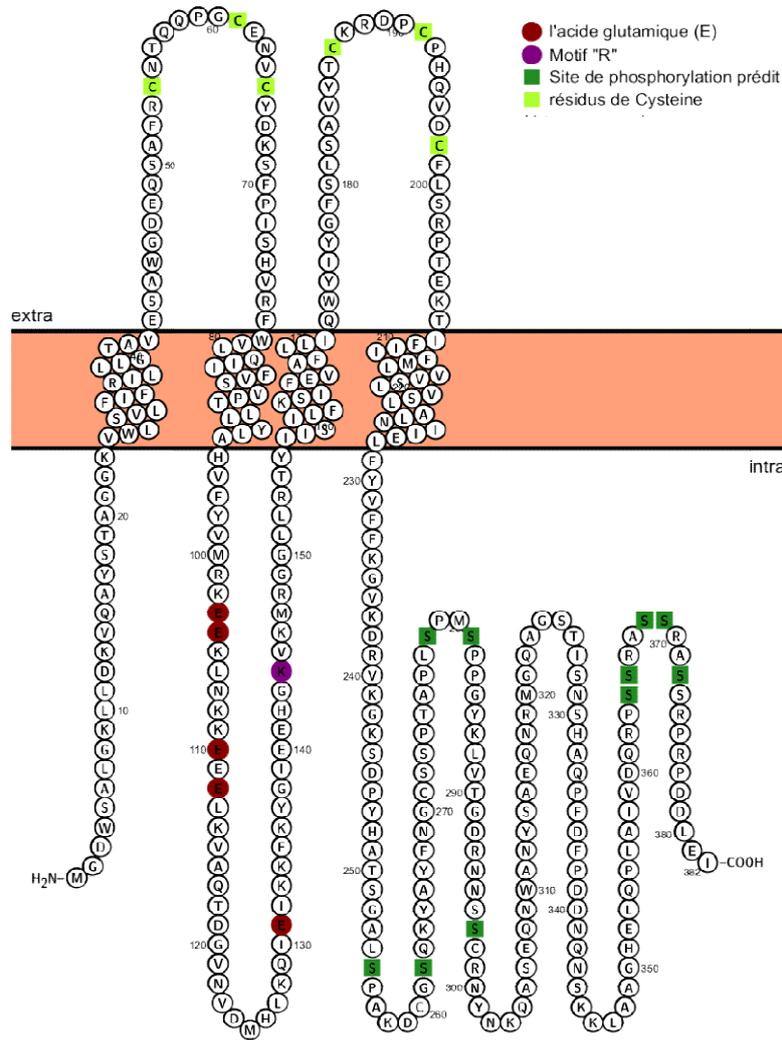


Figure 4.9 : Visualisation de la topologie et des sites de phosphorylation de la Cx43

Représentation de la topologie membranaire de hCx43 et des sites de phosphorylation prédits basés sur des analyses de séquence utilisant différents programmes réalisés dans ce travail. hCx43 est composée de quatre régions transmembranaires hydrophobes (M1, M2, M3 et M4) reliées par une boucle intracellulaire (I) qui présente le motif «R» caractéristique du groupe Alpha. Deux boucles extracellulaires (E1 et E2) où elles contiennent chacune 3 résidus de cystéines conservés. Les extrémités N-terminales et C-terminales sont intracytoplasmiques. Les sites de phosphorylation prédits : S255, S262, S279, S282, S297, S364, S365, S368, S369 et S372 se trouvent au niveau du C-terminal.

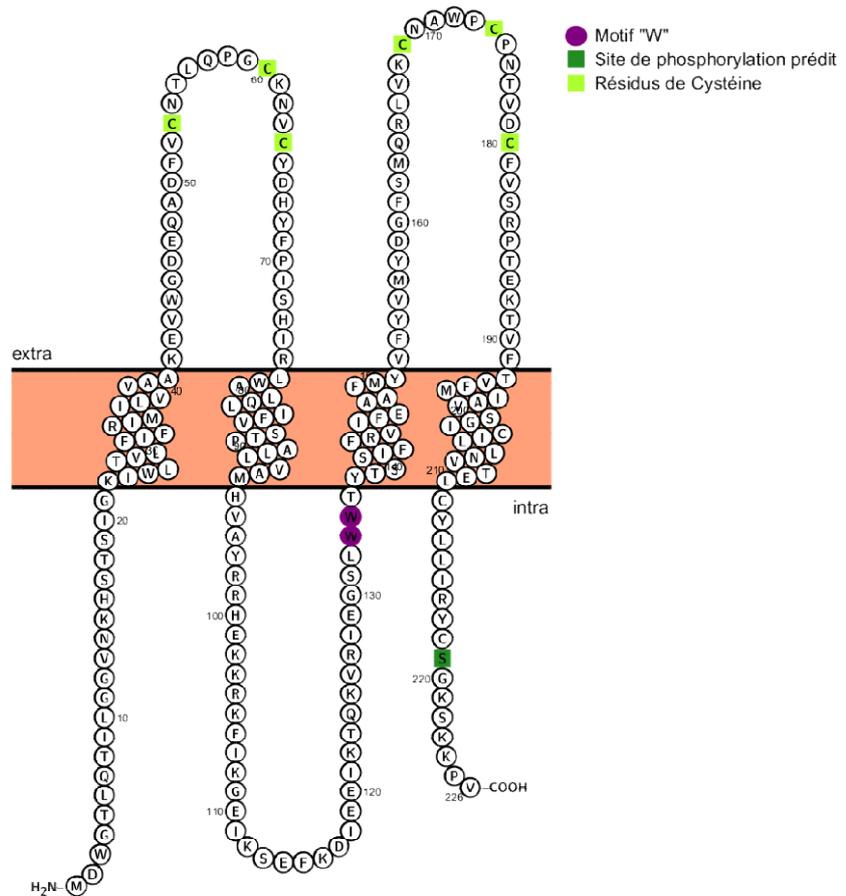


Figure 4.10 : Visualisation de la topologie et des sites de phosphorylation de la Cx26

Représentation de la topologie membranaire de hCx26 et des sites de phosphorylation prédits basés sur des analyses de séquence utilisant différents programmes réalisés dans ce travail. hCx26 est composée de quatre régions transmembranaires hydrophobes (M1, M2, M3 et M4) reliées par une boucle intracellulaire (I) qui présente le motif «WW» caractéristique du groupe Beta. Deux boucles extracellulaires (E1 et E2) où elles contiennent chacune 3 résidus de cystéines conservés. Les extrémités N-terminales et C-terminales sont intracytoplasmiques. Le domaine C-terminal renferme le site de phosphorylation prédit : S219

Conclusion

Les connexines sont des protéines transmembranaires appartenant à la famille des *tetraspans* impliquées dans la formation de structures cellulaires uniques appelées jonctions communicantes. Elles forment une grande famille de vingt-et-une isoformes dans le génome humain, présentant une forte homologie. Les connexines sont largement distribuées dans tous les tissus.

La fonction des jonctions communicantes est étroitement régulée par l'augmentation des ions de calcium (Ca^{2+}) et la baisse du pH. De plus, les modifications post-traductionnelles, notamment la phosphorylation joue aussi un rôle très important dans le transport des connexines jusqu'à la membrane plasmique mais également dans la régulation de la perméabilité jonctionnelle.

Dans ce travail, nous avons effectué une analyse comparative de deux groupes principaux des connexines Alpha et Beta exprimés au niveau du tissu épithélial à l'aide des programmes bioinformatiques disponibles en ligne en employant des séquences protéiques à partir des bases de données spécialisées.

L'étude de la topologie membranaire à l'aide des outils de prédiction ProtScale, ΔG Predictor et TOPCONS a montré que la Cx43 et la Cx26 comportent quatre régions transmembranaires hydrophobes (M1, M2, M3 et M4) reliées par une boucle intracellulaire (I) et deux boucles extracellulaires (E1 et E2). Ces boucles extracellulaires contiennent chacune 3 résidus de cystéine caractéristiques de la famille des connexines. Les extrémités N-terminale et C-terminale se trouvent du côté intracytoplasmique. Ces programmes également ont prédits les limites des segments transmembranaires.

L'alignement multiple des séquences de chaque groupe a présenté des résultats similaires, par rapport à la forte conservation au niveau de leur domaine N-terminal, de leurs domaines transmembranaires M1-M4 et de leurs boucles extracellulaires E1 et E2. En revanche, nous avons constaté aussi une forte divergence au niveau de la boucle intracytoplasmique et du domaine C-terminal en mettant en évidence la différence de longueur de ces domaines non conservés entre les deux groupes. De plus, notre travail montre une différence structurale entre les deux groupes qui est importante dans le phénomène d'oligomérisation des différentes isoformes. Nous avons ainsi identifié des motifs spécifiques à chaque groupe qui sont corrélés à la compatibilité hétéromérique des connexines. Nous avons observé la présence du motif de type "R" contenant un résidu d'arginine ou de lysine conservé, qui distingue les connexines du

groupe Alpha. Nous avons également identifié un motif di-tryptophane hautement conservé appelé motif "W", qui est fortement corrélé avec les connexines du groupe Bêta.

Le domaine C-terminal des connexines renferme de nombreux sites de modifications post-traductionnelles. Dans notre étude, l'utilisation du programme NetPhos3.1 a permis de prédire la présence de plusieurs sites de phosphorylation potentiels. Nous avons identifié trois sites (S364, S365 et S369) dans le domaine C-terminal de la Cx43, ainsi qu'un site S219 dans le domaine C-terminal de la Cx26. Ces sites n'ont pas encore été identifiés expérimentalement et nécessitent des études et des expériences ultérieures pour confirmer ou réfuter nos prédictions.

En revanche, nous avons également identifié sept sites (S255, S262, S279, S282, S297, S368 et S372) au niveau du domaine C-terminal de la Cx43 qui ont été soutenus par des expériences antérieures. Par exemple, la phosphorylation de S368 par la PKC a été démontrée et elle entraîne une réduction de la conductance unitaire du canal, conduisant à sa fermeture.

En perspective, nous anticipons que l'utilisation des méthodes de bioinformatique sera l'un des outils les plus puissants à la disposition des scientifiques pour comprendre la régulation des connexines. Nous envisageons de revoir les prédictions que nous avons effectuées lors de cette étude en utilisant d'autres outils de prédiction et/ou pour d'autres modifications post-traductionnelles, voire même au sein d'autres tissus, afin de mieux comprendre les mécanismes de ces protéines.

Annexes

Annexe 1

Alignement multiple des isoformes Cx37, Cx40, Cx43 et Cx46

sp P36382.3 CXA5_HUMAN	MGDWSFLGNFLEEVEHKHSTVVGKVVLTVLFIFRMLVLGTAAESSWGDEQADFRCDTIQPG	60
sp Q9Y6H8.4 CXA3_HUMAN	MGDWSFLGRLLLENAQEHSTVIGKVVLTVLFIFRILVLGAAAEDVWGDEQSDFTCNTQQPG	60
sp P17302.2 CXA1_HUMAN	MGDWSALGKLLDKVQAYSTAGGKVVLSVLFIFRILLGTAVESAWGDEQSAFRCNTQQPG	60
sp P35212.3 CXA4_HUMAN	MGDWGFLEKLLDQVQEHSTVVGKIWLTVLFIFRILILGLAGESVWGDEQSDFECNTAQQG **** * .:*.:.: ;** .**:**.*****;*:* * . *****: * *:* **	60
sp P36382.3 CXA5_HUMAN	CQNVCYDQAFPIHSHIRYVWLQIIFVSTPSLVYMGHAMHTVRMQEKRLREAERAKEVRGS	120
sp Q9Y6H8.4 CXA3_HUMAN	CENVCYDRAFPISHIRFWALQIIFVSTPTLIYLGHVLIHVRMEEKKEREEREEQLKRESP	120
sp P17302.2 CXA1_HUMAN	CENVCYDKSFPISHVRFVWLQIIFVSVPTLLYLAHVYVVRMKEEKLNKKEELKVAQTDG	120
sp P35212.3 CXA4_HUMAN	CTNVCYDQAFPIHSHIRYVWLQFIFVSTPTLVYLGHVLYLSRREERLRQKEGELRALPAKD * *****:*****;*:*:**.***;*:*:*:*:*:* * :*: . : * *	120
sp P36382.3 CXA5_HUMAN	GSYEE--PVAEKAELSCWEEGNGRIALQGTLLNTYVCSILIRTTMEVGFIVGQYFIYGIF	178
sp Q9Y6H8.4 CXA3_HUMAN	SP-----KEPPQDNPSRDDRGRVRMAGALLRTYVFNIIFKTLFEVGFIAQYFLYGFE	174
sp P17302.2 CXA1_HUMAN	VNVDMLKQIEIKFKYGEIHHGKVKMRGGLLRTYIISILFKSIFEVAFLLIQWYIYGFS	180
sp P35212.3 CXA4_HUMAN	PQVERALAAVERQMAKISVAEDGRLRIRGALMGTYVASVLCCKSVLEAGFLYQWRLYGWT *:: : * * : ** : .:: : : * . * : * : ;**	180
sp P36382.3 CXA5_HUMAN	LTTLHVCRRSPPHPVNCVYSRPTKENVFIVFMLAVAALSLLLLSLAELYHLGWKKIRQRF	238
sp Q9Y6H8.4 CXA3_HUMAN	LKPLYRCDRWPCPNTVDCFISRPTKTIFFIFMLAVACASLLLNMLEIYHLGWKKLQGV	234
sp P17302.2 CXA1_HUMAN	LSAVYTCKRDPCHQVDFLSRPTKTIFFIFMLVVSLVSLALNIELFVHFFKGVKDRV	240
sp P35212.3 CXA4_HUMAN	MEPVFVCQRAPCPYLVDFVSRPTKTIFFIFMLVGLISLVLLNLELVHLLCRCLSRGM : . : * * ** * : : ***** : ** : ** : * . * * * : * : : : : .	240
sp P36382.3 CXA5_HUMAN	VKPRQHMAKQLSGPSV-----GIVQS-----CTPP	264
sp Q9Y6H8.4 CXA3_HUMAN	TSRLGPDASEAPLGTADPPPLPPSSRPPAVAIGFPPYYAHTAAPLGQARAVGYPGAPPPA	294
sp P17302.2 CXA1_HUMAN	KGKSDPYHA-TSG-ALSP-AKDCGS----QKYAYFNGCSSPTAPLSPMSP-----	283
sp P35212.3 CXA4_HUMAN	RARQGQDAPPTQG-TS---SDPYTD----QVFFYLPGVQG-----	272
sp P36382.3 CXA5_HUMAN	PDFNQ-----CLENGPGGKFFNPFNSNMASQNTDNLVTEQVRGQEQTPEGGFI-----	313
sp Q9Y6H8.4 CXA3_HUMAN	ADFKLLALTEARGKQSALKYNGHHLLMTEQNWANQAERQPPALKAYPAASTPAAPSP	354
sp P17302.2 CXA1_HUMAN	PGYKLV---TGDRNNSSCRNYN---K-QASEQNWANYSAEQNRMGQAGSTISN-----	329
sp P35212.3 CXA4_HUMAN	-----PSSPPCPTYN---GLSSSEQNWANLTTEERLASSRPLFLD----- .. :* :** * :*	310
sp P36382.3 CXA5_HUMAN	-----QVRYGQ-----KP---EVPNGVSPGHRLPHGYHSDK	341
sp Q9Y6H8.4 CXA3_HUMAN	VGSSSPPLAHEAEGAAPLLLDGSGSLEGSALAGTPEEEEEQAVTTAAQMHQPPLPLGDP	414
sp P17302.2 CXA1_HUMAN	--SHAQPF-----DFPDDNQNSKKLAAGHELQPLAIVDQ	361
sp P35212.3 CXA4_HUMAN	-----PP-----PQNGQ * .	317
sp P36382.3 CXA5_HUMAN	RR----LSKASSKARSDDLVS	358
sp Q9Y6H8.4 CXA3_HUMAN	GRASKASRASSGRARPEDLAI	435
sp P17302.2 CXA1_HUMAN	RPSSRASSRASSRPRPDDLEI	382
sp P35212.3 CXA4_HUMAN	KPPSRPSSSASKKQYV----- :* :	333

Annexe 2 Aligement multiple des isoformes Cx26, Cx30, Cx30.3, Cx31, Cx31.1 et Cx32

sp P08034.1 CXB1_HUMAN	MNWTGLYTLISGVNRHSTAIGRVWLSVIFIFRIMVLVAAESVWGDEKSSIFICNTLQPGC	60
sp P29033.3 CXB2_HUMAN	MDWGTLQTLILGGVNHSTSIGKIWLTVLFIFRIMILVAAAEVWGDEQADFVFCNTLQPGC	60
sp O95452.2 CXB6_HUMAN	MDWGTLHTFIFGGVNHSTSIGKVWITVIFIFRVMILVAAQEVWGDEQEDFVFCNTLQPGC	60
sp O75712.1 CXB3_HUMAN	MDWKTQALLSGVNKYSTAFGRWLVSFVFRVLVYVAAERVWGDEQKDFDCNTKQPGC	60
sp Q9NTQ9.1 CXB4_HUMAN	MNWAFLQGLLSGVNKYSTVLSRIWLSVVFIFRVLVYVAAEVDDEQKDFVFCNTKQPGC	60
sp O95377.2 CXB5_HUMAN	MNWSIFEGLLSGVNKYSTAFGRWLVSFVIFRVLVYLVTAERVWSDDHKDFDCNTRQPGC	60
	: : : : **:* : : : **:* : : : **:* : : : **:* : : : **:* : : : **:* : : : **	
sp P08034.1 CXB1_HUMAN	NSVCYDQFFPISSHVRLWSLQLILVSTPALLVAMHVAHQHIEKK-MLRLEGHDPLHLEE	119
sp P29033.3 CXB2_HUMAN	KNVCYDHYFPISHIRLWALQLIFVSTPALLVAMHVAYRRHEKRRKFIKGEIKSEFKDIEE	120
sp O95452.2 CXB6_HUMAN	KNVCYDHFPPVSHIRLWALQLIFVSTPALLVAMHVAYRRHETTRKFRRGEKRNDFKDIED	120
sp O75712.1 CXB3_HUMAN	TNVCYDNYFPISNIRLWALQLIFVTCPSLLVILHVAYREERERRRQKHG-DQCA----K	115
sp Q9NTQ9.1 CXB4_HUMAN	PNVCYDEFFPVSHVRLWALQLILVTCPSLLVVMHVAYREERERKHHKKG-PNAP----S	115
sp O95377.2 CXB5_HUMAN	SNVCFDEFFPVSHVRLWALQLILVTCPSLLVVMHVAYREVQEKRRHREAHG-ENSG----R	115
	.*:* : : : **:* : : : **:* : : : **:* : : : **:* : : : **:* : : : **	
sp P08034.1 CXB1_HUMAN	VKRHKVHISGLWWTYVISVFRLLFEAVFMVYFYLPLPGYAMVRLVKCD-VYPCPNTVD	178
sp P29033.3 CXB2_HUMAN	IKTQKVRIEGSLWWTYSSIFFRVIFFAAMVYFVYVMDGFSMQRLVKCN-AWPCPNTVD	179
sp O95452.2 CXB6_HUMAN	IKKQKVRIEGSLWWTYSSIFFRIIFEAAMVYFVFLYNGYHLPWLKCG-IDPCPNLVD	179
sp O75712.1 CXB3_HUMAN	LYDNAGKKGGLWWTYLFSLIFKLIIEFLFLYLLHTLWHGFNMPRLVQCANVAPCPNIVD	175
sp Q9NTQ9.1 CXB4_HUMAN	LYDNLKSKRGGWLWWTYLSLIFKAAVDAGFLYIFHRLYKDYDMPRVVACS-VEPCPHTVD	174
sp O95377.2 CXB5_HUMAN	LYLNPCKKGGWLWWTYVCSLVFKASVDIAFLYFHSFYPKYILPPVVKCH-ADPCPNIVD	174
	: : : * ***** *:.* : : : **:* : : : : : : : : * ****: **	
sp P08034.1 CXB1_HUMAN	CFVSRPTEKTVFTVFMIAASGICILNVAEVVYLIIRACARRAQRSSNPPSRKGSFGHR	238
sp P29033.3 CXB2_HUMAN	CFVSRPTEKTVFTVFMIAVSGICILNVTLECYLLIRYCSGKSKKPV-----	226
sp O95452.2 CXB6_HUMAN	CFISRPEKTVFTIFMISASVICMLLNVAELCYLLKVCFRRSKRAQTQKNHPNH-----	234
sp O75712.1 CXB3_HUMAN	CYIARPEKKIFTYFMV GASAVCIVL TICELCYLICHVRLRGLHKDKPRGGCSPSSASR	235
sp Q9NTQ9.1 CXB4_HUMAN	CYISRPEKTVFTYFMVTTAAICILLNLSEVFLVVGKRCMEIFGPRHRRPRCRECLPDTC	234
sp O95377.2 CXB5_HUMAN	CFISKPSEKNIFTLMVATAAICILLNLVELIYLVSKRCHECLAARKAQAMCTGHHPHGT	234
	*: : : **:* : : : **:* : : : **:* : : : **:* : : : **:* : : : **	
sp P08034.1 CXB1_HUMAN	LSPEYKQ---NEINKLLSEQDQSL-----KDILRRSPGTGAGLAEKSDRCSAC	283
sp P29033.3 CXB2_HUMAN	-----	226
sp O95452.2 CXB6_HUMAN	ALKESKQ---NEMNELISDSGQNA-----IT---GFPS-----	261
sp O75712.1 CXB3_HUMAN	ASTCRCHHLVEAGEVDPDPGNNKLAQ--SAPNLTPI-----	270
sp Q9NTQ9.1 CXB4_HUMAN	PPYVLSQGG-----HPEDGNSVLMKAGSAPVDAGGYP-----	266
sp O95377.2 CXB5_HUMAN	TSSCKQ--D-----DLLSGDLIFLGSDSHPLLPRDRPHVKKTIL-----	273

Annexe 3

Tableau récapitulatif des prédictions des trois programmes

		TM1	TM2	TM3	TM4	Séquence prédite	
α	Cx43	Protscale	24-44	77-97	156-176	208-228	V24-V41 ; W78-A94 ; I156-I172 ; I208-L228
		Δ G predictor	23-41	75-94	151-172	207-229	
		TOPCONS	21-41	78-98	156-176	208-228	
	Cx46	Protscale	20-40	72-92	153-173	202-222	I21-A40 ; F77-F92 ; I153-F169 ; I202-L220
		Δ G predictor	19-41	75-94	153-173	200-220	
		TOPCONS	21-41	77-97	149-169	201-221	
	Cx37	Protscale	21-40	77-99	149-167	208-230	I24-A40 ; Y77-L93 ; A150-L165 ; I208-L226
		Δ G predictor	24-46	75-93	144-165	206-226	
		TOPCONS	21-41	76-96	150-170	208-228	
	Cx40	Protscale	20-40	77-97	165-185	206-226	V21-A40 ; W78-G94 ; V165-I177 ; V206-L226
		Δ G predictor	19-41	75-94	157-179	206-227	
		TOPCONS	21-41	78-98	157-177	206-226	
β	Cx32	Protscale	23-45	76-95	131-153	192-214	V23-A40 ; L76-M93 ; Y135-F153 ; V192-Y211
		Δ G predictor	18-40	74-93	135-157	190-212	
		TOPCONS	22-42	76-96	133-153	191-211	
	Cx26	Protscale	21-40	74-94	136-156	190-210	K22-A40 ; L76-M93 ; Y136-Y152 ; T192-L210
		Δ G predictor	22-44	74-93	132-154	191-213	
		TOPCONS	22-42	76-96	132-152	192-212	
	Cx31	Protscale	21-40	76-98	127-149	188-210	G21-A40 ; L76-A96 ; W129-L149 ; F190-L209
		Δ G predictor	18-40	74-96	127-149	187-209	
		TOPCONS	20-40	76-96	129-149	190-210	
	Cx30.3	Protscale	21-40	74-94	131-151	185-205	S21-A40 ; L76-H94 ; Y131-Y147 ; T187-V205
		Δ G predictor	18-40	75-97	127-149	186-208	
		TOPCONS	20-40	76-96	127-147	187-207	
	Cx31.1	Protscale	21-40	76-98	127-149	188-210	G21-A40 ; L76-V95 ; W129-F149 ; L188-Y207
		Δ G predictor	18-40	75-97	127-149	186-208	
		TOPCONS	21-41	75-95	129-149	187-207	
	Cx30	Protscale	23-45	76-98	132-154	193-215	V23-A40 ; W77-M93 ; Y136-Y152 ; I193-Y212
		Δ G predictor	20-40	74-93	136-158	193-215	
		TOPCONS	22-42	77-97	132-152	192-212	

Références Bibliothèques

- A -

Aasen, T., Johnstone, S., Vidal-Brime, L., Lynn, K. S., & Koval, M. (2018). Connexins: Synthesis, post-translational modifications, and trafficking in health and disease. *International journal of molecular sciences*, 19(5), 1296.

Abbaci, M. (2008). Suivi de la fonctionnalité des jonctions communicantes par la technique de gap-FRAP sur des modèles in vitro (2-D, 3-D) et ex vivo : Intérêt pour le diagnostic du cancer (Thèse de Doctorat). Université Henri Poincaré Nancy 1.

Ahmad, S., & Evans, W. H. (2002). Post-translational integration and oligomerization of connexin 26 in plasma membranes and evidence of formation of membrane pores: implications for the assembly of gap junctions. *Biochemical Journal*, 365(3), 693-699.

Ahmad, S., Diez, J. A., GEORGE, C. H., & EVANS, W. H. (1999). Synthesis and assembly of connexins in vitro into homomeric and heteromeric functional gap junction hemichannels. *Biochemical Journal*, 339(2), 247-253.

Ahmed, R., Rangwala, H., & Karypis, G. (2010). TOPTMH: Topology predictor for transmembrane α -helices. *Journal of Bioinformatics and Computational Biology*, 8(01), 39-57.

Allagnat, F., Krattinger, N., Nicod, P., Meda, P., & Haefliger, J. A. (2005). Jonctions communicantes. *Rev Med Suisse*, 1, 1126-33.

Andreeva, A., Howorth, D., Chothia, C., Kulesha, E., & Murzin, A. G. (2014). SCOP2 prototype: a new approach to protein structure mining. *Nucleic acids research*, 42(D1), D310-D314.

Anumonwo, J. M., Taffet, S. M., Gu, H., Chanson, M., Moreno, A. P., & Delmar, M. (2001). The carboxyl terminal domain regulates the unitary conductance and voltage dependence of connexin40 gap junction channels. *Circulation research*, 88(7), 666-673.

Atkinson, M. M., Lampe, P. D., Lin, H. H., Kollander, R., Li, X. R., & Kiang, D. T. (1995). Cyclic AMP modifies the cellular distribution of connexin43 and induces a persistent increase in the junctional permeability of mouse mammary tumor cells. *Journal of Cell Science*, 108(9), 3079-3090.

Axelsen, L. N., Calloe, K., Holstein-Rathlou, N. H., & Nielsen, M. S. (2013). Managing the complexity of communication: regulation of gap junctions by post-translational modification. *Frontiers in pharmacology*, 4, 130.

- B -

Bagos, P. G., Liakopoulos, T. D., & Hamodrakas, S. J. (2005). Evaluation of methods for predicting the topology of β -barrel outer membrane proteins and a consensus prediction method. *BMC bioinformatics*, 6, 1-13.

Bai, D., Yue, B., & Aoyama, H. (2018). Crucial motifs and residues in the extracellular loops influence the formation and specificity of connexin docking. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1860(1), 9-21.

Baranova, A., Ivanov, D., Petrash, N., Pestova, A., Skoblov, M., Kelmanson, I., ... & Panchin, Y. (2004). The mammalian pannexin family is homologous to the invertebrate innexin gap junction proteins. *Genomics*, 83(4), 706-716.

Baxevanis, A. D., & Bateman, A. (2015). The importance of biological databases in biological discovery. *Current protocols in bioinformatics*, 50(1), 1-1.

Benayad, A. (2018). Caractérisation des domaines des connexines par des méthodes bioinformatiques (Mémoire de master). Université de Tlemcen.

Bennett, M. V., & Zukin, R. S. (2004). Electrical coupling and neuronal synchronization in the mammalian brain. *Neuron*, 41(4), 495-511.

Bennett, M. V., Barrio, L. C., Bargiello, T. A., Spray, D. C., Hertzberg, E. L., & Saez, J. C. (1991). Gap junctions: new tools, new answers, new questions. *Neuron*, 6(3), 305-320.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic acids research*, 41(D1), D36-D42.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2010). GenBank. *Nucleic acids research*, 38(suppl_1), D46-D51.

Bernsel, A., Viklund, H., Hennerdal, A., & Elofsson, A. (2009). TOPCONS: consensus prediction of membrane protein topology. *Nucleic acids research*, 37(suppl_2), W465-W468.

- Berthoud, V. M., Beyer, E. C., Kurata, W. E., Lau, A. F., & Lampe, P. D. (1997). The Gap-Junction Protein Connexin 56 is Phosphorylated in the Intracellular Loop and the Carboxy-Terminal Region. *European Journal of Biochemistry*, 244(1), 89-97.
- Bevans, C. G., Kordel, M., Rhee, S. K., & Harris, A. L. (1998). Isoform composition of connexin channels determines selectivity among second messengers and uncharged molecules. *Journal of Biological Chemistry*, 273(5), 2808-2816.
- Beyer, E. C., & Berthoud, V. M. (2018). Gap junction gene and protein families: Connexins, innexins, and pannexins. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1860(1), 5-8.
- Beyer, E. C., Paul, D. L., & Goodenough, D. A. (1987). Connexin43: a protein from rat heart homologous to a gap junction protein from liver. *The Journal of cell biology*, 105(6), 2621-2629.
- Blom, N., Gammeltoft, S., & Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of molecular biology*, 294(5), 1351-1362.
- Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., & Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, 4(6), 1633-1649.
- Boengler, K., & Schulz, R. (2017). Connexin 43 and mitochondria in cardiovascular health and disease. *Mitochondrial Dynamics in Cardiovascular Medicine*, 227-246.
- Bone, L. J., Deschênes, S. M., Balice-Gordon, R. J., Fischbeck, K. H., & Scherer, S. S. (1997). Connexin32 and X-linked charcot–marie–tooth disease. *Neurobiology of disease*, 4(3-4), 221-230.
- Bosco, D., Haefliger, J. A., & Meda, P. (2011). Connexins: key mediators of endocrine function. *Physiological reviews*, 91(4), 1393-1445.
- Brikci Nigassa, A. (2014). Domaine juxta-membranaire de la connexine43: Détermination par RMN en solution de la structure et de l'interaction avec la tubuline et les microtubules (Thèse de doctorat). Université de Tlemcen.

Brikci-Nigassa, A., Clement, M. J., Ha-Duong, T., Adjadj, E., Ziani, L., Pastre, D., ... & Savarin, P. (2012). Phosphorylation controls the interaction of the connexin43 C-terminal domain with tubulin and microtubules. *Biochemistry*, 51(21), 4331-4342.

Bruzzone, R., White, T. W., & Paul, D. L. (1996). Connections with connexins: the molecular basis of direct intercellular signaling. *European Journal of Biochemistry*, 238(1), 1-27.

Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., & Velankar, S. (2017). Protein Data Bank (PDB): the single global macromolecular structure archive. *Protein crystallography: methods and protocols*, 627-641.

- C -

Cameron, S. J., Malik, S., Akaike, M., Lerner-Marmarosh, N., Yan, C., Lee, J. D., ... & Yang, J. (2003). Regulation of epidermal growth factor-induced connexin 43 gap junction communication by big mitogen-activated protein kinase 1/ERK5 but not ERK1/2 kinase activation. *Journal of Biological Chemistry*, 278(20), 18682-18688.

Cao, D., Lin, G., Westphale, E. M., Beyer, E. C., & Steinberg, T. H. (1997). Mechanisms for the coordination of intercellular calcium signaling in insulin-secreting cells. *Journal of Cell Science*, 110(4), 497-504.

Chanson, M., Watanabe, M., O'Shaughnessy, E. M., Zoso, A., & Martin, P. E. (2018). Connexin communication compartments and wound repair in epithelial tissue. *International Journal of Molecular Sciences*, 19(5), 1354.

Condorelli, D. F., Parenti, R., Spinella, F., Salinaro, A. T., Belluardo, N., Cardile, V., & Cicirata, F. (1998). Cloning of a new gap junction gene (Cx36) highly expressed in mammalian brain neurons. *European Journal of Neuroscience*, 10(3), 1202-1208.

- D -

Doble, B. W., Dang, X., Ping, P., Fandrich, R. R., Nickel, B. E., Jin, Y., ... & Kardami, E. (2004). Phosphorylation of serine 262 in the gap junction protein connexin-43 regulates DNA synthesis in cell-cell contact forming cardiomyocytes. *Journal of cell science*, 117(3), 507-514.

Duffy, H. S., Sorgen, P. L., Girvin, M. E., O'Donnell, P., Coombs, W., Taffet, S. M., ... & Spray, D. C. (2002). pH-dependent intramolecular binding and structure involving Cx43 cytoplasmic domains. *Journal of Biological Chemistry*, 277(39), 36706-36714.

Duthe, F., Dupont, E., Verrecchia, F., Plaisance, I., Severs, N. J., Sarrouilhe, D., & Herve, J. C. (2000). Dephosphorylation agents depress gap junctional communication between rat cardiac cells without modifying the Connexin43 phosphorylation degree. *General physiology and biophysics*, 19(4), 441-450.

- E -

Elfgang, C., Eckert, R., Lichtenberg-Fraté, H., Butterweck, A., Traub, O., Klein, R. A., ... & Willecke, K. (1995). Specific permeability and selective formation of gap junction channels in connexin-transfected HeLa cells. *The Journal of cell biology*, 129(3), 805-817.

Evans, W. H., De Vuyst, E., & Leybaert, L. (2006). The gap junction cellular internet: connexin hemichannels enter the signalling limelight. *Biochemical Journal*, 397(1), 1-14.

- F -

Falk, M. M., Buehler, L. K., Kumar, N. M., & Gilula, N. B. (1997). Cell-free synthesis and assembly of connexins into functional gap junction membrane channels. *The EMBO journal*, 16(10), 2703-2716.

Falk, M. M., Kumar, N. M., & Gilula, N. B. (1994). Membrane insertion of gap junction connexins: polytopic channel forming membrane proteins. *The Journal of cell biology*, 127(2), 343-355.

Fallon, R. F., & Goodenough, D. A. (1981). Five-hour half-life of mouse liver gap-junction protein. *The Journal of cell biology*, 90(2), 521-526.

Foote, C. I., Zhou, L., Zhu, X., & Nicholson, B. J. (1998). The pattern of disulfide linkages in the extracellular loop regions of connexin 32 suggests a model for the docking interface of gap junctions. *The Journal of cell biology*, 140(5), 1187-1197.

Fox, T., & Kriegel, J. M. (2006). Machine learning techniques for in silico modeling of drug metabolism. *Current Topics in Medicinal Chemistry*, 6(15), 1579-1591.

- G -

- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server (pp. 571-607). *Humana press*.
- Gemel, J., Valiunas, V., Brink, P. R., & Beyer, E. C. (2004). Connexin43 and connexin26 form gap junctions, but not heteromeric channels in co-expressing cells. *Journal of cell science*, 117(12), 2469-2480.
- González, D., Gómez-Hernández, J. M., & Barrio, L. C. (2007). Molecular basis of voltage dependence of connexin channels: an integrative appraisal. *Progress in biophysics and molecular biology*, 94(1-2), 66-106.
- Goodenough, D. A., & Paul, D. L. (2003). Beyond the gap: functions of unpaired connexon channels. *Nature reviews Molecular cell biology*, 4(4), 285-295.
- Goodenough, D. A., & Paul, D. L. (2009). Gap junctions. *Cold Spring Harbor perspectives in biology*, 1(1), a002576.

- H -

- Hervé, J. C., Bourmeyster, N., & Sarrouilhe, D. (2004). Diversity in protein-protein interactions of connexins: emerging roles. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1662(1-2), 22-41.
- Hervé, J. C., Derangeon, M., Sarrouilhe, D., Giepmans, B. N., & Bourmeyster, N. (2012). Gap junctional channels are parts of multiprotein complexes. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1818(8), 1844-1865.
- Hervé, J. C., Derangeon, M., Théveniau-Ruissy, M., Miquerol, L., Sarrouilhe, D., & Gros, D. (2008). Connexines et canaux jonctionnels. Leurs rôles dans la propagation de l'activité électrique cardiaque et le développement du cœur. *Pathologie Biologie*, 56(5), 334-341.
- Hervé, J. C., Pluciennik, F., Verrecchia, F., Bastide, B., Delage, B., Joffre, M., & Deleze, J. (1996). Influence of the molecular structure of steroids on their ability to interrupt gap junctional communication. *The Journal of membrane biology*, 149, 179-187.

Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., ... & Von Heijne, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, 450(7172), 1026-1030.

Hirosawa, M., Totoki, Y., Hoshida, M., & Ishikawa, M. (1995). Comprehensive study on iterative algorithms of multiple sequence alignment. *Bioinformatics*, 11(1), 13-18.

Hirst-Jensen, B. J., Sahoo, P., Kieken, F., Delmar, M., & Sorgen, P. L. (2007). Characterization of the pH-dependent interaction between the gap junction protein connexin43 carboxyl terminus and cytoplasmic loop domains. *Journal of biological chemistry*, 282(8), 5801-5813.

Hopperstad, M. G., Srinivas, M., & Spray, D. C. (2000). Properties of gap junction channels formed by Cx46 alone and in combination with Cx50. *Biophysical journal*, 79(4), 1954-1966.

Huang, J. Y., & Brutlag, D. L. (2001). The EMOTIF database. *Nucleic acids research*, 29(1), 202-204.

- I -

Iovine, M. K., Gumpert, A. M., Falk, M. M., & Mendelson, T. C. (2008). Cx23, a connexin with only four extracellular-loop cysteines, forms functional gap junction channels and hemichannels. *FEBS letters*, 582(2), 165-170.

- J -

John, S. A., & Revel, J. P. (1991). Connexon integrity is maintained by non-covalent bonds: intramolecular disulfide bonds link the extracellular domains in rat connexin-43. *Biochemical and biophysical research communications*, 178(3), 1312-1318.

Jordan, K., Chodock, R., Hand, A. R., & Laird, D. W. (2001). The origin of annular junctions: a mechanism of gap junction internalization. *Journal of cell science*, 114(4), 763-773.

- K -

Kardami, E., Dang, X., Iacobas, D. A., Nickel, B. E., Jeyaraman, M., Srisakuldee, W., ... & Spray, D. C. (2007). The role of connexins in controlling cell growth and gene expression. *Progress in biophysics and molecular biology*, 94(1-2), 245-264.

Kemp, B. E., & Pearson, R. B. (1990). Protein kinase recognition sequence motifs. *Trends in biochemical sciences*, 15(9), 342-346.

Kovacs, J. A., Baker, K. A., Altenberg, G. A., Abagyan, R., & Yeager, M. (2007). Molecular modeling and mutagenesis of gap junction channels. *Progress in biophysics and molecular biology*, 94(1-2), 15-28.

Koval, M., Molina, S. A., & Burt, J. M. (2014). Mix and match: investigating heteromeric and heterotypic gap junction channels in model systems and native tissues. *FEBS letters*, 588(8), 1193-1204.

Kumar, N. M., & Gilula, N. B. (1986). Cloning and characterization of human and rat liver cDNAs coding for a gap junction protein. *The Journal of cell biology*, 103(3), 767-776.

- L -

Laird, D. W. (2006). Life cycle of connexins in health and disease. *Biochemical Journal*, 394(3), 527-543.

Laird, D. W., Puranam, K. L., & Revel, J. P. (1991). Turnover and phosphorylation dynamics of connexin43 gap junction protein in cultured cardiac myocytes. *Biochemical Journal*, 273(1), 67-72.

Lampe, P. D., & Lau, A. F. (2000). Regulation of gap junctions by phosphorylation of connexins. *Archives of biochemistry and biophysics*, 384(2), 205-215.

Lampe, P. D., & Lau, A. F. (2004). The effects of connexin phosphorylation on gap junctional communication. *The international journal of biochemistry & cell biology*, 36(7), 1171-1186.

Laroum, S. (2011). Prédiction de la localisation des protéines membranaires: méthodes méta-heuristiques pour la détermination du potentiel d'insertion des acides aminés (Thèse de doctorat). Université d'Angers.

Lastwika, K. J., Dunn, C. A., Solan, J. L., & Lampe, P. D. (2019). Phosphorylation of connexin 43 at MAPK, PKC or CK1 sites each distinctly alter the kinetics of epidermal wound repair. *Journal of Cell Science*, 132(18), jcs234633.

- Leithe, E., & Rivedal, E. (2004). Ubiquitination and down-regulation of gap junction protein connexin-43 in response to 12-O-tetradecanoylphorbol 13-acetate treatment. *Journal of Biological Chemistry*, 279(48), 50089-50096.
- Leonard, S. A., & Littlejohn, T. G. (2004). Common file formats. *Current Protocols in Bioinformatics*, 5(1), A-1B.
- Li, F., Li, C., Marquez-Lago, T. T., Leier, A., Akutsu, T., Purcell, A. W., ... & Chou, K. C. (2018). Quokka: a comprehensive tool for rapid and accurate prediction of kinase family-specific phosphorylation sites in the human proteome. *Bioinformatics*, 34(24), 4223-4231.
- Liao, C. K., Cheng, H. H., Wang, S. D., Yeih, D. F., & Wang, S. M. (2013). PKC ϵ mediates serine phosphorylation of connexin43 induced by lysophosphatidylcholine in neonatal rat cardiomyocytes. *Toxicology*, 314(1), 11-21.
- Likic, V. (2008). The Needleman-Wunsch algorithm for sequence alignment. Lecture given at the 7th Melbourne Bioinformatics Course, Bi021 Molecular Science and Biotechnology Institute, University of Melbourne, 1-46.
- Lim, P. K., Bliss, S. A., Patel, S. A., Taborga, M., Dave, M. A., Gregory, L. A., ... & Rameshwar, P. (2011). Gap Junction-Mediated Import of MicroRNA from Bone Marrow Stromal Cells Can Elicit Cell Cycle Quiescence in Breast Cancer Cells. *miRNA in Breast Cancer Dormancy. Cancer research*, 71(5), 1550-1560.
- Lin, R., Warn-Cramer, B. J., Kurata, W. E., & Lau, A. F. (2001). v-Src phosphorylation of connexin 43 on Tyr247 and Tyr265 disrupts gap junctional communication. *The Journal of cell biology*, 154(4), 815-828.

- M -

- Maeda, S., Nakagawa, S., Suga, M., Yamashita, E., Oshima, A., Fujiyoshi, Y., & Tsukihara, T. (2009). Structure of the connexin 26 gap junction channel at 3.5 Å resolution. *Nature*, 458(7238), 597-602.
- Martin, P. E. M., & van Steensel, M. (2015). Connexins and skin disease: Insights into the role of beta connexins in skin homeostasis. *Cell and Tissue Research*, 360(3), 645-658.

- Matsumura, K., Mayama, T., Lin, H., Sakamoto, Y., Ogawa, K., & Imanaga, I. (2006). Effects of cyclic AMP on the function of the cardiac gap junction during hypoxia. *Experimental & Clinical Cardiology*, 11(4), 286.
- Matsuuchi, L., & Naus, C. C. (2013). Gap junction proteins on the move: connexins, the cytoskeleton and migration. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1828(1), 94-108.
- Meda, P. (1996). Connexines, canaux jonctionnels et communication cellulaire. *Med. Sci.* 12, 909-920.
- Meşe, G., Richard, G., & White, T. W. (2007). Gap junctions: basic structure and function. *Journal of Investigative Dermatology*, 127(11), 2516-2524.
- Meshoul, S., & Layeb, A. (2017). Approche quantique évolutionnaire pour l'alignement multiple de séquences en bioinformatique (Mémoire de Magistère). Université Mentouri de Constantine.
- Milks, L. C., Kumar, N. M., Houghten, R., Unwin, N., & Gilula, N. B. (1988). Topology of the 32-kd liver gap junction protein determined by site-directed antibody localizations. *The EMBO Journal*, 7(10), 2967-2975.
- Morley, G. E., Taffet, S. M., & Delmar, M. (1996). Intramolecular interactions mediate pH regulation of connexin43 channels. *Biophysical Journal*, 70(3), 1294-1302.
- Musil, L. S., & Goodenough, D. A. (1993). Multisubunit assembly of an integral plasma membrane channel protein, gap junction connexin43, occurs after exit from the ER. *Cell*, 74(6), 1065-1077.
- N -
- Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3), 443-453.
- Nielsen, H., Brunak, S., & von Heijne, G. (1999). Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein engineering*, 12(1), 3-9.

Nielsen, H., Engelbrecht, J., Brunak, S., & Von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein engineering*, 10(1), 1-6.

Nitsche, J. M., Chang, H. C., Weber, P. A., & Nicholson, B. J. (2004). A transient diffusion model yields unitary gap junctional permeabilities from images of cell-to-cell fluorescent dye transfer between *Xenopus* oocytes. *Biophysical Journal*, 86(4), 2058-2077.

- O -

Oh, S., Rivkin, S., Tang, Q., Verselis, V. K., & Bargiello, T. A. (2004). Determinants of gating polarity of a connexin 32 hemichannel. *Biophysical journal*, 87(2), 912-928.

Olk, S., Zoidl, G., & Dermietzel, R. (2009). Connexins, cell motility, and the cytoskeleton. *Cell motility and the cytoskeleton*, 66(11), 1000-1016.

Omasits, U., Ahrens, C. H., Müller, S., & Wollscheid, B. (2014). Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*, 30(6), 884-886.

Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., & Thornton, J. M. (1997). CATH—a hierarchic classification of protein domain structures. *Structure*, 5(8), 1093-1109.

Oyamada, M., Kimura, H., Oyamada, Y., Miyamoto, A., Ohshika, H., & Mori, M. (1994). The expression, phosphorylation, and localization of connexin 43 and gap-junctional intercellular communication during the establishment of a synchronized contraction of cultured neonatal rat cardiac myocytes. *Experimental cell research*, 212(2), 351-358.

- P -

Panchin, Y. I., Kelmanson, M., Matz, K., Lukyanov, N., Usman, S., Lukyanov. (2000) A ubiquitous family of putative gap junction molecules, *Curr. Biol.* 10. R473–R474.

Paulson, A. F., Lampe, P. D., Meyer, R. A., TenBroek, E., Atkinson, M. M., Walseth, T. F., & Johnson, R. G. (2000). Cyclic AMP and LDL trigger a rapid enhancement in gap junction assembly through a stimulation of connexin trafficking. *Journal of cell science*, 113(17), 3037-3049.

- Pei, J. (2008). Multiple protein sequence alignment. *Current opinion in structural biology*, 18(3), 382-386.
- Peracchia, C. (2004). Chemical gating of gap junction channels: roles of calcium, pH and calmodulin. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1662(1-2), 61-80.
- Peracchia, C. (2020). Calmodulin-mediated regulation of gap junction channels. *International Journal of Molecular Sciences*, 21(2), 485.
- Pluciennik, F., Verrecchia, F., Bastide, B., Herve, J. C., Joffre, M., & Deleze, J. (1996). Reversible interruption of gap junctional communication by testosterone propionate in cultured Sertoli cells and cardiac myocytes. *The Journal of membrane biology*, 149, 169-177.
- Procida, K., Jørgensen, L., Schmitt, N., Delmar, M., Taffet, S. M., Holstein-Rathlou, N. H., ... & Braunstein, T. H. (2009). Phosphorylation of connexin43 on serine 306 regulates electrical coupling. *Heart Rhythm*, 6(11), 1632-1638.

- Q -

- Qi, G. J., Chen, Q., Chen, L. J., Shu, Y., Bu, L. L., Shao, X. Y., ... & Tian, B. (2016). Phosphorylation of connexin 43 by Cdk5 modulates neuronal migration during embryonic brain development. *Molecular neurobiology*, 53, 2969-2982.
- Qin, H., Shao, Q., Igdoura, S. A., Alaoui-Jamali, M. A., & Laird, D. W. (2003). Lysosomal and proteasomal degradation play distinct roles in the life cycle of Cx43 in gap junctional intercellular communication-deficient and-competent breast tumor cells. *Journal of Biological Chemistry*, 278(32), 30005-30014.
- Qiu, Y., Zheng, J., Chen, S., & Sun, Y. (2022). Connexin mutations and hereditary diseases. *International Journal of Molecular Sciences*, 23(8), 4255.

- R -

- Račkauskas, M., Neverauskas, V., & Skeberdis, V. A. (2010). Diversity and properties of connexin gap junction channels. *Medicina*, 46(1), 1.
- Ramazi, S., & Zahiri, J. (2021). Post-translational modifications in proteins: resources, tools and prediction methods. *Database*, 2021.

- Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., & Yao, X. (2009). DOG 1.0: illustrator of protein domain structures. *Cell research*, 19(2), 271-273.
- Reynhout, J. K., Lampe, P. D., & Johnson, R. G. (1992). An activator of protein kinase C inhibits gap junction communication between cultured bovine lens cells. *Experimental cell research*, 198(2), 337-342.
- Richards, T. S., Dunn, C. A., Carter, W. G., Usui, M. L., Olerud, J. E., & Lampe, P. D. (2004). Protein kinase C spatially and temporally regulates gap junctional communication during human wound repair via phosphorylation of connexin43 on serine368. *The Journal of cell biology*, 167(3), 555-562.
- Rubin, J. B., Verselis, V. K., Bennett, M. V., & Bargiello, T. A. (1992). A domain substitution procedure and its use to analyze voltage dependence of homotypic gap junctions formed by connexins 26 and 32. *Proceedings of the National Academy of Sciences*, 89(9), 3820-3824.
- S -
- Sáez, J. C., Berthoud, V. M., Branes, M. C., Martinez, A. D., & Beyer, E. C. (2003). Plasma membrane channels formed by connexins: their regulation and functions. *Physiological reviews*, 83(4), 1359-1400.
- Sáez, J. C., Nairn, A. C., Czernik, A. J., Fishman, G. I., Spray, D. C., & Hertzberg, E. L. (1997). Phosphorylation of connexin43 and the regulation of neonatal rat cardiac myocyte gap junctions. *Journal of molecular and cellular cardiology*, 29(8), 2131-2145.
- Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., ... & Sherry, S. T. (2021). Database resources of the national center for biotechnology information. *Nucleic acids research*, 49(D1), D10.
- Segretain, D., & Falk, M. M. (2004). Regulation of connexin biosynthesis, assembly, gap junction formation, and removal. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1662(1-2), 3-21.
- Shi, Y., Li, X., & Yang, J. (2022). Mutations of CX46/CX50 and Cataract Development. *Frontiers in Molecular Biosciences*, 9, 842399.

- Shinde, M. Y., Sidoli, S., Kulej, K., Mallory, M. J., Radens, C. M., Reicherter, A. L., ... & Klein, P. S. (2017). Phosphoproteomics reveals that glycogen synthase kinase-3 phosphorylates multiple splicing factors and is associated with alternative splicing. *Journal of Biological Chemistry*, 292(44), 18240-18255.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), 539.
- Sirnes, S., Kjenseth, A., Leithe, E., & Rivedal, E. (2009). Interplay between PKC and the MAP kinase pathway in Connexin43 phosphorylation and inhibition of gap junction intercellular communication. *Biochemical and biophysical research communications*, 382(1), 41-45.
- Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of molecular biology*, 147(1), 195-197.
- Söhl, G., & Willecke, K. (2004). Gap junctions and the connexin protein family. *Cardiovascular research*, 62(2), 228-232.
- Söhl, G., Nielsen, P. A., Eiberger, J., & Willecke, K. (2003). Expression profiles of the novel human connexin genes hCx30. 2, hCx40. 1, and hCx62 differ from their putative mouse orthologues. *Cell communication & adhesion*, 10(1), 27-36.
- Solan, J. L., & Lampe, P. D. (2005). Connexin phosphorylation as a regulatory event linked to gap junction channel assembly. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1711(2), 154-163.
- Sosinsky, G. (1995). Mixing of connexins in gap junction membrane channels. *Proceedings of the National Academy of Sciences*, 92(20), 9210-9214.
- Srinivas, M., Verselis, V. K., & White, T. W. (2018). Human diseases associated with connexin mutations. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1860(1), 192-201.
- Stains, J. P., & Civitelli, R. (2005). Gap junctions regulate extracellular signal-regulated kinase signaling to affect gene transcription. *Molecular biology of the cell*, 16(1), 64-72.

Stains, J. P., Lecanda, F., Screen, J., Towler, D. A., & Civitelli, R. (2003). Gap junctional communication modulates gene transcription by altering the recruitment of Sp1 and Sp3 to connexin-response elements in osteoblast promoters. *Journal of Biological Chemistry*, 278(27), 24377-24387.

Stoesser, G., Baker, W., van den Broek, A., Camon, E., Garcia-Pastor, M., Kanz, C., ... & Vaughan, R. (2002). The EMBL nucleotide sequence database. *Nucleic acids research*, 30(1), 21-26.

- T -

Talbot, J. (2012). Rôle de la connexine 43 dans l'ostéogenèse et dans le développement tumoral des sarcomes d'ewing (Thèse de Doctorat). Université de Nantes.

Tateno, Y., Imanishi, T., Miyazaki, S., Fukami-Kobayashi, K., Saitou, N., Sugawara, H., & Gojobori, T. (2002). DNA Data Bank of Japan (DDBJ) for genome scale research in life science. *Nucleic acids research*, 30(1), 27-30.

TenBroek, E. M., Lampe, P. D., Solan, J. L., Reynhout, J. K., & Johnson, R. G. (2001). Ser364 of connexin43 and the upregulation of gap junction assembly by cAMP. *The Journal of cell biology*, 155(7), 1307-1318.

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., ... & Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature biotechnology*, 40(7), 1023-1025.

Thompson, J. D., Plewniak, F., & Poch, O. (1999). A comprehensive comparison of multiple sequence alignment programs. *Nucleic acids research*, 27(13), 2682-2690.

- U -

UniProt Consortium. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic acids research*, 47(D1), D506-D515.

Urschel, S., Hoher, T., Schubert, T., Alev, C., Sohl, G., Worsdorfer, P., ... & Willecke, K. (2006). Protein kinase A-mediated phosphorylation of connexin36 in mouse retina results in decreased gap junctional communication between AII amacrine cells. *Journal of biological chemistry*, 281(44), 33163-33171.

- V -

- Van Kempen, M. J., Vermeulen, J. L., Moorman, A. F., Gros, D., Paul, D. L., & Lamers, W. H. (1996). Developmental changes of connexin40 and connexin43 mRNA distribution patterns in the rat heart. *Cardiovascular research*, 32(5), 886-900.
- Verrecchia, F., & Hervé, J. C. (1997). Reversible inhibition of gap junctional communication elicited by several classes of lipophilic compounds in cultured rat cardiomyocytes. *Canadian Journal of Cardiology*, 13(11), 1093-1102.
- Verrecchia, F., Duthe, F., Duval, S., Duchatelle, I., Sarrouilhe, D., & Hervé, J. C. (1999). ATP counteracts the rundown of gap junctional channels of rat ventricular myocytes by promoting protein phosphorylation. *The Journal of Physiology*, 516(Pt 2), 447.
- Verselis, V. K., Ginter, C. S., & Bargiello, T. A. (1994). Opposite voltage gating polarities of two closely related connexins. *Nature*, 368(6469), 348-351.
- Virág, D., Dalmadi-Kiss, B., Vékey, K., Drahos, L., Klebovich, I., Antal, I., & Ludányi, K. (2020). Current trends in the analysis of post-translational modifications. *Chromatographia*, 83, 1-10.

- W -

- Wang, G., & Wu, X. (2021). The potential antiepileptogenic effect of neuronal Cx36 gap junction channel blockage. *Translational Neuroscience*, 12(1), 46-51.
- Wang, L., Fu, Y., Peng, J., Wu, D., Yu, M., Xu, C., ... & Tao, L. (2013). Simvastatin-induced up-regulation of gap junctions composed of connexin 43 sensitize Leydig tumor cells to etoposide: An involvement of PKC pathway. *Toxicology*, 312, 149-157.
- Warn-Cramer, B. J., Lampe, P. D., Kurata, W. E., Kanemitsu, M. Y., Loo, L. W., Eckhart, W., & Lau, A. F. (1996). Characterization of the Mitogen-activated Protein Kinase Phosphorylation Sites on the Connexin-43 Gap Junction Protein (□). *Journal of Biological Chemistry*, 271(7), 3779-3786.
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189-1191.

- Wei, C. J., Francis, R., Xu, X., & Lo, C. W. (2005). Connexin43 associated with an N-cadherin-containing multiprotein complex is required for gap junction formation in NIH3T3 cells. *Journal of Biological Chemistry*, 280(20), 19925-19936.
- Wei, C. J., Xu, X., & Lo, C. W. (2004). Connexins and cell signaling in development and disease. *Annu. Rev. Cell Dev. Biol.*, 20, 811-838.
- White, T. W., Bruzzone, R., Wolfram, S., Paul, D. L., & Goodenough, D. A. (1994). Selective interactions among the multiple connexin proteins expressed in the vertebrate lens: the second extracellular domain is a determinant of compatibility between connexins. *The Journal of cell biology*, 125(4), 879-892.
- White, T. W., Paul, D. L., Goodenough, D. A., & Bruzzone, R. (1995). Functional analysis of selective interactions among rodent connexins. *Molecular biology of the cell*, 6(4), 459-470.
- Willecke, K., Eiberger, J., Degen, J., Eckardt, D., Romualdi, A., Güldenagel, M., ... & Söhl, G. (2002). Structural and functional diversity of connexin genes in the mouse and human genome.
- Wolvetang, E. J., Pera, M. F., & Zuckerman, K. S. (2007). Gap junction mediated transport of shRNA between human embryonic stem cells. *Biochemical and biophysical research communications*, 363(3), 610-615.
- Y -
- Yeager, M., & Gilula, N. B. (1992). Membrane topology and quaternary structure of cardiac gap junction ion channels. *Journal of molecular biology*, 223(4), 929-948.
- Yeager, M., & Nicholson, B. J. (1996). Structure of gap junction intercellular channels. *Current opinion in structural biology*, 6(2), 183-192.
- Young, B., Lowe, J. S., Stevens, A., & Heath, J. W. (2006). *Wheater's Functional Histology: A Text and Colour Atlas* (5th ed.). Page 82.
- Z -
- Zaki, M. J., Karypis, G., & Yang, J. (2007). Data mining in bioinformatics (*BIOKDD*).

- Zhang, H. (2016). Overview of sequence data formats. *Statistical genomics: Methods and protocols*, 3-17.
- Zhang, J. T., & Nicholson, B. J. (1994). The topological structure of connexin 26 and its distribution compared to connexin 32 in hepatic gap junctions. *The Journal of membrane biology*, 139, 15-29.
- Zhou, Y., Yang, W., Lurtz, M. M., Ye, Y., Huang, Y., Lee, H. W., ... & Yang, J. J. (2007). Identification of the calmodulin binding domain of connexin 43. *Journal of biological chemistry*, 282(48), 35005-35017.
- Zimmer, D. B., Green, C. R., Evans, W. H., & Gilula, N. B. (1987). Topological analysis of the major protein in isolated intact rat liver gap junctions and gap junction-derived single membrane structures. *Journal of Biological Chemistry*, 262(16), 7751-7763.
- Zou, J., Salarian, M., Chen, Y., Veenstra, R., Louis, C. F., & Yang, J. J. (2014). Gap junction regulation by calmodulin. *FEBS letters*, 588(8), 1430-1438.