



AIX-MARSEILLE UNIVERSITÉ
ECOLE DOCTORALE DES SCIENCES DE LA VIE
ET DE LA SANTE
UMR 7255 - LISM - Laboratoire d'Ingénierie des
Systèmes Macromoléculaires

Thèse présentée pour obtenir le grade universitaire de
Docteur

Discipline : Biologie
Spécialité : Biochimie structurale

Marlon SIDORE

Etude de l'énergétique de l'assemblage des
protéines membranaires

Energetics of the assembly of membrane proteins

Sous la direction de Jean-Pierre DUNEAU et James STURGIS

Soutenance le 10/12/2018 devant le jury composé de :

Luca MONTICELLI	Rapporteur
Marc BAADEN	Rapporteur
Cécile BREYTON	Examinatrice
Rainer HIENERWADEL	Examineur
Norbert GARNIER	Examineur
Jean-Pierre DUNEAU	Directeur de thèse
James STURGIS	Directeur de thèse

Résumé

Les protéines membranaires occupent en moyenne 50 % de la masse des membranes cellulaires. Cependant, certaines membranes spécialisées peuvent avoir de 20 à 90 % de leur masse en protéines, telles que les membranes du cristallin ou les membranes qui hébergent certains appareils photosynthétiques. Dans ce cadre, l'importance de l'assemblage des protéines membranaires dans des complexes cohérents, dynamiques et fonctionnels n'est plus à démontrer.

Mon projet s'inscrit dans la compréhension des forces qui mènent à l'assemblage des protéines membranaires. J'utilise pour cela le modèle de l'Aquaporine Z (AqpZ) d'*Escherichia coli* qui est étudié expérimentalement au laboratoire et est idéal de part sa structure homotétramérique très conservée dans un contexte de divergence de séquence élevée. Par ailleurs, cette famille de protéines réalise ses fonctions de transport d'eau et de petits solutés au sein d'une très grande variété de membranes hôtes.

Sur ce modèle, des approches de dynamique moléculaire ont pu mettre en évidence des perturbations de la structure des lipides qui pourraient être à l'origine de forces s'exerçant à longue distance. J'utilise en premier lieu une approche de dynamique moléculaire gros grains avec des forces de biais adaptatifs pour étudier les relations entre orientations de deux monomères d'AqpZ. Il existe, de façon surprenante, des forces se propageant à longue distance vraisemblablement par les lipides qui biaisent les orientations relatives entre les protéines.

Un deuxième axe de mon travail est l'étude des enrichissements lipidiques autour de l'AqpZ native ou mutée, à différentes distances, avec l'utilisation d'une membrane complexe rendant compte de la diversité lipidique de la membrane interne d'*E.coli*. Dans cette analyse, la cardiolipine est enrichie à proximité de la protéine. Les résultats obtenus sont comparés avec des analyses de fractions lipidiques qui entourent l'aquaporine Z lors d'extraction dans des nanodisques de Styrene Maleic Acid (SMA). Dans la continuité de cet axe, j'ai développé un polymère gros grains rendant compte des propriétés physico-chimiques du SMA pour interroger l'influence du polymère sur ces enrichissements.

Enfin, j'ai construit un système contenant 125 monomères d'AqpZ dans une membranes simple ou complexe, qui représentent 50 % en masse en protéines. Ce système m'a permis de questionner l'évolution spontanée d'un tel système encombré, mais aussi le devenir des forces à longue distance et des lipides enrichis à la surface de la protéine dans ce contexte.

Abstract

Membrane proteins represent on average 50 % of the mass of cellular membranes. However, specialized membranes can have from 20 to 90 % of their mass in proteins, such as lens or photosynthetic membranes. In this context, the importance of the assembly of membrane proteins in coherent, dynamic and functional complexes isn't to be proven anymore.

The goal of my project is to understand the different forces that lead to the assembly of membrane proteins. For this aim, I am using the Aquaporin Z (AqpZ) model protein from *Escherichia coli*, which is studied in our laboratory. This protein is ideal : it has a homotetrameric structure that is very conserved in a context of heavy sequence divergence. In addition, this protein family performs its functions of transporting water and small solutes in a wide variety of host membranes.

On this model, molecular dynamics approaches have been able to highlight lipid structure perturbations that could be responsible for long-range forces. First, I use a coarsed grain molecular dynamics approach with adaptive biasing forces to study the relations between orientations of two AqpZ monomers. Surprisingly, there are forces propagating at long distance, presumably by the lipids which in turn bias the relative orientations between the proteins.

The second axis of my work is the study of lipid enrichments around native or mutated AqpZ, at different distances, with the use of a complex membrane accounting for the lipid diversity of the inner membrane of *E.coli*. In this analysis, cardiolipin is enriched near the protein. The results obtained are then compared with analyses of the lipid fractions that surround AqpZ during extraction in nanodisks of Styrene Maleic Acid (SMA). In the continuity of this axis, I developed a coarsed grain polymer accounting for the physico-chemical properties of SMA to examine the influence of the polymer on these enrichments.

Finally, I built a system containing 125 AqpZ monomers in a simple or complex membrane, which represents 50 % protein by weight. This system allowed me to examine the spontaneous evolution of such a crowded system, but also to investigate the fate of the long distance forces and the lipid enrichments at the protein surface in this context.

Remerciements

La thèse de doctorat représente un moment privilégié qui se situe entre la vie étudiante et l'apprentissage des codes de la recherche. De nombreuses personnes se sont alors retrouvées au milieu de cette période de flou, pour le meilleur ou pour le pire. Ce sont certaines de ces personnes, qui m'ont supportés durant ces trois ans, que je voudrais remercier plus particulièrement.

Je voudrais tout d'abord remercier mes directeurs et co-directeurs de thèse, *James Sturgis* et *Jean-Pierre Duneau*, pour leur confiance renouvelée à la suite de mon master. James, merci de m'avoir permis d'effectuer ma thèse dans ton laboratoire et de m'avoir fait collaborer sur le code des simulations de Monte-Carlo. Jean-Pierre, merci de m'avoir supporté et de m'avoir fait confiance tout en me laissant une grande liberté. Merci pour tout ce que j'ai appris à vos cotés.

Une thèse n'est pas une thèse sans un jugement par les pairs. Je souhaiterais remercier *Luca Monticelli* et *Marc Baaden* d'avoir accepté la charge d'être rapporteur de thèse et *Cécile Breyton*, *Rainer Hienerwadel* et *Norbert Garnier* d'avoir accepté de faire partie du jury de thèse.

Ce travail a de plus été soutenu par le GENCI-CINES pour l'utilisation du supercalculateur OCCIGEN (Numéro de grant C2015077044, X201607744, A0010707044 et A0030707044), le CNRS et Espoir contre la Mucoviscidose.

Je souhaite remercier les collaborateurs qui m'ont permis d'effectuer des collaborations sur des sujets parfois éloignés du mien, mais aussi de comprendre un peu mieux les rouages de l'interdisciplinarité, *Sinan Al'Attar*, *Axel Magalon*, *Bruno Guigliarelli*, *Frédéric Biaso*, *Stéphane Grimaldi*, *Julia Rendon*, *Eric Cascales*, *Eric Durand* et *Camille Garcia*.

Je remercie aussi le reste de l'équipe "Assemblage et dynamique des protéines membranaires", Valérie, Baptiste, Marcel pour tes discussions politiques à table, mais aussi ceux qui ont quitté l'équipe, en particulier Victoria pour ces pauses café à évacuer notre frustration commune des reviewers et pour avoir partagé une protéine commune (et pour ton mutant!).

Je remercie plus particulièrement Gauthier pour ces années de débauche à discuter de science et Camille pour tes blagues que j'ai du supporter durant ma dernière année de thèse et pour l'environnement de travail "My Little Poney" pendant ton M2.

Je remercie aussi l'ensemble du LISM pour leur accueil et leur bonne humeur.

Enfin, ces remerciements ne seraient pas complets si je ne mentionnais pas ma mère. Merci d'avoir été patiente avec moi durant cette thèse.

Table des matières

1	Introduction	1
1.1	Avant-propos	1
1.2	Approche historique	1
1.2.1	Pré-XX ^{ème} siècle : Balbutiements : la membrane comme hypothèse inutile	1
1.2.2	Jusqu'à la bicouche : d'une simple couche de lipide à une (bi ?)couche lipidique	4
1.2.3	Une (bi ?)couche de lipides entre deux couches de protéines	5
1.2.4	De l'utilisation du microscope électronique à la mosaïque fluide	7
1.2.5	La mosaïque fluide	8
1.3	La membrane biologique : le modèle actuel	10
1.3.1	La mosaïque fluide, un modèle qui a pu intégrer ses critiques depuis 1972	10
1.3.2	Diffusion anormale et aspects moléculaires	11
1.3.3	Interactions à longue distance	13
1.3.4	Mon projet	14
2	Matériel et méthodes	17
2.1	La dynamique moléculaire	17
2.1.1	Principe général	17
2.1.2	Intégration des lois de Newton	18
2.1.3	Contrôle de la température	18
2.1.4	Contrôle de la pression	19
2.1.5	Les fonctions d'interaction	19
2.1.6	Différents types de champs de forces, à différentes échelles	21
2.2	Construction des systèmes	22
2.2.1	Systèmes dans NAMD	23
2.2.2	Systèmes dans GROMACS	24
2.3	Colvars	25
2.4	Intégration thermodynamique	26
2.4.1	L'Adaptive Biasing Force (ABF)	27
2.4.2	Umbrella Sampling	29
2.4.3	Dynamique moléculaire dirigée	30
2.5	Paramètres des simulations	30
2.5.1	MARTINI dans NAMD	30

2.5.2	MARTINI dans GROMACS	30
2.6	Automatisation et Analyses	31
2.6.1	Gestion des fichiers .pdb, .gro et .ndx en python	31
2.6.2	Post-traitement des trajectoires	31
2.6.3	Calcul de l'angle de spin	31
2.6.4	Passage d'un PMF tridimensionnel à un PMF unidimen- sionnel	32
2.6.5	Densités de lipides	32
2.6.6	Comptage des lipides autour de la protéine	32
2.6.7	Temps de résidence des lipides	33
2.6.8	Création d'un polymère gros grain amphiphile	33
2.6.9	Analyse d'une membrane encombrée	33
2.6.10	Développement d'un logiciel de simulation de Monte-Carlo	35
2.6.11	Fonctions de distribution radiale	35
3	Thermodynamique de l'assemblage d'un dimère d'AqpZ	36
3.1	Résultats	37
3.1.1	Deux protéines censées interagir dans une membrane simple	38
3.1.2	Orientations préférentielles à distance constante	40
3.1.3	Un ABF à trois dimensions	41
3.1.4	Comparaison avec un PMF à une dimension	43
3.2	Discussion	45
4	Relations entre la protéine et les lipides qui l'entourent - annulus et tri des lipides	48
4.1	Résultats	49
4.1.1	Une membrane plus "réaliste"	49
4.1.2	Tri des lipides autour de la protéine	51
4.1.3	Préférences lipidiques à la surface de la protéine	53
4.2	Discussion	57
5	Devenir de ces forces dans une membrane encombrée	60
5.1	Résultats	61
5.1.1	Structuration d'un réseau : des chaînettes de protéines .	61
5.1.2	Devenir des orientations relatives entre dimères	62
5.1.3	Préférences lipidiques à la surface de l'Aquaporine Z . .	63
5.1.4	Modèle de Monte-Carlo : de très gros grains	65
5.1.5	Flux de lipides entre les protéines	67
5.2	Discussion	67
6	Conclusion	69
7	Utilisation du gros grain avec l'ABF, dans NAMD	73
7.1	Résultats	74
7.2	Discussion	77

8	Modélisation d'un polymère amphiphile, mimant les propriétés du copolymère de SMA	79
8.1	Résultats	79
8.1.1	Rapprocher la simulation de l'expérience : développement d'un polymère gros grain	79
8.1.2	Trouver l'hydrophobicité parfaite : solubiliser une membrane avec un polymère	83
8.1.3	Un nanodisque	85
8.1.4	Enrichissements différentiels dans le nanodisque	85
8.2	Discussion	87
9	Investigation d'un crash : pourquoi MARTINI n'aime pas NAMD	89
9.1	MARTINI dans NAMD ou GROMACS : différences ?	89
9.2	Etude de cas	90
9.3	Discussion et conclusion	90
9.3.1	Injection d'énergie dans le système	92
9.3.2	Possibles solutions	92
10	Collaboration : Dimérisation de segments transmembranaires du système de sécrétion de type VI	93
10.1	Matériels et Méthodes	95
10.2	Résultats	97
10.3	Discussion	99
11	Collaboration : Découverte d'un canal à eau dans NarGH	101
11.1	Matériels et méthodes	102
11.2	Résultats	104
11.3	Discussion	109
11.3.1	pK_a et acidité des résidus autour du site actif	111
11.3.2	Implication du Glu-581 et de l'Asp-801 dans l'ouverture et la fermeture du canal	111
	Annexes	125
A	Lois de Newton	126
B	Calcul de l'angle de spin	129

Table des figures

1.1	Micrographie de liège, par Robert Hooke	2
1.2	Tension de surface, de Rayleigh	4
1.3	Modèle paucimoléculaire	6
1.4	Théorie des plaques	7
1.5	La mosaïque fluide	8
1.6	Mismatch hydrophobe	12
2.1	Workflow simplifié d'une simulation de dynamique moléculaire	17
2.2	Potentiels de liaison	20
2.3	Représentation du passage à l'échelle gros grain	22
2.4	Les variables collectives	26
2.5	Mécanisme général de l'Adaptive Biasing Force	28
2.6	Umbrella Sampling - histogramme des configurations	29
3.1	RMSD de chaque monomère du dimère d'AqpZ	38
3.2	Variables angulaires non contraintes	39
3.3	PMF à deux dimensions	40
3.4	PMF à trois dimensions	42
3.5	Collapsus du PMF 3D	43
3.6	PMF suivant la distance	44
3.7	Échantillonnage de la simulation du PMF 1D	44
4.1	Une membrane plus complexe	50
4.2	Enrichissement lipidique total	52
4.3	Enrichissement lipidique par feuillet	53
4.4	Densités lipidiques à la surface de la protéine native	54
4.5	Densités lipidiques à la surface de la protéine sauvage dans une membrane simple	54
4.6	Densités lipidiques à la surface de la protéine mutée	55
4.7	Densités lipidiques des insaturations	56
5.1	Système initial des simulations de la membrane encombrée . .	61
5.2	Membrane encombrée après 20 μ s de simulation	62
5.3	Échantillonnage des orientations de la membrane encombrée . .	63
5.4	Cartes de densité moyenne des cardiolipines et des protéines . .	64
5.5	Fonctions de distribution radiale et oligomérisation en chaînettes des protéines de la membranes simple et complexe	65

5.6	Fonctions de distribution radiale issues des simulations de Monte-Carlo	66
5.7	Représentation de la membrane encombrée complexe en stream-lines	67
7.1	PMF par ABF	75
7.2	PMF par Umbrella Sampling	76
7.3	Configuration problématique dans l'Umbrella Sampling	77
8.1	Composition lipidique de la membrane interne d' <i>E. coli</i> dans des nanodisques de SMA	80
8.2	Présentation du SMA	81
8.3	Résidus gros grains mimant les propriétés du SMA	82
8.4	SMILES et SMA linéaire	83
8.5	Un polymère sur une membrane	84
8.6	Solubilisation partielle d'une membrane par un polymère amphiphile	84
8.7	Le nanodisque de SMA	86
9.1	Evolutions temporelles de l'énergie des liaisons, des angles dièdres et de Van der Waals	91
10.1	Architecture et mécanisme du système de sécrétion de type VI	94
10.2	Le complexe membranaire	96
10.3	Cartes de proximités et densités de présence des peptides	98
11.1	Site actif de NarGH	103
11.2	La Glu-loop	105
11.3	Densités de l'eau autour du site actif	107
11.4	Diffusion des molécules d'eau autour du site actif	108
11.5	Coordination des molécules d'eau dans le canal dynamique menant au site actif	110

Liste des tableaux

2.1	Systèmes construits	23
2.2	Composition relative des lipides du modèle de la membrane interne gros grain d' <i>E.coli.</i>	25
2.3	Variables collectives	27
5.1	Contacts natifs et non-natifs dans la membrane encombrée . . .	62
8.1	Enrichissement en cardiolipines dans la membrane et dans le nanodisque	85
11.1	Liaisons hydrogènes impliquant les résidus d'intérêt	105

*Il n'y a pas de mauvais modèle ...
quand on a connaissance de ses limites.*

Jean-Pierre Duneau

Liste des abréviations

ABF	Adaptive Biasing Force
AqpZ	Aquaporine Z
CL	Cardiolipine
DFT	Théorie de la Fonctionnelle de la Densité
DPPC	Dipalmitoyl Phosphatidylcholine
<i>E.coli</i>	<i>Escherichia coli</i>
EPR	Résonance Paramagnétique Electronique
Fe-S	Fe ₄ S ₄
GpA	Glycophorine A
GROMACS	GRoningen MACHine for Chemical Simulations
HS-AFM	Microscopie à force atomique à haute vitesse
MIP	Major Intrinsic Proteins
MoCo	Cofacteur à molybdène
MS	Spectrométrie de Masse
NAMD	NANoscale Molecular Dynamics
Nar	Nitrate réductase A
OPM	Orientations of Proteins in Membranes
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PMF	Potentiel de Force Moyenne
POPE	Palmitoyloleoyl Phosphatidylethanolamine
POPG	Palmitoyloleoyl Phosphatidylglycerol
rédox	Oxydo-Réduction
RMSD	Root Mean Square Deviation
SMA	Styrene Maleic Acid
SMA-AqpZ	nanodisques de SMA contenant l'AqpZ
SMD	Steered Molecular Dynamics
T6SS	Système de sécrétion de type VI
US	Umbrella Sampling
VMD	Visual Molecular Dynamics

Section 1

Introduction

1.1 Avant-propos

En raison de l'ambiguïté du terme de "preuve" en science, terme polysémique dans les domaines des mathématiques et des sciences naturelles que je considère comme source de confusion, j'utiliserai l'anglicisme "évidence" pour désigner les "preuves" du domaine des sciences naturelles. En effet, "preuve" peut être traduit en anglais par "proof" dans le domaine des mathématiques pour une preuve formelle et par "evidence" pour la preuve reliée à une théorie/hypothèse. L'utilisation de "preuve" masquerait la nature inductive de la preuve scientifique et du bagage philosophique sous-jacent, notamment le problème de l'induction (HUME, 1748).

1.2 Approche historique de la membrane biologique : de la simple barrière à la mosaïque fluide

1.2.1 Pré-XX^{ème} siècle : Balbutiements : la membrane comme hypothèse inutile

La première "membrane" a été découverte par Robert Hooke en 1665 qui a observé la première "cellule" végétale de liège à partir d'un microscope (HOOKE, 1665). La cellule était alors décrite comme un espace creux entouré d'une paroi, que l'on connaît maintenant comme la paroi cellulaire, et était alors comparée aux alvéoles d'abeille (figure 1.1). Dans cette vision, les cellules sont les canaux dans lesquels sont transportés les "jus" de la plante. Dans les 200 ans qui ont suivi, la "membrane" était alors synonyme de la "paroi" cellulaire et désignait la seule délimitation visible au microscope : la paroi cellulaire des plantes. Schleiden et Schwann, en 1837 et 1839, ont ensuite précisé le concept de cellule, et c'est l'idée d'une cellule composée d'une substance amorphe, le protoplasme, qui est restée (SCHLEIDEN et al., 1847). Des méthodes chimiques permettant de séparer la paroi cellulaire de ce protoplasme (DUTROCHET, 1824)

et la délimitation entre la cellule et l'extérieur était vue comme le résultat d'une différence de phase entre ce protoplasme et l'eau extérieure.

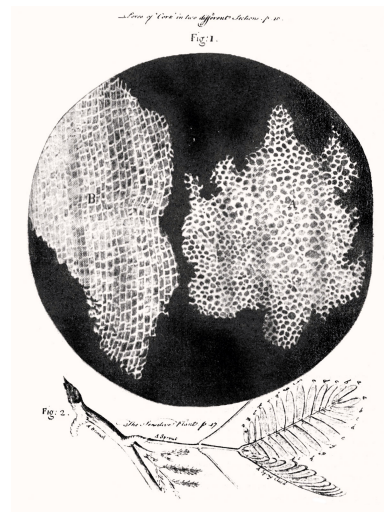


FIGURE 1.1 – . Description de Hooke des cellules de liège, en "alvéoles d'abeille" (HOOKE, 1665).

Les expériences d'osmose ont également échoué à déterminer l'existence d'une membrane différente du protoplasme. A partir de la deuxième moitié du XIX^{ème} siècle, le protoplasme était vu comme un colloïde (GRAHAM, 1861) et la limite entre la cellule et l'extérieur était vue comme un précipité de Traube (TRAUBE, 1867).

Colloïde. Graham sépara les molécules en deux catégories selon si elles pouvaient traverser une feuille de papier (GRAHAM, 1861). Les sels inorganiques et les sucres pouvaient facilement traverser et furent appelés crystalloïdes alors que les composés qui ressemblaient à de la gélatine et qui ne pouvaient pas traverser furent appelés colloïdes. La structure du protoplasme, visqueuse et gélatineuse, a rapidement été appelée colloïde.

Précipité de Traube. Un précipité de Traube est la membrane semi-perméable qui se crée à l'interface entre une solution de ferrocyanide de potassium et une solution de sulfate de cuivre. Se formant spontanément à une interface, ce précipité confortait l'hypothèse de l'absence de membrane, le protoplasme se durcissant simplement à son interface avec l'eau (TRAUBE, 1867).

Par exemple, les expériences d'osmose sur des cellules végétales ont permis l'observation de la rétraction de la cellule dans un milieu hypertonique (DETMER, 1895 ; WILKIE, 1961 ; SMITH, 1962). Ces résultats sont simples à analyser aujourd'hui, mais étaient plus ambigus dans le contexte de la deuxième moitié du XIX^{ème} siècle : la membrane restait invisible, la paroi cellulaire n'était pas indispensable et l'idée du précipité de Traube confirmait

l'inutilité d'une membrane supplémentaire car les propriétés observées pouvaient simplement être le résultat du durcissement des limites du protoplasme. Ce n'est qu'à partir du XX^{ème} siècle que l'existence d'une membrane, en plus de la paroi cellulaire, s'est imposée.

Trois lignes d'évidence sont généralement mises en avant pour l'origine de la membrane : les études de perméabilité d'Overton (OVERTON, 1895 ; OVERTON, 1899), les expériences de microinjection de Chamber (CHAMBERS, 1922) et les mesures électriques de Höber (PETHIG et al., 2012).

Overton immergea des cellules dans des solutions de plus de 500 substances différentes et remarqua que les substances "solubles dans l'éther" (ce qu'on appelle aujourd'hui apolaires) n'amenèrent pas de réduction de la taille des cellules, à l'opposé des substances solubles dans l'eau. La conclusion était que les molécules apolaires ("solubles dans l'éther") rentraient plus facilement dans les cellules et qu'il existait donc une membrane distincte de la paroi cellulaire faite de composés "solubles dans l'éther". Il a suggéré le cholestérol et les phospholipides (lécithines) comme composant de cette membrane, même si la difficulté était ensuite d'expliquer comment les molécules hydrophiles pouvaient passer cette membrane.

Après l'amélioration des techniques de microinjection, Chamber quant à lui injecta une solution pouvant digérer le cytoplasme dans différentes parties de la cellule.

Réinterpréter des résultats antérieurs selon nos propres concepts peut poser problème quand il s'agit de comprendre les cheminements qui ont amené à une découverte. Ainsi, la découverte de la membrane cellulaire est souvent attribuée à Overton en 1895 : puisque des substances "solubles dans l'éther" (apolaires, ou lipides) forment cette membrane, cela devrait être évident ...

L'article de Chamber (CHAMBERS, 1922) nous rappelle par contre que, même en 1922, l'existence d'une membrane n'était toujours pas acquise : "It is well known that selective permeability, or semipermeability, is one of the essential characteristics of the living cell. So far, however, there is no evidence as to whether the semipermeability of protoplasm is a property of its entire mass or of its surface only."

Lorsque cette solution était déposée sur le protoplasme, il ne se passait rien, alors que si la solution était injectée dans la cellule, le protoplasme lysait et aucun dommage n'était fait à la membrane avant la lyse complète. Cela indiquait que le protoplasme était fait d'une matière différente de sa surface. Höber quant à lui mesura la conductivité interne des cellules et obtint des mesures qui étaient incompatibles avec l'attachement du protoplasme aux électrolytes du milieu extérieur : les électrolytes devaient au contraire être à l'intérieur des cellules. La nature différente, et hydrophobe, de la limite protoplasme/eau semble donc acquise à partir de ces trois évidences. Il faut cependant noter que ces conclusions ne sont à ce moment valables uniquement

chez les eucaryotes, il faudra encore attendre de pouvoir observer les membranes avec un microscope électronique pour admettre leur existence chez les bactéries.

1.2.2 Jusqu'à la bicouche : d'une simple couche de lipide à une (bi ?)couche lipidique

Le concept de bicouche lipidique est en général attribué à Gorter et Grendel (GORTER et al., 1925), en se basant sur les travaux de Langmuir (LANGMUIR, 1917) sur le caractère amphiphile des molécules lipidiques.

Une partie du débat porte d'abord sur l'épaisseur des films d'huile déposés sur l'eau. Avant de proposer son interprétation, Langmuir (LANGMUIR, 1917) décrit les travaux précédents de Rayleigh et ensuite de Devaux sur la tension de surface de l'eau lorsqu'elle est contaminée par de l'huile. Par exemple, Rayleigh mesura la tension de surface de l'eau selon la quantité d'huile ajoutée (Figure 1.2).

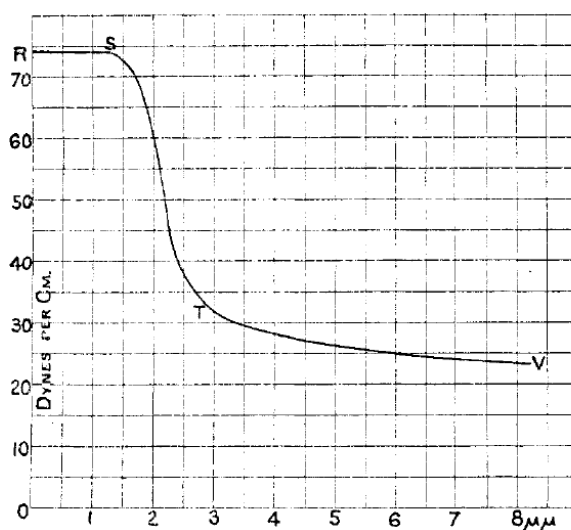


Fig. 1

FIGURE 1.2 - . Tension de surface de l'eau selon la quantité d'huile ajoutée. Le point S est interprété comme une couche simple d'huile et le point T est interprété comme une couche double.

La tension de surface diminue rapidement à partir d'une épaisseur d'huile critique. De là, Rayleigh conclua que le point S, avant que la tension de surface ne chute, correspondait à une monocouche d'huile alors que le point T correspondait à une double couche.

A partir des propriétés des films d'huiles sur de l'eau et d'un point de vue "de chimiste" (LANGMUIR, 1917), une monocouche d'huile sur de l'eau ne pouvait qu'avoir les molécules d'huiles qui s'orientaient, le groupement carboxylique se solubilisant dans l'eau alors que le groupement hydrophobe pointant

vers l'extérieur. Avant de passer à la bicouche, notons la manière particulière avec laquelle Langmuir effectue un glissement de concept (LANGMUIR, 1917), à mon avis indispensable à toute entreprise d'histoire des sciences (LATOUR, 1987 ; KUHN et al., 2012). Tout un chapitre est en effet intitulé "« Physical » versus « Chemical » Forces" et présente un conflit entre "la vision des physiciens" et "la vision des chimistes" sur les forces entre les molécules. Selon Langmuir, les physiciens considéraient uniquement des forces non spécifiques dépendantes de la distance entre les molécules, qui s'étendaient à l'infini, comme la gravité et l'électrostatique. Les chimistes, quant à eux, considéraient uniquement la spécificité des interactions avec un rayon d'action à courte portée et son origine atomique, sur lesquels Gurvich et Langmuir ajoutent une vision moléculaire que l'on qualifierait aujourd'hui de structurale car dépendante des atomes voisins. Même en étant conscient du contexte, il est difficile de ne pas y voir une caricature, un homme de paille : en effet, quel physicien du début du XX^{ème} siècle ignorait que la gravité était négligeable à très courte distance ou que l'attraction dépendait du type d'atome ? Malgré ce préambule, Langmuir se base sur la théorie des valences de Lewis et propose alors l'origine du comportement de l'huile sur l'eau : comme l'huile est en monocouche sur l'eau, le groupement carboxyle a une affinité pour cette dernière (de par la valence secondaire de l'oxygène) et est donc solubilisé alors que le reste du lipide (la chaîne d'hydrocarbure) a une plus grande affinité pour les chaînes d'hydrocarbures et ne se tournera donc pas vers l'eau.

Gorter et Grendel (GORTER et al., 1925) se sont ensuite basés sur les travaux de Langmuir (bien que brièvement cités) pour déduire la bicouche lipidique. Gorter et Grendel ont extrait les lipides d'érythrocytes (connus pour n'avoir aucune membrane interne) et ont utilisé ces lipides pour couvrir une surface d'une monocouche qui avait deux fois la surface des érythrocytes. Cette expérience est créditée aujourd'hui comme étant la plus concluante sur la caractérisation de la bicouche lipidique. La démonstration semble convainquante aujourd'hui mais n'a pas fait consensus avant des travaux bien plus tardifs du fait des problèmes techniques liés à l'expérience. Par exemple, et jusqu'en 1965, bien qu'une bicouche soit admise peu après l'article de 1925, l'emballage de la membrane faisait encore débat (le caractère "complet" ou "partiel" de la bicouche n'était pas encore tranché (MADDY et al., 1965)). Rétrospectivement, la démonstration de Gorter et Grendel ne laissait pas non plus la place aux protéines membranaires. Malgré ces interrogations, ces travaux ont tout de même ouvert la voie aux descriptions moléculaires de la membrane et ont amené au premier modèle influent de la membrane, le modèle paucimoléculaire.

1.2.3 Une (bi ?)couche de lipides entre deux couches de protéines

Dans les années 1930s, Danielli et Davson (DANIELLI et al., 1935) proposent le modèle paucimoléculaire. Comme les membranes contenaient une grande quantité de protéines (sans information sur leur position) et la tension de surface semblait plus faible qu'attendu par une simple interface eau/lipides,

le modèle proposé consistait en une couche lipidique entre deux couches de protéines (voir figure 1.3).

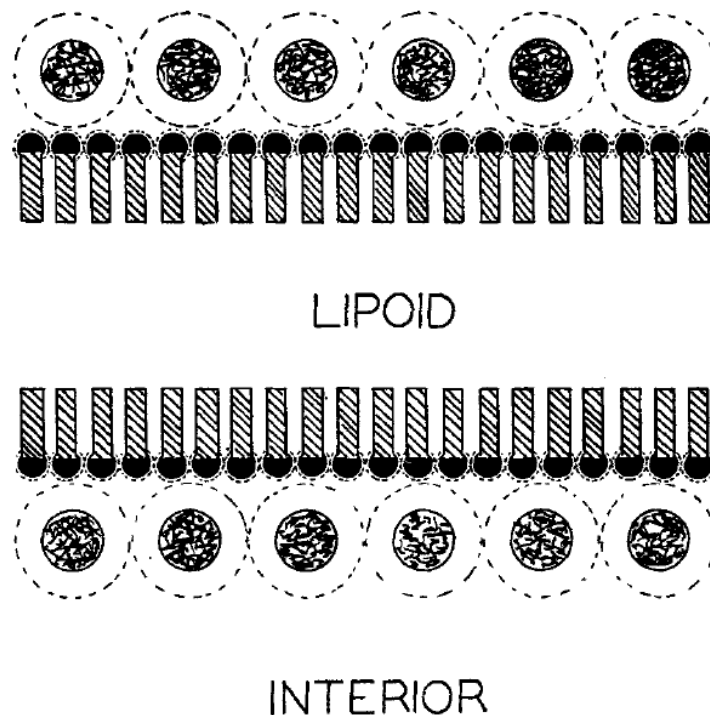


FIGURE 1.3 – . Schéma du modèle paucimoléculaire (DANIELLI et al., 1935), constitué de deux couches de protéines de part et d'autre d'une couche de lipides d'une épaisseur inconnue.

Un résumé du consensus en 1935 est aussi accessible au début de l'article de Danielli : "Il y a une quantité considérable d'évidences qui soutient l'hypothèse que les cellules vivantes sont entourés d'une couche mince de matériel lipoïde. [Les travaux précédents] montrent de manière convainquante que cette couche qui sépare électriquement l'intérieur des cellules de l'extérieur est d'une épaisseur uni à tri-moléculaire." De fait, 10 ans après l'expérience sur les globules rouges indiquant une formation en bicouche, le doute persiste quand à la structure en mono, bi ou tricouche lipidique.

Ainsi, en 1935, la nature exacte des lipides qui constituaient la couche de lipide était inconnue mais certaines de ses propriétés l'étaient : ces lipides sont des lipides connus, ils possèdent une proportion de groupements acides (comme les groupements carboxyles), des groupes basiques (amines) et une couche de protéine est absorbée sur la couche lipoïde. Ce modèle posait cependant la question de la perméabilité : comment des molécules pouvaient traverser cette couche continue de protéines-lipides ?

L'idée que des protéines puissent alors traverser la membrane et permettre le transport de molécules hydrophiles, même si brièvement évoqué, n'est pas

encore crédible, au profit de modèles où les ions eux-mêmes permettaient la création de pores. De plus, les expériences examinant la tension de surface des membranes, qui constituaient l'argument principal du modèle, ont rapidement été invalidées par les mêmes auteurs. Le résultat, bien que ce modèle ait été dominant, a permis le développement de modèles compétiteurs (même développés par Danielli), le plus souvent proposant des modèles plus mosaïques qui permettaient le passage de molécules de différentes manières pour résulter en des modèles mixtes.

1.2.4 De l'utilisation du microscope électronique à la mosaïque fluide

D'un point de vue général, l'utilisation du microscope électronique devrait être décisive dans la caractérisation de la structure de la membrane cellulaire. En effet, simplement voir la membrane devrait alors être suffisant. En pratique, cela n'a pas été si simple et illustre très bien la sous-détermination de la théorie par les faits (DUHEM, 1954 ; QUINE, 1951). En effet, les structures visibles en microscopie électronique ont été interprétées à la lumière des différentes théories sur la membrane-mosaïque. Par exemple, des micrographes électroniques sont interprétés dans les années 1950s comme étant un mélange du modèle paucimoléculaire en mosaïque avec des pores (HILLIER et al., 1953), des "plaques" qui s'alternent avec des pores (voir figure 1.4 où l'on voyait des "plaques" de protéines, du modèle paucimoléculaire, séparées par des pores).

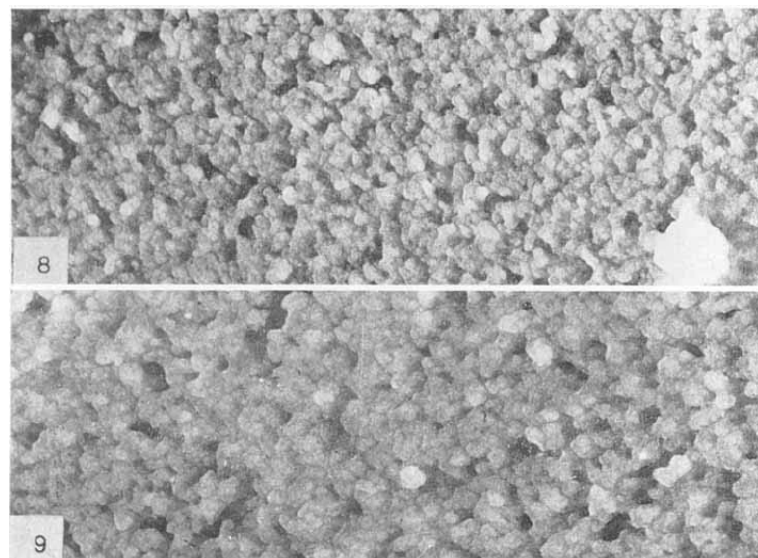


FIGURE 1.4 – . Micrographe électronique de la surface d'un érythrocyte. L'accent est mis sur les "plaques" de différentes tailles (ici les plaques les plus courantes en haut et les plus larges en bas), alternées de pores, mêlant le modèle paucimoléculaire et la théorie des pores (HILLIER et al., 1953).

En cohabitation avec les difficultés d'interprétation des observations de microscopie électronique, les observations déterminantes pour le développement

du modèle de la mosaïque fluide sont les expériences de cryofracture et de diffusion. La cryofracture consiste à geler l'échantillon puis à le fracturer, séparant les deux feuilletts de la bicouche lipidique, avant observation par microscopie électronique. L'arrivée de la cryofracture dans les années 1960s a donc permis l'observation de protubérances (d'un coté de la fracture) et de cavités (de l'autre coté), qui a forcé à considérer l'existence d'objets (les protéines membranaires) qui traversaient la membrane. Le caractère fluide de la membrane a ensuite été dérivé des expériences de diffusion où, par exemple, les membranes de deux cellules (humain/souris) ont été fusionnées et les anticorps de chacun suivis par fluorescence (FRYE et al., 1970) ; le mélange rapide de ces anticorps a permis d'établir la fluidité de la membrane. Rapidement, le modèle de la mosaïque fluide en a découlé.

1.2.5 La mosaïque fluide

En 1971 et 1972, Singer et Nicholson présentent le modèle de la mosaïque fluide (SINGER et al., 1972). La membrane est alors "une solution bi-dimensionnelle de protéines globulaires orientées et de lipides" (voir figure 1.5).

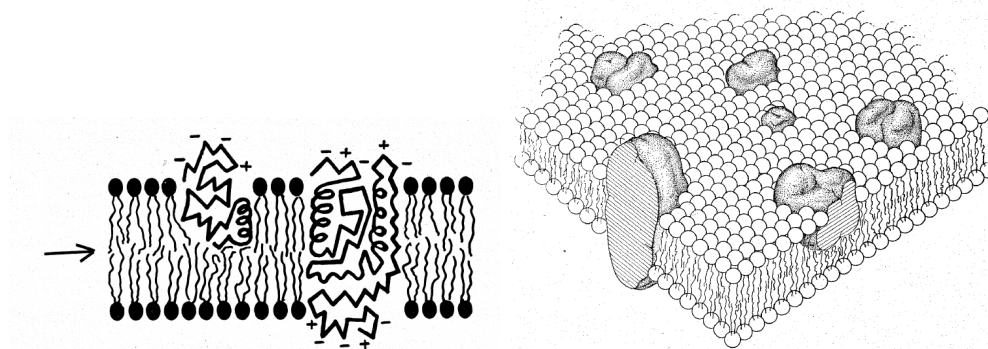


FIGURE 1.5 – . Modèle de la mosaïque fluide (SINGER et al., 1972) de Singer et Nicholson. A gauche, un schéma d'une coupe où les lipides sont arrangés en bicouche discontinue avec leur tête polaire au contact de l'eau. A droite, un schéma tri-dimensionnel.

Déterminer le moment où l'épaisseur de la couche de lipide formant la membrane est admise comme étant bimoléculaire, une bicouche, est difficile.

On peut cependant affirmer que le modèle de la mosaïque fluide met fin à des années d'interrogations car le modèle permet surtout d'expliquer la quantité manquante de lipides dans une bicouche complète par la place occupée par les protéines membranaires.

Ce modèle intègre les considérations thermodynamiques et expérimentales décrites jusque là dans la littérature.

Il décrit l'existence de deux catégories de protéines associées aux membranes :

- Les protéines périphériques, associées de manière faible à la bicouche. Ces protéines sont facilement dissociable par des traitements ioniques sans lipides associés, tout en restant solubles dans les solvants aqueux ensuite, comme le cytochrome c des membranes mitochondriales.
- Les protéines intégrales qui pénètrent partiellement ou entièrement la membrane et représentent plus de 70% des protéines membranaires. Ces protéines nécessitent des traitements plus drastiques pour les dissocier, comme des détergents, avec des lipides qui restent associés, et sont souvent insolubles ou agrégés dans l'eau s'ils ne sont pas associés avec des lipides/détergents.

Ces dernières protéines sont hétérogènes en taille et en majorité en hélices α .

Concernant les phospholipides, ils sont en majorité sous forme de bicouche. Les transitions de phases observées ne permettent pas de conclure si la membrane est totalement sous forme d'une bicouche continue ou interrompue, et si certaines parties de la membranes sont sous forme de mono ou tricouche, même si la majorité est sous forme de bicouche.

Les interactions protéines-lipides dans les membranes semblent alors jouer un rôle direct dans un grand nombre de fonctions cellulaires, avec des protéines qui nécessitent des phospholipides particuliers pour leur activité. Cependant, l'association de lipides avec les protéines ne change pas la transition de phase des lipides, laissant supposer que les interactions entre les phospholipides et les protéines sont faibles et largement indépendantes. Ce paradoxe est résolu en considérant que la majorité des lipides de la membrane est sous forme de bicouche peu couplée aux protéines avec une petite fraction de lipides qui est en interaction beaucoup plus forte avec certaines protéines.

Le modèle de la mosaïque fluide est alors accepté comme un modèle permettant d'appréhender toutes les membranes cellulaires. Depuis 1972, ce modèle a évolué pour prendre en compte la richesse des données dont nous disposons aujourd'hui.

De la manière de faire un historique en science

L'histoire de la membrane biologique n'est pas présentée ici comme dans les manuels scolaires. En effet, ces derniers et notre formation nous habituent à une présentation graduelle de l'histoire des sciences qui se résumerait à une liste de découvertes dans le temps. Cette manière de présenter les choses n'est cependant pas pertinente pour s'intéresser à l'histoire d'un sujet, tout comme présenter Einstein comme la continuation de Newton manque à la fois les différents contextes historiques et les changements de concepts et de paradigmes qui ont traversé l'histoire (KUHN et al., 2012). Kuhn notait en

effet que "L'histoire, si on la considérait comme autre chose que des anecdotes ou des dates, pourrait transformer de façon décisive l'image de la science dont nous sommes actuellement empreints".

Pour cette raison, dans cette présentation historique, je me suis attaché à brièvement replacer chaque découverte dans son contexte, pour faire ressortir aussi bien les concepts principaux mis en jeu mais aussi les doutes et les anomalies qui traversaient, traversent et traverseront la science.

De plus, comme pour beaucoup de concepts acquis en science, il est difficile de pointer un "moment" qui correspondrait à la découverte, ou même de créditer une personne en particulier et la membrane biologique en est un bon exemple.

L'existence du concept de membrane biologique est relativement récent et il est nécessaire de revisiter chaque expérience du point de vue des concepts de son époque, à chaque fois que nous disons "c'est la membrane qui est responsable de ce comportement".

Le terme de "membrane" a lui-même évolué au cours du temps.

1.3 La membrane biologique : le modèle actuel

1.3.1 La mosaïque fluide, un modèle qui a pu intégrer ses critiques depuis 1972

Peu après l'énonciation originale du modèle de la mosaïque fluide, une avalanche de données est venue remettre en question certains présupposés.

Dans le modèle originel, la fluidité de la membrane permet la libre diffusion latérale des protéines. Le modèle a, dès 1976 (NICOLSON, 1976), été enrichi par les nombreux exemples où des éléments extérieurs viennent limiter la diffusion des protéines, comme les éléments du cytosquelette et la matrice extracellulaire, ou par des séquestration de protéines intégrales dans des domaines fonctionnels moins mobiles, bien que les mécanismes précis n'aient été disséqués que plus tard.

Les lipides qui forment ce fluide sont distribués dynamiquement de manière inégales dans le plan de la membrane (OHVO-REKILÄ et al., 2002 ; ANISHKIN et al., 2013), avec la présence de certains lipides (cholestérol, cardiolipines) ou chaînes latérales qui en changent la fluidité et peuvent participer à la formation de domaines dynamiques. La potentielle fonction de tri effectuée par ces lipides et l'existence, liée, de rafts lipidiques, en particulier leur taille et leur persistance, reste toutefois encore en débat à ce jour.

Le modèle actuel est cependant toujours une mosaïque fluide, bien que la

notion de fluidité ait été amoindrie au profit des aspects mosaïques et dynamiques.

1.3.2 Diffusion anormale et aspects moléculaires

Le modèle originel de la mosaïque fluide incluait une diffusion aléatoire des protéines, dite Brownienne, bien que des limitations étaient originellement connues. Expérimentalement, ce sont des expériences de suivi de molécule unique (SCHÜTZ et al., 1997), de spectroscopie de corrélation de fluorescence, de redistribution de fluorescence après photoblanchiment et en particulier de microscopie à force atomique à haute vitesse (HS-AFM) qui ont permis le suivi de la diffusion limitée des protéines membranaires. Un nombre important de modèles théoriques de ces diffusions anormales ont été développés et ne seront pas discutés ici (METZLER et al., 2016).

En suivant les protéines grâce à ces techniques, nous pouvons ainsi observer trois types de comportement (MUNGUIRA et al., 2016) :

- Des protéines immobiles, qui vibrent autour de leur position aléatoirement.
- Aux frontières de cette phase solide, des protéines qui glissent les unes par rapport aux autres, en étant fixes parfois sur l'échelle de la minute.
- Des protéines au mouvement Brownien mais qui sont enfermées dans des espaces définis.
- Des protéines qui diffusent avec un mouvement Brownien

En particulier, dans le contexte encombré qui est le contexte cellulaire, la diffusion des protéines dépend des interactions moléculaires (CASUSO et al., 2012; MUNGUIRA et al., 2016). L'hétérogénéité moléculaire présente dans les membranes favorise une dynamique du verre.

Ces différents régimes dynamiques sont difficilement observés expérimentalement car il faut suivre et la protéine d'intérêt et son environnement pour les mettre en évidence. Ce sont en particulier des expériences d'HS-AFM qui permettent alors d'observer cette dynamique avec une résolution temporelle en dessous de la seconde (ANDO et al., 2014).

La diffusion translationnelle et rotationnelle est alors limitée, avec une diffusion rotationnelle presque nulle lorsque les protéines sont impliquées dans des interactions à courte distance jusqu'à une rotation de plus de 120° sur une demi-seconde (CASUSO et al., 2012).

Les sources de ces diffusions anormales se retrouvent donc à l'échelle moléculaire et comprennent les différentes interactions lipides/lipides et lipides/protéines ainsi que leurs aspects thermodynamiques (HÆLMS, 2002). A cette échelle, cette diffusion anormale se traduit alors par des interactions privilégiées, de certaines configurations de protéines/lipides à courte ou longue distance, dans des environnements la plupart du temps encombrés. De plus, la présence même de protéines dans une membrane amène à des perturbations

telles qu'une courbure de la membrane et un mésappariement hydrophobe.

L'hydrophobicité des résidus à la surface de la protéine est déterminante dans le positionnement de la protéine par rapport à la membrane. Lorsque l'épaisseur hydrophobe ne correspond pas à l'épaisseur de la membrane, cette dernière va répondre par une contrainte (figure 1.6).



FIGURE 1.6 – . Schéma général du mismatch hydrophobe. La protéine est schématisée par un cylindre. Lorsque l'épaisseur hydrophobe de la protéine est différente de l'épaisseur de la membrane, cette dernière est stressée et les lipides se réorganisent.

En réponse à un mésappariement hydrophobe, plusieurs réponses de l'environnement ou de la protéine sont possibles pour réduire ce stress et réduire l'énergie potentielle de cette configuration problématique (KILLIAN, 1998). Dans le cas d'une membrane simple avec une seule protéine et un seul type de lipide, ceux-ci peuvent s'étirer ou s'écraser pour s'adapter à l'épaisseur hydrophobe de la protéine. Lorsque la membrane est plus complexe et possède des lipides avec des chaînes latérales de taille différente, les lipides les plus adaptés à l'épaisseur hydrophobe de la protéine peuvent être recrutés à sa périphérie et ainsi réduire le stress de surface. Lorsque plusieurs protéines sont présentes, leur aggrégation permet de réduire le nombre de surfaces stressées. D'autres solutions peuvent aussi se mettre en place mais sont des cas plus particulier, par exemple par un changement conformationnel qui modifierait l'épaisseur hydrophobe de la protéine ou un basculement de la protéine s'il ne s'agit que d'une seule hélice transmembranaire.

De l'adaptation des lipides à la surface de la protéine naît aussi une courbure. En fonction simplement de la taille et de la forme de la protéine insérée, la protéine induit une courbure de la membrane. Elle est révélée à la fois *in vivo*, *in vitro* et *in silico* (FRESE et al., 2008 ; HSIN et al., 2009).

En conséquence de ces perturbations, la diffusion des lipides va aussi être affectée. Outre le tri de certains lipides selon la taille de leur chaîne latérale pour s'adapter au mésappariement hydrophobe, des lipides vont être plus lents lorsqu'ils sont proches de la protéine. Cette zone, du contact de la protéine jusqu'à 4-5 couches de lipides autour, a été appelée l'annulus

lipidique (KILLIAN, 1998; GOOSE et al., 2013; J.-P. DUNEAU et al., 2017). La composition de la première couche de lipides autour de la protéine est dépendante des propriétés physico-chimiques de la protéine et, par extension, les couches suivantes le sont aussi.

Des perturbations à longues distances des protéines sur les lipides ont été mises en évidence, principalement *in silico* (J.-P. DUNEAU et al., 2017). Dans ce contexte, les lipides sont alors perturbés par la présence de la protéine et se relaxent de manière différente suivant la distance et la partie du lipide (tête polaire, phosphate, insaturation). La perturbation est dépendante de la protéine, et donc de la nature physico-chimique des atomes de surface, mais aussi de la topographie de cette surface.

1.3.3 Interactions à longue distance

L'importance biologique des interactions à courte distance entre protéines n'est plus à démontrer. De plus, les interactions entre protéines et lipides sont aussi importantes pour la fonction de certaines protéines; par exemple, des cardiolipines sont indispensables à la bonne fonction de l'Aquaporine Z (LAGANOWSKY et al., 2014), l'ATP synthase (ACEHAN et al., 2011) ou d'autres protéines (SHEN et al., 2017; CHRISTIE et al., 2011; CLAYPOOL, 2009), toutes possédant plusieurs segments transmembranaires. L'accessibilité des données sur les interactions à courte distance est due en majorité à leur (relative) facilité d'acquisition : les interactions directes peuvent être étudiées grâce à une diversité de méthodes biochimiques et l'on va retrouver, par exemple, les lipides les plus ancrés à la surface d'une protéine après la procédure de cristallisation. Malgré cela, le couplage d'études biochimiques et bioinformatiques est souvent nécessaire car les protocoles biochimiques entraînent, par exemple, une délipidation importante, masquant les lipides qui ne sont pas ancrés très fortement mais importants pour la fonction. L'annulus lipidique est par conséquent difficile d'accès expérimentalement. Au delà de la courte distance, les techniques expérimentales montrent des limites de résolution et les interactions sont beaucoup moins claires.

D'un point de vue biologique, les protéines doivent trouver un équilibre entre une aggrégation complète et une séparation totale, privilégier les interactions utiles et, surtout, éloigner les interactions non spécifiques.

Ce problème est à rapprocher d'une expression en dynamique moléculaire : les champs de forces, qui définissent les interactions entre les différents atomes, sont trop "collants". Du point de vue du modèle moléculaire que l'on peut se faire des protéines, il est en effet difficile d'obtenir un modèle ayant des propriétés qui satisfassent certains critères empiriques (température de fusion, comportement de chaînes latérales, vibrations ...) sans derrière avoir un modèle où les atomes collent trop entre eux.

On peut rationaliser ce problème de différentes façons (c'est une question d'échelle de temps, etc.) mais le concept initial reste : il nous est difficile de modéliser et donc d'imaginer des molécules qui ne collent pas plus entre elles que ce qu'il se passe dans la réalité.

De ce point de vue, la vision d'une simple diffusion brownienne où les interactions importantes se mettraient en place à courte distance et les autres seraient ignorées semble, conceptuellement, impossible. Il faut donc, pour compléter notre modèle, des forces à plus grande distance pour contrecarrer ces atomes que l'on ne peut imaginer que trop "collants".

Les perturbations induites par la présence de protéines dans une membrane, sur la membrane et sur d'autres protéines, sont alors le candidat idéal. Plus qu'en permettant aux interactions favorables de se mettre en place, des interactions à longue distance pourraient cette fois forcer et du coup expliquer pourquoi les protéines membranaires ne se tombent pas simplement dessus.

Ces forces à longue distances n'ont été que peu décrites jusque là. Des travaux précédents au laboratoire ont observé des forces à longue distance qui se propageraient au travers des lipides à partir d'observations d'anomalies dans le profil énergétique de la dissociation d'OmpF (CASUSO et al., 2012), une porine bactérienne. Des travaux sur l'annulus de lipide ont de plus permis d'observer des perturbations qui dépendent du type de protéine et qui s'étendent jusqu'à plusieurs couches de lipides (J.-P. DUNEAU et al., 2017), ce qui correspond à plusieurs dizaines d'angströms.

1.3.4 Mon projet

Le but de ce manuscrit s'inscrit dans l'exploration et la description de ces forces à longue distance entre les protéines membranaires avec une approche *in silico* de dynamique moléculaire.

Pour cela, une protéine modèle utilisée au laboratoire est l'Aquaporine Z (AqpZ). Elle est formée de 6 hélices transmembranaires formant un pore, dont le rôle est d'assurer la circulation stricte des molécules d'eau. Dans les membranes, l'AqpZ oligomérisé et forme des homo-tétramères. Cette association en tétramère ainsi que la topologie du monomère sont retrouvés pratiquement chez tous les membres de la superfamille des Major Intrinsic Proteins (MIP). Plus particulièrement, la structure des Aquaporines est très conservée dans l'ensemble du vivant, des bactéries jusqu'aux plantes et mammifères, malgré une faible conservation de séquence (ZARDOYA, 2005). L'état oligomérique des protéines de cette superfamille semble donc résider non pas dans sa séquence mais dans sa topographie.

De plus, le monomère semble être l'unité fonctionnelle : chacun semble fonctionner indépendamment au sein du tétramère. En revanche, nous avons

des évidences pour penser qu'un monomère isoler ne peut pas assurer sa fonction (SCHMIDT et al., 2017). Il semble donc que chaque monomère, traduit individuellement dans la membrane, doive "trouver" d'autres monomères pour être sous sa forme tétramérique, la seule forme observée expérimentalement et qui semble alors nécessaire à sa fonction.

Ce monomère va nous permettre d'étudier les relations qui existent entre les différentes faces de la protéine et les lipides l'entourant ainsi que les relations entre différents monomères, à différentes échelles, à l'aide de simulations de dynamique moléculaire gros grain.

Les simulations apportent une vision moléculaire des modèles et permettent à la fois de produire des explications sur les phénomènes observés et des prédictions sur ceux à observer. Dans ces simulations, les paramètres des atomes et les formes fonctionnelles, qui définissent les interactions entre les particules, sont regroupés dans un champ de forces. Le champ de forces est habituellement paramétré pour reproduire des résultats expérimentaux comme des enthalpies ou des paramètres spectroscopiques variés. Ils peuvent de plus exister à diverses échelles, avec des particules qui peuvent contenir un groupe d'atomes. Dans le cas des membranes, le plus utilisé est le champ de forces gros grain MARTINI (MARRINK et al., 2007; MONTICELLI et al., 2008) qui regroupe environ 4 atomes "lourds" dans une particule (environ 10 atomes en comptant les hydrogènes). C'est un champ de forces simple originellement paramétré pour reproduire les séparations de phase de divers mélanges de lipides. Alors qu'un champ de forces dit "tout atome", où chaque particule représente un atome, ne permet d'atteindre que l'échelle de la centaine de nanosecondes en routine, un tel champ de forces gros grain permet, de par la réduction du nombre de particules, de calculer jusqu'à l'échelle de la dizaine de microsecondes au prix d'une approximation, tout en gardant la topographie de surface des protéines. Atteindre ces temps de simulation est important pour les protéines membranaires car les temps caractéristiques dans les membranes sont plus grands que dans le cytoplasme.

En utilisant cet outil ainsi que notre protéine modèle, l'AqpZ, la question naïve qui sous-tend ce manuscrit restera "Comment les protéines trouvent-elles le partenaire idéal?" au travers de 3 axes principaux :

- Tout d'abord, en reprenant les problèmes rencontrés dans l'énergétique de la dissociation d'OmpF (CASUSO et al., 2012), j'ai cherché à explorer l'énergétique de la dissociation de deux monomères d'AqpZ en prenant en compte les orientations relatives des protéines. En effet, un écueil de l'étude sur OmpF était le manque d'échantillonnage des orientations relatives, malgré un potentiel de force moyenne qui faisait sens.
- Ensuite, je me suis intéressé aux lipides présents autour de l'AqpZ, sous sa forme tétramérique, monomérique, mais aussi d'un mutant

monomérique caractérisé au laboratoire (SCHMIDT et al., 2017), pour questionner plus en détails la nature de l'annulus de lipides. Pour cela, j'ai développé un modèle gros grain de la membrane interne d'*Escherichia coli* contenant des cardiolipines pour étudier les interactions et le tri des lipides effectué par la protéine à différentes distances de sa surface.

- Pour finir, après avoir étudié les relations entre deux monomères puis entre la protéine et les lipides, j'ai construit des membranes encombrées de 125 monomères d'AqpZ pour observer la dynamique des protéines et des lipides dans un contexte se rapprochant plus d'une membrane réaliste, les membranes biologiques contenant en moyenne, en masse, 50% de protéines.

Après avoir exploré ces axes, plusieurs chapitres supplémentaires suivront qui concernent des développements plus méthodologiques ainsi que des travaux réalisés dans le cadre de collaborations.

Section 2

Matériel et méthodes

2.1 La dynamique moléculaire

2.1.1 Principe général

La dynamique moléculaire consiste à simuler l'évolution d'un système de particules au cours du temps en utilisant les lois de Newton (figure 2.1). Classiquement, les forces entre les particules sont calculées et intégrées dans les équations du mouvement dans le temps avec un temps d'intégration le plus petit possible (2 à 20 fs suivant le type de simulation). Un contrôle de la température et de la pression est aussi présent.

Schéma simplifié de l'algorithme de dynamique moléculaire

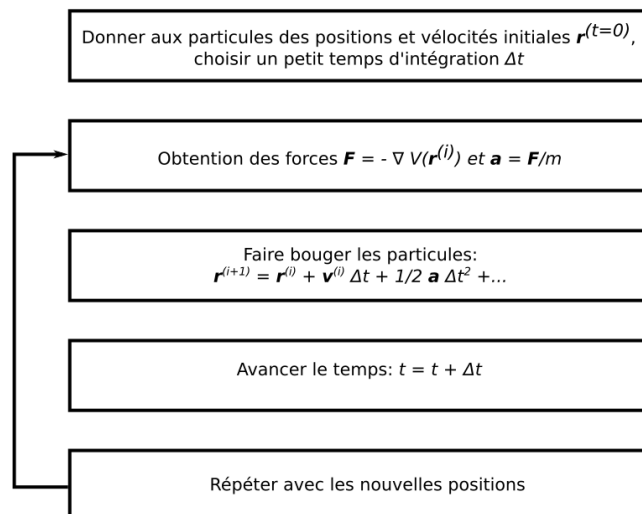


FIGURE 2.1 – Workflow simplifié d'une simulation de dynamique moléculaire. Après avoir attribué à chaque particule une position et une vitesse, la simulation consiste à intégrer les équations du mouvement dans le temps. Un contrôle de la température et de la pression, non affiché, est aussi présent.

2.1.2 Intégration des lois de Newton

Plusieurs méthodes d'intégration des lois de Newton sont possibles. Dans ce travail, nous avons utilisé la méthode de Velocity Verlet (VERLET, 1967; SWOPE et al., 1982) ainsi que la méthode de leap-frog (*voir annexe A et encadré*).

L'algorithme de leapfrog permet de mettre à jour les positions ($\mathbf{r}$) et les vitesses ($\mathbf{v}$) des particules selon les forces qui s'appliquent. La forme classique est :

$$\begin{aligned}\mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \Delta t \cdot \mathbf{v}_r(t + \frac{1}{2}\Delta t) \\ \mathbf{v}_r(t + \frac{1}{2}\Delta t) &= \mathbf{v}_r(t - \frac{1}{2}\Delta t) + \Delta t \cdot \frac{\mathbf{F}_r}{m}(t)\end{aligned}\tag{2.1}$$

Ces équations seront ensuite modifiées par un contrôle de la température et de la pression.

2.1.3 Contrôle de la température

Contrôler la température est conceptuellement simple du point de vue macroscopique - on a un système de mesure et on chauffe ou refroidit pour maintenir la température. Du côté microscopique, du point de vue des atomes, c'est moins simple.

Pour contrôler la température, plusieurs méthodes ont été proposées. La plus simple est la méthode de Berendsen (BERENDSEN, POSTMA et al., 1984) et consiste à simplement recalibrer périodiquement les vitesses des atomes pour que le changement de la température soit lié à la différence de température :

$$\frac{dT}{dt} = \frac{1}{\tau}(T_0 - T)\tag{2.2}$$

Où la déviation de la température voulue T_0 diminue exponentiellement avec une constante de temps τ . Cette méthode supprime simplement les fluctuations de l'énergie cinétique (et est de plus très efficace pour cela) et ne produit pas un ensemble thermodynamique correct, c'est pourquoi elle est uniquement utilisée lors de l'équilibration. Lors des simulations, j'ai utilisé plusieurs thermostats qui permettent d'obtenir un ensemble thermodynamique correct : le thermostat dit velocity rescale (BUSSI et al., 2008) implémenté dans GROMACS (BERENDSEN, SPOEL et al., 1995; ABRAHAM et al., 2015; VAN DER SPOEL et al., 2005; HESS et al., 2008) (*voir encadré*) et le thermostat de Langevin (SCHLICK, 2010b) implémenté dans NAMD (PHILLIPS et al., 2005; JIANG et al., 2014). L'idée générale de ces thermostats reste identique : obtenir un ensemble thermodynamique correct en corrigeant la température du système.

Le thermostat dit de velocity rescale est une extension du thermostat de Berendsen. Le principe est de corriger les vitesses pour que la distribution de l'énergie cinétique soit correcte.

Pour l'énergie cinétique, Berendsen correspond à

$$dK = (K_0 - K) \frac{dt}{\tau} \quad (2.3)$$

Où K est l'énergie cinétique, τ est la constante de temps et K_0 l'énergie cinétique correspondant à la température de référence. Velocity rescale ajoute un terme stochastique à cette équation :

$$dK = (K_0 - K) \frac{dt}{\tau} + 2 \sqrt{\frac{K K_0}{N_f}} \frac{dW}{\sqrt{\tau}} \quad (2.4)$$

Avec N_f le nombre de degré de liberté du système et dW un procédé de Wiener (terme stochastique lié au mouvement Brownien).

2.1.4 Contrôle de la pression

Le contrôle de la pression est conceptuellement identique au contrôle de la température, bien que considérer à quoi correspond la pression dans un système microscopique n'est pas évident. Schématiquement, le contrôle de la pression peut se faire par la méthode de Berendsen, similaire au contrôle de la température avec :

$$\frac{dP}{dt} = \frac{1}{\tau} (P_0 - P) \quad (2.5)$$

Toujours dans le soucis de simuler un ensemble thermodynamique correct que la méthode de Berendsen ne permet pas, j'ai utilisé le barostat de Parinello-Rahman (PARRINELLO et al., 1981) implémenté dans GROMACS et le barostat de Langevin (SCHLICK, 2010b) implémenté dans NAMD.

Une fois que la température et la pression sont sous contrôle, un point clé de la dynamique moléculaire est $\mathbf{F}$, les forces qui vont s'appliquer sur les particules et qui sont regroupées dans un **champ de forces**.

Le champ de forces regroupe deux composantes :

- Un nombre d'équations, utilisées pour obtenir les énergies potentielles et leurs dérivées, les forces.
- Les paramètres de ces équations.

2.1.5 Les fonctions d'interaction

La fonction d'énergie potentielle est essentielle et regroupe des termes liés et non liés qui reproduisent les interactions intra et inter-moléculaires du système. Elle se présente en général comme une somme.

Les termes liés

Les termes liés correspondent à toutes les interactions entre particules liées : étirement des liaisons covalentes, torsion des angles et angles dièdres (propres et impropres). Conceptuellement, autour d'un minimum d'énergie, ces termes peuvent être vus comme de simples ressorts (potentiels harmoniques) entre les atomes :

$$\begin{aligned} V_{bond} &= \frac{1}{2}k(r - r_0)^2 \\ V_{angle} &= \frac{1}{2}k(\theta - \theta_0)^2 \\ V_{improper} &= \frac{1}{2}k(\zeta - \zeta_0)^2 \end{aligned} \quad (2.6)$$

Bien qu'il y ait d'autres potentiels possibles (figure 2.2 pour les potentiels de liaison), ils sont tous des approximations autour d'un minimum d'énergie (SCHLICK, 2010a).

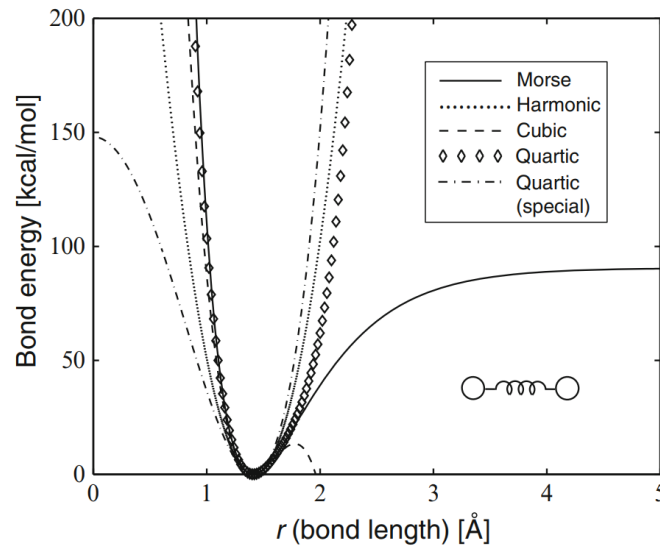


FIGURE 2.2 – Les différents potentiels de liaison covalente possibles. Le potentiel de Morse est le modèle le plus précis mais d'autres approximations, harmoniques et moins coûteuses, sont possibles (SCHLICK, 2010a).

Les termes non-liés

Les termes non liés correspondent à toutes les autres interactions et regroupent en général un terme de répulsion, un terme d'attraction et un terme électrostatique (de Coulomb). Ces deux premiers sont regroupés dans le potentiel de Lennard-Jones (JONES, 1924) :

$$V_{LJ} = 4\epsilon \left[\left(\frac{\sigma^{12}}{r} \right) - \left(\frac{\sigma^6}{r} \right) \right] \quad (2.7)$$

Le premier terme, en r^{-12} , est le terme répulsif et est la répulsion de Pauli à courte distance. Ce terme correspond à la rencontre des orbitales électroniques. Le deuxième terme, attractif, correspond à une attraction à longue distance, ou force de Van der Waals.

L'interaction de Coulomb est décrite classiquement par :

$$V_c(r_{ij}) = f \frac{q_i q_j}{r_{ij}} \quad (2.8)$$

Avec $f = \frac{1}{4\pi\epsilon_0}$, q_i et q_j les charges des particules considérées et r_{ij} la distance entre ces particules.

2.1.6 Différents types de champs de forces, à différentes échelles

En plus des différentes approximations dans les équations utilisées, les paramètres qui gouvernent les interactions entre les particules viennent aussi en plusieurs versions.

A ce jour, il n'y a pas de champ de forces universel qui pourrait être utilisé partout et qui pourrait converger d'une manière ou d'une autre. A la place, il y a différents champs de forces qui sont choisis suivant leur application et suivant les préférences de la communauté. Traditionnellement, on voudra par exemple utiliser le champ de forces CHARMM (HUANG et al., 2017) développé par Martin Karplus et al. pour travailler sur des protéines solubles, le champ de forces AMBER (MAIER et al., 2015) développé par Peter Kollman et al. pour travailler avec des acides nucléiques, le champ de forces OPLS-AA (ROBERTSON et al., 2015) développé par William Jorgensen pour travailler avec de petites molécules et le champ de forces gros grain MARTINI (MARRINK et al., 2007; MONTICELLI et al., 2008) développé par Siewert Marrink et al. pour travailler sur des lipides, membranes ou protéines membranaires. Cependant, les différents champs de forces sont de plus en plus utilisés indépendamment de la nature des objets simulés, le répertoire des molécules disponibles étant devenu la contrainte principale qui dicte son choix.

Dans ce travail, nous nous sommes concentrés sur le champ de forces gros grain MARTINI.

Le champ de forces gros grain MARTINI

Dans ce champ de forces, plusieurs atomes sont regroupés dans une seule particule (ou bille). Chaque particule représente environ 4 atomes "lourds" avec leurs hydrogènes (figure 2.3). C'est un champ de forces qui permet, grâce à la réduction du nombre de particules et au prix d'approximations (voir encadré), de calculer plus longtemps et d'atteindre des temps de simulations pertinents pour les membranes biologiques (en effet, les mouvements dans les membranes

sont plus lents et il faut calculer plus longtemps pour avoir accès à des phénomènes intéressants). Malgré ces approximations, ce champ de forces retient les propriétés de partition des membranes et permet de conserver la topographie de surface des protéines, supposée importante dans leur assemblage et leurs interactions avec les lipides.

Le champ de forces MARTINI est un champ de forces gros grain qui permet de calculer plus longtemps au prix d'approximations :

- Il est empirique, basé originellement sur la partition des lipides.
- Il a une résolution inférieure à l'atome, rassemblant plusieurs atomes dans une bille. En conséquence, certains degrés de libertés sont figés, ce qui a des implications au niveau thermodynamique, notamment du point de vue entropique. Par ailleurs, les liaisons hydrogènes sont de plus impossibles : l'eau gèle à 300 K et les structures secondaires des protéines doivent être maintenues grâce à un réseau élastique.

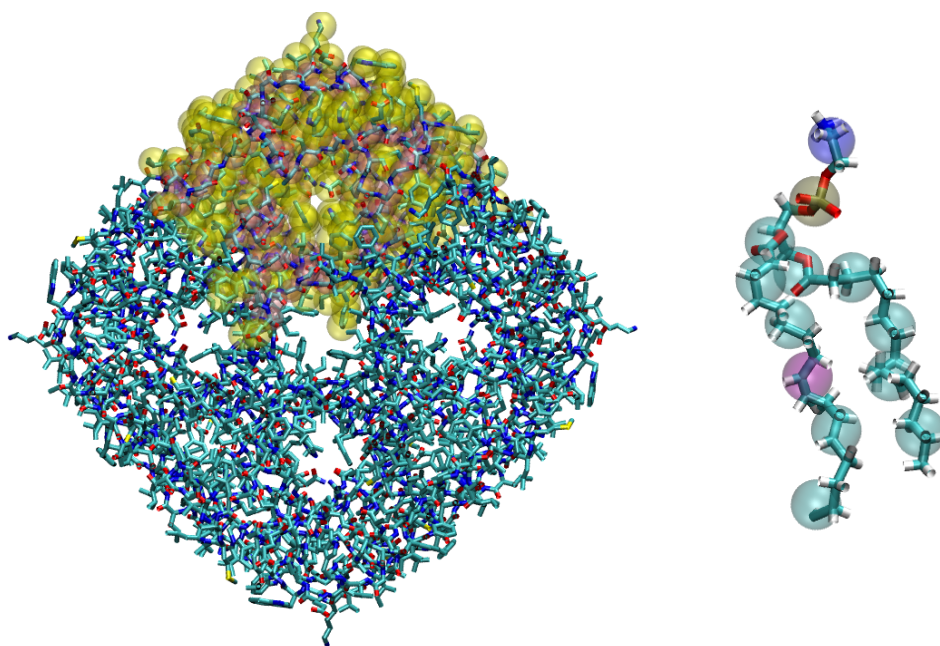


FIGURE 2.3 – Représentation tout atome (En bâtons) avec superposition gros grain (Sphères Van Der Waals transparentes) de l'AqpZ (gauche) et d'un lipide POPE (droite).

2.2 Construction des systèmes

J'ai construit des systèmes gros grains pour les logiciels de dynamique moléculaire NAMD (PHILLIPS et al., 2005 ; JIANG et al., 2014) et GROMACS (BERENDSEN, SPOEL et al., 1995 ; ABRAHAM et al., 2015 ; VAN DER SPOEL et al., 2005 ; HESS et al., 2008). Les détails des systèmes (protéine, taille, nombre

de particules) sont dans la table 2.1. Les protéines ont tout d'abord été obtenues à partir de la PDB ou de la base de données de protéines membranaires orientées (OPM) (LOMIZE et al., 2012) puis, lorsque cela a été nécessaire, les chaînes latérales manquantes ont été rajoutées grâce à SwissProt PDB Viewer (GUEX et al., 1997).

TABLE 2.1 – Systèmes gros grain MARTINI construits et simulés. La membrane complexe correspond à une membrane possédant 9 types de lipides différents détaillés plus bas. Les lipides notés .o indiquent l'ancienne version du champ de forces MARTINI. Les systèmes avec polymère ne sont pas indiqués ici.

Software	Protéine (PDB ID)	Taille ($\text{\AA}^3$)	Membrane	Particules
NAMD	Dimère de GpA (1AFO)	85x85x85	DPPC	4349
	Dimère d'AqpZ (1RC2)	200*200*100	POPE :POPG	31955
GROMACS	Dimère de GpA	85x85x85	DPPC	4272
	Mono-AqpZ	200*200*85	Complexe	29900
	Mono-AqpZ _{W14A}	200*200*85	Complexe	29917
	Tétra-AqpZ	200*200*100	Complexe	36012
		200*200*100	POPE :POPG	34867
		200*200*150	POPE.o :POPG.o	50572
	Tétra-AqpZ _{W14A}	200*200*85	Complexe	30099
	125 AqpZ	500*500*150	POPE :POPG	318790
		500*500*200	Complexe	453275
	None	210*210*75	Complexe	30735

2.2.1 Systèmes dans NAMD

CHARMM-GUI (E. L. Wu et al., 2014) est une interface graphique en ligne qui permet la construction d'un système membranaire tout-atome utilisable dans les logiciels de dynamique moléculaire les plus utilisés. La bicouche modélisée contient la protéine orientée et les deux lipides majoritaires de la membrane d'*E.coli*, les phospholipides POPE et POPG, dans un ratio respectif 4 :1 (MOREIN et al., 1996). La taille du système est de 200*200*100 $\text{\AA}^3$ et a été choisie pour éviter l'influence des protéines les unes sur les autres à travers les conditions périodiques. Les lipides sont insérés par la méthode du remplacement, qui consiste à placer des pseudo-atomes de rayon 5.4 $\text{\AA}$ autour de la protéine puis à les remplacer par les lipides qui sont dans des conformations prises aléatoirement parmi une banque de conformations générées par des simulations de dynamique moléculaire libres. Lorsque la distance lipide-protéine est inférieure à 1.2 $\text{\AA}$, le lipide est remplacé par une autre conformation jusqu'à obtenir au moins 1.2 $\text{\AA}$ entre le lipides et la protéine. La solvation et l'ionisation du système par CHARMM-GUI n'est pas importante car le système sera transformé en système gros-grain et les molécules d'eau et d'ions seront alors enlevées.

Le système est ensuite transformé en système gros grain MARTINI en utilisant les scripts fournis par l'équipe de NAMD (RAMYA GAMINI, 2009) et modifiés au laboratoire (disponibles à l'adresse <https://tinyurl.com/y75y37me>). Ces scripts consistent, dans l'ordre, à découper le système pour obtenir un ensemble système/topologie sans l'eau et les ions, transformation du système en système gros-grain, création d'une topologie gros-grain, correction de cette topologie gros-grain avec les structures secondaires en utilisant DSSP (KABSCH et al., 1983), solvation et ionisation dans 0.15 M NaCl.

2.2.2 Systèmes dans GROMACS

Le script *martinize.py* a été utilisé pour obtenir les protéines gros grains. Les membranes sont ensuite construites grâce à INSANE (WASSENAAR, INGÓLFSSON et al., 2015). Le programme génère une membrane en distribuant les billes des lipides sur une grille prédéfinie. Un tutorial permettant de générer un système membranaire avec ou sans protéine est disponible sous format Jupyter Notebook sur https://framagit.org/msidore/tutorial_insane.

Les systèmes contenant une Aquaporine Z tétramérique ont aussi deux lipides dans le trou central, un POPE du côté des charges positives et un POPE de l'autre côté. Sans ces lipides et sans réseau élastique entre les monomères, le tétramère s'écrase en comblant ce vide et forme un losange.

Une membrane complexe d'*E.coli*

Malgré une richesse de lipides différents dans la membrane interne d'*E.coli* (KITO et al., 1972), l'approximation gros grain permet une augmentation de la complexité avec seulement 9 types de lipides différents. Après avoir dupliqué et modifié la cardiolipine existante (DAHLBERG, 2007) pour obtenir, à la fin, une cardiolipine avec deux insaturations et une cardiolipine avec quatre insaturations, la membrane peut être générée par insane avec les ratios et types lipidiques présents dans la table 2.2.

Construction d'une membrane encombrée

Pour construire une membrane encombrée, la stratégie utilisée a été la duplication d'un patch de membrane encombrée.

Cette première membrane a utilisé le script *daft.py* (WASSENAAR, PLUHACKOVA et al., 2015), qui est un adaptateur pour INSANE, pour insérer 5 monomères d'Aquaporine Z dans une membrane simple (POPE :POPG) ou complexe la plus petite possible, dans un ratio en masse 1 :1 lipides :protéines. Après duplication du système (5 fois en x et 5 fois en y) et de la topologie (25 fois), nous obtenons un système encombré et prêt à être simulé.

TABLE 2.2 – Composition relative des lipides du modèle de la membrane interne gros grain d' *E. coli*.

Lipide	Notation	Billes acyles	Insaturations	Correspondance	Abondance (%)
PE	PE8 :0	8	0	32 to 36	6.2
PE	PE8 :1	8	1	32 to 36	50.2
PE	PE8 :2	8	2	32 to 36	12.5
PE	PE6 :0	6	0	24 to 28	6.2
PG	PG8 :0	8	0	32 to 36	1.3
PG	PG8 :1	8	1	32 to 36	12
PG	PG8 :2	8	2	32 to 36	3.3
CL	CL20 :4	20	4	64 to 72	3.7
CL	CL10 :2	10	2	64 to 72	4.6

Les billes acyles prennent en compte le nombre total de billes acyles du phospholipide.

Les insaturations sont le nombre total d'insaturation du phospholipide.

La correspondance informe sur la longueur de la chaîne acyle correspondante, en nombre d'atome de carbone par phospholipide.

Les Notebook permettant la création et l'analyse de cette membrane encombrée reprennent cette construction (https://framagit.org/msidore/thesis_energetics_lipids/tree/master/Notebooks/Crowded).

2.3 Suivi des orientations relatives et coordonnées de réaction dans NAMD

Le module Colvars (FIORIN et al., 2013) a été utilisé pour définir des coordonnées de réaction. Les coordonnées de réactions, ou variables collectives, utilisées ici vont permettre de contrôler la distance entre les centres de masses des monomères ainsi que leurs orientations relatives (figure 2.4).

Les variables d'angles sont particulièrement critiques dans le suivi des orientations relatives des monomères. Pour obtenir des orientations relatives non ambiguës, le premier monomère a dû être figé à l'aide d'un potentiel harmonique.

- Angle de spin ou ϕ : orientation du second monomère par rapport à sa configuration de référence.
- Angle dièdre ou θ : rotation du second monomère autour du premier.

L'utilisation de ces deux angles permet le suivi exhaustif des orientations relatives d'un monomère par rapport à l'autre (*Une discussion sur les autres possibilités est présente dans l'encadré*).

Intuitivement, pour contrôler les orientations des monomères, nous utilisons deux angles de spin. Cependant, cette possibilité masque un problème :

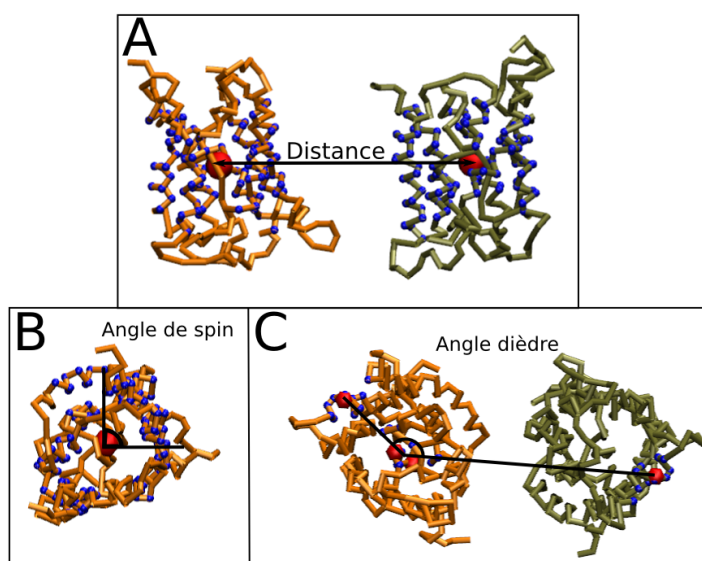
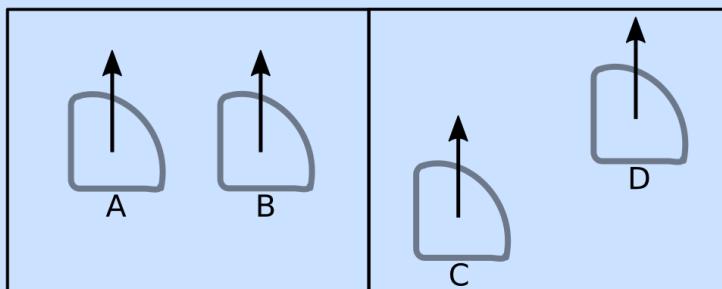


FIGURE 2.4 – Les variables collectives utilisées avec le dimère d'Aquaporine Z. Les grains en bleu représentent les grains utilisés par la variable à travers leur centre de masse, qui est lui représenté en rouge. A) Distance entre les monomères. B) L'orientation, ou angle de spin, du second monomère. C) L'angle dièdre, ou l'orientation du second monomère autour du premier.

plusieurs orientations relatives différentes auraient le même couple $\phi_1-\phi_2$ (Figure 2.3).



Orientations relatives des protéines membranaires. Les deux couples de protéines (A-B et C-D) ont ici le même couple $\phi_1-\phi_2$, les mêmes orientations, mais pas les mêmes orientations relative par rapport au second monomère.

La table 2.3 résume l'ensemble des variables collectives utilisées dans ce travail.

2.4 Intégration thermodynamique

Pour obtenir une mesure énergétique correspondant à l'énergie libre dans une simulation, il faut effectuer une intégration thermodynamique le long

TABLE 2.3 – Résumé des variables collectives utilisées.

Système	Type	Largeur	Limites	Atomes
GpA	Distance	1Å	[7 :27]Å	Squelette
Dimère d'AqpZ	Distance	5Å	[35 :65]Å	Squelette
	Spin angle	4°	[-180 :180]°	Squelette
	Angle dièdre	4°	[-180 :180]°	Squelette des hélices
	Angle dièdre	4°	[-4 :56]°	Squelette des hélices

d'un chemin de réaction. Entre deux configurations, il faut avoir accès à l'ensemble des états du système. Dans un monde parfait, nous pourrions simuler un temps infini et avoir accès à tous les états du système, même les moins probables ; dans le contexte des simulations, où le temps effectif de simulation dépasse rarement la dizaine de μ s, il est nécessaire d'utiliser des astuces thermodynamiques pour pouvoir échantillonner ces états moins probables et avoir accès à ces énergies.

Une limitation commune à l'ensemble des méthodes d'intégration thermodynamique concerne les barrières d'énergies sur des coordonnées de réaction orthogonales à celles suivies : si ces coordonnées ne sont pas moyennées durant le temps de la simulation, l'estimation du Potentiel de Force Moyenne (PMF) sera biaisée.

2.4.1 L'Adaptive Biasing Force (ABF)

L'Adaptive Biasing Force (COMER et al., 2015; DARVE et al., 2001; LELIÈVRE et al., 2008) permet d'explorer de façon plus exhaustive l'espace conformationnel accessible à un système. Après avoir défini une (ou plusieurs) coordonnée de réaction, sur laquelle vont s'appliquer des forces de biais, l'algorithme va permettre d'échantillonner plus souvent les régions de l'espace configurationnel qui ne seraient visitées qu'infréquemment mais qui sont importantes pour obtenir une estimation pertinente de l'énergie libre.

L'énergie libre le long d'une coordonnée de réaction peut être vue comme un potentiel résultant de la force moyenne agissant le long de cette coordonnée de réaction (l'opposé du gradient de ce potentiel). C'est la force moyenne qui est calculée et ensuite intégrée pour obtenir le potentiel.

Dans ce but, l'ABF utilise un biais adaptatif qui correspond à l'opposé de la force moyenne instantanée qui s'applique le long de la coordonnée de réaction étudiée : il permet au système d'échapper les pièges cinétiques dans lesquels il serait resté très longtemps (figure 2.5). De plus, ces forces instantanées moyennées sont à leur tour moyennées au cours du temps pour donner une estimation de la dérivée de l'énergie libre à chaque étape.

Rapidement, au cours de la simulation, la force moyenne qui s'applique sur la coordonnée de réaction est calculée à la volée et contrebalancée par une force

de biais inverse. A terme, la force moyenne instantanée va tendre vers 0 et le système possèdera alors un potentiel de force moyenne presque plat : le système aura alors une dynamique accélérée le long de la coordonnée de réaction, une diffusion libre, sans puits ni barrière d'énergie potentielle (figure 2.5).

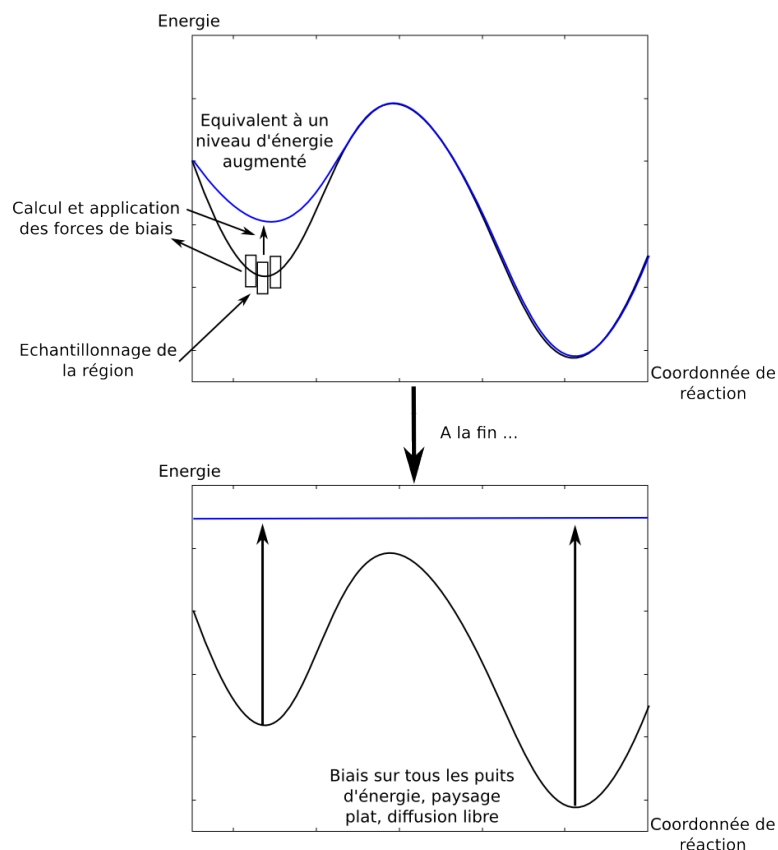


FIGURE 2.5 – Mécanisme général de l'ABF. Lorsqu'une région est échantillonnée un certain temps, un biais est calculé et appliqué sous forme de forces, ce qui revient à relever le niveau d'énergie de la configuration correspondante. A la fin, le paysage énergétique tend donc à s'aplanir et le système à avoir une diffusion libre.

Le post-traitement des forces de biais de cette simulation, via l'outil *abf_integrate*, va ensuite permettre d'obtenir le potentiel de force moyenne le long de la coordonnée de réaction.

Au total, 4 types d'ABF différents ont été mis en place :

- La dissociation de la Glycophorine A (GpA), ABF sur la distance du contact à 26Å entre le backbone des hélices
- Le dimère d'Aquaporine Z, ABF unidimensionnel sur la distance entre 35 et 65Å entre le centre de masse du squelette, avec l'angle dièdre restreint de -4 à 56°
- Le dimère d'Aquaporine Z, 5 ABF bidimensionnel sur l'angle de spin et l'angle dièdre en maintenant la distance à 40, 50, 60, 70 et 80Å entre les centres de masses du squelette

- Le dimère d'Aquaporine Z, ABF tridimensionnel sur la distance entre 35 et 65Å, l'angle de spin et l'angle dièdre entre -4 et 56°

2.4.2 Umbrella Sampling

L'Umbrella Sampling est une technique qui consiste à s'affranchir des barrières d'énergies le long d'une coordonnée de réaction pour calculer le PMF le long de celle-ci, sans ajouter de potentiel de biais. Après avoir obtenu un échantillonnage suffisant des valeurs de la coordonnée de réaction d'intérêt (par dynamique moléculaire dirigée la plupart du temps), l'Umbrella Sampling consiste à simuler librement, en parallèle et indépendamment, chaque échantillon en restreignant la variable d'intérêt à sa valeur par un potentiel harmonique. En simulant en parallèle toutes les valeurs de la coordonnée de réaction, on s'affranchit des barrières d'énergie présentes sur celle-ci.

Ainsi, en considérant l'ensemble de ces simulations, nous obtiendront un échantillonnage de l'espace configurationnel de la coordonnée de réaction exhaustif (figure 2.6), composé des échantillonnages chevauchant des simulations indépendantes.

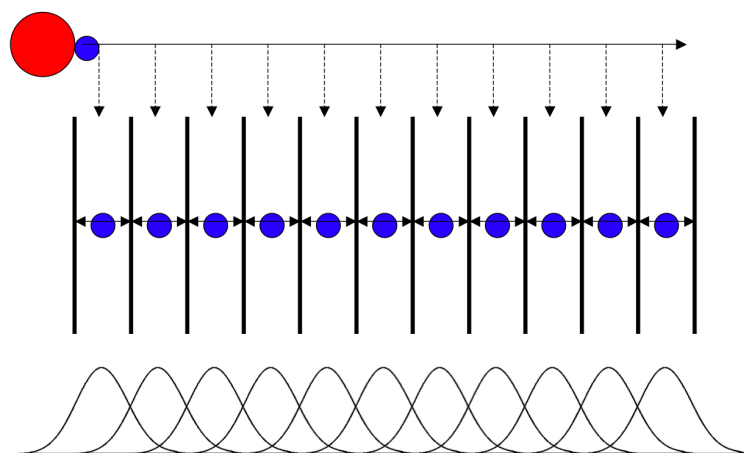


FIGURE 2.6 – Mécanisme général de l'Umbrella Sampling. Le cercle rouge est le point initial de l'Umbrella Sampling. Chaque simulation indépendante (point violet) est restreinte autour d'une valeur de la coordonnée de réaction et va ainsi échantillonner son espace proche. L'histogramme idéal des configurations, en bas, est constitué par l'ensemble des simulations indépendantes dans des fenêtres chevauchantes. Image par Justin Lemkul <http://www.bevanlab.biochem.vt.edu>.

Une fois que l'échantillonnage est assez chevauchant pour permettre une intégration (bas de la figure 2.6 pour un cas idéal), la méthode d'analyse d'histogrammes pondérés, ou WHAM (KUMAR et al., 1992), est utilisée pour obtenir le PMF.

2.4.3 Dynamique moléculaire dirigée

La dynamique moléculaire dirigée (Steered Molecular Dynamics, SMD) est utilisée pour parcourir un système le long d'une coordonnée de réaction. Plus classiquement, cette coordonnée est une distance et la SMD est utilisée pour éloigner deux molécules en appliquant des forces. Bien qu'il soit possible de retrouver des énergies libres à partir de SMD, notre utilisation de cette technique s'est restreinte à la préparation des systèmes pour l'Umbrella Sampling, en obtenant des snapshots des systèmes à différentes distances entre nos protéines d'intérêt.

2.5 Paramètres des simulations

2.5.1 MARTINI dans NAMD

Les simulations de MARTINI dans NAMD ont utilisé NAMD version 2.10 (avec l'Aquaporine Z) ou 2.11 (avec la Glycophorine A) avec la dernière version du module colvars (FIORIN et al., 2013) avec le champ de forces gros grain MARTINI 2.0 (MARRINK et al., 2007) et son extension 2.1 pour les protéines (MONTICELLI et al., 2008), avec l'approche de réseau élastique ElNeDyn.

Chaque système a été équilibré rapidement avec 4 ns de NVT et 4 ns de NPT avant d'être simulé, pour continuer l'équilibration, pendant 0.5 à 1 μ s, dans un ensemble thermodynamique NPT à 325 K. La température a été maintenue grâce à un contrôle de la température de Langevin (constante de temps de 1 ps). La pression a été maintenue à 1 bar en utilisant un barostat de Nosé-Hoover couplé à la dynamique de Langevin avec des constantes de temps de 2 et 1 ps. Le temps d'intégration a été limité à 10 fs, à la place du classique 20 fs pour les systèmes gros grains MARTINI, pour limiter les crash dues aux instabilités numériques résultantes de l'import de MARTINI dans NAMD (voir section 9).

Un fichier de configuration type commenté est présent en annexe.

2.5.2 MARTINI dans GROMACS

Toutes les simulations ont utilisé GROMACS version 5.1.4 (BERENDSEN, SPOEL et al., 1995; ABRAHAM et al., 2015). Après une courte minimisation, les systèmes ont été équilibrés 1 ns dans l'ensemble NVT puis 1 à 4 ns dans l'ensemble NPT. Le contrôle de la température de Berendsen a permis d'éliminer les fluctuations importantes de la température au début de l'équilibration. Les simulations ont ensuite utilisé les paramètres correspondant au "New-RF" (DE JONG et al., 2016) précédemment optimisé. Ces paramètres utilisent un temps d'intégration de 20 fs, la liste des voisins est mise à jour toutes les 20 étapes suivant un "cut-off" automatiquement déterminée. Les potentiels et forces de Lennard-Jones et de Coulomb sont stoppés à 1.1 nm, avec les

potentiels déplacés à 0 au cut-off. Les schèmes de couplage de Velocity rescale (BUSSI et al., 2008) et de Parinello-Rahman (PARRINELLO et al., 1981) ont été utilisés avec des paramètres de couplage de 1 et 12 ps⁻¹. L'intégrateur leapfrog a été utilisé avec une température et une pression constantes de 325 K et 1 bar respectivement.

2.6 Automatisation et Analyses

La plupart des analyses et constructions de cette thèse ont été automatisés, lorsque cela est possible, par des scripts, commentés de manière extensive, en python, bash ou C/C++ et dans le cahier de laboratoire électronique, pouvant accepter du code, Jupyter Notebook (KLUYVER et al., 2016). Chaque Notebook est commenté et est disponible à l'adresse https://framagit.org/msidore/thesis_energetics_lipids/Notebooks.

Les scripts python utilisent numpy, matplotlib (HUNTER, 2007) pour générer des graphes et MDAnalysis (MICHAUD-AGRAWAL et al., 2011) pour la gestion des trajectoires.

L'ensemble des scripts est disponible sous la license CRAPL, license encourageant le partage de code, même moche, dans le milieu académique (<https://matt.might.net/articles/crapl>).

2.6.1 Gestion des fichiers .pdb, .gro et .ndx en python

J'ai développé une bibliothèque permettant de gérer les fichiers .pdb, .gro et surtout .ndx suivant mes besoins. Cette bibliothèque, nommée humoristiquement "Friand_Dur", est disponible à l'adresse https://framagit.org/msidore/thesis_energetics_lipids/Friand_Dur.

2.6.2 Post-traitement des trajectoires

Les trajectoires GROMACS ont été traitées avec les outils fournis. En particulier, trjconv a été utilisé de manière extensive pour centrer le système ou pour s'affranchir des conditions périodiques. Chaque traitement est précisé dans le Notebook correspondant.

2.6.3 Calcul de l'angle de spin

Dans le cas où l'angle de spin n'a pas été défini dans l'utilisation du module colvars, ou simplement dans le calcul de l'angle de spin dans une simulation utilisant GROMACS, j'ai développé un package Python/C pour suivre l'angle de spin (voir annexe B), disponible à l'adresse <https://framagit.org/msidore/spinpack> sous la licence CRAPL.

Brièvement, une matrice de rotation est obtenue à partir de la superposition de la structure sur la structure initiale puis est utilisée pour peupler une matrice 4*4 caractéristique de la rotation. La diagonalisation de cette matrice nous donne ensuite le vecteur propre correspondant à la valeur propre maximale, qui est un quaternion équivalent à la rotation caractéristique. Ce quaternion $Q(w, x, y, z)$ est ensuite décomposable en angle de spin et angle de tilt, deux composantes de l'orientation, et nous récupérons l'angle de spin.

2.6.4 Passage d'un PMF tridimensionnel à un PMF unidimensionnel

Pour passer d'un PMF tridimensionnel à un PMF unidimensionnel, il faut revenir aux populations présentes dans chaque tranche du PMF unidimensionnel. Nous pouvons revenir à la densité de population entre états à l'équilibre grâce à l'équation :

$$\omega(\zeta) - \omega(\zeta') = -k_B T \ln \frac{P(\zeta)}{P(\zeta')} \quad (2.9)$$

Où $\omega(\zeta)$ est le PMF dans l'état ζ ; $\frac{P(\zeta)}{P(\zeta')}$ est la densité d'états $\frac{\zeta}{\zeta'}$; k_B est la constante de Boltzmann et T est la température.

Cette analyse est disponible dans le Jupyter Notebook "From 3D to 1D PMF" (https://framagit.org/msidore/thesis_energetics_lipids/tree/master/Notebooks/Energetics).

2.6.5 Densités de lipides

Les densités de lipides sont calculées grâce au module volmap de VMD (HUMPHREY et al., 1996). Lorsque le calcul des densités est inclus dans une autre analyse, il sera présent explicitement dans celle-ci.

Une normalisation des densités est aussi parfois effectuée par le nombre de billes présentes dans chaque densité dans le but de pouvoir les comparer si elles proviennent de lipides différents. Cette normalisation est disponible à l'adresse https://framagit.org/msidore/thesis_energetics_lipids/scripts/density/normalize_density.py.

2.6.6 Comptage des lipides autour de la protéine

Des scripts TCL dans VMD ont été utilisés pour compter les lipides. Ces scripts, ainsi que les scripts python utilisés pour analyser ces comptages, sont disponibles à l'adresse https://framagit.org/msidore/thesis_energetics_lipids/scripts/tri_des_lipides/.

2.6.7 Temps de résidence des lipides

Les temps de résidence des lipides au niveau de l'interface du tétramère d'Aquaporine Z ont utilisé un comptage du nombre de frame dans lequel un lipide restait au même endroit.

Un "patch" d'interaction à la jonction des monomères a été défini par les résidus 2, 3, 9, 13 et 226 pour le tétramère WT, par les résidus 2, 3, 9, 13, 226, 10, 13, 14, 17, 122, 140 pour le mutant $W_{14}A$. Le patch d'interaction correspondant aux lipides coincés entre les interfaces du mutant $W_{14}A$ est défini par les résidus 50 et 172.

L'analyse consiste en un comptage des lipides présents à chaque "frame" dans un rayon de 6 Å autour des résidus ci-dessus en gardant en mémoire la présence de ces lipides dans les "frame" précédentes.

Cette analyse est regroupée dans le Jupyter Notebook "Residency times at specific locations around AqpZ" (https://framagit.org/msidore/thesis_energetics_lipids/tree/master/Notebooks/Tri_des_lipides).

2.6.8 Création d'un polymère gros grain amphiphile

La création du polymère gros grain a nécessité, dans l'ordre :

- Une version modifiée de auto-martini
- L'obtention d'une chaîne de caractère aléatoire de S et M rendant compte des distributions expérimentales
- Des classes pythons permettant de gérer un champ de forces
- La génération d'un fichier de topologie et de coordonnées

Le Jupyter Notebook "Coarse grained SMA" contient l'ensemble de ces étapes ainsi que les paramètres et scripts nécessaires pour l'équilibrer dans l'eau (https://framagit.org/msidore/thesis_energetics_lipids/tree/master/Notebooks/Tri_des_lipides).

2.6.9 Analyse d'une membrane encombrée

Visualisation en "streamlines"

Pour visualiser les mouvements de lipides dans un grand système, nous avons utilisé une approche utilisant des lignes de flux (streamlines), initialement développé en météorologie puis étendu aux systèmes membranaires (CHAVENT, REDDY et al., 2014).

Une partie importante des scripts fournis pour faire cette représentation ont du être retravaillés ou refaits. Cette analyse se trouve dans le Notebook "Crowded membrane - lipid streamlines analysis" et est disponible dans le script `lipid_meteo.py` à l'adresse https://framagit.org/msidore/thesis_energetics_lipids/scripts/crowded/lipid_meteo.py.

Densités des lipides autour de chaque monomère

Les densités de lipide autour de chaque monomère sont obtenues après "fitting" de chaque monomère sur un monomère de référence, qui est le premier monomère de la première "frame". Le Notebook "Crowded membrane - Analysis protocol" détaille chaque étape (https://framagit.org/msidore/thesis_energetics_lipids/tree/master/Notebooks/Crowded).

Analyse des orientations

Les orientations entre chaque paire de monomères ont été obtenues grâce aux calculs de l'angle de spin décrits précédemment. Le script permettant cette analyse est disponible à l'adresse https://framagit.org/msidore/thesis_energetics_lipids/blob/master/scripts/crowded/orientation_analysis.py.

Comptage des interfaces en contact

Les interfaces du monomère de l'Aquaporine Z ont été définies selon les résidus à sa surface :

- La face 1 correspond aux résidus 68, 70, 71, 73, 74, 78, 79, 80, 83, 90, 255, 274, 280, 282, 286, 287, 288, 294, 296, 313, 317, 318, 320, 322, 331, 332, 334, 335, 337, 338, 340, 342, 344, 350, 352, 354, 358, 360, 446, 450, 461, 463, 465, 467, 470
- La face 2 correspond aux résidus 16, 21, 32, 41, 45, 49, 50, 55, 57, 60, 85, 91, 97, 99, 101, 104, 108, 109, 196, 204, 206, 208, 211, 212
- La face 3 correspond aux résidus 139, 144, 149, 150, 151, 155, 159, 161, 240, 244, 256, 257, 261, 263, 265, 267, 270, 272, 278, 404, 406, 412, 417, 425, 430, 432, 436, 438, 440, 442, 444, 448
- La face 4 correspond aux résidus 1, 3, 7, 13, 15, 16, 19, 162, 165, 167, 169, 171, 172, 176, 178, 180, 183, 185, 187, 189, 191, 195, 196, 200, 204, 222, 224, 225, 382, 388, 390, 394, 396, 397, 404

Le script de comptage est disponible à l'adresse https://framagit.org/msidore/thesis_energetics_lipids/scripts/counting_the_interfaces.py.

La matrice d'interaction résultante de ce comptage a ensuite été transformée en matrice PMF-like en considérant que chaque comptage correspondait à une population grâce à l'équation, pour chaque état :

$$E_i = -\frac{\ln(x_i)}{S} \quad (2.10)$$

Avec $S = \sum_i x_i$

2.6.10 Développement d'un logiciel de simulation de Monte-Carlo

Se basant sur une version précédente développée par James Sturgis, j'ai contribué à l'amélioration d'un logiciel de simulation de Monte-Carlo permettant de simuler des disques dans un espace 2D avec un potentiel simple (J. P. DUNEAU et al., 2013). Plus particulièrement, l'objectif a été de passer le code en C++ ainsi que l'utilisation d'objets pour permettre de simuler plus que des disques simples. Le code est disponible à l'adresse https://github.com/jnsturgis/hard_discs2.

2.6.11 Fonctions de distribution radiale

Pour analyser l'organisation des monomères à la fin des simulations de membrane encombrée, un résidu central du monomère (Asn-186) a été sélectionné comme particule de référence pour mesurer des fonctions de distribution radiale.

Afin de comparer ces résultats avec ceux des simulations de Monte-Carlo, les configurations finales ont été chargées dans VMD (format natif XYZ) pour être analysées.

Ces fonctions ont été calculées en utilisant un script .tcl sous VMD.

Section 3

Thermodynamique de l'assemblage d'un dimère d'AqpZ

La formation des complexes fonctionnels de protéines membranaires, indispensables à la vie de la cellule, reste cependant difficile à décrire. Pour beaucoup de protéines, nous connaissons les surfaces d'interactions et d'autres déterminants observés dans les structures expérimentales. Cependant, nous n'avons que peu d'informations sur le processus d'assemblage en lui-même et sur les forces qui sont responsables de cet assemblage.

Le processus d'assemblage requiert une proximité entre les protéines qui doivent interagir mais aussi un ensemble de forces pour ne pas former d'agrégats de protéines. Expérimentalement, la dispersion et le clustering des protéines membranaires ont été observés avec des expériences de transfert d'énergie par résonance en fluorescence, de tracking de molécules uniques ou de microscopie à force atomique à haute vitesse (HS-AFM) (BADER et al., 2009; CHADDA et al., 2016; HYDE et al., 2012). Les organisations observées dépendent de la thermodynamique sous-jacente ; cependant, la mesure directe de ces interactions est rare et limitée à l'étude de peptides transmembranaires dans des micelles de détergent (EBIE et al., 2007; FLEMING et al., 2004; L. E. FISHER et al., 2003; L. E. FISHER et al., 1999), avec quelques explorations de la thermodynamique des interactions entre des peptides dans des bicouches (CHEN, MERZLYAKOV et al., 2009; CHEN, NOVICKY et al., 2010).

Pour des protéines plus complexes, les études sont encore plus rares. Dans une étude précédente sur la protéine OmpF, une analyse statistique de HS-AFM a révélé des orientations préférentielles entre des protéines distantes (CASUSO et al., 2012). Ces études restent limitées car la résolution spatiale et temporelle nécessaire est difficilement accessible ; malgré cela, des interactions à longue distance semblent importantes.

Due aux difficultés pour obtenir des informations expérimentales, les simulations constituent une voie complémentaire pour comprendre les interactions moléculaires des protéines membranaires. Pour cela, l'approche

classique consiste à calculer la variation d'énergie libre liée à l'agrégation de deux protéines. Ces approches utilisent une variété de méthodes (PRAKASH et al., 2011) d'intégration thermodynamique, avec particulièrement la plus utilisée, l'Umbrella Sampling (PRAKASH et al., 2010; PRASANNA et al., 2013; SENGUPTA et al., 2010; KUZNETSOV et al., 2015; JANOSI et al., 2010) (US). Ces méthodes d'intégration forcent l'échantillonnage le long de la coordonnée d'intérêt, souvent la distance entre les deux protéines d'intérêt, avec la nécessité d'échantillonner toutes les autres coordonnées, orthogonales, suivant la fonction de partition du système. Dans ces intégrations donc, seule la distance entre les protéines est habituellement contrôlée pour être échantillonnée correctement. Les autres coordonnées, non contrôlées par l'algorithme, sont alors simplement hypothétisées comme étant moyennées et donc suffisamment échantillonnées, principalement dû au fait qu'il est difficile, si ce n'est impossible, de connaître à priori les coordonnées les plus lentes. Cependant, nous avons de bonnes raisons d'émettre l'hypothèse que certains degrés de liberté se moyennent lentement, par exemple lors des dernières étapes de l'association entre deux protéines membranaires où les lipides adsorbés ont une diffusion réduite.

Ainsi, après l'investigation du profil d'énergie libre d'OmpF (CASUSO et al., 2012), et bien que la distance entre les deux protéines soit correctement échantillonnée, leurs orientations étaient peu explorées. Pour évaluer l'importance de ce paramètre, nous nous tournons ici vers un autre algorithme d'intégration thermodynamique, l'ABF, qui va permettre de parcourir l'espace selon la distance mais aussi selon les orientations, sans avoir besoin de connaître la rugosité du potentiel sous-jacent.

La méthode ABF est actuellement implémentée seulement dans le logiciel de dynamique moléculaire NAMD. Pour combiner le champ de force gros grain MARTINI avec NAMD, nous avons donc utilisé une implémentation de ce dernier. Une discussion sur cette implémentation est présente dans ce manuscrit dans les sections 7 et 9.

Dans ce cadre et en utilisant deux sous-unités de notre protéine modèle, l'AqpZ, j'ai tout d'abord étudié le comportement de ces protéines dans une membrane simple de POPE/POPG et plus particulièrement au travers de leurs orientations relatives, avant d'investiguer le paysage énergétique des orientations relatives entre les protéines de ce système.

3.1 Résultats

Rappel des méthodes

Les orientations relatives entre deux protéines sont suivies dans cette section grâce à deux variables :

- θ , un angle dièdre, qui correspond à la rotation du second monomère autour du premier
- ϕ , un angle de spin, qui correspond à la rotation du second monomère autour de son axe

La distance entre les centres de masses des monomères est aussi suivie et appelée d .

De plus, les simulations de cette partie ont été effectuées avec NAMD.

3.1.1 Deux protéines censées interagir dans une membrane simple

Vérification de la stabilité du dimère

Avant d'interpréter le comportement de deux monomères d'AqpZ éloignés dans une membrane, il faut tout d'abord s'assurer que le dimère entier est stable (même si l'instabilité du dimère et la non interaction des monomère poserait plus de problème au niveau biologique et/ou au niveau de la fiabilité des champs de forces).

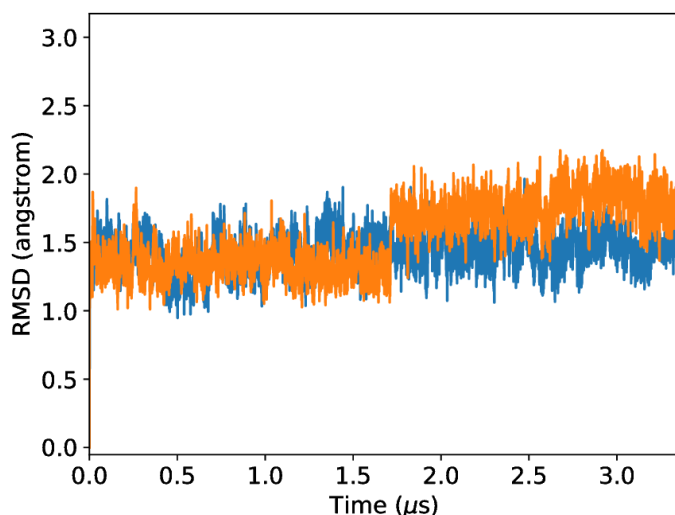


FIGURE 3.1 – RMSD de chaque monomère du dimère d'AqpZ, en orange et bleu.

Le RMSD (Root Mean Square Deviation) de chaque monomère du dimère (figure 3.1) reste stable sur l'échelle de la simulation. Nous observons cependant un petit changement de conformation unique qui semble se répercuter d'un monomère à l'autre. Bien qu'ayant un air de coopérativité, le caractère

unique de l'évènement, la présence du réseau élastique ainsi que l'existence des problèmes liés à l'utilisation de MARTINI dans NAMD ne nous permettent pas de le commenter d'avantage. Nous pouvons donc dissocier ce dimère et étudier le comportement des deux protéines.

Deux monomères qui s'influencent librement

En fixant la distance à 80 Å et en laissant les protéines évoluer librement, leurs orientations peuvent être suivies grâce à un angle de spin qui définit la rotation de chaque monomère par rapport à un axe qui passe par son centre et qui est orienté suivant la normale au plan de la membrane (figure 3.2, gauche).

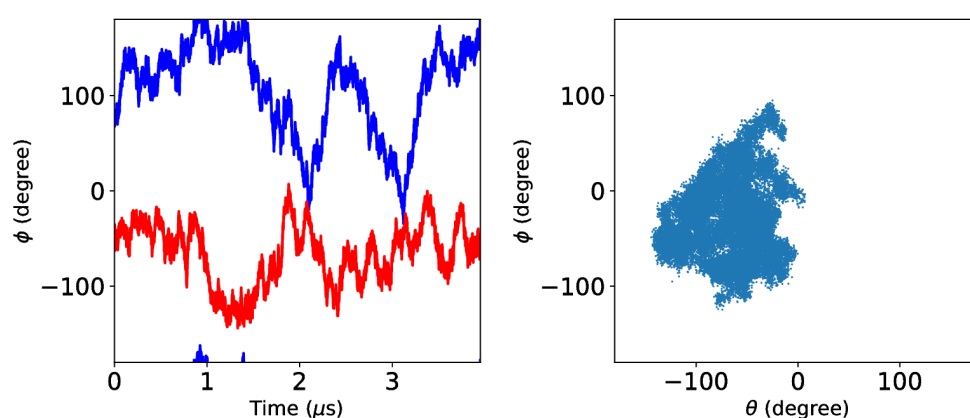


FIGURE 3.2 – Gauche. Angle de spin, θ , de chaque monomère au cours de la simulation. Droite. Orientations relatives, θ et ϕ , échantillonnés lors de la simulation.

Sur l'échelle de la microseconde, les orientations possibles entre les deux monomères ne s'échantillonnent que très partiellement, ce qui sous-entend une possible influence des protéines entre elles à une distance où elles ne sont pas censées interagir.

Cependant, cette représentation des orientations individuelle ne rend pas compte des orientations relatives. Grâce aux deux angles ϕ et θ (définis dans les méthodes), l'échantillonnage de ces orientations peut être suivi sur un graphique en deux dimensions (figure 3.2, droite). Sur 3 μ s, dans une simulation à une distance prédéfini, l'échantillonnage des orientations relatives reste partiel. Le même comportement est observé à une distance de 70 Å (données non montrées). Cela illustre bien que l'orientation relative est une grandeur qui spontanément ne s'échantillonne pas suffisamment pour donner une intégration thermodynamique adéquate. j'ai donc réalisé des simulations à distances constantes en mettant les deux coordonnées d'orientation sous le contrôle de l'algorithme d'Adaptive Biasing Force (ABF) afin d'accéder à l'énergétique sous-jacente.

3.1.2 Orientations préférentielles à distance constante

Rappel des méthodes

Idéalement, si nous avions accès à la fonction de partition du système (si nous pouvions simuler assez longtemps pour que le système échantillonne ses états selon sa fonction de partition), nous aurions directement accès aux potentiels de force moyenne de ce système. En réalité, à cause principalement des échelles de temps limitées des simulations, nous devons utiliser des "astuces" thermodynamiques pour avoir accès au potentiel sous-jacent.

L'Adaptive Biasing Force (ABF) est un algorithme d'intégration thermodynamique qui va permettre d'obtenir l'énergétique le long d'une ou plusieurs coordonnées de réactions. Plus simplement, lorsque le système est bloqué dans un puits d'énergie, l'algorithme va appliquer des forces qui vont faire sortir le système de ce puits pour aller échantillonner le reste de l'espace configurationnel.

Chaque simulation a duré 20 μ s et a permis d'obtenir des PMF le long de θ et ϕ avec une distance de 40, 50, 60, 70 et 80 Å (figure 3.3).

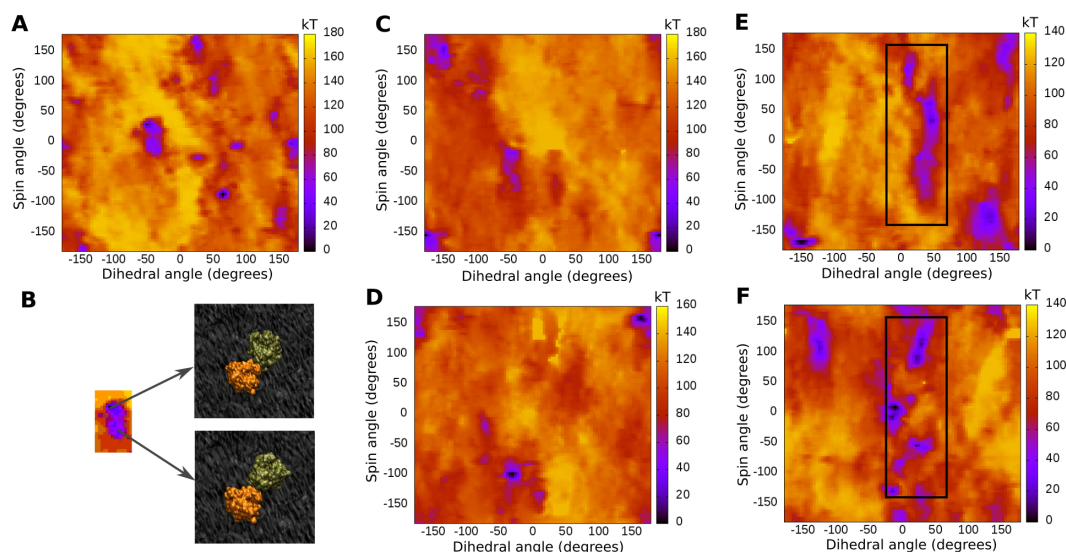


FIGURE 3.3 – PMF utilisant l'ABF sur les deux variables collectives angulaires, θ et ϕ . La distance est fixée à 40 (A), 50 (C), 60 (D), 70 (E) et 80 Å (F). (B) montre les orientations relatives correspondantes aux positions distales du puits d'énergie principal de la carte à 40 Å. Les contours noirs de (E) et (F) représentent la tendance de conservation relative des deux dernières cartes.

De manière surprenante, mais comme attendu après l'observation des simulations spontanées très contraintes, l'intégration thermodynamique a mis à jour un Potentiel de Force Moyenne (PMF) granuleux. Chaque paysage énergétique est constellé de zones de plus basse énergie potentielle qui correspondent à des orientations relatives préférées entre les monomères. A courte distance

(figure 3.3A), ce PMF montre un puits d'énergie qui correspond à la rotation du second monomère autour de son axe quand les deux pointes des protéines interagissent, ce qui serait potentiellement une partie d'un chemin d'assemblage. Ces cartes montrent aussi, qualitativement, une conservation avec déplacement des puits d'énergie de différents PMF (3.3E et 3.3F), ce qui indique la présence probable de vallées énergétiques reliant ces cartes.

Les maxima énergétique de ces cartes sont très élevés, ce qui traduit une très grande difficulté pour échantillonner ces conformations. Plus de la moitié du temps de simulation a été consacré à cet échantillonnage qui reste probablement insuffisant. En conséquence, ces grandes valeurs ne peuvent pas être considérées comme justes et l'interprétation de ces cartes restera qualitative.

3.1.3 Un ABF à trois dimensions

Pour comparer ces PMF, idéalement, il faudrait appliquer l'ABF sur les trois variables. Un ABF à deux dimensions doit remplir $90 \times 90 = 8100$ fenêtres d'échantillonnage; un ABF à trois dimensions, même avec une fenêtre aussi large que $5 \text{ \AA}$ pour la distance, nécessiterait $90 \times 90 \times 8 = 64800$ fenêtres d'échantillonnage et donc environ $160 \text{ }\mu\text{s}$ de temps de simulation pour le même échantillonnage que ces cartes 2D (qui en manquent déjà beaucoup). Une vitesse maximale de $0.6 \text{ }\mu\text{s/jour}$ rend cette simulation inenvisageable.

Pour surmonter ce problème, et comme nous voulons établir la possibilité d'explorer ce PMF, j'ai limité l'exploration de θ entre -2 et 54° et d entre 35 et $65 \text{ \AA}$ par pas de $5 \text{ \AA}$, ce qui laisse "seulement" 5040 fenêtres d'échantillonnage à explorer. θ étant la variable la plus lente à s'échantillonner (figure 3.2), cette réduction devrait accélérer grandement l'échantillonnage.

L'exploration de la distance commence à $35 \text{ \AA}$ ce qui est plus grand que la distance initiale dans le dimère natif, pour s'affranchir de problèmes liés à l'exploration des états où des lipides pourraient être limitants. En effet, nous avons de bonnes raisons de penser qu'il est difficile d'explorer les dernières étapes de l'assemblage, impliquant une diffusion diminuée et l'ancrage des lipides entre les surfaces, dans nos échelles de temps; s'affranchir de ces étapes permet d'éviter tout problème lié à ces degrés de libertés non moyennés et de se concentrer sur les étapes précoces de l'assemblage, là où des interactions à longue portée pourraient jouer un rôle.

Le résultat de cette simulation de $25 \text{ }\mu\text{s}$ est montré dans la figure 3.4 qui représente des coupes du PMF à différentes distances.

Ce PMF est toujours granuleux. L'espace conformationnel est aussi hautement structuré et cohérent à travers les différentes distances. En effet, en commençant éloigné, le potentiel est majoritairement homogène malgré des zones de plus haute énergie. Quand la distance diminue, le potentiel devient hétérogène et des puits d'énergies se définissent jusqu'à une distance de $40 \text{ \AA}$

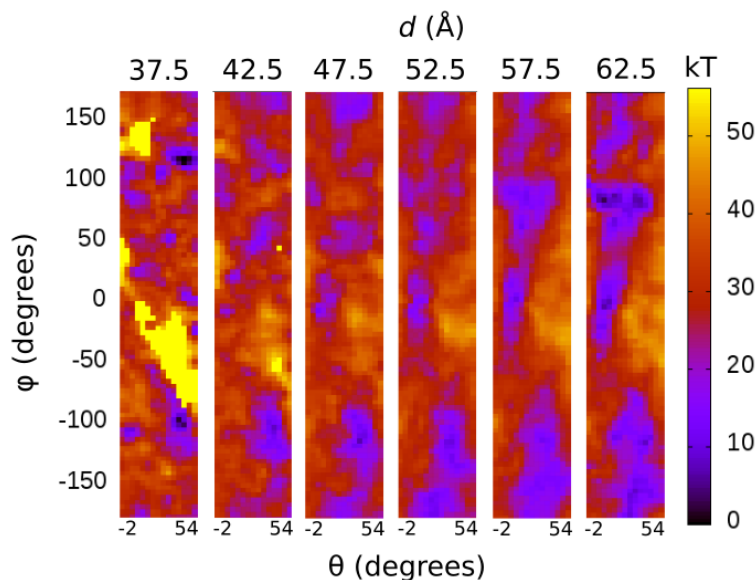


FIGURE 3.4 – Coupes du PMF tri-dimensionnel utilisant les trois variables collectives. Les distances affichées sont les centres des fenêtres d'échantillonnage - par exemple, la carte à 37.5 Å représente la fenêtre de 35 à 40 Å. Chaque point de ces cartes représente une orientation relative entre les monomères.

où presque tout le potentiel est constellé de puits d'énergies séparés par des barrières. La première carte montre une zone non échantillonnée due à des gêne stériques entre monomères.

La gamme d'énergies de ce PMF, jusqu'à ~56 kT, est aussi plus basse que les ~180 kT des simulations précédentes, ce qui montre qu'un meilleur échantillonnage des configurations les plus défavorables permet de diminuer la valeur de ces hautes énergies. Dans l'ensemble, la variation d'énergie associée à ces cartes suggère que l'association s'accompagne du passage d'une barrière d'énergie potentielle.

La réduction de cette carte en une dimension (figure 3.5) confirme l'existence de cette barrière et montre une variation d'énergie de quelques kT seulement, ce qui est conforme aux échelles d'énergie habituellement publiées pour ce type de simulation et est une indication de convergence. Cela illustre que ces état spontanément peu probables ne contribuent pas à l'énergétique de l'association du dimère lorsqu'elle est suivie sur une seule dimension. Les états qui comptent le plus pour les calculs de thermodynamique sont donc suffisamment échantillonnés même si la valeur maximale du PMF reste haute.

Réduction du PMF à plusieurs dimensions sur une seule peut se faire en

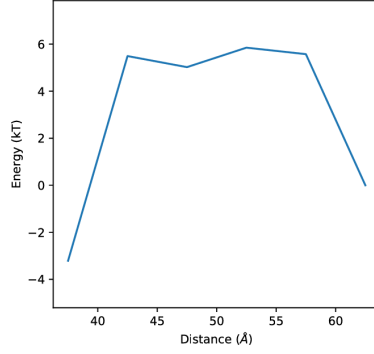


FIGURE 3.5 – PMF 3D collapsé à une dimension selon la distance. Le zéro est arbitrairement positionné à la plus longue distance.

revenant aux populations grâce à :

$$\omega(\zeta) - \omega(\zeta') = -k_B T \ln \frac{P(\zeta)}{P(\zeta')} \quad (3.1)$$

Où $\omega(\zeta)$ est le PMF dans l'état ζ ; $\frac{P(\zeta)}{P(\zeta')}$ est la densité d'états $\frac{\zeta}{\zeta'}$; k_B est la constante de Boltzmann et T est la température.

Une observation particulière du PMF 3D est qu'il y a plusieurs puits d'énergie équivalents séparés par des barrières énergétiques. Dans le contexte classique de simulations à une dimension, l'échantillonnage de ces puits doit être compromis.

3.1.4 Comparaison avec un PMF à une dimension

Nous avons réalisé une étude ABF classique en appliquant les forces de biais uniquement sur la distance. Comme dans la simulation précédente, l'espace conformationnel de l'angle θ est limité entre -2 et 54° , ce qui nous permettra de comparer le PMF de cette simulation ABF 1D, classique, avec le PMF réduit à une dimension de la simulation précédente. Le PMF obtenu après simulation de 20 μ s est montré sur la figure 3.6.

Avec uniquement 6 fenêtres échantillonnée, cette simulation a convergé en quelques microsecondes. L'échelle d'énergie est similaire au précédent PMF réduit. Cependant, par rapport aux autres PMF de dissociation obtenus avec d'autres intégrations thermodynamiques sur d'autres systèmes, il n'y a pas de plateau à la distance maximale, ce qui indique qu'une dimension de 20 nm de coté n'est pas suffisante pour avoir une situation où les protéines n'interagissent pas. De plus, aux petites distances, nous n'avons pas le puits d'énergie qui devrait correspondre au dimère natif car, comme indiqué précédemment, nous avons pris le parti de commencer nos simulations au point de premier contact

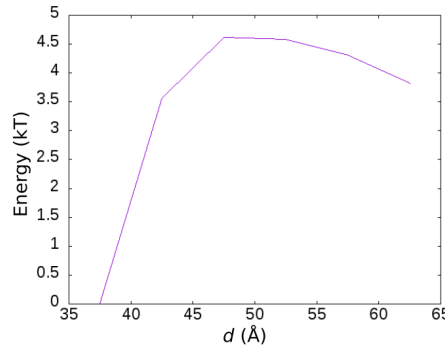


FIGURE 3.6 – PMF unidimensionnel selon la distance.

protéine-protéine. Au final, le profil de ce PMF apparaît conforme à ce que l'on pouvait attendre d'un système tel que le notre et nous permet d'étudier l'échantillonnage spontanée des orientations relatives.

La figure 3.7 montre l'échantillonnage des deux variables angulaires qui n'ont pas été biaisées par l'ABF lors de cette simulation.

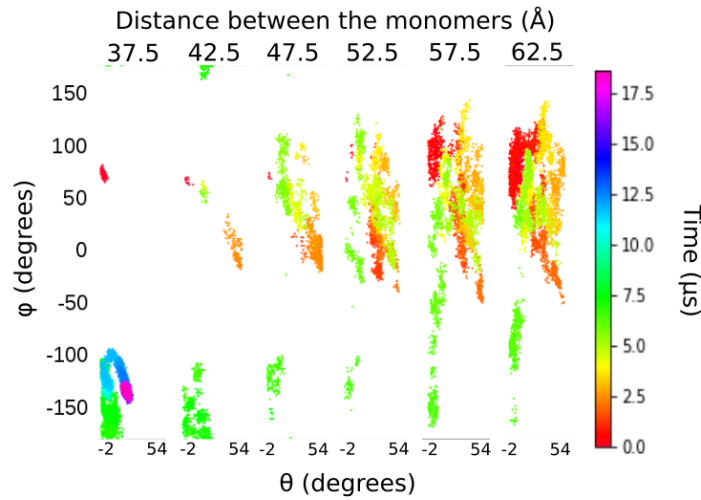


FIGURE 3.7 – Coupes de l'échantillonnage des orientations relatives au cours de la simulation d'ABF unidimensionnel.

Les orientations relatives échantillonnées au cours de cette simulation sont très limitées et se regroupent seulement autour d'une gamme d'angle restreinte. Bien que la simulation commence à 37.5 Å ($\phi = 75$, rouge), les premières réorientations spontanées ne se font qu'après 55 Å. On observe de plus, au cours de la simulation, des va-et-viens avec réorientations entre les différentes distances. Pour revenir dans la zone de contact, il a fallu attendre 7 à 10 μ s de simulation et une réorientation non contrôlée ($-180 < \phi < -100$, vert à violet). La grande majorité de l'espace des orientations n'est cependant pas

exploré dans nos échelles de temps.

Dans cette simulation, qui ne contrôle que la distance entre les monomères, il y'a donc au moins 2 coordonnées de réaction qui ne sont pas correctement échantillonnées durant la simulation.

A ce niveau, nous pouvons affirmer le biais présent dans les PMF classiques, mais surtout, et bien que cette conclusion ne soit que qualitative, l'interaction des protéines au travers de la membrane jusqu'à des longue distance est maintenant difficile à réfuter dans le cadre de ce type de modélisation.

Il y a donc des forces qui se propagent dans la membrane, à partir de la surface de la protéine et jusqu'à une distance bien supérieure aux "cut-off" des simulations. Ces forces doivent donc se propager grâce aux lipides, seul élément présent possible. La deuxième partie de cette thèse s'intéresse donc aux relations entre la surface de la protéine et les lipides.

3.2 Discussion

La rotation des protéines membranaires est un processus lent qui se passe sur l'échelle de la microseconde (MACDONALD et al., 1988; SAFFMAN et al., 1975; THOMAS, 1986). Spontanément, seule une petite partie des orientations relatives entre les protéines est moyennée lors d'une simulation si ces orientations ne sont pas contrôlées. Dans ce cas, l'existence de multiples bassins de basse énergie séparés par des barrières d'énergie potentielle limite considérablement l'exploration de ces orientations.

Dans le contexte d'une intégration thermodynamique où seule la distance est contrôlée, l'échantillonnage de ces orientations manque donc considérablement. Théoriquement, l'échantillonnage de l'ensemble de l'espace n'est pas nécessaire pour obtenir un PMF valide mais seulement les bassins de plus basse énergie. Ainsi, dans le cas théorique où il n'y a qu'une seule vallée énergétique sur le PMF, l'intégration donnera un résultat valide.

Si cependant il existe plusieurs bassins de basse énergie à une même distance, le PMF calculé ne sera pas bon. La dépendance à l'orientation du PMF a déjà été mentionnée dans plusieurs publications, où les auteurs imposent explicitement des contraintes sur les orientations de leurs protéines (PÉRIOLE et al., 2012; DUNTON et al., 2014). Dans ces publications, le PMF dépend de la surface de la protéine. Des différences sont aussi observées à longue distance, montrant encore que des interactions différentielles existent quand des lipides sont présents entre les surfaces.

Obtenir la convergence dans les calculs d'énergie libre de protéines membranaires reste une tâche difficile qui demande l'inclusion de plus de degrés de libertés que la simple distance, même dans des systèmes modèles simples (LI et al., 2014; ЧИПОТ, 2008). Par exemple, pour le dimère de GpA, la distance

s'avère être une coordonnée insuffisante à courte distance où il existerait 4 différents sous-états sur la position de plus basse énergie (DOMAŃSKI et al., 2017). Avec l'AqpZ et ces simulations, nous ajoutons des évidences que les orientations relatives sont aussi la source d'une hétérogénéité à moyenne et longue distance.

Méthodologiquement, ce travail apporte donc une caution : il faut surveiller ces orientations pour obtenir une estimation valable de l'énergie libre. Ces orientations peuvent être suivies de différentes manières, non limitées aux angles θ et ϕ (CASUSO et al., 2012) choisis dans cette étude, mais il faut au moins qu'une des deux variables prennent en compte les deux protéines (voir méthodes).

Dans notre cas, une protéine a dû être fixée pour garantir l'orthogonalité des deux variables. Cette situation garde cependant une pertinence biologique lorsque les protéines diffusent et s'ancrent sur un réseau de protéines agrégées, comme observée avec OmpF par AFM à haute vitesse. Notre PMF prend donc en compte les conséquences de cette immobilisation, qui incluent aussi l'adaptation des lipides le long des surfaces protéiques et une possible potentialisation des perturbations à longue distance induites à travers les lipides. Ces perturbations, observées précédemment dans notre étude sur 12 protéines membranaires (J.-P. DUNEAU et al., 2017), sont influencées directement par la première couche de lipides au contact de la protéine. A leur tour, ces perturbations vont ensuite aboutir aux orientations préférentielles observées dans cette étude.

Une barrière d'énergie est de plus présente dans le PMF 1D. Cette barrière n'a jamais été décrite auparavant mais correspond à des phénomènes déjà observés au laboratoire (J.-P. DUNEAU et al., 2017; CASUSO et al., 2012). Ce comportement devra être étudié plus en détail dans des travaux futurs mais s'ancrer dans l'idée que les protéines doivent maintenir un potentiel répulsif pour ne pas s'agréger. La source de ce potentiel répulsif pourrait alors venir de la combinaison d'un faible mismatch hydrophobe et de la composition en lipides, qui inclut des PE, qui favorise l'existence de phases hexagonales, mais cela reste spéculatif.

Un certain nombre de limitations est cependant présent. La structuration de notre PMF à la distance la plus longue montre que le système n'est pas assez grand pour obtenir un état où les protéines n'interagissent pas et est observable sur les PMF 3D et 1D, avec une absence de plateau à la plus grande distance. Ainsi, aucun zéro n'a pu être placé. La description des chemins à travers ce PMF n'en est cependant pas affectée.

Les niveaux d'énergie observés dans nos PMF multidimensionnels indiquent qu'ils ne sont globalement pas convergés. Pour combler ce déficit, des temps de simulations supérieurs aux 20 μ s utilisés auraient été nécessaire mais ont été techniquement limités par les instabilités liées à l'implémentation de MARTINI

dans NAMD.

En complément de ce travail, j'ai effectué une étude plus approfondie sur les problèmes techniques liés à l'utilisation du champ de force MARTINI dans NAMD au travers d'un crash ainsi qu'un travail préliminaire d'intégration thermodynamique en utilisant l'ABF et l'Umbrella Sampling. Ces compléments sont présentés dans les section 7 et 9.

Malgré ces limitations, nous pouvons avoir confiance dans les conclusions de cette étude sur l'existence de forces à longues distance entre les protéines membranaires qui vont affecter leurs orientations. Ces forces, qui partent de la surface de la protéine, se propagent à travers les lipides et, comme elles diffèrent suivant les orientations relatives des protéines, dépendent donc des propriétés des surfaces en contact avec les lipides. Cette dépendance est alors un facteur à prendre en compte lorsque nous voulons calculer la variation d'énergie libre de la dissociation de deux protéines. Cela ajoute surtout des questions quant aux PMF présents dans la littérature : est-ce que les orientations sont moyennées ? Et est-ce que d'autres degrés de liberté, non examinés, sont effectivement moyennés ?

Ces forces, qui se propagent jusqu'à longue distance à travers les lipides, dépendent donc de la surface de la protéine. La question qui se pose alors concerne les lipides : comment est-ce qu'ils s'organisent autour de la protéine ? Est-ce que ces forces proviennent des propriétés physico-chimiques des atomes de surface ou de la topographie de cette dernière qui structurerait les lipides ? Dans la section suivante, nous avons cherché à répondre à ces questions en étudiant le comportement de l'Aquaporine Z au sein d'une membrane complexe.

Section 4

Relations entre la protéine et les lipides qui l'entourent - annulus et tri des lipides

L'annulus est un anneau lipidique de solvation dynamique autour des protéines membranaires (LEE, 2011). Bien que des lipides très ancrés à la surface des protéines soient parfois vus en cristallographie aux rayons X ou en cryo microscopie électronique (GONEN et al., 2005; GAO et al., 2016; FRAUENFELD et al., 2011), l'observation des lipides un peu moins ancrés à la surface où leur différenciation avec les molécules de détergent est souvent difficile.

La complexité des lipides présents dans les membranes biologiques est plus importante que dans notre modèle précédent. En effet, bien qu'une membrane de POPE/POPG soit un modèle simple de la membrane interne d'*E. coli*, un lipide particulier manque : la cardiolipine (CL). Les CL sont indispensables pour la fonction de certaines protéines de la membrane interne d'*E. coli* ou de la mitochondrie (SHEN et al., 2017; CLAYPOOL, 2009; ACEHAN et al., 2011). De manière imagée, ce lipide correspond à "deux PG fusionnés par leur tête".

Cette diversité lipidique est particulièrement mise en évidence dans l'étude de l'annulus (J.-P. DUNEAU et al., 2017). Au laboratoire, ces lipides peuvent être analysés expérimentalement en utilisant des nanodisques. Pour former ces derniers, une méthode utilisée est l'utilisation du polymère de Styrene Maleic Acid (SMA) (SCHMIDT et al., 2018) qui permet de dissoudre les membranes sans utiliser de détergent et de former des inclusions membranaires qui peuvent contenir des protéines membranaires avec un anneau de lipide, qui rendrait alors compte de cet annulus. Dans des études précédentes sur le translocon Sec (PRABUDIANSYAH et al., 2015) et le canal à potassium KcsA (DÖRR et al., 2014), les nanodisques formés par le SMA contenant la protéine avaient un enrichissement en lipides chargés négativement, en particulier en CL, qui suggère une potentielle importance fonctionnelle. Dans le cas de l'AqpZ et de son mutant monomérique, l'exploration de la composition en

lipides de son annulus peut amener à des considérations sur sa fonction mais aussi sur l'origine et la propagation des forces responsables de l'assemblage et de la ségrégation dans les membranes.

Au laboratoire, V.Schmidt a effectué les expériences sur les nanodisques de SMA contenant l'AqpZ, sauvage ou son mutant monomérique. Grâce à la spectrométrie de masse native, il est en effet possible d'analyser la composition en lipides des nanodisques et de caractériser les interactions lipide-protéines spécifiques. La spectrométrie de masse (MS) native a en effet déjà été utilisée pour identifier les lipides endogènes co-purifiés avec des complexes membranaires ou suivre la cinétique et la thermodynamique de l'ajout de lipides exogènes (LAGANOWSKY et al., 2014). Cependant, il reste difficile d'avoir une vision structurale précise des sites de fixation.

Pour pouvoir ensuite intégrer ses résultats à une échelle moléculaire et questionner les modèles actuels, j'ai donc cherché à construire un modèle de bicouche de membrane interne d'*E.coli* le plus complexe possible dans la représentation gros grain MARTINI et prenant en compte la diversité lipidique connue de cette membrane, dans le but d'étudier *in silico* cet annulus.

Pour cela, le champ de force gros grain MARTINI est idéal et permet la simulation d'une grande variété de lipides.

Dans le but de promouvoir la création et simulation de membranes, un tutoriel est présent à l'adresse https://framagit.org/msidore/tutorial_insane.

En utilisant une double approche, mêlant expériences et simulations, nous avons pu étudier les préférences lipidiques de l'AqpZ (monomère, tétramère ainsi qu'un mutant monomérique) et le rôle particulier des cardiolipines, lipides important mais pas toujours pris en compte dans les modèles membranaires.

4.1 Résultats

4.1.1 Une membrane plus "réaliste"

La membrane interne d'*E.coli* possède 3 types de lipides majoritaires : les PE, les PG et les CL. La membrane simple souvent utilisée comme modèle, une membrane de POPE/POPG dans un ratio de 3 :1 ou 4 :1, ne prend pas en compte les CL bien qu'étant un constituant aussi important en masse que les PG. De plus, les CL interagissent souvent avec les protéines membranaires et sont parfois indispensables à leur bonne fonction, y compris pour l'AqpZ (SHEN et al., 2017; CLAYPOOL, 2009; ACEHAN et al., 2011).

A partir de la composition de la membrane interne d'*E.coli*, j'ai donc produit un modèle plus "réaliste" contenant 9 lipides gros grains, dont 4 PE, 3 PG et 2 CL, pour rendre compte de la diversité lipidique de cette membrane (Figure 4.1).

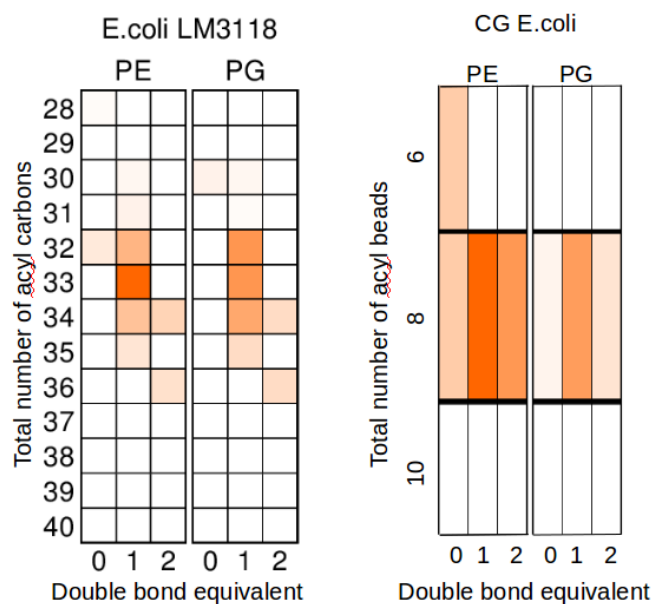


FIGURE 4.1 – Haut : Proportions des chaînes latérales des phospholipides de la membrane interne d'*E. coli* (KITO et al., 1972) en équivalent double liaison et comparaison au modèle gros grain. L'intensité de la case est proportionnelle à la quantité de chaque phospholipide. Bas : Snapshot de la membrane complexe équilibrée. Les PE sont en rouge, les PG sont en bleu et les CL sont en vert.

La largeur de cette membrane est de 32 Å et la surface par lipide est de 68.38 Å², en accord avec les paramètres classiques d'une membrane dans la littérature.

J'ai ensuite construit plusieurs systèmes (voir méthodes) en utilisant soit cette membrane complexe, soit la membrane simple de POPE/POPG dans un ratio 4 :1, avec soit le monomère d'AqpZ, soit le tétramère d'AqpZ. Un mutant, créé par V.Schmidt, a aussi été simulé avec ces différentes membranes ; ce mutant a la particularité d'être monomérique sur gel. Chaque simulation a duré 20 µs.

Ces simulations ont été effectuées sous GROMACS 5.1.4 pour s'affranchir des problèmes de stabilité discutés précédemment et dans la section 9.

De plus, dans le but de faciliter la création et la simulation de systèmes membranaires simples, j'ai développé un tutoriel utilisant les outils de GROMACS et MARTINI permettant de construire une membrane simple, avec ou sans protéines. Ce tutoriel est présent à l'adresse https://framagit.org/msidore/tutorial_insane/.

4.1.2 Tri des lipides autour de la protéine

Pour questionner l'effet des forces induites par la protéine au travers des lipides sur un possible tri de ces derniers, nous pouvons passer par des comptages des lipides dans différents rayons autour de la protéine. La figure 4.2 résume les enrichissements lipidiques (ratio par rapport à la composition totale de la membrane) dans un rayon de 6 Å autour de la protéine selon le type de lipide lors des simulations avec la membrane complexe.

A cette distance, plusieurs effets sont présents : en fonction de la tête polaire du lipide et en fonction du nombre d'insaturation. A insaturations égales, les PG ont globalement un enrichissement supérieur aux PE. L'enrichissement des CL à 2 insaturations, en comparaison des PG et PE à 2 insaturations, semble moins élevé. L'effet le plus important reste dépendant des insaturations, avec systématiquement un enrichissement plus élevé au plus le nombre d'insaturation est grand, jusqu'au maximum des CL à 4 insaturations dans notre système.

Les préférences pour les types de lipides sont cependant dépendantes du feuillet que l'on observe, et donc de la physico-chimie des acides aminés à la surface de la protéine. Toujours dans ce même rayon de 6 Å, on observe des enrichissements en type de lipide qui correspondent à la "positive inside rule", avec un enrichissement en lipides chargés négativement sur le feuillet cytoplasmique (PG et CL) et un enrichissement en lipides neutres (PE) sur le feuillet périplasmique (Figure 4.3).

La "positive inside rule" est une règle empirique concernant la répartition des charges à la surface des protéines membranaires (HEIJNE, 1992). Classiquement, la face cytoplasmique est enrichie en acides aminés chargés positivement.

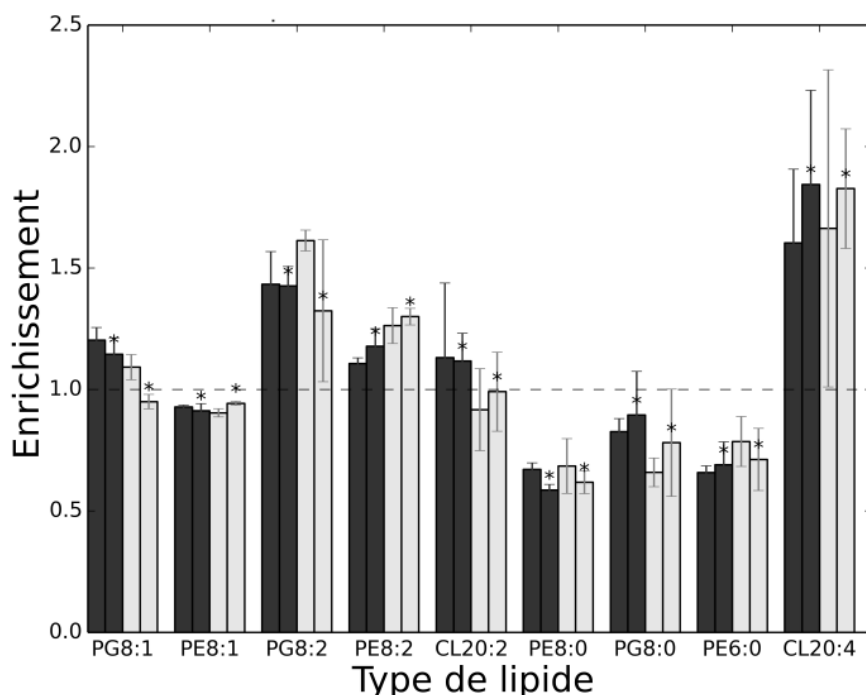


FIGURE 4.2 – Enrichissement des différents lipides dans un rayon de 6 Å autour de la protéine. Les résultats de 4 simulations sont affichés pour chaque lipide : le tétramère est en gris foncé, le monomère est en gris clair et le mutant W₁₄A est indiqué par une *. Les types de lipides utilisent la notation décrite dans les méthodes avec le nombre de billes de la chaîne acyle suivi du nombre d'insaturation ; ainsi, un POPE est noté PE8 :1. Les écarts types sont indiqués par des barres d'erreur.

En s'éloignant de la protéine, les enrichissements diminuent simplement jusqu'au bulk (jusqu'à un rayon de 18 Å autour de la protéine) en gardant les mêmes tendances. Ainsi, une partie des forces qui se répandent à partir de la surface de la protéine peuvent prendre leur source dans la physico-chimie des acides aminés de surface qui vont privilégier l'interaction avec certains types de lipides.

Cependant, un simple comptage et une représentation des enrichissements dans un rayon de la protéine peut cacher d'autres effets plus structurels : en effet, les insaturations sont enrichies mais sont des groupements de lipides qui s'adaptent plus facilement à la topographie de surface des protéines. J'ai donc questionné cette possible structuration en étudiant les lipides à la surface immédiate de la protéine.

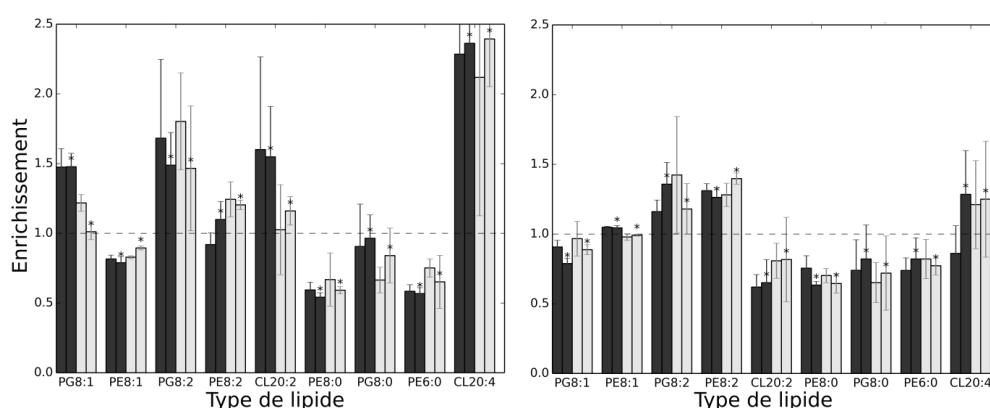


FIGURE 4.3 – Enrichissement des différents lipides dans un rayon de 6 Å autour de la protéine sur le feuillet cytoplasmique (gauche) et périplasmique (droite). Les résultats de 4 simulations sont affichés pour chaque lipide : le tétramère est en gris foncé, le monomère est en gris clair et le mutant W₁₄A est indiqué par une ★. Les types de lipides utilisent la notation décrite dans les méthodes avec le nombre de billes de la chaîne acyle suivi du nombre d'insaturation ; ainsi, un POPE est noté PE8 :1. Les écarts types sont indiqués par des barres d'erreur. Les enrichissements en types lipidiques suivent la "positive inside rule", avec les lipides chargés négativement sur le feuillet cytoplasmique et les lipides neutres sur le feuillet périplasmique.

4.1.3 Préférences lipidiques à la surface de la protéine

Afin d'interroger les structurations de lipides autour de la protéine, on peut moyenner leur présence au cours de la simulation. Une fois ces densités de lipide calculées, on peut les coucher sur la structure de la protéine pour observer les zones de plus haute densité, qui correspondent aux zones où les lipides vont avoir une plus grande affinité. La figure 4.4 montre les densités des 3 types de lipides autour du tétramère et du monomère d'AqpZ.

Tout d'abord, une observation rassurante sur le tétramère : les densités sont symétriques tout autour du tétramère, ce qui indique que le temps de simulation était suffisant.

Ensuite, les densités des lipides, qui correspondent aux positions préférentielles des lipides, se placent à des emplacements précis autour de la protéine. Sur le tétramère d'AqpZ, les PE se positionnent donc en majorité sur la face cytoplasmique et les PG et CL sur la face périplasmique, en accord avec la "positive inside rule". Les CL vont cependant se placer sur la zone liant les monomères d'AqpZ alors que les PG vont occuper le reste des monomères. Sur le monomère seul, on retrouve la répartition des PG sur le côté du monomère originellement en contact avec les lipides (le côté où les PG se placent sur le tétramère) alors que les CL, perdant leur zone d'ancrage, vont se placer sur les deux faces qui originellement sont à l'intérieur du tétramère. Les lipides épousent de plus la topographie de la surface de la protéine.

Nous observons aussi, pour les PG et PE, les mêmes tendances sur le tétra-

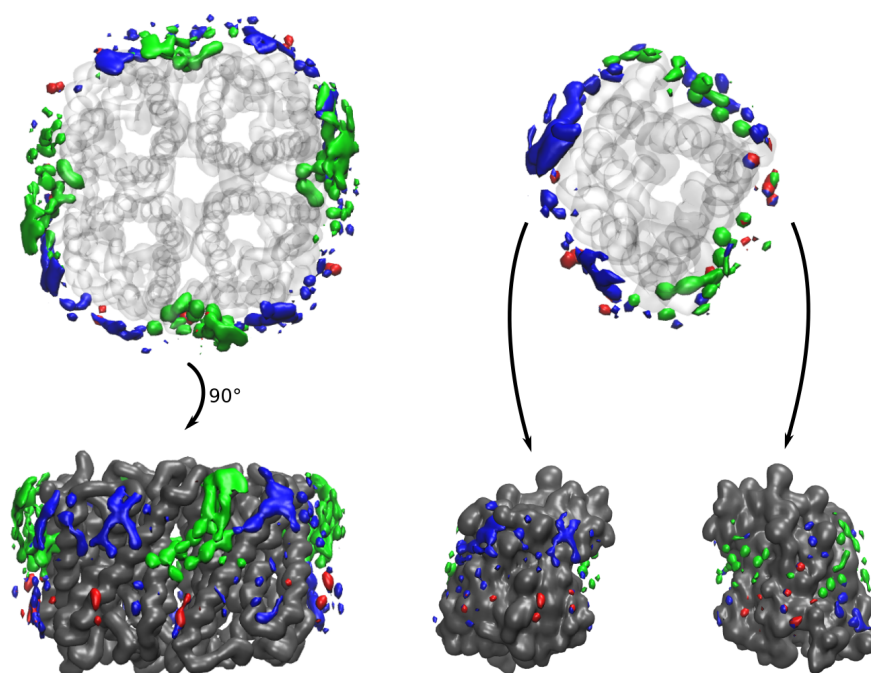


FIGURE 4.4 – Densités lipidiques normalisées par le nombre de particules à la surface du tétramère (gauche) ou du monomère (droite) d'AqpZ. La protéine est en représentation cartoon transparente en haut et en surface opaque en bas, les densités des PE sont en rouge, celles PG sont en bleu et celles CL sont en vert.

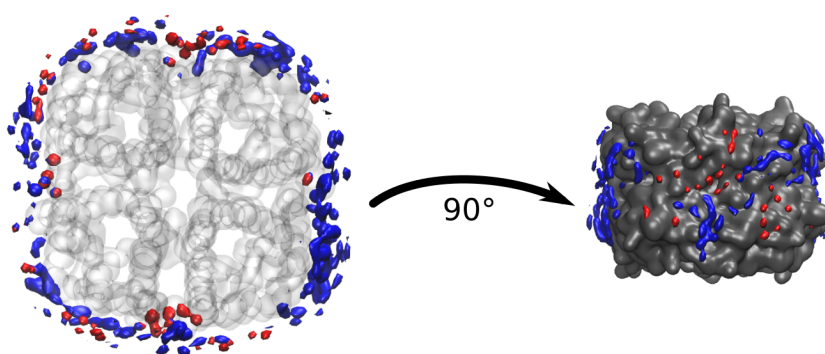


FIGURE 4.5 – Densités lipidiques normalisées par le nombre de particules à la surface du tétramère sauvage d'AqpZ dans une membrane simple. La protéine est en représentation cartoon transparente en haut et en surface opaque en bas, les densités des PE sont en rouge et celles PG sont en bleu.

mère dans une membrane plus simple de POPE/POPG (figure 4.5).

Le résultat des simulations du mutant monomérique, $W_{14}A$, est montré dans la figure 4.6.

Les densités observées sur le tétramère mutant ne sont pas symétriques : la densité de PE observée s'explique par le temps d'échange des lipides coincés dans ces interfaces : quand le temps de résidence des lipides à la surface de la protéine sauvage est de 1 à 3 ns, les lipides coincés entre les monomères y restent en moyenne plusieurs centaines de nanosecondes et il faudrait simuler beaucoup plus longtemps pour moyennner ces événements.

Malgré ces défauts, nous pouvons être confiants sur ce résultat : dans la protéine sauvage, ce sont les CL qui ont une meilleure affinité pour la zone interfaciale entre les monomères et on observe les CL qui pénètrent le tétramère mutant. La densité de PE est donc très probablement due à la proximité d'une PE lors de la simulation et n'importe quel lipide, peu importe son type, pourrait alors entrer dans cette interface.

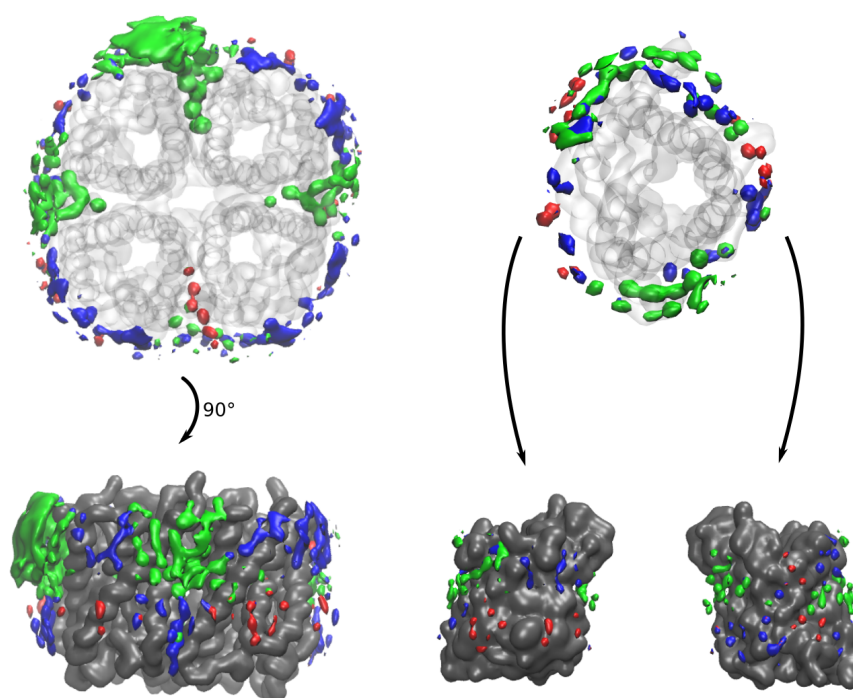


FIGURE 4.6 – Densités lipidiques normalisées par le nombre de particules à la surface du tétramère (gauche) ou du monomère (droite) mutant d'AqpZ. La protéine est en représentation cartoon transparente en haut et en surface opaque en bas, les densités des PE sont en rouge, celles PG sont en bleu et celles CL sont en vert.

Les différents types de lipide se répartissent sur le tétramère mutant de la même manière que sur le tétramère natif : les PE sur la face cytoplasmique et

les PG et CL sur la face périplasmique. Les PG sont cependant plus présent sur la zone interfaciale entre les monomères, du côté cytoplasmique.

La mutation provoquant un "trou" dans la structure, les CL, qui ont une plus grande affinité pour la zone interfaciale entre les monomères, peuvent alors s'immiscer entre les monomères (Figure 4.6, gauche). Ce mutant étant expérimentalement monomérique, cette observation permet de proposer une explication : les lipides, et incidemment les cardiolipines, perturbent le tétramère en entrant dans le trou provoqué par la mutation.

Le monomère mutant présente quant à lui des densités différentes du sauvage. La différence principale se situe au niveau des CL : alors que les CL se répartissaient uniquement le long des faces originellement à l'intérieur du tétramère dans le monomère sauvage, elles se répartissent maintenant sur les arêtes de la protéine, qui correspondent à la zone interfaciale entre les monomères. La mutation semble donc avoir favorisé la présence de CL sur cette zone.

Au delà des têtes polaires et si l'on regarde les billes correspondant aux insaturations (jaune, figure 4.7) par rapport à celles correspondant aux saturations (orange), elles vont aussi se retrouver préférentiellement à la surface de la protéine. Le tri à la surface de la protéine s'effectue non seulement par les charges, mais aussi par les insaturations : une chaîne insaturée va en effet avoir plus de facilité à s'adapter à la surface de la protéine.

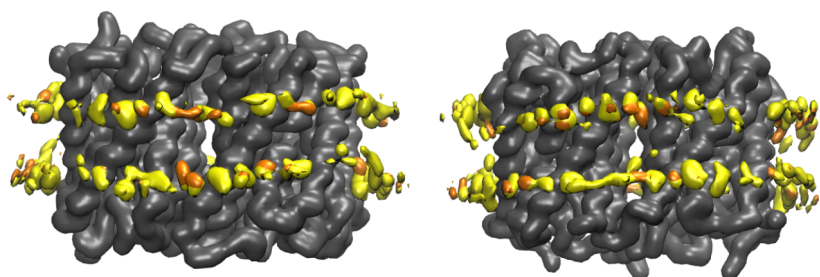


FIGURE 4.7 – Densités lipidiques des billes correspondant aux insaturations ou saturations, normalisées par le nombre de particules à la surface du tétramère sauvage (gauche) ou muté (droite) d'AqpZ. La protéine est en représentation surface opaque, les densités des insaturations en jaune et des saturations en orange.

A ce niveau, nous avons donc certaines préférences lipidiques autour de la protéine. Tout d'abord à sa surface, les lipides vont être triés en fonction des propriétés physico-chimiques des acides aminés de surface et épouser sa topographie. Dans cette adaptation, les doubles liaisons aident et sont donc

aussi favorisées. En s'éloignant de la protéine, ces tendances persistent et se propagent jusqu'à une distance de 20 Å de la surface de la protéine.

4.2 Discussion

Nous avons simulé ici un modèle gros grain complexe de la membrane interne d'*E.coli* avec l'AqpZ, notre protéine modèle, sauvage et mutante. La particularité de notre système est la prise en compte des cardiolipines, un lipide peu représenté dans les simulations de dynamique moléculaire malgré sa présence, en masse, dans une proportion similaire aux PG, ainsi que la variabilité des insaturations.

Nous avons tout d'abord observé des enrichissements lipidiques à la surface de la protéine. En particulier, les cardiolipines sont enrichies à proximité et à quatre positions autour de la protéine. Toujours à la surface et comme attendu par la "positive-inside rule", les PG et CL sont enrichis du côté cytoplasmique et les PE sont enrichis du côté périplasmique. Un résultat important est que ces enrichissements s'étendent ensuite au delà de la zone d'ancrage en diminuant jusqu'à arriver au bulk.

Le mutant $W_{14}A$ montre les mêmes tendances avec l'infiltration des cardiolipines à l'emplacement du trou laissé par la mutation. Cette protrusion de lipides à l'interface entre les monomères qui contribue à une déstabilisation du tetramère est cohérente avec la forme monomérique observée expérimentalement sur gel.

Ces résultats exemplifient alors la relation qui pourrait exister entre l'existence de forces à longue distance précédemment décrites et les problématiques de tri des lipides au voisinage d'une surface protéique. Dépendante de la surface de la protéine, deux phénomènes s'additionnent :

- Le tri du type de lipide à la surface de la protéine. Ce tri va s'opérer principalement à partir des charges de la tête polaire du lipide et des charges présente sur le bord de la protéine (Figure 4.4). Ici, les PG et CL, lipides chargés négativement, vont se rassembler sur la face cytoplasmique de la protéine, face globalement chargée positivement, et les PE, lipides neutres, vont plutôt se rassembler sur la face périplasmique, elle chargée négativement et repoussant les lipides chargés de la même manière. Ces enrichissements vont ensuite se répercuter jusqu'à une distance de 12 à 18 Å depuis la surface de la protéine.
- L'adaptation des lipides à la surface de la protéine, dépendante du type du lipide, de la topographie et des propriétés physico-chimiques de cette surface. Ici, les insaturations vont en particulier faciliter l'adaptation du lipide sur la surface de la protéine (Figure 4.7) et permettre à ce lipide de rester plus longtemps. C'est ce qui est observé ici, avec un enrichissement des lipides insaturés par rapport au même lipide possédant moins d'insaturation.

En conclusion, les règles de la positive-inside rule sont importantes pour accorder la physico-chimie de la surface de la protéine à la nature des lipides ancrés, mais notre modèle de membrane complexe montre que d'autres règles, notamment liées à l'ancrage différentiel des insaturations, sont également à l'oeuvre. La longueur de chaînes pourrait aussi être une autre source de modulation, mais notre modèle de membrane complexe ne permet pas de questionner cette hypothèse.

Nos études sur ce modèle montre que la montée en complexité de la composition de la membrane ne s'oppose pas à la formation de perturbations longue distance de l'organisation des lipides.

Parallèlement, des nanodisques de SMA contenant l'AqpZ (SMA-AqpZ) sauvage ou mutée ont été analysés par lipidomique, chromatographie sur couche mince et spectrographie de masse (MS). Ces nanodisques sont composés d'un polymère anionique entourant un mélange de protéine et lipides ; l'analyse a montré que chaque nanodisque contenait une seule protéine, du polymère ainsi qu'un annulus de lipides, possiblement perturbés par le polymère. La MS native a ensuite permis d'observer des lipides individuels fixés sur la surface de la protéine, qui résistent à la dissociation par compétition avec les molécules de détergent ; en particulier, elle mesure les préférences des lipides par rapport aux détergents qu'elle déplace.

La composition lipidique des nanodisques SMA-AqpZ est cependant différente des enrichissements vus en simulation, avec en particulier une déplétion en cardiolipines. Pour rapprocher les simulations de l'expérience, il faut alors souligner les différences entre ces deux situations. Les simulations mettent en jeu une protéine unique dans une mer de lipides, où les CL sont en excès par rapport aux sites de fixation de la protéine. Dans le contexte d'une membrane biologique où l'AqpZ est surexprimée, la quantité de CL disponible par nanodisque est alors bien plus limitée et la déplétion observée pourrait refléter cette disponibilité et limiter le nombre de CL par nanodisque. Dans ces conditions, la MS-native nous informe sur le nombre de sites de haute affinité pour les CL. Il est ressorti de ces analyses que l'AqpZ possède un site de haute affinité par monomère pour la CL, dont l'affinité est réduite dans le cas du mutant.

Ces résultats sont compatibles avec l'expérience : l'AqpZ tétramérique possède quatre sites de haute affinité pour la CL que nous situons, grâce aux simulations, sur la zone entre les monomères. Dans les conditions de la MS native, l'affinité pour la CL diminue pour le mutant, monomérique dans ces conditions, car des CL s'engouffrent entre les monomères et déstabilisent le tétramère. Les enrichissements observés dans nos simulations ne reflètent alors pas le cas d'une membrane biologique où les enrichissements vont aussi dépendre des autres protéines présentes.

Néanmoins, ces enrichissements, dans un système modèle comme celui-ci, rendent compte des forces caractéristiques de cette protéine qui vont se propager à distance, même si nous sommes loin des influences observées à plus

grande distance dans la partie précédente qui pouvaient aller jusqu'à 50 Å de la surface de la protéine. Dans une membrane biologique, c'est alors la somme des influences individuelles de toutes les protéines au travers des lipides de la membrane, sur les autres protéines et sur les lipides, qui va alors être importante. Pour étudier le devenir de ces influences, nous avons alors construit un système modèle plus grand, avec un ratio AqpZ/lipides de 1 :1 en masse, et ainsi avoir un modèle d'une membrane encombrée, plus proche d'une membrane biologique.

Pour pousser la comparaison entre l'expérience et la simulation, nous avons de plus développé un modèle de polymère gros grain mimant les propriétés du SMA. Ce modèle et son impact sur le tri des lipides est présenté dans la section 8

Section 5

Devenir de ces forces dans une membrane encombrée

Nous avons mis en évidence dans la section 3 l'existence de forces à longue distance entre les protéines membranaires. Ces forces se propagent à travers les lipides et, comme vu dans la section 4, dépendent des propriétés physico-chimiques et de la topographie de la surface de la protéine.

La première tendance pour étudier ce type de système est de réduire la complexité au minimum suivant un réductionnisme classique (KAUFFMAN, 1971). Dans ces modèles, la protéine est étudiée isolée dans une mer de lipides plus ou moins complexes.

Cependant, en moyenne, les membranes biologiques sont composées en masse de 50% de protéines : en plus de la diversité des lipides, c'est un environnement plus complexe et surtout plus encombré. Dans ces conditions, les conséquences de la combinaison des forces à longue distance sont plus difficiles à prédire. Dans ce contexte, l'évolution des enrichissements lipidiques et des sites à haute affinités, notamment pour les cardiolipines (section 4), doit être réinterrogé alors qu'il n'y aura pas nécessairement suffisamment de lipides pour satisfaire les préférences de toutes les protéines de la membrane.

Cette montée en complexité représente la deuxième tendance dans les simulations de dynamique moléculaires membranaires (CHAVENT, DUNCAN et al., 2016) et est favorisée par l'utilisation des champs de forces gros grain. Par exemple, un modèle de membrane plasmique contenant 144 copies d'un récepteur couplé aux protéines G a révélé des contacts entre les protéines et une diffusion des lipides modifiée (KOLDSO et al., 2015). L'impact du cytosquelette dans la diffusion des protéines et des lipides a aussi été exploré (CORRADI et al., 2018).

Cependant, ces modèles n'ont jamais été utilisés pour savoir ce qu'il adviendrait de l'annulus par rapport à un système plus simple. Par ailleurs, on peut aussi se demander si le champ de forces gros grain est pertinent pour

questionner les conséquences de l'encombrement sur l'assemblage des protéines.

5.1 Résultats

Afin de mesurer ces impacts, deux systèmes comprenant 125 monomères d'AqpZ distribués sur une surface membranaire de 2500 nm^2 ont été construits, soit avec la membrane complexe décrite précédemment, soit avec une membrane plus simple de POPE/POPG, simulées sur un temps de $20 \mu\text{s}$. L'encombrement de cette membrane est d'environ 50% en masse en protéine.

Le système initial avec la membrane complexe est montré sur la figure 5.1.

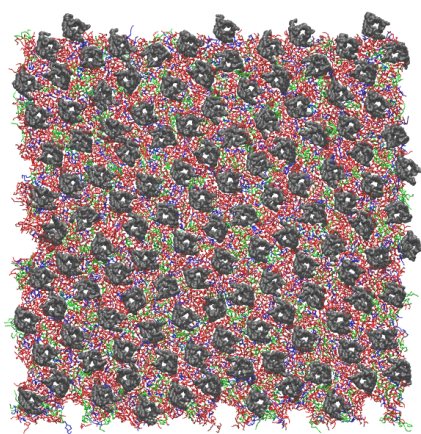


FIGURE 5.1 – Système initial des simulations de la membrane encombrée. Les lipides sont en représentation bâtons, les CL sont en vert, les PG sont en bleu et les PE sont en rouge. Les 125 monomères d'AqpZ sont en représentation surface du squelette pour mieux les distinguer.

5.1.1 Structuration d'un réseau : des chaînettes de protéines

Après $2 \mu\text{s}$ de simulation, dans la membrane encombrée ou simple, les protéines vont avoir tendance à se toucher pour former rapidement un réseau de chaînettes (figure 5.2). Ce réseau reste ensuite presque figé sur l'échelle de la dizaine de microseconde. La mise en place de ce réseau dans la membrane complexe est visualisable en vidéo à l'adresse <https://drive.google.com/open?id=1YGNRSwfkuJ53oAuyGa4gNHsLlz7FqFFw>.

Un tel réseau, en chaînettes, n'était pas attendu. On peut noter qu'aucun tétramère natif d'AqpZ ne s'est reformé, même si cela n'est pas forcément attendu dans ces échelles de temps.

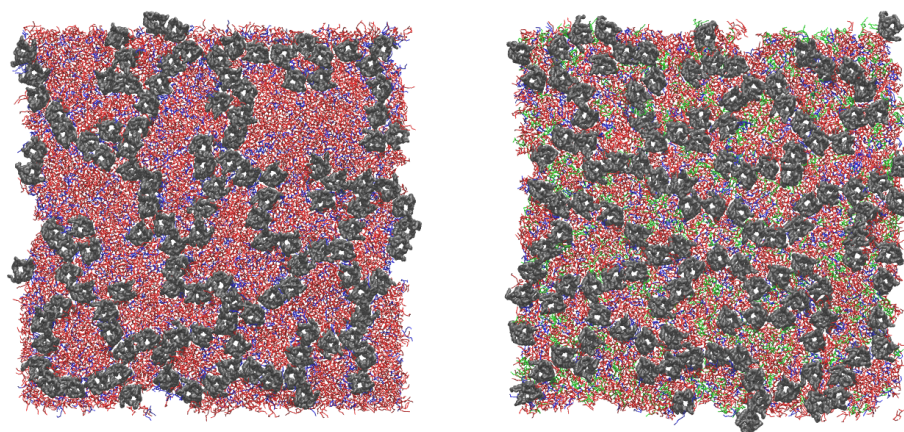


FIGURE 5.2 – Membrane encombrée simple (gauche) ou complexe (droite) après 20 μ s de simulation. Les lipides sont en représentation bâtons, les CL sont en vert, les PG sont en bleu et les PE sont en rouge. Les 125 monomères d'AqpZ sont en représentation surface du squelette.

Avec 4 faces potentielles d'interaction par monomère, il y a 16 possibilités de contact dont seule une définit l'interface native. On observe dans la table 5.1 que la majorité des dimères interagissent par des interfaces non-natives dans les deux membranes. Le nombre de contacts est en revanche plus élevé dans la membrane complexe.

TABLE 5.1 – Nombre de contacts natifs et non-natifs dans la membrane simple et complexe.

Type de membrane	Contacts natifs	Contacts non-natifs
Simple	5	161
Complexe	10	193

Dans la continuité des sections 3 et 4, les analyses qui suivent mettent à jour les orientations relatives des dimères et les lipides à proximité des monomères lors des dernières 500 ns de simulation, où le réseau de protéine est stable.

5.1.2 Devenir des orientations relatives entre dimères

Dans un premier temps, nous avons analysé les orientations relatives des dimères les plus proches (distance de 35 à 40 Å entre monomères) de ce système. Un histogramme de ces orientations en fonction de la distance est présenté dans le panel le plus à gauche de la figure 5.3. On observe que seul un ensemble restreint d'orientations relative est présent. De plus, ces orientations sont différentes dans la membrane simple et complexe. On observe ensuite une

plus grande diversité d'orientations relatives dans la membrane complexe.

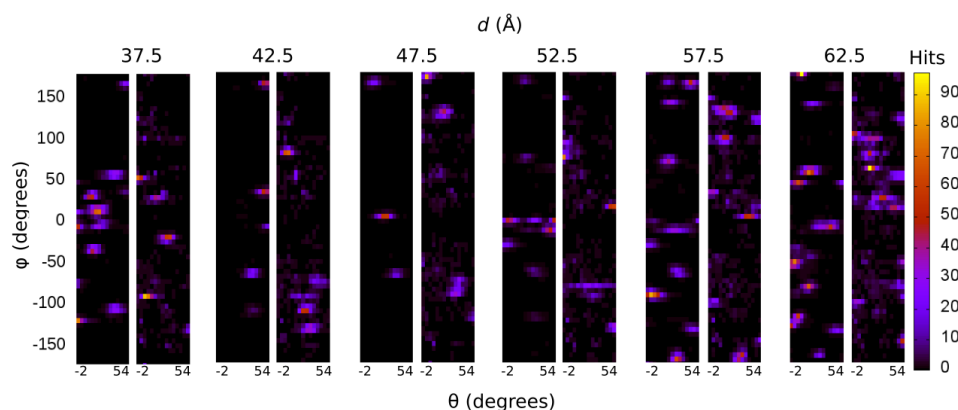


FIGURE 5.3 – Échantillonnage des orientations relatives entre dimères de la membrane encombrée simple (gauche des panneaux) et complexe (droite) des derniers 500 ns de simulation. Seules les distances entre 35 et 65 Å sont visualisées.

Au fur et à mesure que l'on considère des distances croissantes entre les monomères, ces différences persistent jusqu'à 65Å, distance qui continue de n'impliquer majoritairement que des monomères deux à deux même si certains effets coopératifs ne sont pas exclus.

Pour étudier ce réseau plus en détails et observer le rôle des lipides et le devenir des préférences lipidiques à la surface des protéines dans ce contexte, nous avons calculé les cartes de densités des lipides autour de chaque monomère.

5.1.3 Préférences lipidiques à la surface de l'Aquaporine Z

Nous avons calculé les cartes de densités des lipides et des protéines autour de chaque monomère puis les avons moyenné. Ainsi, seules les contributions des lipides les plus ancrés et des protéines les plus présentes seront décrites. La figure 5.4 présente ces cartes de densités.

Concernant la membrane simple, on observe une distribution homogène de la PG autour des protéines. En revanche, dans le cas de la membrane complexe, on observe que la CL est distribuée majoritairement sur seulement deux des quatre faces dont une est une interface protéine-protéine dans la structure native et l'autre une interface protéine-lipide. Concernant les probabilités de présence du deuxième monomère, on observe des distribution très semblables à l'exception d'une densité additionnelle dans le cas de la membrane simple sur une interface native. Un artefact de la procédure, non résolu à ce jour, empêche d'observer le deuxième monomère sur l'interface non-native de la figure 5.4, gauche bien que l'absence de densité lipidique et

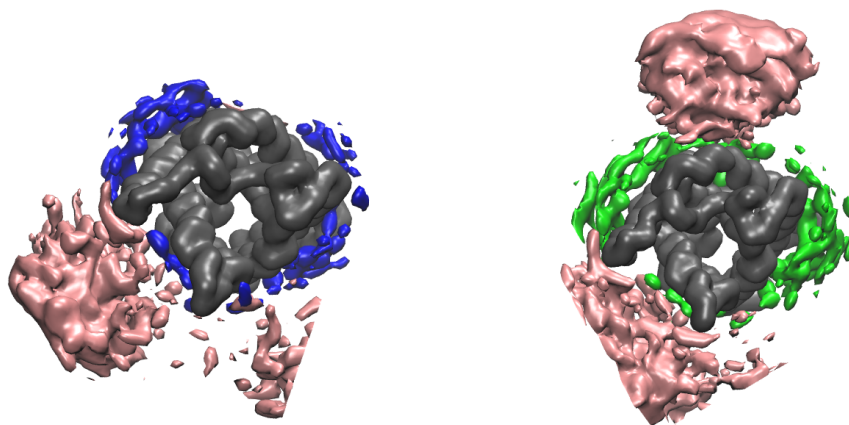


FIGURE 5.4 – Cartes de densité moyennes des cardiolipines et protéines moyennées de tous les monomères et visualisées autour d'un monomère de référence de la membrane simple (gauche) et complexe (droite). Le monomère de référence est en représentation de surface du squelette en gris, la densité des protéines est en rose, la densité des PG est en bleu et la densité des CL est en vert. Une densité rose est manquante en haut à gauche dû à un artefact non résolu.

une petite densité protéique soit présente. On peut cependant remarquer que l'interaction privilégiée n'est pas forcément définie par une face mais plutôt par une arête.

L'organisation des protéines peut ensuite être visualisée par une fonction de distribution radiale (figure 5.5). On observe dans les plus courtes distances la présence de trois pics distincts qui indiquent de fortes corrélations dans le mode d'interaction entre les protéines voisines. Dans le détail, des différences de densité sont observés entre les deux systèmes, ce qui traduit une différence d'organisation. A moyenne et à longue distance, une succession de valeurs hautes et basses montrent qu'il existe encore des structures régulières avec des répétitions spatiales. La position des pics est encore différente entre les deux systèmes.

Bien que se structurant tous les deux en chaîne, ces données indiquent que les deux systèmes s'organisent différemment en fonction de la composition en lipide.

Nous avons ensuite voulu exploiter la capacité des différentes surfaces à accueillir d'autres monomères pour développer un modèle à plus basse résolution basé sur des simulations de Monte-Carlo dans le but de comprendre les forces à l'origine de l'organisation en chaîne de ces protéines.

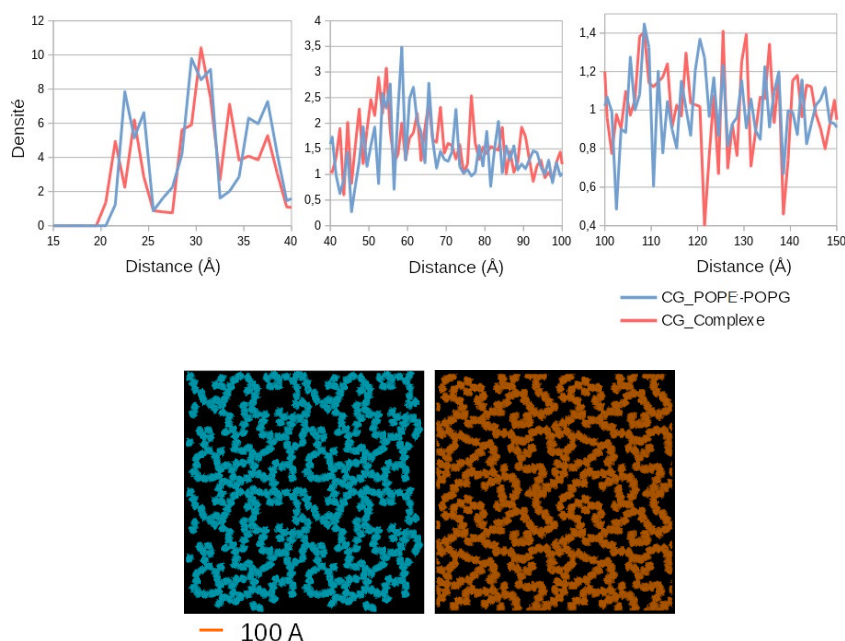


FIGURE 5.5 – Fonctions de distribution radiale (en haut) de la membranes simple (bleu) et complexe (rouge) pour les grandes proximités (à gauche), les proximités intermédiaire (au centre) et les plus grandes distances (à droite).

5.1.4 Modèle de Monte-Carlo : de très gros grains

Dans cette modélisation, les quatre faces d'interaction sont représentées chacune par une particule. Ainsi, chacune peut être paramétrée indépendamment de ses voisines par un potentiel attractif ou répulsif avec différentes profondeurs de puits et de portée d'interaction (J. P. DUNEAU et al., 2013). Nous avons utilisé les comptages des contacts entre faces pour paramétrer ces différentes intensités d'interaction, la portée étant maintenue constante à 3 fois le rayon de chaque bille. Différentes simulations ont ensuite été réalisées en faisant varier la profondeur du puits et le nombre de contacts enregistrés à partir duquel le potentiel devient répulsif.

Après 50000 étapes de simulation, la configuration finale est analysée et comparée aux simulations gros grain. Dans ce cas, des fonctions de distribution radiale très différentes sont obtenues lorsque les paramètres sont modifiés (figure 5.6). On peut observer soit des objets relativement agrégés avec la présence d'espaces importants entre eux (figure 5.6, gauche), soit des interactions plus relâchées qui correspondent à une plus grande solubilisation des monomères dans la membrane. Dans tous les cas, ce modèle simple n'est pas suffisant pour rendre compte de l'organisation en chaînettes observées dans nos simulations gros grain.

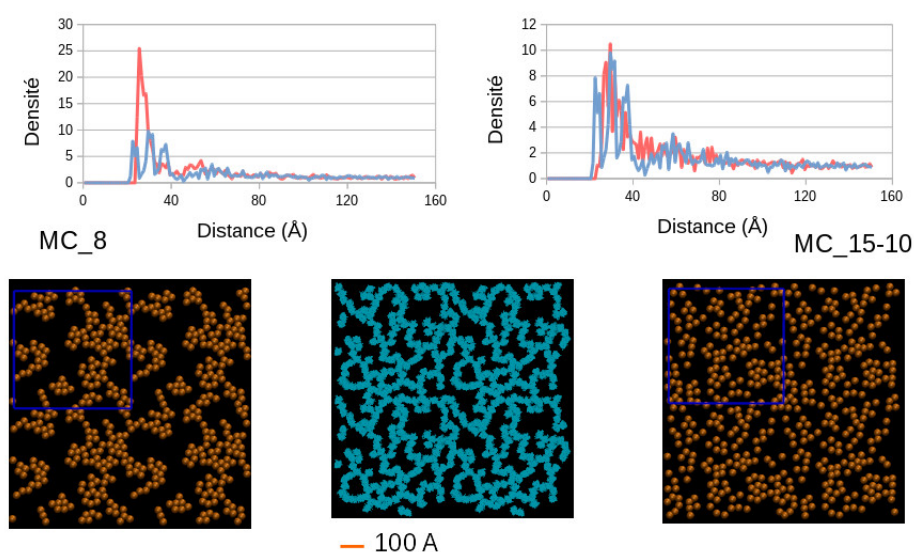


FIGURE 5.6 – Fonction de distribution radiale des protéines de la membrane simple comparée à celles issues de simulation de Monte-Carlo. Les diagrammes du haut renvoient à des jeux de potentiel d'interaction différents qui ont été construits à partir des comptages des contacts entre les faces des monomères dans la simulation de la membrane simple encombrée : à gauche, le potentiel "MC 8" est attractif à partir de 8 contacts enregistrés, à droite 15 contacts sont nécessaire mais l'intensité du potentiel a été réduite d'un facteur 10 (potentiel MC 15-10). En bas les structures obtenue à l'issue des simulations de Monte-Carlo (à gauche et à droite), comparée à la structure de référence (au centre).

5.1.5 Flux de lipides entre les protéines

Grâce à une méthode inspirée des modélisations des vents ou courants océaniques (CHAVENT, REDDY et al., 2014), nous pouvons relever les mouvements larges, de flux, des lipides pour comprendre comment, dans une telle membrane encombrée, peuvent s'organiser les échanges de lipides entre les protéines. Une question est de savoir dans quelle mesure, dans une membrane encombrée, il peut exister des lipides "libres" d'interactions avec une surface protéique. Le résultat de ces "streamlines" est montré sur la figure 5.7 pour chaque feuillet au début de la simulation, sur une période qui correspond à la mise en place du réseau en chaîne.

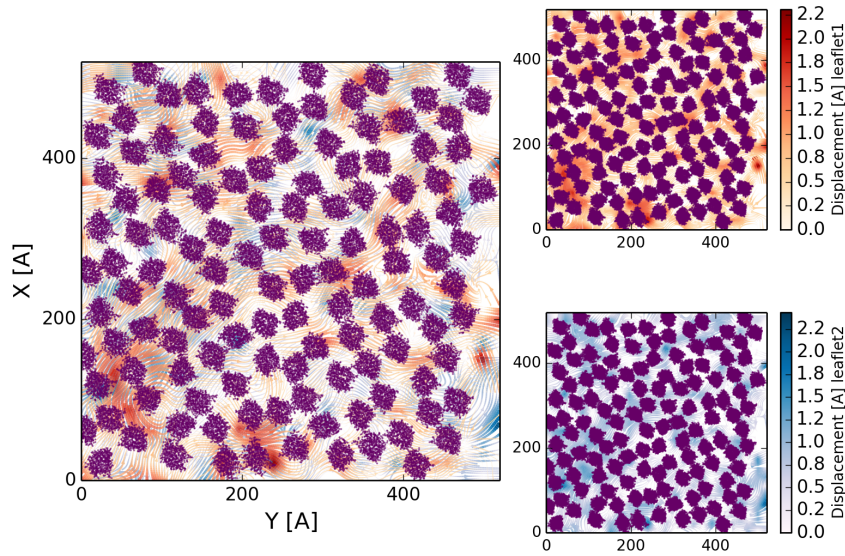


FIGURE 5.7 – Représentation en "streamlines" à 0.1 ns.

Les lignes de flux se répartissent inégalement sur la surface membranaire, ce qui indique que des contraintes variables s'appliquent aux lipides de la membrane. De plus, ces contraintes s'appliquent différemment suivant le feuillet.

5.2 Discussion

Les simulations des systèmes encombrés ont montré que la composition en lipide de la membrane influence l'agrégation des protéines en biaisant les possibilités d'orientation relative entre les monomères. Ceci est corrélé aux différents enrichissements des lipides sur la surface des protéines, qui peuvent aboutir à la modulation des zones privilégiées pour l'établissement de contacts protéine-protéine.

Notre modèle de Monte-Carlo ne nous a pas permis de comprendre les forces à l'origine de l'organisation des protéines dans nos systèmes gros grain. Cependant, nous avons vu au travers des densités de protéine que, plus que les contacts entre les faces, les contacts entre les pointes semblent déterminants. Nous devons donc modifier le modèle d'interaction afin de pouvoir intégrer ce mode privilégié d'association. Par ailleurs, nous avons simplement essentiellement joué sur des différentiels d'interaction entre les faces, mais pas encore sur la portée de l'interaction. Nous pensons que celle-ci pourrait être critique et pourra être modifiée pour prendre également en compte l'existence de barrières d'énergie potentielle à longue distance.

Notre étude sur la densité moyenne des lipides à la surface des protéines ne rend pas compte des variations qui pourraient exister entre les différents monomères. Aussi, il serait intéressant de pouvoir étudier ces fluctuations en relation avec les flux de lipides que nous avons pu mettre en évidence grâce à l'approche des lignes de flux.

Section 6

Conclusion

Les protéines membranaires évoluent dans un environnement complexe. Afin d'accomplir leur fonction, elles auront parfois besoin de lipides particulier (comme les cardiolipines), mais elles doivent aussi faire avec un ensemble de contraintes variées. Parmi ces contraintes, la plus grande est probablement la compétition avec les autres protéines : comment est-ce qu'une protéine particulière va s'approprier un lipide, essentiel, alors que d'autres protéines en auront aussi besoin ? Et au delà du lipide, comment une protéine va-t-elle réussir à oligomériser avec ses bons partenaires dans une membrane remplie d'autres protéines ?

Dans cette thèse, nous avons utilisé des méthodes *in silico* telles que la dynamique moléculaire gros grain et les simulations de Monte-Carlo pour répondre à ces questions au mieux. Nous avons de plus utilisé deux approches principales dans la manière d'appréhender ces problèmes. La première est d'isoler une ou deux protéines dans un système simple et d'étudier les relations avec les lipides ou la deuxième protéine. La deuxième est d'au contraire aller vers une augmentation de la complexité et tenter de comprendre comment ce que l'on a observé sur des systèmes plus simples se combinent et peuvent amener à une compréhension plus globale de ce qu'il se passe dans la membrane.

Dans la section 3, nous avons tout d'abord calculé le potentiel de force moyenne entre deux monomères d'Aquaporine Z et observé l'existence de forces s'exerçant à longue distance entre deux monomères. Ces forces s'appliquent différemment suivant l'orientation relative des protéines et s'étendent jusqu'à une distance minimale de 50 Å depuis la surface de la protéine ; en effet, la taille de nos systèmes ne nous a pas permis d'arriver à une distance où les monomères ne s'influençaient plus. Dans ces calculs, nous apportons de plus une caution méthodologique dans les calculs d'énergie libre des protéines membranaires qui concerne les degrés de libertés orthogonaux. En effet, certaines variables comme les orientations relatives entre les deux protéines doivent être au moins surveillées et doivent être pris en compte par l'algorithme d'intégration thermodynamique dans le cas où elles ont une diffusion trop lente pour les échelles de temps des simulations.

Sans prendre en compte ces orientations relatives, l'estimation de l'énergie libre pourra ne pas être correcte. La section 7 montre que, même dans le cadre d'un protocole classique d'Umbrella Sampling appliqué au modèle de référence du segment transmembranaire de la Glycophorine A, des biais sur l'échantillonnage de ces orientations mènent à des distorsions dans le PMF calculé.

Ces résultats montrent aussi que le protocole ABF, parce que moins dépendant du chemin originalement établi sera moins soumis aux blocages qui conduisent au biais décrits. La section 9 suivantes montre cependant que le protocole que nous avons développer pour pallier ces défauts souffre de défauts techniques qui limite son usage en routine.

Dans la section 4, nous nous sommes intéressés à l'origine de ces forces à longue distance dans le cadre d'un monomère ou d'un tétramère d'Aquaporine Z isolé dans une membrane. Nous avons étudié plus particulièrement l'organisation des lipides autour de la protéine, seul candidat pouvant transporter ces forces au delà du cut-off des simulations réalisées. Dans cette étude, nous avons mis en relation les forces à longue distance précédemment décrites avec l'adaptation des lipides à la surface de la protéine et le tri qui s'opère sous la dépendance de cette surface. Nous avons de plus complexifié notre membrane en y ajoutant des cardiolipines, un lipide majeur de la membrane interne d'*E.coli*, et en faisant varier le degré d'insaturation. Dépendant des propriétés physico-chimiques de la surface de la protéine, les lipides vont alors être triés selon leur charge, faisant écho à la "positive-inside rule" mais aussi selon leur degré d'insaturation qui permet une adaptation à la topographie de cette surface. Ainsi, la complexification de la membrane ne s'oppose pas à la formation de perturbations qui là encore s'étendent sur de longues distances.

Dans la section 5, nous avons ajouté un degrés supplémentaire de complexification en augmentant la taille de la membrane et le nombre de protéines. Nous nous sommes ici intéressé à l'impact des forces entre les monomères d'Aquaporine Z et l'organisation des lipides sur la structuration des protéines dans un contexte de membrane encombré. Dans le contexte de ces membranes, qu'elle soit de composition "simple" ou "complexe", les protéines s'organisent dans un réseau de chaînettes sur l'échelle de la dizaine de microseconde. Nous avons observé que la composition en lipide de la membrane influence l'agrégation des protéines en biaisant les possibilités d'orientation relative entre les monomères. En même temps, les différents enrichissements des lipides sur la surface des protéines aboutissent à des zones privilégiées pour l'établissement de contacts protéine-protéine.

Comme spécifié dans mon introduction, Langmuir (LANGMUIR, 1917) insistait sur une différence entre des "forces physiques" qui s'appliqueraient jusqu'à une distance infinie de manière non spécifique et des "forces chimiques" qui au contraire auraient une portée limitée et dépendrait de la nature des atomes qui

forment les molécules et de leurs voisins. Plus particulièrement, il décompose ensuite ces "forces chimiques" en deux catégories qui ont encore une pertinence maintenant :

- La première est une force très spécifique qui agit d'atome en atome, de manière similaire dans toutes les directions. Cette catégorie peut alors accueillir les forces qui gouvernent l'ancrage des lipides à la surface de la protéine et les spécificités d'interaction à courte distance entre les protéines.
- La deuxième, bien que moins spécifique, provient de la combinaisons des différents atomes et peuvent alors agir dans des directions spécifiques. Les forces à plus longue distance, qui dépendent des faces des protéines et qui gouvernent leurs probabilités d'orientations relatives peuvent entrer dans cette catégorie. Ces forces vont aussi correspondre aux effets de tri des lipides qui se répercutent à plus longue distance à partir de la surface de la protéine.

Nous pouvons ainsi décrire les différentes influences dans "la vie" d'une protéine membranaire. Au moment de sa traduction et de son repliement, cette protéine va avoir trois problèmes majeurs : ne pas agréger avec d'autres protéines, trouver ses partenaires et récupérer les lipides nécessaires à sa fonction.

A peine produite, elle va être confrontée aux interactions à longue distance émanant des effets "lipophobes" que nous avons décrit (J.-P. DUNEAU et al., 2017). Inversement, elle va aussi être à l'origine de ce même effet ; ces processus vont alors être indispensable pour assurer la dispersion des protéines dans le cadre naturel d'une membrane native très encombrée. Cet effet va prendre son origine dans le tri et l'adaptation des lipides sur la surface de la protéine, tout en sélectionnant les lipides nécessaires à sa fonction comme les cardiolipines. Ces cardiolipines vont aussi permettre à la protéine de se réorienter plus facilement dans la membrane. A partir de ces lipides à la surface et se propageant de couche de lipide en couches de lipides, des forces à longues distances vont aussi être à l'origine du recrutement des partenaires de cette protéine en favorisant des réorientations, aboutissant à des chemins privilégiés amenant au premier contact.

A ce moment, où mon travail s'arrête, un nouveau problème se présente : le remplacement des lipides ancrés sur les surfaces par les surfaces des protéines partenaires.

Dans la continuité de ce travail, un objectif optimiste sera d'étudier l'influence de mutations de l'AqpZ sur les PMF tridimensionnels obtenus dans la section 3. Cependant, cette perspective sera dépendante des moyens mis en oeuvre pour résoudre les problèmes techniques liées à l'exploration des orientations relatives couplée à l'ABF.

Un deuxième objectif sera d'analyser plus en détails le comportement de la membrane encombrée afin de pouvoir reproduire l'organisation en chaînettes dans des simulations de Monte-Carlo, afin de comprendre par quelle combinaison d'interaction à courte et à longue distance ces organisations vont naître.

Plus ambitieusement, nous pourrions alors trouver un jeu de paramètres pour retrouver les faces natives dans les simulations de Monte-Carlo.

Il nous faudra aussi analyser les flux de lipides et les propriétés de transport d'une surface à l'autre en fonction de l'état d'oligomérisation de la protéine : est-ce que les lipides disponibles peuvent contribuer à organiser cinétiquement l'assemblage des protéines ?

Enfin, un objectif plus éloigné serait de trouver des coordonnées généralisées permettant de comprendre les dernières étapes de l'assemblage après le premier contact, qui mettent en jeu le remplacement des lipides par des surfaces protéiques. Tout cela pour finalement arriver à comprendre toutes les étapes qui amènent à un tétramère.

Section 7

Utilisation du gros grain avec l'ABF, dans NAMD

Dans la section 3, nous avons vu que, pour étudier la thermodynamique de l'assemblage de deux monomères d'Aquaporine Z, l'utilisation de coordonnées généralisées supplémentaires est indispensable. Alors que la distance était correctement échantillonnée, les orientations relatives entre les protéines ont dûes être contrôlées pour permettre l'intégration thermodynamique. Cependant, les calculs réalisés concernent un couple de monomère dont la taille est bien supérieure à celle habituellement étudiée dans ce type de calcul. Le plus souvent, les assemblages étudiés se limitent à ceux de segments transmembranaire uniques dont on cherche à caractériser les interactions. Dans ce cadre, celui de la Glycophorine A constitue le modèle de référence. Nous avons donc voulu tester sur ce système notre implémentation du couplage de l'ABF et du champ de forces MARTINI au sein du logiciel NAMD. En comparant notre approche à celle classique de l'umbrella sampling, nous avons de nouveau observé comment un défaut d'exploration de l'espace configurationnel et notamment orientationnel peut altérer la qualité du PMF produit. Dans la section d'après, nous reviendrons sur les difficultés techniques et algorithmiques qui limitent l'usage de cette implémentation.

Les schémas d'intégration thermodynamique requièrent une hypothèse fondamentale : les degrés de libertés orthogonaux, qui ne sont pas sous le contrôle de l'algorithme d'intégration, doivent être échantillonnés correctement ; c'est à dire échantillonnés selon la fonction de partition sous-jacente. Alors qu'il est moins problématique de sous-échantillonner des états de faible probabilité d'existence, qui sont des zones de haute énergie sur le PMF, cela redevient un problème lorsque des zones de faible énergie sont cachées derrière des barrières énergétiques qui ne peuvent pas être surmontées dans les échelles de temps de la simulation. Cette préoccupation est d'autant plus critique dans les systèmes gros grains membranaires lorsque, malgré une diffusion facilitée, les temps des mouvements d'intérêt sont encore plus grand.

Nous avons en effet vu précédemment (section 3) que les réorientations des protéines sont des mouvements qui ne sont pas échantillonnés spontanément dans les échelles de temps des simulations. La possibilité de surveiller des

coordonnées de réaction autres que la coordonnée d'intérêt devient donc indispensable. Alors que cela est déjà possible en utilisant le module colvars avec une variété importante de coordonnées de réaction prédéfinies, cette possibilité est restreinte à l'utilisation du logiciel NAMD. Cependant, dans le contexte de simulations membranaires, le modèle gros grain MARTINI est utilisé et optimisé avec le logiciel GROMACS.

Comme nous sommes conscients que la facilité d'utilisation des algorithmes est une des deux raisons principales de leur diffusion (l'autre étant leur utilité), nous sommes convaincus que le couplage du module colvars, qui permet de suivre une grande variété de coordonnées de réaction et d'utiliser plusieurs algorithmes d'intégration thermodynamique, avec le champ de forces gros grain MARTINI permettrait de fondamentalement améliorer la poursuite de l'énergétique de systèmes gros grain et d'apporter un plus grand focus sur les coordonnées orthogonales.

Cette section s'intéresse donc à valider l'utilisation du module colvars, dans NAMD, avec le champ de force gros grain MARTINI en utilisant la classique dissociation du dimère de Glycophorine A (GpA). Nous comparerons le PMF obtenu par l'utilisation de l'ABF dans NAMD avec celui obtenu par US dans GROMACS d'après le protocole de SENGUPTA et al. (2010).

7.1 Résultats

Pour étudier l'association d'un dimère de GpA, nous avons cherché à nous placé dans des conditions aussi proches que possible de celles proposées par SENGUPTA et al. (2010). Le dimère de GpA a été inséré dans une membrane de DPPC. Après une équilibration d'une microseconde, nous pouvons alors effectuer l'intégration thermodynamique selon la distance entre les monomères.

L'ABF mono-dimensionnel en utilisant NAMD a utilisé la distance entre le squelette des monomères comme coordonnée de réaction. Dans ce calcul, la distance a été explorée de 5 à 25 Å par pas de 1 Å pour mimer les 21 fenêtres d'US réalisé dans l'article de référence. Le PMF de l'association du dimère de GpA obtenu après 20 μ s de simulation, est présenté dans la figure 7.1.

Ce PMF présente un minimum global unique à 5 Å. Un premier épaulement est présent vers 12 Å suivi d'un second minimum, puis un dernier épaulement à 15 Å avant d'arriver à un plateau à 25 Å. La différence d'énergie libre entre l'état dimérique et l'état totalement dissocié est d'environ 12.5 kJ.mol⁻¹.

L'US a suivi le protocole de SENGUPTA et al. (2010) avec GROMACS 5.1.4. Dans cette simulation, le potentiel parapluie agit sur les billes du squelette du premier peptide avec une constante de force de 1000 kJ.mol⁻¹nm⁻². Le calcul a consisté originellement en 21 fenêtres d'échantillonnage qui correspondent à un décalage de 0.1 nm par simulation. Afin d'assurer un échantillonnage le meilleur possible, le nombre de simulations a ensuite été augmenté jusqu'à 97. Dans l'article original, les structures initiales ont été extraite de simulation

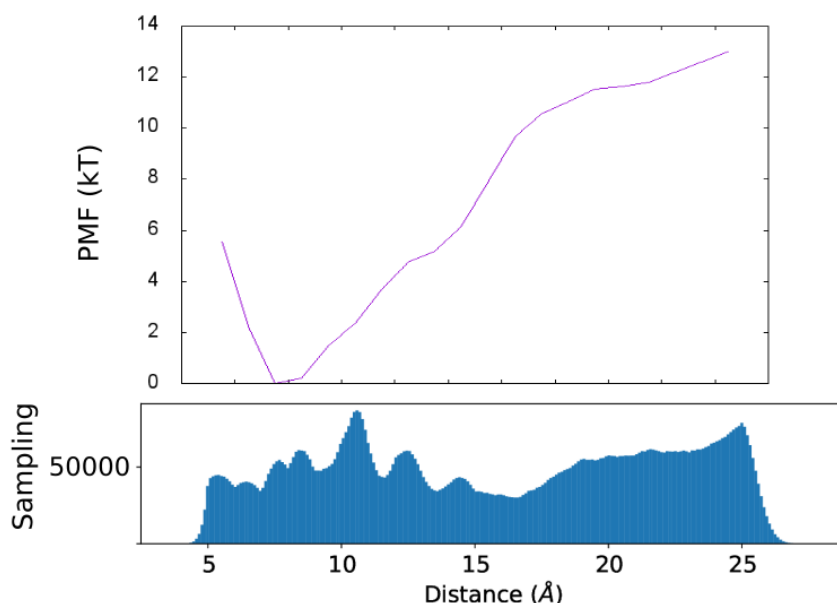


FIGURE 7.1 – Potentiel de force moyenne de l'association du dimère de GpA selon la distance entre les monomères par ABF. L'échantillonnage correspondant est montré en dessous.

provenant d'assemblages spontanés. Pour cette étude je suis parti d'une simulation de dynamique dirigée à partir de l'état dimérique et équilibrés pendant 200 ns. Chaque fenêtre a ensuite été simulée pendant 2 μ s. La méthode WHAM a été utilisée pour débiaiser le potentiel parapluie. Le PMF correspondant est présenté dans la figure 7.2 (haut).

Sur ce profil, il y a deux minimums à 6 et 9 Å qui correspondent à des états associés. Ensuite, au plus les segments transmembranaires de la GpA s'éloignent, au plus l'énergie libre augmente jusqu'à un épaulement vers 15 Å pour arriver à un plateau à 25 Å. Il n'y a pas de barrière à l'association. La différence d'énergie libre entre l'état dimérique et l'état totalement dissocié est d'environ 11.5 kJ.mol⁻¹.

Une différence majeure avec la simulation de référence de SENGUPTA et al. (2010) et notre étude ABF précédente est la présence de ce deuxième minimum à 9 Å. L'échantillonnage correspondant est montré dans la figure 7.2 (bas). Ce deuxième minimum est dû à deux simulations qui présentent une distribution Gaussienne plus resserrée autour de 9 Å. Une de ces configurations est présentée dans la figure 7.3.

A cette distance, le système est bloqué dans une configuration où les hélices se touchent sur la portion N-terminale.

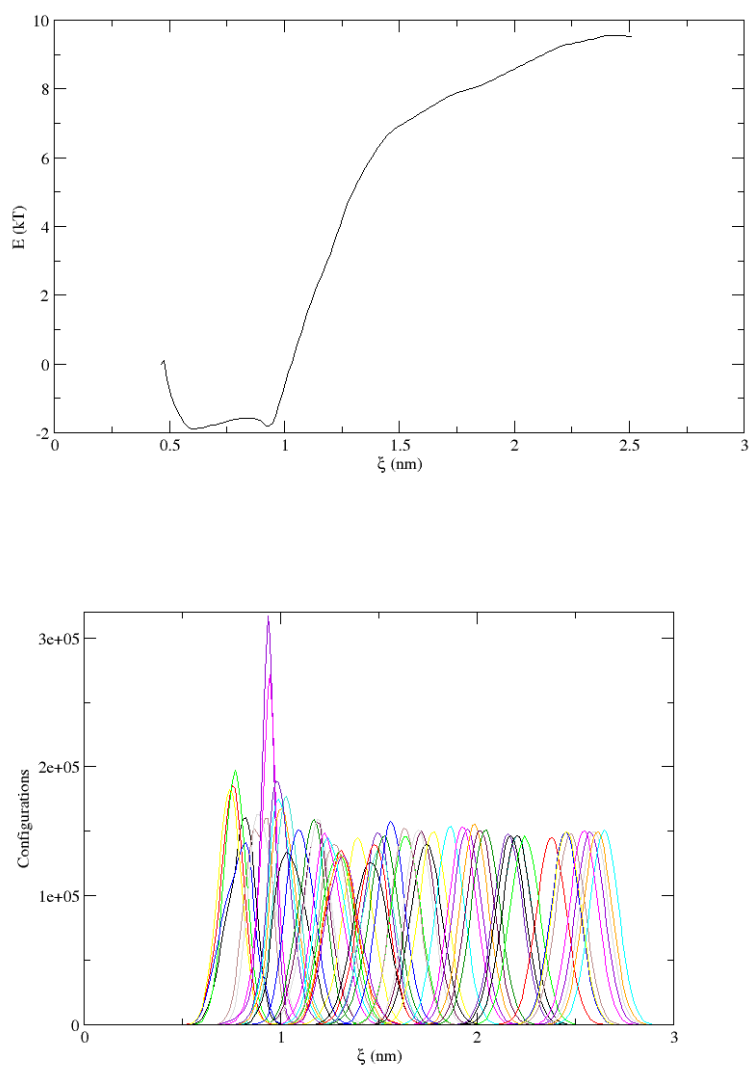


FIGURE 7.2 – Potentiel de force moyenne de l'association du dimère de GpA selon la distance par US (haut) et échantillonnage de cette simulation (bas).

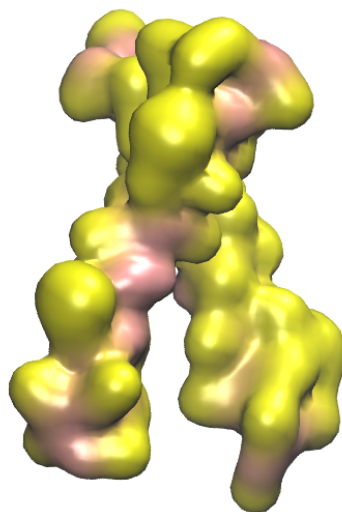


FIGURE 7.3 – Configuration problématique de l’Umbrella Sampling. Chaque hélice de Glycophorine A est représentée en surface avec le squelette en rose et les chaînes latérales en jaune.

7.2 Discussion

Nous avons calculé le PMF de l’association du dimère gros grain de GpA en utilisant l’US (classique, dans GROMACS) et l’ABF (dans NAMD). L’ABF en particulier présente une bonne adéquation avec les PMF publiés malgré de petites différences (SENGUPTA et al., 2010). De manière surprenante, une meilleure adéquation est même trouvée avec un PMF calculé pour un système tout atome (KUZNETSOV et al., 2015).

En revanche, le PMF réalisé en utilisant l’US montre des limites. En effet le deuxième minimum observé dans notre étude, à 9 Å, est absents dans les autres PMF publiés. Ce minimum, qui correspond à des configurations où les protéines sont bloquées semble provenir d’un artefact lié aux conditions de production des coordonnées initiales. Pour notre US, elles sont issues d’une simulation de dynamique moléculaire dirigée démarrée à partir de la structure cristallographique du dimère. Cette simulation a favorisé le maintien de contacts inter-protéines y compris a de grandes distances. Aussi, le chemin emprunté et l’échantillonnage qui a suivi a pu privilégier des configurations à forte enthalpie supportant même la formation d’un tilt important. Cependant, de tels états devraient être entropiquement défavorisés notamment à cause de l’exclusion des lipides de la surface concernée. Dans un contexte où le gros grain favorise notoirement les contacts entre hélices membranaires (JAVANAINEN et al., 2017), le risque est alors de rester bloqué pendant tout le temps de l’échantillonnage.

Au contraire, dans l’ABF, les protéines font des allées et venues entre

les différentes positions de la coordonnée de réaction, créant ainsi de multiples événements de dissociation et d'association. De même, pour leur US, SENGUPTA et al. (2010) ont généré leurs configurations initiales à partir de peptides séparés qui se sont auto-assemblés. De cette manière, les conformations initiales étaient probablement moins stabilisées et donc plus facilement dissociables.

Le souci rencontré avec notre protocole d'Umbrella Sampling souligne de nouveau le risque associé au mauvais échantillonnage des coordonnées orthogonales à celles contrôlées (et en particulier celles liées à l'orientation relative). Dans cette situation, à partir du moment où un blocage opère, les autres orientations ne sont plus explorées menant à une mauvaise appréciation du PMF. Dans ce cadre, l'ABF a permis de s'affranchir de ces problèmes. Comme dans la section 3, cette étude souligne le rôle que peuvent jouer les coordonnées de réaction orthogonales dans le calcul d'un PMF.

Section 8

Modélisation d'un polymère amphiphile, mimant les propriétés du copolymère de SMA

Nous avons vu dans la section 4 que des enrichissements en lipides dépendent de la nature des surfaces protéiques. En particulier, les Cardiolipines (CL) se positionnent de façon préférentielle à l'interface entre les sous-unités d'Aquaporine Z en accord avec des données expérimentales de spectrométrie de masse native. Cependant, les enrichissements que nous observons sur la surface ne sont pas compatibles avec ceux observés au laboratoire par V.Schmidt à partir de l'utilisation d'un polymère de SMA (Styrene Maleic Acid) (figure 8.1).

L'analyse de la composition en lipide des nanodisques montre une déplétion en CL. Nous avons donc des simulations qui montrent un enrichissement en CL et des expériences qui montrent une déplétion. Le polymère de SMA est décrit dans la littérature comme permettant de capturer la protéine et son environnement comme le ferait un emporte pièce (Figure 8.2) afin d'obtenir des nanodisques.

Est-ce que le modèle gros grain *in silico* n'est pas à même de capturer les enrichissements des lipides autour la protéine ? Ou est-ce que les expériences biaisent notre vision de l'environnement immédiat de la protéine à cause du polymère ? Pour répondre à ces question, j'ai donc développé un modèle gros grain d'un polymère amphiphile similaire au SMA.

8.1 Résultats

8.1.1 Rapprocher la simulation de l'expérience : développement d'un polymère gros grain

Le développement d'un polymère gros grain doit suivre les contraintes du champ de forces MARTINI avec des billes prédéfinies qui représentent environ 4 atomes "lourds". Pour cette raison, un polymère réaliste de SMA

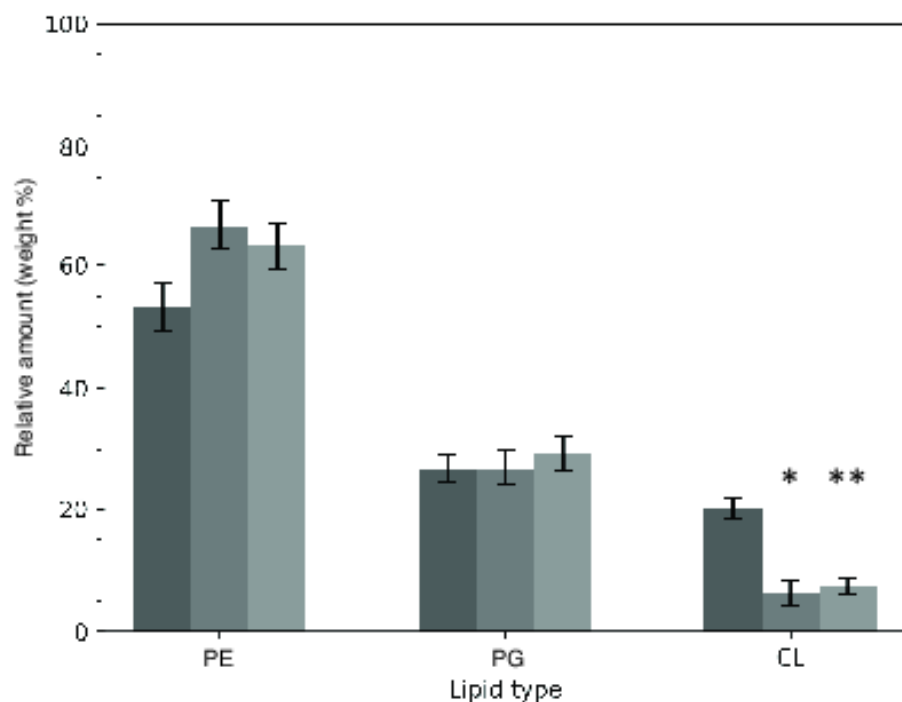


FIGURE 8.1 – Compositions en têtes polaires des lipides isolés de la membrane interne d'*E. coli* en association avec l'AqpZ et le mutant déstabilisant W₁₄A dans des nanodisques de SMA. Les compositions montrées sont les moyennes et erreurs standard déterminées en utilisant trois réplicats biologiques. La composition en CL des nanodisques est significativement réduite (*, $P < 0.05$; **, $P < 0.01$) relativement à la composition de la membrane. La barre foncée à gauche indique les mesures sur la membrane totale; la barre centrale, les mesures sur les lipides co-purifiant avec l'AqpZ dans les nanodisques; et la barre claire à droite les lipides co-purifiant avec l'AqpZ-W₁₄A dans les nanodisques.

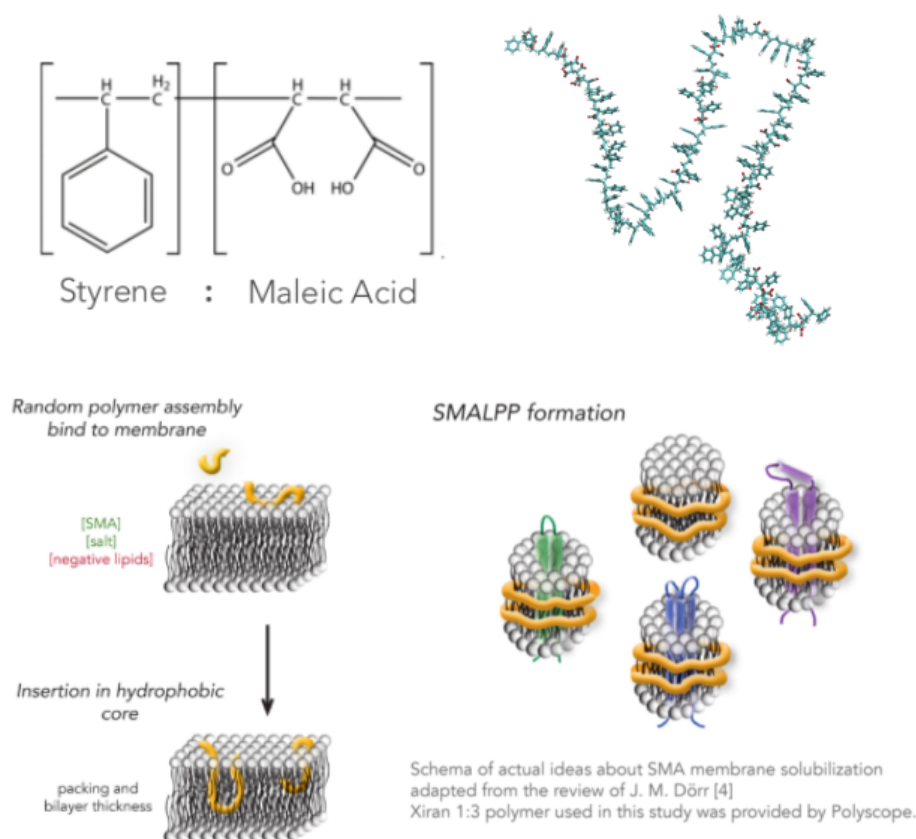


FIGURE 8.2 – Unités du SMA et représentation atomique d'un polymère (haut). Schéma possible de la solubilisation de membranes biologiques par un polymère de SMA (bas).

est impossible à modéliser et nous avons plutôt cherché à modéliser l'impact d'un polymère amphiphile, mimant les propriétés du SMA (un agencement de billes négatives et de styrènes) suivant différents paramètres de billes.

J'ai donc développé un polymère gros grain mimant les propriétés du polymère Styrene Maleic Acid grâce à la modularité du champ de forces MARTINI. Pour le styrene, j'ai utilisé 4 billes hydrophobes (de notation C et SC pour les billes plus petites des cycles dans MARTINI) en me basant sur la paramétrisation existante du polystyrène (ROSSI et al., 2011); pour l'acide maléique, sa structure impose deux billes chargées négativement car il n'y a aucune particule avec deux charges négatives dans le champ de forces. La représentation de ces deux résidus est montrée dans la figure 8.3.

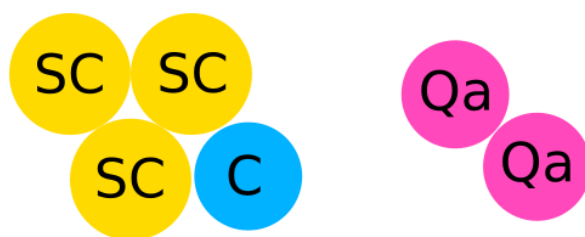


FIGURE 8.3 – Schéma gros grain du styrène (gauche) et acide maléique (droite) formant le polymère de SMA. Les billes C et SC sont hydrophobes alors que la bille Qa est chargée négativement. Ces billes permettent de reproduire les propriétés principales d'un polymère amphiphile tel que le SMA.

Une première paramétrisation a ensuite été obtenue grâce à automartini (BEREAU et al., 2015) pour les dimères de résidus (styrène/acide maléique et styrène/styrène) en utilisant la notation SMILES correspondante et en bloquant le nombre de billes voulues dans le script. Un polymère de 101 résidus est ensuite généré par un script (disponible à l'adresse https://framagit.org/msidore/thesis_energetics_lipids/tree/master/Notebooks/Tri_des_lipides) qui va randomiser l'ordre styrène/acide maléique, en s'assurant de ne pas avoir deux acides maléiques proches, car le SMA ne semble pas posséder deux acides maléiques côte à côte (SCHEIDELAAR et al., 2015), et générer une topologie. Pour faciliter la génération, le polymère résultant est linéaire; il sera ensuite minimisé et simulé dans l'eau (Figure 8.4).

Le choix de 101 résidus par polymère correspond à la moyenne du nombre de résidus par polymère déterminée expérimentalement au laboratoire. J'ai choisi 101 résidus au lieu de 100, alors que ce dernier aurait été un arrondi prenant en compte la variabilité de l'échantillon tout en satisfaisant un besoin humain compulsif d'arrondi, car ce nombre me semblait tout autant arbitraire

Type	SMILES
S	<chem>[C@H][C@@H](c1ccccc1)</chem>
SS	<chem>[C@H][C@@H](c1ccccc1)[C@H][C@@H](c1ccccc1)</chem>
SM	<chem>[C@@H][C@H](c1ccccc1)[C@H](C([O-])=O)[C@H](C([O-])=O)</chem>
MS	<chem>[C@@H](C([O-])=O)[C@@H](C([O-])=O)[C@@H](c1ccccc1)[C@@H]</chem>



FIGURE 8.4 – SMILES utilisés et exemple de SMA linéaire produit par le script.

que 100 dans ces circonstances.

Les systèmes contenant ce polymère seront détaillés dans les résultats. Brièvement, la construction s'est faite en plaçant manuellement des polymères au dessus d'une membrane puis en utilisant INSANE pour la solubilisation. Un nanodisque constituant une parcelle de membrane a quant à lui été construit en plaçant manuellement des polymères non collapsés autour d'un patch de membrane contenant l'Aquaporine Z.

8.1.2 Trouver l'hydrophobicité parfaite : solubiliser une membrane avec un polymère

Les billes hydrophobes du champ de forces MARTINI peuvent prendre 5 niveaux d'hydrophobicité différents, nommés de C1 à C5, pour le squelette du polymère ou les billes plus petites des cycles qui appartiennent au styrène. Nous avons alors mis en contact des polymères avec une membrane dans un ratio polymères/lipides similaire à un nanodisque de SMA.

L'hydrophobicité de ces billes est le paramètre qui va déterminer comment le polymère va solubiliser la membrane. Lorsque l'hydrophobicité est maximale (billes C5, figure 8.5, gauche), le polymère ne pénètre pas dans la membrane alors que lorsque l'hydrophobicité est minimale (billes C1, figure 8.5, droite), le polymère se mélange complètement à la membrane. Cependant, dans un nanodisque de SMA, il faut un équilibre entre le mélange et l'exclusion totale avec les lipides. Nous avons donc construit un polymère plus équilibré, avec un mélange de billes peu hydrophobes au niveau du squelette et de billes moyennement hydrophobe pour les cycles (ce mélange sera noté C1/SC3, avec des billes C1 pour le squelette et des billes C3 pour les cycles). Lorsque ce polymère est mis en contact avec une membrane (figure 8.6, gauche), il ne va se mélanger que partiellement avec celle-ci, ce qui est le comportement recherché pour un nanodisque. Sur une membrane plus grande contenant une AqpZ, le comportement est similaire (figure 8.6, droite).

Cependant, les échelles de temps des simulations ne permettent pas d'arriver à la formation de nanodisques (une solubilisation complète des membranes prend plusieurs heures au laboratoire). Comme nous ne pouvons

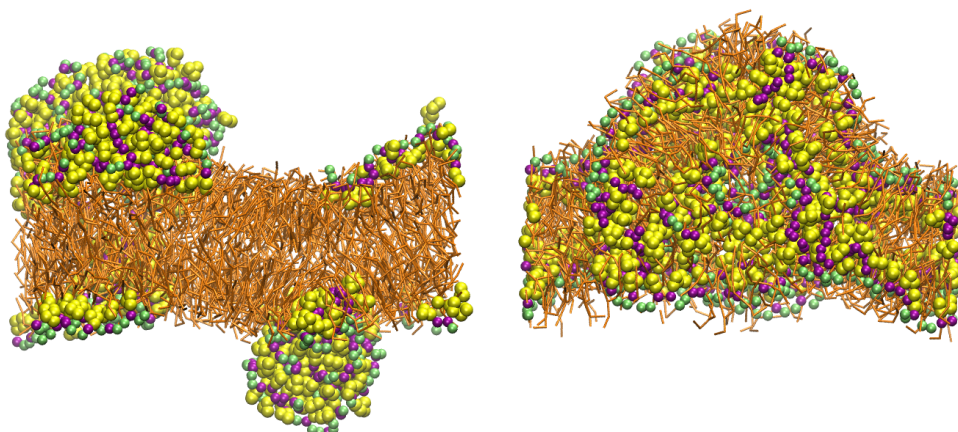


FIGURE 8.5 – Mise en contact du polymère très hydrophobe (billes C5, gauche) ou peu hydrophobe (billes C1, droite) avec une membrane. Le polymère très hydrophobe ne solubilise pas la membrane alors que le polymère moins hydrophobe se mixe totalement avec elle. Les lipides de la membrane sont représentés en bâtons oranges, le polymère est représenté en sphères de Van der Waals avec les billes des cycles des styrènes en jaune, les billes hydrophobes du squelette en violet et les billes de l'acide maléique, négatives, sont en cyan.

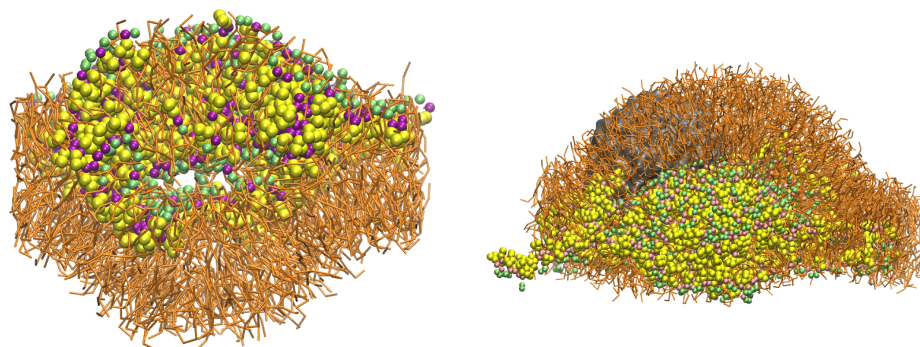


FIGURE 8.6 – Mise en contact d'un polymère équilibré (billes C1 pour le squelette des styrènes et C3 pour les cycles) avec une membrane (gauche) et avec une plus grosse membrane contenant une AqpZ (droite). Ce polymère ne solubilise pas totalement la membrane. Les lipides de la membrane sont représentés en bâtons oranges, le polymère est représenté en sphères de Van der Waals avec les billes des cycles des styrènes en jaune, les billes hydrophobes du squelette en rose et les billes de l'acide maléique, négatives, sont en vert.

pas savoir si ces paramètres permettraient d'obtenir des nanodisques, nous nous contenterons donc de ce comportement intermédiaire entre la mixité complète et l'exclusion complète pour le choix des paramètres du polymère. D'autres tests ont été effectués avec des combinaisons de billes plus ou moins hydrophobes (données non présentées), mais nous retiendrons la combinaison C1/SC3 qui semble garantir le comportement voulu dans les limites indiquées précédemment.

Pour obtenir un nanodisque, nous sommes donc partis d'une portion de membrane contenant l'AqpZ provenant de la section 4 et avons placé des polymères autour.

8.1.3 Un nanodisque

La construction du nanodisque et le nanodisque équilibré est présenté dans la figure 8.7.

Le nanodisque a été construit en positionnant 4 polymères de SMA aux pourtours d'une parcelle de membrane complexe équilibrée contenant 232 lipides et un tétramère d'AqpZ. On observe le positionnement spontané du polymère autour du patch de membrane en une dizaine de nanosecondes. Le nanodisque a ensuite été simulé pendant 1 μ s.

8.1.4 Enrichissements différentiels dans le nanodisque

La table 8.1 indique la présence des cardiolipines à différentes distances de la surface de la protéine pour la membrane complexe décrite dans la section 4 et pour le nanodisque. On observe une forte diminution de la quantité de cardiolipines dans la couche la plus externe, celle qui est en contact avec le polymère. En revanche, cela se traduit par un fort excès de CL au contact de la protéine mais aussi dans la tranche intermédiaire.

TABLE 8.1 – Proportions (%) et enrichissement des cardiolipines dans la membrane complexe et dans le nanodisque en fonction des différentes couches de lipides autour de la protéine.

Distance (Å)	Membrane	Nanodisque	Ratio nanodisque/membrane
0-6	11	25	2.3
6-12	6.7	11	1.6
12-18	7.5	4	0.5

Ceci confirme la possibilité d'une influence du polymère sur la répartition des lipides au sein du nanodisque.

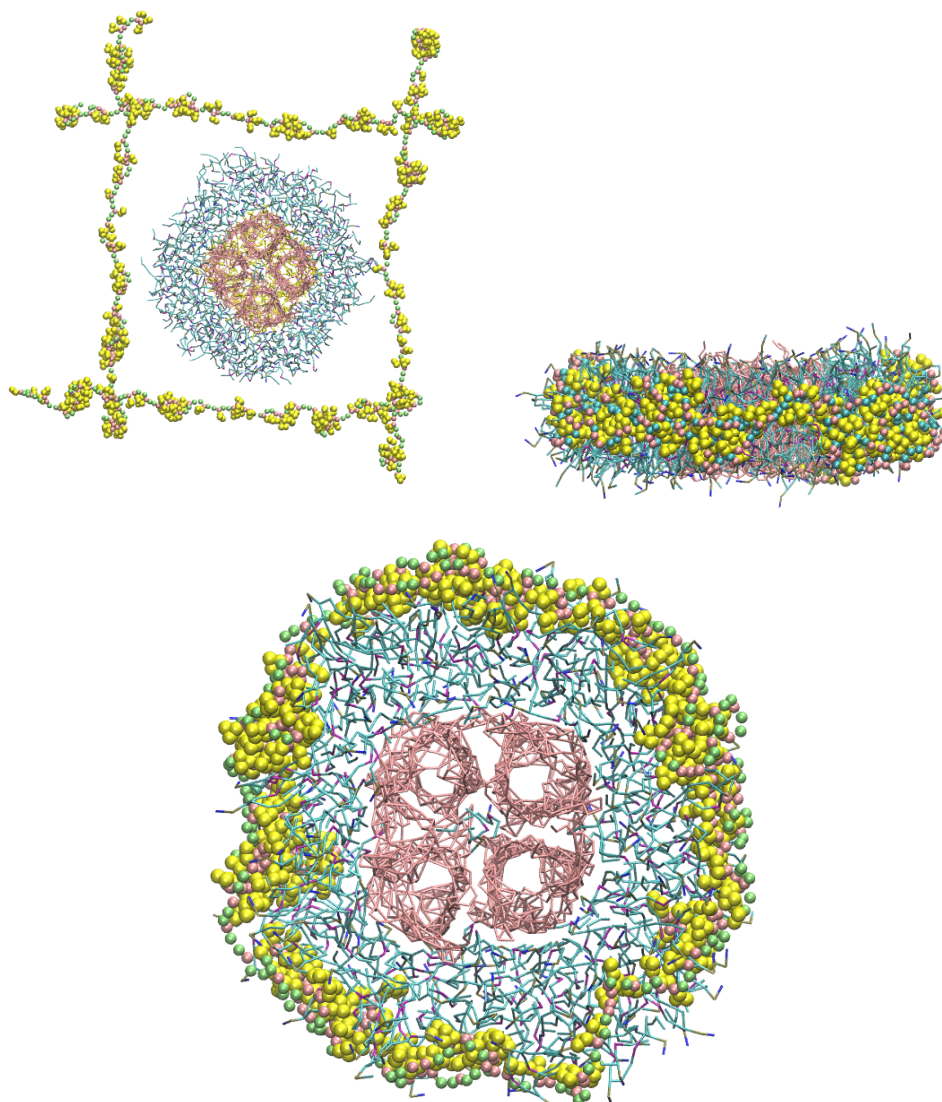


FIGURE 8.7 – Configuration initiale du patch membranaire comprenant un tétramère d'AqpZ, son annulus et 4 polymères de SMA (en haut à gauche) et nanodisque obtenu après 1 μ s de simulation (en haut à droite et en bas). Les lipides de la membrane sont représentés en bâtons, le polymère est représenté en sphères de Van der Waals avec les billes des cycles des styrènes en jaune, les billes hydrophobes du squelette en rose et les billes de l'acide maléique, négatives, sont en vert et le squelette de la protéine est en bâtons roses.

8.2 Discussion

Afin de résoudre un conflit entre la simulation et l'expérience concernant l'enrichissement de certains lipides au contact de l'Aquaporine Z et dans son annulus, nous avons développé un modèle simple de polymère amphiphile mimant les propriétés du SMA. Nous avons trouvé un jeu de paramètres qui permet de former un complexe stable dans lequel le polymère se maintient sur le pourtour d'un patch membranaire dont la taille a été imposée par des contraintes expérimentales (clichés de microscopie électronique (SCHMIDT et al., 2018; DÖRR et al., 2014)).

Au cours d'une simulation de seulement une microseconde, nous avons ensuite observé une modification des enrichissements en lipide, et en particulier en cardiolipine, entre le bord du nanodisque et la surface de la protéine. Il était attendu que, par simple répulsion électrostatique, le polymère repousse les cardiolipines vers la protéine.

Ce résultat indique que, même dans une situation où le disque serait globalement déplété en cardiolipines, il peut favoriser le maintien des CL à la surface de la protéine. Cela pourrait jouer un rôle de stabilisation structurale ou fonctionnelle pour les nombreuses protéines qui montrent une dépendance à ce lipide.

Nos résultats ne permettent cependant pas de prédire l'effet de la solubilisation des membranes par le SMA sur la composition finale du nanodisque. Ce paramètre peut également dépendre des autres protéines membranaires, des conditions et de la cinétique de la solubilisation. Dans le cadre de l'Aquaporine Z, il n'y a donc pas d'incompatibilité entre l'observation d'une déplétion globale en CL dans les nanodisques et le maintien d'un enrichissement de ce lipide en surface.

Récemment, l'équipe de Peter Tieleman a présenté un autre modèle de SMA gros grain compatible avec le champ de forces MARTINI (XU et al., 2018). Leur modèle est basé sur l'utilisation d'une bille supplémentaire qui concerne le squelette de l'acide maléique. De plus, ils ont utilisé des billes plus petites pour le squelette (SC1) et les billes des cycles sont plus hydrophobes que dans notre modèle (SC5). J'ai étudié ce dernier point et constaté dans mes conditions que le polymère avait tendance à se collapser plus fortement et à avoir une moindre miscibilité avec la membrane, de façon analogue à celle observée dans la figure 8.5. Les cinétiques en jeu dans les simulations sont de toutes façons difficilement comparables à celles mises en oeuvre dans les protocoles expérimentaux. Cela se traduit pour cette étude comme pour la notre par l'impossibilité d'obtenir un nanodisque à partir de la mise en contact avec une membrane. Ce critère cinétique est donc peu pertinent pour juger de l'utilité du modèle. Dans tous les cas, ces changements au niveau des parties hydrophobes de la molécule n'auront pas de conséquences importantes sur la répulsion électrostatique avec

les cardiolipines mise en évidence dans notre étude.

Section 9

Investigation d'un crash : pourquoi MARTINI n'aime pas NAMD

Dans le domaine de la simulation numérique gros grain des membranes biologiques, le champ de forces qui s'est imposé est le champ de force MARTINI. Ce champ de forces a été développé pour le logiciel de dynamique moléculaire GROMACS mais une version existe, bien que nécessitant des modifications, pour le logiciel NAMD (voir méthodes) et qui est disponible à l'adresse <https://www.ks.uiuc.edu/Training/Tutorials/martini/rbcg-tutorial.pdf>.

Cependant, un problème récurrent dans la mailing list de NAMD et dans nos simulations est la présence de crash plus ou moins fréquents. La terminologie consacrée associée est "explosion" (http://www.gromacs.org/Documentation/Terminology/Blowing_Up) et est le résultat ici d'atomes se déplaçant trop rapidement.

Cette section est destinée à comprendre et expliquer ces explosions.

9.1 MARTINI dans NAMD ou GROMACS : différences ?

Tout d'abord, nous devons comprendre les différences d'implémentation de MARTINI dans NAMD et GROMACS.

MARTINI est un champ de force modulable possédant 16 particules différentes. Cependant, les paramètres des acides aminés individuels viennent se superposer aux paramètres des billes individuelles. NAMD ne permet pas cette combinatoire : la version du champ de force dans NAMD contient donc 111 particules différentes.

MARTINI utilise un réseau élastique pour contrebalancer la perte des structures fines des protéines comme les liaisons hydrogènes. Ce réseau élastique est modélisé dans NAMD par des liaisons additionnelles de même

constantes de force et de distance que le réseau élastique dans GROMACS.

Une dernière différence est l'existence de liaisons contraintes. Dans GROMACS, certaines liaisons (du cycle de l'histidine, du tryptophane et de la tyrosine et de la première bille de la chaîne latérale de l'isoleucine, de la thréonine et de la valine) sont contraintes par l'algorithme LINCS. Dans NAMD, des contraintes n'existent que pour les molécules d'eau par les algorithmes SHAKE ou SETTLE et ne peuvent s'étendre aux acides aminés. Une modification simple du code ne permet pas non plus d'étendre cet algorithme à des systèmes de plus de deux particules. Ces contraintes sont alors modélisées dans NAMD par des liaisons avec une grande constante de force.

Il semble alors que ce soit cette dernière différence qui soit la source de l'instabilité. Un petit écart de la position contrainte, due à un point chaud momentané, entraînerait alors un grand écart énergétique en réaction.

9.2 Etude de cas

Pour voir plus en détails ce qu'il se passe lors d'une explosion, j'ai repris une simulation d'un dimère d'AqpZ préalablement simulé et j'ai sauvegardé une configuration par étape jusqu'à l'explosion. Une vidéo de cette explosion est disponible à l'adresse <https://drive.google.com/open?id=17FoNtu0zBlktIcn1IJgq1DlBoU32zT3h> et les évolutions temporelles de l'énergie des liaisons, des angles dièdres et de Van der Waals sont présentés sur la figure 9.1.

Cette simulation a explosé au bout d'environ 2700 steps, ce qui correspond à 27ps. Le temps de simulation n'est ici pas pertinent pour interpréter cette explosion car je suis parti d'un système précédent qui avait déjà été simulé pendant 20 μ s. Les énergies des liaisons et des angles dièdres augmentent énormément à la dernière étape sauvegardée (juste avant le crash), ce qui indique qu'une bille au moins s'est largement éloignée de sa position d'équilibre. L'énergie de Van der Waals augmente tout autant et indique que cette bille s'est alors retrouvée à la même position qu'une autre. L'étape d'après, l'explosion, résulte du contrecoup des fonctions du champ de forces avec un écart à la position d'équilibre si important.

La vidéo révèle aussi un jeu de "boules de billard" où la bille incriminée et les billes alentours se déplacent loin de leur position d'équilibre et finissent par causer l'explosion.

9.3 Discussion et conclusion

Des explosions, comme dans l'étude de cas précédente, sont cependant courantes dans nos simulations. Pour cette raison, le temps d'intégration est

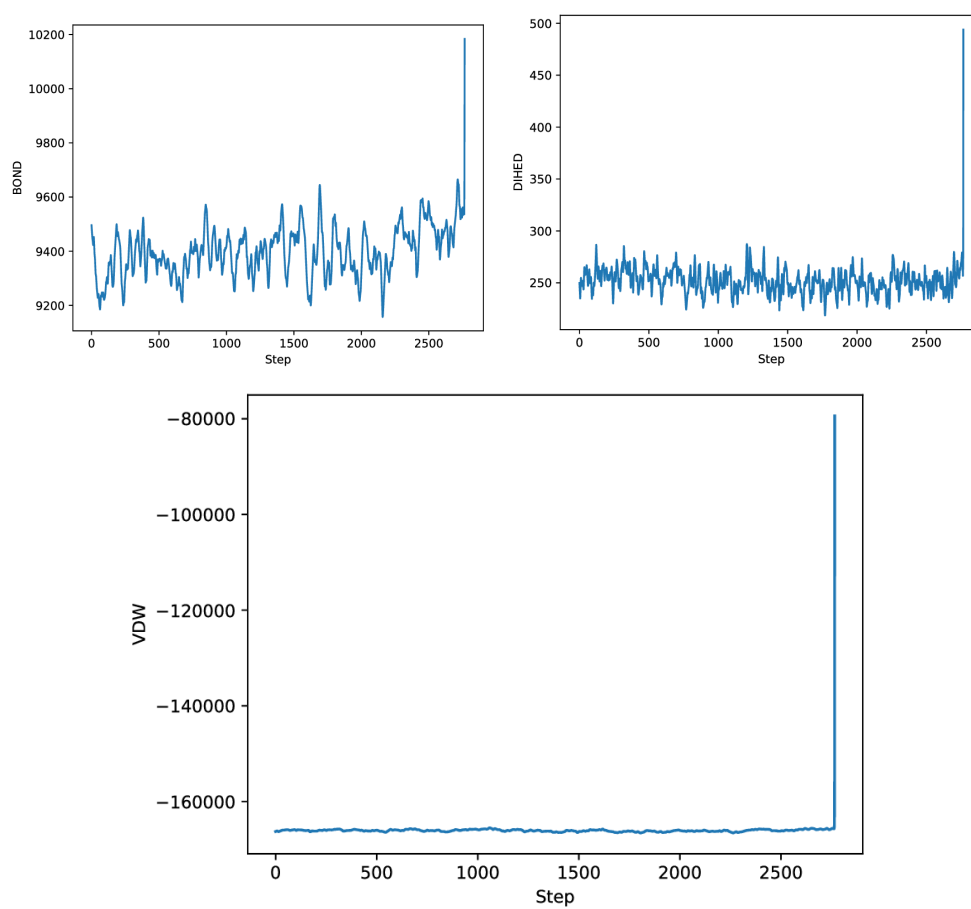


FIGURE 9.1 – Evolutions temporelles de l'énergie des liaisons, des angles dièdres et de Van der Waals.

réduit à 10 fs ; tout temps d'intégration supérieur entraîne une augmentation de la fréquence des explosions.

L'implémentation de MARTINI dans NAMD a déjà été utilisée dans la littérature (SHIH, FREDDOLINO, ARKHIPOV et al., 2008 ; SHIH, FREDDOLINO, SLIGAR et al., 2007 ; ARKHIPOV et al., 2008) sans noter ce problème. Malgré ce silence et depuis mon intervention dans la mailing list de NAMD, je reçois régulièrement des e-mails mentionnant ce problème, ce qui indique que de tels crash sont répandus.

9.3.1 Injection d'énergie dans le système

En plus de limiter le temps d'intégration à 10 fs, le remplacement de contraintes par des liaisons avec une forte constante de force peut alors injecter de l'énergie dans le système lorsque l'énergie ajoutée n'est pas suffisante pour faire exploser la simulation. Cette énergie est ensuite évacuée par le couplage au bain thermique et amène à un possible effet d'amortissement global de la vitesse des atomes, réduisant alors les mouvements.

Cette instabilité rendrait alors le système non-physique et permettrait au système, suivant les moments, de passer des barrières d'énergie ou au contraire, après amortissement de la température, de ne pas avoir assez d'énergie pour les passer.

9.3.2 Possibles solutions

Pour utiliser MARTINI dans NAMD, il faut actuellement faire avec cette instabilité. La solution la plus simple est donc simplement de relancer la simulation d'une configuration précédente et espérer que le système n'explose pas, ou au moins explose plus tard. C'est la solution mise en oeuvre dans les simulations utilisant NAMD de ce manuscrit.

Une solution possible est la diminution du temps d'intégration. Moins ce temps est grand, au moins le système explose. Cette solution n'est cependant pas envisageable car nous perdrons alors l'intérêt principal de l'utilisation du champ de force gros grain.

Une autre possibilité serait d'adapter l'algorithme LINCS dans NAMD, mais cela demanderait un travail de développement important dans NAMD.

Comme la contrainte d'utilisation de NAMD vient du module colvars (FIORIN et al., 2013), la meilleure solution serait d'adapter le module à GROMACS pour utiliser nativement le champ de forces MARTINI. Une ébauche est actuellement disponible mais nécessite encore un travail important pour la rendre fonctionnelle. En effet, les développements du champ de force MARTINI se font actuellement sur GROMACS et cela permettrait l'utilisation des dernières versions du champ de forces avec le module colvars.

Section 10

Collaboration : Dimérisation de segments transmembranaires du système de sécrétion de type VI

Le système de sécrétion de type VI (T6SS) est une machine macromoléculaire très répandue chez les bactéries à gram négative comme les Proteobacteries et les Bacteroidetes. Plus particulièrement, c'est une nano-arbalète moléculaire impliquée dans la pathogénicité envers les cellules procaryotes et eucaryotes et la conquête de nouvelles niches écologiques (HOOD et al., 2010 ; HACHANI et al., 2016 ; BINGLE et al., 2008). La perte de sa fonction impacte généralement la pathogénicité envers l'hôte (HACHANI et al., 2016 ; ALTERI et al., 2016), le T6SS a donc un rôle crucial dans la virulence en plus de son rôle antibactérien.

Il s'agit d'une machine moléculaire ancre au niveau des deux membranes. Cette machine est constituée de trois sous-complexes : le complexe d'ancrage aux membranes qui interagit avec la plateforme d'assemblage dans lequel se loge le complexe tubulaire cytoplasmique (ZOUED, BRUNET et al., 2014) (Figure 10.1). Un minimum de 13 gènes est nécessaire pour l'assemblage d'un T6SS fonctionnel, bien que l'opéron du système comporte de 16 à 36 gènes suivant les espèces. Dans notre modèle *E.coli* entéroaggrégative, les protéines composant le T6SS sont toutes nommées "Tss" (CASCALES et al., 2012).

Nous nous intéressons ici aux protéines qui composent le complexe membranaire, qui permet le recrutement de la plateforme d'assemblage et le début de la polymérisation de la queue contractile. C'est la seule partie du T6SS à ne pas être désassemblée après utilisation de la machinerie, ce qui en fait une cible thérapeutique. Ce complexe membranaire est composé de trois protéines : TssL, TssM et TssJ, assemblées en décamères (ZOUED, BRUNET et al., 2014) (Figure 10.2). TssJ est une lipoprotéine associée à la membrane externe qui possède un motif lipobox (ZOUED, CASSARO et al., 2016) et effectue une liaison avec TssM via son domaine périplasmique. TssM est ancrée à la membrane interne via 3 hélices transmembranaires (M1, M2 et

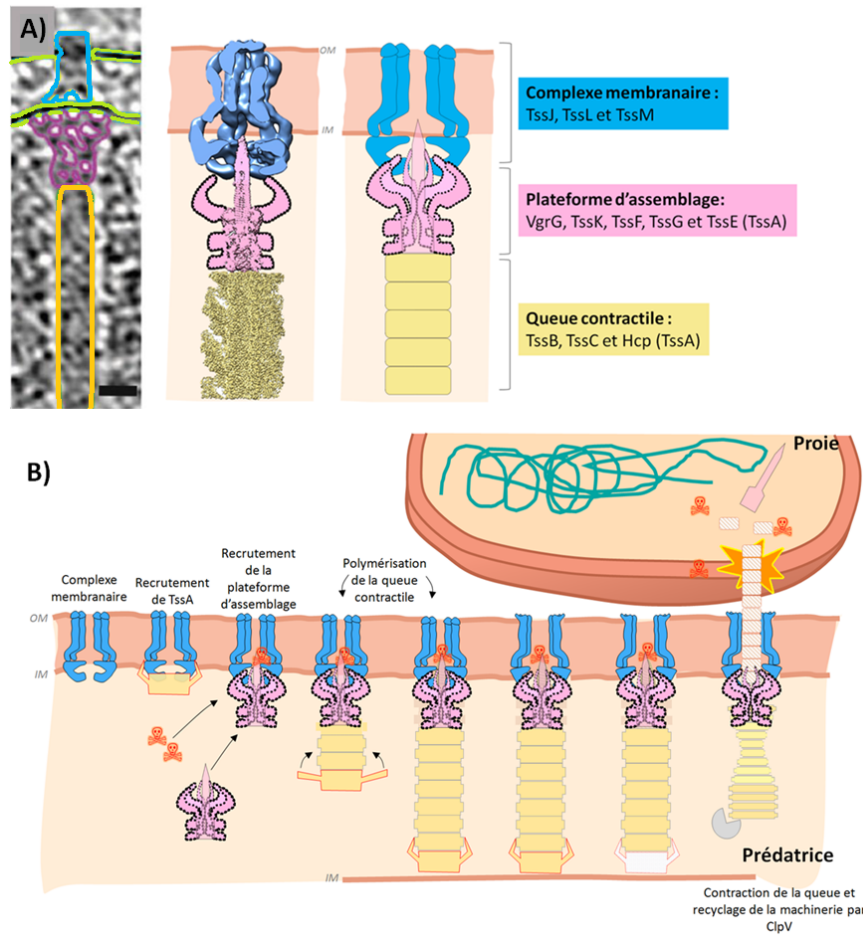


FIGURE 10.1 – Généralités sur le T6SS. Représentation schématique A) de l'architecture globale du T6SS, avec une photo de cryotomographie électronique (échelle 9.5 nm) et B) de son mécanisme.

M3) localisées dans le N-terminal, et traverse le périplasma où elle se lie avec TssJ. Il est proposé que TssM soit multimérique et forme un pore traversant la membrane interne et s'étendant jusqu'au périplasma (ZOUED, BRUNET et al., 2014). Les segments transmembranaires de TssM effectuent des contacts avec TssL (DURAND et al., 2015).

Cette dernière possède un domaine globulaire N-terminal suivi d'une hélice transmembranaire C-terminale (L) traversant la membrane interne. Bien que le domaine C-terminal de TssL soit monomérique en solution, il semblerait que la protéine dimérise au niveau de son segment transmembranaire : ce dimère est résistant en SDS (ZOUED, J.-P. DUNEAU et al., 2018).

La plateforme d'assemblage est constituée de 4 protéines (TssE, TssF, TssG et TssK) qui vont former un complexe avec VgrG, qui permet la perforation de la membrane cible (ZOUED, BRUNET et al., 2014). Les toxines vont aussi se fixer via divers mécanismes de transport à VgrG pour être éjectées dans la cellule cible. Le tube interne du fourreau contractile est formé de la protéine Hcp qui s'assemble spontanément en anneaux hexamériques à partir de la plateforme d'assemblage. Autour de ce tube interne, un fourreau de deux protéines (TssB et TssC) stabilisent l'assemblage.

Après chaque "tir" de la machinerie, le tube d'Hcp et son fourreau doivent se reformer tandis que le complexe membranaire est réutilisé. Le complexe TssJ-TssL-TssM est donc crucial au fonctionnement du T6SS car il permet l'assemblage du reste de la machinerie. Comme il est en permanence dans les membranes des cellules, il est aussi une cible thérapeutique de choix : l'inhiber revient à bloquer le T6SS. Ce projet s'inscrit dans la compréhension des interactions entre les protéines de ce complexe membranaire pour développer des peptides inhibiteurs qui viendront inhiber les interactions protéine-protéine au sein de la membrane. Nous nous sommes particulièrement intéressé à la dimérisation de TssL (L) et à son interaction hypothétique avec le segment transmembranaire M3 de TssM.

Cette collaboration pluridisciplinaire mêle des approches *in vitro* et *in vivo*, via des études de ToxCat et de compétition bactérienne menées par Camille Garcia dans le cadre de son M2, Eric Durand et Eric Cascales au Laboratoire d'Ingénierie des Systèmes Macromoléculaires, ainsi que des simulations de dynamique moléculaire permettant d'étudier l'homodimérisation de L et l'interaction entre L et M3 où j'interviens. Je présenterai donc ici les méthodes et résultats *in silico* et discuterai en faisant appel aux résultats des collaborateurs.

10.1 Matériels et Méthodes

Afin de trouver des interfaces d'interaction préférentielles pour les dimères de L et les hétérodimères L-M3, nous avons utilisé le protocole DAFT (WASSENAAR, PLUHACKOVA et al., 2015). Pour chaque couple de peptide, 500 systèmes membranaires gros grain utilisant le champ de force MARTINI

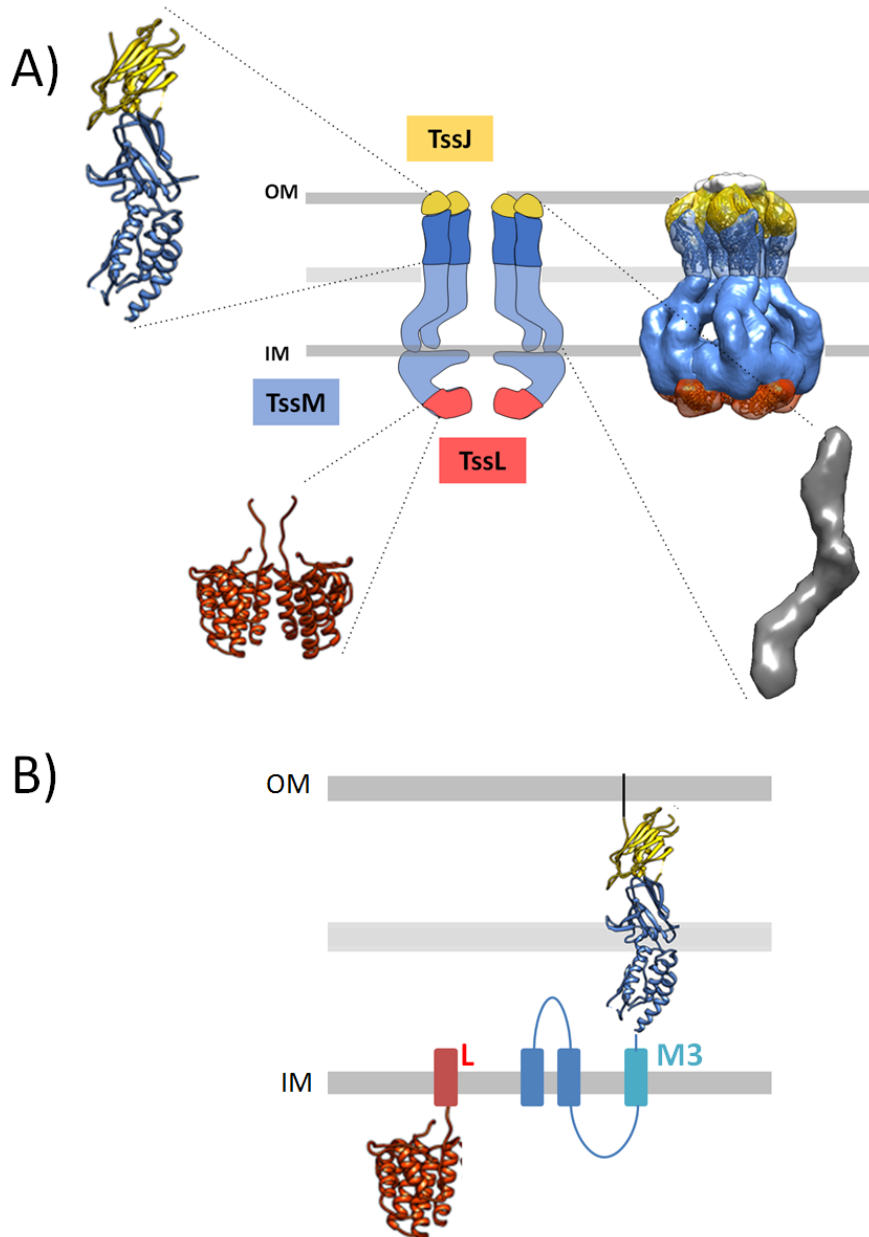


FIGURE 10.2 – A) Représentation schématique du complexe membranaire du T6SS. Ce complexe est composé de trois protéines, TssJ (jaune), TssM (bleue) et TssL (rouge). Les structures résolues des différents domaines du complexe sont représentées. B) Représentation des monomères composant le T6SS.. TssK (jaune) est une lipoprotéine ancrée à la membrane externe (OM), TssM (bleue) est ancrée à la membrane interne (IM) par trois segments transmembranaires (M1, M2 et M3) et TssL (rouge) ancrée à la membrane interne par son segment transmembranaire (L).

(MONTICELLI et al., 2008; MARRINK et al., 2007) ont été créés, chacun contenant deux peptides possédant des orientations aléatoires dans le plan de la bicouche lipidique. Cette dernière se compose de POPG et POPE avec un ratio de 1 :4 (MOREIN et al., 1996) pour modéliser la composition de la membrane interne de *E.coli* et d'une concentration en NaCl de 0.15 M.

L'intégrateur leapfrog a été utilisé avec une température de 310 K et une pression de 1 bar. La température a été maintenue grâce à l'algorithme de Velocity rescale (BUSSI et al., 2008) avec une constante de couplage de 1.0 ps^{-1} et la pression a été maintenue grâce à l'algorithme de Berendsen avec une constante de couplage de 1.0 ps^{-1} . Les potentiels de Lennard-Jones et de Coulomb utilisent une fonction de shift au cut-off à 1.1 nm.

Chaque simulation a été équilibrée pendant 10 ps de NVT avec un temps d'intégration de 2 fs et 100 ps de NPT avec un temps d'intégration de 20 fs. Les simulations de production ont ensuite utilisé un temps d'intégration de 20 fs et un temps de simulation de 0.5 μs . Toutes les simulations ont été effectuées à l'aide de GROMACS 5.1.4 (BERENDSEN, SPOEL et al., 1995; ABRAHAM et al., 2015).

Les analyses ont été conduites selon les recommandations de DAFT. Dans un premier temps, les peptides ont été recentrés dans les boîtes grâce aux outils de GROMACS. Dans un second temps, pour chaque configuration finale des 500 simulations, une carte de contact a été calculée à partir des paires de distance entre deux peptides et a été convertit en mesure de proximité, entre 0 et 1, selon la formule (10.1).

$$p_{i,j} = e^{0.1*(d_{i,j})^4} \quad (10.1)$$

Avec $p_{i,j}$ la proximité entre les résidus i et j et $d_{i,j}$ la distance entre les résidus i et j. Seules les proximités entre 0.7 et 1 sont ensuite visualisées. Les cartes de contacts entre peptides ont été réalisées à l'aide du package MATPLOTLIB sous python 2.7.4 et les densités protéiques ont été calculées en utilisant le plugin VOLMAP de VMD (HUMPHREY et al., 1996), puis les images ont été produites à l'aide de VMD.

Les scripts utilisés dans ce travail sont disponibles à l'adresse https://framagit.org/msidore/thesis_energetics_lipids/tree/master/scripts/daft.

10.2 Résultats

Afin de trouver les interfaces d'interaction préférentielles entre les peptides M3, L et le mutant utilisé en RMN appelé 2K, qui possède deux arginines en amont et en aval du segment transmembranaire, le protocole DAFT a été utilisé. nous avons ainsi obtenus des cartes de proximité entre les peptides L-L, L-M3, 2K-M3, L-2K et 2K-2K (Figure 10.3). La visualisation spatiale des

peptides correspond à la densité moyenne des configurations finales sur 500 simulations de chaque dimère.

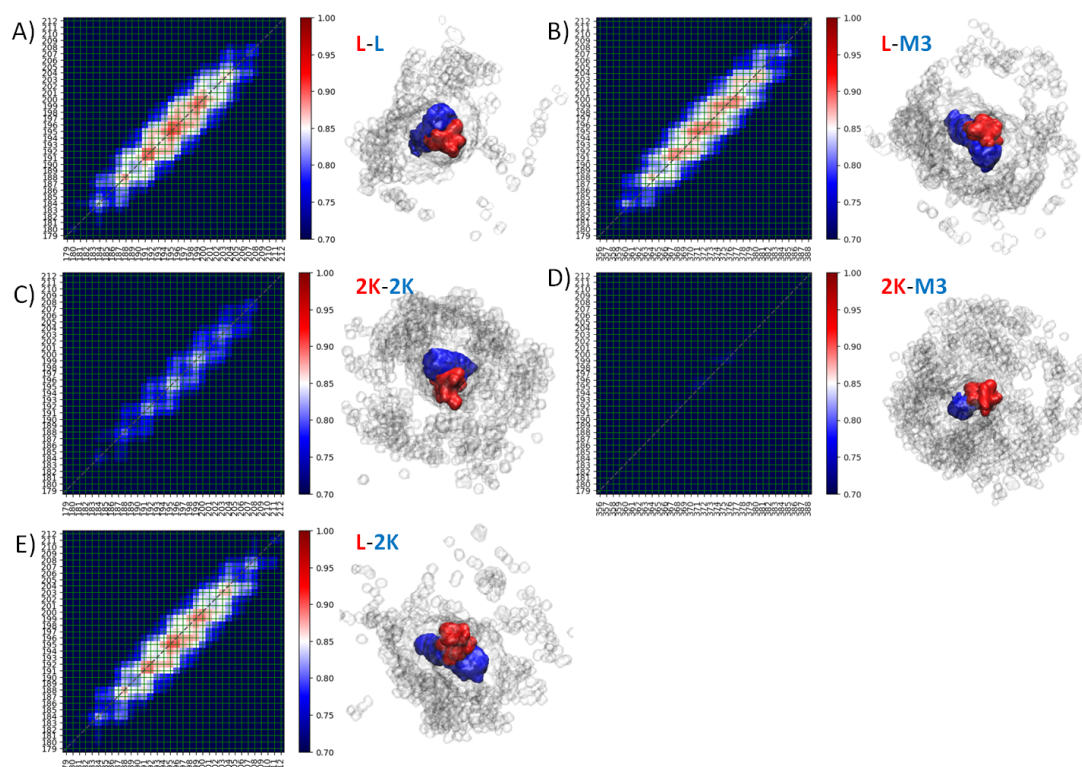


FIGURE 10.3 – Cartes de proximité entre les peptides et densités de présence des peptides *in silico*. Étude dynamique des interactions A) homo-L, B) entre L et M3, C) homo-2K, D) entre 2K et M3 et E) entre L et 2K. L'intensité du rouge des pixels est proportionnelle à la moyenne de proximité entre résidus. La représentation des interactions après simulation des peptides, où tous les peptides rouges sont orientés de la même façon, est présente à droite des cartes. Les densités bleues correspondent à la densité moyenne de présence des peptides partenaires dans les configurations finales des 500 simulations effectuées.

Nous observons une interaction importante entre L et lui-même et entre L et M3 (Figure 10.3A et B). Les données de proximité de la simulation homo-L sont conformes avec les hypothèses des précédentes études et la simulation entre L et M3 renforce une hypothèse établie au laboratoire sur la base d'études préliminaires de mutagenèse et de RMN.

Nous observons ensuite une interaction plus faible entre 2K et lui-même et 2K et M3 ((Figure 10.3C et D). Cependant, l'interaction entre L et 2K est presque équivalente à l'interaction homo-L (Figure 10.3E).

Le peptide L est donc capable de se lier à L et M3 alors que 2K montre une plus grande sélectivité en faveur de sa dimérisation avec L. Cette expérience soutient aussi l'hypothèse d'une interaction entre L et M3.

10.3 Discussion

Le système de sécrétion de type VI (T6SS) est une machine moléculaire majeure impliquée dans la virulence et la compétition bactérienne. Ce système est similaire à une nano-arbalète, traverse les deux membranes et est composé d'un complexe membranaire, d'une plateforme d'assemblage et d'une queue contractile. Nous nous sommes intéressés aux interactions des segments transmembranaires des protéines du système membranaire, plus particulièrement des segments transmembranaires L et M3. Un variant utilisé précédemment en RMN qui possède deux arginines en amont et en aval, 2K, a aussi été utilisé.

Les résultats obtenus par dynamique moléculaire suggèrent que les peptides modifiés qui inhibent le plus serait moins sensible à l'homodimérisation et donc plus disponibles pour l'interférence. En effet, le peptide 2K interagit faiblement avec lui-même alors qu'il interagit favorablement avec une séquence sauvage (Figure 10.3C et D). Cependant, il y a une bonne interaction entre les peptides L et 2K, mais sur une interface de L différente de l'interaction homo-L, comme indiqué sur les densités de présence A et E de la figure 10.3. Pour l'interaction homo-L, les cartes de proximité sont conformes avec les hypothèses des précédentes études et nous retrouvons la même interface d'interaction que celle observée par spectroscopie RMN (ZOUED, J.-P. DUNEAU et al., 2018).

Ce résultat est cohérent avec les expériences *in vitro* via un test double hybride ToxCat (Russ et al., 1999). Dans cette expérience, nous avons observé la capacité du peptide 2K à interférer avec l'homodimérisation de L alors que le peptide L n'avait aucun effet interférant sur la dimérisation. Ce résultat contre intuitif peut être expliqué par une homodimérisation de L qui limiterait sa disponibilité pour l'inhibition. Cela suggère que les peptides modifiés qui inhibent le plus serait moins sensible à l'homodimérisation et donc plus disponibles pour l'interférence.

Les résultats des simulations montrent une interaction importante entre L et M3 (Figure 10.3A et B) qui consolide notre hypothèse d'interaction entre eux. De plus, il semblerait que le peptide 2K fixe bien L mais fixe plus faiblement M3. En effet, l'interaction L-2K a lieu au niveau de l'interface sur laquelle l'interaction L-M3 a lieu. Cela peut s'expliquer par des interactions électrostatiques différentes : lorsque l'on substitue les résidus Q209 et R210 par des lysines sur le peptide L, on favorise les ponts salins entre les lysines chargées positivement et les résidus Q et R chargés négativement dans le segment transmembranaire.

Ces travaux de collaboration seront poursuivis dans le cadre d'un financement ANR dans le but d'identifier des interactions possibles au sein du complexe membranaire L et M. Il s'agira de raffiner un modèle de cryo-microscopie électronique, actuellement disponible à 4Å de résolution, pour lequel la partie

membranaire ne présente pas de densité exploitable. Ces interactions seront également étudiées pour d'autres bactéries pathogènes qui présentent des T6SS d'architectures semblables.

Section 11

Collaboration : Découverte d'un canal à eau dans NarGH

La superfamille des enzymes à Mo/W-*bis*PGD comprend un vaste répertoire de protéines à molybdène et tungstène. Ces protéines catalysent une grande variété de réactions vitales, impliquées dans des voies métaboliques, chez les procaryotes. De plus, plusieurs membres de cette famille étaient vraisemblablement présents chez LUCA, selon des études de phylogénétiques et de paléogéochimie (NITSCHKE et al., 2009).

Les réactions catalysées comprennent des réactions d'oxydo-réduction (rédox) avec une grande diversité de substrats et de potentiels rédox. L'atome de molybdène du centre actif est lié covalamment à quatre thiolates faisant partis de deux pyranoptérines substitués par une guanosine. Cette sphère de coordination hexavalente du métal est complétée par un ligand acide aminé (Aspartate, sérine ou (séléno-)cystéine) et un atome de soufre inorganique ou un atome d'oxygène (GRIMALDI, SCHOEPP-COTHENET et al., 2013). Une énigme persiste dans cette superfamille protéique en considérant la diversité chimique limitée des sites actifs du molybdène par rapport à la grande diversité des réactivités et des substrats (GRIMALDI, SCHOEPP-COTHENET et al., 2013).

Des études récentes impliquant de la spectroscopie par résonance paramagnétique électronique et des caractérisations enzymologiques ont porté l'attention sur la participation d'un réseau de liaisons hydrogène autour du site actif dans la modulation du potentiel rédox au niveau du molybdène (S.-Y. WU et al., 2015 ; DUVAL et al., 2016). Ces études amènent à explorer l'environnement du centre à molybdène dans la détermination des interactions responsables du réglage fin de la réactivité du molybdène.

Dans cette étude, nous nous intéressons à la Nitrate réductase A d'*E.coli* (Nar). Elle est composée de trois sous-unités, NarI recevant les électrons via ses deux hemes et ancrant la protéine à la membrane, NarH transférant les électrons grâce à son réseau de complexes Fe₄S₄ (Fe-S) et NarG contenant un complexe Fe-S et un cofacteur à molybdène (MoCo) permettant la réduction du nitrate (JORMAKKA et al., 2004 ; BERTERO, ROTHERY, PALAK et al., 2003).

Cette enzyme peut être considérée comme un modèle représentatif de

la superfamille de par son repliement prototypique et le cofacteur de sa sous-unité catalytique. Du point de vue de l'évolution, Nar est une des seules enzymes à avoir été proposée comme présente chez LUCA et partage un clade majeur avec la perchlorate réductase (SCHOEPP-COTHENET et al., 2012). Nar est de plus insensible à l'oxygène et hautement stable même sans sa sous-unité NarI. Elle est en conséquence très étudiée (GRIMALDI, BIASO et al., 2016; BERTERO, ROTHERY, BOROUMAND et al., 2005).

Le site actif de NarG présente deux signatures pH-dépendantes du molybdène (V) en résonance paramagnétique électronique (EPR), dénotées "haut pH" et "bas pH" (VINCENT et al., 1978). Cependant, la relation entre ces signatures et les structures cristallographiques accessibles n'est pas claire. Une étude récente, combinant les techniques d'EPR pulsée et de calculs de théorie de la fonctionnelle de la densité (DFT), a montré que la coordination du métal se faisait par l'aspartate 222 et un oxygène, soit un groupement OH soit une molécule d'eau (RENDON et al., 2017). Le site actif est situé profondément dans l'enzyme. Un entonnoir qui débouche à la surface de la protéine (Figure 11.1) doit permettre au substrat de diffuser vers le coeur mais aucun chemin mène au voisinage du centre actif. Cependant, à proximité de cette région, une série de résidus polaire pourraient être importantes pour la réactivité de l'enzyme et pour la sélection du substrat.

Cette étude fait partie d'une collaboration entre notre équipe, le laboratoire de Bioénergétique et Ingénierie des Protéines avec Julia Rendon, Stéphane Grimaldi, Frédéric Biaso et Bruno Guigliarelli et le Laboratoire de Chimie Bactérienne avec Sinan Al-Attar et Axel Magalon, combinant de la mutagenèse dirigée, des mesures enzymatiques, des études d'EPR et de la dynamique moléculaire dans l'élucidation des déterminants du site actif de NarGH. En particulier, l'utilisation de la dynamique moléculaire a pu mettre en évidence l'existence d'un canal à eau dynamique, absent de la structure cristallographique, reliant le bulk au site actif.

11.1 Matériels et méthodes

Le champ de force AMBER ff03 (BAYLY et al., 1995) a été utilisé pour tous les systèmes, en incluant les paramètres des centres FeS et du GDP qui ont été extraits et réimplémentés à partir de SMITH et al. (2012) et MEAGHER et al. (2003) respectivement. Malgré une paramétrisation du cofacteur à molybdène récemment effectuée, l'utilisation d'une espèce versatile en électrons pose des problèmes de fiabilité dans des simulations Newtoniennes. Nous avons donc utilisé la structure de l'apo-NarGHI (PDB ID : 1SIW), qui possède une différence structurale négligeable de 0.24 Å par rapport à NarGHI (PDB ID : 1Q16). La sous-unité I, membranaire, a été supprimée. La sous-unité catalytique de l'apo-NarGHI comprend un GDP, lié non covalamment à l'enzyme, à la même position que la guanosine d'une des deux pyranoptérines du complexe natif. Les molécules d'eau cristallographiques ont toutes été

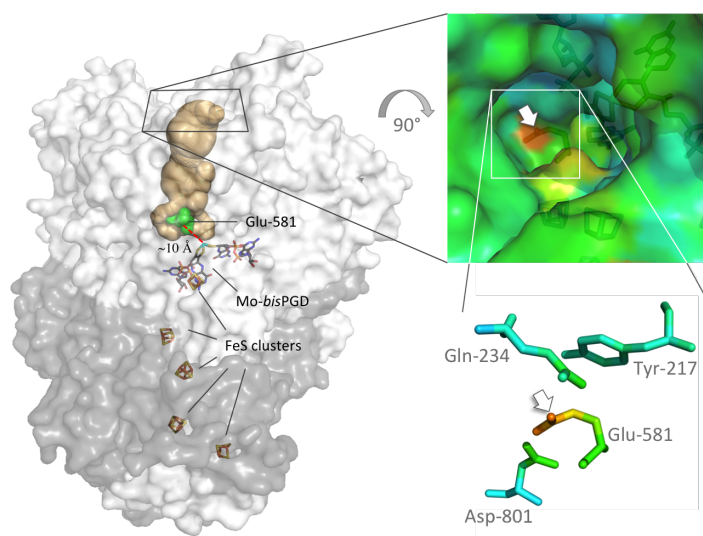


FIGURE 11.1 – Le site actif de NarGH est protégé du bulk par le résidu conservé Glu-581. Gauche : Représentation de la surface de NarG (gris clair) et NarH (gris foncé) et leurs cofacteurs. L'entonnoir du substrat (or) a été esquissé grâce à l'outil Caver, et se termine sur le Glu-581 (vert). Droite : vue d'en haut montrant la position du Glu-581 (flèche blanche) au fond de l'entonnoir du substrat et une vue simplifiée affichant le Glu-581 et les acides aminés proximaux. Les images à droites ont été colorés par le facteur b où le bleu indique une valeur faible et le orange une valeur forte. Ces images ont été créées avec PyMol par Sinan Al-Attar en utilisant le PDB ID : 1Q16.

maintenues. Propka (SØNDERGAARD et al., 2011; OLSSON et al., 2011) a été utilisé pour déterminer les états de protonation de chaque acide aminé protonable. La protéine a été solvatée par de l'eau TIP3P, au centre d'une boîte cubique de 15 nm³ comprenant 0.15 nM de NaCl, avec un excès d'ions Na⁺ pour neutraliser la charge du système.

Les simulations de dynamique moléculaire ont été effectuées avec GROMACS 5.1.4 en utilisant l'électrostatique de Particle Mesh Ewald avec un cutoff de Coulomb dans l'espace réel de 1.2 nm et un cutoff des interactions de Van der Waals de 1.2 nm avec une correction pour la dispersion à longue distance. L'algorithme de recherche de voisin a été mis à jour toutes les 10 étapes. Le temps d'intégration était de 2 fs avec un thermostat de Berendsen modifié, une constante de temps de 0.1 ps et une température de 300 K. La pression a été contrôlée avec un barostat isotropique de Parinello-Rahman avec une constante de temps de 2 ps et une compressibilité de 4.5x10⁻⁵ bar⁻¹ à une pression de 1 atm. Après une minimisation de 3000 étapes d'algorithme du gradient, une équilibration en deux étapes a été effectuée, d'abord en stabilisant la température pendant 0.1 ns puis en établissant le contrôle de la pression pendant 0.5 ns. Deux simulations ont été effectuées pendant 20 ns et une troisième pendant 200 ns. Tous les calculs ont été effectués sur 96 processeurs sur le supercalculateur OCCIGEN au CINES (Montpellier, France). Après avoir centré et ajusté la trajectoire sur la protéine en utilisant l'outil trjconv de GROMACS, les simulations ont été visualisées en utilisant VMD. L'outil Volmap a été utilisé pour calculer les cartes de densité de l'eau. Les liaisons hydrogènes ont été calculé en utilisant l'outil hbond de GROMACS. Tous les scripts utilisés dans ce travail sont disponibles à l'adresse <https://framagit.org/msidore/nargh> sous la license WTFPL.

11.2 Résultats

L'environnement du site actif de NarGH, entre le molybdène et le fond de l'entonnoir, possède la plupart de ses acides aminés conservés. Cet entonnoir mène directement sur deux acides aminés acides, le Glu-581 et l'Asp-801, dont le premier possède un facteur d'agitation thermique élevé sur les structures cristallographiques, suggérant une certaine mobilité. Cet acide aminé est de plus prédit pour avoir un pKa proche de 9, ce qui est éloigné de son pKa classique (3-4) mais qui n'est pas inhabituel dans des environnements de site actif; il est donc protoné dans nos simulations. Il se situe sur une boucle en forme de L (His-576 jusqu'à Pro-585), qui sera appelé Glu-loop (Figure 11.2, gauche)

Trois simulations ont été effectuées en partant de différentes distributions de vitesses et seront dénommées simulation 1, 2 et 3. Comme indiqué dans ces trois simulations, la Glu-loop et le Glu-581 subissent d'importants changements de conformation (figure 11.2, droite). De plus, l'analyse du réseau de liaisons hydrogènes (Table 11.1) entourant le Glu-581 indique que la dynamique locale de la Glu-loop influence fortement ce réseau.

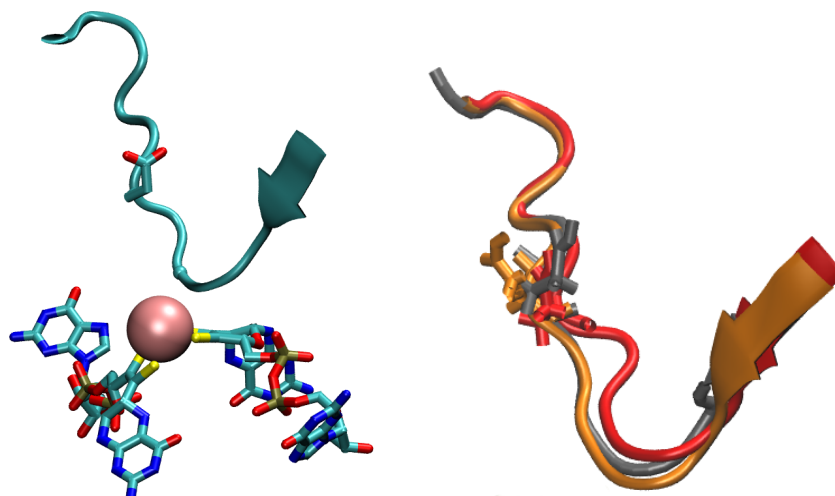


FIGURE 11.2 – Gauche : Le backbone de la Glu-loop (représentation cartoon, bleu) avec le Glu-581 (bâton), dans la structure cristallographique de l'apoprotéine (PDB ID : 1SIW). Le MoCo (bâton) n'est pas présent dans les simulations mais indique la position de la Glu-loop par rapport au site actif. Droite : La structure moyennée du backbone de la Glu-loop (cartoon) avec le Glu-581 (bâton) dans les trois simulations ; la simulation 1 est en gris, la simulation 2 est en rouge et la simulation 3 est en orange.

Résidus impliqués	Simulation 1	Simulation 2	Simulation 3	PDB ID : 1Q16
E581-D801	93.2	0.2	2.5	1
E581-Q234	3.4	18.2	94.2	0
E581-Y217	0	0	0	0
Q234-Eau	279	259	169	1
E581-Eau	215	307	291	1
D801-Eau	104	323	472	3
Y217-Eau	9.5	4.8	0.3	0

TABLE 11.1 – Occurrence (%) des liaisons hydrogènes impliquant des résidus hydrophiles enfouis dans les 3 simulations de l'apo-NarGH et nombre de liaisons dans la structure cristallographique (PDB ID : 1Q16). L'occurrence des liaisons hydrogènes est définie quand la distance entre donneur (D) et accepteur (A) est inférieure à 3.5 Å et l'angle Hydrogène-D-A est inférieur à 30°.

Trois comportements différents ont été observés. La structure de la Glu-loop de la simulation 1 ne dévie pas de la structure cristallographique, avec une liaison hydrogène en moyenne entre Glu-581 et Asp-801. A l'inverse, dans la simulation 2, la Glu-loop subit un changement de conformation qui a deux conséquences immédiates. Premièrement, la distance entre Glu-581 et Asp-801 augmente et la liaison hydrogène entre ces résidus disparaît et est remplacée par une liaison hydrogène entre Glu-581 et Gln-234. Deuxièmement, le changement de structure local de la Glu-loop éloigne la boucle du site actif. La simulation 3 montre quant à elle une situation intermédiaire; le backbone de la Glu-loop reste à la même position que dans la simulation 1 mais la chaîne latérale du Glu-581 se réoriente vers le Gln-234 pour faire une liaison hydrogène stable. Ni les simulations, ni la structure cristallographique ne montrent de liaison hydrogène entre Glu-581 et Tyr-217 (voir discussion pour ce résultat "surprenant"). Ces résultats soutiennent la vision d'une Glu-loop dynamique dans NarG, comme suggère aussi le fort facteur d'agitation thermique observé à ce niveau dans la structure cristallographique.

La table 11.1 montre aussi les degrés d'hydratation des résidus Glu-581, Asp-801 et Gln-234. Ces degrés d'hydratation dépendent ici de la position de la Glu-loop et du Glu-581. En particulier, quand la liaison hydrogène entre Glu-581 et Asp-801 est cassée dans les simulations 2 et 3, cet aspartate se retrouve hautement hydraté. La reconfiguration de l'hydratation des résidus proches du site actif n'était pas anticipée; pour évaluer la relation entre la dynamique de la protéine et l'hydratation, j'ai calculé la densité moyenne de l'eau (Figure 11.3) et suivi la diffusion de molécules d'eau uniques au cours des simulation (Figure 11.4).

Dans la simulation 1, les molécules d'eau peuvent être séparées entre 3 groupes spatiaux (Figure 11.4, simulation 1) : le premier est localisé au niveau de l'entonnoir du substrat (cyan), le deuxième se positionne dans un espace entre le Glu-581 et l'Asp-801 (vert) et le troisième se situe dans la cavité du site actif (bleu). Dans cette simulation, les molécules d'eau ne bougent pas entre ces différentes cavités, ce qui est cohérent avec les densités d'eau (Figure 11.3) qui montre que les différentes cavités ne sont pas interconnectées. Après le changement de conformation local de la simulation 2, une molécule d'eau individuelle (cyan) peut maintenant diffuser à travers les trois cavités (Figure 11.4, simulation 2). Tout d'abord, cette molécule explore la cavité du site actif, puis traverse un tunnel intermédiaire nouvellement ouvert pour finir sa diffusion dans le bulk à travers l'entonnoir du substrat. Une autre molécule d'eau (bleu), dont la trajectoire chevauche celle de la première, est capable de diffuser jusqu'à une zone où le molybdène devrait se trouver et avoir un site de coordination proche. Un tel comportement est rendu possible par l'ouverture d'un tunnel intermédiaire qui prend son origine dans le déplacement de la Glu-loop décrit dans la figure 11.2. L'existence de ce tunnel est confirmé par les calculs de densité de l'eau (Figure 11.3). Quand la structure du cofacteur MoCo est superposée sur cette trajectoire, aucun chevauchement n'est observé,

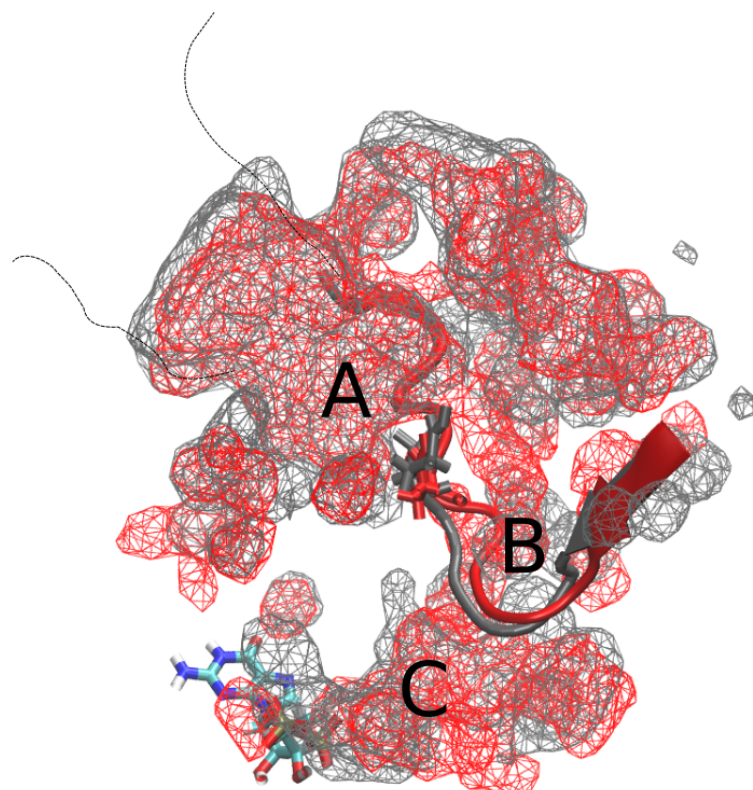


FIGURE 11.3 – Densités de l'eau moyennées sur les trajectoires dans les simulations 1 et 2. Le GDP (bâton) est présent dans la cavité du site actif de l'apoprotéine. Dans le profil de densité de l'eau de la simulation 1 (grille grise), aucune jonction n'existe entre le tunnel du substrat (A) et la zone du site actif (C). Dans la simulation 2 (grille rouge), un canal s'ouvre au milieu dans le compartiment intermédiaire (B) qui connecte les cavités A et C.

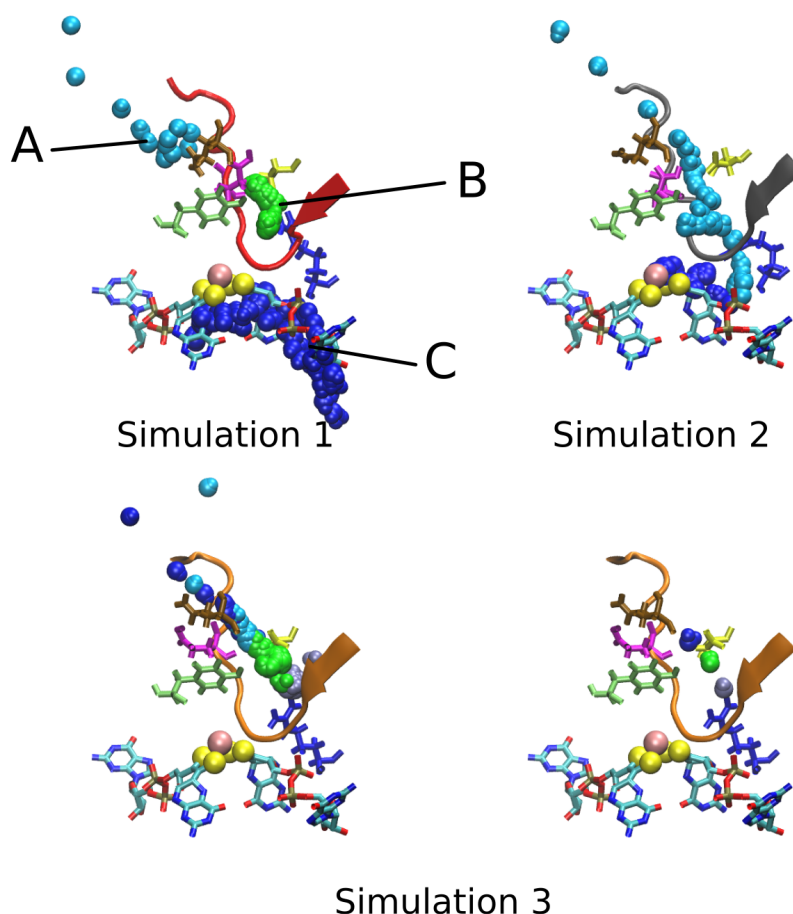


FIGURE 11.4 – Diffusion des molécules d'eau autour du site actif dans les trois simulations. Le MoCo (sphères et bâton), absent des simulations, a été ajusté sur la protéine pour dénoter le site actif. Sur chaque panneau, les résidus qui entourent un canal aqueux potentiel sont représentés en bâtons (Gln-234 en marron, Glu-581 en violet, Asp-801 en jaune, Tyr-217 en vert et Arg-260 en bleu). La traces de molécules d'eau unique est indiquée par une couleur unique : elles ne pénètrent pas dans la cavité du site actif (C) dans la simulation 1 mais traversent totalement dans la simulation 2. La simulation 2 permet le passage de molécules jusqu'à la cavité intermédiaire (B) uniquement.

ce qui indique que l'entrée de ce canal est compatible avec la structure de la holoprotéine. La structure de la simulation 2, permettant l'arrivée de molécules d'eau jusqu'à la cavité du site actif, peut alors être considérée comme une forme "ouverte".

Dans la simulation 3, où seule la chaîne latérale du Glu-581 est renversée, deux molécules d'eau présentes initialement dans le tunnel intermédiaire peuvent atteindre l'entonnoir du substrat et le bulk du solvant (Cyan et bleu, figure 11.4). Dans l'échelle de temps de cette simulation, qui a été étendu à 200ns, deux autres molécules d'eau plus profondément enfouies dans ce tunnel restent prisonnières proche de leur position initiale (gris et vert). Ces simulations suggèrent que le mouvement de la chaîne latérale du Glu-581 permet de connecter le bulk au tunnel intermédiaire, mais que le mouvement de la Glu-loop est indispensable pour permettre l'entrée jusqu'à la cavité du site actif comme vu dans la simulation 2.

Dans la forme fermée, deux molécules d'eau sont piégées dans le canal (figure 11.5 A-B). Ce phénomène apparaît amplifié dans la forme ouverte, où les molécules d'eau sont prises en charge individuellement par différents résidus hydrophiles (figure 11.5 C-F) dont certains ont fait l'objet de l'étude enzymatique. Cela permet de comprendre comment les mutations de ces résidus peuvent influencer le k_{cat} en limitant potentiellement l'accès des substrats au site actif et le k_M en modifiant la captation de ce substrat au sein de cette zone qui pourrait être vue comme un filtre de sélectivité. Par ailleurs, la figure 11.5 G montre la conservation des fonctions physico-chimiques importantes pour la formation de ce canal au sein de la perchlorate réductase, une enzyme capable de prendre en charge les mêmes substrats. Finalement, la figure 11.5 H illustre comment cette coordination successive des molécules d'eau pourrait permettre le transport des protons de l'extérieur vers le centre actif par un mécanisme de Grotthuss (CUKIERMAN, 2006).

11.3 Discussion

Ce travail s'inscrit dans une démarche interdisciplinaire mêlant enzymologie, EPR et simulations dans le but de découvrir de nouveaux facteurs qui contrôlent la réactivité du centre actif des enzymes de la superfamille des Mo/W-*bis*PGD. Dans ce but, nous avons utilisé Nar, une protéine représentante modèle de cette superfamille qui est de plus étudiée dans l'institut et où une grande quantité de données biochimiques et structurales est disponible.

La focalisation de ce travail a été sur la fonction de la Glu-581 ainsi que de son environnement polaire, tous strictement conservés, qui se situent au fond d'un entonnoir où passe le substrat et se termine à environ 10Å au dessus du molybdène du site actif. Même si cette région est éloignée du cofacteur, elle est importante pour sa réactivité. Dans cette discussion, j'intégrerai aussi les résultats obtenus par les collaborateurs, non présentés ici.

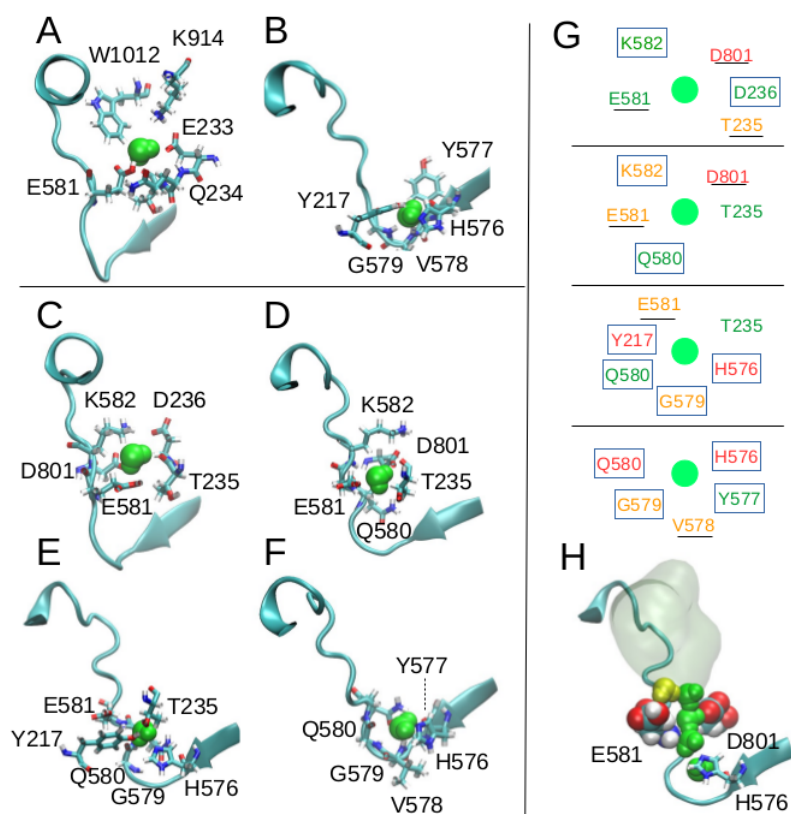


FIGURE 11.5 – Coordination des molécules d'eau dans le tunnel dynamique menant à la cavité du site actif. Dans l'état fermé (A-B, structure finale de la simulation 1), deux molécules d'eau sont piégées entre le bulk et la cavité du site actif sans possibilité de s'échapper. Les chaînes latérales qui sont à distance de contact ($< 3.5 \text{ \AA}$) sont affichées. (C à F) Quatre positions successives des molécules d'eau sont observées dans la forme ouverte (structure finale de la simulation 2). Chaque position est contrainte par des interactions qui impliquent majoritairement des résidus polaires. (G) Diagrammes de la coordination autour de chacune des molécules d'eau dans le canal. Les contacts avec des résidus polaires qui impliquent les chaînes latérales sont en rouge et ceux qui impliquent le squelette sont en orange. Les interactions de Van der Waals (en vert) contribuent à la délimitation stérique de chaque site. Les labels entourés correspondent aux résidus qui sont conservés dans la perchlorate réductase et ceux soulignés correspondent aux résidus ayant des propriétés physico-chimiques similaires. (H) Les quatre molécules d'eau (vert) sont affichées ensemble avec une molécule d'eau additionnelle (jaune). La file continue de molécules d'eau fait le pont entre l'entrée principale de l'entonnoir (transparent) au niveau du Glu-581 et de l'Asp-801 jusqu'aux résidus proches de la position théorique du molybdène comme l'His-576.

11.3.1 pK_a et acidité des résidus autour du site actif

L'analyse des pK_a nous a conduit à considérer dans nos simulation que le Glu-581 est protoné et pas l'Asp-801. Ce dernier semble pourtant faire une liaison hydrogène avec cet acide glutamique; cependant, les outils de prédictions peinent actuellement à prédire la protonation des aspartates (S. J. FISHER et al., 2009), ce qui laisse ouvert au doute l'état de ce résidu.

Une série de mutations ont été effectuées et les paramètres cinétiques de l'enzyme résultante ont été déterminés. Des mutants E581Q et D801N ont un k_{cat} drastiquement diminué et un k_M similaire à l'enzyme native indiquant d'une part que le transfert de proton émanant de ces résidus est important et d'autre part que la protonation de ces résidus est impliquée dans la fixation du substrat.

11.3.2 Implication du Glu-581 et de l'Asp-801 dans l'ouverture et la fermeture du canal

Nos résultats de simulation sont également compatibles avec la suggestion que des interactions directes entre ces deux résidus pourraient contrôler l'ouverture et la fermeture de ce canal en promouvant la réorganisation d'un réseau de liaisons hydrogènes et le mouvement de la Glu-loop qui aboutit à la mise en continuité du tunnel avec la zone du site actif.

Des molécules d'eau peuvent alors librement circuler du solvant vers le centre actif et être stabilisées dans le canal par des résidus conservés aussi dans la perchlorate réductase (Figure 11.5). Une chaîne de molécules d'eau peut alors faire le pont entre le solvant et une position proche de la position théorique du molybdène. De plus, ce canal est suffisamment large pour pouvoir accommoder des substrats comme le nitrate et le chlorate, et dans ce cadre constituer un filtre de sélectivité pour ces substrats.

Finalement, ces molécules d'eau constituent aussi un "câble à proton" qui pourrait prendre en charge le transport des protons jusqu'au centre actif par un mécanisme de Grotthuss (CUKIERMAN, 2006).

Bibliographie

- ABRAHAM, M. J., T. MURTOLA, R. SCHULZ, S. PALL, J. C. SMITH, B. HESS et al. (2015). « Gromacs : High performance molecular simulations through multi-level parallelism from laptops to supercomputers ». In : *SoftwareX* 1-2, p. 19-25. DOI : 10.1016/j.softx.2015.06.001.
- ACEHAN, D., A. MALHOTRA, Y. XU, M. REN, D. L. STOKES et M. SCHLAME (mai 2011). « Cardiolipin affects the supramolecular organization of ATP synthase in mitochondria. » In : *Biophysical journal* 100.9, p. 2184-92. DOI : 10.1016/j.bpj.2011.03.031.
- ALTERI, C. J. et H. L. MOBLEY (avr. 2016). « The Versatile Type VI Secretion System ». In : *Virulence Mechanisms of Bacterial Pathogens, Fifth Edition*. T. 4. 2. American Society of Microbiology, p. 337-356. DOI : 10.1128/microbiolspec.VMBF-0026-2015.
- ANDO, T., T. UCHIHASHI et S. SCHEURING (mar. 2014). « Filming Biomolecular Processes by High-Speed Atomic Force Microscopy ». In : *Chemical Reviews* 114.6, p. 3120-3188. DOI : 10.1021/cr4003837.
- ANISHKIN, A. et C. KUNG (mar. 2013). « Stiffened lipid platforms at molecular force foci. » In : *Proceedings of the National Academy of Sciences of the United States of America* 110.13, p. 4886-92. DOI : 10.1073/pnas.1302018110.
- ARKHIPOV, A., Y. YIN et K. SCHULTEN (sept. 2008). « Four-scale description of membrane sculpting by BAR domains. » In : *Biophysical journal* 95.6, p. 2806-21. DOI : 10.1529/biophysj.108.132563.
- BADER, A. N., E. G. HOFMAN, J. VOORTMAN, P. M. van BERGEN EN HENEGOUWEN et H. C. GERRITSEN (nov. 2009). « Homo-FRET Imaging Enables Quantification of Protein Cluster Sizes with Subcellular Resolution ». In : *Biophysical Journal* 97.9, p. 2613-2622. DOI : 10.1016/J.BPJ.2009.07.059.
- BAYLY, C. I., K. M. MERZ, D. M. FERGUSON, W. D. CORNELL, T. FOX, J. W. CALDWELL et al. (mai 1995). « A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules ». In : *Journal of the American Chemical Society* 117.19, p. 5179-5197. DOI : 10.1021/ja00124a002.
- BEREAU, T. et K. KREMER (juin 2015). « Automated Parametrization of the Coarse-Grained Martini Force Field for Small Organic Molecules ». In : *Journal of Chemical Theory and Computation* 11.6, p. 2783-2791. DOI : 10.1021/acs.jctc.5b00056.

- BERENDSEN, H. J. C., D. van der SPOEL et R. van DRUNEN (1995). « GRO-MACS : A message-passing parallel molecular dynamics implementation ». In : *Computer Physics Communications* 91.1-3, p. 43-56. DOI : 10.1016/0010-4655(95)00042-E.
- BERENDSEN, H. J. C., J. P. M. POSTMA, W. F. van GUNSTEREN, A. DiNOLA et J. R. HAAK (oct. 1984). « Molecular dynamics with coupling to an external bath ». In : *The Journal of Chemical Physics* 81.8, p. 3684-3690. DOI : 10.1063/1.448118.
- BERTERO, M. G., R. A. ROTHERY, N. BOROUHAND, M. PALAK, F. BLASCO, N. GINET et al. (avr. 2005). « Structural and biochemical characterization of a quinol binding site of Escherichia coli nitrate reductase A. » In : *The Journal of biological chemistry* 280.15, p. 14836-43. DOI : 10.1074/jbc.M410457200.
- BERTERO, M. G., R. A. ROTHERY, M. PALAK, C. HOU, D. LIM, F. BLASCO et al. (sept. 2003). « Insights into the respiratory electron transfer pathway from the structure of nitrate reductase A ». In : *Nature Structural & Molecular Biology* 10.9, p. 681-687. DOI : 10.1038/nsb969.
- BINGLE, L. E., C. M. BAILEY et M. J. PALLAN (fév. 2008). « Type VI secretion : a beginner's guide ». In : *Current Opinion in Microbiology* 11.1, p. 3-8. DOI : 10.1016/J.MIB.2008.01.006.
- BUSSI, G., D. DONADIO et M. PARRINELLO (mar. 2008). « Canonical sampling through velocity-rescaling ». In : DOI : 10.1063/1.2408420.
- CASCALES, E. et C. CAMBILLAU (avr. 2012). « Structural biology of type VI secretion systems. » In : *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 367.1592, p. 1102-11. DOI : 10.1098/rstb.2011.0209.
- CASUSO, I., J. KHAO, M. CHAMI, P. PAUL-GILLOTEAUX, M. HUSAIN, J.-P. DUNEAU et al. (août 2012). « Characterization of the motion of membrane proteins using high-speed atomic force microscopy ». In : *Nature Nanotechnology* 7.8, p. 525-529. DOI : 10.1038/nnano.2012.109.
- CHADDA, R. et J. ROBERTSON (jan. 2016). « Measuring Membrane Protein Dimerization Equilibrium in Lipid Bilayers by Single-Molecule Fluorescence Microscopy ». In : *Methods in Enzymology* 581, p. 53-82. DOI : 10.1016/BS.MIE.2016.08.025.
- CHAMBERS, R. (nov. 1922). « A MICRO INJECTION STUDY ON THE PERMEABILITY OF THE STARFISH EGG. » In : *The Journal of general physiology* 5.2, p. 189-93.
- CHAVENT, M., A. L. DUNCAN et M. S. SANSOM (oct. 2016). « Molecular dynamics simulations of membrane proteins and their interactions : from nanoscale to mesoscale ». In : *Current Opinion in Structural Biology* 40, p. 8-16. DOI : 10.1016/J.SBI.2016.06.007.
- CHAVENT, M., T. REDDY, J. GOOSE, A. C. E. DAHL, J. E. STONE, B. JOBARD et al. (2014). « Methodologies for the analysis of instantaneous lipid diffusion in md simulations of large membrane systems ». In : *Faraday Discuss.* 169, p. 455-475. DOI : 10.1039/C3FD00145H.

- CHEN, L., M. MERZLYAKOV, T. COHEN, Y. SHAI et K. HRISTOVA (juin 2009). « Energetics of ErbB1 Transmembrane Domain Dimerization in Lipid Bilayers ». In : *Biophysical Journal* 96.11, p. 4622-4630. DOI : 10.1016/J.BPJ.2009.03.004.
- CHEN, L., L. NOVICKY, M. MERZLYAKOV, T. HRISTOV et K. HRISTOVA (mar. 2010). « Measuring the Energetics of Membrane Protein Dimerization in Mammalian Membranes ». In : *Journal of the American Chemical Society* 132.10, p. 3628-3635. DOI : 10.1021/ja910692u.
- CHIPOT, C. (2008). « Free Energy Calculations Applied to Membrane Proteins ». In : Humana Press, p. 121-144. DOI : 10.1007/978-1-59745-177-2_7.
- CHRISTIE, D. A., C. D. LEMKE, I. M. ELIAS, L. A. CHAU, M. G. KIRCHHOFF, B. LI et al. (sept. 2011). « Stomatin-like protein 2 binds cardiolipin and regulates mitochondrial biogenesis and function. » In : *Molecular and cellular biology* 31.18, p. 3845-56. DOI : 10.1128/MCB.05393-11.
- CLAYPOOL, S. M. (oct. 2009). « Cardiolipin, a critical determinant of mitochondrial carrier protein assembly and function ». In : *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1788.10, p. 2059-2068. DOI : 10.1016/J.BBAMEM.2009.04.020.
- COMER, J., J. C. GUMBART, J. H ?NIN, T. LELI ?VRE, A. POHORILLE et C. CHIPOT (jan. 2015). « The Adaptive Biasing Force Method : Everything You Always Wanted To Know but Were Afraid To Ask ». In : *The Journal of Physical Chemistry B* 119.3, p. 1129-1151. DOI : 10.1021/jp506633n.
- CORRADI, V., E. MENDEZ-VILLUENDAS, H. I. INGÓLFSSON, R.-X. GU, I. SIUDA, M. N. MELO et al. (juin 2018). « Lipid-Protein Interactions Are Unique Fingerprints for Membrane Proteins ». In : *ACS Central Science* 4.6, p. 709-717. DOI : 10.1021/acscentsci.8b00143.
- CUKIERMAN, S. (août 2006). « Et tu, Grotthuss ! and other unfinished stories ». In : *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1757.8, p. 876-885. DOI : 10.1016/J.BBABI0.2005.12.001.
- DAHLBERG, M. (2007). « Polymorphic phase behavior of cardiolipin derivatives studied by coarse-grained molecular dynamics ». In : *Journal of Physical Chemistry B* 111.25, p. 7194-7200. DOI : 10.1021/jp071954f.
- DANIELLI, J. F. et H. DAVSON (fév. 1935). « A contribution to the theory of permeability of thin films ». In : *Journal of Cellular and Comparative Physiology* 5.4, p. 495-508. DOI : 10.1002/jcp.1030050409.
- DARVE, E. et A. POHORILLE (2001). « Calculating free energies using average force ». In : *Journal of Chemical Physics* 115.20, p. 9169-9183. DOI : 10.1063/1.1410978.
- DE JONG, D. H., S. BAOUKINA, H. I. INGÓLFSSON et S. J. MARRINK (2016). « Martini straight : Boosting performance using a shorter cutoff and GPUs ». In : *Computer Physics Communications* 199, p. 1-7. DOI : 10.1016/j.cpc.2015.09.014.
- DETMER, W. (1895). *Das pflanzenphysiologische Praktikum : Anleitung zu pflanzenphysiologischen Untersuchungen für Studierende und Lehrer der Naturwissenschaften sowie der Medicin, Land- und Forstwirth-*

- schaft / von W. Detmer*. Jena : G. Fischer, DOI : 10.5962/bhl.title.16662.
- DOMAŃSKI, J., G. HEDGER, R. B. BEST, P. J. STANSFELD et M. S. P. SANSOM (avr. 2017). « Convergence and Sampling in Determining Free Energy Landscapes for Membrane Protein Association ». In : *The Journal of Physical Chemistry B* 121.15, p. 3364-3375. DOI : 10.1021/acs.jpcc.6b08445.
- DÖRR, J. M., M. C. KOORENGEVEL, M. SCHÄFER, A. V. PROKOFYEV, S. SCHEIDELAAR, E. A. W. van der CRUIJSEN et al. (déc. 2014). « Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K⁺ channel : The power of native nanodiscs ». In : *Proceedings of the National Academy of Sciences* 111.52, p. 18607-18612. DOI : 10.1073/pnas.1416205112.
- DUHEM, P. (1954). *The Aim and Structure of Physical Theory*. Princeton University Press.
- DUNEAU, J. P. et J. N. STURGIS (déc. 2013). « Lateral organization of biological membranes role of long-range interactions ». In : *European Biophysics Journal* 42.11-12, p. 843-850. DOI : 10.1007/s00249-013-0933-x.
- DUNEAU, J.-P., J. KHAO et J. N. STURGIS (jan. 2017). « Lipid perturbation by membrane proteins and the lipophobic effect ». In : *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1859.1, p. 126-134. DOI : 10.1016/J.BBAMEM.2016.10.014.
- DUNTON, T. A., J. E. GOOSE, D. J. GAVAGHAN, M. S. P. SANSOM et J. M. OSBORNE (jan. 2014). « The Free Energy Landscape of Dimerization of a Membrane Protein, NanC ». In : *PLoS Computational Biology* 10.1. Sous la dir. de H. GRUBMÜLLER, e1003417. DOI : 10.1371/journal.pcbi.1003417.
- DURAND, E., V. S. NGUYEN, A. ZOUEB, L. LOGGER, G. PÉHAU-ARNAUDET, M.-S. ASCHTGEN et al. (juil. 2015). « Biogenesis and structure of a type VI secretion membrane core complex ». In : *Nature* 523.7562, p. 555-560. DOI : 10.1038/nature14667.
- DUTROCHET, H. (1824). *Recherches anatomiques et physiologiques sur la structure intime des animaux et des végétaux, sur leur mobilité* /. Paris : Baillière, DOI : 10.5962/bhl.title.117365.
- DUVAL, S., J. M. SANTINI, D. LEMAIRE, F. CHASPOUL, M. J. RUSSELL, S. GRIMALDI et al. (sept. 2016). « The H-bond network surrounding the pyranopterins modulates redox cooperativity in the molybdenum-bisPGD cofactor in arsenite oxidase ». In : *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1857.9, p. 1353-1362. DOI : 10.1016/J.BBABI0.2016.05.003.
- EBIE, A. Z. et K. G. FLEMING (fév. 2007). « Dimerization of the Erythropoietin Receptor Transmembrane Domain in Micelles ». In : *Journal of Molecular Biology* 366.2, p. 517-524. DOI : 10.1016/J.JMB.2006.11.035.
- FIORIN, G., M. L. KLEIN et J. HÉNIN (2013). « Using collective variables to drive molecular dynamics simulations ». In : *Molecular Physics* 111.22-23, p. 3345-3362. DOI : 10.1080/00268976.2013.813594.
- FISHER, L. E., D. M. ENGELMAN et J. N. STURGIS (oct. 1999). « Detergents modulate dimerization, but not helicity, of the glycophorin A transmem-

- brane domain ». In : *Journal of Molecular Biology* 293.3, p. 639-651. DOI : 10.1006/JMBI.1999.3126.
- FISHER, L. E., D. M. ENGELMAN et J. N. STURGIS (nov. 2003). « Effect of Detergents on the Association of the Glycophorin A Transmembrane Helix ». In : *Biophysical Journal* 85.5, p. 3097-3105. DOI : 10.1016/S0006-3495(03)74728-6.
- FISHER, S. J., J. WILKINSON, R. H. HENCHMAN et J. R. HELLIWELL (oct. 2009). « An evaluation review of the prediction of protonation states in proteins versus crystallographic experiment ». In : *Crystallography Reviews* 15.4, p. 231-259. DOI : 10.1080/08893110903213700.
- FLEMING, K. G., C.-C. REN, A. K. DOURA, M. E. EISLEY, F. J. KOBUS et A. M. STANLEY (mar. 2004). « Thermodynamics of glycophorin A transmembrane helix dimerization in C14 betaine micelles ». In : *Biophysical Chemistry* 108.1-3, p. 43-49. DOI : 10.1016/J.BPC.2003.10.008.
- FRAUENFELD, J., J. GUMBART, E. O. v. d. SLUIS, S. FUNES, M. GARTMANN, B. BEATRIX et al. (mai 2011). « Cryo-EM structure of the ribosome-SecYE complex in the membrane environment ». In : *Nature Structural & Molecular Biology* 18.5, p. 614-621. DOI : 10.1038/nsmb.2026.
- FRESE, R. N., J. C. PÂMIES, J. D. OLSEN, S. BAHATYROVA, C. D. van der WEIJ-DE WIT, T. J. AARTSMA et al. (jan. 2008). « Protein shape and crowding drive domain formation and curvature in biological membranes. » In : *Biophysical journal* 94.2, p. 640-7. DOI : 10.1529/biophysj.107.116913.
- FRYE, L. D. et M. EDIDIN (1970). « The Rapid Intermixing of Cell Surface Antigens After Formation of Mouse-Human Heterokaryons ». In : *Journal of Cell Science* 7.2.
- GAO, S., M. PAN, Y. ZHENG, Y. HUANG, Q. ZHENG, D. SUN et al. (nov. 2016). « Monomer/Oligomer Quasi-Racemic Protein Crystallography ». In : *Journal of the American Chemical Society* 138.43, p. 14497-14502. DOI : 10.1021/jacs.6b09545.
- GONEN, T., Y. CHENG, P. SLIZ, Y. HIROAKI, Y. FUJIYOSHI, S. C. HARRISON et al. (déc. 2005). « Lipid-protein interactions in double-layered two-dimensional AQP0 crystals ». In : *Nature* 438.7068, p. 633-638. DOI : 10.1038/nature04321.
- GOOSE, J. E. et M. S. P. SANSOM (avr. 2013). « Reduced lateral mobility of lipids and proteins in crowded membranes. » In : *PLoS computational biology* 9.4, e1003033. DOI : 10.1371/journal.pcbi.1003033.
- GORTER, E. et F. GRENDL (mar. 1925). « ON BIMOLECULAR LAYERS OF LIPOIDS ON THE CHROMOCYTES OF THE BLOOD. » In : *The Journal of experimental medicine* 41.4, p. 439-43.
- GRAHAM, T. (jan. 1861). « Liquid Diffusion Applied to Analysis ». In : *Philosophical Transactions of the Royal Society of London* 151, p. 183-224. DOI : 10.1098/rstl.1861.0011.
- GRIMALDI, S., F. BIASO, B. BURLAT et B. GUIGLIARELLI (sept. 2016). « CHAPTER 3. Electron Paramagnetic Resonance Studies of Molybdenum Enzymes ». In : *Molybdenum and Tungsten Enzymes*. Cambridge : Royal Society of Chemistry, p. 68-120. DOI : 10.1039/9781782628842-00068.

- GRIMALDI, S., B. SCHOEPP-COTHENET, P. CECCALDI, B. GUIGLIARELLI et A. MAGALON (août 2013). « The prokaryotic Mo/W-bisPGD enzymes family : A catalytic workhorse in bioenergetic ». In : *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1827.8-9, p. 1048-1085. DOI : 10.1016/J.BBABI0.2013.01.011.
- GUÉX, N. et M. C. PEITSCH (jan. 1997). « SWISS-MODEL and the Swiss-Pdb Viewer : An environment for comparative protein modeling ». In : *Electrophoresis* 18.15, p. 2714-2723. DOI : 10.1002/elps.1150181505.
- HACHANI, A., T. E. WOOD et A. FILLOUX (fév. 2016). « Type VI secretion and anti-host effectors ». In : *Current Opinion in Microbiology* 29, p. 81-93. DOI : 10.1016/J.MIB.2015.11.006.
- HEIJNE, G. von (mai 1992). « Membrane protein structure prediction : Hydrophobicity analysis and the positive-inside rule ». In : *Journal of Molecular Biology* 225.2, p. 487-494. DOI : 10.1016/0022-2836(92)90934-C.
- HELMS, V. (déc. 2002). « Attraction within the membrane. Forces behind trans-membrane protein folding and supramolecular complex assembly. » In : *EMBO reports* 3.12, p. 1133-8. DOI : 10.1093/embo-reports/kvf245.
- HESS, B., C. KUTZNER, D. VAN DER SPOEL et E. LINDAHL (2008). « GRG-MACS 4 : Algorithms for highly efficient, load-balanced, and scalable molecular simulation ». In : *Journal of Chemical Theory and Computation* 4.3, p. 435-447. DOI : 10.1021/ct700301q.
- HILLIER, J. et J. F. HOFFMAN (oct. 1953). « On the ultrastructure of the plasma membrane as determined by the electron microscope ». In : *Journal of Cellular and Comparative Physiology* 42.2, p. 203-247. DOI : 10.1002/jcp.1030420205.
- HOOD, R. D., P. SINGH, F. HSU, T. GÜVENER, M. A. CARL, R. R. S. TRINIDAD et al. (jan. 2010). « A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. » In : *Cell host & microbe* 7.1, p. 25-37. DOI : 10.1016/j.chom.2009.12.007.
- HOOKE, R. (1665). *Micrographia, or, Some physiological descriptions of minute bodies made by magnifying glasses :with observations and inquiries thereupon /by R. Hooke ...* London : Printed by Jo. Martyn et Ja. Allestry, printers to the Royal Society ... , DOI : 10.5962/bhl.title.904.
- HSIN, J., J. GUMBART, L. G. TRABUCO, E. VILLA, P. QIAN, C. N. HUNTER et al. (juil. 2009). « Protein-induced membrane curvature investigated through molecular dynamics flexible fitting. » In : *Biophysical journal* 97.1, p. 321-9. DOI : 10.1016/j.bpj.2009.04.031.
- HUANG, J., S. RAUSCHER, G. NAWROCKI, T. RAN, M. FEIG, B. L. de GROOT et al. (jan. 2017). « CHARMM36m : an improved force field for folded and intrinsically disordered proteins ». In : *Nature Methods* 14.1, p. 71-73. DOI : 10.1038/nmeth.4067.
- HUME, D. (1748). *Enquete sur l'entendement humain*. classiques des sciences sociales. Chicoutimi : J.-M. Tremblay. DOI : 10.1522/cla.hud.enq.
- HUMPHREY, W., A. DALKE et K. SCHULTEN (fév. 1996). « VMD : Visual molecular dynamics ». In : *Journal of Molecular Graphics* 14.1, p. 33-38. DOI : 10.1016/0263-7855(96)00018-5.

- HUNTER, J. D. (2007). « Matplotlib : A 2D Graphics Environment ». In : *Computing in Science & Engineering* 9.3, p. 90-95. DOI : 10.1109/MCSE.2007.55.
- HYDE, H. C., W. SANDTNER, E. VARGAS, A. T. DAGCAN, J. L. ROBERTSON, B. ROUX et al. (oct. 2012). « Nano-Positioning System for Structural Analysis of Functional Homomeric Proteins in Multiple Conformations ». In : *Structure* 20.10, p. 1629-1640. DOI : 10.1016/J.STR.2012.08.022.
- JANOSI, L., A. PRAKASH et M. DOXASTAKIS (juil. 2010). « Lipid-Modulated Sequence-Specific Association of Glycophorin A in Membranes ». In : *Biophysical Journal* 99.1, p. 284-292. DOI : 10.1016/J.BPJ.2010.04.005.
- JAVANAINEN, M., H. MARTINEZ-SEARA et I. VATTULAINEN (nov. 2017). « Excessive aggregation of membrane proteins in the Martini model ». In : *PLOS ONE* 12.11. Sous la dir. d'E. PAPALEO, e0187936. DOI : 10.1371/journal.pone.0187936.
- JIANG, W., J. C. PHILLIPS, L. HUANG, M. FAJER, Y. MENG, J. C. GUMBART et al. (2014). « Generalized scalable multiple copy algorithms for molecular dynamics simulations in NAMD ». In : *Computer Physics Communications* 185.3, p. 908-916. DOI : 10.1016/j.cpc.2013.12.014.
- JONES, J. E. (oct. 1924). « On the Determination of Molecular Fields. II. From the Equation of State of a Gas ». In : *Proceedings of the Royal Society A : Mathematical, Physical and Engineering Sciences* 106.738, p. 463-477. DOI : 10.1098/rspa.1924.0082.
- JORMAKKA, M., D. RICHARDSON, B. BYRNE et S. IWATA (mar. 2004). « Architecture of NarGH Reveals a Structural Classification of Mo-bisMGD Enzymes ». In : *Structure* 12.1, p. 95-104. DOI : 10.1016/J.STR.2003.11.020.
- KABSCH, W. et C. SANDER (déc. 1983). « Dictionary of protein secondary structure : Pattern recognition of hydrogen-bonded and geometrical features ». In : *Biopolymers* 22.12, p. 2577-2637. DOI : 10.1002/bip.360221211.
- KAUFFMAN, S. A. (1971). « Articulation of Parts Explanation in Biology and the Rational Search for them ». In : p. 257-272. DOI : 10.1007/978-94-010-3142-4_18.
- KILLIAN, J. (nov. 1998). « Hydrophobic mismatch between proteins and lipids in membranes ». In : *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* 1376.3, p. 401-416. DOI : 10.1016/S0304-4157(98)00017-3.
- KITO, M., S. AIBARA, M. KATO et T. HATA (mar. 1972). « Differences in fatty acid composition among phosphatidylethanolamine, phosphatidylglycerol and cardiolipin of Escherichia coli ». In : *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 260.3, p. 475-478. DOI : 10.1016/0005-2760(72)90062-8.
- KLUYVER, T., B. RAGAN-KELLEY, F. PÉREZ, B. GRANGER, M. BUSSONNIER, J. FREDERIC et al. (2016). *Jupyter Notebooks—a publishing format for reproducible computational workflows*, p. 87-90. DOI : 10.3233/978-1-61499-649-1-87.

- KOLDSO, H. et M. S. P. SANSOM (nov. 2015). « Organization and Dynamics of Receptor Proteins in a Plasma Membrane ». In : *Journal of the American Chemical Society* 137.46, p. 14694-14704. DOI : 10.1021/jacs.5b08048.
- KUHN, T. S. et I. HACKING (2012). *The Structure of Scientific Revolutions : 50th Anniversary Edition*. University of Chicago Press.
- KUMAR, S., J. M. ROSENBERG, D. BOUZIDA, R. H. SWENDSEN et P. A. KOLLMAN (oct. 1992). « THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method ». In : *Journal of Computational Chemistry* 13.8, p. 1011-1021. DOI : 10.1002/jcc.540130812.
- KUZNETSOV, A. S., P. E. VOLYNSKY et R. G. EFREMOV (2015). « Role of the Lipid Environment in the Dimerization of Transmembrane Domains of Glycophorin A. » In : *Acta naturae* 7.4, p. 122-7.
- LAGANOWSKY, A., E. READING, T. M. ALLISON, M. B. ULMSCHNEIDER, M. T. DEGIACOMI, A. J. BALDWIN et al. (2014). « Membrane proteins bind lipids selectively to modulate their structure and function. » In : *Nature* 510.7503, p. 172-5. DOI : 10.1038/nature13419.
- LANGMUIR, I. (sept. 1917). « THE CONSTITUTION AND FUNDAMENTAL PROPERTIES OF SOLIDS AND LIQUIDS. II. LIQUIDS. 1 ». In : *Journal of the American Chemical Society* 39.9, p. 1848-1906. DOI : 10.1021/ja02254a006.
- LATOUR, B. (1987). *Science in Action*. T. 6. 1, p. 274. DOI : 10.1017/CB09781107415324.004.
- LEE, A. G. (juin 2011). « Lipid-protein interactions. » In : *Biochemical Society transactions* 39.3, p. 761-6. DOI : 10.1042/BST0390761.
- LELIÈVRE, T., M. ROUSSET et G. STOLTZ (2008). « Long-time convergence of an adaptive biasing force method ». In : *Nonlinearity* 21.6, p. 1155-1181. DOI : 10.1088/0951-7715/21/6/001.
- LI, P.-C., N. MIYASHITA, W. IM, S. ISHIDO et Y. SUGITA (fév. 2014). « Multidimensional umbrella sampling and replica-exchange molecular dynamics simulations for structure prediction of transmembrane helix dimers ». In : *Journal of Computational Chemistry* 35.4, p. 300-308. DOI : 10.1002/jcc.23494.
- LOMIZE, M. A., I. D. POGOZHEVA, H. JOO, H. I. MOSBERG et A. L. LOMIZE (jan. 2012). « OPM database and PPM web server : resources for positioning of proteins in membranes ». In : *Nucleic Acids Research* 40.D1, p. D370-D376. DOI : 10.1093/nar/gkr703.
- MACDONALD, P. M. et J. SEELIG (avr. 1988). « Dynamic properties of gramicidin A in phospholipid membranes ». In : *Biochemistry* 27.7, p. 2357-2364. DOI : 10.1021/bi00407a017.
- MADDY, A. H., B. R. MALCOLM et D. G. CORNWELL (déc. 1965). « Protein conformations in the plasma membrane. » In : *Science (New York, N.Y.)* 150.3703, p. 1616-8. DOI : 10.1126/science.150.3703.1616.
- MAIER, J. A., C. MARTINEZ, K. KASAVAJHALA, L. WICKSTROM, K. E. HAUSER et C. SIMMERLING (août 2015). « ff14SB : Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB ». In : *Journal of*

- Chemical Theory and Computation* 11.8, p. 3696-3713. DOI : 10.1021/acs.jctc.5b00255.
- MARRINK, S. J., H. J. RISSELADA, S. YEFIMOV, D. P. TIELEMAN et A. H. DE VRIES (2007). « The MARTINI force field : Coarse grained model for biomolecular simulations ». In : *Journal of Physical Chemistry B* 111.27, p. 7812-7824. DOI : 10.1021/jp071097f.
- MEAGHER, K. L., L. T. REDMAN et H. A. CARLSON (juil. 2003). « Development of polyphosphate parameters for use with the AMBER force field ». In : *Journal of Computational Chemistry* 24.9, p. 1016-1025. DOI : 10.1002/jcc.10262.
- METZLER, R., J.-H. JEON et A. CHERSTVY (oct. 2016). « Non-Brownian diffusion in lipid membranes : Experiments and simulations ». In : *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1858.10, p. 2451-2467. DOI : 10.1016/J.BBAMEM.2016.01.022.
- MICHAUD-AGRAWAL, N., E. J. DENNING, T. B. WOOLF et O. BECKSTEIN (juil. 2011). « MDAnalysis : A toolkit for the analysis of molecular dynamics simulations ». In : *Journal of Computational Chemistry* 32.10, p. 2319-2327. DOI : 10.1002/jcc.21787.
- MONTICELLI, L., S. K. KANDASAMY, X. PERIOLE, R. G. LARSON, D. P. TIELEMAN et S. J. MARRINK (2008). « The MARTINI coarse-grained force field : Extension to proteins ». In : *Journal of Chemical Theory and Computation* 4.5, p. 819-834. DOI : 10.1021/ct700324x.
- MOREIN, S., A. ANDERSSON, L. RILFORS et G. LINDBLOM (mar. 1996). « Wild-type Escherichia coli cells regulate the membrane lipid composition in a "window" between gel and non-lamellar structures. » In : *The Journal of biological chemistry* 271.12, p. 6801-9. DOI : 10.1074/JBC.271.12.6801.
- MUNGUIRA, I., I. CASUSO, H. TAKAHASHI, F. RICO, A. MIYAGI, M. CHAMI et al. (fév. 2016). « Glasslike Membrane Protein Diffusion in a Crowded Membrane ». In : *ACS Nano* 10.2, p. 2584-2590. DOI : 10.1021/acsnano.5b07595.
- NICOLSON, G. L. (avr. 1976). « Transmembrane control of the receptors on normal and tumor cells : I. Cytoplasmic influence over cell surface components ». In : *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* 457.1, p. 57-108. DOI : 10.1016/0304-4157(76)90014-9.
- NITSCHKE, W. et M. J. RUSSELL (nov. 2009). « Hydrothermal Focusing of Chemical and Chemiosmotic Energy, Supported by Delivery of Catalytic Fe, Ni, Mo/W, Co, S and Se, Forced Life to Emerge ». In : *Journal of Molecular Evolution* 69.5, p. 481-496. DOI : 10.1007/s00239-009-9289-3.
- OHVO-REKILÄ, H., B. RAMSTEDT, P. LEPPIMÄKI et J. PETER SLOTTÉ (jan. 2002). « Cholesterol interactions with phospholipids in membranes ». In : *Progress in Lipid Research* 41.1, p. 66-97. DOI : 10.1016/S0163-7827(01)00020-0.
- OLSSON, M. H. M., C. R. SØNDERGAARD, M. ROSTKOWSKI et J. H. JENSEN (fév. 2011). « PROPKA3 : Consistent Treatment of Internal and Surface

- Residues in Empirical pKa Predictions ». In : *Journal of Chemical Theory and Computation* 7.2, p. 525-537. DOI : 10.1021/ct100578z.
- OVERTON, E. (1895). « Ueber die osmotischen Eigenschaften der lebenden Pflanzen und Thierzelle ». In : *Vierteljahrschr Naturf Ges Zurich*.
- (1899). « Ueber die Allgemeinen Osmotischen Eigenschaften der Zelle ». In : *Vierteljahrschr Naturf Ges Zurich*.
- PARRINELLO, M. et A. RAHMAN (1981). « Polymorphic transitions in single crystals : A new molecular dynamics method ». In : *Journal of Applied Physics* 52.12, p. 7182-7190. DOI : 10.1063/1.328693.
- PERIOLE, X., A. M. KNEPP, T. P. SAKMAR, S. J. MARRINK et T. HUBER (juil. 2012). « Structural Determinants of the Supramolecular Organization of G Protein-Coupled Receptors in Bilayers ». In : *Journal of the American Chemical Society* 134.26, p. 10959-10965. DOI : 10.1021/ja303286e.
- PETHIG, R. et I. SCHMUESER (nov. 2012). « Marking 100 years since Rudolf Höber's discovery of the insulating envelope surrounding cells and of the beta-dispersion exhibited by tissue ». In : *Journal of Electrical Bioimpedance* 3.1, p. 74-79. DOI : 10.5617/jeb.401.
- PHILLIPS, J. C., R. BRAUN, W. WANG, J. GUMBART, E. TAJKHORSHID, E. VILLA et al. (2005). *Scalable molecular dynamics with NAMD*. DOI : 10.1002/jcc.20289.
- PRABUDIANSYAH, I., I. KUSTERS, A. CAFORIO et A. J. DRIESSEN (oct. 2015). « Characterization of the annular lipid shell of the Sec translocon ». In : *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1848.10, p. 2050-2056. DOI : 10.1016/J.BBAMEM.2015.06.024.
- PRAKASH, A., L. JANOSI et M. DOXASTAKIS (déc. 2010). « Self-Association of Models of Transmembrane Domains of ErbB Receptors in a Lipid Bilayer ». In : *Biophysical Journal* 99.11, p. 3657-3665. DOI : 10.1016/J.BPJ.2010.10.023.
- (oct. 2011). « GxxxG Motifs, Phenylalanine, and Cholesterol Guide the Self-Association of Transmembrane Domains of ErbB2 Receptors ». In : *Biophysical Journal* 101.8, p. 1949-1958. DOI : 10.1016/J.BPJ.2011.09.017.
- PRASANNA, X., P. J. PRAVEEN et D. SENGUPTA (oct. 2013). « Sequence dependent lipid-mediated effects modulate the dimerization of ErbB2 and its associative mutants ». In : *Physical Chemistry Chemical Physics* 15.43, p. 19031. DOI : 10.1039/c3cp52447g.
- QUINE, W. V. (jan. 1951). « Main Trends in Recent Philosophy : Two Dogmas of Empiricism ». In : *The Philosophical Review* 60.1, p. 20. DOI : 10.2307/2181906.
- RAMYA GAMINI, D. C. (2009). *Residue-Based Coarse Graining using MARTINI Force Field in NAMD*.
- RENDON, J., F. BIASO, P. CECCALDI, R. TOCI, F. SEDUK, A. MAGALON et al. (avr. 2017). « Elucidating the Structures of the Low- and High-pH Mo(V) Species in Respiratory Nitrate Reductase : A Combined EPR, 14,15 N HYSCORE, and DFT Study ». In : *Inorganic Chemistry* 56.8, p. 4422-4434. DOI : 10.1021/acs.inorgchem.6b03129.

- ROBERTSON, M. J., J. TIRADO-RIVES et W. L. JORGENSEN (juil. 2015). « Improved Peptide and Protein Torsional Energetics with the OPLS-AA Force Field ». In : *Journal of Chemical Theory and Computation* 11.7, p. 3499-3509. DOI : 10.1021/acs.jctc.5b00356.
- ROSSI, G., L. MONTICELLI, S. R. PUISTO, I. VATTULAINEN et T. ALA-NISSILA (jan. 2011). « Coarse-graining polymers with the MARTINI force-field : polystyrene as a benchmark case ». In : *Soft Matter* 7.2, p. 698-708. DOI : 10.1039/C0SM00481B.
- RUSS, W. P. et D. M. ENGELMAN (fév. 1999). « TOXCAT : a measure of trans-membrane helix association in a biological membrane. » In : *Proceedings of the National Academy of Sciences of the United States of America* 96.3, p. 863-8. DOI : 10.1073/PNAS.96.3.863.
- SAFFMAN, P. G. et M. DELBRÜCK (août 1975). « Brownian motion in biological membranes. » In : *Proceedings of the National Academy of Sciences of the United States of America* 72.8, p. 3111-3. DOI : 10.1073/PNAS.72.8.3111.
- SCHEIDELAAR, S., M. C. KOORENGEVEL, J. D. PARDO, J. D. MEELDIJK, E. BREUKINK et J. A. KILLIAN (2015). « Molecular Model for the solubilization of membranes into nanodisks by styrene maleic acid copolymers ». In : *Biophysical Journal* 108.2, p. 279-290. DOI : 10.1016/j.bpj.2014.11.3464.
- SCHLEIDEN, M. J., T. SCHWANN et H. SMITH (1847). *Microscopical researches into the accordance in the structure and growth of animals and plants. Translated from the German of Theodor Schwann, by Henry Smith.* London, Sydenham Society, DOI : 10.5962/bhl.title.17276.
- SCHLICK, T. (2010a). « Force Fields ». In : Springer, New York, NY, p. 265-298. DOI : 10.1007/978-1-4419-6351-2_9.
- (2010b). « Molecular Dynamics : Basics ». In : p. 425-461. DOI : 10.1007/978-1-4419-6351-2_13.
- SCHMIDT, V. et J. N. STURGIS (juin 2017). « Making Monomeric Aquaporin Z by Disrupting the Hydrophobic Tetramer Interface ». In : *ACS Omega* 2.6, p. 3017-3027. DOI : 10.1021/acsomega.7b00261.
- (mar. 2018). « Modifying styrene-maleic acid co-polymer for studying lipid nanodiscs ». In : *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1860.3, p. 777-783. DOI : 10.1016/J.BBAMEM.2017.12.012.
- SCHOEPP-COTHENET, B., R. van LIS, P. PHILIPPOT, A. MAGALON, M. J. RUSSELL et W. NITSCHKE (déc. 2012). « The ineluctable requirement for the trans-iron elements molybdenum and/or tungsten in the origin of life ». In : *Scientific Reports* 2.1, p. 263. DOI : 10.1038/srep00263.
- SCHÜTZ, G., H. SCHINDLER et T. SCHMIDT (août 1997). « Single-molecule microscopy on model membranes reveals anomalous diffusion ». In : *Biophysical Journal* 73.2, p. 1073-1080. DOI : 10.1016/S0006-3495(97)78139-6.
- SENGUPTA, D. et S. J. MARRINK (oct. 2010). « Lipid-mediated interactions tune the association of glycophorin A helix and its disruptive mutants in membranes ». In : *Physical Chemistry Chemical Physics* 12.40, p. 12987. DOI : 10.1039/c0cp00101e.

- SHEN, Z., Y. LI, A. N. GASPARI, H. ABELIOVICH et M. L. GREENBERG (fév. 2017). « Cardiolipin Regulates Mitophagy through the Protein Kinase C Pathway. » In : *The Journal of biological chemistry* 292.7, p. 2916-2923. DOI : 10.1074/jbc.M116.753574.
- SHIH, A. Y., P. L. FREDDOLINO, A. ARKHIPOV, S. G. SLIGAR et K. SCHULTEN (jan. 2008). « Chapter 11 Molecular Modeling of the Structural Properties and Formation of High-Density Lipoprotein Particles ». In : *Current Topics in Membranes* 60, p. 313-342. DOI : 10.1016/S1063-5823(08)00011-2.
- SHIH, A. Y., P. L. FREDDOLINO, S. G. SLIGAR et K. SCHULTEN (2007). « Disassembly of nanodiscs with cholate ». In : *Nano Letters* 7.6, p. 1692-1696. DOI : 10.1021/nl0706906.
- SINGER, S. J. et G. L. NICOLSON (fév. 1972). « The Fluid Mosaic Model of the Structure of Cell Membranes ». In : *Science* 175.4023, p. 720-731. DOI : 10.1126/science.175.4023.720.
- SMITH, D. M. A., Y. XIONG, T. P. STRAATSMAN, K. M. ROSSO et T. C. SQUIER (juin 2012). « Force-Field Development and Molecular Dynamics of [NiFe] Hydrogenase ». In : *Journal of Chemical Theory and Computation* 8.6, p. 2103-2114. DOI : 10.1021/ct300185u.
- SMITH, H. W. (nov. 1962). « The plasma membrane, with notes on the history of botany. » In : *Circulation* 26.5, p. 987-1012. DOI : 10.1161/01.CIR.26.5.987.
- SØNDERGAARD, C. R., M. H. M. OLSSON, M. ROSTKOWSKI et J. H. JENSEN (juil. 2011). « Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values ». In : *Journal of Chemical Theory and Computation* 7.7, p. 2284-2295. DOI : 10.1021/ct200133y.
- SWOPE, W. C., H. C. ANDERSEN, P. H. BERENS et K. R. WILSON (1982). « A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules : Application to small water clusters ». In : *The Journal of Chemical Physics* 76.1, p. 637-649. DOI : 10.1063/1.442716.
- THOMAS, D. D. (1986). « Rotational diffusion of membrane proteins ». In : *Techniques for the Analysis of Membrane Proteins*. Dordrecht : Springer Netherlands, p. 377-431. DOI : 10.1007/978-94-009-4085-7_13.
- TRAUBE, M. (1867). *Experimente zur Theorie der Zellenbildung und Endosmose*. Archiv für Anatomie Physiologie und wissenschaftliche Medizin.
- VAN DER SPOEL, D., E. LINDAHL, B. HESS, G. GROENHOF, A. E. MARK et H. J. C. BERENDSEN (2005). « GROMACS : Fast, flexible, and free ». In : *Journal of Computational Chemistry* 26.16, p. 1701-1718. DOI : 10.1002/jcc.20291.
- VERLET, L. (juil. 1967). « Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules ». In : *Physical Review* 159.1, p. 98-103. DOI : 10.1103/PhysRev.159.98.
- VINCENT, S. P. et R. C. BRAY (avr. 1978). « Electron-paramagnetic-resonance studies on nitrate reductase from Escherichia coli K12 ». In : *The Biochemical journal* 171.3, p. 639-647. DOI : 10.1042/bj1710639.

- WASSENAAR, T. A., H. I. INGÓLFSSON, R. A. BÖCKMANN, D. P. TIELEMAN et S. J. MARRINK (mai 2015). « Computational Lipidomics with insane : A Versatile Tool for Generating Custom Membranes for Molecular Simulations ». In : *Journal of Chemical Theory and Computation* 11.5, p. 2144-2155. DOI : 10.1021/acs.jctc.5b00209.
- WASSENAAR, T. A., K. PLUHACKOVA, A. MOUSSATOVA, D. SENGUPTA, S. J. MARRINK, D. P. TIELEMAN et al. (2015). « High-throughput simulations of dimer and trimer assembly of membrane proteins. The DAFT approach ». In : *Journal of Chemical Theory and Computation* 11.5, p. 2278-2291. DOI : 10.1021/ct5010092.
- WILKIE, J. S. (juin 1961). « Carl Nägeli and the fine Structure of Living Matter ». In : *Nature* 190.4782, p. 1145-1150. DOI : 10.1038/1901145a0.
- WU, E. L., X. CHENG, S. JO, H. RUI, K. C. SONG, E. M. DÁVILA-CONTRERAS et al. (oct. 2014). « CHARMM-GUI Membrane Builder toward realistic biological membrane simulations ». In : *Journal of Computational Chemistry* 35.27, p. 1997-2004. DOI : 10.1002/jcc.23702.
- WU, S.-Y., R. A. ROTHERY et J. H. WEINER (oct. 2015). « Pyranopterin Coordination Controls Molybdenum Electrochemistry in Escherichia coli Nitrate Reductase. » In : *The Journal of biological chemistry* 290.41, p. 25164-73. DOI : 10.1074/jbc.M115.665422.
- XUE, M., L. CHENG, I. FAUSTINO, W. GUO et S. J. MARRINK (août 2018). « Molecular Mechanism of Lipid Nanodisk Formation by Styrene-Maleic Acid Copolymers. » In : *Biophysical journal* 115.3, p. 494-502. DOI : 10.1016/j.bpj.2018.06.018.
- ZARDOYA, R. (juin 2005). « Phylogeny and evolution of the major intrinsic protein family ». In : *Biology of the Cell* 97.6, p. 397-414. DOI : 10.1042/BC20040134.
- ZOUED, A., Y. R. BRUNET, E. DURAND, M.-S. ASCHTGEN, L. LOGGER, B. DOUZI et al. (août 2014). « Architecture and assembly of the Type VI secretion system ». In : *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1843.8, p. 1664-1673. DOI : 10.1016/J.BBAMCR.2014.03.018.
- ZOUED, A., C. J. CASSARO, E. DURAND, B. DOUZI, A. P. ESPAÑA, C. CAMBILLAU et al. (nov. 2016). « Structure-Function Analysis of the TssL Cytoplasmic Domain Reveals a New Interaction between the Type VI Secretion Baseplate and Membrane Complexes ». In : *Journal of Molecular Biology* 428.22, p. 4413-4423. DOI : 10.1016/J.JMB.2016.08.030.
- ZOUED, A., J.-P. DUNEAU, E. DURAND, A. P. ESPAÑA, L. JOURNET, F. GUERLESQUIN et al. (mar. 2018). « Tryptophan-mediated Dimerization of the TssL Transmembrane Anchor Is Required for Type VI Secretion System Activity ». In : *Journal of Molecular Biology* 430.7, p. 987-1003. DOI : 10.1016/J.JMB.2018.02.008.

Annexes

Annexe A

Lois de Newton

Newton et le mouvement en physique

L'énonciation des lois de Newton marque un tournant dans l'histoire des sciences en étendant le principe d'inertie de Galilée. C'est à partir de Newton que l'on peut savoir comment un objet change sa vitesse et sa trajectoire quand un autre objet l'affecte. Newton a écrit trois lois dans les *Principes mathématiques de la philosophie naturelle* :

- La première loi est le principe d'inertie de Galilée. "Tout corps persévère dans l'état de repos ou de mouvement uniforme en ligne droite dans lequel il se trouve, à moins que quelque force n'agisse sur lui, et ne le contraigne à changer d'état." ou, si un objet n'est pas dérangé, il continue à avancer avec une vitesse constante avec un mouvement rectiligne et uniforme (ou ne bouge pas s'il n'était pas initialement en mouvement).
- La troisième loi est le principe d'action-réaction. "L'action est toujours égale à la réaction ; c'est-à-dire que les actions de deux corps l'un sur l'autre sont toujours égales et de sens contraires." ou, si un objet exerce une force sur un autre objet, ce dernier exerce une force égale de signe opposé sur ce premier.
- La deuxième loi est aussi appelée principe fondamental de la dynamique, et c'est celle qui est à la base du calcul du mouvement chez Newton. "Les changements qui arrivent dans le mouvement sont proportionnels à la force motrice ; et se font dans la ligne droite dans laquelle cette force a été imprimée." ou, dit autrement, le changement au cours du temps d'une quantité de mouvement est proportionnel à la force.

La deuxième loi de Newton parle de "quantité de mouvement", ou *momentum*, et est le produit de la masse et de la vitesse d'un objet. Cette loi peut s'écrire

$$\mathbf{F} = m \frac{d\mathbf{v}}{dt} = m\mathbf{a} \quad (\text{A.1})$$

où m est la masse, $\mathbf{a}$ est l'accélération, $\mathbf{v}$ est la vitesse et $\mathbf{F}$ est la quantité de mouvement (ou *momentum*).

Parler de vitesse n'est cependant pas très intuitif. On va donc formuler plus précisément comment les coordonnées x -, y - et z - d'un objet changent au cours du temps. Dans un intervalle de temps Δt , un objet va bouger d'une certaine

distance dans les trois dimensions Δx , Δy et Δz en fonction de la vitesse dans ces trois mèmes dimensions. On peut donc écrire le déplacement :

$$\begin{aligned}\Delta x &= v_x \Delta t \\ \Delta y &= v_y \Delta t \\ \Delta z &= v_z \Delta t\end{aligned}\tag{A.2}$$

Et ainsi reformuler la deuxième loi de Newton :

$$\begin{aligned}\mathbf{F}_x &= m \frac{dv_x}{dt} = m \frac{d^2 x}{dt^2} = m \mathbf{a} \\ \mathbf{F}_y &= m \frac{dv_y}{dt} = m \frac{d^2 y}{dt^2} = m \mathbf{a} \\ \mathbf{F}_z &= m \frac{dv_z}{dt} = m \frac{d^2 z}{dt^2} = m \mathbf{a}\end{aligned}\tag{A.3}$$

Par commodité, ces trois dimensions sont habituellement représentées par un vecteur unique $\mathbf{r}$ et l'on a :

$$\mathbf{F}_r = m \frac{dv_r}{dt} = m \frac{d^2 \mathbf{r}}{dt^2} = m \mathbf{a}\tag{A.4}$$

Dans la suite, seule une dimension (ou degré de liberté) sera traitée par simplicité.

Intégration des lois de Newton

Maintenant, nous pouvons intégrer cette loi pour connaître de nouvelles positions et vitesses à un temps $t + \Delta t$. La méthode d'Euler va donner, en développant la première équation de A.2 :

$$x(t + \Delta t) = x(t) + \Delta t \cdot v_x(t)\tag{A.5}$$

La méthode leapfrog, utilisée dans les schémas classiques de mécanique (aussi bien en dynamique moléculaire que dans le calcul des trajectoires des astres), utilise, pour le calcul de la position à $t + \Delta t$, comme meilleure approximation, la vitesse à $t + \frac{1}{2} \Delta t$. On a donc :

$$x(t + \Delta t) = x(t) + \Delta t \cdot v_x(t + \frac{1}{2} \Delta t)\tag{A.6}$$

Pour calculer la nouvelle position, il nous faut donc la vitesse à $t + \frac{1}{2} \Delta t$. En se référant directement à l'équation A.3, on retrouve :

$$\begin{aligned}v_x(t + \frac{1}{2} \Delta t) &= v_x(t - \frac{1}{2} \Delta t) + \Delta t \cdot a_x(t) \\ &\text{avec } a(t) = \frac{\mathbf{F}_x}{m} \\ \text{et } v_x(t_{\frac{1}{2}}) &= v_x(0) + \frac{1}{2} \Delta t \cdot a_x(0)\end{aligned}\tag{A.7}$$

En calculant les positions tous les Δt et les vitesses tous les $t + \frac{1}{2}\Delta t$, on peut ainsi calculer la trajectoire d'un objet. La première vitesse pour démarrer l'algorithme utilise de plus une demi-étape. C'est l'algorithme principal utilisé dans GROMACS et dans les simulations de ce manuscrit qui utilisent GROMACS.

De son côté, NAMD utilise l'algorithme dit de Velocity Verlet. La différence majeure est l'évaluation simultanée de v et x en évaluant entre deux étapes la vitesse à $t + \frac{1}{2}\Delta t$. Cette évaluation intermédiaire est allégée en pratique car les forces de l'étape précédente sont gardées en mémoire et peuvent être réutilisées à l'étape suivante.

$$\begin{aligned}
 v_x(t + \frac{1}{2}\Delta t) &= v_x(t) + \frac{1}{2}\Delta t \cdot \frac{F_x}{m_x}(t) \\
 x(t + \Delta t) &= x(t) + \Delta t \cdot v_x(t + \frac{1}{2}\Delta t) \\
 v_x(t + \Delta t) &= v_x(t + \frac{1}{2}\Delta t) + \frac{1}{2}\Delta t \cdot \frac{F_x}{m_x}(t + \frac{1}{2}\Delta t)
 \end{aligned} \tag{A.8}$$

Chaque méthode d'intégration est ensuite modifiée pour prendre en compte le contrôle de la température et de la pression.

Pourquoi intégrer avec des "demi-pas" ?

Intégrer avec des "demi-pas" permet une propriété importante pour nos systèmes : la réversibilité temporelle. Après n étapes de leapfrog, on peut inverser le sens de l'intégration et revenir exactement au point de départ, ce qui n'est pas le cas de l'intégration d'Euler. On dit que l'intégration d'Euler est d'ordre 1, avec une erreur proportionnelle à l'erreur effectuée en une étape ; l'intégration leapfrog est quant à elle d'ordre 2, avec une erreur qui ne dépassera jamais l'erreur effectuée en une étape, ce qui permet la réversibilité temporelle.

Annexe B

Calcul de l'angle de spin

Ce calcul reproduit le calcul de l'angle de spin effectué par le module *colvars*. Il est utile notamment lorsque ce module n'est pas utilisé. Une limitation de ce calcul est qu'il ne s'occupe que de la rotation autour d'un axe prédéfini ; cependant, la réutilisation de ce code pour s'intéresser à d'autres aspects de l'orientation entre deux objets est possible. Ce code est disponible en tant que package python, qui utilise l'API Python/C pour la partie du calcul la plus intensive, à l'adresse <https://framagit.org/msidore/spinpack>.

Le problème de la superposition

Une structure moléculaire de N particules peut être représentée comme une matrice $N \times 3$ et représentant les coordonnées x , y et z des particules. Le but de la superposition de deux structures, donc de deux matrices $N \times 3$, est d'obtenir la translation et la rotation orthogonale (ou matrice de rotation) qui minimize le carré des distances entre les deux matrices $N \times 3$.

La translation est rapidement éliminée du problème en enlevant le centroïde de chaque conformation, ce qui correspond à ramener les centroïdes des deux structures à l'origine.

Le problème de superposition d'une structure B sur une structure A revient alors à minimiser la somme des moindres carrés E par rapport à la rotation orthogonale R .

$$E = \|BR - A\|_F^2 \quad (\text{B.1})$$

où $\|X\|_F$ représente la norme euclidienne de la matrice X . En développant cette équation, il peut être montré que la minimisation de E peut être obtenue uniquement à partir du produit interne des matrices A et B (B^T est la transposée de B) :

$$M = B^T A = \begin{bmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{bmatrix} \quad (\text{B.2})$$

Où

$$S_{xy} = \sum_i^N x_{B,i} y_{A,i} \quad (\text{B.3})$$

Cette matrice est ensuite utilisée pour peupler un quaternion caractéristique de la rotation :

$$\begin{bmatrix} (S_{xx} + S_{yy} + S_{zz}) & S_{yz} - S_{zy} & S_{zx} - S_{xz} & S_{xy} - S_{yx} \\ S_{yz} - S_{zy} & (S_{xx} - S_{yy} - S_{zz}) & S_{xy} + S_{yx} & S_{zx} + S_{xz} \\ S_{zx} - S_{xz} & S_{xy} + S_{yx} & (-S_{xx} + S_{yy} - S_{zz}) & S_{yz} + S_{zy} \\ S_{xy} - S_{yx} & S_{zx} + S_{xz} & S_{yz} + S_{zy} & (-S_{xx} - S_{yy} + S_{zz}) \end{bmatrix} \quad (\text{B.4})$$

Cette matrice est ensuite diagonalisée. Le vecteur propre correspondant à la valeur propre maximale de cette matrice est ensuite un quaternion équivalent à la rotation maximale. Ce quaternion $Q(w, x, y, z)$ est ensuite décomposable en angle de tilt et de spin. Nous avons donc directement l'angle de spin :

$$angle = 2 * \pi * \arctan2(Q_z, Q_w) \quad (\text{B.5})$$