

# 1

## Introduction

## 1.1 La communication chez les organismes cellulaires

Il existe plusieurs manières pour les cellules de communiquer entre elles. Cette communication est essentielle pour coordonner leurs fonctions et s'adapter à leur environnement. Nous allons dans un premier temps présenter les principales stratégies utilisées par les cellules pour transmettre et recevoir de l'information, avant de nous concentrer sur les mécanismes de la communication intercellulaire spécifique aux plantes.

### 1.1.1 Organismes unicellulaires et multicellulaires

La communication entre cellules est commune à tous les organismes. Selon la nature du signal à transmettre et de la cellule réceptrice, elle prend des formes variées qui impactent la vitesse de transmission, la distance parcourue, ou encore le type de molécule messagère. Chez les organismes unicellulaires, la communication peut être intraspécifique (entre cellules de la même espèce) ou interspécifique (entre cellules d'espèces différentes). Par exemple, les bactéries utilisent des molécules messagères comme l'AMPc pour coordonner des comportements collectifs, tels que la formation de biofilms protecteurs contre les antibiotiques. [1]. Cependant, ces signaux peuvent également servir à d'autres fins, comme la compétition pour les ressources ou la détection de changement dans les conditions environnementales du milieu. Chez les organismes multicellulaires, la communication est encore plus complexe en raison de la différenciation cellulaire. Les cellules spécialisées doivent interagir de manière précise pour assurer le bon fonctionnement de l'organisme dans son ensemble. Par exemple, chez les animaux, les neurones communiquent rapidement via des synapses sur de courtes distances, tandis que les cellules endocrines utilisent des hormones pour transmettre des signaux sur de longues distances en se servant du réseau sanguin. Chez les plantes, la multicellularité clonale implique que les cellules issues de la division restent interconnectées, formant des tissus différenciés aux fonctions spécialisées. Cette continuité physique facilite les échanges directs entre cellules voisines, permettant une coordination fine au sein des tissus. Ainsi, la communication chez les multicellulaires repose sur des réseaux sophistiqués, où les cellules d'un même tissu interagissent différemment selon leur spécialisation, et où les échanges entre tissus distincts nécessitent des mécanismes de signalisation adaptés.

### 1.1.2 Les différents types de communication

La communication intercellulaire peut être classée en deux grandes catégories : directe et indirecte, selon que les cellules interagissent physiquement ou via des molécules de signalisation. Ces mécanismes reposent sur des voies de signalisation spécifiques, illustrées dans la figure 1.1.

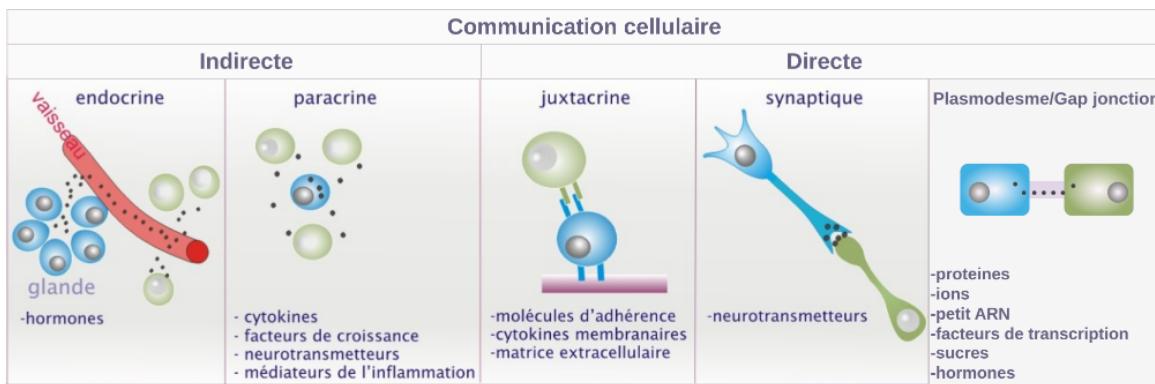


FIGURE 1.1 – Type de communication cellulaire. Sous chaque schéma, on trouve une liste des exemples de molécules ou messagers impliqués dans chaque voie de communication, tels que les hormones, les facteurs de croissance, ou les ions, qui jouent un rôle clé dans la transmission des signaux. Adapté de IJsbrand Kramer.

## Communication indirecte

Dans la communication indirecte, les cellules échangent des informations via des molécules de signalisation libérées dans l'environnement. On distingue trois modes principaux :

**Communication autocrine** : Une cellule produit elle-même des molécules de signalisation qui se lient à ses propres récepteurs, lui permettant de réguler son activité de manière autonome. Ce mécanisme est crucial dans des processus comme la prolifération cellulaire ou la réponse immunitaire.

**Communication paracrine** : Une cellule libère des molécules de signalisation qui agissent sur des cellules voisines. Par exemple, les facteurs de croissance stimulent la division cellulaire dans les tissus environnants. Ce type de communication est souvent localisé et rapide.

**Communication endocrine** : Une hormone est sécrétée dans la circulation sanguine (chez les animaux) ou dans d'autres fluides, permettant d'atteindre des cellules cibles éloignées. Ce mode de communication est plus lent mais permet une coordination à l'échelle de l'organisme entier.

## Communication directe

La communication directe implique un contact physique entre les cellules, permettant un échange rapide et précis d'informations. Deux mécanismes principaux existent :

**Communication juxtacrine** : Les cellules interagissent directement via des molécules membranaires, comme les protéines de surface, qui se lient à des récepteurs sur une cellule adjacente. Ce type de communication est essentiel dans des processus comme la reconnaissance cellulaire ou la réponse immunitaire.

**Communication par jonctions** : Des canaux spécialisés permettent un échange di-

rect de molécules entre cellules adjacentes. Chez les animaux, les gap junctions facilitent le passage d'ions et de petites molécules, assurant une coordination rapide au sein des tissus. Chez les plantes, les plasmodesmes maintiennent une continuité cytoplasmique entre cellules, permettant un flux direct de nutriments et de signaux.

Ainsi, la distinction entre communication directe et indirecte ne repose pas seulement sur la présence ou l'absence de contact physique entre les cellules, mais aussi sur la nature des molécules impliquées. Ces mécanismes complémentaires permettent aux organismes de s'adapter à des besoins variés, allant de la coordination locale à la régulation globale.

### 1.1.3 Communications intercellulaires chez les plantes

Au cours de l'évolution, les plantes ont développé deux voies principales de communication intercellulaire.(voir figure 1.2) :

- **Le transport symplasmique**, qui permet un échange direct entre cytoplasmes de cellules adjacentes via des canaux spécialisés appelés plasmodesmes.
- **Le transport apoplasmique**, qui repose sur la diffusion de molécules à travers les parois cellulaires et les espaces intercellulaires.

Ces deux voies diffèrent essentiellement de par la manière dont les signaux vont traverser la MP. Les cellules végétales possèdent une paroi cellulaire rigide, composée de cellulose et d'autres polysaccharides complexes, qui entourent et protègent la MP (MP). Bien qu'elle assure un rôle de soutien et de protection contre les agressions extérieures, elle constitue également une barrière physique aux échanges intercellulaires. Alors que le transport apoplasmique est principalement passif et dépend des propriétés de la paroi, le transport symplasmique est finement contrôlé par l'ouverture et la perméabilité des plasmodesmes. Ce contrôle dynamique permet d'ajuster le passage des molécules en fonction des besoins physiologiques de la plante en réponse à des stress biotiques ou abiotiques. Les plasmodesmes jouent un rôle fondamental en établissant des connexions directes entre cellules adjacentes. Ils permettent l'échange direct de molécules entre cellules tout en régulant ces échanges . Ils sont essentiels aux processus physiologiques des plantes. Nous allons maintenant examiner leur structure et leur fonction pour mieux comprendre leur rôle dans la communication intercellulaire végétale.

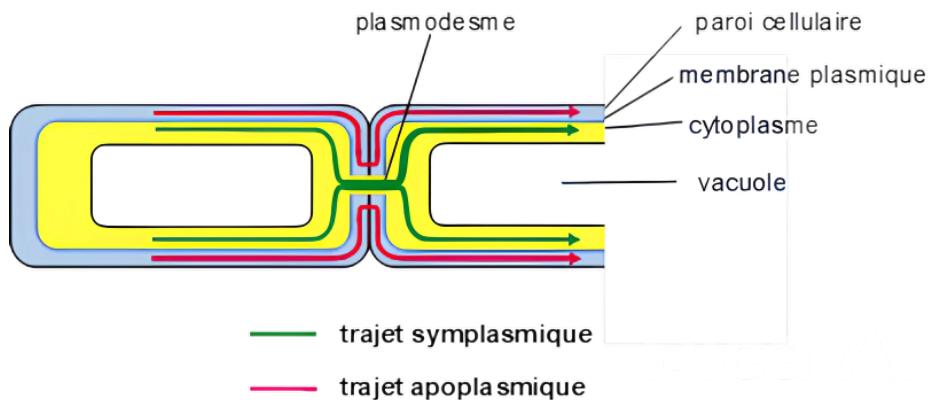


FIGURE 1.2 – Les transports symplasmique et apoplasmique chez les plantes. D'après : <https://cdn.futura-sciences.com/buildsv6/images>

## 1.2 Le rôle clé du plasmodesme dans la communication cellulaire chez les plantes

### 1.2.1 Plasmodesmes : structure et fonction

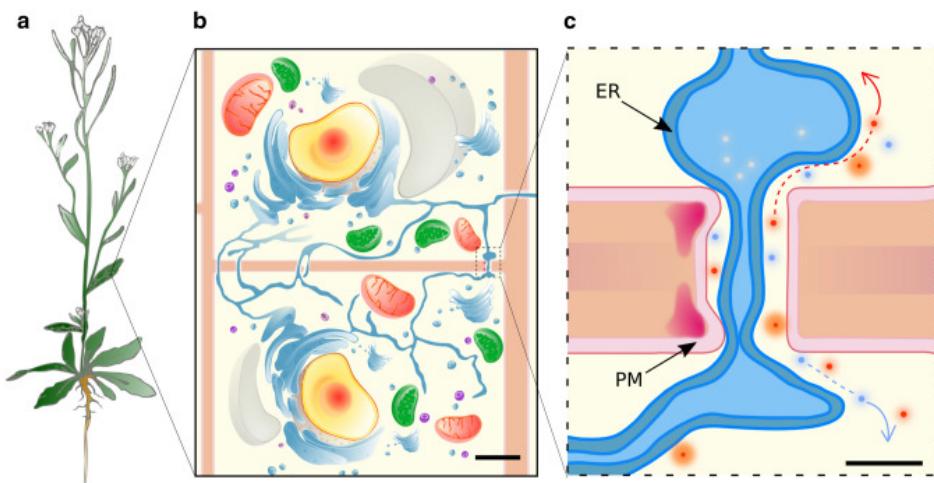


FIGURE 1.3 – Trois points de vue d'*Arabidopsis thaliana*. a) Vue extérieure de la plante. b) Vue cellulaire, réseau endoplasmique (ER) en bleu et la membrane plasmique (PM) en rose. c) Vue d'un plasmodesme. Image issue de [2]

Les cellules végétales possèdent une structure unique. Comme expliqué précédemment, leur paroi cellulaire rigide, composée principalement de cellulose et d'autres sucres complexes, entoure et protège la membrane plasmique (MP). En plus de son rôle structurel, elle intervient dans la signalisation cellulaire (voie apoplasmique), notamment en réponse aux attaques de pathogènes et aux stress abiotiques. Cependant, cette structure limite les échanges intercellulaires, restreignant la diffusion passive de nombreuses molécules. Ainsi elles utilisent en parallèle, la voie symplasmique via les plasmodesmes. Les plasmodesmes sont des canaux, bordés par la MP et traversés par un desmotubule, une extension du RE (RE)(Voir figure 1.3). L'espace entre le RE et la MP, appelé manchon cytoplasmique, permet le transport de molécules telles que des nutriments, des hormones, des facteurs de croissance et des protéines, des petits ARN, des ions, des protéines et facteurs de transcription, des sucres issus de la photosynthèse. Ces signaux mobiles sont nécessaires à la survie et au développement des plantes. [3, 4, 5] (voir figure 1.3).

Les plasmodesmes ont une composition moléculaire spécifique en protéines et lipides qui restent encore mal comprise, en particulier pour les lipides.

En dehors des plasmodesmes, la MP présente une composition complexe, incluant des phosphatidylcholines (PC), des phosphatidylséries (PS), du sitostérol et des phosphatidylinositols (PI)[6]. Cette composition peut varier en fonction de l'état cellulaire, notamment sous l'effet du stress, influençant par exemple les niveaux de phosphatidylinositol-4-phosphate (PI4P), qui augmentent lorsque la concentration de sel augmente [7]. La membrane du RE est principalement composée de phosphocholine (PC) et de phosphoéthanolamine (PE). Cependant, la composition du desmotubule, la partie du RE dans le plasmodesme, reste incertaine. Pour la suite de notre projet, nous avons supposé qu'elle était relativement proche de celle du RE. En outre, l'espace entre les membranes MP et RE, varie entre 2 et 10 nm et permet le transport de molécules de différentes tailles, des ions jusqu'à des complexes protéiques. [3]. Ainsi, la régulation de cet espace est essentielle pour la régulation dynamique des échanges intercellulaires et donc du bon fonctionnement des cellules et de la plante. Cette capacité est cruciale pour coordonner les réponses à l'échelle de la plante, notamment parce que des organes comme les feuilles et les racines se trouvent dans des environnements très différents. Par exemple, lorsque les racines détectent un manque d'eau, elles doivent transmettre cette information aux feuilles pour ajuster leur taux de transpiration. Pour ce faire, l'information doit parfois parcourir de longues distances, traversant des milliers de cellules interconnectées.

### 1.2.2 Rôle essentiel des plasmodesmes dans le développement et l'adaptation des plantes

Les PD sont des structures clés de la communication intercellulaire chez les plantes, influençant leur physiologie, leur développement et leur capacité d'adaptation aux stress environnementaux. Ils permettent l'échange de signaux, de nutriments et d'hormones, coordonnant ainsi les processus physiologiques et développementaux. Durant le développement, les PD régulent la formation des organes et la différenciation cellulaire en contrôlant le transport d'auxines, de cytokinines et de facteurs de transcription essentiels à l'établissement des domaines symplasmiques. [8]

En réponse aux stress abiotiques, comme les variations de température ou le manque d'eau, leur ouverture ou fermeture, modulée entre autres par la callose, un polymère de glucose ( $\beta$ -1,3-glucan) déposé autour des PD impactent la distribution des ressources et la tolérance aux conditions extrêmes. Par exemple, chez *Arabidopsis thaliana*, une exposition à des températures élevées (30°C pendant 3–4 jours) bloque le déchargement du phloème, ce qui perturbe l'approvisionnement en produits photosynthétiques vers les racines et inhibe leur croissance. Ce phénomène implique la callose synthase CALS8, qui intervient dans la régulation du trafic des PD en réponse au stress thermique.[9] De manière similaire, le stress froid entraîne une accumulation de callose dans les méristèmes apicaux et floraux, affectant le transport intercellulaire du facteur de

transcription WUSCHEL (WUS) et perturbant le développement floral.

Dans le contexte des stress biotiques, les PD sont également des cibles des pathogènes, qui manipulent leur ouverture pour favoriser l'invasion des tissus végétaux. Des agents pathogènes comme *Pseudomonas syringae*, *Fusarium oxysporum* et divers virus exploitent les PD pour propager leurs infections. [10][11][12] . Par exemple, *Phytophthora brassicae* sécrète un effecteur, RxRL3, qui cible les callose synthases et inhibe l'accumulation de callose, favorisant ainsi son mouvement intercellulaire [13]. Cependant, les plantes activent une réponse immunitaire en restreignant la perméabilité des PD via la signalisation hormonale et l'accumulation de callose.

Ainsi, les PD jouent un rôle central dans la régulation de la croissance, la défense et l'adaptation des plantes aux conditions environnementales. Leur régulation fine est donc essentielle pour maintenir un équilibre entre développement et réponse aux stress environnementaux.

### 1.2.3 Biogenèse des plasmodesmes

Contrairement aux cellules animales, dont la division cellulaire entraîne une séparation complète des cellules filles, la cytokinèse des plantes maintient une connexion physique entre cellules adjacentes via des plasmodesmes. Voir figure 1.4. Ce processus, appelé cytokinèse incomplète, permet la formation des plasmodesmes qui se forment dans la plaque cellulaire naissance. Ainsi dès que les cellules filles se forment elles sont déjà connectées et peuvent communiquer.

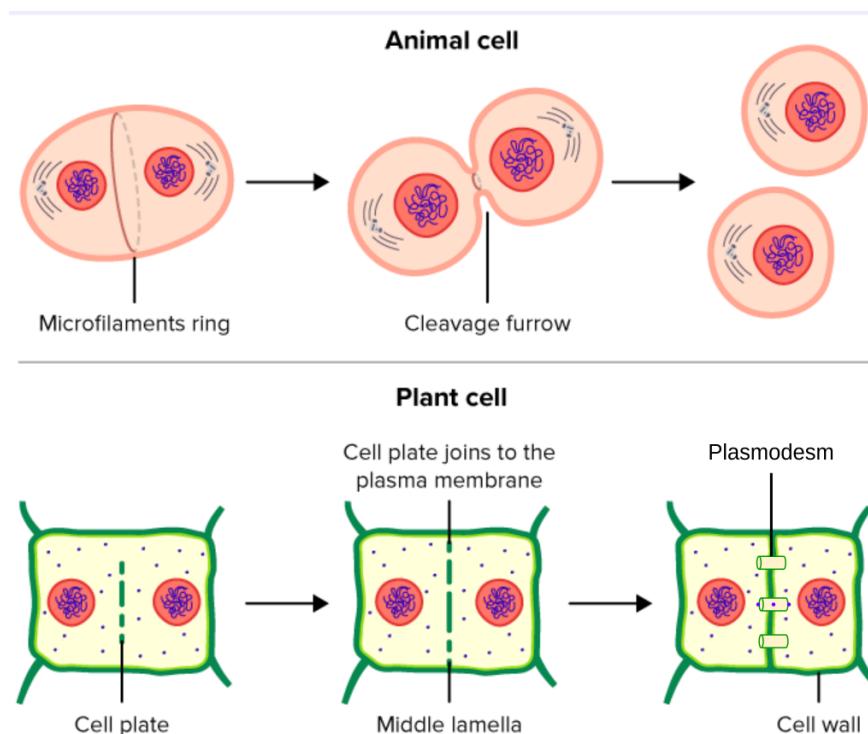


FIGURE 1.4 – Division cellulaire : comparaison de la cytokinèse entre les mammifères et les plantes (*image par CK-12 Foundation*) (Faire apparaître les PD sur le schéma)

Selon les travaux récents de Patrick Li et al. [14], chez *Arabidopsis thaliana*, le RE joue un rôle central dans la formation et la stabilisation des plasmodesmes. La formation des plasmodesmes à la fin de la division cellulaire dépend de l'infiltration du RE, qui s'intègre à leur structure. Ce processus est stabilisé par les protéines *Multiple C2 domain and Transmembrane region Proteins* (MCTP), qui assurent le lien entre le RE et la membrane plasmique, garantissant ainsi la pérennité des plasmodesmes. À l'issue du processus, des centaines de plasmodesmes sont formés, établissant une continuité cytoplasmique directe. Voir figure A,B et C de 1.5.

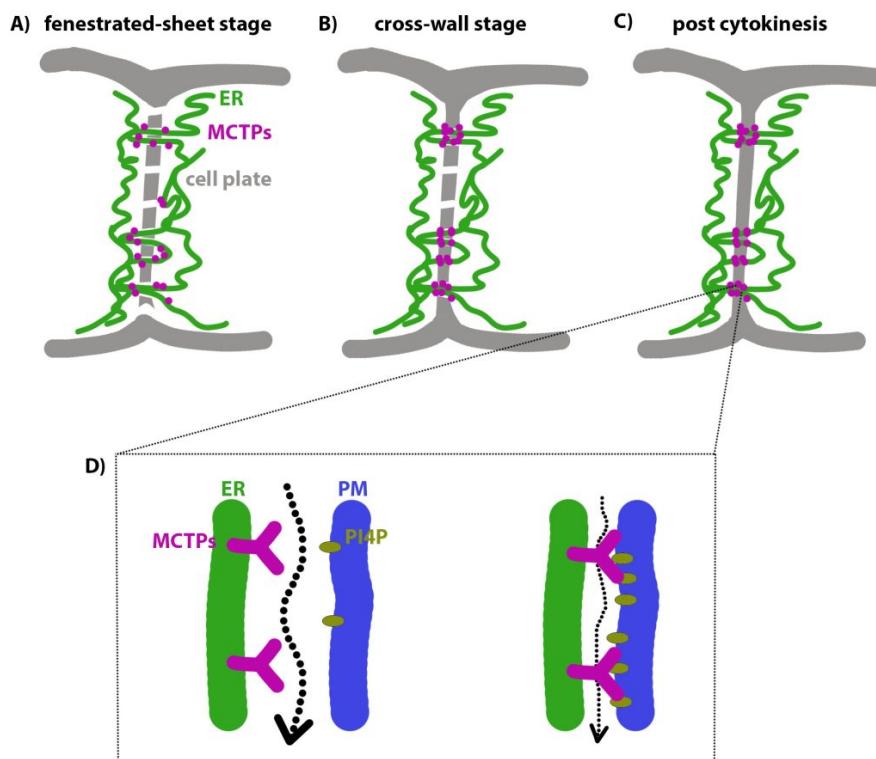


FIGURE 1.5 – Formation des plasmodesmes (PDs) par abscission incomplète du réticulum endoplasmique (RE). Les MCTPs (multiple C2 domaines et transmembrane protéine) (magenta) localisés au RE (vert) de la plaque cellulaire (gris) (A) s’accumulent aux PDs au stade de paroi transversale (B) et empêchent l’abscission du RE lors de la cytokinèse, stabilisant ainsi les ponts intercellulaires (C). Les MCTPs régulent les sites de contact membranaire (MCS) des PDs en reliant la membrane plasmique (MP, bleu) et le RE via le lipide PI4P (jaune) (D). Extrait de <https://prelights.biologists.com>

### 1.2.4 Régulation du plasmodesme

Les plasmodesmes sont des structures dynamiques, permettant aux cellules végétales de contrôler la communication intercellulaire. Comme expliqué précédemment, les cellules peuvent ajuster l'ouverture ou la fermeture des plasmodesmes en fonction des besoins physiologiques de chaque cellule. Ainsi, les cellules peuvent fermer certains plasmodesmes ou en ouvrir d'autres, dirigeant la communication. Un facteur clé du contrôle de l'ouverture et de la fermeture des plasmodesmes est la régulation de la callose, dont nous avons vu l'importance dans la réponse aux stress. Ce polysaccharide (-1,3-glucane) s'accumule ou se dégrade sur la paroi végétale entourant les plasmodesmes, ajustant ainsi leur perméabilité. Lors de périodes de stress ou de différenciation, le niveau de callose est régulé par deux activités enzymatiques : les callose synthases (ou glucane synthase-like) et les  $\beta$ -1,3-glucanases [15]. Cette accumulation ou dégradation, proche des extrémités des ouvertures des plasmodesmes, permet de réguler la perméabilité des plasmodesmes à priori en poussant la MP contre le RE est fermant le manchon cytoplasmique [16] (voir Figure 1.6).

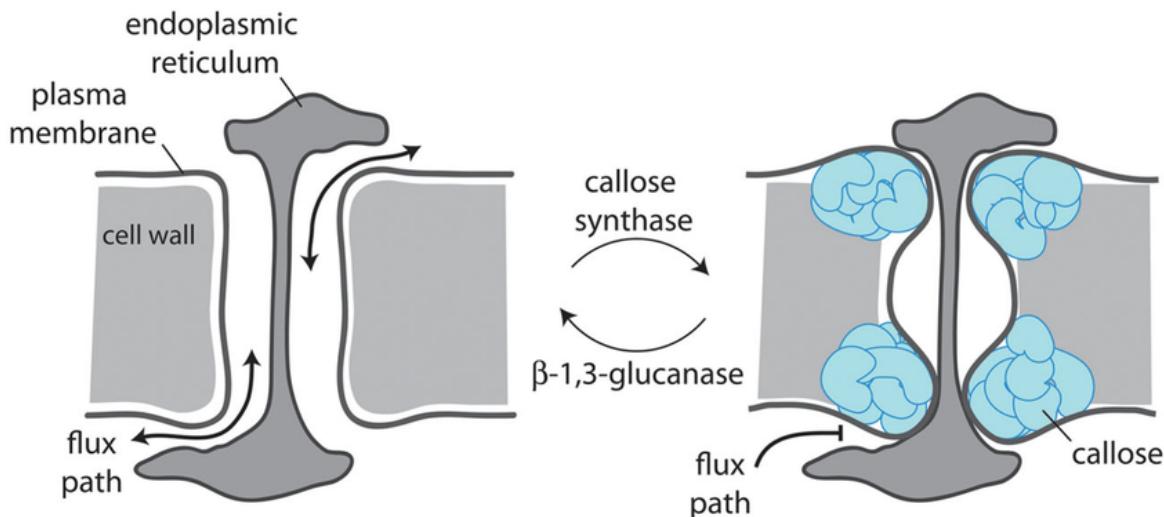


FIGURE 1.6 – Développement de la callose dans les régions de la membrane plasmique entourant les plasmodesmes en fonction des calloses synthases et des  $\beta$ -1,3-glucanases. Image issue de [17].

Toutefois, des études récentes démontrent que l'accumulation de callose n'est pas le seul mécanisme de régulation des nanopores, et qu'un autre mécanisme, plus central, repose sur des acteurs internes aux plasmodesmes situés entre le RE et la MP. En particulier, la famille de protéines nommée MCTP (Multiple C2 domain and Trans-membrane region Proteins) joue un rôle crucial dans la régulation et la stabilité de l'espace ER-MP dans le plasmodesme[5]. Des problèmes de transport moléculaire apparaissent même lorsque la callose est produite normalement, ce qui montre que seul la

callose ne suffit pas à fermer les plasmodesmes. De plus, forcer la production de callose dans un mutant MCTP ne permet pas leur fermeture, suggérant que la callose a besoin des MCTP pour bien fonctionner.[18].

### 1.3 Le rôle des protéines MCTP

Les protéines MCTP (Multiple C2 domain and Transmembrane region Proteins) régulent l'espace entre les membranes cellulaires en remplissant plusieurs fonctions essentielles. Les travaux de Brault et al. ont identifié les MCTP comme des protéines d'attachement entre ces deux membranes. Leur région transmembranaire permet leur ancrage au RE, tandis que leurs domaines C2 interagissent avec des lipides spécifiques de la MP, assurant ainsi leur localisation aux plasmodesmes et leur fonction de stabilisation [18]. Dans le cadre de ma thèse, j'ai contribué à l'étude de Perez-Sancho et al., en réalisant des simulations des interactions entre les domaines C2 de MCTP4 et la MP contenant du PI4P Voir figure D 1.5. Cette étude a montré que MCTP3, MCTP4 et MCTP6 sont exprimés dans presque tous les tissus et spécifiquement localisés aux plasmodesmes. Les résultats expérimentaux ont révélé que ces trois protéines forment un complexe fonctionnel, essentiel pour la régulation de l'espace intermembranaire et la stabilité des plasmodesmes. Parmi ces trois protéines, MCTP4 joue un rôle central. En effet, seuls les mutants doubles ou triples impliquant la mutation de MCTP4 présentent des défauts majeurs de développement, suggérant que cette protéine est essentielle à la croissance de la plante. Ces observations montrent que la déstabilisation de l'espace entre le RE et la MP entraîne des conséquences sévères sur le développement, soulignant ainsi l'importance de sa régulation (voir Figure 1.7). Ensemble nous avons montrer les protéines MCTP régulent l'espace entre les membranes en collaboration avec le lipide PI4P.

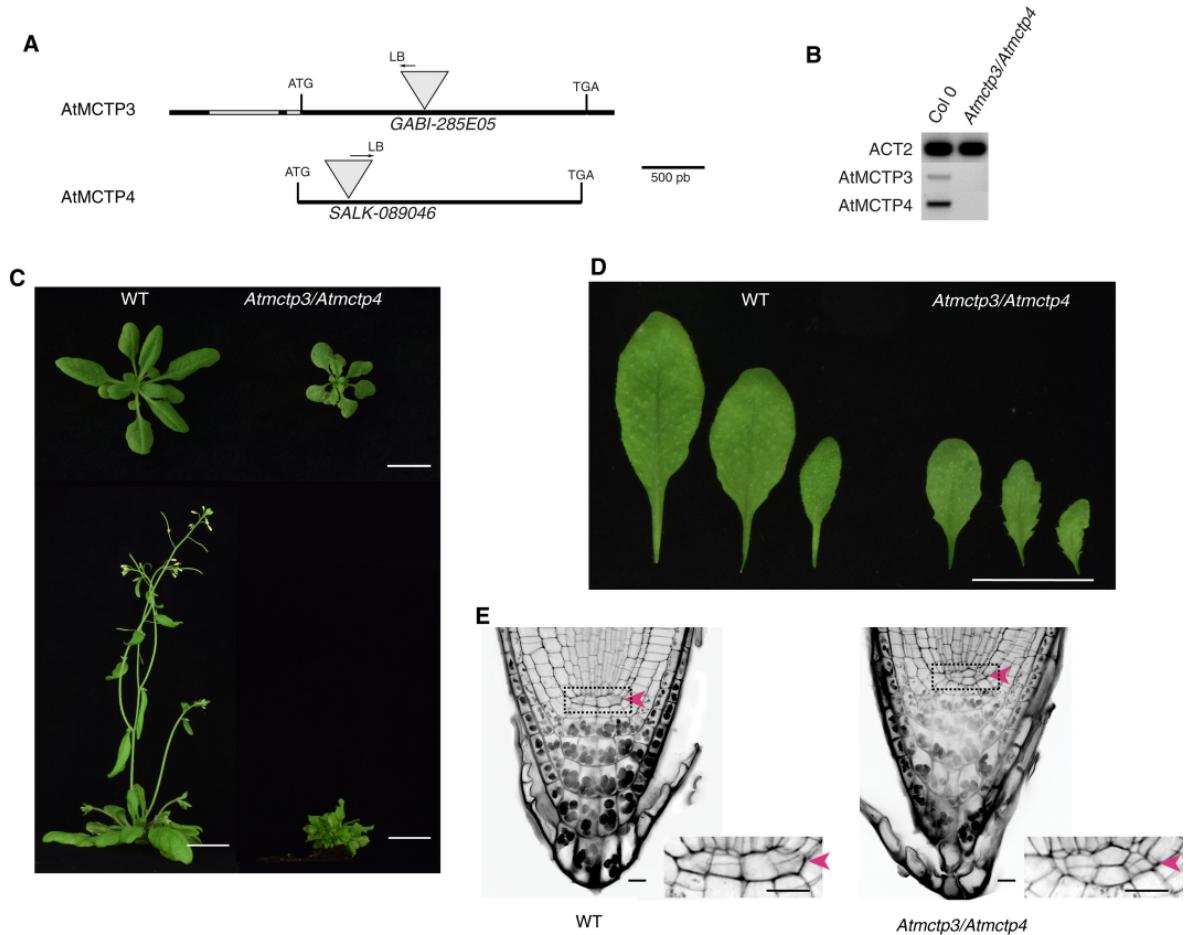


FIGURE 1.7 – A) Les mutations des protéines MCTP3 et MCTP4. B) Chez le double mutant les protéines MCTP3 et MCTP4 ne sont pas exprimées. C) Le double mutant entraîne un retard de développement majeur chez *Arabidopsis thaliana*. D) Phénotype des feuilles WT et mutante. E) Phénotype des racine WT et mutante. Figure extraite de [18]

### 1.3.1 MCTP structure

Les MCTP sont présents chez de nombreux organismes et partagent une structure similaire. Par exemple, les métazoaires tels que la drosophile possèdent deux MCTP (MCTP1 et MCTP2), tandis que l'Homme n'en possède qu'un seul (Shin et al., 2005) [19]. Chez les plantes, les MCTP forment de grandes familles, avec 16 membres identifiés chez *Arabidopsis thaliana* [20] et 31 chez le coton [21]. Les MCTP présentent des caractéristiques structurelles conservées au cours de l'évolution entre les espèces : trois ou quatre domaines C2 à l'extrémité N-terminale et un ou deux domaines transmembranaires à l'extrémité C-terminale [19] (voir figure ??). Cependant, aucune structure expérimentale n'est actuellement disponible pour ces protéines.



FIGURE 1.8 – Organisation des domaines de la famille de protéines MCTP d’Arabidopsis. Alignement des 16 protéines MCTP d’*A. thaliana*. Les domaines C2 sont représentés en bleu et les domaines transmembranaires (TMD) en jaune. Extrait de [22]

Néanmoins, des tentatives de modélisation de la structure de MCTP4 chez *Arabidopsis thaliana* à été réalisée par Jules Petit. Il a permis de définir séparément les différentes parties de la protéine, notamment les deux "épingles" d'hélices (HP1 et HP2) qui constituent la région transmembranaire (TMR) ainsi que les hélices membranaire AH1, TM0, AH2 et les domaines C2 (C2B, C2C et C2D). Dans un premier temps, une combinaison de plusieurs outils de modélisation ont été utilisés pour délimiter les régions transmembranaires. Pour déterminer les régions transmembranaires (TMR), deux outils complémentaires ont été utilisés : PSIPRED (Position-Specific Iterated Prediction of Secondary Structure), un logiciel de prédiction de la structure secondaire des protéines, et l’analyse des clusters hydrophobes (HCA) (Hydrophobic Cluster Analysis), qui visualise et analyse les motifs hydrophobes dans les séquences protéiques (voir Figure 1.9). Une fois les régions transmembranaires identifiées, chaque sous-domaine a été modélisé individuellement à l'aide du serveur Robetta. Ces modèles ont ensuite été intégrés dans une boîte de simulation de dynamique moléculaire (DM) avec une membrane, où des simulations ont été réalisées pour étudier les interactions des épingles transmembranaires et des hélices avec la membrane (voir Figure 1.9). Les modèles des domaines C2 ont été obtenus par une combinaison d'outils de modélisation comparative avec Modeller [23] et par repliement *de novo* avec le serveur Robetta, selon la qualité des alignements disponibles. Grâce à ces modélisations, les différentes parties de la protéine MCTP ont été caractérisées, ouvrant la voie à une meilleure compréhension de sa structure et de ses fonctions. Malgré ces avancées, aucun modèle 3D complets des protéines MCTP n’avaient été construits avant ma thèse.

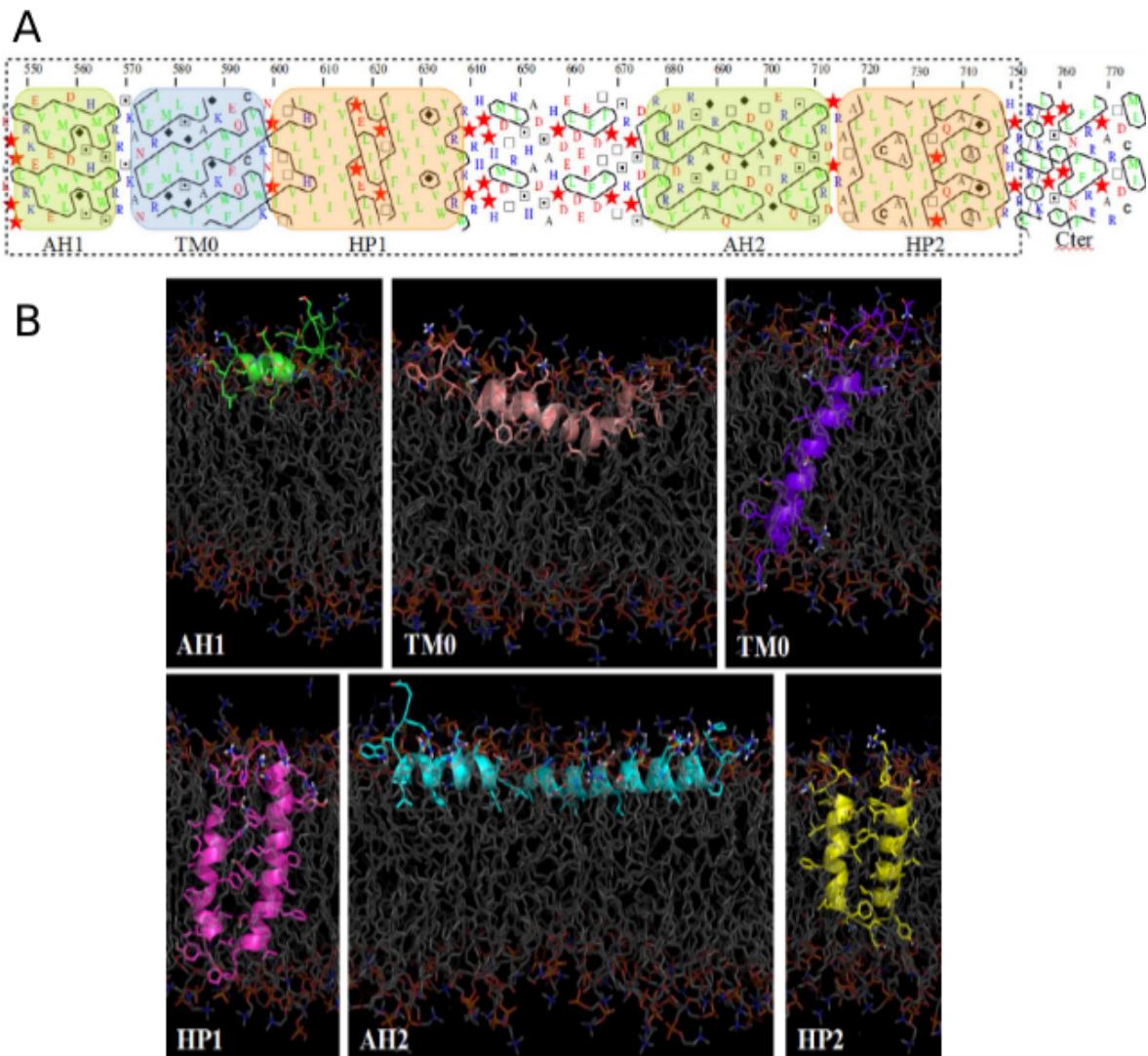


FIGURE 1.9 – A) Analyses sur les hélices qui composent la région transmembranaire de MCTP4. B) Les hélices de la partie transmembranaire en épingle et les hélices au dessus de cette partie ont été simulées en CG et ensuite un backmapping a été réalisé en fin du simulations.

### 1.3.2 Interaction avec deux membranes spécifique

Les protéines MCTP interagissent spécifiquement avec les deux membranes MP et RE : leur région transmembranaire (TMR) s'intègre dans le RE, tandis que leurs domaines C2 se lient à la MP. Cette double interaction est rendue possible par les propriétés uniques (différentes charges, longueurs de chaînes lipidiques ...) de ces membranes. Les lipides chargés négativement de la MP peuvent favoriser l'ancrage des domaines C2, tandis que la structure du RE soutient l'intégration des régions transmembranaires. La présence de lipides anioniques sur la MP est potentiellement un élément de reconnaissance par les protéines MCTP permettant un ancrage dynamique et régulé. De nombreuses protéines qui ont des domaines C2 peuvent interagir de différentes manières les membranes biologiques. Les travaux de Brault et al. chez *Arabidopsis thaliana* ont montré que les domaines C2 chez MCTP4 interagissent individuellement avec des membranes à charge négative, surtout en présence de phosphoinositol-4-phosphate (PI4P) [18]. Les expériences *in vivo* confirment qu'une réduction de PI4P diminue l'attachement des MCTP à la MP, redistribuant ces protéines dans le cytoplasme[18]. De plus, bien que l'étude de Hinderliter et al. soit menée dans un autre système (Synaptotagmins), elle suggère que le motif C2 des protéines membranaires se lie aux phospholipides en réponse au calcium ( $\text{Ca}^{2+}$ ), ce qui pourrait favoriser leur ancrage sur des membranes riches en lipides négatifs[24]. Il est possible que, dans le cas des MCTP, la présence de calcium contribue également à renforcer leur interaction avec les lipides anioniques de la membrane. La reconnaissance des lipides anioniques ou du calcium passe par la structure tridimensionnelle des protéines MCTP. Cependant, ces protéines sont difficiles à purifier et cristalliser. C'est pourquoi nous avons utilisé des outils de modélisation pour prédire la structure 3D des MCTP.

## 1.4 Modélisation structure 3D

### 1.4.1 La modélisation par Homologie

La modélisation par homologie permet de prédire la structure 3D d'une protéine en comparant sa séquence d'acides aminés à celle d'une autre protéine dont la structure est déjà connue. Cette méthode repose sur le principe que la structure 3D des protéines ayant des fonctions importantes est souvent conservée au cours de l'évolution[25]. Ainsi, la construction des modèles de structures 3D par homologie peut être très efficace si la séquence de la protéine d'intérêt présente une similarité supérieure à 30 % avec celle de la protéine modèle, comme par exemple la prédiction hétérodimère complexe formé par Ferredoxin-NADP(+) Reductase (FNR)[26]. Cette approche est facilitée par la disponibilité de grandes bases de données de structures expérimentales accessibles,

telles que la Protein Data Bank (PDB)[27], et par des serveurs qui automatisent la procédure, comme SWISS-MODEL [28].

Le processus de modélisation par homologie est composé de plusieurs étapes, qui impliquent :

1. La sélection de la séquence d'acides aminés de la protéine cible (séquence d'intérêt).
2. L'utilisation d'une matrice de substitution des résidus (par exemple, BLOSUM ou PAM) pour évaluer la probabilité de substitutions entre acides aminés.
3. L'application d'un algorithme d'alignement pour aligner les séquences de manière optimale, à l'aide de programmes tels que BLAST ou FASTA [29]
4. La construction du modèle : Construire le modèle 3D de la protéine cible en utilisant un logiciel de modélisation comme MODELLER, qui génère la structure en fonction de l'alignement de séquences et de la structure connue de la protéine modèle.[23]
5. L'évaluation du modèle : Évaluer la qualité du modèle généré, en vérifiant la géométrie et la conformité avec des critères structuraux (PROCHECK).[30]

Voir figure 1.10

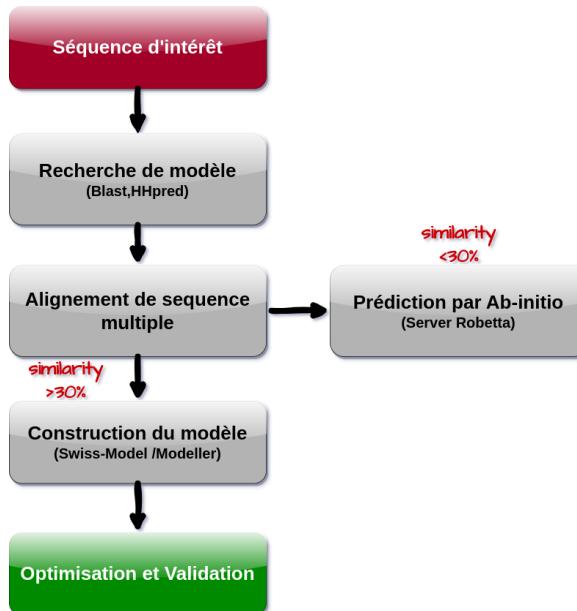


FIGURE 1.10 – Résumé des étapes de la modélisation par homologie

Concernant les protéines MCTP, la similarité des séquences présentes dans la PDB est inférieure à 30 % , la fiabilité des modèles obtenus par homologie est donc très faible[31]. Malgré ces limitations, de nouveaux outils combinant l'homologie et l'apprentissage profond ont récemment émergé, apportant de nouvelles perspectives. Des

méthodes comme AlphaFold et RoseTTAFold utilisent des réseaux de neurones profonds et intègrent des informations évolutives provenant de multiples alignements de séquences pour prédire les structures protéiques avec une précision remarquable. Ces approches ont démontré leur efficacité sur un large éventail de protéines, y compris des protéines membranaires complexes, ouvrant ainsi la voie à une modélisation plus précise des MCTP.

### 1.4.2 Prédiction des structures 3D par méthode de deep learning

Récemment, les avancées en apprentissage profond ont révolutionné le domaine de la prédiction de structure 3D de protéine, notamment avec le développement d'outils tels qu'AlphaFold2 de DeepMind [32], RoseTTAFold [33], OmegaFold [34] et ESMFold [35]. Ces avancées, accessibles au grand public depuis la mise à disposition d'AlphaFold en 2021, ont permis la prédiction de centaines de millions de structures protéiques. En 2024, Demis Hassabis, John Jumper et David Baker, figures clés derrière AlphaFold et RoseTTAFold, ont partagé le prix Nobel, soulignant l'impact scientifique majeur de ces outils. Toutefois, malgré leur succès, il reste important de considérer leurs limites. Leur précision peut diminuer dans le cas des familles de protéines sous-représentées dans les ensembles de données, des protéines transmembranaires, des protéines désordonnées ou des systèmes impliquant des interactions complexes, notamment avec des ions dans la version 3 d'AlphaFold. Nous développerons les limites de ces outils dans le chapitre 2. AF2 a été entraîné sur les séquences d'acide aminé et de structure 3D connus. Et un vocabulaire qui se limite à 20 acides aminés des protéines. Le fonctionnement d'AF2 repose sur un workflow structuré en plusieurs étapes clés. Voir 1.11 Lorsqu'on fournit la séquence d'intérêt dont on ne connaît pas la structure 3D de la protéine, cette séquence va être analysée. Un alignement multiple de séquences (MSA) à partir de bases de données de séquences protéiques est réalisé. Ce MSA, combiné à des bases de données structurelles existantes, fournit des informations essentielles sur les relations évolutives et structurelles entre les résidus. L'architecture d'AlphaFold2 est construite autour d'un réseau attentionnel profond, conçu pour capturer les relations spatiales et contextuelles entre les résidus. Ce réseau intègre les données des MSA et les représente sous forme de matrices de distances et d'angles. Enfin, un module spécifique appelé "structure module" affine ces prédictions pour produire une structure tridimensionnelle finalisée. Ce module combine des techniques d'optimisation géométrique et d'apprentissage automatique pour garantir la cohérence spatiale des structures. AlphaFold3 introduit des améliorations significatives par rapport à ses prédecesseurs, en élargissant son vocabulaire pour inclure non seulement les protéines, mais aussi des molécules telles que l'ARN, l'ADN, ainsi qu'une liste de ligands et d'ions. Ce modèle se distingue par des

changements fondamentaux dans son approche. Contrairement à AlphaFold2, où l’alignement de séquences (MSA) jouait un rôle central, AlphaFold3 réduit son importance, mettant l’accent sur une architecture innovante. Celle-ci intègre un modèle de diffusion, remplaçant le transformer utilisé précédemment. Ces deux versions ont des limitations que nous avons explorées dans le chapitre 2.

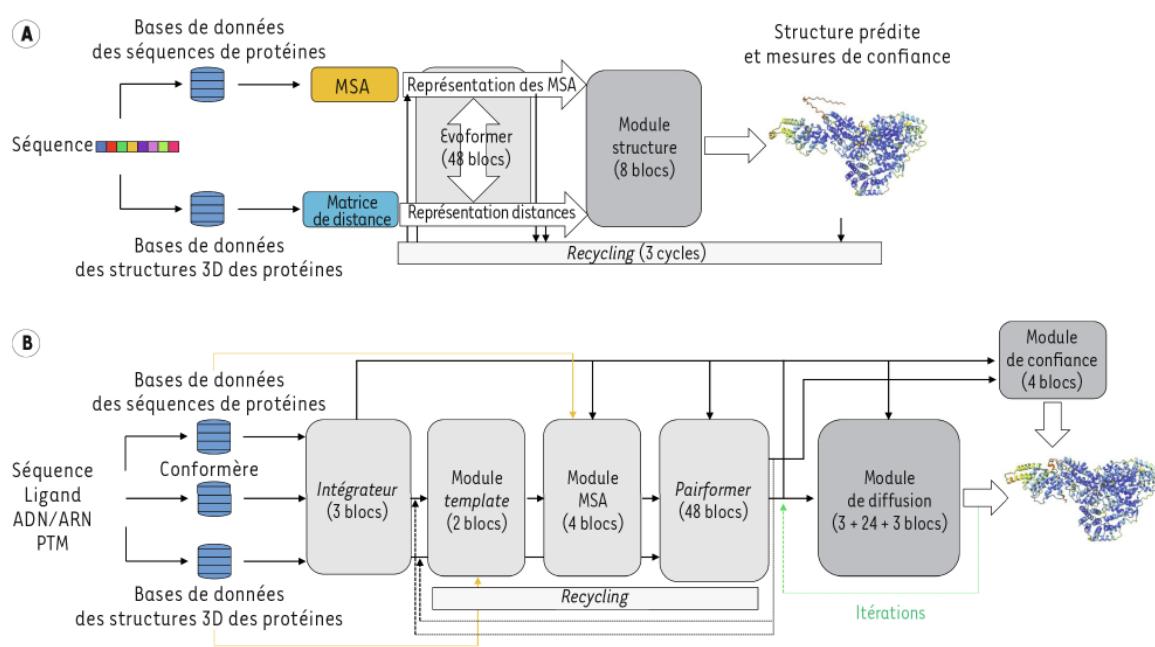


FIGURE 1.11 – A ) Architecture d'alphaFold 2 et B) Architecture d'alphaFold 3 [36]

## 1.5 Dynamique Moléculaire

### 1.5.1 Généralités

Étudier et caractériser expérimentalement le comportement des protéines membranaires, notamment via la cristallographie ou la RMN, n'est pas toujours possible en raison des difficultés liées à leur nature hydrophobe et leur taille. Cependant, à partir d'un modèle structural de la protéine, il est possible de simuler des conditions proches de la réalité en utilisant la dynamique moléculaire, qui repose sur les principes fondamentaux de la physique, notamment les lois de la mécanique newtonienne. Ces simulations permettent de modéliser les mouvements et interactions des atomes en fonction des forces qui les gouvernent, offrant ainsi une compréhension des interactions et des changements conformationnels des protéines. La dynamique moléculaire permet de calculer et de suivre la trajectoire d'une ou d'un ensemble de particules au cours du temps. Pour initialiser la simulation, nous avons besoin de la position initiale du modèle, définie par ses coordonnées cartésiennes ( $x$ ,  $y$ ,  $z$ ), qui peuvent être obtenues,

par exemple, à partir d'une prédiction issue d'outils comme AlphaFold. Les vitesses initiales de chaque particule sont ensuite attribuées de manière aléatoire. En connaissant les forces qui s'exercent sur les particules, nous pouvons résoudre les équations de Newton pour obtenir leur trajectoire en fonction du temps. (Voir figure 1.12)

### — Calcul des Forces

- À partir de l'énergie potentielle  $U(\vec{r})$ , on calcule les forces :

$$\vec{F} = -\nabla U(\vec{r}) \quad (1.1)$$

### — Application des Lois du Mouvement

- En utilisant la deuxième loi de Newton pour une particule de masse  $m$  :

$$\vec{F} = m\vec{a} \quad (1.2)$$

### — Détermination de la Position en Fonction du Temps

- En intégrant l'accélération, on obtient la vitesse  $\vec{v}(t)$  :

$$\vec{v}(t) = \vec{v}_0 + \int_{t_0}^t \vec{a}(t') dt' \quad (1.3)$$

- En intégrant la vitesse, on obtient la position  $\vec{r}(t)$  :

$$\vec{r}(t) = \vec{r}_0 + \int_{t_0}^t \vec{v}(t') dt' \quad (1.4)$$

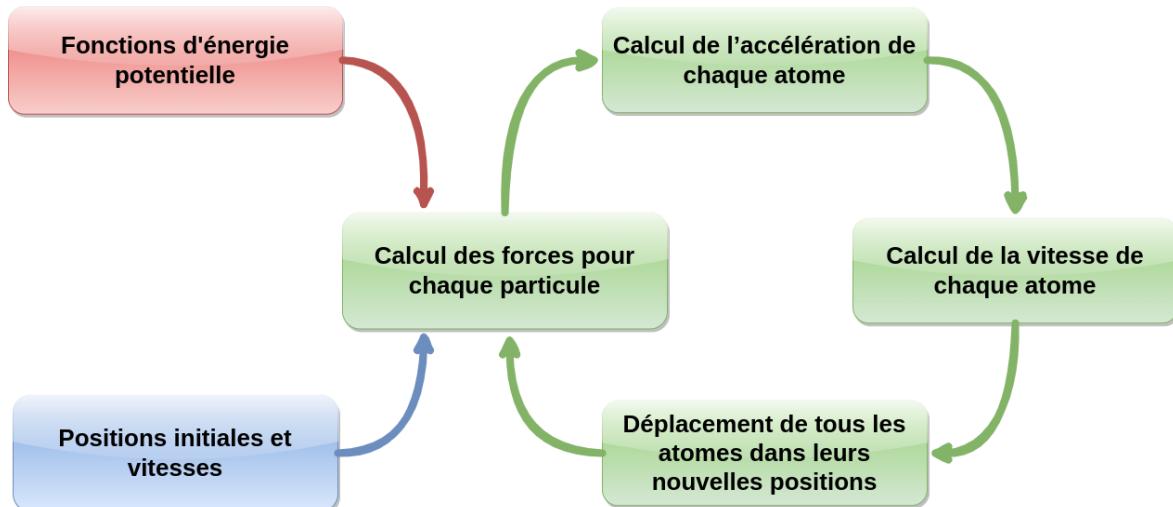


FIGURE 1.12 – Cycle d'intégration en dynamique moléculaire : du potentiel aux nouvelles positions

Pour modéliser les systèmes biologiques complexes, tels que les protéines membranaires, nous devons prendre en compte les différentes interactions et contraintes entre

les particules. Dans les simulations, ces interactions sont paramétrées soit à partir de données expérimentales, soit à partir de calculs *ab initio*, ou bien par une combinaison des deux. Elles sont ensuite décrites par des équations mathématiques, regroupées sous le terme de champ de force, qui permettent de les représenter avec un certain degré de précision.

### 1.5.2 Champs de force

Un champ de force est un ensemble de fonctions mathématiques et de paramètres, issus soit de calculs *ab initio*, soit de données empiriques, ou d'une combinaison des deux, utilisés pour modéliser les interactions au sein d'un système moléculaire, tels que les protéines, les lipides ou d'autres (bio)molécules. Ces fonctions permettent de calculer l'énergie potentielle  $U$  du système, qui correspond à l'énergie stockée dans les interactions entre les particules. Ces interactions se divisent en deux catégories : les interactions liées, celles qui concernent les liaisons covalentes, et qui sont représentées par des valeurs de distances, des angles et des angles dièdres, et les interactions non liées (comme les forces électrostatiques et de Van der Waals).

Potentiel d'énergie totale :

$$U_{\text{potentielle}} = U_{\text{liaison}} + U_{\text{angle}} + U_{\text{dièdre}} + U_{\text{impropre}} + U_{\text{VdW}} + U_{\text{électrostatique}}$$

Potentiel d'énergie de liaison :

$$U_{\text{liaison}} = \frac{1}{2}k_l(l - l_0)^2$$

Où :  $l$  est la distance entre deux atomes liés.  $l_0$  est la distance d'équilibre de la liaison.  $k_l$  est la constante de force de la liaison.

Potentiel d'énergie d'angle

$$U_{\text{angle}} = \frac{1}{2}k_\theta(\theta - \theta_0)^2$$

Où :  $\theta$  est l'angle entre trois atomes.  $\theta_0$  est l'angle d'équilibre.  $k_\theta$  est la constante de force de l'angle.

Potentiel d'énergie de dièdre (torsion)

$$U_{\text{dièdre}} = \frac{1}{2}V_m [1 + \cos(m\omega - \gamma)]$$

Où :  $V_m$  est l'amplitude du potentiel de dièdre.  $m$  est la multiplicité du dièdre.  $\omega$

est l'angle dièdre.  $\gamma$  est le décalage de phase.

Potentiel d'énergie de dièdre impropre

$$U_{\text{impropre}} = \frac{1}{2} k_\omega (\omega - \omega_0)^2$$

Où :  $k_\omega$  est la constante de force du dièdre impropre.  $\omega$  est la valeur actuelle du dièdre impropre.  $\omega_0$  est la valeur d'équilibre.

Potentiel d'énergie de Van der Waals (potentiel d'énergie de Lennard-Jones)

$$U_{\text{VdW}} = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

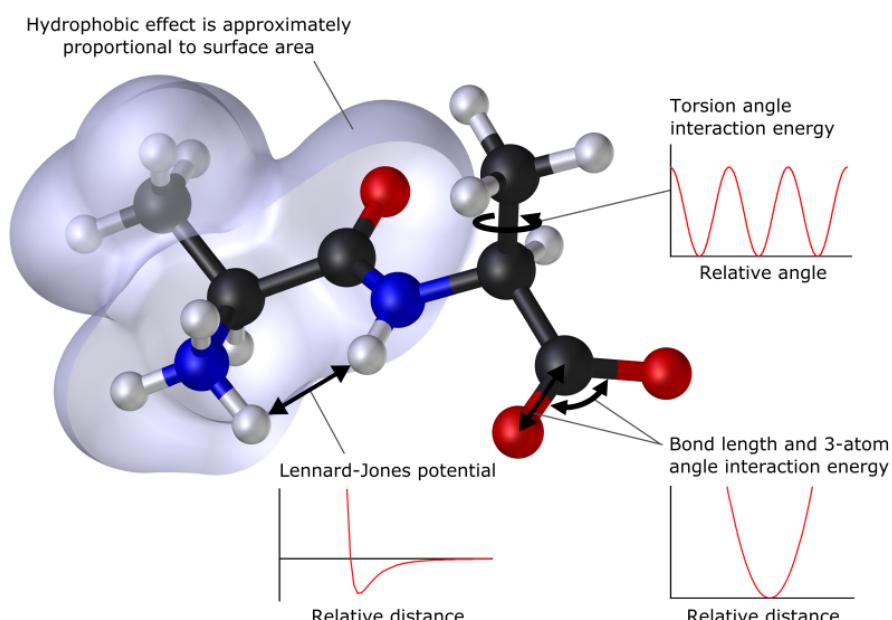


FIGURE 1.13 – Description des potentiels d'énergies des interactions liées et non liées . *Wikipédia :Molecular mechanics*

Une fois que l'on connaît l'énergie potentielle appliquée sur le système, on peut déterminer les mouvements des atomes en résolvant les équations de Newton. Cependant, il n'est pas possible de résoudre ces équations de manière analytique pour un système à N corps ( $N > 2$ ). Or, les systèmes biologiques sont composés de milliers de particules. Néanmoins, on peut les approcher numériquement en utilisant des algorithmes spécifiques avec un pas de temps ( $\delta t$ ) petit et discret. L'algorithme de Verlet et sa version améliorée, le leapfrog, sont de très bons exemples d'intégration. L'équation de l'algorithme de Verlet :

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + a(t)\delta t^2$$

où : -  $\mathbf{r}(t)$  est la position de la particule à l'instant  $t$ , -  $\delta t$  est le pas de temps,

L'algorithme leapfrog calcule les **vitesses** à des instants intermédiaires, c'est-à-dire à  $t + \frac{\delta t}{2}$ , et les **positions** aux instants entiers  $t$ .

$$v\left(t + \frac{\delta t}{2}\right) = v\left(t - \frac{\delta t}{2}\right) + a(t)\delta t$$

Où  $a(t) = F(t)m\dot{a}(t) = mF(t)$ , avec  $F(t)F(t)$  étant la force agissant sur la particule à l'instant  $t$ .

$$r(t + \delta t) = r(t) + v\left(t + \frac{\delta t}{2}\right)\delta t$$

### 1.5.3 Les conditions périodiques aux limites (PBC)

Dans une simulation moléculaire, les particules évoluent dans une boîte de taille définie par l'utilisateur. Pour éviter que les bords de cette boîte n'influencent les trajectoires des particules ou que celles-ci ne s'éloignent indéfiniment, on utilise des **conditions périodiques aux limites (PBC)**. Cela signifie que lorsqu'une particule sort d'un côté de la boîte, elle réapparaît de l'autre côté, créant ainsi un environnement virtuellement infini. Cependant, cette approche peut poser des problèmes pour le calcul des interactions de longue portée, comme les forces électrostatiques et de van der Waals. En effet, les particules peuvent interagir avec leurs propres images périodiques, ce qui peut entraîner des erreurs dans le calcul de l'énergie potentielle. Pour résoudre ce problème, des algorithmes comme **PME (Particle Mesh Ewald)** sont utilisés pour calculer ces interactions de manière précise en tenant compte des conditions périodiques. De plus, pour éviter les artefacts ou les interactions indésirables qui pourraient influencer le comportement d'une protéine simulée, on peut maintenir une distance minimale d'un nm entre la protéine et les bords de la boîte. Cela permet de s'assurer que la protéine n'interagit pas avec ses propres images périodiques, simulant ainsi un environnement plus réaliste.

### 1.5.4 Thermostat et Barostat

Pour simuler une membrane de manière réaliste, il est important de gérer la pression et la température dans la boîte de simulation. Un thermostat est un algorithme qui gère la température. Dans les cellules végétales, la température est de 300 K. Cette température est maintenue dans notre simulation grâce à l'algorithme V-rescale, qui ajuste les vitesses des particules pour atteindre la température cible dans le système avec un

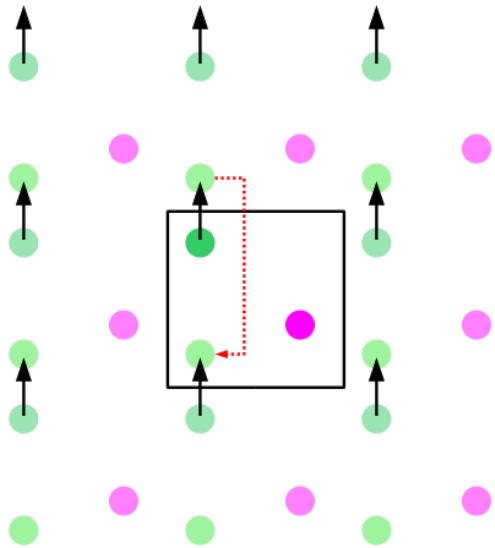


FIGURE 1.14 – Conditions périodiques aux limites (PBC). *Image issue de Wikipédia*

barostat adapté. Un barostat est un algorithme qui permet de maintenir une pression cible en ajustant le volume de la boîte de simulation. Pour reproduire fidèlement le comportement d'une membrane, la pression doit être gérée de manière semi-isotrope afin de maintenir une tension de surface réaliste. Ainsi, la pression sur les axes **x** et **y** (dans le plan de la membrane) est contrôlée de la même manière, tandis que la pression sur l'axe **z** (perpendiculaire à la membrane) est générée différemment pour permettre à l'épaisseur de la membrane de fluctuer naturellement sans contraintes artificielles. L'algorithme de barostat choisi est celui de Parrinello-Rahman.

## 1.6 Simulation Coarse-Grained (CG)

Décrire l'ensemble des caractéristiques d'un système présente des limites, notamment en ce qui concerne la taille du système : plus celui-ci est grand, plus il demandera des ressources de calcul pour résoudre les équations au cours du temps. Selon la stratégie choisie pour étudier le système, différentes approches peuvent être utilisées, comme la simplification de la représentation du système atomique en modèle "gros grain".

### 1.6.1 Généralités

Les simulations de dynamique moléculaire à CG offrent une approche efficace pour étudier des systèmes biomoléculaires complexes. En simplifiant la représentation des molécules, elles permettent de simuler des phénomènes à grande échelle, tant en taille qu'en durée, tout en minimisant les ressources de calcul. Dans cette thèse, l'utilisation du champ de force Martini a été privilégié.

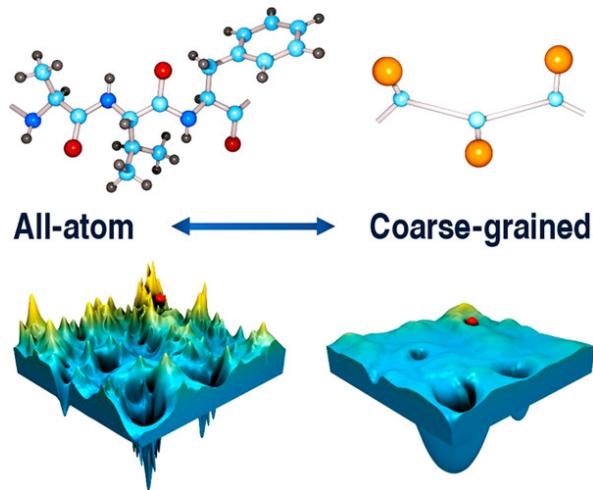
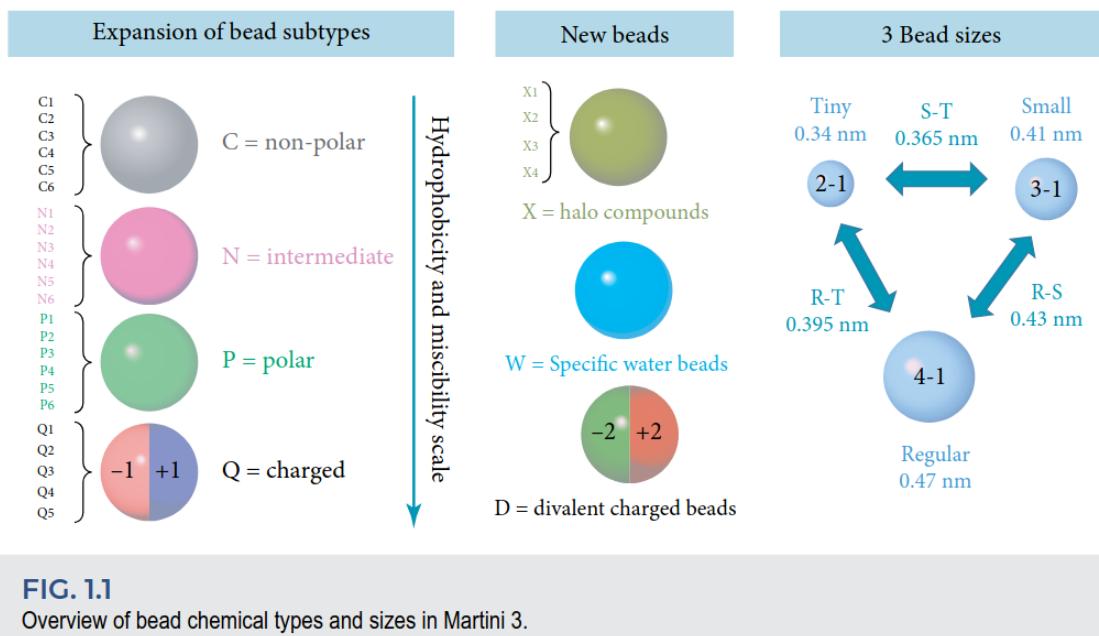


FIGURE 1.15 – Lissage du paysage conformationnelle avec un champs de force gros grain. Source : [37]

Le développement de MARTINI a commencé avec la modélisation des chaînes lipidiques des membranes par Marrink et al [38]. Les lipides sont des structures qui se composent d'une tête polaire et d'une chaîne lipidique à deux ramifications. Cependant, simuler de grands assemblages membranaires avec une précision atomique demande des ressources de calcul considérables. C'est pourquoi la simplification de la représentation atomique en "grains" a été introduite. Cette approche permet de réduire significativement la taille des calculs tout en augmentant le pas de temps (timestep) des simulations, rendant ainsi possible l'exploration plus large des conformations du système (Voir figure 1.15). MARTINI 2 a ensuite étendu son champ d'application aux systèmes protéiques [39], mais a révélé certaines limitations, notamment la surestimation des interactions entre les protéines et les lipides, ainsi qu'une tendance à agréger les systèmes, ce qui biaise certaines simulations [40]. Ce phénomène, décrit comme le problème de « stickiness » a été observé dans divers contextes, notamment pour les protéines solubles et membranaires ainsi que pour les glucides [41, 40]. Les travaux de Alessandri et al., 2019 ont identifié l'origine du problème, un mauvais paramétrage des petites (S) et moyenne (R) billes, avec des distances de liaisons trop faibles dans la description des biomolécules [42]. Pour pallier ces problèmes, MARTINI 3 a été développé en reparamétrisant de nombreux types de particules, afin de mieux représenter la diversité chimique des lipides et des protéines [43]. Cette version a montré son utilité pour modéliser des systèmes biologiques complexes, comme l'interaction du SARS-CoV-2 avec des membranes cellulaires[44]. Néanmoins, malgré ces avancées, le champ de force MARTINI reste un modèle simplifié qui nécessite souvent des ajustements pour des simulations plus spécifiques.

**Martini 3**

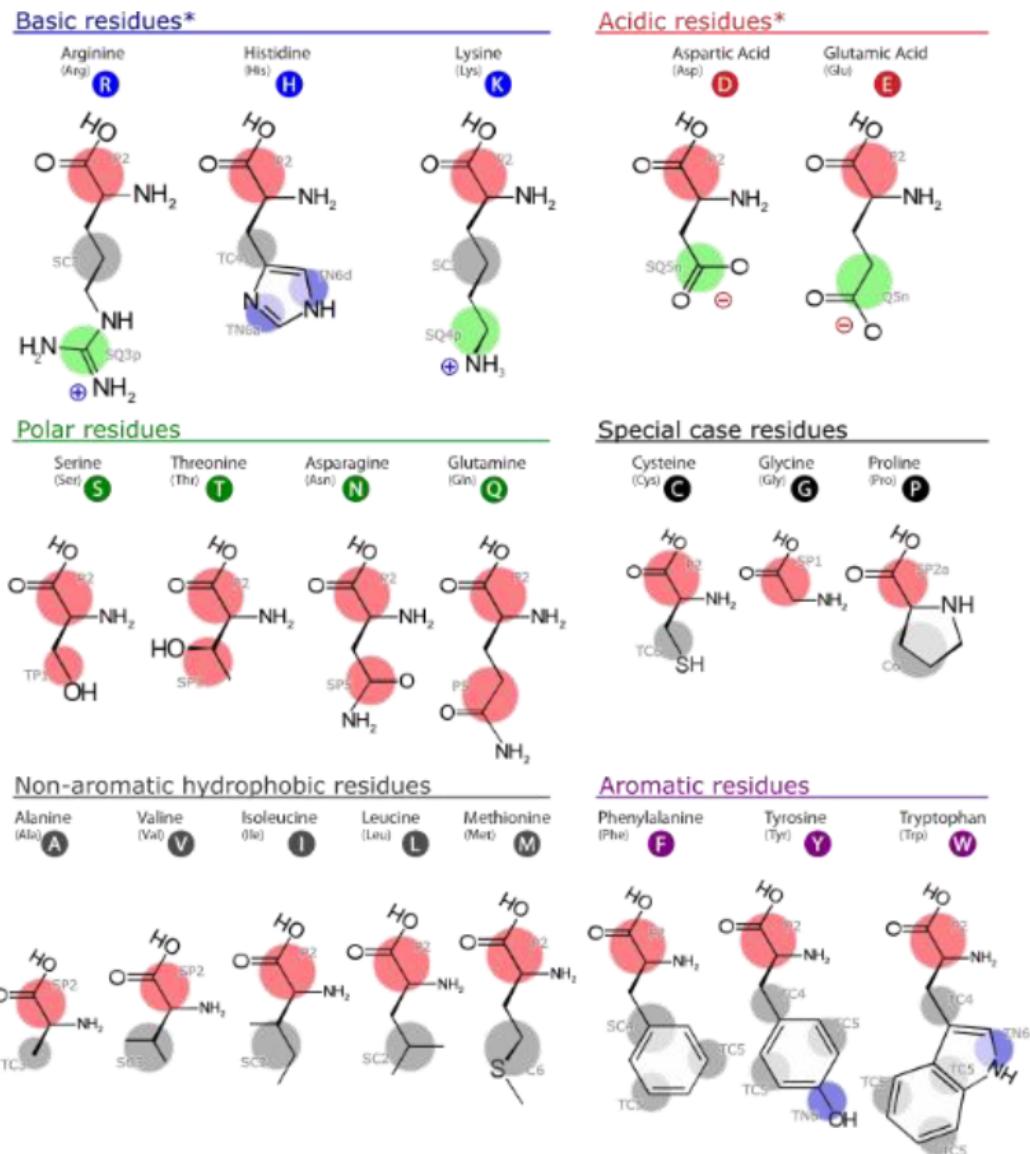
Dans la version Martini 3, les atomes sont définis selon des billes spécifiques. Il y a sept catégories de billes chimiques : polaire (P), intermédiaire/non-polaire (N), apolaire (C), halo-composés (X), ions monovalents (Q), ions divalents (D) et eau (W). Chacune de ces billes existe en plusieurs tailles différentes : Taille S (petite), Taille R (standard) et Taille T (très petite). Les billes comprennent des sous-types différenciés par des numéros, qui indiquent leur niveau de polarité relative, allant de 1 (faible polarité) à un maximum de 6 (haute polarité). Les billes W et D, n'ont pas de sous type. Voir Figure 1.16.

**FIG. 1.1**

Overview of bead chemical types and sizes in Martini 3.

FIGURE 1.16 – Définitions des billes dans Martini

Par exemple, pour le résidu aspartate, nous trouvons les billes P2 et SQn5. Comme ce résidu possède des propriétés acides et est chargé négativement dans un environnement à pH 7, il sera composé de 2 billes : une bille pour le squelette carboné (P2) et une bille chargée négativement pour le groupement carboxylate (SQn5). Les différents acides aminés sont décrits dans la Figure 1.17 et le Tableau 1.1.



\* Most common protonation state is shown.

FIGURE 1.17 – Topologie des acides aminés avec Martini 3

Amino Acid	Beads
GLY	SP1
ALA	SP2, TC3
CYS	P2, TC6
VAL	SP2, SC3
LEU	P2, SC2
ILE	P2, SC2
MET	P2, C6
PRO	SP2a, SC3
HYP	P1, P1
ASN	P2, SP5
GLN	P2, P5
ASP	P2, SQ5n
GLU	P2, Q5n
THR	P2, SP1
SER	P2, TP1
LYS	P2, SC3, SQ4p
ARG	P2, SC3, SQ3p
HIS	P2, TC4, TN6d, TN5a
PHE	P2, SC4, TC5, TC5
TYR	P2, TC4, TC5, TC5, TN6
TRP	P2, TC4, TN6d, TC5, TC5, TC5

TABLE 1.1 – Représentation des acides aminés en billes dans Martini 3

La matrice d’interaction dans Martini 3 est un tableau qui définit comment différents types de billes (ou beads) interagissent entre elles. Chaque bead représente un groupe d’atomes, et la matrice attribue à chaque paire de beads une force d’attraction ou de répulsion, déterminée par des paramètres appelés  $\varepsilon_{ij}$  (intensité de l’interaction) et  $\sigma_{ij}$  (taille des beads). Plus  $\varepsilon_{ij}$  est grand, plus les beads s’attirent fortement. Dans Martini 3, il existe 22 niveaux d’interaction, ce qui permet une modélisation plus précise des différentes interactions chimiques, comme celles entre lipides, protéines ou solvants. Voir Figure 1.18.

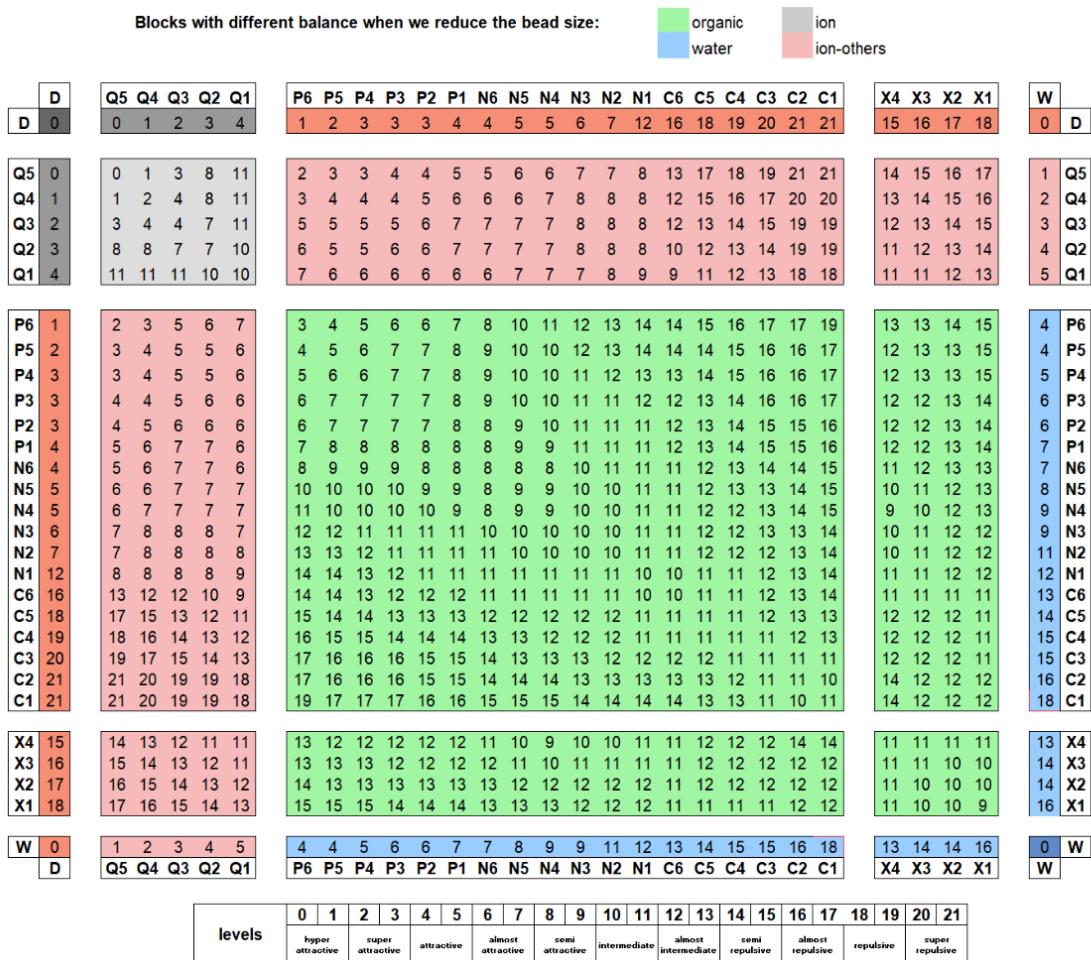


FIGURE 1.18 – Matrice d’interactions entre les billes dans Martini 3.

### 1.6.2 Membrane en Martini

Dans les membranes des plantes, on trouve principalement des lipides qui présentent une tête polaire hydrophile attachée à un squelette de glycérol, ainsi qu’une queue hydrophobe composée de deux acides gras. Ces lipides forment une barrière hydrophobe, isolant les cellules et les organites de leur environnement. Les acides gras sont essentiellement constitués d’une chaîne d’hydrocarbures terminée par un groupe carboxyle (-COOH). En fonction de la longueur de leur chaîne carbonée, ils sont classés en :

- **Acides gras à chaîne courte** : moins de 6 carbones,
- **Acides gras à chaîne moyenne** : de 6 à 12 carbones,
- **Acides gras à chaîne longue** : de 13 à 21 carbones,
- **Acides gras à très longue chaîne** : 22 carbones ou plus.

Différents types de lipides assurent des fonctions essentielles dans la membrane. Par exemple, le sitostérol apporte de la fluidité à la membrane. Le phosphatidylinositol-4-

phosphate (PI4P) est le principal lipide anionique qui détermine l'identité électrostatique de la membrane [45]. La phosphatidylcholine (PC) et la phosphatidylsérine (PS) sont les constituants majoritaires de la MP. Martini dispose d'une large base de données de modèles de lipides. Nous avons pu utiliser plusieurs lipides déjà paramétrés comme POPC, DOPC, PIPE, PIPC et PI4P. Voir tableau 1.2 . Cependant, deux composés étaient manquants dans Martini 3 : le sitostérol et un lipide PS (18 :2 24 :1). Pour le lipide PS, nous avons réussi à créer sa topologie par analogie, en nous basant sur deux autres lipides existants dans Martini 3 (PIPS et DNPS). Concernant le modèle du sitostérol, bien que nos tentatives initiales de paramétrisation n'aient pas abouti, nous avons pu obtenir le modèle auprès de l'équipe de développement de Martini 3, notamment grâce à Paolo Souza et Luis Borges. La paramétrisation du cholestérol, réalisée par Luis Borges, a servi de base à ce travail \cite{borges-araujo\_improved\_2022}. L'obtention de ce modèle nous a permis de réaliser des simulations avec des compositions réalistes des membranes du RE RE et de la MP[46].

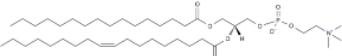
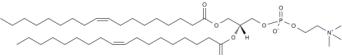
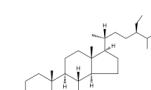
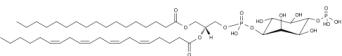
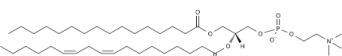
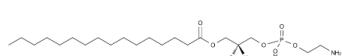
Lipide	Structure chimique	Longueur de chaîne
POPC		16 :0/18 :1
DOPC		18 :1/18 :1
Sitostérol		
PI4P		17 :0/20 :4
PIPC		16 :0/18 :2
PIPE		16 :0 18 :2
PS (PIPS + DNPS) image non dispo		18 :2 /24 :1

TABLE 1.2 – Tableau des lipides associés à la MP et au RE

### 1.6.3 Martini 3 : Un modèle élastique gros-grain pour la simulation des protéines

Dans les simulations de dynamique moléculaire gros grains utilisant le champ de force Martini, la stabilité structurale des protéines est un défi majeur. Cela est principalement dû à l'absence de représentation explicite des liaisons hydrogène, qui jouent un rôle crucial dans le maintien de la structure tertiaire des protéines. Pour surmonter cette limitation, il est essentiel d'incorporer un réseau de contraintes supplémentaires qui soutient la structure protéique. L'approche principale utilisée à cet effet est le modèle élastique (Elastic Network Model). Le **modèle élastique** consiste à introduire des liaisons harmoniques entre les billes du squelette protéique (*BB*) pour maintenir la conformation native. Une liaison élastique est établie entre deux billes du squelette si :

- Les résidus correspondants sont séparés par au moins deux résidus dans la séquence,
- La distance entre les billes du squelette (*d*) est inférieure à une distance de coupure

Rc, généralement comprise entre 0,8 et 1,0 nm.

Ces liaisons sont modélisées par un potentiel harmonique défini par :

**Potentiel harmonique pour les liaisons élastiques :**

$$U(d) = \frac{1}{2} \times k \times (d - d_0)^2$$

où :

- $U(d)$  est l'énergie potentielle de la liaison.
- $k$  est la constante de force (typiquement entre 500 et 1000  $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{nm}^{-2}$ ).
- $d$  est la distance actuelle entre les billes du squelette.
- $d_0$  est la distance d'équilibre (généralement la distance initiale entre les billes).

Ainsi le **modèle élastique** est simple à implémenter et efficace pour maintenir la structure globale de la protéine.

Dans le cadre de mes travaux sur la modélisation des MCTP, l'utilisation de ces approches est cruciale pour assurer la stabilité des protéines lors des simulations CG. Par exemple, pour la modélisation de la région transmembranaire de MCTP4, le maintien de la structure hélicoïdale est essentiel pour étudier correctement ses interactions avec la membrane. En appliquant un réseau élastique avec des paramètres appropriés, nous pouvons préserver la conformation native tout en permettant des fluctuations réalistes.

## 1.7 OBJECTIF DE LA THÈSE

Cette thèse vise à explorer comment la protéine MCTP4 interagit avec les deux membranes qui composent les plasmodesmes, en utilisant des approches de modélisation moléculaire. L'objectif est de mieux comprendre comment la famille MCTP structure ces jonctions intercellulaires spécialisées chez les plantes. Ce travail s'inscrit dans le cadre du projet ANR DIVCON, au sein duquel notre groupe de modélisation cherche à élucider le rôle des protéines MCTP dans l'organisation des plasmodesmes. Pour cela, nous réalisons des prédictions structurelles et des simulations de dynamique moléculaire afin de modéliser ces protéines dans des conditions proches de leur état naturel. Ce projet est issu d'une collaboration entre le Laboratoire de Biochimie Théorique, dirigé par Antoine Taly, où se déroule cette recherche, et l'équipe expérimentale dirigée par Emmanuelle Bayer au sein du Laboratoire de Biogénèse Membranaire (CNRS et Université de Bordeaux).

### Chapitre 2

Dans ce chapitre, nous explorons la prédition des structures 3D des protéines MCTP à l'aide des outils de deep learning, en mettant en évidence leurs forces et leurs limites. Afin d'avoir une analyse croisée de ces approches, nous avons organisé un workshop réunissant expérimentateurs et théoriciens. Lors de cet événement, nous avons discuté de divers systèmes, qu'il s'agisse de protéines désordonnées, d'oligomères ou de protéines membranaires. Ce workshop a permis non seulement de présenter nos résultats, mais aussi de favoriser des discussions approfondies grâce à des sessions de visualisation 3D impliquant divers experts. En particulier, nous avons étudié l'utilisation de méthodes complémentaires telles que DREAMM, un outil de prédition des régions transmembranaires, MembraneFold/DeepTMHMM, ainsi que la dynamique moléculaire afin d'améliorer la robustesse de nos modèles. L'évaluation critique des prédictions obtenues a révélé plusieurs limitations des approches actuelles, notamment en ce qui concerne la modélisation des domaines transmembranaires. Pour pallier ces lacunes, nous avons adopté une approche combinatoire intégrant plusieurs méthodes de prédition et en croisant les résultats obtenus avec des données expérimentales. Cette stratégie nous a permis d'améliorer la fiabilité des modèles et d'identifier des pistes pour optimiser les futures prédictions des structures des MCTP. Cette stratégie sera développée dans le chapitre 3.

### Chapitre 3

Bien que de nombreuses espèces expriment des protéines MCTP, aucune structure 3D expérimentale complète n'a été déterminée à ce jour. Notre objectif dans ce travail était donc de proposer un modèle fiable de MCTP4. Nous avons d'abord utilisé

AlphaFold2 pour prédire un modèle de cette protéine. Cependant, l'incertitude concernant la partie transmembranaire a révélé les limites de cette approche. Pour obtenir un modèle plus robuste, nous avons adopté une approche combinée, intégrant des outils de deep learning et des simulations de dynamique moléculaire afin d'affiner la structure transmembranaire de MCTP4. Pour cela, nous avons utilisé le champ de force coarse-grain Martini 3, qui permet d'explorer la flexibilité et l'insertion des domaines transmembranaires dans la membrane. Cette approche nous a permis de comprendre que, individuellement, les outils de deep learning ne permettent pas d'explorer un ensemble de conformations possibles, mais qu'ils offrent néanmoins des points de départ pertinents pour les simulations. Grâce aux simulations, nous avons également pu observer des conformations non identifiées par les outils de deep learning, soulignant ainsi la complémentarité entre ces deux approches.

## **Chapitre 4**

L'ancrage des protéines à la membrane est un processus clé dans leur fonction et leur régulation. Pour les MCTP, la présence de phosphatidylinositol-4-phosphate (PI4P) dans la MP pourrait jouer un rôle déterminant dans leur ciblage et leur dynamique d'interaction. Nous avons cherché à comprendre comment les domaines C2 des MCTP interagissent avec la MP enrichie en PI4P et si ces lipides modulent leur affinité et leur stabilité. Pour répondre à cette question, nous avons collaboré avec l'équipe du LBM afin de croiser données expérimentales et simulations. Cette approche nous a permis d'analyser les interactions spécifiques entre PI4P et les domaines C2 et d'explorer leur impact sur la dynamique du plasmodesme.

## **Chapitre 5**

Certains domaines C2 sont connus pour interagir avec la membrane en présence de calcium, modulant ainsi leur affinité pour certains lipides. Cependant, nous ne savons pas si MCTP interagit avec le calcium et si cela influence la liaison des MCTP à la membrane via une compétition avec le PI4P. Nous avons adopté une approche combinant simulations de dynamique moléculaire et proposer des mutant pour déstabiliser l'interaction avec le calcium. Nos simulations permettront d'évaluer si la présence de calcium modifie les préférences de liaison des MCTP, en particulier leur affinité pour le PI4P. Une éventuelle compétition entre ces deux éléments pourrait avoir des implications fonctionnelles majeures, influençant le ciblage membranaire des MCTP et leur rôle dans la signalisation cellulaire.