

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche
scientifique

Université Abou Bakr Belkaid –Tlemcen

Faculté des Sciences de la Nature et de la Vie et des Sciences de la Terre et de l'Univers

Département de Biologie



Laboratoire :

« Antibiotiques Antifongiques : Physico-Chimie, Synthèse et Activité Biologique »

LRM : laboratoire de recherche sur les macromolécules

Mémoire en vue de l'obtention du diplôme de
Master en Biologie

Option : Biochimie Appliquée

Présenté par : Amal Hachim

Thème

*Caractérisation de la connexine23 humaine
par
méthodes bioinformatiques*

Soutenu le : 29/10/2020 devant le jury composé de :

Présidente	Pr Chahrazed BEKHECHI	Professeur	Université de Tlemcen
Examinatrice	Dr Meriem BENMANSOUR	Maître de conférences	Université de Tlemcen
Examinatrice	Dr Nabila BENARIBA	Maître de conférences	Université de Tlemcen
Encadrante	Dr Amal BRIKCI NIGASSA	Maître de conférences	Université de Tlemcen

Année universitaire : 2019-2020

Dédicaces

À ma chère mère, Fouzia

À mon cher père, Mohamed

Qui n'ont jamais cessé de formuler des prières à mon égard, de me soutenir et de m'épauler pour que je puisse atteindre mes objectifs.

À mon frère Abdessamad

À mes sœurs Kawther et Mimi

Pour leur soutien moral, leur appui et encouragements

À toute ma famille

Pour leur soutien tout au long de mon parcours universitaire

À mes chères amies Hayat, Wissem et Asma

Pour leurs aides et supports dans les moments difficiles

À mon cher « Y »

Pour le bonheur, le sourire, le soutien et l'encouragement

Que ce travail soit l'accomplissement de votre soutien infailible,

Merci d'être toujours là pour moi

Remerciements

En préambule à ce mémoire je remercie Allah tout Puissant de m'avoir donné la patience et le courage durant ces longues années d'études.

Ce travail n'aurait pas pu voir le jour sans l'aide et l'encadrement de Dr Amal BRIKCI NIGASSA née SAIDI que je remercie pour la qualité de son encadrement exceptionnel, pour sa patience et sa disponibilité tout au long de la réalisation de ce mémoire.

Je tiens à remercier profondément Pr Chahrazed BEKHECHI pour l'honneur qu'elle me fait en acceptant de présider ce jury.

J'exprime également mes sincères remerciements à Dr Nabila BENARIBA d'avoir accepté d'examiner et discuter ce travail.

Je tiens à remercier également Dr Meriem BENMANSOUR d'avoir accepté d'examiner et discuter ce travail.

Mes remerciements s'adressent aussi à tous mes professeurs pour leurs efforts et leur grande patience.

Mes profonds remerciements vont également à toutes les personnes qui m'ont aidé et soutenu de près ou de loin.

Merci à tous et à toutes.

Résumé

Dans le génome humain 21 isoformes de connexines ont été identifiées servant à la formation des canaux jonctionnels. Cependant, il n'existe que trois isoformes de pannexines servant à la formation des hémicanaux.

Dans ce travail, nous avons fait une analyse structurale et fonctionnelle de la connexine23 (CX23) aux moyens d'outils informatiques. En premier temps, nous avons caractérisé les différents domaines de la CX23 en mettant l'accent sur les résidus essentiels par la prédiction des modifications post-traductionnelles. Puis, nous avons utilisé les séquences de la CX23 et les trois *hPanx* afin de montrer les régions conservées entre ces protéines.

L'étude de la topologie avec les programmes ProtScale, ΔG predictor et TOPCONS a montré que la CX23 est une protéine membranaire à quatre passages transmembranaires exposant deux boucles extracellulaires et une boucle intracytoplasmique avec des extrémités N- et C-terminales cytoplasmiques.

Les résultats de l'étude des modifications post-traductionnelles ont montré que la CX23 semble contenir des sites de phosphorylation, palmitoylation et protéolyse. Cependant, elle semble ne contenir aucun site de glycosylation.

L'alignement multiple des séquences de la CX23 avec les trois *hPanx* a été réalisé par le programme T-Coffee révélant que les deux résidus cystéines dans chaque boucle extracellulaire de ces protéines sont très conservés.

Mots clés : connexines, pannexines, CX23, topologie membranaire, modifications post-traductionnelles, alignement multiple

Abstract

In the human genome, 21 connexin isoforms have been identified that are used for the formation of junction channels. However, there are only three pannexin isoforms for the formation of hemichannels.

In this work, we performed a structural and functional analysis of the connexin23 (CX23) using computer tools. First, we characterised the different domains of CX23, focusing on the essential residues by predicting post-translational modifications. Then, we used the sequences of CX23 and the three *hPanx* to show the conserved regions between these proteins.

The study of the topology with the ProtScale, ΔG predictor and TOPCONS programmes showed that CX23 is a four-pass transmembrane membrane protein exposing two extracellular loops and a intracytoplasmic loop with cytoplasmic N- and C-terminal extremities.

The results of the post-translational modification study suggested that CX23 appears to contain sites of phosphorylation, palmitoylation and proteolysis. However, it does not appear to contain any sites of glycosylation.

The multiple alignment sequences of the CX23 with the three *hPanx* was performed by the T-Coffee programme, revealing that the two cysteine residues in each extracellular loop of these proteins are highly conserved.

Keywords: connexins, pannexins, CX23, membrane topology, post-translational modifications, multiple alignment

ملخص

عند الانسان يوجد 21 عضو من الكناكسينات التي تشكل قنواة ناقلة. بينما يوجد ثلاثة اعضاء من البانكسينات التي تقوم بتشكيل قنوات نصفية.

في هذا العمل قمنا باجراء تحليل هيكلي ووظيفي للكناكسين 23 باستخدام ادوات معلوماتية. اولاً قمنا بتحديد المقاطع المختلفة للكناكسين 23 مع دراسة التعديلات مابعد الترجمة. ثم قمنا باستخدام المتتاليات البشرية للكناكسين 23 والبنكسينات الثلاث من اجل اظهار المناطق المحفوظة بين هذه البروتينات.

دراسة طبولوجيا عن طريق ثلاثة برامج وضحت ان الكناكسين 23 عبارة عن بروتين غشائي مع اربعة ممرات عبر الغشاء بالاضافة الى حلقتين خارجيتين وحلقة داخلية مع النهايات الامينة والكربوكسيلية من الجانب الداخلي للسيتوبلازم.

اظهرت نتائج التعديلات مابعد الترجمة ان الكناكسين 23 يحتوي على مواقع اضافة الفسفرة حمض البالمتيك وتحلل البروتين الا انه لا يحتوي على أي موقع لاضافة السكر.

محاذاة المتتاليات البشرية للكناكسين 23 والبنكسينات وضحت بان عنصري السيستين في كل حلقة خارجية لهذه البروتينات يكونان جد محفوظين.

الكلمات المفتاحية: محاذاة متعددة- الكانكسينات- البانكسينات- الكوناكسين 23- تعديلات مابعد الترجمة.

Table des matières

Introduction	1
Partie 1 : Synthèse bibliographique	4
Chapitre I : Connexines & Pannexines	5
1 Classes des protéines membranaires	6
1.1 Les protéines membranaires intégrales	6
1.2 Les protéines membranaires à ancrage lipidique	7
1.3 Les protéines membranaires périphériques	8
2 Fonctions des protéines membranaires	9
3 Connexines et pannexines	9
3.1 Expression génique des connexines et des pannexines	10
3.2 Structure des connexines et pannexines	11
3.3 Modifications post-traductionnelles des connexines	14
4 Connexine 23	15
4.1 Expression génique de la connexine23.....	15
4.2 Structure de la connexine23	16
Chapitre II : Analyse des protéines membranaires par bioinformatique	18
1 Introduction.....	19
2 Banques de données.....	19
2.1 Banques de données primaires/généralistes	19
2.2 Banques de données secondaires/ spécialisées.....	20
2.3 Banques de données composites	20
3 Étude de la topologie des protéines membranaires par des techniques informatiques.....	21
4 Étude des modifications post-traductionnelles.....	22
5 Alignement des séquences	24
5.1 Alignement par paires	24
5.2 Alignement multiple.....	24
Partie II : Travaux expérimentaux & Résultats	26
Chapitre III : Matériel & Méthodes	27
1 Choix de la CX23	28
2 Étude de la topologie de la CX23	29
2.1 ProtScale	29
2.2 ΔG predictor 1.0.....	31
2.3 TOPCONS.....	31
3 Étude des modifications post-traductionnelles de la CX23	32
3.1 Étude de la glycosylation avec NetNGlyc 1.0 et NetOGlyc 4.0.....	32
3.2 Étude de la phosphorylation avec NetPhos 3.1	32
3.3 Étude de la palmitoylation avec CSS-Palm 4.0.....	33
3.4 Étude de la protéolyse avec PROSPEROUS	34

4	Réalisation de la topologie membranaire présumée de la CX23 par Protter	35
5	Alignement de la CX23 avec les trois <i>hPanx</i>	35
5.1	Alignement des séquences avec T-Coffee	35
5.2	Edition de l'alignement avec Jalview version 2.10.4b1	38
Chapitre IV : Résultats & Interprétation		40
1	La CX23 possède quatre domaines transmembranaires	41
1.1	Analyse de la CX23 par ProtScale	41
1.2	Analyse de la CX23 par ΔG predictor :	42
1.3	Analyse de la CX23 par TOPCONS	43
1.4	Prédiction de la topologie de la CX23 par les trois méthodes	44
2	Sites des modifications post-traductionnelles de la CX23	45
2.1	La CX23 ne présente aucun site de glycosylation	45
2.2	La CX23 est phosphorylée	45
2.3	La CX23 est palmitoylée	46
2.4	La CX23 présente des sites de clivage par les caspases	46
3	Topologie membranaire présumée de la CX23	46
4	Alignement de la CX23 avec les trois <i>hPanx</i>	48
4.1	Alignement de la CX23 et les <i>hPanx</i> montrant les domaines Nt, M1, boucle E1, M2	49
4.2	Alignement de la CX23 et les <i>hPanx</i> montrant la boucle I et les domaines M3, boucle E2 et M4 .	50
Conclusion		52
Références bibliographiques		54

Liste des abréviations

- Akt** : Kinase appelée aussi protéine kinase B
- AMPc** : Adénosine monophosphate cyclique
- ATP** : Adénosine triphosphate
- CDK1** : Kinases dépendantes des cyclines
- Ct** : Extrémité carboxyle-terminale
- CX** : Connexine humaine
- Cx23** : Connexine23 de la souris
- CX23** : Connexine23 humaine
- E** : Boucle extracytoplasmique
- Fasta** : Fast alignement
- GPCR** : Récepteur couplé à la protéine G
- GPI** : Glycosylphosphatidylinositol
- hPanx** : Pannexine humaine
- I** : Boucle intracytoplasmique
- IP₃** : Inositol triPhosphate
- kDa** : Kilodalton
- M** : Domaine transmembranaire
- MAPK** : Protéines kinases activées par des mitogènes
- Mpb** : Million de paires de bases.
- Nt** : Extrémité amino-terminale
- PKA** : Protéine kinase A
- PKC** : Protéine kinase C
- RE** : Réticulum endoplasmique
- RTK** : Récepteur à activité Tyrosine Kinase
- Δ Gapp** : Variation d'énergie libre apparente

Liste des figures

Figure 1.1: Classes des protéines membranaires.....	6
Figure 1.2: Structures des connexines et des pannexines.....	12
Figure 1.3: Alignement multiple des 21 isoformes des connexines.....	17
Figure 2.1: Interface de banque de données UniProt.....	20
Figure 3.1: Requête pour la base de données UniProtKB/Swiss-Prot.....	28
Figure 3.2: Séquence de la CX23 au format Fasta.....	29
Figure 3.3: Interface de programme ProtScale.....	30
Figure 3.4: Interface du programme ΔG predictor.....	31
Figure 3.5: Interface du programme TOPCONS.....	32
Figure 3.6: Interface du programme NetPhos.....	33
Figure 3.7: Interface du programme CSS-Palm.....	34
Figure 3.8: Interface du programme PROSPEROUS.....	34
Figure 3.9: Processus d'alignement multiple par T-Coffee.....	36
Figure 3.10: Interface du programme T-Coffee.....	37
Figure 3.11: Résultat de la requête d'alignement fourni par T-Coffee.....	37
Figure 3.12: Programme Jalview avec les différentes options de couleurs.....	38
Figure 3.13: Couleurs selon le pourcentage d'identité « PID ».....	38
Figure 4.1: Profil d'hydrophobicité de la CX23 avec la représentation en domaine.....	41
Figure 4.2: Diagramme d'hydrophobicité de la CX23.....	42
Figure 4.3: Topologie membranaire de la CX23 prédite par TOPCONS.....	43
Figure 4.4: Représentation en domaine de la CX23.....	44
Figure 4.5: Topologie membranaire présumée de la CX23 représentée par le programme Protter.....	47
Figure 4.6: Matrice de pourcentage d'identité de la CX23 et les trois <i>hPanx</i>	48
Figure 4.7: Alignement multiple de la CX23 et les trois <i>hPanx</i>	49
Figure 4.8: Alignement multiple de la CX23 avec les <i>hPanx</i> montrant les domaines Nt, M1, boucle E1, et le domaine M2.....	51
Figure 4.9: Alignement multiple de la CX23 avec les <i>hPanx</i> montrant la boucle I, les domaines M3, boucle E2, et le domaine M4.....	51

Liste des tableaux

Tableau 2.1: Les principaux outils utilisés pour l'étude des modifications post-traductionnelles.	23
Tableau 2.2: Les principaux programmes utilisés pour les différents types d'alignement.	25
Tableau 3.1: Nomenclature de la CX23.	29
Tableau 3.2: Échelle de Kyte et Doolittle.	30
Tableau 4.1: La prédiction des domaines transmembranaires de la CX23 par les trois programmes.	44

Introduction

Les protéines sont classées en protéines globulaires, protéines fibreuses, et protéines membranaires. Le type et la quantité de ces dernières varient en fonction du type cellulaire et la localisation subcellulaire. L'analyse de séquence génomique complète a montré qu'environ 30% des protéines codées par ces génomes sont des protéines membranaires (**Krogh et al., 2001**).

Les protéines membranaires jouent un rôle crucial dans de nombreux processus biologiques comme le transport membranaire, l'adhérence cellulaire et la communication cellulaire (**Zhang et al., 2019**).

Parmi les protéines qui assurent la communication cellulaire, nous avons les protéines des jonctions communicantes. Elles correspondent à deux familles de protéines appelées *connexines* (*Cx*) chez les vertébrés et *innexines* (*Inx*) chez les invertébrés. Par la suite une troisième famille appelée *pannexines* (*Panx*) a été identifiée chez les vertébrés par **Panchin et collaborateur** en l'an **2000**.

Dans le génome humain, les connexines constituent une famille multigénique de 21 isoformes réparties en cinq groupes (α , β , γ , δ , ϵ). Par ailleurs, la famille des pannexines est représentée par trois isoformes, *hPanx1*, *hPanx2*, *hPanx3*. Les connexines et les pannexines sont des protéines membranaires polytopiques partageant une topologie similaire de quatre domaines transmembranaires. Les connexines servent à la formation des canaux jonctionnels reliant le cytoplasme de cellules voisines, tandis que les pannexines ne peuvent former que des hémicanaux qui relient le cytoplasme au milieu extracellulaire (**Bao et al., 2004 ; Oyamada et al., 2005**).

La caractéristique clé qui distingue ces deux familles de protéines est la présence de six résidus cystéines dans les deux boucles extracytoplasmiques chez les connexines contrairement aux pannexines qui comportent seulement quatre résidus cystéine. Il s'est avéré que la connexine23 contient également quatre résidus cystéine d'où l'intérêt de la comparer avec les pannexines. Par ailleurs, il existe très peu de travaux consacrés à la Connexine23 (**Penuela et al., 2013**).

Ainsi, le premier objectif de notre étude est la prédiction de la topologie de la connexine23 humaine (CX23) afin de caractériser les domaines transmembranaires et les domaines intra- et extracellulaires. Dans ce but, trois programmes ont été utilisés : ProtScale

(Gasteiger *et al.*, 2005), Δ G Predictor (Hessa *et al.*, 2007) et TOPCONS (Bernsel *et al.*, 2009).

Le deuxième objectif est de prédire quelques sites de modifications post-traductionnelles que peut subir la CX23 à l'aide des outils informatiques tels que les programmes NetNGlyc (Gupta *et al.*, 2004) et NetOGlyc (Steentoft *et al.*, 2013) pour la glycosylation, NetPhos (Blom *et al.*, 2004) pour la phosphorylation, CSS-Palm (Ren *et al.*, 2008) pour la palmitoylation et PROSPEROus (Song *et al.*, 2018) pour le clivage.

Le dernier objectif de notre travail est de comparer la séquence de la CX23 avec les trois séquences des pannexines humaines par la réalisation d'un alignement multiple de séquences. Pour cela, nous avons utilisé le programme T-Coffee (*Tree-based Consistency Objective Function for Alignment Evaluation*) (Notredame *et al.*, 2000).

Ce manuscrit commence par une introduction générale, présentant la problématique du sujet, suivie de deux parties :

Dans la première partie, une synthèse bibliographique présente deux chapitres :

- Le chapitre I rassemble les connaissances concernant les protéines membranaires, les connexines, les pannexines et la connexine23.
- Le chapitre II introduit la notion de bioinformatique ainsi que les bases de données et les différentes méthodes informatiques utilisées dans ce travail.

La deuxième partie, expérimentale, comporte deux chapitres :

- Le chapitre III explique les méthodes et techniques appliquées dans ce travail.
- Le chapitre IV présente les résultats obtenus au cours de notre travail et leur interprétation.

Enfin, notre travail se termine par une conclusion.

Partie 1 :
Synthèse bibliographique

Chapitre I :
Connexines & Pannexines

1 Classes des protéines membranaires

En fonction de leurs modes d'interaction avec la bicouche lipidique, les protéines membranaires peuvent être classées en trois catégories, (i) les protéines intégrales, (ii) les protéines à ancrage lipidique, (iii) les protéines périphériques (**Lodish *et al.*, 2008**).

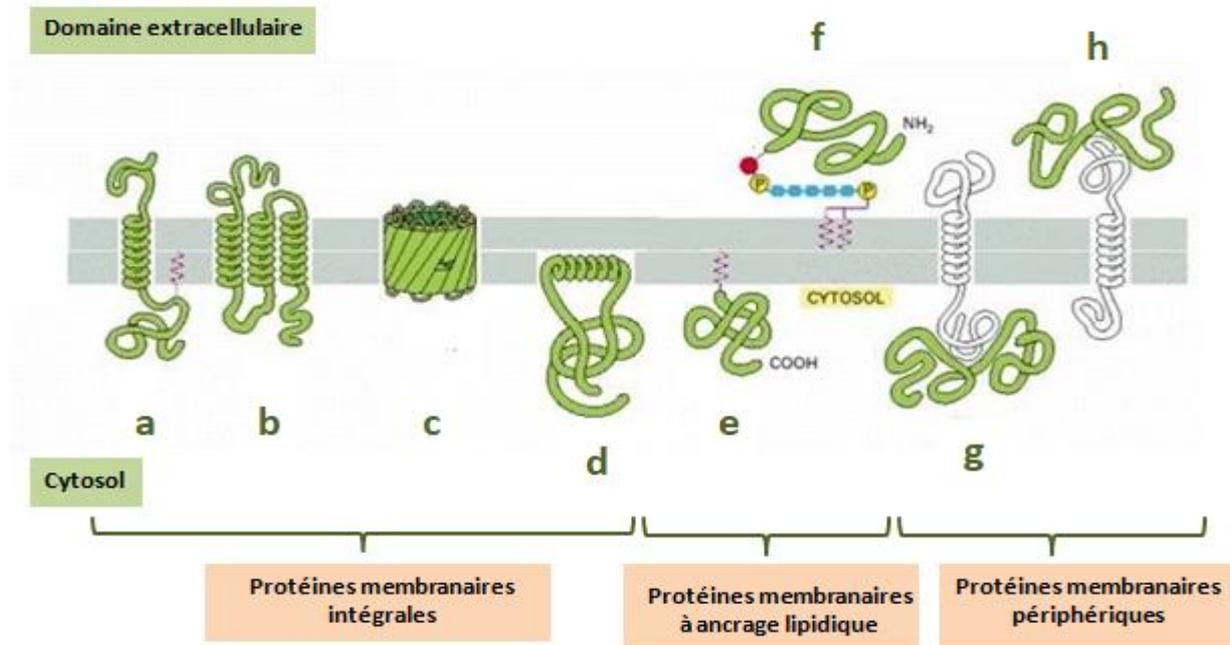


Figure 1.1 : Classes des protéines membranaires (D'après **Alberts *et al.*, 2018**).

La figure représente les trois classes des protéines membranaires. La première classe désigne les protéines intégrales (en vert) interagissant directement avec la bicouche lipidique (en gris) de quatre façons différentes selon la caractéristique du domaine transmembranaire qui peut être : (a) une seule hélice α , (b) plusieurs hélices α , (c) plusieurs brins β , (d) une hélice α parallèle au plan de la membrane. La deuxième classe désigne les protéines à ancrage lipidique représentées par deux types : des protéines acylées (e) par un acide gras myristique ou palmitique situées à la face cytosolique de la membrane, des protéines ancrées par GPI (f) situées à la face extracellulaire. La dernière classe désigne les protéines périphériques (en vert) qui peuvent s'associer à l'une des deux faces de la membrane, par un contact indirecte (g ; h) *via* des interactions non covalentes avec les protéines intégrales.

1.1 Les protéines membranaires intégrales

Elles constituent la classe la plus importante des protéines membranaires. Elles interagissent directement avec la membrane plasmique par un segment transmembranaire. Les protéines transmembranaires traversent la bicouche lipidique et comprennent trois types de domaines. Les domaines intra- et extracytoplasmique sont en contact avec l'environnement aqueux donc comportent des acides aminés hydrophiles. Cependant le domaine

transmembranaire est constitué de nombreux acides aminés hydrophobes dont les chaînes latérales interagissent avec les queues hydrophobes des phospholipides de la bicouche lipidique (Lodish *et al.*, 2016).

Selon le type de structure secondaire qui constitue les domaines transmembranaires et le nombre de ces domaines, trois groupes de protéines intégrales sont distingués.

- Les protéines bitopiques : le domaine transmembranaire est constitué d'une seule hélice α (figure 1.1a).

- Les protéines polytopiques : les domaines transmembranaires sont constitués soit de plusieurs hélices α dont le nombre varie de 2 à 19 (figure 1.1b), soit de plusieurs brins β dont le nombre varie de 8 à 22 (figure 1.1c) (Lehninger *et al.*, 2017).

- Les protéines monotopiques : qui peuvent être ancrées dans un des deux feuilletts de la membrane par le biais d'une hélice α parallèle au plan de la membrane (figure 1.1d) (Alberts *et al.*, 2018)

Les protéines à hélices α transmembranaires se trouvent dans toutes les membranes biologiques et sont plus diversifiées sur le plan fonctionnel par rapport aux protéines à brin β qui se localisent majoritairement dans les membranes externes des bactéries, mitochondries, et chloroplastes (Fagerberg *et al.*, 2010). Sur le plan structural, les protéines membranaires se replient dans leur conformation native et les sous-unités peuvent se regrouper pour former des canaux pour les protéines à hélice α comme le récepteur nicotinique de l'acétylcholine qui est un canal pentamérique. Tandis que les protéines à brin β ont tendance à former un tonneau β comme celui des porines constitué de 16 brins.

1.2 Les protéines membranaires à ancrage lipidique

Ce sont des protéines hydrosolubles qui se situent sur les deux faces de la membrane plasmique. Elles sont liées par covalence à une molécule de lipide dont la queue hydrophobe est incorporée dans l'une ou l'autre des feuilletts (Lodish *et al.*, 2016 ; Lehninger *et al.*, 2017 ; Alberts *et al.*, 2018).

Il existe trois principaux types de protéines membranaires à ancrage lipidique, les protéines acylées, les protéines prénylées, et les protéines à glycosylphosphatidylinositol (GPI).

- Les protéines acylées

Ce sont des protéines ayant subi une modification post-traductionnelle de type acylation qui conduit à la liaison covalente d'un acide gras. L'acide myristique (C14) se lie par une liaison amide à un résidu glycine N-terminal et l'acide palmitique (C16) se lie par une liaison thioester à un résidu cystéine. Les protéines acylées se localisent dans la face interne de la membrane plasmique (**figure 1.1e**).

- **Les protéines prénylées**

Ce sont des protéines cytosoliques ayant subi une prénylation. Cela est le résultat de la liaison covalente thioéther d'un polymère hydrophobe d'isoprène sur un résidu cystéine près de l'extrémité C-terminale. Ce résidu cystéine appartient au motif «CAAX» dont C représente un résidu cystéine, A représente un acide aminé aliphatique et X représente un résidu dont la nature détermine le type de prénylation qui doit survenir sur la protéine. Une unité d'isoprène est composée de cinq atomes de carbones et les polymères les plus connus sont le farnésyl (C15) et le géranylgeranyle (C20).

- **Les protéines à glycosylphosphatidylinositol (GPI)**

Elles se trouvent exclusivement sur la face externe de la membrane plasmique. Le GPI est un complexe constitué d'une phosphoéthanolamine liée d'une part au C-terminal de la protéine par une liaison amide covalente, et de l'autre part à un tétraholoside de trois mannoses et une glucosamine. Cette dernière est attachée à un phosphatidylinositol au moyen d'une liaison osidique (**figure 1.1f**).

1.3 Les protéines membranaires périphériques

Elles peuvent être associées à l'une des deux faces de la membrane plasmique, *i.e.*, la face cytosolique ou la face extracellulaire. Certaines protéines membranaires périphériques ne sont pas en contact direct avec la bicouche lipidique. Elles peuvent établir des interactions non covalentes de type électrostatiques ou hydrogènes avec les domaines hydrophiles des protéines intégrales (**figure 1.1g, h**), comme elles peuvent être en contact direct avec des groupements de têtes des phospholipides membranaires (**Lodish et al., 2016 ; Alberts et al., 2018**).

2 Fonctions des protéines membranaires

Bien que les membranes biologiques présentent la même structure de base de la bicouche lipidique, les protéines associées à une membrane particulière sont responsables de ses activités distinctives (**Lodish *et al.*, 2016**).

Les protéines membranaires notamment les protéines intégrales peuvent servir de transporteurs, de canaux et de pompes ioniques qui permettent la création et /ou le maintien des gradients électrochimique.

Elles jouent aussi le rôle de récepteurs pour les molécules de signalisation via leur domaine extracytoplasmique (réception des signaux). Cela permet de transmettre des signaux à des molécules intracytoplasmiques (transduction des signaux). Les protéines les plus connues de cette catégorie sont les récepteurs couplés à la protéine G (RCPG) et les récepteurs à activité tyrosine kinase (RTK). Les récepteurs membranaires exercent également des activités catalytiques enzymatiques comme le récepteur à activité tyrosine kinase (RTK) et le récepteur sérine/thréonine kinase (**Muller *et al.*, 2008**). Les protéines membranaires agissent aussi comme des molécules de reconnaissance du système immunitaire.

En outre, les protéines membranaires permettent l'adhérence cellulaire du fait qu'elles sont incluses dans la formation des différents types des jonctions cellulaires. Trois types des jonctions cellulaires sont distinguées, les jonctions étanches appelées aussi jonctions serrées, les jonctions d'ancrage adhérentes et desmosomales et les jonctions communicantes par la formation des canaux. Ces jonctions exercent un rôle fonctionnel par la transmission des signaux intracellulaires.

3 Connexines et pannexines

La famille des pannexines a été identifiée pour la première fois dans le génome des mammifères en 2000 par **Panchin et collaborateurs**, grâce à leur homologie de séquence (25% à 33%) avec les innexines qui constituent les protéines des canaux jonctionnels chez les invertébrés (**Penuela *et al.*, 2013**).

Par ailleurs les pannexines partagent des caractéristiques fonctionnelles et structurales avec les connexines qui sont les sous-unités des jonctions communicantes chez les vertébrés.

En effet, les connexines, pannexines et innexines constituent trois familles des protéines membranaires polytopiques appartenant à la superfamille des protéines nommées *tetraspans* de fait qu'elles partagent une topologie similaire de quatre domaines transmembranaires.

3.1 Expression génique des connexines et des pannexines

Les connexines et les pannexines sont exprimées dans le génome humain et aussi dans le génome des autres organismes tels que la souris, le rat, la vache, le chien et le poisson zèbre.

- **Expression génique des connexines**

Dans le génome humain, il existe 21 isoformes de connexines qui constituent une famille des protéines multigéniques (Sohl et Willecke, 2004). Deux nomenclatures différentes sont utilisées pour la classification de ces protéines. La première est basée sur la masse moléculaire théorique en kDa. La deuxième nomenclature divise les 21 isoformes en plusieurs groupes en fonction de la similarité de séquence et de la longueur de leurs boucles cytoplasmiques (Byer et Berthoud, 2009 ; Eiberger *et al.*, 2001). Initialement, les connexines ont été classées en trois groupes, le groupe α , le groupe β , le groupe γ . Par la suite et grâce à des analyses de l'évolution de ces protéines, deux groupes ont été rajoutés, le groupe δ et le groupe ϵ (Begandt *et al.*, 2017). Dans chacun de ces groupes, les connexines sont numérotées en fonction de l'ordre de découverte.

Ces 21 isoformes de connexines sont codées par des gènes situés sur différents chromosomes. Ainsi, les CX37, CX40, CX50, CX59 sont codées par des gènes situés sur le chromosome 1, les gènes de la CX43 et la CX62 sont situés sur le chromosome 6, la CX46 est codée par un gène situé sur le chromosome 13 (Sohl et Willecke, 2004).

La CX43 de groupe α , la CX32 et la CX26 de groupe β , sont les connexines les mieux caractérisées (Yeager, 1998). En outre, les connexines sont exprimées dans différents tissus et types cellulaires à l'exception des muscles squelettiques différenciés, des érythrocytes et des spermatozoïdes matures. La CX43 est la connexine la plus ubiquitaire. Elle est exprimée de façon endogène dans plus de 34 tissus englobant 46 types de cellules y compris les hépatocytes, les cardiomyocytes, les kératinocytes, les astrocytes (Račkauskas *et al.*, 2010).

- **Expression génique des pannexines**

Contrairement aux connexines, il n'existe que trois isoformes des pannexines dans le génome humain nommées *pannexine 1* (hPanx1), *pannexine 2* (hPanx2), et *pannexine 3* (hPanx3) (Ishikawa *et al.*, 2011).

La *hPanx 1* et la *hPanx 3* sont codées par des gènes situés sur le chromosome 11, et le gène de la *hPanx3* est situé sur le chromosome 22 (Yen et Saier, 2007 ; Penuela *et al.*, 2013).

La *hPanx1* est la plus ubiquitaire. Elle est exprimée dans différentes cellules comme les neurones, les cardiomyocytes, les kératinocytes et les cellules épithéliales. En revanche, la *hPanx2* est plus répandue dans le système nerveux central. Alors que la *hPnx3* se trouve principalement dans les ostéoblastes, les fibroblastes, et les chondrocytes (Penuela *et al.*, 2013, Baranova *et al.*, 2004).

3.2 Structure des connexines et pannexines

- Structure des connexines

Plusieurs techniques ont permis de mettre en évidence l'organisation topologique des connexines au sein de la membrane plasmique. Des études basées sur l'analyse topologique, la microscopie électronique, les données de diffraction des rayons X, la protéolyse, et le marquage par des anticorps dirigés contre des épitopes extracellulaires ont été établies notamment sur la Cx26, la Cx32, et la Cx43 (Goodenough *et al.*, 1988).

Les connexines humaines présentent une topologie commune. Elles sont constituées de quatre domaines transmembranaires (M1, M2, M3, M4) α -hélicoïdaux, deux boucles extracellulaires (E1, E2) formées par un empilement des feuillettes β antiparallèles (tonneau β), avec une boucle intracellulaire (I) (Foote *et al.*, 1998). Les extrémités amino- et carboxy-terminales (Nt, Ct) sont exposées du côté cytoplasmique (figure 1.2).

Sur le plan structural, les connexines de groupe α présentent une boucle cytoplasmique (I) avec la région C-terminale plus longue par rapport aux connexines du groupe β (Yeager, 1998). Toutes les isoformes conservent des régions du second domaine transmembranaire (M2). Tandis que la boucle intracytoplasmique et l'extrémité C-terminale sont les moins conservées. Cela permet une régulation différente de chaque isoforme, et donc une participation à des cascades de signalisation uniques en particulier lorsque plusieurs isoformes sont exprimées dans la même cellule.

L'intérêt s'est porté aussi sur les domaines extracellulaires des connexines qui sont impliqués dans la spécificité de l'arrimage de l'hémicanal. Ces domaines sont hautement conservés en raison de la présence de trois résidus cystéines dans chaque boucle chez toutes les isoformes à l'exception de la Cx23 qui présente seulement deux cystéines dans chaque boucle. Ces cystéines servent à la formation des ponts disulfures intramoléculaires. Ces

derniers servent à lier et maintenir les deux boucles extracellulaires en une conformation compacte et rigide pour qu'elles puissent interagir avec les boucles des connexines voisines en formant un canal jonctionnel. Ainsi, la mutation de ces résidus cystéines entraîne la perte de la fonction du canal. (Foote *et al.*, 1998 ; Penuela *et al.*, 2013 ; Brikci Nigassa, 2014)

Les connexines permettent la formation des structures hexamériques appelées connexons qui peuvent être homomériques constitués par des connexines identiques, ou bien hétéromériques si les connexines sont différentes. Après leur translocation jusqu'à la membrane plasmique, ces connexons forment en arrimage avec d'autres connexons des cellules voisines des canaux jonctionnels qui peuvent également être homotypiques ou hétérotypiques selon la nature des connexons qui les constituent (Oyamada *et al.*, 2005 ; dykes *et al.*, 2004) (figure 1.2).

Les canaux jonctionnels peuvent se regrouper par milliers pour former des jonctions gap. Ces canaux sont spécialisés dans la communication cellule-cellule en faisant lier le cytoplasme de deux cellules adjacentes. Ainsi, ils permettent un échange de certaines molécules de faible poids moléculaire ($M < 1000$ Da) comme les ions, l'AMPc et l'IP₃ via ces canaux (signalisation paracrine). Par conséquent, ils servent au maintien de diverses fonctions physiologiques telles que la synchronisation de la réponse cellulaire comme la contraction musculaire et la libération de l'insuline par les cellules β pancréatiques. Ils jouent également un rôle crucial dans le maintien de l'homéostasie tissulaire comme celle des hépatocytes. Par ailleurs, plusieurs découvertes des maladies génétiques dues à des mutations au niveau des jonctions gap ont été mentionnées (Oyamada *et al.*, 2005 ; Brikci Nigassa, 2014).

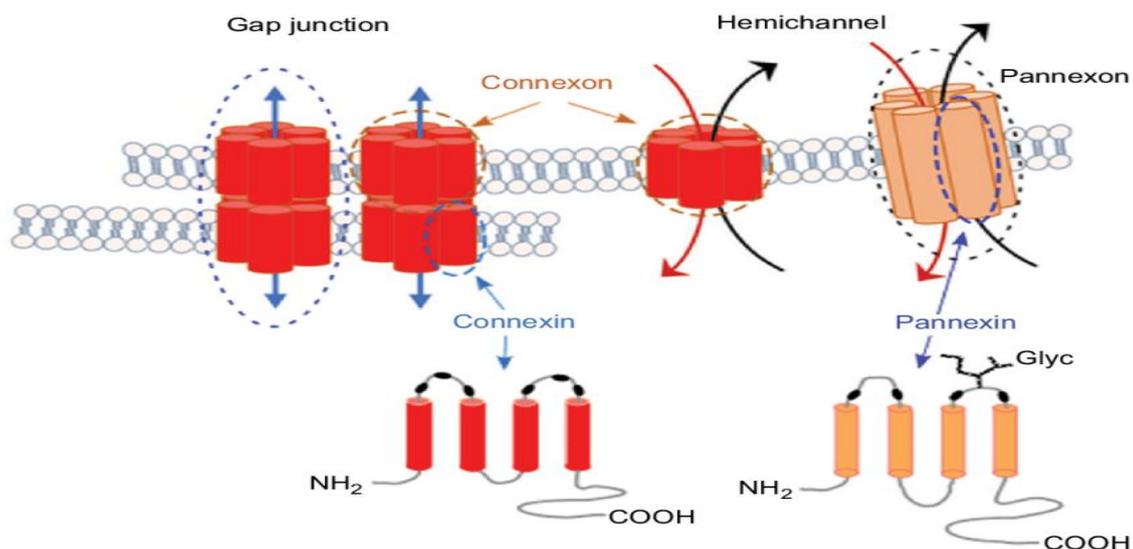


Figure 1.2 : Structures des connexines et des pannexines (Makarenkova *et al.*, 2018).

- **Structure des pannexines**

Les pannexines humaines présentent une topologie similaire à celle des connexines. Elles présentent quatre segments transmembranaires (M1, M2, M3, M4), deux boucles extracytoplasmiques (E1, E2) avec une boucle intracytoplasmique (I), les extrémités N- et C-terminales se situent dans la face cytoplasmique (**Baranova *et al.*, 2004**) (**figure 1.2**).

De plus, chez les trois pannexines, les boucles extracytoplasmiques sont les plus conservées, contenant deux résidus cystéines dans chacune, tandis que le domaine C-terminal contient les séquences les plus variables (**Penuela *et al.*, 2013**).

La *hPanx1* dont le poids moléculaire est d'environ 47,6 kDa, et la *hPanx3* avec un poids moléculaire de 44,7 kDa, partagent 30% d'homologie de séquence. Cependant, la *hPanx2* présente un poids moléculaire nettement plus grand d'environ 74,4 kDa en raison de la longueur de sa queue C-terminale (**Penuela *et al.*, 2013**).

L'oligomérisation des *hPanx1* forme un canal hexamérique, tandis que les *hPanx2* peuvent former des canaux heptamériques ou octamériques. Suivant la nomenclature établie pour les connexons des connexines, les oligomères formés par les pannexines sont appelés pannexons. Alors que, les connexons désignent des hémicanaux, les pannexons désignent des canaux à membrane unique (**figure 1.2**). Par contre, de nombreux auteurs déconseillent fortement d'utiliser l'appellation hémicanaux pour les pannexons car cela implique qu'ils sont destinés à s'assembler en un canal intercellulaire (**Penuela *et al.*, 2013, Ambrosi *et al.*, 2010**).

Les canaux formés par les *hPanx1* jouent un rôle essentiel dans la signalisation paracrine entre les cellules par la libération de l'ATP dans l'espace extracellulaire. Cela conduit à l'activation des récepteurs purinergiques ionotropiques P2X et métabotropiques P2Y impliqués dans la propagation des vagues de calcium. La liaison de l'ATP aux P2Y augmente l'inositol 1, 4, 5 triphosphate et libère le Ca^{2+} emmagasiné dans le RE. Ensuite, l'ouverture des canaux à membrane unique sera activée en libérant plus d'ATP dans le milieu extracellulaire. Cette dernière permet de remplir diverses fonctions physiologiques telles que la sensation gustative, la clairance mucociliaire des voies respiratoires, la capture insulino-dépendante du glucose par les adipocytes, la réponse immunitaire, l'apoptose et la différenciation des kératinocytes et des neurones (**Penuela *et al.*, 2013**).

3.3 Modifications post-traductionnelles des connexines

Les modifications post-traductionnelles sont des transformations chimiques que peuvent subir les protéines après leur traduction en produisant des changements au niveau de la fonction, de la demi-vie et de la localisation cellulaire de ces protéines. Deux catégories de modifications post-traductionnelles sont distinguées, (i) le clivage de la protéine de manière auto-catalytique ou par des protéases (la protéolyse), (ii) l'ajout réversible ou irréversible des molécules ou des groupements fonctionnels sur des sites spécifiques.

L'analyse des séquences des connexines a montré que ces protéines présentent des sites pour différents types de modifications post-traductionnelles telles que la glycosylation, la phosphorylation, la palmitoylation et la protéolyse.

- **La glycosylation**

La glycosylation consiste en l'addition d'un glycane sur un résidu asparagine de la protéine cible conduisant à la *N*-glycosylation, ou sur des résidus sérine/thréonine conduisant à la *O*-glycosylation.

Pour les connexines aucun site de glycosylation n'a été trouvé (**Martin et Evans, 2004** ; **Saez et al., 2003**). Cependant, l'étude de **Rahman et collaborateurs (1993)** a montré que la CX32 est *N*-glycosylée.

- **La phosphorylation**

La phosphorylation est un processus réversible qui constitue une voie de régulation d'une protéine par l'addition des groupements phosphates par des kinases sur des résidus sérine (S), thréonine (T) ou tyrosine (Y) d'une protéine cible.

La phosphorylation des connexines est largement étudiée notamment la phosphorylation de la connexine43 dont au moins 19 sites ont été identifiés par plusieurs kinases (Src, MAPK, CDK1, PKC, PKA, AKT) (**Solan et Lampe, 2014**). Les sites de phosphorylation se localisent majoritairement dans les domaines intracellulaires des protéines cibles notamment le domaine C-terminal. La phosphorylation des connexines semble avoir un impact sur l'altération de la communication jonctionnelle intercellulaire conduisant à la fermeture du canal, l'assemblage de la plaque jonctionnelle, et le trafic vers la membrane plasmique (**Brikci Nigassa, 2014**).

- **La palmitoylation**

Elle consiste à la fixation d'un acide palmitique (C16) sur un résidu cystéine d'une protéine cible.

L'étude de la palmitoylation chez les connexines a montré que seule la connexine32 est palmitoylée sur le résidu cystéine C277 et C283 au niveau de la queue C-terminale (**Locke et al., 2006**).

- **La protéolyse**

La protéolyse est une modification post-traductionnelle qui consiste à hydrolyser une ou plusieurs liaisons peptidiques par des protéases telles que les caspases impliquées dans l'inflammation et l'apoptose.

4 Connexine23

4.1 Expression génique de la connexine23

Parmi les 21 isoformes des connexines qui ont été identifiées dans le génome humain, la connexine23 (CX23) représente la séquence protéique la plus courte comportant 205 résidus d'acides aminés avec un poids moléculaire de 23,76 kDa. Par conséquent, la CX23 a été classée comme le seul représentant du groupe ϵ qui est le dernier des cinq groupes des connexines (**Byer et Berthoud, 2009**).

Le gène (*GJF1*) de la CX23 est codé par le chromosome 6 (**Sohl et Willecke, 2004**). De plus, des orthologues de la CX23 ont été identifiés chez d'autres organismes tels que la connexine23 de la souris (Cx23), et celle du poisson zèbre qui sont les connexines23 les plus étudiées avec des sources très limitées dans la bibliographie.

Des études en protéomique sur le cristallin de la souris adulte ont montré que la Cx23 est un composant des membranes des fibres du cristallin avec une abondance similaire à celle de la Cx46 et inférieure à celle de la Cx50 (**Gustincich et al., 2003**). Par ailleurs, la Cx23 du poisson zèbre a été identifiée comme la quatrième connexine exprimée dans le cristallin. **Puk et collaborateurs (2008)** ont montré qu'une mutation dans la Cx23 est associée à une perturbation complète de la formation du cristallin indiquant que cette connexine joue un rôle important dans son développement.

4.2 Structure de la connexine23

La séquence de la connexine 23 présente des particularités par rapport aux autres connexines (**Iovine et al., 2007 ; Sonntag et al., 2009**). La caractéristique principale que présente la connexine23 est la présence de deux résidus cystéines dans chaque boucle extracellulaire au lieu de trois cystéines comme il est connu chez les autres connexines. **Sonntag et collaborateurs (2009)** ont montré par un alignement de la Cx23 avec la Cx31 que les motifs des cystéines dans les deux boucles extracellulaires de la Cx23 sont caractérisés par l'absence de résidus cystéines centraux en les comparant avec les motifs canoniques des autres connexines, cela est montré aussi dans **la figure 1.3 (Benayad, 2018)**. Ainsi, cette caractéristique est partagée avec la famille des pannexines qui contiennent également quatre résidus cystéines.

En outre, les études *in vitro* établies sur la Cx23 (souris et poisson zèbre) ont montré que celle-ci ne forme pas des canaux jonctionnels mais forme uniquement des hémicanaux (**Sonntag et al., 2008 ; Berthoud et al., 2016**). Cela peut être similaire à ce qui était déjà montré pour les canaux à membrane unique des pannexines.

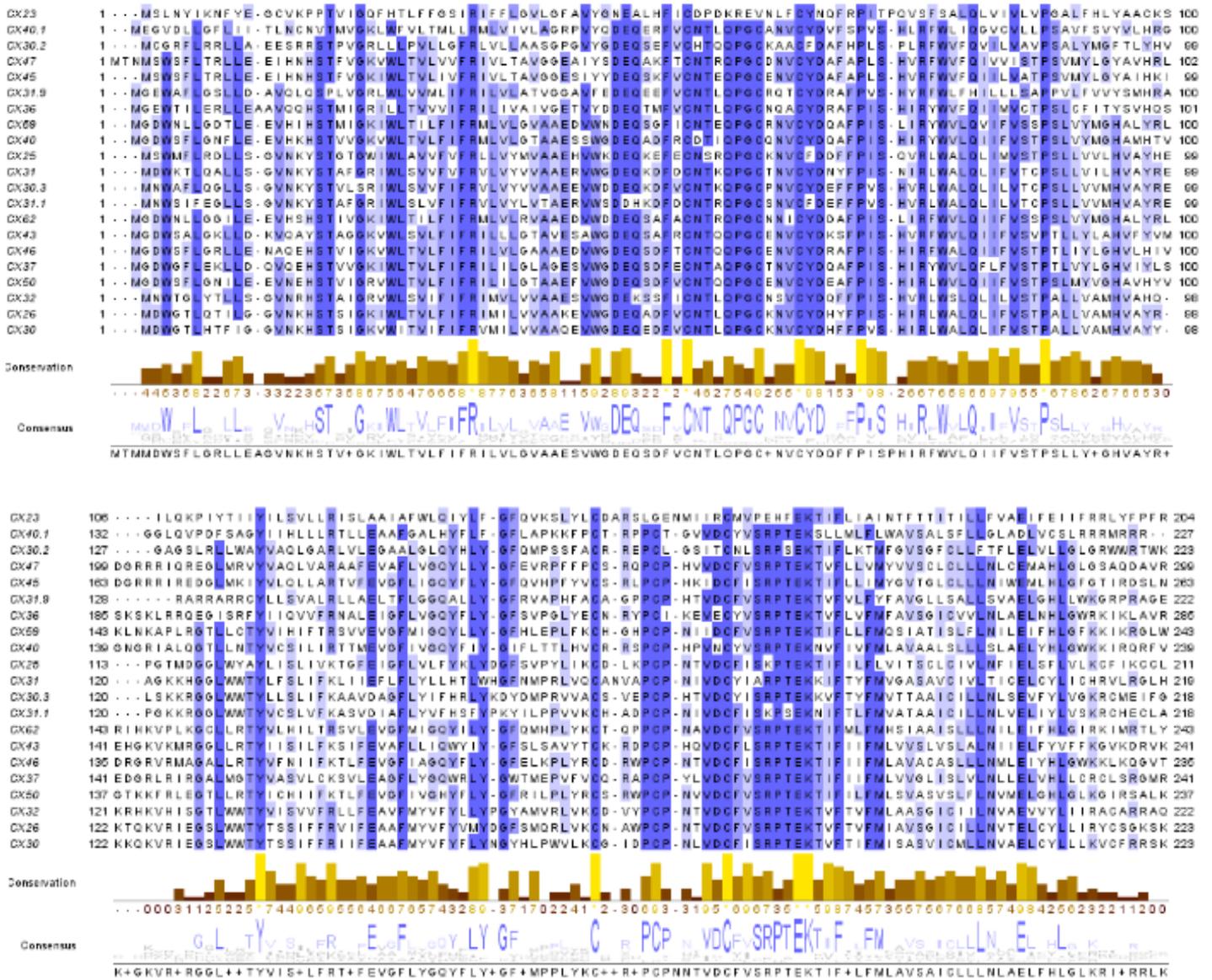


Figure 1.3 : Alignement multiple des 21 isoformes des connexines (Benayad, 2018).

Chapitre II :
Analyse des protéines membranaires
par bioinformatique

1 Introduction

La bioinformatique est un domaine de recherche pluridisciplinaire qui utilise les techniques informatiques afin d'analyser, comprendre et organiser les informations associées aux biomolécules. Autrement dit, elle constitue une discipline qui étudie les outils de gestion de l'information biologique en s'étendant sur un large spectre d'application.

A l'échelle chronologique, les débuts de la bioinformatique remontent à plus de 50 ans (**Gauthier et al., 2019**), *i.e.*, le début des années 70.

Plusieurs définitions ont été adoptées pour introduire ce terme comme celle proposée par le centre national de l'information biotechnologique (**NCBI 2001**) qui définit la bioinformatique comme suit :

« *Bioinformatics is the field of science in which biology, computer science, and information technology merge into a single discipline* ».

2 Banques de données

Une banque de données est un vaste ensemble d'informations organisées. Elle constitue une source de données biologiques en permettant de collecter, stocker, archiver, ainsi que faciliter l'exploitation et la diffusion de ces données d'une manière structurée et consultable.

Les données biologiques peuvent être sous forme de séquences pour les acides nucléiques, ou sous forme de séquences et de structures pour les protéines.

La biochimiste américaine **Margaret Dayhoff** fut la première à voir l'intérêt de rassembler toutes les données sur les séquences protéiques. Pour cela, elle a créé avec ses collègues le premier atlas de séquences protéiques en 1965, qui est devenu en 1989 la banque de données PIR. En parallèle, deux banques de données de séquences nucléiques ont pris leur essor, GenBank aux États-Unis, et EMBL en Europe. Par la suite, le logiciel BLAST a été mis en œuvre en 1990, suivi par un enchaînement de procédés de séquençage des différents génomes jusqu'à l'an 2001 où la première séquence complète du génome humain (plus de 3000 Mpb) a été publiée.

Les bases de données biologiques peuvent être classées en trois groupes.

2.1 Banques de données primaires/généralistes

Elles correspondent à une collection de données plutôt hétérogènes sur les séquences nucléotidiques et les séquences protéiques.

- GenBank (**Benson et al., 2012**), EMBL « *European Molecular Biology Laboratory* » (**Stoesser et al., 2003**), DDBJ « *DNA Data Bank of Japan* » (**Tateno et al., 2002**), sont les principales banques de données des séquences nucléotidiques.

- UniProt « *Universal Protein Resource* » (**UniProt Consortium, 2007**), est la principale banque de données des séquences protéiques. Elle regroupe deux sections:
 - Swiss-Prot (**Boeckmann *et al.*, 2003**), pour les données corrigées manuellement.
 - TrEMBL (**Boeckmann *et al.*, 2003**), pour les données non corrigées.

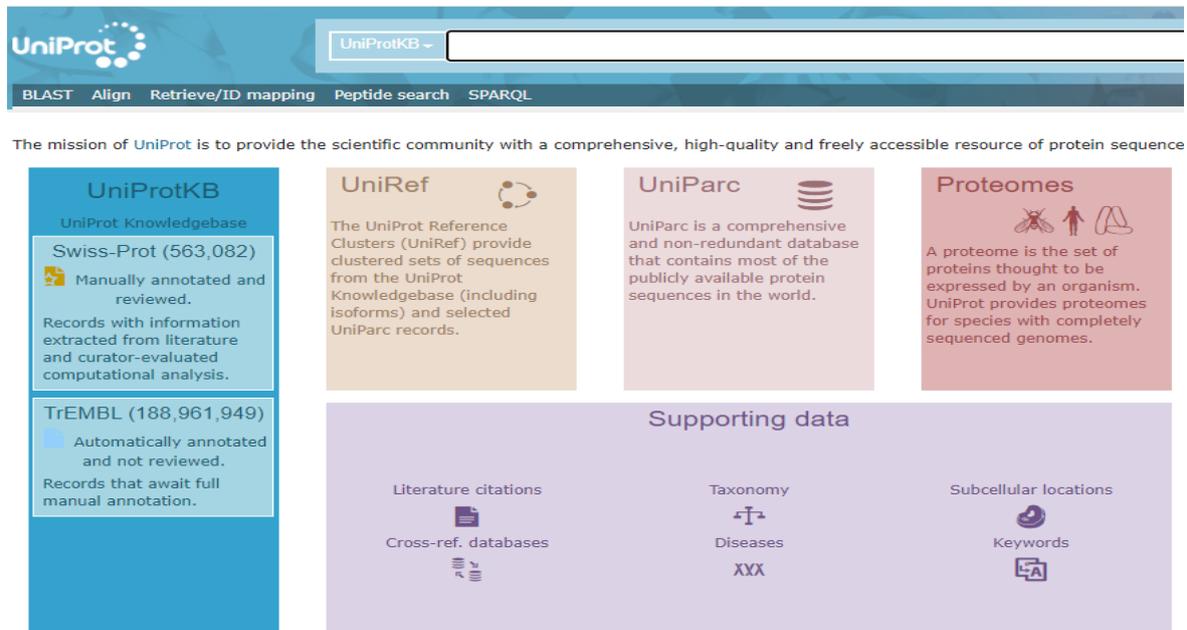


Figure 2.1 : Interface de la banque de données UniProt.

2.2 Banques de données secondaires/ spécialisées

Elles correspondent à une collection de données homogènes établies autour d'une thématique particulière. Elles présentent des informations sur les séquences et les structures.

- PROSITE (**Bairoch *et al.*, 1997**) présente les motifs protéiques .
- CATH (**Sillitoe *et al.*, 2015**) classifie les domaines protéiques.
- SCOP2 (**Murzin *et al.*, 1995**) classifie les structures protéiques.
- TFD (**Ghosh, 1992**) regroupe les facteurs de transcription.
- Pfam (**Mistry *et al.*, 2007**) regroupe les familles des protéines et leurs domaines.

2.3 Banques de données composites

Elles hébergent une variété de bases de données simultanément comme NCBI (**NCBI Resource Coordinators 2016**) qui regroupe des sources de données nucléotidiques et protéiques.

3 Étude de la topologie des protéines membranaires par des techniques informatiques

Les protéines membranaires sont importantes dans différents processus cellulaires. Leur analyse est essentielle surtout qu'il a été prouvé que ceci est hautement relatif à leur fonction (**Zhang et al., 2019**). Ceci implique l'identification des structures secondaires des domaines transmembranaires de ces protéines ainsi que leur nombre (bitopique/polytopique), leur orientation « in/out » et la localisation des extrémités N- et C-terminales, *i.e.*, l'analyse de sa topologie.

Deux structures basiques des domaines transmembranaires des protéines intégrales ont été identifiées, les hélices α et les brins β . Cependant, il est plus facile de prédire les segments transmembranaires en hélice α longs et hydrophobes que ceux plus courts et moins hydrophobes qui sont en brins β . Pour cette raison, les méthodes dédiées à la prédiction de ces derniers sont beaucoup plus rares que celles développées pour les hélices α . De plus, la topologie de 400 protéines membranaires en hélice α a été déterminée expérimentalement (**Elofsson et von Heijne, 2007**).

De nombreuses méthodes expérimentales conventionnelles ont été utilisées. Cependant, ces techniques sont très coûteuses et prennent beaucoup de temps d'où la nécessité de développer les méthodes *in silico*¹. Ces dernières sont des approches théoriques efficaces et fiables avec un gain de temps considérable, permettant de guider les expériences *in vivo* et *in vitro* pour l'étude des protéines.

Au fil du temps, plusieurs méthodes de prédiction de la topologie ont été mises au point pour les protéines membranaires en hélice α .

De plus, les algorithmes de prédiction étaient basés sur la détermination des segments transmembranaires présentant les meilleurs scores d'hydrophobicité utilisant différentes échelles comme celles de **Eisenberg et collaborateurs (1984)**, **Wolfenden et collaborateurs (1981)**, et l'échelle de **Kyte et Doolittle (1982)**. Cette dernière est utilisée dans le programme ProtScale (**Gasteiger et al., 2005**) qui présente le résultat sous forme d'un diagramme appelé *profil d'hydrophobie*.

D'autres caractéristiques ont été prises en considération afin d'optimiser la prédiction de la topologie comme l'utilisation de la règle « *positive inside rule* » qui montre la distribution

1 *In silico* : désigne une recherche effectuée au moyen de modèles informatiques.

de résidus chargés positivement comme l'arginine et la lysine au niveau des boucles cytoplasmiques (**Sipos et von Heijne, 1993 ; Nilsson *et al.*, 2005**).

De plus, d'autres auteurs ont montré la tendance des résidus aromatiques tryptophane et tyrosine à se localiser près des extrémités des segments transmembranaires (**Engelman et Steitz, 1986 ; Sipos et von Heijne, 1993 ; Muller *et al.*, 2008**).

Au cours des dernières décennies, plusieurs programmes ont été développés pour la prédiction, en utilisant des approches d'apprentissage automatique. Ces logiciels utilisent des modèles statistiques tels que le modèle de Markov caché (MMC) e.g., HMMTOP (**Tusnady *et al.*, 2001**), et TMHMM (**Krogh *et al.*, 2001**), des réseaux de neurones (RN) (**Rost *et al.*, 1995**), e.g., PHDhtm (**Rost *et al.*, 1996**). La majorité de ces programmes sont disponibles gratuitement sur le web.

Il existe ainsi des logiciels combinant les résultats de plusieurs méthodes comme TOPCONS (**Bernsel *et al.*, 2009**), qui intègre le résultat de cinq méthodes de prédiction. Il sert aussi à différencier les segments transmembranaires de la séquence du peptide signal. De même, on peut citer ConPred II (**Arai *et al.*, 2004**), qui combine les résultats de huit méthodes et CCTOP (**Dobson *et al.*, 2015**), qui fait intervenir dix méthodes.

Un autre outil, ΔG predictor développé par **Hessa et collaborateurs (2005)** s'avère aussi très intéressant pour la prédiction des hélices transmembranaires. Pour une séquence en acides aminés d'une hélice α transmembranaire donnée, le programme prédit la différence d'énergie libre pour que cette séquence soit insérée dans la membrane de réticulum endoplasmique (RE) au moyen d'un translocon.

4 Étude des modifications post-traductionnelles

Les modifications post-traductionnelles jouent un rôle crucial dans divers processus biologiques y compris le maintien de la structure des protéines, la signalisation, la régulation et la défense immunitaire. Pour cela, l'identification des sites de modifications post-traductionnelles s'avère très intéressante.

Plus de 32 logiciels de prédiction sont disponibles sur le web pour les différents types de modifications post-traductionnelles (**tableau 2.1**). Sachant qu'environ 350 types de celles-ci ont été découverts (**Zhou *et al.*, 2006**),

Tableau 2.1 : Les principaux outils utilisés pour l'étude des modifications post-traductionnelles.

Type de MPT	Programme	Site web du programme	Référence
Glycosylation	YingOYang	http://www.cbs.dtu.dk/services/YinOYang/	Gupta et Brunak, 2001
	NetNGlyc	http://www.cbs.dtu.dk/services/NetNGlyc/	Gupta <i>et al.</i> , 2004
	NetOGlyc	http://www.cbs.dtu.dk/services/NetOGlyc/	Steentoft <i>et al.</i> , 2013
	GPP	https://comp.chem.nottingham.ac.uk/glyco/	Hamby et Hirst, 2008
	O-GlycBase	http://www.cbs.dtu.dk/databases/OGLYCBASE/	Gupta <i>et al.</i> , 1999
	GLYCOPP	http://crdd.osdd.net/raghava/glycopp/	Chauhan <i>et al.</i> , 2012
Phosphorylation	NetPhos 3.1	http://www.cbs.dtu.dk/services/NetPhos/	Blom <i>et al.</i> , 2004
	GPS	http://973-proteinweb.ustc.edu.cn/gps_web/	Xue <i>et al.</i> , 2008
	KinasePhos 2.0	http://kinasephos2.mbc.nctu.edu.tw/	Wong <i>et al.</i> , 2007
	PPRED	http://www.cse.univdhaka.edu/~ashis/ppred/	Biswas <i>et al.</i> , 2010
	PhosphoSVM	http://sysbio.unl.edu/PhosphoSVM/	Dou <i>et al.</i> , 2006
	ScanSite	https://scansite4.mit.edu/4.0/	Obenauer <i>et al.</i> , 2003
	Musite	http://musite.sourceforge.net/	Gao <i>et al.</i> , 2010
Palmitoylation	DISPHOS	http://www.dabi.temple.edu/disphos/	Iakoucheva <i>et al.</i> , 2004
	CSS-palm	http://csspalm.biocuckoo.org/	Ren <i>et al.</i> , 2008
	SwissPalm	https://swisspalm.org/	Blanc <i>et al.</i> , 2015
	SeqPalm	http://lishuyan.lzu.edu.cn/seqpalm/	Li <i>et al.</i> , 2015
Protéolyse	CKSAAP-PALM	http://doc.aporc.org/wiki/CKSAAP-Palm	Wang <i>et al.</i> , 2009
	GPS-CCD	http://ccd.biocuckoo.org/	Liu <i>et al.</i> , 2011
	PROSPER	https://prosper.erc.monash.edu.au/	Song <i>et al.</i> , 2012
	PROSPERous	http://prosperous.erc.monash.edu/	Song <i>et al.</i> , 2018
	CaSPredictor	http://icb.usp.br/farmac/Jose/CaSPredictorfiles	Garay <i>et al.</i> , 2005
Ubiquitinylation	CasCleave 2.0	http://www.structbioinform.org/cascleave2/	Wang <i>et al.</i> , 2014
	UbPred	http://www.ubpred.org/	Radivojac <i>et al.</i> , 2010
	UbiSite	http://csb.cse.yzu.edu.tw/UbiSite/predict.php	Akimov <i>et al.</i> , 2018
	UbiNet	http://csb.cse.yzu.edu.tw/UbiNet/	Nguyen <i>et al.</i> , 2016
Methylation	UbiProber	http://bioinfo.ncu.edu.cn/UbiProber.aspx	Chen <i>et al.</i> , 2014
	MeMO	http://bioinfo.tsinghua.edu.cn/memo.html	Chen <i>et al.</i> , 2006
	BPB-PPMS	http://www.bioinfo.bio.cuhk.edu.hk/bpbppms	Shao <i>et al.</i> , 2009
	MASA	http://masa.mbc.nctu.edu.tw/	Shien <i>et al.</i> , 2009
Acetylation	iMethyl-PseAAC	http://www.jci-bioinfo.cn/iMethyl-PseAAC	Qiu <i>et al.</i> , 2014
	NetAcet	http://www.cbs.dtu.dk/services/NetAcet/	Kiemer <i>et al.</i> , 2005
	PAIL	http://bioinformatics.lcd-ustc.org/pail	Li <i>et al.</i> , 2006
	N-Ace	http://n-ace.mbc.nctu.edu.tw/	Lee <i>et al.</i> , 2010
SUMOylation	PSKacePred	http://bioinfo.ncu.edu.cn/PSKacePred.aspx	Suo <i>et al.</i> , 2012
	SUMOSP	http://bioinformatics.lcd-ustc.org/sumosp/	Xue <i>et al.</i> , 2006
	SUMOHydro	http://protein.cau.edu.cn/others/SUMOHydro/	Chen <i>et al.</i> , 2012
	GPS-SUMO	http://sumosp.biocuckoo.org/	Zhao <i>et al.</i> , 2014

5 Alignement des séquences

L'alignement des séquences protéiques est une manière de représenter deux ou plusieurs séquences les unes sous les autres afin d'obtenir le plus de correspondances possible.

En plus de rechercher les similarités entre les séquences, le processus d'alignement permet également de détecter la présence des motifs caractéristiques, de déterminer les liens évolutifs et les relations phylogénétiques entre les protéines, ainsi que d'améliorer la prédiction de leurs structures secondaires et tertiaires, et de même de leurs fonctions biologiques.

5.1 Alignement par paires

Il consiste à comparer des séquences deux à deux. Dans ce but, il existe des méthodes qui s'appuient sur la programmation dynamique². L'algorithme de **Needleman et Wunsch (1970)** réalise l'alignement global optimal pour aligner les séquences sur toute leur longueur. L'algorithme de **Smith et collaborateurs (1981)** réalise l'alignement local optimal où les séquences seront alignées sur une portion de leur longueur.

5.2 Alignement multiple

Il consiste à comparer simultanément plusieurs séquences homologues entre elles. Il permet entre autre de caractériser les régions conservées et les régions variables au sein d'une famille de séquences, en construisant leur séquence consensus³. Il faut noter que l'alignement multiple est plus puissant que l'alignement par paires, comme il peut aussi être global ou local.

Plusieurs méthodes et programmes ont été développés et mis en ligne pour effectuer les différents types d'alignement des séquences (**tableau 2.2**).

² La programmation dynamique : méthode algorithmique introduite par **R.Bellman (1950)** consiste à résoudre un problème en le décomposant en sous-problèmes.

³ Séquence consensus : séquence peptidique établie après un alignement multiple dans laquelle le résidu d'acide aminé le plus fréquent à chaque position sera noté.

Tableau 2.2 : Les principaux programmes utilisés pour les différents types d'alignement.

Type d'alignement	Méthode /programme	Référence
Recherche d'homologie	BLAST/PSI-BLAST FASTA	Atschul <i>et al.</i> , 1997 Pearson, 1994
Alignement par paires	EMBOSS Needle EMBOSS Water K-mer K-tuple	Needlemen et Wunsch, 1970 Smith et Waterman, 1981 Edgar, 2004 Wilbur et Lipman, 1983
Alignement multiple exact	MCA DCA	Lipman et Pearson, 1988 Stoye <i>et al.</i> , 1997
Alignement multiple progressif	Clustal Omega Clustal w MUSCLE MAFFT T-Coffee TM-Coffee ProbCons	Sievers <i>et al.</i> , 2011 Thompson <i>et al.</i> , 1994 Edgar, 2004 Katoh et Standley, 2013 Notredame <i>et al.</i> , 2000 Flodem <i>et al.</i> , 2016 Do <i>et al.</i> , 2005
Alignement itératif	PRRP SAGA DIALIGN	Gotoh, 1993 Notredame et Higgins, 1996 Morgenstern, 2004
Alignement intégrant la structure	STAMP 3D-Coffee Expresso	Russell et Barton, 1992 O'Sullivan <i>et al.</i> , 2004 Armougom, 2006
Alignement de motifs	PHI-Blast GLAM2	Zhang <i>et al.</i> , 1998 Frith <i>et al.</i> , 2008

Partie II :
Travaux expérimentaux
&
Résultats

Chapitre III :

Matériel & Méthodes

1 Choix de la CX23

Dans ce travail, nous avons choisi d'étudier la CX23 l'une des 21 isoformes de la famille multigénique des connexines qui ont été alignées par **Benayad (2018)**. Par ailleurs, la CX23 a été comparée aux pannexines qui ont été étudiées par **Metalsi (2019)**.

Nous nous sommes intéressés à l'analyse de la séquence peptidique de la CX23 qui contient seulement quatre résidus cystéines, deux dans chacune des deux boucles extracytoplasmiques, contrairement aux autres connexines. En revanche, cette particularité que présente la CX23 est partagée avec la famille des pannexines.

Tout d'abord, nous avons étudié la topologie membranaire de la CX23. Ensuite, nous avons étudié la prédiction des sites potentiels des modifications post-traductionnelles au moyen de différents programmes bioinformatiques. Enfin, nous avons réalisé un alignement de la CX23 avec les trois *hPanx* afin de déterminer la présence de caractéristiques communes entre ces protéines.

La séquence de la CX23 a été obtenue à partir de la banque de données UniProtKB/Swiss-Prot (<https://www.uniprot.org/>) qui contient des séquences corrigées manuellement par des scientifiques (**figure 3.1**).



The image shows a screenshot of the UniProtKB/Swiss-Prot website. At the top, there is a search bar with 'UniProtKB' selected. Below the search bar, there are navigation links: BLAST, Align, Retrieve/ID mapping, Peptide search, and SPARQL. The main heading is 'UniProtKB - A6NN92 (CXE1_HUMAN)'. Below this, there is a 'Display' section with a menu of options: BLAST, Align, Format, Add to basket, and History. The main content area shows the following information:

Entry	Protein	Putative gap junction epsilon-1 protein
Publications	Gene	GJE1
Feature viewer	Organism	<i>Homo sapiens (Human)</i>
Feature table	Status	Reviewed - Annotation score: ●●●○ - Protein uncertain ¹

Figure 3.1 : Requête pour la base de données UniProtKB/Swiss-Prot.

La banque de données UniProtKB/Swiss-Prot fournit plusieurs informations (en plus des séquences protéiques) telles la nomenclature complète de la protéine, le numéro d'accèsion, l'organisme et la localisation cellulaire (**tableau 3.1**).

Tableau 3.1 : Nomenclature de la CX23.

	Nom selon UniprotKB	Numéro d'accèsion	Acides aminés
CX23	Full=Putative gap junction epsilon-1 protein; AltName: Full=Connexin-23; Short=Cx23	A6NN92.1	205

Le fichier texte téléchargé à partir de cette base de données est sous un format particulier appelé « Fasta » qui représente la séquence sous un format à une seule lettre (**figure 3.2**). Ce format est supporté par tous les programmes utilisés dans ce travail.

FASTA ▾

**RecName: Full=gap junction epsilon-1 protein;AltName: Full=Connexin-23;
Short=Cx23**

UniProtKB/Swiss-Prot: A6NN92.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>sp|A6NN92.1|CXE1_HUMAN RecName: Full=Putative gap junction epsilon-1
protein; AltName: Full=Connexin-23; Short=Cx23
MSLNYIKNFYEGCVKPTVIGQFHTLFFGSIIRIFFLGVLFVAVYGNALHFIKDPDKREVNLFQYNQFRP
ITPQVSFSAQLVIVLVPALFHLAACKSINQECILQKPIYTIYIILSVLLRISLAAIAFWLQIYLFQF
QVKSLYLCDARSLGENMIIRCIVPEHFETIFLIAINTFTTITILLFVAEIFEIIFFRRLYFPFRQ
```

Figure 3.2 : Séquence de la CX23 au format Fasta.

2 Étude de la topologie de la CX23

Dans ce travail, nous avons réalisé la prédiction de la topologie de la CX23 dans le but de discriminer les domaines transmembranaires de ceux qui ne sont pas transmembranaires, de définir le nombre de ces domaines et de montrer l'orientation de cette protéine par rapport à la membrane. Pour cela, nous avons utilisé trois programmes : ProtScale, ΔG predictor et TOPCONS.

2.1 ProtScale

ProtScale (<https://web.expasy.org/protscale/>) est un programme qui permet de calculer les indices d'hydrophatie d'une protéine donnée en fournissant une représentation graphique sous forme d'un tracé en 2D appelé « *profil d'hydrophatie* » (**figure 3.3**).

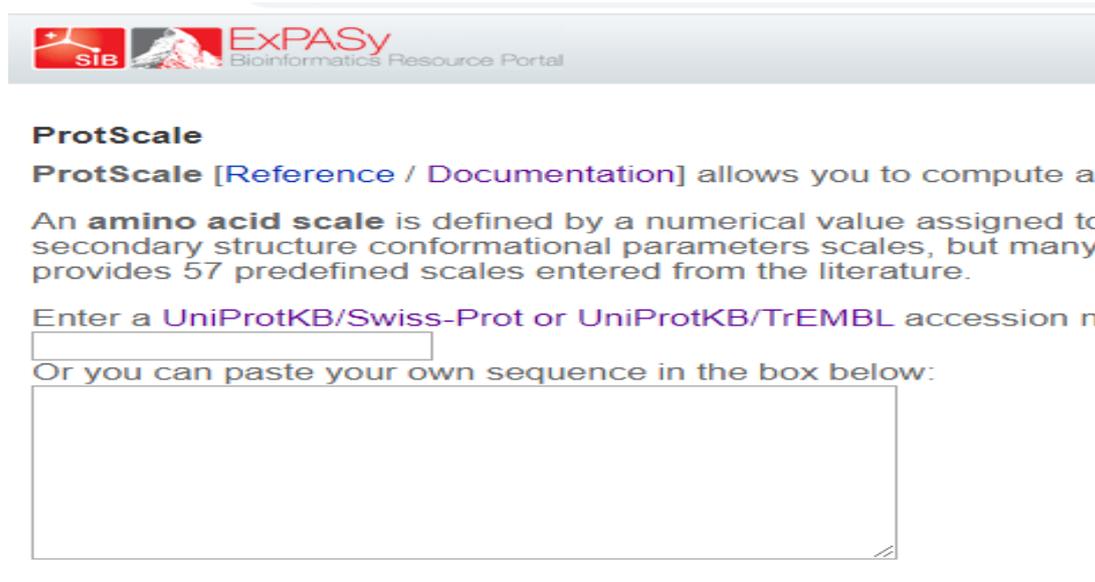


Figure 3.3 : Interface du programme ProtScale.

Pour réaliser ce profil, ProtScale peut utiliser plusieurs échelles disponibles sur le portail de ressources bioinformatiques ExPasy.

Dans ce travail, nous avons utilisé l'échelle de **Kyte et Doolittle (1982)** où à chacun des 20 acides aminés correspond une valeur d'hydropathie (**Tableau 3.2**).

Tableau 3.2 : Échelle de Kyte et Doolittle.

Acide aminé	Acide aminé	Acide aminé	Acide aminé
Ile (I) : +4,5	Met (M) : +1,9	Trp (W) : -0,9	Gln (Q) : -3,5
Val (V) : +4,2	Ala (A) : +1,8	Tyr (Y) : -1,3	Asp (D) : -3,5
Leu (L) : +3,8	Gly (G) : -0,4	Pro (P) : -1,6	Asn (N) : -3,5
Phe (F) : +2,8	Thr (T) : -0,7	His (H) : -3,2	Lys (K) : -3,9
Cys (c) : +2,5	Ser (S) : -0,8	Glu (E) : -3,5	Arg (R) : -4,5

La séquence protéique en question est parcourue avec une fenêtre glissante d'une taille donnée, Le programme utilise une fenêtre de 19 résidus qui permet de distinguer les domaines hydrophobes. Pour chaque position, la valeur moyenne d'hydropathie des acides aminés est calculée et attribuée au résidu central de la fenêtre. Les valeurs positives indiquent le comportement hydrophobe donc le segment traité est supposé être transmembranaire. Les valeurs négatives indiquent le comportement hydrophile donc le segment traité est supposé être non transmembranaire (**Gasteiger et al., 2005**).

2.2 ΔG predictor 1.0

ΔG predictor (<http://dgpred.cbr.su.se/>) est un programme de prédiction qui grâce à l'option « *Full protein scan* » permet de rechercher une hélice α transmembranaire putative au sein d'une séquence protéique. Les travaux expérimentaux de **Hessa et collaborateurs (2005)** ont testé l'insertion de séquences d'acides aminés de 19 résidus dans la membrane de réticulum endoplasmique au moyen du translocon Sec61. Cela a permis d'évaluer quantitativement cette insertion par l'étude de la différence de l'énergie libre apparente ΔG_{app} . Une valeur négative de ΔG_{app} indique que la séquence est reconnue par le translocon comme étant une hélice transmembranaire qui doit être intégrée dans la membrane. Les résultats sont exprimés sous forme d'une représentation graphique bidimensionnelle nommée *diagramme d'hydrophobicité* qui utilise l'échelle de **Hessa et collaborateurs (2005)**. Ce diagramme est composé de plusieurs graphiques superposés montrant les valeurs prévues de ΔG_{app} en fonction de la position centrale de l'hélice. Les graphiques sont de longueurs variables allant de 19 à 23 acides aminés (**Hessa et al., 2007**) (**figure 3.4**).

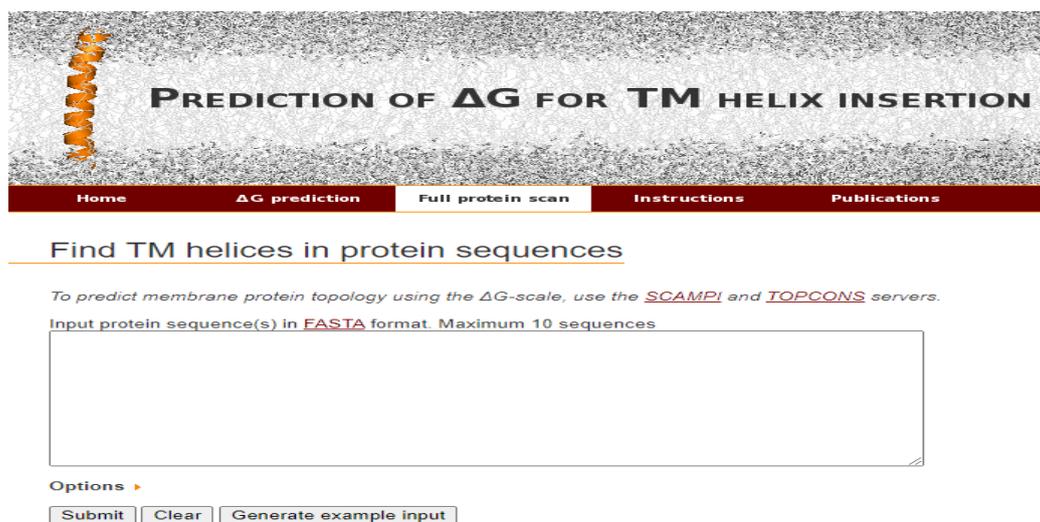


Figure 3.4 : Interface du programme ΔG predictor.

2.3 TOPCONS

TOPCONS (<http://topcons.cbr.su.se/>) est un programme qui prédit la topologie d'une protéine membranaire. Il caractérise les différents domaines transmembranaires avec leurs orientations (extérieur \longrightarrow intérieur et vice versa). Le programme discrimine le segment hydrophobe du peptide signal des segments transmembranaires. Cela a fait de TOPCONS un outil plus puissant et plus que les autres programmes existants. Il implémente cinq méthodes différentes de prédiction. L'une d'elles utilise la séquence de la protéine sous format Fasta

tandis que les autres méthodes ont besoin d'un profil de séquence obtenu à partir de bases de données telles que UniProt comme fichier d'entrée. Ces profils sont utilisés par ces méthodes afin de réaliser la prédiction de la position de chaque résidu de la séquence protéique. Nous avons quatre possibilités de localisation (i pour *inside*, o pour *outside*, m pour *membrane region*, s pour *signal peptide*). TOPCONS combine les résultats de ces méthodes et calcule la topologie consensus finale (Brensel *et al.*, 2009) (figure 3.5).

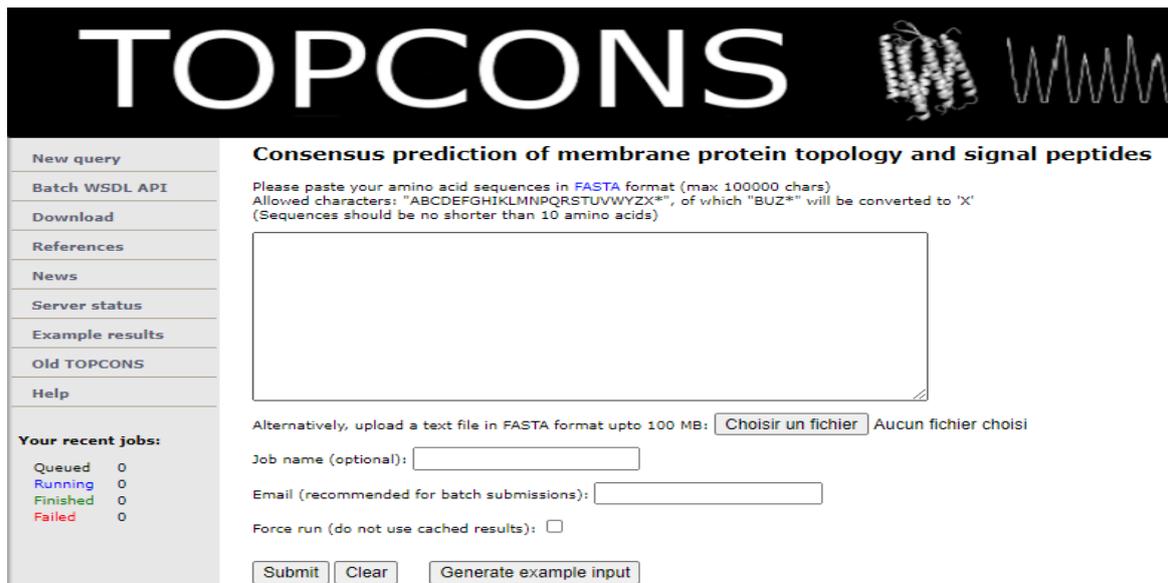


Figure 3.5 : Interface du programme TOPCONS.

3 Étude des modifications post-traductionnelles de la CX23

3.1 Étude de la glycosylation avec NetNGlyc 1.0 et NetOGlyc 4.0

Le programme NetNGlyc (<http://www.cbs.dtu.dk/services/NetNGlyc/>) prédit les sites de N-glycosylation des protéines humaines en examinant le motif N-X-S/T (X acide aminé quelconque sauf la proline) qui est seulement un prérequis pour cette modification. Il utilise des réseaux de neurones artificiels pour discriminer les séquons accepteurs des séquences ne présentant aucun site accepteur (Gupta *et al.*, 2004). Par ailleurs, le programme NetOGlyc (<http://www.cbs.dtu.dk/services/NetOGlyc/>) prédit les sites de la O-glycosylation des protéines des mammifères. Il reconnaît les résidus d'acides aminés S/T susceptibles de lier le N-acétylgalactosamine en utilisant également des réseaux de neurones artificiels (Steentoft *et al.*, 2013).

3.2 Étude de la phosphorylation avec NetPhos 3.1

NetPhos est un outil web (<http://www.cbs.dtu.dk/services/NetPhos/>) qui utilise les réseaux de neurones artificiels pour prédire les sites de phosphorylation qui sont la sérine (S),

la thréonine (T) et la tyrosine (Y) des protéines eucaryotes par des différentes kinases. Pour déterminer la spécificité des sites des substrats à une kinase donnée, il est essentiel de disposer de données importantes concernant les sites de phosphorylation. Six kinases sont sélectionnées de fait qu'elles présentent le plus grand nombre de sites accepteurs connus qui varient entre 22 et 258. Ces kinases sont la PKA (*cyclic AMP dependent protein kinase*), la PKC (*proteine kinase C*), la PKG (*cyclic GMP-dependent proteine kinase*), la cdc2 (*cyclin-dependent kinase*), la CKII (*caseine kinase 2*) et la CaM-II (Ca^{+2} / *calmoduline-dependent protein kinase*) (Blom *et al.*, 1999 ; Blom *et al.*, 2004). Ces informations concernant les sites de phosphorylation validées à partir de données expérimentales sont classées dans la base de données *Phosphobase* (Kreegipuu *et al.*, 1999) (figure 3.6).

Figure 3.6 : Interface du programme NetPhos.

3.3 Étude de la palmitoylation avec CSS-Palm 4.0

CSS-Palm (<http://csspalm.biocuckoo.org/>) a été implémenté en 2006 par Zhou et collaborateurs comme étant le premier programme permettant la prédiction de site de palmitoylation. La dernière version CSS-Palm 4.0 utilise l'intelligence artificielle et réalise les prédictions à haute performance à l'aide des données expérimentales montrant les sites de palmitoylation. Il définit pour chaque résidu cystéine un score final. Celui qui présente le score final le plus élevé est défini comme étant un site potentiel de palmitoylation (Zhou *et al.*, 2006 ; Ren *et al.*, 2008 ; Xue *et al.*, 2011 ; Xue et Ren, 2013) (figure 3.7).

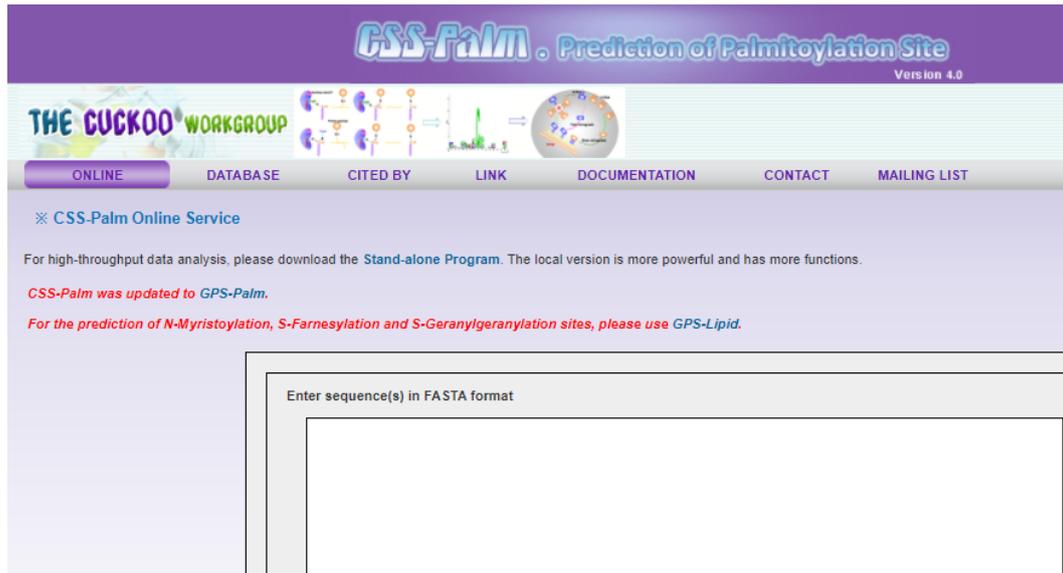


Figure 3.7 : Interface du programme CSS-Palm.

3.4 Étude de la protéolyse avec PROSPERous

PROSPERous (<http://prosperous.erc.monash.edu/>) est un programme qui sert à prédire *in silico* des sites de clivage spécifiques aux protéases à partir des séquences en acides aminés des substrats. Les données des substrats spécifiques utilisées sont extraites de la base de données MEROPS qui contient des informations sur les protéases, leurs substrats et leurs inhibiteurs. Seules les séquences des substrats validées expérimentalement avec leurs sites de clivages sont sélectionnées. Cet outil utilise des modèles de régression logistiques pour réaliser la prédiction. Cela fait de PROSPERous un programme performant pour la prédiction rapide des sites de clivage potentiels pour 90 protéases. (Song *et al.*, 2018) (Figure 3.8).

PROSPERous: an integrative tool for predicting substrate cleavage sites of 90 proteases

PROSPERous is an integrative server for *in silico* prediction of protease-specific substrates and their cleavage sites from amino acid sequences. PROSPERous is primarily based on amino acid weights derived from amino acid occurrences, through a combination of multiple scoring functions to score, predict and rank potential cleavage sites of proteases. Therefore, PROSPERous is a powerful tool for substrate identification in protease systems biology that complements the existing prediction tools.

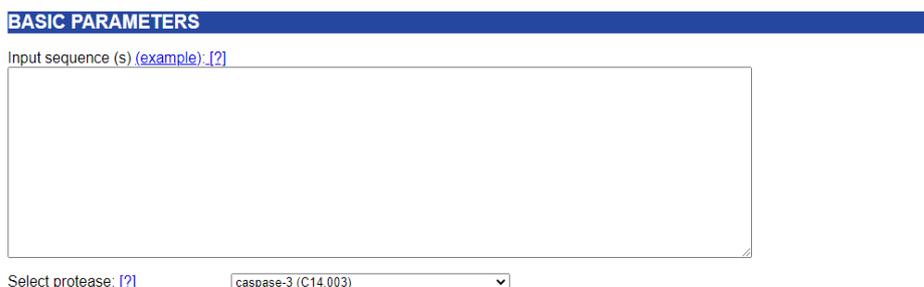


Figure 3.8 : Interface du programme PROSPERous.

4 Réalisation de la topologie membranaire présumée de la CX23 par Protter

Protter (<http://wlab.ethz.ch/protter>) est un outil web qui permet de générer des visualisations personnalisées de la topologie membranaire de la protéine étudiée. De plus, plusieurs caractéristiques peuvent aussi être identifiées comme les modifications post-traductionnelles, la structure secondaire et les motifs (Omasits *et al.*, 2014).

5 Alignement de la CX23 avec les trois hPanx

L'alignement des séquences peptidiques consiste à les mettre les unes sous les autres d'une manière à avoir le maximum de résidus partagés entre les séquences. Pour cela, il existe plusieurs programmes ou algorithmes permettant de réaliser des alignements par paires ou des alignements multiples qui peuvent être locaux ou globaux. Ces algorithmes d'alignement utilisent le système de score. Les identités (résidus identiques) et les similarités (résidus qui possèdent des propriétés physico-chimiques semblables) augmentent le score d'alignement. Par ailleurs, les insertions des espaces dites « gap » ou les suppressions des résidus conduisant à des délétions diminuent ce score (pénalité).

5.1 Alignement des séquences avec T-Coffee

T-Coffee (*Tree-based Consistency Objective Function for Alignment Evaluation*) (Notredame *et al.*, 2000), est un moyen simple pour effectuer des alignements multiples des séquences en utilisant des sources de données hétérogènes. Le programme fournit aux utilisateurs plusieurs méthodes selon certains critères comme les séquences peptidiques, les informations sur les structures 3D (la méthode Expresso), etc. T-Coffee peut aligner un total de 150 séquences dont chacune peut contenir 10 000 résidus et 2500 résidus pour les autres modes (Di Tommaso *et al.*, 2011).

T-Coffee aligne les protéines en utilisant une approche progressive. Il génère d'abord une bibliothèque primaire des alignements par paires des séquences. Cette bibliothèque comporte des alignements globaux en utilisant le programme ClustalW, et des alignements locaux produits par Lalign. A partir de cela, une matrice de distance est produite qui à son tour construit un arbre guide en utilisant la méthode appelée « *Neighbour-joining* ». Il y a également l'établissement d'une bibliothèque étendue qui s'appuie sur le résultat de la bibliothèque primaire. Les scores obtenus dans la bibliothèque étendue guident alors l'alignement progressif final (Notredame *et al.*, 2000) (figure 3.9).

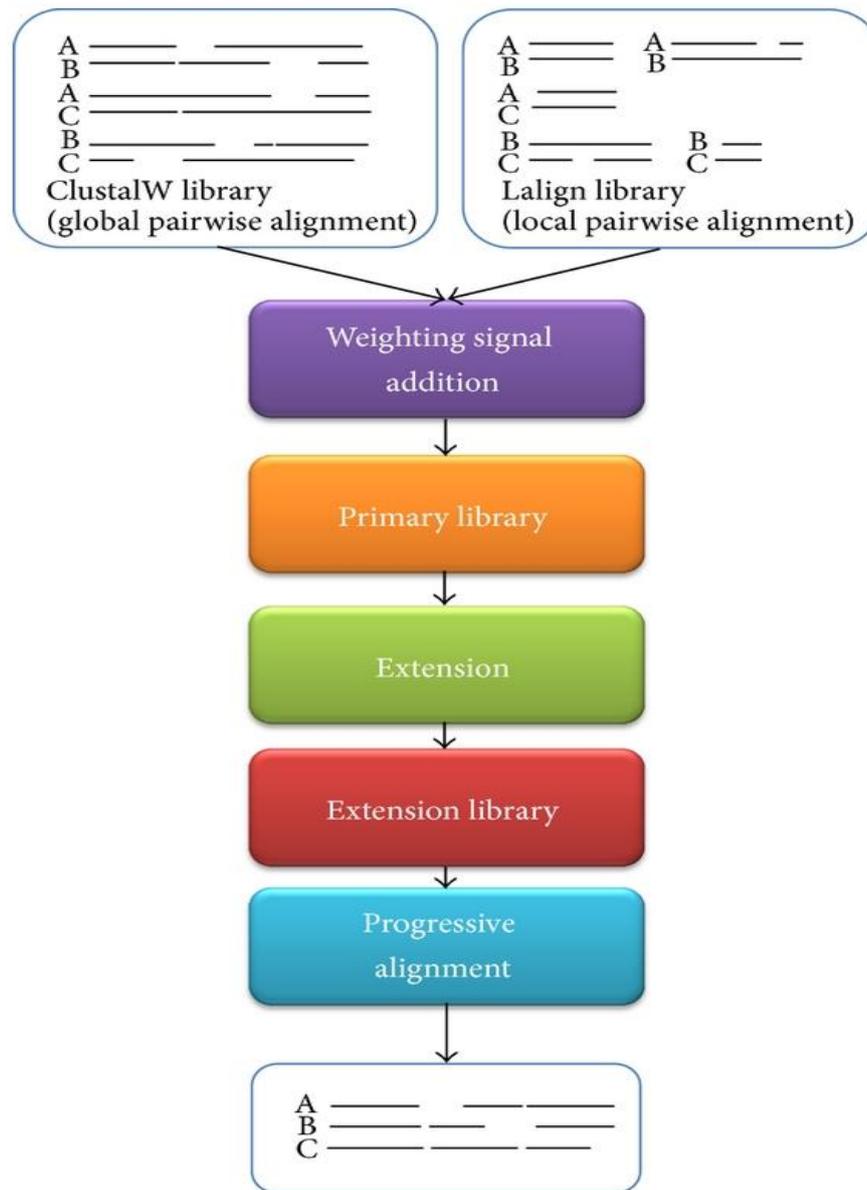


Figure 3.9 : Processus d'alignement multiple par T-Coffee.

Pour réaliser notre alignement, nous avons accédé au programme T-Coffee via son site web (<https://www.ebi.ac.uk/Tools/msa/tcoffee/>) (figure 3.10). Puis nous avons introduit les séquences comme fichier d'entrée « *input* ». Les séquences introduites correspondent à celles de la CX23 humaine (205 acides aminés) et des trois pannexines humaines, *hPanx1* (426 acides aminés), *hPanx2* (677 acides aminés) et *hPanx3* (392 acides aminés) sous format Fasta à partir de la base de données UniProtKB/Swiss-Prot.

Figure 3.10 : Interface du programme T-Coffee.

Le processus d'alignement est lancé en cliquant sur « *submit* », et les séquences alignées sont affichées par défaut sous format ClustalW « *output* ». La dernière étape a consisté à éditer l'alignement obtenu. Pour cela nous avons utilisé le programme Jalview en cliquant sur « *Results Viewers* » (figure 3.11).

Figure 3.11 : Résultat de la requête d'alignement fourni par T-Coffee.

5.2 Édition de l'alignement avec Jalview version 2.10.4b1

Jalview (Waterhouse *et al.*, 2009), est un programme libre et gratuit développé en langage informatique Java. Il sert à l'édition interactive, l'analyse, et la visualisation des alignements de séquences multiples. Il peut fonctionner avec des annotations des séquences, des informations sur la structure secondaire, des arbres phylogénétiques, et des structures moléculaires en 3D.

Dans notre travail, nous avons utilisé ce programme comme éditeur d'alignement, notamment pour colorer les résidus des acides aminés. Pour cela, Jalview fournit une multitude de choix de couleurs classées dans le menu « *colour* » (figure 3.12).

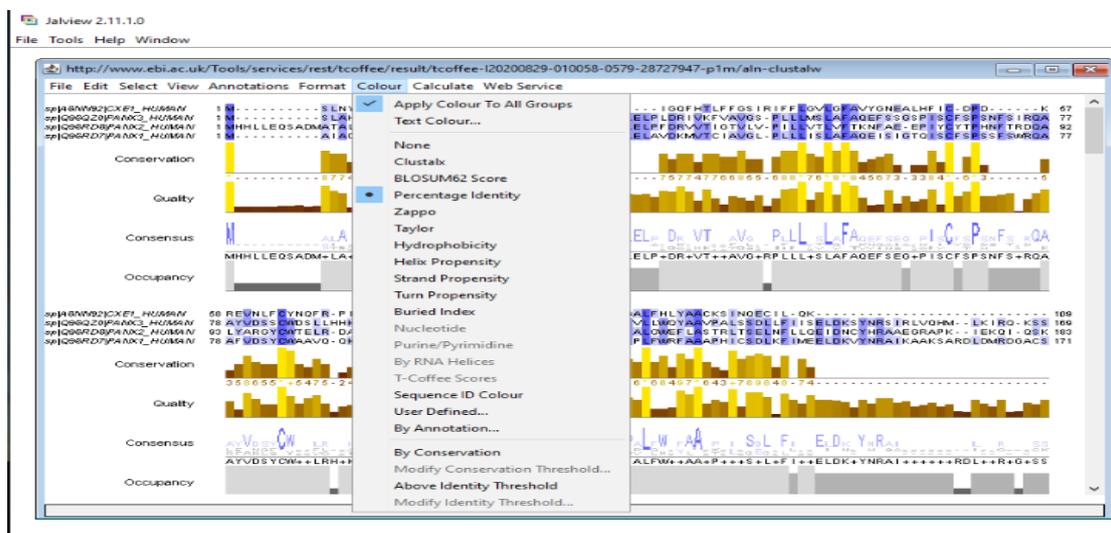


Figure 3.12 : Programme Jalview avec les différentes options de couleurs.

Nous avons choisi l'option « *Percentage Identity* » dans laquelle les résidus sont colorés selon leurs pourcentages d'identité dans chaque colonne, qui est en accord avec la séquence logo consensus représentée par les différentes hauteurs des résidus en fonction de leurs fréquences par colonne. Seuls les résidus qui sont en accord avec le résidu consensus sont colorés (figure 3.13).

Percentage agreement	Colour
> 80 %	Mid blue
> 60 %	Light blue
> 40 %	Light grey
<= 40%	White

Figure 3.13 : Couleurs selon le pourcentage d'identité « PID ».

De plus, le programme permet d'analyser la conservation de chaque groupe de séquence en se basant sur la méthode de **Livingstone et Barton (1993)**, prenant en considération les propriétés physico-chimiques des résidus. Le calcul de la conservation sera représenté sous l'alignement sous forme d'histogrammes.

Chapitre IV :
Résultats & Interprétation

1 La CX23 possède quatre domaines transmembranaires

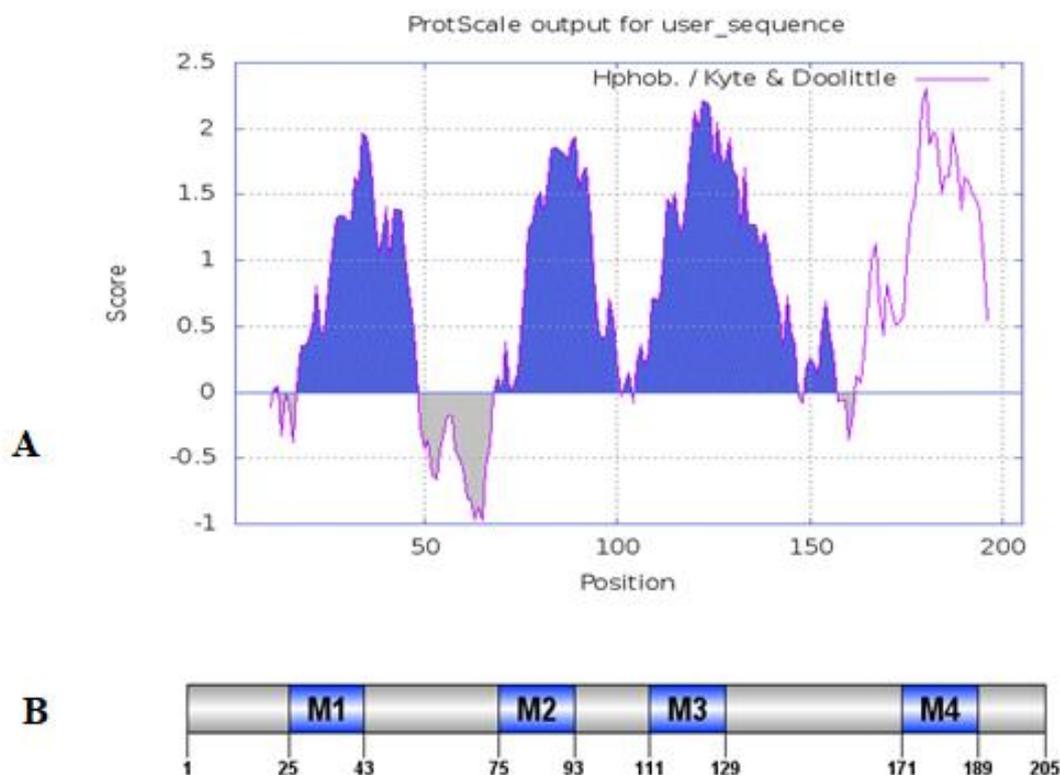
1.1 Analyse de la CX23 par ProtScale

Le programme ProtScale a permis d'obtenir les résultats suivants sous forme de diagramme appelé « *profil d'hydropathie* ».

Au niveau de ce diagramme, l'axe des ordonnées représente l'indice d'hydropathie, et celui des abscisses montre les différentes positions des acides aminés de la séquence protéique analysée.

L'analyse du profil d'hydropathie de la CX23 indique la présence de quatre régions hydrophobes correspondant à quatre séquences notées : 25-43 ; 75-93 ; 111-129 ; 171-189. De plus, ce profil montre la présence de cinq régions hydrophiles : 1-24 ; 44-74 ; 94-110 ; 130-170 ; 190-205 (**figure 4.1 A**).

La représentation des différents domaines hydrophobes de la CX23 (**figure 4.1 B**) est faite à l'aide du logiciel DOG2.0 (**Ren *et al.*, 2009**), servant à la visualisation des structures en domaine des protéines.



A: Profil d'hydropathie de la CX23. **B:** Représentation en domaine de la CX23.

Figure 4.1 : Profil d'hydropathie de la CX23 avec la représentation en domaine.

1.2 Analyse de la CX23 par ΔG predictor :

En plus du programme ProtScale utilisé pour la caractérisation des régions transmembranaires, nous avons utilisé un deuxième programme qui prédit les hélices transmembranaires d'une séquence protéique appelé ΔG predictor.

Les résultats sont obtenus sous forme d'un diagramme où l'axe des ordonnées représente ΔG_{app} pour la différence d'énergie libre apparente correspondante, et l'axe des abscisses représente la position centrale de l'hélice.

L'analyse du diagramme d'hydrophobicité de la CX23 montre également la présence de quatre domaines hydrophobes (**figure 4.2**). Les couleurs présentes sur la figure montrent les différentes tailles d'hélice utilisées (19-23).

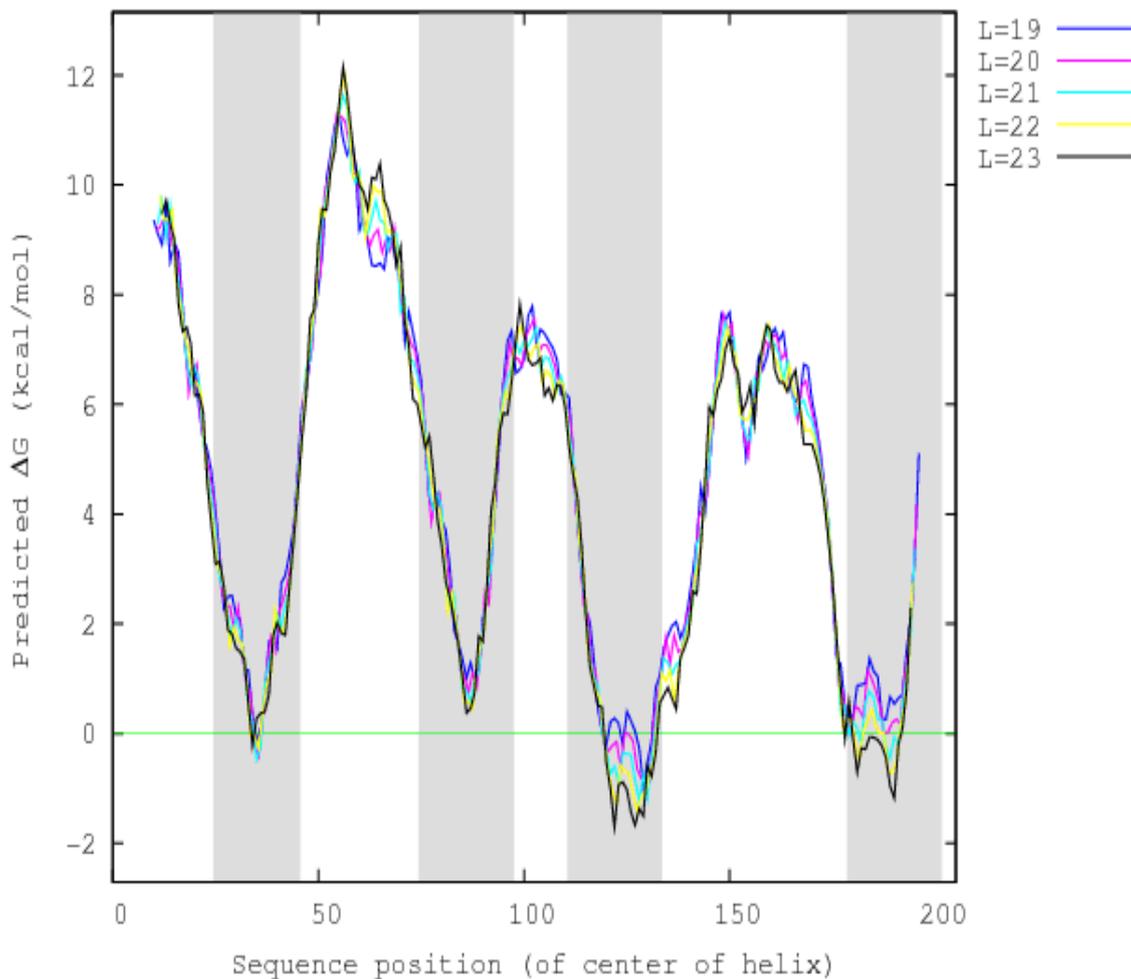


Figure 4.2 : Diagramme d'hydrophobicité de la CX23.

1.3 Analyse de la CX23 par TOPCONS

TOPCONS est le troisième programme utilisé dans ce travail pour la prédiction de la topologie de la CX23. Ce programme implémente cinq algorithmes différents. La prédiction est présentée sous forme de schéma qui montre les domaines de la CX23 en précisant leur localisation (intérieur/extérieur) et leur orientation (intérieur → extérieur et vice versa).

L'analyse de la représentation schématique de la CX23 montre la présence de quatre régions transmembranaires hydrophobes avec une indication sur les extrémités N- et C-terminales qui sont intracytoplasmiques (**figure 4.3**).

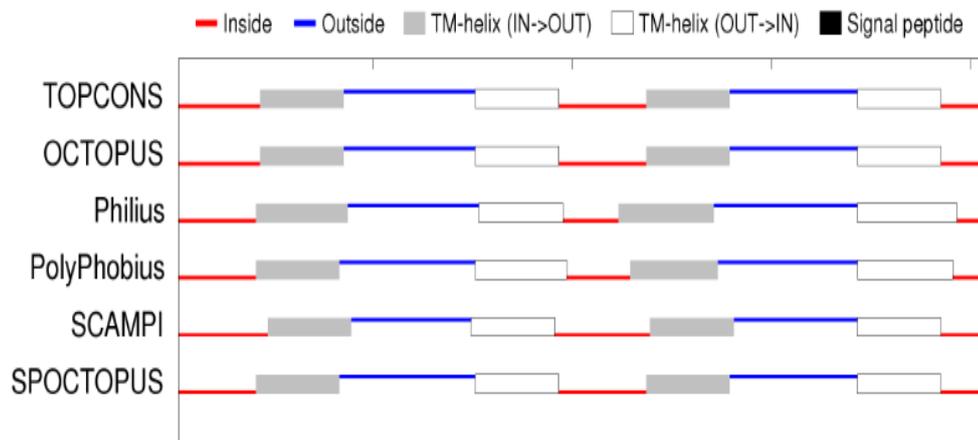


Figure 4.3 : Topologie membranaire de la CX23 prédite par TOPCONS.

Inside désigne intérieur ; *outside* désigne extérieur ; *TM-helix* (IN → OUT) désigne hélice α transmembranaire disposée de l'intérieur vers l'extérieur ; *TM-helix* (OUT → IN) désigne hélice α transmembranaire disposée de l'extérieur vers l'intérieur ; *signal peptide* désigne peptide signal. OCTOPUS, Philius, PolyPhobius, SCAMPI et SPOCTOPUS représentent les cinq programmes qu'utilise TOPCONS.

1.4 Prédiction de la topologie de la CX23 par les trois méthodes

Les prédictions de la CX23 à l'aide des trois programmes ProtScale, ΔG predictor et TOPCONS sont données dans le **tableau 4.1**. Ces derniers ont tous montré que la CX23 possède quatre domaines transmembranaires hydrophobes en hélice α notés : M1, M2, M3, M4.

Après l'analyse des résultats des trois programmes, nous avons choisi les séquences des domaines transmembranaires de la CX23 prédites par TOPCONS correspondent aux segments : Q22-A42 ; S76-A96 ; S119-G139 ; F172-F192. Ces résultats nous ont permis de représenter les différents domaines transmembranaires à l'aide du logiciel de visualisation des structures en domaines des protéines DOG2.0 (**Ren et al., 2009**) (**figure 4.4**).

Tableau 4.1 : La prédiction des domaines transmembranaires de la CX23 par les trois programmes.

		M1	M2	M3	M4
CX23	<i>ProtScale</i>	25-43	75-93	111-129	171-189
	<i>ΔG predictor</i>	25-45	75-97	111-133	179-201
	<i>OCTOPUS</i>	22-42	76-96	119-139	172-192
	<i>Philius</i>	21-43	77-97	112-135	172-196
	<i>PolyPhobius</i>	21-41	76-98	115-136	172-195
	<i>SCAMPI</i>	24-44	75-95	120-140	172-192
	<i>SPOCTOPUS</i>	21-41	76-96	119-139	172-192
	<i>TOPCONS</i>	22-42	76-96	119-139	172-192

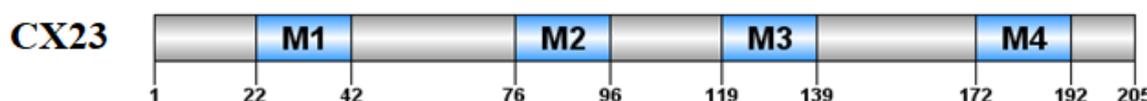


Figure 4.4 : Représentation en domaine de la CX23.

2 Sites des modifications post-traductionnelles de la CX23

Il faut noter qu'aucune étude n'a été faite concernant la prédiction des modifications post-traductionnelles de la CX23.

2.1 La CX23 ne présente aucun site de glycosylation

L'analyse de la CX23 par les programmes NetNGlyc et NetOGlyc de la *N*-glycosylation et la *O*-glycosylation respectivement n'a montré aucun site de glycosylation. Ce résultat est en accord avec les études déjà établies sur la glycosylation des autres isoformes des connexines (**Martin et Evans, 2004 ; Saez et al., 2003**). Cependant, l'étude de **Rahman et collaborateurs (1993)** a montré que la CX32 est *N*-glycosylée.

2.2 La CX23 est phosphorylée

L'analyse de la CX23 par le programme NetPhos a prédit certains sites de phosphorylation. En se basant sur l'alignement des 21 isoformes (**figure 1.3**) nous avons choisi parmi les sites prédits par le programme deux sites qui sont les plus conservés chez les connexines et notamment les connexines 43, 32 et 26 qui sont les plus étudiées et les plus ubiquitaires.

Le premier site prédit, le résidu Y65 (tyrosine 65) de la première boucle extracellulaire (boucle E1), appartient au motif conservé $\underline{V}CY^*D$ conservé dans la majorité des connexines. La phosphorylation sur des résidus tyrosine dans les domaines extracellulaires n'est bien élucidée. Cependant, **Bordoli et collaborateurs (2014)** ont identifié une protéine VLK « *vertebrate lonesome kinase* » comme étant la première protéine tyrosine kinase extracellulaire qui phosphoryle les protéines dans la voie sécrétoire et les domaines extracellulaires des protéines. Cela nécessite plus d'études pour confirmer si certainement le résidu Y65 de la CX23 est phosphorylé par cet enzyme. Le motif $\underline{V}CY^*D$ présente le résidu C64 conservé dans toutes les connexines qui sert avec les autres cystéines des deux boucles extracellulaires à les lier et les maintenir en une conformation compacte et rigide pour qu'elles puissent interagir avec les boucles des connexines voisines en formant un canal jonctionnel.

Le deuxième site prédit, le résidu T72 (thréonine 72) de la première boucle extracellulaire (boucle E1) qui appartient au motif conservé $P-[IVL]-[ST^*]$ dont le résidu thréonine de la CX23 est remplacé par un résidu sérine chez les autres isoformes. Plusieurs protéines sérine/thréonine kinases extracellulaire appelées « *ecto-kinases* » ont été identifiées comme la PKA, PKC et CKII (**Yalak et al., 2014**). Comme il a été reporté que la CX26 est

phosphorylée sur des résidus T177, S183 et T186 de la deuxième boucle extracellulaire (boucle E2) par des ectokinases (**D'hondt et al., 2013**).

2.3 La CX23 est palmitoylée

La recherche des sites putatifs de palmitoylation par l'utilisation du programme CSS-Palm a montré que la CX23 est palmitoylée sur le résidu C13 (cystéine 13) qui se trouve au milieu du domaine N-terminal. Par ailleurs, il a été constaté que la CX32 présente également deux sites de palmitoylation qui sont les résidus C277 et C283 de la queue C-terminale de cette connexine (**Locke et al., 2006**).

2.4 La CX23 présente des sites de clivage par les caspases

Afin de prédire les sites de clivage de la connexine23 humaine nous avons opté pour le programme PROSPEROUS qui offre une grande spécificité et une grande fiabilité. Nous avons choisi les deux meilleurs résultats en utilisant la caspase-3.

De nombreux travaux ont identifié le site de clivage des protéines DXXD↓ par la caspase-3 où D est l'acide aspartique, X acide aminé quelconque et ↓ désigne le site de clivage (**Talaniane et al., 2011**).

La protéolyse des connexines est associée surtout à la CX46 et la CX50 exprimées dans les fibres du cristallin comme la connexine23 (**D'hondt et al., 2013**). **Yin et collaborateurs (2001)** ont montré que la Cx45.6 (poule) qui est un orthologue de la CX50, exprimée également dans les fibres du cristallin, constitue un substrat direct pour la caspase-3 et elle est clivée entre les résidus G361 et E367 du domaine C-terminal.

Pour la CX23 le programme a prédit deux sites de clivage. Le premier site est le résidu D56 (acide aspartique 56) appartenant à la séquence ⁵⁴DPD↓k⁵⁷ qui se trouve dans la première boucle extracellulaire. Le second site prédit est le résidu D149 (acide aspartique 149) situé dans la séquence ¹⁴⁸CD↓AR¹⁵¹ qui se trouve dans la deuxième boucle extracellulaire.

D'autres études devront être envisagées pour confirmer ces sites de clivage car les séquences prédites ne correspondent pas parfaitement au motif DXXD↓ connu.

3 Topologie membranaire présumée de la CX23

Le programme Protter nous a permis de tracer la topologie tout en combinant les résultats obtenus des différentes analyses effectuées concernant la prédiction des domaines transmembranaires et les modifications post-traductionnelles de la CX23.

La topologie de la CX23 montre la présence de quatre domaines transmembranaires, deux boucles extracytoplasmiques, une boucle intracytoplasmique et des extrémités N- et C-terminales intracytoplasmiques. Les deux boucles extracellulaires sont caractérisées par la présence de deux résidus cystéines conservés dans chacune. Le domaine C-terminal est très court (13 acides aminés) comparé au domaine C-terminal des autres connexines tel que celui de la CX43 (155 acides aminés) et celui de la CX32 (75 acides aminés) (Willecke *et al.*, 2002). Par ailleurs, les sites des différentes modifications post-traductionnelles sont aussi indiqués, à savoir la phosphorylation (Y65, T72), la palmitoylation (C13) et le clivage par les caspases (D56, D149) (figure 4.5).

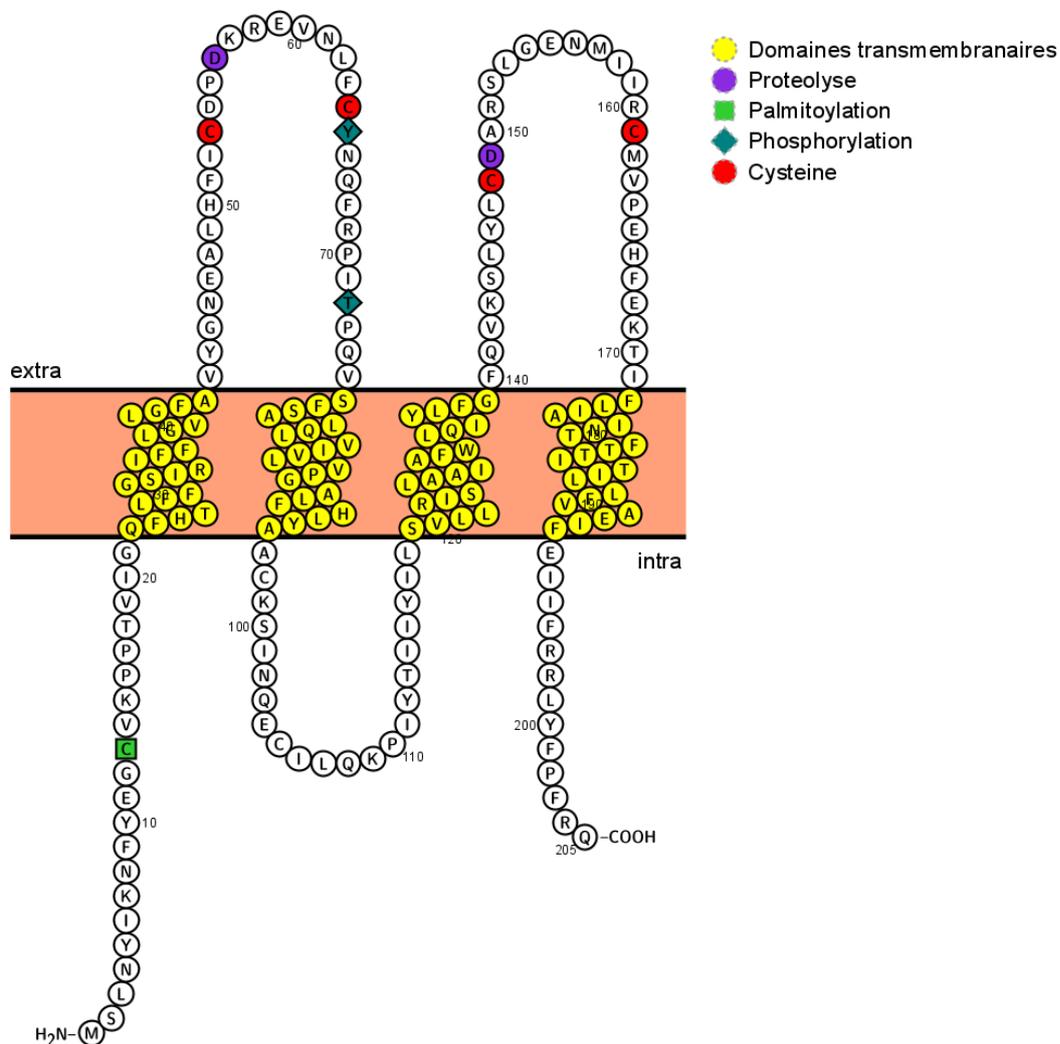


Figure 4.5 : Topologie membranaire présumée de la CX23 représentée par le programme Protter.

4 Alignement de la CX23 avec les trois *hPanx*

A l'aide du programme T-Coffee (Notredame *et al.*, 2000) qui sert à l'alignement multiple des protéines, nous avons analysé la séquence de la CX23 avec les séquences des trois *hPanx* afin de rechercher les éléments qui définissent la présence de l'homologie entre ces protéines.

L'alignement effectué par T-Coffee a été édité par le programme *Jalview* version 2.10.4b1 (Waterhouse *et al.*, 2009). Les acides aminés identiques sont ombrés en bleu foncé et les acides aminés similaires sont ombrés en bleu pâle.

Le programme T-Coffee a permis d'obtenir une matrice qui contient les scores des alignements par paires des quatre protéines. Pour chaque paire de séquences est donné un pourcentage d'identité. La diagonale présentant les 100% sépare la matrice en deux triangles symétriques (figure 4.6).

Afin de pouvoir comparer les pourcentages d'identité de la matrice en fonction des paires des séquences de la CX23 avec chaque *hPanx*, nous avons utilisé la couleur rouge pour les distinguer. À partir de cela, nous avons constaté que l'alignement de la CX23 et la *hPanx2* présente le pourcentage d'identité le plus élevé avec 17,24%.

	1	2	3	4
1.CX23	100.00	13.79	17.24	16.75
2.PANX3	13.79	100.00	28.46	43.09
3.PANX2	17.24	28.46	100.00	26.01
4.PANX1	16.75	43.09	26.01	100.00

La diagonale de la matrice

Figure 4.6 : Matrice de pourcentage d'identité de la CX23 et les trois *hPanx*.

En examinant la figure 4.7, nous constatons que la CX23 présente une homologie de séquence avec les *hPanx* dans le domaine Nt, les boucles extracellulaires et les domaines transmembranaires. Cependant, la boucle intracytoplasmique et le domaine Ct ne montrent aucune similarité et présentent des séquences et des longueurs très diversifiées. Cela est

traduit par le grand nombre des insertions « gap » introduites de fait que la CX23 est très courte par rapport aux pannexines.

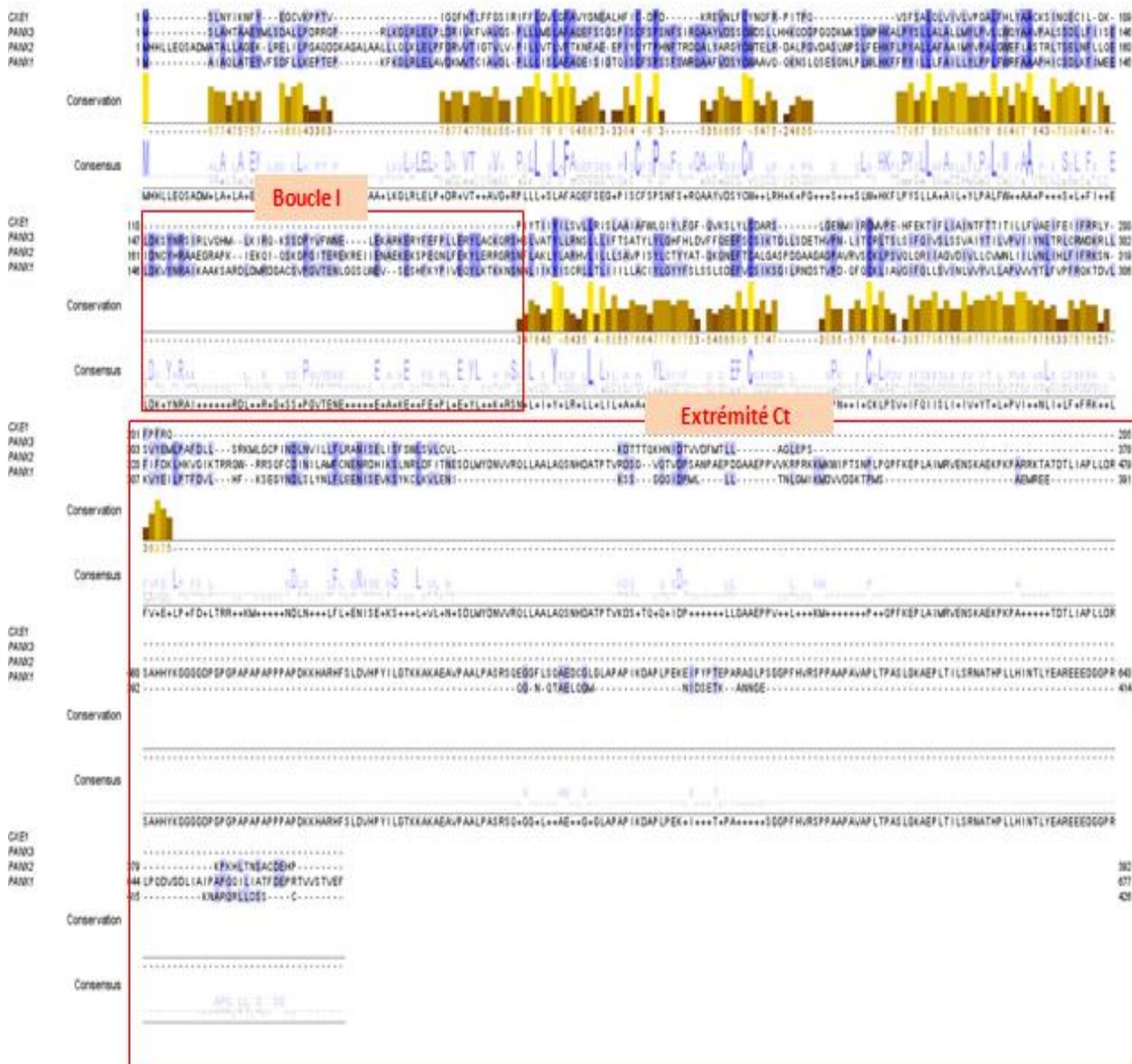
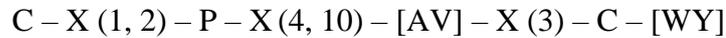


Figure 4.7 : Alignement multiple de la CX23 et les trois hPanx.

4.1 Alignement de la CX23 et les hPanx montrant les domaines Nt, M1, la boucle E1 et M2

L'analyse de l'alignement au niveau du domaine Nt, M1, la boucle E1 et le domaine M2 (figure 4.8), montre que ces régions présentent certains acides aminés conservés. Les acides aminés les plus conservés sont T, L, F et A dans le domaine M1 et L, A et Y dans le domaine M2 comme on peut l'observer dans l'histogramme et les séquences logo. Ces acides aminés sont pour la plupart de nature hydrophobe.

La boucle E1 présente elle aussi quelques conservations, elle contient deux cystéines très conservées avec la présence de quelques insertions au niveau de la CX23 indiquant que la boucle E1 est plus courte que celle des *hPanx*. Les conservations ont permis de définir une séquence consensus au niveau de cette boucle qui correspond à la séquence suivante :

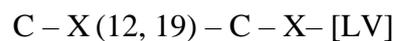


4.2 Alignement de la CX23 et les *hPanx* montrant la boucle I, les domaines M3, la boucle E2 et M4

L'analyse de la boucle intracytoplasmique de la CX23 avec les trois *hPanx* ne présente aucune similarité comme le montre l'histogramme des conservations donc les séquences de cette boucle sont très différentes (**figure 4.9**).

En examinant l'alignement au niveau du domaine M3, la boucle E2 et le domaine M4 (**figure 4.9**), nous constatons que ces régions sont moins conservées. Les acides aminés les plus conservés dans le domaine M3 sont L et Y. Le domaine M4 ne contient aucun acide aminé conservé.

La boucle E2 présente moins de conservation par rapport à la boucle E1. Cependant les deux résidus cystéines sont très conservés avec la présence des insertions au niveau de cette région de la CX23 qui indiquent la différence de la longueur de sa boucle par rapport à celles des *hPanx*. Ces conservations permettent de définir une séquence consensus au niveau de cette boucle qui correspond à la séquence suivante :



Les deux cystéines hautement conservées des boucles extracellulaires E1 et E2 semblent être impliquées dans la formation de ponts disulfures intramoléculaires qui stabilisent ces domaines sous forme de tonneau β (**Foote et al., 1998**).

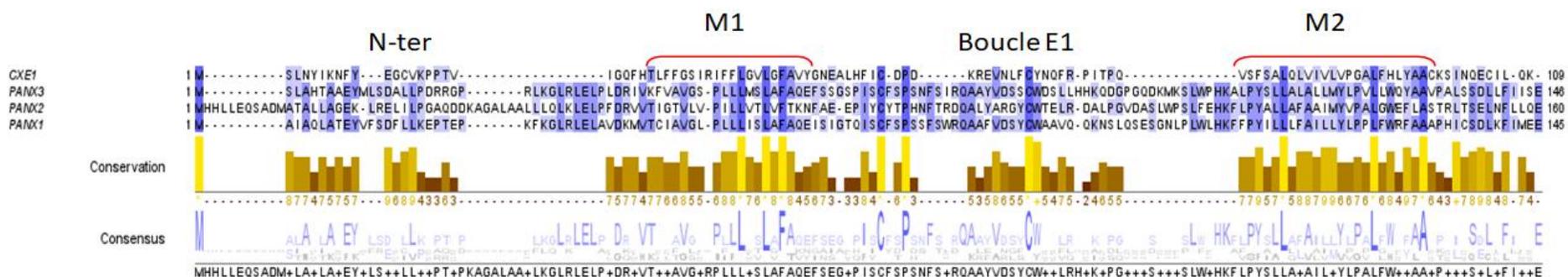


Figure 4.8 : Alignement multiple de la CX23 avec les hPanx montrant les domaines Nt, M1, boucle E1, et le domaine M2.

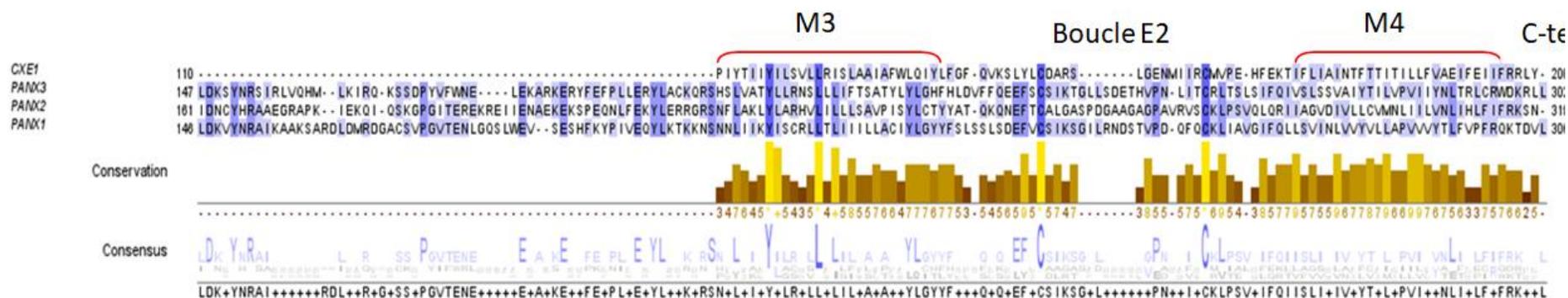


Figure 4.9 : Alignement multiple de la CX23 avec les hPanx montrant la boucle I, les domaines M3, boucle E2, et le domaine M4.

Conclusion

Les protéines membranaires sont très importantes de par les fonctions qu'elles assurent. Elles sont impliquées dans le transport des molécules à travers la membrane plasmique et la signalisation cellulaire. Dans le génome humain 21 isoformes de connexines ont été identifiées. Cependant, il n'existe que trois isoformes de pannexines. La CX23 est une isoforme particulière qui appartient à la famille des connexines en partageant des caractéristiques avec les pannexines telles que le nombre des résidus cystéines dans les boucles extracellulaires et la formation des hémicanaux.

Dans ce travail, nous avons fait une analyse structurale et fonctionnelle de la CX23 au moyen d'outils informatiques. En premier temps, nous avons caractérisé les différents domaines de la CX23 en mettant l'accent sur les résidus essentiels par la prédiction des modifications post-traductionnelles. Puis, nous avons utilisé les séquences de la CX23 et les trois *hPanx* afin de montrer les régions conservées entre ces protéines.

L'étude de la topologie avec les programmes ProtScale, ΔG predictor et TOPCONS a montré que la CX23 est une protéine membranaire à quatre passages transmembranaires exposant deux boucles extracellulaires et une boucle intracytoplasmique avec des extrémités N- et C-terminales intracytoplasmiques dont l'extrémité C-terminale est très courte.

Les résultats de l'étude des modifications post-traductionnelles ont montré que la CX23 contient des sites de phosphorylation, palmitoylation et protéolyse. Cependant, elle semble ne contenir aucun site de glycosylation.

L'alignement multiple des séquences de la CX23 avec les trois *hPanx* réalisé par le programme T-Coffee a révélé que les deux résidus cystéines dans chaque boucle extracellulaire de ces protéines sont très conservés.

En perspective, ces résultats ouvrent la voie à de nouvelles recherches concernant la connexine23. Pour cela, nous suggérons de reprendre la prédiction des différents sites des modifications post-traductionnelles avec d'autres programmes. Nous pourrions aussi prédire la structure secondaire et tertiaire de la CX23 qui n'a pas encore été élucidée.

Références bibliographiques

A

Akimov, V., Barrio-Hernandez, I., Hansen, S. V., Hallenborg, P., Pedersen, A. K., Bekker-Jensen, D. B., ... & Kratchmarova, I. (2018). UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites. *Nature structural & molecular biology*, 25(7), 631-640.

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., & Keith Roberts, P. W. (2018). *Molecular biology of the cell*.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), 3389-3402.

Ambrosi, C., Gassmann, O., Pranskevich, J. N., Boassa, D., Smock, A., Wang, J., ... & Sosinsky, G. E. (2010). Pannexin1 and Pannexin2 channels show quaternary similarities to connexons and different oligomerization numbers from each other. *Journal of Biological Chemistry*, 285(32), 24420-24431.

Arai, M., Mitsuke, H., Ikeda, M., Xia, J. X., Kikuchi, T., Satake, M., & Shimizu, T. (2004). ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability. *Nucleic acids research*, 32(suppl_2), W390-W393.

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., ... & Notredame, C. (2006). Espresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic acids research*, 34(suppl_2), W604-W608.

B

Bairoch, A., Bucher, P., & Hofmann, K. (1997). The PROSITE database, its status in 1997. *Nucleic Acids Research*, 25(1), 217-221.

Bao, L., Locovei, S., & Dahl, G. (2004). Pannexin membrane channels are mechanosensitive conduits for ATP. *FEBS letters*, 572(1-3), 65-68.

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

- Begandt, D., Good, M. E., Keller, A. S., DeLalio, L. J., Rowley, C., Isakson, B. E., & Figueroa, X. F. (2017). Pannexin channel and connexin hemichannel expression in vascular function and inflammation. *BMC Cell Biology*, 18(1), 1-16.
- Benayad, A. (2018). Caractérisation des domaines des connexines par des méthodes bioinformatiques (Mémoire de master). Université de Tlemcen.
- 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.
- 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., Minogue, P. J., Snabb, J. I., Dzhashiashvili, Y., Novak, L. A., Zoltoski, R. K., ... & Beyer, E. C. (2016). Connexin23 deletion does not affect lens transparency. *Experimental eye research*, 146, 283-288.
- Beyer, E. C., & Berthoud, V. M. (2009). The family of connexin genes. In *Connexins* (pp. 3-26). Humana Press.
- Biswas, A. K., Noman, N., & Sikder, A. R. (2010). Machine learning approach to predict protein phosphorylation sites by incorporating evolutionary information. *BMC bioinformatics*, 11(1), 273.
- Blanc, M., David, F., Abrami, L., Migliozzi, D., Armand, F., Bürgi, J., & van der Goot, F. G. (2015). SwissPalm: protein palmitoylation database. *F1000Research*, 4.
- 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.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., ... & Pilbout, S. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic acids research*, 31(1), 365-370.

Bordoli, M. R., Yum, J., Breitkopf, S. B., Thon, J. N., Italiano Jr, J. E., Xiao, J., ... & Keller, T. L. (2014). A secreted tyrosine kinase acts in the extracellular environment. *Cell*, 158(5), 1033-1044.

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.

C

Chauhan, J. S., Bhat, A. H., Raghava, G. P., & Rao, A. (2012). GlycoPP: a webserver for prediction of N-and O-glycosites in prokaryotic protein sequences. *PloS one*, 7(7), e40155.

Chen, H., Xue, Y., Huang, N., Yao, X., & Sun, Z. (2006). MeMo: a web tool for prediction of protein methylation modifications. *Nucleic acids research*, 34(suppl_2), W249-W253.

Chen, X., Qiu, J. D., Shi, S. P., Suo, S. B., Huang, S. Y., & Liang, R. P. (2013). Incorporating key position and amino acid residue features to identify general and species-specific Ubiquitin conjugation sites. *Bioinformatics*, 29(13), 1614-1622.

Chen, Y. Z., Chen, Z., Gong, Y. A., & Ying, G. (2012). SUMOhydro: a novel method for the prediction of sumoylation sites based on hydrophobic properties. *PloS one*, 7(6), e39195.

D

Dayhoff, M. O., Schwartz, R. M., & Orcutt, B. C. (1965). National Biomedical Research Foundation. Atlas of protein sequence and structure, 1.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobittg, M., Montanyola, A., Chang, J. M., ... & Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic acids research*, 39(suppl_2), W13-W17.

D'hondt, C., Iyyathurai, J., Vinken, M., Rogiers, V., Leybaert, L., Himpens, B., & Bultynck, G. (2013). Regulation of connexin-and pannexin-based channels by post-translational modifications. *Biology of the Cell*, 105(9), 373-398.

Do, C. B., Mahabhashyam, M. S., Brudno, M., & Batzoglou, S. (2005). ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome research*, 15(2), 330-340.

Dobson, L., Reményi, I., & Tusnády, G. E. (2015). CCTOP: a Consensus Constrained TOPology prediction web server. *Nucleic acids research*, 43(W1), W408-W412.

Dou, Y., Yao, B., & Zhang, C. (2014). PhosphoSVM: prediction of phosphorylation sites by integrating various protein sequence attributes with a support vector machine. *Amino acids*, 46(6), 1459-1469.

Dykes, I. M., Freeman, F. M., Bacon, J. P., & Davies, J. A. (2004). Molecular basis of gap junctional communication in the CNS of the leech *Hirudo medicinalis*. *Journal of Neuroscience*, 24(4), 886-894.

E

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5), 1792-1797.

Eiberger, J., Degen, J., Romualdi, A., Deutsch, U., Willecke, K., & Söhl, G. (2001). Connexin genes in the mouse and human genome. *Cell communication & adhesion*, 8(4-6), 163-165.

Eisenberg, D., Schwarz, E., Komaromy, M., & Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *Journal of molecular biology*, 179(1), 125-142.

Elofsson, A., & Heijne, G. V. (2007). Membrane protein structure: prediction versus reality. *Annu. Rev. Biochem.*, 76, 125-140.

Engelman, D. M., Steitz, T. A., & Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annual review of biophysics and biophysical chemistry*, 15(1), 321-353.

F

Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M., & Berglund, L. (2010). Prediction of the human membrane proteome. *Proteomics*, 10(6), 1141-1149.

Floden, E. W., Tommaso, P. D., Chatzou, M., Magis, C., Notredame, C., & Chang, J. M. (2016). PSI/TM-Coffee: a web server for fast and accurate multiple sequence alignments of regular and transmembrane proteins using homology extension on reduced databases. *Nucleic acids research*, 44(W1), W339-W343.

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.

Frith, M. C., Saunders, N. F., Kobe, B., & Bailey, T. L. (2008). Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol*, 4(5), e1000071.

G

Gao, J., Thelen, J. J., Dunker, A. K., & Xu, D. (2010). Musite, a tool for global prediction of general and kinase-specific phosphorylation sites. *Molecular & Cellular Proteomics*, 9(12), 2586-2600.

Garay-Malpartida, H. M., Occhiucci, J. M., Alves, J., & Belizário, J. E. (2005). CaSPredictor: a new computer-based tool for caspase substrate prediction. *Bioinformatics*, 21(suppl_1), i169-i176.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607). Humana press.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* (pp. 571-607). Humana press.

Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2019). A brief history of bioinformatics. *Briefings in bioinformatics*, 20(6), 1981-1996.

Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2019). A brief history of bioinformatics. *Briefings in bioinformatics*, 20(6), 1981-1996.

Ghosh, D. (1992). TFD: the transcription factors database. *Nucleic acids research*, 20(Suppl), 2091.

Goodenough, D. A., Paul, D. L., & Jesaitis, L. (1988). Topological distribution of two connexin32 antigenic sites in intact and split rodent hepatocyte gap junctions. *The Journal of Cell Biology*, 107(5), 1817-1824.

Gotoh, O. (1993). Optimal alignment between groups of sequences and its application to multiple sequence alignment. *Bioinformatics*, 9(3), 361-370.

Gupta, R., & Brunak, S. (2001, December). Prediction of glycosylation across the human proteome and the correlation to protein function. In *Pac Symp Biocomput* (Vol. 7, pp. 310-22).

Gupta, R., Jung, E., & Brunak, S. (2004). Prediction of N-glycosylation sites in human proteins. *Protein Science*, 13(2), 203-206.

Gustincich, S., Batalov, S., Beisel, K. W., Bono, H., Carninci, P., Fletcher, C. F., ... & Kawasaki, Y. (2003). Analysis of the mouse transcriptome for genes involved in the function of the nervous system. *Genome research*, 13(6b), 1395-1401.

H

Hamby, S. E., & Hirst, J. D. (2008). Prediction of glycosylation sites using random forests. *BMC bioinformatics*, 9(1), 500.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., ... & von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, 433(7024), 377-381.

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.

I

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic acids research*, 32(3), 1037-1049.

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.

Ishikawa, M., Iwamoto, T., Nakamura, T., Doyle, A., Fukumoto, S., & Yamada, Y. (2011). Pannexin 3 functions as an ER Ca²⁺ channel, hemichannel, and gap junction to promote osteoblast differentiation. *Journal of Cell Biology*, 193(7), 1257-1274.

J

Jochmann, R., Holz, P., Sticht, H., & Stürzl, M. (2014). Validation of the reliability of computational O-GlcNAc prediction. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1844(2), 416-421.

K

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780.

Kiemer, L., Bendtsen, J. D., & Blom, N. (2005). NetAcet: prediction of N-terminal acetylation sites. *Bioinformatics*, 21(7), 1269-1270.

Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, Sowa ME, Rad R, Rush J, Comb MJ, Harper JW, Gygi SP (2011). Systematic and quantitative assessment of the ubiquitin-modified proteome. *Molecular cell*. 44:325–40.

Kreegipuu, A., Blom, N., & Brunak, S. (1999). PhosphoBase, a database of phosphorylation sites: release 2.0. *Nucleic Acids Research*, 27(1), 237-239.

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, 305(3), 567-580.

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, 157(1), 105-132.

L

- Lee, T. Y., Hsu, J. B. K., Lin, F. M., Chang, W. C., Hsu, P. C., & Huang, H. D. (2010). N-Ace: Using solvent accessibility and physicochemical properties to identify protein N-acetylation sites. *Journal of computational chemistry*, 31(15), 2759-2771.
- Li, A., Xue, Y., Jin, C., Wang, M., & Yao, X. (2006). Prediction of Nε-acetylation on internal lysines implemented in Bayesian Discriminant Method. *Biochemical and biophysical research communications*, 350(4), 818-824.
- Li, S., Li, J., Ning, L., Wang, S., Niu, Y., Jin, N., ... & Xi, L. (2015). In silico identification of protein S-palmitoylation sites and their involvement in human inherited disease. *Journal of chemical information and modeling*, 55(9), 2015-2025.
- Liu, Z., Cao, J., Gao, X., Ma, Q., Ren, J., & Xue, Y. (2011). GPS-CCD: a novel computational program for the prediction of calpain cleavage sites. *PLoS One*, 6(4), e19001.
- Livingstone, C. D., & Barton, G. J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Bioinformatics*, 9(6), 745-756.
- Locke, D., Koreen, I. V., and Harris, A. L. (2006). Isoelectric points and post-translational modifications of connexin26 and connexin32. *FASEB J*, 20, 1221-1223.
- Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Scott, M. P., Bretscher, A., ... & Matsudaira, P. (2016). *Molecular cell biology*. Macmillan.

M

- Makarenkova, H. P., Shah, S. B., & Shestopalov, V. I. (2018). The two faces of pannexins: new roles in inflammation and repair. *Journal of inflammation research*, 11, 273.
- Martin PE, Evans WH (2004). Incorporation of connexins into plasma membranes and gap junctions. *Cardiovascular Research*. 62:378–87.
- Mistry, J., Bateman, A., & Finn, R. D. (2007). Predicting active site residue annotations in the Pfam database. *BMC bioinformatics*, 8(1), 298.
- Morgenstern, B. (2004). DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. *Nucleic acids research*, 32(suppl_2), W33-W36.

Müller, D. J., Wu, N., & Palczewski, K. (2008). Vertebrate membrane proteins: structure, function, and insights from biophysical approaches. *Pharmacological reviews*, 60(1), 43-78.

Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of molecular biology*, 247(4), 536-540.

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.

Nelson, D. L., Lehninger, A. L., & Cox, M. M. (2017). *Lehninger principles of biochemistry*. Macmillan.

Nguyen, V. N., Huang, K. Y., Weng, J. T. Y., Lai, K. R., & Lee, T. Y. (2016). UbiNet: an online resource for exploring the functional associations and regulatory networks of protein ubiquitylation. *Database*, 2016.

Nilsson, J., Persson, B., & von Heijne, G. (2005). Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes. *Proteins: Structure, Function, and Bioinformatics*, 60(4), 606-616.

Notredame, C., & Higgins, D. G. (1996). SAGA: sequence alignment by genetic algorithm. *Nucleic acids research*, 24(8), 1515-1524.

Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology*, 302(1), 205-217.

O

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.

O'Sullivan, O., Suhre, K., Abergel, C., Higgins, D. G., & Notredame, C. (2004). 3DCoffee: combining protein sequences and structures within multiple sequence alignments. *Journal of molecular biology*, 340(2), 385-395.

Oyamada, M., Oyamada, Y., & Takamatsu, T. (2005). Regulation of connexin expression. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1719(1-2), 6-23.

P

Panchina, Y., Kelmanson, I., Matz, M., Lukyanov, K., Usman, N., & Lukyanov, S. (2000). A ubiquitous family of putative gap junction molecules. *Current biology*, 10(13), R473-R474.

Pearson, W. R. (1994). Using the FASTA program to search protein and DNA sequence databases. In *Computer Analysis of Sequence Data* (pp. 307-331). Humana Press.

Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences*, 85(8), 2444-2448.

Penuela, S., Bhalla, R., Gong, X. Q., Cowan, K. N., Celetti, S. J., Cowan, B. J., ... & Laird, D. W. (2007). Pannexin 1 and pannexin 3 are glycoproteins that exhibit many distinct characteristics from the connexin family of gap junction proteins. *Journal of cell science*, 120(21), 3772-3783.

Penuela, S., Gehi, R., & Laird, D. W. (2013). The biochemistry and function of pannexin channels. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1828(1), 15-22.

Puk, O., Löster, J., Dalke, C., Soewarto, D., Fuchs, H., Budde, B., ... & Graw, J. (2008). Mutation in a novel connexin-like gene (*Gjfl*) in the mouse affects early lens development and causes a variable small-eye phenotype. *Investigative ophthalmology & visual science*, 49(4), 1525-1532.

Q

Qiu, W. R., Xiao, X., Lin, W. Z., & Chou, K. C. (2014). iMethyl-PseAAC: identification of protein methylation sites via a pseudo amino acid composition approach. *BioMed research international*, 2014.

R

- Račkauskas, M., Neverauskas, V., & Skeberdis, V. A. (2010). Diversity and properties of connexin gap junction channels. *Medicina*, 46(1), 1.
- Radivojac, P., Vacic, V., Haynes, C., Cocklin, R. R., Mohan, A., Heyen, J. W., ... & Iakoucheva, L. M. (2010). Identification, analysis, and prediction of protein ubiquitination sites. *Proteins: Structure, Function, and Bioinformatics*, 78(2), 365-380.
- Rahman S, Carlile G, Evans WH (1993). Assembly of hepatic gap junctions. Topography and distribution of connexin 32 in intracellular and plasma membranes determined using sequence-specific antibodies. *Journal of Biological Chemistry*. 268:1260–5.
- Ren, J., Wen, L., Gao, X., Jin, C., Xue, Y., & Yao, X. (2008). CSS-Palm 2.0: an updated software for palmitoylation sites prediction. *Protein Engineering, Design & Selection*, 21(11), 639-644.
- Rost, B., Fariselli, P., & Casadio, R. (1996). Topology prediction for helical transmembrane proteins at 86% accuracy—Topology prediction at 86% accuracy. *Protein Science*, 5(8), 1704-1718.
- Rost, B., Sander, C., Casadio, R., & Fariselli, P. (1995). Transmembrane helices predicted at 95% accuracy. *Protein Science*, 4(3), 521-533.
- Russell, R. B., & Barton, G. J. (1992). Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins: Structure, Function, and Bioinformatics*, 14(2), 309-323.

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.
- Shao, J., Xu, D., Tsai, S. N., Wang, Y., & Ngai, S. M. (2009). Computational identification of protein methylation sites through bi-profile Bayes feature extraction. *PloS one*, 4(3), e4920.
- Shien, D. M., Lee, T. Y., Chang, W. C., Hsu, J. B. K., Horng, J. T., Hsu, P. C., ... & Huang, H. D. (2009). Incorporating structural characteristics for identification of protein methylation sites. *Journal of computational chemistry*, 30(9), 1532-1543.

- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Thompson, J. D. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), 539.
- Sillitoe, I., Lewis, T. E., Cuff, A., Das, S., Ashford, P., Dawson, N. L., ... & Lehtinen, S. (2015). CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic acids research*, 43(D1), D376-D381.
- SIPOS, L., & VON HEIJNE, G. (1993). Predicting the topology of eukaryotic membrane proteins. *European Journal of Biochemistry*, 213(3), 1333-1340.
- Smith, T. F., & Waterman, M. S. (1981). Comparison of biosequences. *Advances in applied mathematics*, 2(4), 482-489.
- Söhl, G., & Willecke, K. (2004). Gap junctions and the connexin protein family. *Cardiovascular research*, 62(2), 228-232.
- Solan, J. L., & Lampe, P. D. (2014). Specific Cx43 phosphorylation events regulate gap junction turnover in vivo. *FEBS letters*, 588(8), 1423-1429.
- Song, J., Li, F., Leier, A., Marquez-Lago, T. T., Akutsu, T., Haffari, G., ... & Pike, R. N. (2018). PROSPERous: high-throughput prediction of substrate cleavage sites for 90 proteases with improved accuracy. *Bioinformatics*, 34(4), 684-687.
- Song, J., Tan, H., Perry, A. J., Akutsu, T., Webb, G. I., Whisstock, J. C., & Pike, R. N. (2012). PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PloS one*, 7(11), e50300.
- Sonntag, S., Söhl, G., Dobrowolski, R., Zhang, J., Theis, M., Winterhager, E., ... & Willecke, K. (2009). Mouse lens connexin23 (Gje1) does not form functional gap junction channels but causes enhanced ATP release from HeLa cells. *European journal of cell biology*, 88(2), 65-77.
- Stentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T. B., ... & Gupta, R. (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The EMBO journal*, 32(10), 1478-1488.

Stoesser, G., Baker, W., van den Broek, A., Garcia-Pastor, M., Kanz, C., Kulikova, T., ... & Mancuso, R. (2003). The EMBL nucleotide sequence database: major new developments. *Nucleic Acids Research*, 31(1), 17-22.

Stoye, J., Moulton, V., & Dress, A. W. (1997). DCA: an efficient implementation of the divide-and-conquer approach to simultaneous multiple sequence alignment. *Bioinformatics*, 13(6), 625-626.

Suo, S. B., Qiu, J. D., Shi, S. P., Sun, X. Y., Huang, S. Y., Chen, X., & Liang, R. P. (2012). Position-specific analysis and prediction for protein lysine acetylation based on multiple features. *PloS one*, 7(11), e49108.

T

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.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, 22(22), 4673-4680.

Tusnady, G. E., & Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics*, 17(9), 849-850.

W

Wagner SA, Beli P, Weinert BT, Nielsen ML, Cox J, Mann M, Choudhary C (2011). A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Molecular and cellular Proteomics*.; 10:M111. 013284.

Wang, M., Zhao, X. M., Tan, H., Akutsu, T., Whisstock, J. C., & Song, J. (2014). Cascleave 2.0, a new approach for predicting caspase and granzyme cleavage targets. *Bioinformatics*, 30(1), 71-80.

Wang, X. B., Wu, L. Y., Wang, Y. C., & Deng, N. Y. (2009). Prediction of palmitoylation sites using the composition of k-spaced amino acid pairs. *Protein Engineering, Design & Selection*, 22(11), 707-712.

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.

Wilbur, W. J., & Lipman, D. J. (1983). Rapid similarity searches of nucleic acid and protein data banks. *Proceedings of the National Academy of Sciences*, 80(3), 726-730.

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. *Biological chemistry*, 383(5), 725-737.

Wolfenden, R., Andersson, L., Cullis, P. M., & Southgate, C. C. B. (1981). Affinities of amino acid side chains for solvent water. *Biochemistry*, 20(4), 849-855.

Wong, Y. H., Lee, T. Y., Liang, H. K., Huang, C. M., Wang, T. Y., Yang, Y. H., ... & Hwang, J. K. (2007). KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic acids research*, 35(suppl_2), W588-W594.

X

Xue, Y., Liu, Z., Cao, J., & Ren, J. (2011). Computational prediction of post-translational modification sites in proteins. *Yang NS, Ed*, 105-124.

Xue, Y., Ren, J., Gao, X., Jin, C., Wen, L., & Yao, X. (2008). GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Molecular & cellular proteomics*, 7(9), 1598-1608.

Xue, Y., Zhou, F., Fu, C., Xu, Y., & Yao, X. (2006). SUMOsp: a web server for sumoylation site prediction. *Nucleic acids research*, 34(suppl_2), W254-W257.

Y

Yalak, G., Ehrlich, Y. H., & Olsen, B. R. (2014). Ecto-protein kinases and phosphatases: an emerging field for translational medicine. *Journal of translational medicine*, 12(1), 165.

Yeager, M. (1998). Structure of cardiac gap junction intercellular channels. *Journal of structural biology*, 121(2), 231-245.

Yen, M. R., & Saier Jr, M. H. (2007). Gap junctional proteins of animals: the innexin/pannexin superfamily. *Progress in biophysics and molecular biology*, 94(1-2), 5-14.

Z

Zhang, X., Chen, L., Guo, Z. H., & Liang, H. (2019). Identification of human membrane protein types by incorporating network embedding methods. *IEEE Access*, 7, 140794-140805.

Zhang, Z., Miller, W., Schäffer, A. A., Madden, T. L., Lipman, D. J., Koonin, E. V., & Altschul, S. F. (1998). Protein sequence similarity searches using patterns as seeds. *Nucleic acids research*, 26(17), 3986-3990.

Zhao, Q., Xie, Y., Zheng, Y., Jiang, S., Liu, W., Mu, W., ... & Ren, J. (2014). GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic acids research*, 42(W1), W325-W330.

Zhou, F., Xue, Y., Yao, X., & Xu, Y. (2006). A general user interface for prediction servers of proteins' post-translational modification sites. *Nature protocols*, 1(3), 1318-1321.

Zhou, F., Xue, Y., Yao, X., & Xu, Y. (2006). CSS-Palm: palmitoylation site prediction with a clustering and scoring strategy (CSS). *Bioinformatics*, 22(7), 894-896.

Résumé

Dans le génome humain 21 isoformes de connexines ont été identifiées servant à la formation des canaux jonctionnels. Cependant, il n'existe que trois isoformes de pannexines servant à la formation des hémicanaux. Dans ce travail, nous avons fait une analyse structurale et fonctionnelle de la connexine23 (CX23) aux moyens d'outils informatiques. En premier temps, nous avons caractérisé les différents domaines de la CX23 en mettant l'accent sur les résidus essentiels par la prédiction des modifications post-traductionnelles. Puis, nous avons utilisé les séquences de la CX23 et les trois *hPanx* afin de montrer les régions conservées entre ces protéines. L'étude de la topologie avec les programmes ProtScale, ΔG predictor et TOPCONS a montré que la CX23 est une protéine membranaire à quatre passages transmembranaires exposant deux boucles extracellulaires et une boucle intracytoplasmique avec des extrémités N- et C-terminales cytoplasmiques. Les résultats de l'étude des modifications post-traductionnelles ont montré que la CX23 semble contenir des sites de phosphorylation, palmitoylation et protéolyse. Cependant, elle semble ne contenir aucun site de glycosylation. L'alignement multiple des séquences de la CX23 avec les trois *hPanx* a été réalisé par le programme T-Coffee révélant que les deux résidus cystéines dans chaque boucle extracellulaire de ces protéines sont très conservés.

Mots clés : connexines, pannexines, CX23, modifications post-traductionnelles, alignement multiple

Abstract

In the human genome, 21 connexin isoforms have been identified that are used for the formation of junction channels. However, there are only three pannexin isoforms for the formation of hemichannels. In this work, we performed a structural and functional analysis of the connexin23 (CX23) using computer tools. First, we characterised the different domains of CX23, focusing on the essential residues by predicting post-translational modifications, Then, we used the sequences of CX23 and the three *hPanx* to show the conserved regions between these proteins. The study of the topology with the ProtScale, ΔG predictor and TOPCONS programmes showed that CX23 is a four-pass transmembrane membrane protein exposing two extracellular loops and a intracytoplasmic loop with cytoplasmic N- and C-terminal extremities. The results of the post-translational modification study suggested that CX23 appears to contain sites of phosphorylation, palmitoylation and proteolysis. However, it does not appear to contain any sites of glycosylation. The multiple alignment sequences of the CX23 with the three *hPanx* was performed by the T-Coffee programme, revealing that the two cysteine residues in each extracellular loop of these proteins are highly conserved.

Keywords: connexins, pannexins, CX23, post-translational modifications, multiple alignment

ملخص

عند الإنسان يوجد 21 عضو من الكناكسينات التي تشكل قنوات ناقلة. بينما يوجد ثلاثة اعضاء من البانكسينات التي تقوم بتشكيل قنوات نصفية. في هذا العمل قمنا باجراء تحليل هيكلي ووظيفي للكناكسين 23 باستخدام ادوات معلوماتية. اولاً قمنا بتحديد المقاطع المختلفة للكناكسين 23 مع دراسة التعديلات ما بعد الترجمة. ثم قمنا باستخدام المتتاليات البشرية للكناكسين 23 والبنكسينات الثلاث من اجل اظهار المناطق المحفوظة بين هذه البروتينات. دراسة طبولوجيا عن طريق ثلاثة برامج وضحت ان الكناكسين 23 عبارة عن بروتين غشائي مع اربعة ممرات عبر الغشاء بالاضافة الى حلقتين خارجيتين وحلقة داخلية مع النهايات الامينة والكربوكسيلية من الجانب الداخلي للسيتوبلازم اظهرت نتائج التعديلات ما بعد الترجمة ان الكناكسين 23 يحتوي على مواقع اضافة الفسفرة حمض البالمتيك وتحلل البروتين الا انه لا يحتوي على أي موقع لاضافة السكر. محاذاة المتتاليات البشرية للكناكسين 23 والبنكسينات وضحت بان عنصري السيستين في كل حلقة خارجية لهذه البروتينات يكونان جد محفوظين.

الكلمات المفتاحية: محاذاة متعددة- الكناكسينات- البانكسينات- الكوناكسين 23- تعديلات ما بعد الترجمة.