HAL
open science

Simulations de dynamique moléculaires du complexe collecteur de lumière de type 2 d'une bactérie pourpre dans différents environnements micellaires et membranaire

Esra Karakas

To cite this version:

Esra Karakas. Simulations de dynamique moléculaires du complexe collecteur de lumière de type 2 d'une bactérie pourpre dans différents environnements micellaires et membranaire. Biophysique [physics.bio-ph]. Université Paris Saclay (COmUE), 2016. Français. NNT : 2016SACLX012 . tel-01461588

HAL Id: tel-01461588

https://pastel.hal.science/tel-01461588

Submitted on 8 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2016SACLX012

THÈSE DE DOCTORAT
DE
L'UNIVERSITÉ PARIS-SACLAY
PRÉPARÉE À
L'ÉCOLE POLYTECHNIQUE

ÉCOLE DOCTORALE N° 573
Interfaces - Approches interdisciplinaires : fondements, applications et innovations

Spécialité de doctorat : Biologie

Par

Mme Esra Karakas Dagli

Simulations de dynamique moléculaire du complexe collecteur de lumière de type 2 d'une bactérie pourpre dans différents environnements micellaires et membranaire

Thèse présentée et soutenue à Gif-sur-Yvette, le 9 mai 2016:

Composition du Jury :

M. Miroux Bruno	Directeur de recherche, CNRS	Président
M. Sturgis James	Professeur, Université Aix-Marseille 2	Rapporteur
M. Baaden Marc	Directeur de recherche, CNRS	Rapporteur
M. Cogdell Richard	Professeur, Université de Glasgow	Examinateur
M. Bunker Alex	Directeur de recherche, Université d'Helsinki	Examinateur
M. Robert Bruno	Directeur de recherche, CEA Saclay	Directeur de thèse
M. Abel Stéphane	Chercheur, CEA Saclay	Co-directeur de thèse

À mes chers parents et grands-parents,

Remerciements

Je tiens tout d'abord à remercier chaleureusement les Drs. Bruno Robert et Stéphane Abel de m'avoir accueillie dans leur laboratoire (LBSM) et pour m'avoir encadrée pendant ces trois ans et demi de thèse. Ces années ont été très enrichissantes à la fois professionnellement et humainement. Je remercie aussi les personnes de la Maison de la Simulation et tout particulièrement Edouard Audit qui ont permis l'existence de ce travail. Je souhaite remercier le Dr. Massimo Marchi de m'avoir accueillie dans son équipe de modélisateurs et pour ses précieuses aides et connaissances qu'il m'a apportée. Je remercie le Dr. Stéphane Abel avec qui j'ai énormément appris et je lui en suis reconnaissante.

Je remercie le Dr. François-Yves Dupradeau de l'université d'Amiens pour son aide dans le calcul des charges atomiques RESP. Je remercie le Dr. Javier Pérez et le Dr. Andrew Gall qui m'ont grandement aidée pour les expériences de SAXS au synchrotron SOLEIL. Au LLB, je remercie le Dr. Sophie Combet-Jeancenel pour son aide dans le traitement des données de SAXS.

J'exprime toute ma gratitude aux membres du jury M. le professeur James Sturgis et M. Marc Baaden qui ont accepté d'être les rapporteurs de ce travail, ainsi que M. le professeur Richard Cogdell et M. Alex Bunker de participer à mon jury en tant qu'examinateurs et enfin, M. Bruno Miroux pour avoir accepté d'être le président.

À la Maison de la Simulation, je remercie Mme Valérie Belle pour ses nombreuses aides avec l'administration du CEA, les permanents (Julien Derouillat, Pierre Kestener, Martial Mancip, et Julien Bigot) et les non-permanents (Jeaniffer, Sophie, Malyka, Karen, Fabien, Ralitsa).

À l'iBiTec-S, dans l'administration, je souhaite remercier Sophie, Karine, Céline et Isabelle pour leurs aides. Dans les équipes de recherche, je remercie les Drs. Maïté Paternostre, Stéphane Bressanelli, Yves Boulard, François André, Annie Barrand-Frelet, Stéphane Orlowski, Ghada Ajlani, Francis Haraux et tous les autres (Sana, Céline, Thanh, Liz, Adrien, Adjélé, Nahuel, Cyntia, Thibault, Pierre, Marine et Xujun) pour leur bonne humeur et les discussions qu'on a eues. Je remercie toutes les personnes que j'ai peut-être oubliées de mentionner ici et avec qui j'ai eu des contacts pendant ces trois ans et demi.

Enfin, je remercie toute ma famille, belle-famille et mes ami(e)s pour m'avoir encouragée et soutenue pendant ces années de thèse. Je souhaite remercier tout particulièrement mon mari, Eren, pour son aide précieuse, et pour avoir été très patient et compréhensif.

Table des matières

Introduction générale **2**

Chapitre 1 : Le complexe collecteur de lumière de type 2 des bactéries pourpres **6**

I. Introduction 6

II. Le photosystème chez les bactéries pourpres 8

1. Les phospholipides membranaires 9

2. Le complexe LH2 10

3. Le complexe CR-LH1 12

III. Les variants naturels des complexes antennaires périphériques 14

IV. Les interactions bactériochlorophylles-a/peptides et entre bactériochlorophylles-a . 15

V. L'absorption des pigments 16

VI. Transfert d'énergie dans les complexes LH 17

VII. Expériences de spectroscopie à molécule unique (*Single-molecule*) 20

VIII. Conclusion du chapitre 22

Chapitre 2 : La Dynamique Moléculaire **24**

I. Introduction 24

1. Bref historique 24

2. Les simulations moléculaires 26

II. Représentations des interactions entre atomes 27

1. La fonction d'énergie potentielle ou champ de force 27

2. Les interactions intramoléculaires ou interactions liantes 28

3. Les interactions intermoléculaires ou interactions non liantes 30

4. Le code de simulation moléculaire GROMACS 33

III. Les techniques de simulation en dynamique moléculaire 33

1. Les conditions périodiques aux bords 33

2. Calculs des interactions électrostatiques 35

3. Les ensembles thermodynamiques et moyennes statistiques 37

4. Simulation à pression et température constantes 39

5. L'intégration des équations du mouvement 41

6. Le volume de Voronoï 42

IV. Le calcul des charges partielles atomiques 43

V. Mise au point de nos simulations en dynamique moléculaire 47

Chapitre 3 : Simulations des environnements hydrophobes biomimétiques 50

I. Les micelles de LDAO 50

1. Introduction à l'article 50

2. Article 50

3. Les principales conclusions de l'article 94

Chapitre 4 : Simulations du complexe LH2 dans les environnements biomimétiques 96

I. Construction des systèmes 96

1. Champs de force utilisés 96

2. Simulations des différents systèmes 97

3. Le complexe LH2 dans une micelle de LDAO 97

4. Le complexe LH2 dans une micelle de β octyle glucoside 100

5. Validations expérimentales des données des nombres d'agrégations de LDAO et bOG par LH2 101

6. Le complexe LH2 dans une membrane de POPC 105

II. Résultats des simulations 107

1. Caractéristiques globales des différents systèmes 107

1.1. Les diamètres du complexe LH2 dans nos simulations 107

1.2. Les dimensions et la stabilité du complexe LH2 entouré de détergents 109

1.3. Le recouvrement membranaire du complexe LH2 112

2. Analyse des caractéristiques des détergents autour du complexe LH2 115

2.1. Conformation des détergents 115

2.2. Surfaces accessibles au solvant et hydratation des têtes polaires des détergents 116

2.3. Épaisseur des têtes polaires 117

3. Analyses structurales de la stabilité des peptides 118

4. Analyses des interactions entre peptides, pigments BChls-a et peptides-pigments BChls-a 119

4.1. Dynamique entre peptides 119

4.2. Dynamique entre bactériochlorophylles-a 121

4.3. Analyse des distances fonctionnelles clefs 124

5. Analyse des corrélations entre les différentes distances clefs 129

III. Conclusions de cette partie 131

Chapitre 5 : Calculs quantiques de l'absorbance 135

I. Modélisation du spectre d'absorbance 135

II. Conclusion de ce chapitre 139

Conclusions et perspectives 141

Annexe 1 : Test et validation des paramètres pour le β-Octyle glucoside 147

Annexe 2 : Validation des paramètres de simulation de la membrane pure de POPC . 161

Annexe 3 : Calculs des charges RESP du résidu CXM et de la RG 170

Annexe 4 : Nos expériences de SAXS avec les complexes LH2-détergents 178

Annexe 5 : Structure des peptides du complexe LH2 182

Annexe 6 : Analyse en Composantes Principales (ACP) 188

Annexe 7 : Participation à d'autres travaux 197

Références 291

Introduction générale

En 1870, Victor Hugo écrivait « c'est une triste chose de songer que la nature parle et que le genre humain n'écoute pas » (Carnets, albums, journaux). Environ un siècle plus tard, en 1960, naissait la bionique marquant officiellement l'intérêt des scientifiques pour la découverte de mécanismes inspirés du vivant. En effet, la plupart des inventions humaines sont grandement inspirées par l'observation de la nature. L'avant du TGV japonais (Shinkansen) a un nez qui mime le bec du martin pêcheur, ce qui lui a permis de gagner 10 % de vitesse en économisant 15 % d'énergie. En natation, les combinaisons calquées sur la peau de requin augmentent l'hydrodynamisme et actuellement, des tests de vernis recouvert sur les avions imitant cette peau de requin sont réalisés pour augmenter l'aérodynamisme des avions. Cette liste d'exemples non exhaustive s'enrichie d'année en année et l'application de ces connaissances devient bénéfique dans divers domaines.

Les exemples d'organisme vivant cités ci-dessus sont « facilement » remarquables à l'œil nu, mais il existe aussi d'autres organismes vivants de tailles microscopiques qui ont réussi à s'adapter et à perdurer au cours de l'évolution avec des mécanismes fascinants. Certains de ces microorganismes sont capables de faire la photosynthèse pour vivre dans des fonds marins, comme par exemple les bactéries pourpres. Ce processus bioénergétique vital pour ces organismes utilise l'énergie lumineuse pour synthétiser de la matière organique essentielle à leur survie. Pour ce faire, ces bactéries pourpres possèdent un photosystème composé de deux antennes collectrices de lumière (LH1 et LH2) et d'un centre réactionnel (CR), le tout ancré dans leur membrane plasmique. La lumière est principalement absorbée par des pigments photosynthétiques liés au complexe LH2 et l'énergie d'excitation résultante est ensuite transférée au complexe LH1 et, de là, au centre réactionnel où elle est transformée en énergie chimique. Depuis plusieurs décennies, les premières étapes de l'évolution de l'excitation pendant le processus photosynthétique ont suscitées un vif intérêt. De plus, nous savons que l'énergie fossile représente la grande majorité de l'énergie dans le monde, cependant elle s'épuise d'année en année et une solution pour continuer à avoir de l'énergie serait l'énergie renouvelable comme par exemple l'énergie solaire.

De ce fait, le photosystème des bactéries pourpres étant plus simple à étudier que celui des plantes, le complexe LH2 est vite devenu un système modèle. Chez *Rhodopseudomonas acidophila*, par exemple, l'étude clef en cristallographie de McDermott *et al.* (McDermott *et al.* 1995) ont mis en avant une symétrie parfaite en C_9 du complexe LH2 (pour *light*

harvesting) constitué d'une sous-unité protéique associée à des pigments photosynthétiques. Des études de spectroscopie de fluorescence en molécule unique de LH2 en fonction du temps ont montré que l'intensité et la position de transition électronique du complexe pouvaient fortement fluctuer avec le temps, décrivant ainsi un « désordre dynamique ».

Cette symétrie parfaite a un lien étroit avec la fonction du complexe LH2. Une étude récente a analysé, *in silico*, différentes tailles du complexe LH2 allant d'une symétrie en 5 à 12 protomères (Jang *et al.* 2015). Les auteurs ont montré que la formation du complexe LH2 non-naturel était énergétiquement possible, mais ces complexes présentaient des liaisons hydrogènes instables entraînant un plus grand désordre dynamique de l'énergie d'excitation des bactériochlorophylles-a et donc, une incapacité à relayer correctement l'exciton. Dans la même année, une autre étude a réussi à assembler des complexes LH2 en introduisant des pigments artificiels conjugués à ces complexes (Yoneda *et al.* 2015). Les résultats ont montré que ces nouveaux pigments révélaient un transfert d'énergie efficace aux chromophores intrinsèques, sans altérer la fonction du complexe LH2. Cela montre que la construction de systèmes photosynthétiques artificiels est capable de transférer l'énergie efficacement et en augmentant la gamme des longueurs d'onde (650 nm) comparé au complexe naturel. Ces complexes sont donc d'une grande importance pour la mise en place des systèmes de production d'énergie renouvelable.

Le désordre dynamique au sein des complexes LH2 a été largement observé en spectroscopie de fluorescence en molécule unique (LH2-détergents), cependant, du fait des limitations des méthodes, aucune étude n'a pu examiner au niveau moléculaire ces fluctuations au cours du temps. L'objectif de ce projet de thèse est de mieux comprendre l'origine de ce « désordre » à l'échelle atomique en employant des approches de dynamiques moléculaires classiques. Pour ce faire, nous avons modélisé le complexe LH2 dans différents environnements biomimétiques constitués de détergents (diméthyldodecylamine-N-oxide (LDAO) et β octyle glucoside (bOG)) et d'une membrane de POPC. Pour valider nos modèles, des expériences de diffraction des rayons X aux petits angles (SAXS) ont été réalisées avec les complexes LH2-LDAO et LH2-bOG. Dans un second temps, nous avons plus spécifiquement étudié les interactions peptide-pigment et pigment-pigment en fonction de l'environnement. Enfin, afin de relier les variations des interactions entre les différents composants du complexe décrits dans nos simulations, aux variations d'absorption du complexe LH2 et au désordre dynamique, des calculs quantiques de l'absorbance des pigments ont été réalisés à partir de structures atomiques représentatives de la simulation du complexe LH2-LDAO.

Le manuscrit est organisé de la manière suivante :

- Dans le premier chapitre, nous introduirons le contexte biologique de ce travail de thèse avec des rappels sur la photosynthèse chez les bactéries pourpres et une description structurale et fonctionnelle du complexe LH2 chez *Rhodopseudomonas acidophila* strain 10050. Nous détaillerons l'importance des expériences de spectroscopie de fluorescence en molécule unique de LH2 entouré de détergents qui ont permis de mettre en évidence le désordre dynamique au sein des complexes LH2.
- Dans le deuxième chapitre, nous ferons une introduction générale sur la dynamique moléculaire et les techniques que nous avons utilisées durant cette thèse pour modéliser le complexe dans les différents environnements biomimétiques de façon réaliste. Puis, nous terminerons par une description des charges partielles atomiques RESP et l'approche des blocs modulaires que nous avons utilisée pour calculer les charges des détergents et des différents composants du complexe LH2 compatibles avec les champs de force dérivés d'AMBER.
- Dans le troisième chapitre, nous présenterons les simulations des environnements hydrophobes micellaires constitués de détergents (LDAO et bOG) et d'une membrane de POPC. Ce chapitre reprendra essentiellement les résultats des simulations de micelles pures de LDAO et celles avec un fragment de lipide kinase de l'article publié dans *Journal of Biomolecular Structure and Dynamics* (Karakas *et al.* 2016) qui nous ont permis de valider nos paramètres pour le LDAO.
- Dans le quatrième chapitre, nous décrirons les simulations et les résultats des complexes LH2 entourés de LDAO, bOG et de POPC. Nous analyserons l'effet de l'environnement sur la structure globale et interne du complexe LH2.
- Dans le cinquième chapitre, nous présenterons quelques résultats préliminaires sur l'absorbance réalisés à partir des configurations de notre simulation du complexe LH2-LDAO.

Ce manuscrit contient aussi différentes annexes où sont détaillées les paramètres de simulations, les modèles utilisés pour les valider et les autres travaux effectués au cours de mes 3 ans et demi de thèse.

Enfin, nous terminerons le manuscrit avec nos conclusions et perspectives de ce travail.

Chapitre 1 : Le complexe collecteur de lumière de type 2 des bactéries pourpres

I. Introduction

La photosynthèse est le processus bioénergétique qui permet à certains organismes vivants, appelés organismes photosynthétiques ou phototrophes, de synthétiser de la matière organique essentielle à leur survie en utilisant l'énergie lumineuse du soleil. Tous ces organismes possèdent un appareil photosynthétique membranaire mais ne réalisent pas le même type de photosynthèse. Cependant, le processus commence toujours par l'absorption de photons par les pigments assimilateurs des antennes collectrices de lumière (*light harvesting* ou LH) qui vont transférer ces particules au centre réactionnel (CR). Cela va entraîner une séparation des charges transmembranaires. À partir de là, les premières réactions d'oxydoréduction photochimique ont lieu décrivant ainsi le centre réactionnel comme le lieu de conversion de l'énergie des photons en énergie chimique.

Les organismes photosynthétiques se divisent en deux grands groupes, ceux qui pratiquent la photosynthèse oxygénique et anoxygénique, c'est-à-dire qu'ils produisent ou non du dioxygène, respectivement. Parmi les organismes photosynthétiques anoxygéniques se trouvent les bactéries pourpres. Ces dernières possèdent un système antennaire qui a la particularité d'être riche en pigment caroténoïde (jaune-rouge) masquant la couleur bleu-vert des pigments bactériochlorophylles. La niche écologique de certaines de ces bactéries les place dans les milieux aquatiques anaérobiques (Figure 1). À ce niveau, le spectre de l'énergie solaire qui atteint ces bactéries pourpres a été filtré par les pigments contenus dans les phototrophes oxygéniques qui sont situés juste au-dessus, plus proche du soleil. Ce filtrage élimine la lumière bleue et rouge (jusqu'à 750 nm). Cependant, les bactéries pourpres peuvent assurer leur balance énergétique en capturant les photons de la lumière du proche infrarouge, au-delà de 750 nm, grâce à ses pigments photosynthétiques spécifiques (les bactériochlorophylles). Le mode de croissance privilégié est phototrophique en anaérobiose, cependant les bactéries pourpres sont capables de développer une activité respiratoire en fonction des conditions de cultures et des changements environnementaux imposés.

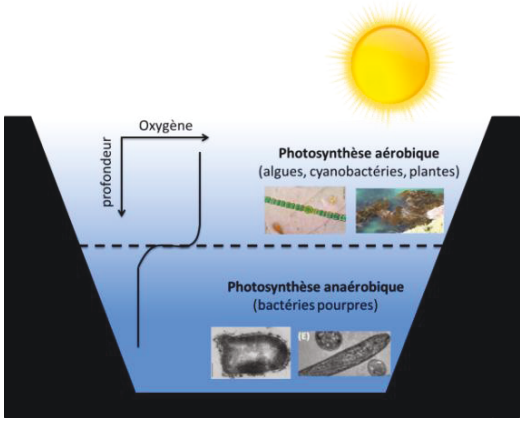


Figure 1 : Schématisation d'un lac avec les différents phototrophes. Les organismes ayant une activité respiratoire de la strate supérieure consomment l'oxygène, ce qui entraîne le développement anaérobie dès 4-5 mètres de profondeur. Refaite à partir de (Cogdell *et al.* 2006)

Le système antennaire des bactéries pourpres photosynthétiques est relativement plus simple comparé à celui des plantes. De ce fait, depuis plusieurs décennies, il est souvent utilisé comme un système modèle pour étudier les premières étapes de d'évolution de l'excitation pendant le processus photosynthétique. Durant cette thèse, nous nous sommes intéressés à la première antenne collectrice de lumière de type 2 (LH2) de la bactérie pourpre *Rhodopseudomonas (Rps.) acidophila* appelée aussi *Rhodoblastus acidophilus* qui fait partie des bactéries pourpres non-sulfureuses photosynthétiques. Le terme « non-sulfureuse » vient du fait qu'elles n'utilisent pas de composés sulfurés pour la fixation du CO_2 (Figure 2).

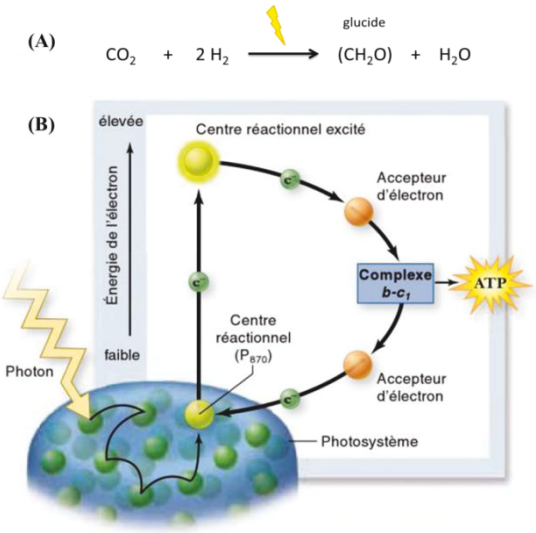


Figure 2 : Équation bilan générale de la photosynthèse chez les bactéries pourpres non sulfureuse photosynthétiques (A) et le transfert cyclique d'électrons : parcours d'un électron dans le photosystème d'une bactérie pourpre non sulfureuse (B). Le complexe b-c_1 est le complexe de cytochrome b-c assurant le recyclage des électrons libérés. Tirée du livre (Peter *et al.* 2011).

En 1996, une étude a montré que le procaryote qui est devenu un endosymbionte des eucaryotes, et a évolué en mitochondrie, est un proche parent des bactéries pourpres non-sulfureuses. Cette conclusion repose sur les ressemblances métaboliques entre les mitochondries et ces bactéries, ainsi que sur les comparaisons de séquences de bases dans leurs petites sous-unités d'ARN ribosomique (Bui *et al.* 1996).

II. Le photosystème chez les bactéries pourpres

Chez les bactéries photosynthétiques, l'appareil photosynthétique est localisé dans les invaginations de la membrane cytoplasmique. Il est constitué de complexes protéiques membranaires avec des pigments spécialisés. Cet appareil photosynthétique comprend deux types de complexe antennaire collecteur de lumière appelés LH1 et LH2, et d'un centre réactionnel (RC). Le complexe LH1 est l'antenne interne étroitement associée au centre réactionnel et l'ensemble est entouré de complexes LH2 périphériques (Figure 3).

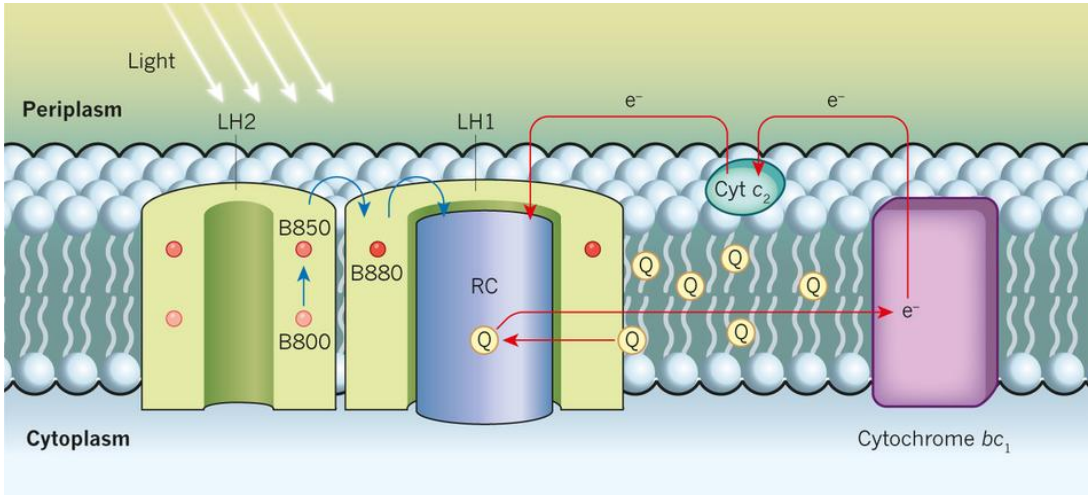


Figure 3 : Photosystème membranaire des bactéries pourpres. Le RC excité réduit les molécules d'ubiquinone (Q). Ces dernières quittent le complexe LH1-RC et passent dans la membrane, transférant ainsi les électrons au cytochrome bc1. Le transfert cyclique des électrons est complété par une protéine soluble (cytochrome c2). Tirée de (Cogdell *et al.* 2014).

La lumière est principalement absorbée par les pigments photosynthétiques liés au complexe LH2 et l'énergie d'excitation résultante est ensuite transférée au complexe LH1 et, de là, au centre réactionnel où elle est transformée en énergie chimique.

1. Les phospholipides membranaires

Les lipides membranaires regroupent une large famille de lipides avec des structures chimiques très variées dont un exemple est donné sur la Figure 4.

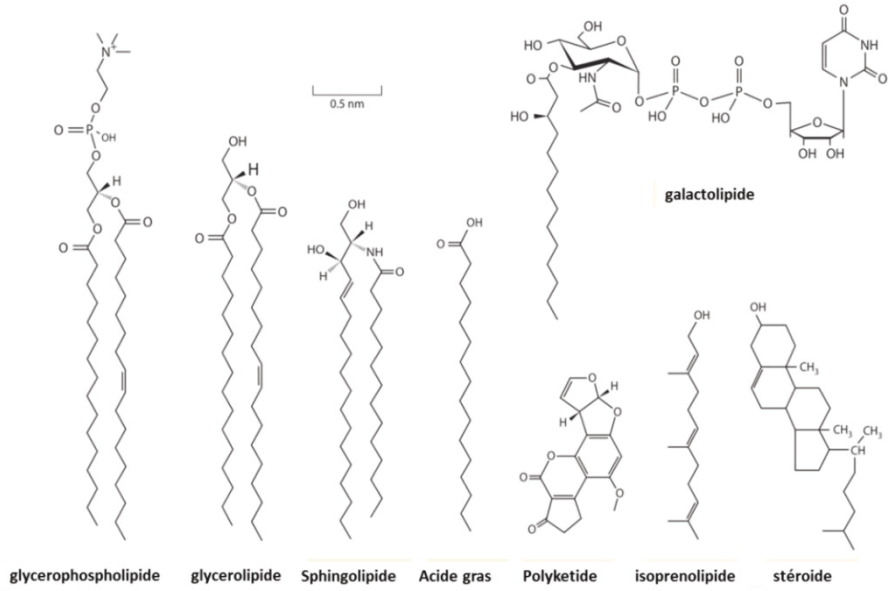


Figure 4 : Exemple de lipides membranaires. Adaptée de (Fahy *et al.* 2005).

Parmi les lipides membranaires, les phospholipides (PL) représentent une majorité des lipides membranaires. Ils sont constitués d'un groupement polaire constitué du groupement sucre, amine ou alcool, d'un phosphate, glycérol et deux chaînes alcanes plus ou moins saturées. Associés à d'autres molécules (par exemple les stérols), ils forment une « barrière » qui sépare et régule (avec les protéines membranaires) les échanges l'environnement intracellulaire de l'extérieur. La composition en PL varie en fonction du type d'organisme (animale, bactérie ou plante) et de sa localisation et fonction (Tableau 1).

Tableau 1 : Comparaison de la composition plasmique dans différents organismes (SM : sphingomyèline).

Cellule animale		Cellule végétale	Cellule bactérienne
Membrane plasmique	**Membrane mitochondriale interne**		
30 % cholestérol 43 % PL 14 % SM	3 % cholestérol 77 % PL 2 % SM 18 % cardiolipine	10 % PL 80 % glycolipide 10 % sulfo-lipide	95 % PL 5 % cardiolipine

En ce qui concerne la membrane photosynthétique de la bactérie pourpre *Rps. acidophila* strain 10050, celle-ci est constituée de phosphatidylethanolamine (PE) à 49 % et de phosphatidylcholine (PC) à ~31 %, et minoritairement de cardiolipine à ~4 % et de phosphatidylglycerol (PG) à 16 % avec des chaînes grasses saturées (16:0-18 :0) (Russell *et al.* 2002).

2. Le complexe LH2

La première résolution de la structure cristallographique par diffraction de rayons X a été celle du complexe LH2 chez *Rps. acidophila* strain 10050 en 1993 avec une résolution de 3 Å (McDermott *et al*. 1995). Depuis, trois nouvelles structures tridimensionnelles de plus haute résolution ont été résolues (1KZU (Prince *et al.* 1997), 2FKW (Cherezov *et al.* 2006), 1NKZ (Papiz *et al.* 2003) à 2,50 ; 2,45 et 2 Å, respectivement). Ces structures ont permis de montrer une remarquable symétrie circulaire de neuf sous-unités (Figure 5B). Un protomère de ce complexe est constitué d'un hétérodimère α et β (de 53 et 41 résidus, respectivement) associé à quatre pigments photosynthétiques, un caroténoïde (rhodopine glucoside noté RG1) et trois bactériochlorophylles-a (BChls-a notés B800 et B850)(Figure 5C).

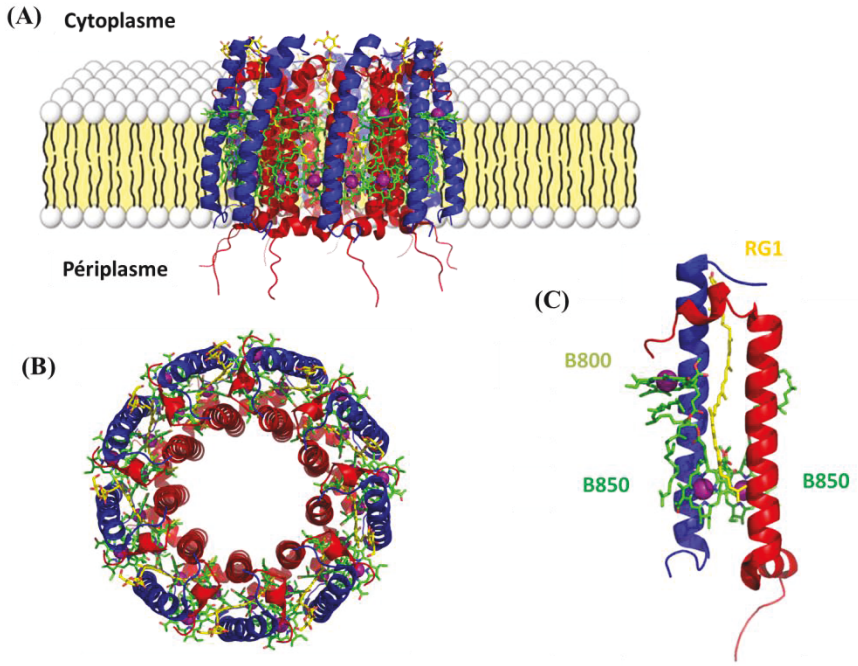


Figure 5 : Organisation du complexe LH2 vue de profil (A) et d'au-dessus (B) et d'un protomère (C) composé des peptides α (rouge) et β (bleu) et des pigments (RG1, B800 et B850).

Ce complexe LH2 a une structure annulaire avec un anneau externe d'environ 68 Å de diamètre formée par les peptides α, et un anneau interne 36 Å de diamètre formé par les peptides β. Chaque peptide est formé de trois régions distinctes en hélices-α : les domaines polaires N- et C- terminales, et un domaine central hydrophobe qui traverse la membrane. La partie N- terminale des peptides est située du côté cytoplasmique et a une longueur de 12–14 résidus pour le peptide α et de 13–15 résidus pour le peptide β. La région C-terminale, localisée du côté périplasmique, est constituée d'environ 16-20 et 5-12 résidus pour les polypeptides α et β, respectivement. Quant à la partie transmembranaire, elle est constituée d'une vingtaine de résidus majoritairement hydrophobes. Cette région contient également des histidines et arginines très conservées (Zuber 1985)(Zuber *et al*. 1991) qui interagissent avec les BChls-a *via* des interactions de types van der Waals ou liaisons hydrogènes (Figure 6).

```
                1         10         20         30         40         50
                |          |          |          |          |          |
peptide α       MNQGKIWTV VNPAIGIPAL LGSVTVIAIL VHLAILSHTT WFPAYWQGGV KKAA

peptide β       ATLTAEQSE ELHKYVIDGT RVFLGLALVA HFLAFSATPW LH
```

Figure 6 : Séquences en acide aminé des peptides α et β, avec les parties en hélices alpha (rouge) et en *random coil* (vert). Les résidus conservés sont mis en rouge et gras.

L'absorption de photons de la lumière, par les pigments photosynthétiques à une certaine longueur d'onde, produit des transitions électroniques génératrices d'états excités. Cela se traduit par des bandes d'absorption situées à ces longueurs d'onde. Dans le cas du complexe LH2, la rhodopine glucoside absorbe à une longueur d'onde d'environ 500 nm et les BChls-a à 800 et 850 nm, qu'on écrira B800 et B850. Les bactériochlorophylles-a sont des porphyrines dont le noyau de base est un macrocycle plan constitué de quatre pyrroles reliés par des ponts méthines (-CH=) et possédant un métal de magnésium lié à deux pyrroles (Figure 7A). Ce macrocycle est relié à une chaîne phytyle. Le système d'électron π du cycle bactériochlorine qui coordonne l'ion Mg^{2+} aux quatre atomes N pyrroliques, est responsable des bandes d'absorption des BChls-a (Robert *et al.* 1985).

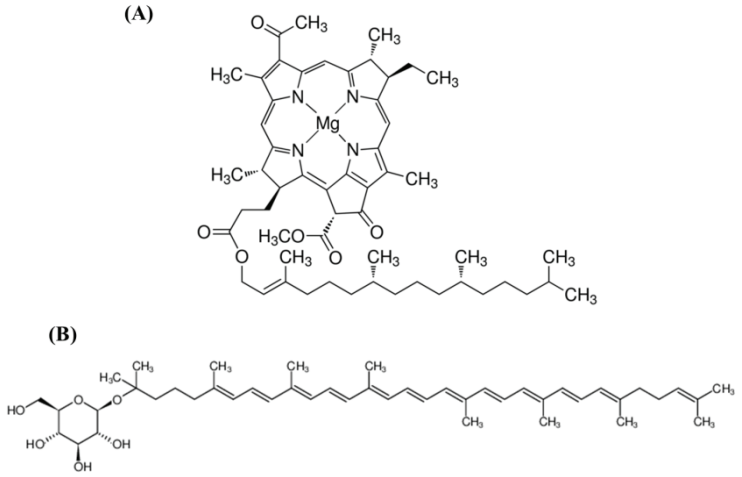


Figure 7 : Représentation des pigments bactériochlorophylles-a (A) et rhodopine glucoside (B).

La rhodopine glucoside est constituée d'un glucoside relié à une chaîne de onze doubles liaisons conjuguées en configuration *trans* (Figure 7B). Cette longue chaîne carbonée peut former un grand nombre d'interactions hydrophobes au sein des complexes. De ce fait, elles ont un rôle important dans la cohésion de l'ensemble du complexe. En effet, une étude a montré qu'un mutant délété de rhodopines glucosides ne parvenait pas à s'assembler en complexe (Hunter *et al.* 1994). Ces pigments sont aussi capables de transférer l'énergie absorbée du visible aux BChls-a, ce qu'ils leur confèrent un rôle de collecteur de lumière accessoire. Ces caroténoïdes ont aussi un rôle important dans la photoprotection contre l'oxydation photo induit. En effet, ils inhibent l'énergie en excès des bactériochlorophylles-a afin d'éviter la formation de radicaux d'oxygène (i.e oxygène singulet) capables de détruire la cellule (Papiz *et al.* 2003). La structure tridimensionnelle du LH2 résolue à 2 Å suggère la présence d'un deuxième caroténoïde (Papiz *et al.* 2003). Mais des expériences de spectroscopie Raman et la dernière structure 3D résolue à 2,45 Å ont montré qu'il n'y avait qu'une seule molécule de rhodopine glucoside présente par hétérodimère αβ. Chacune de ces molécules joue un rôle important dans le transfert de l'énergie lumineuse au sein du complexe (voir la partie VI de ce chapitre).

3. Le complexe CR-LH1

Des expériences de cristallographie électronique et cryo-EM ont permis d'obtenir pour la première fois en 1995 une structure à 8,5 Å de résolution d'un cristal 2D de CR-LH1 purifié à partir de la bactérie *Rhodopsirillum rubrum* (Karrasch *et al.* 1995). Ces complexes LH1 présentent une structure circulaire composée de seize hétérodimères αβ avec un « trou » au centre assez large capable de contenir le centre réactionnel. Contrairement au

complexe LH2, le complexe CR-LH1 peut adopter différentes conformations comme par exemple circulaire, sous forme de S, elliptique ou bien en arc (Gerken *et al.* 2003)(Scheuring *et al.* 2005)(Fotiadis *et al.* 2004)(Bahatyrova *et al.* 2004)(Figure 8). Selon l'espèce, le complexe CR-LH1 peut être sous forme monomérique (*Rps. rubrum*) ou dimérique (*Rps. spheroides*). Du fait de son instabilité en présence de détergent, il existe peu de structures cristallographiques à haute résolution du complexe CR-LH1. De ce fait, ces structures ne permettent pas de distinguer les détails moléculaires et notamment les interactions pigments-protéines essentielles pour la compréhension de la fonction de ce complexe.

Les différentes structures du complexe CR-LH1 présentées dans la Figure 8 montrent que les deux premiers LH1 (Figure 8a et 8b) ne sont pas complètement refermés. En effet, dans ces deux représentations, une hélice transmembranaire PufX ou W interrompt l'anneau formé par les polypeptides α au niveau du site de fixation de la quinone Q_B du centre réactionnel. Cela laisse supposer que cette protéine pourrait être impliquée dans le passage de Q_B à travers le LH1 et faciliter ainsi sa diffusion dans la membrane. La protéine PufX que l'on trouve chez *Rhodobacter (Rb.) sphaeroides* est un analogue du polypeptide PufX.

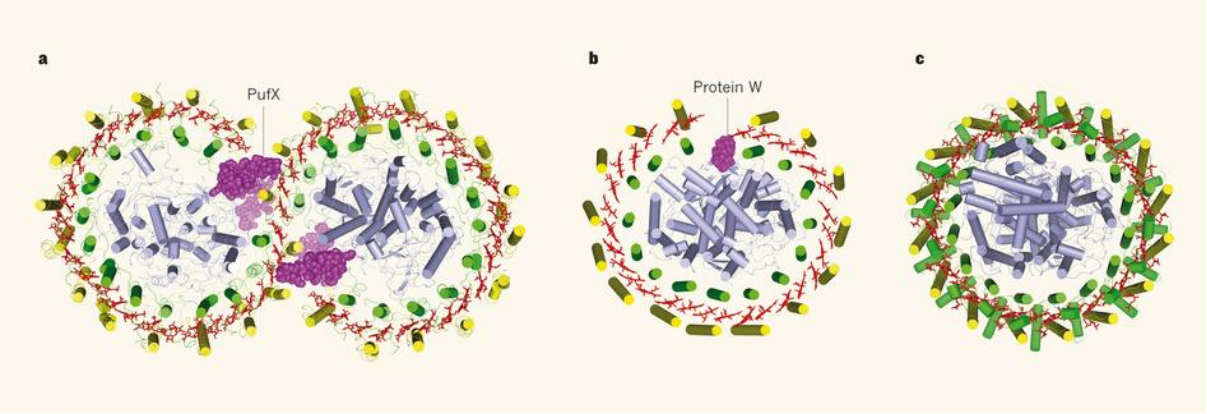


Figure 8 : Différentes structures du complexe CR-LH1, avec les hélices représentées par des cylindres pour les polypeptides αβ du LH1 (en jaune et vert) et les hélices du RC (en bleu), et les bactériochlorophylles du LH1 en rouge. Les caroténoïdes du LH1 et les cofacteurs du RC n'ont pas été représentés par soucis de clarté. (a) un modèle dimérique chez *Rb. sphaeroides* (8 Å de résolution) avec la protéine PufX qui crée des gaps dans l'anneau du LH1, (b) un modèle monomérique chez *Rps. palustris* (4,8 Å) avec la protéine W qui crée un gap et enfin un modèle monomérique chez *Thermochromatium tepidum* (3 Å). Tirée de (Cogdell *et al.* 2014).

Le centre réactionnel est composé de trois sous-unités polypeptidiques L, M et H structurés en hélices α. La sous-unité H est constituée d'une seule hélice transmembranaire et d'un domaine globulaire cytoplasmique, tandis que les deux autres sous-unités contiennent cinq hélices chacune. Ces dernières lient de façon non covalente des cofacteurs qui assurent sa

fonction : un dimère de bactériochlorophylles-a, deux bactériochlorophylles-a accessoires et deux bacteriopheophytines-a (similaire à la bactériochlorophylle-a mais sans Mg^{2+}) absorbant respectivement à 870, 800 et 750 nm, deux molécules de quinone Q_A et Q_B et un caroténoïdes qui complète le spectre d'absorption dans le visible entre 450 et 550 nm.

III. Les variants naturels des complexes antennaires périphériques

La plupart des chercheurs qui étudient les bactéries pourpres travaillent avec un nombre assez restreint d'espèces (i.e *Rb. spheroides* ou *Rps. acidophila*). Ces espèces sont connues pour être historiquement les plus faciles à cultiver avec un génome maniable. Cependant, les complexes antennaires de ces espèces montrent des spectres d'absorption différents. Il y a plusieurs années, il a été montré que lorsque certaines souches de *Rps. acidophila* (i.e strain 7050) étaient cultivées sous de faible intensité lumineuse, elles synthétisaient un nouveau type de complexe LH2 (Cogdell *et al*. 1985) appelé complexe LH3 (McLuskey *et al.* 2001) dont les bactériochlorophylles-a absorbent à des longueurs d'onde de 800 et 820 nm au lieu de 800 et 850 nm. La comparaison des séquences protéiques α de ces deux complexes ont mis en évidence le changement de deux résidus clefs en position 44 et 45 (Figure 9) comme étant corrélés avec le déplacement de la bande d'absorption de 850 à 820 nm. Des études ont montrés que la liaison hydrogène, entre ces résidus et les groupes carbonyles du macrocycle des bactériochlorines, était responsable du déplacement de la bande. En effet, les résidus 44 et 45 formant potentiellement des liaisons hydrogènes (i.e tyrosine et tryptophane) sont remplacés par des résidus hydrophobes comme la phénylalaline et la leucine.

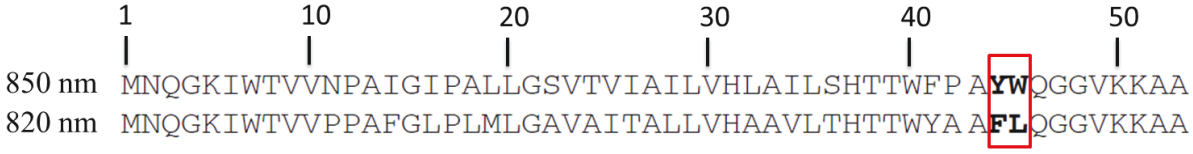


Figure 9 : Comparaison des séquences amino acides primaires des peptides α des LH2 complexes absorbant à 850 et 820 nm chez *Rps. acidophila*. La tyrosine 44 et le tryptophane 45 de la souche 10050 sont remplacés par une phénylalaline et une leucine dans la souche 7050.

De plus, il a été montré que la concentration de LH2 dépend essentiellement des conditions de cultures qui contribuent à augmenter la surface accessible d'absorption de la lumière et par conséquent à optimiser le rendement de la photosynthèse.

IV. Les interactions bactériochlorophylles-a/peptides et entre bactériochlorophylles-a

Nous venons de voir que les liaisons hydrogènes entre les bactériochlorophylles-a et les polypeptides étaient impliquées dans le déplacement des bandes d'absorption. Des études *in vivo* chez *Rb. sphaeroides* ont montré qu'en enlevant une partie des extrémités N-terminales, les complexes LH2 formés étaient dépourvus de B800 (Koolhaas *et al.* 1997), et que l'altération des extrémités C-terminales conduisaient à la formation d'anneau non fonctionnels (Braun *et al.* 2002). La présence de résidus histidines conservés (α-His30 et β-His31) dans le domaine transmembranaire sont indispensables pour les liaisons aux bactériochlorophylles-a et donc indispensables à l'assemblage du LH2 (Olsen *et al.* 2003). La coordination du magnésium central avec les histidines entraîne une légère distorsion du macrocycle bactériochlorine que l'on retrouve dans le complexe LH2 et LH1 (Papiz *et al.* 2003)(Lapouge *et al.* 1999). Cette distorsion n'entraîne qu'un faible déplacement de la bande d'absorption et influe plutôt sur le temps de vie de l'état excité des bactériochlorophylles (Gentemann *et al.* 1997).

Ces interactions sont sensibles à certains facteurs tels que la taille de l'agrégat, la distance entre les pigments, les angles entre les dipôles (Figure 10) et le désordre statique et dynamique (voir la partie V de chapitre).

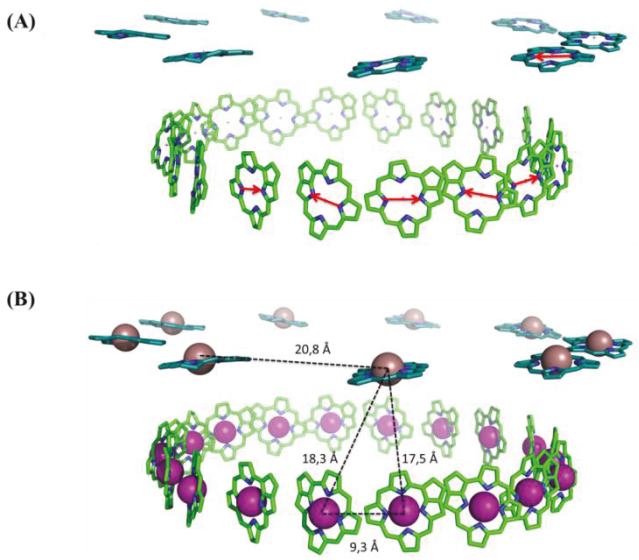


Figure 10 : Relation entre les distances BChls et le moment dipolaire dans le complexe LH2 chez *Rps. acidophila* (PDB 2FKW). Les 9 B800 sont représentés en cyan foncé et les 18 B850 en vert clair, seuls les atomes lourds des macrocycles ont été représentés. Les dipôles de quelques BChls sont représentés par des flèches rouges (A). Les distances entre les magnésiums des BChls sont représentées par des lignes noires discontinues (B).

V. L'absorption des pigments

L'observation du spectre d'absorption des antennes LH permet d'obtenir des informations sur leur structure et leur fonction. En effet, le spectre d'absorption de ces molécules peut en effet être décomposé en régions distinctes, correspondant aux transitions électroniques. Les rhodopines glucosides et les bactériochlorophylles-a sont connues pour avoir deux pics d'absorption caractéristiques S_1 et S_2 et, Q_x et Q_y respectivement (Figure 11.2)(Cogdell *et al.* 2006). En présence de la membrane photosynthétique, l'interaction des pigments avec la partie protéique induit un décalage vers le rouge ou *red shift* des bandes majeures d'absorption (Figure 11.2c). Cela est d'autant plus remarquable pour la transition Q_y qui se décale de 772 nm à 800-900 nm. Pour tenter d'expliquer le déplacement et l'élargissement de la bande d'absorption de la transition Q_y, il est important de prendre en compte la force d'interactions des bactériochlorophylles-a. Ces dernières interagissent fortement et la transition Q_y pourrait être représentée par une unique transition électronique provenant de l'ensemble des bactériochlorophylles-a, comme un multimère figé et ordonné. Cependant, les propriétés électroniques de la transition Q_y dépendent également des variations des propriétés individuelles de chaque bactériochlorophylle-a de l'anneau, induites par le désordre. Cette notion de désordre considère qu'à un instant donné, toutes les bactériochlorophylles-a n'ont pas les mêmes propriétés optiques. Cela pourrait expliquer l'élargissement inhomogène de la bande d'absorption. En effet, dans l'assemblage de molécules identiques tels que les bactériochlorophylles-a, les différences physico-chimiques peuvent provenir de différences de conformations des molécules et de différences structurales dans leur environnement proche. Ces différences peuvent être liées au désordre intrinsèque de la structure (désordre statique) et à ses fluctuations dynamiques (désordre dynamique).

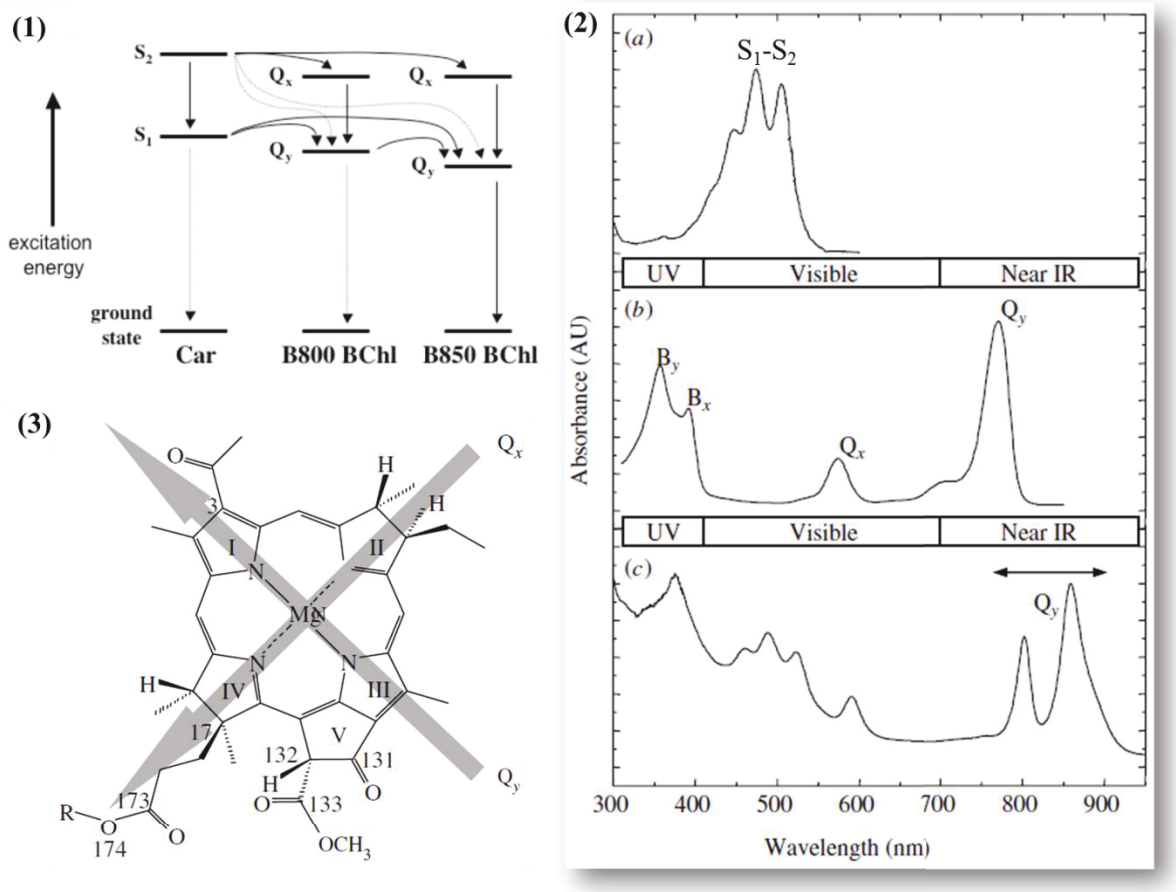


Figure 11 : Représentation des niveaux d'énergie des états d'excitation des pigments avec les chemins possibles du transfert excitonique (1) (Sundström *et al.* 1999). L'absorbance des pigments (2) (Cogdell *et al.* 2006) : RG (a), BCL (b), d'un LH2 (c) avec Bx et By, les transitions de plus haute énergie (bandes de Soret). Une structure de la bactériochlorine avec les dipôles de transition Qx et Qy représentés par des flèches (3) (Cogdell, *et al.* 2006).

VI. Transfert d'énergie dans les complexes LH

Un atome est dans un état excité lorsqu'un des électrons d'une couche électrique inférieure passe, grâce à l'énergie reçu de l'atome, à une couche électronique supérieure (Frenkel 1931). L'état excité étant instable, l'électron va revenir à sa couche électronique initiale. Cela entraîne une libération d'énergie sous forme de photon qui équivaut au photon reçu (fluorescence). Dans le cas du complexe LH2, après absorption des photons par les pigments du LH2, ces derniers vont transmettre leur énergie d'excitation au LH1 puis au centre réactionnel. Un pigment donneur excité va transmettre son excitation à un autre pigment accepteur non excité. En d'autres termes, l'excitation est transférée d'une molécule avec un niveau d'énergie haut à une molécule de niveau d'énergie plus basse (Figure 11.1).

Le transfert d'énergie entre les complexes antennaires et le centre réactionnel est décrit par le mécanisme de Förster (Förster 1948). En d'autres termes, ce mécanisme décrit le transfert d'énergie entre molécules fluorescentes appelé aussi le transfert d'énergie par résonance de Förster (FRET, *Förster resonance energy transfer*) (Figure 12). Le FRET est défini comme un transfert d'énergie non radiatif (sans émission de lumière) résultant d'une interaction dipôle-dipôle entre deux molécules (donneur et accepteur d'énergie).

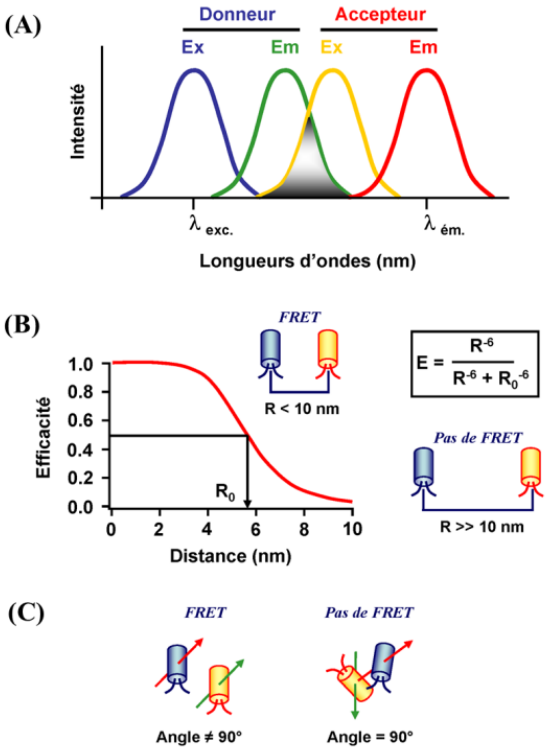


Figure 12 : Condition du FRET (*Förster Resonance Energy Transfer*). Le couple de particule interagit via l'interaction de Coulomb. Tirée de wikipedia.org.

Dans la description la plus simple du mécanisme de Förster, le taux de transfert d'énergie est déterminé par le chevauchement du spectre de fluorescence du donneur et le spectre d'absorption de l'accepteur (Figure 12). La durée de vie de l'état excité est très courte, de l'ordre de la nanoseconde. C'est pourquoi, dans le cadre de la photosynthèse, le transfert d'énergie jusqu'au centre réactionnel doit avoir lieu dans ce laps de temps (Figure 14).

Le transfert d'énergie est orienté vers le centre réactionnel par un gradient d'énergie crée par la position de transition électronique Q_y. Cette unique direction confère au photosystème son efficacité avec un rendement supérieur à 95 % (Fleming *et al.* 2008)(Pullerits *et al.*1996). Les caroténoïdes, qui sont en interaction de van der Waals avec

les bactériochlorophylles, peuvent transférer leur énergie aux bactériochlorophylles avec une efficacité variable (entre 30 à 90 % selon les antennes et les espèces bactériennes). Dans les LH2, l'énergie d'excitation migre entre B800 en 500 fs, puis entre B800 et B850 en environ 1,2 ps. Delà, le transfert d'énergie est très rapide entre les molécules B850 qui sont en forte interaction (100-200 fs) et se trouve délocaliser sur une partie de l'anneau. De ce fait, les anneaux voisins en contact direct (autres LH2 ou LH1) peuvent recevoir l'énergie d'excitation, quel que soit la localisation des bactériochlorophylles excités sur l'anneau du LH2. Entre les B850 du LH2 et B875 du LH1, le passage de l'énergie d'excitation s'effectue en 3-5 ps. Enfin, cette énergie termine son chemin entre les B750 et le dimère de bactériochlorophylles du centre réactionnel dans un temps plus lent de 35 ps.

Cependant, dans le prérequis fondamental du FRET, pour définir cette interaction dipôle-dipôle, il faut que la taille de la partie optiquement active *d* soit plus petite que la distance entre les deux molécules fluorescentes *r* (Figure 13).

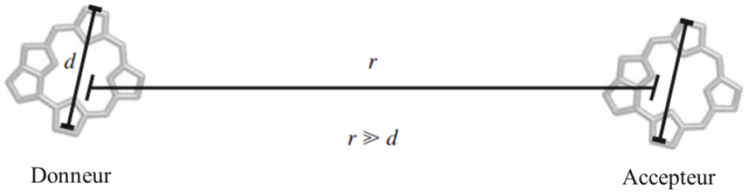


Figure 13 : Illustration de la condition à respecter pour avoir une interaction dipôle-dipôle avec d = 8,8 Å. Tirée de (Cogdell *et al.* 2006).

De ce fait, cette approximation ne fonctionne pas dans le cas des B850 et d'autres approches ont été discutées dans la littérature pour décrire le transfert de l'énergie d'excitation avec : la résonance généralisée de Förster (Scholes *et al.* 2001), la résonance multichromophorique de Förster ainsi que la théorie de *Redfield* (Chernyak *et al*. 1996)(Zhang *et al.* 1998)(Grondelle *et al*. 2006).

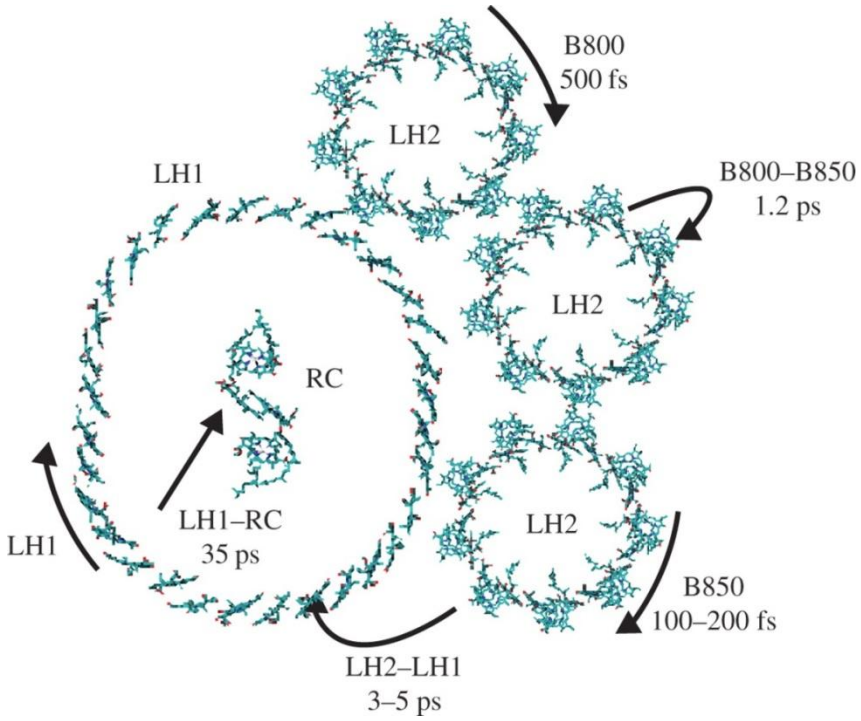


Figure 14 : Modèle de photosystème chez *Rps. acidophila* avec les différents temps de transfert d'énergie (Smyth *et al*. 2012).

VII. Expériences de spectroscopie à molécule unique (*Single-molecule*)

Généralement, la spectroscopie optique permet d'avoir une information sur les paramètres qui déterminent les structures électroniques des complexes collecteurs de lumière. Cependant, il est difficile d'étudier le processus de transfert d'énergie du fait de l'élargissement inhomogène des lignes d'absorption optique en raison de l'hétérogénéité de l'ensemble des pigments. En effet, dans un grand nombre de molécule hétérogène, chaque molécule présente un site d'absorption qui va contribuer à l'élargissement de la ligne d'absorption moyenne. De ce fait, les détails spectroscopiques seront cachés dans cet élargissement inhomogène des lignes d'absorption. Pour contourner ces difficultés, des techniques de spectroscopie de fluorescence en molécule unique ont été appliquées pour étudier le complexe LH2. Cette technique analyse une à une des molécules fluorescentes et permet l'observation de processus dynamique (Figure 15).

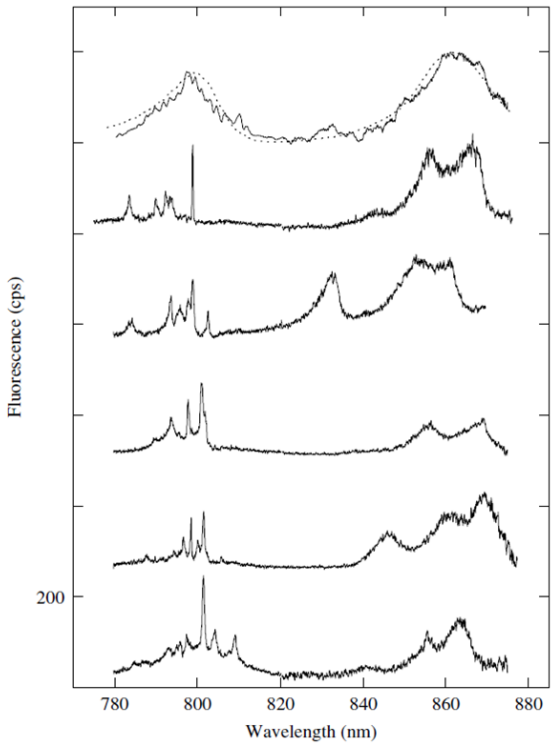


Figure 15 : Spectres d'excitation-fluorescence *Rps. acidophila* pour un ensemble de complexes LH2 (en haut) et plusieurs LH2 individuels à 1,2 K (Cogdell *et al.* 2006).

Dans la Figure 15, l'observation des différents spectres d'excitation-fluorescence pour chaque complexe LH2 montre deux bandes différentes pour les B800 et les B850. En effet, les bandes des B800 consistent en plusieurs raies spectrales relativement étroites alors que celles des B850 sont constituées de 2-3 bandes spectrales plus larges. Afin de comprendre ces différences, il est utile de considérer les forces d'interaction intermoléculaire entre bactériochlorophylles voisines dans un anneau. Des expériences ont montrés que le couplage parmi les B800 était beaucoup plus petit que celui des B850 (van Oijen *et al.* 1998)(van Oijen *et al.* 2000). À partir de ces informations spectrales, Sauer et ses collaborateurs ont pu approximer la force de couplage des B800 à 30 cm^{-1} et celle des B850 à environ 300 cm^{-1} (Sauer *et al.* 1996).

Les expériences de spectroscopie décrites ci-dessus ont été réalisées à très basse température. On pourrait se poser la question de savoir si ces résultats sont valables dans des conditions physiologiques. Des expériences de spectrométrie de fluorescence ont également été réalisées à température ambiante mais avec des temps d'exposition plus courts, comparé aux températures cryogéniques (Jimenez *et al.* 1996)(Gall *et al.* 2015). Dans toutes ces expériences de spectroscopie en molécule unique, le même phénomène de désordre dynamique a été observé dans les spectres de fluorescence.

Dans notre laboratoire, des expériences de spectroscopie de fluorescence en molécule unique de LH2 chez *Rps. acidophila* strain 10050 ont été réalisées dans des conditions ambiantes et en fonction du temps d'exposition (Figure 16). Les spectres obtenus pour deux LH2 différents entourés de molécule de détergents (N,N-dimethyl-dodecylamine oxide ou LDAO) ont aussi montré un désordre dynamique au sein de ces complexes. En effet, comme nous pouvons le voir sur la Figure 16, le spectre d'émission peut visiter un large éventail d'états d'émission parfois rapidement et basculer dans d'autres états d'émission durant parfois quelques millisecondes (les spectres étant collectés durant quelques millisecondes).

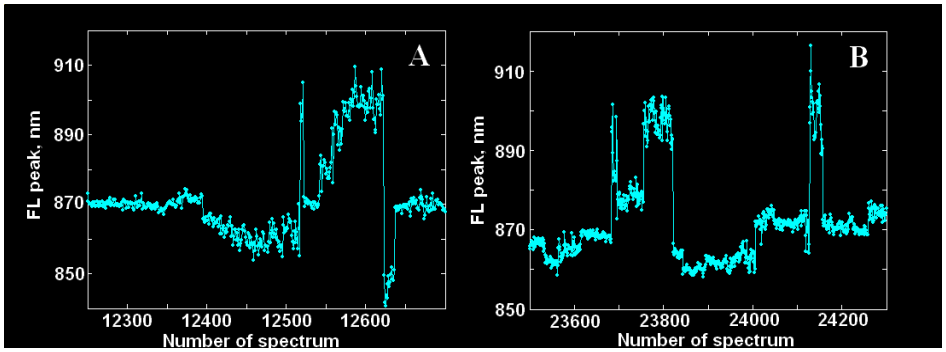


Figure 16 : Deux spectres de fluorescence en molécule unique obtenus pour deux complexe LH2 en fonction du temps d'exposition chez *Rps. acidophil*a strain 10050 et réalisés au laboratoire, dans des conditions ambiantes (Rutkauskas *et al.* 2006).

VIII. Conclusion du chapitre

Dans ce premier chapitre, nous avons vu que certaines bactéries étaient capable de faire la photosynthèse pour vivre dans leur niche écologique. Ce processus bioénergétique vital pour ces organismes se fait grâce à un photosystème composé de deux antennes collectrices de lumière (LH1 et LH2) et d'un centre réactionnel (CR), le tout ancré dans leur membrane plasmique. Chez *Rps. acidophila*, les études cristallographiques et microscopiques ont mis en avant une symétrie parfaite en C_9 constitué d'une sous-unité protéique et de pigments photosynthétiques (B800, B850 et RG1). Des études de spectroscopie de fluorescence en molécule unique de LH2 en fonction du temps ont montré que l'intensité et la position de transition électronique du complexe pouvaient fortement fluctuer avec le temps. Ces observations décrivent un « désordre dynamique » en lien avec la fonction biologique du complexe LH2, qui est caractérisée par une remarquable efficacité d'utilisation de l'énergie lumineuse. À ce jour, beaucoup d'études mettent en avant l'existence de ce désordre mais aucune n'a pu examiner au niveau moléculaire, les fluctuations locales ou globales au sein du complexe LH2 qui gouvernent ce désordre.

Chapitre 2 : La Dynamique Moléculaire

I. Introduction

La compréhension du fonctionnement des systèmes biologiques passe par l'étude de différents phénomènes biologique, physique ou chimique avec des échelles de temps et de résolutions spatiales très variables qui ne peuvent être accessibles qu'en utilisant des approches expérimentales spécifiques (Figure 1). En ce qui concerne les approches de modélisation et de simulation moléculaires, celles-ci permettent de caractériser des phénomènes avec des résolutions spatiales de l'ordre de l'Angström – 100 nanomètres et avec des échelles de temps de l'ordre de la femto (10^{-15}) à la milliseconde (10^{-3} s), ce qui rend ces techniques complémentaires aux approches expérimentales.

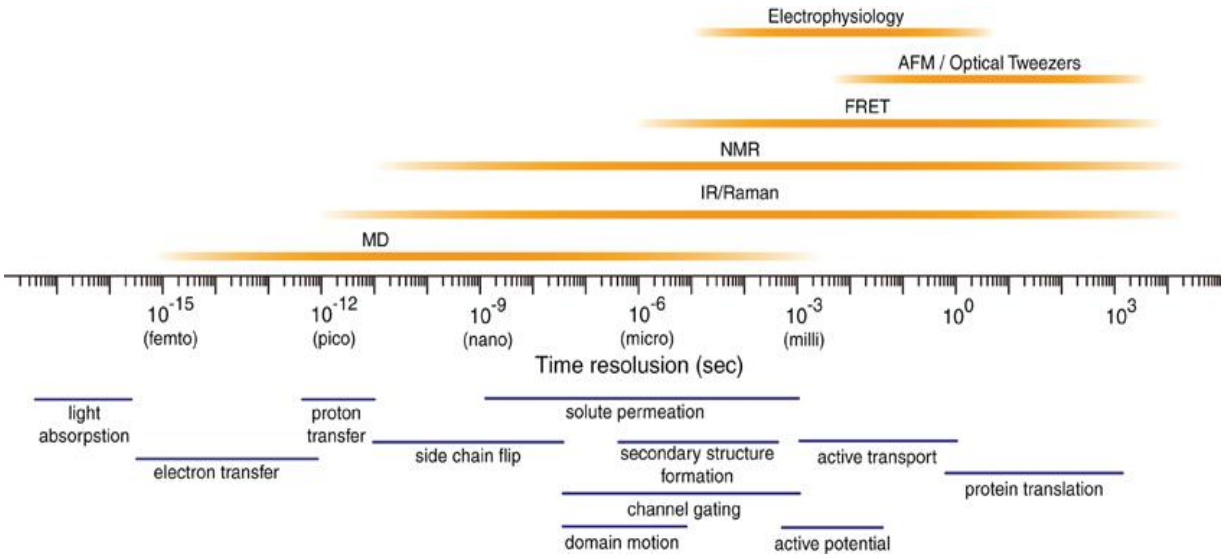


Figure 1 : Résolution temporelle dans différentes techniques biophysiques. Tirée de (Ode *et al.* 2012).

1. Bref historique

Les premiers travaux de simulations moléculaires remontent aux années 1950 par Metropolis et ses collaborateurs pour étudier les propriétés d'un liquide simple représenté par un modèle de disques durs en deux dimensions (Metropolis *et al.* 1953). Cette méthode a donné le nom de Monte-Carlo (MC) Metropolis et est devenue la base des simulations MC reposant sur des techniques probabilistes. Plus tard, en 1956, Alder et Wainwright, ont posé les bases de la Dynamique Moléculaire (DM) actuelle, dite « déterministe », en réalisant une simulation de sphères dures avec une résolution des équations du mouvement (Alder *et al.* 1957). En ce qui nous concerne, les premières simulations moléculaires de systèmes

biologiques datent de la fin des années 70 avec les travaux pionniers de Martin Karplus avec l'étude d'une petite protéine, la BPTI (McCammon *et al.* 1977) avec 885 atomes, simulée pendant 2 ps. Au même moment, Michael Levitt et Arieh Warshel posent les bases des approches des simulations de mécaniques moléculaires et quantiques (MM/QM) pour étudier l'hydrolyse d'un modèle de liaison peptidique par la lysozyme (Warshel *et al*. 1976). Ces travaux fondateurs ont servis de base pour les simulations que nous connaissons et pour cela Karplus, Levitt et Warshel ont reçu en 2013 le prix Nobel de Chimie. D'année en année, avec l'augmentation des moyens informatiques et le développement d'algorithmiques de plus en plus efficace, la complexité et la taille des systèmes étudiés par simulation moléculaire ont accrues (Figure 2). Ainsi, il est possible de simuler actuellement de très gros systèmes comme le chromatophore de la bactérie pourpre *Rhodospirillum photometricum* (~23 millions d'atomes) (Chandler *et al.* 2014) ou un modèle de membrane photosynthétique chez *Rhodobacter sphaeroides* (~100 millions d'atomes) (Cartron *et al.* 2014). Enfin, tout récemment en 2015, des simulations gros grain (~31 millions particules) de ~5 µs d'une particule de virus grippal sur une membrane de mammifère ont été réalisées (Reddy *et al.* 2015).

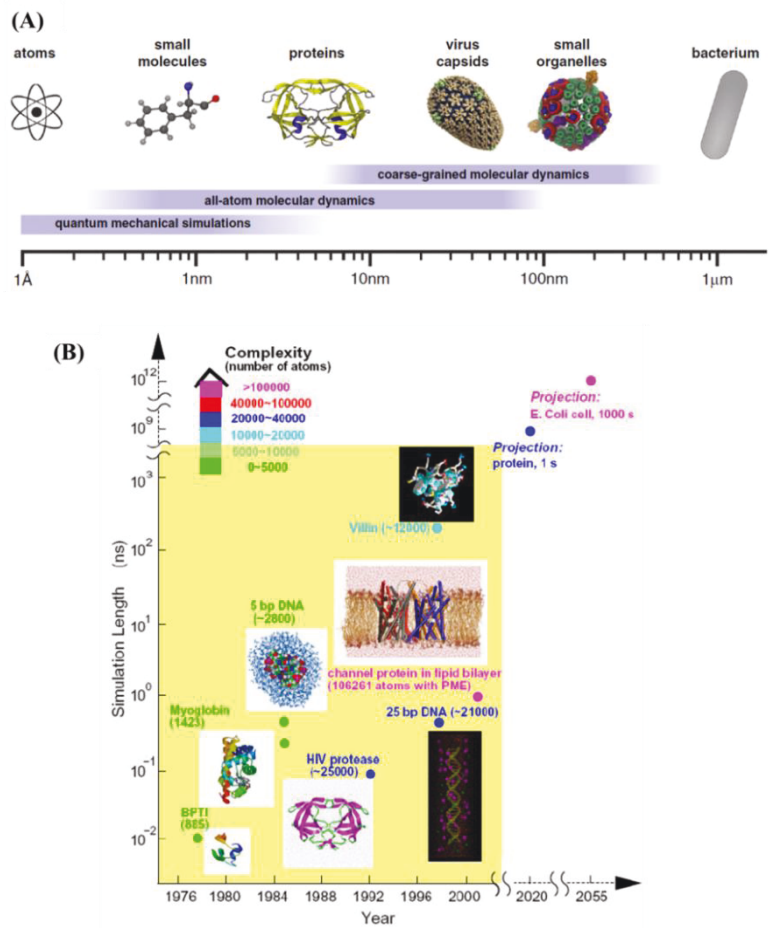


Figure 2 : Échelle de taille selon différentes simulations moléculaires (A) tirée de (Perilla *et al.* 2015), et l'évolution des dynamiques moléculaires par rapport à la taille et au temps de simulation, tirée du cours du Dr. Prabal K Maiti.

2. Les simulations moléculaires

Les biomolécules sont constituées de centaines voire de milliers d'atomes et l'étude de ces systèmes, composés de différentes particules (protons, neutrons et électrons), nécessiterait la résolution de l'équation de Schrödinger dépendant du temps. Or, cette dernière est, à l'heure actuelle, hors de portée pour étudier des systèmes supérieurs à plusieurs centaines d'atomes. Pour aller au-delà de cette limite, la mécanique moléculaire (MM) ou mécanique classique est utilisée et repose sur l'approximation de Born-Oppenheimer (Born *et al*. 1927) : les mouvements des électrons autour des noyaux peuvent être séparés, car leur masse est bien plus importante que celles des électrons. Dans le cadre de cette approximation, la modélisation d'un système moléculaire est décrite par un ensemble de points de masse fixes dont l'effet des électrons est décrit par une fonction d'énergie potentielle. Des approches mixtes combinant les approches de mécanique

quantique et classique, peuvent être aussi utilisées pour combiner la « qualité » de la QM (meilleure description des interactions entre particules) et la « rapidité » des simulations de dynamiques moléculaires. Elles sont utilisées pour décrire par exemple les réactions enzymatiques ou les évènements liés au transfert d'électrons dans le cas des réactions liées à la photosynthèse. Dans ce cas, les atomes appartenant au site catalytique, le ligand, le chromophore, etc. et son proche environnement situé à une distance de quelques Å sont décrits complètement (i.e. en tenant compte de l'ensemble des particules, incluant les électrons) à l'aide d'une approche quantique, et le reste du système (par exemple protéine et solvant) par une approche classique de mécanique moléculaire.
Dans ce qui suit, nous décrierons les principes et méthodes utilisés pour réaliser des simulations de dynamiques moléculaires classiques.

II. Représentations des interactions entre atomes

1. La fonction d'énergie potentielle ou champ de force

Contrairement à la mécanique quantique, la mécanique moléculaire ne tient pas compte explicitement de l'environnement électronique et les atomes sont représentés par des masses ponctuelles chargées. Les interactions interatomiques sont décrites en fonction du temps en résolvant l'équation du mouvement (seconde loi de Newton) :

$$F_i = m_i . a_i \qquad \text{(II.1)}$$

où F_i sont les forces exercées sur l'atome i de masse m_i et a_i son accélération.

À chaque pas de temps, les nouvelles positions et vitesses de chaque particule sont calculées en fonction des interactions auxquelles elles sont soumises. Ces interactions sont modélisées de manière empirique (issus des expériences) grâce à une fonction d'énergie potentielle (ou champ de force). Un exemple de fonction d'énergie est donné dans la Figure 3 où l'énergie potentielle du système E_{total} est égale la somme des interactions intramoléculaires dites liantes ($E_{liantes}$), et intermoléculaires ou non liantes ($E_{non\text{-}liantes}$) :

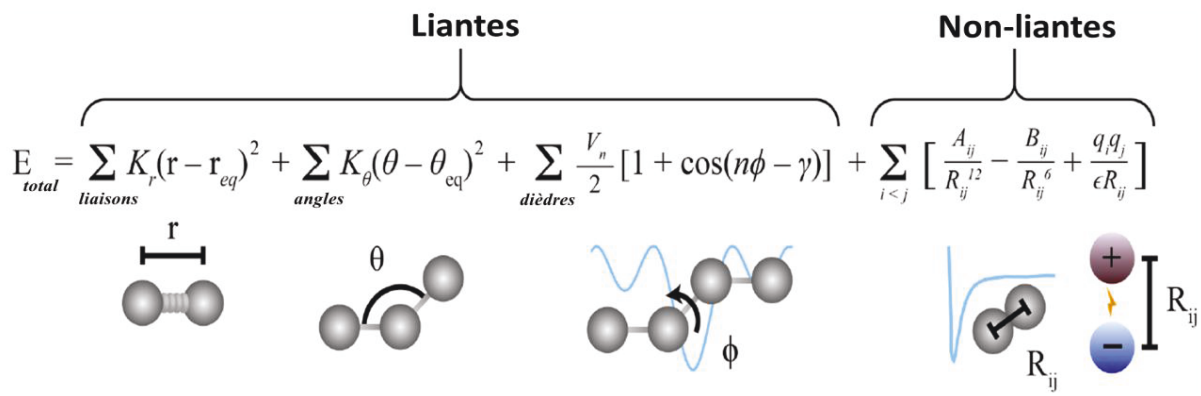


Figure 3 : Un exemple des différents termes d'un champ de force. Adaptée de (Durrant *et al.* 2011).

Les différents termes de la Figure 3 seront décrits plus en détails dans les sections suivantes. Le calcul de l'énergie potentielle E_{total} est la base de la dynamique moléculaire puisque les forces F_i localisées sur un atome i sont la dérivée de l'énergie du système par rapport à la position i de l'atome :

$$F_i = -\frac{\partial E}{\partial r_i} = m_i . a_i = m_i . \frac{d^2 r_i}{dt^2} \qquad \text{(II.2)}$$

2. Les interactions intramoléculaires ou interactions liantes

Intéressons-nous maintenant à la première partie de la fonction d'énergie potentielle qui correspond aux interactions intramoléculaires ou dites « liantes » dépendantes :

$$E_{liés} = \sum_{liaisons} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dièdres} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \qquad \text{(II.3)}$$

Elles sont définies par trois types d'interactions : les interactions d'élongation des liaisons covalentes entre deux atomes (« *stretching*»), de courbure des angles de valence (« *bending* ») entre deux liaisons covalentes successives, et de torsion des angles dièdres entre quatre atomes successifs. Les deux premiers termes sont généralement modélisés par des potentiels harmoniques, comme dans l'équation II.3, où K_r et K_θ représentent les constantes de force de liaison associées aux liaisons et aux angles de valence, respectivement, et r_{eq} et θ_{eq} sont leur valeur d'équilibre respective. Dans le cas des liaisons covalentes impliquant l'atome d'hydrogène, les fréquences de vibration étant très élevées, ces liaisons sont modélisées par des liaisons rigides fixées à leur valeur d'équilibre grâce à la méthode des contraintes (Ciccotti *et al.* 1986). Le dernier terme est souvent modélisé par un potentiel périodique avec un développement en série de Fourrier où V_n est la hauteur de la barrière d'énergie de rotation, ϕ l'angle dièdre, n la périodicité de rotation et γ l'angle de

phase. Par exemple, pour n=1, n=2 et n=3, la torsion sera périodique par 360°, 180° et 120°, respectivement. Ainsi, le premier ordre de la série de Fourier (n=1) va agir sur la stabilisation des formes de conformation *trans* et *cis*, le deuxième ordre (n=2) sur la stabilisation planaire des doubles liaisons entre atomes de type sp^2, et le troisième ordre (n=3) sur la stabilisation des formes de conformations décalées ou éclipsées. Ceci est illustré dans le cas simple de la molécule d'éthane (Figure 4) pour laquelle l'énergie de torsion autour de la liaison Csp^3-Csp^3 est définie avec n = 3 et γ = 0°. La courbe présente trois minima aux angles 60°, 180° et 300° représentant les conformations décalées, et trois maxima aux angles 0°, 120° et 240° pour les conformations éclipsées des différents rotamères.

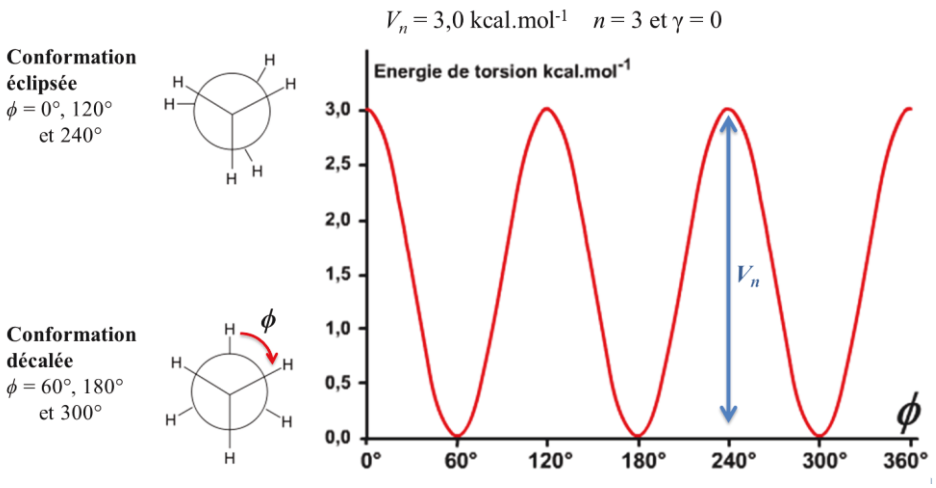


Figure 4 : Variation de l'énergie de torsion du dièdre H-C-C-H de l'éthane.

Parfois, il est aussi nécessaire d'imposer que certains atomes se trouvent dans le même plan. C'est le cas par exemple des liaisons peptidiques et des cycles aromatiques, on ajoute alors un terme additionnel à la fonction d'énergie appelé angle dièdre « impropre » qui est modélisé avec une fonction harmonique :

$$E_{impropre} = \sum_{impropre} K_{\omega}(\omega - \omega_{eq})^{2} \qquad \text{(II.4)}$$

où K_{ω} est la constante de force, ω_{eq} une valeur de référence.

Toutes les valeurs de références et les constantes de force utilisées dans le calcul des interactions liantes sont basées sur des données expérimentales (par exemple cristallographique, spectroscopie vibrationnelle, etc.).

3. Les interactions intermoléculaires ou interactions non liantes

La deuxième partie de la fonction de l'énergie potentielle concerne les interactions intermoléculaires, ou non liantes, appartenant aux atomes de molécules éloignées ou aux atomes d'une même molécule mais séparés entre eux par 3 liaisons covalentes consécutives (on parle couramment d'interaction « 1-4 »). Dans la plupart des champs de force, ces interactions sont représentées par une somme de deux termes de tous les couples d'atomes *i* et *j*. Ces termes comprennent les interactions de courte portée, dites de van-der-Waals (E_{vdW}), qui sont généralement modélisées sous la forme d'un potentiel de Lennard-Jones (E_{LJ}), et les interactions de longue portée représentées par un potentiel électrostatique coulombien ($E_{élec}$) :

$$E_{non-liante} = E_{vdW} + E_{élec} \quad \text{(II.5)}$$

$$E_{non-liante} = \sum_{i<j}\left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right) + \sum_{i<j}\left(\frac{q_i q_j}{\epsilon R_{ij}}\right) \quad \text{(II.6)}$$

où A_{ij} et B_{ij} sont les paramètres du potentiel de Lennard-Jones (Allen and Tildesley 1989), R_{ij} est la distance entre les atomes *i* et *j*, q_i et q_j sont les charges partielles des atomes i et j obtenues par calcul quantique *ab initio* ou semi-empirique, et ϵ est la permittivité relative ou constante diélectrique.

- ***Le potentiel de Lennard-Jones***

Les interactions de courte portée pour des atomes séparés par plus de deux liaisons covalentes sont modélisées avec le potentiel de Lennard-Jones dit en « 6-12 » qui est la résultante des forces répulsives dues à l'exclusion de Pauli (Pauli 1925) et des forces attractives dites de London (London 1930) :

$$E_{LJ} = \sum_{i<j}\left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right) = \sum_{i<j}\left[\left(\frac{(4\sigma_{ij}\varepsilon_{ij})^{12}}{R_{ij}^{12}}\right) - \left(\frac{(4\sigma_{ij}\varepsilon_{ij})^{6}}{R_{ij}^{6}}\right)\right] \quad \text{(II.7)}$$

$$E_{LJ} = \sum_{i<j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}}\right)^{6}\right] \quad \text{(II.8)}$$

où R_{ij} est la distance entre les atomes i et j, σ_{ij} est la distance entre les atomes i et j où $E_{LJ} = 0$, et ε_{ij} est le puits d'énergie qui est égale à $2^{1/6}\sigma_{ij}$ (ou largeur d'interaction dipolaire).

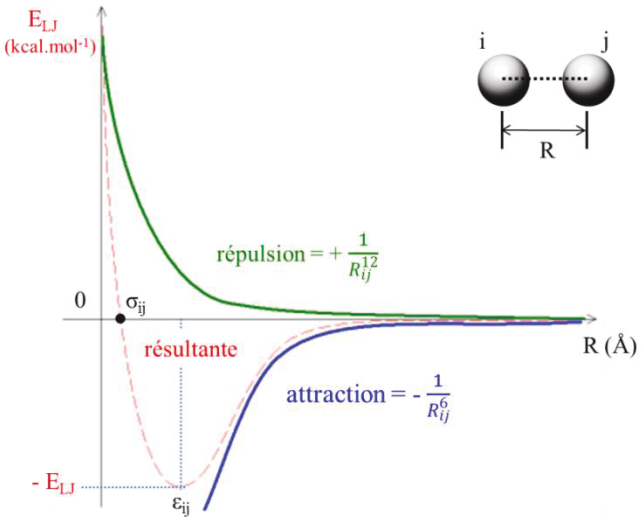


Figure 5 : Représentation du potentiel de Lennard-Jones (ou interaction 6-12) pour un couple d'atomes ij.

On exprime souvent le terme répulsif par une hyperbole en $1/r^n$ avec n compris entre 9 et 12, avec r la distance de séparation des deux atomes. Les termes à $1/r^{12}$ et $1/r^6$ reproduisent avec une bonne approximation la partie répulsive et attractive du potentiel (Figure 5). Dans le cas où deux atomes i et j sont de nature différente, on applique les règles de combinaisons de Lorentz-Berthelot pour calculer le rayon de collision σ_{ij} et le puits de potentiel ε_{ij} :

$$\sigma_{ij} = \frac{(\sigma_{ii} + \sigma_{jj})}{2} \text{ et } \varepsilon_{ij} = \sqrt{(\varepsilon_{ii}\varepsilon_{jj})} \qquad \text{(II.9)}$$

- ***Le potentiel coulombien***

Les interactions de longue portée entre une paire d'atomes i et j sont modélisées par le potentiel coulombien qui prend en compte la distance entre les deux atomes (R_{ij}), leur charge respective (q_i et q_j) (Figure 6) ainsi que la constante diélectrique ϵ :

$$E_{élec} = \sum_{i<j}\left(\frac{q_i q_j}{\epsilon R_{ij}}\right) = \sum_{i<j}\left(\frac{q_i q_j}{4\pi\epsilon_0\epsilon_r R_{ij}}\right) \qquad \text{(II.10)}$$

où ϵ_0 est la permittivité diélectrique du vide et ϵ_r, la permittivité diélectrique relative au milieu.

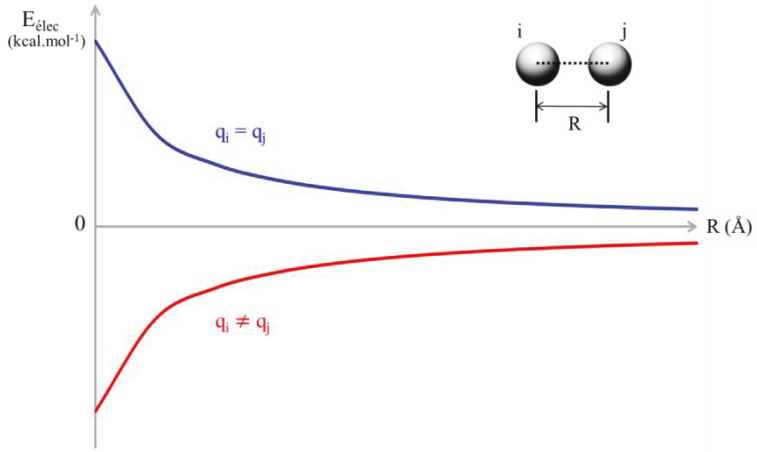


Figure 6 : Représentation du potentiel coulombien avec ϵ_r égale à 80 (eau) et avec des charges de même signe (en bleu) ou des charges de signe opposé (en rouge).

Notons que l'utilisation de la loi de Coulomb avec des charges fixes pour traiter les interactions électrostatiques est une grande simplification des champs de force additifs. En effet, la distribution des charges des atomes étant fixée, leur distribution est indépendante des fluctuations du champ électrique environnant. Pour compenser, les charges atomiques sont augmentées de telle manière à reproduire la polarisation des atomes dans un environnement polaire en phase condensée (par exemple en phase aqueuse, le model d'eau TIP3P à un moment dipolaire de l'ordre de 2,35 D en comparaison avec sa valeur en phase gazeuse, 1,85 D) (Jorgensen *et al.* 1983). Cette approximation fonctionne bien pour les simulations courantes, mais peut être défaillante dans le cas d'environnement très chargé. Pour remédier à ce problème, différentes alternatives sont utilisées comme les approches dites « de charges fluctuantes » qui modifient le potentiel du champ de force non-polarisable au niveau du potentiel coulombien. Ainsi, l'interaction coulombienne varie en fonction de l'environnement. Citons par exemple l'approche de l'oscillateur de Drude (basée sur le modèle d'électron mobile (Harder *et al.* 2006)) où la modélisation de la polarisation est faite par l'intermédiaire de dipôles, grâce à des paires de charges ponctuelles reliées par des ressorts harmoniques. Dans le cas de la molécule d'eau, la charge fictive est reliée à l'atome d'oxygène par un ressort. La charge totale de la molécule est assurée par le fait que la charge de l'atome d'oxygène soit réduite par la valeur de la charge portée par la particule fictive.

Les champs de force classiques utilisés pour la simulation des protéines et des acides nucléiques, tels que AMBER (*Assisted Model Building with Energy Refinement*) (Cornell *et al.* 1995), OPLS (*Optimized Potentials for Liquid Simulations*)(Jorgensen *et al.* 1988), CHARMM (*Chemistry at HARvard Macromolecular Mechanics*)(MacKerell *et al.* 1998) ou

GROMOS (*GROningen Molecular Simulation*) (Schmid *et al.* 2011), reproduisent efficacement les propriétés structurales et thermodynamiques d'un système dans la plage des valeurs de température et de pression où elles ont été paramétrées. Ces champs de force diffèrent par leurs paramètres « liants » et « non-liants » qui interviennent dans l'expression de l'énergie potentielle. Les champs de force se divisent en trois grands groupes selon le niveau de détail moléculaire : tout-atome qui traite tous les atomes explicitement (AMBER, CHARMM, OPLS-AA (Jorgensen *et al.* 1996), etc.), atome-unifié qui regroupe les atomes liés aux atomes d'hydrogène apolaire dans une particule (GROMOS, etc.) et enfin gros-grain qui regroupe plusieurs atomes (généralement 4) dans une unique particule (MARTINI (Marrink *et al.* 2007) etc.).

4. Le code de simulation moléculaire GROMACS

Pour ce travail de thèse, nous avons essentiellement utilisé le code GROMACS (GROningen Machine for Chemical Simulations, http://www.gromacs.org/) (Abraham *et al.* 2015), pour nos simulations et analyses. GROMACS est une suite de programme de simulation et d'analyse sous licence libre et hautement parallélisé, développé initialement au début des années 90 par le groupe d'Herman Berendsen à l'université de Gröningen (Pays-Bas). GROMACS utilise des algorithmes très efficaces permettant la simulation de larges systèmes (bio)moléculaires (plusieurs centaines de millions d'atomes) sur des calculateurs hautement parallèles. GROMACS permet aussi l'utilisation des GPU *boostant* significativement les performances du code.

III. Les techniques de simulation en dynamique moléculaire

Pour simuler un système moléculaire dans les conditions les plus réalistes possibles, on emploie différentes approches que nous décrierons brièvement dans les sections suivantes.

1. Les conditions périodiques aux bords

Les systèmes étudiés en DM sont de taille finie alors que dans l'expérience, ils sont intégrés au sein de systèmes beaucoup plus vastes, dont la simulation totale par les méthodes numériques est complètement impossible. Cependant, les propriétés du système microscopique peuvent être corrélées avec celles de la phase macroscopique en s'affranchissant des effets de bords. En conséquence, de manière à pouvoir simuler des

systèmes « infinis » (liquides, cristaux, etc.), il est possible de modéliser ces derniers par des systèmes de taille finie, délimités par une cellule unitaire de différentes formes géométriques (cubiques, octaédriques, sphériques…) et répliquée (virtuellement) à l'infini selon les trois directions de l'espace, qu'on appelle les conditions périodiques aux bords ou PBC (*Periodic Boundary Conditions*)(Figure 7). Avec les conditions périodiques aux bords, si une particule/molécule *i* quitte l'un des bords du domaine de simulation (boîte centrale **a**), cette particule/molécule rentre simultanément dans la boîte par celui d'en face avec la même quantité de mouvement (voir Figure 7A, flèches noires).

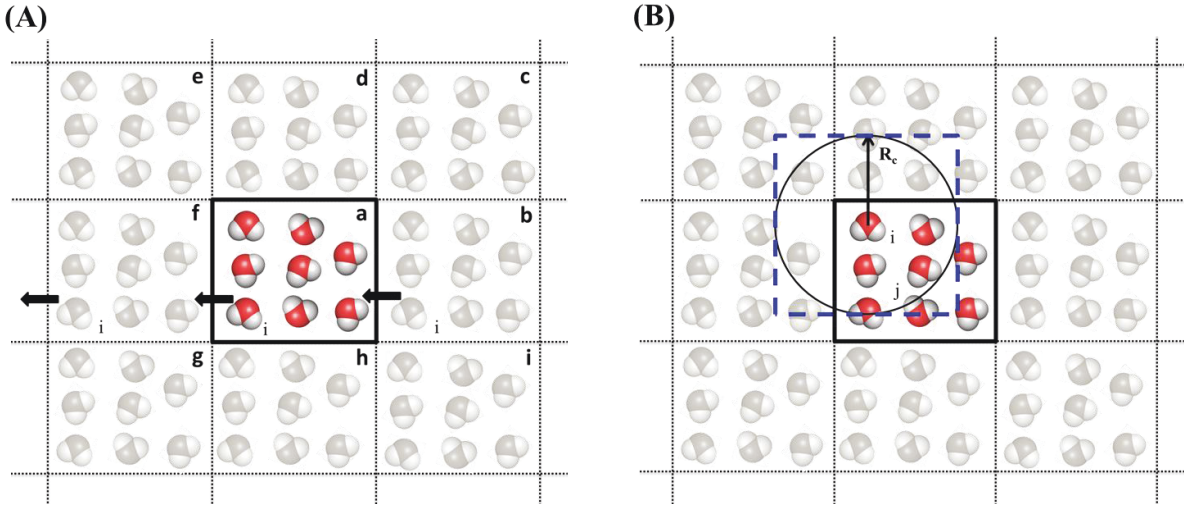


Figure 7 : Représentation bidimensionnelle d'une boîte de simulation répliquée dans les trois dimensions de l'espace des PBC (A) et une représentation de la sphère de troncature (B).

Les conditions périodiques aux bords contraint de faire des approximations pour le traitement des interactions entre molécules. Notamment, l'approximation dite de « l'image minimale » (Figure 7B, carré pointillé en bleu,) suppose que chaque particule *i* de la boîte de simulation, interagit avec l'image la plus proche de toutes les autres particules *j*. Le calcul des interactions pour la particule *i* située dans la boîte de simulation s'effectue à partir des contributions des autres atomes de la même boîte ou des boîtes avoisinantes, situées à l'intérieur de la sphère de troncature R_c ou *cut-off* (voir Figure 7B, cercle noir). Au-delà de cette distance R_c, les interactions ne sont pas explicitement prises en compte. Il est important de préciser qu'il faut éviter qu'un atome puisse interagir simultanément avec sa propre image. C'est pourquoi, la distance de troncature R_c est définie comme étant la moitié du plus petit côté de la boîte de simulation. Dans les simulations de fluides, la position des atomes voisins ne change pas significativement au-delà de 10 - 20 pas de simulation. De ce fait, pour gagner du temps, on utilise la méthode dite des « *non-bonded neighbour list* » proposé

initialement par Verlet en 1967 (Verlet 1967). Cette méthode consiste à stocker, dans un tableau indicé appelé aussi liste de voisin de Verlet, la position des atomes voisins *j* d'un atome *i* qui sont à l'intérieur de la troncature. Cette liste est remise périodiquement à jour en fonction de la distance que parcourt un atome dans le système périodique.

Notons que bien que la boîte cubique est fréquemment utilisée dans les simulations, d'autres formes géométriques sont également possibles pour répliquer de façon infinie un système périodique (par exemple hexagonal, octaèdre tronquée, etc.). Le choix du type de boîte dépend du système à simuler pour réduire le nombre d'atome dans la boîte de simulation.

2. Calculs des interactions électrostatiques

L'introduction d'un rayon de troncature permet de réduire considérablement le temps de calcul mais il faut tout de même estimer les énergies d'interactions entre les paires d'atomes *i* et *j*. Ces interactions seront d'autant plus valables que la portée des interactions sera réduite. À l'inverse, dans le cas des interactions de longues distances, le système de sphère de troncature sera vraisemblablement erroné car il peut entraîner une discontinuité de l'énergie au voisinage de R_c. Pour contourner ce problème, différentes approches ont été développées. La plus utilisée pour les systèmes périodiques que nous simulons couramment est l'approche basée sur les travaux de Paul Peter Ewald (Ewald 1921), qui consiste à évaluer les interactions d'une molécule avec toutes les autres situées dans la boîte centrale, ainsi que dans toutes les boîtes images. Ewald a introduit une approche que l'on nomme la somme d'Ewald qui consiste à séparer le potentiel d'interactions en deux termes U_0 à courte portée et U_1 à longue portée qui convergent rapidement :

$$E_{élec} = \mathrm{U}_0 + \mathrm{U}_1 \qquad \text{(III.1)}$$

La distribution de charge auxiliaire est choisie de manière à rendre les termes U_0 à courte portée dans l'espace direct, et le terme U_1 périodique dans l'espace réciproque. Ces espaces peuvent être définis avec l'identité suivante :

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1-f(r)}{r} \qquad \text{(III.2)}$$

Dans la sommation d'Ewald, chaque particule chargée i est entourée par une distribution de charge auxiliaire $\rho_i(r)$ afin d'en rendre la charge nulle. Cette distribution est généralement représentée par une distribution gaussienne centrée sur chaque site d'interaction (Figure 8).

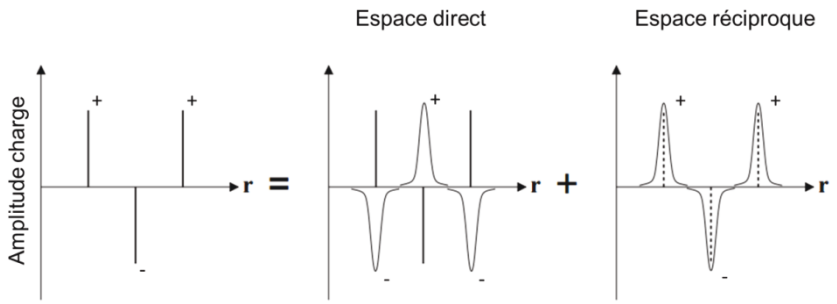


Figure 8 : Composantes d'une somme d'Ewald où chaque charge initiale est entourée par une distribution de charge gaussienne α dans l'espace réel, d'amplitude égale mais de signe opposé. Cette contribution est contre balancée dans l'espace réciproque par une gaussienne ρ(r) de signe contraire.

La méthode d'Ewald est couteuse en temps de calcul en fonction du nombre de particule (sa complexité est l'ordre N^2, ou N est le nombre de particules du système). D'autres approches ont été proposées pour améliorer la performance de la somme dans l'espace réciproque grâce à une transformée de Fourier rapide (FFT), comme par exemple la méthode dite « *Particle Mesh Ewald* » ou PME développée par Darden et Essman (Darden *et al.* 1993)(Essmann *et al.* 1995). Elle consiste, comme la méthode d'Ewald « classique », à calculer l'énergie électrostatique finale des particules grâce à une interpolation des charges sur une grille tridimensionnelle (Figure 9).

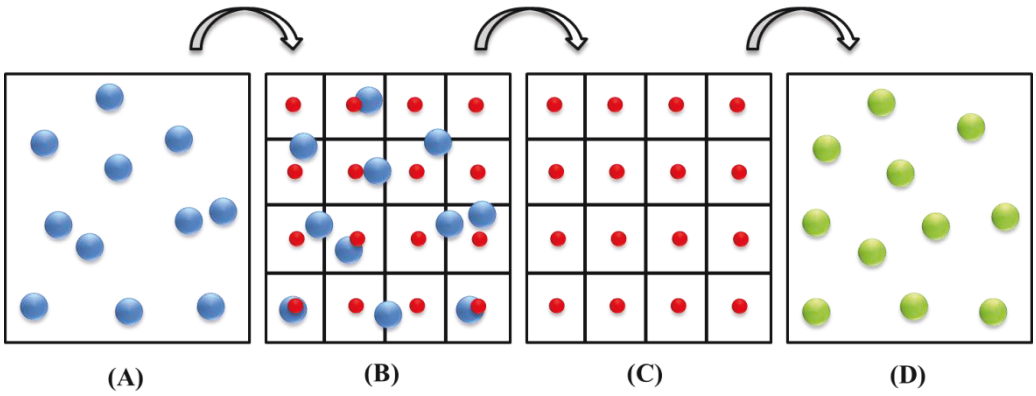


Figure 9 : Représentation de la méthode PME sur une grille bidimensionnelle. Soit (A), un système de particules chargées. Ces charges sont interpolées sur une grille bidimensionnelle (B) où le potentiel et les forces sont calculés sur chaque point de la grille (C). Puis, les forces sont interpolées en retour vers les particules, dont la position est mise à jour.

Avec cette approche, le gain numérique est considérable puisqu'elle fait passer la dépendance en N de $N^{3/2}$ à Nlog(N).

3. Les ensembles thermodynamiques et moyennes statistiques

- ***La notion d'ensemble***

La physique statistique (ou mécanique statistique) et les simulations numériques permettent de définir des grandeurs thermodynamiques macroscopiques (comme la pression, l'énergie libre, l'enthalpie, etc.) à partir de la modélisation de systèmes à l'échelle microscopique. En physique statistique, c'est la connaissance de la fonction de partition $\boldsymbol{Q}$ d'un système qui permet d'accéder à l'ensemble des grandeurs thermodynamiques. L'apport des simulations numériques est de permettre le calcul des moyennes d'ensemble sans avoir à calculer la fonction de partition explicitement. En mécanique statistique, il existe plusieurs ensembles thermodynamiques accessibles *via* les simulations numériques :

- ensemble NVE ou *microcanonique* : le nombre de particules, le volume et l'énergie restent constants
- ensemble NVT ou *canonique* : le nombre de particules, le volume et la température restent constants
- ensemble NPT ou *isobare-isotherme* : à nombre de particules, pression et température constants
- ensemble μVT ou *grand-canonique* : à potentiel chimique (μ), volume et température constants.

Par exemple, si l'on définit l'Hamiltonien du système où $H = E_{cin} + U$ est la somme de l'énergie cinétique et de l'énergie potentielle du système, la fonction de partition dans l'ensemble microcanonique est :

$$Q(N,V,E) = \sum_{\alpha} \delta(H_{\alpha} - E) \qquad \text{(III.3)}$$

où la somme est écrite sur tous les états α possibles du système. On écrit ainsi que seuls les états α qui ont un Hamiltonien H_{α} égale à l'énergie E sont des états permis du système. La grandeur thermodynamique naturelle de l'ensemble microcanonique est l'entropie S qui est définie comme :

$$S = k \ln Q(N,V,E) = k ln \sum_{\alpha} \delta(H_{\alpha} - E) \qquad \text{(III.4)}$$

où k est la constante de Boltzmann.

Les équations de Newton de la mécanique classique conservent l'énergie totale : la mécanique Newtonienne est donc la méthode naturelle pour générer les états α de l'ensemble microcanonique. Nous verrons plus loin quelles méthodes peuvent être employées pour simuler dans les ensembles *canoniques* et *isobare-isotherme.*

- ***Les moyennes statistiques***

Par définition de la mécanique statistique, les variables conjuguées aux grandeurs constantes de l'ensemble fluctuent. Pour l'ensemble microcanonique, il s'agit de la pression (variable conjuguée de V), de la température (variable conjuguée de E) et du potentiel chimique (variable conjuguée de N). Soit A un observable d'un système constitué de N particules dépendantes de leurs vecteurs position, r^N et impulsion p^N. L'ensemble des variables $\{r^N, p^N\}$ représente un point dans l'espace des phases. Cet espace des phases est un espace multidimensionnel de 6N dimensions (3N cordonnées et 3N vitesses) qui décrit l'ensemble des états microscopiques possibles d'un système. L'ensemble des points dans l'espace des phases satisfait un certain état thermodynamique.

D'après le principe d'*ergodicité*, la valeur d'une grandeur moyennée au cours du temps A_t, coïncide avec la moyenne effectuée dans un ensemble statistique <A> :

$$\lim_{t\to\infty} \overline{A_t} = \langle A \rangle = \int\int drdp\, A(r^N, p^N)\, \rho(r^N, p^N) \qquad \text{(III.5)}$$

où $\rho(r^N, p^N)$ est la densité de probabilité sur l'ensemble des états. De ce fait, l'ensemble dont l'espace des phases est parcouru par la DM est a priori l'ensemble microcanonique (ou NVE) car les valeurs de N et du volume V ne changent pas lors de l'intégration de l'équation du mouvement, et que celles-ci conservent également l'énergie E. Dans ce système, la température et la pression ne sont pas fixées mais elles peuvent être estimées comme moyenne sur le temps des observables microscopiques (température ou pression) :

$$T = \langle T \rangle = \frac{2}{3Nk_b} \langle \sum_{i=1}^{N} \frac{p_i^2}{2m_i} \rangle \qquad \text{(III.6)}$$

$$P = \langle P \rangle = \frac{1}{3V} \langle \sum_{i=1}^{N} \left(\frac{p_i^2}{m_i} + r_i . E_i \right) \rangle \qquad \text{(III.7)}$$

Il est donc souvent plus intéressant de simuler dans des conditions où la température et/ou la pression sont fixées.

4. Simulation à pression et température constantes

Pour simuler d'autres ensembles que l'ensemble NVE, il faut donc coupler le système considéré à un autre système (appelé « réservoir » ou « bain ») virtuellement infini. Il est possible de faire cela en utilisant la méthode du système étendu introduite par Andersen en 1980 (Andersen 1980). Elle consiste à écrire de nouvelles équations du mouvement avec un degré de liberté additionnel (réservoir) couplé uniformément aux degrés de liberté originaux.

- ***Le thermostat pour simuler à température constante***

La température T du système est reliée à l'énergie cinétique et donc aux vitesses par la formule :

$$E_{cinétique} = \sum_{i=1}^{N_{atomes}} \frac{1}{2} m_i v_i^2 = \frac{1}{2} N_{ddl} kT \qquad \text{(III.8)}$$

où m_i est la masse de l'atome i, v_i est la vitesse de l'atome i, N_{ddl} est le nombre de degré de liberté, *k* la constante de Boltzmann et T la température. Ainsi, pour maintenir la température du système autour de la température désirée, les algorithmes vont se baser sur l'ajustement des vitesses des particules. Dans GROMACS, trois thermostats sont couramment utilisés pour simuler à température constante. La méthode basée sur le couplage à un bain thermique « *weak coupling bath* » de Berendsen (Berendsen *et al.* 1984) qui est la plus simple et est très efficace pour atteindre très rapidement la température voulue et utile pour équilibrer un système. Cependant, ce thermostat supprime les fluctuations de l'énergie cinétique ce qui signifie que l'on ne génère pas l'ensemble strictement canonique. Ce problème peut être résolu en multipliant toutes les vitesses par un terme stochastique $(E_{voulue}/E_{système})^{1/2}$ qui assure une distribution correct de l'énergie cinétique, appelé le thermostat de l'ajustement des vitesses « *velocity-rescaling* » (Bussi *et al.* 2007). Le troisième thermostat est l'approche de Nosé-Hoover (Nosé 1984)(Hoover 1985) qui se base sur l'approche du système étendu. L'Hamiltonien H décrivant le système est alors « étendu » en introduisant un terme de frottement ζ dans l'équation du mouvement :

$\zeta = (E_{système}-E_{voulue})/Q_T$ où Q est la masse du réservoir.

Pratiquement, au début de chaque simulation, pour que la température du système soit fixée à la température désirée, des vitesses initiales sont assignées sur chaque atome du système selon une distribution de Maxwell-Boltzmann. Et au cours de la simulation, les fluctuations

de la température du système sont contrôlées pour rester autour de la valeur moyenne $\langle T \rangle$ à l'aide des thermostats mentionnés ci-dessus.

- ***Le barostat pour simuler à pression constante***

Dans le même esprit du maintien de la température à une valeur constante, il est possible de coupler le système à un « bain de pression ». Dans une simulation à pression constante le volume est une variable dynamique où le volume de la boîte de simulation peut changer en fonction du temps (et par conséquence la pression). Il existe plusieurs approches pour ajuster la pression : l'isotropie, où le volume change mais pas la forme de la boîte (Figure 10A), la semi-isotropie où seuls les axes x et y changent de façon concertée (Figure 10B) et l'anisotropie, où tous les paramètres de la boîte peuvent changés indépendamment (Figure 10C). Dans le cas particulier d'une simulation de membrane, par exemple, on peut coupler le système à une pression semi-isotropique ou anisotropique pour permettre des fluctuations de l'aire de la membrane (Figure 10).

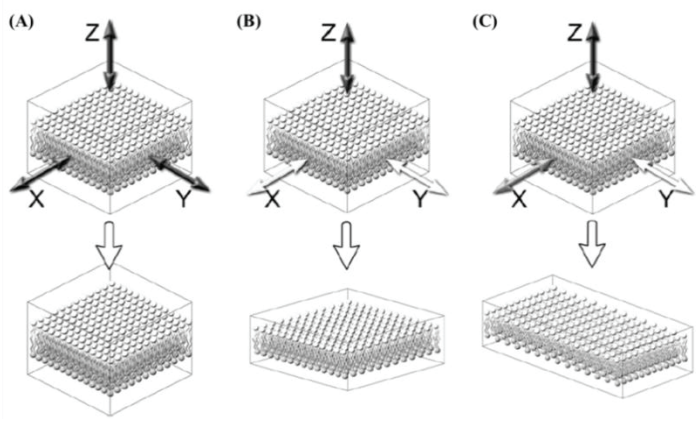


Figure 10 : Représentation des méthodes de couplages de la pression. (A) isotropique, les axes X, Y et Z sont couplés, (B) semi-isotropique, les axes X et Y sont couplés et (C) anisotropique ne couple aucune direction (Kandt *et al.* 2007).

Dans ce travail, avec GROMACS, nous avons utilisé la méthode de Berendsen (Berendsen *et al.* 1984), similaire au thermostat de Berendsen, qui ajuste les coordonnées et les vecteurs de la boîte à chaque pas. La méthode Parrinello-Rahman (Parrinello *et al*.1981)(Rahman *et al.* 1971) se base sur un système étendu avec neuf degrés de liberté supplémentaire qui correspondent aux trois vecteurs définissant la boîte de simulation. Comme pour le thermostat de Nosé-Hoover, le barostat de Parrinello-Rahman génère exactement l'ensemble NPT.

Pratiquement, si la pression est très loin de l'équilibre, la méthode Parrinello-Rahman entraînera de grandes oscillations des paramètres de la boîte. C'est pourquoi, il est préférable d'équilibrer au préalable avec la méthode de Berendsen pour se rapprocher de l'équilibre et d'utiliser ensuite la méthode de Parrinello-Rahman.

5. L'intégration des équations du mouvement

En dynamique moléculaire, l'évolution temporelle d'un système mécanique est obtenue en intégrant les équations du mouvement de Newton. Pour ce faire, nous avons utilisé l'algorithme de *leap-frog* (Hockney *et al.* 1974)(Figure 11), qui dérive de l'algorithme de Verlet, et qui permet de calculer les positions atomiques $\mathbf{r}_i$ à tous les temps t, et toutes les vitesses atomiques $\mathbf{v}_i$ aux temps intermédiaires $t+\frac{\Delta t}{2}$:

$$r_i(t+\Delta t) = r_i(t) + v_i\left(t+\frac{\Delta t}{2}\right)\Delta t \qquad \text{(III.9)}$$

$$v_i\left(t+\frac{\Delta t}{2}\right) = v_i\left(t-\frac{\Delta t}{2}\right) + a_i(t)\Delta t \qquad \text{(III.10)}$$

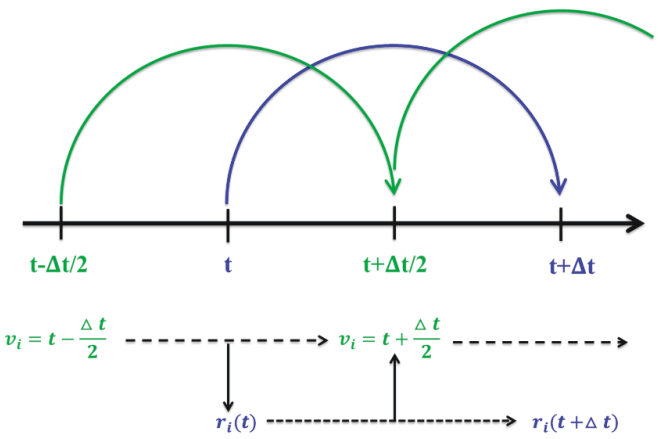


Figure 11 : Principe de l'algorithme de *leap-frog*.

Le choix du pas de temps Δt est un paramètre clef des algorithmes d'intégration. Un pas de temps trop court entraîne un temps de simulation trop court. À l'inverse, un pas de temps trop grand impliquera une simulation instable du fait d'une augmentation importante de l'énergie potentielle entre chaque pas d'intégration (Figure 12).

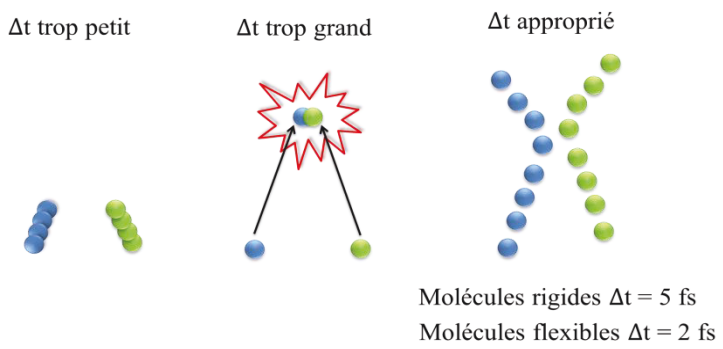


Figure 12 : Représentation des différents pas de temps d'une simulation tout atome.

En général, le pas de temps doit être plus petit que les mouvements de plus haute fréquence du système, typiquement celui des atomes d'hydrogène. En pratique, on contraint ces mouvements de haute fréquence à leur valeur d'équilibre avec un algorithme comme SHAKE (Ryckaert *et al.* 1977) ou LINCS (Hess *et al.* 1997) pour ainsi augmenter le pas d'intégration et aussi gagner en temps de calcul. Dans nos simulations, nous avons utilisé un pas de temps de 2 fs, mais pour d'autres représentations (par exemple de type gros grain, des sites virtuels) du fait que les interactions entre particules et que le paysage énergétique (« *energy landscape* ») sont très simplifiés, de plus grand pas de temps (20 - 40 fs) peut être utilisé sans affecter la stabilité de la simulation.

6. Le volume de Voronoï

Durant cette thèse, pour calculer l'ensemble des surfaces et volumes, nous avons utilisé l'approche de Voronoï en utilisant le programme « *trjVoronoi* » développé au laboratoire par Massimo Marchi (Abel *et al.* 2012) basé sur les classes et les librairies Voro^{++} (Rycroft 2009). Elles consistent à calculer sur chaque configuration atomique du système, le volume de Voronoï de chaque atome (Figure 13) en utilisant la tessellation de Voronoï (Voronoi 1908). Le plus petit polyèdre défini par l'intersection de tous les plans bissecteurs représente le volume de Voronoï de l'atome, appelé aussi le polyèdre de Voronoï. Ces derniers arrangés périodiquement dans un ensemble de sites coïncident avec les mailles d'une cellule de Wigner-Seitz.

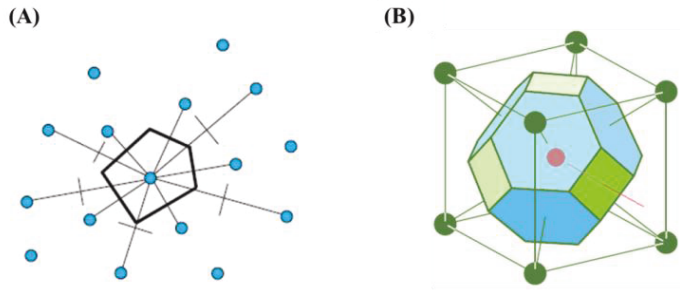


Figure 13 : Représentation en deux dimensions du polyèdre de Voronoï (A) et d'une cellule de Wigner-Seitz (B) donnant un octaèdre tronqué.

En général, seuls les atomes lourds sont pris en compte dans le calcul du volume car il a été montré que les volumes de Voronoï de tous les atomes ont une erreur négligeable (<0,5 % de différences). Cette approche diffère de celle implémentée par défaut dans GROMACS basée sur la surface de Connolly (Connolly 1983)(Figure 14). Dans cette méthode, chaque sphère atomique est représentée par le rayon de van-der-Waals de l'atome qui définit la surface de van-der-Waals. Une autre sphère *s* est « roulée » sur cette surface et recouvre les creux pour ainsi obtenir la surface de Connolly. Cependant, la précision de cette méthode dépend du rayon de la sphère qui va rouler.

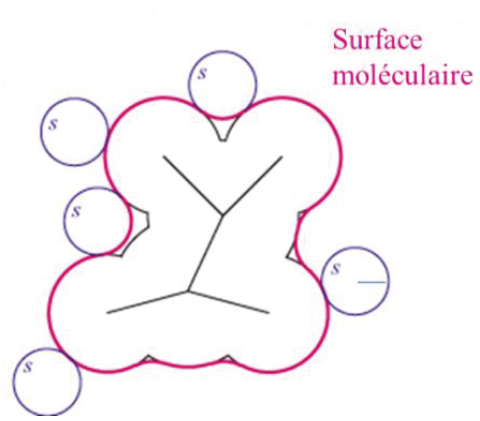


Figure 14 : Représentation de la surface de Connolly.

IV. Le calcul des charges partielles atomiques

Nous avons vu que la distribution de charge d'une molécule est simulée par des charges ponctuelles placées à l'endroit des noyaux atomiques. Quand un atome électriquement neutre se lie chimiquement à un autre atome plus électronégatif, ses électrons sont partiellement attirés par ce dernier le long de la liaison (Figure 15).

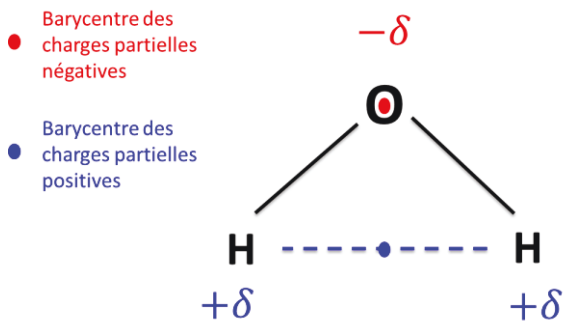


Figure 15 : Représentation des charges partielles dans le cas d'une molécule d'eau.

Ces charges partielles, qui interviennent dans le calcul des forces électrostatiques, sont évaluées à partir de calculs de chimie quantique sur de petites structures. Ces calculs donnent une approximation de la densité électronique de la molécule. Il existe plusieurs méthodes pour calculer ces charges, comme par exemple :

- Mulliken (Mulliken 1955) : basée sur l'analyse de la population de la fonction d'onde, qui divise en deux les populations de chevauchement entre les deux atomes d'une liaison. Cette méthode est assez simple mais dépend beaucoup de la base utilisée (cf. Figure 16) et par conséquent il arrive que les valeurs soient loin des valeurs réelles
- Gasteiger (Gasteiger *et al.* 1978) : empirique, basée sur l'électronégativité des atomes de la liaison
- RESP (*Restrained ElectroStatic Potential*)(Bayly *et al.* 1993) : attribue des charges ponctuelles au noyau des atomes pour reproduire au mieux le potentiel électrostatique. Cette méthode dépend beaucoup de la conformation de la molécule (Vanquelef *et al.* 2011).

Il a été montré que l'approche RESP donne de bons résultats pour simuler les biomolécules en phase condensée ainsi que pour les propriétés intermoléculaires (Dupradeau *et al.* 2008) (Bayly *et al.* 1993). Cette approche est aussi utilisée dans le champ de force AMBER.

- ***La méthode Hartree-Fock***

La méthode *ab initio* la plus simple est la méthode Hartree-Fock (HF) qui est une méthode de résolution approchée de l'équation de Schrödinger d'un système quantique, dans laquelle la répulsion coulombienne électron-électron n'est pas spécifiquement prise en compte. On choisit différentes fonctions gaussiennes pour décrire les orbitales atomiques et l'ensemble

de ces fonctions forme « la base » (Figure 16). On utilise généralement, une base composée d'un nombre fini d'orbitales atomiques centrées au niveau du noyau atomique.

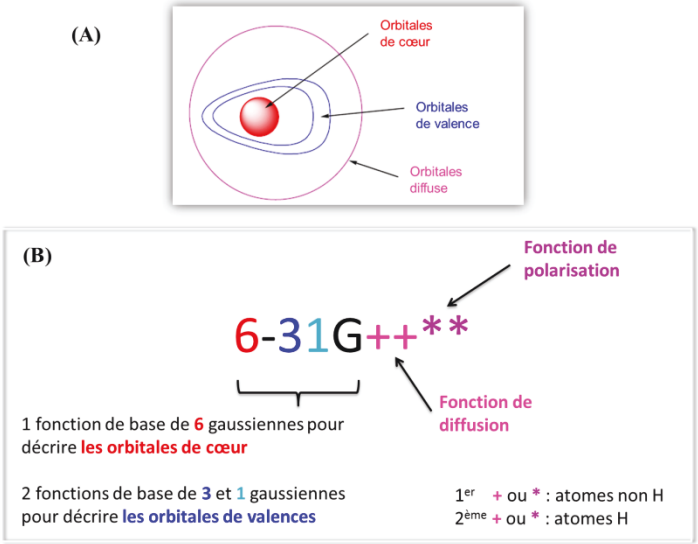


Figure 16 : Les zones à traiter dans la conception d'une base (A) et la notation de Pople avec les fonctions de polarisation et de diffusion (B).

Dans la plupart des liaisons chimiques, les électrons de valences (ou externes) sont plus impliqués que les électrons de cœur. Pour être précis dans le calcul, il est courant de représenter les orbitales de valences par plusieurs fonctions de base. John Pople et son équipe (Hehre *et al.* 1969) ont mis en place une notation pour ces bases à valence séparée : X-YZg où X et YZ représentent le nombre de fonction de base décrivant les électrons de cœur et les électrons de valences, respectivement. Il est possible d'ajouter à la couche de valence des fonctions de polarisation et de diffusion pour donner une flexibilité à la densité électronique (Figure 16B).

- ***Protocole de dérivation des charges RESP***

Les champs de force incluent leurs propres charges atomiques dérivées selon une méthodologie bien définie (par exemple dans le cas du champ de force CHARMM, les charges atomiques sont obtenues en calculant les interactions d'énergie avec des molécules d'eau afin de reproduire le minimum d'énergie d'interaction, la géométrie et le moment dipolaire par rapport aux calculs de QM). Dans le cas du champ de force AMBER, la dérivation des charges partielles atomiques se fait en phase gazeuse après optimisation de la molécule et calcul du potentiel moléculaire électrostatique (MEP). Les charges atomiques RESP sont calculées de manière à reproduire au mieux la distribution des électrons autour de la molécule (Bayly *et al.* 1993).

La phase d'optimisation QM étant couteuse en temps de calcul, il est souvent nécessaire de « découper » les molécules de grandes tailles (≥80 atomes) en plusieurs fragments moléculaires avec des structures chimiques bien définies. Cette approche dite en brique modulaire (ou « *building block* ») a été utilisée la première fois par Cornell et Bayly (Cornell *et al.* 1995)(Bayly *et al.* 1993) pour dériver les première charges RESP des acides-aminées dans la première version des paramètres AMBER pour les protéines. Concrètement, l'approche « *building block* » consiste à « découper » une molécule d'intérêt en différents fragments moléculaires avec des conformations bien définies (des exemples de découpages sont donnés dans l'annexe 3). Les fragments moléculaires passent par une optimisation géométrique sur lesquels un calcul de potentiel électrostatique moléculaire (MEP) est effectué et les charges RESP sont alors dérivées (Figure 17).

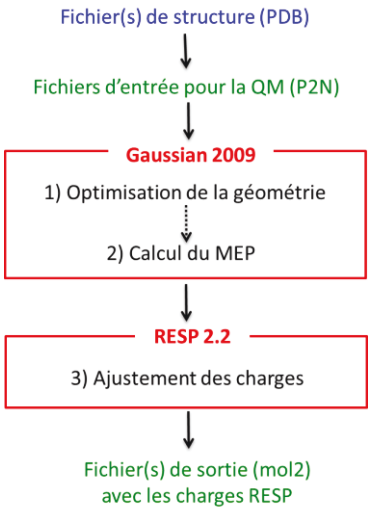


Figure 17 : Description simplifiée du serveur R.E.D. Un fichier P2N est l'équivalent d'un fichier PDB contenant les paramètres d'optimisation (e.g. charges totale, paramètre de contraintes).

Pour cette étape, nous avons utilisé le programme Gaussian version 2009 (www.gaussian.com) (Frisch *et al.* 2009), avec une base de type HF/6-31G* en phase gazeuse (Hehre 1976). Pour calculer la grille de point impliquée dans le calcul du MEP, nous avons utilisé l'algorithme de Connolly (Connolly 1983) pour le champ de force AMBER, et CHELPG (Breneman and Wiberg 1990) pour le champ de force GLYCAM. Ces deux algorithmes diffèrent par la position des points sur la grille : l'algorithme de Connolly utilise la surface de Connolly avec des points sur différentes couches concentriques autour de la molécule, alors que l'algorithme de CHELPG utilise un cube. Ce calcul introduit une légère erreur (une surestimation de la polarisation de la molécule de 10-15 %) qui rend les

charges compatibles avec les simulations en phase condensée (Bayly *et al.* 1993). L'ajustement des charges est réalisé avec le programme RESP (Bayly *et al.* 1993) (version 2.2 disponible sur R.E.D.S) en deux étapes avec des valeurs de contraintes hyperboliques de 0,0005 et 0,001 pour AMBER, et en une étape avec une valeur de contrainte hyperbolique de 0,01 pour GLYCAM. Ces contraintes hyperboliques permettent un meilleur ajustement des charges au MEP.

Dans le cas de la dérivation des charges pour le potentielle GLYCAM, on considère que les hydrogènes aliphatiques liés aux carbones sp^3 ne participent pas aux liaisons hydrogènes et en général ont des charges partielles quasi-nulles (Kirschner *et al.* 2008). On applique des contraintes internes consistant à donner au groupement CH, une charge unique.

Enfin, l'orientation moléculaire de chaque conformation optimisée est contrôlée par un algorithme de réorientation de corps rigide (RBRA) pour avoir des charges reproductibles indépendamment du programme quantique utilisé. Toute cette procédure de dérivation de charges partielles est automatisée dans le serveur Web R.E.D (http://upjv.q4md-forcefieldtools.org/REDS/)(Vanquelef *et al.* 2011).

Cette approche en « *building block* » à plusieurs avantages :

- de réduire significativement les calculs quantiques d'optimisation (facteur 2, voire plus) par rapport à la molécule entière
- de contrôler, en utilisant un protocole standardisé, la conformation de départ et par conséquent d'obtenir des charges RESP reproductibles.
- de créer une bibliothèque homogène de charges pour des molécules avec des structures similaires.

La Figure 17 montre les étapes importantes que nous avons réalisées pour dériver les charges partielles manquantes dans nos systèmes (i.e détergents, rhodopine glucoside et de la méthionine modifiée (N-carboxy-L-méthionine)).

V. Mise au point de nos simulations en dynamique moléculaire

Dans cette dernière section de ce chapitre, nous décrierons les différentes étapes utilisées pour construire et simuler nos systèmes moléculaires (Figure 18).

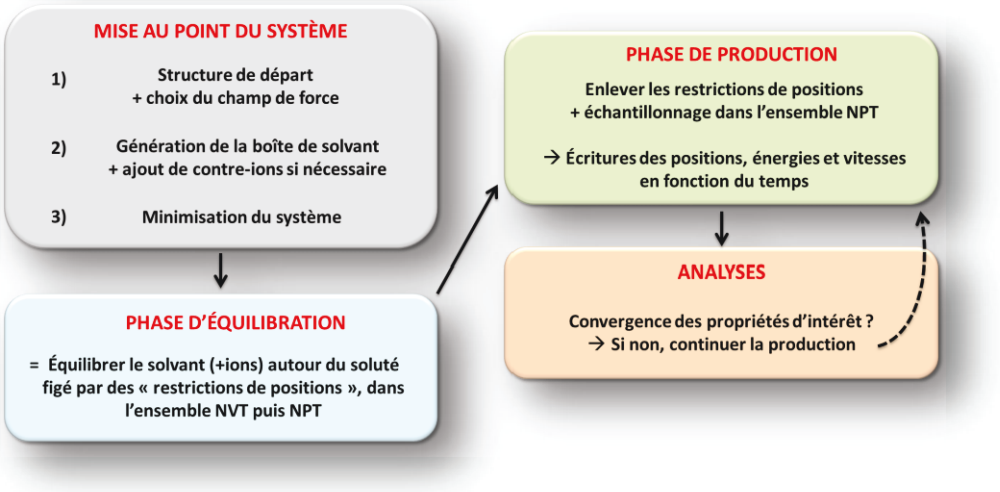


Figure 18 : Protocole de base que nous avons utilisé pour nos simulations de dynamique moléculaire.

1. *Construction du système.* Cette étape commence par une structure de départ, généralement un fichier PDB construit soit avec une approche *in-silico* ou à partir des données de diffraction de rayon X ou de RMN. Puis, à partir de cette structure, le programme *pdb2gmx* gère les fichiers paramètres selon le champ de force choisi. On construit une boîte de simulation avec des molécules de solvant. La neutralisation des systèmes se fait en ajoutant des contre-ions en utilisant le programme *genbox*.
2. *L'étape de minimisation* du système a pour objectif d'obtenir une conformation qui correspond au minimum de l'énergie d'interaction. Durant cette étape, les longueurs des liaisons et des angles des molécules seront ajustées selon les paramètres du champ de force.
3. *L'étape d'équilibration* se fait généralement en deux phases : une première dans un ensemble NVT pour « thermaliser » le système à une température fixée et une seconde étape dans un ensemble NPT où le volume de la boîte est ajusté pour arriver à la densité choisie du système.
4. *L'étape de production* consiste à collecter les configurations atomiques du système simulé en fonction d'une périodicité choisie. Cette étape peut être plus ou moins longue en fonction du système considéré et la convergence des observables souhaités.

Chapitre 3 : Simulations des environnements hydrophobes biomimétiques

Dans ce chapitre, nous allons décrire les simulations qui nous ont permis de valider nos paramètres pour le LDAO, qui seront utilisés par la suite dans les simulations avec le complexe LH2. Les simulations avec le bOG et la membrane de POPC (1-palmitoyl-2-oléoylphosphatidylcholine (POPC) sont présentés dans les annexes 1 et 2, respectivement. Le LDAO est un détergent zwitteronique utilisé notamment dans les expériences de spectroscopie de fluorescence en molécule unique du complexe LH2. Il présente un avantage car il est chimiquement homogène et a montré une utilité dans l'étude des protéines membranaires (Sardet *et al.* 1976)(Seddon *et al.* 2004), et enfin il est peu onéreux comparé au bOG. Ce dernier est aussi utilisé dans l'étude des protéines membranaires car il peut être facilement séparé par dialyse.

I. Les micelles de LDAO

1. Introduction à l'article

Les micelles de détergents ont un rôle très important dans l'étude des protéines membranaires, elles miment l'environnement membranaire, indispensable pour éviter que les protéines membranaires qui ont de larges régions hydrophobes se déstructurent ou s'agrègent dans l'eau. Dans cet article, publié dans « *Journal of Biomolecular Structure and Dynamics* », nous décrivons la dérivation des charges partielles atomiques RESP en utilisant l'approche des « *buildings blocks* » (voir chapitre 2). Ce protocole a été utilisé pour obtenir des charges compatibles avec les champs de force AMBER pour les protéines. Elles ont été validées en effectuant des simulations de micelles de LDAO dans de l'eau pure et en présence d'un fragment de lipide kinase humaine PI4KA appelé DI (Taveneau *et al.* 2015). Différents protocoles de simulations (pré-assemblées et auto-assemblées), champs de force (Amber99sb (Hornak *et al.* 2006) et Lipid14 (Dickson *et al.* 2014)), et différentes concentrations ioniques (0 et 300 mM de NaCl) ont été utilisés. Le modèle tridimensionnel de la protéine DI a été construit par modélisation *ab initio* avec le serveur web QUARK (Xu *et al*. 2012).

2. Article

Journal of Biomolecular Structure and Dynamics, 2016
http://dx.doi.org/10.1080/07391102.2015.1135822

Taylor & Francis
Taylor & Francis Group

Derivation of original RESP atomic partial charges for MD simulations of the LDAO surfactant with AMBER: applications to a model of micelle and a fragment of the lipid kinase PI4KA

Esra Karakas[a,b,c,†], Cyntia Taveneau[c,†], Stéphane Bressanelli[c*], Massimo Marchi[b,c], Bruno Robert[b,c] and Stéphane Abel[b,c*]

[a]*Maison de la Simulation, USR 3441, CEA – CNRS – INRIA – Univ. Paris-Sud – Univ. de Versailles, 91191, Gif sur Yvette, France;* [b]*Commissariat à l'Energie Atomique et aux Energies Alternatives, DRF/IBITECS/SB2SM/LBMS & CNRS UMR 9198, Saclay, France;* [c]*Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris Sud, Université Paris-Saclay, 91198 Gif sur Yvette, France*

Communicated by Ramaswamy H. Sarma

(*Received 8 October 2015; accepted 21 December 2015*)

In this paper, we describe the derivation and the validation of original RESP atomic partial charges for the N, N-dimethyl-dodecylamine oxide (LDAO) surfactant. These charges, designed to be fully compatible with all the AMBER force fields, are at first tested against molecular dynamics simulations of pure LDAO micelles and with a fragment of the lipid kinase PIK4A (DI) modeled with the QUARK molecular modeling server. To model the micelle, we used two distinct AMBER force fields (i.e. Amber99SB and Lipid14) and a variety of starting conditions. We find that the micelle structural properties (such as the shape, size, the LDAO headgroup hydration, and alkyl chain conformation) slightly depend on the force field but not on the starting conditions and more importantly are in good agreement with experiments and previous simulations. We also show that the Lipid14 force field should be used instead of the Amber99SB one to better reproduce the C(sp3)C(sp3)C(sp3)C(sp3) conformation in the surfactant alkyl chain. Concerning the simulations with LDAO-DI protein, we carried out different runs at two NaCl concentrations (i.e. 0 and 300 mM) to mimic, in the latter case, the experimental conditions. We notice a small dependence of the simulation results with the LDAO parameters and the salt concentration. However, we find that in the simulations, three out of four tryptophans of the DI protein are not accessible to water in agreement with our fluorescence spectroscopy experiments reported in the paper.

Keywords: micelle; MD simulation; molecular modeling; fluorescence spectroscopy; LDAO surfactant; lipid kinase PI4KA

Downloaded by [CEA Saclay] at 02:18 22 March 2016

1. Introduction

In addition to being routinely used by pharmaceutical industry and cosmetics, surfactant molecules are also important in the investigation of membrane proteins. Indeed, since surfactants mimic a membrane environment, they are used to solubilize and to keep membrane proteins biologically active (Privé, 2007). Among these surfactants, N,N-dimethyl-dodecylamine oxide (LDAO, Figure 1) is an amphiphilic surfactant with a high hydrophilicity character (Kocherbitov, Veryazov, & Söderman, 2007). In an aqueous environment, depending on the experimental conditions and concentrations, LDAO can form different mesostructures such as micelles or lamellar phases (Lutton, 1966). To the best of our knowledge, the first work that examined the structural properties of LDAO micelles was done by Herrmann et al. in the early 80s (Herrmann, 1962) with static light scattering experiments. In this work, the authors showed that 76 LDAO monomers are needed to form a stable micelle. More recently, other groups reported aggregation numbers in the range of 67–110 molecules (Kakitani, Imae, & Furusaka, 1995; Thiyagarajan & Tiede, 1994; Timmins, Leonhard, Weltzien, Wacker, & Welte, 1988; Timmins, Hauk, Wacker, & Welte, 1991). In more recent small-angle neutron scattering (SANS) experiments, Barlow, Lawrence, Zuberi, Zuberi, and Heenan (2000) found an aggregation number around 104. The large variations of the aggregation numbers estimated in experiments can be explained by the fact that interparticle interactions were not taken into account or the presence of ions that lead to a partial ionization of the LDAO headgroup (Kakitani et al., 1995).

Molecular dynamics (MD) simulations have been used in the past to gain insights in the structures of surfactant assemblies (Bongiorno, Ceraulo, Indelicato, Liveri, & Indelicato, 2015). In particular, Marchetti,

*Corresponding authors. Email: stephane.abel@cea.fr (Stéphane Abel); stephane.bressanelli@i2bc.paris-saclay.fr (Stéphane Bressanelli)
Esra Karakas and Cyntia Taveneau contributed equally to this work

© 2016 Informa UK Limited, trading as Taylor & Francis Group

Marchi, and Le Maire (2006) and Sterpone, Marchetti, Pierleoni, and Marchi (2006) examined the structure and dynamics of water in the solvation shell of a model of LDAO micelles. LDAO was modeled with the CHARMM22 (Kamath, Guvench, & MacKerell, 2008) force field and its charges were computed from density functional theory calculations. More recently, (Lorenz, Hsieh, Dreiss, & Lawrence, 2011) used the OPLS-AA force field (Jorgensen, Maxwell, & Tirado-Rives, 1996) to examine the structural properties of micelles containing LDAO alone and LDAO in association with a mixture of C4/C8 ethyl ester oils. All these authors, using an aggregation number of 104 (Barlow et al., 2000), found that the LDAO micelles had an ellipsoidal shape with a size of around 19 Å and average hydration number per LDAO around 8, in good agreement with experiments (Barlow et al., 2000; Thiyagarajan & Tiede, 1994). Although the different AMBER force fields are widely used to simulate membrane proteins (see for instance Cordomí, Caltabiano, & Pardo, 2012; Kumar, Hajjar, Ruggerone, & Ceccarelli, 2010), to the best of our knowledge, no AMBER parameter (Restrained ElectroStatic Potential, RESP charges) is available for the LDAO. To fill the gap, in this paper we have derived original RESP atomic partial charges for this surfactant with a theory level compatibles with the AMBER force field for proteins and its successive adaptations (Cieplak, Cornell, Bayly, & Kollman, 1995; Dickson et al., 2014; Hornak et al., 2006; Wang, Wolf, Caldwell, Kollman, & Case, 2004). To test the transferability of these charges and validate our approach, we used two test cases.

In the first test case, we simulated a model of pure LDAO micelle, where the LDAO RESP charges were combined with Amber99SB (Hornak et al., 2006) and Lipid14 (Dickson et al., 2014) force fields. We chose the former parameter, since we have shown previously that it can be used to model reasonably well a model of dodecyl phosphocholine (DPC) surfactant in pure micelle (Abel, Dupradeau, & Marchi, 2012) or with membrane peptides (Abel, Lorieau, de Foresta, Dupradeau, & Marchi, 2014). We also found in the former study that the Amber99SB parameters tend to overestimate the *trans* conformation of the C(sp3)C(sp3)C(sp3)C(sp3) torsion in the alkyl chain, in contrast to other force fields such as CHARMM36 (Klauda et al., 2010), GROMOS53A6 (Oostenbrink, Soares, van der Vegt, & van Gunsteren, 2005) or GROMOS54A7 (Poger, Van Gunsteren, & Mark, 2010). To solve this problem, Dickson et al. (2014) developed new parameters for simulations of long alkyl chain lipids (called hereafter Lipid14) where the C(sp3)C(sp3)C(sp3)C(sp3) torsion parameters and the CH_2 proton Lennard-Jones values were recomputed to better match experiments.

In the second test case, we simulated a model of protein–LDAO complex containing a fragment of the human lipid kinase PI4KA called through the paper "DI" and studied in our recent paper (Taveneau, Blondeau, & Bressanelli, 2015). PI4KA is a 2102-residue protein that transiently associates with the cytosolic leaflet of membranes to perform its phosphoryl transfer activity to the headgroup of phosphatidyl-inositol (PI) (Balla, 2013). We previously reported that the first 1100 N-terminal residues of PI4KA are dispensable for this function and particularly for the support of hepatitis C virus (HCV) replication (Harak et al., 2014). By sequence analysis, we also defined a likely pivotal domain of some 131 residues, which we term PI4KA "DI", located right after the 1100 N-terminal residues. We further verified that DI is part of the functional C-terminus of PI4KA and thus potentially involved in membrane recruitment (Taveneau et al., 2015). In this work, we also show that the use of LDAO is essential to solubilize and characterize the DI protein structure. As the DI protein three-dimensional structure is not known, we construct the first model of DI with the QUARK protocol (Xu & Zhang, 2012) and use it in MD simulations with LDAO to examine the biophysical properties of the protein–surfactant complex.

The paper is organized as follows: In Section 2, we describe the derivation of the RESP charges for the LDAO, the construction of our micellar systems with two approaches called "pre-assembled" (PA) and "self-assembled" (SA) and the gathering of the LDAO models with QUARK. Furthermore, we discuss the experimental details for the LDAO-solubilized DI purification and characterization by tryptophan fluorescence and size exclusion chromatography-multi-angle laser light scattering (SEC-MALLS). In Section 3, we present our results on micellar characteristic properties, such as size, shape, surfactant conformation, and hydration, and on the characteristics of the interactions between our PI4KA DI model and the LDAO surfactants.

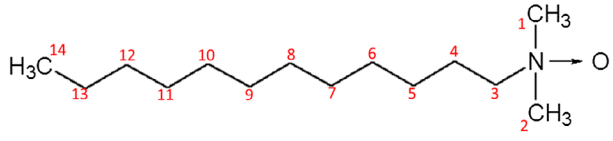


Figure 1. N, N-dimethyldodecylamine-N-oxide (LDAO) molecule with the carbon numeration used for our analysis.

2. Material and methods

2.1. *RESP charges derivation*

The RESP charges derivation for the LDAO was carried out using the "building blocks" approach. In this approach, a complex molecule (i.e. with different chemical groups) is split into different elementary chemical

Downloaded by [CEA Saclay] at 02:18 22 March 2016

Downloaded by [CEA Saclay] at 02:18 22 March 2016

fragments with well-defined and controlled conformations (Dupradeau et al., 2010; Gouin et al., 2007). The "building blocks" approach has many advantages over the whole molecule approach to derive RESP charges: (i) the computational time required for geometry optimization and molecular electrostatic potential (MEP) computation is drastically reduced, (ii) the optimized geometry of the conformation(s) of each molecular fragment is fully defined and controlled, (iii) optimized conformations presenting intramolecular hydrogen bonds are discarded from charge derivation to avoid over-polarization effects, and finally (iv) a large set of analog molecules can be simultaneously involved in charge derivation leading to a homogeneous Force Field Topology DataBase (FFTopDB). This approach was successfully used in the past to obtain RESP charges compatible with the Amber force fields for various type of molecules (e.g. glucose polymers (Gouin et al., 2007), cyclodextrines (Cézard, Trivelli, Aubry, Djedaïni-Pilard, & Dupradeau, 2011), calixarenes (Vanquelef et al., 2011), peptide nucleic acid (Sanders et al., 2013), or phosphocholine and glycolipid surfactants (Abel, Dupradeau, Raman, MacKerell, & Marchi, 2011; Abel et al., 2012). Here, we chose to split the LDAO molecule into two molecular fragments between the C3 and C4 bond (see Figure 2) and add a methyl cap to these fragments to obtain the ethyl-N,N-dimethyl amine oxide (Et-DAO) and n-dodecane (DOD). These two fragments with an *all-trans* conformation were involved in charge derivation procedure. The charge constraint for each cap was set to 0 since total charge of the adjacent methyl is close to *0e* (see Supplementary Tables S1–11) allowing the

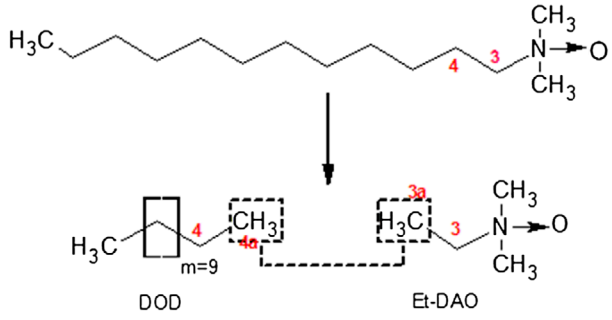


Figure 2. Building block scheme used to derive the RESP charges for LDAO molecule and its derivative. Et-DAO and DOD stand for the ethyl-N,N-dimethyl amine-oxyde and n-dodecane, respectively, involved in the intermolecular charge constraints. The methyl caps are show with dashed boxes. Note that carbon atoms were numbered according to the RESP charge derivation procedure described in the main text. Intermolecular charge constraints set to 0 for CH_3 used during the fitting step are represented using dashed boxes (see the main text for details). This procedure allows defining the molecular fragments required for construction of the N,N-dimethyl amine oxide-based surfactants with different alkyl chain lengths (here, with $m = 2$–12).

definition of the molecular fragments and force field libraries required to build the LDAO and analogs.

Two molecular orientations based on the C1, N, and C2, and C3, C4, and C5 sets of atom names for N, N-dimethylamine-N-oxide and the n-dodecane, respectively, were involved in charge derivation. The geometry optimization, frequency calculations, and MEP computations of each block were carried out in the gas phase with the Gaussian 2009 program (Frisch et al., 2009) using the HF/6-31G* level of theory (Hehre, 1976). This level of theory in the MEP computation leads to an implicit polarization required for condensed phase simulations with the AMBER force fields (Bayly, Cieplak, Cornell, & Kollman, 1993). Only the building block conformations with the lowest minima were involved in the charge derivation procedure. The charge fitting was performed in two steps with hyperbolic constraint values of .0005 and .001 with the RESP program (Bayly et al., 1993) (version 2.2). The Connolly algorithm was used to compute the grid point involved in the MEP calculation (Connolly, 1983). The molecular orientation of each optimized geometry was controlled using the rigid-body reorientation algorithm implemented in the RED IV program leading to a RESP charges derivation procedure fully controlled (i.e. independent of the QM program and the choice of the initial structure, see discussion in Dupradeau et al. (2010). To automatize the RESP charge derivation procedure, we used the RED server (Vanquelef et al., 2011). Note that the RESP charges of the alkyl chain are found in same range that we obtained previously for the alkyl chain of the DPC and its analog (see Abel et al., 2012). The computed RESP charges for the each fragment used to construct the LDAO and its analog are reported in Supplementary Tables S1–S11 and the corresponding mol2 files (with the molecule topologies and the atomic partial charges) will be freely available in R.E.DD.B (Dupradeau et al., 2008). The RESP atomic charges obtained in this work differ from ones available in literature (for instance Neale et al. (2013) or Sterpone, Marchetti, Pierleoni, and Marchi (2006)) probably because these authors used different charge derivation methodology as ours. In addition, to verify that NO group of our model is enough hydrophilic to ensure an amphiphilic character of our LDAO model, we carried out >120 ns long simulations with different LDAO concentrations and we found that in all the case, LDAO surfactants form quickly stable aggregates.

As we mentioned in the introduction to evaluate the transferability of our RESP charges to different AMBER force fields, we combined them with the Amber99SB (Hornak et al., 2006) and the Lipid14 (Dickson et al., 2014) parameters. We point out that for modeling the N–O atoms in the LDAO headgroup, we chose by analogy the N4 (nA) and O (Oc) atom types for the four substituted nitrogen and carbonyl oxygen available in

GAFF (Wang et al., 2004) and Lipid14, respectively. A similar choice was done for the N and O atom types by Neale et al. (2013) for modeling the LDAO headgroup with the Berger, Edholm, and Jähnig (1997) united-atom force field.

2.2. Simulation setup

2.2.1. Test case 1: simulation of a pure LDAO micelle

To compare our simulation results with previous ones, we chose to construct a micelle of 104 LDAO monomers. The surfactants were solvated in a cubic box of 100 Å of side containing 32,000 TIP3P (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983) water molecules (96,000 atoms in total) given a system with concentration of ~.18 M. This is ~180 times the LDAO experimental critic micellar concentration of 1 mM (Herrmann, 1962) and similar to the concentration simulated by Lorenz et al. (2011).

In this work, we examine the influence of the simulation initial conditions using "pre-assembled" (PA) and "self-assembled" (SA) protocols to generate the initial system coordinates. In the PA approach, a spherical micelle is constructed with the Packmol program (Martínez, Andrade, Birgin, & Martínez, 2009). This program builds an initial MD starting point by packing the surfactant molecules in defined regions of space, thus forming a spherical micelle. We used two spheres in this construction: Initially, the hydrophobic end of LDAO was oriented inside a first sphere with the C1 carbon of the alkyl chain (see Supplementary Figure S1) placed on the outer sphere of 17 Å, whereas C12 was placed inside an inner sphere of radius of 4.0 Å. The SA simulations, instead, were constructed by randomly placing the 104 LDAO monomers within the simulation box. These simulations were done twice (called hereafter SA1 and SA2 in Tables and Figures) with different random seeds to examine the influences of the simulation starting conditions.

The simulations presented in this paper were carried out with GROMACS (v4.5.5 and 4.6.6) MD package (Pronk et al., 2013). To prepare the system prior to the simulation production runs, each micellar system was minimized with the steepest descent algorithm and an energy tolerance (energy difference between two consecutive steps) lower than 1000 kJ.mol^{-1} nm^{-1}. In the second stage, the minimized systems were equilibrated in the NVT ensemble at 298 K for 400 ps using the Berendsen thermostat (Berendsen, Postma, van Gunsteren, & Hermans, 1981) with a coupling constant $\tau_T = .1$ ps. The resulting molecular systems were then equilibrated in the NPT ensemble at ambient conditions ($T = 298$ K and $p = 1.015$ bar) for 500 ps using the Bussi et al. thermostat (Bussi, Donadio, & Parrinello, 2007) ($\tau_T = .1$ ps) and the Parrinello–Rahman barostat (Parrinello & Rahman, 1981; Rahman & Stillinger, 1971) ($\tau_p = 3.0$ ps),with a compressibility value of $4.5.10^{-5}$ bar^{-1}. The LDAO and the water were separated in two thermostatic baths. During the NVT and NPT equilibration steps, the LDAO molecules were harmonically restrained with a force constant of 1000 kJ.mol^{-1} nm^{-2} to equilibrate the solvent. The equilibrated systems were then simulated for production in the NPT ensemble ($T = 298$ K and $p = 1.015$ bar) using the Nosé–Hoover thermostat (Hoover, 1985; Nosé, 1984) ($\tau_T = .4$ ps) and the Parrinello–Rahman barostat ($\tau_p = 3.0$ ps). Periodic boundary conditions were used and the time step for integration was fixed at 2 fs with the neighbor list updated every 10 fs. The particle-mesh Ewald method (Essmann et al., 1995) was used to evaluate the electrostatic interactions and the P-LINCS algorithm (Hess, 2007) was applied to restrain bond lengths to their equilibration values. We carried out production runs of ~100–150 ns for the PA and the SA simulations, respectively, with the atomic configurations collected every 2 ps for subsequent analysis (Table 1).

2.3. Test case 2: simulation of the DI protein with and without LDAO

2.3.1. PI4KA DI ab initio modeling

In the absence of the three-dimensional (3D) structure of PI4KA DI and available homologs, we used the QUARK (Xu & Zhang, 2012) webserver with the PI4KA DI amino acid sequence as an input to create a molecular model of the protein (see Section 2.4). We used the default parameters without specifying additional restraints, secondary structure or excluding some templates to guide the modeling. We chose this modeling server, since it was shown previously (Tai, Bai, Taylor, & Lee, 2014) that it provides excellent results for obtaining 3D models and performing atomic-level structure refinements of proteins with unknown 3D structure. QUARK uses a semi-reduced model to represent protein residues by considering in the computation only the full backbone atoms and the sidechains centers of mass. A neuronal network is used to predict the additional structural features. The global fold is then generated by assembling the small fragments and additionally performing replica-exchange Monte Carlo simulations with a composite knowledge-based force field. In this last step, QUARK uses small fragments of 1 to 20 residues. We finally compared the quality of the DI model obtained with the secondary structure computed from other predictor webservers (such as Robetta (Kim, Chivian, & Baker, 2004), GenThreader (Jones, 1999) and PredictProtein (Rost, Yachdav, & Liu, 2004)) and found that in agreement with them, QUARK predicts a compact protein an

Downloaded by [CEA Saclay] at 02:18 22 March 2016

Table 1. Simulation parameters: N_{LDAO}, N_{H2O}, n_{Na^+}/n_{Cl^-}, and N_{prot} are the numbers of LDAO, water, sodium, chloride ions, and protein in each simulation box, respectively. t_{sim} is the productive simulation time in ns. Note that the protein was, in all the case, modeled with Amber99SB-ILDN force field and that the self-assembly (SA) Protein–LDAO simulations were carried out twice as mentioned in the main text.

Force Field	MD	Pure LDAO micelle				
Amber99SB		N_{LDAO}	N_{H2O}	n_{Na+}/n_{Cl}	$N_{Protein}$	t_{sim}
	PA	104	31,989	–	–	95
	SA1					130
	SA2					150
Lipid14	PA					100
	SA1					140
	SA2					145
		LDAO micelle + DI Protein				
Protein alone**	0 mM	0	9997	3/0	1	185
	300 mM		9879	120/117		215
Amber99SB	SA1	41	20,000	3/0	1	180
	SA2		19,763	120/117	1	
Lipid14	SA1		20,000	3/0	1	180
	SA2		19,763	120/117	1	

high-α-helix content (see Supplementary Figure S2 and Section 3.2.1).

2.3.2. *Simulation of the DI protein with and without LDAO*

We simulated the DI protein model with the Amber99SB-ILDN parameters (Lindorff-Larsen et al., 2010), an improvement of the Amber99SB force field for proteins, in presence and in absence of the LDAO surfactants modeled with the same force fields used in the pure micelle simulations. The initial conformations of these simulations were obtained by placing the DI protein in the center of a cubic box (81 Å of side) with 41 randomly placed LDAO (see Section 3.2.1 for the choice of this number) and ~20 000 TIP3P waters (Jorgensen et al., 1983). With this system, we carried out two sets of simulation with 0 mM and 300 mM of NaCl to examine the effect of the salt concentration on the simulation results and also mimic in the latter case the experimental conditions. For this, we added in the former system only 3 Na^+ (Joung & Cheatham, 2008) to neutralize the system whereas in the latter one, 120 Na^+ and 117 Cl^- ions were added instead. The LDAO concentration of these systems is 114 mM, ~57 times larger than in our experiments (2 mM, see Section 2.4.3).

The simulations of the DI with and without surfactant were carried out during 185–215 ns (Table 1), excluding the equilibration periods (~900 ps) carried out in the NVT and NPT ensembles with the same temperature and pressure conditions as in the pure micelle simulations. During the equilibration periods, the protein and the LDAO were kept fixed to randomize water and ions. As for the pure micelle simulations, the LDAO-DI simulations were done twice to examine the reproducibility of the simulation results.

2.4. *Expression and purification of DI constructs*

2.4.1. *Cloning of DI into pGEX-6P-1 and pET28(a+) expression vectors*

We used the same procedures as detailed in our previous work (Taveneau et al., 2015). Briefly, we amplified by PCR DNA sequences corresponding to DI from the full-length PI4KA gene sequence (Gene Bank accession no: AB384703). We then introduced the DI DNA coding for residues 1135-1265 of PI4KA in two different bacterial expression plasmids: pGEX-6P-1, containing an N-terminal Glutathione S-transferase (GST)-tag sequence; and a customized pET28(a+) vector containing an N-terminal 6-histidine tag with a TEV protease cleavage site. As a result, the expressed proteins comprise 230 and 21 extra residues upstream the 131 DI residues for GST-DI and his6-DI, respectively. The DI inserts were sequenced prior to expression tests (GATC biotech, Germany).

2.4.2. *Expression of GST-DI and his6-DI*

The bacterium strains used for protein expression were *E. coli* BL21(DE3) Rosetta for his6-DI and *E. coli* BL21(DE3)C41 for GST-DI. After transformation with the recombinant pGEX-6P-1 or pET28(a+), a single colony was subcultured first into 5 mL and then into 500 mL of fresh LB medium supplemented with 100 µg/mL ampicillin for PGEX-6P-1 construction, 50 µg/mL kanamycin for pET28(a+) construction, and 34 µg/mL chloramphenicol for both subcultures at 37°C under agitation. On the third day, this subculture was diluted in 5 L of fresh medium with appropriate antibiotics to a final optical density about .1 at a wavelength of 600nM (OD600). Cells were cultivated at 37°C with shaking until they reached an OD600 about .8. Protein expression was then induced at 12°C with .3 mM

Downloaded by [CEA Saclay] at 02:18 22 March 2016

Isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were then incubated overnight and finally harvested by centrifugation of 20 min at 3500×*g*. Pellets were stored at −20°C.

2.4.3. *Purification of GST-DI and his6-DI*

Purification protocols were substantially the same for the two constructs. The entire process was realized at 4°C and all buffers were supplemented with LDAO. Pellets from a 4-L culture were suspended in 40 mL of lysis buffer implemented with 1 tablet of protease inhibitor cocktail (ref 11836170001, Roche) and about 250 U of benzonase (ref 70746-3, Merck-Millipore). Lysis was performed using sonication 3 × 22 min with pulses of 3 s at 30% amplitude. The lysate was clarified by two successive centrifugations: the first one for 12 min at 8.000 g, and the second one for 30 min at 40.000 g. The supernatant was then filtered with a .45 µm PES membrane filter. The lysate was charged at .5 mL/min in a GSTrap FF or Histrap HP column of 1 mL, depending on the fusion tag, connected to an AKTApurifier 10 liquid chromatography system. The column was washed at 1 ml/min until OD280 nm baseline was reached. Proteins were eluted with elution buffer at a flow rate of .5 ml/min and collected in fractions of 1 mL. Buffer compositions are described in Supplementary Table S12. The sample was concentrated in 500 µl and loaded on a Superdex 200 10/300 GL column (GE Heathcare), pre-equilibrated in 30 mM Tris-HCl pH 7.4, 300 mM NaCl, 10% glycerol, 2 mM LDAO. The purified protein in the flow through was concentrated by ultrafiltration (with a cut off of 50 kDa to alleviate LDAO micelle concentration) to a final protein concentration of .53 mg/mL for GST-DI and with a cut-off of 10 kDa to a final concentration of .1 mg/mL for his6-DI. Each purification step was analyzed by 14% tris tricine SDS-PAGE. The samples were then flash frozen and stored at −80°C.

2.4.4. *Biochemical and biophysical analyses*

2.4.4.1. Gel filtration analysis. In order to check the importance of detergent in GST-DI solubility, three purifications were carried out following the purification protocol above. Only the amount of LDAO was changed from the wash step of the affinity column to the end of the purification. Three conditions were tested: without LDAO, 1 mM LDAO (i.e. below the CMC) and 2 mM LDAO (i.e. slightly above the CMC). 500 µL of the more concentrated fraction of the affinity column were injected on Superdex 10/300 GL in order to avoid the intermediate concentration step. Gel filtration chromatograms were superimposed to evaluate the solubility of DI under these conditions.

2.4.4.2. SEC-MALLS analysis of GST-DI. For DI-LDAO stoichiometry determination, SEC-MALLS-RI (refractive index) analysis of GST-DI was carried out using Superdex 200 10/300 GL Increase column *(GE Healthcare)* connected to Shimadzu HPLC system. In addition to absorbance at 280 nm (UV detection), light scattering was detected with a *mini DAWN TREOS* system (Wyatt Technology) and refractive index measured with an Optilab T-rEX differential refractometer (Wyatt Technology). A total of 30 µg of protein were injected at a flow rate of .5 ml/min on the column equilibrated in 30 mM Tris-HCl pH 7.4, 300 mM NaCl, 10% glycerol, 2 mM LDAO. Protein was collected in 500 µL fractions, precipitated with TCA, and analyzed by 14% tris-tricine SDS-PAGE. Molar mass of proteins, micelles, and protein–detergent complexes were determined from the triple detection data (MALLS, UV, and refractive index) with ASTRA 6.1 software (Wyatt Technology) using protein conjugate analysis with a refractive index increment (dn/dc) .1804 ml g^{-1} predicted thanks to SEDFIT software calculation module (Schuck, 2000). A control experiment was performed by injecting 48 µg of GST alone in the same buffer as GST-DI. A total of 200 µg of Bovine Serum Albumin was used to calibrate the detector constant.

2.4.4.3. Tryptophan fluorescence analysis. Degree of tryptophan burial was evaluated for his6-DI by measuring tryptophan fluorescence on a PerkinElmer LS50B fluorimeter at 18°C. Excitation was performed at 295 nm and emission was measured between 305 and 400 nm (both with slit width 2.5 nm) at a speed of 120 nm/min.

3. Results and discussions

3.1. *Simulation of pure LDAO micelle*

In the next six subsections, we will focus on the validation of our LDAO RESP charges by comparing the simulations results with available simulation and experimental data.

3.1.1. *Aggregation process of the LDAO monomers into a single micelle*

In Figure 3, we present representative snapshots of the aggregation process of the 104 LDAO monomers into a single micelle as a function of the simulation time for the first Lipid14 SA simulation (Figure 3). As shown in this figure, it takes ~5 ns for all of the LDAO monomers to form small clusters and an additional ~30 ns for these small clusters to aggregate into a single micelle. To compare the LDAO aggregation process in the Amber99SB and Lipid14 simulations, we report in Figure 4 the number of LDAO clusters, N(t), as a function of time

Downloaded by [CEA Saclay] at 02:18 22 March 2016

Downloaded by [CEA Saclay] at 02:18 22 March 2016

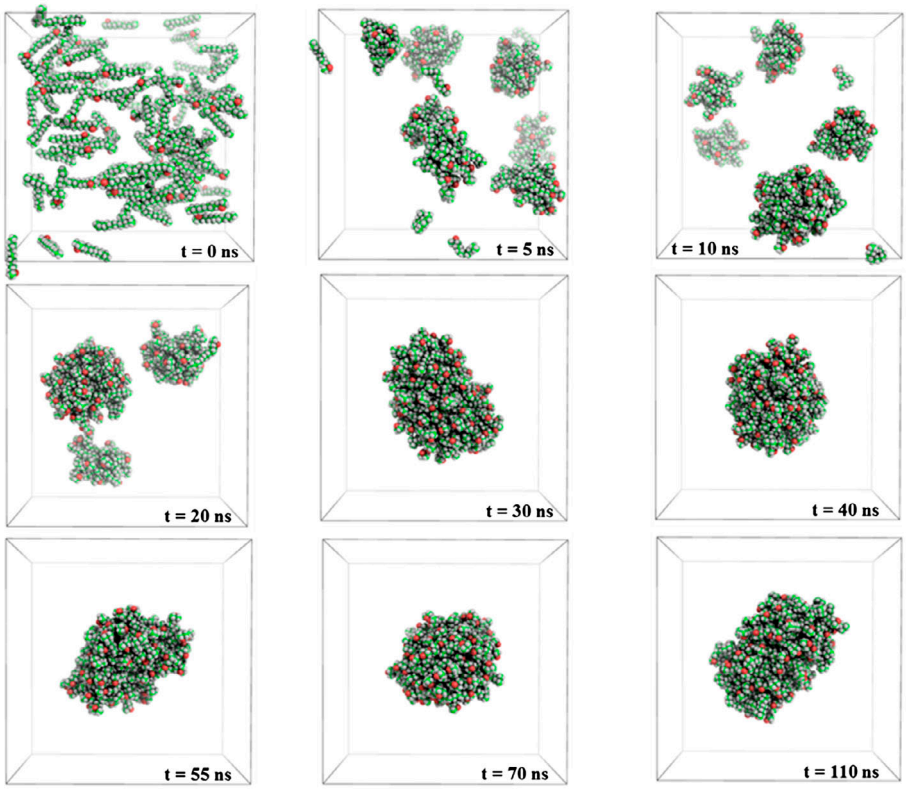
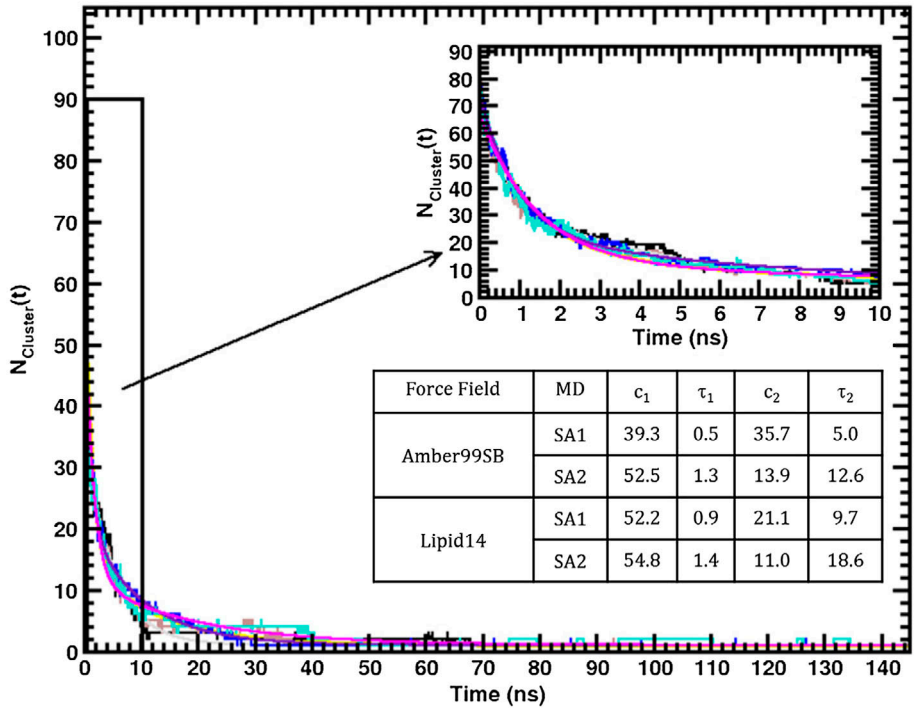


Figure 3. Representative snapshots of the LDAO self-aggregation into a single micelle obtained with the Lipid14 SA1 simulation. Figure was drawn with Pymol (Schrodinger, 2010).

for each simulation (Figure 4). As previously done for micelles with DPC (Abel et al., 2012) and β-dodecyl maltoside (DDM) surfactants (Abel et al., 2014), we consider that two LDAO molecules belong to the same cluster if any of their alkyl chain carbon atoms are within 4.1 Å of each other. As shown in Figure 4, the LDAO aggregation kinetics differs among the two force fields and the simulation replica. Indeed, it took approximatively 50 ns for the LDAO form a stable micelle with all the surfactant monomers. The aggregation kinetics obtained in this work is similar to the results obtained by Lorenz et al. (2011) for the same system modeled with the OPLS-AA force field. As previously observed for other micellar systems (see for instance (Abel et al., 2012) and reference cited therein), in our simulations the micelle aggregation process occurs with two different time scales. To characterize these time scales (Abel et al., 2012), we fit the $N_{cluster}(t) - {}_{cluster}(t_\infty)$ with a double exponential function: $c_1 e^{-t/\tau_1} + c_2 e^{-t/\tau_2}$, where $N_{cluster}(t_\infty) = 1$. In this equation, τ_1 and τ_2 correspond to the relaxation times for a faster and for a slower aggregation process, respectively (Nyrkova & Semenov, 2005) (Figure 4). As for the exponential parameters, if one hand the sum of c_1 and c_2 corresponds approximately to the initial number of monomers, given the separation of time scales between τ_1 and τ_2, c_2 can be identified as the initial number of clusters for the slower self-aggregation kinetics. The τ_1, τ_2, c_1, and c_2 values computed from the four self-assembled simulations are reported in the inset in Figure 4. We first notice that the faster kinetics has a timescale $\tau_1 \sim 1$ns for all the simulations done with the two force fields. This faster aggregation process is followed by a relaxation time of τ_2 = 5–18.6 ns. It worth to mention that except one Lipid14 simulations where we observed a short escape (~5 ns) of one LDAO in the aqueous phase (see blue and cyan curves in Figure 5), in all the simulations, the 104 LDAO monomers form a stable micelle. For sake of comparison, the $\tau_1$and τ_2 values are found slightly larger than the values previously obtained by us (Abel et al., 2014) for pure DDM micelle with 132 monomers (τ_1 = .6 and τ_2 = 3.5 ns) modeled with the GLYCAM force field (Kirschner et al., 2008).

3.1.2. Size and shape of the micelle

The sizes of the micelles in the SA and PA simulations were compared by computing their radii of gyration, R_g, and their effective radius, R_M. Figure 5 depicts the time evolution of the six micelles R_g obtained from the PA/SA simulations (Figure 5). By opposition to the number



Force Field	MD	c_1	τ_1	c_2	τ_2
Amber99SB	SA1	39.3	0.5	35.7	5.0
	SA2	52.5	1.3	13.9	12.6
Lipid14	SA1	52.2	0.9	21.1	9.7
	SA2	54.8	1.4	11.0	18.6

Figure 4. The LDAO self-assembly (SA) as a function of time for the Amber99SB (black and brown colors) and the Lipid14 simulations (blue and cyan colors). The double-exponential fitting values for each SA simulations are reported in the inset and drawn with gray, brown, cyan, and blue for the Amber99SB and Lipid14 first and second replica, respectively (see main text for details).

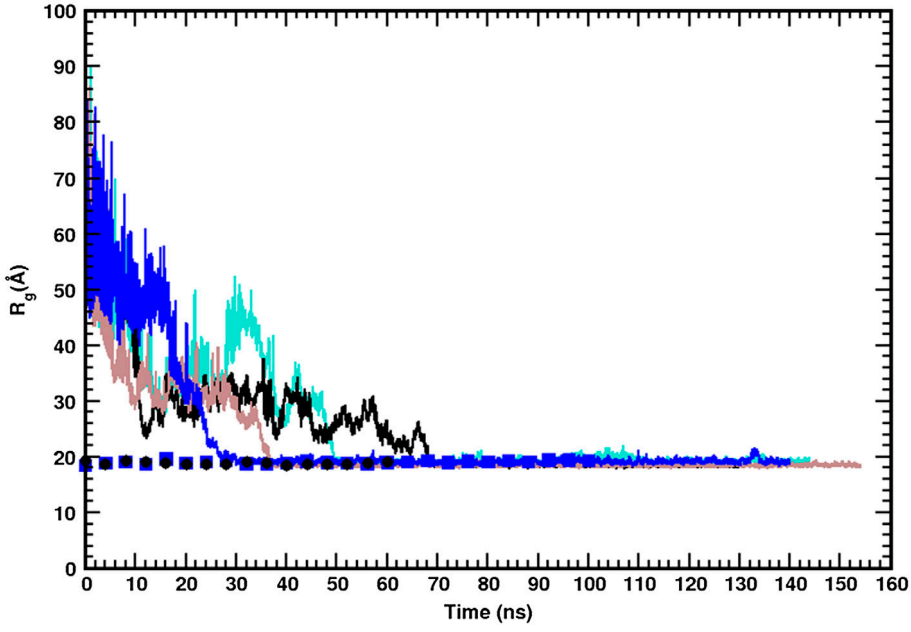


Figure 5. Time evolution of the micelle radius of gyration in the PA and SA simulations. The Amber99SB and Lipid14 PA runs are shown with black and blue squares, respectively, whereas the Amber99SB and (Lipid14) SA are shown in black and brown and (blue and cyan) continue lines, respectively.

Downloaded by [CEA Saclay] at 02:18 22 March 2016

Downloaded by [CEA Saclay] at 02:18 22 March 2016

of clusters $N_{cluster}$, the LDAO micelle R_g values display large oscillations just before settling to the stable value around 18.7 Å, close the micelle R_g obtained in the PA Amber99SB and Lipid14 simulations (Figure 5, black and blue squares, respectively), after 30–70 ns depending on the force field and simulation runs, demonstrating that formation of the final cluster occurs rather abruptly within 2−3 ns. The average R_g values computed from the last 50 ns of each Amber99SB and Lipid14 simulations are reported in the 3rd column of Table 2. The values show that the Amber99SB LDAO micelles are slightly smaller than the Lipid14 ones (18.5 ± .2 Å and 19.2 ± .3 Å, respectively) and these results do not depend on the simulation starting conditions. The average micelle R_g values obtained in our simulations are in good agreement with previous simulations of LDAO micelle modeled with the CHARMM22 (18.7 ± .1 Å) (Marchetti et al., 2006) and OPLS-AA (17.6 Å) (Lorenz et al., 2011) force fields.

To compare the micelle sizes obtained in our simulations with those obtained from the SANS experiments (see by Barlow et al. (2000) and Sardet, Tardieu, and Luzzati (1976)) and from previous simulations (Lorenz et al., 2011), we also computed the micelle mean effective radius, R_M (4th column Table 2), for a solid sphere of uniform density (Karjiban, Basri, Rahman, & Salleh, 2012) using the following expression:

$$R_M = \sqrt{\frac{5}{3}} < R_g > \tag{1}$$

The Amber99SB and Lipid14 micelle mean R_M values are in range of 23.9–24.7 ± .2 Å and agree well with SANS experiments (22–24.4 Å) (Barlow et al., 2000; Sardet et al., 1976) and the simulation of Lorenz et al. (2011) with the OPLS-AA force field (22.7 Å).

In Figure 6, we present a pictorial view of the micelle at the end of one of the Lipid14 SA simulation (Figure 6). As shown in this figure, the micelle presents pronounced ellipsoidal shape. In order to obtain the micelle shape parameters, we compute the three ellipsoid semi-axis average values, a, b, and c (Supplementary Table S13) of the equivalent ellipsoid from the principal moment of inertia (MOI) (Abel, Sterpone, Bandyopadhyay, & Marchi, 2004):

$$\begin{aligned} I_1 &= \tfrac{M}{5}(b^2 + c^2) \\ I_2 &= \tfrac{M}{5}(a^2 + c^2) \sim \\ I_3 &= \tfrac{M}{5}(a^2 + b^2) \end{aligned} \tag{2}$$

where M is the total mass of the micelle and I_1, I_2, and I_3 (with $I_3 > I_2 > I_1$) are the micelle MOI. In addition to the I_1/I_3 ratio, we have also computed the asymmetry parameter, α, (Tieleman, van der Spoel, & Berendsen, 2000) a SAXS/SANS metric used to evaluate the deviation of the shape of an aggregate from the sphericity, defined as:

$$\alpha = \frac{(2I_1 - I_2 - I_3)}{(I_1 + I_2 + I_3)} \tag{3}$$

An elliptical micellar shape corresponds to an α larger than 0 (Huang & Somasundaran, 1997; Tieleman et al., 2000). The I_1/I_3 and α values are reported in 8th and 9th columns of Table 2. They show that the micelles simulated with the Lipid14 parameters are found slightly more ellipsoidal that those obtained from the Amber99SB simulations and this result depends slightly on the simulation starting conditions. Indeed, for the Amber99SB and (Lipid14) simulations, we obtained mean I_1/I_3 and α values of 1.44 ± .16 and .16 ± .03 and (1.55 ± .18 and .19 ± .04). The mean ratios I_1/I_3 are close to the experimental estimate (1.45) (Kakitani et al., 1995; Thiyagarajan & Tiede, 1994). Remarkably, the computed α's are very

Table 2. Average dimensions and shape parameters of the micelles obtained from the pre-assembled (PA) and the two self-assembled (SA) simulations (see main text). R_g and R_M stand to the micelle gyration and effective radii, respectively. The R_M values were computed according to the expression (1) given in the main text. The maximum errors for the radii of gyration (in Å), the MOI I_1, I_2, and I_3 (in a.m.u.nm^2), the axial ration I_1/I_3 and the asymmetry parameter α are .3 Å, .4.10^4 a.m.u.nm^2, .10 and .07, respectively.

Force Field	MD	R_g	R_M	I_1	I_2	I_3	I_1/I_3	α
Amber99SB	PA	18.5	23.9	6.9	5.9	4.8	1.43	.16
	SA1	18.4	23.8	6.0	5.2	4.0	1.50	.17
	SA2	18.5	23.9	6.5	5.7	4.7	1.38	.15
Lipid14	PA	19.1	24.6	6.9	6.1	4.5	1.53	.19
	SA1	19.1	24.6	7.0	6.1	4.4	1.59	.19
	SA2	19.2	24.7	6.9	5.9	4.6	1.50	.17

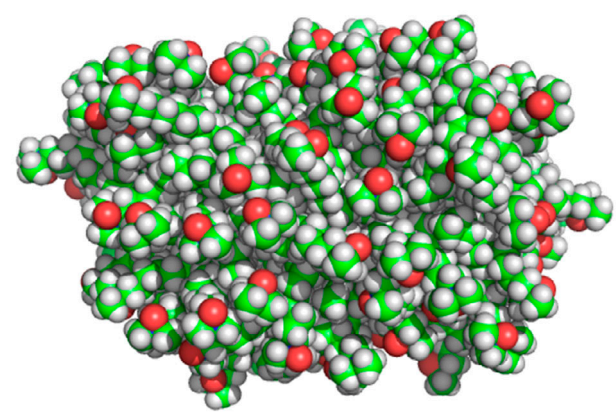


Figure 6. Final snapshot of the Lipid14 SA1 simulation represented as van-der-Waals spheres. The carbon, oxygen, and the hydrogen atoms are in green, red, and white colors, respectively. The nitrogen atoms are hidden by two methyl groups of the headgroup and thus not shown in the Figure. This figure was produced with PyMOL (Schrodinger, 2010).

similar to the experimental values (~.20) obtained from SANS experiments (Barlow et al., 2000; Lorenz et al., 2011; Sterpone et al., 2006). Finally, we notice that the average computed semi-axis of the Amber99SB and the (Lipid14) micelles (a = 28.6 ± .8 Å, b = 24.5 ± .8 Å and c = 20.1 ± .3 Å and (a = 29.8 ± 1.0 Å, b = 23.8 ± .8 Å and c = 19.3 ± .3 Å) (see also Supplementary Table S13) are consistent with the values obtained by Marchetti et al. (2006) from simulations with the CHARMM22 potential (a = 26.6 ± .6 Å b = 20.4 ± .3 and c = 23.9 ± .2) and by SANS experiments (a = 30.5 ± .5 Å and b = c 19.4 ± .2) (Thiyagarajan & Tiede, 1994).

3.1.3. *Micelle microstructure*

To investigate the spatial distribution of the components of the micelle system, we computed their mean radial mass density profile, or $\rho(r)$, as function of the distance r from the micelle center of mass (COM, r = 0 Å, Figure 7). Here, the $\rho(r)$ functions for the solvation water, the LDAO molecules, their heads and tails were computed every 2 ps over the final 50 ns of each simulation. As depicted in Figure 7, the $\rho(r)$'s of these species are found similar independently of the simulation starting conditions but slightly differ with the force field in case of their positions and amplitudes. These differences are caused by the different micelle sizes, shapes and the surfactant alkyl chain conformations (see our discussion in the Section 3.1.6) between the Amber99SB and Lipid14 simulations. More specifically, the computed alkyl chain $\rho(r)$'s extend from 0 to ~20 Å and reach a maximum density of .78 g cm^{-3} near the center of the micelle, which is close to the n-dodecane bulk density at room temperature (.74–.79 g cm^{-3}) (Aicart, Tardajos, & Diaz Pena, 1980; Barlow et al., 2000). We also find that the LDAO headgroup region extends from 16 to ~26 Å in all simulations and its corresponding $\rho(r)$ has a strong peak near ~24 Å, which agrees well with the R_M average value (23.9–24.7 Å) obtained in the previous section. We also remark that a similar result (21.5 Å) was obtained by Lorenz et al. (2011). As far as water is concerned, its corresponding $\rho(r)$'s in Figure 7 shows that the solvent penetrates deeply into the micelle headgroup and shares some contacts with the micelle hydrophobic surface (see the discussion in Section 3.1.5). Finally, differences in the $\rho(r)$ amplitudes between the two force fields is probably caused by the different micelle sizes, shapes, and the surfactant alkyl chain conformations (see Section 3.1.6).

3.1.4. *Volume and Interfacial properties of the micelle*

By neutron reflectometry and SANS experiments, Barlow et al. (2000) and Kakitani et al. (1995) estimated the molecular volume of a LDAO monomer inside the micelle, or V^V_{LDAO}, within 410.2–432.9 Å^3. In order to compare these values with those obtained from our six simulations, we used the *trjVoronoi* program (Abel et al., 2012) that uses some of the voro++ classes and libraries (Rycroft, 2009) to compute, at each simulation time frame, the Voronoi volume of each atom of the system using the Voronoi–Delauney tessellation (Voronoi, 1908). As previously (Abel et al., 2012), explicit hydrogen atoms were excluded in this calculation. Thus, we obtained similar V^V_{LDAO} values for the micelles simulated with the same force field. The average Amber99SB V^V_{LDAO} value (434.2 ± 3.3 Å^3) agrees well with experiment, whereas the V^V_{LDAO} values in the Lipid14 are found, on average, 5.6% larger (451.5 ± 3.9 Å^3). This difference may be related to the readjustment of the methylene proton Lennard-Jones parameters in Lipid14.Using the *trjVoronoi* program, we also compute the solvent accessible surface area (SASA) of the whole LDAO molecule ($\mathrm{SASA_{LDAO}}$) and the headgroup ($\mathrm{SASA_{Head}}$) (Table 3). The SASA was calculated by summing up the surfaces of the Voronoi polyhedral facets shared between the LDAO atoms and water. As suggested by the Figure 6 and the corresponding SASA values reported in the 4th and 5th columns in Table 3, the six micelles have a large part of their surface in contact with water. The computed $\mathrm{SASA_{LDAO}}$ is found around 131.3 Å² for all the six micelles. The average surface per surfactant headgroup ($\mathrm{SASA_{Head}}$) is between 75.9 – 78.2 ± 1.0 Å² and is in agreement with experimental data (70 ± 10 Å²) (Barlow et al., 2000). In the last column of Table 3, we also reported the average fraction of surface, f_{tail}, shared between the LDAO alkyl chains and the solvent (where $f_{\mathrm{tail}} = 1 - (\frac{SASA_{\mathrm{Head}}}{SASA_{\mathrm{LDAO}}})$). The f_{tail} values (~41%) are found similar between the six micelles and thus confirm that the LDAO headgroup are largely solvated by water. For sake of comparison, the $\mathrm{SASA_{Head}}$ value found here is ~20% lower than in Sterpone et al. (2006) for the same micelle size simulated with the CHARMM22 force field (Mackerell, Feig, & Brooks, 2004).

3.1.5. *Micelle hydration properties*

The micelle and the LDAO headgroup hydration was characterized by computing the number of hydration waters per LDAO molecule, n^w_{LDAO}, as a function of the force field and the simulation protocol. To do this, similarly to Lorenz et al. (2011), we counted the number of oxygen water within a distance of 3.5 Å of any LDAO atom. While doing so, we took care to not double count same water atom hydrating two LDAO atoms. The average n^w_{LDAO} value does not depend significantly on the force field and the simulation protocol and is equal to 8.9 ± .2 for all the six simulations. This value is in agreement with SANS experiment (8–12) (Barlow et al., 2000; Kakitani et al., 1995) and previous simulation

Downloaded by [CEA Saclay] at 02:18 22 March 2016

Downloaded by [CEA Saclay] at 02:18 22 March 2016

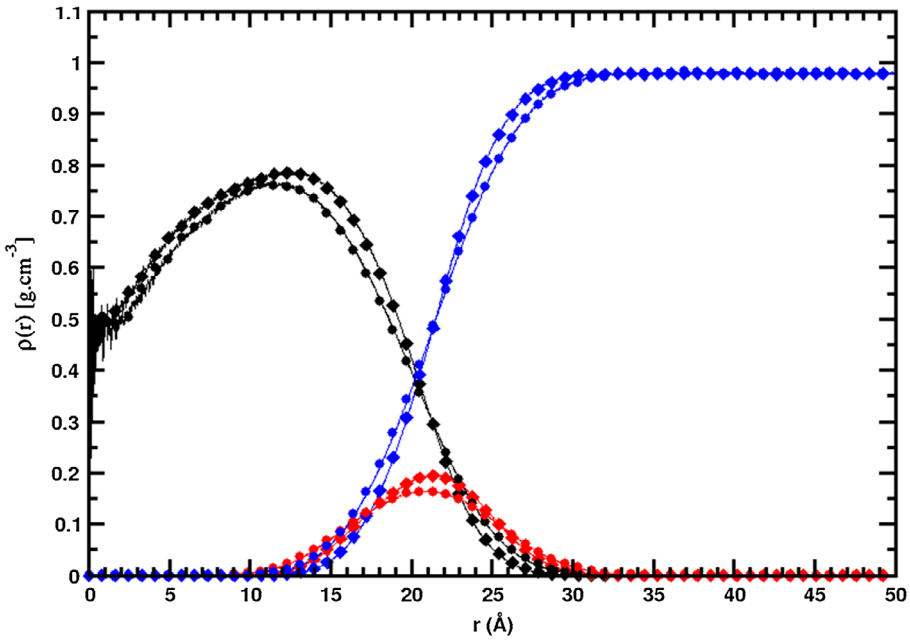


Figure 7. Radial density profiles for the Amber99SB (•) and Lipid14 (♦) micelles. Note that the values for the PA and SA simulations are superposed. Water, LDAO headgroup, and alkyl tail are in blue, red, and black colors, respectively.

Table 3. Average of LDAO volume (in Å^3) and water surface area per LDAO and headgroup (in Å^2) obtained for the PA and the SA simulations f_{tail} is the LDAO fraction (in %) in contact with water (see main text for details). The maximum errors of V^V_{LDAO}, SASA and f_{tail} are 4.0 Å^3, 1.0 Å^2 and .3, respectively.

Force Field	MD	V^V_{LDAO}	$SASA_{LDAO}$	$SASA_{Head}$	f_{tail}
Amber99SB	PA	429.1	131.5	77.0	41.5
	SA1	439.2			
	SA2	434.2	131.3	76.9	
Lipid14	PA	451.5	130.1	75.9	41.6
	SA1	451.6			
	SA2	451.6	133.3	78.2	41.3

(~8–10) (Lorenz et al., 2011). To investigate further the hydration of the LDAO headgroup, we have also computed the mean pair radial density function between the oxygen atom of N-oxide group, O_N and the water oxygen atom, O_w, or $g_{(O_N-O_W)}$(Supplementary Figure S3). Since we used different LJ parameters for the LDAO N and O atoms in the Amber99SB and Lipid14 simulations, it is expected that the $g_{(O_N-O_W)}$curves slightly differ with the force filed. Indeed, we found a strong peak at ~2.5 and 2.7 Å for Amber99SB and Lipid14 simulations, respectively. Moreover, the first peak width, which is related to hydration of the N–O group, is also found slightly higher in the Lipid14 simulations than in the Amber99SB ones. The number of waters in the first shell of the N–O group in the Amber99SB and lipid14 micelle were found near 2.5 and 2.7, respectively, estimated by integrating $g_{(O_N-O_W)}$to its first minimum at ~3.4 Å and do not depend on the simulation starting conditions. These values are found slightly lower than the hydration value computed for trimethylamine N-oxide from *ab initio* MD (~2.9) (Palazzesi, Calvaresi, & Zerbetto, 2011).

Finally, we have also investigated the hydrogen bonds formed between the oxygen of the LDAO N–O group and water. We counted a hydrogen bonds (HB) only if the distance O_w–O_N is within a 3.5 Å cut-off and the O_w–H_w – O_N angle is greater than 150° (Lorenz et al., 2011). We find similar mean HB values of 2.6 ± .2 for all the simulations, slightly higher than Lorenz et al. (2.1) obtained in their simulations with the OPLS-AA force field (Lorenz et al., 2011).

3.1.6. Conformational properties of the LDAO molecules

Differences in the micelle size and alkyl chain $\rho(r)$ functions between the Amber99SB and Lipid14 force fields may suggest distinct conformations for surfactant alkyl tail inside the micelle core. To examine this aspect in more detail, we have first computed the LDAO alkyl chain end-to-end distance, $d_{endToEnd}$, between the first (C_3) and the last carbon (C_{14}). We find that $d_{endToEnd}$ for the LDAO alkyl chain modeled with the Amber99SB force field is significantly higher than the n-dodecane in the Lipid14 micelle (mean values of 12.0 ± .1 and 11.2 ± .1 Å, respectively) which is smaller than the length of an extended n-dodecane alkyl chain (13.8 Å) (Sterpone, Briganti, & Pierleoni, 2009), but close to the values

obtained in previous MD simulations (~13 Å) for different surfactants with the same alkyl chain (Abel et al., 2011, 2012; Holler & Callis, 1989; Palazzesi et al., 2011). A higher value for $d_{endToEnd}$ in the Amber99SB micelles indicates that the LDAO alkyl chains are less folded (i.e. more extended) that in the Lipid14 ones (see a representative example in Supplementary Figure S4). To verify this, we examined the conformations of the alkyl chain for each PA/SA simulation by computing the mean normalized dihedral angle distribution of CCCC dihedral, $P(\Phi)$. From these functions, we computed the percentage of *trans* conformation, *%t* for which the angle Φ is in an interval of 120 and 240° (Figure 8).

The 9 CCCC dihedrals named with a number between 1 and 9 correspond to the sequence of all the dihedrals between dihedral $C_3C_4C_5C_6$ and $C_{11}C_{12}C_{13}C_{14}$ (see Figure 1). As shown in Figure 8, the percentage of *trans* conformation for each CCCC dihedrals are very similar between the SA/PA simulations but vary significantly with the two force fields. Indeed, the *%t* and the *trans/-gauche* ratio (*t/g*) values are found significantly higher in case of the Amber99SB simulations ~75.0% (3.0), compared to in the Lipid14 ones 63.0% (1.7) confirming that in the former micelle the LDAO alkyl chain is less folded (or more ordered) that in the latter ones. The *%t* and the *t/g* values can be compared to the experimental value obtained for the dodecane alkyl chain of SDS (65% and 1.85) obtained by Holler and Callis (1989) from Fourier transform infrared spectroscopy experiment and by Dickson et al. (2014) from MD simulations for the n-tridecane in MD simulations modeled with Lipid14 parameters. Marchetti et al. (2006) found that the LDAO alkyl chain

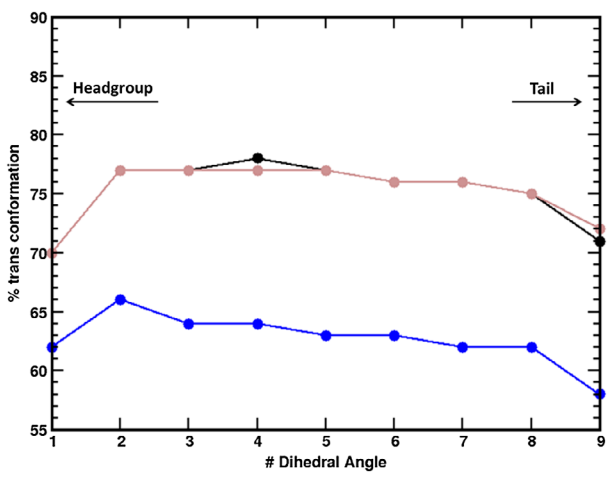


Figure 8. The average percentage of *trans* conformation, for each CCCC dihedral of the LDAO alkyl chain obtained from the Amber99SB SA and PA (in brown and black colors, respectively) and Lipid14 SA and PA (in blue and cyan colors, respectively) simulations. Note that the values of the PA and SA Lipid14 simulations are superposed. Each point was obtained from the last 50 ns of each run.

in micelle modeled with the CHARMM force field *%t* and the *t/g* values of ~70% and 2.33, respectively. The higher *%t* and *t/g* values obtained in the overall Amber99SB simulation are slightly lower than the values we obtained for DPC micelle in our previous report (Abel et al., 2012) (85.4% and *t/g* ratio ≈ 5.84). Nevertheless, this study confirms that the Amber99SB tends to overestimate the CH^2–CH_2–CH_2–CH_2 *trans* conformation and therefore suggests that the Lipid14 force field should be used for modeling surfactant alkyl chain.

Following Palazzesi et al. (2011), we have also examined the orientation of the LDAO molecule inside the micelle core by computing the angle (θ) formed by the vector that joins C3 to C14 carbon atoms and the C3 atom with the micelle center of mass (Figure 9(A)). The average distribution profiles of the θ angle, $P(\theta)$ for the PA/SA Amber99SB and Lipid14 simulations are depicted in Figure 9(B).

As we can see, the LDAO alkyl chain in the Lipid14 and Amber99SB micelles has nearly the same orientation with a θ angle of 20–30° indicating that the LDAO chain are not perfectly aligned to the micelle radial ax. This result does not depend on the force field and the simulation starting conditions. The maximum θ angle distributions found in our simulations are comparable to the value obtained for SDS (~30°) (Palazzesi et al., 2011).

Finally, we have also investigated the mobility of each carbon atom in the LDAO alkyl chain of each micelle as a function of the carbon position, by computing their average translational diffusion coefficients (D_c) from different parts of each trajectory (Figure 10). The diffusion coefficients were obtained from the slope of the mean-square displacement vs. time for each carbon atoms via the Einstein's relation ($D_c = |r(t) - r(0)|^2/6t$). The curves obtained from the simulations have a similar asymmetric "U" shape where the higher D_c values are obtained for the first (C3) and the last (C14) of the alkyl chain carbons indicating that these carbons are the most mobile of the micelle core. A similar trend was also obtained by Palazzesi et al. (2011) and Marchetti et al. (2006) from previous SDS and LDAO micelle simulations, respectively. The Figure 10 also shows that the D_c values do not change significantly (within the statistical error of ±.15 cm² s^{-1}) with the simulation starting conditions or the force field and are in range .4–.9 × 10^{-5} ± .2.10^{-5} cm² s^{-1}. The D_c values are also consistent with those obtained by Sharma et al. from Quasi Elastic Neutron Scattering experiments (.1–1.5 × 10^{-5} cm² s^{-1}) for SDS in micelle (Sharma et al., 2010).

3.2. *Simulations of the DI-LDAO complex*

After the validation of our LDAO RESP charges against experiments and simulations, we have used them to simulate the DI-LDAO complex with different NaCl concentration.

Downloaded by [CEA Saclay] at 02:18 22 March 2016

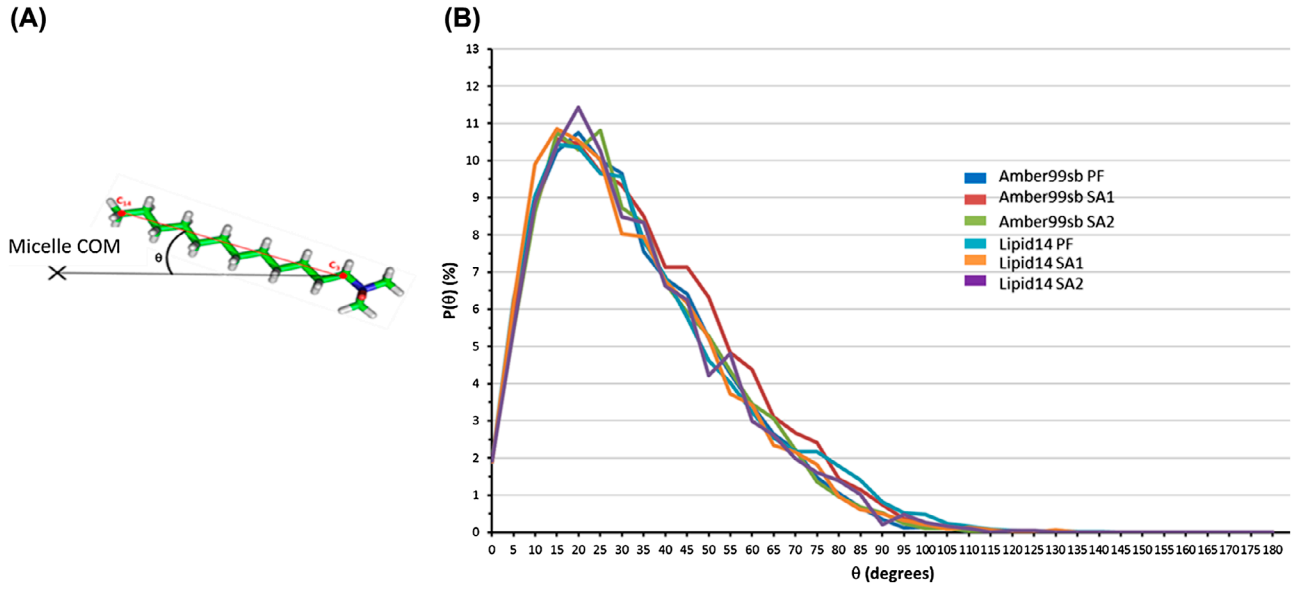


Figure 9. (A and B) Probability distribution $P(\theta)$ of the LDAO alkyl chain angle θ inside the micelle core (see main text for details).

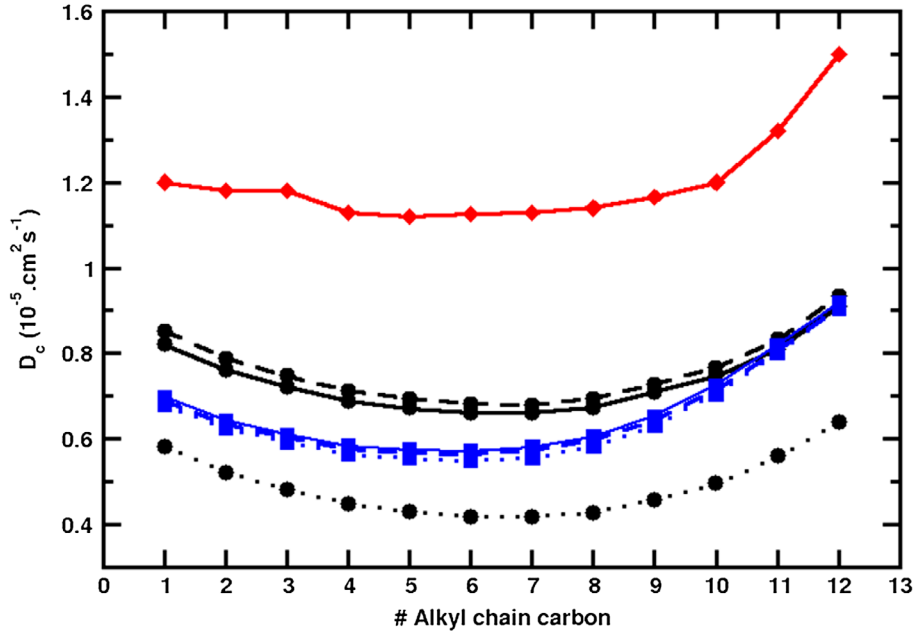


Figure 10. Average diffusion coefficients of the LDAO alkyl carbon atoms in the PA (continue line) and SA (dashed line) Amber99SB (black) and Lipid14 simulations (blue). The 1st and 12th carbons are near the LDAO headgroup and at end of tail, respectively. We compared our results with the MD study of a SDS micelle (Palazzesi et al., 2011) (in the red color).

3.2.1. Construction and simulations of the DI protein–LDAO complex

Since the atomic structure of DI is not available, we turned to the protein modeling for generating atomic models for our simulations. Fortunately, secondary structure prediction algorithms reliably catalog DI as all-helical protein (Harak et al., 2014), well in agreement with UV circular dichroism measurements (data not shown). For a 130-residue all-helical protein, *ab initio* modeling generates a limited range of possible structures

Downloaded by [CEA Saclay] at 02:18 22 March 2016

(Supplementary Figure S5) that usually includes the correct fold. We describe below simulations on the QUARK model (last model of Supplementary Figure S5), that was validated *post hoc* by our experimental results. The QUARK protein model contains six α-helices, five internal coils and has a helix content of 83% (computed with the DSSP program (Kabsch & Sander, 1983; Joosten et al., 2011)) (Figure 11). The position and the lengths of six α-helices are consistent with those obtained from other modeling webservers such as Robetta (Kim et al., 2004), GenThreader (Jones, 1999) or PredictProtein (Rost et al., 2004) (Supplementary Figure S2). Also, to scrutinize the dependency of our simulations on the protein model, we carried out different runs with different starting and simulation conditions (see Table 1) with a number of LDAO molecules determined by experimental means. Indeed, the DI protein can be purified in a non-aggregated form only in the presence of LDAO detergent above its CMC (Figure 12(A)). We used a combination of multi-angle light scattering, refractive index detection, and UV detection coupled to size exclusion chromatography (Figure 12(B)) to determine the mass of LDAO actually associated to DI (Figure 12). The size exclusion peaks of excess LDAO micelles and LDAO-associated his6-DI proved to be too poorly separated to do this (data not shown). We therefore used GST-DI and obtained a protein molecular mass of 90.0 and 18.9 kDa of associated detergent for the main peak (Figure 12(C)). A control with GST only yields a protein molecular mass of 51.1 kDa with no associated detergent. GST is known to form dimers and these values are in good agreement with the molecular masses computed from the protein sequences of dimeric GST-DI (83.7 kDa) and dimeric GST (53.2 kDa). Thus 18.9 kDa of LDAO are associated with 2 DI molecules, or 41 LDAO molecules per DI molecule that justify the use of 41 LDAO in the simulations of the LDAO-protein complex. Based on these findings, we report here of simulations on the DI-LDAO complex where the protein is modeled with the Amber99SB-ILDN parameters (Lindorff-Larsen et al., 2010) and the surfactant with Amber99SB and Lipid14. Figures 13 and S6 depict final snapshots of the LDAO-DI complex with the protein modeled by QUARK in presence of the 41 LDAO molecules and at different NaCl concentrations (Figure 13). As we will see below, concerning the conformation of the protein and its LDAO binding sites, the simulations produce slight different results depending on the simulations starting conditions, the force fields and NaCl concentration. To examine this aspect in details, we firstly analyzed the stability of the DI-LDAO complex and compare with the DI protein in pure water by computing the time evolution of their gyration radii (Figure 14 and S7, respectively).

We first notice that the aggregation of the LDAO around the DI protein depends slightly on the force field and the salt concentration and it took approximatively

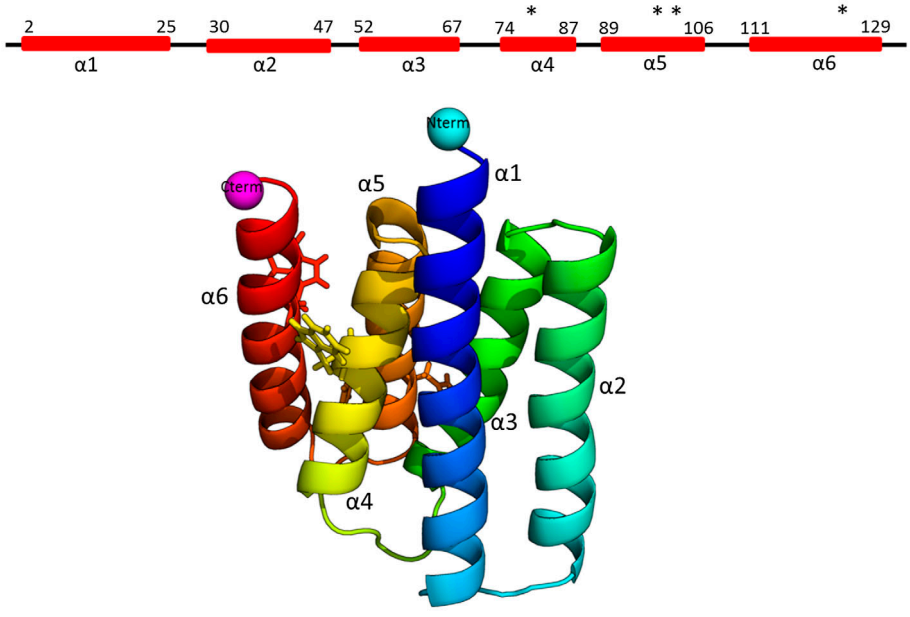


Figure 11. 3D representations of the DI protein model obtained with the QUARK web server. In the upper panel, the six α-helices and five coils are shown in red rectangles and black lines, respectively. The relative positions of the four tryptophans (i.e. Trp82, Trp100, Trp102, and Trp122 in the DI sequence are highlighted with black asterisk and shown in yellow, orange, red sticks in the lower panel. Note that Trp102 is hidden in the DI representation. The lower Figure was produced with PyMOL (Schrodinger, 2010).

Downloaded by [CEA Saclay] at 02:18 22 March 2016

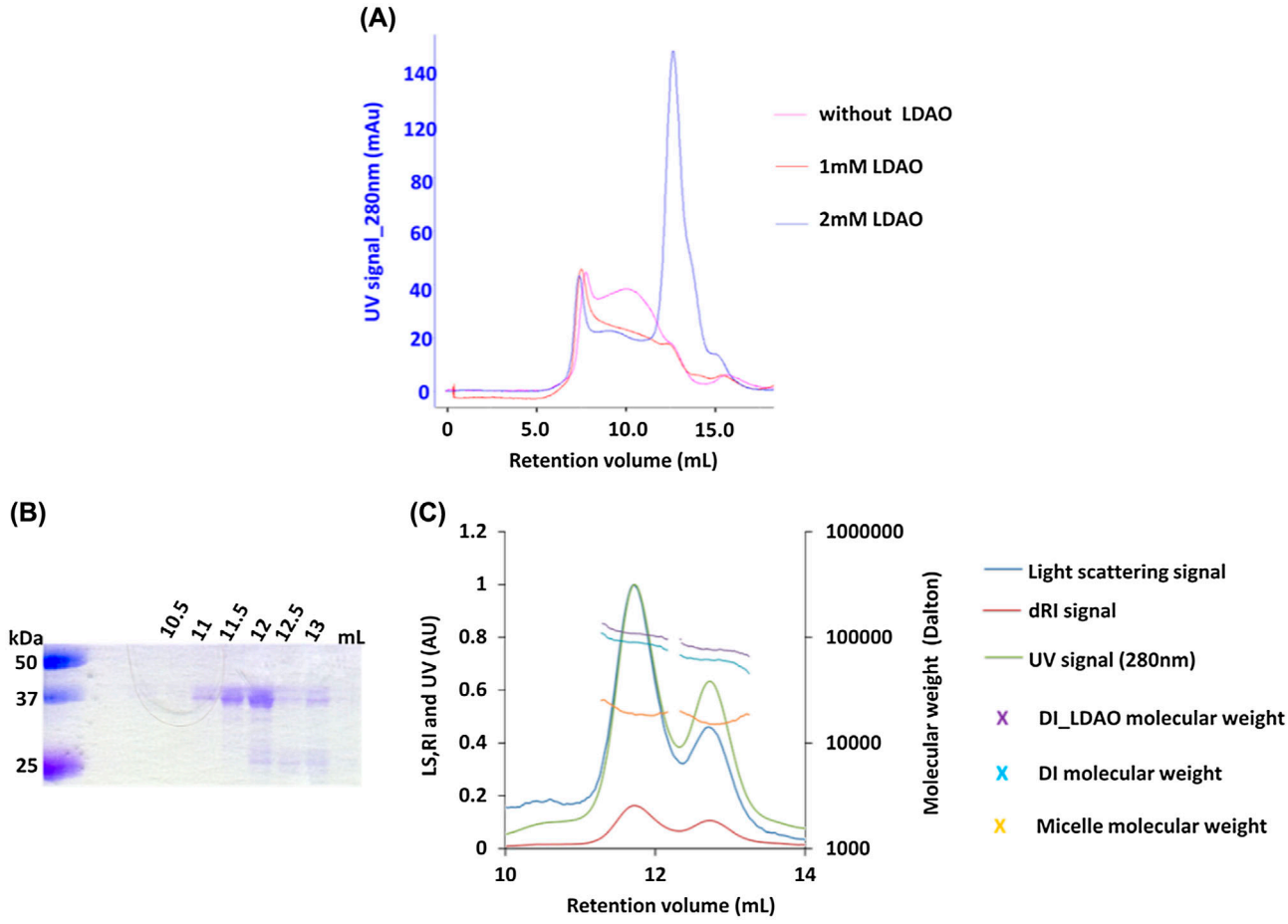


Figure 12. Experimental characterization of LDAO-DI interaction. In Panel A, the size-exclusion chromatography experiments of GST-DI with different amounts of LDAO. The protein was eluted under three conditions: no detergent (pink), detergent under the CMC (1 mM, red) and at the CMC (2 mM, blue). The experiments were carried out on Superdex 200 10/300 GL column during the purification of DI. In Panel C, determination of the aggregation number of the detergent molecules in the complex GST-DI with LDAO using SEC-MALLS experiment. GST-DI alone was first eluted following by a heterogeneous mix of GST-DI plus DI composing the second picture of the chromatogram. The corresponding polyacrylamide SDS-Page is displayed on the left.

Downloaded by [CEA Saclay] at 02:18 22 March 2016

30–45 ns for that the DI and 41 LDAO form a stable complex with constant R_g values around 19 Å. An examination of the snapshots at simulation times (not shown) show that in all the simulations the LDAO forms stable aggregates around the protein and LDAO binding sites do not change significantly during the course of the runs. The sizes of the protein alone (i.e. no LDAO included in the R_g calculations) are found similar to that of the protein R_g (15 Å) in pure water in the two NaCl concentrations (Supplementary Figure S7). We also notice that in one Lipid14 simulation carried out at 0 mM (see Figure 14(B)), the DI-LDAO R_g slightly increases indicating significant changes of the protein initial structure in this system. These results indicate that, here, R_g is not sufficient to discriminate the structural differences between the simulations. Thus, we have turned to examining other structural properties of the DI model such as the stability of the protein initial conformations; the secondary structure changes and the accessibility of water of specific protein residues (see next sections).

3.2.2. *Stability of the model initial structures*

The stability of the protein initial conformation was examined by computing the time evolution of the proteins C_α root mean square deviations (RMSD) in pure water and with LDAO at different NaCl concentration. The minimized DI protein structure was used as a reference in the RMSD calculations (Figure 15).

In pure water (Figure 15A), the protein initial conformation at 0 mM is less conserved than at 300 mM of NaCl with a RMSD values around 4 and 5 Å, respectively. In presence of LDAO in the medium, except two simulations (black and red curves in Figure 15(C) and

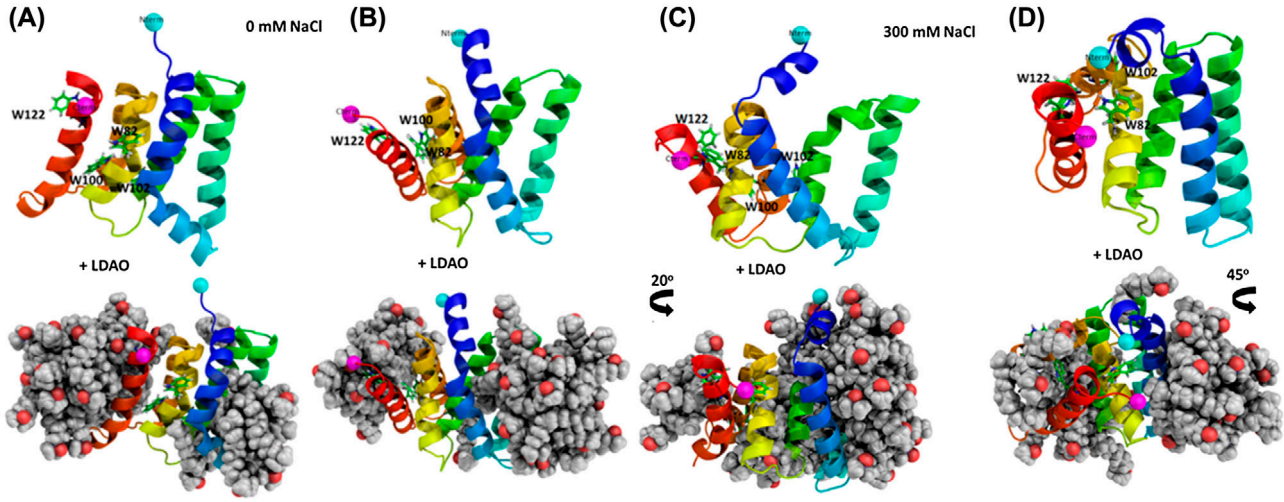


Figure 13. Final snapshots of the first Amber99SB(A,C) and Lipid14(B,D) simulations at different NaCl concentration. The four tryptophans are showed in sticks. In gray, the LDAO molecules and the DI protein is colored blue to red from N- to C-terminus. These figures were produced with PyMOL (Schrodinger, 2010).

Downloaded by [CEA Saclay] at 02:18 22 March 2016

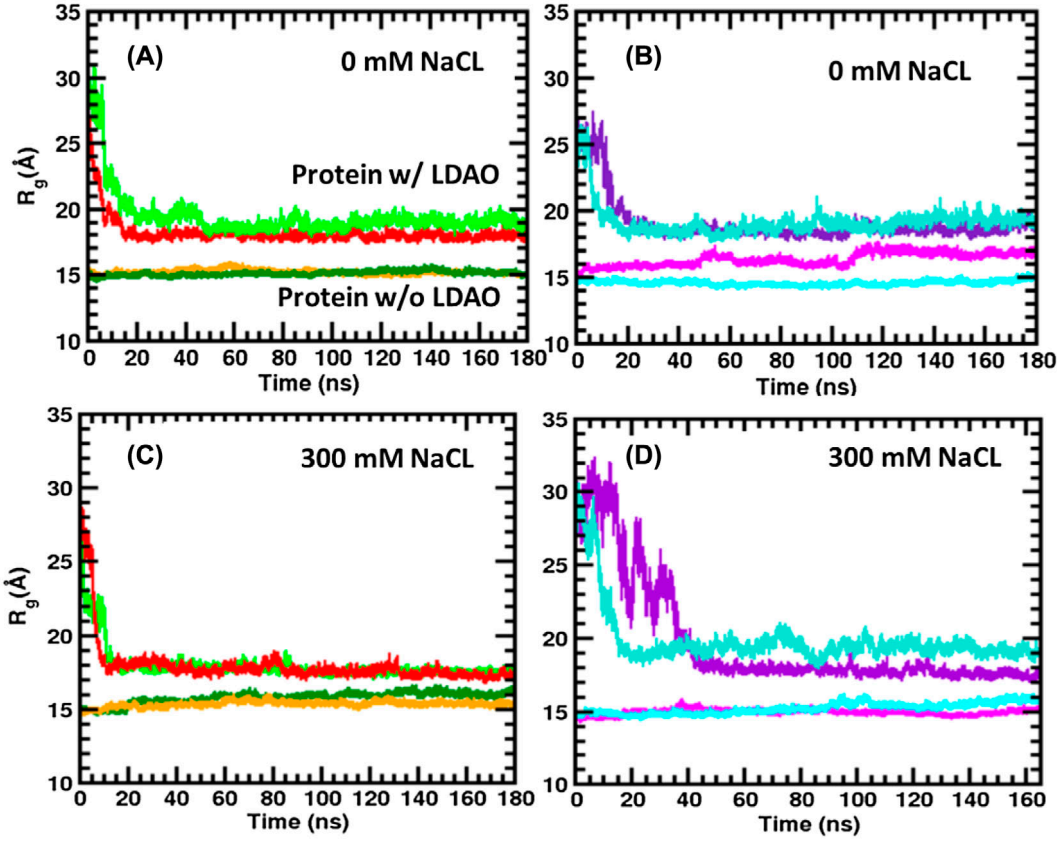


Figure 14. Time evolution of radius of gyration of the protein alone and the protein–LDAO complex with Amber99SB (A and C) and Lipid14 force fields (B and D). The R_g, the first and second replica of SA Amber99SB and the Lipid14 simulations are shown in green (dark green), red (orange), dark cyan (cyan), and violet (pink), respectively.

Downloaded by [CEA Saclay] at 02:18 22 March 2016

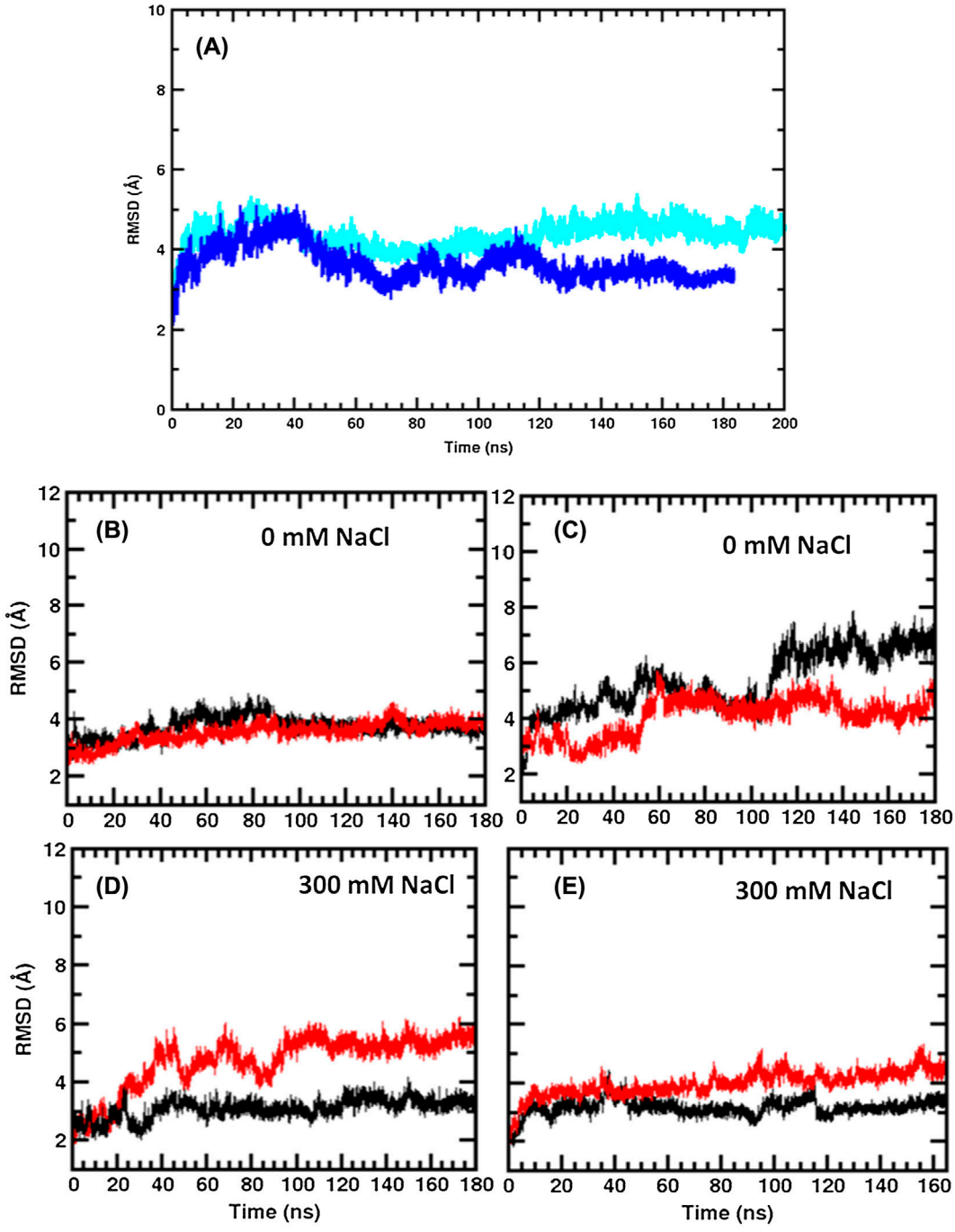


Figure 15. Time evolution of the RMSD of the α-carbons of the protein alone (A) and in the LDAO complex simulated with Amber99SB (B and D) and Lipid14 force fields (C and E). In the panel A, the RMSD of the protein simulated at 0 mM and 300 mM are shown in blue and cyan colors, respectively. In the panels B-E, the first and second SA simulations are shown in red and black colors, respectively.

(D)), the protein initial conformation is stabilized during all the course of the simulation time with a RMSD around 4.0 Å. The high change in the protein RMSD during the last 40 ns of the Lipid14 SA2 simulation (black curve in Figure 15(C)) is consistent with the increase in the R_g of the protein depicted in Figure 14(B).

3.2.3. *Time evolution of the DI secondary structure*

To link the protein initial conformation changes with the protein secondary structure (SS) change during the simulations, we computed the time evolution of the protein SS in pure water and with LDAO at different NaCl

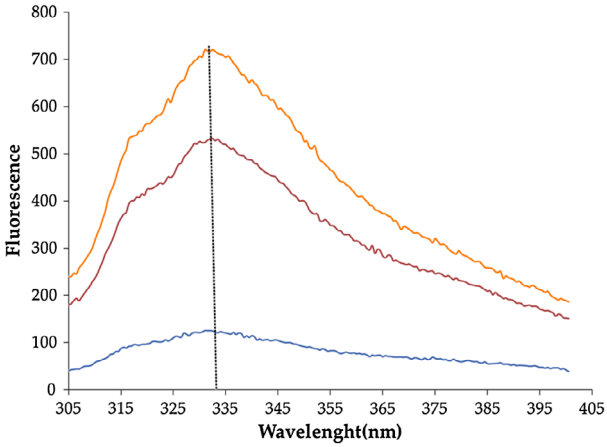


Figure 16. Intrinsic fluorescence emission spectra of his6-DI in LDAO buffer upon excitation at 295 nm. The emission wavelength seems to be independent from the protein concentration with a maximum consistently found at 331 nm. Concentrations are 2.7, 1.8, and .45 μg/ml for the curves in orange, maroon and blue, respectively.

concentration (Supplementary Figures S8–11). The figures show that the protein SS in pure water is significantly changed at 300 mM of salt with a decrease in the helix content of 67 to 57% (computed from the last 40 ns of each run), respectively. Specifically, the denaturation of the protein SS occurs mainly at the beginning of the 1st and in the 2nd α-helix (see Figure 13 and Supplementary Figure S6). In presence of surfactant, and at the two salt concentrations, the protein SS is stabilized with a helix contents around 68%. This value does not change; the LDAO force field or the simulation starting conditions (see Supplementary Figures S10 and S11).

3.2.4. Interfacial properties of the Protein–LDAO micelle complex

According to Figure 13 and Supplementary Figure S6, the LDAO molecules do not bind exactly to the same protein region depending on the force field and the salt concentration. In order to examine this, we analyzed the number of unique contacts (N_{LDAO}) between the LDAO and the protein. To obtain these numbers, we counted a unique contact between a protein residue and LDAO if the distance, d, between the LDAO heavy alkyl chain atoms and a protein residue is smaller than 4.0 Å (Böckmann & Caflisch, 2005). The total N_{LDAO} values computed from the last 30 ns do not change significantly with force field and the starting conditions but slightly increase with the salt concentration. Indeed, we found that in the simulation with 0 mM of NaCl, the protein in the Amber99SB and Lipid14 simulations has, on average, 26.5 and 27.5 ± 1.0 LDAO in direct contact, respectively. With the increase in the salt concentration, the number of contacts increases slightly up to 30.5 and 29.0 ± 1.0, respectively. Comparison of the local coverage of the protein at 0 and 300 mM and between the different Amber99SB and the Lipid14 replicas is shown in Supplementary Figures S12–13. The figures show that the LDAO molecules bind mainly the 1st, 3rd, 5th, and 6th helix.

Figure 13 also shows that the four DI tryptophans (Trp 82, Trp100, Trp102, and Trp122) are also covered by the surfactant. In Supplementary Figure S14, we present the hydropathicity plot of the DI protein computed with the Kyte–Doolittle approach and with a default window size of 9 residues (Kyte & Doolittle, 1982), as obtained from the ExPASy server (Wilkins et al., 1999) with the positions of the four tryptophan residues (i.e. Trp82 Trp100, Trp102, and Trp122) highlighted with red triangles. A positive hydropathicity score signals hydrophobicity; conversely, negative values are found for hydrophilic residues. By examining the hydropathicity plot, we predict a higher hydrophobic character for Trp100, Trp102, Trp122, whereas for Trp82 the hydropathicity score is close to 0. We then expect that these three latter tryptophans will be in an apolar environment in contrast to the former one. Experimentally, the

Table 4. Accessibility of the four protein tryptophans computed from the last 30 ns of each run. +, ±, and − mean that the corresponding tryptophan is buried, partially buried or not buried, respectively, during the last 30 ns of each run (see main text).

	Trp 82	Trp 100	Trp 102	Trp 122
Amber99SB 0 mM NaCl				
SA1	−	+	+	+
SA2	−	+	+	+
Amber99SB 300 mM NaCl				
SA1	+	+	+	+
SA2	+	+	+	−
Lipid14 0 mM NaCl				
SA1	±	+	+	+
SA2	±	+	+	+
Lipid14 300 mM NaCl				
SA1	±	+	+	+
SA2	±	+	+	+

Downloaded by [CEA Saclay] at 02:18 22 March 2016

Downloaded by [CEA Saclay] at 02:18 22 March 2016

hydrophobic/hydrophilic character (i.e. buried/covered by LDAO or not) of these tryptophans was investigated by fluorescence spectroscopy with a his6-DI in LDAO buffer.

The emission spectrum upon excitation at 295 nm displays a clear maximum at 331 nm (Figure 16). To relate these results with our simulations, we computed the time evolution of the surface area in contact with water of the four tryptophans, $SASA_x$ (where x stands for the tryptophan in the DI sequence) during the last 30 ns of each run with the *trjVoronoi* program (Supplementary Figures S15 and S16). We consider that a given tryptophan is buried if it is covered by more than one LDAO and its SASA in the complex is smaller than its SASA value in pure water during the last 30 ns of each run. By comparing the protein–LDAO binding sites in Supplementary Figures S12–S13 and the SASA in Supplementary Figures S15–S16 for each Trp, we constructed the Table 4 that summarizes the results regarding the Trp accessibility to water (Table 4). Table 4 indicates that the three tryptophans located in 5th and 6th helix (i.e. Trp100, Trp102, and Trp122) are mainly buried and not exposed to water. That three out of four Trp are not accessible to water is consistent with the fluorescence experiments described above. Indeed, an emission maximum at 331 nm rules out more than one exposed tryptophan (the maximum would be above 340 nm, and above 346 nm for four fully exposed tryptophans) but also indicates that not all tryptophans are in completely apolar environments (the maximum would be below 325 nm) (Shen et al., 2008). Finally, we point out that the tryptophan accessibility to water can sporadically change during the simulation time (see for instance Trp102 see Supplementary Figures S15 and S16 (G)).

4. Conclusion

This paper has been devoted to the development and testing of original RESP atomic charges for the LDAO surfactant and its analog compatible with the AMBER force fields for proteins. We tested and validated these charges with MD simulations of two systems. In the first system, we simulated a pure micelle in water composed of 104 LDAO molecules modeled with the Amber99SB and Lipid14 force fields. We performed simulations with different initial conditions using pre-assembled micelles and self-assembling random distributions of LDAO molecules. We show that micelle characteristics agree well with previous experimental and simulation results. In particular, we find that the simulated micelles have an ellipsoidal shape with a major over minor axis ratio around 1.4 and an effective size around 22 Å. On average, the number of water molecules bound to a single LDAO molecule is ~9 and each LDAO headgroup has ~2.6 waters in their hydration shell, thus forming ~2.5 hydrogen bonds. In case of the alkyl chain conformation, simulations show significant differences between the two force fields. Moreover, in agreement with a previous report, we show that the Amber99SB force field tends to overestimate the *trans* population of the alkyl chain CH_2–CH_2–CH_2–CH_2 torsion compared to experiments (~75.0% vs. 65%) and this problem is resolved using the Lipid14 parameters.

Furthermore, we have examined the structural properties of a complex formed by LDAO and the DI protein, a segment of PI4KA in contact with the cellular membrane. In absence of an experimental 3D structure of the DI protein, a model (the last on Supplementary Figure S5) was constructed with the QUARK protein modeling protocol. This webserver predicts a protein with six helices and five coils, in agreement, with results obtained with other modeling server such as Robetta, GenThreader, and PredictProtein. Thus, we carried out two sets of MD simulations of >180 ns each with 41 LDAO molecules and one protein. The protein was modeled with the Amber99SB-ILDN force field and LDAO with the same force field used in the pure micelle simulations. We carried out different runs with different starting conditions at two NaCl concentrations of 0 and 300 mM to examine the effect of excess of salt on the simulation results and mimic the experimental conditions. The simulations show a small dependence of the simulation results with the starting conditions and the NaCl concentration. Finally, we find that three out of four tryptophans (i.e. Trp100, Trp102, and Trp122) located in the 5th and 6th helix of the protein, are not accessible to water and are considered as buried, in agreement with our fluorescence spectroscopy experiments. We actually performed simulations with another possible model of DI with a different fold (the first on Supplementary Figure S5) generated from the I-TASSER server. Unlike the results with the QUARK model, the results with the I-TASSER model (not shown) were not consistent with our SEC-MALLS and fluorescence measurements, allowing this alternate model to be discarded.

To conclude, the usefulness of the present derivation is attested both by its ability to recapitulate published measurements of pure LDAO micelle parameters and by its use in the simulations of DI-LDAO models. In the latter case, simulations allow discrimination between alternate models with little experimental data in a case with some ambiguity on the protein's fold. The charges here derived are now available for the simulation of further LDAO-containing systems.

Supplementary material

The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2015.1135822.

Author Contributions

E.K. and C.T. performed the experiments, analyzed the data, S.A, S.B, MM, and B.R conceived and designed the project. S.A, EK C.T, M.M and S.B wrote the manuscript. The manuscript was finalized through contributions of all the authors.

Acknowledgments

We acknowledge Christophe Velours of the I2BC core biophysical platform for SEC-MALLS analyses. This work has benefited from the facilities and expertise of the Macromolecular Interaction Platform of I2BC (UMR 9198 – http://www.i2bc-saclay.fr/spip.php?article280). We also acknowledge Dr Stéphane Roche for advice and assistance in fluorescence spectroscopy and Pr. François-Yves Dupradeau (U. of Amiens, France) for helpful discussions and critical reading of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the granted access to the HPC resources of CCRT/CINES made by GENCI (Grand Equipement National de Calcul Intensif) under the allocations [grant number t2013077082], [grant number t2014077082]. We thank also "Maison de la Simulation" and the CEA for the PhD financial support of EK. This work was also supported by ANRS (France REcherche Nord & sud SIDA – hiv Hépatites: FRENSH) through a grant to SB and a predoctoral fellowship to CT and by the French Infrastructure for Integrated Structural Biology (FRISBI) [grant number ANR-10-INSB-05-01].

ORCID

Stéphane Abel http://orcid.org/0000-0002-1980-0839

References

Abel, S., Dupradeau, F.-Y., & Marchi, M. (2012). Molecular dynamics simulations of a characteristic DPC micelle in water. *Journal of Chemical Theory and Computation, 8*, 4610–4623. Retrieved September 7, 2015, doi:10.1021/ct3003207

Abel, S., Dupradeau, F.-Y., Raman, E. P., MacKerell, A. D., & Marchi, M. (2011). Molecular simulations of dodecyl-β-maltoside micelles in water: Influence of the headgroup conformation and force field parameters. *The Journal of Physical Chemistry B, 115*, 487–499.

Abel, S., Lorieau, A., de Foresta, B., Dupradeau, F.-Y., & Marchi, M. (2014). Bindings of hMRP1 transmembrane peptides with dodecylphosphocholine and dodecyl-β-d-maltoside micelles: A molecular dynamics simulation study. *Biochimica et Biophysica Acta (BBA) – Biomembranes, 1838*, 493–509. Retrieved March 1, 2014, from http://linkinghub.elsevier.com/retrieve/pii/S0005273613003659

Abel, S., Sterpone, F., Bandyopadhyay, S., & Marchi, M. (2004). Molecular modeling and simulations of AOT−water reverse micelles in isooctane: Structural and dynamic properties. *The Journal of Physical Chemistry B, 108*, 19458–19466. Retrieved July 27, 2014, from http://dx.doi.org/10.1021/jp047138e

Aicart, E., Tardajos, G., & Diaz Pena, M. (1980). Isothermal compressibility of cyclohexane + N-hexane, cyclohexane + N-heptane, cyclohexane + N-octane, and cyclohexane + N-nonane. *Journal of Chemical & Engineering Data, 25*, 140–145.

Balla, T. (2013). Phosphoinositides: Tiny lipids with giant impact on cell regulation. *Physiological Reviews, 93*, 1019–1137.

Barlow, D. J., Lawrence, M. J., Zuberi, T., Zuberi, S., & Heenan, R. K. (2000). Small-angle neutron-scattering studies on the nature of the incorporation of polar oils into aggregates of N, N-dimethyldodecylamine-N-oxide. *Langmuir, 16*, 10398–10403. Retrieved October 8, 2012, from http://pubs.acs.org/doi/abs/10.1021/la0002233

Bayly, C. I., Cieplak, P., Cornell, W. D., & Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *The Journal of Physical Chemistry, 97*, 10269–10280. Retrieved March 9, 2012, from http://pubs.acs.org/doi/abs/10.1021/j100142a004

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., & Hermans, J. (1981). Interaction models for water in relation to protein hydration. In B. Pullman (Ed.), *Intermolecular forces, The Jerusalem Symposia on quantum chemistry and biochemistry*. (pp. 331–342). Netherlands: Springer.

Berger, O., Edholm, O., & Jähnig, F. (1997). Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical Journal, 72*, 2002–2013. Retrieved April 7, 2012, from http://www.sciencedirect.com/science/article/B94RW-4V8X6TH-4/2/32a85e710fd97894bf4e2e872fd51285

Böckmann, R. A., & Caflisch, A. (2005). Spontaneous formation of detergent micelles around the outer membrane protein OmpX. *Biophysical Journal, 88*, 3191–3204.

Bongiorno, D., Ceraulo, L., Indelicato, S., Liveri, V. T., & Indelicato, S.. (2015). Charged supramolecular assemblies of surfactant molecules in gas phase. *Mass Spectrometry Reviews*. Retrieved from http://doi.wiley.com/10.1002/mas.21476

Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics, 126*, 014101. Retrieved May 21, 2013, from http://link.aip.org/link/?JCP/126/014101/1

Cézard, C., Trivelli, X., Aubry, F., Djedaïni-Pilard, F., & Dupradeau, F.-Y. (2011). Molecular dynamics studies of native and substituted cyclodextrins in different media: 1. Charge derivation and force field performances. *Physical Chemistry Chemical Physics, 13*, 15103–15121. Retrieved January 5, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/21792425

Cieplak, P., Cornell, W. D., Bayly, C. I., & Kollman, P. A. (1995). Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *Journal of Computational Chemistry, 16*, 1357–1377. Retrieved March 31, 2012, from http://dx.doi.org/10.1002/jcc.540161106

Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science, 221*, 709–713.

Cordomí, A., Caltabiano, G., & Pardo, L. (2012). Membrane protein simulations using AMBER force field and berger

Downloaded by [CEA Saclay] at 02:18 22 March 2016

Downloaded by [CEA Saclay] at 02:18 22 March 2016

lipid parameters. *Journal of Chemical Theory and Computation, 8*, 948–958. Retrieved February 12, 2012, from http://pubs.acs.org/doi/abs/10.1021/ct200491c

Dickson, C. J., Madej, B. D., Skjevik, Å. A., Betz, R. M., Teigen, K., Gould, I. R., & Walker, R. C. (2014). Lipid14: The Amber lipid force field. *Journal of Chemical Theory and Computation, 10*, 865–879. Retrieved March 25, 2014, from http://pubs.acs.org/doi/abs/10.1021/ct4010307

Dupradeau, F. Y., Cézard, C., Lelong, R., Stanislawiak, É., Pêcher, J., Delepine, J. C., & Cieplak, P. (2008). R.E.DD.B.: A database for RESP and ESP atomic charges, and force field libraries. *Nucleic Acids Research, 36*, D360–67. doi: 10.1093/nar/gkm887

Dupradeau, F.-Y., Pigache, A., Zaffran, T., Savineau, C., Lelong, R., Grivel, N., ... Cieplak, P. (2010). The R.E.D. tools: Advances in RESP and ESP charge derivation and force field library building. *Physical Chemistry Chemical Physics, 12*, 7821–7839.Retrieved from http://dx.doi.org/10.1039/C0CP00111B

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics, 103*, 8577–8593. doi: 10.1063/1.470117

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., … Fox, D. J. (2009). *Gaussian 09, Revision D.01*.

Gouin, S. G., Vanquelef, E., Fernández, J. M. G., Mellet, C. O., Dupradeau, F. Y., & Kovensky, J. (2007). Multi-mannosides based on a carbohydrate scaffold: Synthesis, force field development, molecular dynamics studies, and binding affinities for lectin Con A. *The Journal of Organic Chemistry, 72*, 9032–9045. Retrieved from http://pubs.acs.org/doi/abs/10.1021/jo071248a

Harak, C., Radujkovic, D., Taveneau, C., Reiss, S., Klein, R., Bressanelli, S., & Lohmann, V. (2014). Mapping of functional domains of the lipid kinase phosphatidylinositol 4-kinase type III alpha involved in enzymatic activity and hepatitis C virus replication. *Journal of Virology, 88*, 9909–9926. doi: 10.1128/JVI.01063-14

Hehre, W. J. (1976). Ab initio molecular orbital theory. *Accounts of Chemical Research, 9*, 399–406.

Herrmann, K. W. (1962). Non-ionic-cationic micellar properties of dimethyldodecylamine oxide. *The Journal of Physical Chemistry, 66*, 295–300.

Hess, B. (2007). P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation, 4*, 116–122. Retrieved from http://dx.doi.org/10.1021/ct700200b

Holler, F., & Callis, J. B. (1989). Conformation of the hydrocarbon chains of sodium dodecyl sulfate molecules in micelles: An FTIR study. *The Journal of Physical Chemistry, 93*, 2053–2058. Retrieved from http://dx.doi.org/10.1021/j100342a068

Hoover, W. G. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Physical Review. A, 31*, 1695–1697.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics, 65*, 712–725. Retrieved March 29, 2012, from http://dx.doi.org/10.1002/prot.21123

Huang, L., & Somasundaran, P. (1997). Theoretical model and phase behavior for binary surfactant mixtures. *Langmuir, 13*, 6683–6688.

Jones, D. T. (1999). GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences. *Journal of Molecular Biology, 287*, 797–815.

Joosten, R. P., te Beek, T. A. H., Krieger, E., Hekkelman, M. L., Hooft, R. W. W., Schneider, R., ... Vriend, G. (2011). A series of PDB related databases for everyday needs. *Nucleic Acids Research, 39*, D411–D419. doi: 10.1093/nar/gkq1105

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics, 79*, 926–935. Retrieved from http://link.aip.org/link/?JCP/79/926/1

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS All-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society, 118*, 11225–11236. Retrieved from http://pubs3.acs.org/acs/journals/doilookup?in_doi=10.1021/ja9621760

Joung, I. S., & Cheatham, T. E. (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated Biomolecular simulations. *The Journal of Physical Chemistry B, 112*, 9020–9041. Retrieved November 15, 2012, from http://dx.doi.org/10.1021/jp8001614

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers, 22*, 2577–2637.

Kakitani, M., Imae, T., & Furusaka, M. (1995). Investigation of mixed micelles of dodecyldimethylamine oxide and sodium dodecyl sulfate by SANS: Shape, size, charge, and interaction. *The Journal of Physical Chemistry, 99*, 16018–16023.

Kamath, G., Guvench, O., & MacKerell, Jr., A. D. (2008). CHARMM additive all-atom force field for acyclic carbohydrates and inositol. *Journal of Chemical Theory and Computation, 4*, 765–778. doi:10.1021/ct800019u.

Karjiban, R. A., Basri, M., Rahman, M. B., & Salleh, A. B. (2012). Molecular dynamics simulation of palmitate ester self-assembly with diclofenac. *International Journal of Molecular Sciences, 13*, 9572–9583.

Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research, 32*, W526–W531.

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., & Woods, R. J. (2008). GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. Journal of Computational Chemistry, 29*, 622–655. Retrieved October 28, 2013, from http://dx.doi.org/10.1002/jcc.20820

Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., ... Pastor, R. W. (2010). Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *The Journal of Physical Chemistry B, 114*, 7830–7843. Retrieved July 31, 2012, from http://dx.doi.org/10.1021/jp101759q

Kocherbitov, V., Veryazov, V., & Söderman, O. (2007). Hydration of trimethylamine-N-oxide and of dimethyldodecylamine-N-oxide: An ab initio study. *Journal of Molecular Structure: THEOCHEM, 808*, 111–118. Retrieved October 20, 2012, from http://linkinghub.elsevier.com/retrieve/pii/S016612800700019X

Kumar, A., Hajjar, E., Ruggerone, P., & Ceccarelli, M. (2010). Molecular simulations reveal the mechanism and the determinants for ampicillin translocation through OmpF. *The Journal of Physical Chemistry B, 114*, 9608–9616.

Retrieved January 25, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/20590090

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology, 157*, 105–132. Retrieved February 19, 2013, from http://www.ncbi.nlm.nih.gov/pubmed/7108955

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., & Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins, 78*, 1950–1958. doi: 10.1002/prot.22711

Lorenz, C. D., Hsieh, C.-M., Dreiss, C. A., Lawrence, M., & Lawrence, J. M. (2011). Molecular dynamics simulations of the interfacial and structural properties of dimethyldodecylamine-N-oxide micelles. *Langmuir, 27*, 546–553. Retrieved August 25, 2011, from http://www.ncbi.nlm.nih.gov/pubmed/21166438

Lutton, E. S. (1966). Phase behavior of the dimethyldodecyl amine oxide-H2O system. *Journal of the American Oil Chemists Society, 43*, 28–30.

Mackerell, A. D., Feig, M., & Brooks, C. L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry, 25*, 1400–1415. Retrieved July 22, 2012, from http://dx.doi.org/10.1002/jcc.20065

Marchetti, G., Marchi, M., & Maire, M. L. (2006). Modélisation Moléculaire Du Phénomène Du Transport Du Calcium Dans La Protéine ATPAse-Ca2+ (SERCA1a) (PhD). pages 165.

Martínez, L., Andrade, R., Birgin, E. G., & Martínez, J. M. (2009). PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry, 30*, 2157–2164. Retrieved October 16, 2014, from http://www.ncbi.nlm.nih.gov/pubmed/19229944

Neale, C., Ghanei, H., Holyoake, J., Bishop, R. E., Privé, G. G., & Pomès, R. (2013). Detergent-mediated protein aggregation. *Chemistry and Physics of Lipids, 169*, 72–84 doi: 10.1016/j.chemphyslip.2013.02.005

Nosé, S. (1984). A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics, 81*, 511–519.

Nyrkova, I. A., & Semenov, A. N. (2005). On the theory of micellization kinetics. *Macromolecular Theory and Simulations, 14*, 569–585.

Oostenbrink, C., Soares, T. A., van der Vegt, N. F. A., & van Gunsteren, W. F. (2005). Validation of the 53A6 GROMOS force field. *European Biophysics Journal, 34*, 273–284. Retrieved July 6, 2011, from http://dx.doi.org/10.1007/s00249-004-0448-6

Palazzesi, F., Calvaresi, M., & Zerbetto, F. (2011). A molecular dynamics investigation of structure and dynamics of SDS and SDBS micelles. *Soft Matter, 7*, 9148–9156.

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics, 52*, 7182–7190.

Poger, D., Van Gunsteren, W. F., & Mark, A. E. (2010). A new force field for simulating phosphatidylcholine bilayers. *Journal of Computational Chemistry, 31*, 1117–1125. Retrieved from http://dx.doi.org/10.1002/jcc.21396

Privé, G. G. (2007). Detergents for the stabilization and crystallization of membrane proteins. *Methods, 41*, 388–397. Retrieved November 5, 2012, from http://www.ncbi.nlm.nih.gov/pubmed/17367711

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., ... Lindahl, E. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics, 29*, 845–854 doi: 10.1093/bioinformatics/btt055

Rahman, A., & Stillinger, F. H. (1971). Molecular dynamics study of liquid water. *The Journal of Chemical Physics, 55*, 3336–3359.

Rost, B., Yachdav, G., & Liu, J. (2004). The predictprotein server. *Nucleic Acids Research, 32*, W321–W326.

Rycroft, C. H. (2009). VORO++: A three-dimensional Voronoi cell library in C++. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 19*, 041111. Retrieved December 3, 2011, from http://www.ncbi.nlm.nih.gov/pubmed/20059195

Sanders, J. M., Wampole, M. E., Chen, C.-P., Sethi, D., Singh, A., Dupradeau, F.-Y., ... Wickstrom, E. (2013). Effects of hypoxanthine substitution in peptide nucleic acids targeting KRAS2 oncogenic mRNA molecules: Theory and experiment. *The Journal of Physical Chemistry B, 117*, 11584–11595. doi: 10.1021/jp4064966

Sardet, C., Tardieu, A., & Luzzati, V. (1976). Shape and size of bovine rhodopsin: A small-angle X-ray scattering study of a rhodopsin-detergent complex. *Journal of Molecular Biology, 105*, 383–407.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical Journal, 78*, 1606–1619.

Schrodinger, L. (2010). *The PyMOL molecular graphics system*. DeLano Scientific San Carlos CA. San Carlos, CA: Schrödinger, LLC. Retrieved from http://www.pymol.org/

Sharma, V. K., Mitra, S., Verma, G., Hassan, P. A., Garcia Sakai, V., & Mukhopadhyay, R. (2010). Internal dynamics in SDS micelles: Neutron scattering study. *The Journal of Physical Chemistry B, 114*, 17049–17056. Retrieved from http://dx.doi.org/10.1021/jp108274y

Shen, C., Menon, R., Das, D., Bansal, N., Nahar, N., Guduru, N., ... Reshetnyak, Y. K. (2008). The protein fluorescence and structural toolkit: Database and programs for the analysis of protein fluorescence and structural data. *Proteins: Structure, Function, and Genetics, 71*, 1744–1754. doi: 10.1002/prot.21857

Sterpone, F., Briganti, G., & Pierleoni, C. (2009). Sphere versus cylinder: The effect of packing on the structure of non-ionic C 12 E 6 micelles. *Langmuir, 25*, 8960–8967. Retrieved from http://dx.doi.org/10.1021/la900591e

Sterpone, F., Marchetti, G., Pierleoni, C., & Marchi, M. (2006). Molecular modeling and simulation of water near model micelles: Diffusion, rotational relaxation and structure at the hydration interface. *The Journal of Physical Chemistry B, 110*, 11504–11510. Retrieved November 22, 2011, from http://www.ncbi.nlm.nih.gov/pubmed/16771426

Tai, C.-H., Bai, H., Taylor, T. J., & Lee, B. (2014). Assessment of template-free modeling in CASP10 and ROLL. *Proteins: Structure, Function, and Bioinformatics, 82*, 57–83.

Taveneau, C., Blondeau, K., & Bressanelli, S. (2015). Definition and expression in *E. coli* of large fragments from the human lipid kinase phosphatidylinositol 4-kinase type III alpha, and purification of a 1100-residue N-terminal module. *Protein Expression and Purification, 114*, 121–127. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S1046592815300061

Thiyagarajan, P., & Tiede, D. M. (1994). Detergent micelle structure and micelle-micelle interactions determined by

Downloaded by [CEA Saclay] at 02:18 22 March 2016

small-angle neutron scattering under solution conditions used for membrane protein crystallization. *The Journal of Physical Chemistry, 98*, 10343–10351. Retrieved April 30, 2013, from http://pubs.acs.org/doi/abs/10.1021/j100091a058

Tieleman, D. P., van der Spoel, D., & Berendsen, H. J. C. (2000). Molecular dynamics simulations of dodecylphosphocholine micelles at three different aggregate sizes: Micellar structure and chain relaxation. *The Journal of Physical Chemistry B, 104*, 6380–6388. Retrieved May 28, 2013, from http://dx.doi.org/10.1021/jp001268f

Timmins, P. A., Hauk, J., Wacker, T., & Welte, W. (1991). The influence of heptane-1,2,3-triol on the size and shape of LDAO micelles Implications for the crystallisation of membrane proteins. *FEBS Letters, 280*, 115–120. Retrieved October 20, 2012, from http://www.sciencedirect.com/science/article/pii/001457939180217Q

Timmins, P. A., Leonhard, M., Weltzien, H. U., Wacker, T., & Welte, W. (1988). A physical characterization of some detergents of potential use for membrane protein crystallization. *FEBS Letters, 238*, 361–368.

Vanquelef, E., Simon, S., Marquant, G., Garcia, E., Klimerak, G., Delepine, J. C., ... Dupradeau, F.-Y. (2011). R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Research, 39*, (Web Server issue), W511–W517. doi: 10.1093/nar/gkr288

Voronoi, G. (1908). Nouvelles Applications Des Paramètres Continus À La Théorie Des Formes Quadratiques. Deuxième Mémoire. Recherches Sur Les Parallélloèdres Primitifs. *Journal für die reine und angewandte Mathematik (Crelles Journal), 1908*, 198–287. Retrieved January 11, 2012, from http://www.reference-global.com/doi/abs/10.1515/crll.1908.134.198

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry, 25* 1157–1174. Retrieved July 20, 2012, (citeulike-article-id:197425)

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology (Clifton, N.J.), 112*, 531–552.

Xu, D. & Zhang, Y. (2012). Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins, 80*, 1715–1735.

Downloaded by [CEA Saclay] at 02:18 22 March 2016

SUPPORTING INFORMATION

Derivation of original RESP atomic partial charges for MD simulations of the LDAO surfactant with AMBER: Applications to a model of micelle and a fragment of the lipid kinase PI4KA

Esra Karakas[1,2,3]‡, Cyntia Taveneau[3]‡, Stéphane Bressanelli[3], Massimo Marchi[2,3], Bruno Robert[2,3] and Stéphane Abel[2,3]*

[1]Maison de la Simulation, USR 3441, CEA – CNRS – INRIA – Univ. Paris-Sud – Univ. de Versailles, 91191,

[2]Commissariat à l'énergie atomique et aux énergies alternatives, Gif-sur-Yvette Cedex, France DSV/iBiTec-S/SB2SM/LBMS & CNRS UMR 9198

[3] Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris Sud, Université Paris-Saclay, 91198 Gif sur Yvette, France

‡ Equal contributions

* Corresponding authors: stephane.abel@cea.fr & stephane.bressanelli@vms.cnrs-gif.fr

Additional Tables and Figures

Group	Atom type		Charges
	Amber99sb	Lipid14	
CH_3 Head Group	CT	cD	-0.0961
CH_3 Head Group	HP	hE	0.0790
N-Oxide Head Group	N4	nA	0.1569
N-Oxide Head Group	O	oC	-0.6041

Table S1: Computed RESP atomic partial charges and atom types used in the Amber99sb and Lipid14 simulations for the LDAO headgroup.

Group	Atom type		Charges
	Amber99sb	Lipid14	
1st CH_2 Hydrophobic Tail	CT	cD	0.0269
1nd CH_2 Hydrophobic Tail	HC	hL	0.0449
2rd CH_2 Hydrophobic Tail	CT	cD	0.0586
2th CH_2 Hydrophobic Tail	HC	hL	-0.0075
3th CH_2 Hydrophobic Tail	CT	cD	-0.0002
3th CH_2 Hydrophobic Tail	HC	hL	0.0074
4th CH_2 Hydrophobic Tail	CT	cD	0.0220
4th CH_2 Hydrophobic Tail	HC	hL	-0.0011
5th CH_3 Hydrophobic Tail	CT	cD	-0.0764
5th CH_3 Hydrophobic Tail	HC	hL	0.0157

Table S2: Computed RESP atomic partial charges and atom types used in the Amber99sb and Lipid14 simulations for the n-pentane alkyl chain. Note that the 1st CH_2 is connected to nitrogen atom of the LDAO headgroup are the same in each fragment.

Group	Atom type		Charges
	Amber99sb	Lipid14	
1st CH_2 Hydrophobic Tail	CT	cD	0.0269
1nd CH_2 Hydrophobic Tail	HC	hL	0.0449
2rd CH_2 Hydrophobic Tail	CT	cD	0.0762
2th CH_2 Hydrophobic Tail	HC	hL	-0.0119
3th CH_2 Hydrophobic Tail	CT	cD	-0.0000
3th CH_2 Hydrophobic Tail	HC	hL	0.0004
4th CH_2 Hydrophobic Tail	CT	cD	-0.0093
4th CH_2 Hydrophobic Tail	HC	hL	0.0008
5th CH_2 Hydrophobic Tail	CT	cD	0.0482
5th CH_2 Hydrophobic Tail	HC	hL	-0.0073
6th CH_2 Hydrophobic Tail	CT	cD	-0.0742
6th CH_2 Hydrophobic Tail	HC	hL	0.0145

Table S3: Same legend as Table S2 for the n-hexane alkyl chain.

Group	Atom type		Charges
	Amber99sb	Lipid14	
1st CH_2 Hydrophobic Tail	CT	cD	0.0269
1nd CH_2 Hydrophobic Tail	HC	hL	0.0449
2rd CH_2 Hydrophobic Tail	CT	cD	0.0692
2th CH_2 Hydrophobic Tail	HC	hL	-0.0081
3th CH_2 Hydrophobic Tail	CT	cD	0.0116
3th CH_2 Hydrophobic Tail	HC	hL	-0.0025
4th CH_2 Hydrophobic Tail	CT	cD	-0.0081
4th CH_2 Hydrophobic Tail	HC	hL	-0.0063
5th CH_2 Hydrophobic Tail	CT	cD	0.0118
5th CH_2 Hydrophobic Tail	HC	hL	-0.0019
6th CH_2 Hydrophobic Tail	CT	cD	0.0458
6th CH_2 Hydrophobic Tail	HC	hL	-0.0044
7th CH_2 Hydrophobic Tail	CT	cD	-0.0856
7th CH_2 Hydrophobic Tail	HC	hL	0.0168

Table S4: Same legend as Table S2 for the n-heptane.

Group	Atom type		Charges
	Amber99sb	Lipid14	
1st CH_2 Hydrophobic Tail	CT	cD	0.0269
1nd CH_2 Hydrophobic Tail	HC	hL	0.0449
2rd CH_2 Hydrophobic Tail	CT	cD	0.0579
2th CH_2 Hydrophobic Tail	HC	hL	-0.0064
3th CH_2 Hydrophobic Tail	CT	cD	0.0209
3th CH_2 Hydrophobic Tail	HC	hL	-0.0010
4th CH_2 Hydrophobic Tail	CT	cD	-0.0002
4th CH_2 Hydrophobic Tail	HC	hL	-0.0086
5th CH_2 Hydrophobic Tail	CT	cD	0.0114
5th CH_2 Hydrophobic Tail	HC	hL	-0.0097
6th CH_2 Hydrophobic Tail	CT	cD	0.0122
6th CH_2 Hydrophobic Tail	HC	hL	0.0004
7th CH_2 Hydrophobic Tail	CT	cD	0.0367
7th CH_2 Hydrophobic Tail	HC	hL	-0.0026
8th CH_2 Hydrophobic Tail	CT	cD	-0.0857
8th CH_2 Hydrophobic Tail	HC	hL	0.0171

Table S4: Same legend as Table S2 for the n-octane alkyl chain.

Group	Atom type		Charges
	Amber99sb	Lipid14	
1st CH_2 Hydrophobic Tail	CT	cD	0.0269
1nd CH_2 Hydrophobic Tail	HC	hL	0.0449
2rd CH_2 Hydrophobic Tail	CT	cD	0.0641
2th CH_2 Hydrophobic Tail	HC	hL	-0.0083
3th CH_2 Hydrophobic Tail	CT	cD	0.0108
3th CH_2 Hydrophobic Tail	HC	hL	-0.0003
4th CH_2 Hydrophobic Tail	CT	cD	0.0080
4th CH_2 Hydrophobic Tail	HC	hL	-0.0079
5th CH_2 Hydrophobic Tail	CT	cD	0.0180
5th CH_2 Hydrophobic Tail	HC	hL	-0.0113
6th CH_2 Hydrophobic Tail	CT	cD	0.0133
6th CH_2 Hydrophobic Tail	HC	hL	-0.0071
7th CH_2 Hydrophobic Tail	CT	cD	0.0050
7th CH_2 Hydrophobic Tail	HC	hL	0.0007
8th CH_2 Hydrophobic Tail	CT	cD	0.0389
8th CH_2 Hydrophobic Tail	HC	hL	-0.0042
9th CH_2 Hydrophobic Tail	CT	cD	-0.0806
9th CH_2 Hydrophobic Tail	HC	hL	0.0160

Table S5: Same legend as Table S2 for the n-nonane alkyl chain.

Group	Atom type		Charges
	Amber99sb	Lipid14	
1st CH_2 Hydrophobic Tail	CT	cD	0.0269
1nd CH_2 Hydrophobic Tail	HC	hL	0.0449
2rd CH_2 Hydrophobic Tail	CT	cD	0.0645
2th CH_2 Hydrophobic Tail	HC	hL	-0.0081
3th CH_2 Hydrophobic Tail	CT	cD	0.0144
3th CH_2 Hydrophobic Tail	HC	hL	-0.0015
4th CH_2 Hydrophobic Tail	CT	cD	-0.0031
4th CH_2 Hydrophobic Tail	HC	hL	-0.0068
5th CH_2 Hydrophobic Tail	CT	cD	0.0283
5th CH_2 Hydrophobic Tail	HC	hL	-0.0111
6th CH_2 Hydrophobic Tail	CT	cD	0.0204
6th CH_2 Hydrophobic Tail	HC	hL	-0.0088
7th CH_2 Hydrophobic Tail	CT	cD	0.0049
7th CH_2 Hydrophobic Tail	HC	hL	-0.0065
8th CH_2 Hydrophobic Tail	CT	cD	0.0064
8th CH_2 Hydrophobic Tail	HC	hL	-0.0005
9th CH_2 Hydrophobic Tail	CT	cD	0.0423
9th CH_2 Hydrophobic Tail	HC	hL	-0.0047
10th CH_2 Hydrophobic Tail	CT	cD	-0.0822
10th CH_2 Hydrophobic Tail	HC	hL	0.0163

Table S6: Same legend as Table S2 for the n-decane alkyl chain.

Group	Atom type		Charges
	Amber99sb	Lipid14	
1st CH_2 Hydrophobic Tail	CT	cD	0.0269
1nd CH_2 Hydrophobic Tail	HC	hL	0.0449
2rd CH_2 Hydrophobic Tail	CT	cD	0.0612
2th CH_2 Hydrophobic Tail	HC	hL	-0.0073
3th CH_2 Hydrophobic Tail	CT	cD	0.0167
3th CH_2 Hydrophobic Tail	HC	hL	-0.0011
4th CH_2 Hydrophobic Tail	CT	cD	0.0026
4th CH_2 Hydrophobic Tail	HC	hL	-0.0083
5th CH_2 Hydrophobic Tail	CT	cD	0.0152
5th CH_2 Hydrophobic Tail	HC	hL	-0.0094
6th CH_2 Hydrophobic Tail	CT	cD	0.0185
6th CH_2 Hydrophobic Tail	HC	hL	-0.0069
7th CH_2 Hydrophobic Tail	CT	cD	0.0138
7th CH_2 Hydrophobic Tail	HC	hL	-0.0099
8th CH_2 Hydrophobic Tail	CT	cD	0.0095
8th CH_2 Hydrophobic Tail	HC	hL	-0.0079
9th CH_2 Hydrophobic Tail	CT	cD	0.0079
9th CH_2 Hydrophobic Tail	HC	hL	-0.0001
10th CH_2 Hydrophobic Tail	CT	cD	0.0393
10th CH_2 Hydrophobic Tail	HC	hL	-0.0038
11th CH_2 Hydrophobic Tail	CT	cD	-0.0833
11th CH_2 Hydrophobic Tail	HC	hL	0.0166

Table S7: Same legend as Table S2 for the n-undecane alkyl chain.

Group	Atom type		Charges
	Amber99sb	Lipid14	
1st CH_2 Hydrophobic Tail	CT	cD	0.0269
1nd CH_2 Hydrophobic Tail	HC	hL	0.0449
2rd CH_2 Hydrophobic Tail	CT	cD	0.0612
2th CH_2 Hydrophobic Tail	HC	hL	-0.0073
3th CH_2 Hydrophobic Tail	CT	cD	0.0167
3th CH_2 Hydrophobic Tail	HC	hL	-0.0011
4th CH_2 Hydrophobic Tail	CT	cD	-0.0011
4th CH_2 Hydrophobic Tail	HC	hL	-0.0077
5th CH_2 Hydrophobic Tail	CT	cD	0.0220
5th CH_2 Hydrophobic Tail	HC	hL	-0.0114
6th CH_2 Hydrophobic Tail	CT	cD	0.0185
6th CH_2 Hydrophobic Tail	HC	hL	-0.0069
7th CH_2 Hydrophobic Tail	CT	cD	0.0222
7th CH_2 Hydrophobic Tail	HC	hL	-0.0081
8th CH_2 Hydrophobic Tail	CT	cD	0.0138
8th CH_2 Hydrophobic Tail	HC	hL	-0.0099
9th CH_2 Hydrophobic Tail	CT	cD	0.0095
9th CH_2 Hydrophobic Tail	HC	hL	-0.0079
10th CH_2 Hydrophobic Tail	CT	cD	0.0079
10th CH_2 Hydrophobic Tail	HC	hL	0.0001
11th CH_2 Hydrophobic Tail	CT	cD	0.0393
11th CH_2 Hydrophobic Tail	HC	hL	-0.0037
12th CH_3 Hydrophobic Tail	CT	cD	-0.0834
12th CH_3 Hydrophobic Tail	HC	hL	0.0166

Table S8: Same legend as Table S2 for the n-dodecane alkyl chain.

Group	Atom type		Charges
	Amber99sb	Lipid14	
1st CH_2 Hydrophobic Tail	CT	cD	0.0269
1nd CH_2 Hydrophobic Tail	HC	hL	0.0449
2rd CH_2 Hydrophobic Tail	CT	cD	0.0639
2th CH_2 Hydrophobic Tail	HC	hL	-0.0078
3th CH_2 Hydrophobic Tail	CT	cD	0.0118
3th CH_2 Hydrophobic Tail	HC	hL	-0.0008
4th CH_2 Hydrophobic Tail	CT	cD	0.0044
4th CH_2 Hydrophobic Tail	HC	hL	-0.0080
5th CH_2 Hydrophobic Tail	CT	cD	0.0172
5th CH_2 Hydrophobic Tail	HC	hL	-0.0102
6th CH_2 Hydrophobic Tail	CT	cD	0.0244
6th CH_2 Hydrophobic Tail	HC	hL	-0.0090
7th CH_2 Hydrophobic Tail	CT	cD	0.0116
7th CH_2 Hydrophobic Tail	HC	hL	-0.0068
8th CH_2 Hydrophobic Tail	CT	cD	0.0230
8th CH_2 Hydrophobic Tail	HC	hL	-0.0094
9th CH_2 Hydrophobic Tail	CT	cD	0.0167
9th CH_2 Hydrophobic Tail	HC	hL	-0.0098
10th CH_2 Hydrophobic Tail	CT	cD	0.0084
10th CH_2 Hydrophobic Tail	HC	hL	-0.0074
11th CH_2 Hydrophobic Tail	CT	cD	0.0070
11th CH_2 Hydrophobic Tail	HC	hL	-0.0000
12th CH_2 Hydrophobic Tail	CT	cD	0.0402
12th CH_2 Hydrophobic Tail	HC	hL	-0.0041
13th CH_3 Hydrophobic Tail	CT	cD	-0.0827
13th CH_3 Hydrophobic Tail	HC	hL	0.0164

Table S9: Same legend as Table S2 for the n-tridecane alkyl chain.

Group	Atom type		Charges
	Amber99sb	Lipid14	
1st CH_2 Hydrophobic Tail	CT	cD	0.0269
1nd CH_2 Hydrophobic Tail	HC	hL	0.0449
2rd CH_2 Hydrophobic Tail	CT	cD	0.0619
2th CH_2 Hydrophobic Tail	HC	hL	-0.0074
3th CH_2 Hydrophobic Tail	CT	cD	0.0160
3th CH_2 Hydrophobic Tail	HC	hL	-0.0012
4th CH_2 Hydrophobic Tail	CT	cD	-0.0023
4th CH_2 Hydrophobic Tail	HC	hL	-0.0074
5th CH_2 Hydrophobic Tail	CT	cD	0.0240
5th CH_2 Hydrophobic Tail	HC	hL	-0.0111
6th CH_2 Hydrophobic Tail	CT	cD	0.0204
6th CH_2 Hydrophobic Tail	HC	hL	-0.0077
7th CH_2 Hydrophobic Tail	CT	cD	0.0162
7th CH_2 Hydrophobic Tail	HC	hL	-0.0085
8th CH_2 Hydrophobic Tail	CT	cD	0.0116
8th CH_2 Hydrophobic Tail	HC	hL	-0.0080
9th CH_2 Hydrophobic Tail	CT	cD	0.0265
9th CH_2 Hydrophobic Tail	HC	hL	-0.0094
10th CH_2 Hydrophobic Tail	CT	cD	0.0160
10th CH_2 Hydrophobic Tail	HC	hL	-0.0093
11th CH_2 Hydrophobic Tail	CT	cD	0.0071
11th CH_2 Hydrophobic Tail	HC	hL	-0.0074
12th CH_2 Hydrophobic Tail	CT	cD	0.0076
12th CH_2 Hydrophobic Tail	HC	hL	-0.0002
13th CH_2 Hydrophobic Tail	CT	cD	0.0404
13th CH_2 Hydrophobic Tail	CT	hL	-0.0040
14th CH_3 Hydrophobic Tail	CT	cD	-0.0836
14th CH_3 Hydrophobic Tail	HC	hL	0.0166

Table S10: Same legend as Table S2 for the n-tetradecane alkyl chain.

I.	GST-DI	his6-DI
Lysis buffer	30mM Tris pH7.4, 500mM NaCl, 20% glycerol, 43mM LDAO, 1mM DTT	30mM Tris pH7.4, 500mM NaCl, 20% glycerol, 43mM LDAO, 20mM imidazole, 1mM DTT
Wash buffer	30mM Tris pH7.4, 300mM NaCl, 10% glycerol, 2mM LDAO, 1mM DTT	30mM Tris pH7.4, 1M NaCl, 10% glycerol, 2mM LDAO, 1mM DTT and 50mM imidazole
Elution buffer	30mM Tris pH7.4, 300mM NaCl, 10% glycerol, 2mM LDAO, 25 mM gluthation, 1mM DTT	30mM Tris pH7.4, 300mM NaCl, 10% glycerol, 2mM LDAO, 500mM imidazole, 1mM DTT

Table S11: List of the biochemical used in experiment

Force Field	MD	*a*	*b*	*c*
Amber99sb	PA	29.0	19.9	24.9
	SA1	29.1	19.9	24.7
	SA2	28.0	20.2	24.2
Lipid14	PA	29.9	19.3	23.6
	SA1	30.1	19.3	23.6
	SA2	29.5	19.3	24.1

Table S12: Average semi-axis obtained from last 50 ns of the pre-assembled (PA) and self-assembled (SA) simulations (see main text). The maximum error is never greater than 1.7 Å.

List of Figures

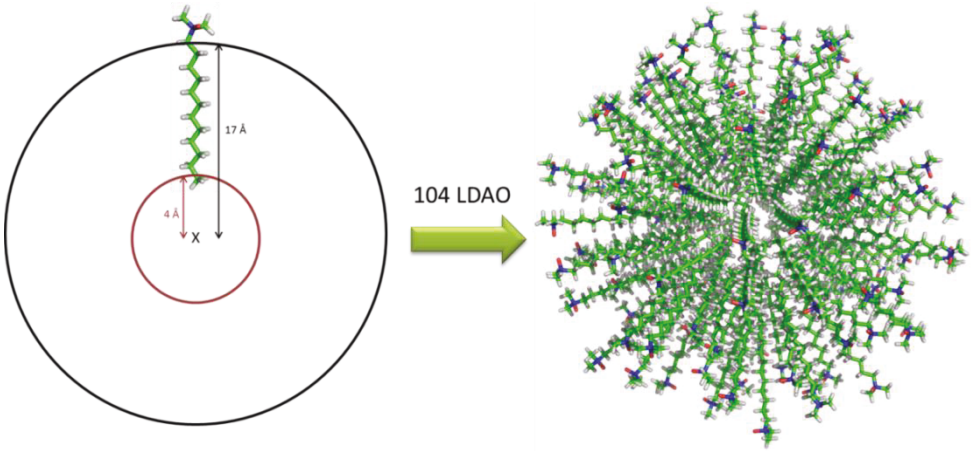


Figure S1: Geometric model used to construct the preformed LDAO with Packmol [38] see main text for details.

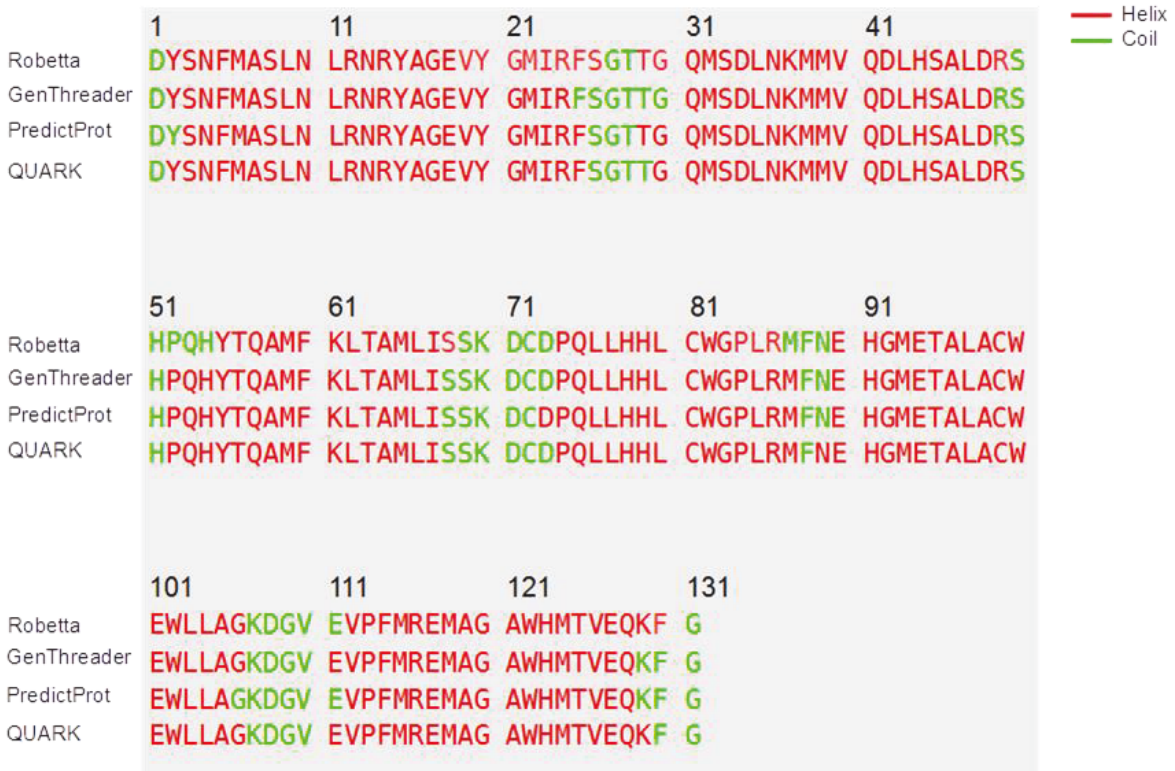


Figure S2: Sequence alignment of DI protein with the secondary structure prediction of different methods (see main text). In red and green the helix and random coil predictions.

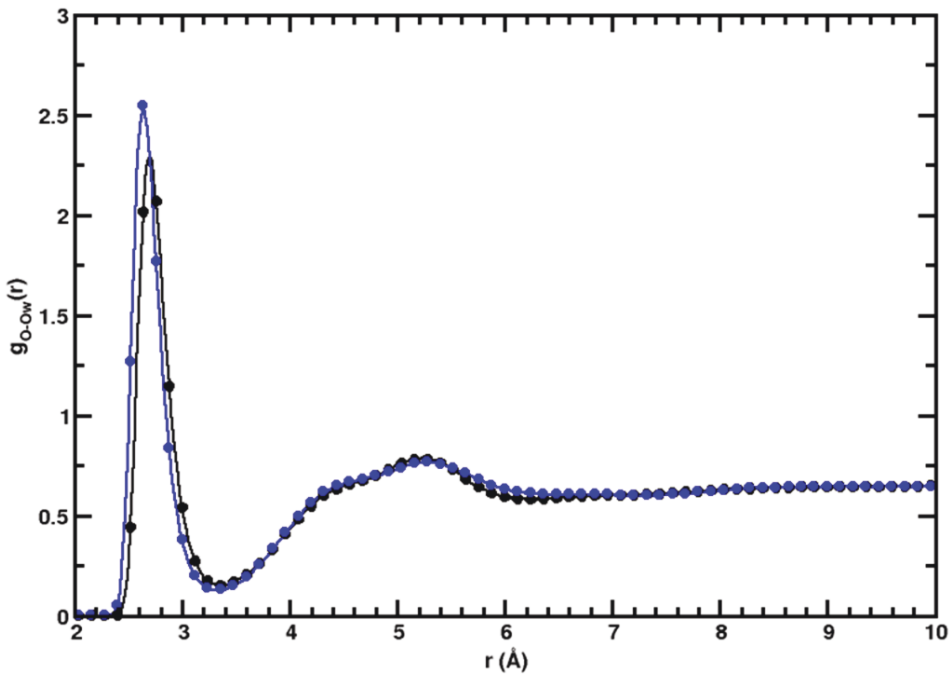


Figure S3: Average radial density function between the LDAO and water oxygen atoms. Results obtain for PA and SA Amber99sb and lipid14 are in black and blue colors.PA and SA simulations are highlighted with continue line, sphere and square, respectively.

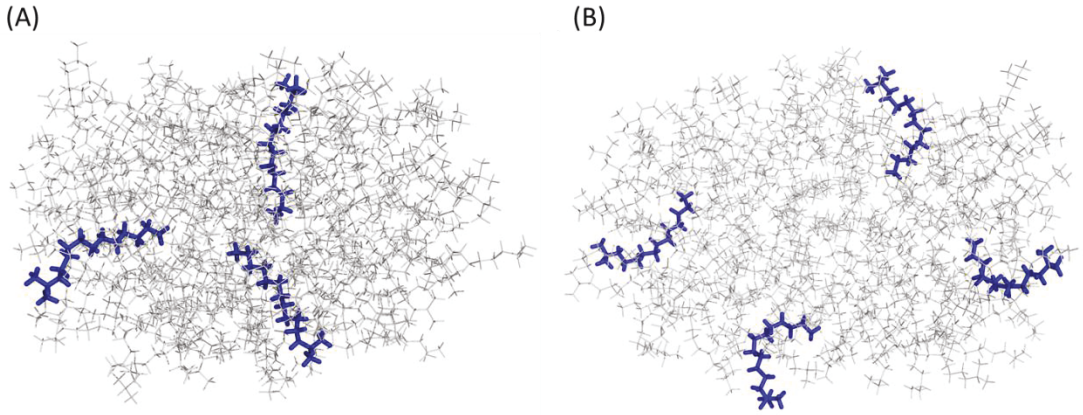


Figure S4: Representative LDAO conformations in the Amber99sb (A) and Lipid14 micelles PA micelles (B).

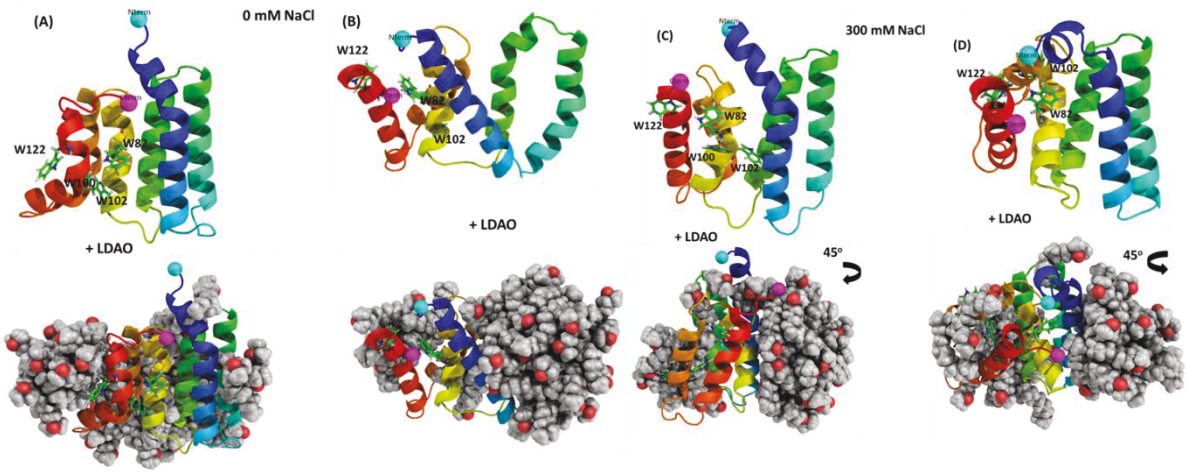


Figure S5: Final snapshot of the second Amber99sb(A) and Lipid14(B) simulations at two NaCl concentration. The 4 tryptophans are showed in sticks. In grey, the LDAO molecules and the DI protein is colored blue to red from N- to C-terminus. These figures were produced with PyMOL [1].

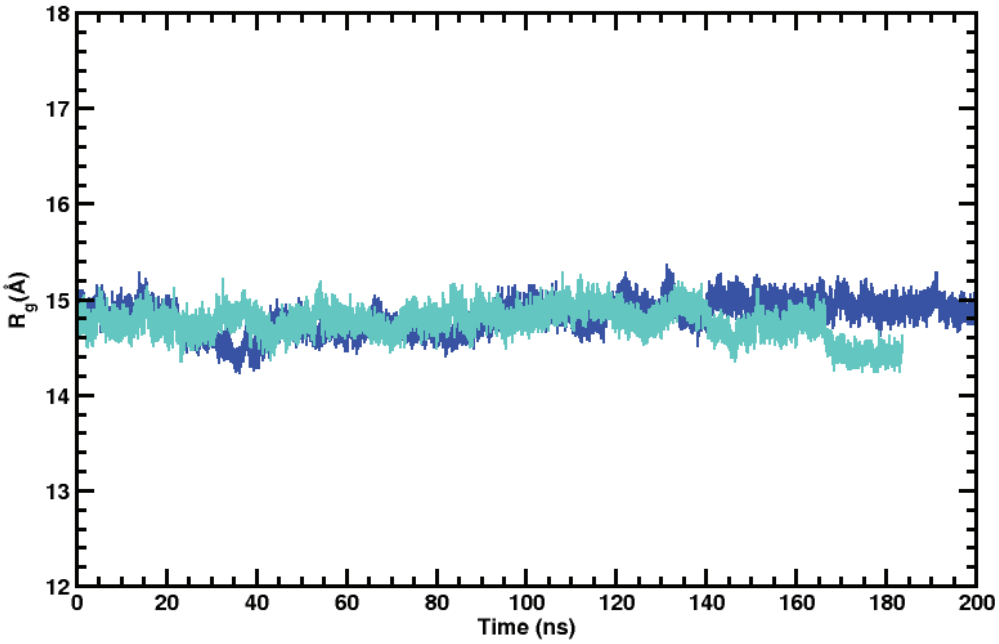


Figure S6: Time evolution of the radius of gyration of the protein in the systems without LDAO at 0 mM (blue) and 300 mM (cyan) of NaCl.

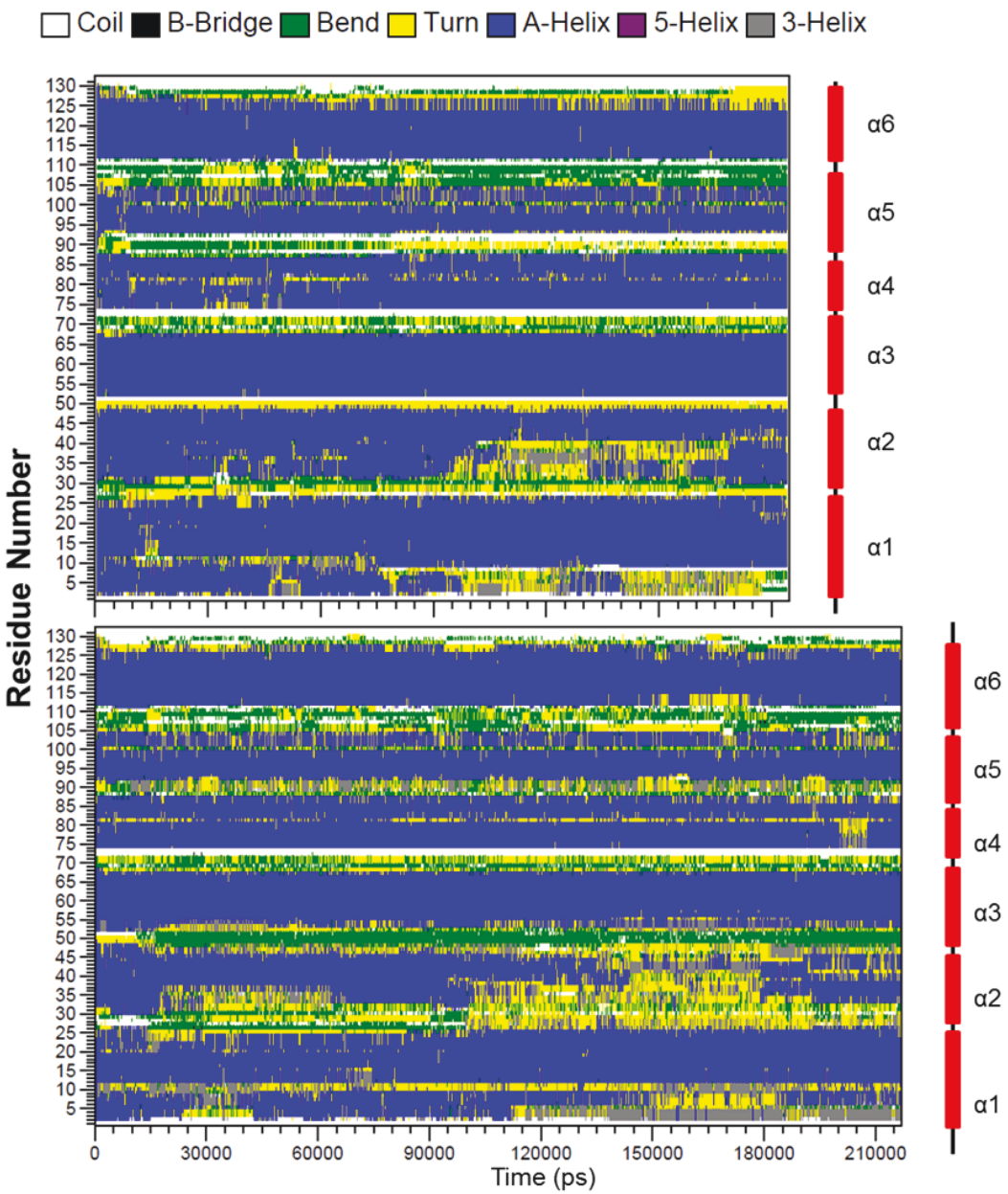


Figure S7: Time evolution of protein secondary structures in pure water at 0 mM (upper panel) and 300 mM (lower panel) of NaCl.

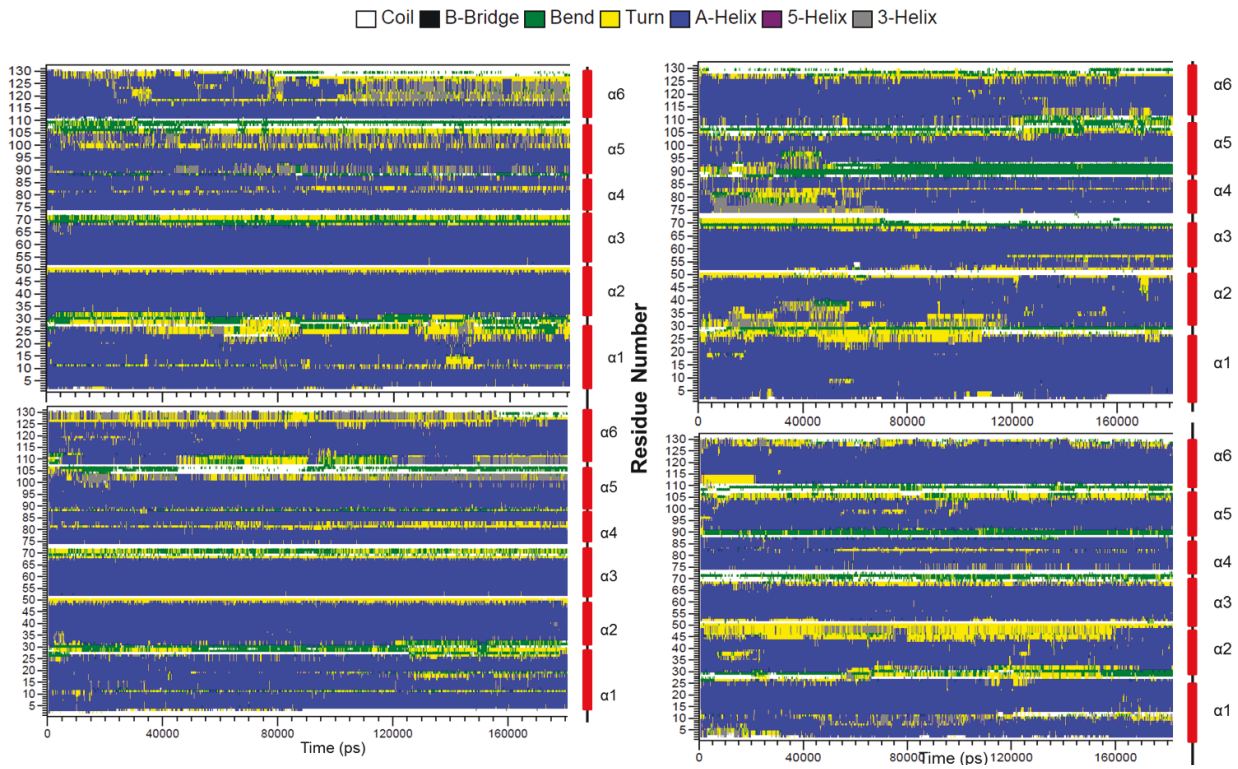


Figure S8: Time evolution of protein secondary structures in the complex simulated with Amber99sb (left panel) and Lipid14 (right panel) force fields at 0 mM of NaCl.

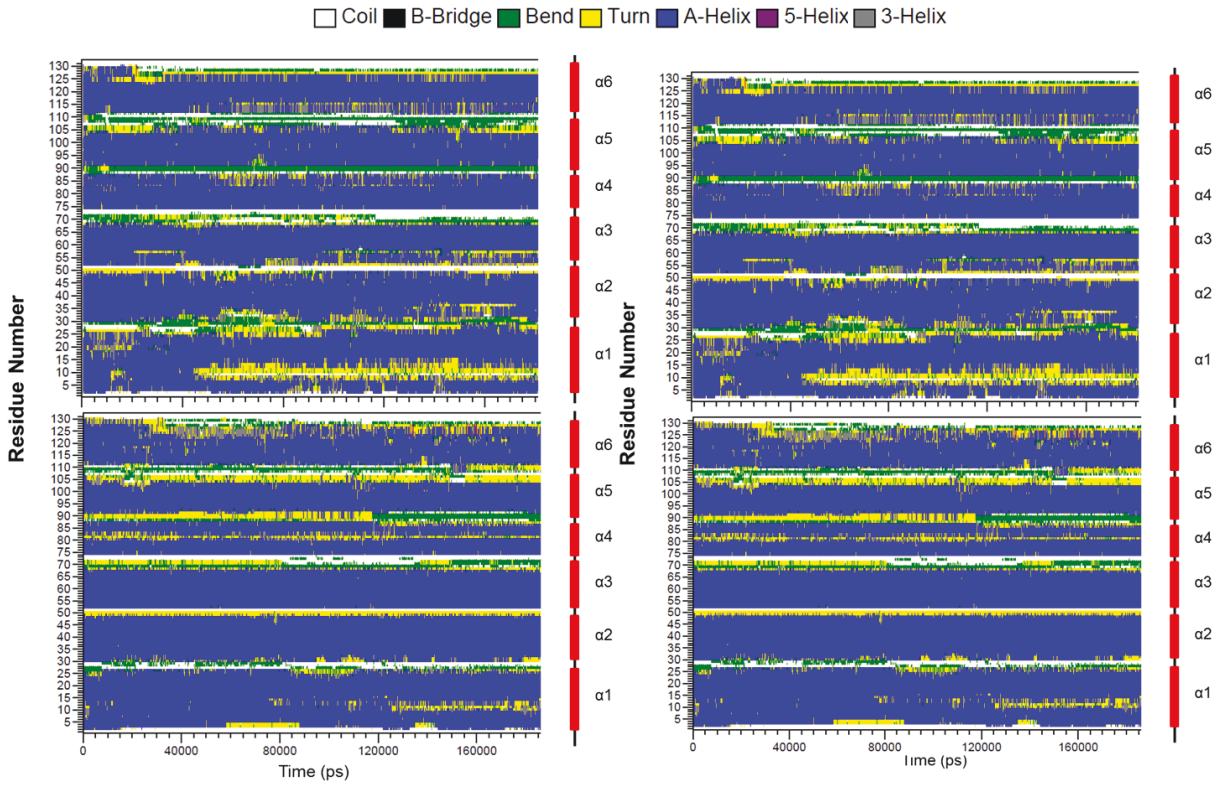


Figure S9: Time evolution of protein secondary structures in the complex simulated with Amber99sb (left panel) and Lipid14 (right panel) force fields at 300 mM of NaCl.

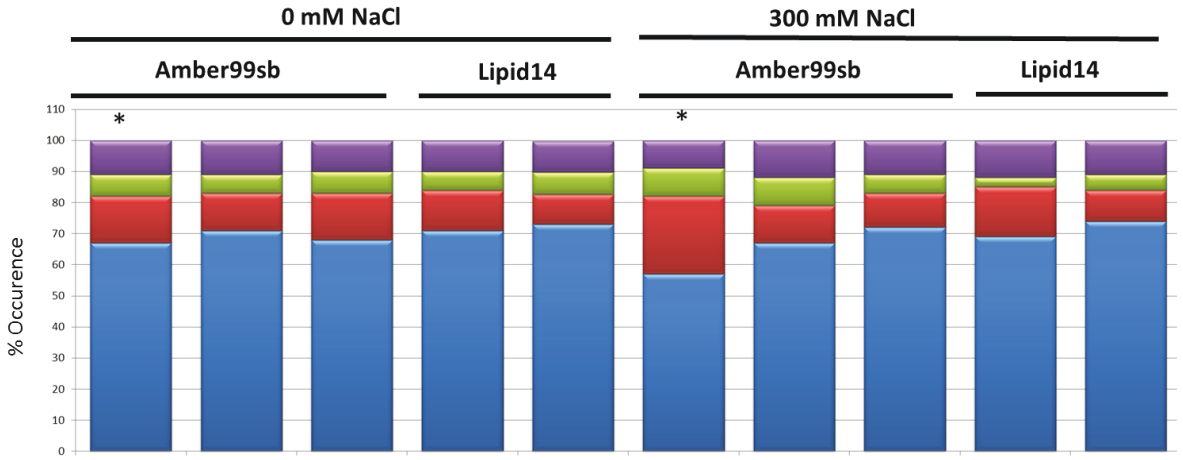


Figure S10: Average percentages of helix (blue), turn (red), bend (green) and coil (violet) for each protein simulated with and without LDAO at different salt concentration. These values were computed from the last 40 ns of each run. The results obtained from the protein without LDAO are shown with an asterisk.

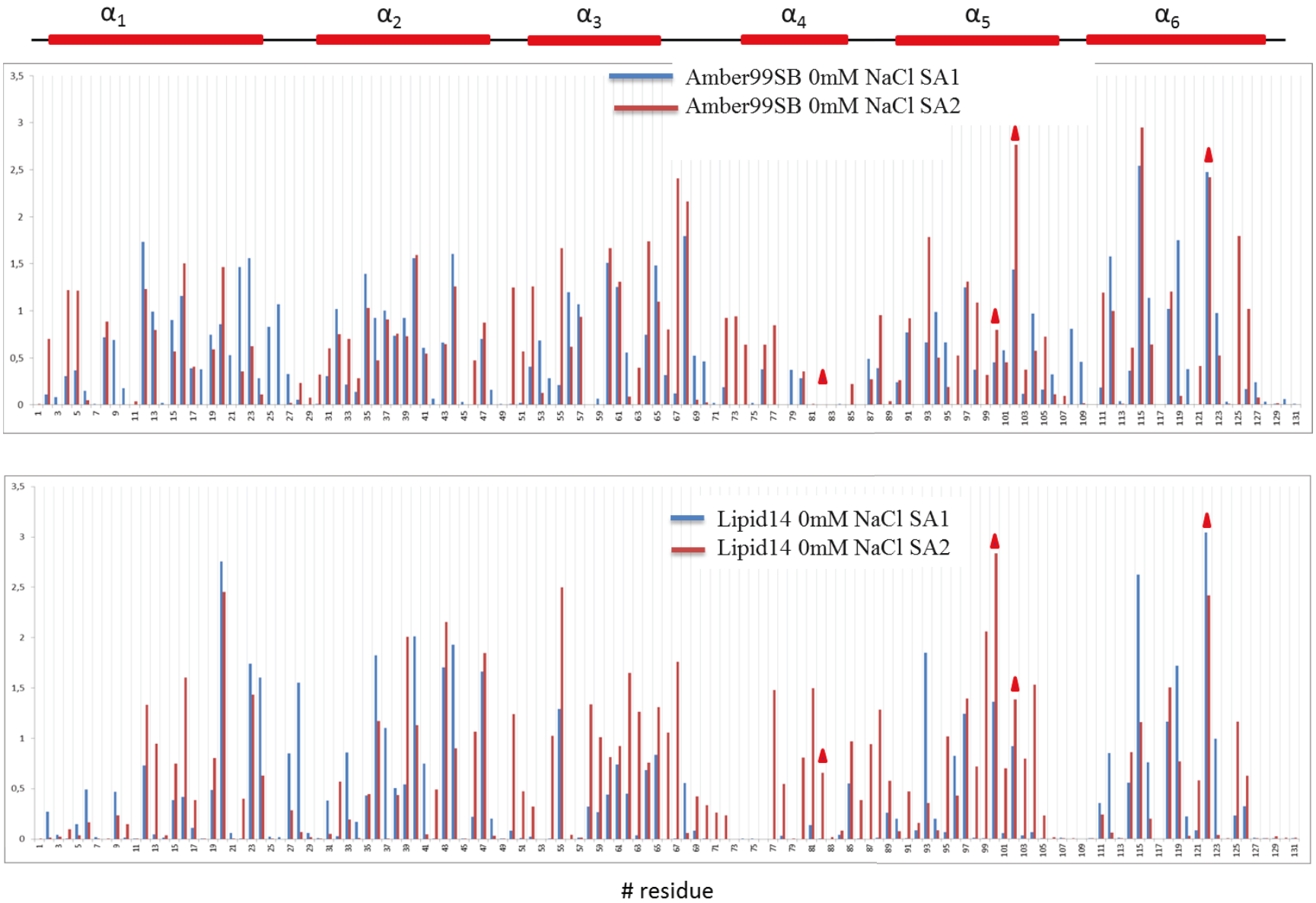


Figure S11: Average number of unique contact between the LDAO alkyl chain and the protein residues 0mM of NaCl. The values were computed from the last 30 ns of each run. The four tryptophans in the DI protein are highlighted with red triangles (see main text for details).

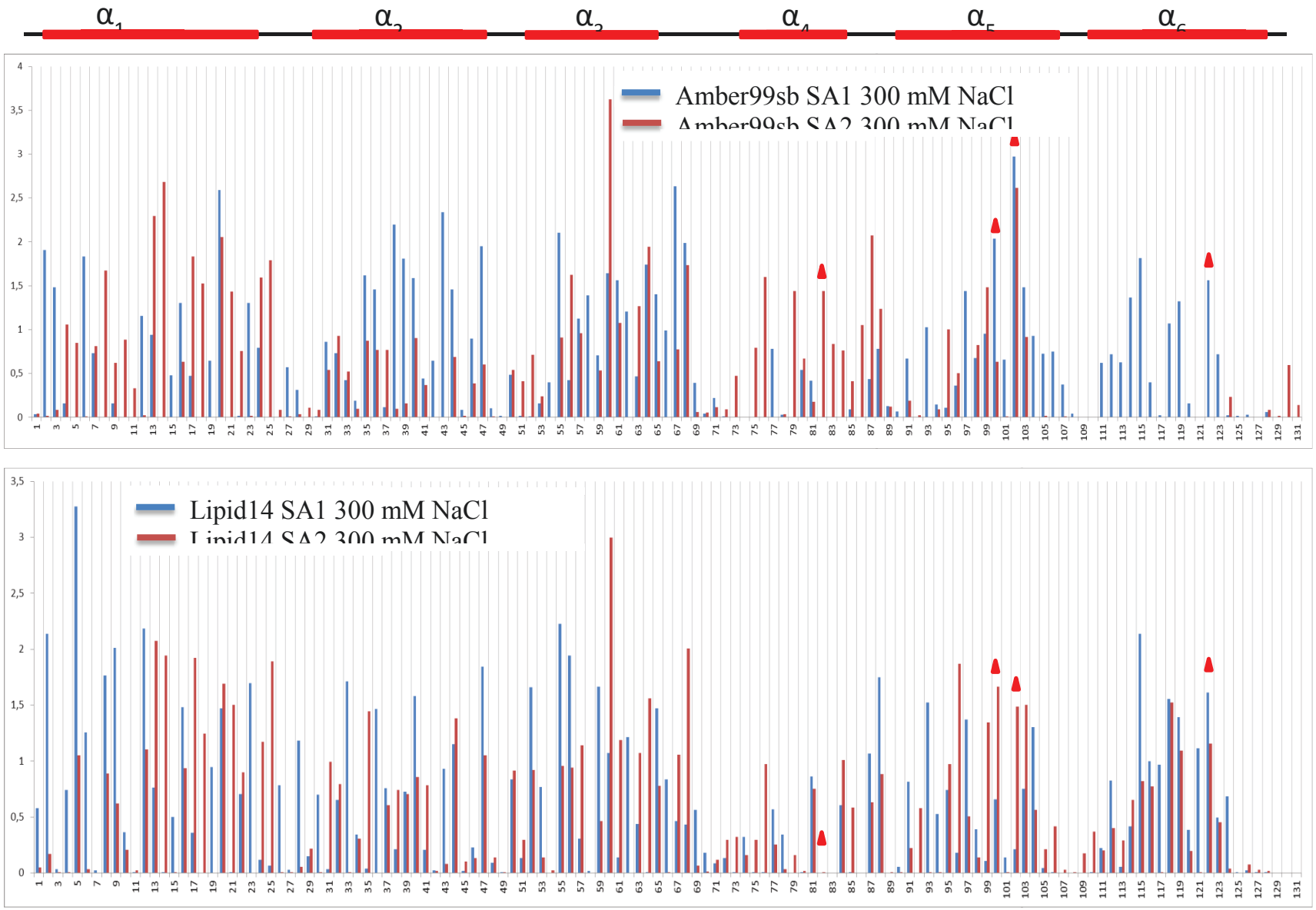


Figure S12 : Average number of unique contact between the LDAO alkyl chain and the protein residues at 300mM of NaCl. The values were computed from the last 30 ns of each run. The four tryptophans in the proteins are highlighted with red triangles (see main text for details).

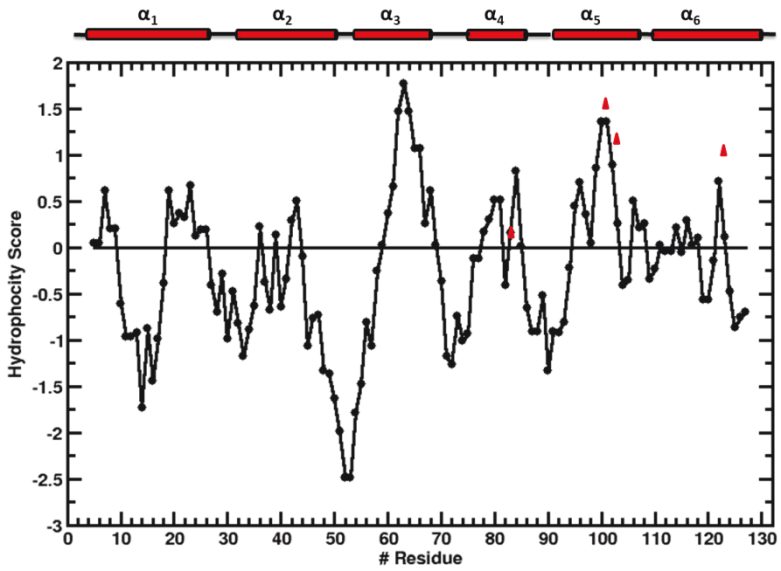


Figure S13: Hydrophobicity plot of the DI protein computed with the Kyte-Doolittle approach obtained from the ExPASy server [2]

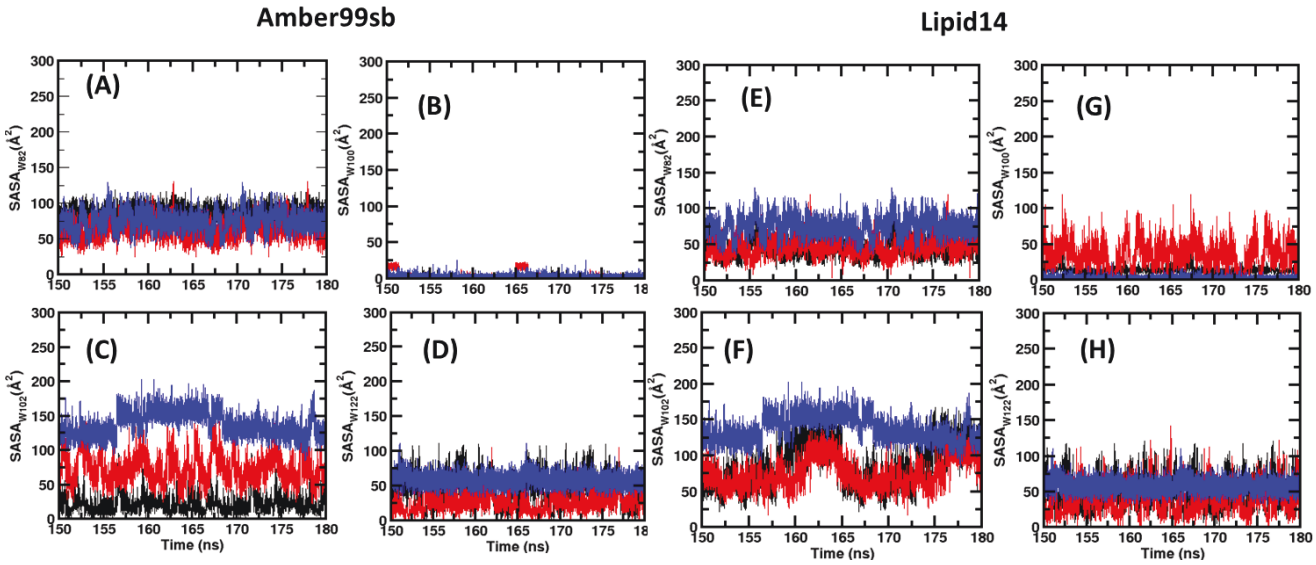


Figure S14: Time evolution of the SASA for each tryptophan at 0 mM for the protein in pure water (blue) and of the first (red) and second (black) replicas. Note that only the last 30 ns of each run used to examine the tryptophan accessibilities are shown (see main text for details).

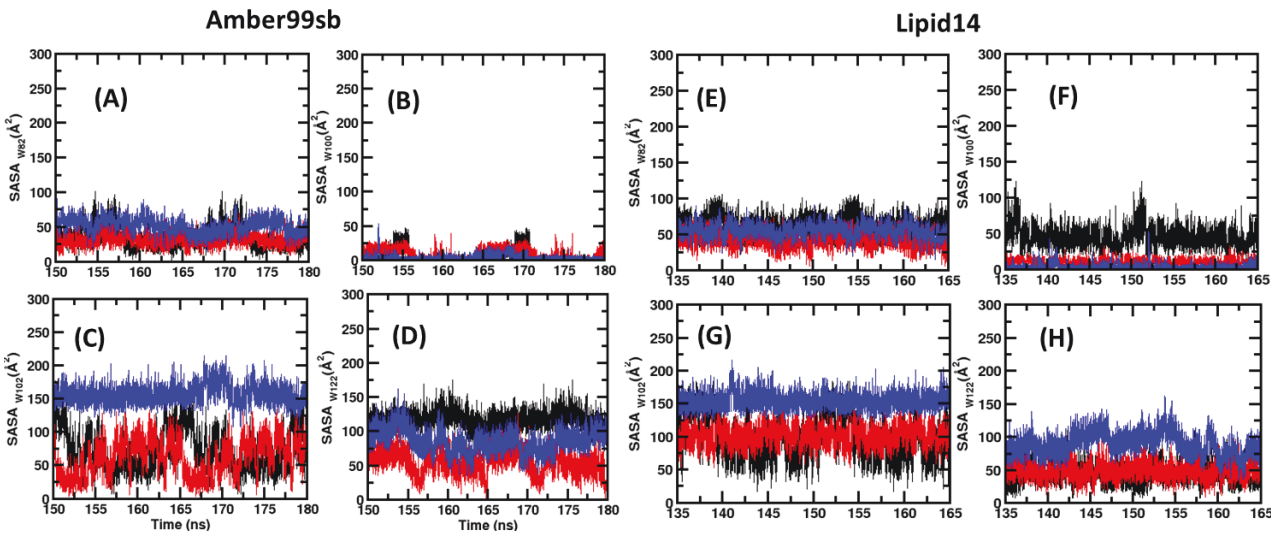


Figure S15: Time evolution of the SASA for each tryptophan at 300 mM for the protein in pure water (blue) and of the first (red) and second (black) replicas. Note that only the last 30 ns of each run used to examine the tryptophan accessibilities are shown (see main text for details).

References

[1] The PyMOL Molecular Graphics System, Version 1.7.4, Schrödinger, LLC, n.d.

[2] M.R. Wilkins, E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel, et al., Protein identification and analysis tools in the ExPASy server, Methods Mol. Biol. 112 (1999) 531–552.

3. Les principales conclusions de l'article

Les résultats des propriétés structurales des micelles pures de LDAO ont montré que la forme, la taille, l'hydratation et la conformation des chaînes alkyles dépendaient légèrement du champ de force mais pas des conditions initiales. De plus, ces résultats sont en bon accord la littérature expérimentale et d'autres simulations moléculaires. Les micelles obtenues ont une forme ellipsoïdale (ratio des demi-axes majeur sur mineur de 1,40) avec un rayon de giration d'environ 19 Å qui dépend peu du champ de force et de la construction des micelles. La structure interne des micelles est aussi peu modifiée avec les différentes conditions de simulations. Nos calculs d'hydratation ont montré qu'en moyenne neuf molécules d'eau interagissent avec une molécule de LDAO et ~3 avec la tête polaire.

Concernant la conformation de la chaîne alkyle du LDAO, nous avons montré que le champ de force Lipid14 reproduisait mieux les résultats obtenues par FTIR (Holler *et al*. 1989) alors que le potentiel Amber99sb à tendance à surestimer la conformation *trans* des dièdres des chaînes carbonées CH_2-CH_2-CH_2-CH_2.

L'analyse des propriétés structurales du LDAO en présence de la protéine DI ont montré que les résultats sont en bon accord avec les valeurs expérimentales (dichroïsme circulaire, SEC-MALLS (chromatographie d'exclusion stérique couplée à la diffusion de la lumière multi-angulaire) et la spectroscopie de fluorescence des tryptophanes). En particulier, le modèle prédit par QUARK présente six régions en hélice α qui se retrouvent aussi dans la prédiction faite par d'autres logiciels de modélisation (i.e Robetta, GenThreader et PredictProtein). Nous avons montré que les simulations dépendent faiblement des conditions initiales et de la concentration de sel. L'étude approfondit de l'accessibilité des tryptophanes avec l'eau ont montré que trois des quatre tryptophanes ne sont pas accessibles au solvant et sont considérés comme enfouis, ce qui est bon accord avec les expériences de spectroscopie de fluorescence.

Finalement, ces deux exemples de simulations avec des systèmes de complexité croissante nous ont permis de valider nos charges RESP en combinaison avec les champs de force Amber99sb et Lipid14.

Chapitre 4 : Simulations du complexe LH2 dans les environnements biomimétiques

Ce chapitre sera consacré à la mise au point et à la discussion des simulations du complexe LH2 en présence de LDAO, bOG et POPC. Dans une première partie, nous détaillerons le choix des paramètres et le protocole des simulations. La deuxième partie sera consacrée à l'analyse structurale des différents systèmes.

I. Construction des systèmes

1. Champs de force utilisés

Comme on l'a vu dans le chapitre 1, le complexe LH2 est composé de dipeptides αβ et de pigments photosynthétiques (RG et BChl-a). Les structures cristallographiques disponibles à haute résolution (PDB : 1KZU (Prince *et al.* 1997), 1NKZ (Papiz *et al.* 2003), 2FKW (Cherezov *et al.* 2006)) ont montré que le premier résidu des peptides α est une méthionine modifié, une N-carboxy-L-méthionine (CXM). Pour modéliser cette acide aminé et en l'absence de paramètres (i.e. charges partielles atomiques RESP), nous avons dérivé des charges RESP en utilisant l'approche décrite par Cornell et Bayly (Cornell *et al.* 1995)(Bayly *et al.* 1993) (cf. annexe 3).

Dans le cas du BChl-a, nous avons utilisé les paramètres développés par Ceccarelli *et al.* (Ceccarelli *et al.* 2003), basés sur Amber94 (Cornell *et al.* 1995) précédemment utilisés pour modéliser le centre réactionnel d'une bactérie pourpre (*Rb. sphaeroides*) obtenue par DFT (*Density Functional Therory*).

Pour le RG, nous avons aussi calculé les charges RESP avec la méthodologie Glycam (cf. annexe 3). Les paramètres liants et non-liants sont issus de Tessier *et al.* (Tessier *et al.* 2008) pour les glycolipides insaturés.

Les champs de force Amber et Glycam utilisent différentes valeurs pour le facteur multiplicatif des interactions 1-4 (électrostatique (fudgeQQ) et Lennard-Jones (fudgeLJ)) qui sont 0,83 et 0,5, et 1,0 et 1,0, respectivement. Pour combiner ces deux champs de force avec GROMACS, nous avons utilisé la méthode « double paire de liste et moitié d'epsilon » (Chakrabarti *et al*. 2010) avec les paramètres fudgeQQ=0,1666 et fudgeLJ=1,0. La liste des interactions 1-4 a été répliquée 5 fois pour Amber et 6 fois pour Glycam.

2. Simulations des différents systèmes

Pour simuler les systèmes du complexe LH2 dans les différents environnements micellaires de détergents et dans la membrane de POPC, nous avons adopté plusieurs stratégies que nous allons décrire dans les sous parties suivantes.

Pour toutes ces simulations, nous avons combiné différents champs de force, à savoir, Amber99sb-ILDN (Lindorff-Larsen *et al.* 2010) pour la partie protéique, Amber99sb et Lipid14 pour le LDAO, Amber94 pour les bactériochlorophylles-a, Glycam pour la rhodopine glucoside et le β octyle glucoside, et Lipid14 pour le POPC.

Pour ce travail, nous avons effectués 6 simulations dans les conditions ambiantes de température (300 K et 303 K pour la membrane) et de pression (1,015 bar). Les caractéristiques des systèmes simulées sont données dans le Tableau 1, ci-dessous :

Tableau 1 : Paramètres des simulations.

	Complexe LH2-LDAO				**Complexe LH2-bOG**	**Complexe LH2-POPC**
	Amber99sb			**Lipid14**		
	Système A	**Système B**	**Système C**	**Système C'**	**Système D**	**Système E**
$N_{détergent/lipides}$	260	290	308	308	243	409
N_{eau}	110 000	110 000	95 000	95 000	95 000	50 000
N_{atomes}	360 301	361 711	317 557	317 557	314 745	222 887
Taille de la boîte (Å)	140x140x140					140x130 x120
Temps (ns)	300	200	200	300	200	200

3. Le complexe LH2 dans une micelle de LDAO

Une première étude de solubilisation, réalisée par Gast et ses collaborateurs (Gast *et al.* 1994), a montré que le centre réactionnel de *Rhodopseudomonas viridis*, qui a une forme cylindrique avec une section transversale elliptique de 70 et 30 Å de diamètre (Deisenhofer *et al.* 1984), pouvait s'associer à 260 ± 30 molécules de LDAO. Plus tard, Jirsakova et ses collaborateurs (Jirsakova *et al.* 1996) ont montré par des expériences de centrifugation analytique que le complexe LH2 de *Rubrivivax gelatinosus* (bactérie pourpre), dont les

dimensions sont similaires à *Rps. acidophila* (Ranck *et al.* 2001), s'associait à un nombre de molécules de LDAO compris entre 252 et 315.

En prenant les données de Gast *et al.*, nous avons effectués des simulations auto-assemblées du complexe LH2 dans 260 LDAO (Système A). Nous avons minimisé le système puis nous avons équilibré dans l'ensemble NVT à 300 K avec le thermostat Berendsen et une constante de couplage de τ_T = 0,1 ps pendant 1 ns. L'équilibration du système a continué dans l'ensemble NPT en isotropique pendant 5 ns avec le thermostat *v-rescale* (*velocity rescale*) et le barostat de Parrinello-Rahman (τ_P = 3 ps), avec une compressibilité de $4{,}5.10^{-5}$ bar^{-1}. Le système équilibré a été simulé pour la production dans l'ensemble NPT en changeant le thermostat par celui de Nosé-Hoover (τ_T = 0,3 ps) et le barostat par celui de Parrinello-Rahman (τ_P = 1 ps). La stratégie d'auto-assemblage adoptée ici a consisté à bloquer tout le complexe LH2 pendant l'équilibration et la production jusqu'à ce que toutes les molécules de LDAO s'agrègent autour du complexe LH2, puis nous avons relâché le complexe pour produire environ 300 ns. Nous avons simulé ~30 ns par jour sur 192 CPUs du supercalculateur CURIE. En visualisant la trajectoire, nous avons constaté que le nombre de LDAO choisis n'était pas suffisant pour recouvrir complètement la partie transmembranaire du complexe (Figure 1). En parallèle, nous avons aussi simulé le même système avec une autre méthode qui consistait à ne pas contraindre le complexe LH2. La comparaison de ces deux méthodes a montré qu'en ne figeant pas le complexe, nous devions simuler quatre fois plus longtemps pour que toutes les molécules de LDAO s'agrègent autour du complexe LH2 (~ 640 ns). De plus, quelques analyses sur la structure finale obtenue ont montré que globalement, les deux systèmes étaient similaires.

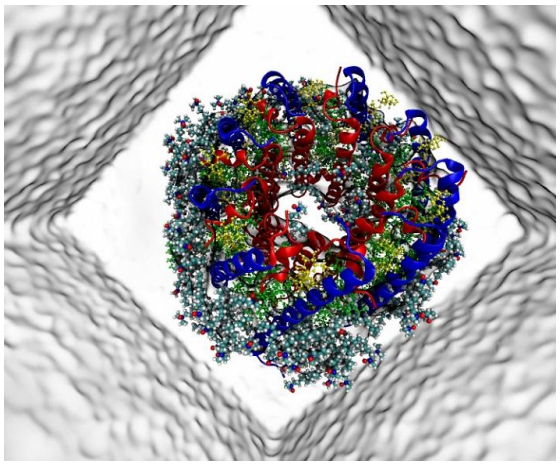


Figure 1 : Représentation de la dernière configuration de la simulation du complexe LH2-260 LDAO dans une boîte d'eau. Les molécules de LDAO sont représentées en sphère bleue, rouges blanches et cyan, les molécules d'eau en surface grise, et les peptides en hélice bleue et rouge, les BCL en sphère verte et les RG en sphère jaune.

Par conséquent, nous avons réalisé une deuxième simulation en adoptant notre première stratégie avec les mêmes paramètres de simulation mais avec 290 molécules de LDAO (Gast *et al.* 1994) (System B).

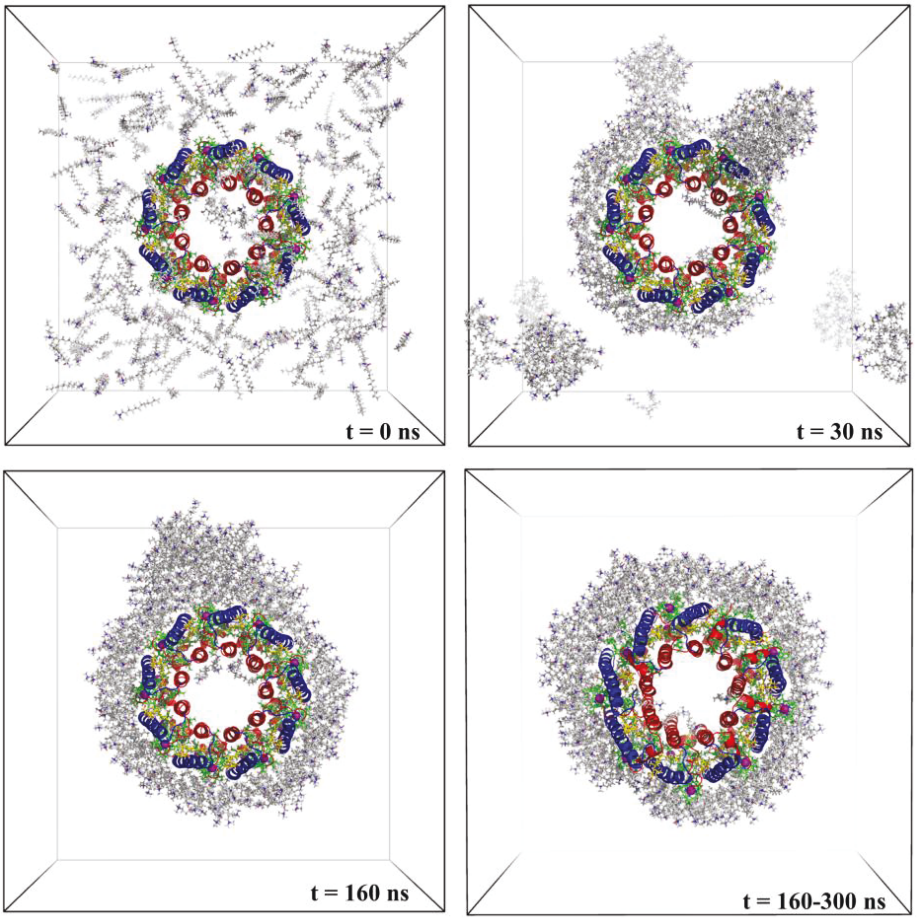


Figure 2 : Représentation des différentes configurations du système complexe LH2-LDAO. De 0 à 160 ns, agrégation des molécules de LDAO autour du complexe LH2 figé puis défigé de 160 à 300 ns. Les molécules d'eau ne sont pas représentées pour plus de clarté (Système B).

Comme nous pouvons le voir sur la Figure 2, quelques molécules de LDAO (~4) s'associent aux parties hydrophobes des peptides α, à l'intérieur de l'anneau central. Ceci est en bon accord avec les expériences de RX de Prince et ses collaborateurs (Prince *et al.* 2003) qui ont montré que l'anneau central pouvait aussi contenir des molécule de détergents. Nous pouvons penser que le temps de simulation (et le nombre monomères de LDAO) n'était pas suffisant pour obtenir une pénétration suffisante des molécules de LDAO dans l'anneau.

Pour accélérer l'étude de notre système, nous avons utilisé préférentiellement la structure PDB 2FKW de Gast *et al.* qui contient 18 molécules de LDAO à l'intérieur de l'anneau central et simulé ce système dans une boîte de simulation plus petite (Système C).

Dans notre article sur les micelles de LDAO, nous avons montré que le champ de force Amber99sb surestimait la conformation *trans* des dièdres des chaînes carbonées du détergent comparé à Lipid14. Nous avons donc réalisé une quatrième simulation de 300 ns en partant de la configuration avec des molécules de LDAO placées à l'intérieur de l'anneau du complexe LH2 et avec le champ de force Lipid14 (Système C').

4. Le complexe LH2 dans une micelle de β octyle glucoside

Pour étudier l'effet du détergent sur la structure du complexe LH2, nous avons utilisé le β octyle glucoside (bOG) (dont nous donnons les paramètres validés dans l'annexe 1). Gast et ses collaborateurs ont montré que 360 ± 100 molécules de bOG pouvaient s'agrégeaient autour du centre réactionnel de *Rps. viridis* (Gast *et al.* 1994). L'erreur sur ce nombre d'agrégation étant assez importante, nous avons utilisé une approche alternative consistant à calculer le nombre de monomères de bOG à partir des données géométrique de la couronne de détergent autour du complexe (assumé comme un tore, Figure 3) suggéré par le Dr. Miroslav Papiz (communications personnelles).

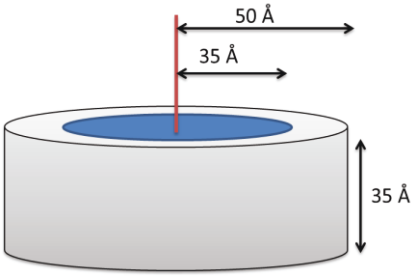


Figure 3 : Représentation de la partie recouverte par le bOG (gris) du complexe LH2 (bleu) par des cylindres.

En prenant comme paramètres du cylindre creux, les paramètres géométriques de 50 Å de rayon et 35 Å de hauteur (Gast *et al.* 1994) et la partie du complexe LH2 comme un autre cylindre de rayon et de hauteur de 35 Å (Figure 3). La différence de volume de ces deux cylindres nous donne le volume du cylindre creux qui représente le volume des détergents autour du complexe LH2 : $V_{det}=(\pi\, 50^2 35) - (\pi\, 35^2 35) = 140193{,}6\ Å^3$. En approximant ce volume par le volume d'une demi sphère (~2/3 de V_{det}) et en divisant par le volume moléculaire d'un bOG (417 $Å^3$ (Prince *et al.* 2003)) nous obtenons un nombre d'agrégation par LH2 de 225 bOG. À partir du nombre d'agrégation obtenu, nous avons effectué une simulation préassemblée du complexe LH2 en utilisant 225 molécules de bOG. La couronne de bOG qui entoure le complexe LH2 a été créée avec une approche de pré-assemblage en utilisant le serveur web CHARMM-GUI (Cheng *et al.* 2013) (Figure 4).

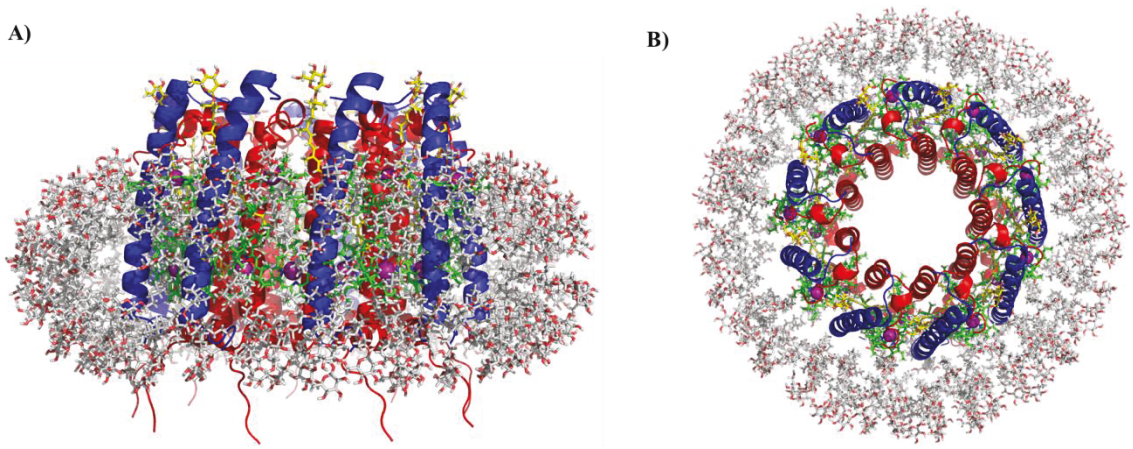


Figure 4 : Représentation du complexe LH2 entouré de bOG construit avec CHARMM-GUI de profil (A) et de haut (B).

Nous avons considéré que le bOG pouvait aussi être présent dans l'anneau central du complexe LH2 (Prince *et al.* 2003). Ainsi, nous avons remplacé manuellement les 18 molécules de LDAO présentes dans la structure 2FKW par des molécules de bOG et échantillonner les conformations pendant ~200 ns (Figure 5) (Système D).

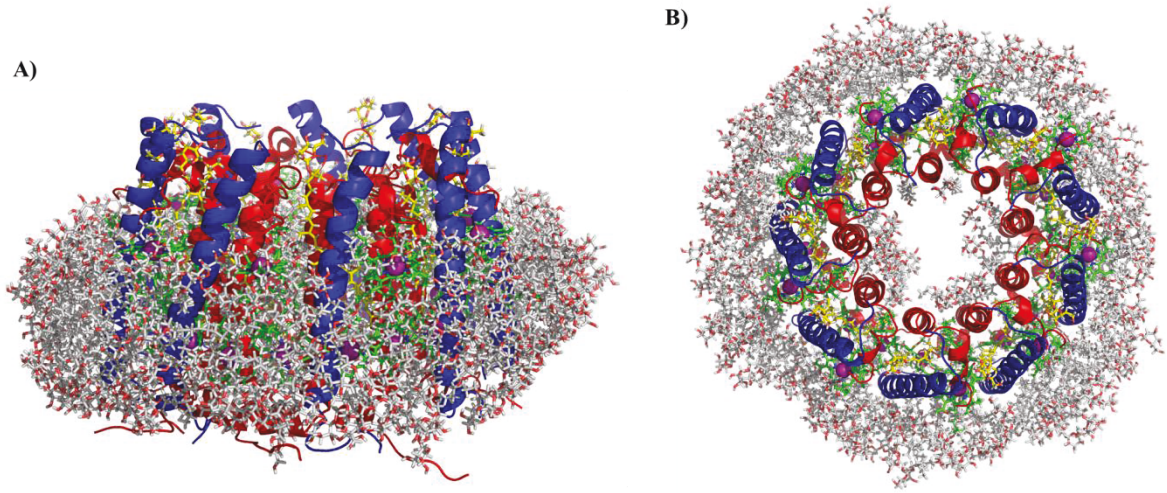


Figure 5 : Représentation final du complexe LH2 entouré de bOG et avec des 18 molécules de bOG dans l'anneau central.

5. Validations expérimentales des données des nombres d'agrégations de LDAO et bOG par LH2

Pour valider nos modèles complexes LH2-détergents, il est apparu intéressant de faire des expériences de SAXS sur la ligne de lumière SWING avec le Dr. Javier Pérez au Synchroton SOLEIL (voir annexe 4).

À partir des courbes de SAXS, nous pouvons calculer la taille du complexe LH2-détergent et obtenir une estimation du nombre de monomères de détergents autour du complexe LH2. Pour ce faire, nous avons utilisé le programme Memprot développé par Koutsioubas et Pérez (Pérez *et al*. 2015) qui se base sur une représentation géométrique d'un ellipsoïde (Figure 6) pour modéliser le complexe LH2 et le détergent. Les régions hydrophobes des détergents sont modélisées comme un tore elliptique de hauteur *a*, des demi-axes majeurs *be* et mineurs *b*/*e*. Quant à la partie hydrophile des détergents, elle est définie avec le paramètre *t*. Les détergents sont représentés par deux types de pseudo-atome avec des densités électroniques correspondant à la tête et à la queue du détergent, respectivement. Ces densités sont obtenues en divisant le nombre d'électron n_{e-} par le volume de van-der-Waals de la partie hydrophile/hydrophobe (Oliver *et al.* 2013).

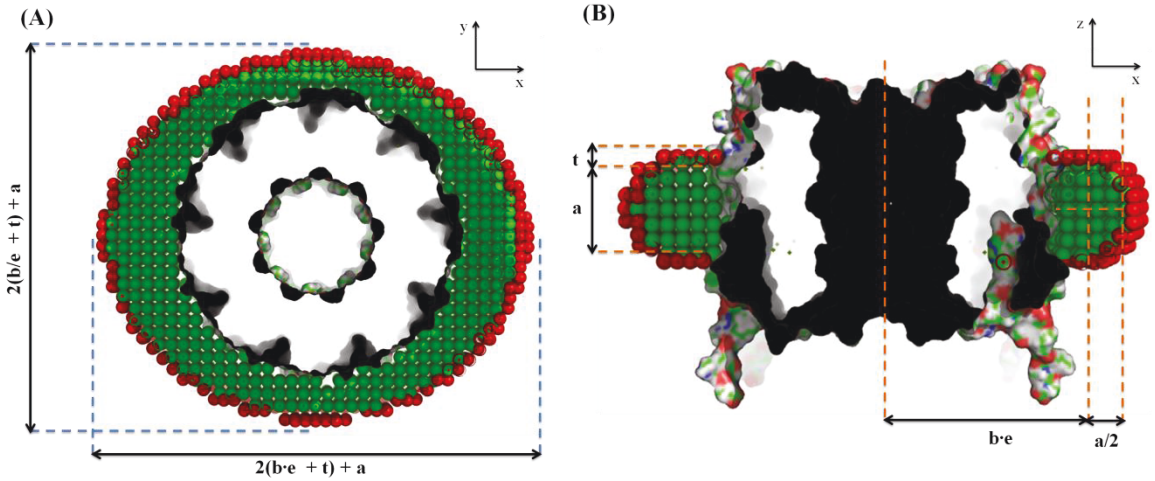


Figure 6 : Modèle du complexe LH2-bOG avec les paramètres géométriques utilisés par Memprot vu de haut (A) et de profil (B). Les sphères vertes correspondent aux parties hydrophobes des détergents et celles en rouge aux parties hydrophiles. Adaptée de (Pérez *et al.* 2015).

Le programme prend en entrée : une structure cristallographique centrée en (x,y,z)=0, un fichier avec les différents paramètres géométriques et électroniques (cités ci-dessous) et enfin le fichier avec la courbe de SAXS expérimentale. Memprot fait appel à un autre logiciel appelé CRYSOL (Svergun *et al.* 1995) où deux autres paramètres sont aussi pris en compte concernant l'optimisation des densités électroniques et le contraste de la couche d'hydratation. Nous avons testé différents paramètres de forme et le Tableau 2 décrit les valeurs obtenues avec la meilleure courbe ajustée à la courbe expérimentale (Figure 7). Précisons que les paramètres de forme connus dans la littérature, comme l'épaisseur de la couronne de détergents ou l'épaisseur des têtes hydrophiles des détergents, n'ont pas permis d'avoir un ajustement correct à la courbe expérimentale de SAXS.

Tableau 2 : Valeurs des paramètres géométriques et électroniques obtenus avec Memprot après ajustement sur la courbe expérimentale avec le LDAO et le bOG (Figure 7).

	Complexe LH2-LDAO	Complexe LH2-bOG
Rayon de la partie transmembranaire (Å)	32	32
a (Å)	23	20
b (Å)	40	40
t (Å)	1	2
e (Å)	0,9	1,1
$\rho_{hydrophile}$ (e.Å^{-3})	0,489	0,535
$\rho_{hydrophobe}$ (e.Å^{-3})	0,226	0,264
χ^2	22,19	10,43
$N_{détetgents}$	258	224

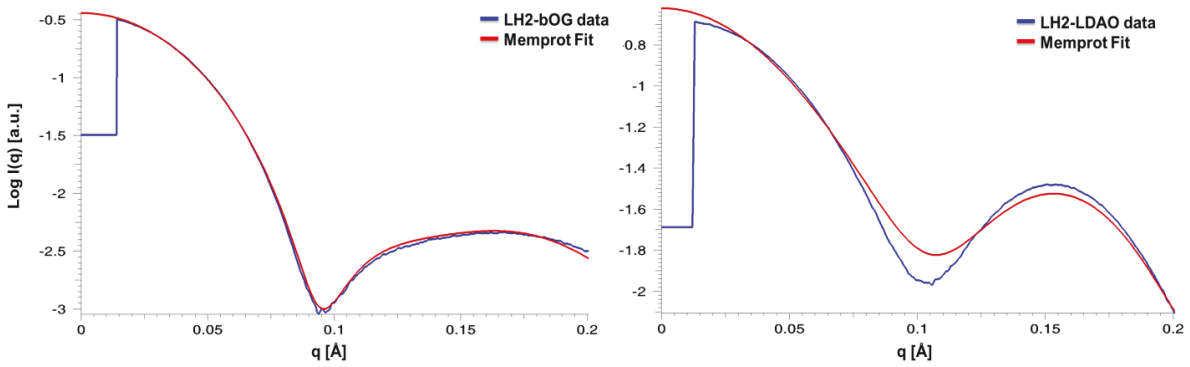


Figure 7 : Courbes de SAXS expérimentale (en bleu) ajustées par la courbe de SAXS de Memprot avec les meilleurs paramètres géométriques (en rouge).

Nous remarquons que dans le cas du LH2-bOG, l'ajustement de la courbe expérimentale avec les paramètres du Tableau 2 est excellent ($\chi^2 \approx 10$), ce qui n'est pas le cas pour le complexe LH2-LDAO ($\chi^2 \approx 22$). Le meilleur ajustement des courbes de SAXS expérimentales donne des nombres d'agrégation de 224 et 258 avec le bOG et LDAO, respectivement. La valeur pour le bOG est en bon accord avec celle utilisé dans nos simulations (225), au contraire de celle du système avec le LDAO (290, différence de 32 pour la couronne).

Nous pouvons expliquer cela par le fait que la partie hydrophile du LDAO a un contraste assez proche de celui de l'eau. Ainsi, lorsque le programme ajoute les gros grains hydrophiles du LDAO, l'ajustement ne change pas beaucoup. Pour la densité électronique des parties hydrophiles et hydrophobes du bOG, nous avons pris comme valeurs initiales

0,55 et 0,28 e.Å^{-3} (Oliver *et al.* 2013) et l'estimation par Memprot est de 0,53 et 0,26 après l'ajustement des courbes.

La Figure 8 représente la couronne de bOG estimée par Memprot (A, B, C) avec les têtes hydrophiles et les queues hydrophobes en sphère rouge et verte, respectivement. Nous avons superposé cette structure avec la dernière configuration de notre simulation LH2-bOG (D, E, F) pour comparer les deux couronnes de détergents. Ces dernières se recouvrent parfaitement et permet de valider notre modèle LH2-bOG.

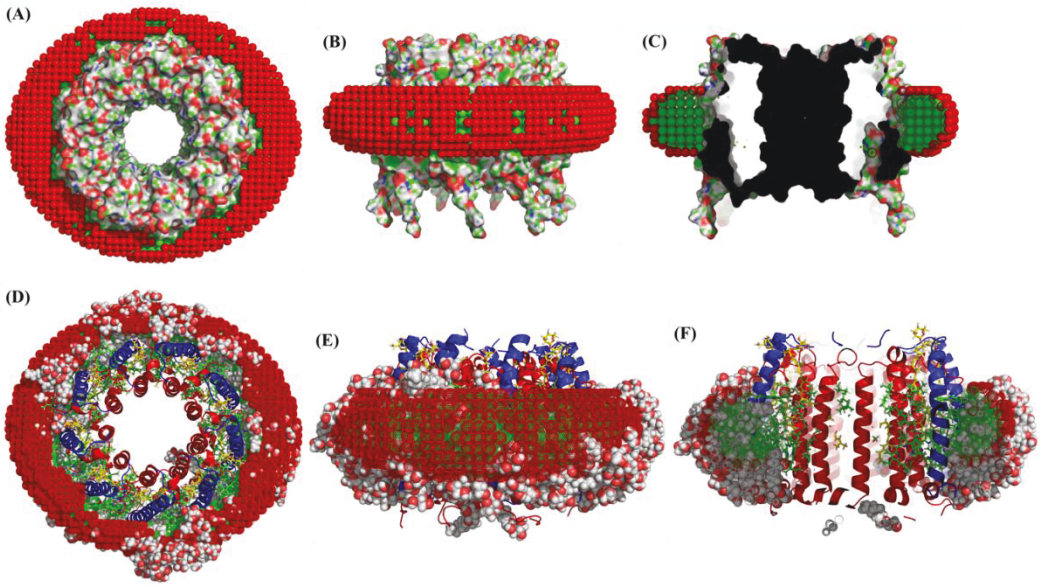


Figure 8 : Représentations de la couronne de bOG prédite par Memprot avec une vue de haut (A), de profil (B) et une coupe transversale (C). Le complexe LH2 est représenté en surface et les parties hydrophobes et hydrophiles du détergent par des sphères vertes et rouges, respectivement. Ce modèle a été superposé à la dernière configuration de notre simulation LH2-bOG représenté en (D)(E)(F) avec les bOG en sphères grises et rouge.

Malheureusement, dans le cas du LDAO (Figure 9), les problèmes de contrastes de densité électronique du détergent (0,42 et 0,27 e.Å^{-3} des parties hydrophiles et hydrophobes (Oliver *et al.* 2013), respectivement) nous limitent dans l'exploitation de la courbe expérimentale.

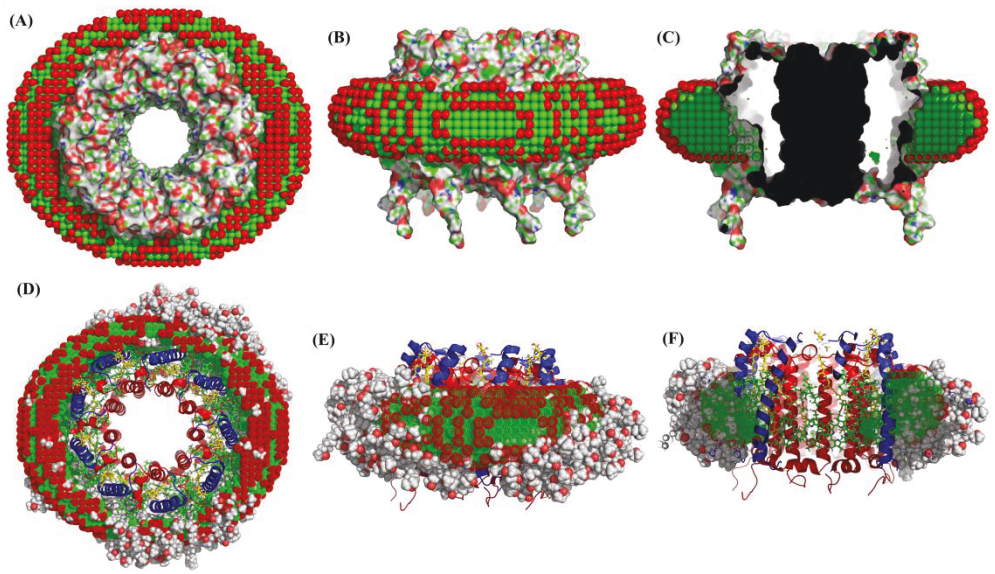


Figure 9 : Représentations de la couronne de LDAO prédite par Memprot avec une vue de haut (A), de profil (B) et une coupe transversale (C). Le complexe LH2 est représenté en surface et les parties hydrophobes et hydrophiles du détergent par des sphères vertes et rouges, respectivement. Ce modèle a été superposé à la dernière configuration de notre simulation LH2-LDAO représenté en (D)(E)(F) avec les LDAO en sphères grises et rouge.

6. Le complexe LH2 dans une membrane de POPC

Le dernier système que nous avons simulé consiste au complexe LH2 dans un environnement membranaire constitué de POPC (Système E). Pour ce faire, nous avons inséré le complexe LH2 dans une membrane de POPC en utilisant la méthodologie InflatGRO (Kandt *et al.* 2007). Cette approche consiste à « tasser » les lipides autour du complexe LH2 en étirant et en comprimant les lipides dans le plan transversal. En pratique, nous sommes partis de la dernière conformation de notre membrane pure de 128 POPC (voir annexe 2) que nous avons répliqué deux fois dans chacune des directions x et y, et une fois dans z pour que la nouvelle membrane (512 POPC) soit suffisamment grande pour y insérer le complexe LH2. Puis, nous avons placé le complexe LH2 au centre de la membrane de POPC. Pour empêcher que le complexe ne bouge pendant le processus de « tassage » nous avons utilisé, comme suggéré par les auteurs, une très grande force de contrainte (100 000 $kJ.mol^{-1}.nm^{-1}$) sur les atomes lourds de la protéine. Nous avons utilisé le programme inflatgro avec un facteur d'ajustement de la position des lipides de 1,3, un *cutoff* de rayon de 11,5 Å pour enlever les lipides, et un espacement de la grille de 5 Å pour calculer la surface des lipides. Cette opération a permis d'enlever 109 molécules de POPC et d'obtenir le système de la Figure 10.

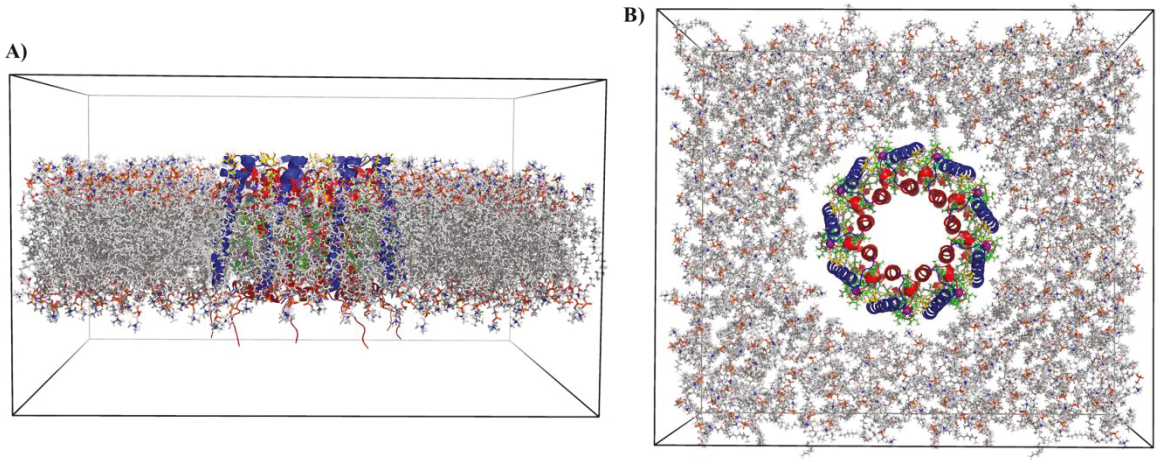


Figure 10 : Représentation du complexe LH2 inséré dans une membrane de 403 POPC, vue de profil (A) et de haut (B).

Nous avons « solvaté » le système avec 50 000 molécules d'eau dans une boîte d'environ 140x130x120 Å de dimension. Puis, en gardant le LH2 complexe figé, nous avons minimisé et équilibré dans l'ensemble NVT à 303 K (au-dessus de la température de transition de phase du POPC). L'équilibration du système a continué dans l'ensemble NPT en anisotropique pendant 135 ns avec le thermostat *v-rescale* et le barostat de Berendsen (τ_P = 3 ps), avec une compressibilité de $4{,}5.10^{-5}$ bar^{-1}. Le système équilibré a été simulé pour la production pendant ~ 200 ns dans l'ensemble NPT en anisotropique (Figure 11).

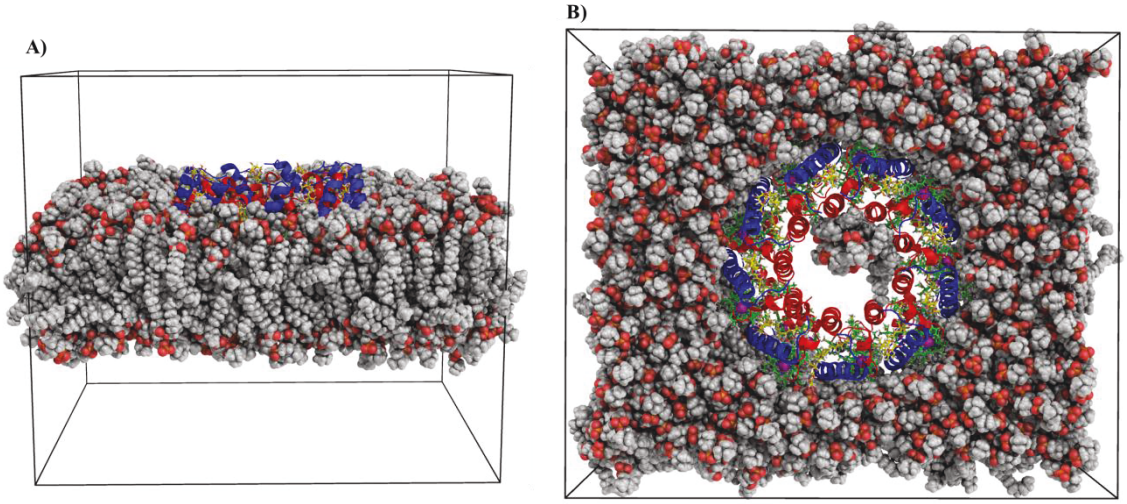


Figure 11 : Représentation de la configuration finale du système POPC-LH2 complexe avec des lipides (en sphère) à l'intérieur de l'anneau central.

Comme pour les autres systèmes, nous avons ajouté manuellement 6 molécules de POPC au centre de l'anneau central et simulé pendant environ 200 ns avec les mêmes paramètres décrits précédemment.

II. Résultats des simulations

Dans cette partie, nous présenterons les résultats obtenus des différentes propriétés structurales du complexe LH2 en présence de LDAO, bOG et POPC.

1. Caractéristiques globales des différents systèmes

1.1. Les diamètres du complexe LH2 dans nos simulations

Tout d'abord nous avons calculé le diamètre de l'anneau interne et externe du complexe LH2 à trois différents niveaux des hélices (Figure 12). Pour ce faire, nous avons choisi les carbones α des résidus proches de la partie N-terminale, C-terminale et dans la partie transmembranaire des peptides α et β. Sur les graphiques de la Figure 12, nous avons calculés les distances de l'anneau interne (31 Å, 31,5 Å et 30 Å) et de l'anneau externe (52,4 Å, 60 Å et 62,4 Å) dans la structure cristallographique et nous les avons comparées à celles calculées dans nos différents systèmes (courbes bleues, jaunes et marrons pour l'anneau interne et courbes noires, rouges et vertes pour l'anneau externe).

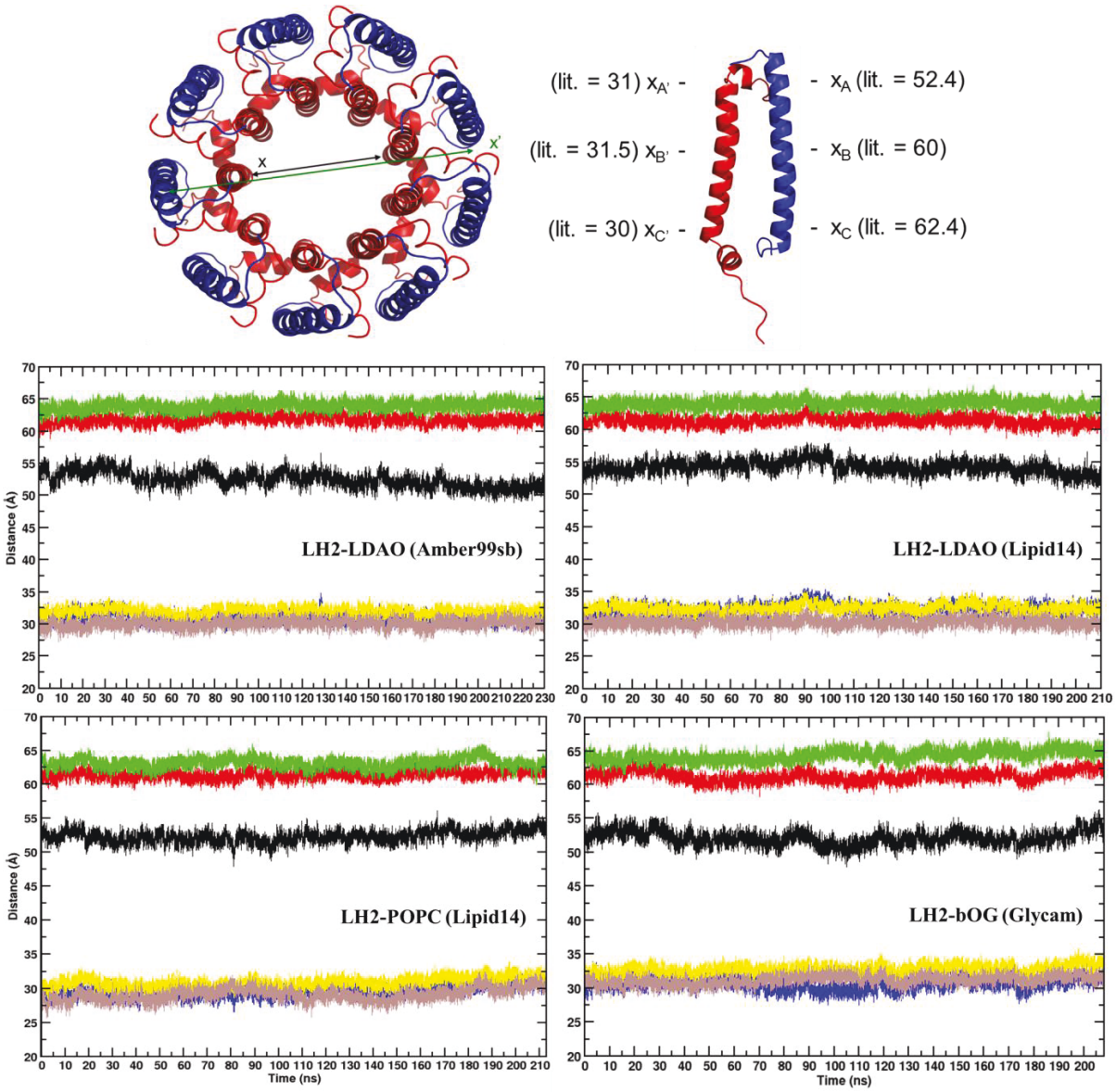


Figure 12 : Graphiques des diamètres de l'anneau interne et externe du complexe LH2 dans les différents systèmes comparées aux valeurs de la structure cristallographique. Les valeurs XA', XB' et XC' sont les trois diamètres de l'anneau interne calculés à partir de la structure cristallographique, et XA, XB et XC sont ceux pour l'anneau externe. Par comparaison aux valeurs dans nos simulations avec les courbes bleues, jaunes et marrons pour l'anneau interne et, noires, rouges et vertes pour l'anneau externe.

La Figure 12, montre que les distances restent stables au cours de la simulation et sont en bon accord avec les valeurs de la structure cristallographique 2FKW (Cherezov *et al.* 2006) reportées dans la dernière colonne du Tableau 3.

Tableau 3 : Diamètres calculés dans nos simulations comparées aux valeurs de la structure cristallographique (en Å).

	LH2-LDAO		LH2-POPC	LH2-bOG	Structure RX
	Amber99sb	Lipid14			2FKW
X_A	52,5 ± 0,1	54,2 ± 0,1	52,3 ± 0,8	52,1 ± 0,1	52,4
X_B	61,7 ± 0,6	61,3 ± 0,6	61,4 ± 0,6	61,3 ± 0,8	60
X_C	64,0 ± 0,6	64,0 ± 0,6	63,1 ± 0,7	64,3 ± 0,8	62,4
$X_{A'}$	31,1 ± 0,7	32,4 ± 0,7	29,6 ± 0,8	30,6 ± 0,8	31
$X_{B'}$	32,0 ± 0,6	32,4 ± 0,6	31,0 ± 0,7	32,8 ± 0,7	31,5
$X_{C'}$	30,1 ± 0,6	30,3 ± 0,6	29,1 ± 0,9	31,2 ± 0,7	30

1.2. Les dimensions et la stabilité du complexe LH2 entouré de détergents

Nous avons aussi analysé la structure globale des différents systèmes en calculant les dimensions du complexe entouré de la couronne de détergents. Pour ce faire, nous avons calculé les principaux moments d'inertie des différents composants des systèmes (peptides, peptides et détergents, complexe LH2, complexe LH2 et détergents) avec l'outil *g_gyrate* de GROMACS :

$$I_1 = \frac{M}{5}(b^2 + c^2) \quad \text{(II.1)}$$

$$I_2 = \frac{M}{5}(a^2 + c^2) \quad \text{(II.2)}$$

$$I_3 = \frac{M}{5}(a^2 + b^2) \quad \text{(II.3)}$$

où M est la masse total du groupement étudié, et a, b, et c, les trois demi-axes de l'ellipsoïde.

À partir de ces principaux moments d'inertie, nous avons extrait les valeurs des trois demi-axes et calculé l'excentricité *e*=1-a/c, avec a et c les demi-axes mineurs et majeurs respectivement (Tableau 4). Nous remarquons que toutes les formes sont oblates (a < b ≈ c) et que le complexe LH2-LDAO est légèrement plus excentré comparé à celui avec du bOG (0,47 et 0,45 ± 0,6 respectivement). De plus, la valeur de l'excentricité *e* du complexe LH2-LDAO est légèrement plus faible que celle obtenue dans l'étude de spectroscopie en molécule unique avec une simulation MC (0,52) (Oijen *et al.* 1999) (Hong *et al.* 2004).

Dans le Tableau 5, nous avons comparé nos valeurs de forme avec celles obtenues avec Memprot. Globalement, les modèles générés par Memprot surestiment le demi-axe mineur

et sous-estiment le demi-axe majeur comparé à nos modèles issus des simulations. Cependant, dans le cas du complexe LH2-bOG, le rapport des demi-axes majeur sur mineur montre une forme ellipsoïde marquée avec Memprot et dans nos simulations (1,22 ± 0,1 et 1,83 ± 0,3, respectivement). Dans le cas du complexe LH2-LDAO, le modèle proposé par Memprot suggère une forme beaucoup moins ellipsoïde (0,80 ± 0,6) en comparaison à nos simulations (1,90 ± 0,4 et 1,89 ± 0,2). Cette différence de forme est probablement la conséquence du mauvais ajustement de la courbe de SAXS effectué par Memprot.

Tableau 4 : Valeurs moyennées des demi-axes (Å) des peptides, peptides avec détergents/lipides, complexe LH2 et du complexe LH2 avec les détergents/lipides dans les différents systèmes. L'erreur sur les valeurs avec les complexe LH2/peptides-détergents est de ± 0,6 et de ± 0,2 pour les peptides/complexe LH2.

	LH2-LDAO (Amber99sb)				**LH2-LDAO (Lipid14)**				**LH2-bOG (Glycam)**			
	a	b	c	*e*	a	b	c	*e*	a	b	c	*e*
Peptides	34,9	40,1	41,1	0,15	35,1	40,1	40,8	0,14	34,4	40,6	41,4	0,17
Peptides-détergents	30,3	54,2	57	0,47	30,7	54,2	57,5	0,47	30,5	51,6	55,1	0,45
Complexe LH2	32,4	41	42,1	0,23	32,5	41,1	41,8	0,22	31,8	41,4	42,4	0,25
Complexe LH2-détergents	29,2	52,4	55,4	0,47	29,6	52,8	55,8	0,47	29,3	50,6	53,7	0,45

Tableau 5 : Comparaison des paramètres géométriques (en Å) issus de nos simulations et ceux obtenus avec le meilleur ajustement de la courbe de SAXS par Memprot.

	LH2-bOG simulation	**LH2-bOG Memprot**	**LH2-LDAO simulation**		**LH2-LDAO Memprot**
			Amber99sb	**Lipid14**	
Demi-axe mineur	29,3 ± 0,2	36,2 ± 0,1	29,2 ± 0,2	29,6 ± 0,4	44 ± 0,6
Demi-axe majeur	53,7 ± 0,5	44,2 ± 0,1	55,4 ± 0,6	55,8 ± 0,6	35,6
Rapport des demi-axes	1,83 ± 0,3	1,22 ± 0,1	1,90 ± 0,4	1,88 ± 0,2	0,80 ± 0,6

Dans la continuité de l'analyse de la stabilité des systèmes micellaires, nous avons calculé le rayon de giration des différents composants des systèmes (i.e peptides, RG, BChl-a, complexe LH2, des détergents seuls, des peptides et des détergents, du complexe LH2 et des détergents). Pour ce faire, nous avons utilisé l'outil *g_gyrate* de GROMACS qui utilise la formule :

$$R_g = \left(\frac{\sum_i \|r_i\|^2 m_i}{\sum_i m_i}\right)^{1/2} \qquad \text{(II.4)}$$

avec m_i la masse de l'atome i et r_i la position de l'atome i par rapport au centre de la molécule.

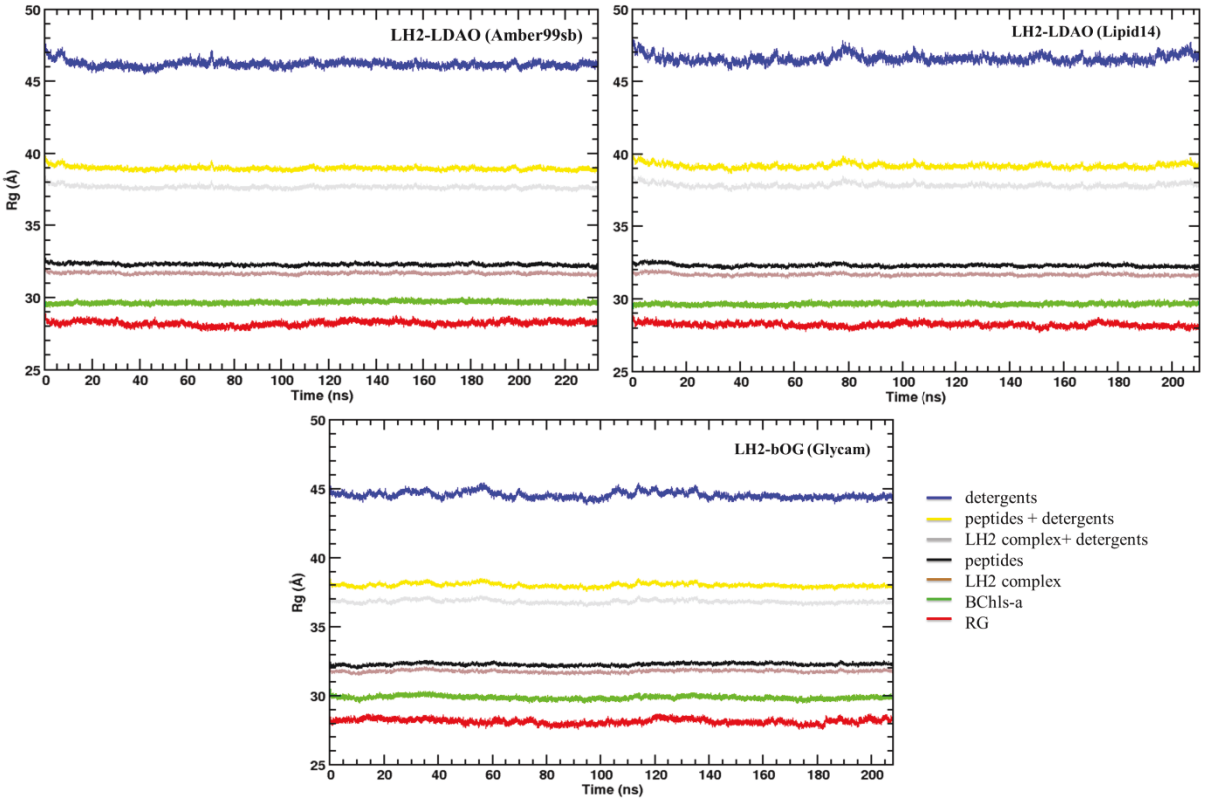


Figure 13 : Rayons de giration des détergents (bleu), des peptides et des détergents (jaune), du complexe LH2 et des détergents (gris), peptides (noir), du complexe LH2 (marron), des BChls-a (vert) et des RG (rouge) dans les différents systèmes.

Les graphiques de la Figure 13 représentent les rayons de giration des différents groupements dans les systèmes micellaires en fonction du temps de simulation. Nous remarquons que les valeurs des R_g restent très stables tout au long de la simulation. De plus, les valeurs du R_g du complexe LH2 entouré de bOG (36,8 ± 0,1 Å) ont des valeurs légèrement plus petites comparé à celles du complexe LH2 entouré de LDAO (37,7 ± 0,1 Å).

À partir des courbes de SAXS, nous avons calculé la pente de la droite aux très faibles valeurs de q selon la loi de Guinier (Guinier 1937) ($P(q) \approx \exp\left(-\frac{q^2 R_g^2}{3}\right)$) qui nous donne le rayon de giration R_g. Nous obtenons des valeurs de 38,1 ± 0,2 Å et 28,8 ± 3,4 Å pour le système LH2-bOG et LH2-LDAO, respectivement. Nous remarquons que le rayon de giration pour le complexe LH2-bOG est assez proche de la valeur que nous avons calculé dans nos simulations avec du bOG (36,8 ± 0,1 Å) alors que le rayon de giration du complexe LH2-LDAO est beaucoup petit que celui de nos simulations (37,7 ± 0,1 Å).

1.3. Le recouvrement membranaire du complexe LH2

Le LH2 étant un complexe membranaire, nous avons aussi examiné le recouvrement de la partie hydrophobe du complexe en fonction du type de détergent et de la membrane. La Figure 14 montre une coupe transversale de la dernière configuration des systèmes (A, B, C et D). Nous pouvons voir que dans le cas du POPC, le complexe LH2 est bien recouvert par les lipides. Concernant les systèmes avec le bOG et LDAO, la Figure 14 indique que, qualitativement, le LDAO et le bOG recouvrent la majorité de la partie transmembranaire du complexe LH2.

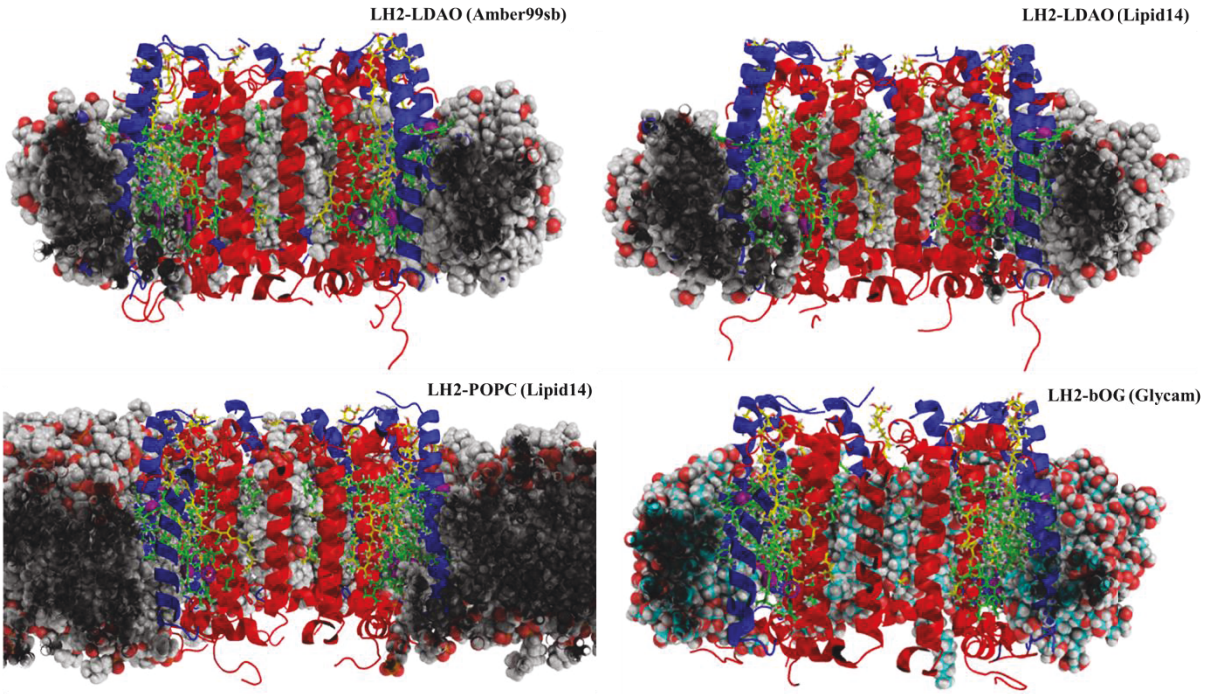


Figure 14 : Coupe transversale des dernières configurations du système LH2-LDAO (Amber99sb), LH2-LDAO (Lipid14), LH2-POPC (Lipid14) et LH2-bOG (Glycam). Les molécules de LDAO et de POPC sont représentées par des sphères grises et le bOG par des sphères de couleur cyan.

Pour l'analyse du recouvrement globale des parties transmembranaires par les détergents, nous avons calculé les profils de densité des différents composants des systèmes LH2-détergents orientés selon l'axe Z (Figure 15). Pour ce faire, nous avons réorienté et centré les systèmes selon l'axe Z et nous avons utilisé l'outil de calcul de densité *g_density* de GROMACS. La distance entre les deux pics dans la courbe de densité des têtes hydrophiles des détergents (en rouge) nous donne l'épaisseur de notre couronne de détergents. Nous obtenons une valeur d'environ 30 Å dans nos trois systèmes (LDAO Amber99sb, LDAO Lipid14 et bOG Glycam) qui est en bon accord avec la valeur d'OPM (*Orientation Protein Membrane*) (31,8 ± 0,6 Å) (Lomize *et al.* 2006). Cette dernière regroupe les protéines membranaires de la PDB sur lesquelles des calculs théoriques d'arrangements spatiaux dans une bicouche lipidique ont été réalisés et comparés avec les données expérimentales. Nous remarquons aussi que l'amplitude des LDAO est légèrement plus élevée que celle des bOG, ce qui nous suggère que dans l'axe Y, la couronne de LDAO est légèrement plus grande que celle formée par le bOG.

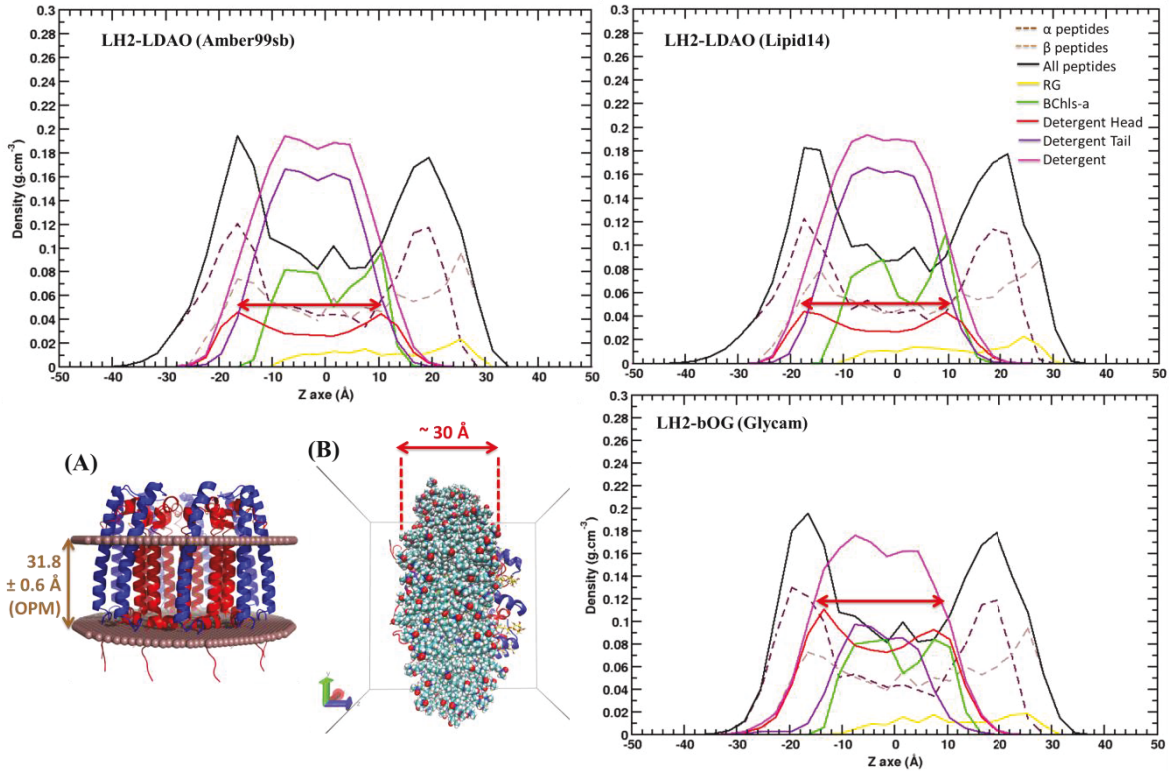


Figure 15 : Densité des différents composants dans les différents systèmes avec une représentation du recouvrement de la partie transmembranaire du LH2 par les détergents (flèches rouges). (A) l'épaisseur de la partie transmembranaire extrait de la base de données OPM et (B) une représentation du complexe LH2-détergents pour illustrer les différents graphiques.

Nous avons aussi étudié le recouvrement au niveau de chaque peptide. Pour ce faire, nous avons calculé la différence relative Σ_x des surfaces accessibles au solvant (SAS) à l'aide du programme *trjVoronoi* (Abel *et al.* 2012), en présence et absence de détergents ($SAS_x^{détergent}$ et SAS_x^{eau}, respectivement) :

$$\sum_x = \frac{\left(SAS_x^{eau} - SAS_x^{détergent}\right)}{SAS_x^{eau}} \quad \text{(II.5)}$$

Les graphiques A, B et D de la Figure 16 représentent les différences relatives en fonction du numéro de résidu pour le système avec le LDAO Amber99sb, LDAO Lipid14 et bOG Glycam, respectivement. Les résidus des peptides α (en rouge) et β (en bleu) sont dans un environnement hydrophobe pour des valeurs de $\Sigma x > 0$ et dans un environnement hydrophile pour des valeurs de $\Sigma x < 0$. Nous avons comparé nos valeurs avec le degré d'hydrophobicité théorique des résidus issu de la PDBTM (*Protein Data Bank of Transmembrane proteins*) (Kozma *et al.* 2012) représenté par des barres roses. Nous remarquons que ces dernières se superposent aux pics des Σx. De plus, nous avons comparé nos valeurs avec le score d'hydropathie de l'approche Kyte & Doolittle (Kyte *et al*. 1982) sur le serveur ExPASy (Wilkins *et al.* 1999). Cette approche évalue l'hydropathie des résidus à partir d'une séquence en acide aminé et à l'aide une fenêtre de résidu à définir. Dans notre cas, nous avons choisi la fenêtre par défaut de 9 résidus (Figure 16C). Nous voyons que les valeurs de Σx sont aussi en bon accord avec les scores Kyte & Doolittle. Ces résultats nous montrent que la couronne de détergents recouvre bien les parties transmembranaires du complexe LH2.

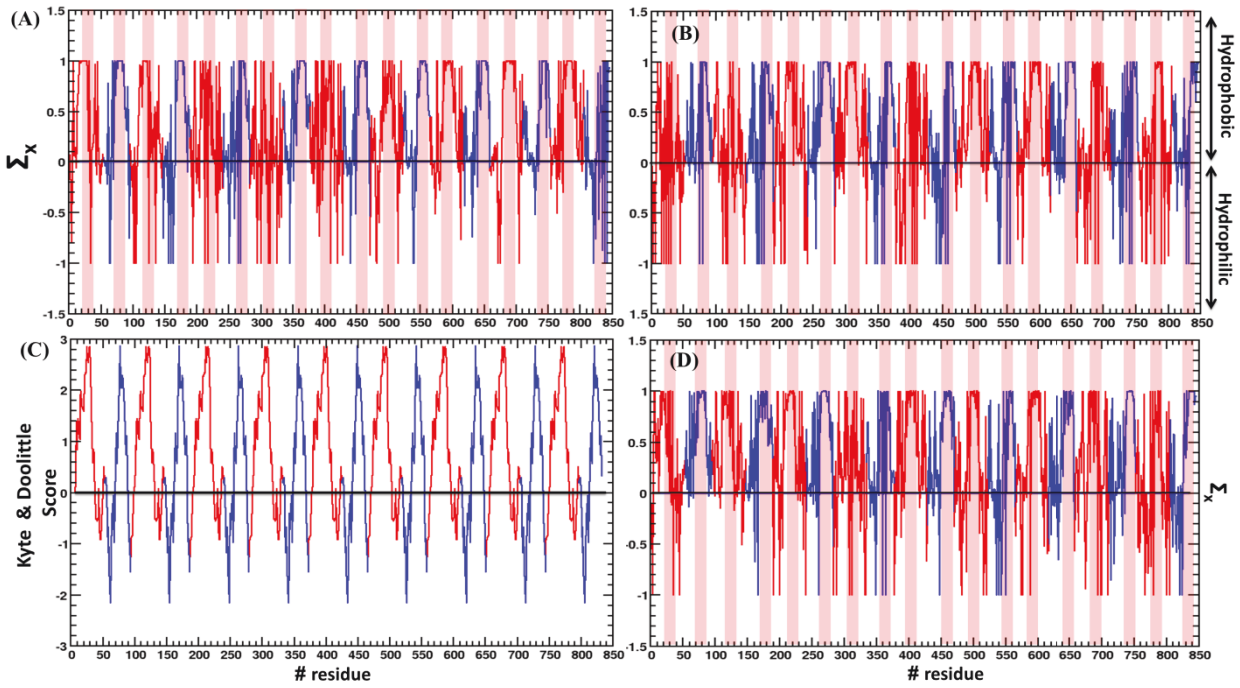


Figure 16 : Différences relatives Σx avec le LDAO Amber99sb (A), le LDAO Lipid14 (B) et le bOG Glycam (D) en fonction du résidu. Les barres roses représentent le degré d'hydrophobicité des résidus issue de la PDBTM. (C) le score Kyte & Doolittle. Les courbes rouges représentent les peptides α et celles en bleu les peptides β.

2. Analyse des caractéristiques des détergents autour du complexe LH2

Dans cette sous partie, nous allons présenter les caractéristiques des détergents autour du complexe LH2 et les comparer avec celles des micelles pures de LDAO (Karakas *et al.* 2016) et de bOG (cf. annexe 1) que nous avons aussi simulées.

2.1. Conformation des détergents

Nous avons analysé la conformation des détergents en calculant le pourcentage des conformations *trans* des angles dièdres des chaînes alkyles CCCC. Les valeurs présentées dans le Tableau 6 nous montrent que les pourcentages obtenus dans les simulations du complexe LH2-détergents sont similaires à ceux obtenus dans les simulations de micelles pures.

Tableau 6 : Pourcentage des conformations *trans* des angles dièdres des chaînes alkyles CCCC. *moyenne des micelles pré-assemblées et auto-assemblées.

LDAO				bOG	
Amber99sb		**Lipid14**		**Glycam**	
Micelle pure*	LH2	Micelle pure*	LH2	Micelle pure*	LH2
75	76,2	63	62,3	74,5	74,3

2.2. Surfaces accessibles au solvant et hydratation des têtes polaires des détergents

Puis nous avons étudié les surfaces des détergents en calculant les surfaces accessibles au solvant (SAS) à l'aide du programme *trjVoronoi* (Abel *et al*. 2012). Le Tableau 7 présente les résultats obtenus dans les différentes simulations en présence de LDAO et de bOG. Nous remarquons que les valeurs des SAS dans les simulations du complexe LH2-détergents sont assez proches des valeurs obtenues dans les simulations de micelles pures.

Tableau 7 : Surface accessible au solvant des chaînes alkyles et des têtes dans les différentes simulations (en Å^2), avec une erreur maximale de 1 Å² pour les micelles pures de LDAO, de 3,4 Å² pour les micelles pures de bOG, de 2 Å² pour le complexe LH2-LDAO et de 3 Å² pour le complexe LH2-bOG. *moyenne des micelles pré-assemblées et auto-assemblées.

	LDAO				bOG	
	Amber99sb		**Lipid14**		**Glycam**	
	Micelle pure*	LH2	Micelle pure*	LH2	Micelle pure*	LH2
$SAS_{chaînes}$	54,5	47,1	54,6	50,8	27,7	26,3
$SAS_{têtes}$	76,9	74,9	77,1	77,9	115,3	107,5
SAS_{total}	131,4	122,0	131,7	128,7	143	133,8

Pour aller plus loin dans l'analyse structurale des détergents, nous avons analysé la première couche d'hydratation des têtes polaires des détergents dans les différentes simulations. Pour ce faire, nous avons calculé la fonction de densité radiale $g_{O_N-O_W}(r)$ entre les atomes d'oxygènes des têtes polaires (O_N) et ceux des molécules d'eau (O_W) :

$$g_{O_N-O_W}(r) = \frac{1}{N_{O_N}\langle\rho_{O_W}\rangle}\sum_{i\in O_N}^{N_{O_N}}\sum_{j\in O_W}^{N_{O_W}}\frac{\delta(r_{ij}-r)}{4\pi r^2} \qquad \text{(II.6)}$$

avec N_{O_N} le nombre d'atomes O_N, ρ_{O_W} la densité des atomes O_W et r la distance entre les atomes O_N et O_W.

En intégrant les valeurs de $g_{O_N-O_W}(r)$ à leur premier minimum pour le bOG et le LDAO (3,5 et 3,4 Å, respectivement), nous avons pu estimer le nombre de molécules d'eau dans la première couche d'hydratation des têtes polaires du bOG et du LDAO (Tableau 8 et 9, respectivement). Nous remarquons que le nombre de molécules d'eau entre les simulations de micelles pures de détergents et de micelles associées au complexe LH2 sont très similaires.

Tableau 8 : Nombre de molécules d'eau dans la première couche d'hydratation des oxygènes des têtes polaires du bOG. *moyenne des micelles pré-assemblées et auto-assemblées.

	bOG	
	Micelle pure*	Complexe LH2
O1A	0,4	0,4
O2A	1,6	1,6
O3A	1,7	1,7
O4A	2,5	2,2
O5A	0,5	0,5
O6A	2,2	1,9

Tableau 9 : Nombre de molécules d'eau dans la première couche d'hydratation de l'oxygène des têtes polaires du LDAO. *moyenne des micelles pré-assemblées et auto-assemblées.

	LDAO			
	Amber99sb		**Lipid14**	
	Micelle pure*	LH2	Micelle pure*	LH2
O1	2,5	2,4	2,7	2,5

L'ensemble de ces résultats, nous montrent que les molécules de LDAO et de bOG associées au complexe LH2 reproduisent bien les propriétés caractéristiques du LDAO et du bOG décrites dans nos simulations de micelles pures, et sont donc dans des états proches de ceux trouvés dans les micelles pures.

2.3. Épaisseur des têtes polaires

Puis, nous avons calculé l'épaisseur des têtes polaires du LDAO et du bOG dans nos différentes simulations afin de les comparer avec les valeurs estimées par Memprot (Tableau

10). Nous remarquons que dans le cas du bOG, Memprot sous représente la taille des têtes hydrophiles comparé à celle du LDAO.

Tableau 10 : Épaisseurs des têtes polaires dans nos simulations et avec Memprot (en Å).

	LH2-bOG simulation	LH2-bOG Memprot	LH2-LDAO simulation		LH2-LDAO Memprot
			Amber99sb	Lipid14	
Têtes hydrophiles	4,4 ± 1,0	2	1,1 ± 0,3	1,1 ± 0,3	1

3. Analyses structurales de la stabilité des peptides

Pour aller plus loin et avant d'étudier plus finement le recouvrement membranaire des peptides par les détergents, nous avons analysé la stabilité de la partie transmembranaire des différents peptides au cours du temps de simulation. Pour ce faire, nous avons calculé le RMSD (*Root Mean Square Deviation*) ou l'écart quadratique moyen des carbones α de la partie transmembranaire des peptides α et β (résidus 19-37 et 18-36, respectivement) par rapport à la structure minimisée. Ce RMSD est défini dans GROMACS (*g_rms*) par l'équation :

$$RMSD\ (t_1, t_2) = \left[\frac{1}{M}\sum_{i=1}^{N} m_i \|\boldsymbol{r_i}(t_1) - \boldsymbol{r_i}(t_2)\|^2\right]^{1/2} \qquad \text{(II.7)}$$

avec $M = \sum_{i=1}^{N} m_i$ et $\mathbf{r}_i(t_x)$ la position de l'atome i au temps t_x.

Les graphiques de la Figure 17 montrent les valeurs obtenues pour le RMSD au cours du temps pour chaque partie transmembranaire des 9 peptides α (4 graphiques du haut) et des 9 peptides β (4 graphiques du bas) dans nos différentes simulations. La partie transmembranaire a été définie grâce aux valeurs théoriques de la base de données OPM (Lomize *et al.* 2006). Dans notre cas, nous avons utilisé les données pour la structure PDB 1NKZ.

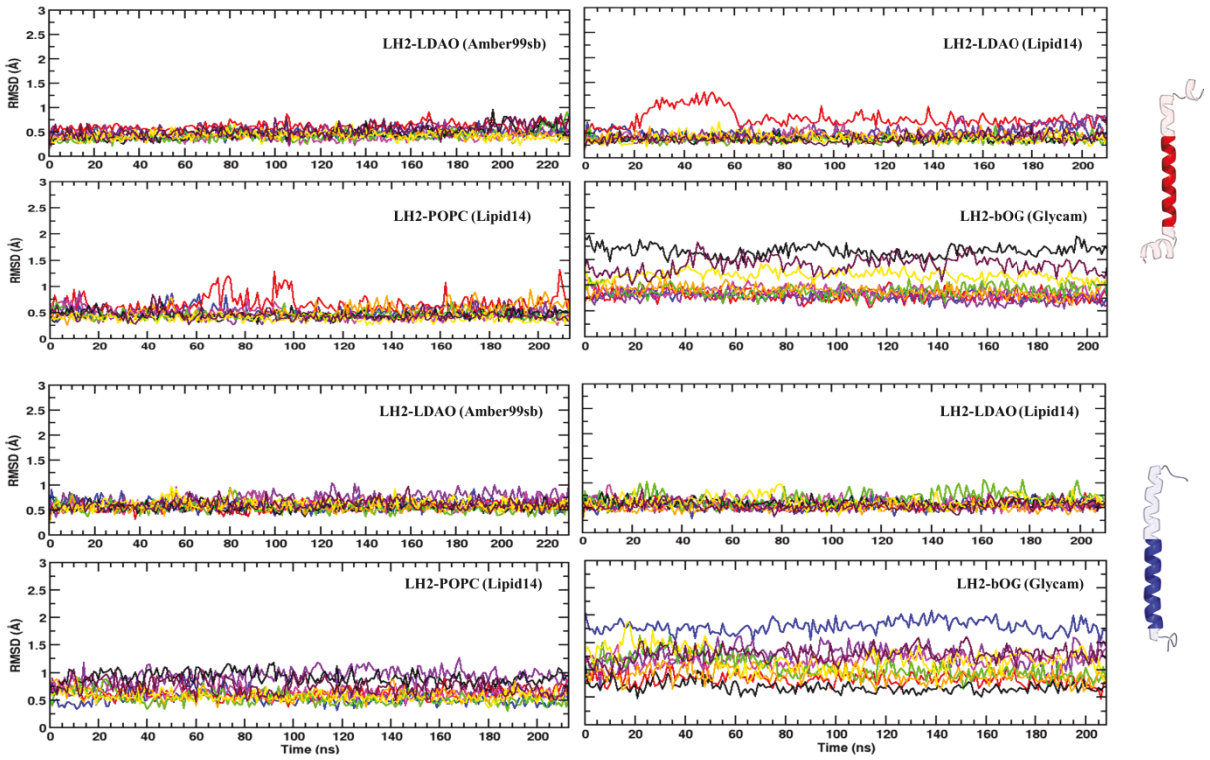


Figure 17 : RMSD sur la carbones α de la partie transmembranaire des 9 peptides α (4 graphiques du haut) et des 9 peptides β (4 graphiques du bas) dans nos 4 quatre simulations. Chaque couleur représente un peptide et la partie hélicale mis en avant dans la représentation des deux peptides montre la partie transmembranaire définie par la base de données OPM (voir le corps du texte).

Nous remarquons que globalement, les valeurs du RMSD restent stables au cours de la simulation hormis quelques fluctuations dans deux peptides α (courbes rouges). Ces fluctuations peuvent s'expliquer par une perte locale de la structure secondaire observée dans les résultats du calcul des structures secondaires en fonction du temps (voir annexe 5). Dans le cas de la simulation avec le bOG, les valeurs du RMSD des peptides α et β sont légèrement plus élevées (valeurs entre 1-2 Å) comparé à celles des systèmes avec le LDAO (valeurs entre 0,5-1 Å). Cette observation est en bon accord avec la littérature où il a été montré que le bOG influençait sur la structure du complexe LH2 (Liu *et al.* 2011).

4. Analyses des interactions entre peptides, pigments BChls-a et peptides-pigments BChls-a

4.1. Dynamique entre peptides

Nous avons en premier lieu caractérisé les mouvements des peptides et des bactériochlorophylles-a au sein des complexes LH2. Pour ce faire, nous avons calculé les corrélations de mouvement entre paires d'atomes grâce à l'outil *g_correlation* développé en C

par Lange et ses collaborateurs (Lange *et al*. 2006) pour GROMACS v3.3. Cette méthode se base sur le calcul du coefficient de corrélation de Pearson généralisé d'une matrice de covariance des fluctuations atomiques :

$$C_{ij} = \frac{\langle x_i \cdot x_j \rangle}{\sqrt{\langle x_i^2 \rangle \langle x_j^2 \rangle}} \quad \text{et} \quad r_{MI}[\mathbf{x_i}, \mathbf{x_j}] = \left\{1 - e^{\left(-2I[\mathbf{x_i}, \mathbf{x_j}]/d\right)}\right\}^{-1/2} \qquad \text{(II.8)}$$

où $\mathbf{x}_i$ et $\mathbf{x}_j$ sont les vecteurs de fluctuations des atome i et j.

La Figure 18 montre les différentes matrices des corrélations croisées des mouvements atomiques des carbones α des peptides α (numéros en orange) et β (numéro en cyan) dans nos différentes simulations. Les parties en bleu représentes les mouvements les moins corrélés ou opposés et celles en rouge représentes les mouvements les plus corrélés ou identiques. Nous remarquons que dans le cas du LDAO, en plus de la grande diagonale rouge foncée qui correspond aux mouvements des atomes identiques, des mouvements corrélés existent entre les peptides α adjacents (petite diagonale rouge-blanche). En revanche, dans le cas de la simulation avec le POPC et le bOG, ces mouvements sont moins corrélés et nous montre que les peptides α adjacents ont des mouvements opposés. Entre les peptides β voisins, dans toutes les simulations, nous remarquons que les mouvements de leur squelette peptidique ne sont pas corrélés comparé aux peptides α adjacents. Nous pouvons expliquer cette observation par la présence des bactériochlorophylles-a 800 entre les peptides β qui augmentent l'espace entre ces peptides. Dans le cas des dipeptides αβ d'une même paire, dans toutes les simulations, nous remarquons que les parties N-terminales (αβ) et C-terminales (αβ) ont les mêmes mouvements. Ces résultats sont en bon accord avec les données cristallographiques (Papiz *et al.* 2003) qui montrent la présence de liaisons hydrogènes entre les parties N- et C-terminale des peptides α et β d'une même paire. Enfin, il est intéressant de noter qu'au sein d'un même peptide α ou β, la partie N-terminale a un mouvement opposé à la partie C-terminale. Nous observons aussi ce mouvement opposé dans notre analyse en composantes principales (ACP) sur le squelette peptidique (voir annexe 6).

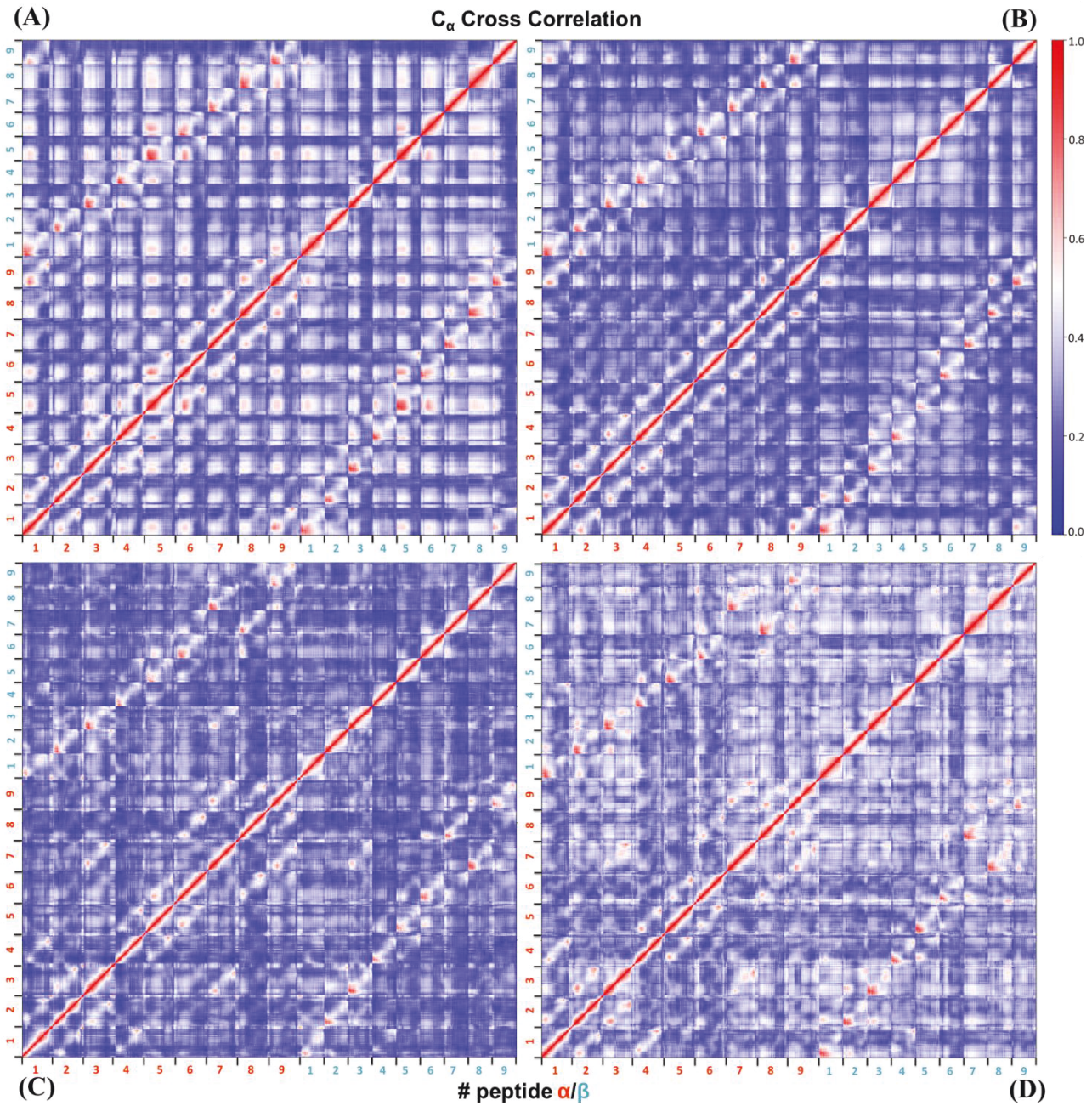


Figure 18 : Matrices des corrélations croisées des fluctuations atomiques des carbones α dans la simulation avec le LH2-LDAO Amber99sb (A), le LH2-LDAO Lipid14 (B), le LH2-POPC (C) et le LH2-bOG (D). Les numéros en orange correspondent aux peptides α et ceux en cyan aux peptides β. L'échelle de couleur représente les mouvements du moins corrélé (bleu) au plus corrélé (rouge). Les calculs ont été réalisés tout au long des trajectoires et la figure a été générée avec le package Bio3d (Grant *et al.* 2006) de R (R Development Core Team 2008).

4.2. Dynamique entre bactériochlorophylles-a

Le système d'électron π du cycle bactériochlorine qui coordonne l'ion Mg^{2+} aux quatre atomes N pyrroliques, est responsable des bandes d'absorption des BChls-a. C'est pourquoi nous nous sommes intéressés plus particulièrement aux mouvements des atomes du magnésium des bactériochlorophylles-a 800 et 850. La Figure 19 montre les différentes

matrices des corrélations croisées des mouvements des magnésiums dans les B800. Nous remarquons que dans les simulations avec du LDAO Lipid14, POPC et bOG (Figure 19B, C et D, respectivement), le mouvement des atomes de magnésium n'est pas très corrélé. En revanche, dans la simulation avec le LDAO Amber99sb, certains atomes de magnésium proches (7 et 6) ou éloignés (1 et 5) montrent des mouvements relativement similaires.

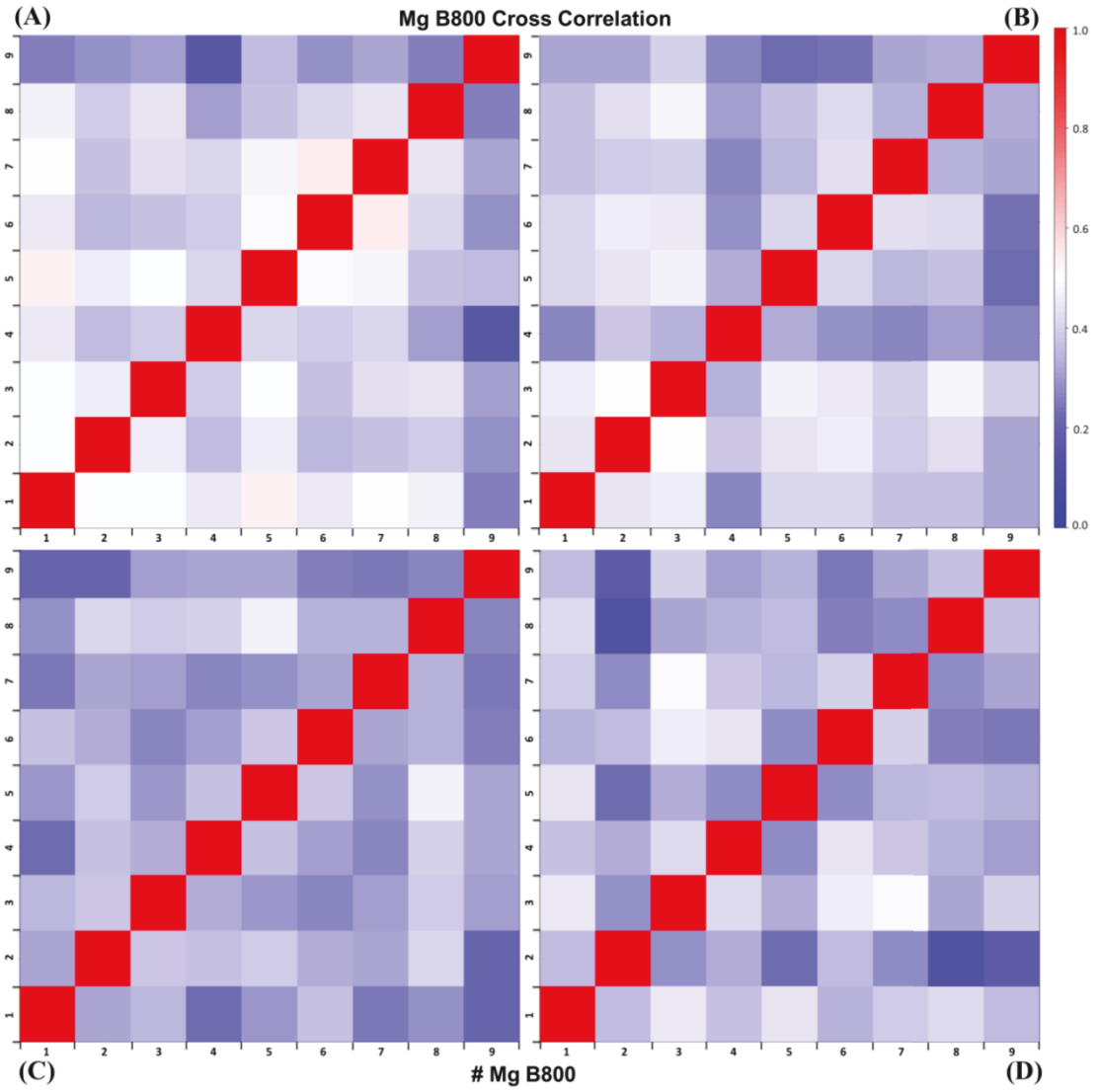


Figure 19 : Matrices des corrélations croisées des fluctuations atomiques des magnésiums B800 dans les simulations de LH2-LDAO Amber99sb (A), le LH2-LDAO Lipid14 (B), le LH2-POPC (C) et le LH2-bOG (D).

La Figure 20 montre les différentes matrices des corrélations croisées des mouvements des magnésiums dans les B850. Dans le cas des simulations avec le LDAO (Figure 20A et B), nous remarquons que tous les magnésiums adjacents ont un mouvement similaire. En revanche, dans le cas des simulations avec le POPC et le bOG (Figure 20C et D), certains magnésiums qui sont voisins ont des mouvements opposés (par exemple, α3 et β4, α6 et β7)

alors que d'autres qui sont éloignés ont des mouvements plus similaires (par exemple, α9 et β4, α6 et β1).

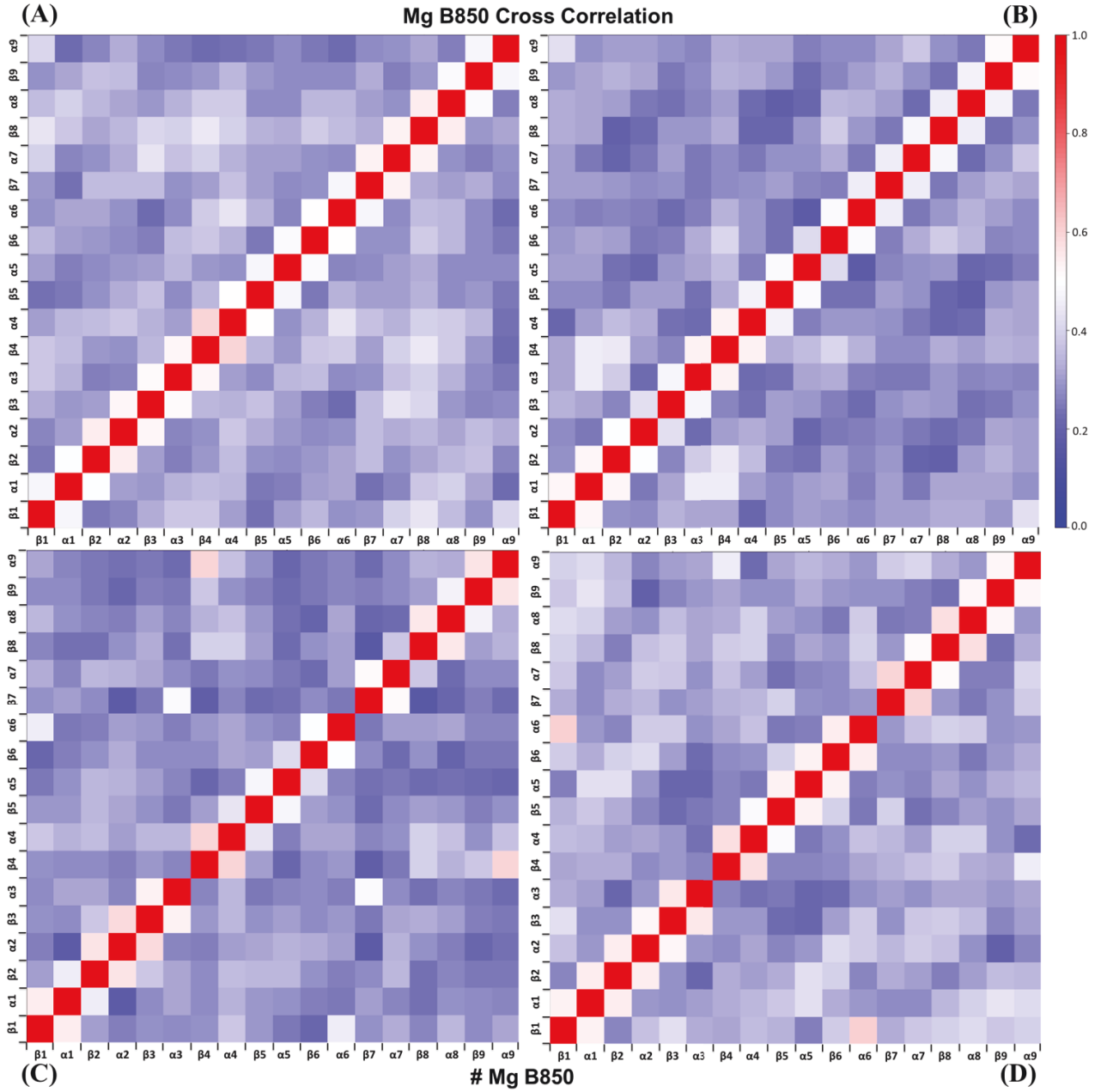


Figure 20 : Matrices des corrélations croisées des fluctuations atomiques des magnésiums αB850 et βB850 dans les simulations de LH2-LDAO Amber99sb (A), le LH2-LDAO Lipid14 (B), le LH2-POPC (C) et le LH2-bOG (D).

Ces corrélations de fluctuations nous ont permis de voir le mouvement atomique des carbones α ou des magnésiums par rapport aux autres atomes Cα ou magnésiums tout au long de nos simulations. Nous constatons que l'environnement micellaire ou membranaire a un effet sur ces mouvements. Pour aller plus loin, dans notre analyse nous avons aussi calculé les ACP

(cf. annexe 6) des distances clefs connues dans la littérature pour avoir un lien avec la fonction biologique des complexes LH2.

4.3. Analyse des distances fonctionnelles clefs

Interactions impliquant les B800

Il a été montré que des distances et des liaisons hydrogènes entre peptides-pigments et pigments-pigments étaient importantes dans le maintien et le couplage excitonique des BChls-a au sein du complexe LH2 (Robert *et al.* 1985). Nous avons calculé les distances entre les B800, les arginines 20 et les B800, et les CXM et les B800 (Figure 21).

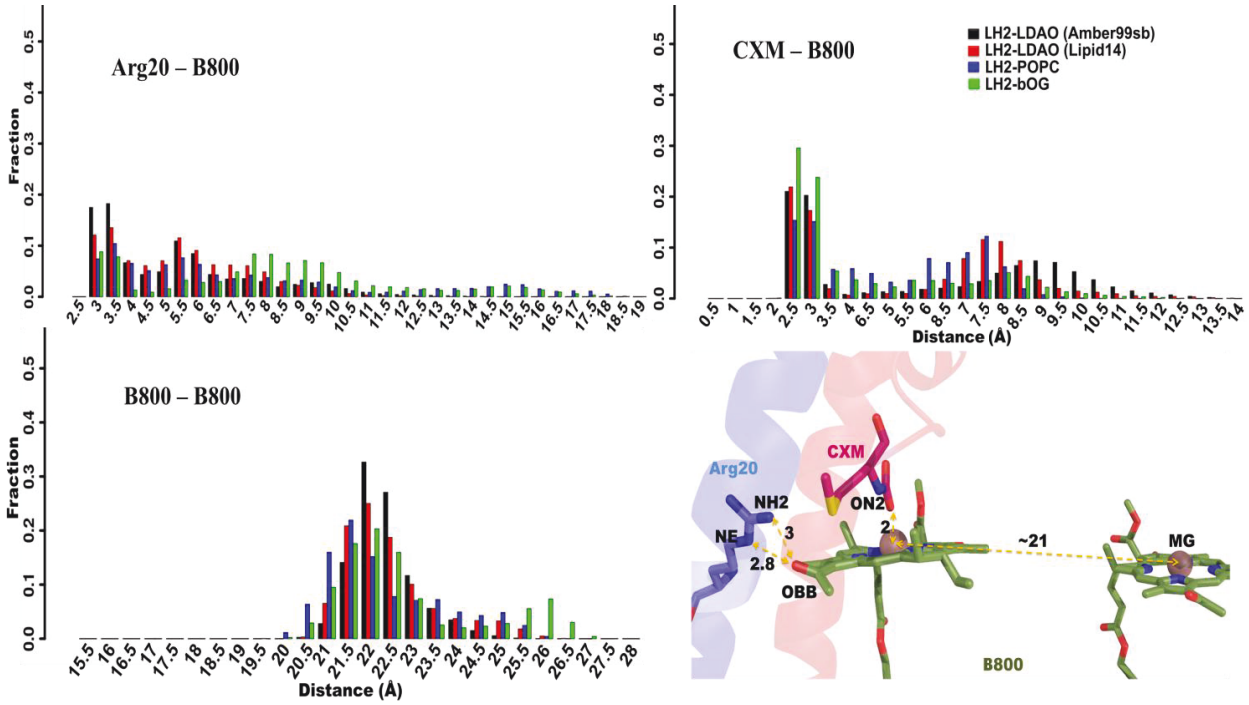


Figure 21 : Histogrammes des distances impliquant les B800 et une représentation de ces interactions avec les distances en ligne jaune discontinue (en Å), tirée de la littérature (Papiz *et al.* 2003).

Nous remarquons que globalement les distributions sont assez étalées, ce qui nous montre une assez grande variabilité dans les distances. Pour la distance entre l'arginine 20 (atomes d'azote NE ou NH2) et le B800 (atome d'oxygène OBB), nous pouvons voir deux pics, l'un à 3 - 3,5 Å dans toutes les simulations et l'autre plus loin à 5,5-6 Å avec du LDAO ou du POPC, et à 7,5 - 8 Å dans le cas du bOG. De plus, nous remarquons que dans les simulations avec du POPC et du bOG, la queue de distribution s'étend jusqu'à 18 Å. Nous retrouvons aussi cette tendance avec la distance entre les CXM (atomes d'oxygène ON2) et les B800 (atome Mg) avec un premier pic à 2,5-3 Å dans toutes les simulations et un deuxième pic à

7,5 Å avec le LDAO Lipid14 et POPC, et à 8-9 Å pour le bOG et le LDAO Amber99sb. En revanche, dans le cas des distances entre les magnésiums des B800, toutes les distributions de distance sont centrées à 22 Å avec un deuxième petit pic à 26 Å dans le cas de la simulation avec le bOG.

Nous avons aussi calculé le pourcentage de présence de chaque liaison hydrogène entre les amines NE-HE/NH2-HH21 des arginines 20 et l'oxygène OBB des B800 illustré dans la Figure 21. Le calcul a été réalisé avec l'outil de GROMACS *g_hbond* avec un rayon et un angle maximal de 3,4 Å et 30° respectivement, pour la liaison N-H---O (Jang *et al.* 2015).

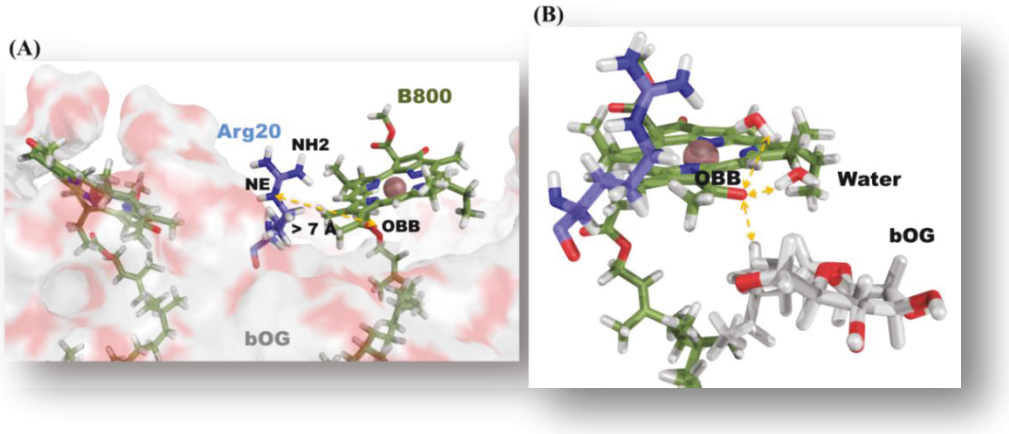


Figure 22 : Représentations des interactions Arg20-B800 avec un exemple d'éloignement (A) et les causes de ces éloignements (B) qui sont dus aux molécules de détergents.

Tableau 11 : Pourcentage de présence des liaisons hydrogènes entre les 9 Arg20(NE-HE)-B800(OBB)/Arg20(NH2-HH21) et les B800(OBB) au cours des différentes simulations. Les calculs ont été réalisés tout au long des trajectoires.

Peptide	LH2-LDAO (Amber99sb)		LH2-LDAO (Lipid14)		LH2-POPC		LH2-bOG	
	NE	NH2	NE	NH2	NE	NH2	NE	NH2
# 1	15	17	50	5	49	3	0	0
# 2	59	23	43	16	1,1	0,2	87	50
# 3	81	43	15	5	0	0	21	21
# 4	3	2	22	14	30	19	0	0
# 5	43	16	25	13	0	0	0	0
# 6	77	23	2,5	3	0	0	26	21
# 7	0,3	0	20	1	60	17	0	0
# 8	22	7	24	19	0	0	0	0
# 9	25	8	29	8	53	3	0	0
Total	325	139	231	84	193	42	134	92

Les résultats du Tableau 11 nous montre que chaque liaison hydrogène peut être perturbée très localement dans un peptide et presque jamais perturbée dans un autre (i.e LH2-LDAO Amber99sb avec 3 et 81 % dans le peptide 4 et 3, respectivement). En revanche, dans la simulation avec le POPC et le bOG, la majorité des liaisons hydrogènes n'existent pas pendant la simulation. La Figure 22 montre que les détergents viennent s'insérer autour des B800, ce qui empêche la formation de liaisons hydrogènes avec les arginines 20 et favorisent les interactions avec les molécules de détergents ou d'eau.

Interactions impliquant les B850

Dans le même esprit, nous avons aussi analysé les distances entre les B850, et entre les peptides et les B850. Pour ce faire, nous avons calculé les distances entre les magnésiums des B850 (Figure 23A), l'azote des histidines α31/β30 (NE1) et les magnésiums des B850 (Figure 23B), et entre l'oxygène (OBB) des B850 et l'azote des tryptophanes 45 (αW45) ou l'oxygène des tyrosines 44 (αY44) (Figure 23A).

Dans le cas des histidines, les distributions sont parfaitement centrées à 3 Å et sont très similaires. Cela montre le maintien de cette interaction durant les simulations et ce, quel que soit l'environnement micellaire ou membranaire. De même, pour les distances entre les magnésiums des B850 avec des distributions similaires centrées entre 9-10 Å. Dans les distances entre les Trp45/Tyr44 et les B850, nous remarquons que toutes les distributions sont centrées à 3-3,5 Å mais avec les Trp45, un deuxième pic est observable dans les simulations avec du POPC/bOG et du LDAO Amber99sb, avec une valeur de 5 et 8 Å, respectivement. Pour les distances entre les Tyr44 et les B850, la majorité des valeurs se situe à 3 Å avec quelques valeurs autour de 5,5 et entre 9-13 Å dans les simulations avec du bOG et du POPC, respectivement. Cela explique les faibles pourcentages de présence des liaisons hydrogènes observés dans le Tableau 12 (voir plus loin).

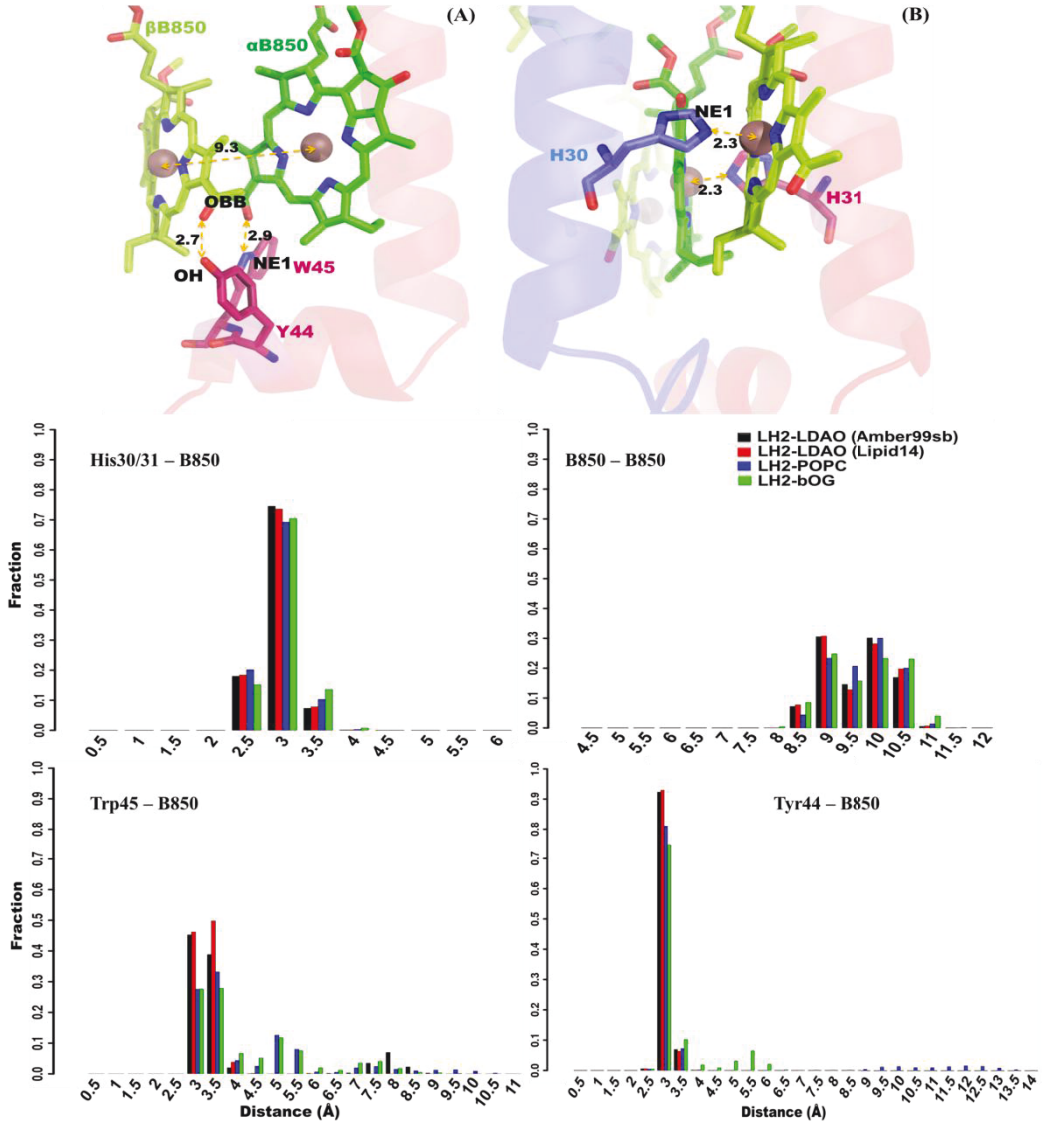


Figure 23 : Histogrammes des distances impliquant les B850 et deux représentations de ces interactions à partir de la structure cristallographique 1NKZ.

Il existe de fortes interactions entre les histidines 30/31 et les B850 alors qu'avec les tyrosines 44 et les tryptophanes 45, il arrive que leur interaction avec les B850 soit perturbée. Cela nous suggère que les magnésiums des B850 sont bien maintenus par les histidines 30/31 qui se situent dans un environnement totalement hydrophobe empêchant ainsi de fortes perturbations entre les magnésiums des B850. En revanche, les résidus Trp45 et Tyr44 qui sont un peu moins confinés comparé aux histidines 30/31, peuvent perdre leur liaison hydrogène avec le groupement OH des macrocycles des B850. Pour les Tyr44, la Figure 24 nous montre que des interactions à moins de 3 Å avec le détergent/lipide et d'autres résidus ou l'eau peuvent entrainer un éloignement entre les Tyr44 et les B850.

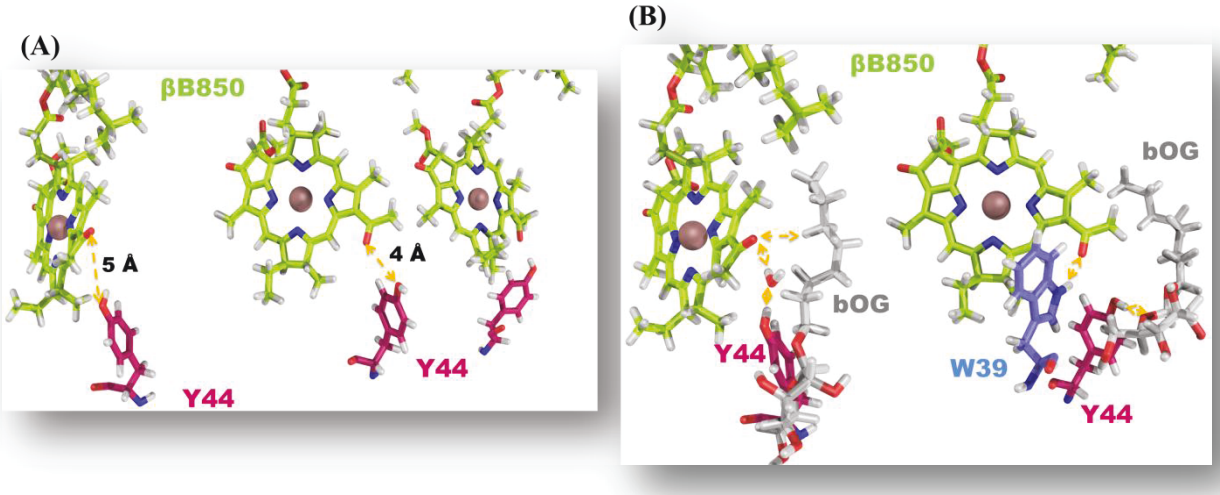


Figure 24 : Représentations des interactions Tyr44-B850 dans nos simulations avec un exemple de deux éloignements (A) et les causes de ces éloignements (B) qui sont dus aux molécules de bOG, d'autres résidus du peptide β ou à des molécules d'eau.

Pour les Trp45, des interactions à moins de 3 Å avec les rhodopines glucosides, d'autres résidus ou l'eau peuvent entrainer un éloignement entre les Trp45 et les B850 (Figure 25). Dans cet exemple, nous remarquons que le changement de conformation du tryptophane 45 explique les grandes distances de l'histogramme à ~ 7 Å de la Figure 23.

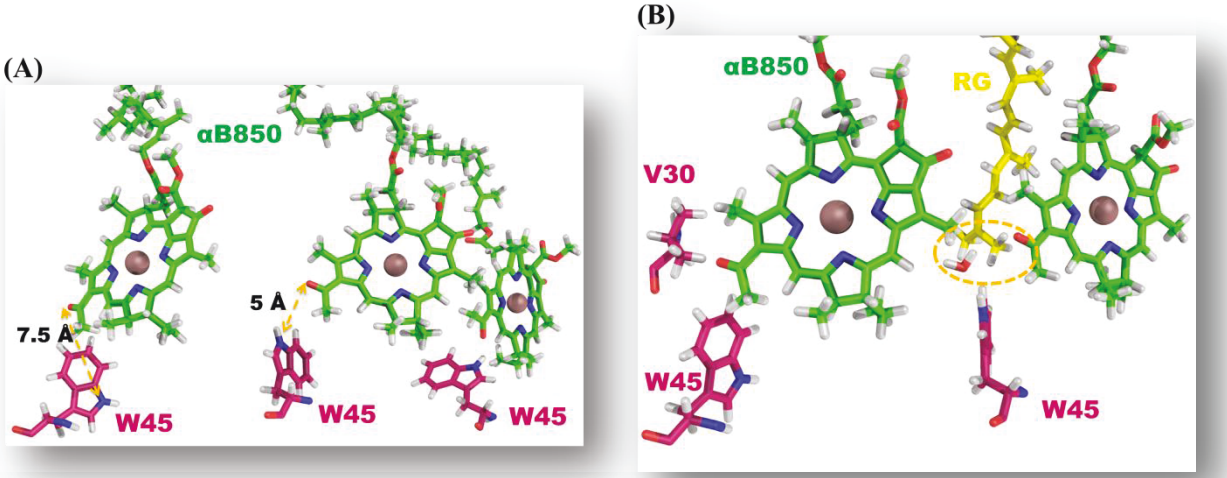


Figure 25 : Représentations des interactions Trp45-B850 avec un exemple de deux éloignements (A) et les causes de ces éloignements (B) qui sont dus aux molécules de RG, d'autres résidus du peptide α ou à des molécules d'eau.

Ces changements d'interaction entre les résidus et les bactériochlorophylles-a changent considérablement l'orientation globale des macrocycles ainsi que celle des groupements hydroxyles. De plus, cela est aussi observable dans l'analyse ACP que nous avons faite sur les

atomes lourds du macrocycle des B850 et B800 qui montrent des fluctuations au cours des simulations (voir annexe 6).

Tableau 12 : Pourcentage de présence des liaisons hydrogènes entre les 9 Tyr44(OH-HH)-βB850(OBB) et les 9 Trp45(NE1-HE1)-αB850(OBB) au cours des différentes simulations.

Peptide	LH2-LDAO (Amber99sb)		LH2-LDAO (Lipid14)		LH2-POPC		LH2-bOG	
	Y44	W45	Y44	W45	Y44	W45	Y44	W45
# 1	100	90	99	85	99	79	98	51
# 2	99	88	100	90	99	0	99	91
# 3	99	94	100	88	99	0	96	86
# 4	100	94	99	87	99	85	100	89
# 5	99	16	99	89	99	0	96	90
# 6	99	95	100	88	99	89	99	0
# 7	100	88	100	89	100	74	19	25
# 8	100	44	100	93	99	80	100	95
# 9	100	94	100	76	0	72	96	20
Total	896	703	897	785	793	479	803	547

5. Analyse des corrélations entre les différentes distances clefs

À partir des fluctuations des différentes distances clefs que nous avons calculées, nous voulions allez plus loin en essayant d'analyser les possibles corrélations entre ces différentes interactions. Pour ce faire, nous avons calculé les matrices de corrélations croisées sur l'ensemble des distances grâce au logiciel R (R Development Core Team 2008) et des scripts écrits au laboratoire. Les quatre matrices de la Figure 26 montrent un exemple de corrélation dans les différentes simulations, entre deux types de distances impliquant des atomes assez éloignés : distances Trp-B850 et CXM-B800 (noté TrpY et MgCXMY, respectivement, avec Y allant de 1 à 9). Les corrélations positives sont en bleu et celles négatives en rouge. La partie supérieure des diagonales montre l'intensité de la couleur et la taille des cercles qui sont proportionnelles aux coefficients de corrélation. Ces derniers sont inscrits dans la partie inférieure des diagonales.

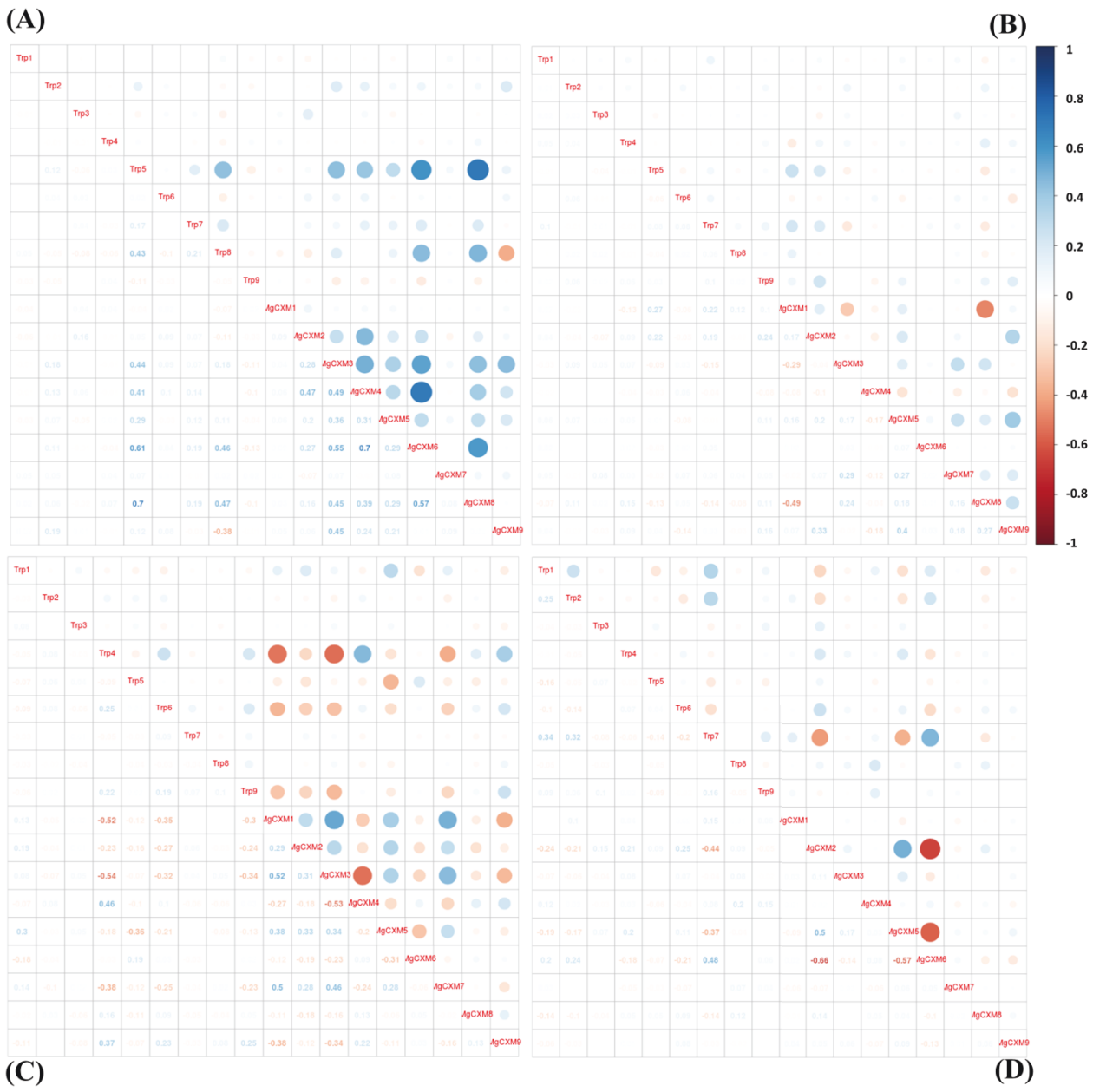


Figure 26 : Matrices croisées des corrélations des 9 distances Trp-B850 (noté Trp1-9) et CXM-B800 (noté MgCXM1-9) dans les systèmes LH2-LDAO Amber99sb (A), LH2-LDAO Lipid14 (B), LH2-POPC (C) et LH2-bOG (D). L'intensité de la couleur et la taille des cercles sont proportionnelles aux coefficients de corrélation. Le dégradé de bleu correspond aux corrélations positives ou mouvements similaires et celui en rouge, aux anti-corrélations ou mouvements opposés.

Nous remarquons qu'il existe très peu de mouvements corrélés dans les simulations avec le LH2-LDAO Lipid14 et bOG contrairement aux simulations avec le LDAO Amber99sb et POPC. Parmi ces corrélations, dans la simulation du LH2-LDAO Amber99sb, nous pouvons voir qu'il y a plus de corrélations positives entre les distances CXM-B800 comparé à la simulation avec du POPC où les corrélations sont négatives. Prenons l'exemple de la distance Trp5 dans la simulation avec le LDAO Amber99sb, nous remarquons que ces mouvements

sont similaires à ceux des MgCXM6 et MgCXM8 avec un coefficient de corrélation de 0,51 et 0,7 respectivement. En revanche dans la simulation avec du POPC, les mouvements Trp4 sont anti-corrélés ou opposés avec ceux des MgCXM1 et MgCXM3, avec un coefficient de corrélation de -0,52 et -0,54 respectivement. Cette analyse nous montre que des distances impliquant des atomes éloignés peuvent avoir des mouvements assez similaires ou opposé selon l'environnement micellaire ou membranaire. De plus, cela nous suggère que ces environnements biomimétiques influent localement sur la dynamique interne des complexes LH2.

III. Conclusions de cette partie

Dans ce chapitre, nous avons montré que :

1. Le nombre de monomères de détergent autour du complexe LH2, que nous avons tiré de la littérature (dans le cas du LDAO) ou estimé à partir de la littérature (bOG), nous ont permis de simuler nos complexes LH2 dans les environnements micellaires. Dans le but de valider nos modèles, nous avons réalisé des mesures de SAXS du complexe LH2-LDAO et LH2-bOG et calculé le nombre de détergents autour du complexe LH2 grâce au logiciel Memprot. Dans le cas du bOG, l'estimation du nombre de monomère de bOG issue des expériences de SAXS (224) est très proche de la valeur utilisée dans notre simulation (225). Par contre dans le cas du LDAO, les problèmes de contrastes de densité électronique du détergent avec le solvant nous ont limités dans l'exploitation de la courbe expérimentale de SAXS, et la valeur d'un nombre d'agrégation estimée à partir des courbes de SAXS est éloignée de celle utilisée dans nos simulations.
2. Les calculs des propriétés structurales pour les différents complexes LH2-détergents/lipides ont montré que les diamètres des anneaux internes et externes du complexe LH2 ne varient pas et restent stables quel que soit l'environnement micellaire et membranaire indiquant une stabilité de la forme générale concentrique du complexe. L'analyse des paramètres de forme des complexes LH2-détergents ont montré une forme ellipsoïde oblate ($a < b \approx c$) avec une légère différence de forme, qui est un peu plus excentrique en présence de LDAO comparé au bOG. Cette conclusion se retrouve aussi dans les valeurs du Rg du complexe LH2 entouré de LDAO (37,7 ± 0,1 Å) et de bOG (36,8 ± 0,1 Å). Ce résultat est en bon accord avec les résultats des mesures de SAXS réalisées en présence de complexe LH2-bOG (38,1 ± 0,2 Å). Malheureusement, dans le cas du LDAO, les problèmes de contrastes de densité électronique du détergent avec le

solvant nous limitent dans l'exploitation de la courbe expérimentale de SAXS et nous obtenons des rayons de giration plus faible (28,8 ± 3,4 Å) difficilement explicables.

3. Dans le cas du recouvrement des parties hydrophobes par les détergents, les analyses du recouvrement global des parties transmembranaires ainsi que le recouvrement au niveau de chaque peptide montrent que la couronne de LDAO et de bOG recouvrent parfaitement les parties transmembranaires du complexe LH2. De plus, l'analyse des monomères de LDAO et de bOG associés au complexe LH2 a montré qu'ils reproduisent bien les propriétés caractéristiques du LDAO et du bOG (conformation des chaînes alkyles, hydratation et épaisseur des têtes polaires) décrites dans nos simulations de micelles pures. Nous avons comparé l'épaisseur des têtes polaires des bOG dans nos simulations (4,4 ± 1,0 Å) avec celle obtenue avec Memprot (2 Å) qui nous montre que ce dernier sous-estime l'épaisseur des têtes.
4. Concernant la stabilité des peptides α/β constituant le complexe, le calcul du RMSD sur les carbones α des parties transmembranaires ont montré que dans le cas de la simulation avec le bOG, les valeurs du RMSD sont légèrement plus élevées (valeurs entre 1-2 Å) comparé à celles des systèmes avec le LDAO (valeurs entre 0,5-1 Å). Cette observation est en bon accord avec la littérature où il a été montré que le bOG influençait sur la structure du complexe LH2. Nous retrouvons aussi des résultats similaires dans l'analyse des mouvements du squelette peptique dans laquelle les peptides α adjacents ont des mouvements opposés dans le cas du bOG et aussi en présence de POPC. Dans cette même étude, nous avons remarqué qu'au sein d'un même peptide α ou β, la partie N-terminale a un mouvement opposé à la partie C-terminale. Nous observons aussi ce mouvement opposé dans notre analyse en composantes principales (ACP) sur le squelette peptidique.

Enfin nous avons analysé des distances clefs impliquant les B800 d'une part, et les B850 d'autre part. Dans le cas des B800, en présence de LDAO, les résultats ont montré que chaque liaison hydrogène entre les arginines 20 et les B800 peut être perturbée très localement dans un peptide et presque jamais perturbée dans un autre. En revanche, dans la simulation avec le POPC et le bOG, la majorité des liaisons hydrogènes n'existent pas pendant la simulation. Les raisons de ces perturbations sont que les détergents/lipides viennent s'insérer autour des B800 qui sont peu confinés, ce qui empêche la formation de liaisons hydrogènes avec les arginines 20 et favorise les interactions avec les molécules de détergents ou d'eau.

Dans le cas des B850, les histidines 30/31 qui se situent dans un environnement totalement hydrophobe maintiennent les B850, empêchant ainsi de fortes perturbations entre les magnésiums des B850. En revanche, les résidus tryptophane 45 et tyrosine 44 qui sont un peu moins confinés comparé aux histidines 30/31 peuvent perdre leur liaison hydrogène avec le groupement OH des macrocycles des B850. Nous avons observé que des bOG ou POPC, d'autres résidus ou l'eau peuvent entrainer un éloignement entre les Tyr44 et les B850. Dans le cas des Trp45, le changement de conformation du tryptophane 45 explique les grandes distances observées quel que soit l'environnement. La corrélation de ces distances clefs nous a montré que des distances impliquant des atomes éloignés peuvent avoir des mouvements assez similaires ou opposé selon l'environnement micellaire ou membranaire.

Chapitre 5 : Calculs quantiques de l'absorbance

Au cours de nos simulations, nous avons pu observer la dynamique du complexe LH2. Les transitions d'absorption électronique des bactériochlorophylles-a dans ce complexe présentent un élargissement inhomogène important, dont l'origine a été attribuée, depuis les années 90 à son désordre dynamique. À partir de nos simulations, nous pouvons donc essayer de calculer comment les variations des interactions entre les différents composants du complexe observées, peuvent se traduire par des variations de l'absorption de ce complexe, et donc en déduire l'origine moléculaire de l'élargissement inhomogène des transitions d'absorption de ce complexe. Pour ce faire, des calculs quantiques des propriétés d'absorption des bactériochlorophylles-a dans les complexes LH2 ont été réalisés à partir des structures atomiques représentatives de nos simulations. Dans un premier temps, nous avons commencé avec le système composé du complexe LH2-LDAO (Amber99sb).

I. Modélisation du spectre d'absorbance

La modélisation du spectre d'absorbance a été réalisée en collaboration avec l'équipe du Dr. Olga Rancova, de l'Université de Vilnius en Lituanie, qui utilise un modèle électrostatique basé sur les interactions intermoléculaires afin de calculer les propriétés d'absorption et d'émission des bactériochlorophylles-a. En théorie, ces propriétés peuvent être décrites grâce au model excitonique de Frenkel (Frenkel 1931) par un Hamiltonien (Rancova *et al.* 2012) :

$$\hat{\mathrm{H}} = \sum_m E_m \hat{B}_m^\dagger \hat{B}_m + \sum_{m,n}^{m \neq n} V_{mn}\, \hat{B}_m^\dagger \hat{B}_n \qquad \text{(I.1)}$$

où E_m est l'énergie d'excitation du pigment m (ou énergie de site), $\hat{B}_m^\dagger$ représente l'opérateur de création d'une excitation sur le pigment m, $\hat{B}_m$ est l'opérateur conjugué d'annihilation (réaction entre une particule et son antiparticule), et V_{mn} représente le couplage entre les pigments m et n. Le spectre d'absorption A est obtenu après diagonalisation de cet Hamiltonien :

$$A(\omega) = \omega \sum_{i=1}^{27} \mu_i^2 \cdot \frac{\gamma}{(\omega-\omega_i)^2+\gamma^2} \qquad \text{(I.2)}$$

avec ω_i, les valeurs propres de l'Hamiltonien, μ_i représentent les moments de transition des dipôles et γ est la largeur spectrale Lorentzienne choisie en cm^{-1}.

Dans nos simulations, nous avons utilisé la forme protonée des chaînes latérales des arginines car dans les conditions physiologiques (Garrett *et al.* 2000), c'est la forme qui

prédomine (pK_A = 12,5). Cependant, au sein des protéines, les pK de protonation/déprotonation des acides aminés peuvent être profondément influencés par l'environnement. Le calcul des transitions d'absorption en présence d'une arginine protonée en position 20 sur le polypeptide β, à proximité des bactériochlorophylles-a responsables de l'absorption à 800 nm des LH2, induit dans le calcul, un déplacement des transitions d'absorption vers le rouge de ces molécules de près de 700 cm^{-1}, et ne permet en aucun cas d'approcher pour ces transitions, une position proche de celle observée expérimentalement. De ce fait, il est nécessaire d'imposer à la chaîne latérale des arginines 20, un état déprotoné pour obtenir par le calcul, une position de la transition des B800 plus conforme à celle observée expérimentalement.

Pour être en bon accord avec leur modèle, nous avons réalisé une simulation avec les paramètres compatibles pour AMBER développés par Wu *et al.* (Wu *et al.* 2013) afin de rendre les chaînes latérales des arginines neutres.

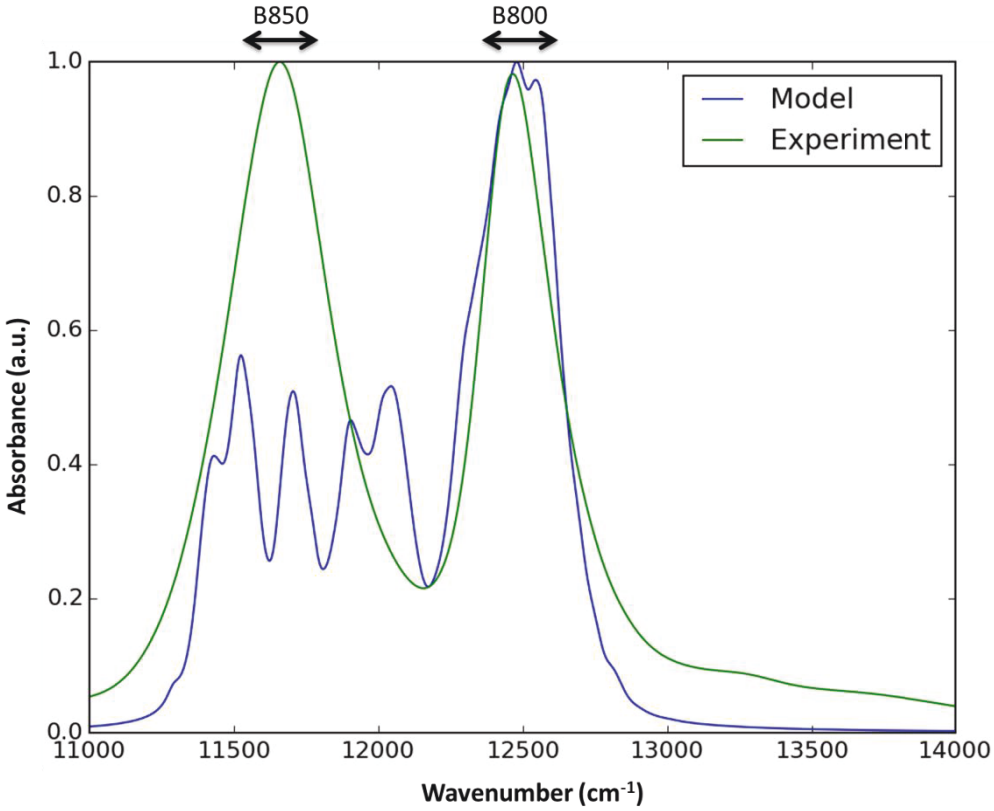


Figure 1 : Absorbance des BChls-a dans notre simulation (bleu) comparée avec les valeurs expérimentales (vert).

Nous sommes partis de la dernière configuration de notre simulation LH2-LDAO en modifiant l'arginine positive en arginine neutre, puis nous avons simulé 70 ns avec les mêmes paramètres de simulation décrits dans le chapitre 4. De cette trajectoire, nous avons extrait 140 configurations (sans les molécules d'eau) de 0 à 70 ns toutes les 500 ps au format PQR (fichier PDB avec nos valeurs de charge et de rayon pour chaque atome) qui ont servis dans le calcul de l'absorbance (Figure 1). Cette figure a été générée en calculant les positions et les intensités des composantes de l'absorption de chaque configuration, en leur imposant une largeur spectrale Lorentzienne de 30 cm^{-1} pour tenir compte de l'élargissement homogène de ces composantes, et en faisant la somme de l'ensemble des spectres d'absorption obtenus sur l'ensemble des configurations.

- ***Absorbance des B800***

Nous remarquons d'une part que, cette démarche aboutit à une modélisation raisonnable de l'absorbance des B800, tant du point de vue de la position comme du point de vue de la largeur (Figure 1). Ce résultat suggère que les fluctuations protéiques autour du site des B800, même échantillonnées sur une dynamique relativement courte, sont suffisantes pour induire l'élargissement inhomogène de cette transition. Nous avons calculé la probabilité de présence de chaque liaison hydrogène entre les arginines 20 et les B800, dans les simulations utilisant des arginines neutres ou chargées (Tableau 1). Nous remarquons que les liaisons hydrogènes entre les B800 et les arginines neutres sont environ deux fois moins présentes que dans le cas des arginines positives.

Tableau 1 : Pourcentage de présence des liaisons hydrogènes entre : les 9 Arg20(NE-HE)-B800(OBB)/Arg20(NH2-HH21) et les B800(OBB), et les 9 Tyr44(OH-HH)-βB850(OBB) et les 9 Trp45(NE1-HE1)-αB850(OBB) au cours des simulation avec les arginines positives ou neutres.

Peptide	LH2-LDAO (Amber99sb, Arg positive)		LH2-LDAO (Amber99sb, Arg neutre)		LH2-LDAO (Amber99sb, Arg positive)		LH2-LDAO (Amber99sb, Arg neutre)	
	NE	NH2	NE	NH2	Y44	W45	Y44	W45
# 1	15	17	0	0	99,7	90,3	99,6	87,1
# 2	59	23	26	10	98,9	87,7	99,2	76,8
# 3	81	43	20	4	99,5	93,9	99,6	88,9
# 4	3	2,3	4	7	99,6	94	99,8	91,3
# 5	43	16	16	13	99,5	16,3	99,2	85,2
# 6	77	23	16	16	99,5	94,7	99,5	93,1
# 7	0,3	0,3	34	17	99,7	87,8	99,7	93,6
# 8	22	7	13	7	99,6	44	99,8	78,3
# 9	25	8	1	18	99,6	93,8	99,5	95,6

En comparant les spectres calculés à partir des différentes dynamiques et/ou en prenant en compte différents paramètres des structures obtenues durant ces dynamiques, on peut noter d'une part que, si la présence d'arginines positives a un effet dramatique sur la position de la transition des B800, elle n'a qu'un effet très limité sur l'élargissement inhomogène de cette transition. Comme la statistique de présence des liaisons hydrogènes est différente selon que les dynamiques sont effectuées en présence d'arginine positives ou neutres, nous pouvons en déduire que la dynamique des liaisons hydrogènes des B800 a un effet marginal sur l'élargissement inhomogène de cette transition. De même, on peut effectuer les mêmes calculs en prenant ou non en compte le solvant (solvant impliquant les molécules d'eau mais aussi les molécules de détergents au voisinage des bactériochlorophylles-a). Il apparait que la prise en compte du solvant induit un déplacement limité de la transition des B800 (quelques dizaines de cm^{-1}) mais n'a quasiment aucun effet sur la largeur de cette transition. Nous pouvons donc déduire de ces calculs que, pour la transition des B800, la position est déterminée par l'électrostatique du site d'ancrage des bactériochlorophylles-a et la présence des liaisons hydrogènes, tandis que l'élargissement inhomogène de cette transition est essentiellement lié à la dynamique générale de la poche.

- ***Absorbance des B850***

Pour ce qui concerne la transition des B850, la situation est bien plus compliquée, car les positions et largeurs de cette transition dépendent d'une part, des énergies de site des 18 bactériochlorophylles-a, et d'autre part de l'ensemble des couplages excitoniques entre ces molécules. Les calculs entrepris à partir de nos simulations prédisent une largeur inhomogène un peu supérieure à celle observée expérimentalement, et la répartition des transitions électroniques montre un manque important des contributions autour de 11700 cm^{-1}. Ce résultat est un peu étonnant pour un échantillonnage obtenu à partir de trajectoires limitées, et traduit plus probablement l'imprécision des calculs pour ce qui concerne le calcul de la position de transitions provenant de pigments en forte interaction excitonique. Avant de lancer des trajectoires plus longues, nous nous proposons, avec nos collègues lituaniens, d'analyser l'origine des variations observées. On peut noter toutefois, comme dans le cas de la transition à 800 nm, que la prise en compte des molécules de solvants n'ont qu'un effet mineur sur la largeur inhomogène des B850. Par ailleurs, le calcul des énergies de site des différentes molécules participant à la transition à 850 nm, prédit une inhomogénéité de celles-ci voisine de celle des B800 (Figure 2). La largeur inhomogène excessive que nous observons dans la

prédiction des B850 provient donc de la manière dont les interactions entre bactériochlorophylles sont estimées.

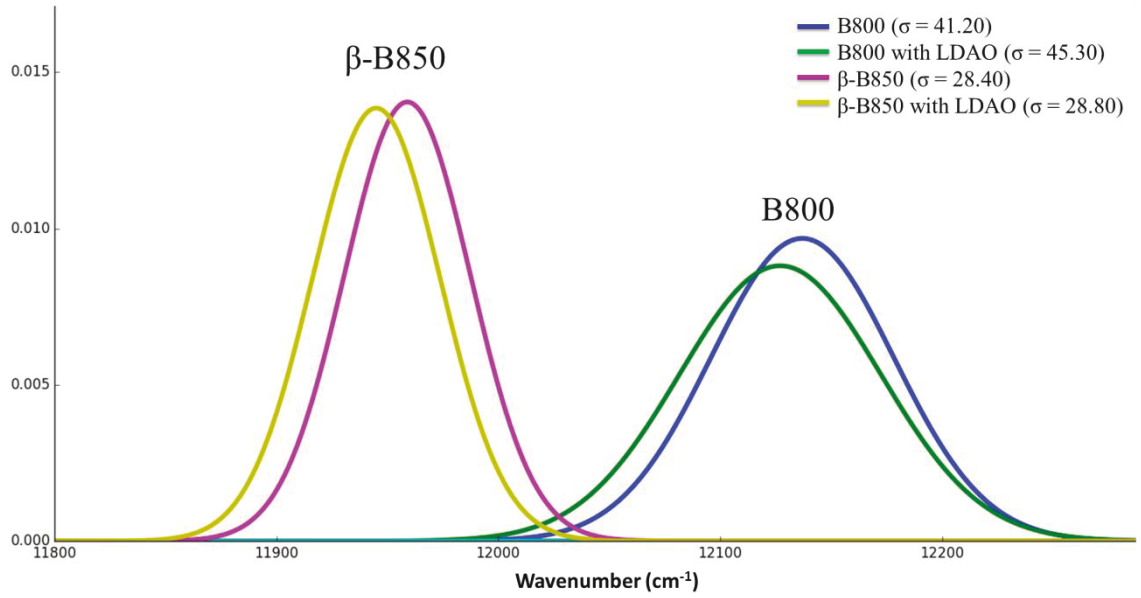


Figure 2 : Distribution Gaussienne des énergies de site prédites pour les bactériochlorophylles-a absorbant à 800 nm, comparées à celles prédites pour les bactériochlorophylles-a absorbant 850 nm liées au polypeptide beta des LH2. Les valeurs σ représentent la déviation standard de chaque Gaussienne.

II. Conclusion de ce chapitre

Ces premiers résultats obtenus dans la modélisation quantique de l'absorbance des bactériochlorophylles-a, en utilisant comme base de calcul à la fois la structure cristallographique des LH2 et les variations de cette structure lors de trajectoires obtenues à partir de dynamique moléculaire, reproduit relativement bien l'absorbance des B800 mesurée expérimentalement. Dans le cas des B850, le spectre d'absorbance prédit est très inhomogène et traduit clairement que les interactions entre pigments ne sont pas correctement prises en compte dans le calcul. Ces résultats sont assez encourageants et nous montre que les 140 configurations de notre simulation, bien que sûrement insuffisantes pour décrire correctement le spectre d'absorption des LH2, permettent déjà une prise en compte assez significative du désordre dynamique des LH2.

Conclusions et perspectives

Conclusions

Dans ce travail de thèse, nous avons mêlé les approches théoriques de simulation de dynamiques moléculaires et expérimentales de SAXS, pour étudier les propriétés structurales du complexe LH2 de la bactérie pourpre *Rps. acidophila* strain 10050 dans différents environnements biomimétiques. Notre laboratoire a montré, dans des conditions ambiantes et en fonction du temps d'exposition, une variation de l'intensité d'absorption de la fluorescence que l'on pouvait relier à un « désordre dynamique » et qui, à l'échelle de la molécule, correspondait à des fluctuations locales des interactions entre les pigments. Le but de ce travail de thèse a été de mieux comprendre l'origine de ce désordre à l'échelle atomique.

Dans un premier temps, pour simuler le complexe LH2 dans les différents environnements biomimétiques, nous avons développé les paramètres (charges partielles atomiques RESP) pour le LDAO et le bOG compatibles avec le champ de force AMBER pour les protéines. Ces paramètres ont été testés en effectuant des simulations de micelles pures et en présence d'un modèle de fragment de la protéine kinase humaine PI4KA. Le bon accord de ces simulations avec les données expérimentales et de simulations disponibles nous a permis de valider notre approche. Dans le cas de la membrane de POPC, nous avons utilisé les paramètres disponibles pour AMBER que nous avons convertis au format utilisé par GROMACS. Cette conversion a été validée grâce à une simulation de membrane pure de POPC en présence de 0,15 M de KCl. Nos résultats ont bien reproduit les propriétés structurales du POPC, disponibles dans la littérature.

Dans un second temps, nous avons paramétré les différents composants du complexe LH2 en dérivant des charges partielles atomiques RESP pour le résidu CXM et la rhodopine glucoside, compatibles pour le champ de force Amber99sb et Glycam, respectivement. Pour les BChls-a, nous avons utilisé les paramètres développés par Ceccarelli *et al.* (Ceccarelli *et al.* 2003), compatibles pour Amber94. Le nombre de monomères de détergent autour du complexe LH2 a été choisi à partir des données expérimentales et affiné par des simulations moléculaires auto-assemblées (LDAO) ou estimé directement à partir de modèle géométrique (tore) tiré de la littérature (bOG). L'incertitude sur les valeurs du nombre d'agrégation du LDAO et du bOG autour du complexe LH2 étant assez importante, nous avons réalisé des expériences de SAXS du complexe LH2-LDAO et LH2-bOG, et calculé le nombre théorique de détergents autour du complexe LH2. Dans le cas du bOG, l'estimation du nombre de

monomères de bOG issue des expériences de SAXS (224) est très proche de la valeur utilisée dans notre simulation (225). En revanche, dans le cas du LDAO, les contrastes de densité électronique du détergent proches avec ceux du solvant nous ont limités dans l'exploitation de la courbe expérimentale de SAXS, et la valeur obtenue (258) est relativement éloignée de la valeur utilisée dans nos simulations (290).

Dans un troisième temps, nous avons spécifiquement étudié l'influence de la nature du détergent et de l'environnement (micellaire, membrane). Les calculs des propriétés structurales des différents complexe LH2-détergents/lipides ont montré que :

- les dimensions des anneaux peptidiques internes et externes du complexe LH2 ne varient pas et restent stables au cours des différentes simulations. L'analyse des paramètres de forme des complexes LH2-détergents ont montré une forme ellipsoïde oblate avec une légère différence de forme, qui est un peu plus excentrique en présence de LDAO, comparé au bOG.
- localement, nous avons montré que le bOG influençait un peu plus la structure des peptides comparé au LDAO.
- nous avons noté que, quel que soit l'environnement, au sein d'un même peptide α ou β, la partie N-terminale a un mouvement opposé à la partie C-terminale.
- le macrocycle des B800 explore plus de conformations comparé à ceux des B850. Les raisons de ces variations peuvent s'expliquer par le fait que les détergents/lipides viennent s'insérer autour des B800 qui sont peu confinés, comparé aux B850.

Tous ces résultats nous suggèrent que, même si la structure globale du complexe LH2 reste stable tout au long des simulations, ces environnements biomimétiques influent localement sur la dynamique interne des complexes LH2 et notamment au niveau de l'orientation des macrocycles des BChls-a.

Enfin, en ce qui concerne l'analyse du spectre d'absorption du complexe LH2-LDAO. Les premiers résultats obtenus dans la modélisation quantique de l'absorbance des bactériochlorophylles-a reproduit bien l'absorbance des B800 mesurée expérimentalement. Dans le cas des B850, le spectre d'absorbance prédit est inhomogène et traduit clairement que les interactions entre pigments ne sont pas correctement prises en compte dans le calcul. Ces résultats sont assez encourageants et nous montre que les 140 configurations de notre simulation sont sûrement insuffisantes pour décrire correctement le spectre d'absorption des B850.

Perspectives

Les perspectives de ce travail de thèse sont nombreuses du point de vue expérimental et théorique. Par exemple, les expériences de SAXS que nous avons réalisées avec les complexes LH2 de *Rps. acidophila* strain 10050 entourés de détergents (LDAO ou bOG) représentent, à ce jour, les premières mesures effectuées sur ce complexe chez *Rps. acidophila*. Nous avons noté cependant les limites de cette méthode dans le cas des monomères de LDAO qui ont un contraste électronique proche de l'eau. Il serait donc intéressant de compléter ces résultats par des mesures de diffusion des neutrons aux petits angles (SANS) qui ont une sensibilité aux éléments légers et qui offrent la possibilité de faire du marquage isotopique.

Enfin, les résultats préliminaires que nous avons obtenus avec le calcul quantique de l'absorbance sur quelques conformations de notre complexe LH2-LDAO sont assez encourageants et montrent cependant qu'il faudrait probablement d'avantage de conformations pour améliorer le spectre d'absorbance. Ces conformations devront être obtenues soit en diminuant la fréquence de collecte des configurations, soit en faisant des simulations plus longues pour avoir plus d'échantillonnage. Nous avons aussi noté que dans nos simulations, l'environnement (détergents ou lipides) influençait sur la dynamique interne du complexe LH2, c'est pourquoi il serait très intéressant de comparer les spectres d'absorbance obtenus pour les complexes LH2 dans les autres environnements bOG et POPC.

Les annexes

Annexe 1 : Test et validation des paramètres pour le β-Octyle glucoside

Dans cette annexe, nous présenterons les résultats des simulations moléculaires qui nous ont permis de tester et valider notre modèle de potentiel pour le β-Octyle-glucoside (bOG, Figure A1-1) utilisé dans les simulations du complexe LH2. Pour ce faire, nous avons effectué deux simulations de micelles directes pures dans les conditions ambiantes avec différentes conditions initiales.

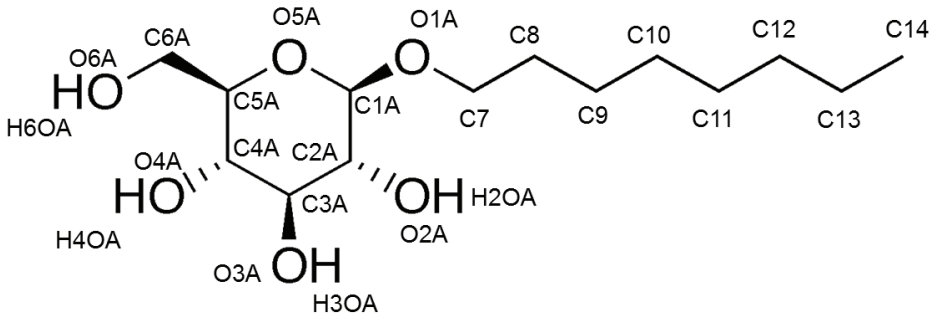


Figure A1-1 : Représentation moléculaire du bOG avec les différents noms des atomes (les hydrogènes apolaires ne sont pas montrés sur la figure).

I. Description du modèle pour le bOG

Pour modéliser la tête polaire b-D-glucopyranose et la chaîne alkyle de la molécule de bOG, nous avons utilisé le champ de force basé sur la version f de GLYCAM06 (Kirschner *et al.* 2008)(Tessier *et al.* 2008). Ce potentiel est compatible avec tous les champs de force AMBER. Les charges partielles atomiques du bOG sont issues d'un précédant travail (Abel *et al.* 2011) où les auteurs ont dérivé les charges RESP pour une large famille de glycolipides avec différentes longueurs de chaîne d'alcane et d'unités α et β glucose (projet F72 dans la base de données RED.DB http://upjv.q4md-forcefieldtools.org/REDDB/projects/F-72/)(Abel *et al.* 2011). Les charges et les différents atomes types de la molécule de bOG sont donnés dans le Tableau A1-1.

Tableau A1-1 : Charges partielles atomiques RESP et types d'atomes utilisés pour modéliser la molécule de bOG.

Atome	Atome Types	Charge
C1A	cg	0,3005
H1A	h2	0,0000
O1A	os	-0,4456
C2A	cg	0,4091
H2A	h1	0,0000
O2A	oh	-0,7057
H2OA	ho	0,4094
C3A	cg	0,2633
H3A	h1	0,0000
O3A	oh	-0,7099
H3OA	ho	0,4342
C4A	cg	0,3272
H4A	h1	0,0000
O4A	oh	-0,6674
H4OA	ho	0,4056
C5A	cg	0,0974
H5A	h1	0,0000
O5A	os	-0,4067
C6A	cg	0,2389
H61, H61	hc	0,0000
O6A	os	-0,6200
H6OA	ho	0,4079
C7	cg	0,1843
H71, H72	hc	0,0000
C8	cg	0,1245
H81, H82	hc	0,0000
C9	cg	-0,0533
H91, H92	hc	0,0000
C10	cg	-0,0091
H101, H102	hc	0,0000
C11	cg	0,0097
H111, H112	hc	0,0000
C12	cg	0,0169

H121, H122	hc	0,0000
C13	cg	0,0058
H131, H132	hc	0,0000
C14	cg	-0,0170
H141, H142, H143	hc	0,0000

1. Protocole et paramètres des simulations

1.1. Choix du nombre de bOG par micelle

Pour évaluer notre modèle de bOG, nous avons donc effectué deux simulations d'une micelle pure avec un nombre d'agrégation (n_{ag}) de 75 dans une boîte contenant 32 000 molécules d'eau TIP3P (Jorgensen *et al*. 1998). Cela correspond à une concentration en bOG de l'ordre de 130 mM, soit 5,2 fois la concentration micellaire critique de 25 mM déterminée expérimentalement par Mesa *et al.* (Mesa *et al.* 1993) à 25°C. Le nombre d'agrégation (n_{agg}) choisi a pour objectif de rendre possible la comparaison de nos résultats avec ceux de Bogusz *et al.* (Bogusz *et al.* 2000)(Bogusz *et al.* 2001) où le bOG a été modélisé avec la version de CHARMM27 pour les lipides, et Hal *et al.* (Ha *et al.* 1988) pour la partie carbohydrate. En ce qui concerne le nombre d'agrégation de 75, celui-ci est dans la fourchette des valeurs expérimentales comprises entre 27 – 110 par extension de fluorescence (Aoudia *et al*. 1998), diffusion de la lumière et SAXS (Lipfert *et al.* 2007)(Lorber *et al.* 1990). Cette large plage de valeurs est liée au fait que n_{ag} varie beaucoup en fonction de la méthode expérimentale et des conditions expérimentales (concentration et température).

1.2. Construction des systèmes

De façon similaire à la validation du potentiel pour le LDAO décrit dans l'article 1 (Karakas *et al.* 2016), nous avons effectué des dynamiques moléculaires avec deux systèmes de tailles identiques construits en utilisant une approche pré-assemblée et auto-assemblée.

- Le système pré-assemblé (PA) a été construit en introduisant dans la boîte de simulation une micelle sphérique constituée de 75 bOG, construit avec le programme PACKMOL (Martínez *et al.* 2009). Les paramètres géométriques utilisés avec PACKMOL sont donnés sur la Figure A1-2.
- Le système auto-assemblé (SA pour *self-assembled* en anglais) a été simulé en introduisant dans la boîte de simulation 75 bOG avec des orientations différentes.

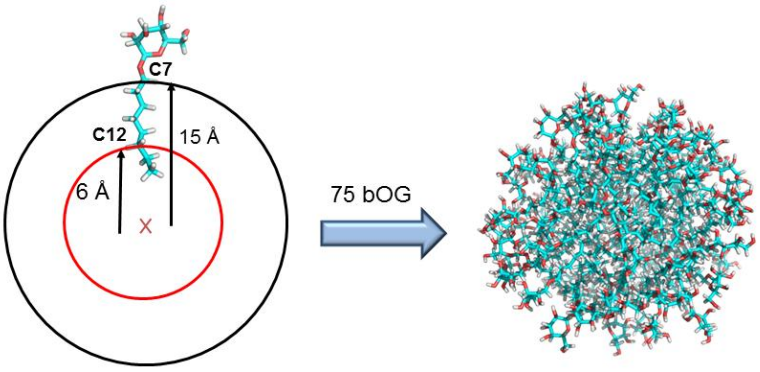


Figure A1-2 : Modèle géométrique utilisé pour construire la micelle pré-assemblée avec le programme PACKMOL. Les rayons des sphères 6 et 15 Å ont été choisis afin d'obtenir un *packing* optimal pour les 75 bOG et de minimiser les *clashes* stériques pour les atomes situés à la fin de la chaîne n-octyle et inter-chaînes. La valeur de 15 Å est légèrement plus élevée que celle déduite à partir de la relation de Tanford (Tanford 1972), 11,6 Å avec la relation l_c=1,5+1,265 n_c et n_c=8).

1.3. Paramètres des simulations

Ces deux systèmes ont été simulés dans les conditions ambiantes (T = 298K et P = 1,015 bar) durant ~140 ns chacun avec la pression et la température contrôlées avec les algorithmes de Nosé-Hoover (Nosé 1984)(Hoover 1985) (τ_T = 0,3 ps) et Parrinello-Rahman (Parrinello *et al.* 1981)(Rahman *et al.* 1971) (τ_T = 1,0 ps et une compressibilité isotherme $4{,}5.10^{-5}$ bar^{-1}). Les systèmes ont été minimisés avec l'algorithme de *steepest-descent* et des périodes d'équilibration de 1 ns dans les ensembles NVT et NPT. La méthode d'Ewald (SPME) (Essmann *et al.* 1995) a été utilisée pour évaluer les forces électrostatiques avec un *cut-off* de 10 Å pour les interactions courtes et longues distances. Les configurations atomiques ont été collectées toutes les 2 ps pour analyse.

II. Résultats

1. Agrégation des monomères de bOG en fonction du temps

Comme pour les micelles de LDAO, nous avons examiné l'agrégation des monomères de bOG en fonction du temps (Figure A1-3) en examinant la variation du nombre de *clusters* en fonction du temps. De façon similaires aux micelles de LDAO, nous avons considéré l'existence d'un *cluster* si la distance entre les atomes de carbones de la chaîne n-octyle de deux molécules de bOG est inférieure à 4,1 Å (Sanders *et al.* 2012)(Abel *et al.* 2012). On peut constater sur cette Figure A1-3 que, comme pour les micelles de LDAO, le processus d'agrégation s'effectue en plusieurs échelles de temps (courte et lente) que l'on peut

caractériser en utilisant l'équation (A1.II.1) basée sur un modèle cinétique simplifié d'agrégation/dissociation (Nyrkova *et al*. 2005)(Cueille *et al*. 1997) .

$$N(t) - N(\infty) = c_1 e^{(-t/\tau 1)} + c_2 e^{(-t/\tau 2)} \quad \text{avec} \quad N(\infty) = 1 \qquad \text{(A1.II.1)}$$

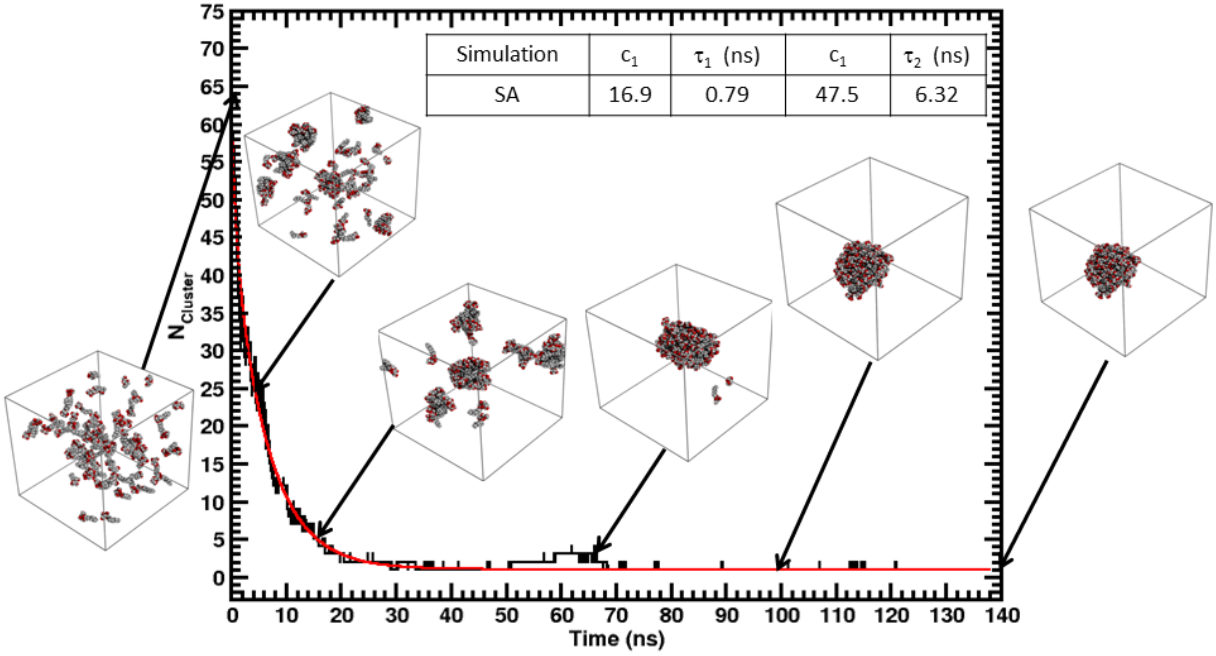


Figure A1-3 : Assemblage des monomères de bOG en fonction du temps. La courbe rouge correspond au *fit* de la double exponentielle $N_{cluster}(t)$-$N(\infty)$ qui donne deux constantes de temps τ_1 et τ_2 correspondant à la formation rapide de petits *cluster* de bOG et leur coalescence plus lente.

Comme pour d'autres systèmes micellaires (par exemple, avec le LDAO (Karakas *et al.* 2016), DDM (Abel *et al.* 2014), n-dodecyle phosphocholine (Abel *et al.* 2012)), la cinétique d'agrégation peut être caractérisée selon deux échelles de temps: rapide (τ_1 < 1ns) et lente (τ_2 > 1ns). Dans le cas de la simulation du bOG, les valeurs de τ_1 et τ_2, valent 0,79 et τ_2 = 6,32 ns, respectivement et comparables à celles obtenues pour le LDAO. Les représentations finales des micelles obtenues à partir des simulations pré-assemblées et auto-assemblées sont données sur la Figure A1-4.

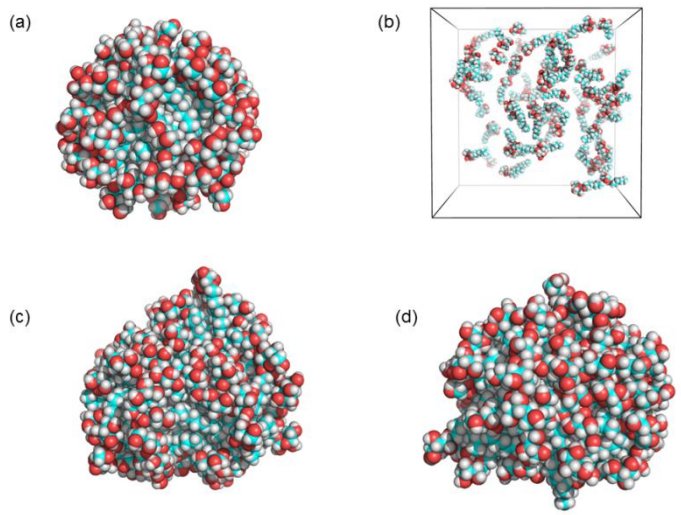


Figure A1-4 : Représentations atomiques des micelles au début et à la fin des simulations pré-assemblées (a, c) et auto-assemblées (b, d), respectivement.

2. Forme des micelles

La forme des micelles a été caractérisée en calculant les moments d'inerties et la valeur des semi-axes *a*, *b* et *c*, caractérisant l'ellipsoïde qui modélise la micelle. L'évolution en fonction du temps des valeurs des semi-axes *a*, *b* et *c* pour les deux simulations est montrée sur la Figure A1-5.

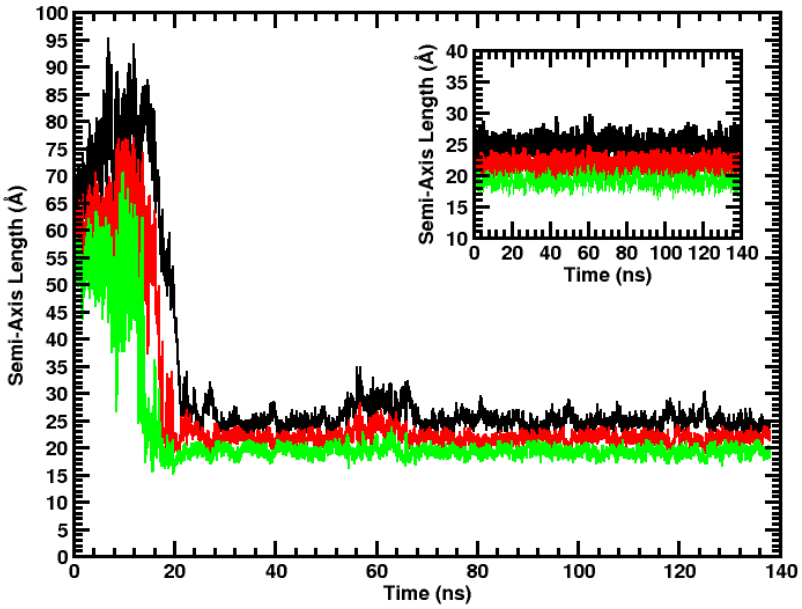


Figure A1-5 : Évolution en fonction du temps de la longueur des semi-axes *a* (noir), *b* (rouge), *c* (vert) de l'ellipsoïde représentant des micelles auto-assemblées (grande figure) et pré-assemblées (figure encadrée).

La Figure A1-5 montre qu'après une décroissance de la valeur des semi-axes *a, b* et *c* correspondant à la période d'agrégation des monomères de bOG, les valeurs des semi-axes se stabilisent (après 20 ns de simulation) et restent constantes indiquant que la forme des micelles ne change pas significativement jusqu'à la fin des simulations. La « bosse » observée, entre 56 – 68 ns, correspond à la période où un monomère de bOG s'échappe de la micelle. Les caractéristiques des formes reportées dans le Tableau A1-2 indiquent que les micelles sont ellipsoïdales (plus précisément de forme *prolate-like*) et leurs formes ne dépendent pas significativement du protocole de simulation utilisé. La valeur de l'ellipticité ($e = \sqrt{1 - \frac{c^2}{a^2}}$) égale à 0,6 est similaire à celles obtenues par Bogusz *et al.* (Bogusz *et al.* 2000) et Konidala *et al.* (Konidala *et al.* 2006) pour des micelles avec 75 et 92 monomères de bOG, respectivement.

3. Dimension des micelles

Les dimensions des micelles pré-assemblées et auto-assemblées ont été évaluées en calculant leur rayon de giration (R_g) et le rayon de la sphère équivalente de densité uniforme $R_s = \sqrt{\frac{5}{3}} R_g$. Sur la Figure A1-6, nous avons représenté l'évolution en fonction du temps de R_g des deux micelles.

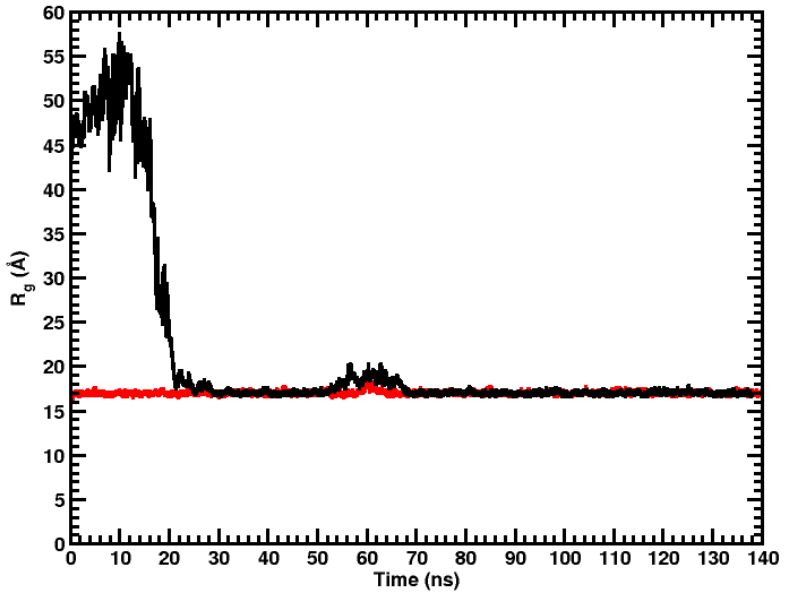


Figure A1-6 : Dimension du rayon de giration des bOG en fonction du temps dans le cas des simulations auto-assemblées (noir) et pré-assemblées (rouge).

La diminution de la valeur du R_g dans le cas de la simulation auto-assemblée correspond à la formation de *clusters* de bOG. De plus, il faut approximativement 20 ns pour former une micelle unique à partir de tous les monomères de bOG. La petite « hausse » du R_g entre 56 et 68 ns, correspond à la période où on observe l'échappement de deux monomères de bOG dans la phase aqueuse (Figure A1-3).

Tableau A1-2 : Dimension avec les semi-axes et l'ellipticité moyenne des micelles. Une valeur de 0 et 1 de l'ellipticité correspond à une sphère et à un ellipsoïde aplati, respectivement. [a](Bogusz *et al.* 2000) [b](Konidala *et al.* 2006) [c](Lorber *et al.* 1990).

DM	R_g (Å)	R_s (Å)	*a* (Å)	*b* (Å)	*c* (Å)	*a/c*	*e*
PA	17,0 ± 0,1	21,9	24,8 ± 1,2	21,7 ± 0,8	19,0 ± 0,7	1,30 ± 0,5	0,6
SA	17,0 ± 0,1	21,9	24,8 ± 1,2	21,6 ± 0,8	19,2 ± 0,7	1,29 ± 0,5	0,6
Ref.[a, b]	17,6	22,7	1,3	~1,2	~1,1	-	~0,6 – 0,62
Exp. [c]	-	23,3 ± 3	-	-	-	-	-

En premier lieu, la Figure A1-6 indique que la taille des micelles ne dépend pas du protocole de simulation utilisé. Ensuite, la valeur de R_g se stabilise autour de 17,0 Å qui est proche de celle obtenue pour la micelle de bOG dans la simulation pré-assemblée. Cette valeur est en bon accord avec la valeur obtenue par Bogusz *et al.* (Bogusz *et al.* 2000) (17,6 Å) ou expérimentalement, dans le cas du R_s par De Lucas *et al.* (23,3 Å) en utilisant la spectroscopie de fluorescence, la mesure des variations de tension de surface ou la chromatographie d'exclusion de taille.

4. Structure interne des micelles

La structure interne des deux micelles ont été comparée en calculant le profil de densité radiale pour la micelle entière, la chaîne alkyle, la tête polaire du détergent et l'eau (Figure A1-7).

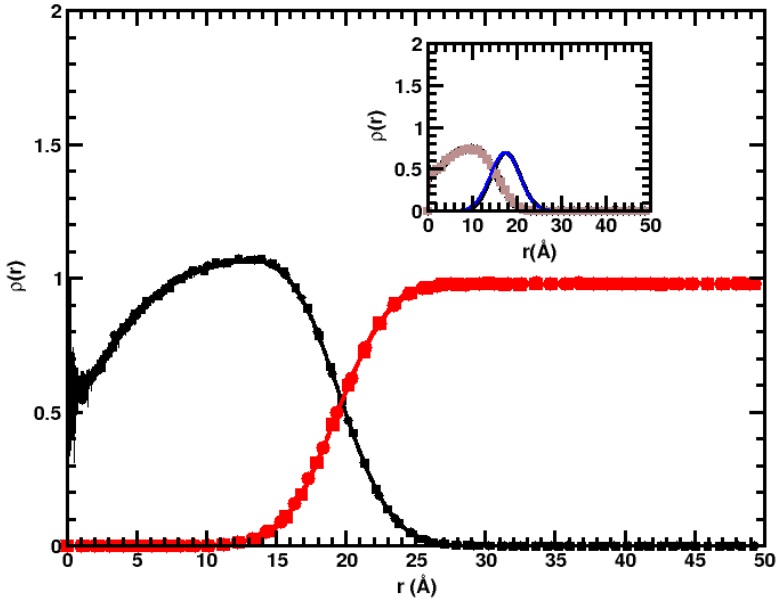


Figure A1-7 : Profil de densité radiale des micelles obtenues dans les simulations auto-assemblées (cercle) et pré-assemblées (carré). Avec la micelle entière (noir), la chaîne alkyle (marron), la tête beta-glucoside (violet) et l'eau (en rouge).

Comme on peut le constater sur la Figure A1-7, la structure interne de la micelle ne change pas avec la méthode de construction des micelles. Pour les deux micelles, le groupement sucre (en violet, sur la courbe encadrée) se trouve à environ 19 Å du centre de la masse de la micelle. L'eau pénètre significativement dans la couche polaire de la micelle délimitée par le groupement glucopyranose. L'intersection entre les profils de densité de l'eau (en rouge) et de la micelle (en noir) est ~20 Å, valeur très proche du rayon effectif, R_s, de la micelle calculée à partir du rayon de giration.

5. Surface accessibles au solvant

La surface partagée entre l'eau et, la tête polaire et la chaîne alkyle du détergent a été calculée avec la surface de Voronoï et l'outil *trjVoronoi* (Abel *et al.* 2012) développé au laboratoire (cf. section III.6, chapitre 2). Le Tableau A1-3 présente les valeurs moyennes de SAS pour la tête polaire, SAS_{GLC}, la chaîne aliphatique octyle (SAS_{OCT}).

Tableau A1-3 : Accessibilité moyenne au solvant des deux micelles pour la tête polaire SAS_{GLC} et la chaîne octyle SAS_{OCT}. La fraction de surface en contact entre le solvant et les chaînes aliphatiques sont données dans la 4ème colonne. La valeur A_{Rg} est calculée à partir du volume d'une sphère avec le rayon de giration divisé par le nombre de monomères de bOG. [a](Bogusz *et al.* 2000) [b](Konidala *et al.* 2006) [c](Nilsson *et al.* 1996)

Système	SAS_{GLC} (Å^2)	SAS_{OCT} (Å^2)	f_{tail} (%)	A_{Rg} (Å^2)
PF	116,2 ± 3,6	28,5 ± 2,2	19,6 ± 0,1	48,5
SA	114,4 ± 3,4	26,9 ± 2,0	19,0 ± 0,1	48,4
Ref.[a, b]	130 - 139	34 - 52	20 – 27,5	55,0
Exp.[c]	-	-	-	44

À la lecture du tableau, on constate que 1) les SAS sont similaires entre les deux micelles construites avec des approches différentes, 2) les SAS de la tête polaire et de la queue hydrophobe du détergent sont environ 20 % plus faibles que celles calculées par Bogusz *et al.* (Bogusz *et al.* 2000) et Konidala *et al.* (Konidala *et al.* 2006) en utilisant la surface de Connolly. Notons que nous avons aussi effectué les calculs de SAS en utilisant l'approche de Connolly de l'outil *g_sas* de GROMACS avec une sphère de rayon 1,4 Å, et nous n'observons pas de différence significative avec les valeurs de SAS obtenues avec les surfaces de Voronoï. En ce qui concerne la fraction du détergent en contact avec l'eau. Nos calculs indiquent aussi qu'environ 19 % de la chaîne alkyle est en contact avec l'eau, de façon similaire aux résultats obtenues par Bogusz *et al.* (Bogusz *et al.* 2000) et Konidala *et al.* (Konidala *et al.* 2006). Enfin en ce qui concerne la surface par tête polaire, nos valeurs autour de 48 Å^2, sont en bon accord avec les données de Nilsson *et al.* calculées à partir des expériences de SAXS et de RMN.

6. Hydrations des sites accepteurs du bOG

Dans les sections précédentes, nous avons montré que la tête glucopyranose du bOG était largement hydratée par l'eau. Pour caractériser ce phénomène plus en détail, nous avons calculés les fonctions de densités radiales g(r) entre l'eau et les oxygènes des hydroxyles (O2A,…, O6A) et éther (O1A) de la tête polaire du bOG. Les g(r) obtenues pour les systèmes sont représentées sur la Figure A1-8. À l'exception de l'oxygène éther (O1A) où on observe une hydration plus faible (par rapport aux oxygènes hydroxyles), les nombres d'hydratation ne changent pas significativement entre les deux systèmes.

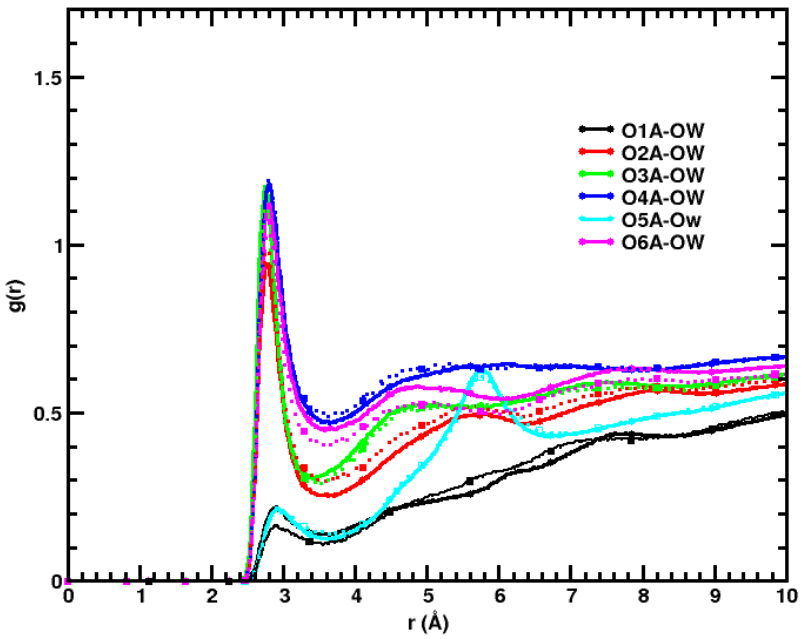


Figure A1-8 : Fonction de densité radiale entre l'eau et les différents sites accepteurs de la tête polaire du bOG (voir sur la Figure A1-1, la localisation des atomes d'oxygène sur le glucose). Les g(r) obtenues à partir des simulations pré-assemblées et auto-assemblées sont représentées avec des lignes continues (●---●) et pointillées (■ ... ■).

Dans le Tableau A1-4, nous avons reporté le nombre de molécule d'eau dans le premier voisinage des atomes d'oxygène en intégrant les g(r), jusqu'à leur premier minimum (~3,5 Å) ou 3,0 Å comme Konidala *et al.* (Konidala *et al.* 2006).

Tableau A1-4 : Nombre moyen de molécules d'eau dans le premier voisinage des atomes d'oxygène hydroxyle et éther de la tête polaire. Les valeurs entre parenthèse ont été obtenues en intégrant les g(r) jusqu'à 3,0 Å. [a](Konidala *et al.* 2006) [b](He *et al.* 2000) [c](Nilsson *et al.* 1996) [d](Mesa *et al.* 1993)

Système	$O1_A$-O_W	O_2A-O_W	O_3A- O_W	O_4A- O_W	O_5A-O_W	O_6A-O_W	bOG-O_W
PF	0,4 (0,2)	1,7 (1,0)	1,8 (1,1)	2,5 (1,2)	0,5 (0,2)	2,1 (1,2)	9,0 ± 0,3
SA	0,5 (0,2)	1,6 (0,9)	1,7 (1,2)	2,5 (1,2)	0,5 (0,2)	2,3 (1,2)	9,9 ± 0,3
Ref.[a]	0,3	0,9	1,3	1,2	0,2	1,2	7,0
Exp.	-	-	-	-	-	-	4[b], 6[c] , 12[d]

Enfin, le calcul du nombre de molécules d'eau unique au voisinage du groupement glucose montre des valeurs relativement similaires entre les deux systèmes et dans la plage de valeur estimée expérimentalement ou dans les simulations précédentes du bOG dans les micelles.

7. Conformation des molécules de bOG

Nous avons aussi examiné la conformation des molécules de bOG entre les deux simulations en calculant les pourcentages de conformation *trans* des angles dièdres CCCC de la chaîne octyle, quelques dièdres représentatifs du groupement glucopyranose (ϖ_1= O6A-C6A-C5A-O5A) et ceux impliqués dans la liaison entre la tête polaire et la chaîne alkyle (O5A-C1A-O1A-C7, C1A-O1A-C7-C8 et O1A-C7-C8-C9) (Tableau A1-5 et Figure A1-9).

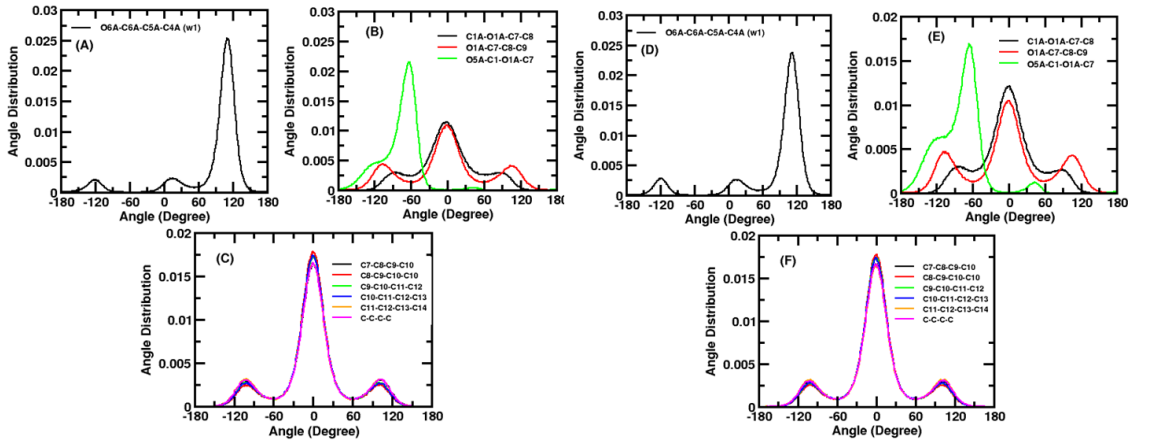


Figure A1-9 : Distributions moyennes des angles dièdres dans le cas des simulations pré-assemblées (A, B et C) et auto-assemblées (D, E et F).

Tableau A1-5 : Pourcentages moyens de conformation *trans*. Un dièdre est considéré dans une conformation *trans* si la valeur de l'angle Φ est dans l'intervalle -120 < Φ < 120°. [a](Bogusz *et al.* 2001).

Système	O6A-C6A-C5A-O5A	C1A-O1A-C7-C8	O1A-C7-C8-C9	O5A-C1A-O1A-C7	CCCC
PF	11,5	74,2	60,3	70,0	75,3
SA	11,2	65,0	66,7	80,0	73,7
Ref.[a]	15,0	78,1	69,1	85,0	80,4

En ce qui concerne les conformations du dièdre ϖ_1, les résultats sont similaires entre les deux simulations pré-assemblées et auto-assemblées, avec une préférence pour une conformation gauche$^+$. Ces résultats sont en bon accord avec ceux de Bogusz (Bogusz *et al.* 2001). Dans le cas des conformations des dièdres impliquées dans la liaison éther entre la tête polaire et chaîne octyle (O5A-C1A-O1A-C7, C1A-O1A-C7-C8 et O1A-C7-C8-C9), on observe une légère différence entre les deux types de simulations avec des conformations plus *gauche* dans

le cas des chaînes du bOG dans les micelles auto-assemblées (88 %, 10 % et 57 %) (Bogusz *et al.* 2001).

Dans le cas de la chaîne aliphatique du bOG, celle-ci est trouvée majoritairement *trans* (75 %) dans les deux types simulations et sont légèrement plus faibles que celles obtenues par Bogusz et al (Bogusz *et al.* 2000) et Konidala *et al.*(Konidala *et al.* 2006) autour de 80,4 % avec le champ de force CHARMM de Ha *et al.* (Ha *et al.* 1988), mais en bon accord avec les résultats obtenus pour la chaîne alkyle du dodecyl-maltoside avec le champ de force GLYCAM (84,7 %) ou CHARMM36 (72,1 %)(Abel *et al.* 2011).

Le repliement partiel de la chaîne alkyle du détergent occasionne une diminution de la longueur de la chaîne octyle puisque l'on calcul une distance entre les atomes C7 et C14 autour de 7,7 Å qui est ~25 % plus faible que la longueur d'une chaîne octyle étendue (10,4 Å) calculée avec la formule de Tanford (Tanford 1972).

Annexe 2 : Validation des paramètres de simulation de la membrane pure de POPC

Dans cette annexe, nous décrierons la mise au point et les résultats des simulations pour la membrane pure de POPC avec 0,15 M de KCl qui nous ont permis de valider notre modèle de membrane de POPC utilisé pour modéliser le complexe LH2.

I. Simulations moléculaires de membranes

1. Brèves considérations

Pour simuler des protéines membranaires, il est nécessaire d'insérer celle-ci dans un environnement biomimétique constitué soit de détergents soit de phospholipides (PL) membranaires. Comme il est encore difficile (mais pas impossible, voir par exemple le récent travail de van Eerden *et al.* (van Eerden *et al.* 2015)) de simuler une membrane avec une composition proche de celle des membranes cellulaires et que la plupart des expériences sur le complexe LH2 ont été faites en présence de POPC (Figure A2-1) (Chandler *et al.* 2008)(Janosi *et al.* 2006), nous avons utilisé préférentiellement ce PL comme modèle membranaire.

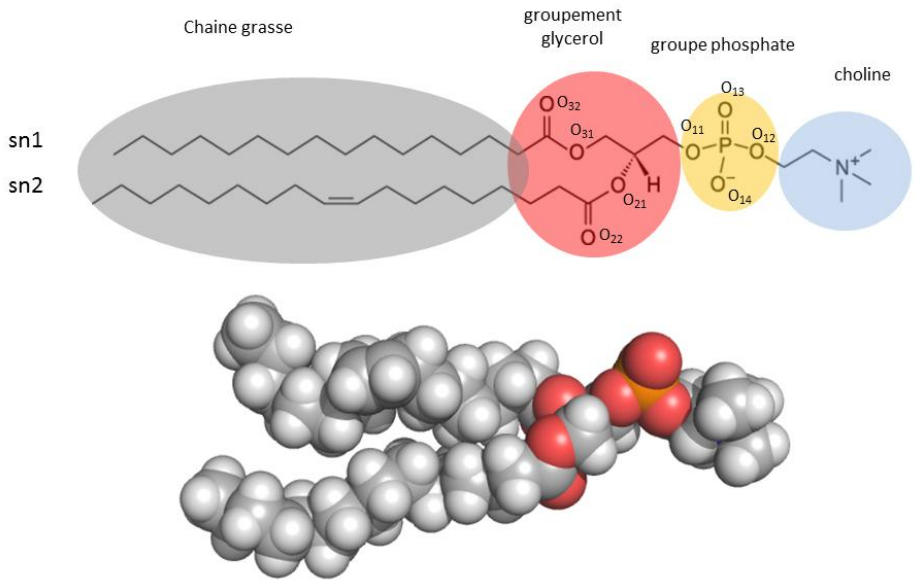


Figure A2-1 : Formule chimique du POPC avec les groupes constituants la molécule (en haut) et sa représentation en sphères de van-der-Waals (en bas).

Le POPC est un phopholipide mono-insaturé avec une chaîne 1-palmitoyl et oleyol en sn2 (18 :1 (9), 18 :0) utilisé couramment dans les expériences et simulations pour mimer les membranes biologiques.

2. Mise au point du système membranaire

- ***Conversion des paramètres Lipid14***

Pour créer la membrane de POPC, nous avons utilisé l'outil de construction de membrane disponible dans le logiciel de visualisation VMD (*Membrane Builder*) (Humphrey *et al.* 1996). La bicouche de POPC est composée de 128 phospholipides divisés en deux monocouches de 64 POPC (Figure A2-2).

Pour modéliser la molécule de POPC, nous avons utilisé les récents paramètres de Dickson *et al.* (Dickson *et al.* 2014) compatibles avec les champs de force dérivés d'AMBER qui sont une mise à jour des paramètres Lipid11 (Skjevik *et al.* 2012) permettant de faire des simulations de systèmes membranaires avec un couplage de pression de type anisotropique. Ces paramètres ont été développés avec une approche en briques modulaires (chaînes alkyles, groupe polaires) qui sont assemblées pour former différents phospholipides (e.g. DOPC, POPE, etc.).

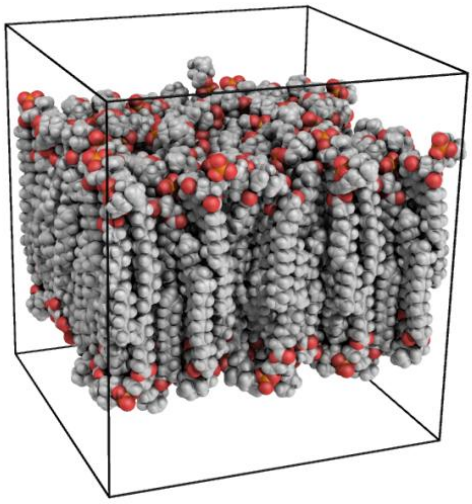


Figure A2-2 : Représentation de la boîte de simulation de départ avec 128 molécules de POPC, les molécules d'eau ont été masquées pour plus de clarté.

Les paramètres de Lipid14 pour le POPC n'étant pas disponibles par default dans la liste des champs de force de GROMACS, nous avons, en premier lieu, construit les paramètres du POPC au format AMBER (*.prmtop and *.inpcrd) et convertis ces fichiers au format de topologie et paramètre de GROMACS (*.top) en utilisant le code (AnteChamber PYthon Parser interfacE) ACPYPE (Silva *et al*. 2012). Pour vérifier la bonne conversion des paramètres, nous avons calculé un point d'énergie en prenant une configuration avec le module *Sander* disponible dans AmberTools14 (Case *et al.* 2015) et GROMACS. Le Tableau

A2-1 donne les valeurs d'énergie pour les différents termes de la fonction d'énergie et les différences calculées entre les deux codes.

Tableau A2-1 : Comparaison des énergies entre AMBER et GROMACS après conversion.

	AMBER (kJ/mol)	GROMACS (kJ/mol)	Différence
$E_{potentielle}$	229,27	230,73	+1,46
$E_{liaison}$	122,06	122,11	+0,05
E_{angle}	330,37	331,74	+1,37
$E_{dièdre}$	210,51	210,69	+0,18
E_{LJ}	60,36	60,29	+0,07
$E_{Coulomb}$	-76,58	-76,38	+0,20

Les valeurs des énergies sont similaires (aux arrondis près) permettant de valider notre conversion.

3. Simulation de la membrane

Dans un deuxième temps, pour valider nos paramètres de simulations (e.g. *cutoff* d'interaction, système de couplage de pression et température), nous avons simulé le même système que Dickson *et al.* (Dickson *et al.* 2014) contenant 128 POPC, 31 molécules d'eau TIP3P par POPC et 0,15 M de KCl. Nous avons minimisé le système en 10 000 pas avec la méthode de *steepest-descent* et une énergie de tolérance de 1000 $kJ.mol^{-1}.nm^{-1}$. Puis, nous avons équilibré le système dans l'ensemble NVT à 303 K avec le thermostat *v-rescale* et une constante de couplage de τ_T = 0,1 ps pendant 400 ps. L'équilibration du système a continué dans l'ensemble NPT en semi-isotropique pendant 500 ps avec le même thermostat et le barostat de Berendsen (τ_P = 3 ps), avec une compressibilité de $4{,}5.10^{-5}$ bar^{-1} et une pression de 1,0135 bar. En plus de ces paramètres, nous avons appliqué des corrections de dispersion à longue portée de l'énergie et de la pression. Enfin, le système a été équilibré en semi-isotropique pendant 127 ns dans l'ensemble NPT en changeant le barostat par celui de Parrinello-Rahman (τ_P = 5 ps). Nous avons ensuite changé de système de couplage (anisotropique) et collecté les trajectoires (toutes les 2 ps) pendant 207 ns (Figure A2-3). Nous avons simulé ~ 60 ns par jour sur 64 CPUs du supercalculateur CURIE (TGCC-CCRT).

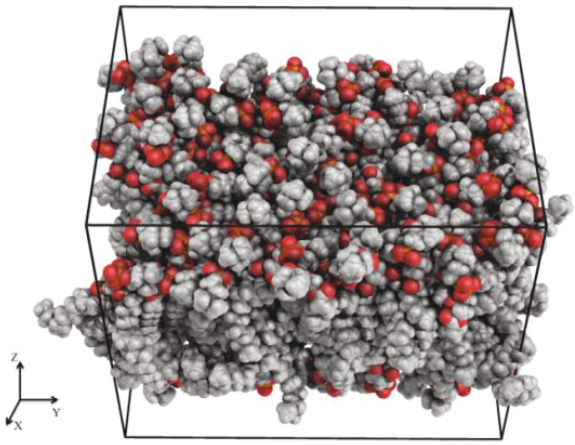


Figure A2-3 : Dernière configuration de la simulation NPT en anisotropique (t = 200 ns).

II. Résultats

Pour valider notre approche, nous avons calculés différents paramètres de structures de la membrane et comparer avec la littérature expérimentale et la simulation de Dickson *et al.*

1. Aire par lipide et compressibilité de surface

La surface par lipide (A_L) est une propriété importante qui permet de vérifier si la membrane a convergé vers un état stable. Pour calculer la valeur moyenne de surface et le volume par lipide, nous avons utilisé les paramètres de la boîte de simulation toutes les 10 ps:

$$A_L = \frac{dimension\ X \times dimension\ Y}{N_{lipide\ par\ monocouche}} \quad \text{(A2.II.1)}$$

$$V_L = \frac{(V_{boîte}) - (N_{eau} \times V_{eau})}{N_{total\ de\ lipide}} \quad \text{(A2.II.2)}$$

en prenant un volume de 30,80 Å^3 pour la molécule d'eau TIP3P (calculé à partir d'une simulation de boîte d'eau pure de 1000 molécules dans les mêmes conditions de température et de pression). Cette valeur est en bon accord avec celle obtenue dans l'article, qui est de 30,53 Å^3.

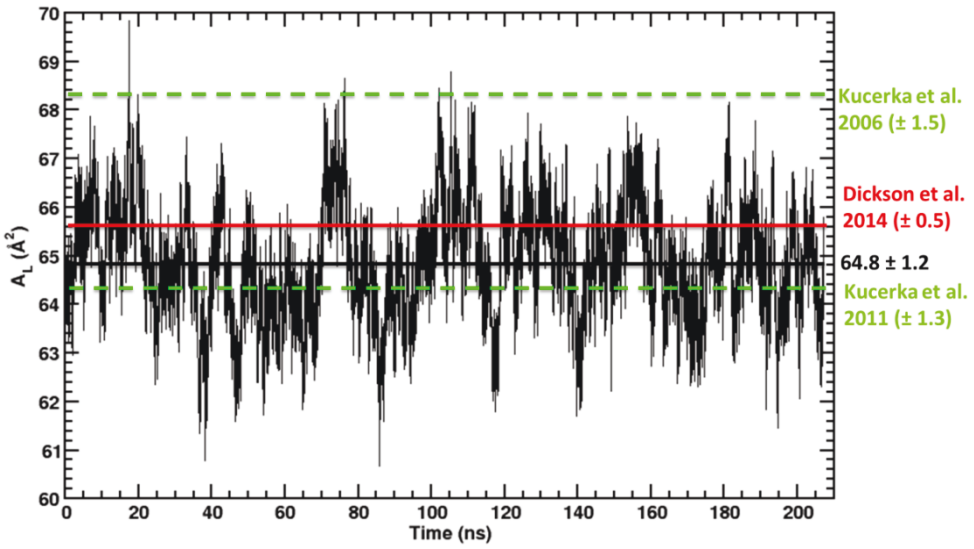


Figure A2-4 : Surface par lipide (S_L ou A_L en anglais) en fonction du temps de simulation (en noir) comparer aux valeurs expérimentales en vert : (Kučerka *et al.* 2006) en SAXS et (Kučerka *et al.* 2011) en combinant SAXS et SANS, et théorique (Dickson *et al.* 2014) en rouge.

La Figure A2-4 montre l'évolution des surfaces par lipide en fonction du temps de simulation. La valeur moyenne obtenue de 64,8 ± 1,2 Å^2 est en bon accord avec les valeurs expérimentales (64,3 et 68,3 Å^2) et celle de Dickson *et al.* (65,6 ± 0,5 Å^2).

Le module de compressibilité de surface K_A a été calculé à partir de la formule :

$$K_A = \frac{2k_bT\langle SL\rangle}{\sigma_S^2\, N_{lipide\ par\ monocouche}} \qquad \text{(A2.II.3)}$$

avec k_b la constante de Boltzmann, T la température et σ_S^2 la variance de la surface par lipide. La valeur de K_A est rassemblée dans le Tableau A2-2 avec d'autres paramètres de structures.

Tableau A2-2 : Propriétés structurales de la membrane de POPC comparées aux résultats expérimentaux et ceux de l'article (Dickson *et al.* 2014). Les données expérimentales sont issues de (Kučerka *et al.* 2006) (Kučerka *et al.* 2011).

	Surface par lipide A_L (Å²)	Volume par lipide V_L (Å^3)	Module de compressibilité isothermal de surface K_A (mN.m^{-1})	Épaisseur de la bicouche D_{HH} (Å)	Épaisseur de la bicouche de Luzzati D_B (Å)
Expériences	64,3-68,3	1256	180-330	37	36,8-39,1
Article Lipid14	65,6 ± 0,5	1205,4 ± 0,4	257 ± 47	36,9 ± 0,6	36,8 ± 0,3
Notre simulation	64,8 ± 1,2	1189,7 ± 6,1	338	37,2	36,6

2. Profil de densité électronique

L'épaisseur de la bicouche de Luzzati D_B a été calculée *via* la formule 2VL/AL (Nagle *et al.* 2000). Enfin, l'épaisseur de la bicouche a été examinée à partir du graphique de la densité électronique en mesurant la distance D_{HH} entre les deux pics de la membrane (Figure A2-5), l'épaisseur de la membrane, $D'_B/2$ et de la chaîne grasse.

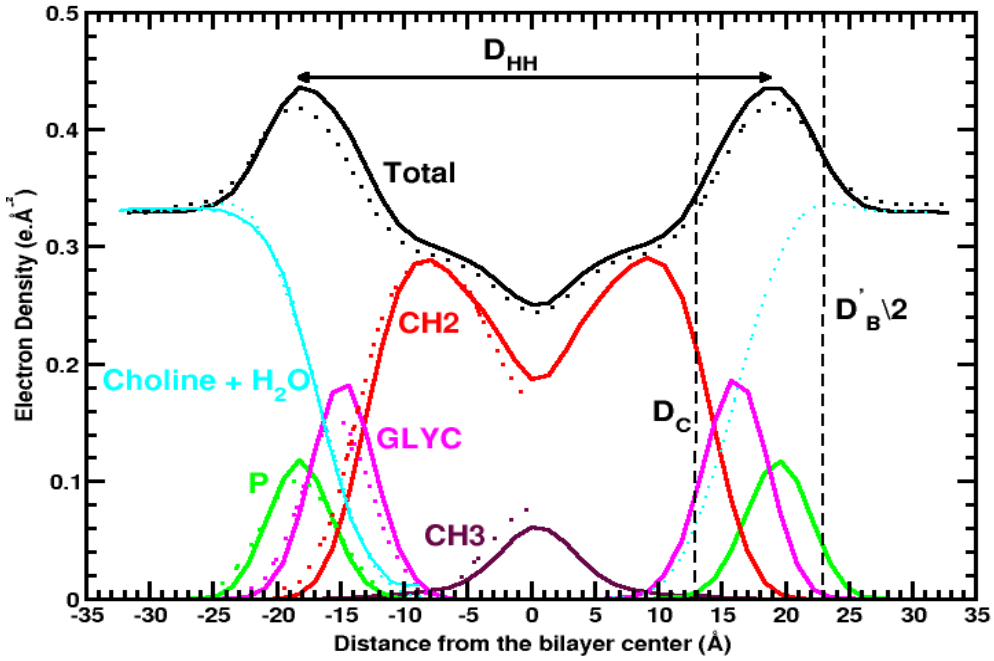


Figure A2-5 : Profil de densité électronique moyen des différents composants de la membrane de POPC dans notre simulation. En ligne continue, les données issues de notre simulations en point celles de (Kučerka *et al.* 2006) à 303 K avec une surface par lipide A_L de 64,3 Å^2.

Hormis l'amplitude des pics, notre simulation reproduit bien les caractéristiques structurales de la membrane de POPC expérimentales.

3. Facteur d'ordre des chaînes alkyles du POPC

L'orientation moyenne des chaînes lipidiques par rapport à la normale à la membrane a été obtenue en calculant le paramètre d'ordre S_{CD}. Ce dernier peut être comparé aux mesures expérimentales de RMN au deutérium qui décrivent les liaisons Carbone-Deutérium. Nous avons calculé le paramètre d'ordre du carbone i selon l'expression:

$$|S_{CD}| = \left|\frac{1}{2}\langle 3cos^2\theta_i - 1\rangle\right| \qquad \text{(A2.II.4)}$$

avec θ_i l'angle entre la normale à la membrane z et l'axe du vecteur passant par le carbone i-1 et i+1.

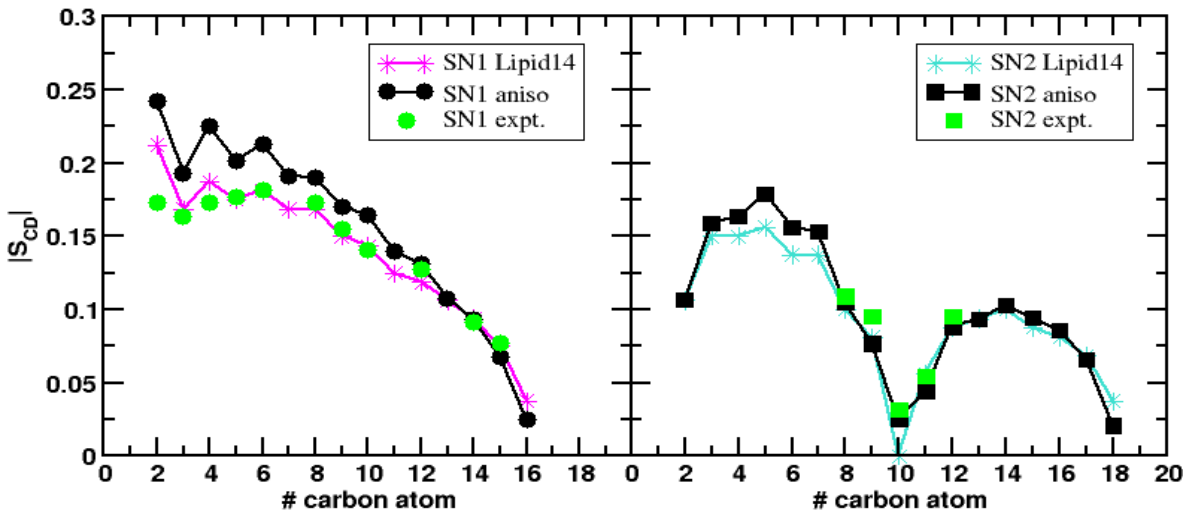


Figure A2-6 : Paramètre d'ordre pour la chaîne SN1 (à gauche) et SN2 (à droite) dans nos simulations (noir) comparé aux valeurs expérimentales (vert) et théoriques (magenta et cyan)(Dickson *et al.* 2014).

Les graphiques de la Figure A2-6 montrent les valeurs du paramètre d'ordre S_{CD} pour les chaînes saturées SN1 et insaturées SN2 dans notre simulation. Ces valeurs ont été calculées grâce à un programme codé en Python dans notre laboratoire. Nous avons comparé nos résultats avec les valeurs expérimentales et celles obtenues par Dickson et ses collaborateurs. Nous remarquons que nos valeurs sont en bon accord avec la littérature expérimentale et théorique.

4. Hydratation de la tête polaire

Enfin, nous avons analysé l'hydratation des têtes polaires des POPC en calculant le nombre de molécule d'eau au voisinage des groupes polaires du POPC en intégrant les valeurs de $g_{N-O_W}(r)$ à leur premier minimum (5,9, 4,5 and ~3,3 Å pour l'azote, le phosphate et les oxygènes, respectivement) (Tableau A2-3) et en les comparant à la littérature expérimentale et théorique.

Tableau A2-3 : Nombre de molécules d'eau dans la première couche d'hydratation de l'azote du phosphate et des oxygènes liés au phosphate ($P-O_{11}$, $P-O_{12}$, $P=O_{13}$ et $P-O_{14}$) et du groupe carboxyle des chaînes sn1 ($C-O_{31}$ et $C=O_{32}$) et sn2 ($C-O_{21}$ et $C=O_{22}$) du POPC. La localisation des atomes est donnée sur la Figure A2-1. [a](Lopez *et al.* 2004), [b](Tieleman *et al.* 1996), [c](Pan *et al.* 2012), [d](Kucerka *et al.* 2008), [e](Poger *et al.* 2010).

Atomes	Notre simulation	Littérature
N	18,5	19,0 – 23,0[a,b]
P	6,2	8,0[c,d]
O_{12}	0,4	0,6 – 0,4[e]
O_{13}	2,8	3,4[c]
O_{14}	2,8	
O_{11}	0,4	0,6 – 0,4[e]
O_{21}	0,2	0,6 – 0,4[e]
O_{22}	0,9	-
O_{31}	0,1	0,6 – 0,4[e]
O_{32}	1,1	-

En accord avec la littérature, nous trouvons que l'hydratation du POPC se fait majoritairement avec les groupes phosphate et choline 6,2 et 18,5 molécules d'eau, respectivement. Les oxygènes éthers sont moins hydratés que les oxygènes carboxyles (0,4 *versus* 1,0).

Annexe 3 : Calculs des charges RESP du résidu CXM et de la RG

Dans cette annexe, nous décrierons les approches utilisées pour dériver les charges partielles atomiques RESP pour le résidu N-carboxy-L-méthionine (CXM) et la rhodopine glucoside (RG).

I. Le résidu N-carboxy-L-méthionine (CXM)

Pour le calcul des charges partielles du résidu CXM, nous avons suivi la méthode décrite dans l'article de Cornell et ses collaborateurs (Cornell *et al.* 1995) pour les acides aminées et en utilisant le serveur web RED. Cette approche consiste à utiliser des groupements acétyle (ACE) et méthylamide (NME) fixés à la partie N and C terminale du résidu à modifier en appliquant des contraintes de charges (*INTRA-MCC*) égale à zéro sur les groupements ACE et NME (Figure A3-1). Les charges ont été couplées aux paramètres du champ de force Amber99sb-ILDN (Lindorff-Larsen *et al.* 2010). Les charges RESP et le type d'atome pour chaque atome du CXM sont donnés dans le Tableau A3-1 (RRMS = 0,16).

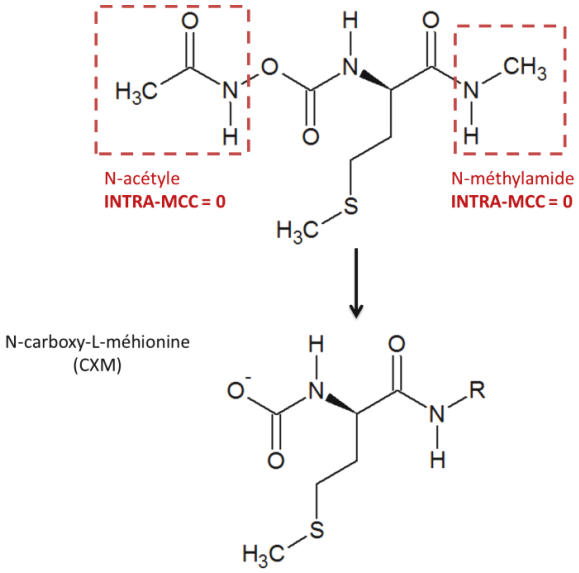


Figure A3-1 : Représentation des blocs modulaires utilisés dans le calcul des charges partielles RESP du résidu CXM. Les groupements N-acétyle (ACE) et méthylamide (NME) sont soumis à des contraintes internes de charge 0 (INTRA-MCC=0).

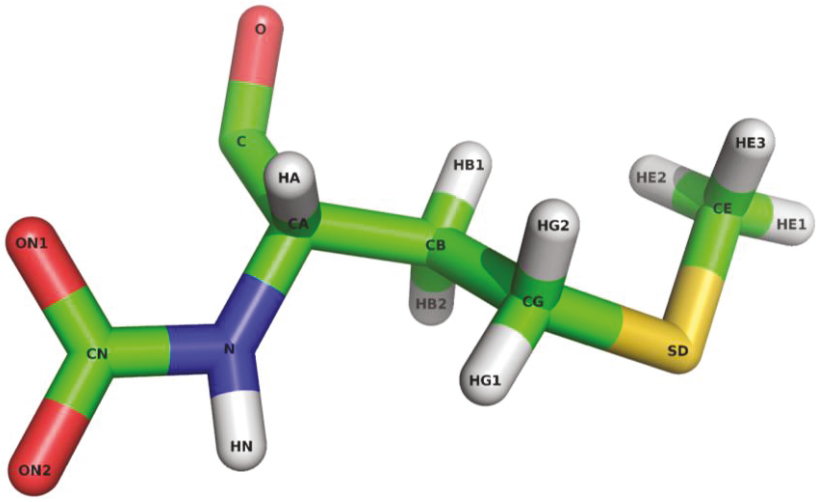


Figure A3-2: Représentation du résidu CXM avec le nom des atomes.

Tableau A3-1 : Charges partielles atomiques RESP du résidu CXM avec le type d'atome utilisé dans Amber99sb.

Atome	Type d'atome	Charges
N	N	-0,4793
HN	H	0,2777
CA	CT	0,0347
HA	H1	0,0671
C	C	0,5853
O	O	-0,5376
CB	CT	0,0108
HB1, HB2	HC	0,0545
CG	CT	-0,2053
HG1, HG2	H1	0,1405
SD	S	-0,2822
CE	CT	-0,0869
HE1, HE2, HE3	H1	0,0790
CN	C	0,7262
ON1	O	-0,1576
ON22	O	-0,5799

II. La rhodopine glucoside (RG)

Dans le cas de la rhodopine glucoside, nous avons employé l'approche en brique modulaire (Figure A3-3), faisant appel à un méthyle glucopyranoside et l'hydrométhyle lycopène.

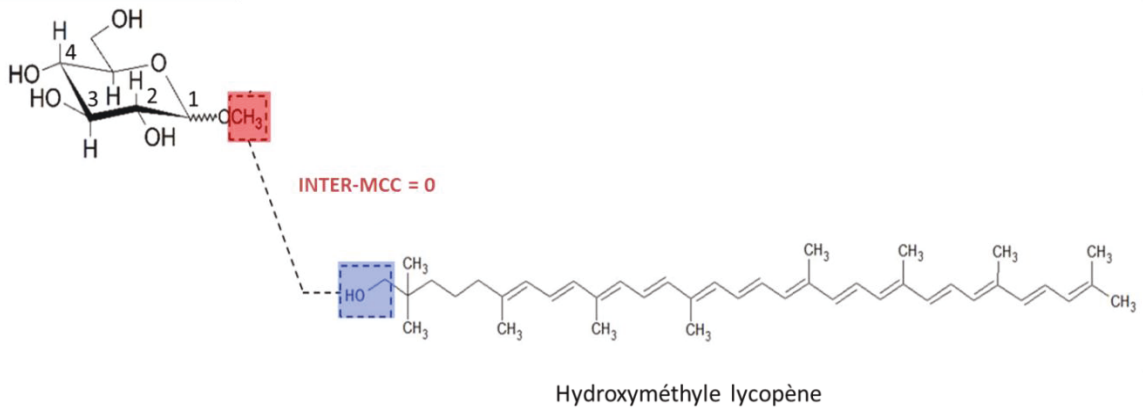


Figure A3-3 : Représentation des blocs modulaires utilisés dans le calcul des charges partielles RESP du RG avec le méthyle glucopyranoside à gauche et le lycopène à droite. La contrainte de charge nulle entre molécule est représentée par l'INTER-MCC=0 et celles sur les hydrogènes des carbone sp^3 ne sont pas représentés mais elles ont été spécifiées par des INTRA-MCC=0.

La dérivation des charges RESP pour la partie polaire méthyle glucopyranoside et de la chaîne lycopène a été réalisée avec le protocole GLYCAM (Woods *et al.* 2000) (Kirschner *et al.* 2012). Les groupements méthyles et hydroxyméthyles ont été contraints à une charge nulle (*INTER-MCC*). De plus, des contraintes internes ont été ajoutées sur les hydrogènes liés aux carbones aliphatiques pour que leur charge soit égale à zéro. Conformément au prescription original de GLYCAM , pour éviter la formation de liaisons hydrogènes intramoléculaires et un effet d'hyperpolarisation, une contrainte angulaire entre les hydroxyles OH2, OH3 et OH4 du glucose ont été fixé à 180° (« *ModRedundant F* ») du programme GAUSSIAN03 (Frisch *et al.* 2004) lors de l'optimisation géométrique des glucoses (Abel *et al.* 2011). De plus, pour tenir compte du fait qu'en solution, l'angle dièdre ω des molécules de glucopyranose coexistent dans des conformations *gauche-gauche* (*gg*, ω = −60°) et *gauche−trans* (*gt*, ω = +60°), nous avons contraints ces deux conformations (« *ModRedundant F* ») lors de la dérivation des charges. Les charges RESP et le type d'atome de chaque atome du RG sont donnés dans le Tableau A3-2 et A3-3 (RRMS = 0,25).

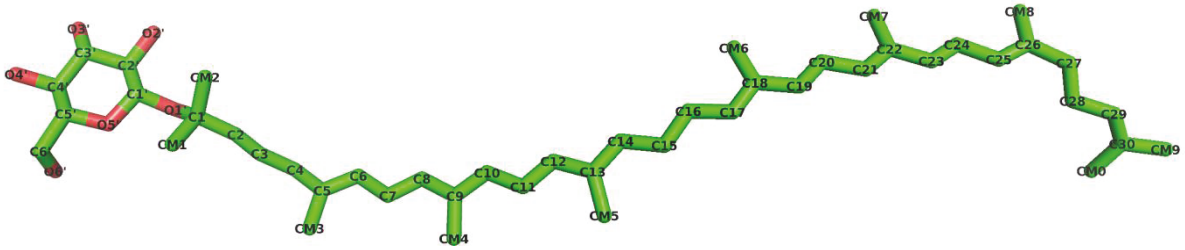


Figure A3-4 : Représentation du résidu RG avec le nom des atomes. Les hydrogènes ne sont pas montrés pour clarifier la figure.

Tableau A3-2 : Charges partielles atomiques RESP de la tête glucoside de la RG avec le type d'atome utilisé dans Glycam.

Atome	Type d'atome	Charges
C6'	cg	0,2371
H6'1, H6'2	h1	0,0000
O6'	oh	-0,6182
HO6	ho	0,4061
C5'	cg	0,1106
H5'	h1	0,0000
O5'	os	-0,4095
C4'	cg	0,3274
H4'	h1	0,0000
O4'	oh	-0,6777
HO4	ho	0,4009
C3'	cg	0,3090
H3'	h1	0,0000
O3'	oh	-0,7177
HO3	ho	0,4428
C2'	cg	0,3125
H2'	h1	0,0000
O2'	oh	-0,6858
HO2	ho	0,4061
C1'	cg	0,3154
H1'	h2	0,0000
O1'	os	-0,2992

Tableau A3-3 : Charges partielles atomiques RESP du lycopène de la RG avec le type d'atome utilisé dans Glycam.

Atome	Type d'atome	Charges
C1	cg	0,9297
CM1	cg	-0,2937
HM11, HM12, HM13	hc	0,0000
CM2	cg	-0,2937
HM21, HM22, HM23	hc	-0,6182
C2	cg	-0,2627
H21, H22	hc	0,0000
C3	cg	0,0606
H31, H32	hc	0,0000
C4	cg	0,0971
H41, H42	hc	0,0000
C5	ck	-0,2391
CM3	cg	0,1254
HM31, HM32, HM33	hc	0,0000
C6	ck	0,0578
H61	ha	0,0000
C7	cj	-0,1318
H71	ha	0,0000
C8	cj	0,1436
H81	ha	0,0000
C9	ck	-0,3496
CM4	cg	0,2091
HM41, HM42, HM43	hc	0,0000
C10	ck	0,1586
H101	ha	0,0000
C11	cj	-0,1897
H111	ha	0,0000
C12	cj	0,1944
H121	ha	0,0000
C13	ck	-0,4271
CM5	cg	0,2318
HM51, HM52, HM53	hc	0,0000
C14	ck	0,1973

H141	ha	0,0000
C15	cj	-0,0834
H151	ha	0,0000
C16	cj	-0,0870
H161	ha	0,0000
C17	ck	0,2045
H171	ha	0,0000
C18	ck	-0,4371
CM6	cg	0,2349
HM61, HM62, HM63	hc	0,0000
C19	cj	0,1999
H191	ha	0,0000
C20	cj	-0,1883
H201	ha	0,0000
C21	ck	0,1528
H211	ha	0,0000
C22	ck	-0,3427
CM7	cg	0,2090
HM71, HM72, HM73	hc	0,0000
C23	cj	0,1615
H231	ha	0,0000
C24	cj	-0,1907
H241	ha	0,0000
C25	ck	0,1694
H251	ha	0,0000
C26	ck	-0,3599
CM8	cg	0,2098
HM81, HM82, HM83	hc	0,0000
C27	cj	0,1529
H271	ha	0,0000
C28	cj	-0,1395
H281	ha	0,0000
C29	ck	0,0434
H291	ha	0,0000
C30	ck	-0,1867
CM9	cg	0,0834

HM91, HM92, HM93	hc	0,0000
CM0	cg	0,1160
HM01, HM02, HM03	hc	0,0000

Annexe 4 : Nos expériences de SAXS avec les complexes LH2-détergents

Dans le chapitre 4, nous avons proposé des modèles de complexes LH2 entourés de molécules de détergent LDAO ou de bOG. Le nombre de molécules de détergent a été choisi en se basant sur les données de la littérature qui comportaient des incertitudes assez importantes. Pour vérifier la justesse de nos modèles, nous avons donc effectué des expériences de diffusion de rayons X aux petits angles (SAXS ou *Small Angle X-rays Scattering*) sur la ligne de lumière SWING dirigée par le Dr. Javier Pérez (Synchroton SOLEIL).

- ***Principes de la SAXS***

La diffusion des rayons X aux petits angles est une technique expérimentale qui permet d'étudier les propriétés structurales de macromolécules (de quelques dizaines à quelques centaines de nm) en se basant sur l'interaction des photons avec les nuages électroniques. C'est la diffusion des photons traversant l'échantillon qui fournit des informations sur la fluctuation des densités électroniques des macromolécules.

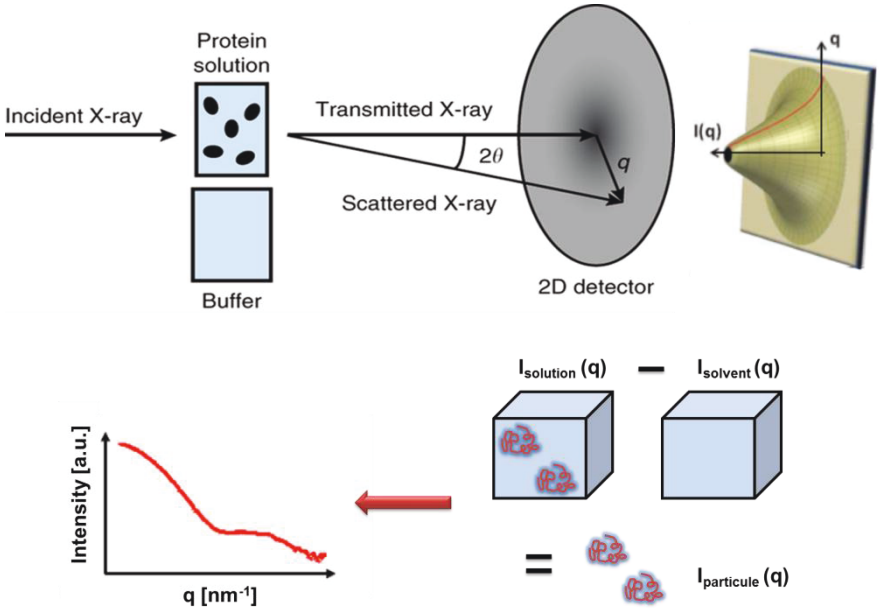


Figure A4-1 : Principe des expériences de SAXS. En haut : schéma du dispositif mis en place pour une expérience de SAXS ; en bas : le traitement des données pour obtenir la courbe de SAXS pour la particule étudiée. Redessinée à partir de (Skou *et al.* 2014)

La Figure A4-1 montre le principe d'une expérience de SAXS qui consiste à mesurer et analyser les modulations de l'intensité diffusée par un faisceau de rayons X traversant un

échantillon à des angles très proches du faisceau direct. La ligne de lumière SWING a la particularité de collecter les données de SAXS en combinant une mesure de SEC-HPLC (chromatographie d'exclusion stérique-liquide à haute performance) (Pérez *et al*. 2012). Cette méthode permet de soustraire à l'échantillon mesuré la contribution des micelles libres en solution.

- ***Nos expériences de SAXS***

Les expériences de SAXS ont été réalisées avec des complexes LH2 en présence de LDAO et de bOG. Les échantillons de 200 µL ont été préparés au laboratoire par le Dr. Andrew Gall avec des complexes LH2 dans 1 % bOG ou 0,1 % de LDAO, et avec 200 mM Tris.Cl et 100mM NaCl selon le protocole décrit dans (Gall *et al*. 1999) et (Rutkauskas *et al.* 2006). La concentration finale des complexes LH2 est de 258 et 148 µM en présence de bOG et LDAO, respectivement. La Figure A4-2 montre les profils d'élution réalisés au cours de l'expérience avec le LH2-bOG et le LH2-LDAO, permettant d'obtenir I(0) après extrapolation de l'intensité à q=0 par l'approximation de Guinier, et les rayons de giration des complexes (courbes bleue et noire, respectivement).

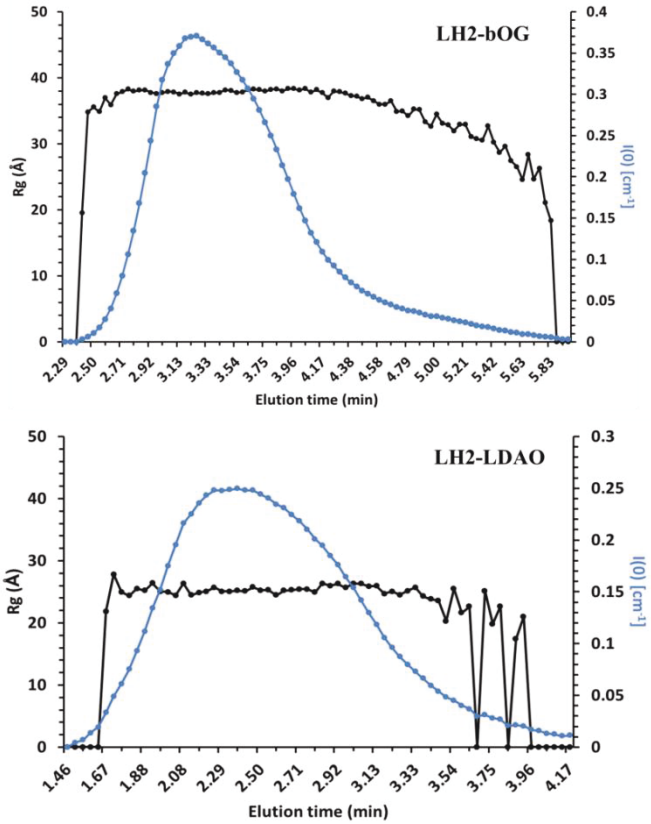


Figure A4-2 : Profils d'élution du complexe LH2-bOG et LH2-LDAO, avec l'extrapolation de l'intensité à q=0, I(0) en bleu et en cm^{-1}, et le rayon de giration en noir en fonction du temps d'élution en minute.

Nous remarquons qu'au niveau du pic de I(0) qui correspond à la présence de complexe LH2-détergents, le rayon de giration reste pratiquement constant tout au long de l'élution. Les rayons de giration du complexe LH2-bOG et LH2-LDAO sont égaux à 38,1 ± 0,6 Å et 25,6 ± 0,1 Å, respectivement. Les courbes de diffraction extraites pendant cette période d'élution ont été moyennées pour obtenir les courbes de la Figure A4-3. Ces courbes ont un aspect assez différent ce qui nous suggère des formes assez différentes entre le complexe LH2-LDAO et bOG.

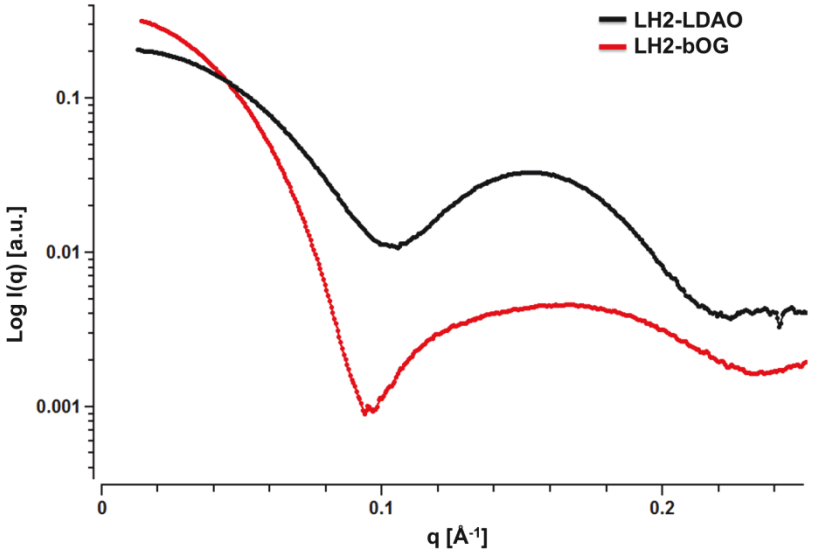


Figure A4-3 : Courbes de diffraction finale pour le LH2-bOG et LH2-LDAO.

Annexe 5 : Structure des peptides du complexe LH2

Dans le but d'analyser l'évolution des conformations des peptides du complexe LH2 dans nos différentes simulations (avec du LDAO, POPC et bOG), nous avons calculé les structures secondaires grâce au programme DSSP (Joosten *et al.* 2011) (Kabsch *et al*. 1983). Puis, nous avons calculé le facteur d'agitation thermique ou facteur de Debye-Waller (facteur B) du carbone α de chaque résidu. Globalement, nous remarquons que les hélices α (résidus 12-37 et 5-36 des peptides α et β, respectivement) sont conservées tout au long des différentes simulations.

I. Structures secondaires

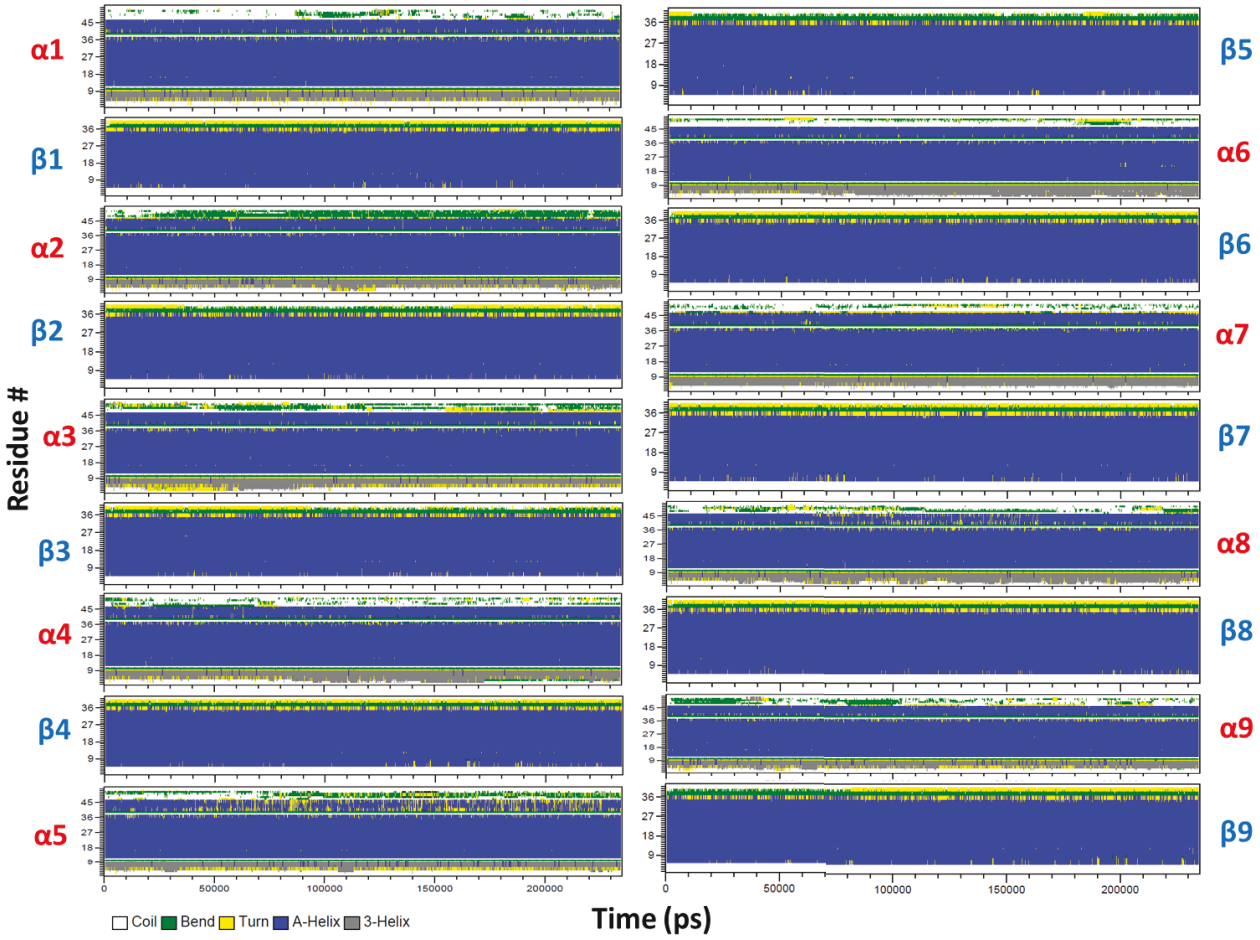


Figure A5-1 : Structures secondaires des peptides α et β du complexe LH2 au cours de la simulation avec du LDAO (Amber99sb).

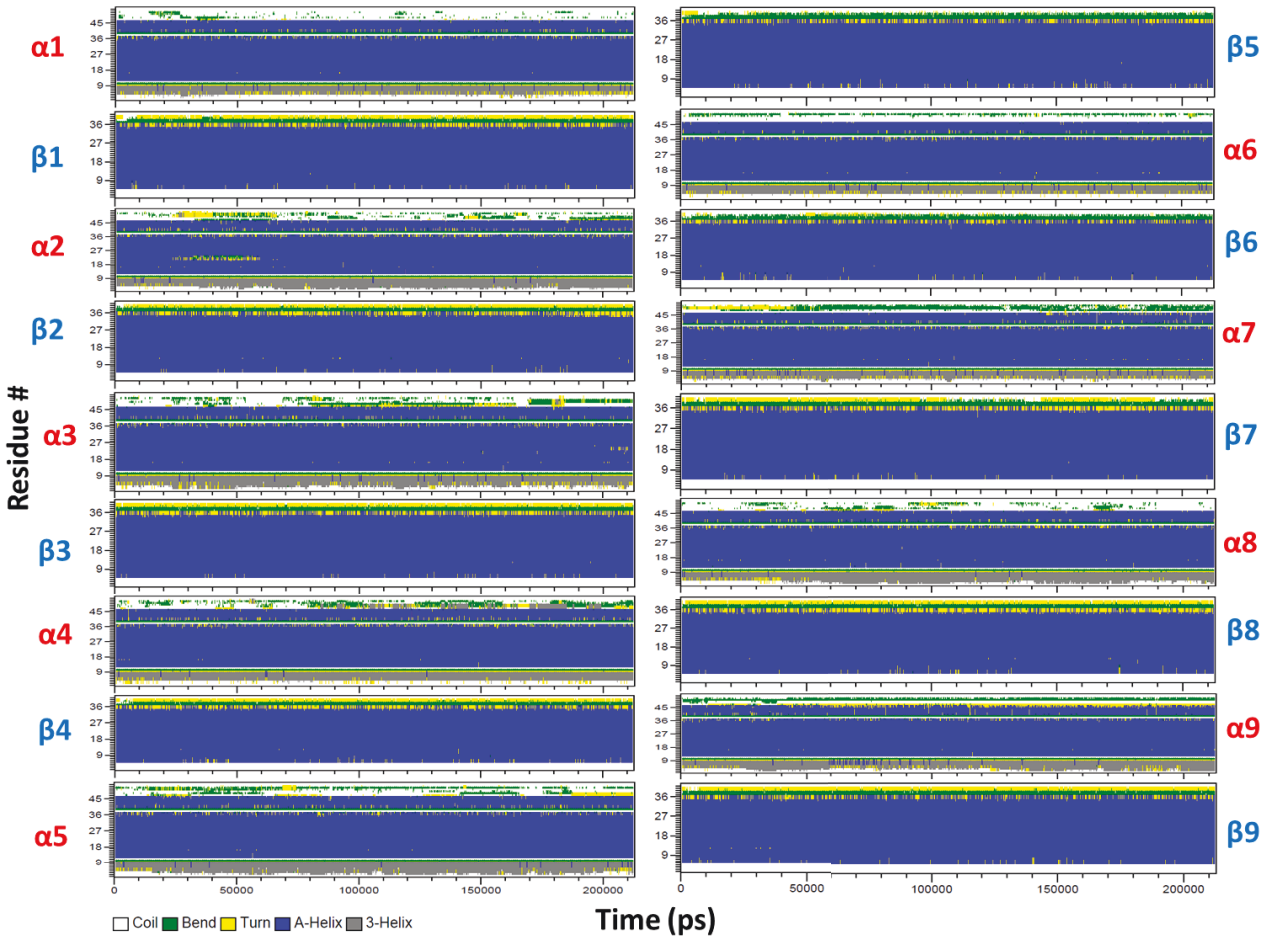


Figure A5-2 : Structures secondaires des peptides α et β du complexe LH2 au cours de la simulation avec du LDAO (Lipid14).

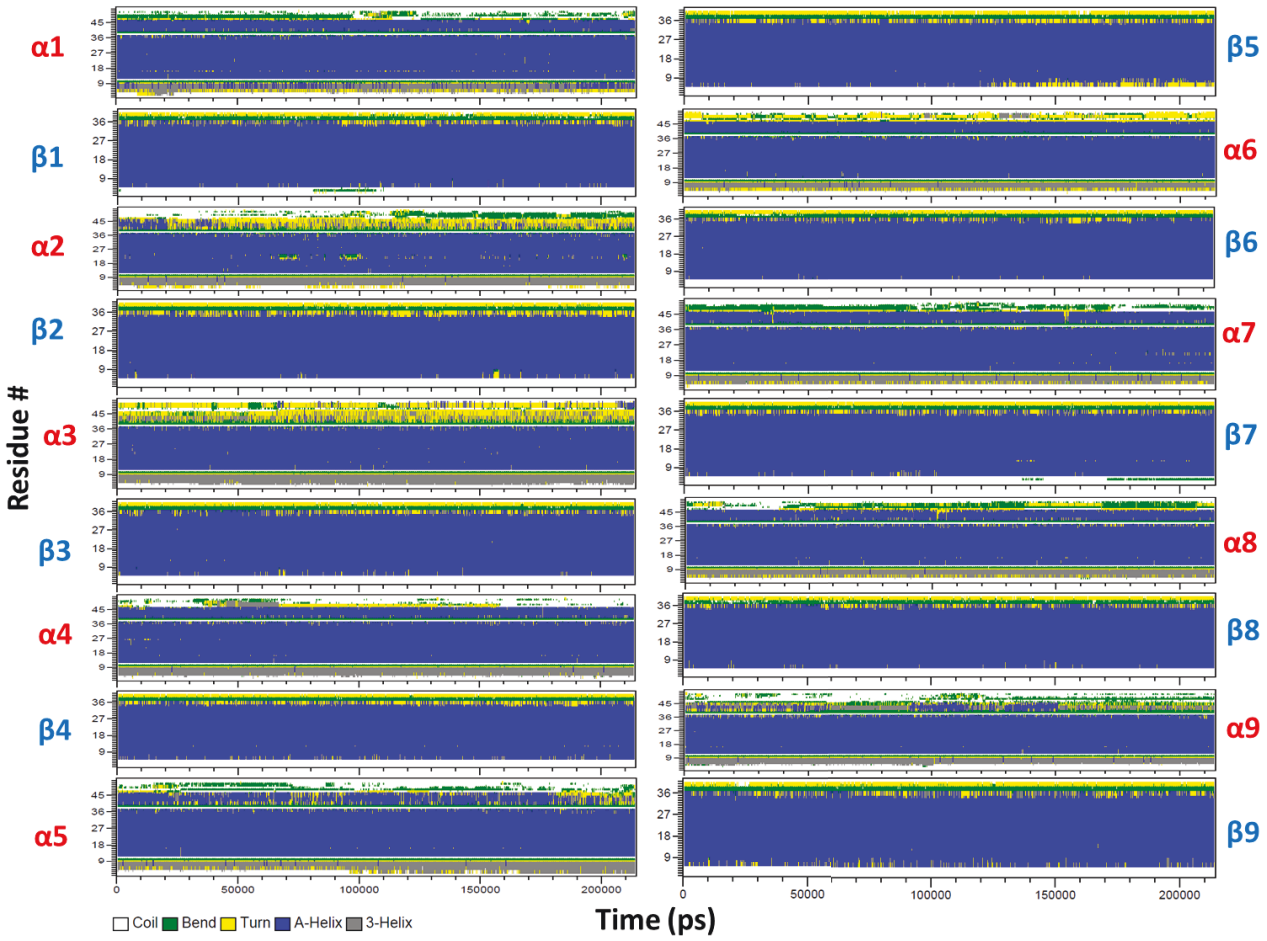


Figure A5-3 : Structures secondaires des peptides α et β du complexe LH2 au cours de la simulation avec du POPC.

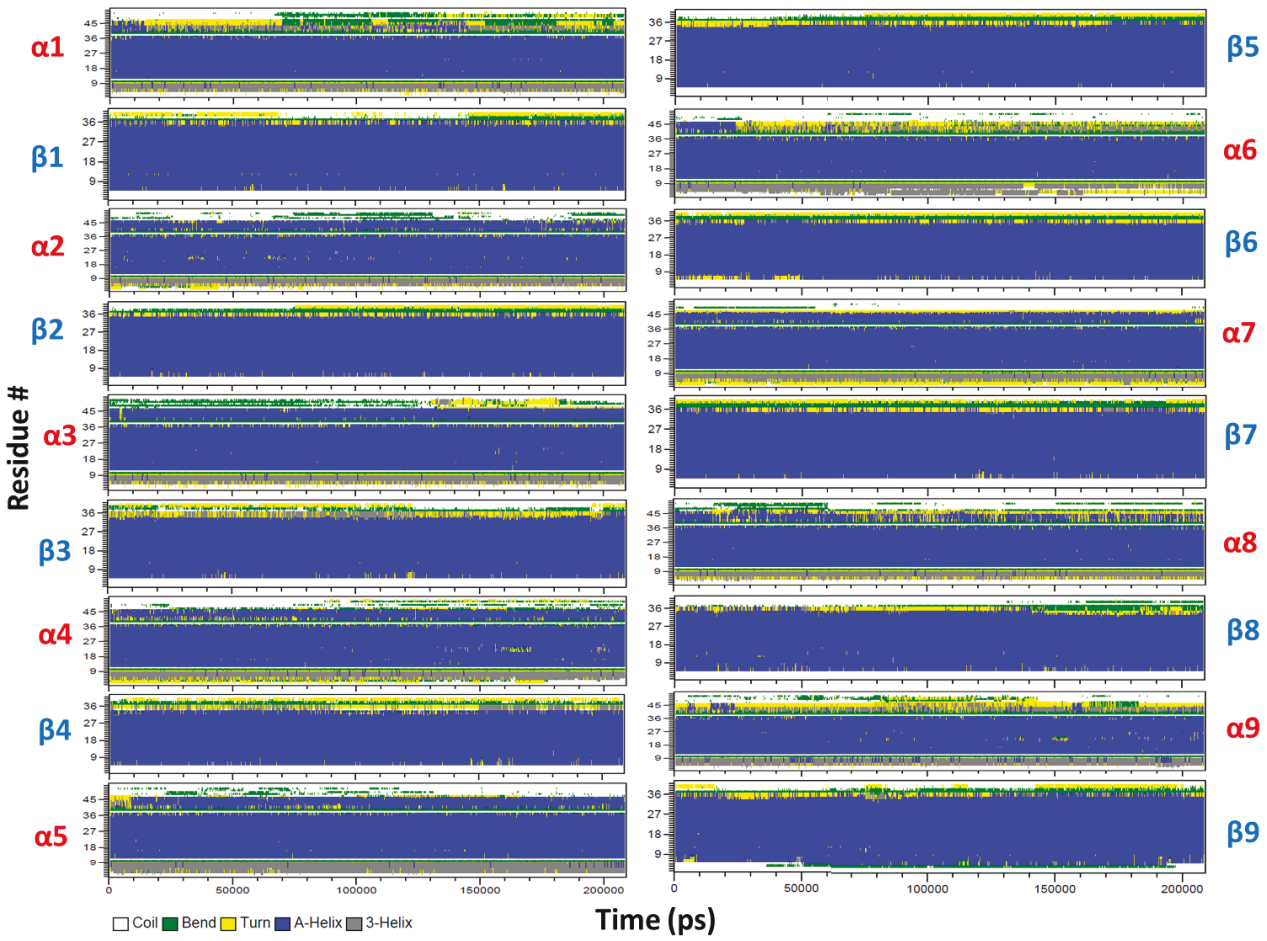


Figure A5-4 : Structures secondaires des peptides α et β du complexe LH2 au cours de la simulation avec du bOG.

II. Le facteur d'agitation thermique

Pour le calcul du facteur B des carbones α de tous les résidus, nous avons utilisé la commande *g_rmsf* de GROMACS sur les 30 dernières ns des différentes simulations (Figure A5-5). Les mouvements de grandes amplitudes correspondent aux régions non-structurées (boucles) des peptides.

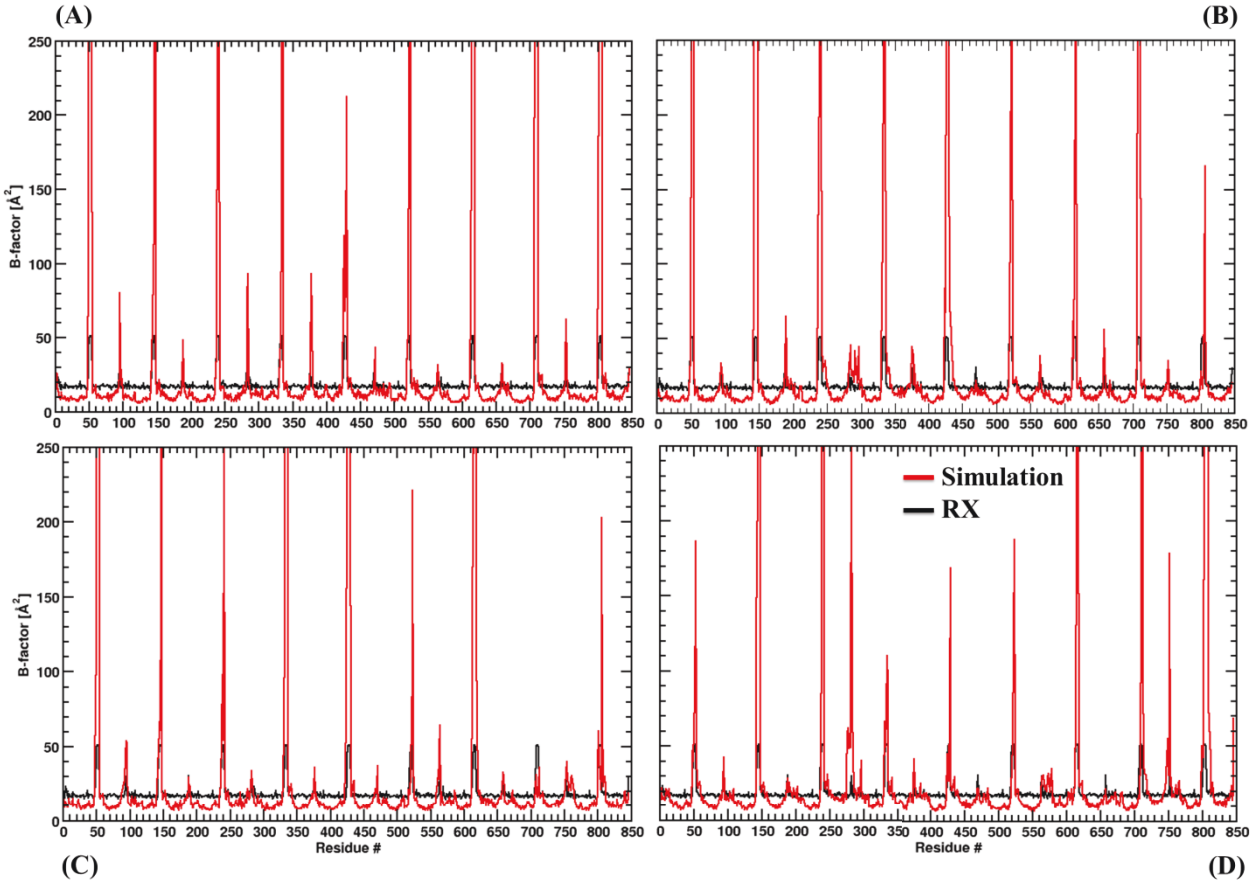


Figure A5-5 : Facteur B du carbone α de chaque résidu dans les quatre simulations (courbes rouges): LH2-LDAO Amber99sb (A), LH2-LDAO Lipid14 (B), LH2-POPC (C) et LH2-bOG (D). La courbe en noir correspond aux valeurs extraites de la structure cristallographique 1KZU.

Annexe 6 : Analyse en Composantes Principales (ACP)

L'analyse en composantes principales (ACP) d'une trajectoire nous permet d'identifier les plus grands mouvements collectifs des atomes et ainsi révéler les structures sous-jacentes aux fluctuations atomiques. Avec GROMACS, nous avons réalisé une ACP de la trajectoire en générant une matrice de covariance (avec *g_covar*) qui va capturer le degré de colinéarité des mouvements atomiques pour chaque paire d'atomes. Puis, nous avons diagonalisé cette matrice (avec *g_anaeig*) afin d'analyser les vecteurs et les valeurs propres. Chacun des vecteurs propres décrit un mouvement collectif des particules et la valeur propre associée donne la somme de la fluctuation décrit par le mouvement collectif de l'atome.

I. L'ACP des peptides α et β

Pour le calcul de l'ACP des peptides α et β, nous avons sélectionné le squelette peptidique (N-Cα-C) de la partie correspondant aux résidus 15 à 37 et 6 à 35, respectivement. Les graphiques de la Figure A6-1 représentent le pourcentage des fluctuations expliquées par chaque vecteur dans les différentes simulations. Nous remarquons que globalement 5 vecteurs suffisent pour expliquer les fluctuations des carbones α tout au long de la simulation et parmi ces 5 vecteurs les deux premiers vecteurs décrivent le plus de variabilité : 29 et 7 % avec le LDAO Amber99sb, 17 et 9 % avec le LDAO Lipid14, 12 et 7 % avec le POPC, et 19 et 11 % avec le bOG. C'est pourquoi pour la suite de l'analyse nous n'avons utilisé que ces deux premiers vecteurs pour représenter les principaux mouvements du squelette peptidique.

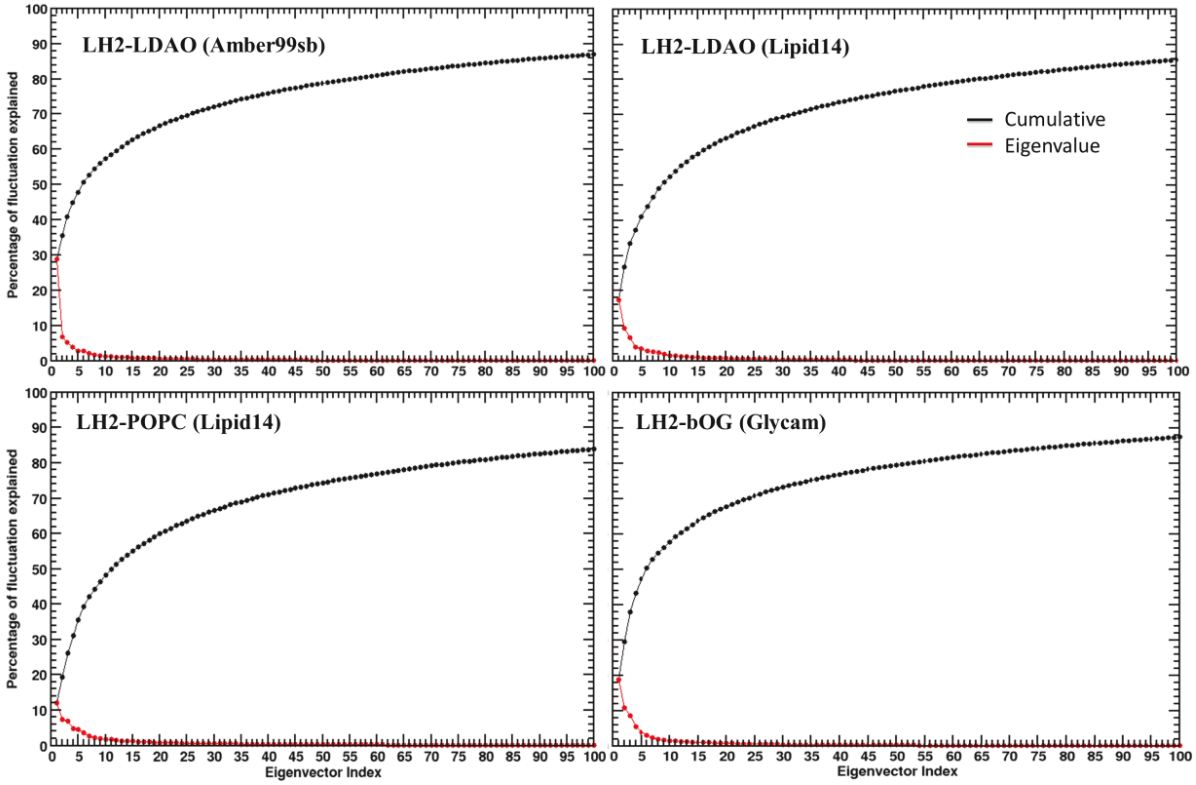


Figure A6-1 : Pourcentage des fluctuations décrit par les vecteurs propres dans les différents systèmes pour les atomes du squelette peptidique des hélices transmembranaires. Les courbes rouges représentent les valeurs propres associées aux vecteurs propres et les courbes noires montrent les pourcentages cumulés des valeurs propres.

Afin de visualiser l'espace conformationnel exploré par le squelette peptidique, nous avons projeté le vecteur 2 contre le vecteur 1 (Figure A6-2) dans les différents systèmes. Nous remarquons que le squelette peptidique explore un large éventail de conformations en présence de LDAO et de bOG comparé à la simulation avec du POPC. De plus, les conformations des peptides en présence de bOG sont assez différentes comparé à celles en présence de LDAO.

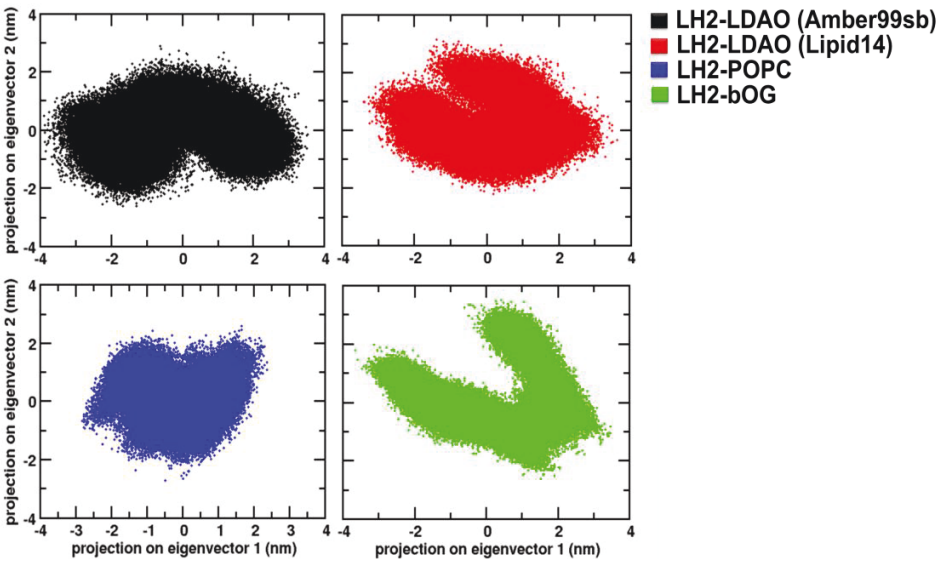


Figure A6-2 : Projection du vecteur 1 contre le vecteur 2 dans les différents systèmes pour visualiser l'espace conformationnel du squelette peptidique.

Nous pouvons aussi visualiser les fluctuations dans chaque vecteur grâce au RMSF (*Root Mean Square Fluctuation*). Les courbes de la Figure A6-3 montrent les fluctuations de chaque atome du squelette peptidique des hélices transmembranaires dans les deux premiers vecteurs. Nous remarquons que dans ces vecteurs, les fluctuations peuvent fortement varier d'un peptide à un autre et selon l'environnement.

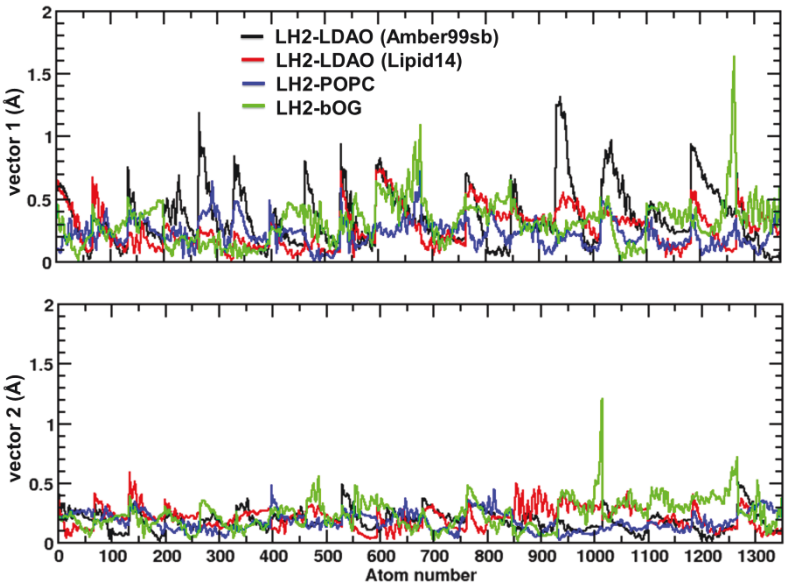


Figure A6-3 : RMSF des atomes du squelette peptidique des peptides α (1-594) et β (595-1350) dans les 2 premiers vecteurs et dans les différentes simulations.

La Figure A6-4 montre les configurations les plus extrêmes du squelette peptidique des hélices transmembranaires extraites à partir de l'ACP réalisée dans les différentes simulations.

Les flèches noires indiquent le sens du mouvement de chaque extrémité des hélices transmembranaires et nous montrent que ces mouvements sont opposés.

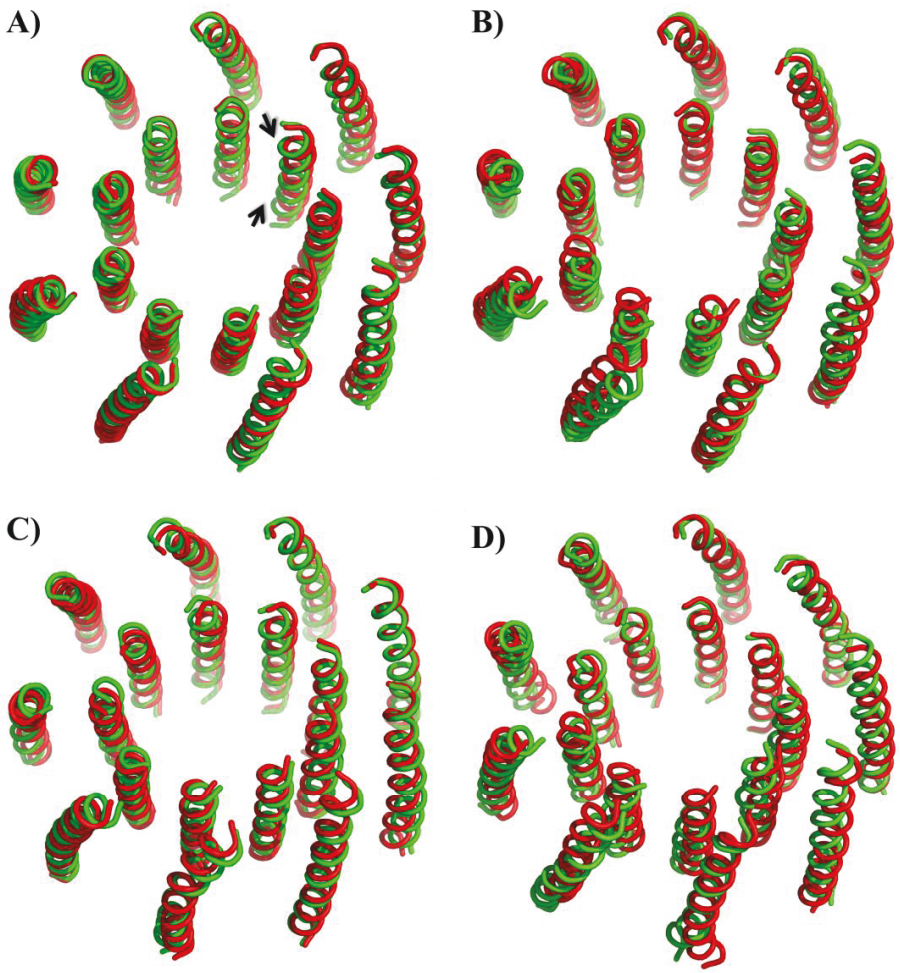


Figure A6-4 : Représentation de la première conformation (en vert) et de la conformation la plus extrême (en rouge) extraites de l'ACP dans la simulation du LH2-LDAO Amber99sb (A), LDAO Lipid14 (B), POPC (C) et bOG (D). Les deux flèches en noire montrent le sens du mouvement de chaque extrémité des hélices transmembranaires (qu'on retrouve dans les différents systèmes).

II. L'ACP des bactériochlorophylles-a

Comme pour le squelette peptidique, nous avons réalisé une ACP sur les 30 atomes lourds du macrocycle des B800 et B850.

- ***ACP du macrocycle des B800***

Les graphiques de la Figure A6-5 montrent l'espace conformationnel exploré par les atomes lourds du macrocycle des B800 dans les différents systèmes. Les deux premiers vecteurs décrivent le plus de variabilité : 29 et 11 % avec le LDAO Amber99sb, 22 et 15 % avec le LDAO Lipid14, 17 et 15 % avec le POPC, et 23 et 19 % avec le bOG. Nous remarquons que

les macrocycles forment deux groupes de conformations assez différents en présence de LDAO Amber99sb et de bOG comparé aux deux autres simulations.

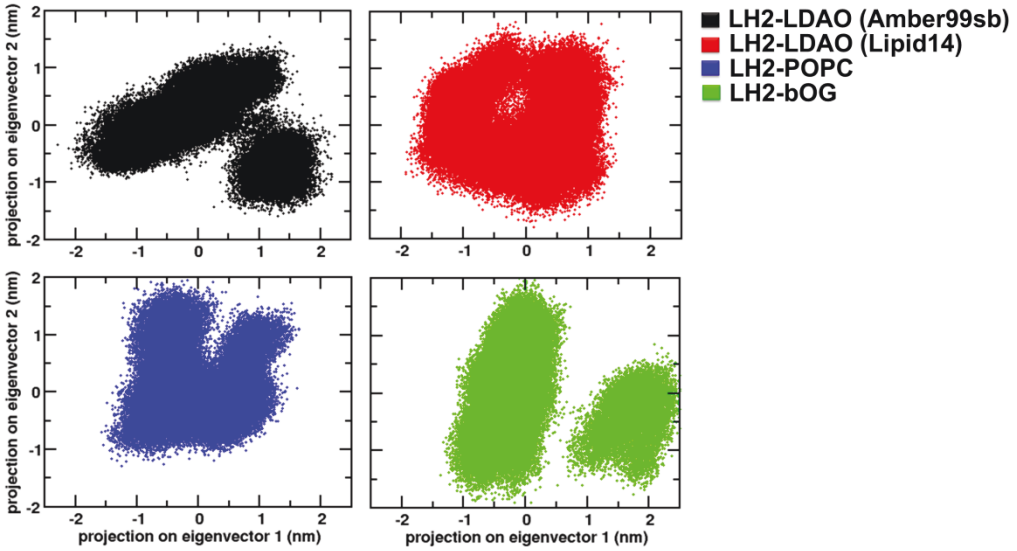


Figure A6-5 : Projection du vecteur 1 contre le vecteur 2 dans les différents systèmes pour visualiser l'espace conformationnel des atomes lourds du macrocycle des B800.

Nous pouvons aussi visualiser les fluctuations dans chaque vecteur grâce au RMSF (*Root Mean Square Fluctuation*). Les courbes de la Figure A6-6 montrent les fluctuations de chaque atome lourd des macrocycles B800 dans les deux premiers vecteurs. Nous remarquons que dans ces vecteurs, les fluctuations peuvent fortement varier d'un macrocycle à un autre.

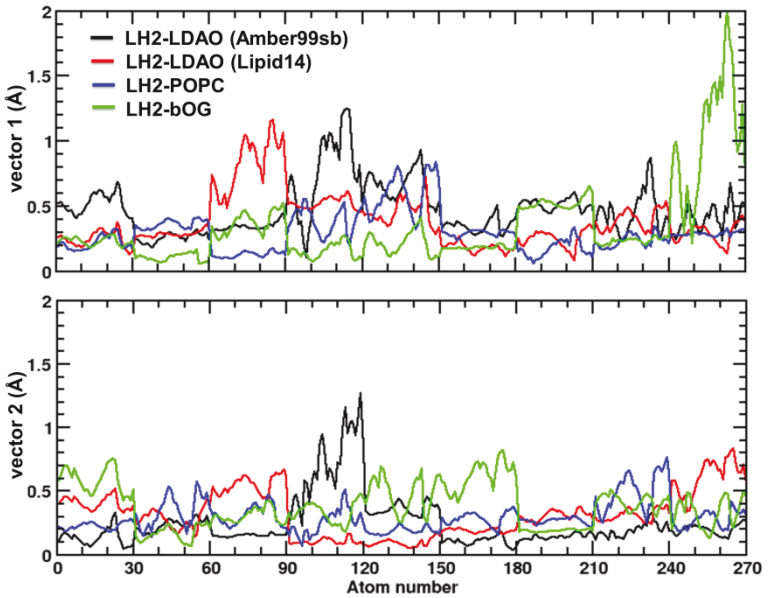


Figure A6-6 : RMSF des atomes du macrocycle des B800 dans les 2 premiers vecteurs et dans les différentes simulations. Les 30 premiers atomes correspondent aux atomes lourds du macrocycle du premier B800 et ainsi de suite jusqu'au neuvième B800.

La Figure 7 montre les configurations les plus extrêmes des atomes lourds du macrocycle des B800 extraites à partir de l'ACP réalisée dans les différentes simulations.

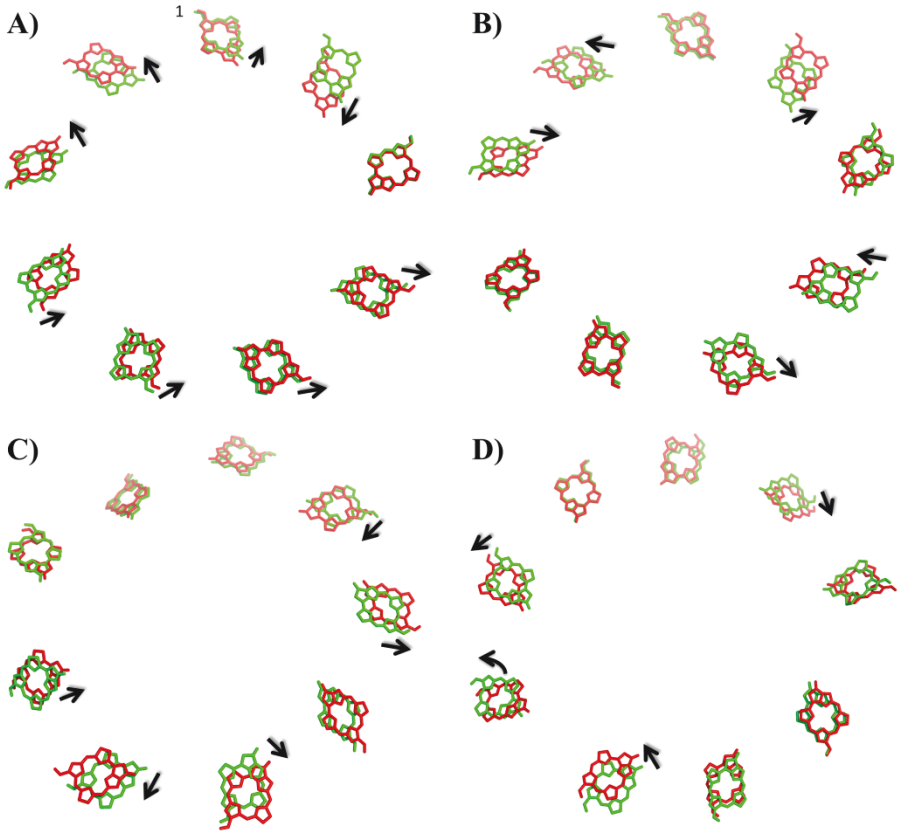


Figure A6-7 : Représentation de la première conformation (en vert) et de la conformation la plus extrême (en rouge) extraites de l'ACP dans la simulation du LH2-LDAO Amber99sb (A), LDAO Lipid14 (B), POPC (C) et bOG (D). Les deux flèches en noire montrent le sens du mouvement macrocycle. Le premier B800 est indiqué par le numéro 1, puis la numération suit le sens des aiguilles d'une montre.

- ***ACP du macrocycle des B850***

Dans le même esprit, les graphiques de la Figure A6-8 montrent l'espace conformationnel exploré par les atomes lourds du macrocycle des B850 dans les différents systèmes. Les deux premiers vecteurs décrivent le plus de variabilité : 21 et 8 % avec le LDAO Amber99sb, 14 et 10 % avec le LDAO Lipid14, 13 et 11 % avec le POPC, et 15 et 10 % avec le bOG. Nous remarquons que globalement le macrocycle des B850 explore beaucoup de conformations mais il reste un peu plus stable comparé au macrocycle des B800 (Figure A6-5). Les nuages de points s'étendent un peu plus dans le cas du POPC et bOG ce qui nous suggère qu'avec ces environnements, le macrocycle des B850 peut explorer d'autres conformations comparé aux environnements avec du LDAO.

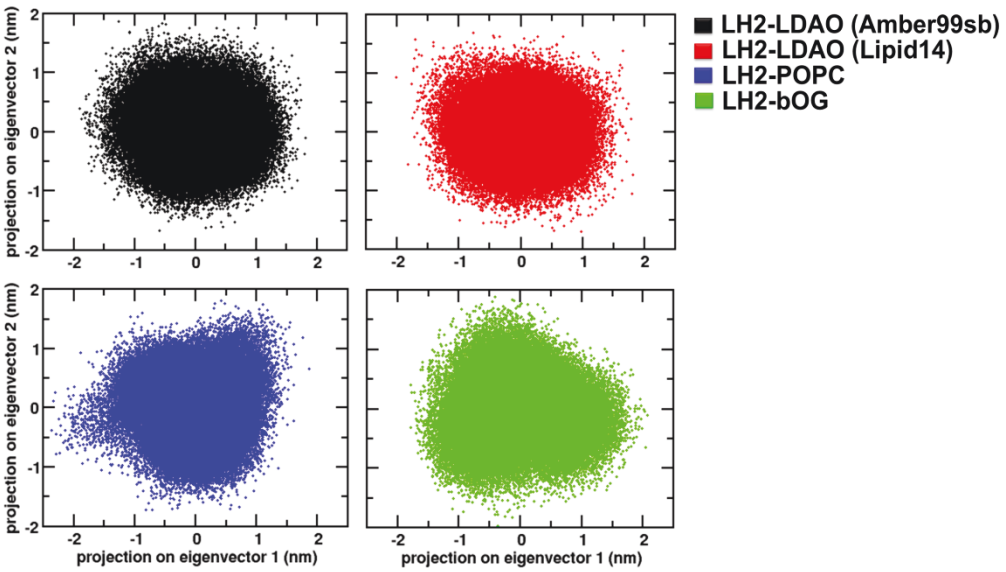


Figure A6-8 : Projection du vecteur 1 contre le vecteur 2 dans les différents systèmes pour visualiser l'espace conformationnel des atomes lourds du macrocycle des B850.

Les courbes de la Figure A6-9 montrent les fluctuations de chaque atome lourd des macrocycles B850 dans les deux premiers vecteurs. Nous remarquons que dans ces vecteurs, les fluctuations peuvent légèrement varier d'un macrocycle à un autre comparé aux fluctuations des atomes lourds du macrocycle des B800.

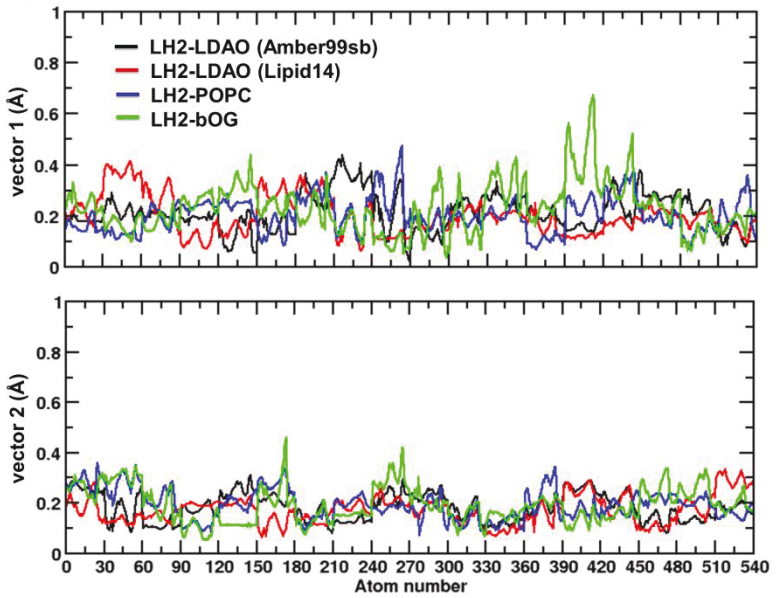


Figure A6-9 : RMSF des atomes du macrocycle des B850 dans les 2 premiers vecteurs et dans les différentes simulations. Les 30 premiers atomes correspondent aux atomes lourds du macrocycle du premier B800 et ainsi de suite jusqu'au seizième B850.

La Figure A6-10 montre les configurations les plus extrêmes des atomes lourds du macrocycle des B850 extraites à partir de l'ACP réalisée dans les différentes simulations.

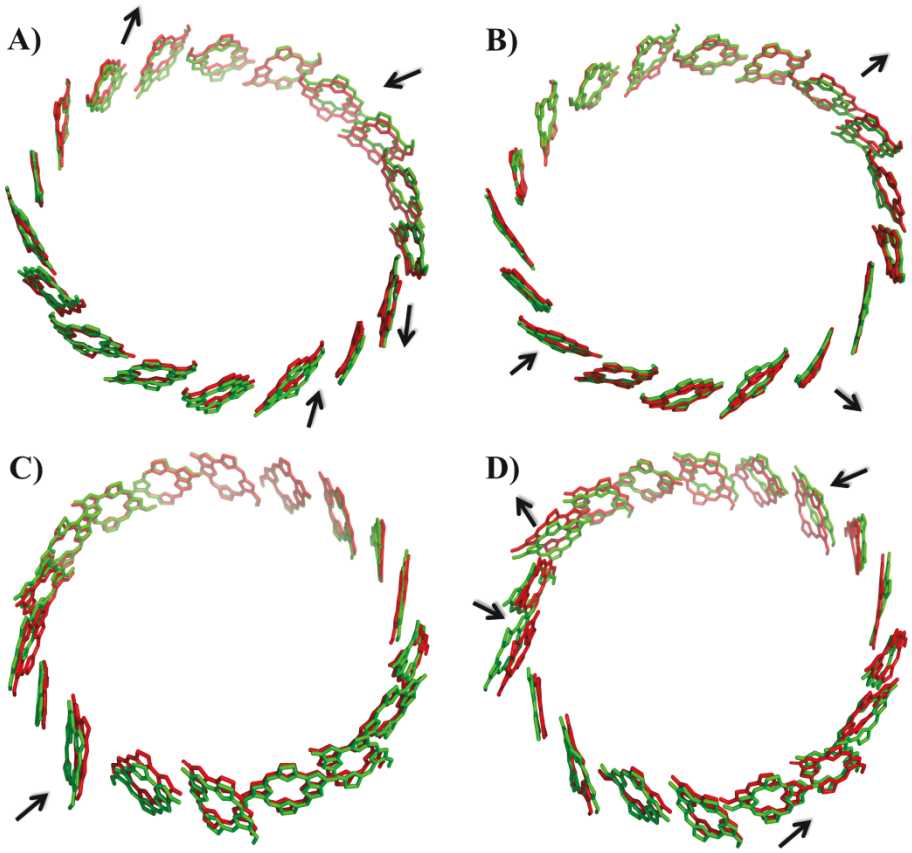


Figure A6-10 : Représentation de la première conformation (en vert) et de la conformation la plus extrême (en rouge) extraites de l'ACP dans la simulation du LH2-LDAO Amber99sb (A), LDAO Lipid14 (B), POPC (C) et bOG (D).

Annexe 7 : Participation à d'autres travaux

Dans cette annexe, nous introduirons les articles issus des collaborations que nous avons eues dans d'autres projets de recherche et dans lesquelles j'ai pu participer.

- ***Article 1 : travail sur la structure micellaire des sophorolipides*** (Manet *et al.* 2015)

Les sophorolipides sont des biosurfactants principalement produits par des microorganismes et font partie des glycolipides bolaamphiphile (amphiphile bipolaires). La distribution atypique des régions amphiphiles et lipophiles donne lieu à des agrégations assez intéressantes de différentes formes : micelles, bicelles, vésicules, cônes, fibres, tubes etc. Beaucoup d'études ont été réalisées sur la synthèse et la caractérisation des systèmes « bola », mais peu se sont intéressées à l'étude structurale. Dans ce contexte, nous avons analysés la structure micellaire des sophorolipides en combinant l'expérimental et la théorie avec des expériences de SAXS, SANS et des simulations de dynamiques moléculaires.

Les résultats obtenus ont montré que les systèmes non-chargés (pH < 5) avaient une forme ellipsoïde prolate. Comparé à la plupart des systèmes de surfactants, l'épaisseur de la couche hydrophile est pratiquement de 0 Å dans la direction axiale de l'ellipsoïde et d'environ 12 Å dans la coupe transversale, formant ainsi un « grain de café ». Nous supposons que la micelle de sophorolipides peut se décomposer en 3 domaines : un cœur hydrophobe, une couche extérieure hydrophile et une région axiale hydrophile/hydrophobe. Ces résultats nous indiquent que les sophorolipides acquièrent différentes configurations (courbée et linéaire) au sein d'une micelle.

Enfin, des expériences d'*anomalous* SAXS (ASAXS), pour la spécification chimique de la structure étudiée, ont été réalisées pour investiguer la distribution des groupements COO^- (5 < pH < 7). Le faible signal de l'ASAXS montre que les groupements COO^- sont répartis de manière diffuse dans la grande couche hydrophile plutôt qu'à l'interface micelle-eau.

- ***Article 2 : travail sur les propriétés structurales et dynamiques de micelles inverses de DOPC*** (Abel *et al.* soumis à Langmuir)

Dans cet article, nous nous sommes intéressés aux micelles inverses (MI) de 1,2-dioleoyl-sn-phosphatidylcholine (DOPC). Les MI sont des gouttelettes d'eau dispersées dans une phase organique. Le contenue en eau est un paramètre expérimentalement contrôlable qui dépend de la fraction eau /détergent (ici W_0 = [H_20]/[DOPC]). Du fait de cette propriété intéressante, les MI sont des systèmes modèles très intéressants pour étudier l'effet du confinement et de l'hydratation sur les propriétés structurales et dynamiques de biomolécules confinées (par exemple l'équilibre dépliement/repliement des protéines). Nous avons examiné la structure des micelles de DOPC dans le benzène à l'aide des simulations de dynamique moléculaire. Nous avons d'abord construit des modèles de DOPC avec W_0 = 1 – 16 dont la dimension est en accord avec les données de diffusion de la lumière. L'influence de la quantité d'eau dans les MI sur la structure (par exemple forme et taille) des micelles, l'hydratation de la tête polaire du lipide et la dynamique de l'eau ont été comparés aux résultats obtenues avec une bicouche en phase liquide cristal de DOPC. Nous montrons, en particulier, que la forme des micelles change significativement avec la valeur de W_0. L'hydratation des groupes polaires du DOPC augmente progressivement avec W_0 jusqu'à arriver à un taux d'hydratation du DOPC similaire à celui trouvé dans la membrane. En ce qui concerne les caractéristiques dynamique de l'eau confinée, nous observons que celles-ci tendent vers celles de la membrane et de l'eau pure pour les MI les plus grandes.

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Structure of Bolaamphiphile Sophorolipid Micelles Characterized with SAXS, SANS, and MD Simulations

Sabine Manet,[†] Anne-Sophie Cuvier,[†] Claire Valotteau,[†] Giulia C. Fadda,[‡] Javier Perez,[§] Esra Karakas,[⊥,∥] Stéphane Abel,[∥] and Niki Baccile*,[†]

[†]Sorbonne Universités, UPMC Univ Paris 06, CNRS, Collège de France, UMR 7574, Chimie de la Matière Condensée de Paris, F-75005 Paris, France

[‡]Laboratoire Léon Brillouin, LLB, CEA Saclay, F-91191 CEDEX Gif-sur-Yvette, France

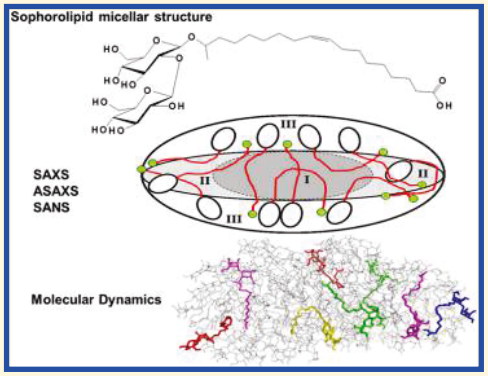
[§]SWING, Synchrotron Soleil, BP 48, F-91192 Gif-sur-Yvette, France

[⊥]Maison de la Simulation, USR 3441, CEA − CNRS − INRIA, Univ. Paris-Sud − Univ. de Versailles, 91191 CEDEX Gif-sur-Yvette, France

[∥]Institute for Integrative Biology of the Cell (I2BC), Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Centre National de la Recherche Scientifique (CNRS), Université Paris-Sud, CEA-Saclay, F-91191 Gif-sur-Yvette, France

Supporting Information

ABSTRACT: The micellar structure of sophorolipids, a glycolipid bolaamphiphile, is analyzed using a combination of small-angle X-ray scattering (SAXS), small-angle neutron scattering (SANS), and molecular dynamics (MD) simulations. Numerical modeling of SAXS curves shows that micellar morphology in the noncharged system (pH< 5) is made of prolate ellipsoids of revolution with core−shell morphology. Opposed to most surfactant systems, the hydrophilic shell has a nonhomogeneous distribution of matter: the shell thickness in the axial direction of the ellipsoid is found to be practically zero, while it measures about 12 Å at its cross-section, thus forming a "coffee bean"-like shape. The use of a contrast-matching SANS experiment shows that the hydrophobic component of sophorolipids is actually distributed in a narrow spheroidal region in the micellar core. These data seem to indicate a complex distribution of sophorolipids within the micelle, divided into at least three domains: a pure hydrophobic core, a hydrophilic shell, and a region of less defined composition in the axial direction of the ellipsoid. To account for these results, we make the hypothesis that sophorolipid molecules acquire various configurations within the micelle including bent and linear, crossing the micellar core. These results are confirmed by MD simulations which do show the presence of multiple sophorolipid configurations when passing from spherical to ellipsoidal aggregates. Finally, we also used Rb^+ and Sr^{2+} counterions in combination with anomalous SAXS experiments to probe the distribution of the COO^- group of sophorolipids upon small pH increase (5 < pH < 7), where repulsive intermicellar interactions become important. The poor ASAXS signal shows that the COO^- groups are rather diffused in the broad hydrophilic shell rather than at the outer micellar/water interface.



INTRODUCTION

Surfactants are compounds employed in a very large panel of applications in which amphipathic interfaces are made compatible. The formation of micelles above a critical micellar concentration is one of the most common aspects among amphiphilic molecules in water, and a thorough description of their structure is one of the first steps when studying the properties of new compounds.[1] Although the aggregation of classical cationic (e.g., alkylammonium salts),[2−4] anionic (e.g., alkylsulfates salts),[4] or nonionic (e.g., block copolymers)[5] surfactants into micelles has been studied, both experimentally and theoretically, for decades, less conventional compounds are much less studied. Among these, bolaform lipids[6] constitute a class of interesting molecules with two functional end-groups. If the end-groups are chemically equivalent, one speaks of symmetric bolaamphiphiles (bolas); otherwise, one refers to asymmetric bolas. The atypical distribution of amphiphilic and lypophilic regions within the same molecule is often responsible for interesting aggregation phenomena which lead to the formation of micelles, bicelles,[7] vesicles,[8] cones,[9] fibers,[10] tubes,[11,12] and so on. Many significant systems have been reviewed in 2004 by Fuhrhop,[6] whereas Fariya et al.[13] have recently reviewed bolaamphiphiles in the context of pharmaceutical applications. A large amount of work has been dedicated to the synthesis and characterization of bola systems; however, much less work has been devoted to a thorough

Received: June 5, 2015
Revised: September 16, 2015
Published: October 2, 2015

ACS Publications © 2015 American Chemical Society

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

structural study. In the context of micellar objects composed of bolaamphiphiles,[8,10,14−20] advanced structural considerations of the micellar packing and bola distribution are rare.[19,20] Indeed, Nagarajan[21] predicted the formation and structure of micelles composed of bolaamphiphiles, basing his hypotheses on the analysis of the packing parameter. Interestingly, he predicted the formation of spherical, cylindrical, and discoidal micelles, in which he imagined that the bolaform compound has two configurations: elongated, which induces transradial crossing of the end-groups; folded, under which the end-groups lie both on the same side. Yan et al.[14] have made progress with macroscopical verification of Nagarajan's assumptions by observing a pH/cosolvent-driven tube-to-vesicle-to-micelle transition in disodium phenyl-1,4-bis(oxyhexanoate) solutions, although the actual bola distribution within the micelle was never elucidated. Additional insights in the structural composition of bola-composed micelles have been discussed in refs 7, 14−20, mainly using small-angle neutron scattering (SANS) and/or self-diffusion nuclear magnetic resonance (NMR) experiments. However, in practically all cases, these experiments have been used to describe the micellar shape and size. Only few authors, with little consensus in their conclusions, have tried to determine the molecular distribution of the bolaamphiphile within the micellar aggregate. Caponetti et al.,[16] for instance, have used advanced modeling analyses of SANS spectra to study the bola *N*-aza-18-crown-6 ether compound. Although Davey et al.[19] and Shinde et al.[20] agree on the fact that their respectively studied asymmetric bolaamphiphiles, alcohol derivatives of alkylammonium and oleate salts, form uniform spherical/cylindrical micelles in which both polar groups are in contact with the solvent (water), they do not agree on the molecular arrangements. Davey proposes an elongated cross-micellar conformation, whereas Shinde proposes a bent configuration. Interestingly, these recent studies disagree with older works in which it was proposed that only one of the two polar groups is actually located at the micellar/solvent palisade.[22,23]

Such a lack of deep characterization of bola-micelles and disagreement on their structure and molecular conformation is quite odd because of the importance in terms of applications for many bola systems. In fact, the chemical nature of end-groups in bolas can be as wide as imagination can allow: polymerizable,[24] complexing,[16,25] pH-responsive,[11,14] fullerenic,[26] chiral,[27] and glycosidic[18] functions are just some examples of the broad variety that one can find in the literature. The range of applications of bolas in general goes from one-pot porous material synthesis[24] to supramolecular materials with temperature-driven ferroelectric−paraelectric transitions,[25] ion-channel supramolecular membranes,[28] metallo-hydrogels for dye adsorption and water purifier,[16,29] viscoelastic fluids,[30] biomedical carriers,[13] and many more. For this reason, it is of paramount importance to characterize the spatial distribution and accessibility of end-functions within the supramolecular aggregates and, in particular, within micelles. Laying one of the end-groups within the interior of the micellar aggregate can be of relevance for drug-delivery systems and conception of nanoreactors. Whether charged end-groups concentrate at the micellar surface or not can be of relevance in nanomaterial synthesis procedure in which charge matching drives the material formation, such as mesostructured oxides or layer-by-layer approaches, but it can also have drastic relevance in gelling properties, for instance. The distribution and accessibility of biologically relevant moieties like peptides or carbohydrates is important for biomedical applications like gene-antigene recognition for cancer treatment. Even in the system studied by Caponetti,[16] in which authors looked at metal-binding affinity of the crown ether end-groups, it is necessary to know the distribution of the crown ether groups and their accessibility in view of metal ion removal in solution.

For these reasons, in the present work, we try to set up a combination of complementary analytical approaches based on advanced small-angle scattering techniques in order to prove both indirectly (by way of model functions) and directly the spatial distribution of a glycolipid asymmetric bolaamphiphile. We study here the micellar structure of sophorolipids, a yeast-derived bolaform glycolipid composed of sophorose (glucose $\beta(1,2)$) attached to the C17 atom of oleic acid via acetal bond, the carboxylic group being free of access at the opposite side of the molecule (Figure 1).[31] This compound is part of the glycolipid family and the presence of sophorose-COOH end-groups make it, de facto, a bolaform compound with pH-responsive properties,[32] which confer to it a double neutral/anionic or neutral/neutral nature.[33] Recent works on the self-assembly of the acidic open chain form of sophorolipids have shown its ability to form assemblies such as fibers[34,35] and micelles.[32,36] In the latter case, our group has shown that the micellar charge can be adjusted using pH.[33] However, no information on the charge distribution could be obtained. Considering the fact that this compound is easily accessible and considering its importance in various fields like detergency,[37] cosmetics,[38] nanoscience,[39] disinfection,[40] medicine,[41] it becomes a valuable candidate to use for a detailed study of the sophorose/COOH end-groups distribution within the micelle. To do so, we combine several small-angle scattering techniques: small-angle X-ray scattering (SAXS) is used to precisely define the micellar morphology and its size; anomalous-SAXS (ASAXS), a technique that probes the distribution of counterions around a charged macroion, is employed to follow the distribution of the carboxilate end-group in the presence of specific ASAXS probe counterions, Rb^+ and Sr^{2+}; contrast-matching SANS is used to visualize the distribution of the hydrophobic fraction of the molecule. This approach shows that, in the mild acidic pH region (pH< 5), the micelle is likely to be described as an ellipsoid of revolution having a "coffee bean"-like morphology. In this structure, the hydrophobic region only occupies a small inner spheroidal core. The outer shell in the central region of the ellipsoid is mainly made of sophorose and COOH groups. The electron density distribution in the axial direction is more difficult to define, and it probably reflects a more complex distribution of matter. We identify at least two different configurations of the sophorolipid molecules: bent and elongated, with the possibility of crossing the ellipsoid section. We also show that these hypotheses, based on experimental work, are confirmed by molecular dynamics (MD) simulations specifically performed in this work. Upon formation of COO^- groups in the neutral pH range (5 < pH < 7), micelles become negatively charged: ASAXS experiments seem to indicate that the carboxylates are rather diffused in the broad hydrophilic shell rather than localized at the micelle/solvent palisade. To the best of our knowledge, such a complex analysis has never been carried neither on a glycolipid-based compound nor on any other bolaamphiphile compound.

■ EXPERIMENTAL SECTION

Acidic sophorolipids (SL) have been prepared from a commercial batch of a sophorolipid mixture (Sopholiance,

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

Soliance, France; sophorolipids are derived from rapeseed oil, batch number: 11103A, dry content: 60 ± 6%) using alkaline hydrolysis to convert the lactonic/acidic mixture into a fully acidic sophorolipid compound. Extraction and purification have been performed using the method N°2 described in ref 42, to which one should refer for a typical ^{1}H solution NMR spectrum of the compound used. As a general observation on purity, one should note that our compound has a very low content of residual fatty acids (typically less than 5 mol % by ^{1}H NMR). Nevertheless, biobased glycolipids in general are known to contain residual congeners. In the case of sophorolipids, mainly composed of a C18:1 fatty acid tail attached to sophorose, congeners generally represent a low fraction of the actual compound, generally being C16 and/or C18:2 fatty acids variations. A typical HPLC chromatogram was reported in ref 43, showing that C18:1 does constitute the large majority of our sample (>90%), thus being in agreement with the ^{1}H NMR data. Congeners are also detected, but in the present work, we make the hypothesis that, given the low amount, their influence on the self-assembly of C18:1 sophorolipid is negligible.

Sophorolipid solutions are generally prepared by simply dispersing the compound in water (50 mg/mL = 80 mM). If needed, a slightly warm (30−35 °C) ultrasonic bath is used for few seconds to help solubilize the sample. After solubilization, the solutions are clear and stable over time. Specific sample preparation is detailed in each subsection below. Milli-Q quality water was used for the experiments, and all solutions are freshly prepared. RbCl, $SrCl_2$, RbOH, and $Sr(OH)_2$ have been purchased at Sigma-Aldrich.

Small-Angle X-ray Scattering (SAXS). The sample composition is presented in Table 1 and for all of them the concentration of SL is 50 mg/mL. Rb^+ and Sr^{2+} have been used instead of the more classical Na^+ and Ca^{2+} ions due to their activity in the ASAXS experiments, presented later. Each sample is prepared by mixing the salt in water; this solution is eventually split in half: one-half is used for background acquisition, and the indicated amounts of SL are introduced in the second half.

Table 1. List of Samples Studied in This Work[a]

sample	compound	[Sr^{2+}] mM	[Rb^+] mM
1	RbCl	-	10
2	RbCl	-	50
3	RbCl	-	100
4	RbOH	-	40
5	RbOH	-	80
6	$SrCl_2$	5	-
7	$SrCl_2$	25	-
8	$SrCl_2$	50	-
9	$Sr(OH)_2$	10	-
10	$Sr(OH)_2$	20	-

[a]Sophorolipid concentration is constant (80 mM) for all samples.

These experiments have been done at the SWING beamline of the SOLEIL synchrotron at Saint-Aubin (France). Sample-to-detector distance was set to 2526 mm, and the energy of the X-ray beam was E = 14.6 keV. The experimental environment was provided at the beamline and consisted of a flow-through quartz capillary, which it was possible to fill using a syringe. The X-ray beamline hits the same capillary position, which does not change between background and sample acquisitions, thus making the background subtraction highly accurate. For the background, we used the same solution used to prepare a given sophorolipid sample, as also mentioned above. Absolute scaling was calibrated on the signal of water. All data have been divided by the sample transmitted intensity. Data treatment (mask generation, integration) was done using the Foxtrot software provided at the beamline.

Modeling SAXS Data. The analysis of the SAXS profiles for the determination of micelle size and shape parameters has been made with the models displayed by default in the SASview software, provided free of charge on the developer's Web site.[44] The general equation treating the scattered intensity versus the momentum transfer q is eq 1

$$I(q) = \frac{\text{scale}}{V}(\rho - \rho_{\text{solv}})^2 P(q) S(q) + \text{bkg} \tag{1}$$

where scale is the volume fraction, V is the volume of the scatterer, ρ is the scattering length density (SLD) of the object, ρ_{solv} is the SLD of the solvent, $P(q)$ is the form factor of the object, $S(q)$ is the structure factor and bkg is a constant accounting for the background level.

Form Factor. The analytical expressions of the $P(q)$ discussed below are directly implemented in the SASview software and their analytical form is provided in ref 45. The $\chi 2$ test is employed to characterize the quality of the fit, the lower value indicating the best fit. χ^2 is defined as the Pearson's cumulative test statistic, $\chi^2 = \sum_{i=1}^{\text{Npts}}((O_i - E_i)^2/E_i)$, where O_i is an observed measured value; E_i is an expected (theoretical) value, asserted by the null hypothesis. The χ^2 is then normalized by the number of points in the distribution, χ^2/Npts. Some fit parameters can be simply estimated from the sample composition and have been fixed for the fitting process. Among them, the volume fraction of scattering objects has been set at 0.05, reflecting the SL mass concentration of 50 mg/mL (80 mM) used for all samples.

The core and solvent SLD's have been fixed and calculated using eq 2:

$$\rho = \frac{\sum_i^j Z_i r_e}{V_M} \tag{2}$$

where Z_i is the atomic number of the ith of j atoms in a molecule of molecular volume V_M and r_e is the classical electron radius or Thomson scattering length (2.8179×10^{-15} m). Sophorolipids are mainly composed of a C18 aliphatic chain with one unsaturated C═C bond (position $C_{9,10}$), one COOH group (position C_1), the CH in position C_{17} covalently bonded to sophorose, and a CH_3 in position C_{18} (Figure 1). Considering the fact that C_{17} and C_{18} atoms are located at the sophorose/aliphatic frontier, it is a difficult choice to decide which is their exact location. For this reason, we simplify our model and make the assumption that the hydrophobic core starts at the C_{16} carbon, neglecting the contribution of the C_{17} and C_{18} positions. Under these conditions, we define the molecular volume, V_M, of the hydrocarbon chain to be 13 $V_{CH_{2,mic}} + 2V_{CH}$, where $V_{CH_{2,mic}}$ is the volume of one CH_2 group inside a micelle, 27.5 Å^3, and V_{CH} is the volume of a CH group, 22.0 Å^3.[46] These estimations provide a SLD core value, constant for all fitting procedures in this study, of 8.3×10^{-6} Å^{-2} (294 e$^-$/nm^3). The solvent SLD has been calculated taking into account the salt or base concentration influencing the molecular volume.[47] The resulting SLD values are very close and lie between 9.4 and 9.5×10^{-6} Å^{-2}, (334 and 337 e$^-$/nm^3, see full list in Table S1 in the Supporting Information). For

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

core–shell models, the shell SLD is always a variable parameter. We make the hypothesis that the shell is most likely composed of hydrated sophorose; in this case, the shell SLD value should lie between the solvent and hydrated sophorose SLD values, roughly between 1.0 and 1.2 × 10^{-5} Å^{-2}, (355 to 425 e$^-$/nm^3). As a result, the electron density profile from the center of the micelle to the solvent can be drawn as a hat-like shape, in which the shell SLD is the highest value. The choice of the form factor has been determined by comparing the best fits on the basis of the lowest χ^2/Npts among different model functions, as critically discussed on Page S2 in the Supporting Information and shown in Figure S1. We find that a core–shell prolate ellipsoid of revolution, schematized in Figure 2, best describes our SAXS data.

Structure Factor. The structure factor, $S(q)$, which quantifies the intermicellar interactions/correlations, is included in our fitting process using the "Hayter-MSA-Structure" model displayed in the SASview software. This accounts for a repulsive screened Coulombic intermicellar interaction potential,[48,49] as previously used on similar systems.[33] $S(q)$ has only been used for samples containing RbOH (samples 4,5) and $Sr(OH)_2$ (samples 9,10) according to the previously discussed assumption that deprotonation of the COOH introduces negative charges at the micellar/solvent palisade,[33] thus being responsible of the broad interaction peak observed in small-angle neutron scattering spectra and reproduced in this work. $S(q)$ has been set equal to 1 for base-free samples only containing salts (samples 1–3 and samples 6–8 in Table 1). Six parameters are needed to compute the Hayter and Penfold structure factor: the dielectric constant, the volume fraction, the effective radius of interacting objects, the temperature, the surface charge of the micelle, and the salt concentration, used to calculate the ionic strength of the solution, which in turn is used to compute the Debye screening length. As this model is made for monovalent ions, we modified the salt concentration parameter of the samples, otherwise fixed and equal to the experimental concentration, with divalent ions so that the resulting Debye length is correct. The volume fraction of the scatterers is calculated from V_{mic} and n_{mic} ($\eta = V_{mic}n_{mic}$), defined more precisely later on in this same section, and inserted in the Hayter-MSA-Structure routine with R_E, effective radius of interacting objects ($R_E = (1/2)\ ((6\eta/\pi\eta_{mic}))^{1/3}$, see ref 48) as fixed parameters. The temperature, dielectric constant, and salt concentration are also fixed parameters so that the micelle charge is the only refinable parameter for the calculation of $S(q)$.

Fit Strategy Using a Core–Shell Prolate Ellipsoid of Revolution Model and a Hayter-MSA Structure Factor. In order to reduce the number of parameters used for the fitting process, we used a customized version of the core–shell prolate ellipsoid form factor proposed in the 3.0.0 version of SASview. In particular, we introduce some molecular constraints inspired by the work of Hayter and Penfold.[4] By making some assumptions described thereafter, only five fitting parameters are needed: N_{agg} (aggregation number), R_C (equatorial core radius, see Figure 2), ρ_S (scattering length density of the shell, see Figure 2), X_C (axial core ratio, see Figure 2) and z, the surface charge of the micelle. A small background correction (<0.005 cm^{-1}) has also been applied. In the Hayter and Penfold model, if N_{agg} is greater than the one expected to accommodate a sphere, the radius of which is $R_C = l_c$ (length of the fully extended hydrocarbon chain of the surfactant), then the object is allowed to become elliptical so to accommodate the molecular excess. This model has been adapted to the sophorolipid surfactants as follows. For a given N_{agg}, the spherical hard core micelle radius R_{hc} is defined as

$$R_{hc} = \left[\frac{3N_{agg}(13V_{CH_{2,mic}} + 2V_{CH})}{4\pi}\right]^{1/3} \tag{3}$$

If $R_{hc} > l_c$, R_C is set to l_c, and a parameter α is introduced, defined as the fraction of CH_2 inserted in the dry core of the micelle:

$$\alpha = \frac{V_C - 2V_{CH}}{N_{agg}13V_{CH_{2,mic}}} \tag{4}$$

with $V_C = (4/3)\ \pi R_C^3$. If $R_{hc} < l_c$, R_C is set to R_{hc} and $\alpha = 1$.

Sophorolipids are bolaform amphiphiles with two different polar heads (carboxylic acid and sophorose sugar moiety) connected to a 18:1 carbons lipid chain (Figure 1), which reduces to a C16:1 chain if one makes the hypothesis that C_{17} (CH) and C_{18} (CH_3) do not contribute to l_c, which in this case can be estimated to be around 10 Å, half the length of the 16 carbons chain. We first try to use this value of l_c as R_{test}, but the results obtained were not reliable, leading to high χ^2/Npts values, so we let this constraint to relax, and R_C and X_C have been set as refinable parameters. Recently, Penfold et al.[36] have made the same choice when analyzing the SANS profiles of sophorolipid micelles, arguing that the particular alkyl chain geometry of the sophorolipid introduces some uncertainty into what value to which the inner radius should be constrained.

Once α is known, the number of water molecules per sophorolipid in the shell, $n_{w/SL}$, and consequently the composition of the shell, can be calculated using the shell scattering length density ρ_s as a refinable parameter, assuming that the shell hydration should not be a fixed parameter, because it depends on many effects like the number of hydrated CH_2 or the micelle surface charge. From eq 5, we calculate the molar volume, $V_{m,hg}$, and hence the SLD, ρ_{hg}, of the total dry headgroup. Composition of the dry headgroup is not easy to estimate; it certainly includes sophorose, as stated earlier, and most likely the COOH group, but it can also include the C_{17} (CH) and C_{18} (CH_3) atoms, previously excluded from the core region of the micelle, as well as the fraction of the CH_2 aliphatic chain excluded by the α parameter presented in eq 4. This is not surprising as hydration of the CH_2 groups close to the hydrophilic head was reported before both experimentally[50,51] and predicted theoretically using MD simulation.[52] In all cases, the ionic species coming from the added salts or bases are neglected. The general expressions for the headgroup volume and SLD are then:

$$V_{m,hg} = \sum_i V_i + 13(1-\alpha)V_{CH_{2,w}} \tag{5}$$

$$\rho_{hg} = \frac{\sum_i \rho_i V_i + 13(1-\alpha)\rho_{CH_{2,w}} V_{CH_{2,w}}}{V_{m,hg}} \tag{6}$$

where i denotes the sophorose, carboxylate and the aliphatic moieties, CH, CH_3 and possibly CH_2 groups. $V_{CH_{2,w}}$ is the molar volume of hydrated CH_2 smaller than $V_{CH_{2,mic}}$, taken as 26.7 Å^3,[53] while $\rho_{CH_{2,w}}$ is the SLD of the hydrated CH_2. The volume fraction of water in the shell $x_{v,w}$ and the number of water molecules per sophorolipid in the shell $n_{w/SL}$ are obtained by

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113–13133

considering the fraction of water molecules ($n_{w/SL}$) needed to equalize the calculated ρ_{hg} eq 6 and the experimental ρ_{shell} (obtained from the fit). This is summarized in eq 7 and eq 8:

$$x_{v,w} = \frac{\rho_{shell} - \rho_{hg}}{\rho_w - \rho_{hg}} \tag{7}$$

$$n_{w/SL} = \frac{x_{v,w} V_{m,hg}}{(1 - x_{v,w}) V_{m,w}} \tag{8}$$

with $V_{m,w}$ the molar volume of water (29.9 Å^3) and ρ_w the water SLD taken as 9.4×10^{-6} Å^{-2}. As it did not seem straightforward to evaluate the ionic composition and concentration within the shell, the presence of ions coming from added salts, acid or base has been neglected. However, the same fitting process applied using ρ_{solv} instead of ρ_w in eq 7 did not show a deep impact of the ionic composition of the shell on the results (results not shown).

The total micelle volume V_{mic} is then calculated as the sum of all the components in the core and in the shell:

$$V_{mic} = N_{agg}(13\alpha V_{CH_2,mic} + 2V_{CH} + V_{m,hg} + n_{w/SL} V_{m,w}) \tag{9}$$

As V_{mic} can also be defined as the volume of an ellipsoid described by the equatorial core radius R_C, the axial core ratio X_C, the axial shell ratio X_S (see Figure 2), and the equatorial shell radius R_S [$V_{mic} = (4/3)\ \pi R_S^2(R_C X_C + (R_S - R_C)X_S)$, see Figure 2], the latter geometric parameter can then be obtained as the real root of this polynomial equation:

$$X_S R_S{}^3 + R_C(X_C - X_S)R_S{}^2 - 3\frac{V_{mic}}{4\pi} = 0 \tag{10}$$

The core−shell prolate ellipsoid form factor is finally calculated by

$$P(q) = n_{mic}\int_0^1 |F(q, \mu)|^2 d\mu + \text{background}$$

$$F(q, \mu) = V_C(\rho_C - \rho_S)\left[\frac{3j1(x_C)}{x_C}\right] + V_{mic}(\rho_S - \rho_{solv})\left[\frac{3j_1(x_S)}{x_S}\right] \tag{11}$$

$$j_1(x) = \frac{\sin x - x\cos x}{x^2} \tag{12}$$

$$x_C(q, \mu) = q(R_C{}^2 X_C{}^2 \mu^2 + R_C{}^2(1 - \mu^2)) \tag{13}$$

$$x_S(q, \mu) = q(R_S{}^2 X_C{}^2 \mu^2 + R_S{}^2(1 - \mu^2)) \tag{14}$$

The intensity scattered by interacting micelles defined as $I(q) = n_{mic}P(q)S(q)$ can be calculated by relating the micelle density n_{mic} to the known sophorolipid surfactant concentration [SL] (80 mM) and the refinable aggregation parameter N_{agg}, with N_a being the Avogadro's number:

$$n_{mic} = \frac{[SL]N_a}{N_{agg}} \tag{15}$$

Anomalous Small Angle X-ray Scattering (ASAXS). *Elements of Theory.* Extraction of the counterion scattering contribution around sophorolipid micelles was done via recording the anomalous scattering contribution of rubidium and strontium cations. The anomalous effect is measured when the energy of the X-ray beam approaches the absorption edge of a given element. In this case, the scattering factor becomes a complex function and it is no more a constant. The contribution of the scattering factor far from the absorption edge is

$$f_{ion} = f_0 - \rho_m V_{ion} \tag{16}$$

where f_0 equals the atomic number of the element, ρ_m is the electron density of water and V_{ion} is the volume of the counterion. Upon approaching of the edge, f_{ion} becomes

$$f_{ion} = f_0 - \rho_m V_{ion} + f'(E) + if''(E) \tag{17}$$

where $f'(E)$ and $f''(E)$ are, respectively, the real and imaginary part of the energy-dependent scattering vector. Values for $f'(E)$ and $f''(E)$ are tabulated,[54,55] but they can also be calculated from the experimental absorption spectrum.[56] When approaching the absorption edge by tuning the synchrotron incident beam energy, f_{ion} is neither constant nor negligible anymore and the scattered intensity must take that into account. The resulting general expression for the scattering intensity then becomes

$$I_0(q, E) = F_0^2(q) + 2f'(E)F_0(q)v(q) + \lfloor f'^2(E) + f''^2(E)\rfloor v^2(q) \tag{18}$$

where $F_0{}^2(q)$ is the nonresonant intensity of the macroion (= micelle, polymer, etc.) measured far from the absorption edge; $F_0(q)$ and $v(q)$ are the nonresonant and resonant amplitudes respectively of the macroion and the counterion; $v^2(q)$ is the pure resonant contribution of the counterion measured at the absorption edge. A constant term due to fluorescence, affecting the intensity above the absorption edge, has been omitted. For a more extended discussion on the SAXS theory near the absorption edge of a given element, one can refer to refs 57−59. From eq 18, one is generally interested in extracting the pure resonant curve $v^2(q)$, which gives the scattering profile of the counterion alone. Because $I_0(q,E)$ contains three independent variables, one cannot extract directly $v^2(q)$. Two complementary methods are proposed in the literature to operate in such a way. In all cases, one must record the $I_0(q,E)$ at various values of the incident beam energy, far and close to the absorption edge. The classical SAXS experiment described in the previous section must then be repeated by changing the value of E. The first method, called the "matrix" method, consists of a numerical solution of $I_0(q,E)$ needing at least three different energy values.[60] The second approach, called the "fit" (or "Ballauf") method,[61] consists of neglecting the $f''(E)$ term before the absorption edge and performing a quadratic interpolation of $I_0(E)$ for each value of the scattering vector. For a detailed discussion on the problem of solving $I_0(q,E)$, one can refer to refs 60−63. The experimental approach and data treatment used to treat our ASAXS experiments is detailed on Page S6 in the Supporting Information.

Contrast Matching Small-Angle Neutron Scattering (SANS). These experiments have been performed to look at the contribution of the micellar core only, thus matching the scattering of the sophorose headgroup (molecular formula: $C_{12}H_{21}O_{11}$, molar mass: 341.30 g/mol) in the simple ion-free sophorolipid micellar system at 80 mM. For the SLD calculations, we used the bulk sophorose density (1.68 g/mL),[64] and we considered the fact that labile COH groups could be exchange into COD. The estimated SLD for sophorose under these conditions is ∼2.6 × 10^{-6} Å^{-2}, corresponding to a ∼46:54 = D_2O:H_2O mixture. These values are to be considered as an approximation because the exact sophorose density in sophorolipid is not exactly known. Two

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

samples have then been prepared at concentration of the sophorolipid of 50 mg/mL: (a) in fully deuterated water (100:0 = D_2O:H_2O) and (b) in 46:54 = D_2O:H_2O mixture.

SANS experiments have been performed on the PACE beamline at the Laboratoire Léon Brillouin (LLB) facilities at CEA-Saclay (France). Quartz cells (1 nm) have been used as sample holder. The low-q portion of the data was recorded using a sample-to-detector distance of 4565.95 mm and wavelength of 6 Å, whereas the high-q portion was recorded using 866.14 mm and 6 Å. The scattering intensity is obtained from the determination of the number of neutrons in the incident beam and the detector cell solid angle. Data were corrected for the ambient background, empty cell scattering, neutron beam transmission and detector efficiency and normalized to the neutron beam flux to get the scattered intensity $I(q)$ in absolute units according to ref 65. Incoherent signal was substracted by measuring the background value at high-q values for both, sample-free, 100:0 = D_2O:H_2O and 46:54 = D_2O:H_2O mixtures. All data have been treated using the Pasinet 2 software[66] provided at the beamline free of charge. Pair distribution function analysis, $P(r)$, was done using the SASView software package[44] using estimated Dmax values of 40 and 140 Å, respectively, for the 46:54 and 100:0 = D_2O:H_2O mixtures.

Molecular Dynamics Simulations. All the MD simulations discussed in this paper were carried out with GROMACS (v4.6.6)[67,68] and with the protonated acidic sophorolipid (SL) molecule presented in Figure 1. Because the SL ionization degree is found to be <5% (see the Experimental Section) and similar to Prasad et al.,[69] we only simulated the protoned form of the SL and without ions. As in a previous work[69] to model the surfactant, we used an "united-atom" force field based on the GROMOS53A6.[70] In particular, the parameters of the sophorose group were taken from the GROMOS53A6 force field for carbohydrates ($56A6_{CARBO}$) developed by Hansen et al.[71] Concerning the alkyl chain (with the double bond) and the COOH groups, the parameters were taken from Kukol[72] and the glutamic acid parameters available in the GROMOS53A6[70] force field, respectively. To be consistent with the GROMOS force field, we used the simple charge water model (SPC)[73] to model water with the SETTLE algorithm[74] to keep its geometry rigid.

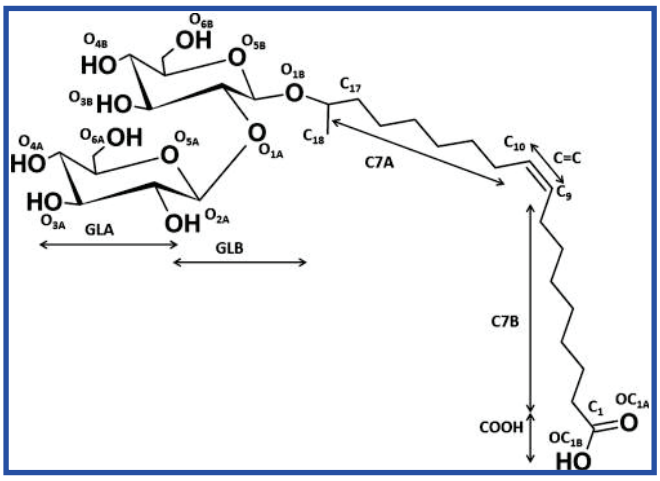


Figure 1. Acidic sophorolipid surfactant with the atom-numbering scheme and part used in the work (see main text for details).

We have to mention that by not taking into account in the simulation the effects of the pH and presence of the ions in the solution, our simulations may have some limitations. However, as we will see further, the simulation results are consistent with the reported experimental data, and we think they provide a good model basis of the micelle as a function of their aggregation numbers. To have a large overview of the micelle characteristics as a function of the SL aggregation number, N_{agg}, we carried out six MD simulations with different SL N_{agg} values (Table 2). Specifically, we chose 28 and 37, values obtained by Penfold et al.[36] for a mixture of sophorolipids (sample S4 in ref 36: 52.2% non acetylated, 39% monoacetylated, and 8.5% diacetylated), respectively, at 5 mM and 30 mM; 56 and 68 adapted from the SAXS data in this work (Table 3, RbCl system). In addition to these simulations, we also performed two more simulations with N_{agg} = 80 and 112, to examine the effect of high N_{agg} values of the SL aggregate structure. These two N_{agg} values correspond to an intermediary point between 56 and 68 and a value estimated by Penfold.[36] The effective concentrations of the simulated systems are in a range of 110 mM and 456 mM (given SL mass fraction, SL wt % of 6.2–12.6%), larger than the experimental concentration (80 mM) to reduce the number of water molecules and, consequently, the computational cost of the simulations. To construct the system for simulation, we used a *self-assembled* approach, where the corresponding number of sophorolipids with a random orientation was inserted in a SPC water cubic box. All the systems were minimized, equilibrated and finally simulated at ambient conditions (P = 1.015 bar and T = 303 K). Specifically, the temperature and the pressure were controlled with the Bussi et al.[75] thermostat (τ_T = 0.1 ps) and the Parrinello–Rahman[76,77] barostat (τ_p = 3.0 ps and with a compressibility of 45 × 10^{-6} bar^{-1}). During the equilibration stages, the SL molecules were harmonically restrained with a force constant of 1000 kJ mol·nm^{-1} to equilibrate the solvent around the

Table 2. Overview of the Simulated Systems[a]

MD no.	N_{agg}	n_{H_2O}	n_{atom}	n_{Water}/SL	SL wt %	[SL] M	L_{box} (Å)	t_{sim} (ns)
MD1	28	13600	42228	485.7	6.6	0.114	76.3	112
MD2	37	18000	55887	486.5	6.6	0.114	77.7	110
MD3	56	13600	43656	242.9	12.4	0.228	83.7	162
MD4	68	16500	52968	242.6	12.4	0.228	82.9	164
MD5	80	19400	62280	242.5	12.4	0.228	87.5	162
MD6	112	13600	46512	121.4	22.0	0.456	80.4	144

[a] N_{agg}, n_{H_2O}, and n_{atom} are the numbers of sophorolipid monomers, water, and atoms composing each systems. n_{Water}/SL, SL wt% [SL], L_{box}, and t_{sim} are the number of water per sophorolipid molecule, the sophorolipid concentration, the total mass fraction, the box size, and the simulation time (in ns), respectively.

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113–13133

Table 3. Fit Parameters[a] Obtained from the Core–Shell Ellipsoid of Revolution Form Factor Using $X_S = 0$ (Defined in Figure 2)

sample	N_{agg}	R_C (Å)	X_C	T_S (Å)	$\rho_S \times 10^{-5}$ (Å^{-2})	α	z
RbCl-10 mM	55	8.0	6.5	12.0	1.0741	0.70	
RbCl-50 mM	64	8.0	7.7	11.7	1.0768	0.72	
RbCl-100 mM	69	7.9	8.1	11.5	1.0829	0.69	
$SrCl_2$-5 mM	52	7.7	6.9	12.4	1.0608	0.71	
$SrCl_2$-25 mM	72	7.7	9.4	12.3	1.0611	0.72	
$SrCl_2$-50 mM	73	7.8	8.9	11.8	1.0751	0.69	
RbOH-40 mM	34	6.6	5.5	12.3	1.0618	0.55	6.8
RbOH-80 mM	24	6.1	4.7	11.0	1.0707	0.53	9.9
$Sr(OH)_2$-10 mM	50	7.2	6.8	12.6	1.0632	0.59	4.6
$Sr(OH)_2$-20 mM	41	7.0	5.4	12.3	1.0785	0.52	5.0

[a] N_{agg} = aggregation number; R_C = equatorial core radius; X_C = axial core ratio; T_S = equatorial shell thickness; ρ_s = scattering length density of the shell; α = fraction of CH_2 inserted in the dry core of the micelle eq 4; z = charge per micelle.

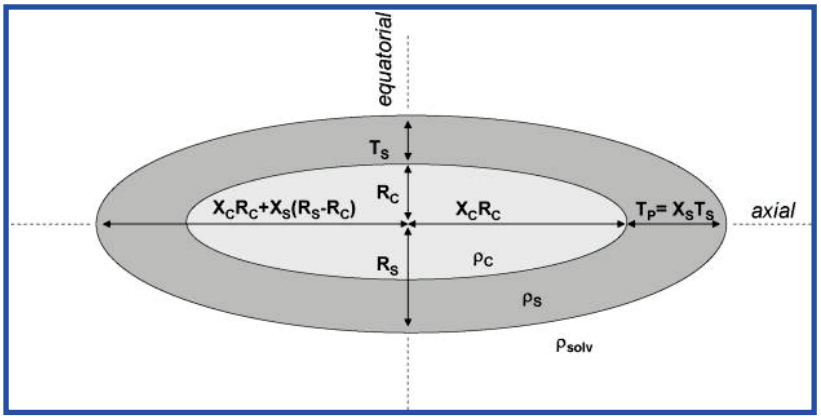


Figure 2. Core–shell prolate ellipsoid of revolution model chosen to fit the SAXS data.

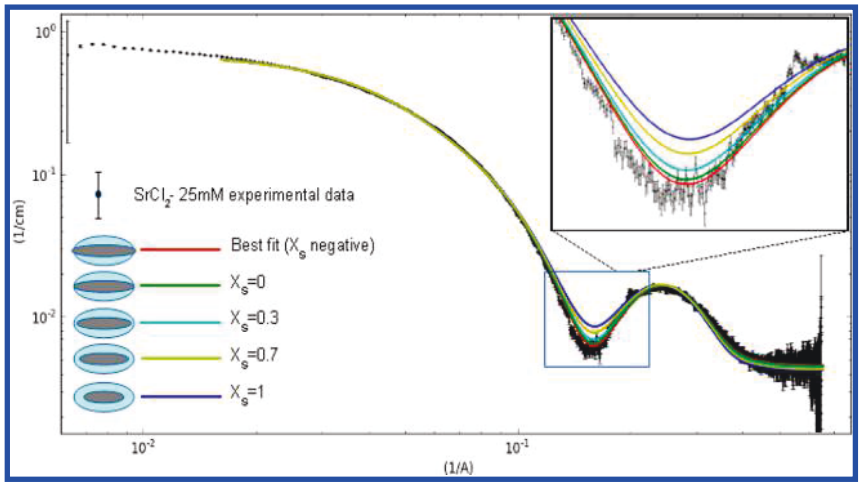


Figure 3. SAXS curve recorded on the $SrCl_2$-25 mM system: influence of X_S (polar shell ratio) value on the calculated curves obtained with a core–shell ellipsoid of revolution form factor. Best fit is given in red.

sophorolipids. SL and water molecules were coupled separately with two thermostatic baths as a standard rule. To integrate the equations of motions, we used a time step of 2 fs with the P-LINCS algorithm[78] to restrain bond lengths to their equilibration values. Electrostatic interactions were treated with the reaction field approach[79] with the nonbonded interactions evaluated with a twin range cutoff scheme, with a short- and long-range cutoff distances of 8 and 14 Å, respectively, and an update frequency of two timesteps for the short-range pair list. To correct the truncation of electrostatic interactions outside the cutoff of 14 Å, a reaction field term[79] corresponding to a relative dielectric permittivity of the SPC water (61)[80] was added. Finally, the production simulations were performed during 110–162 ns with the atomic data collected every 2 ps for subsequent analysis

RESULTS AND DISCUSSION

Description of the "Coffee Bean"-Like Prolate Ellipsoid of Revolution Model. As critically discussed on Page S2 in Supporting Information, the best shape model to fit

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113–13133

SAXS data is the core−shell prolate ellipsoid of revolution function, which gives good χ^2/Npts values below 15. The geometric parameters of this model are the equatorial core radius, R_C, the axial core ratio X_C ($X_C > 1$ for prolate micelles), the equatorial shell thickness, T_S, and the polar shell ratio in the axial direction, X_S (see Figure 2).

For a constant shell thickness, $X_S = 1$, the fit gives a χ^2/Npts value around 13. However, the best fit (χ^2/Npts = 6.7) is obtained for a slightly negative X_S value ($X_S \sim -0.2$) corresponding to a small and negative polar shell thickness in the axial direction, T_P, of the core−shell ellipsoid of revolution model.

In Figure 3, we report a typical SAXS curve recorded on sample 7 ($SrCl_2$-25 mM), which shows a series of fits using the same core−shell prolate ellipsoid of revolution form factor model and a variable X_S, allowed to vary between 0 and 1. It can be clearly seen that the bottom of the first minimum is best fitted with a negative X_S value, a fact which has no clear physical meaning, as it supposes that there is a small part of the hydrocarbon core which is in close contact with the aqueous solvent at the end-tips of the ellipsoid in the axial direction. However, as the best fit is always obtained for $X_S < 0$, we impose $X_S = 0$ for all samples' profiles. We interpret this assumption as follows: the end-tips in the axial direction of the sophorolipid aggregates are somewhat difficult to describe and may consist of a quite disorganized assembly, containing COOH groups, sophorose, salt, water, and part of the aliphatic chain.

It is important to note that whatever the absolute value of X_S, the trends obtained for the other variables is unchanged, as it can be seen from the comparison between the fit parameters for $X_S = 0$ presented in Table 3 (discussed later in more detail) and the fit parameters obtained for $X_S = 0.3$, shown in Table S2/Table S3 in the Supporting Information. Except the surprising feature of $X_S = 0$, all other variable parameters give realistic values: the equatorial core radius and shell thickness are around 7.7 and 12.3 Å, respectively, and the shell SLD is 1.1×10^{-5} Å^{-2} (387 e$^-$/nm^3), corresponding roughly to 73% v/v of water in a simple water−sophorose composition of the shell (Table 4). The X_C ratio of 9.4 characterizing the elongation of the ellipsoid is slightly high, if compared with previous results,[33] and will be discussed thereafter.

The resulting shape that best describes the sophorolipid micelles is very close to a "coffee bean"-like, with a variable shell thickness from T_S to 0, which is quite atypical with respect to the majority of ionic, nonionic, and bolaform[16] surfactant systems, generally described by core−shell form factors with a homogeneous shell thickness.[19−21] Nevertheless, this peculiar shape is the best suited to fit at best all systems described in this work and a large number of complementary other systems.

All SAXS measurements were made at the same sophorolipid concentration, 80 mM, and at concentrations of salt and/or base in the range from 5 to 100 mM. Table 3 shows the model parameters (N_{agg}, R_C, X_C, z, ρ_s) for core−shell ellipsoidal micelles of the sophorolipids with either salt or a base at various concentrations, but also the calculated shell thickness (T_S) and volume fraction of CH_2 in the dry core (α). Table 4 presents the volume fraction of water ($x_{v,w}$), and number of water molecules per sophorolipid ($n_{w/SL}$) in the shell; these have been indicated for three possible headgroup (hg) scenarios: hg_1 considers only sophorose and COOH (or COO^-); hg_2 includes the hg_1 hypothesis and the C_{17} and C_{18} atoms; hg_3 considers the hg_2 scenario as well as the contribution of CH_2 groups from the aliphatic chain excluded from the core (13(1-α), where α has been defined in eq 4 and experimental values given in Table 3).

Table 4. $x_{v,w}$ = Volume Fraction of Water in the Shell (eq 7) and $n_{w/SL}$ = Number of Water Molecules Per Sophorolipid[a] (See eq 8)

sample	$x_{v,w}$			$n_{w/SL}$		
	hg_1	hg_2	hg_3	hg_1	hg_2	hg_3
RbCl-10 mM	0.70	0.65	0.54	32	29	23
RbCl-50 mM	0.69	0.64	0.54	31	28	22
RbCl-100 mM	0.68	0.63	0.51	29	26	20
$SrCl_2$-5 mM	0.73	0.68	0.59	37	34	28
$SrCl_2$-25 mM	0.73	0.68	0.59	37	34	28
$SrCl_2$-50 mM	0.70	0.65	0.54	32	29	22
RbOH-40 mM	0.73	0.68	0.53	37	34	24
RbOH-80 mM	0.71	0.66	0.49	33	30	21
$Sr(OH)_2$-10 mM	0.72	0.68	0.54	36	33	24
$Sr(OH)_2$-20 mM	0.69	0.64	0.46	31	28	18

[a]Several compositions of the headgroup are considered: headgroup 1, hg_1 = sophorose + COOH; headgroup 2, hg_2 = sophorose + COOH + CH (C_{17}) + CH_3 (C_{18}); headgroup 3, hg_3 = sophorose + COOH + CH (C_{17}) + CH_3 (C_{18}) + 13(1-α)CH_2. hg_3 hypothesis best describes our system (refer to text for discussion).

First of all, the results show that the equatorial core radius R_C, which corresponds to the dry part of the micelle, is relatively small, between 6.0 and 8.0 Å, a value which is the lower limit of what is acceptable for the model. In return, the thickness of the shell, between 11.0 and 12.5 Å, is quite large if this value is compared with those reported in the literature for similar systems studied by SANS (∼8 Å, ∼ 3 Å),[33,36] a discrepancy that may be explained by the different contrasts inherent in each technique. At the same time, the value of α is contained in the range between 0.52 and 0.72 depending on the system, and it mainly indicates that a significant proportion of the hydrocarbon chain is in an aqueous environment, which is the counterpart of the small core and large shell of these aggregates. This is expected for a bolaamphiphile, as quantitatively demonstrated here below.

Shell Hydration vs Ion Condensation. $n_{w/SL}$ gives an estimate of the number of water molecules per sophorolipid in the shell. As shown in Table 4, this parameter strongly depends on the supposed composition of the dry headgroup. If one only considers the sophorose and COOH groups to be part of the shell (hg_1 hypothesis in Table 4), $n_{w/SL}$ varies between 29 and 37. In the hg_2 scenario (hg_1 plus CH and CH_3), $n_{w/SL}$ does not vary much, as the value is between 26 and 34, indicating that inclusion of CH and CH_3 in the headgroup does not have a substantial impact on the shell hydration. Finally, if one includes the fraction of CH_2 not included in the core (13(1-α)), hg_3 hypothesis in Table 4, the hydration is sensibly lower and $n_{w/SL}$ now varies between 18 and 28. As a first comment, we must state that these estimations should be taken with caution, because our SAXS model does not include the presence of ions in the shell, which can also be responsible for the variation of the shell SLD. The proportion of ions in the shell could be controlled in the fit through, for example, an adjustable parameter k = number of cations per sophorolipid in the shell. However, k cannot be determined without fixing the number of water molecules per sophorolipid, both parameters influencing the value of ρ_S in opposite ways: ionic species increase the value of ρ_S, whereas water molecules decrease it.

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

By neglecting the influence of ion concentration, the number of water molecules per sophorolipid is adjusted to account for both effects, so that an increase of ρ_S leads to a decrease of $n_{w/SL}$, even if it can come, in reality, from an increase of the ion condensation. As a result, the number of water molecules per sophorolipid reported in this case can be interpreted as minimum values, even if the space filling requirement may also have an influence. This approximation will be confirmed by the ASAXS experiments, which will show that the amount of counterions in the shell is practically negligible, unless their concentration becomes important.

Second, values of $n_{w/SL}$ above ~25 are higher than expected from the literature about glycolipids and disaccharides. It is well-known that hydration properties of disaccharides with the same chemical formula as sophorose (maltose, sucrose, and trehalose, $C_{12}H_{22}O_{11}$) but different structures are different, depending on the solute-water interactions.[81,82] Effects coming from the linkage type between the two glucose unit (sophorose is a glucose β(1,2) disaccharide), the position of hydroxyl groups, the number of hydrogen bonds may influence the hydration properties of disaccharides and have been extensively studied. For example, the headgroup hydration number of maltoside surfactants is reported to be 8 in the case of SAXS studies[83] and in Abel et al.[84] with MD simulations. For the trehalose, which is known as being slightly more hydrated than maltose,[81,82] Lupi et al.[85] report values taken from literature that lie between 4 and 18. It is well-known that hydration numbers reported depend on the experimental method and/or assumption adopted. Ultrasonic measurements give values among the highest, and MD simulations for which the hydration number is defined as the number of H-bound water leads to similar values (around 15 to 18), suggesting that these experimental hydration numbers mainly refer to water molecules directly involved in H-bonds with the solute. In contrast, in their study on hydration and mobility of trehalose in aqueous solution, Revsbech Winther et al.[86] used a geometric hydration number, identified as the number of water molecules required to cover the solute with one layer (47 water molecules per trehalose).

The values obtained in this work are for a micellar shell composed of a sophorose group, that is, a disaccharide and the hydration properties of which, if they are reasonably expected to differ from maltose and trehalose, should still not be excessively high. For this reason, the hg_3 hypothesis provided in Table 4 is most likely to be real, as reasonable values are certainly the ones including the COO^- (or COOH), the CH and CH_3 groups but also 13(1-α)CH_2 groups and an unknown, and variable, number of ions, depending on the sample composition. More than reasonable, the hg_3 hypothesis is actually expected: if hydration numbers of COOH, CH_2, and CH_3 are reported to be small, 1.153, 0.999, and 0.586, respectively,[87] it is expected the aliphatic chain portion close to the headgroup to be hydrated. Typical estimations (both experimental and theoretical) for alkylammonium salt and SDS surfactants are that up to 3 CH_2 groups from the polar head can be hydrated.[50−52] In this work, we find between 4 and 7 hydrated CH_2 groups per sophorolipid (calculated as 13(1-α)CH_2, where α is given in Table 3); these values can be reasonably expected for a system containing two polar headgroups. On the other hand, hydration number of ions is higher, and depending on the ionic concentration, it can influence the SLD of the shell quite heavily: reported hydration numbers lie between 6 and 8 for Rb^+, between 4 and 8 for Na^+,[88] and between 7.3 to 10.3 for Sr^{2+}.[89] According to the ASAXS experiments reported in the next section and showing a negligible amount of counterions in the shell, one can consider the amount of hydration water to be also negligible.

If the contribution to the hydration from all species can reasonably explain the experimental values of $n_{w/SL}$ found here, one should still take these values with caution, given the wide range of variation of the theoretical values. Moreover, there is no indication that the micelle shell does not contain free water that is not accounted for in the hydration numbers reported. The shell hydration may also be strongly related to the conformation of sophorolipids within the micelle, depending on the molecular structure of the sophorolipid and its impacts on the micellar structure.

As we can see from Table 3, the shell thicknesses reported here are always greater than 10.5 Å and usually between 11.0 and 12.5 Å, a range of values which seems to be high compared to the size of the sophorose alone, estimated to be below 10 Å,[32] 7.3 Å in poly sophorolipids,[90] or 9 Å for trehalose.[86] The shell thickness of SL micelles determined by SANS was also reported to be around 8 Å.[33] The quite thick shell found here may take into account the "roughness",[4] confirmed by the MD simulation in the next sections, of the apolar core and hydrated shell interface coming from the presence of CH_2 in the hydrated shell environment. We can estimate the typical separation distance between sophorose molecules in the shell from the sophorose molar concentration (0.93−1.34 mol·L^{-1}.) to be around 3.5−5.1 Å, by using molecular parameters of trehalose (see ref 86). This space, filled by the other components of the shell, can accommodate one water layer, and even if there are some alkyl groups in close contact with the sophorose, with respect to the geometric hydration number of 47 for trehalose in water, the number of water molecules per sophorolipid in the shell presented here, which lie from 20 to 30, in the hg_3 hypothesis, does not seem outrageous.

Area per Sophorolipid. Another interesting piece of information is given by the analysis of the area per sophorolipid at the micelle/solvent interface ($A_{mic,w/SL}$), and at the dry core/hydrated shell interface ($A_{C,S/SL}$, see Table 5). These are calculated using the classical formula for the surface area of a prolate ellipsoid divided by N_{agg}; the values of the semiaxes, as defined in Figure 2, are given in Table 3. Micellar structure of dodecyl maltoside has been described by Cecutti et al.[91] showing that ellipticity of the micelle comes from the bulky

Table 5. Area Parameters[a]

sample	$A_{C,S/SL}$ (Å)	$A_{alkyl/sopho}$ (Å)	$A_{mic,w/SL}$ (Å)	L (Å)
RbCl-10 mM	75	104	196	3.76
RbCl-50 mM	76	103	194	3.69
RbCl-100 mM	74	102	187	3.49
$SrCl_2$-5 mM	78	108	213	4.21
$SrCl_2$-25 mM	79	106	208	4.19
$SrCl_2$-50 mM	75	103	193	3.72
RbOH-40 mM	71	124	219	4.04
RbOH-80 mM	75	141	231	3.72
$Sr(OH)_2$-10 mM	70	111	204	4.03
$Sr(OH)_2$-20 mM	64	112	190	3.48

[a] $A_{mic,w/SL}$ = area per sophorolipid at the micelle/solvent interface; $A_{C,S/SL}$ = dry core/hydrated shell interface; $A_{alkyl/sopho}$ = area per sophorolipid between the alkyl chain and the sugar/carboxylate headgroup; L = average length between the headgroups.

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

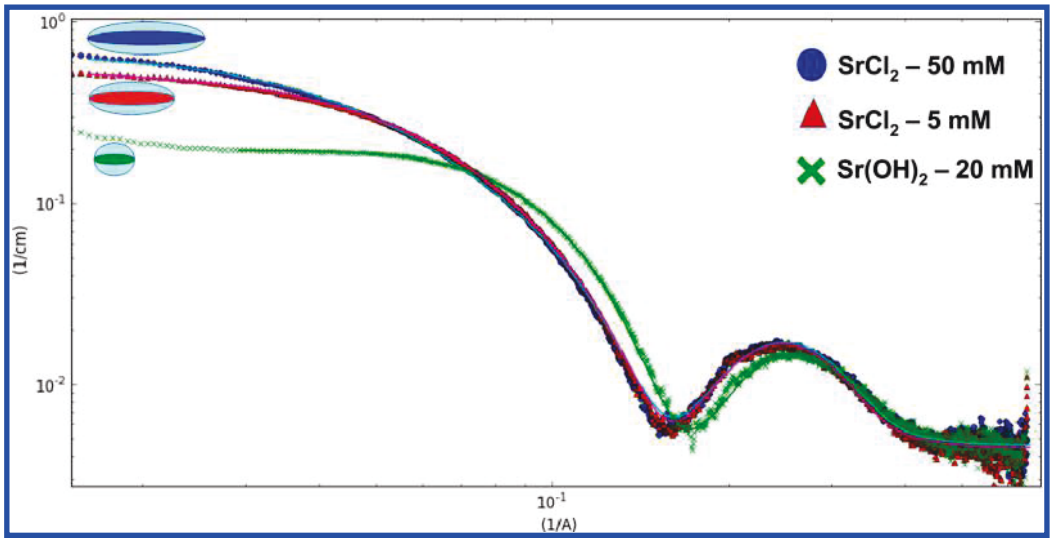


Figure 4. SAXS spectra of sophorolipids in the presence of $SrCl_2$ and $Sr(OH)_2$. The pictures represent the ellipsoidal nature of the micelle with a X_S = 0, as commented before. The difference in longitudinal dimension among the ellipsoids is based on the data obtained from the best fits and presented in Table 3.

hydrated maltose headgroups and their perpendicular conformation to the interface. Area per dodecyl maltoside at the micelle/solvent interface and at hydrophobic core/sugar headgroup interface are 87 and 50 Å^2, respectively. In the case of sophorolipids, however, one must keep in mind that there is one alkyl chain and two different polar heads per surfactant, leading to values lying from 187 to 230 Å^2 per sophorolipid at the micelle/water interface and from 64 to 79 Å^2 per SL at the core/shell interface, depending on the sample composition. Another point to consider is that the core/shell interface in our case is located quite far from the sugar headgroup, considering the α value lying around 0.52−0.72. If the alkyl chains are perpendicular to the core/shell interface, we can estimate the area per sophorolipid between the alkyl chain and the sugar/carboxylate headgroup $A_{alkyl/sopho}$ to be the surface per SL of an ellipsoid of equatorial radius R_C + 6.5(1-α)1.265 and axial radius $R_C X_C$ + 6.5(1-α)1.265 (noticing that this surface may be slightly overestimated as the influence of the X_S parameter is neglected). In this case, areas reported lie between 102 to 141 Å^2 for the most ionized micelles, comparable to that reported by Chen et al.[92] for SL at the air−water interface (104 Å^2).

Ionic and Basic Effects. According to the results in Table 3, some general observations can be made regarding the effect of salts on sophorolipids micelles. First, the salts cause an elongation of the micellar aggregates and subsequent increase of the aggregation number, while the diameter of the core remains constant. For instance, increasing RbCl from 10 mM to 100 mM, it produces an increase of X_C from 6.5 to 8.1, whereas R_C is practically 8 Å. However, it appears that the elongation due to the presence of salt has a limit; this is observed through N_{agg} and X_C values, which increase rapidly between 10 and 50 mM of RbCl (5 and 25 mM of $SrCl_2$) and then seem to reach a plateau at higher concentration, around 73 for N_{agg} and 8.5−9.0 for X_C. Another fairly clear effect of adding salt is the slight shell shrinkage and the concomitant dehydration observed through the shell thickness (T_S), the volume fraction of water in the shell ($x_{v,w}$), and the decreasing number of water molecules per sophorolipid ($n_{w/SL}$).

The shell SLD, ρ_S, is in the range between 1.06 and 1.09 × 10^{-5} Å^{-2}, where the smaller value is systematically observed for the $SrCl_2$ salt system. Small differences between the two salts is also observed on the values of R_C, T_S, $x_{v,w}$, and $n_{w/SL}$. The sophorolipids micelles in the presence of $SrCl_2$ have slightly smaller cores, a shell slightly thicker and hydrated than in the presence of RbCl. These differences have no impact on the value of α that remains constant for all the salt systems and quite high (around 0.69−0.72). The aggregation number N_{agg} is slightly smaller for $SrCl_2$ at low salt concentration (52 vs 55), while at higher concentration the elongation of sophorolipids micelles is more pronounced with $SrCl_2$ than with RbCl (X_C reached a value of 8.9 for $SrCl_2$ and 8.1 only for RbCl). Areas per sophorolipid and the average length between the headgroups, L, reflect these small differences, being slightly higher for $SrCl_2$ than for RbCl (Table 5).

Adding a base to sophorolipids micelles' solutions considerably changes the physicochemical parameters. First of all, one observes the appearance of intermicellar interactions, accounted here with the introduction of a structure factor described with a repulsive screened Coulombic intermicellar interaction potential. This is consistent with what it was previously reported.[33] In terms of micellar size, the equatorial core radius is dramatically reduced, from 7.2 Å for the less basic system ($Sr(OH)_2$ − 10 mM) to 6.1 Å for the most basic system (80 mM RbOH), while the X_C ratio is much smaller than those reported above, decreasing to values smaller than 6 for all [OH^-] higher than 40 mM. The aggregation number follows this trend, decreasing to 24 for RbOH 80 mM. α also undergoes a fall, reaching values around 0.52, which indicates that in basic medium, half of the total amount of CH_2 are located in an aqueous environment. These trends are consistent with an increase of the micelle curvature due to the increased electrostatic repulsion taking place between more ionized headgroups. The shell thickness is practically invariable (12.3 Å) for the less basic samples (RbOH-40 mM and $Sr(OH)_2$-20 mM) and substantially decreases (11.0 Å) at high concentration of RbOH. Finally, Figure 4 shows the typical SAXS profiles of sophorolipids in the presence of $SrCl_2$ and $Sr(OH)_2$, showing the experimentally

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

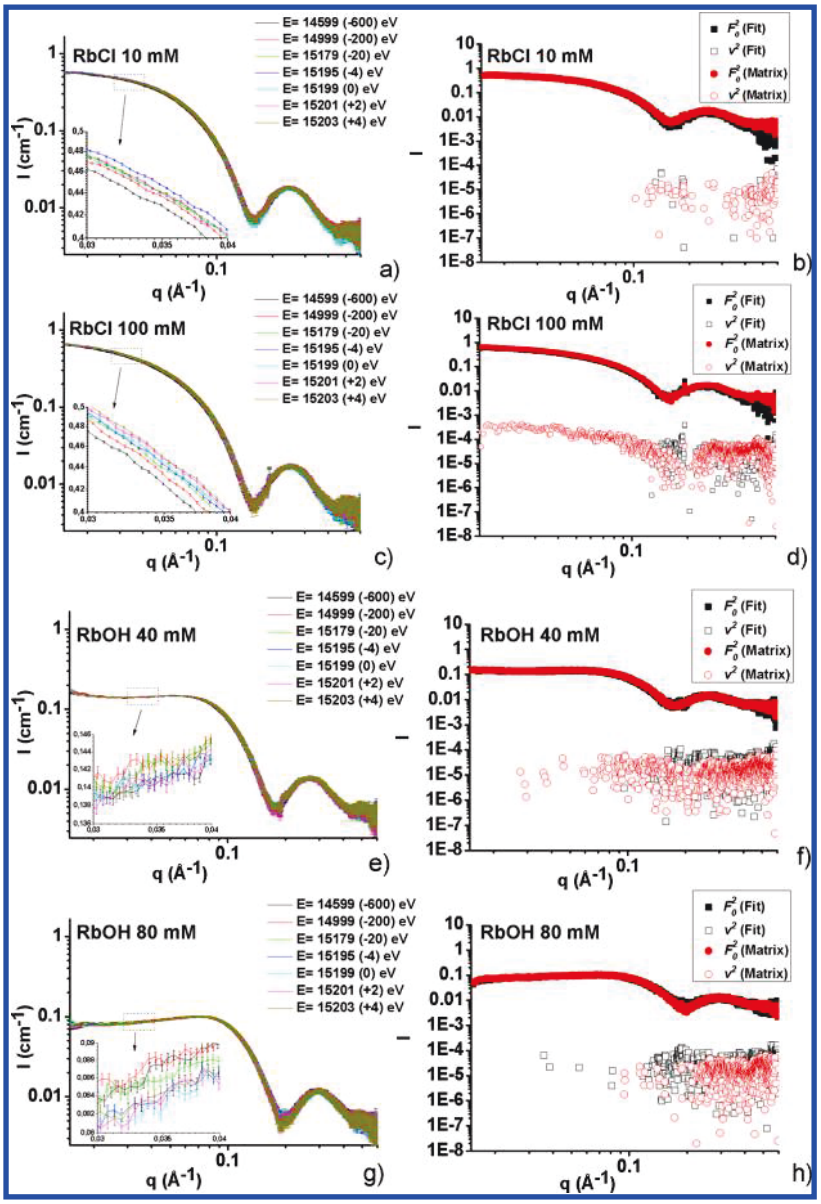


Figure 5. (a, c, e, g) Energy-dependent SAXS spectra for the sophorolipid − RbCl (10, 100 mM) and RbOH (40, 80 mM) systems. The energies at which the spectra have been recorded are indicated in the figure. (b, d, f, h) $F^2(q)$ and $v^2(q)$ scattering profile of the cation (Rb^+ and Sr^{2+}), as obtained from the ASAXS treatment of the spectra. More experimental details are presented on Page S6 in the Supporting Information.

found spectral differences and a pictorial imaging of the salt and base effect on the micellar size.

Counterion Distribution in the Salt Systems Studied by ASAXS. ASAXS is a powerful method to put in evidence the distribution of counterions in soft matter and in particular around polyelectrolytes and micelles.[57−62] It was employed to describe the charge distribution around charged polymers,[58,61] DNA and RNA[93] and CTAB micelles.[60,62] In this study, ASAXS is employed with the goal of revealing indirectly the distribution of the COO^- groups, if one assumes that its negative charge is neutralized by a counterion upon increasing the solution pH. In this case, the SAXS spectrum shows a broad scattering hump, which can be nicely fitted using a Screened Coulomb potential, as it was previously done in SANS experiments.[33] The resulting negative charge of the micelles settles between −5 and −10 (Table 3), and it clearly increases upon increasing the amount of base for both Rb^+ and Sr^{2+}. ASAXS can be used both in a qualitative and quantitative way.[62,63] If the former tells about the space distribution of counterions, the latter provides data on the amount of counterion in the proximity of charged macroions and even their binding degree. However, the quantification process can be very delicate,[60,61,94] and it was mainly used so far on either polyelectrolytes or robust micellar systems.

Figure 5 shows the typical scattering profile, $I(q)$, recorded for the RbCl (10, 100 mM) and RbOH (40, 80 mM) systems at the incident energy E = 14599, 14999, 15179, 15195, 15199, 15201, 15203 eV, corresponding to the ΔE_{Rb+} = −600, −200, −20, −4, 0, + 2, + 4 eV (treatment of background is shown in Figure S2 in the Supporting Information). The ASAXS results are only shown for the RbCl system but very similar conclusions can be drawn for the $SrCl_2$-containing samples, and for this reason, they will not be reported here. As explained in the Experimental Section, fluorescence contribution has been subtracted for all spectra. At a first glance, spectra are very close to each other, but a closer look in the inset of Figure 5a,c, for

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

instance, indicate that $I(q)$ appreciably (with respect to the error bars) increases with increasing energy (color sequence: black, red, green, blue···) for both the RbCl 10 mM and 100 mM systems. This effect is expected in the case of an anomalous effect upon varying energy through the edge, and it demonstrates that Rb^+ ions decorate the micellar outer surface. Nevertheless, variation in $I(q,E)$ is very mild; at 100 mM RbCl (Figure 5c), a concentration in counterions at which the anomalous effect was reported to be very important for polyelectrolytes and ionic surfactants systems,[62,63] $I(q,E)$ only varies by a factor 1.04 when measured far below and at the absorption edge [$I(0.03,14599) = 0.47$ cm^{-1} compared to $I(0.03,15199) = 0.49$ cm^{-1} at the Rb^+ edge], while expected variations should not be less than a factor 2. This is further demonstrated by the splitting of $I(q)$ into the $F_0^2(q)$ and $\nu^2(q)$ components, respectively, the nonresonant intensity of the macro-ion and the pure resonant contribution of the counter-ion, shown in Figure 5b,d. The splitting shows no noticeable effect on the 10 mM system using both the matrix and fit methods, as well as a mild anomalous effect on the 100 mM system, but only employing the matrix method. This last result is coherent with the higher salt amount present in this system, thus indicating that part of the Rb^+ cation in solution decorates the micellar shell. This result helps better understand the evolution of the shell SLD (ρ_s) commented in the SAXS section (Table 3). One can now affirm that the amount of salt in the shell is below the limit of detection of the ASAXS technique, which can reasonably be settled in the mM range. This shows that the very small increase in the ρ_s, found for the RbCl, but also the $SrCl_2$ systems, is most likely due to a true variation in the number of water molecules per sophorolipid. However, for the higher RbCl concentrations (e.g., 100 mM), one cannot exclude the presence of larger amounts of salt in the shell, a fact which may also, as commented before, account for the mild increase in ρ_s. Interestingly, in the presence of salt alone, the equilibrium pH of a sophorolipid solution is slightly below 5. The corresponding experimental ionization degree (([COO^-]/([COOH] + [COO^-]))), calculated from the titration curve (Figure S3 in the Supporting Information), is below 5%. This means that at high salt content, counterions decorate the hydrophilic shell independently from the presence of carboxylate groups. In terms of quantification, the mild anomalous effect recorded does not allow a reasonable quantification of the amount of counterions-per-sophorolipid.

Upon addition of RbOH, carboxylic groups deprotonate into carboxylates, which carry their respective counterions, our hypothesis[33] being that the negative charges are settled at the micellar/water palisade. Under these assumptions, this effect should be detectable by ASAXS at a larger extent with respect to the base-free medium. The corresponding experiments in Figure 5e,g show, on the contrary, that $I(q)$ undergoes no sensible evolution with the incident energy, as all curves practically lie within a small range represented by the error bars. The poor anomalous scattering effect is confirmed by the signal splitting using both the fit and matrix methods (Figure 5f,h), which do not provide a significant $\nu^2(q)$ component. Because similar results are obtained with $Sr(OH)_2$, these will not be discussed further. How should this result be interpreted? One of the main results so far is the fact that the bolaform morphology of sophorolipids promotes the formation of "non-standard" micelles. By this we intend a micelle that does not have a homogeneous hydrophobic core well-separated from the hydrophilic corona with constant thickness. The size of micelles and the aggregation number sensibly decrease upon addition of a base (Table 3), and a partial solubilization of sophorolipids in solution should not be excluded. Such a scenario could suggest that not all COO^- groups do actually settle in a narrow region at the micellar/water interface, but they could be randomly distributed in the thick shell region, being closer to the hydrophobic core. If resonant species (the counterions) penetrate deeper inside the micelle, the difference in terms of core−shell SLD at the element absorption edge will be attenuated, and one can expect the ASAXS signal to be attenuated. This scenario could be suggested by the larger number of hydrated CH_2 (about six per sophorolipid, for an α parameter close to 0.5, Table 3) and not excluded with respect to old works on ionic/neutral bolaamphiphiles[22,23] and according to which water can penetrate deep inside a bola's micelle. However, a comparison with previous ASAXS studies is difficult because all of them have been carried out on either polyelectrolytes or micellar surfactant systems with a well-defined core−(charged)shell interface (e.g., CTAB). In both cases, there are no doubts about the localization of the counterions. Even if fluctuations in the counterion distribution is probably amplified in the case of sophorolipids due to a redistribution of the carboxylate groups away from the outer micellar layer, this phenomenon has been already reported to potentially affect ASAXS measurements on other systems. In micellar systems (TTAB), fluctuation of the counterion cloud was shown to affect the scattering intensity of the macroion (TTA^+) at $q > 0.2$ Å^{-1}.[95] In a rod-like polyelectrolyte, it was shown that longitudinal fluctuations of the number density of the counterions affect SAXS data at $q > 0.3$ Å^{-1}, and they can actually be measured experimentally.[59,61] Finally, fluctuations due to the polydispersity of the number of acrylic acid (PAA) chains on polystyrene (PS) was also shown to affect the pure resonant curve $\nu^2(q)$ at $q > 0.04$ Å^{-1}. According to these works, it seems possible that, if fluctuations of the COO^- distribution occurs in sophorolipid micelles upon addition of a base, the corresponding $\nu^2(q)$ profile will inevitably be affected. It seems that in all systems where the macroion (polyelectrolyte, micelle) charge density is well-defined and located at the surface, only the counterions number density fluctuations at the macroion surface should be considered. In this case, the $\nu^2(q)$ profiles will be affected at (relatively) large q values above 0.2 Å^{-1}, which, in the case of sophorolipid micelles, may not play an important role afterall. On the contrary, if fluctuations of the macroion charge distribution are experienced, as seen in the case of PAA−PS system, the affected q-range may be much larger, starting at $q < 0.1$ Å^{-1}. In this specific case, the effect on the $\nu^2(q)$ profile for sophorolipid micelles may not be negligible. We then expect that if the COO^- distribution will occur at the micellar surface only, the $\nu^2(q)$ will be enhanced, while if the COO^- distribution will occur in the interior of the micelle, the $\nu^2(q)$ will be reduced. We believe that the second scenario does occur in the sophorolipid-based system upon addition of a base.

Contrast Matching by Small-Angle Neutron Scattering. In the SAXS section, we have shown that sophorolipid micelles at 80 mM adopt a nonconventional "coffee bean"-like structure, the length of which is slightly influenced by the amount of salt or base. By "coffee bean"-like structure, we have intended the fact that the composition at the extremities in the axial direction of the micelle is hard to describe with a simple core−shell model, and they are most-likely constituted by a mixture of aliphatic chain, sophorose, and COOH groups.

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

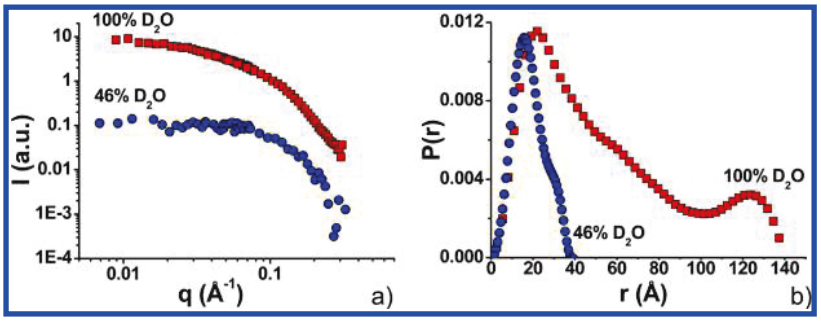


Figure 6. (a) SANS curves recorded on a 80 mM system of SL in 100/0 D_2O/H_2O (red squares) and 46/54 D_2O/H_2O (blue circles); (b) Pair distribution function, *P(r)*, issued from data in (a).

According to the ASAXS experiments, we could also put in evidence that the distribution of the COO^- groups is not only confined at the micellar surface but it is probably random in the thick hydrophilic region. Unravelling the distribution of the sophorolipid molecule inside a micelle can be approached through a direct experiment which consists of "visualizing" the distribution of the oleic acid, or the sophorose group by means of contrast-variation SANS. This experiment consists into matching the SLD of a given section of a particle, or molecule, via a well-chosen mixture of deuterated and hydrogenated solvent, the SLD of which can be tuned by playing on the relative ratio between the two. In the case of SL micelles, it is possible to look at the scattering of the oleic acid, or sophorose, alone by choosing the correct proportion of D_2O and H_2O. Sophorose headgroups can be matched with a 46% D_2O, while oleic acid can be matched with 8% D_2O. Here, we prefer the first approach because experiments with large amounts of H_2O will suffer from the strong incoherent diffusion of hydrogen, inherent to the technique, and which are responsible for a critical increase in the background level. A complete match between the solvent and sophorose will give an idea of the spatial distribution of the pure aliphatic-containing core.

Figure 6 shows the typical SANS curve of a salt-free 80 mM solution of acidic SL in pure D_2O, the description of which is given elsewhere[33] and the geometry of which we now assume to be "coffee bean"-like, as largely commented above. Without going into a further model-based analysis, we simply display the corresponding Indirect Fourier Transformation method (IFT), from which it is possible to evaluate the shape of an aggregate from the corresponding pair distribution function, *P(r)*.[96] The *P(r)* of a 80 mM solution of SL in D_2O (Figure 6b) of course reflects the $I(q)$ data, and it is typical for a highly elongated object, in agreement with the SAXS data presented above and previous data recorded on a similar system.[32] If the hypothesis of a homogeneous core−shell ellipsoid (the shell being of constant thickness and homogeneously composed of sophorose and water only) is correct, then one should expect, after matching the contrast between sophorose and the solvent, a SANS profile describing an ellipsoidal micelle being smaller in size (only the core will contribute). The SANS scattering profile of the contrast-matched (46% D_2O) SL solution is shown in Figure 6a (blue circles), and it presents two interesting features. First of all, the low-*q* intensity is decreased by a factor 100 with respect to the fully deuterated medium; this is a direct proof that the experiment was successful. In fact, according to eq 1, the intensity is directly proportional to the contrast between the object and the solvent. By matching the SLD between the sophorose and the solvent, the contrast only occurs between the core and the solvent, and it is reduced, thus contributing to decrease the scattered intensity. The second point concerns the micellar shape and size after contrast matching. The *P(r)* function at 46% D_2O (Figure 6b) is typical of a much less elongated morphology, closer to a spheroid, if compared to the fully deuterated system. This suggests that a pure aliphatic core is mainly concentrated in a spheroidal region located at the center of the micelle.

This experiment gives credit to, and it actually refines, the sophorolipid micelle model deduced from SAXS analysis. At the equilibrium pH (<5) and at 80 mM, the sophorolipid micelle should be considered as a prolate ellipsoid divided into three regions, as identified in Figure 7: Region I is composed of a fully aliphatic core and the morphology of which is practically spherical; Region II constitutes the end-tips in the axial direction of the ellipsoid, and it is a less defined region composed of aliphatic, sophorose, and COOH groups. Region I is directly identified by contrast-matching SANS experiments while region II is deduced by the fitting SAXS data using a "coffee bean" like model. Finally, Region III is identified as being the outer shell composed of sophorose, COOH groups, water, and possibly counterions. Contribution from vicinal CH_x groups should not be excluded, as detailed in the headgroup, hg_3, hypothesis of Table 4. As commented before, the contribution of salt plays a role on the ellipticity of the micelle only, but it does not play a role on its diameter or its configuration. Furthermore, the ASAXS data show that the amount of salt in the sophorose surrounding is most likely negligible at least up to 100 mM. The orientation of the sophorolipid molecule within the micelle is sketched in Figure 7. Considering the typical micellar dimension with respect to the sophorolipid size, it seems highly probable that molecules can either cross the micelle from one end to the other, or bend in such a way that the COOH and sophorose groups belonging

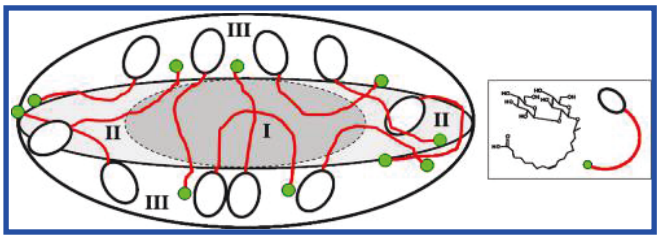


Figure 7. Tentative model structure of an acidic sophorolpid micelle at its equilibrium pH (<5). Region I = full aliphatic; Region II = mixed sophorose/water/COOH/aliphatic; Region III = sophorose/water/COOH. At 5 < pH < 7, the boundaries between the three regions are probably less defined, and the COO^- groups occupy a broader volume in the hydrophilic region.

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

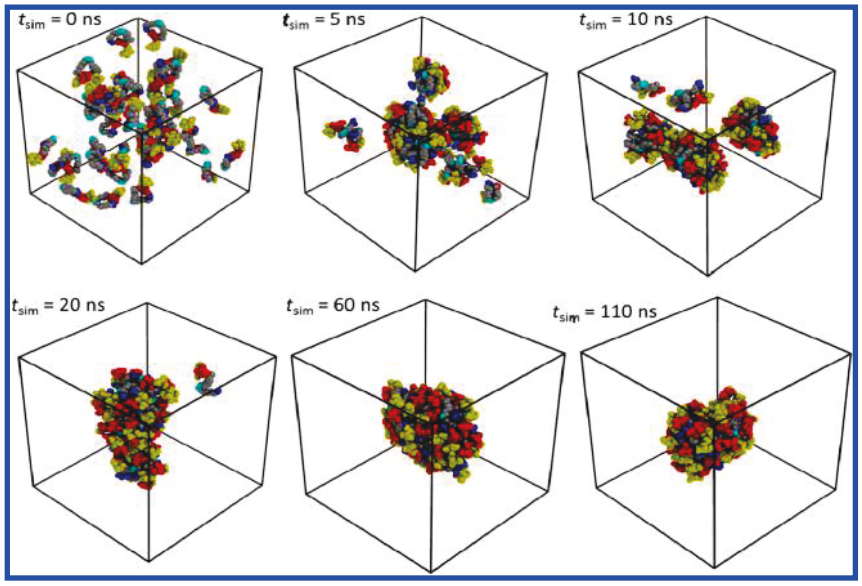


Figure 8. Representative aggregation process of 37 SL monomers into a micelle vs time for the MD2 run. The outermost and innermost glucose are colored in yellow and red, respectively, whereas the alkyl chain, double bonds, and COOH groups are in gray, cyan and blue colors, respectively. Water is not shown for visual clarity. The black line shows the box limits. The figures were drawn with PyMol.[100]

to the same molecule both lie in the same side of Region III. The addition of small amounts of base (pH range comprised between 5 and 7) promotes electrostatic repulsion among micelles induced by the deprotonation of the COOH into COO^-. According to the ASAXS data showing a nonexisting anomalous effect in the presence of both Rb^+ and Sr^{2+}, we believe that in the intermicellar interaction regime, the negative COO^- groups and their respective counterions do not seem to be distributed only at the external micellar-water palisade, as we previously hypothesized,[32,33] but most likely to occupy the entire hydrophilic shell region.

Several authors have tried to study the structure of bolaform amphiphiles and, in particular, the one of glucose-containing units. The group of Vill has published several works on the micellar formation of bolaform glycolipids, mainly symmetric[18] and/or in mixture with anionic or glycolipid surfactants,[97,98] and they concluded that slightly ellipsoidal micelles could form, without going in detail in terms of the molecular distribution within the micelles. If similar works exist on other types of bolaform amphiphiles,[10,14] the precise description of molecular distribution is always a challenging task. Nagarajan,[21] basing his line of thought on geometrical considerations only, predicted the self-assembly behavior of bola amphiphiles, mainly indicating the formation of spherical, cylindrical, and vesicular objects. Interestingly, for spherical micelles, he predicted a micellar radius equal to, or smaller, than half the full length of the hydrophobic chain. He also predicted interpenetration of the molecules and, for cylindrical micelles, local disorder at the end-tips in the axial direction of the cylinder. All these features, especially the disordered Region II shown in Figure 7, nicely confirm Nagarajan's models. Caponetti et al.[16] are the only ones who deeply discussed the distribution of *N*-azo-18-crown-6 ethers in micelles. Even if they did not provide a tentative image of their system, they discussed some interesting parameters. They used a core−shell ellipsoid of revolution form factor and used a homogeneous shell to fit their data. They found that between 15 and 20 water molecules are contained (bound and free) the hydrophilic headgroup but none enters the core. They also found that addition of salt had no major effect on the core radius and shell thickness, even if some differences are found between LiCl and NaCl due to the respective difference in terms for complexation by the crown ether macrocycle. More insight on the structure of bolaamphiphile micelles, including molecular conformation, was also provided by Davey[19] and Shinde,[20] who worked on asymmetric bolaamphiphiles, alcohol derivatives of alkylammonium and oleate salts, which are ionic/neutral bolas analogous in structure to sophorolipids in the neutral pH regime. Interestingly, both authors agreed on the fact that micelles are spherical/ellipsoidal objects and on the fact that both hydrophilic headgroups come in contact with the solvent at the micelle outer palisade, in contrast with older claims according to which only the ionic headgroup is located at the micelle/solvent frontier while the nonionic group is settled in the micelle interior, into which water easily penetrates. However, they do not agree on the molecular arrangements: Davey proposes an elongated cross-micellar conformation, while Shinde proposes a bent configuration. If compared to the studies presented above, our more complex "coffee bean"-like model presented in Figure 7 could be considered as a confluence of Nagarajan's,[21] Davey's,[19] and Shinde's[20] models together, in which the end-caps in the axial direction of the micelle are less defined and where the bolaamphiphile can adopt more than one configuration according to several parameters, such as its position in the micelle itself, the aggregation number, the presence of salt, the ionic/neutral nature of one of the end-groups, the shell hydration state, and so on. In order to verify the micellar model and sophorolipid distribution shown in Figure 7, we have used molecular dynamics simulations shown in the next section.

Molecular Modeling by Molecular Dynamics. In the following sections, we computed representative parameters (e.g., size, shape, surface properties, SL headgroup hydration, and surfactant alkyl chain conformation) to characterize the micelle structure as a function of the SL aggregation number.

I. Aggregation Process of the SL Monomers into Micelle. In Figure 8, we depict the SL aggregation process into a single aggregate for the MD2 simulation with 37 sophorolipids. The

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

Table 6. Aggregate Size and Shape[a]

system	R_g^M	a_M	b_M	c_M	$(a/c)_M$	a_{HC}	b_{HC}	c_{HC}	$(a/c)_{HC}$	l_{pl}
MD1	15.5	22.2	19.9	17.6	1.30	17.3	15.0	12.1	1.48	5.1
MD2	17.3	27.3	21.2	17.5	1.56	23.0	16.3	12.0	1.95	4.9
MD3	20.3	34.3	22.2	20.0	1.68	30.9	17.3	14.9	2.09	4.4
MD4	22.3	38.5	24.0	20.5	1.88	35.7	18.5	14.8	2.41	4.7
MD5	24.4	44.7	25.0	19.2	2.33	41.7	20.5	12.9	3.22	4.5
MD6*	23.4	14.7	40.2	12.2	-	11.1	-	10.1	-	3.3

[a]Values with M and HC subscripts were computed by including all the micelle atoms and those of the hydrophobic core, respectively. The radius of gyration, R_g^M, and the ellipsoid three semi-axis lengths (in Å) were computed from the inertia tensor (e.g., see ref 102). a_M and c_M are the lengths and the radius of the rod. l_{pl} is the average polar layer thickness (including the sophorose and COOH groups) (in Å). The statistical errors (maximum errors) are always lower than 0.3, 0.5, and 0.2 Å for R_g, semi-axis lengths, and polar layer thickness, respectively. *In the case of the MD6 system, due to its rod shape, the $(a/c)_M$ and $(a/c)_{HC}$ values were not computed, and the radius of gyration was obtained for a cylindrical rod of $2b_M$ length (equivalent to the box size) and an average radius of $(a_M+c_M)/2$.

simulations results show that, in general, the aggregation process of the SL monomers into a single and stable micelle is relatively fast (∼10−60 ns), and further analysis (not shown) also suggests that it occurs with two different time scales: one "fast" (t_{sim} < 10 ns) and one "slow" (t_{sim} = 10−60 ns), depending on the system size as previously observed for others micellar systems (see, for instance, ref 99 and references therein). For instance, in the case of the MD2 simulation (Figure 8), it takes ∼10 ns for all SL monomers to form small clusters with various sizes and shapes and ∼55 ns for the clusters to form a single and stable aggregate. In the other systems, the aggregation time depends on the SL numbers and varies between 20−80 ns for the system with 27 and 68 SL, respectively. The stability of the aggregates can be also estimated by obtaining a constant radius of gyration, R_g, values of the aggregate (see next section) and agree with these values.

II. Micelle Shapes and Sizes. In Figure S4 in the Supporting Information, we depicted the final structure of the aggregate as a function of the SL aggregation numbers obtained from the simulations. With the increase of the N_{agg} value, we observe a (quasi)sphere-to−(elongated) ellipsoid transition of the micelle shapes. At high N_{agg} value (i.e., MD6 system), the SL monomers form a spheroidal rod, and this result does not depend on the SL concentration, since we found a similar shape with another simulation (not shown) where SL concentration was set 2 times lower. The micelle shape transition is actually in very good agreement with SANS data obtained on acidic sophorolipids in water as a function of concentration.[101]

To estimate more precisely the micelle shape changes during the course of the MD1−5 simulations, we employ a three-axis ellipsoid of revolution shape to model the micelles. The ellipsoid of mass M_T is identified by three semiaxis a_M, b_M and c_M ($a_M > b_M > c_M$) parameters, and the following equations:[102]

$$I_1 = \frac{1}{5}M_T(a_M^2 + b_M^2)$$

$$I_2 = \frac{1}{5}M_T(a_M^2 + c_M^2)$$

$$I_3 = \frac{1}{5}M_T(b_M^2 + c_M^2) \quad (19)$$

where I_1, I_2, and I_3 (with $I_1 > I_2 > I_3$) are the principal moments of inertia of the micelle. Please note that, with respect to the ellipsoid model in Figure 2 used to fit SAXS data, a_M is associated with the axial ($X_C R_C + T_P$) and b_M, c_M to the equatorial R_S dimensions.

The ratio between the major (a_M) and minor (c_M) semiaxis for the complete 28−80 SL micelle (a_M, b_M, c_M and $(a/c)_M$) are reported in the sixth column of Table 6. The computed micelle a_M, b_M, and c_M values confirm that at low N_{agg} the micelle is nearly spherical and that the micelle changes to a (pronounced) triaxial ellipsoid with the increase of N_{agg}. At N_{agg} = 112, the aggregate forms a rod shape oriented along the y axis, and we may consider that the upper limit to have micelle is around 80 SL. Moreover, as noted previously[84] for similar micellar systems and in agreement with our crude "coffee bean"-like model, the shape of the micellar hydrophobic core is always significantly more ellipsoidal that the overall micelle ($(a/c)_{HC}$ and $(a/c)_M$ in range of 1.48−3.22 and 1.30−2.33, respectively).

By subtracting the semiaxis lengths of the aggregate hydrophobic core from those of the whole micelle, one can also obtain an estimation of the average thickness of the polar outer layer, l_{pl} (including the sophorose and the COOH groups, see last column of Table 6), which is nearly equivalent to the shell thickness (T_S) estimated from the SAXS data (Figure 2). We have to say that the estimation of the l_{pl} values from the semiaxis is quite rough and strongly depends on the presence of alkyl chain carbon atoms at the aggregate surface (and a_M, b_M, and c_M values) that can underestimate the l_{pl} calculations. The average polar layer thickness, l_{pl}, values change a little with the micelle shape and are in range of 4.9−5.1 Å for the aggregates with 28−80 SL monomers. In case of the aggregate with N_{agg} = 112, we estimate a l_{pl} value of 3.3 Å with only the b_M and c_M semiaxis. The computed l_{pl} values here are much smaller than the shell thickness values, T_S, obtained from the SAXS fit (>11.0 Å, see Table 3). The l_{pl} value reflects the average shell thickness, which, according to SAXS data presented above, seems to be highly anisotropic, because T_S varies between ∼11 Å (equatorial) and 0 Å (longitudinal), the average between these values, $\langle T_S \rangle$, being in very good agreement with l_{pl}. The computed l_{pl} values are consistent with those obtained previously for N-dodecyl-β-maltoside (DDM) micelles,[84] where we found an average thickness for the α-maltose group (linear $\beta(1 \rightarrow 4)$ disaccharide) in the range of 7.2−7.7 Å and consequently support the simulation results.

To characterize the aggregate sizes, we computed the time evolution of R_g of the five micellar systems starting from the beginning of the production runs. As shown in Figure S5, after large fluctuations of the R_g due to the SL monomers aggregation processes, the R_g values stabilize after 20−90 ns depending on the number of SL monomers in the box. We consider that all the SL monomers form a single and stable aggregate when the micelle R_g remains constant until the end of

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

the simulation times. In the second column of Table 6, we reported the average R_g for MD1−5 systems computed from the last 40 ns of each simulation. The values are found between 15.5−24.4 Å for the aggregates with 28 and 80, respectively. The values are in relative agreement with data found in this study (18−30 Å from the geometrical parameters detailed in Table 3 or computed from the Penfold values,[36] according to which the R_g (where $R_g^2 = (1/5)(a_M^2 + b_M^2 + c_M^2)$) for an ellipsoid of revolution, with a_M, b_M, c_M being the semiaxes of the ellipsoid obtained from the inertia tensor (see below). In case of the simulation with 112 SL, we obtained a $R_g \approx 23.4$ Å from the expression for a cylindrical rod $R_g^2 = ((R^2/2) + (L_{rod}^2/12))$, where L_{rod} (here, $2b_M$) and R are average lengths of the rod and its radius (i.e., $(a_M + c_M)/2 = 13.85$ Å), respectively.

III. Aggregate Surface Contact Properties and Hydration. As shown in the MD snapshots depicted in Figure S4, different parts of the surfactant (i.e., sophorose headgroup and alkyl chain) are in contact with the solvent. To characterize more precisely the surface contacts shared between the aggregates and water, we computed the solvent accessible surface area (SASA) properties with two approaches. In the first one, we used the whole aggregate semiaxis lengths (a_M, b_M, and c_M) determined in the previous section, and we calculated the SASA for an ellipsoid with a planar surface, $SASA^e_{SL}$, with the Knud Thomsen's Formula: $4\pi((a_M^p b_M^p + a_M^p c_M^p + b_M^p c_M^p)/3)^{(1/p)}$ where $p \approx 1.6075$. Please note that this equation is valid for ellipsoids with three distinct semiaxes, so in case of the aggregate with 112 SL monomers that has a spherical rod shape, the $SASA^e_{SL}$ was estimated from the expression for an open cylinder (i.e., without the lateral areas): $2\pi R L_{rod}$, where R and L_{rod} are the characteristic dimensions of the rod defined just above. In the second approach, we compute the SASA of the SL part in the aggregate (i.e., the headgroup, the double bond, the alkyl chain, and COOH) by splitting the SL molecule into six parts namely, the external and internal glucose ($SASA^V_{GLA}$ and $SASA^V_{GLB}$, respectively), the COOH group ($SASA^V_{COOH}$), the C7 alkyl chains separated by the C═C and bonded to the sophorose and COOH moieties ($SASA^V_{C7A}$ and $SASA^V_{C7B}$, respectively) and finally the two carbon atoms involved in the double bond ($SASA^V_{C=C}$). With these values, one can obtain the total SASA of each aggregate,$SASA^e_{SL}$ with $SASA^V_{GLA} + SASA^V_{GLB} + SASA^V_{COOH} + SASA^V_{C7A} + SASA^V_{C7B} + SASA^V_{C=C} = 100\%$ (Table S5). These values were computed with the *trjVoronoi* program developed in our group (see ref 99 and references cited therein). This program uses a Voronoi−Delauney tessellation[103,104] algorithm to estimate the surface shared between the SL and water atoms (excluding hydrogen).[105] In contrast to the first approach where we assume that aggregate is a planar triaxial ellipsoid, in the latter calculation, the micelle surface irregularity is fully taken into account. This is why the computed $SASA^V_{SL}$ values are always greater than $SASA^e_{SL}$ ones. By comparing the $SASA^V_{SL}$ and $SASA^e_{SL}$ values reported in Table S5 and by visual inspection of the simulation snapshots in Figure S4, one can deduce that the aggregate surface is quite rough, as also estimated by the analysis of the SAXS data above. To quantify this, we compute the surface rugosity factor ($f_{rug} = SASA^V_{SL}/SASA^e_{SL}$), and we found that f_{rug} values are always greater than 1.8 and slightly increase with SL aggregation number. For the aggregate with 112 SL monomers, we found a larger f_{rug} value (3.95) probably because we underestimate the $SASA^e_{SL}$ by considering the aggregate shape as a spherical rod. For the sake of comparison, the computed f_{rug} values are found higher than those obtained in DDM micelles with similar size (1.6−2.0).[84]

From the $SASA^e_{SL}$ values, one can also estimate the area per sophorolipid at the micelle/solvent interface, A^e_{SL} with the expression $A^e_{SL} = (SASA^V_{C7A}/n_{SL})$. We obtained A^e_{SL} values lying from 177.4 to 132.5 Å^2 with SL monomers between 28 and 80 in fair agreement (discrepancy estimated between 10% and 30%) with the experimental estimation given in Table 5. In case of the aggregate with 112 SL monomers, the A^e_{SL} is significantly lower (62.5 Å^2).

Concerning the $SASA^V_{SL}$ of each SL component, the values for the MD1−6 systems are reported in Table S5, columns 2−7. The low SASA values for the alkyl chain and the double bond moieties (∼14 and <1% of the total aggregate surface, respectively) indicate that these groups are protected from the solvent by SL headgroup or are buried in the hydrophobic core of the aggregates. For the sophorose and the COOH groups, they represent ∼78% and 8−9% of the $SASA^V_{SL}$ values and tend to slightly decrease with the increase of the SL aggregation number, which is probably the consequence of the decrease of the micelle curvature and micelle shape changes leading to a different accessibility of the sophorose and COOH solvent to water.

To characterize the aggregate hydration, we have also computed the number of "unique" water (N_w) in the first shell (with a cutoff <4.0 Å)[84] of whole aggregate, the sophorose, the COOH alkyl chain, and double bond atoms (Table S6). The number of water molecules per sophorolipid is found to be ∼27 for the smallest aggregate, and it decreases to ∼17 for the aggregate with 112. These results are in excellent agreement with the experimental data ($n_{w/SL}$), varying between 18 and 28 for the headgroup 3, hg_3, hypothesis, reported in Table 4. Moreover, the N_w values also confirm (as we noted from the corresponding $SASA^V_{SL}$ values), that the hydration of the sophorolipid is mainly due to the sophorose headgroup (as also found experimentally in the headgroup 1 hypothesis in Table 4) and at a less extent to COOH since these groups have 16.9−24.9 and 2.4−4.4 water molecules in their first shells depending on the aggregate size. If small hydration values obtained for the alkyl chain and the C═C bond (3.1−5.0 and <1, respectively), confirming the assumption made above about the accessibility of these groups to water, one should still observe that N_w for the alkyl chain varies between 3 and 5, which are very close estimates to what it is experimentally found in the hg_3 hypothesis presented in Table 4. The slight decrease of the sophorose headgroup hydration as a function of the SL aggregation number may be caused by the slight change of the sophorose headgroup conformation and/or interaction between adjacent headgroup at the micelle surface.

To obtain a more precise picture of the SL headgroup hydration as a function of the SL aggregate sizes, we also computed the average number of nearest waters for each sophorose and COOH oxygen atoms (Figure 1) or $< n_{Ox\text{-}Ow} >$ from the computed radial pair density functions (g(Ox-Ow)) of the sophorose, and COOH (OC_{1A} and OC_{1B})-oxygen and water-oxygen (O_w) atoms (Figure S6). To obtain $< n_{Ox\text{-}Ow} >$ values, we integrated the RDF functions until the first minimum at $r \approx 3.5$ Å after the first peak. No defined peaks at ∼2.5 Å in the (O_{1A}, O_{1B}, O_{5A}, and O_{5B}) RDFs indicate that these nonhydroxyl oxygen atoms are poorly hydrated by water in contrast to the sophorose OH oxygen. The computed $< n_{Ox\text{-}Ow} >$ values in Table S7 in the Supporting Information corroborate these observations. More specifically, the RDFs

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

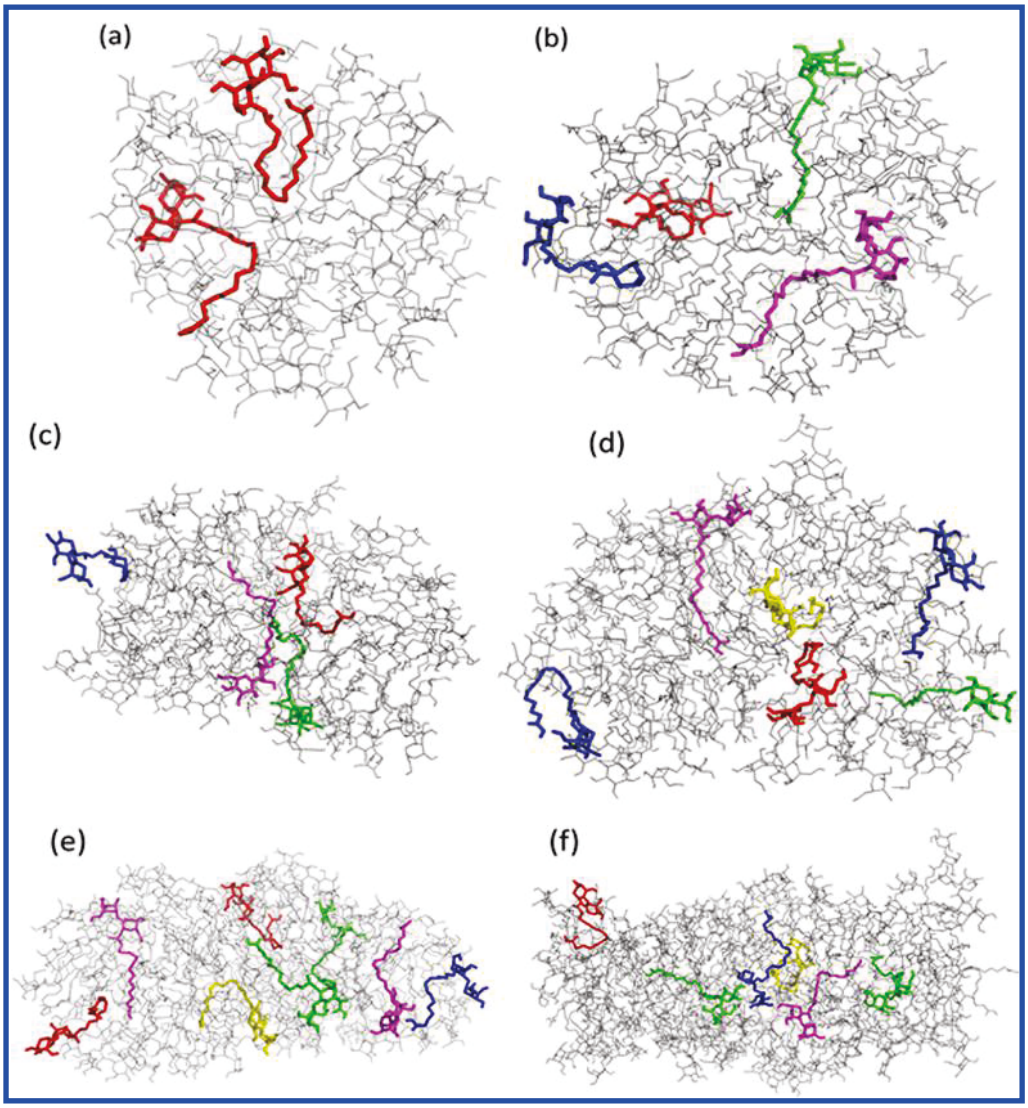


Figure 9. Representative conformation of the sophorolipids in the aggregates with (a): 28, (b) 37, (c) 56, (d) 68 and (e) 80, (f) 112 monomers. Red, blue, green, and yellow colors corresponds to the surfactant conformations where the sophorose and the COOH headgroups are located at the micelle surface with the rest of the surfactant in the hydrophobic core (red), the sophorolipid reside entirely at the micelle surface (blue), alkyl chain is extended and cross the hydrophobic core and with the sophorose and headgroups at the surface (magenta), where COOH in the hydrophobic core (green), where the overall sophorolipid are deeply buried in the hydrophobic core (yellow). The figures were drawn with PyMol.[100]

show that water molecules solvate primarily the hydroxyl oxygen (i.e., O_{2A}, O_{3A}, O_{4A}, O_{6A}, O_{3B}, O_{4B}, and O_{6B}) of the sophorose headgroup and the COOH oxygen atoms (OC_{1A} and OC_{1B}). The only differences in RDFs between the different systems are for the amplitude of the first peak indicating variations of the number of hydration water molecules in the vicinity these oxygen atoms. The $<n_{Ox\text{-}Ow}>$ values for the sophorose hydroxyl oxygen atoms are close (1.8−2.4) between the six aggregates in contrast to the carboxyl ones, where we observe a slight decrease of the $<n_{Ox\text{-}Ow}>$ values with the aggregate size (1.8−1.2 and 1.0−0.7). The hydrations of the ring oxygen (O_{5A} and O_{5B}) of the glucose are around 0.3 and 0.1, respectively. This is close to the value found previously[84] for the same atoms in maltose (∼0.3 water). Finally, concerning the oxygen atoms involved in the glycosidic bond (O_{1A}) and the link between the headgroup and alkyl chain (O_{1B}), the $<n_{Ox\text{-}Ow}>$ values (0.6) indicate these atoms are shielded from the solvent by the headgroup, as noted previously.[84]

IV. Conformation of the SL Alkyl Chain in the Aggregates. Finally, we also examined the conformation of the sophorolipid as a function of the SL aggregation numbers by computing the distribution of the end-to-end distance between the first C1 and the C17 atoms, $P(d_{C1-C17})$, of the SL alkyl chain (Figure 1 and Figure S7 in the Supporting Information). The figure shows a wide distribution of the C1−C17 end-to-end distances (10.5−13.5 Å) with the micelle size indicating a wide different conformation of the SL alkyl chain in the aggregates. The d_{C1-C17} is smaller than the extended oleic acid chain length estimated to be below 26 Å in the "extended" conformation,[106,107] indicating that the SL alkyl chain are bent with the COOH pointing toward the sophorose headgroup or slightly extended in the micelles. Similar observation was found by Prasad et al.[69] with acidic (three double bonds) linoleic acid sophorolipids.

To illustrate this, we highlighted some representative conformations of the sophorolipids in each aggregate (Figure 9) according to their conformations and localizations and the

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

model depicted in Figure 7. The figures clearly show that the sophorolipid molecules adopt different conformations and localizations in the micelles depending on their sizes. In particular, in the smallest micelle (N_{agg} = 28), which has a more spherical shape, there is only one conformation (red-highlighted molecule in Figure 9a−f and region I in Figure 7) for the sophorolipid: the sophorose and the COOH headgroups are located at the micelle surface with the rest of the surfactant partially folded in the hydrophobic core. With the increase of the micelle size, the SL adopt other conformations, for instance, where the all the sophorolipid reside entirely at the micelle surface (blue-highlighted molecule in Figure 9b−f and region II in Figure 7), with the COOH in the hydrophobic core (green-highlighted molecule in Figure 9) or where the alkyl chain in an extended conformation that crosses the hydrophobic core and with the sophorose and headgroups at the surface (magenta-highlighted molecule in Figure 9c−f and region 3 in Figure 7)). In the aggregates with N_{agg} = 68, 80, and 112, we also see that one sophorolipid where the overall sophorolipid is buried in the hydrophobic core (yellow-highlighted molecule in Figure 9d−f). Taken together, the simulation results show that for spherical micelles, the bent conformation of SL is predominant, but as soon as the shape turns into an ellipsoid, various conformations can coexist. These results also put in evidence the fact that the micelle is not constituted by a homogeneous arrangement of sophorose at the surface and oleic acid in the core, but rather a more complex distribution of these chemical groups, which is in agreement with the model picture proposed in Figure 7, deduced from experimental data. In particular, simulation data, both conformational and statistical, support the idea of a disordered region, which was identified as region II in Figure 7, in which molecular conformation and water interpenetration occur, thus explaining the complexity of fitting SAXS data using a standard core−shell model and justifying SANS contrast-matching data showing that a pure aliphatic core only exists in the inner micellar center.

■ CONCLUSION

This work shows a detailed analysis of the micellar structure composed of sophorolipids, an asymmetric bolaform glycolipid containing a pH-sensitive COOH group. If the micellar structure was previously reported to be of prolate ellipsoidal shape, we show here how the asymmetric bolaform compound settles within the micelle itself. Detailed modeling of SAXS spectra show an uneven distribution of matter within the ellipsoid. We find that the core−shell model is by far the most adapted to describe the micelle; however, if the hydrophobic core, the size of which is about 8 × 55 Å, actually corresponds to an ellipsoidal shape, the hydrophilic shell has a nonzero thickness (about 12 Å) only in the equatorial region of the micelle. The actual micellar model rather corresponds to a "coffee bean"-like shape than an actual homogeneous ellipsoid. Contrast matching SANS was used to probe the actual distribution of the hydrophobic components of the micelle. The experiment was run at 100% D_2O, and at 46% D_2O, at which one expects sophorose to be fully contrasted. Interestingly, the hydrophobic core has rather a globular morphology, rather than an elongated ellipsoid. Combining both SAXS and SANS experiments, we make the hypothesis that the ellipsoid of revolution can be divided into three regions: Region I, defining a fully aliphatic core in the micellar center; a less-defined Region II in the axial direction of the ellipsoid and composed of aliphatic, sophorose, and COOH groups; Region III, identifying the outer shell composed of sophorose, COOH groups, water and possibly counterions and in which contribution from vicinal CH_x groups should not be excluded. To account for such a complex structure, one must make the hypothesis that several configurations of sophorolipids are actually found within the micelle: both bent and elongated in the peripheral region but also crossing through the micelle itself. This picture is compatible with both the double polar nature of sophorolipids and the bent conformation of its oleic acid aliphatic chain.

To verify this model issued from experimental data, we carried out explicit MD simulations with 28, 37, 56, 68, 80, and 112 sophorolipid monomers. The simulation results show, in agreement with scattering experiments, that the micelle shape changes from a (quasi)sphere to an elongated ellipsoid with the increase of the N_{agg} value. At high N_{agg} value (here, N_{agg} = 112 SL), the aggregate forms a cylindrical rod, as previously reported in the literature for higher volume fractions. Moreover, all the aggregates present a rough surface with the sophorose and COOH headgroups in direct contact with the solvent leading to a larger hydration of these groups compared to the alkyl chain moiety. Examination of the sophorolipid conformations also shows that when the micelle is small and nearly spherical (here with N_{agg} = 28), the bent conformation of SL is predominant where the headgroup, the sophorose, and the COOH are located at the micelle surface with the alkyl chain partially folded in the hydrophobic core. With the increase of the micelle size, others SL conformations in the aggregate can coexist leading to a complex distribution of sophorolipids in the micelle surface and in a hydrophobic core, also in the agreement with the model of structure deduced from the scattering experiments.

Finally, we have also investigated the sophorolipid configuration upon small increase in pH (<7), when repulsive intermicellar interactions, due to the formation of COO^- groups, characterize the small-angle scattering spectra and, consequently, the micellar network. To do so, we have employed anomalous SAXS, consisting of analyzing the sample using an X-ray incident energy close to the X-ray absorption edge of a given element, in this case Rb^+ (up to 100 mM) and Sr^{2+} (up to 20 mM) counterions, used to neutralize the negative COO^- charges. Contrary to our expectations, the ASAXS effect near the edge was very mild at concentrations in counterions at which it is supposed to be very strong, if we compare the values to ionic surfactant systems. We interpret these results by making the hypothesis that the negatively charged COO^- groups do not preferably distribute at the outer micellar/water palisade but rather throughout the micellar hydrophilic shell volume, thus reducing the scattering effect of the counterions alone.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b05374.

Solvent SLD for all salt compositions; critical determination of the best fit model; fit parameters obtained for X_S = 0.3; molecular area values obtained for X_S = 0.3; experimental approach and data treatment in ASAXS experiments; experimental determination of the ionization degree; additional data obtained from molecular modeling by molecular dynamics (PDF)

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

AUTHOR INFORMATION

Corresponding Author

*E-mail: niki.baccile@upmc.fr. Tel.: +33 1 44 27 15 44.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. Biosurfing/289219. This work was, in part, granted access to the HPC resources of CCRT/CINES under the allocation x2014077322 (project no. gen7322) made by GENCI (Grand Equipement National de Calcul Intensif). SOLEIL Synchrotron (Saint Aubin, France) is kindly acknowledged for providing access to the SWING beamline.

REFERENCES

(1) Tanford, C. *The hydrophobic effect*; Wiley: New York, 1973.

(2) Imae, T.; Ikeda, S. Sphere-Rod Transition of Micelles of Tetradecyl-Trimethylammonium Halides in Aqueous Sodium Halide Solutions and Flexibility and Entanglement of Long Rodlike Micelles. *J. Phys. Chem.* **1986**, *90*, 5216−5223.

(3) Lin, Z.; Cai, J. J.; Scriven, L. E.; Davis, H. T. Spherical-to-Wormlike Micelle Transition in CTAB Solutions. *J. Phys. Chem.* **1994**, *98*, 5984−5993.

(4) Hayter, J. B.; Penfold, J. Determination of Micelle Structure and Charge by Neutron Small-Angle Scattering. *Colloid Polym. Sci.* **1983**, *261*, 1022−1030.

(5) Goldmints, I.; von Gottberg, F. K.; Smith, K. A.; Hatton, T. A. Small-Angle Neutron Scattering Study of PEO−PPO−PEO Micelle Structure in the Unimer-to-Micelle Transition Region. *Langmuir* **1997**, *13*, 3659−3664.

(6) Fuhrhop, J. H.; Wang, T. Bolaamphiphiles. *Chem. Rev.* **2004**, *104*, 2901−2937.

(7) Bonini, M.; Berti, D.; Di Meglio, J. M.; Almgren, M.; Teixeira, J.; Baglioni, P. Surfactant Aggregates Hosting a Photoresponsive Amphiphile: Structure and Photoinduced Conformational Changes. *Soft Matter* **2005**, *1*, 444−454.

(8) Lu, T.; Han, F.; Mao, G.; Lin, G.; Huang, J.; Huang, X.; Wang, Y.; Fu, H. Effect of Hydrocarbon Parts of the Polar Headgroup on Surfactant Aggregates in Gemini and Bola Surfactant Solutions. *Langmuir* **2007**, *23*, 2932−2936.

(9) Douliez, J.-P. Self-Assembly of Hollow Cones in a Bola-amphiphile/Hexadiamine Salt Solution. *J. Am. Chem. Soc.* **2005**, *127*, 15694−15695.

(10) Sirieix, J.; Lauth-de Viguerie, N.; Rivière, M.; Lattes, A. From unsymmetrical bolaamphiphiles to supermolecules. *New J. Chem.* **2000**, *24*, 1043−1048.

(11) Kameta, N.; Masuda, M.; Minamikawa, H.; Shimizu, T. Self-Assembly and Thermal Phase Transition Behavior of Unsymmetrical Bolaamphiphiles Having Glucose- and Amino-Hydrophilic Head-groups. *Langmuir* **2007**, *23*, 4634−4641.

(12) Gubitosi, M.; Travaglini, L.; D'Annibale, A.; Pavel, N. V.; Tato, J. V.; Obiols-Rabasa, M.; Sennato, S.; Olsson, U.; Schillén, K.; Galantini, L. Sugar−Bile Acid-Based Bolaamphiphiles: From Scrolls to Monodisperse Single-Walled Tubules. *Langmuir* **2014**, *30*, 6358−6366.

(13) Fariya, M.; Jain, A.; Dhawan, V.; Shah, S.; Nagarsenker, M. S. Bolaamphiphiles: A Pharmaceutical Review. *Adv. Pharm. Bull.* **2014**, *4*, 483−491.

(14) Yan, Y.; Xiong, W.; Li, X.; Lu, T.; Huang, J.; Li, Z.; Fu, H. Molecular Packing Parameter in Bolaamphiphile Solutions: Adjustment of Aggregate Morphology by Modifying the Solution Conditions. *J. Phys. Chem. B* **2007**, *111*, 2225−2230.

(15) Meister, A.; Bastrop, M.; Koschoreck, S.; Garamus, V. M.; Sinemus, T.; Hempel, G.; Drescher, S.; Dobner, B.; Richtering, W.; Huber, K.; Blume, A. Structure-Property Relationship in Stimulus-Responsive Bolaamphiphile Hydrogels. *Langmuir* **2007**, *23*, 7715−7723.

(16) Caponetti, E.; Chillura-Martino, D.; La Mesa, C.; Muzzalupo, R.; Pedone, L. Structural and Transport Properties of Bola C-16 Micelles in Water and in Aqueous Electrolyte Solutions. *J. Phys. Chem. B* **2004**, *108*, 1214−1223.

(17) Munoz, S.; Mallen, J.; Nakano, A.; Chen, Z.; Gay, I.; Echegoyen, L.; Gokel, G. W. Ultrathin Monolayer Lipid Membranes from a New Family of Crown Ether-Based Bola-Amphiphiles. *J. Am. Chem. Soc.* **1993**, *115*, 1705−1711.

(18) Garamus, V. M.; Milkereit, G.; Gerber, S.; Vill, V. Micellar Structure of a Sugar Based Bolaamphiphile in Pure Solution and Destabilizing Effects in Mixtures of Glycolipids. *Chem. Phys. Lett.* **2004**, *392*, 105−109.

(19) Davey, T. W.; Ducker, W. A.; Hayman, A. R. Aggregation of Omega-Hydroxy Quaternary Ammonium Bolaform Surfactants. *Langmuir* **2000**, *16*, 2430−2435.

(20) Shinde, N.; Narayan, K. S. Studies on Sodium Ricinoleate. 1. Adsorption and Aggregation Behavior. *J. Phys. Chem.* **1992**, *96*, 5160−5165.

(21) Nagarajan, R. Self-Assembly of Bola Amphiphiles. *Chem. Eng. Commun.* **1987**, *55*, 251−273.

(22) Menger, F. M.; Jerkunica, J. M.; Johnston, J. C. The Water Content of a Micelle Interior. The Fjord vs. Reef Models. *J. Am. Chem. Soc.* **1978**, *100*, 4676−4678.

(23) Brown, J. M.; Schofield, J. D. Localised Regions of Reduced Mobility in Micelles; ^{13C} N.M.R. Spin-Lattice Relaxation Times of Functional Surfactants in Aqueous Solution. *J. Chem. Soc., Chem. Commun.* **1975**, *11*, 434.

(24) Besnard, R.; Cambedouzou, J.; Arrachart, G.; Diat, O.; Pellet-Rostaing, S. Self-Assembly of Condensable "Bola-Amphiphiles" in Water/Tetraethoxysilane Mixtures for the Elaboration of Mesostructured Hybrid Materials. *Langmuir* **2013**, *29*, 10368−10375.

(25) Fu, D.-W.; Zhang, W.; Cai, H.-L.; Zhang, Y.; Ge, J.-Z.; Xiong, R.-G.; Huang, S. D. Supramolecular Bola-Like Ferroelectric: 4-Methoxyanilinium Tetrafluoroborate-18-crown-6. *J. Am. Chem. Soc.* **2011**, *133*, 12780−12786.

(26) Sano, M.; Oishi, K.; Ishi-i, T.; Shinkai, S. Vesicle Formation and Its Fractal Distribution by Bola-Amphiphilic [60]Fullerene. *Langmuir* **2000**, *16*, 3773−3776.

(27) Bai, S.; Debnath, S.; Javid, N.; Frederix, P. W. J. M.; Fleming, S.; Pappas, C.; Ulijn, R. V. Differential Self-Assembly and Tunable Emission of Aromatic Peptide Bola-Amphiphiles Containing Perylene Bisimide in Polar Solvents Including Water. *Langmuir* **2014**, *30*, 7576−7584.

(28) Eggers, P. K.; Fyles, T. M.; Mitchell, K. D. D.; Sutherland, T. Ion Channels from Linear and Branched Bola-Amphiphiles. *J. Org. Chem.* **2003**, *68*, 1050−1058.

(29) Ray, S.; Das, A. K.; Banerjee, A. pH-Responsive, Bolaamphiphile-Based Smart Metallo-Hydrogels as Potential Dye-Adsorbing Agents, Water Purifier, and Vitamin B12 Carrier. *Chem. Mater.* **2007**, *19*, 1633−1639.

(30) Graf, G.; Drescher, S.; Meister, A.; Dobner, B.; Blume, A. Self-Assembled Bolaamphiphile Fibers Have Intermediate Properties between Crystalline Nanofibers and Wormlike Micelles: Formation of Viscoelastic Hydrogels Switchable by Changes in pH and Salinity. *J. Phys. Chem. B* **2011**, *115*, 10478−10487.

(31) Van Bogaert, I. N. A.; Saerens, K.; De Muynck, C.; Develter, D.; Soetaert, W.; Vandamme, E. J. Microbial Production and Application of Sophorolipids. *Appl. Microbiol. Biotechnol.* **2007**, *76*, 23−34.

(32) Baccile, N.; Babonneau, F.; Jestin, J.; Pehau-Arnaudet, G.; Van Bogaert, I. N. A. Unusual, pH-Induced, Self-Assembly of Sophorolipid Biosurfactants. *ACS Nano* **2012**, *6*, 4763−4776.

(33) Baccile, N.; Pedersen, J. S.; Pehau-Arnaudet, G.; Van Bogaert, I. N. A. Surface Charge of Acidic Sophorolipid Micelles: Effect of Base and Time. *Soft Matter* **2013**, *9*, 4911−4922.

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

(34) Zhou, S.; Xu, C.; Wang, J.; Gao, W.; Akhverdiyeva, R.; Shah, V.; Gross, R. Supramolecular Assemblies of a Naturally Derived Sophorolipid. *Langmuir* **2004**, *20*, 7926−7932.
(35) Dhasaiyan, P.; Banerjee, A.; Visaveliya, N.; Prasad, B. L. V. Influence of the Sophorolipid Molecular Geometry on Their Self-Assembled Structures. *Chem. - Asian J.* **2013**, *8*, 369−372.
(36) Penfold, J.; Chen, M.; Thomas, R. K.; Dong, C.; Smyth, T. J. P.; Perfumo, A.; Marchant, R.; Banat, I. M.; Stevenson, P.; Parry, A.; Tucker, I.; Grillo, I. Solution Self-Assembly of the Sophorolipid Biosurfactant and Its Mixture with Anionic Surfactant Sodium Dodecyl Benzene Sulfonate. *Langmuir* **2011**, *27*, 8867−8877.
(37) Develter, D. W. G.; Lauryssen, L. M. L. Properties and Industrial Applications of Sophorolipids. *Eur. J. Lipid Sci. Technol.* **2010**, *112*, 628−638.
(38) Maingault, M. Use of Sophorolipids and Cosmetic and Dermatological Compositions. Patent No. WO/1995/034282A.
(39) Singh, S.; Patel, P.; Jaiswal, S.; Prabhune, A. A.; Ramana, C. V.; Prasad, B. L. V. A Direct Method for the Preparation of Glycolipid−Metal Nanoparticle Conjugates: Sophorolipids as Reducing and Capping Agents for the Synthesis of Water Re-Dispersible Silver Nanoparticles and their Antibacterial Activity. *New J. Chem.* **2009**, *33*, 646−652.
(40) Sleiman, J. N.; Kohlhoff, S. A.; Roblin, P. M.; Wallner, S.; Gross, R.; Hammerschlag, M. R.; Zenilman, M. E.; Bluth, M. H. Sophorolipids as Antibacterial Agents. *Ann. Clin. Lab. Sci.* **2009**, *39*, 60.
(41) Fu, S. L.; Wallner, S. R.; Bowne, W. B.; Hagler, M. D.; Zenilman, M. E.; Gross, R.; Bluth, M. H. Sophorolipids and Their Derivatives are Lethal Against Human Pancreatic Cancer Cells. *J. Surg. Res.* **2008**, *148*, 77−82.
(42) Baccile, N.; Cuvier, A.-S.; Valotteau, C.; Van Bogaert, I. N. A. Practical Methods to Reduce Impurities for Gram-Scale Amounts of Acidic Sophorolipid Biosurfactants. *Eur. J. Lipid Sci. Technol.* **2013**, *115*, 1404−1412.
(43) Cuvier, A.-S.; Berton, J.; Stevens, C. V.; Fadda, G. C.; Babonneau, F.; Van Bogaert, I. N. A.; Soetaert, W.; Pehau-Arnaudet, G.; Baccile, N. pH-Triggered Formation of Nanoribbons From Yeast-Derived Glycolipid Biosurfactants. *Soft Matter* **2014**, *10*, 3950−3959.
(44) SasView for Small Angle Scattering Analysis. http://www.sasview.org/.
(45) List of model functions developed in the SasView software. http://danse.chem.utk.edu/downloads/ModelfuncDocs.pdf.
(46) Marsh, D. *Handbook of Lipid Bilayers*, 2nd ed.; CRC Press: Boca Raton, FL, 2013; p448.
(47) Novotny, P.; Sohnel, O. Densities of Binary Aqueous Solutions of 306 Inorganic Substances. *J. Chem. Eng. Data* **1988**, *33*, 49−55.
(48) Hayter, J. B.; Penfold, J. An Analytic Structure Factor for Macroion Solutions. *Mol. Phys.* **1981**, *42*, 109−118.
(49) Hansen, J.-P.; Hayter, J. B. A Rescaled MSA Structure Factor for Dilute Charged Colloidal Dispersions. *Mol. Phys.* **1982**, *46*, 651−656.
(50) Bales, B. L.; Zana, R. Characterization of Micelles of Quaternary Ammonium Surfactants as Reaction Media I: Dodeclytrimethylammonium Bromide and Chloride. *J. Phys. Chem. B* **2002**, *106*, 1926−1939.
(51) Berr, S. S.; Caponetti, E.; Johnson, J. S., Jr.; Jones, R. R. M.; Magid, L. J. Small-Angle Neutron Scattering from Hexadecyltrimethylammonium Bromide Micelles in Aqueous Solutions. *J. Phys. Chem.* **1986**, *90*, 5766−5770.
(52) MacKerell, A. D., Jr. Molecular Dynamics Simulation Analysis of a Sodium Dodecyl Sulfate Micelle in Aqueous Solution: Decreased Fluidity of the Micelle Hydrocarbon Interior. *J. Phys. Chem.* **1995**, *99*, 1846−1855.
(53) Durchschlag, H.; Zipper, P. Calculation of the Partial Volume of Organic Compounds and Polymers *Progr. Colloid Polym. Sci.* **1994**, *94*, 20−39.
(54) Cromer, D. T.; Liberman, D. A. Anomalous Dispersion Calculations Near to and on the Long-Wavelength Side of an Absorption Edge. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1981**, *A37*, 267−268.
(55) Anomalous scattering factors calculated using the subroutine library by Brennan and Cowan: Brennan, S.; Cowan, P. L. *Rev. Sci. Instrum.* **1992**, *63*, 850.
(56) Dreier, P.; Rabe, P.; Malzfeldt, W.; Niemann, W. Anomalous X-Ray Scattering Factors Calculated from Experimental Absorption Spectra. *J. Phys. C: Solid State Phys.* **1984**, *17*, 3123−3136.
(57) Patel, M.; Rosenfeldt, S.; Ballauff, M.; Dingenouts, N.; Pontoni, D.; Narayanan, T. Analysis of the Correlation of Counterions to Rod-Like Macroions by Anomalous Small-Angle X-Ray Scattering. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2962−2967.
(58) Dingenouts, N.; Merkle, R.; Guo, X.; Narayanan, T.; Goerigk, G.; Ballauff, M. Use of Anomalous Small-Angle X-Ray Scattering for the Investigation of Highly Charged Colloids. *J. Appl. Crystallogr.* **2003**, *36*, 578−582.
(59) Guilleaume, B.; Blaul, J.; Ballauff, M.; Wittemann, M.; Rehahn, M.; Goerigk, G. The Distribution of Counterions Around Synthetic Rod-Like Polyelectrolytes in Solution. *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2002**, *8*, 299−309.
(60) Sztucki, M.; Di Cola, E.; Narayanan, T. Instrumental Developments for Anomalous Small-Angle X-Ray Scattering from Soft Matter Systems. *J. Appl. Crystallogr.* **2010**, *43*, 1479−1487.
(61) Ballauff, M.; Jusufi, A. Anomalous Small-Angle X-Ray Scattering: Analyzing Correlations and Fluctuations in Polyelectrolytes. *Colloid Polym. Sci.* **2006**, *284*, 1303−1311.
(62) Sztucki, M.; Di Cola, E.; Narayanan, T. Anomalous Small-Angle X-Ray Scattering from Charged Soft Matter. *Eur. Phys. J.: Spec. Top.* **2012**, *208*, 319−331.
(63) Sztucki, M.; Di Cola, E.; Narayanan, T. New opportunities for Anomalous Small-Angle X-Ray Scattering to Characterize Charged Soft Matter Systems. *J. Phys.: Conf. Series* **2011**, *272*, 012004.
(64) Value obtained from the SciFinder database.
(65) Brulet, A.; Lairez, D.; Lapp, A.; Cotton, J.-P. Improvement of Data Treatment in Small-Angle Neutron Scattering. *J. Appl. Crystallogr.* **2007**, *40*, 165−177.
(66) Pasinet 2 software for SANS raw data treatment. http://didier.lairez.fr/pasinet2/doku.php.
(67) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(68) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; et al. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.
(69) Dhasaiyan, P.; Pandey, P. R.; Visaveliya, N.; Roy, S.; Prasad, B. L. V. Vesicle Structures from Bolaamphiphilic Biosurfactants: Experimental and Molecular Dynamics Simulation Studies on the Effect of Unsaturation on Sophorolipid Self-Assemblies. *Chem. - Eur. J.* **2014**, *20*, 6246−6250.
(70) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.
(71) Hansen, H. S.; Hünenberger, P. H. A Reoptimized GROMOS Force Field for Hexopyranose-Based Carbohydrates Accounting for the Relative Free Energies of Ring Conformers, Anomers, Epimers, Hydroxymethyl Rotamers, and Glycosidic Linkage Conformers. *J. Comput. Chem.* **2011**, *32*, 998−1032.
(72) Kukol, A. Lipid Models for United-Atom Molecular Dynamics Simulations of Proteins. *J. Chem. Theory Comput.* **2009**, *5*, 615−626.
(73) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F. *Intermolecular Forces*; ; Pullman, B., Ed.: Reidel: Dordrecht, 1981; p 331.
(74) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(75) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

(76) Rahman, A.; Stillinger, F. H. Molecular Dynamics Study of Liquid Water. *J. Chem. Phys.* **1971**, *55*, 3336−3359.
(77) Parrinello, M. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(78) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2007**, *4*, 116−122.
(79) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, *102*, 5451.
(80) Heinz, T. N.; van Gunsteren, W. F.; Hünenberger, P. H. Comparison of Four Methods to Compute the Dielectric Permittivity of Liquids from Molecular Dynamics Simulations. *J. Chem. Phys.* **2001**, *115*, 1125.
(81) *Water and Biomolecules: Physical Chemistry of Life Phenomena*, Kuwajima, K., Goto, Y., Hirata, F., Terazima, M., Kataoka, M., Eds.; Springer-Verlag: Berlin/Heidelberg, 2009; p221.
(82) *Advances in Sweeteners*, Grenby, T. H., Ed., Blackie Academics and Professionals: Glasgow, 1996; p192.
(83) Dupuy, C.; Auvray, X.; Petipas, C.; Rico-Lattes, I.; Lattes, A. Anomeric Effects on the Structure of Micelles of Alkyl Maltosides in Water. *Langmuir* **1997**, *13*, 3965−3967.
(84) Abel, S.; Dupradeau, F.-Y.; Raman, E. P.; MacKerell, A. D., Jr.; Marchi, M. Molecular Simulations of Dodecyl-β-maltoside Micelles in Water: Influence of the Headgroup Conformation and Force Field Parameters. *J. Phys. Chem. B* **2011**, *115*, 487−499.
(85) Lupi, L.; Comez, L.; Paolantoni, M.; Perticaroli, S.; Sassi, P.; Morresi, A.; Ladanyi, B. M.; Fioretto, D. Hydration and Aggregation in Mono- and Disaccharide Aqueous Solutions by Gigahertz-to-Terahertz Light Scattering and Molecular Dynamics Simulations. *J. Phys. Chem. B* **2012**, *116*, 14760−14767.
(86) Winther, L. R.; Qyist, J.; Halle, B. Hydration and Mobility of Trehalose in Aqueous Solution. *J. Phys. Chem. B* **2012**, *116*, 9196−9207.
(87) Glinski, J.; Burakowski, A. Is the Hydration Number of a Non-Electrolyte Additive With Length and Constituents of The Solute Molecule? *Eur. Phys. J.: Spec. Top.* **2008**, *154*, 275−279.
(88) Mähler, J.; Persson, I. A Study of the Hydration of the Alkali Metal Ions in Aqueous Solution. *Inorg. Chem.* **2012**, *51*, 425−438.
(89) Moreau, G.; Helm, L.; Purans, J.; Merbach, A. E. Structural Investigation of the Aqueous Eu2+ Ion: Comparison with Sr2+ Using the XAFS Technique. *J. Phys. Chem. A* **2002**, *106*, 3034−3043.
(90) Zini, E.; Gazzano, M.; Scandola, M.; Wallner, S. R.; Gross, R. A. Glycolipid biomaterials: Solid-state properties of a poly(sophorolipid). *Macromolecules* **2008**, *42*, 7463−7468.
(91) Cecutti, C.; Focher, B.; Perly, B.; Zemb, T. Glycolipid Self-Assembly: Micellar Structure. *Langmuir* **1991**, *7*, 2580−2585.
(92) Chen, M.; Dong, C.; Penfold, J.; Thomas, R. K.; Smyth, T. J. P.; Perfumo, A.; Marchant, R.; Banat, I. M.; Stevenson, P.; Parry, A.; Tucker, I.; Campbell, R. A. Adsorption of Sophorolipid Biosurfactants on Their Own and Mixed with Sodium Dodecyl Benzene Sulfonate, at the Air/Water Interface. *Langmuir* **2011**, *27*, 8854−8866.
(93) Pabit, S. A.; Meisburger, S. P.; Li, L.; Blose, J. M.; Jones, C. D.; Pollack, L. Counting Ions Around DNA with ASAXS. *J. Am. Chem. Soc.* **2010**, *132*, 16334−16336.
(94) Huber, K.; Scheler, U. New experiments for the quantification of counterion condensation. *Curr. Opin. Colloid Interface Sci.* **2012**, *17*, 64−73.
(95) Jusufi, A.; Kohlmeyer, A.; Sztucki, M.; Narayanan, T.; Ballauff, M. Self-Assembly of Charged Surfactants: Full Comparison of Molecular Simulations and Scattering Experiments. *Langmuir* **2012**, *28*, 17632−17641.
(96) Glatter, O. A New Method for the Evaluation of Small-Angle Scattering Data. *J. Appl. Crystallogr.* **1977**, *10*, 415.
(97) Gerber, S.; Garamus, V. M.; Milkereit, G.; Vill, V. Mixed Micelles Formed by SDS and a Bolaamphiphile with Carbohydrate Headgroups. *Langmuir* **2005**, *21*, 6707−6711.
(98) Milkereit, G. Investigation of colloidal, biophysical and liquid crystalline properties of synthetic alkyl glycosides and glycolipids, Ph.D. thesis, 2006, University of Hamburg, Germany. http://www.chemie.uni-hamburg.de/bibliothek/2006/DissertationMilkereit.pdf.
(99) Abel, S.; Dupradeau, F.-Y.; Marchi, M. Molecular Dynamics Simulations of a Characteristic DPC Micelle in Water. *J. Chem. Theory Comput.* **2012**, *8*, 4610−4623.
(100) Schrödinger. *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2010.
(101) Baccile, N.; Nassif, N.; Malfatti, L.; Van Bogaert, I. N. A.; Soetaert, W.; Pehau-Arnaudet, G.; Babonneau, F. Sophorolipids: A Yeast-Derived Glycolipid as Greener Structure Directing Agents for Self-Assembled Nanomaterials. *Green Chem.* **2010**, *12*, 1564−1567.
(102) Abel, S.; Sterpone, F.; Bandyopadhyay, S.; Marchi, M. Molecular Modeling and Simulations of AOT−Water Reverse Micelles in Isooctane: Structural and Dynamic Properties. *J. Phys. Chem. B* **2004**, *108*, 19458−19466.
(103) Voronoi, G. F. Nouvelles Applications des Paramètres Continus à la Théorie des Formes Quadratiques. *J. Reine Angew. Math.* **1908**, *134*, 198−287.
(104) Rycroft, C. H. VORO++: A Three-Dimensional Voronoi Cell Library in C++. *Chaos* **2009**, *19*, 041111.
(105) Paci, E.; Marchi, M. Intrinsic Compressibility and Volume Compression in Solvated Proteins by Molecular Dynamics Simulation at High Pressure. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 11609−11614.
(106) Kaibara, K.; Iwata, E.; Eguchi, Y.; Suzuki, M.; Maeda, H. Dispersion Behavior of Oleic Acid in Aqueous Media: from Micelles to Emulsions. *Colloid Polym. Sci.* **1997**, *275*, 777−783.
(107) Garland, E. R.; Rosen, E. P.; Clarke, L. I.; Baer, T. Structure of Submonolayer Oleic Acid Coverages on Inorganic Aerosol Particles: Evidence of Island Formation. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3156−3161.

■ NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on October 2, 2015. The abstract graphic was updated. The revised paper was reposted on October 6, 2015.

DOI: 10.1021/acs.jpcb.5b05374
J. Phys. Chem. B 2015, 119, 13113−13133

Supporting Information

Structure of Bolaamphiphile Sophorolipid Micelles Characterized with SAXS, SANS and MD Simulations

*Sabine Manet,[1] Anne-Sophie Cuvier,[1] Claire Valotteau,[1] Giulia C. Fadda,[2] Javier Perez,[3] Esra Karakas,[4,5] Stéphane Abel,[5] and Niki Baccile[1]**

[1] Sorbonne Universités, UPMC Univ Paris 06, CNRS, Collège de France, UMR 7574, Chimie de la Matière Condensée de Paris, F-75005, Paris, France.

[2] Laboratoire Léon Brillouin, LLB, CEA Saclay, F-91191 Gif-sur-Yvette Cedex, France

[3] SWING, Synchrotron Soleil, BP 48, F-91192 Gif-sur-Yvette

[4] Maison de la Simulation, USR 3441, CEA – CNRS – INRIA – Univ. Paris-Sud – Univ. de Versailles, 91191, Gif-sur-Yvette Cedex, France

[5] Institute for Integrative Biology of the Cell (I2BC), Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Centre National de la Recherche Scientifique (CNRS), Université Paris-Sud, CEA-Saclay, F-91191 Gif-sur-Yvette, France

Table S1: Solvent SLD calculated for all salt compositions

Solvent	ρ_{solv} (10^{-6} Å^{-2})
RbCl-10mM	9.4091
RbCl-50mM	9.4343
RbCl-100mM	9.4656
SrCl2-5mM	9.4083
SrCl2-10mM	9.4301
SrCl2-20mM	9.4572
RbOH-40mM	9.4357
RbOH-80mM	9.4646
$Sr(OH)_2$-10mM	9.4325
$Sr(OH)_2$-20mM	9.4581

Critical determination of the best fit model

Figure S1 in the Supporting Information shows χ^2/Npts value obtained from the fitting of the SAXS profile corresponding to sample 7 ($SrCl_2$-25 mM), chosen as a typical example. First of all, four different morphologies without polydispersity have been tested using the corresponding models proposed in the SASview software: sphere, ellipsoid, cylinder and bicelle, with a single core and core-shell (referred to with C.S. in Figure S1) SLD levels. As it was expected, single core models give poor results with high χ^2/Npts values and they have been immediately discarded. According to our hypothesis, these models cannot account for the structure of sophorolipids micelles, which have to be described by a core-shell model, with an apolar core formed by the oleic acid hydrocarbon moiety and a hydrated shell formed by the hydrophilic sophorose headgroup.

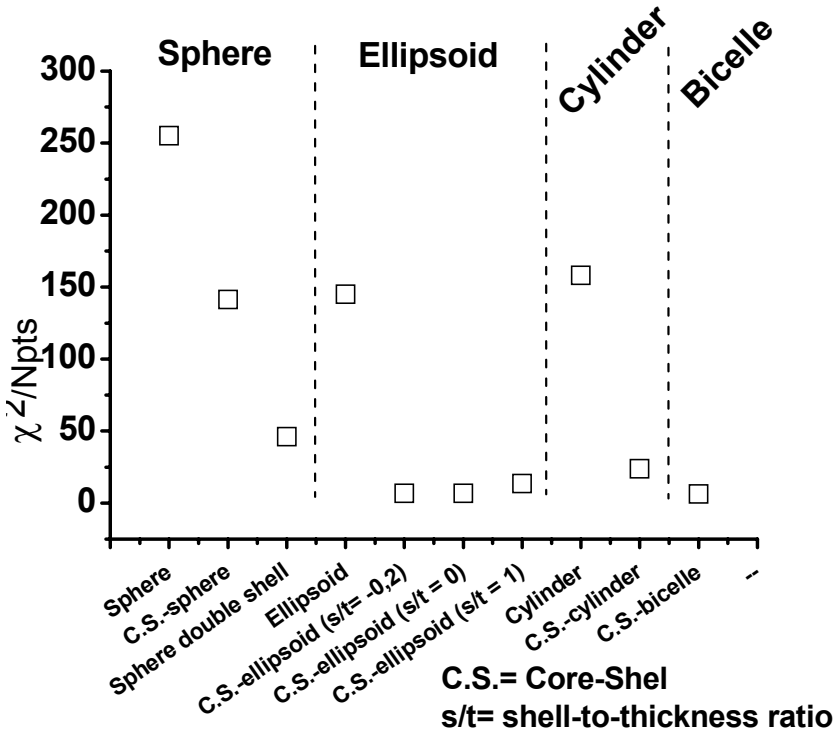


Figure S1 - Evolution of the χ^2/*Npts* for the system $SrCl_2$-25mM as a function of the form factor, *P(q)*.

The sphere model always gives high χ^2/Npts values even for the core shell and core double shell models (χ^2/Npts = 141 and 46 respectively). Best χ^2/Npts values are obtained for core-shell cylinder, ellipsoid and bicelle models (23.8, 13.6 and 6.2 respectively), for a homogeneous shell thickness in the cylinder and ellipsoid models, and two homogeneous shell thicknesses for the bicelle model (face-shell and rim-shell).

The bicelle model seems to provide the best fit, but a closer look to the results shows that the face-shell SLD, a variable parameter of the fitting procedure, is similar, or even higher, than the theoretical pure and dehydrated sophorose SLD (~ 1.4 x 10^{-5} $Å^{-2}$, or 514 e^-/nm^3). This value has no physical meaning since the shell contains sophorose and possibly the carboxylic headgroups with

their hydration sphere; the shell may also contain the methyl group (excluded from the model but present in reality) and, possibly, some hydrated CH_2 groups, as also found by Caponetti.[1] In this scenario, the expected SLD values for the shell are lower than what it is found from the fitting process. Moreover, if the fitted face-shell thickness is correct and corresponds to a realistic value (around 11 Å) the rim-shell thickness is 53 Å, a very high value that cannot be explained with convincing arguments, especially if one considers the very low fitted rim-shell SLD, around 9.6 x 10^{-6} $Å^{-2}$, close to the solvent SLD. Finally, the core thickness of the bicelle, around 10 Å, seems too low if we compare this value to the "extended" oleic hydrocarbon chain length (26 Å).[2] This value is probably slightly lower (22-24 Å)[3] due to the bended configuration induced to the *cis*-double bond but, even if one considers that 3 or 4 methylenes are located in the shell, the core thickness cannot be lower than 12-14 Å. For these reasons and after many trials on various sample scattering profiles without better results, we decided to discard this model.

The core-shell cylinder model gives more realistic sizes and SLD values: the core radius, the shell thickness and the shell SLD are respectively estimated by the fitting procedure as 7.5 Å, 12.3 Å and 1.1 x 10^{-5} $Å^{-2}$, or 383 e^-/nm^3. The latter value corresponds roughly to 75 % v/v of water in the shell, by neglecting any contribution other than water and sophorose. However, in spite of many trials involving polydispersity for different sample profiles, the fit with a core-shell form factor cannot describe properly the bottom of the first oscillation, and the χ^2/Npts values obtained are always higher than 20, so that this model has been discarded, too.

[1] Caponetti, E.; Chillura-Martino, D.; La Mesa, C.; Muzzalupo, R.; Pedone, L. Structural and Transport Properties of Bola C-16 Micelles in Water and in Aqueous Electrolyte Solutions, *J. Phys. Chem. B* **2004**, 108, 1214-1223

[2] Kaibara, K.; Iwata, E.; Eguchi, Y.; Suzuki, M.; Maeda, H. Dispersion Behavior of Oleic Acid in Aqueous Media: from Micelles to Emulsions, *Colloid Polym. Sci* **1997**, 275, 777-783

[3] Garland, E. R.; Rosen, E. P.; Clarke, L. I.; Baer, T. Structure of Submonolayer Oleic Acid Coverages on Inorganic Aerosol Particles: Evidence of Island Formation, *Phys. Chem. Chem. Phys*. **2008**, 10, 3156-3161

Table S2: Fit parameters obtained for X_S= 0.3 using the core-shell ellipsoid of revolution model presented in the main text. For the significance of the symbols, please refer to Figure 2 and to the legends of Table 3 and Table 4 in the main text.

Sample	N_{Agg}	R_C (Å)	X_C	T_S (Å)	ρ_S x 10^{-5} (Å^{-2})	α	$x_{v,w}$ hg_3	$n_{w/SL}$ hg_3	z
RbCl-10mM	55	8.2	5.7	11.7	1.0763	0.67	0.53	22	
RbCl-50mM	64	8.2	6.9	11.5	1.0787	0.69	0.53	21	
RbCl-100mM	68	8.1	7.4	11.3	1.0850	0.67	0.50	19	
$SrCl_2$-5mM	52	7.9	6.1	12.1	1.0623	0.68	0.58	27	
$SrCl_2$-25mM	71	7.9	8.6	12.2	1.0620	0.68	0.58	27	
$SrCl_2$-50mM	73	8.0	8.2	11.7	1.0763	0.67	0.52	22	
RbOH-40mM	34	7.0	4.7	11.8	1.0638	0.55	0.53	23	7.0
RbOH-80mM	24	6.4	4.0	10.6	1.0736	0.50	0.47	19	10.2
$Sr(OH)_2$-10mM	50	7.4	6.1	12.4	1.0644	0.57	0.53	23	4.8
$Sr(OH)_2$-20mM	41	7.2	4.6	12.0	1.0797	0.49	0.44	17	5.2

Table S3: Molecular area values obtained for X_S= 0.3 using the core-shell ellipsoid of revolution model presented in the main text. Symbols are explained in the legend of Table 5 in the main text

Sample	$A_{C,S/SL}$ (Å)	$A_{alkyl/sopho}$ (Å)	$A_{mic,w/SL}$ (Å)	L (Å)
RbCl-10mM	70	100	192	3.67
RbCl-50mM	72	102	191	3.62
RbCl-100mM	71	101	184	3.42
$SrCl_2$-5mM	73	107	209	4.13
$SrCl_2$-25mM	74	105	205	4.15
$SrCl_2$-50mM	71	102	190	3.67
RbOH-40mM	68	122	217	3.96
RbOH-80mM	69	140	226	3.61
$Sr(OH)_2$-10mM	66	110	201	3.96
$Sr(OH)_2$-20mM	59	111	187	3.42

Experimental approach and data treatment in ASAXS experiments

ASAXS measurements in this work have been done using the same experimental apparatus described in the SAXS section above. In addition to that, we have verified the following in order to optimize observation of the anomalous effect: 1) we verified that sophorolipids micelles did not undergo beam damage over about 2 minutes of continuous irradiation, as this is the maximum irradiation time per sample; 2) we averaged three measurements (total acquisition time per measurement is equal to 2 s) per sample; 3) the fluorescence problem was dealt with in two independent ways. In the first method, for each measured energy, the same background level was recorded always using a solution with the exact ion concentration used for a given sample. In practice, a given ion solution is always split in two shares, one is used for background analysis, while the sample is dispersed in the second share for measurement. After background subtraction, there is no residual energy-dependent fluorescence. In the second method, we followed a more classical approach described in the literature (see section III.B2 in Ref. 4). We subtracted to all energies the background (capillary + solvent) recorded far from the adsorption edge, where fluorescence is neglected and verified the presence of fluorescence as a function of the incident energy (Figure S2a in the Supplementary Information). Then, by considering that any shift in the background level close to the edge is solely due to the presence of fluorescence, we subtract it from each spectrum (Figure S2b). Both methods provide equivalent results. 4) We calculated $f'(E)$ and $f''(E)$ directly from the experimental absorption curve measured for standard RbCl (0.5 M) and $SrCl_2$ (0.25 M) water solutions. To do so, we have used the CHOOCH software package[5] directly on our data and interpolated the values using the instrumental resolution ($\Delta E/E = 1.3 \times 10^{-4}$); 5) we rescaled the experimental absorption edge for both Rb^+ and Sr^{2+} to their theoretical values (E_{Rb}= 15199 eV; E_{Sr}= 16105 eV). The measured edge was shifted by a value of approximately 34 eV due small shift of the monochromator main angle, which is a classical, largely documented, effect;[6] 6) we measured $I_0(q,E)$ for eight energy values, shown in Table S4 along with the ΔE, difference between the given energy with respect to the value at the edge, and the corresponding f' and f''. The treatment of $I_0(q,E)$ scattering curves was done using the ASAXS utility in the SAXSUtilities software,[7] available free of charge on the developer's website.[8] This software proposes both the

[4] Dingenouts, N.; Patel, M.; Rosenfeldt, S.; Pontoni, D.; Narayanan, T.; Ballauff, M. Counterion Distribution around a Spherical Polyelectrolyte Brush Probed by Anomalous Small-Angle X-ray Scattering, *Macromolecules* **2004**, 37, 8152-8159

[5] Evans, G.; Pettifer, R. F. CHOOCH: a Program for Deriving Anomalous-Scattering Factors From X-Ray Fluorescence Spectra, *J. Appl. Cryst.* **2001**, 34, 82-86

[6] Ballauff, M.; Jusufi, A. Anomalous Small-Angle X-Ray Scattering: Analyzing Correlations and Fluctuations In Polyelectrolytes, *Colloid Polym Sci*, **2006**, 284, 1303–1311

[7] Sztucki, M.; Di Cola, E.; Narayanan, T. Instrumental Developments for Anomalous Small-Angle X-Ray Scattering from Soft Matter Systems, *J. Appl. Cryst.* **2010**, 43, 1479–1487

[8] http://www.sztucki.de/SAXSutilities

"matrix" and the "fit" methods to extract the pure resonant curve $v^2(q)$ profile. In Table S4 we show the energies at which $I_0(q)$ values were recorded and which were used to extract the $v^2(q)$ scattering profiles.

Table S4 – Energy values of the X-ray beam at which SAXS data have been recorded for each system presented in Table 1 in the main text. ΔE indicates the difference between the given energy and the absorption energy of a given cation. f' and f'' are the real and imaginary part of the energy-dependent scattering vector (please refer to Eq. 17 in the main text).

ΔE (eV)	E_{Rb+} (eV)	f'	f''	ΔE (eV)	E_{Sr2+} (eV)	f'	f''
-600	**14599**	-2.7320	0.5513	-600	**15505**	-2.7636	0.5616
-200	**14999**	-3.7932	0.5248	-200	**15905**	-3.8286	0.5360
-20	**15179**	-6.0708	0.6575	-8	**16097**	-7.9670	0.8747
-4	**15195**	-7.9606	1.6529	-4	**16101**	-8.8549	1.9580
0	**15199**	-8.1150	3.4633	0	**16105**	-8.9079	3.9241
2	**15201**	-7.4691	4.0922	2	**16107**	-8.3326	4.8877
4	**15203**	-6.9082	4.1870	4	**16109**	-7.4539	5.5104

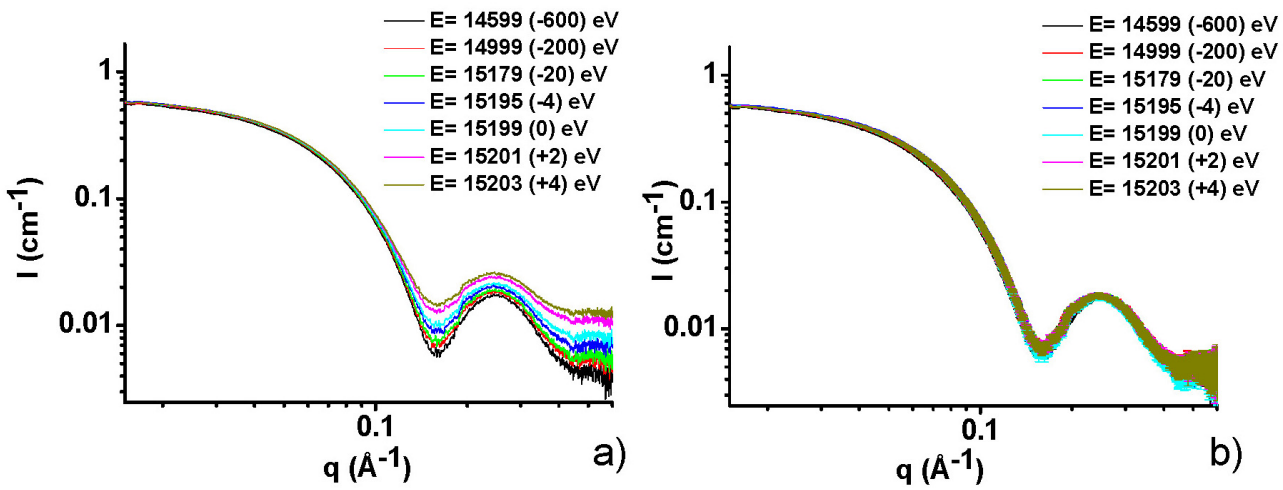


Figure S2 – Illustration of the fluorescence background as a function of the incident energy for the sample SL 80mM and RbCl 10 mM. In a), all spectra have been corrected for the background (capillary, water, RbCl 10mM) recorded at E= 14599 eV. The increasing background as a function of energy (more pronounced at energy values close and above the edge) shows a classical fluorescence effect. In b), we applied the correction described by Dingenouts *et al.* (section III.B2 in N. Dingenouts, M. Patel, S. Rosenfeldt, D. Pontoni, T. Narayanan, M. Ballauff, Macromolecules 2004, 37, 8152-8159).

Experimental determination of the ionization degree.

A 80 mM sophorolipid solution is titrated using micromolar amounts of a RbOH 1M solution (5 μL to 20 μL) from the equilibrium pH to pH= 11:

$$COOH + RbOH \rightarrow COO^{-}Rb^{+} + H_2O \qquad \text{Eq. S1}$$

From the titration curve, one can determine the pKa of sphorolipids as pKa= pH(V_{eq}/2), where V_{eq}/2 is half the equivalent volume at which all COOH are titrated. Then, from Eq. S1, the evolution of the $\frac{[COO^-]}{[COOH]}$ ratio with pH is given by:

$$\frac{[COO^-]}{[COOH]} = 10^{-(pKa-pH)} = k \qquad \text{Eq. S2}$$

The ionization degree, $\frac{[COO^-]}{[COOH]+[COO^-]}$, is obtained from a simple rearrangement of Eq. S2:

$$\frac{[COO^-]}{[COOH]+[COO^-]} = 1 - \frac{1}{k+1} \qquad \text{Eq. S3}$$

where k has been defined in Eq.S2.

At the equilibrium pH of SL in water, which lies between 4.5 and 5 according to concentration, one expects, from the titration curve of SL in water using RbOH (Figure S3), an amount of COO^- groups below 5% of the total carbonyl content. Upon addition of RbOH, the amount of COO^- in solution is expected to increase, as shown by the evolution of the ionization degree with pH (Figure S3).

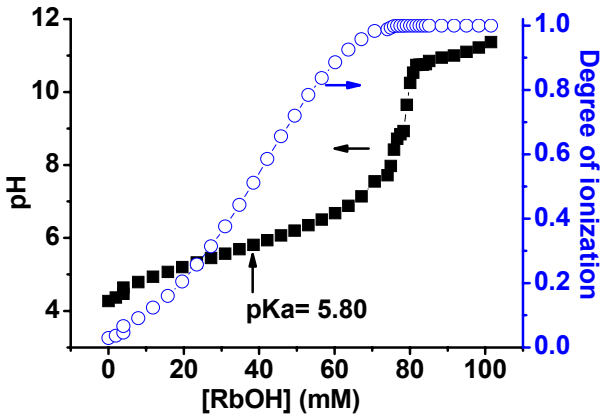


Figure S3 – Titration curve (black squares) of a 80 mM solution of acidic sophorolipids using RbOH. The ionization degree (blue empty circles) is calculated from the titration curve according to Eq. S3.

Additional data obtained from molecular modelling by Molecular Dynamics.

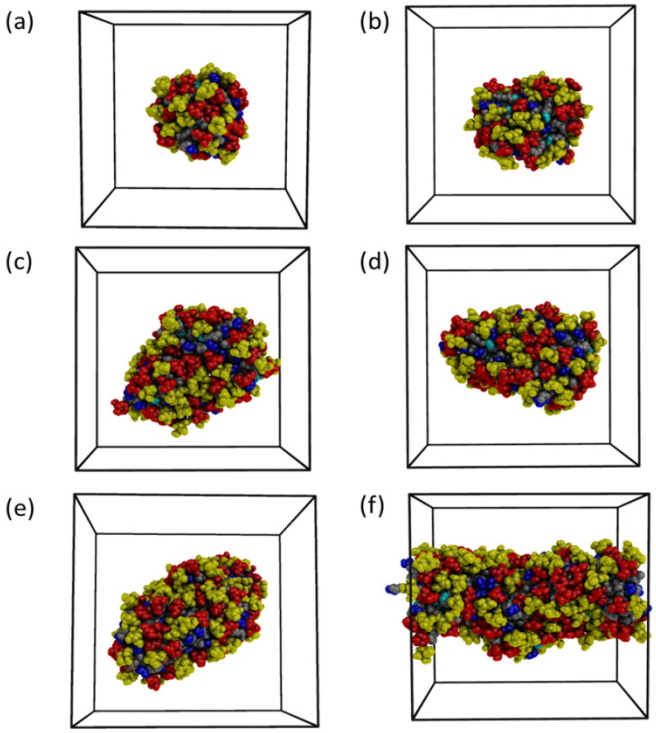


Figure S4: Final snapshots of the simulations obtained for (a) MD1 (with 28 SL at t_{sim} = 112 ns), (b) MD2 (with 37 SL at t_{sim} = 110 ns), (c) MD3 (with 56 SL at t_{sim} = 162 ns), MD4 (with 68 SL at t_{sim} = 164 ns), MD5 (with 80 SL at t_{sim} = 162ns) and MD6 (t_{sim} = 171 ns). The outermost and innermost glucose are colored in yellow and red, respectively, whereas the alkyl chain, double bonds and COOH groups are in grey, cyan and blue colors, respectively. The black line shows the box limits. Water is not shown for visual clarity. The figures were drawn with PyMol.[9]

[9] Schrodinger, L. The PyMOL Molecular Graphics System. *DeLano Scientific San Carlos CA*, 2010

Table S5: Aggregate surface properties. $SASA^V_{GLA}$, $SASA^V_{GLB}$, $SASA^V_{C7A}$, $SASA^V_{C=C}$, $SASA^V_{C7B}$ **and** $SASA^V_{COOH}$ **are the average normalized surface area (in %) of the 6 parts of the sophorolipid molecule and total surface (in Å^2) computed with the Voronoi polyhedron construction (see main).** $SASA^e_{SL}$ **is the total surface areas computed with the Knud Thomsen's Formula, excepted MD6, where we used the formula for a closed cylinder (**$2\pi RL_{box}$ **see main text for more information).** A^e_{SL} **is the surface per sophorolipid (Å^2) and the f_{rug} the rugosity factor. The statistical errors (maximum errors) are always lower than 3.0 and 0.2% for the SASA and the rugosity factor, respectively.**

#MD	$SASA^V_{GLA}$	$SASA^V_{GLB}$	$SASA^V_{C7A}$	$SASA^V_{C=C}$	$SASA^V_{C7B}$	$SASA^V_{COOH}$	$SASA^V_{SL}$	$SASA^e_{SL}$	A^e_{SL}	f_{rug}
MD1	42.6	33.3	8.2	1.2	5.7	9.0	9116.7	4969.0	177.4	1.83
MD2	44.2	33.6	7.4	0.9	4.9	8.9	11401.2	6034.0	163.1	1.89
MD3	45.4	34.1	7.1	0.8	4.4	8.3	15960.3	8024.4	143.4	1.99
MD4	46.3	34.5	6.6	0.6	3.9	8.0	18765.9	9407.5	138.1	2.00
MD5	45.8	34.6	6.3	0.6	3.2	7.9	21820.6	10597.6	132.5	2.05
MD6	47.9	34.7	6.3	0.6	3.2	7.2	27665.6	6996.4	62.5	3.95

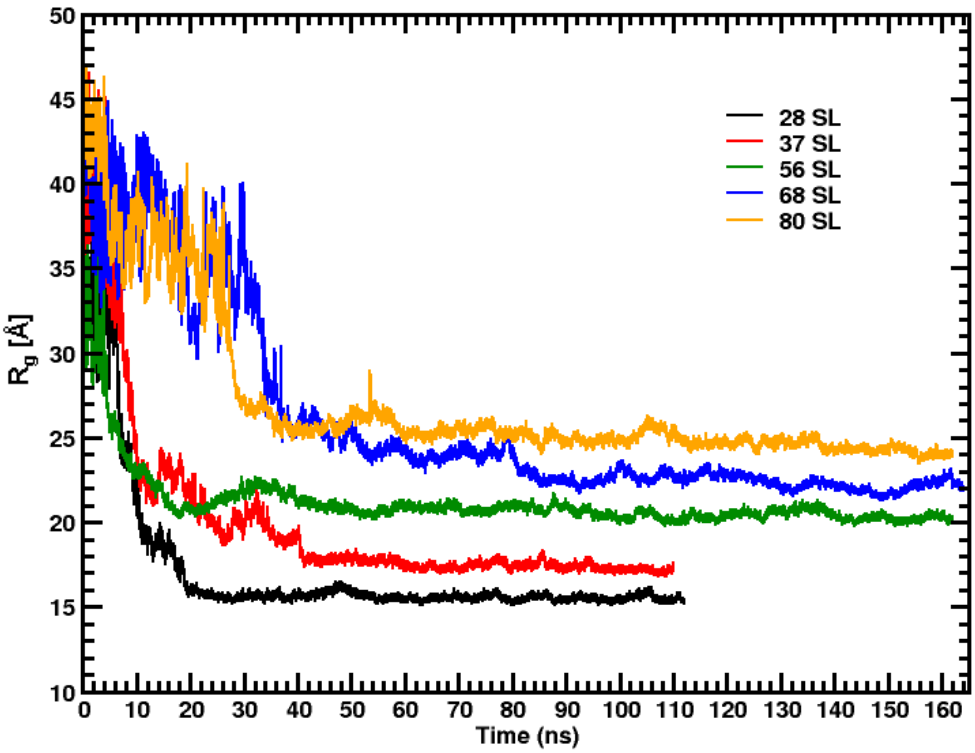


Figure S5: Time evolutions of the SL micelle radius of gyration.

Table S6: Hydration Numbers. N_w/SL, N_w/soph, N_w/COOH, N_w/C=C and N_w/Alkyl give the average number of water molecules at 4.0 Å of the whole aggregate, the sophorose headgroup, the COOH.[10] The statistical errors (maximum errors) are always lower than 0.1%.

System	MD1	MD2	MD3	MD4	MD5	MD6
N_w/SL	26.6	24.2	21.7	20.9	17.8	17.2
N_w/soph	24.9	23.1	20.8	20.2	17.2	16.9
N_w/COOH	4.4	3.9	3.4	3.2	2.7	2.4
N_w/C=C	0.6	0.4	0.4	0.2	0.2	0.2
N_w/Alkyl	5.0	4.0	3.6	3.1	2.7	2.5

[10] S. Abel, F-Y Dupradeau, E. P. Raman, A. D. MacKerell Jr., and M. Marchi, J Phys Chem B., 2011, 115, 487–499

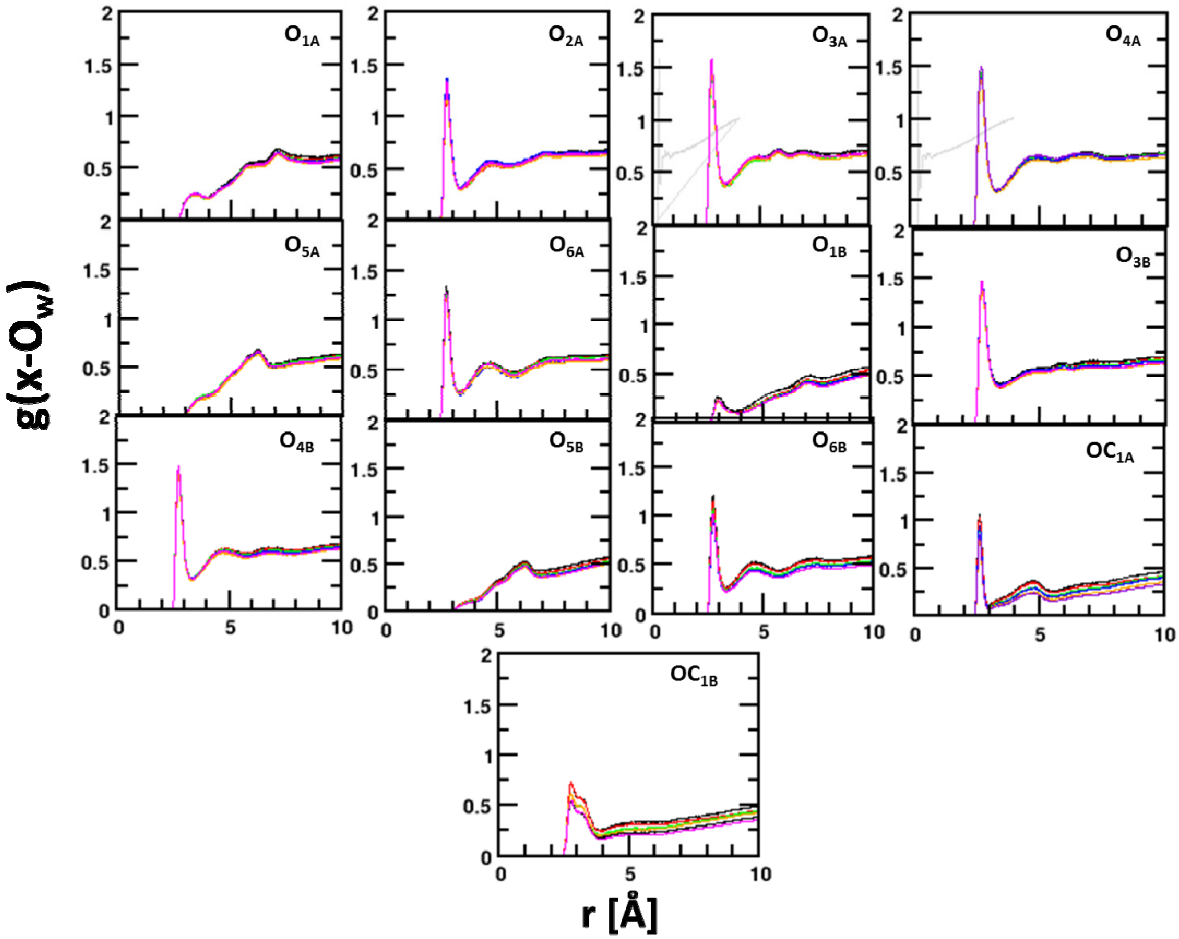


Figure S6: Average radial density functions $g(O_X-O_W)$ between oxygens of water and the sophorolipid for the 28 (black), 37 (red line), 56 (green), 68 (blue), 80 (orange) and 112 (magenta) monomers. See Figure 1 in the main text for the positions of the oxygen atoms in the SL molecule.

Table S7 : Average numbers of water in the first hydration shell $\langle n_{Ox\text{-}Ow} \rangle$ of the sophorolipid oxygen atoms 28 (MD1), 37 (MD2), 56 (MD3), 68 (MD4), 80 (MD5) and 112 (MD6) monomers (see Figure S4). The sophorolipid oxygen atoms are labeled as follows: O_{1A} and O_{1B} are the acetalic oxygens, OC_{1A} and OC_{1B} are the COOH oxygens O_{5A} and O_{5B} are the sophorose ring oxygens, and all the remaining are hydroxyl oxygens. Values have been obtained by integrating the pair correlation functions $g(O_X–O_W)$ up its first minimum at ~3.5 Å (See Figure S6).

$\langle n_{Ox\text{-}Ow} \rangle$	MD1	MD2	MD3	MD4	MD5	MD6
O_{1A}	0.6	0.6	0.6	0.6	0.6	0.5
O_{2A}	2.0	1.9	1.9	2.0	1.8	1.7
O_{3A}	2.4	2.3	2.2	2.3	2.2	2.1
O_{4A}	2.1	2.2	2.0	2.0	1.9	1.9
O_{5A}	0.2	0.3	0.3	0.2	0.2	0.2
O_{6A}	1.9	1.8	1.8	1.7	1.7	1.6
O_{1B}	0.6	0.5	0.5	0.5	0.4	0.4
O_{3B}	2.4	2.3	2.3	2.3	2.2	2.0
O_{4B}	2.2	2.2	2.1	2.0	2	1.9
O_{5B}	0.1	0.1	0.1	0.1	0.2	0.1
O_{6B}	1.8	1.7	1.5	1.4	1.4	1.3
OC_{1A}	1.8	1.8	1.5	1.5	1.3	1.2
OC_{1B}	1.0	1.0	0.9	0.8	0.7	0.7

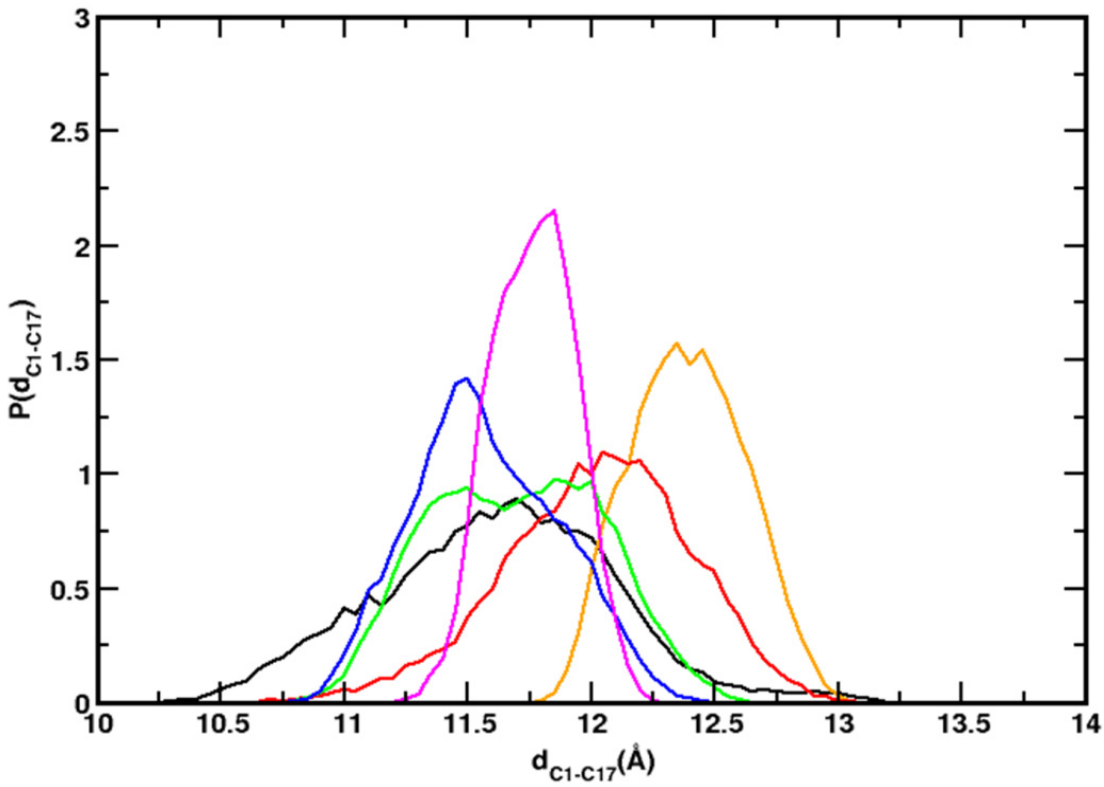


Figure S7: Distribution of the end-to-end distance $P(d_{C1-17})$ between the first and the last carbon of the alkyl of the sophorolipid for the aggregates with 28 (black), 37 (red line), 56 (green), 68 (blue), 68 (orange) and 112 (magenta) monomers

On the Structural and Dynamical Properties of DOPC Reverse Micelles

Stéphane Abel[a, b *], *Nuno Galamba*[d,e], *Esra Karakas*[a, b,c], *Massimo Marchi*[a, b], *Ward H. Thompson*[f], *Damien Laage*[d,e*]

[a] Commissariat à l'Energie Atomique et aux Energies Alternatives, DRF/iBiTEC-S/SB2SM & CNRS UMR 9198, Saclay, France

[b] Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris-Sud, Université Paris-Saclay, F-91198, Gif-sur-Yvette cedex, France.

[c] Maison de la Simulation, USR 3441, CEA-CNRS-INRIA-Univ Paris Sud – Univ Versailles, 91191 Gif-sur-Yvette Cedex, France

[d] Ecole Normale Supérieure, PSL Research University, UPMC Univ Paris 06, CNRS, Département de Chimie, PASTEUR, 24 rue Lhomond, 75005 Paris, France

[e]Sorbonne Universités, UPMC Univ Paris 06, ENS, CNRS, PASTEUR, 75005 Paris, France

[f] Department of Chemistry, University of Kansas, Lawrence, KS 66045, USA

Abstract

The structure and dynamics of phospholipid reverse micelles are studied by molecular dynamics. We report all-atom unconstrained simulations of 1,2-dioleoyl-*sn*-phosphatidylcholine (DOPC) reverse micelles in benzene of increasing sizes, with water to surfactant number ratios ranging from W_o = 1 to 16. The aggregation number, *i.e.*, the number of DOPC molecules per reverse micelle, is determined to fit experimental light-scattering measurements of the reverse micelle diameter. The simulated reverse micelles are found to be approximately spherical. Larger reverse micelles ($W_o > 4$) exhibit a layered structure with a water core and the hydration structure of DOPC phosphate headgroups is similar to that found in phospholipid membranes. In contrast, the structure of smaller reverse micelles ($W_o \leq 4$) cannot be described as a series of concentric layers successively containing water, surfactant headgroups, and surfactant tails, and the headgroups are only partly hydrated and frequently present in the core. The dynamics of water molecules within the phospholipid reverse micelles slow down as the reverse micelle size decreases, in agreement with prior studies on AOT and Igepal reverse micelles. However, the average water reorientation dynamics in DOPC reverse micelles is found to be much slower than in AOT and Igepal reverse micelles with the same W_o ratio. This is explained by the smaller water pool and by the stronger interactions between water and the charged headgroups, as confirmed by the redshift of the computed infrared lineshape with decreasing W_o.

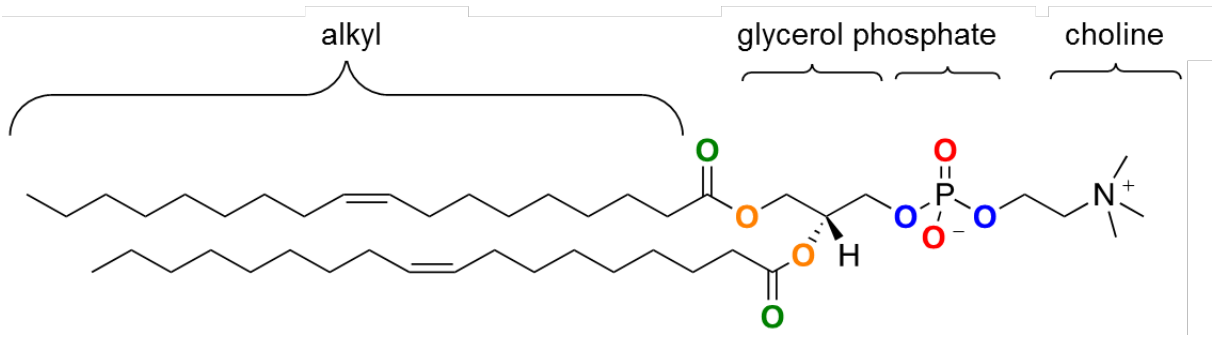
1. Introduction

Reverse micelles (RMs) consist of water droplets sequestered from an organic (*i.e.*, oil) phase by a layer of surfactant molecules. These versatile systems are employed in a broad range of technological applications. They can, for example, serve as vehicles for drug delivery, be used as chemical nanoreactors[1], and be applied to the controlled synthesis of materials and nanoparticles[1,2]. In addition, RMs are paradigm systems to investigate the effects of nanoscale confinement, for example, on the structural and dynamical properties of water [3,4] and on the protein folding and unfolding equilibrium[5-8]. They have been studied with a broad range of techniques, including NMR[9-11], SAXS, SANS [12,13], ultrafast infrared spectroscopy[14-19], and numerical simulations[20-30]. This has, for example, lead to a detailed characterization of the progressive slowdown of water dynamics with decreasing RM sizes and of the specific effects of the surfactant headgroup.

While most of these studies have focused on AOT (dioctyl sodium sulfosuccinate) as a surfactant because of its ability to form stable and monodisperse RMs, phospholipids are another family of surfactants that can form RMs. These RMs are of particular interest because phospholipids are ubiquitous in biochemical systems, as they are the major constituent of cellular membranes. They have also been suggested to have potential functional roles in biological membranes [5].

Among phospholipids, phosphocholine lipids are amphiphilic molecules with a hydrophobic moiety made of two fatty acid chains covalently bonded to a glycerol molecule, and a hydrophilic, zwitterionic headgroup composed of a phosphate anion covalently bonded to a choline cation. Natural phospholipids include lecithin which is a mixture of phosphatidylcholine lipids with acyl chains of various lengths and degrees of saturation, and which is widely used in the pharmaceutical, food and agricultural industries (see, *e.g.*, Refs. [31,32]). Recent series of ultrafast infrared spectroscopy studies[33-38] have focused on the dynamics of water confined in

RMs of DOPC (1,2-dioleoyl-*sn*-phosphatidylcholine, see Scheme 1). These pioneering studies revealed the great similarity between the dynamics of water next to the DOPC phosphate headgroups and the dynamics of water molecules in the vicinity of the phosphate backbone in a DNA double helix.



Scheme 1: Chemical structure of the 1,2-dioleoyl-*sn*-phosphatidylcholine (DOPC) phospholipid. Colors are used to highlight atoms discussed in the text (phosphate oxygen in red, phosphate ether oxygen in blue, glycerol carboxyl oxygen in green, glycerol ester oxygen in orange).

In order to characterize the structure of phospholipid RMs and the dynamics of the confined water molecules, we performed all-atom molecular dynamics (MD) simulations of DOPC RMs in benzene over a broad range of water-to-surfactant, W_0, ratios, from 1 to 16. We determined their structural properties (including their size, shape, and internal structure) and characterized the dynamics of the confined water molecules. We compare our results to existing experimental data on lecithin-based RMs and to other ionic and non-ionic RMs, including the extensively characterized AOT system.

The outline of the remainder of this paper is as follows. In section 2, we describe the construction of the RM systems and the simulation methodology. In section 3, we examine the structure of the RM. In section 4, we compare the water dynamics and infrared spectra obtained from our simulations with the available experimental results. We finally offer some concluding remarks in section 5.

2. Methodology

2.1 Construction of the Reverse Micelle Models

Each RM is characterized by its water-to-surfactant number ratio, W_0. All of the simulated systems have compositions that correspond to locations in the phase diagram where RMs are the most stable structures; RMs thus form spontaneously under these conditions. However, this self-assembling process is complex and too slow to be described with explicit MD simulations at a specific W_0 value (see Ref. [28]). Following prior work[29,30], we therefore adopted a preassembling approach. For this study, we constructed a series of RMs with increasing W_0 ratios from 1 to 16. Details about their composition are given in Tables 1 and S2. The RM models are constructed in three steps. First, a spherical water cluster containing n_{H2O} water molecules is prepared with an initial density chosen to be slightly larger than that of bulk water; this accounts for the fact that some of the water molecules will hydrate the DOPC polar headgroups. Second, n_{DOPC} DOPC molecules are arranged around this water sphere with the fatty acid tails pointing outward in an extended conformation. Third, the RM is inserted in a cubic box of benzene. The number of benzene molecules, n_{benz}, is determined from the lecithin/water/benzene phase diagram in order to be in the L_2 phase region[1], *i.e.*, with a benzene weight fraction $\%wt_{benz} = m_{benz}/(m_{benz}+m_{DOPC}+m_{H2O})$ greater than 80% where m_{benz}, m_{DOPC} , and m_{H2O} are the total masses of benzene, DOPC, and water in the box, respectively. If necessary, n_{benz} is increased to ensure there is more than 15 Å between the outer RM surface and the edges of the simulation box to avoid artifacts due to periodic boundary conditions. The resulting $\%wt_{benz}$ values are between 86.7 and 95.0%, and the effective DOPC concentrations lie between ~0.05 and 0.15 M; the latter values are slightly below the experimental 0.25 M concentration used in Refs. [34-38]).

W_o	Composition	n_{DOPC}	n_{H2O}	$n_{benzene}$	$\%wt_{benz}$	L_{box} (Å)	t_{sim} (ns)
1	Fixed	56	56	4000	87.4	88.0	55
1	Optimized	15	15	2939	95.0	77.7	90
2	Fixed	65	130	4500	86.7	91.7	60
2	Optimized	17	34	3404	95.0	81.6	90
4	Fixed	86	344	6100	86.5	101.6	80
4	Optimized	22	88	4234	95.0	87.1	90
8	Optimized	34	272	7704	95.0	106.6	90
12	Optimized	69	828	16843	95.0	139.2	100
16	Optimized	87	1392	22763	95.0	153.9	110

Table 1: Compositions of the simulated RM systems (see additional systems in Table S2), with the water-to-surfactant ratio W_0, the numbers of DOPC, water, and benzene molecules (n_{DOPC}, n_{H2O}, n_{benz}), the benzene mass fraction ($\%wt_{benzene}$), the size of the simulation box (L_{box}) and the production simulation time (t_{sim}).

A further comparison is provided by the simulation of a fully hydrated DOPC bilayer in the liquid disordered (L_α) phase, with 128 DOPC molecules (64 per leaflet) and 40 water molecules per lipid (see details in the SI).

2.2 Molecular Dynamics Simulations

The MD simulations were performed with GROMACS (v4.6.6)[39-41]. DOPC is described with the CHARMM36 force field[42], benzene with the CHARMM General Force Field (CGenFF)[43], and water with the SPC/E potential[44]. Each micellar system was minimized with the steepest descent algorithm with an energy tolerance of 1000 kJ mol^{-1}. A first constant-temperature equilibration is performed at 300 K for 300 ps with the v-scale thermostat [45] (coupling constant $\tau_T = 0.2$ ps) with the RM (*i.e.*, DOPC and water) and the benzene solvent coupled to two distinct thermostats. An additional equilibration is done in the NPT ensemble

at T = 300 K and P = 1.015 bar with the same thermostat and the Berendsen[46] barostat with a relaxation time of 3.0 ps, and a compressibility of 45×10^{-6} bar^{-1}. During equilibration, the phospholipid and water molecules are harmonically restrained with a 1000 kJ mol^{-1} nm^{-1} force constant. Periodic boundary conditions are applied and long-range electrostatic interactions are evaluated with the particle mesh Ewald method[47]. A switching function for the van der Waals interactions is used between 8.0 Å and 12 Å. Bonds involving hydrogen atoms are constrained with the P-LINCS algorithm[48]. The production runs are propagated at constant temperature and pressure, with the same thermostat as during the equilibration, and the Parrinello-Rahman barostat[49,50]. We stress that no harmonic restraining potential is applied on the RM, so that the RM shape is exclusively determined by the force field. The RM structural properties are calculated from 55 to 110 ns-long NPT trajectories with configurations saved every 2 ps. The dynamical properties are determined from shorter, 0.5 to 6 ns-long, NVT runs at 300 K using the same thermostat as in the NPT runs and with configurations saved every 1 fs.

We note that a recent simulation study[29] of phosphatidylcholine RMs suggested that the CHARMM36 lipid force field employed here underestimates the lipid tail-apolar solvent interactions, leading to distorted RM structures. However, we stress that the aggregation number that was chosen for the only DOPC RM in that study is considerably larger than the optimized aggregation number that we find here (78 for W_0=7 in ref. [29] vs. 34 for W_0 = 8 here; see Table 1) and that lower aggregation number – as will be shown below – is required to yield RM sizes in agreement with experimental measurements. In addition, the main issue with the CHARMM36 force field discussed in ref. [29] was the very limited penetration of the apolar solvent between the lipid tails, whereas, as we will show, our present simulations do not suffer from this problem and a clear solvent penetration is observed.

3. Results and Discussion

3.1 Structure of the Reverse Micelles

3.1.1 Reverse Micelle Size

In experimental studies, only the W_0 water-to-surfactant ratio is controlled and the aggregation number, *i.e.*, the number of DOPC molecules per RM, is not. Here we assume that for a given W_0 ratio, all RMs have the same aggregation number, *i.e.*, that the system is monodisperse (the polydispersity of a DOPC/water/benzene solution, to our knowledge, has not been measured), and we determine this aggregation number to fit the results of dynamic light scattering (DLS) experiments which measured the RM size. DLS measurements on DOPC RMs[34] suggest that the RM hydrodynamic diameter D_H^{exp} (in Å) scales linearly with W_0 as

$$D_H^{exp}(W_O) = 2.8\, W_O + 32.6\ \text{Å} \quad . \tag{1}$$

To determine the aggregation number at every W_0 value, we have used two different strategies. First, following the suggestion of Ref. [34], we calculated the number of DOPC molecules per RM by dividing the RM surface area (assumed to be an ideal sphere) by the surface area of a single DOPC molecule, S_{DOPC}^{exp} = 70 Å^2, estimated from the area-per-lipid in membranes[51-53]. The numbers of DOPC and water molecules in an individual RM are then given by

$$\begin{aligned} n_{DOPC}(W_o) &= \frac{4\pi\left[D_H^{exp}(W_0)/2\right]^2}{S_{DOPC}^{exp}} \\ n_{H_2O}(W_o) &= W_o\, n_{DOPC}(W_o) \end{aligned} \quad . \tag{2}$$

Following this approach, we constructed and simulated three RM systems at W_0 = 1, 2, and 4. These RMs are identified has having a "fixed" aggregation number in Tables 1-2. We then compared their sizes with experimental DLS measurements.[34] DLS probes the decay of the scattered intensity time-correlation function and is thus sensitive to the particle hydrodynamic radius. The hydrodynamic radius, R_H, is calculated from the measured translational diffusion

coefficient and the Stokes-Einstein equation, assuming that the particle is spherical[54]. For non-spherical particles, it yields the apparent hydrodynamic radius, *i.e.*, the radius of the equivalent sphere having the same volume. This apparent radius is a very good approximation to the hydrodynamic radius for moderately distorted spheres, as is the case for the RMs studied here [55,56]. In our simulations, we estimate the RM apparent hydrodynamic diameter D_H^{sim} from the radius of gyration, R_g. For spherical RMs with a uniform density, the radius of gyration is defined to be

$$R_g = \left[\frac{\int_0^{R_H} r^2 4\pi r^2 dr}{\int_0^{R_H} 4\pi r^2 dr}\right]^{1/2} = \sqrt{\frac{3}{5}} R_H$$

where R_H is the hydrodynamic radius, and therefore

$$D_H^{sim} = 2\sqrt{5/3}\, R_g \quad . \tag{3}$$

As shown in Fig. 1 and in Table 2, the diameters of these RMs constructed with the aggregation number determined from Eq. 2 are significantly larger than those determined by DLS. This shows that the aggregation number given by Eq. 2 is overestimated. As discussed below, because of curvature effects, the surface area per DOPC molecule in RMs cannot be assumed to be independent of W_0 and equal to its value in membranes.

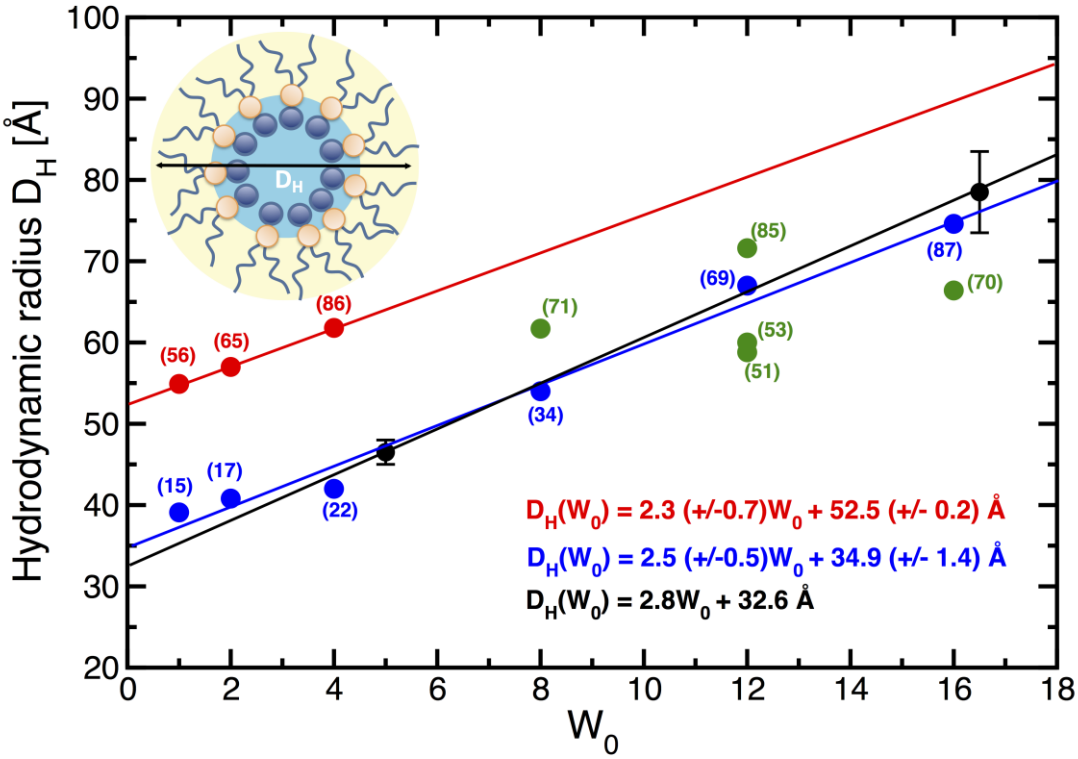


Figure 1: Reverse micelle hydrodynamic diameters D_H versus W_0, respectively from DLS measurements[34] (black), from the simulated RMs constructed with the aggregation number Eq. 2 assuming a fixed headgroup surface area (red), and from our trial-and-error optimized construction (in green our intermediate attempts, in blue the optimized systems). Each system is shown by a dot, and the solid lines are linear regressions for each set. The n_{DOPC} value for each simulated RM is given in parentheses.

We therefore determined the n_{DOPC} aggregation numbers with a second approach, using an iterative trial-and-error procedure to fit the experimental hydrodynamic diameter. We constructed 14 RM systems with W_0 ratios of 1, 2, 4, 8, 12, and 16. The compositions and apparent hydrodynamic diameters, Eq. 3, of the optimized systems are listed in Tables 1-2, where they are identified as having an "optimized" aggregation number.

W_0	Composition	R_g^{RM}	R_g^{W}	D_H^{sim}	D_H^{exp}	R_W^{sim}	d_{Tail}	S_{DOPC}^{outer}	S_{DOPC}^{inner}
1	Optimized	15.2	7.2	39.3	35.4	8.8	10.9	323.5	64.9
1	Fixed	21.4	9.6	55.0		11.7	15.8	170.0	30.7
2	Optimized	15.6	8.2	40.5	38.2	10.1	10.1	302.80	75.4
2	Fixed	22.1	11.2	56.9		13.9	14.6	156.7	37.3
4	Optimized	16.4	9.3	42.3	42.8	11.7	9.5	255.7	78.2
4	Fixed	24.4	15.3	62.9		18.2	13.3	144.6	48.4
8	Optimized	21.1	12.5	55.3	55.8	15.9	11.3	272.7	93.4
12	Optimized	25.4	16.3	65.4	66.2	20.8	11.9	191.9	78.8
16	Optimized	28.9	19.2	74.7	77.4	24.7	12.7	250.4	88.1

Table 2: Structural properties of the simulated RM systems for increasing water/surfactant W_0 ratio and for an aggregation number either fixed to the value given by Eq. 2 or optimized to reproduce the DLS measurements: radii of gyration of the entire RM (including DOPC and water) and of the water core (R_g^{RM} and R_g^{W}), apparent hydrodynamic diameter from our simulations, Eq. 3, (D_H^{sim}), and from the DLS experiments[34] (D_H^{exp}), radius of the water core R_W^{sim}, thickness of the DOPC acyl chain layer d_{Tail}, and surface areas per surfactant at the DOPC/benzene outer interface S_{DOPC}^{outer} (determined from D_H^{sim}) and at the DOPC/water inner interface S_{DOPC}^{inner} (from R_W^{sim}). Distances are in Å, areas in Å^2, and relative standard deviations are less than 2%.

As shown in Fig. 1, the D_H^{sim} hydrodynamic diameters of these optimized RMs are all within experimental error (estimated to be ± 1.5 Å for RM with $W_0 = 5.5$[34]) of the values obtained by DLS[34]. In addition, our results correctly reproduce the linear dependence of the RM size with the water-surfactant-ratio.

Our results in Table 2 also reveal the reason why the aggregation number cannot be deduced from the surface area per DOPC molecule in a membrane (Eq. 2). In every RM, because of the interfacial curvature, the surface area occupied by each DOPC molecule is extremely dif-

ferent at the inner interface with water and at the outer interface with benzene. The surface area determined in a membrane and used in Ref. [34] (approx. 70 Å^2 from our simulations, see SI section 1.2.1, a value consistent with the ~61 – 66 Å^2 area from SANS measurements on phosphatidylcholine microemulsions in cyclohexane[57]) is close to the value found at the inner interface of optimized RMs. However, since the hydrodynamic radius is related to the outer interface, what is needed to determine the aggregation number is the surface area at the outer interface, which is much greater than the value in a membrane.

3.1.2 Reverse Micelle Shape

In the remainder of this paper, we will exclusively discuss the RMs with optimized aggregation numbers that match the experimental hydrodynamic radii. We first analyze the shapes of these RMs. Figure 2 shows typical conformations of these RMs.

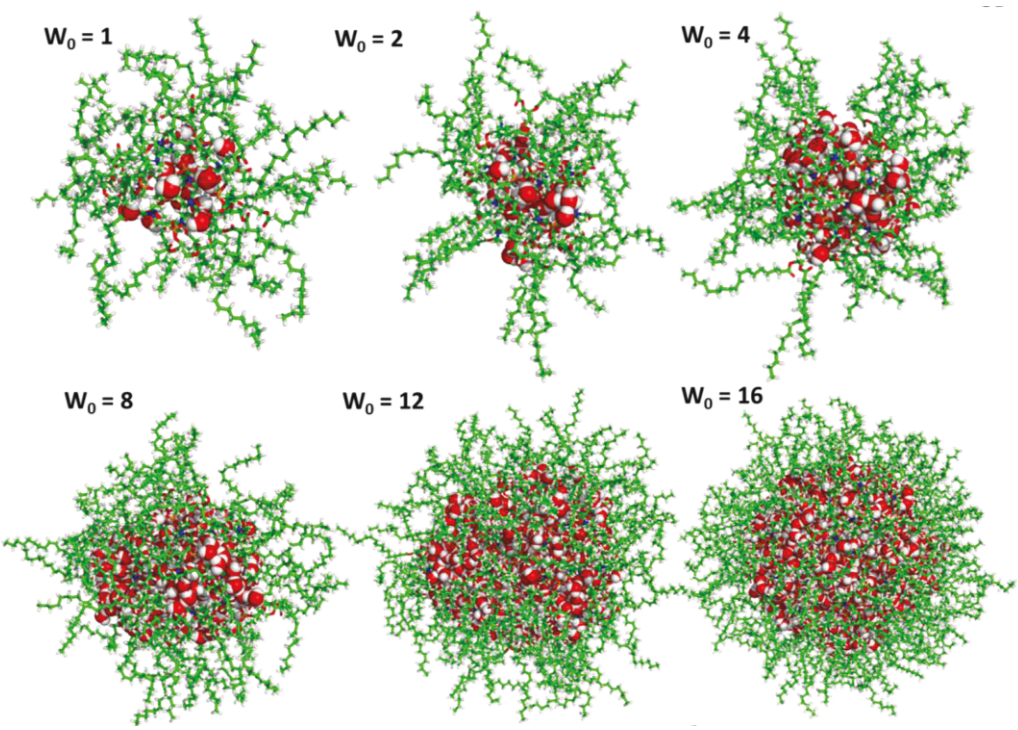


Figure 2: Typical RM configurations from our MD simulations. DOPC molecules are represented as green sticks, water oxygen and hydrogen atoms are shown as red and white spheres, respectively.

To characterize the shape of each RM (including water), we approximate the RM by an ellipsoid of uniform density, whose three semi-axis lengths, a, b, and c can be determined from the principal moments of inertia as[20]

$$\begin{aligned} I_1 &= \frac{1}{5} M_T (a^2 + b^2) \\ I_2 &= \frac{1}{5} M_T (a^2 + c^2) \\ I_3 &= \frac{1}{5} M_T (b^2 + c^2) \end{aligned} \tag{4}$$

where M_T, I_1, I_2, and I_3 (with $I_1 > I_2 > I_3$) are the RM total mass (including water and DOPC molecules) and the principal moments of inertia of the overall RM. For a sphere $a = b = c$, while for an oblate ellipsoid $a = b > c$ and for a prolate one $a > b = c$. We find that for each RM, the RM is approximately spherical and the fluctuations of the semi-axis lengths are very limited during the tens of nanoseconds of our simulations (Fig. S7). For an ellipsoid with three different semi-axes, *i.e.*, when there is no symmetry of revolution, the distortion from the ideal spherical geometry can be measured by the eccentricity e defined here as [27,29,58]

$$e = \sqrt{1 - \frac{c^2}{a^2}} \tag{5}$$

where e ranges from 0 for a sphere to 1 for a rod or a disk (we note that other definitions have been used[59], which would lead here to the same conclusions). Our results in Fig. 3 indicate that the larger (*i.e.*, $W_0 > 4$) RMs are more spherical than the smaller RMs, in agreement with the trend measured experimentally for lecithin RMs[60]. The RMs do not exhibit a clear symmetry of revolution and cannot be characterized as either prolate or oblate.

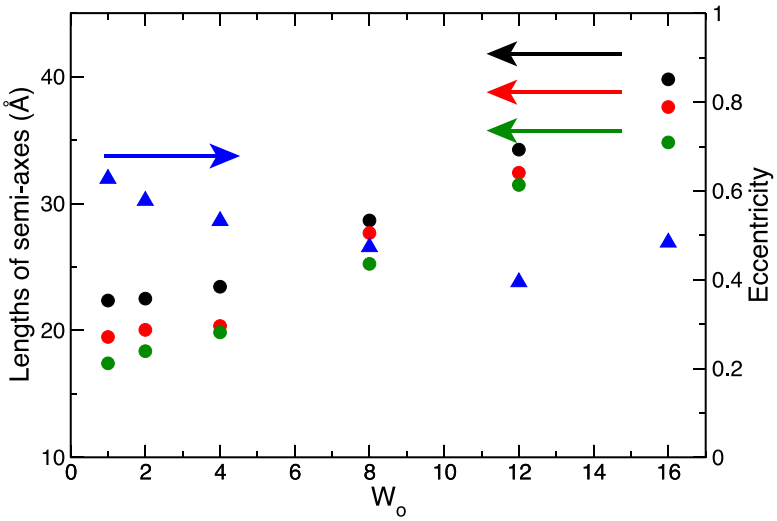


Figure 3: Length of the three RM semi-axes, respectively *a* (black), *b* (red), and *c* (green), Eq. 4, (left axis) and eccentricity, Eq. 5, (blue, right axis) for increasing W_0 (standard deviations are less 0.2 Å for the semi-axis lengths and 0.05 for the eccentricity).

We have also investigated how the shape and the volume of the RMs fluctuate with time. Figure 4 shows the time-correlation functions of the eccentricity eq. 5 and RM volume (defined as $(4\pi/3)abc$) fluctuations with respect to their average value for each RM. The structural fluctuations are found to occur on a slow, nanosecond timescale. While no clear trend is visible with the RM size, very small and very large RMs seem to exhibit slower structural relaxation.

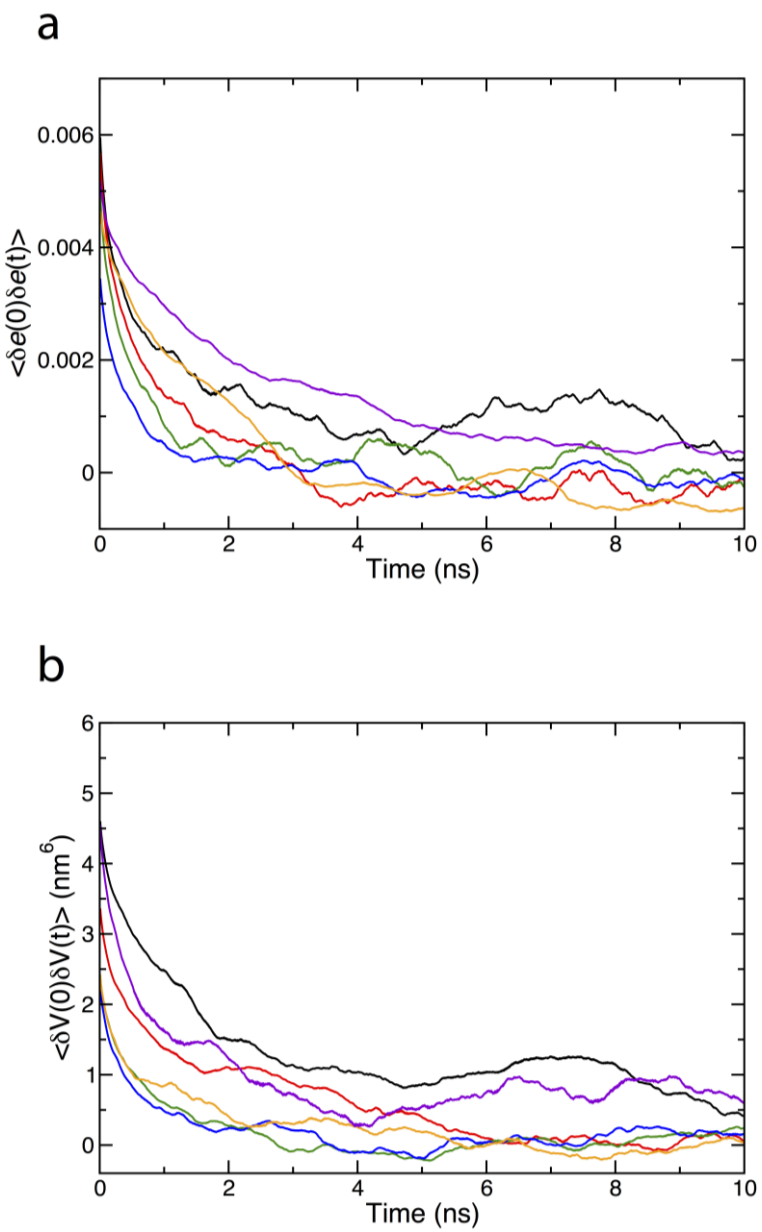


Figure 4: Time-correlation functions of the RM a) eccentricity eq 5 and b) volume fluctuations with respect to their average values (W_o=1 in black, 2 in red, 4 in green, 8 in blue, 12 in orange and 16 in violet).

3.1.3 Deviations from the Ideal Layered Structure

We now turn our attention to the study of the RM internal structure. In particular, we examine the extent to which the RMs differ from the simplified ideal picture of a spherical water core sequestered from the apolar solvent by a shell of surfactant molecules. Figure 5 shows the mean radial mass density profiles along the distance to the RM center-of-mass, for the

different system components, *i.e.*, the water molecules, the DOPC choline ($N^+(CH_3)_3(CH_2)_2$) and phosphate (PO_4^-) moieties, the DOPC glycerol and acyl chains, and the benzene solvent molecules. As discussed in previous reports (*e.g.,* Refs. [20,23,61]), the non-spherical nature of RMs, particularly RMs with $W_0 \leq 8$, may affect the interpretation of such density profiles, which are most meaningful for systems with spherical (or near-spherical) symmetry. The results in Fig. 5 show that the cartoon picture of concentric and separate spherical layers containing successively the water core, the DOPC molecules, and the benzene solvent is too simplified. We find a considerable overlap between the water, DOPC, and benzene components for all W_0 values, whose extent changes with W_0.

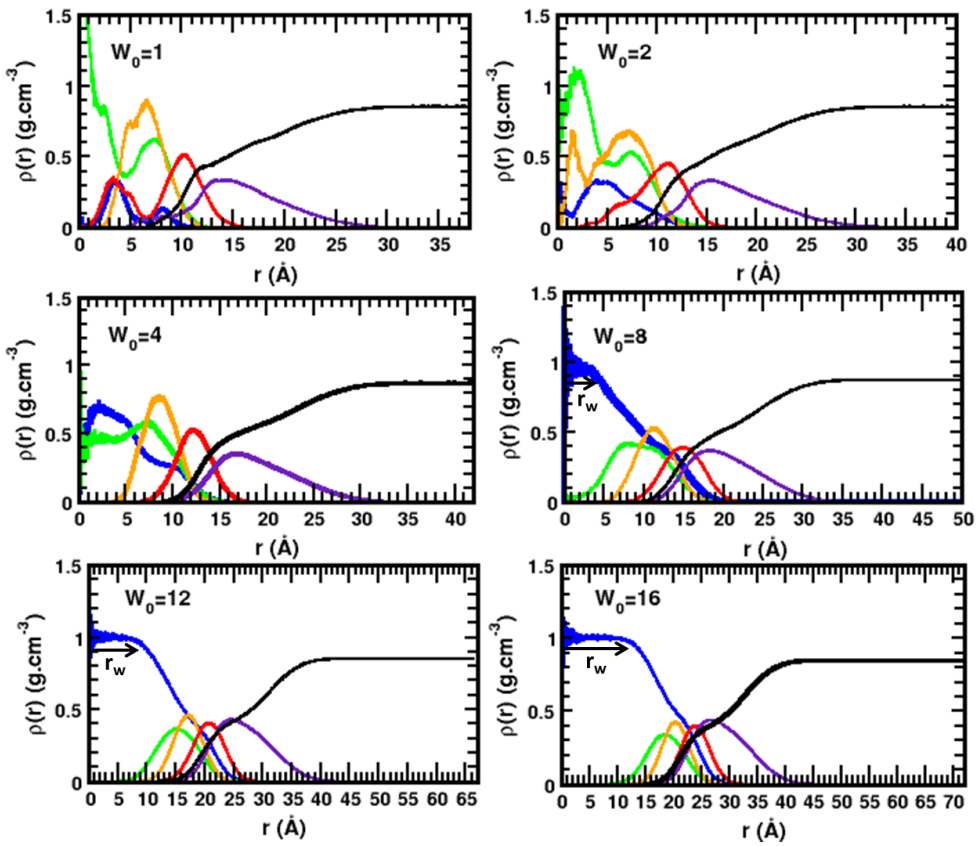


Figure 5: Average radial density profiles, $\rho(r)$, along the distance from the RM center-of-mass for the series of RMs with increasing W_0, for water (blue), choline (green), phosphate (orange), glycerol (red), acyl chains (violet), and benzene (black). The arrows in the plots for the RMs with $W_0 \geq 8$ are approximate indications of the bulk-water core radius, r_w.

First, the benzene and water regions exhibit a very limited overlap, but the benzene solvent does penetrate between the DOPC glycerol and fatty acid chains in agreement with experiments[62-66] and previous MD simulations[29,61]. This can be quantitatively measured via the solvent accessible surface areas (SASA) between the different DOPC groups (*i.e.*, choline, phosphate, glycerol, and acyl chains) and water and benzene. Benzene molecules penetrate deeply in the RM shell and enter in contact with the choline, phosphate, and glycerol groups (Table S4); these contact areas decrease with increasing W_0, and hence with the amount of water inside the RM. We note that at large distances from the RM, the benzene density reaches its bulk value of ~0.88 g cm^{-3},[67] close to the experimental[68] value of 0.92 g cm^{-3}.

Regarding the arrangement of DOPC headgroups, the radial distributions in Fig. 4 indicate that the choline groups are always further inside the RM than the phosphate groups. For the smallest RMs ($W_o \leq 4$), the broad and irregular headgroup distributions suggest a disordered arrangement, while in the larger systems ($W_0 \geq 8$), a layered structure appears more clearly. In the latter case, the thickness of the headgroup region is estimated to be ~10.0 Å, in good agreement with our membrane reference simulation and with experiments (8.0 ± 1.5 Å)[53,69,70].

The confined water radial density profile changes dramatically with the RM size. In the smaller RMs ($W_0 \leq 4$) no well-defined water core can be seen and a large overlap is found between the density distribution of water and those of the phosphate, choline, and glycerol groups. In contrast, in the larger RMs ($W_0 \geq 8$) the water density reaches a value close to that of bulk water in the center of the RM, and other RM components are absent from this region. The radius of the water pool can be estimated from the radius r_w where the water density decreases below its bulk value. As shown in Fig. S8, this radius decreases approximately linearly with the W_o ratio as $r_W(W_0) \approx 1.12\ (\pm 0.03)\ W_0 - 5.33\ (\pm 3.60)$ Å, from ~12 Å for the largest RM (W_0=16), to the disappearance of the water core around $W_0 = 4$, as suggested in experiments[1,62,71,72]. Water molecules are confined within the center of the RM and escape into

the oil phase (*i.e.,* the DOPC tails and benzene) very exceptionally; we registered only a few isolated events for the $W_0 = 8$ and $W_0 = 16$ trajectories and did not consider these isolated water molecules in the calculation of average water properties. In the following, we discuss in more detail the structural properties of the water core, including the water-DOPC headgroup interactions in Section 3.2, and their dynamics in Section 4.

3.2 Properties of the RM Water Core

3.2.1 Water Core Structure

The radial densities in Fig. 5 show that the DOPC headgroups become increasingly hydrated when the number of water molecules per surfactant increases, and we now characterize this progressive hydration. We analyze both the contact area between water and the different moieties of the DOPC surfactants, and the hydration number of key DOPC sites. The contact area between the different components of the systems is calculated by computing the solvent-accessible surface area (SASA) with the *trjVoronoi* tool[73] and we compare the contact areas normalized by the number of DOPC molecules in each RM (aggregation number n_{DOPC} in Table 1). We note that the SASA values computed in this way take into account the corrugation of the RM surface and of the water core caused by the penetration of the water and apolar solvents between the DOPC headgroups and alkyl chains. They are therefore significantly larger than the SASA that can be deduced from the description of the RM as an ideal sphere.

The hydration numbers are determined by integrating the first peak of the water oxygen radial distribution function centered on the DOPC site of interest (Table S5). The contact areas and the hydration numbers plotted in Fig. 6 show that the hydration of DOPC occurs in two main steps when W_o is increased.

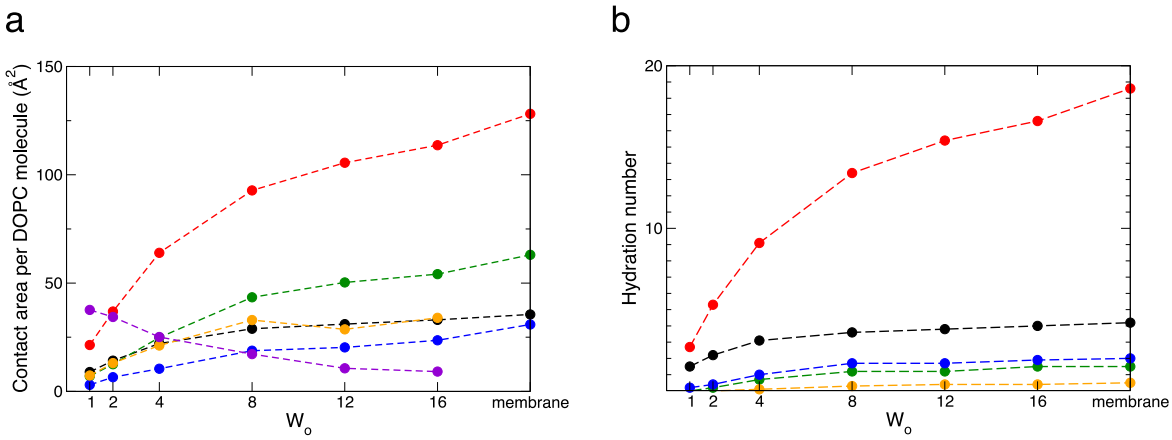


Figure 6: a) Surface area normalized by the number of DOPC molecules between water and phosphate groups (black), water and choline groups (red), water and glycerol (green), water and alkyl chains (blue), water and benzene (orange), and for the contacts between benzene and the phosphate and choline groups (violet) (see Table S4). b) Hydration number for the phosphate oxygen atoms (black), phosphate ester oxygen atoms (green), choline nitrogen atom (red), glycerol carboxyl oxygen atoms (blue), and glycerol ester oxygen atoms (orange) (see Table S5).

First, in smaller RMs ($W_0 \leq 4$), water molecules solvate mostly the polar headgroups (phosphate, choline, glycerol), while the apolar benzene solvent penetrates far between the DOPC chains as shown by the benzene-headgroup contacts. A closer inspection reveals that the phosphate oxygen atoms (red in Scheme 1) are more hydrated than the phosphate ether oxygen sites (blue in Scheme 1), and similarly the glycerol carboxyl oxygen sites (green in Scheme 1) have more contacts with water than the glycerol ester oxygens (orange in Scheme 1) as shown in Fig. 6b.

When W_o increases, the water-headgroup contacts and the phosphate and choline hydration numbers strongly increase, while the headgroup-benzene contacts decrease markedly. For W_0 = 8 the phosphate groups are fully hydrated, and both their contact area with water and their hydration number reach a plateau value (with a total hydration number of the phosphate

group of ~5; see Table S5), in agreement with prior linear infrared spectroscopy results[62] on lecithin RMs. This shows that at $W_0 = 8$, the phosphate hydration shell already resembles that found in phospholipid membranes.

For larger RMs ($W_0 > 8$), the additional water molecules mostly go in the water core (Fig. 5) and a small fraction contributes to a further increase in the hydration of the choline and glycerol groups, although more slowly than for the smaller RMs. A concomitant decrease of the benzene penetration between the RM chains is found. Both the contact areas between water and the DOPC headgroups and the hydration numbers tend towards their values in membranes. We note, however, that the hydration structure of the DOPC molecules in the largest $W_0 = 16$ RM studied here still differs slightly from that in membranes. The hydration numbers of the headgroups (5.6 ± 0.1 and 16.6 ± 0.1 for the phosphate and choline groups) remain slightly below the values found in our membrane simulation (5.8 ± 0.1 and 18.6 ± 0.1 respectively, see Fig. 6 and Table S5) but close to the values in experiments on a model phosphatidylcholine headgroup in aqueous solution (choline hydration number of 20.0 – 23.0)[74] and for phosphate in a phospholipid membrane (6.0 with FTIR[75]).

However, when one focuses specifically on the oxygen atoms within the phosphate group, the hydration number in the $W_0 = 16$ RM (2.0 ± 0.2 per phosphate oxygen and 0.7 ± 0.2 per phosphate ether oxygen) is in good agreement with the value in the membrane (2.1 ± 0.2 per phosphate oxygen and 0.7 ± 0.1 per ether oxygen), and in phosphatidylcholine surfactants in micelles (0.8)[4]. Finally, the hydration number of 1.0 for each glycerol carboxyl site for W_o=16 and in the membrane is in agreement with previous results for fully hydrated phospholipids in solution and in a membrane.[6,7] We note that similar conclusions can be drawn from the analysis of hydrogen bonds in Fig. S12 which is also consistent with recent simulations[30] on RMs of a similar DSPC phospholipid. This suggests that the trends observed for the phosphate headgroup hydration are fairly robust and independent of the nature of the al-

kyl chain and of the apolar solvent, and that in large RMs the headgroup hydration structure and conformation resemble those found in membranes (see SI section S2.3 and Figs. S9-S11).

3.3 Water Dynamics

We now study the reorientation dynamics of the confined water molecules, together with the water OH stretch infrared vibrational spectrum, and connect our results to available experimental data (future work will address the details of the hydrogen-bond (H-bond) dynamics and of the nonlinear vibrational spectra).

3.3.1 Water Reorientation Dynamics

We first calculate the water reorientation time-correlation function (tcf)

$$C_2(t) = \langle P_2[\cos(\boldsymbol{u}_{OH}(0) \cdot \boldsymbol{u}_{OH}(t)]\rangle \,, \tag{6}$$

where P_2 is the second-order Legendre polynomial and $\boldsymbol{u}_{\mathrm{OH}}(t)$ is the unit vector along a water OH bond at time t. The tcfs for the series of RMs and for the DOPC bilayer are shown in Fig. 7.

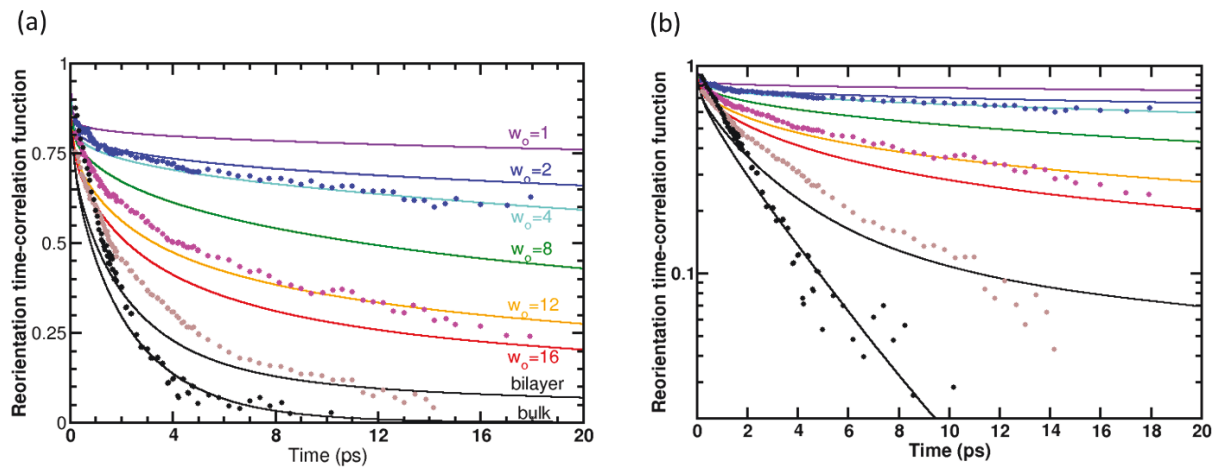


Figure 7: Reorientation time correlation functions (Eq. 6) for water OH bonds in DOPC reverse micelles of increasing sizes, a) on a linear scale and b) on a semi-log scale, from our simulations (solid lines) and from ultrafast infrared spectroscopy measurements[76] on AOT RMs (dots: $W_o = 2$ (blue), 5 (pink), 10 (brown), and bulk water (black)).

Consistent with prior experimental[4,10,19,21,76,77] and simulation[19,21,23,24,27] studies on several RMs, the water reorientational dynamics slows down as the RM size decreases. In addition, our computed tcfs exhibit a pronounced non-exponential relaxation, indicative of a wide heterogeneity in the environments experienced by water molecules. Values for the water reorientation time are estimated by numerical integration of $C_2(t)$ up to a finite time t* and by approximating the residual decay through an exponential fit on the last 10 ps before t*. Figure 8 shows the resulting water reorientation times, together with the water reorientation times measured by ultrafast anisotropy decay in AOT RMs[24] since no experimental data on DOPC RMs are available for comparison. We find that, for similar W_0 ratios, water reorientation dynamics is much slower in DOPC RMs than in corresponding AOT and Igepal RMs. This can be ascribed both to the smaller size of the water pool in the DOPC RM (as shown by the comparison of our distributions in Fig. 5 with the results in Ref. [24] for AOT RMs) and to the stronger interaction of water with the DOPC headgroup than with the AOT sulfonate headgroup and than with the uncharged Igepal headgroup. This latter point is illustrated by the comparison of the water OH stretch vibrational spectra, now described.

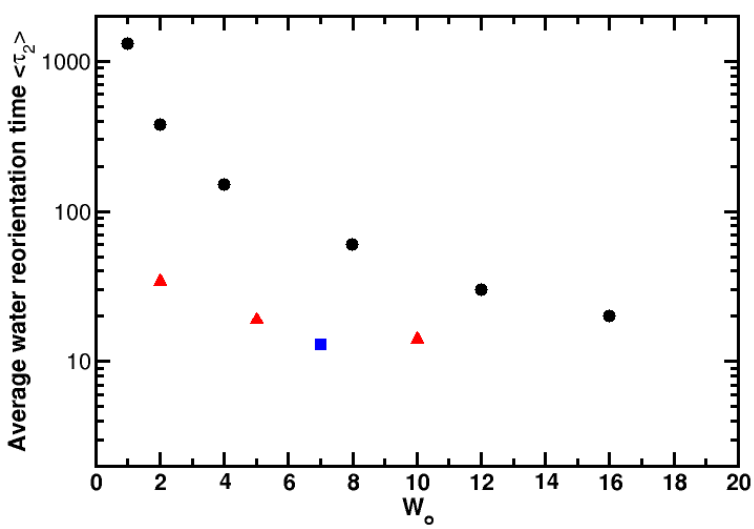


Figure 8: Average water reorientation time $\langle\tau_2\rangle = \int_0^\infty dt\, C_2(t)$ from our DOPC RM simulations (black dots) and from experimental anisotropy decays[17] in AOT (red triangles) and Igepal (blue square) RMs.

3.3.2 Water Infrared Spectrum

The water infrared absorption lineshape is calculated as[78]

$$I(\omega) \sim \int_0^\infty \mathrm{d}t e^{i\omega t} \langle \vec{\mu}(0) \cdot \vec{\mu}(t) e^{-i\int_0^t \mathrm{d}\tau \omega_{10}(\tau)} \rangle\, e^{-t/2T_1} \tag{7}$$

where ω is the frequency, $\vec{\mu}(t)$ is the transition dipole moment between the ground and excited OH stretch vibrational states at time t, ω_{10}(t) is the transition frequency, and T_1 is the vibrational lifetime. The spectra are computed from classical MD NVT trajectories using the quadratic frequency map described in Ref. [78] in which the transition dipole and frequency are obtained from the electric field exerted on the water hydrogen atom by the other atoms in the system. While our simulations are performed with neat H_2O in the RM, our calculation assumes that each water OH stretch is decoupled from the other stretch modes, which is obtained experimentally by isotopic dilution and considering a dilute HOD in D_2O[76].

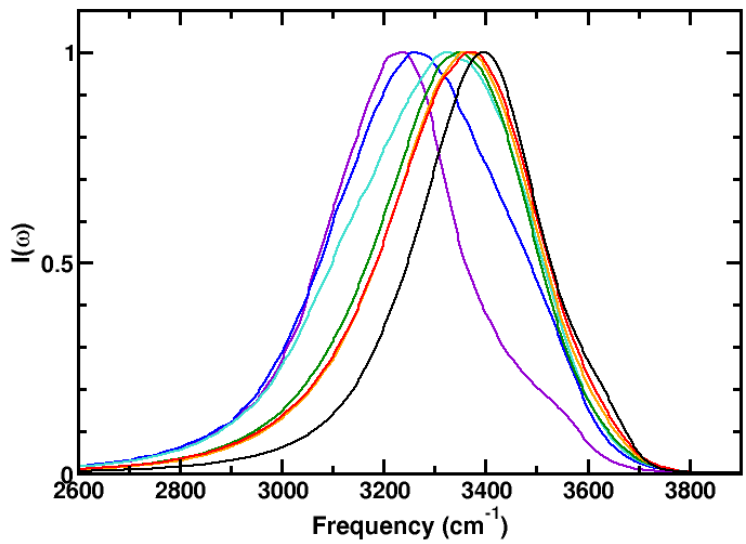


Figure 9: Infrared lineshapes (Eq. 7) in our series of DOPC RMs for W_0 = 1 (violet), 2 (blue), 4 (cyan), 8 (green), 12 (orange), 16 (red), and in the bulk (black).

The computed lineshapes are shown in Fig. 9. They exhibit a shift to the red (lower frequency) as W_0 decreases and the fraction of interfacial waters increases. This suggests that interfacial water molecules have a redshifted frequency distribution, *i.e.*, that water OHs donate stronger hydrogen bonds to acceptors in the DOPC headgroups than to other waters. This redshift is not visible in the experimental lineshapes for neat H_2O in DOPC RMs which only exhibit a narrowing with decreasing W_0.[36] The comparison with the measured lineshapes necessarily include the effects in H_2O due to resonant vibrational coupling that are not present in our simulations of dilute HOD in D_2O. However, both our simulations and these experiments suggest that the water IR lineshape does not exhibit the blueshift seen in AOT RMs for decreasing W_0. This is consistent with the stronger hydrogen-bond interaction between water and the DOPC phosphate headgroup than between water and the AOT sulfonate headgroup.

4. Concluding Remarks

We have presented a detailed molecular dynamics simulation study of the structural and dynamical properties of DOPC phospholipid reverse micelles. We report the first all-atom simulations of DOPC reverse micelles in benzene over a broad range of W_0 ratios (from 1 to 16). We showed that the aggregation number cannot be directly determined from the surface area per lipid in a membrane arrangement and we determined its value by fitting the reverse micelle sizes measured by dynamic light scattering. Our simulations suggest that all investigated reverse micelles are approximately spherical. Smaller reverse micelles ($W_0 \leq 4$) exhibit a disordered core where the headgroups are not fully hydrated. Larger reverse micelles ($W_o > 4$) exhibit a layered structure with a distinct water core.

The dynamics of water molecules within the DOPC reverse micelles slow down as the reverse micelle size decreases, in agreement with prior studies on AOT and Igepal reverse micelles. However, for a given W_0 ratio, the average water reorientation dynamics in DOPC

reverse micelles are found to be much slower than in these other reverse micelles. This is explained by the smaller water pool and by the stronger interaction between water and the charged headgroups in the DOPC reverse micelles. Our calculations of the infrared lineshape for the water OH stretch vibration exhibit a redshift with decreasing W_0, consistent with the strong water-headgroup interaction.

Future work will investigate the water hydrogen-bond dynamics in these DOPC reverse micelles, assess whether they can be described with a two-state core-shell model and connect the results to experimentally accessible two-dimensional infrared spectra.

5. Acknowledgements

We thank Fabio Sterpone (IBPC, Paris, France) for stimulating discussions. This work was granted access to the HPC resources of CCRT/CINES under the allocation t2014077266 made by GENCI (Grand Equipement National de Calcul Intensif). We acknowledge support from the Agence National de la Recherche (Grant No. ANR-2011-BS08-010-01) to D.L and N.G. and from the Chemical Sciences, Geosciences, and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy (Grant No. DE-FG02-05ER15708) to W.H.T.

Supporting Information. DOPC membrane simulations details, additional figures and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

Corresponding Authors

* Stéphane Abel: stephane.abel@cea.fr; Damien Laage: damien.laage@ens.fr

References

(1) Pileni, M. -P. *Structure and Reactivity in Reverse Micelles*; Elsevier Science B.V.: Amsterdam, 1989.

(2) Li, M.; Schnablegger, H.; Mann, S. Coupled synthesis and self-assembly of nanoparticles to give structures with controlled organization. *Nature* **1999**, *402*, 393–395.

(3) Levinger, N. E.; Swafford, L. A. Ultrafast dynamics in reverse micelles. *Annu Rev Phys Chem* **2009**, *60*, 385–406.

(4) Fayer, M. D.; Levinger, N. E. Analysis of water in confined geometries and at interfaces. *Annu Rev Anal Chem* **2010**, *3*, 89-107.

(5) Montal, M. On the structure, dynamics and possible functional roles of inverted micelles in biological membranes. In *Reverse Micelles*; ; Plenum Press: New York, 1984; pp 221-229.

(6) Nicot, C.; Waks, M. Proteins as Invited Guests of Reverse Micelles: Conformational Effects, Significance, Applications. *Biotechnol Genet Eng Rev* **1996**, *13*, 267–314.

(7) Melo, E. P.; Aires-Barros, M. R.; Cabral, J. M. Reverse micelles and protein biotechnology. *Biotechnol Annu Rev* **2001**, *7*, 87-129.

(8) Peterson, R. W.; Anbalagan, K.; Tommos, C.; Wand, A. J. Forced folding and structural analysis of metastable proteins. *J Am Chem Soc* **2004**, *126*, 9498–9499.

(9) Halle, B.; Goeran, C. Hydration of ionic surfactant micelles from water oxygen-17 magnetic relaxation. *J Phys Chem* **1981**, *85*, 2142–2147.

(10) Carlstroem, G.; Halle, B. Water dynamics in microemulsion droplets. A nuclear spin relaxation study. *Langmuir* **1988**, *4*, 1346–1352.

(11) Carlstroem, G.; Halle, B. Shape fluctuations and water diffusion in microemulsion droplets: a nuclear spin relaxation study. *J Phys Chem* **1989**, *93*, 3287–3299.

(12) Eastoe, J.; Cazelles, B. M.; Steytler, D. C.; Holmes, J. D.; Pitt, A. R.; Wear, T. J.; Heenan, R. K. Water-in-CO2 microemulsions studied by small-angle neutron scattering. *Langmuir* **1997**, *13*, 6980–6984.

(13) Hirai, M.; Kawai-Hirai, R.; Sanada, M.; Iwase, H.; Mitsuya, S. Characteristics of AOT microemulsion structure depending on apolar solvents. *J Phys Chem B* **1999**, *103*, 9658–9662.

(14) Dokter, A. M.; Woutersen, S.; Bakker, H. J. Inhomogeneous dynamics in confined water nanodroplets. *Proc Natl Acad Sci U S A* **2006**, *103*, 15355–15358.

(15) Tan, H. S.; Piletic, I. R.; Riter, R. E.; Levinger, N. E.; Fayer, M. D. Dynamics of Water Confined on a Nanometer Length Scale in Reverse Micelles: Ultrafast Infrared Vibrational Echo Spectroscopy. *Phys Rev Lett* **2005**, *94*, 057405.

(16) Moilanen, D. E.; Levinger, N. E.; Spry, D. B.; Fayer, M. D. Confinement or the nature of the interface? Dynamics of nanoscopic water. *J Am Chem Soc* **2007**, *129*, 14311–14318.

(17) Park, S.; Moilanen, D. E.; Fayer, M. D. Water dynamics--the effects of ions and nanoconfinement. *J Phys Chem B* **2008**, *112*, 5279–5290.

(18) Fenn, E. E.; Wong, D. B.; Giammanco, C. H.; Fayer, M. D. Dynamics of Water at the Interface in Reverse Micelles: Measurements of Spectral Diffusion with Two-Dimensional Infrared Vibrational Echoes. *J Phys Chem B* **2011**, *115*, 11658–11670.

(19) Bakulin, A. A.; Cringus, D.; Pieniazek, P. A.; Skinner, J. L.; Jansen, T. L. C.; Pshenichnikov, M. S. Dynamics of Water Confined in Reversed Micelles: Multidimensional Vibrational Spectroscopy Study. *J Phys Chem B* **2013**, doi: 10.1021/jp405853j.

(20) Abel, S.; Sterpone, F.; Bandyopadhyay, S.; Marchi, M. Molecular Modeling and Simulations of AOT−Water Reverse Micelles in Isooctane: Structural and Dynamic Properties. *J Phys Chem B* **2004**, *108*, 19458–19466.

(21) Harpham, M. R.; Ladanyi, B. M.; Levinger, N. E.; Herwig, K. W. Water motion in reverse micelles studied by quasielastic neutron scattering and molecular dynamics simulations. *J Chem Phys* **2004**, *121*, 7855–7868.

(22) Brodskaya, E. N.; Mudzhikova, G. V. Molecular dynamics simulation of AOT reverse micelles. *Mol Phys* **2006**, *104*, 3635–3643.

(23) Chowdhary, J.; Ladanyi, B. M. Molecular dynamics simulation of aerosol-OT reverse micelles. *J Phys Chem B* **2009**, *113*, 15029–15039.

(24) Pieniazek, P. A.; Lin, Y. S.; Chowdhary, J.; Ladanyi, B. M.; Skinner, J. L. Vibrational spectroscopy and dynamics of water confined inside reverse micelles. *J Phys Chem B* **2009**, *113*, 15017–15028.

(25) Biswas, R.; Chakraborti, T.; Bagchi, B.; Ayappa, K. G. Non-monotonic, distance-dependent relaxation of water in reverse micelles: propagation of surface induced frustration along hydrogen bond networks. *J Chem Phys* **2012**, *137*, 014515.

(26) Biswas, R.; Furtado, J.; Bagchi, B. Layerwise decomposition of water dynamics in reverse micelles: A simulation study of two-dimensional infrared spectrum. *J Chem Phys* **2013**, *139*, 144906.

(27) Martinez, A. V.; Dominguez, L.; Małolepsza, E.; Moser, A.; Ziegler, Z.; Straub, J. E. Probing the structure and dynamics of confined water in AOT reverse micelles. *J Phys Chem B* **2013**, *117*, 7345–7351.

(28) Marchi, M.; Abel, S. Modeling the Self-Aggregation of Small AOT Reverse Micelles from First Principles. *J Phys Chem Lett* **2015**, *6*, 170–174.

(29) Vierros, S.; Sammalkorpi, M. Phosphatidylcholine reverse micelles on the wrong track in molecular dynamics simulations of phospholipids in an organic solvent. *J Chem Phys* **2015**, *142*, 094902.

(30) Vierros, S.; Sammalkorpi, M. Role of hydration in phosphatidylcholine reverse micelle structure and gelation in cyclohexane: a molecular dynamics study. *Phys Chem Chem Phys* **2015**, *17*, 14951–14960.

(31) Wagner, A.; Vorauer-Uhl, K. Liposome technology for industrial purposes. *J Drug Deliv* **2011**, *2011*, 591325.

(32) Taylor, T. M.; Davidson, P. M.; Bruce, B. D.; Weiss, J. Liposomal nanocapsules in food science and agriculture. *Crit Rev Food Sci Nutr* **2005**, *45*, 587–605.

(33) Szyc; Yang, M.; Nibbering, E. T. J.; Elsaesser, T. Ultrafast vibrational dynamics and local interactions of hydrated DNA. *Angew Chem Int Ed* **2010**, *49*, 3598–3610.

(34) Levinger, N. E.; Costard, R.; Nibbering, E. T. J.; Elsaesser, T. Ultrafast energy migration pathways in self-assembled phospholipids interacting with confined water. *J Phys Chem A* **2011**, *115*, 11952–11959.

(35) Costard, R.; Levinger, N. E.; Nibbering, E. T. J.; Elsaesser, T. Ultrafast Vibrational Dynamics of Water Confined in Phospholipid Reverse Micelles. *J Phys Chem B* **2012**, *116*, 5752–5759.

(36) Costard, R.; Greve, C.; Heisler, I. A.; Elsaesser, T. Ultrafast Energy Redistribution in Local Hydration Shells of Phospholipids: A Two-Dimensional Infrared Study. *J Phys Chem Lett* **2012**, *3*, 3646–3651.

(37) Costard, R.; Elsaesser, T. Femtosecond OH bending dynamics of water nanopools confined in reverse micelles. *J Phys Chem B* **2013**, *117*, 15338–15345.

(38) Costard, R.; Heisler, I. A.; Elsaesser, T. Structural Dynamics of Hydrated Phospholipid Surfaces Probed by Ultrafast 2D Spectroscopy of Phosphate Vibrations. *J Phys Chem Lett* **2014**, *5*, 506–511.

(39) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **2008**, *4*, 435–447.

(40) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J Chem Theory Comput* **2010**, *6*, 459–466.

(41) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(42) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* **2010**, *114*, 7830–7843.

(43) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P. E. M.; Vorobyov, I.; Alexander, D. D. CHARMM

general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **2010**, *31*, 671–690.

(44) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J Phys Chem* **1987**, *91*, 6269–6271.

(45) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **2007**, *126*, 014101–014101.

(46) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J Chem Phys* **1984**, *81*, 3684–3684.

(47) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J Chem Phys* **1995**, *103*, 8577–8593.

(48) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput* **2007**, *4*, 116–122.

(49) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **1981**, *52*, 7182–7190.

(50) Rahman, A.; Stillinger, F. H. Molecular dynamics study of liquid water. *J Chem Phys* **1971**, *55*, 3336–3359.

(51) Nagle, J. F.; Tristram-Nagle, S. Structure of lipid bilayers. *BBA-Rev Biomembranes* **2000**, *1469*, 159–195.

(52) Kucerka, N.; Nagle, J. F.; Sachs, J. N.; Feller, S. E.; Pencer, J.; Jackson, A.; Katsaras, J. Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data. *Biophys J* **2008**, *95*, 2356–2367.

(53) Pan, J.; Tristram-Nagle, S.; Kucerka, N.; Nagle, J. F. Temperature dependence of structure, bending rigidity, and bilayer interactions of dioleoylphosphatidylcholine bilayers. *Biophys J* **2008**, *94*, 117–124.

(54) Hassan, P. A.; Rana, S.; Verma, G. Making Sense of Brownian Motion: Colloid Characterization by Dynamic Light Scattering. *Langmuir* **2015**, *31*, 3-12.

(55) Perrin, F. Mouvement brownien d'un ellipsoide - I. Dispersion diélectrique pour des molécules ellipsoidales. *J Phys Radium* **1934**, *5*, 497–511.

(56) Hansen, S. Translational friction coefficients for cylinders of arbitrary axial ratios estimated by Monte Carlo simulation. *J Chem Phys* **2004**, *121*, 9111–9115.

(57) Eastoe, J.; Hetherington, K. J.; Sharpe, D.; Steytler, D. C.; Egelhaaf, S.; Heenan, R. K. Droplet Structure in Phosphocholine Microemulsions. *Langmuir* **1997**, *13*, 2490–2493.

(58) Pomata, M. H. H.; Laria, D.; Elola, M. D. D. Molecular dynamics simulations of AOT-water/formamide reverse micelles: structural and dynamical properties. *The Journal of chemical physics* **2008**, *129*, 244503–244503.

(59) Allen, R.; Bandyopadhyay, S.; Klein, M. L. C 12 E 2 Reverse Micelle: A Molecular Dynamics Study . *Langmuir* **2000**, *16*, 10547–10552.

(60) Elworthy, P. H.; McIntosh, D. S. The interaction of water with lecithin micelles in benzene. *J Phys Chem* **1964**, *68*, 3448–3452.

(61) Agazzi, F. M.; Correa, N. M.; Rodriguez, J. Molecular Dynamics Simulation of Water/BHDC Cationic Reverse Micelles. Structural Characterization, Dynamical Properties and the Influence of the Solvent on the Intermicellar Interactions. *Langmuir* **2014**, *30*, 9643–9653.

(62) Maitra, A.; Jain, T. K.; Shervani, Z. Interfacial water structure in lecithin—oil—water reverse micelles. *Colloid Surface* **1990**, *47*, 255–267.

(63) Correa, N. M.; Biasutti, M. A.; Silber, J. J. Micropolarity or Reversed Micelles: Comparison between Anionic, cationic, and Nonionic Reversed Micelles. *J Colloid Interface Sci* **1996**, *184*, 570–578.

(64) Agazzi, F. M.; Falcone, R. D.; Silber, J. J.; Correa, N. M. Solvent blends can control cationic reversed micellar interdroplet interactions. The effect of n-heptane:benzene mixture on BHDC reversed micellar interfacial properties: droplet sizes and micropolarity. *J Phys Chem B* **2011**, *115*, 12076–12084.

(65) Agazzi, F. M.; Rodriguez, J.; Falcone, R. D.; Silber, J. J.; Correa, N. M. PRODAN dual emission feature to monitor BHDC interfacial properties changes with the external organic solvent composition. *Langmuir* **2013**, *29*, 3556–3566.

(66) Abuin, E.; Lissi, E.; Duarte, R.; Silber, J. J.; Biasutti, M. A. Solubilization in AOT-water reverse micelles. Effect of the external solvent. *Langmuir* **2002**, *18*, 8340–8344.

(67) Goodwin, R. D. Benzene Thermophysical Properties from 279 to 900 K at Pressures to 1000 Bar. *J Phys Chem Ref Data* **1988**, *17*, 1541–1541.

(68) Cabaço, M. I.; Danten, Y.; Besnard, M.; Guissani, Y.; Guillot, B. Neutron diffraction and molecular dynamics study of liquid benzene and its fluorinated derivatives as a function of temperature. *J Phys Chem B* **1997**, *101*, 6977–6987.

(69) Kucerka, N.; Kiselev, M. A.; Balgavý, P. Determination of bilayer thickness and lipid surface area in unilamellar dimyristoylphosphatidylcholine vesicles from small-angle neutron scattering curves: a comparison of evaluation methods. *Eur Biophys J* **2004**, *33*, 328–334.

(70) Braun, A. R.; Sachs, J. N.; Nagle, J. F. Comparing simulations of lipid bilayers to scattering data: the GROMOS 43A1-S3 force field. *J Phys Chem B* **2013**, *117*, 5065–5072.

(71) Chaykowski, F. T.; Wan, J. K. S.; Singer, M. A. Interaction of small molecules with phospholipid bilayer membranes: a spin label study. *Chem Phys Lipids* **1979**, *23*, 111–123.

(72) Davenport, J. B.; Fisher, L. R. Interaction of water with egg lecithin in benzene solution. *Chem Phys Lipids* **1975**, *14*, 275–290.

(73) Abel, S.; Dupradeau, F. -Y.; Marchi, M. Molecular Dynamics Simulations of a Characteristic DPC Micelle in Water. *J Chem Theory Comput* **2012**, *8*, 4610–4623.

(74) Foglia, F.; Lawrence, M. J.; Lorenz, C. D.; McLain, S. E. On the hydration of the phosphocholine headgroup in aqueous solution. *J Chem Phys* **2010**, *133*, 145103–145103.

(75) Goñi, F. M.; Arrondo, J. L. R. A study of phospholipid phosphate groups in model membranes by Fourier transform infrared spectroscopy. *Faraday Discuss Chem Soc* **1986**, *81*, 117–126.

(76) Tan, H. S.; Piletic, I. R.; Fayer, M. D. Orientational dynamics of water confined on a nanometer length scale in reverse micelles. *J Chem Phys* **2005**, *122*, 174501.

(77) Fenn, E. E.; Wong, D. B.; Fayer, M. D. Water dynamics at neutral and ionic interfaces. *Proc Natl Acad Sci USA* **2009**, *106*, 15243–15248.

(78) Skinner, J. L.; Auer, B. M.; Lin, Y. -S. Vibrational Line Shapes, Spectral Diffusion, and Hydrogen Bonding in Liquid Water. *Adv. Chem. Phys.* **2009**, *142*, 59-103.

Supporting Information

On the Structural and Dynamical Properties of DOPC Reverse Micelles

Stéphane Abel[a, b *], *Nuno Galamba*[d,e], *Esra Karakas*[a, b,c], *Massimo Marchi*[a, b], *Ward H. Thompson*[f], *Damien Laage*[d,e*]

[a] Commissariat à l'Energie Atomique et aux Energies Alternatives, DRF/iBiTEC-S/SB2SM & CNRS UMR 9198, Saclay, France

[b] Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris-Sud, Université Paris-Saclay, F-91198, Gif-sur-Yvette cedex, France.

[c] Maison de la Simulation, USR 3441, CEA-CNRS-INRIA-Univ Paris Sud – Univ Versailles, 91191 Gif-sur-Yvette Cedex, France

[d] Ecole Normale Supérieure, PSL Research University, UPMC Univ Paris 06, CNRS, Département de Chimie, PASTEUR, 24 rue Lhomond, 75005 Paris, France

[e]Sorbonne Universités, UPMC Univ Paris 06, ENS, CNRS, PASTEUR, 75005 Paris, France

[f] Department of Chemistry, University of Kansas, Lawrence, KS 66045, USA

S1 Construction and simulation of the DOPC bilayer.

The simulations of the bilayer membrane were performed with the GROMACS (v4.6.6) package[1-3]. As for the RM systems, production runs in the NPT and NVT ensembles were propagated to examine the structural and dynamical properties, respectively. The initial conformation of the DOPC bilayer with 128 lipids (64 per leaflet) was constructed with the CHARMM-GUI[4,5] membrane generator module[6-8]. A total of 5120 SPC/E water molecules (40 per DOPC molecule) were added in the box to model a fully hydrated bilayer. A cutoff of 12 Å was used for the pairwise interactions and the neighbor list. Lennard-Jones interactions were smoothly damped to zero between 8.0 and 12 Å, while the electrostatic interactions were truncated at 12 Å with the longer-range interactions treated using the PME method [9]. A time step of 2 fs (resp. 1 fs) was used to integrate the equations of motion with neighbor list updates every 10 fs (resp. 5 fs) in the NPT (resp. NVT) simulations. The covalent bonds involving hydrogen atoms were constrained using the LINCS algorithm [10].

It was shown previously (see, *e.g.,* refs. [11,12]) that the choice of simulation parameters is crucial to obtain correct membrane properties. In the main text, we compare several structural and dynamical properties of DOPC reverse micelles (RM) and membrane modeled with the SPC/E water potential. This water model provides a much better description of the water dynamics (see, e.g., [13,14]), but it differs from the TIP3P potential used to parameterize the CHARMM36 lipid force field [15]. As we will show below, our results on the membrane structure are in good agreement with prior CHARMM36/TIP3P simulations.

To prepare the membrane, we first minimized the membrane-water system obtained with the CHARMM-GUI with the steepest descent algorithm and an energy tolerance of 1000 $kJ.mol^{-1}$. The membrane system was then equilibrated in the NVT ensemble at 303 K for 300 ps with the

v-scale algorithm [16] and a coupling constant τ_T of 0.2 ps; DOPC and water were coupled to distinct thermostatic baths. The resulting configuration was then equilibrated in the NPT ensemble at T = 303 K and P = 1.015 bar by carrying out a long simulation (100 ns) with the semi-isotropic coupling scheme using the Bussi *et al.*[16] and the Parrinello-Rahman[17,18] algorithms; the values for τ_T, τ_p, and the compressibility were 0.2 ps, 5.0 ps, and 45×10^{-6} bar^{-1}, respectively. Finally, the resulting system was subject to another long (155 ns) simulation at 303 K with the same temperature parameters and the pressure maintained anisotropically at 1.015 bar with the Parrinello-Rahman[17,18] barostat using τ_p and compressibility values of 5.0 ps and 45×10^{-6} bar^{-1}, respectively. The data for structural and dynamical analysis were collected every 2 ps and 1 fs, respectively from the NPT and NVT production runs (500 ps). A snapshot of the DOPC membrane system taken from the NPT run, is depicted in Fig. S1. As shown in this figure our simulation parameters and the water model give a disordered membrane structure consistent with a L_α phase.

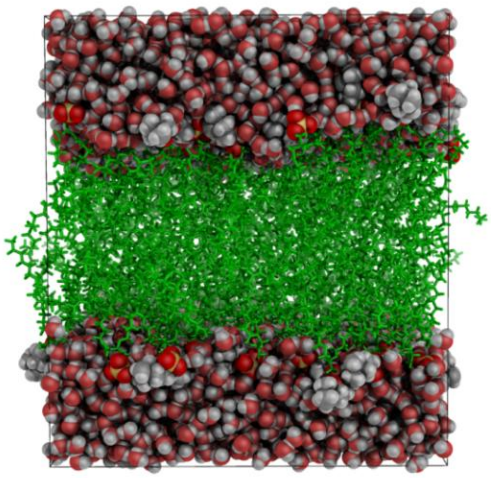


Figure S1: Snapshot of the membrane obtained at the end of the NPT simulation. The headgroup phosphorous, nitrogen, carbon, oxygen and hydrogen atoms are colored in yellow, blue, grey, red and white, respectively, while the acyl chain atoms are in green. The picture was rendered with PyMol.

S2 Structural properties of the membrane

S2.1 Area per lipid, lipid volume, and area compressibity modulus

To examine the stability of the membrane during the course of the NPT simulation, we computed the time evolution of the area per lipid, APL:

$$APL = \frac{L_x L_y}{n_{DOPC}} \tag{1}$$

where L_X, L_Y and n_{DOPC} = 64 are the box x and y vector lengths and the number of lipids per leaflet, respectively.

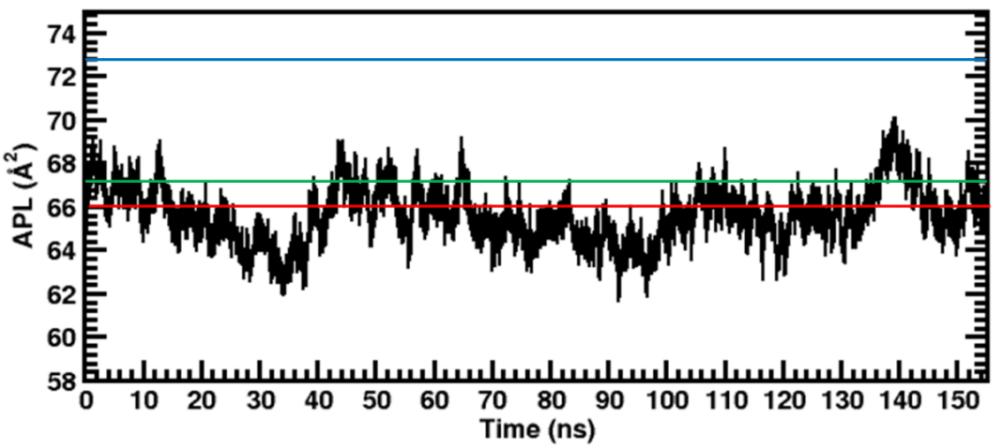


Figure S2: Evolution of area per lipid (APL) during the course of the simulation (black). The lines show the average APL values from the last 50 ns of our simulation (red) and from two experiments[18,19] (green and blue).

Figure S2 shows that the APL is stable during the last ~100 ns of the run and fluctuates around values of ~62.0 – 70 Å^2. The average APL value computed from the last 50 ns of the simulation is equal to 66.0 ± 1.2 Å^2, slightly lower than the experimental APL (67.4 - 72.5 Å^2) [19-21] and than the APL values obtained by Loubet *et al.* (67.3 Å^2) [22] and Klauda *et al.* (69.0 Å^2) [15] with the TIP3P/CHARMM force field. We note that with the same simulation parameters and the TIP3P/CHARMM force field, we were able to reproduce the APL value obtained of Klauda *et*

al. [15] (data not shown). With the simulation box lengths, L_x, L_y, and L_z and the fluctuations of the APL, we computed the average volume of one DOPC, V_{DOPC}, and the area compressibility modulus, K_A, respectively:

$$V_{DOPC} = \frac{L_x L_y L_z}{n_{DOPC} * V_{H2O}} \tag{2}$$

where V_{H2O} is the experimental volume of a water molecule (30.16 Å^3) obtained from the water density at 303K[23], and

$$K_A = \frac{2APL}{\beta n_{DOPC} \sigma_{APL}{}^2} \tag{3}$$

where $\beta = 1/k_bT$, and n_{DOPC}, and σ_{APL} are the number of DOPC and the variance of the APL, respectively. The average values are given in the 6th and 7th rows in Table S1.

Structural Property	This work	Experiments	Simulations
L_x (Å)	63.2 ± 1.2	-	-
L_y (Å)	66.8 ± 1.7	-	-
L_z (Å)	75.2 ± 1.3	-	-
APL (Å^2)	66.0 ± 1.2	67.4 ± 1[19] – 72.5 ± 1[20]	67.3 – 69.0 ± 1 [15,22,24-29]
V_{DOPC} (Å^3)	1273 ± 6.3	1303 [20]	1262 - 1343 [24,25,27,30,31]
K_A (mN.m^{-1})	300	184 – 300	256 - 389
D_{HH} (Å)	38.0	35.3 – 37.1	36.3 – 37.6
D_B (Å)	39.2	35.9 - 38.7	36.8 – 40.4

Table S1: Average membrane structural properties obtained from previous experimental and simulation studies. D_{HH} and D_B are the bilayer and the Luzzati thickness, respectively (see Figure S3 and section S2.2)

The average value of V_{DOPC} obtained from the simulation is 1273 ± 6.3 Å^3, which is slightly smaller than that obtained from experiment (1303 Å^3) [20], but in the range of the values obtained

from previous simulations of same system with different force fields (1262 – 1343 Å^3) [24,25,27,30,31]. Similarly, the K_A value obtained from the simulation, 300 mN.m^{-1}, compares relatively well with previous experimental and simulation studies (for example, 274 mN m^{-1} in ref. [15]).

S2.2 Electron density profile

To examine the membrane characteristics, we computed the average electron density profile (EDP) of the membrane (Fig. S3). We can then obtain characteristics related to the membrane thickness such as the head-head separation, D_{HH}, and the Luzzatti parameter, D_B. The latter quantity is the distance from the membrane center where the water density is half of its bulk value [32] and can be obtained from the lipid volume and the area per lipid from:

$$D_B = \frac{2V_{DOPC}}{APL} \tag{4}$$

a

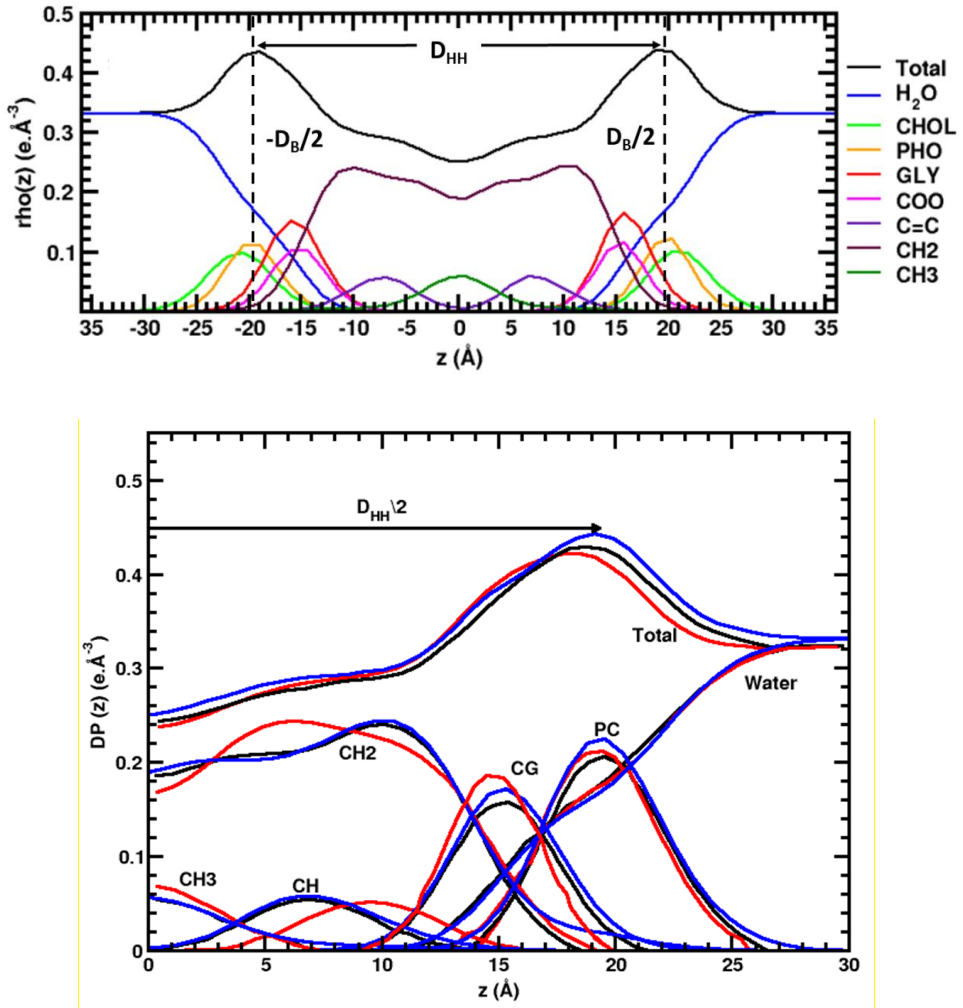


Figure S3: a) Electron density of the DOPC bilayer obtained in this work for the whole system (Total), water (H_2O) and selected components of the lipid (*i.e.*, choline (CHOL), phosphate (PHO), glycerol (GLY), carboxyls (COO), carbons associated with the single (CH_2), double bonds (C=C) and methyl groups (CH_3). b) Comparison of the electron density from an experimental structural density model[19] (red), from prior simulations[15] (black) and from our simulations (blue).

The D_{HH} and D_B values determined from Fig. S3 are listed in Table S1 along with experimental [19–21,31,35] and simulation data [23–27,29,30] from the literature. The D_{HH} and D_B values obtained here suggest that our simulated membrane has a slightly larger thickness. Specifically, we find that

the present D_{HH} values are 2.4 - 7.6% (1.1 - 4.7%) larger than experimental measurements (prior simulations); the present D_B values fall within the range of the reported measured and simulated results. Figure S3b shows that the electron density profiles from our simulations are however in good agreement with those from an experimental model[19] and from prior simulations by Klauda et al. [15], with a shift of <1 Å.

S2.3 Conformations of the DOPC acyl chains and headgroups

The ordering of the two DOPC acyl chains in the membrane was examined by computing the deuterium order parameter, S_{CD}, from the simulation trajectory. This quantity measures the relative orientation of the C-D bonds relative to the bilayer normal (z):

$$S_{CD} = \frac{1}{2}\langle 3\cos^2\theta - 1\rangle \tag{5}$$

Where θ is the angle between the bilayer normal (z) and the C-D bond vector and the bracket denotes the average over all lipids and the simulation time. The calculated S_{CD} can be directly compared with the value obtained from ^{2}H NMR experiments[33]. The average absolute values of the deuterium order $|S_{CD}|$ for the sn1 and sn2 acyl chains are depicted in Figure S4 with experimental value obtained by Warschawski and Devaux[33].

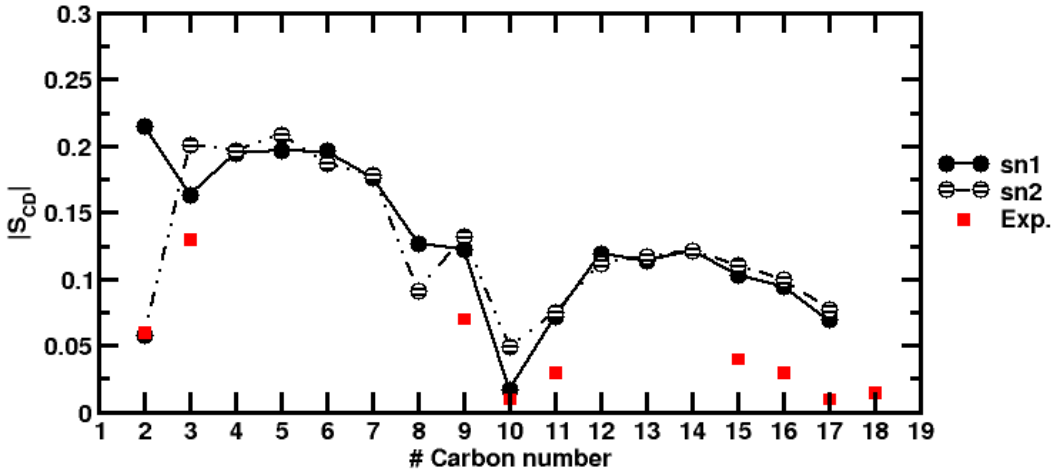


Figure S4: Deuterium order parameter for the DOPC sn1 (filled circles) and sn2 (shaded circles) acyl chains obtained from the 50 ns of the NPT run. The experimental results (red squares) [33] are shown for comparison.

From this figure, it is clear that, in agreement with previous DOPC membrane simulations carried out with the CHARMM36 and the TIP3P CHARMM water model[22,29] or other force fields[24,27], the $|S_{CD}|$ values for the C_2 atoms in the sn1 and sn2 acyl chains are markedly different, indicating different conformations for these atoms in the bilayer. Moreover, the higher $|S_{CD}|$ values for the other carbon atoms relative to experiment indicate that the DOPC chains are slightly too ordered in the simulations. This discrepancy was also found in other simulations[22]. Finally, we note that the unsaturated chains of the DOPC lipid show a distinctive drop around the carbon 10 position due to the *cis* double bond.

The relative arrangement of the DOPC headgroup in the membrane was examined by computing the normalized probability distribution of the $^{-}$P-N^{+} dipole vector with respect to the bilayer normal, which is shown in Figure S5.

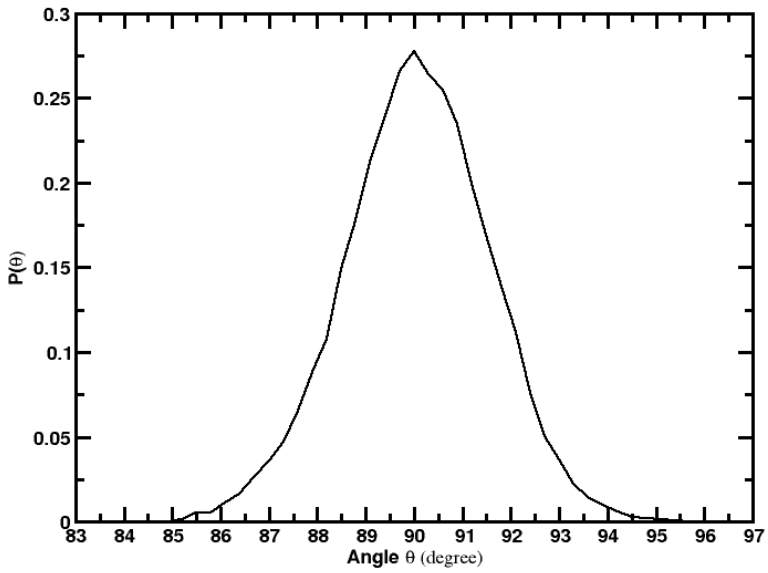


Figure S5: Normalized distribution of angle θ between the $^{-}$P-N^{+} vector and the bilayer normal.

The strong peak in the distribution at ~90° is close to the experimental estimate of 72° obtained with ^{2}H- and ^{31}P-NMR experiments by Akutsu *et al.*, [34] and is in accord with previous simulations[31,35-38] of different phospholipid membranes which obtained values in the range 60-90°.

Experiments[39] and MD simulations[40] have shown that the PL acyl chains and the PL headgroup conformations changed significantly with the increase of the W_0 value the decrease of the penetration of the benzene in the acyl chains see sections 3.1.3 and 3.2). To examine this, we first computed the thickness of the shell of DOPC alkyl chains by computing from the average end-to-end (d_{e-e}) distance[41] between the first (C_{1A} and C_{1B}) and last (C_{18A} and C_{18B}) carbons of the sn1 and sn2 chains (see Scheme 1). The d_{e-e} values averaged over the two acyl chains lies between 14 and 15 Å, shorter than the length for an 18:1 extended chain (~18.0 Å computed with PyMol and the thickness estimated from X-ray measurements (~17.4 Å)[1,42] but close our membrane simulation (~15 Å).

A more complete study can be done by examining the conformation changes within the acyl chains and the DOPC headgroup by computing the normalized probability distribution for each dihedral angle, P(Φ), in the two acyl chains. The dihedral angle is gauche^{+-} where $-180° \leq \Phi \leq -120°$ and $120° \geq \Phi \geq 180°$[43]. The results obtained for the two DOPC acyl chains and the membrane are depicted in Figs. S10 and S11. The figure shows that the conformations of the two chains are approximately equivalent to the conformations found in the membrane and do not noticeably change with W_0 values. The CCCC dihedrals are mainly *trans* when they are located between the headgroup and the C=C double bounds and at the end of the acyl chains. In case of the five dihedral angles involving the C=C double bonds as in the membrane are gauche^{+-}

Regarding the selected dihedral angles within the DOPC headgroup ($NC_5C_6O_{11}$, $C_5C_6O_{11}P$, $C_6O_{11}PO_{12}$, $O_{11}PO_{12}C_{13}$ and $PO_{12}C_{13}C_{14}$ see definitions in Scheme 1) the results presented in Figure S12 show that the *gauche* state is extremely dominant. This is the case for the $N_4C_5C_6O_{11}$ angle where this conformation is favored by the attractive electrostatic interaction between the phosphate and choline groups, and for the two dihedral angles involving the ester oxygen atoms ($C_5C_6O_{11}P$ and $C_6O_{11}PO_{12}$), in agreement with previous observations[34,44-46]. The gauche conformation for these dihedrals slightly increases with the RM water loading to tend to the values obtained in the membrane.

S3 Additional Tables

MD_id	Method	W_0	n_{DOPC}	n_{H2O}	n_{Benz}	R_{H2O}	n_{atm}	$\%wt_{benz}$	L_{box}	ρ_{sys}	t_{sim}
1	Optimized	1	15	15	2939	4.0	37383	95.0	77.7	0.85	90.0
2	Fixed		56	56	4000	6.0	55896	87.4	88.0	0.87	55.0
3	Optimized	2	17	34	3404	4.0	43296	95.0	81.6	0.85	90.0
4	Fixed		65	130	4500	7.0	63360	86.7	91.7	0.87	60.0
5	Optimized	4	22	88	4234	5.0	54108	95.0	87.1	0.88	90.0
6	Fixed		86	344	6100	8.0	86100	86.5	101.6	0.87	80.0
7	Optimized	8	34	272	7704	7.0	97956	95.0	106.6	0.87	90.0
8	Attempt		71	568	5200	10.0	73902	95.0	95.9	0.89	54.0
9	Attempt	12	51	612	9915	11.0	127854	95.0	116.9	0.86	36.0
10	Attempt		53	636	12938	11.0	164478	95.0	127.3	0.85	90.0
11	Optimized		69	828	16843	13.0	214122	95.0	139.2	0.85	100.0
12	Attempt		85	1020	20749	15.0	263778	95.0	149.1	0.86	19.7
13	Attempt	16	70	1120	18315	17.0	232800	95.0	143.0	0.85	90.0
14	Optimized		87	1392	22763	18.0	298338	95.0	153.9	0.85	110.0

Table S2: Overview of the simulations carried out in this work. For each simulation the following parameters are listed: the water-to-surfactant ratio, W_0; the number of DOPC (n_{DOPC}), water (n_{H2O}), and benzene (n_{Benz}) molecules; the initial water core radius, R_{H2O}, (in Å) used to construct the initial conformation of the RM; the total number of particles in each system, n_{atm}; the benzene mass fraction, $\%wt_{benz}$ (in %); the average cubic box side length, L_{box} (in Å) and the system density, ρ_{sys}, (in $g.cm^{-3}$) after the NPT simulations; and the simulation production time (in ns). The "Fixed" systems were constructed with eq. 1 in the main text, while "Attempt" systems were constructed with the "trial and error" strategy (see main text for details), leading to "Optimized" systems.

MD_id	Method	W_0	*a*	*b*	*c*	*a/c*	*e*
1	Optimized	1	22.36	19.49	17.41	1.28	0.63
2	Fixed		29.23	27.71	25.75	1.14	0.47
3	Optimized	2	22.51	20.05	18.37	1.23	0.58
4	Fixed		31.11	28.31	26.20	1.19	0.54
5	Optimized	4	23.45	20.35	19.85	1.18	0.53
6	Fixed		34.97	32.52	27.39	1.28	0.62
7	Optimized	8	28.67	27.68	25.25	1.14	0.47
8	Attempt		36.20	30.70	26.29	1.38	0.69
9	Attempt	12	31.93	29.08	27.53	1.16	0.51
10	Attempt		32.72	29.56	28.01	1.17	0.52
11	Optimized		34.25	32.43	31.47	1.09	0.39
12	Attempt		38.91	35.63	33.52	1.16	0.51
13	Attempt	16	34.43	32.41	30.57	1.13	0.46
14	Optimized		39.79	37.61	34.82	1.14	0.48

Table S3: Average shape parameters for the whole RM (*i.e.*, including both DOPC and water). The a_{RM}, b_{RM}, and c_{RM} are the overall RM average semi-axis lengths (in Å), whereas a_{RM}/c_{RM} and e_{RM} are the ratio between the major (*a*) and the minor (*c*) semi-axis and the eccentricity, respectively. The standard errors for the semi-axis lengths and of the eccentricity values are not greater than 0.2 Å and 0.05, respectively.

W_0	1	2	4	8	12	16	Membrane
Benzene – DOPC/Water							
$SASA_{tot}^{benz}$	12277.7	13921.2	17276.5	25473.7	45947.2	56814.2	-
$\%SASA_{w}^{benz}$	0.9	1.6	2.7	4.4	4.3	5.2	-
$\%SASA_{Chol}^{benz}$	4.0	3.7	2.9	2.1	1.5	1.3	-
$\%SASA_{Phos}^{benz}$	0.6	0.5	0.3	0.2	0.1	0.1	-
$\%SASA_{glyc}^{benz}$	7.8	7.3	7.0	5.8	5.5	5.4	-
$\%SASA_{tail}^{benz}$	86.7	86.9	87.1	86.9	88.5	88.0	-
$\%SASA_{tot}^{benz}$	100.0						-
Water – DOPC							
$SASA_{tot}^{w}$	605.7	1194.3	2670.3	6261.5	14281.2	19542.4	32937.9
$\%SASA_{Chol}^{w}$	53.0	52.5	52.7	50.4	51.0	50.6	49.8
$\%SASA_{phos}^{w}$	22.2	20.2	18.2	15.7	15.0	14.7	13.8
$\%SASA_{glyc}^{w}$	17.3	17.9	20.5	23.6	24.3	24.1	24.5
$\%SASA_{tail}^{w}$	7.5	9.3	8.6	10.2	9.8	10.5	12.0
$\%SASA_{tot}^{w}$	100						

Table S4: Average solvent accessible surface area (*SASA*) in Å^2 for benzene in contact with DOPC and water as well as water in contact with DOPC as a function of the DOPC hydration in the RM and membrane. The percent contribution (*%SASA*) of different kinds of contacts is also given, with the contact indicated by the subscript. Standard errors of the values are not greater than 0.3 %.

W_0	1	2	4	8	12	16	Membrane
N	2.7	5.3	9.1	13.4	15.4	16.6	18.6
P	1.5	2.5	3.8	4.9	5.2	5.6	5.8
O_9	0.8	1.2	1.6	1.8	1.9	2.0	2.1
O_{10}	0.7	1.0	1.5	1.8	1.9	2.0	2.1
O_{11}	~0.0	0.1	0.4	0.7	0.7	0.7	0.8
O_{12}	~0.0	~0.1	0.3	0.5	0.5	0.8	0.7
O_{31}	~0.0	~0.0	~0.0	~0.1	~0.1	~0.1	0.1
O_{32}	0.1	0.2	0.5	0.9	0.9	1.0	1.0
O_{21}	~0.0	0.0	0.1	0.2	0.3	0.3	0.4
O_{22}	~0.1	0.2	0.5	0.8	0.8	0.9	1.0

Table S5: Average hydration numbers for selected atoms of the DOPC headgroup (including the glycerol) obtained in the RM and membrane simulations computed from radial pair density functions, $g_{X\text{-}Ow}(r)$, between the water oxygen, O_W, and selected atoms of the DOPC headgroup (namely, choline (X = N), phosphate (X = P), 4 oxygen atoms bonded to the phosphorus (X = O_9, O_{10}, O_{11}, and O_{12}) and 4 glycerol oxygen atoms (iX = O_{21}, O_{22}, O_{31}, and O_{32}). Values were obtained by integrating the radial density functions, $g_{X\text{-}Ow}$, up to their first minimum at 5.9, 4.5 and ~3.3 Å for nitrogen, phosphate and oxygen, respectively.

S4 Additional Figures

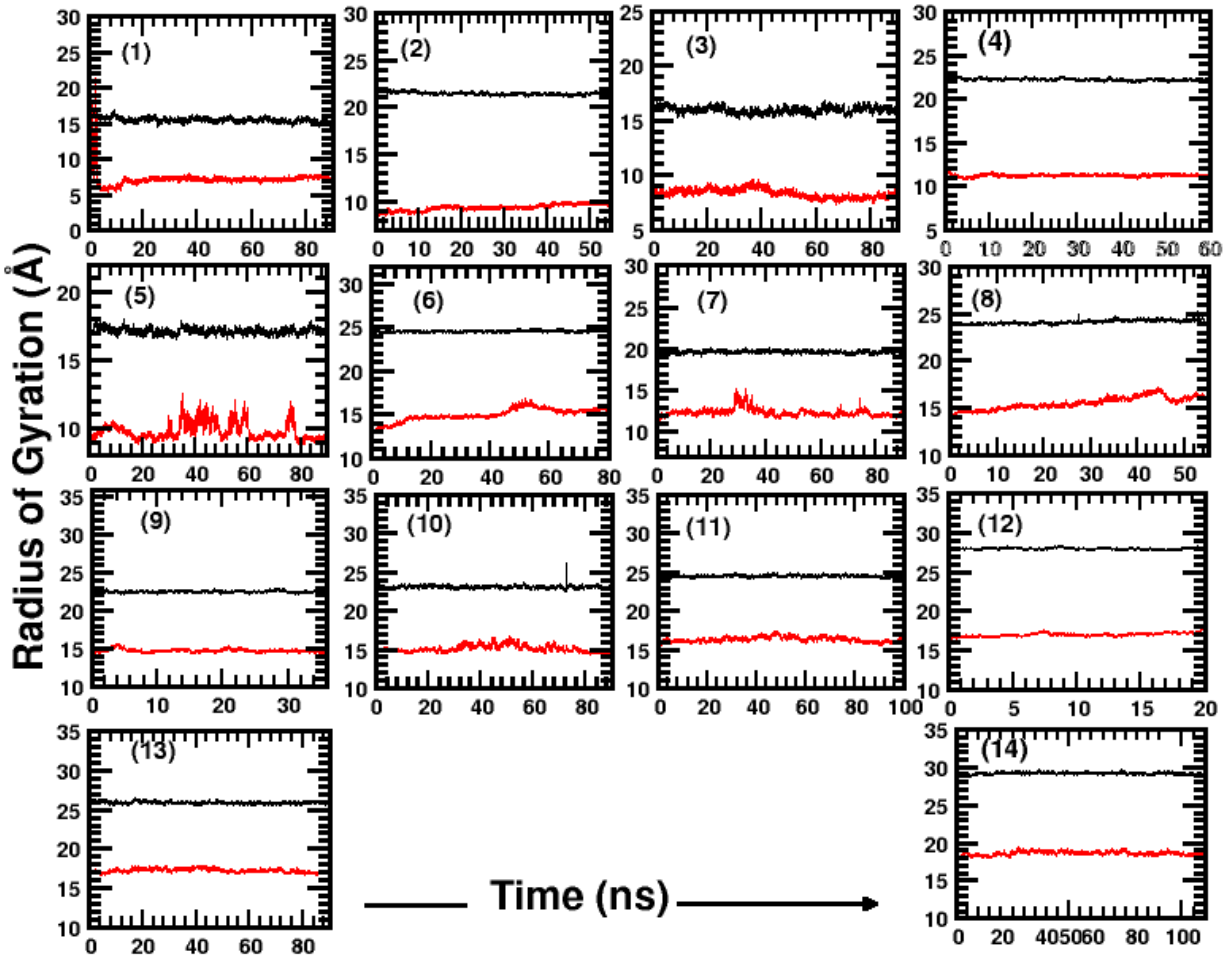


Figure S6: Time evolution of the radii of gyration of the reverse micelles (black) and water cores (red). See Table S2 for the corresponding RM system. The RMs discussed in the main text correspond to the inset number of 1, 3, 5, 7, 11, and 14.

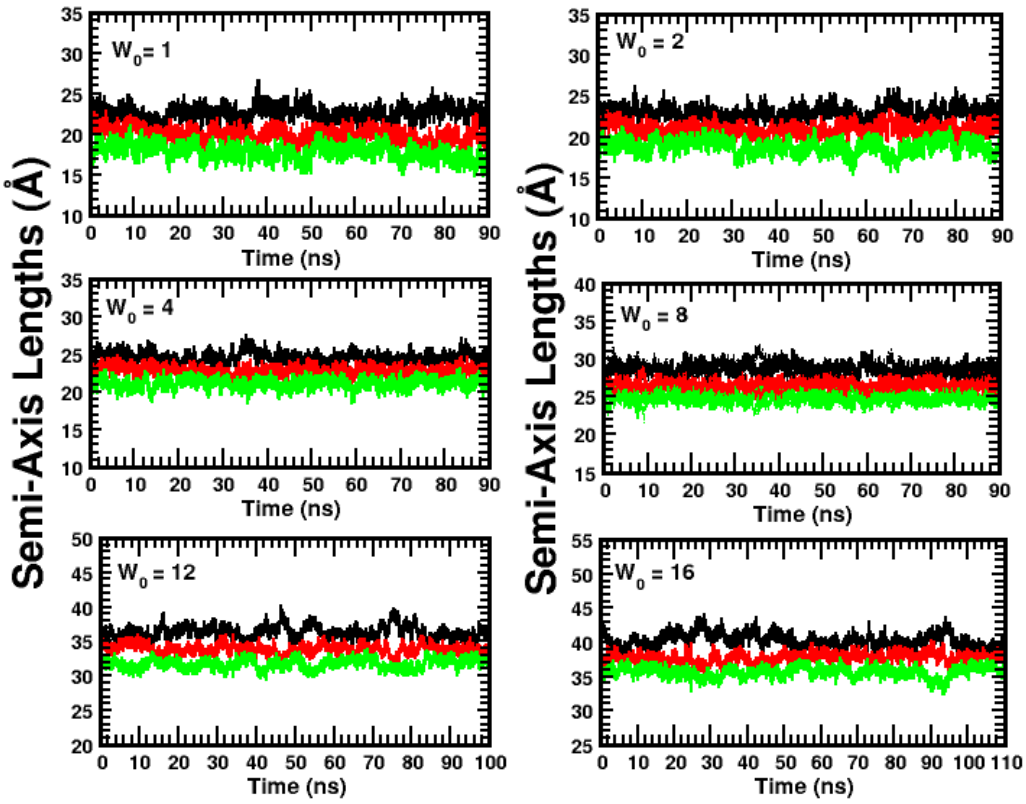


Figure S7: Time evolution of the RM semi-axis lengths *a* (black), *b* (red), and *c* (green).

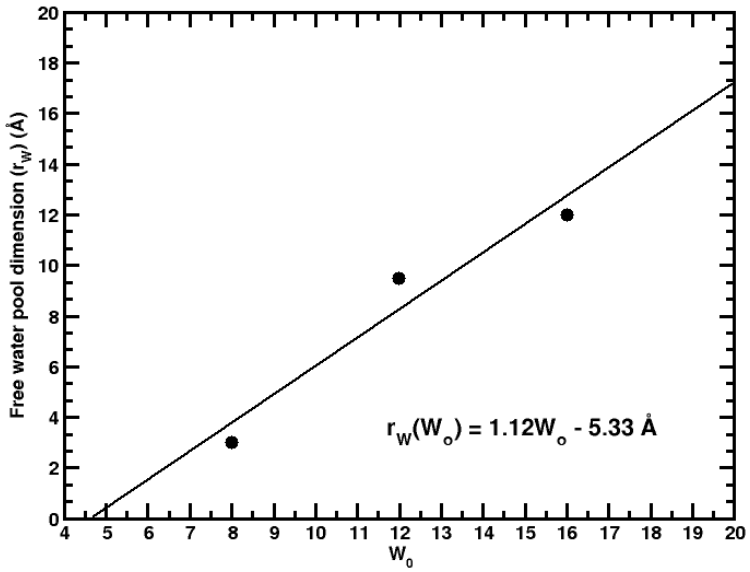


Figure S8: Linear dependence of the RM pool of free water as a function of W_0. The water pool radii (black circles) were obtained from the radial density profiles depicted in Figure 6 in the main text; a linear fit to the data is also shown (black line) and the fit parameters indicated.

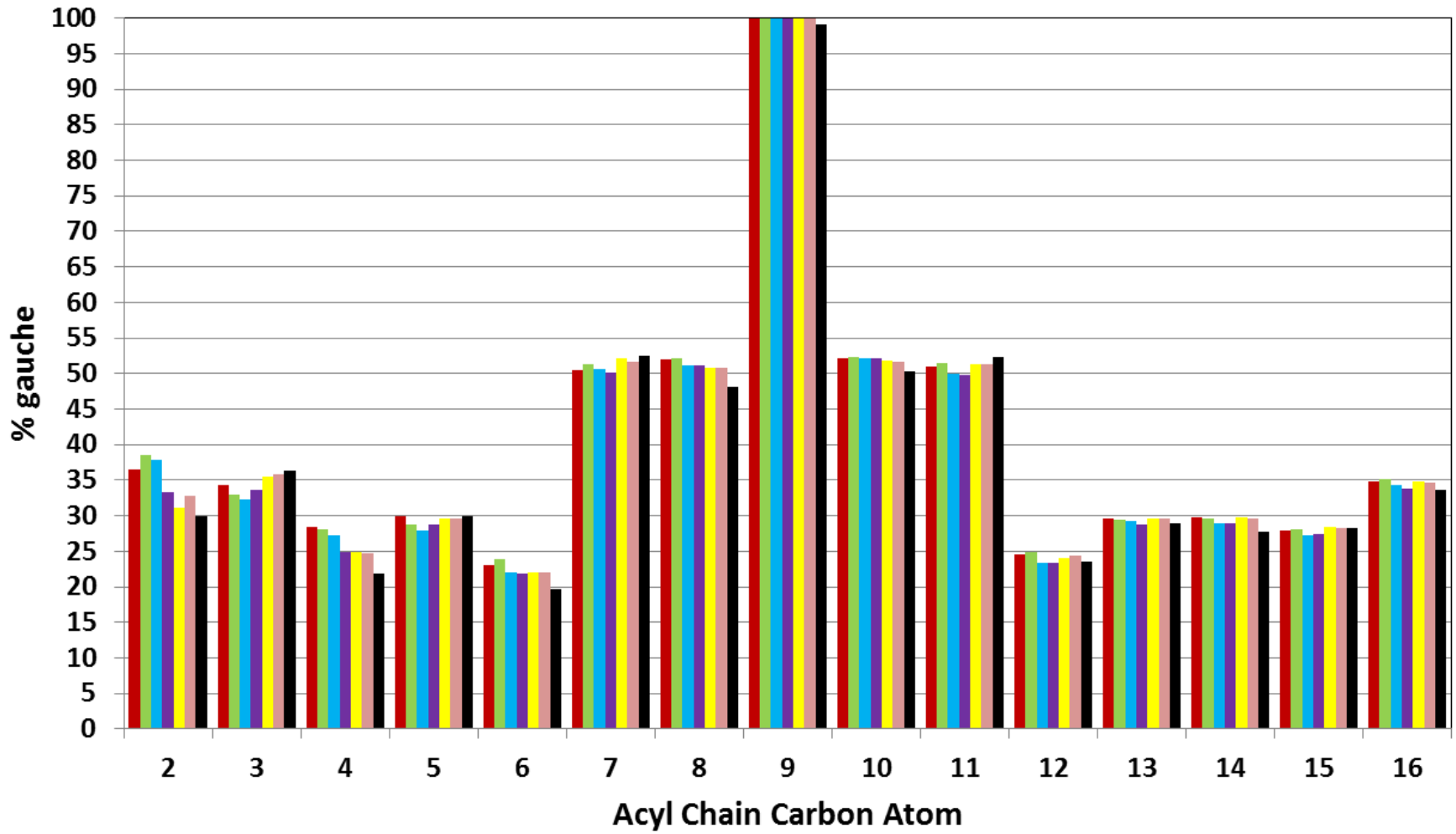


Figure S9: Fractions of *gauche*$^{+-}$ conformations for selected CCCC dihedrals of the DOPC sn1 acyl chains for $W_0 = 1$ (brown), 2 (green), 4 (blue), 8 (magenta), 12 (yellow), and 16 (brown) as well as for a bilayer membrane (black), computed from the second half of each trajectory.

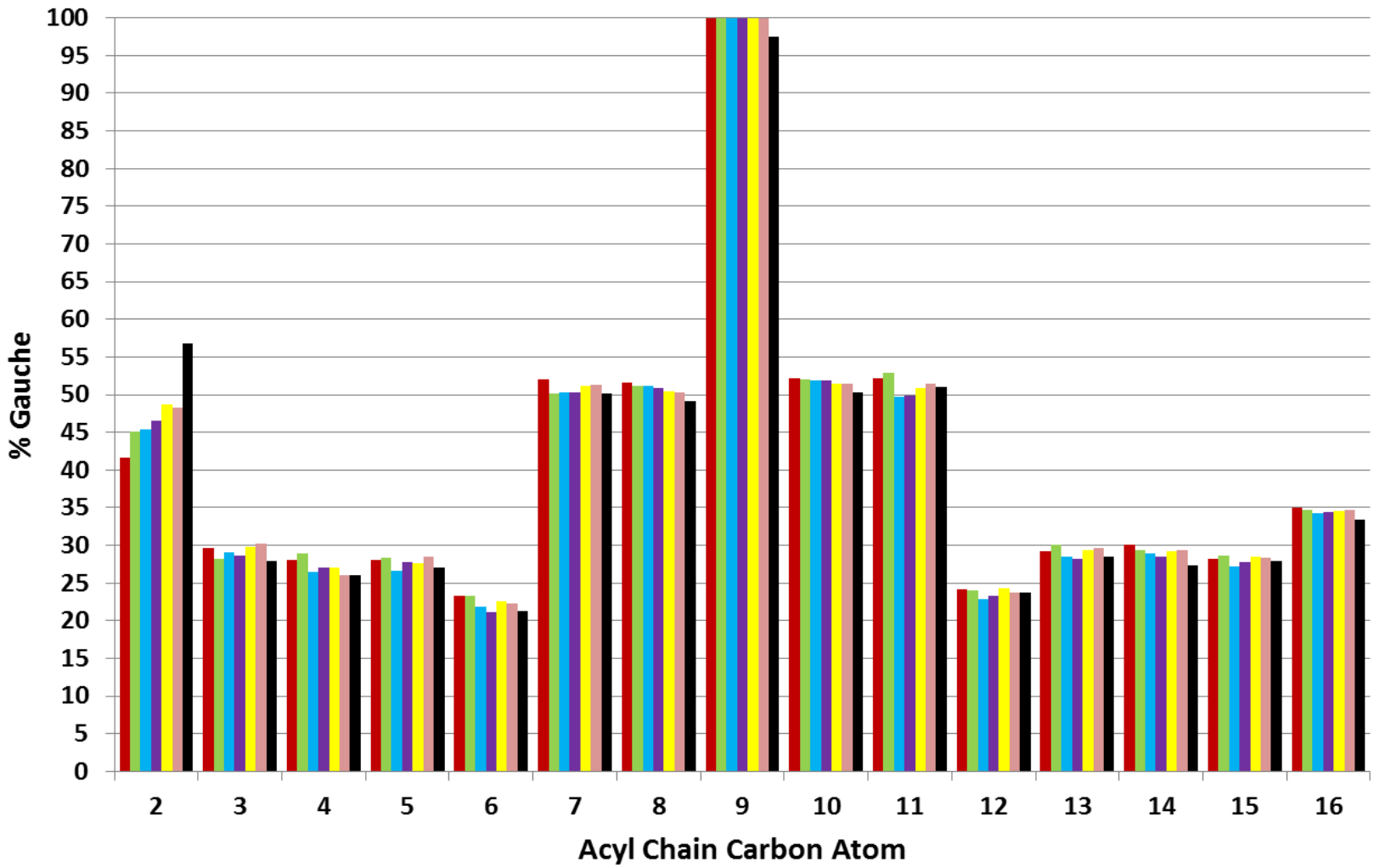


Figure S10: Same as Figure S9, but the conformation of the dihedral angle involving atoms n-1, n, n+1, n+2 is reported for each carbon atom index in the DOPC sn2 acyl chains.

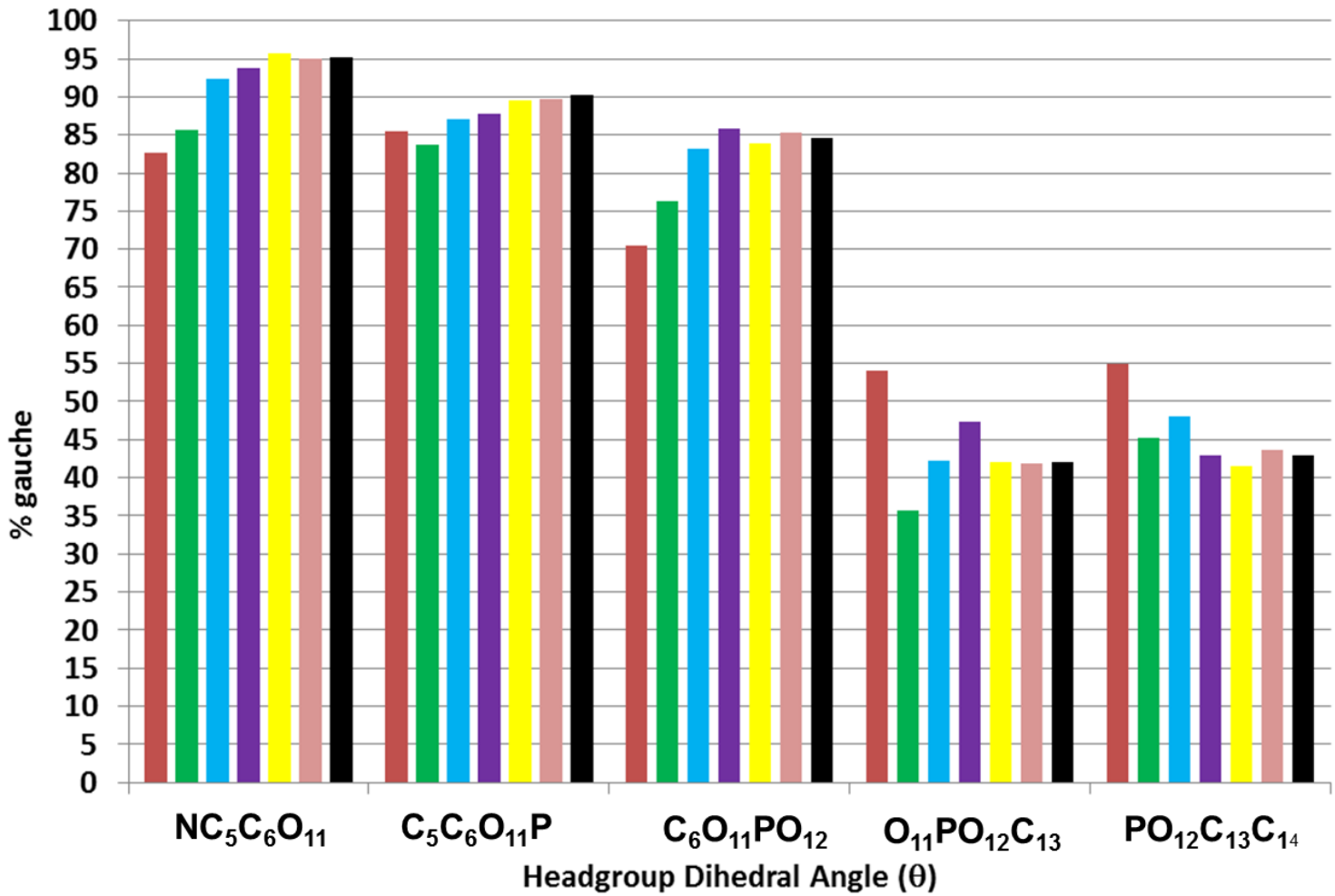


Figure S11: Same as Figure S9, but for headgroup dihedral angles defined by the angles listed for each set of data.

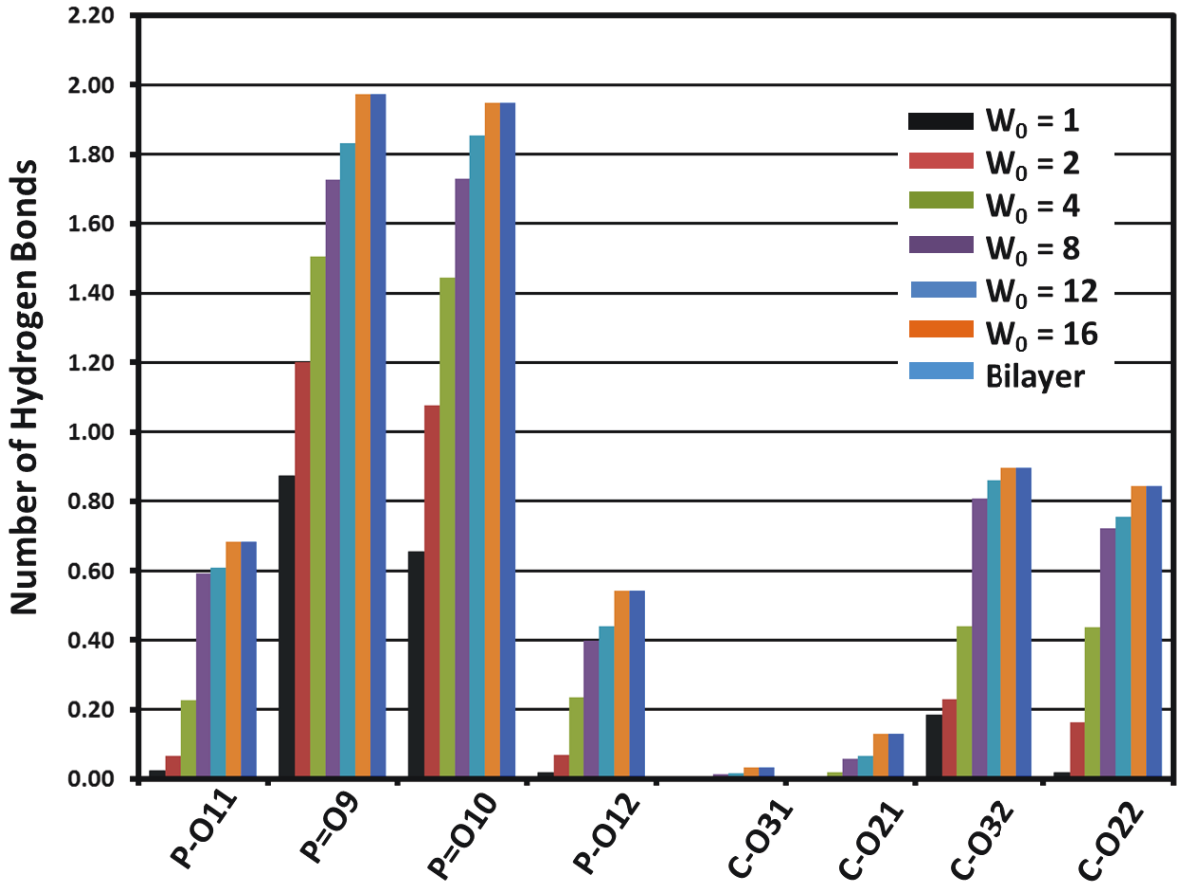


Figure S12: Average number of hydrogen bonds (HBs) for each DOPC acceptor site as a function of the water-to-surfactant ratio. The errors are each less than 0.03.

4. References

(1) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(2) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **2008**, *4*, 435–447.

(3) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J Chem Theory Comput* **2010**, *6*, 459–466.

(4) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **2008**, *29*, 1859–1865.

(5) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S. CHARMM: the biomolecular simulation program. *J Comput Chem* **2009**, *30*, 1545–1614.

(6) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* **2014**, *35*, 1997–2004.

(7) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys J* **2009**, *97*, 50–58.

(8) Jo, S.; Kim, T.; Im, W. Automated builder and database of protein/membrane complexes for molecular dynamics simulations. *PLoS One* **2007**, *2*, e880.

(9) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J Chem Phys* **1995**, *103*, 8577–8593.

(10) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J Chem Theory Comput* **2007**, *4*, 116–122.

(11) Piggot, T. J.; Piñeiro, Á.; Khalid, S. Molecular Dynamics Simulations of Phosphatidylcholine Membranes: A Comparative Force Field Study. *J Chem Theory Comput* **2012**, *8*, 4593–4609.

(12) Pastor, R. W.; Mackerell, A. D. Development of the CHARMM Force Field for Lipids. *J Phys Chem Lett* **2011**, *2*, 1526–1532.

(13) Schmidt, J. R.; Roberts, S. T.; Loparo, J. J.; Tokmakoff, A.; Fayer, M. D.; Skinner, J. L. Are water simulation models consistent with steady-state and ultrafast vibrational spectroscopy experiments? *Chem Phys* **2007**, *341*, 143–157.

(14) Mark, P.; Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *The Journal of Physical Chemistry A* **2001**, *105*, 9954–9960.

(15) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J Phys Chem B* **2010**, *114*, 7830–7843.

(16) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **2007**, *126*, 014101–014101.

(17) Rahman, A.; Stillinger, F. H. Molecular dynamics study of liquid water. *J Chem Phys* **1971**, *55*, 3336–3359.

(18) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* **1981**, *52*, 7182–7190.

(19) Kucerka, N.; Nagle, J. F.; Sachs, J. N.; Feller, S. E.; Pencer, J.; Jackson, A.; Katsaras, J. Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data. *Biophys J* **2008**, *95*, 2356–2367.

(20) Nagle, J. F.; Tristram-Nagle, S. Structure of lipid bilayers. *BBA-Rev Biomembranes* **2000**, *1469*, 159–195.

(21) Pan, J.; Tristram-Nagle, S.; Kucerka, N.; Nagle, J. F. Temperature dependence of structure, bending rigidity, and bilayer interactions of dioleoylphosphatidylcholine bilayers. *Biophys J* **2008**, *94*, 117–124.

(22) Loubet, B.; Kopec, W.; Khandelia, H. Accelerating All-Atom MD Simulations of Lipids Using a Modified Virtual-Sites Technique. *J Chem Theory Comput* **2014**, *10*, 5690–5695.

(23) Jones, F. E.; Harris, G. L. ITS-90 density of water formulation for volumetric standards calibration. *Journal of Research of the National Institute of Standards and Technology* **1992**, *97*, 335–340.

(24) Dickson, C. J.; Rosso, L.; Betz, R. M.; Walker, R. C.; Gould, I. R. GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter* **2012**, *8*, 9617–9627.

(25) Jämbeck, J. P. M.; Lyubartsev, A. P. Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids. *J Phys Chem B* **2012**, *116*, 3164–3179.

(26) Chiu, S. -W.; Pandit, S. A.; Scott, H. L.; Jakobsson, E. An Improved United Atom Force Field for Simulation of Mixed Lipid Bilayers. *J Phys Chem B* **2009**, *113*, 2748–2763.

(27) Dickson, C. J.; Madej, B. D.; Skjevik, A. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J Chem Theory Comput* **2014**, *10*, 865–879.

(28) Bhatnagar, N.; Kamath, G.; Potoff, J. J. Biomolecular simulations with the transferable potentials for phase equilibria: extension to phospholipids. *J Phys Chem B* **2013**, *117*, 9910–9921.

(29) Liu, A.; Qi, X. Molecular dynamics simulations of DOPC lipid bilayers: The effect of Lennard-Jones parameters of hydrocarbon chains. *Comput Mol Biosci* **2012**, *2*, 78.

(30) Jämbeck, J. P. M.; Lyubartsev, A. P. An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes. *J Chem Theory Comput* **2012**, *8*, 2938–2948.

(31) Poger, D.; Mark, A. E. On the Validation of Molecular Dynamics Simulations of Saturated and cis-Monounsaturated Phosphatidylcholine Lipid Bilayers: A Comparison with Experiment. *J Chem Theory Comput* **2010**, *6*, 325–336.

(32) Petrache, H. I.; Feller, S. E.; Nagle, J. F. Determination of component volumes of lipid bilayers from simulations. *Biophys J* **1997**, *72*, 2237–2242.

(33) Warschawski, D. E.; Devaux, P. F. Order parameters of unsaturated phospholipids in membranes and the effect of cholesterol: a 1H--13C solid-state NMR study at natural abundance. *European Biophysics Journal* **2005**, *34*, 987–996.

(34) Akutsu, H.; Nagamori, T. Conformational analysis of the polar head group in phosphatidylcholine bilayers: a structural change induced by cations. *Biochemistry* **1991**, *30*, 4510–4516.

(35) Rosso, L.; Gould, I. R. Structure and dynamics of phospholipid bilayers using recently developed general all-atom force fields. *J Comput Chem* **2008**, *29*, 24–37.

(36) Ulmschneider, J. P.; Doux, J. P.; Killian, J. A.; Smith, J. C.; Ulmschneider, M. B. Peptide partitioning and folding into lipid bilayers. *J Chem Theory Comput* **2009**, *5*, 2202–2205.

(37) Taylor, J.; Whiteford, N. E.; Bradley, G.; Watson, G. W. Validation of all-atom phosphatidylcholine lipid force fields in the tensionless NPT ensemble. *Biochim Biophys Acta* **2009**, *1788*, 638–649.

(38) Siu, S. W.; Vácha, R.; Jungwirth, P.; Böckmann, R. A. Biomolecular simulations of membranes: physical properties from different force fields. *J Chem Phys* **2008**, *128*, 125103.

(39) Maitra, A.; Jain, T. K.; Shervani, Z. Interfacial water structure in lecithin—oil—water reverse micelles. *Colloid Surface* **1990**, *47*, 255–267.

(40) Vierros, S.; Sammalkorpi, M. Phosphatidylcholine reverse micelles on the wrong track in molecular dynamics simulations of phospholipids in an organic solvent. *J Chem Phys* **2015**, *142*, 094902.

(41) Abel, S.; Sterpone, F.; Bandyopadhyay, S.; Marchi, M. Molecular Modeling and Simulations of AOT−Water Reverse Micelles in Isooctane: Structural and Dynamic Properties. *J Phys Chem B* **2004**, *108*, 19458–19466.

(42) Costigan, S. C.; Booth, P. J.; Templer, R. H. Estimations of lipid bilayer geometry in fluid lamellar phases. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2000**, *1468*, 41–54.

(43) Abel, S.; Dupradeau, F. -Y.; Marchi, M. Molecular Dynamics Simulations of a Characteristic DPC Micelle in Water. *J Chem Theory Comput* **2012**, *8*, 4610–4623.

(44) Hauser, H.; Guyer, W.; Pascher, I.; Skrabal, P.; Sundell, S. Polar group conformation of phosphatidylcholine. Effect of solvent and aggregation. *Biochemistry* **1980**, *19*, 366–373.

(45) Pandey, P. R.; Roy, S. Headgroup mediated water insertion into the DPPC bilayer: a molecular dynamics study. *J Phys Chem B* **2011**, *115*, 3155–3163.

(46) Krishnamurty, S.; Stefanov, M.; Mineva, T.; Bégu, S.; Devoisselle, J. M.; Goursot, A.; Zhu, R.; Salahub, D. R. Density functional theory-based conformational analysis of a phospholipid molecule (dimyristoyl phosphatidylcholine). *J Phys Chem B* **2008**, *112*, 13433–13442.

Références

Abel, S., Dupradeau, F.-Y., Marchi, M. (2012). Molecular Dynamics Simulations of a Characteristic DPC Micelle in Water. *Journal of Chemical Theory and Computation*, *8*(11), 4610–4623. http://doi.org/10.1021/ct3003207

Abel, S., Dupradeau, F.-Y., Raman, E. P., MacKerell, A. D., Marchi, M. (2011). Molecular Simulations of Dodecyl-β-maltoside Micelles in Water: Influence of the Headgroup Conformation and Force Field Parameters. *The Journal of Physical Chemistry B*, *115*(3), 487–499. http://doi.org/10.1021/jp109545v

Abel, S., Lorieau, A., de Foresta, B., Dupradeau, F.-Y., Marchi, M. (2014). Bindings of hMRP1 transmembrane peptides with dodecylphosphocholine and dodecyl-β-d-maltoside micelles: A molecular dynamics simulation study. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1838*(1, Part B), 493–509. http://doi.org/10.1016/j.bbamem.2013.10.012

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. http://doi.org/10.1016/j.softx.2015.06.001

Alder, B. J., Wainwright, T. E. (1957). Phase Transition for a Hard Sphere System. *The Journal of Chemical Physics*, *27*(5), 1208–1209. http://doi.org/10.1063/1.1743957

Allen, M. P., Tildesley, D. J. (1989). *Computer Simulation of Liquids*. New York, NY, USA: Clarendon Press.

Andersen, H. C. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics*, *72*(4), 2384–2393. http://doi.org/10.1063/1.439486

Aoudia, M., Zana, R. (1998). Aggregation Behavior of Sugar Surfactants in Aqueous Solutions: Effects of Temperature and the Addition of Nonionic Polymers. *J. Colloid Interface Sci.*, *206*(1), 158–167.

Bahatyrova, S., Frese, R. N., Werf, K. O. van der, Otto, C., Hunter, C. N., Olsen, J. D. (2004). Flexibility and Size Heterogeneity of the LH1 Light Harvesting Complex Revealed by Atomic Force Microscopy Functional Significance For Bacterial Photosynthesis. *Journal of Biological Chemistry*, *279*(20), 21327–21333. http://doi.org/10.1074/jbc.M313039200

Bayly, C. I., Cieplak, P., Cornell, W., Kollman, P. A. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry*, *97*(40), 10269–10280. http://doi.org/10.1021/j100142a004

Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. van, DiNola, A., Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*(8), 3684–3690. http://doi.org/10.1063/1.448118

Bogusz, S., Venable, R. M., Pastor, R. W. (2000). Molecular Dynamics Simulations of Octyl Glucoside Micelles: Structural Properties. *The Journal of Physical Chemistry B*, *104*(23), 5462–5470. http://doi.org/10.1021/jp000159y

Bogusz, S., Venable, R. M., Pastor, R. W. (2001). Molecular Dynamics Simulations of Octyl Glucoside Micelles: Dynamic Properties. *The Journal of Physical Chemistry B*, *105*(35), 8312–8321. http://doi.org/10.1021/jp004475o

Born, M., Oppenheimer, R. (1927). Zur Quantentheorie der Molekeln. *Annalen Der Physik*, *389*(20), 457–484. http://doi.org/10.1002/andp.19273892002

Braun, P., Olsen, J. D., Strohmann, B., Hunter, C. N., Scheer, H. (2002). Assembly of Light-harvesting Bacteriochlorophyll in a Model Transmembrane Helix in its Natural Environment. *Journal of Molecular Biology*, *318*(4), 1085–1095. http://doi.org/10.1016/S0022-2836(02)00192-4

Breneman, C. M., Wiberg, K. B. (1990). Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *Journal of Computational Chemistry*, *11*(3), 361–373. http://doi.org/10.1002/jcc.540110311

Bui, E. T., Bradley, P. J., Johnson, P. J. (1996). A common evolutionary origin for mitochondria and hydrogenosomes. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(18), 9651–9656.

Bussi, G., Donadio, D., Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, *126*(1), 014101. http://doi.org/10.1063/1.2408420

Cartron, M. L., Olsen, J. D., Sener, M., Jackson, P. J., Brindley, A. A., Qian, P., Neil Hunter, C. (2014). Integration of energy and electron transfer processes in the photosynthetic membrane of Rhodobacter sphaeroides. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1837*(10), 1769–1780. http://doi.org/10.1016/j.bbabio.2014.02.003

Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Kollman, P. A. (2015). AMBER 14.

Ceccarelli, M., Procacci, P., Marchi, M. (2003). An ab initio force field for the cofactors of bacterial photosynthesis. *Journal of Computational Chemistry*, *24*(2), 129–142. http://doi.org/10.1002/jcc.10198

Chakrabarti, N., Neale, C., Payandeh, J., Pai, E. F., Pomès, R. (2010). An Iris-Like Mechanism of Pore Dilation in the CorA Magnesium Transport System. *Biophysical Journal*, *98*(5), 784–792. http://doi.org/10.1016/j.bpj.2009.11.009

Chandler, D. E., Hsin, J., Harrison, C. B., Gumbart, J., Schulten, K. (2008). Intrinsic Curvature Properties of Photosynthetic Proteins in Chromatophores. *Biophysical Journal*, *95*(6), 2822–2836. http://doi.org/10.1529/biophysj.108.132852

Chandler, D. E., Strümpfer, J., Sener, M., Scheuring, S., Schulten, K. (2014). Light Harvesting by Lamellar Chromatophores in Rhodospirillum photometricum. *Biophysical Journal*, *106*(11), 2503–2510. http://doi.org/10.1016/j.bpj.2014.04.030

Cheng, X., Jo, S., Lee, H. S., Klauda, J. B., Im, W. (2013). CHARMM-GUI Micelle Builder for Pure/Mixed Micelle and Protein/Micelle Complex Systems. *Journal of Chemical Information and Modeling*, *53*(8), 2171–2180. http://doi.org/10.1021/ci4002684

Cherezov, V., Clogston, J., Papiz, M. Z., Caffrey, M. (2006). Room to Move: Crystallizing Membrane Proteins in Swollen Lipidic Mesophases. *Journal of Molecular Biology*, *357*(5), 1605–1618. http://doi.org/10.1016/j.jmb.2006.01.049

Chernyak, V., Mukamel, S. (1996). Collective coordinates for nuclear spectral densities in energy transfer and femtosecond spectroscopy of molecular aggregates. *The Journal of Chemical Physics*, *105*(11), 4565–4583. http://doi.org/10.1063/1.472302

Ciccotti, G., Ryckaert, J. P. (1986). Molecular dynamics simulation of rigid molecules. *Computer Physics Reports*, *4*(6), 346–392. http://doi.org/10.1016/0167-7977(86)90022-5

Cogdell, R. J., Gall, A., Köhler, J. (2006). The architecture and function of the light-harvesting apparatus of purple bacteria: from single molecules to in vivo membranes. *Quarterly Reviews of Biophysics*, *39*(03), 227–324. http://doi.org/10.1017/S0033583506004434

Cogdell, R. J., Roszak, A. W. (2014). Structural biology: The purple heart of photosynthesis. *Nature*, *508*(7495), 196–197. http://doi.org/10.1038/nature13219

Cogdell, R. J., Scheer, H. (1985). Circular Dichroism of Light-Harvesting Complexes from Purple Photosynthetic Bacteria. *Photochemistry and Photobiology*, *42*(6), 669–678. http://doi.org/10.1111/j.1751-1097.1985.tb01629.x

Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, *221*(4612), 709–713. http://doi.org/10.1126/science.6879170

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Kollman, P. A. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, *117*(19), 5179–5197. http://doi.org/10.1021/ja00124a002

Cueille, S., Sire, C. (1997). Smoluchowski's equation for cluster exogenous growth. *Europhysics Letters*, *40*(3), 239–244. http://doi.org/10.1209/epl/i1997-00454-4

Darden, T., York, D., Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12), 10089–10092. http://doi.org/10.1063/1.464397

Deisenhofer, J., Epp, O., Miki, K., Huber, R., Michel, H. (1984). X-ray structure analysis of a membrane protein complex. *Journal of Molecular Biology*, *180*(2), 385–398. http://doi.org/10.1016/S0022-2836(84)80011-X

Dickson, C. J., Madej, B. D., Skjevik, Å. A., Betz, R. M., Teigen, K., Gould, I. R., Walker, R. C. (2014). Lipid14: The Amber Lipid Force Field. *Journal of Chemical Theory and Computation*, *10*(2), 865–879. http://doi.org/10.1021/ct4010307

Dupradeau, F.-Y., Cézard, C., Lelong, R., Stanislawiak, E., Pêcher, J., Delepine, J. C., Cieplak, P. (2008). R.E.DD.B.: a database for RESP and ESP atomic charges, and force field libraries. *Nucleic Acids Research*, *36*(Database issue), D360–367. http://doi.org/10.1093/nar/gkm887

Durrant, J. D., McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, *9*, 71. http://doi.org/10.1186/1741-7007-9-71

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, *103*(19), 8577–8593. http://doi.org/10.1063/1.470117

Ewald, P. P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen Der Physik*, *369*(3), 253–287. http://doi.org/10.1002/andp.19213690304

Fahy, E., Subramaniam, S., Brown, H. A., Glass, C. K., Merrill, A. H., Murphy, R. C., Dennis, E. A. (2005). A comprehensive classification system for lipids. *Journal of Lipid Research*, *46*(5), 839–862. http://doi.org/10.1194/jlr.E400004-JLR200

Fleming, G. R., Grondelle, R. van. (2008). The Primary Steps of Photosynthesis. *Physics Today*, *47*(2), 48–55. http://doi.org/10.1063/1.881413

Förster, T. (1948). Intermolecular energy transfer and fluorescence. *Ann. Phys.*, *2*, 55–75.

Fotiadis, D., Qian, P., Philippsen, A., Bullough, P. A., Engel, A., Hunter, C. N. (2004). Structural Analysis of the Reaction Center Light-harvesting Complex I Photosynthetic Core Complex of Rhodospirillum rubrum Using Atomic Force Microscopy. *Journal of Biological Chemistry*, *279*(3), 2063–2068. http://doi.org/10.1074/jbc.M310382200

Frenkel, J. (1931). On the Transformation of light into Heat in Solids. I. *Physical Review*, *37*(1), 17–44. http://doi.org/10.1103/PhysRev.37.17

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Fox, D. J. (2009). Gaussian 09, Revision D.01.

Gall, A., Ilioaia, C., Krüger, T. P. J., Novoderezhkin, V. I., Robert, B., van Grondelle, R. (2015). Conformational Switching in a Light-Harvesting Protein as Followed by Single-Molecule Spectroscopy. *Biophysical Journal*, *108*(11), 2713–2720. http://doi.org/10.1016/j.bpj.2015.04.017

Gall, A., Robert, B. (1999). Characterization of the Different Peripheral Light-Harvesting Complexes from High- and Low-Light Grown Cells from Rhodopseudomonas palustris. *Biochemistry*, *38*(16), 5185–5190. http://doi.org/10.1021/bi982486q

Garrett, R. H., Grisham, C. M. (2000). *Biochimie*. De Boeck Supérieur.

Gasteiger, J., Marsili, M. (1978). A new model for calculating atomic charges in molecules. *Tetrahedron Letters*, *19*(34), 3181–3184. http://doi.org/10.1016/S0040-4039(01)94977-9

Gast, P., Hemelrijk, P., Hoff, A. J. (1994). Determination of the number of detergent molecules associated with the reaction center protein isolated from the photosynthetic bacterium Rhodopseudomonas viridis. Effects of the amphiphilic molecule 1,2,3-heptanetriol. *FEBS Letters*, *337*(1), 39–42.

Gentemann, S., Nelson, N. Y., Jaquinod, L., Nurco, D. J., Leung, S. H., Medforth, C. J., Holten, D. (1997). Variations and Temperature Dependence of the Excited State Properties of Conformationally and Electronically Perturbed Zinc and Free Base Porphyrins. *The Journal of Physical Chemistry B*, *101*(7), 1247–1254. http://doi.org/10.1021/jp9631137

Gerken, U., Lupo, D., Tietz, C., Wrachtrup, J., Ghosh, R. (2003). Circular Symmetry of the Light-Harvesting 1 Complex from Rhodospirillum rubrum Is Not Perturbed by Interaction with the Reaction Center†. *Biochemistry*, *42*(35), 10354–10360. http://doi.org/10.1021/bi034969m

Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., Caves, L. S. D. (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, *22*(21), 2695–2696. http://doi.org/10.1093/bioinformatics/btl461

Grondelle, R. van, Novoderezhkin, V. I. (2006). Energy transfer in photosynthesis: experimental insights and quantitative models. *Physical Chemistry Chemical Physics*, *8*(7), 793–807. http://doi.org/10.1039/B514032C

Guinier, A. (1937). Arrangement for obtaining intense diffraction diagrams of crystalline powders with monochromatic radiation. *Compt. Rend.*, *204*, 1115–16.

Harder, E., Anisimov, V. M., Vorobyov, I. V., Lopes, P. E. M., Noskov, S. Y., MacKerell, A. D., Roux, B. (2006). Atomic Level Anisotropy in the Electrostatic Modeling of Lone Pairs for a Polarizable Force Field Based on the Classical Drude Oscillator. *Journal of Chemical Theory and Computation*, *2*(6), 1587–1597. http://doi.org/10.1021/ct600180x

Ha, S. N., Giammona, A., Field, M., Brady, J. W. (1988). A revised potential-energy surface for molecular mechanics studies of carbohydrates. *Carbohydrate Research*, *180*(2), 207–221. http://doi.org/10.1016/0008-6215(88)80078-8

Hehre, W. J. (1976). Ab initio molecular orbital theory. *Accounts of Chemical Research*, *9*(11), 399–406. http://doi.org/10.1021/ar50107a003

Hehre, W. J., Stewart, R. F., Pople, J. A. (1969). Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *The Journal of Chemical Physics*, *51*(6), 2657–2664. http://doi.org/10.1063/1.1672392

He, L.-Z., Garamus, V., Niemeyer, B., Helmholz, H., Willumeit, R. (2000). Determination of micelle structure of octyl-β-glucoside in aqueous solution by small angel neutron scattering and geometric analysis. *Journal of Molecular Liquids*, *89*(1–3), 239–249. http://doi.org/10.1016/S0167-7322(00)90017-5

Hess, B., Bekker, H., Berendsen, H. J. C., Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, *18*(12), 1463–1472. http://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hockney, R. W., Goel, S. P., Eastwood, J. W. (1974). Quiet high-resolution computer models of a plasma. *Journal of Computational Physics*, *14*(2), 148–158. http://doi.org/10.1016/0021-9991(74)90010-2

Holler, F., Callis, J. B. (1989). Conformation of the hydrocarbon chains of sodium dodecyl sulfate molecules in micelles: an FTIR study. *The Journal of Physical Chemistry*, *93*(5), 2053–2058. http://doi.org/10.1021/j100342a068

Hong, X., Weng, Y.-X., Li, M. (2004). Determination of the Topological Shape of Integral Membrane Protein Light-Harvesting Complex LH2 from Photosynthetic Bacteria in the Detergent Solution by Small-Angle X-Ray Scattering. *Biophysical Journal*, *86*(2), 1082–1088.

Hoover, W. G. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, *31*(3), 1695–1697. http://doi.org/10.1103/PhysRevA.31.1695

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, *65*(3), 712–725. http://doi.org/10.1002/prot.21123

Humphrey, W., Dalke, A., Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38, 27–28.

Hunter, C. N., Hundle, B. S., Hearst, J. E., Lang, H. P., Gardiner, A. T., Takaichi, S., Cogdell, R. J. (1994). Introduction of new carotenoids into the bacterial photosynthetic apparatus by combining the carotenoid biosynthetic pathways of Erwinia herbicola and Rhodobacter sphaeroides. *Journal of Bacteriology*, *176*(12), 3692–3697.

Jang, S., Rivera, E., Montemayor, D. (2015). Molecular Level Design Principle behind Optimal Sizes of Photosynthetic LH2 Complex: Taming Disorder through Cooperation of Hydrogen Bonding and Quantum Delocalization. *The Journal of Physical Chemistry Letters*, *6*(6), 928–934. http://doi.org/10.1021/acs.jpclett.5b00078

Janosi, L., Keer, H., Kosztin, I., Ritz, T. (2006). Influence of subunit structure on the oligomerization state of light-harvesting complexes: A free energy calculation study. *Chemical Physics*, *323*(1), 117–128. http://doi.org/10.1016/j.chemphys.2005.08.038

Jimenez, R., Dikshit, S. N., Bradforth, S. E., Fleming, G. R. (1996). Electronic Excitation Transfer in the LH2 Complex of Rhodobacter sphaeroides. *The Journal of Physical Chemistry*, *100*(16), 6825–6834. http://doi.org/10.1021/jp953074j

Jirsakova, V., Reiss-Husson, F., Ranck, J.-L. (1996). Oligomeric state of the light-harvesting complexes B800–850 and B875 from purple bacterium Rubrivivax gelatinosus in detergent solution. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1277*(1–2), 150–160. http://doi.org/10.1016/S0005-2728(96)00087-4

Joosten, R. P., te Beek, T. A. H., Krieger, E., Hekkelman, M. L., Hooft, R. W. W., Schneider, R., Vriend, G. (2011). A series of PDB related databases for everyday needs. *Nucleic Acids Research*, *39*(Database issue), D411–D419. http://doi.org/10.1093/nar/gkq1105

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926–935. http://doi.org/10.1063/1.445869

Jorgensen, W. L., Jenson, C. (1998). Temperature dependence of TIP3P, SPC, and TIP4P water from NPT Monte Carlo simulations: Seeking temperatures of maximum density. *Journal of Computational Chemistry*, *19*(10), 1179–1186.

Jorgensen, W. L., Maxwell, D. S., Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, *118*(45), 11225–11236. http://doi.org/10.1021/ja9621760

Jorgensen, W. L., Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, *110*(6), 1657–1666. http://doi.org/10.1021/ja00214a001

Kabsch, W., Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577–2637. http://doi.org/10.1002/bip.360221211

Kandt, C., Ash, W. L., Peter Tieleman, D. (2007). Setting up and running molecular dynamics simulations of membrane proteins. *Methods*, *41*(4), 475–488. http://doi.org/10.1016/j.ymeth.2006.08.006

Karakas, E., Taveneau, C., Bressanelli, S., Marchi, M., Robert, B., Abel, S. (2016). Derivation of original RESP atomic partial charges for MD simulations of the LDAO surfactant with AMBER: Applications to a model of micelle and a fragment of the lipid kinase PI4KA. *Journal of Biomolecular Structure and Dynamics*, *0*(ja), 1–75. http://doi.org/10.1080/07391102.2015.1135822

Karrasch, S., Bullough, P. A., Ghosh, R. (1995). The 8.5 A projection map of the light-harvesting complex I from Rhodospirillum rubrum reveals a ring composed of 16 subunits. *The EMBO Journal*, *14*(4), 631–638.

Kirschner, K. N., Lins, R. D., Maass, A., Soares, T. A. (2012). A Glycam-Based Force Field for Simulations of Lipopolysaccharide Membranes: Parametrization and Validation. *Journal of Chemical Theory and Computation*, *8*(11), 4719–4731. http://doi.org/10.1021/ct300534j

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., Woods, R. J. (2008). GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *Journal of Computational Chemistry*, *29*(4), 622–655. http://doi.org/10.1002/jcc.20820

Konidala, P., He, L., Niemeyer, B. (2006). Molecular dynamics characterization of n-octyl-β-d-glucopyranoside micelle structure in aqueous solution. *Journal of Molecular Graphics and Modelling*, *25*(1), 77–86. http://doi.org/10.1016/j.jmgm.2005.11.008

Koolhaas, M. H., van der Zwan, G., van Mourik, F., van Grondelle, R. (1997). Spectroscopy and structure of bacteriochlorophyll dimers. I. Structural consequences of nonconservative circular dichroism spectra. *Biophysical Journal*, *72*(4), 1828–1841. http://doi.org/10.1016/S0006-3495(97)78829-5

Kozma, D., Simon, I., Tusnády, G. E. (2012). PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Research*, gks1169. http://doi.org/10.1093/nar/gks1169

Kucerka, N., Nagle, J. F., Sachs, J. N., Feller, S. E., Pencer, J., Jackson, A., Katsaras, J. (2008). Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data. *Biophysical Journal*, *95*(5), 2356–2367. http://doi.org/10.1529/biophysj.108.132662

Kučerka, N., Nieh, M.-P., Katsaras, J. (2011). Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1808*(11), 2761–2771. http://doi.org/10.1016/j.bbamem.2011.07.022

Kučerka, N., Tristram-Nagle, S., Nagle, J. F. (2006). Structure of Fully Hydrated Fluid Phase Lipid Bilayers with Monounsaturated Chains. *The Journal of Membrane Biology*, *208*(3), 193–202. http://doi.org/10.1007/s00232-005-7006-8

Kyte, J., Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*(1), 105–132. http://doi.org/10.1016/0022-2836(82)90515-0

Lange, O. F., Grubmüller, H. (2006). Generalized correlation for biomolecular dynamics. *Proteins: Structure, Function, and Bioinformatics*, *62*(4), 1053–1061. http://doi.org/10.1002/prot.20784

Lapouge, K., Näveke, A., Gall, A., Ivancich, A., Seguin, J., Scheer, H., Robert, B. (1999). Conformation of Bacteriochlorophyll Molecules in Photosynthetic Proteins from Purple Bacteria. *Biochemistry*, *38*(34), 11115–11121. http://doi.org/10.1021/bi990723z

Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins*, *78*(8), 1950–1958. http://doi.org/10.1002/prot.22711

Lipfert, J., Columbus, L., Chu, V. B., Lesley, S. A., Doniach, S. (2007). Size and shape of detergent micelles determined by small-angle X-ray scattering. *Journal of Physical Chemistry B*, *111*(43), 12427–12438. http://doi.org/10.1021/jp073016l

Liu, L.-N., Duquesne, K., Oesterhelt, F., Sturgis, J. N., Scheuring, S. (2011). Forces guiding assembly of light-harvesting complex 2 in native membranes. *Proceedings of the National Academy of Sciences*, *108*(23), 9455–9459. http://doi.org/10.1073/pnas.1004205108

Lomize, M. A., Lomize, A. L., Pogozheva, I. D., Mosberg, H. I. (2006). OPM: Orientations of Proteins in Membranes database. *Bioinformatics*, *22*(5), 623–625. http://doi.org/10.1093/bioinformatics/btk023

London, F. (1930). Zur Theorie und Systematik der Molekularkräfte. *Zeitschrift für Physik*, *63*(3-4), 245–279. http://doi.org/10.1007/BF01421741

Lopez, C. F., Nielsen, S. O., Klein, M. L., Moore, P. B. (2004). Hydrogen Bonding Structure and Dynamics of Water at the Dimyristoylphosphatidylcholine Lipid Bilayer Surface from a Molecular Dynamics Simulation. *The Journal of Physical Chemistry B*, *108*(21), 6603–6610. http://doi.org/10.1021/jp037618q

Lorber, B., Bishop, J. B., DeLucas, L. J. (1990). Purification of octyl β-d-glucopyranoside and re-estimation of its micellar size. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1023*(2), 254–265. http://doi.org/10.1016/0005-2736(90)90421-J

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Karplus, M. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B*, *102*(18), 3586–3616. http://doi.org/10.1021/jp973084f

Manet, S., Cuvier, A.-S., Valotteau, C., Fadda, G. C., Perez, J., Karakas, E., Baccile, N. (2015). Structure of Bolaamphiphile Sophorolipid Micelles Characterized with SAXS, SANS, and MD Simulations. *The Journal of Physical Chemistry B*, *119*(41), 13113–13133. http://doi.org/10.1021/acs.jpcb.5b05374

Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., de Vries, A. H. (2007). The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B*, *111*(27), 7812–7824. http://doi.org/10.1021/jp071097f

Martínez, L., Andrade, R., Birgin, E. G., Martínez, J. M. (2009). PACKMOL: a package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, *30*(13), 2157–2164. http://doi.org/10.1002/jcc.21224

McCammon, J. A., Gelin, B. R., Karplus, M. (1977). Dynamics of folded proteins. *Nature*, *267*(5612), 585–590. http://doi.org/10.1038/267585a0

McDermott, G., Prince, S. M., Freer, A. A., Hawthornthwaite-Lawless, A. M., Papiz, M. Z., Cogdell, R. J., Isaacs, N. W. (1995). Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. *Nature*, *374*(6522), 517–521. http://doi.org/10.1038/374517a0

McLuskey, K., Prince, S. M., Cogdell, R. J., Isaacs, N. W. (2001). The Crystallographic Structure of the B800-820 LH3 Light-Harvesting Complex from the Purple Bacteria Rhodopseudomonas Acidophila Strain 7050†. *Biochemistry*, *40*(30), 8783–8789. http://doi.org/10.1021/bi010309a

Mesa, C. L., Bonincontro, A., Sesta, B. (1993). Solution properties of octyl β-D glucoside. Part 1: Aggregate size. shape and hydration. *Colloid and Polymer Science*, *271*(12), 1165–1171. http://doi.org/10.1007/BF00657072

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, *21*(6), 1087–1092. http://doi.org/10.1063/1.1699114

Mulliken, R. S. (1955). Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *The Journal of Chemical Physics*, *23*(10), 1833–1840. http://doi.org/10.1063/1.1740588

Nagle, J. F., Tristram-Nagle, S. (2000). Structure of lipid bilayers. *Biochimica et Biophysica Acta*, *1469*(3), 159–195.

Nilsson, F., Söderman, O., Johansson, I. (1996). Physical−Chemical Properties of the n-Octyl β-d-Glucoside/Water System. A Phase Diagram, Self-Diffusion NMR, and SAXS Study. *Langmuir*, *12*(4), 902–908. http://doi.org/10.1021/la950602+

Nosé, S. (1984). A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, *52*(2), 255–268. http://doi.org/10.1080/00268978400101201

Nyrkova, I. A., Semenov, A. N. (2005). On the Theory of Micellization Kinetics. *Macromolecular Theory and Simulations*, *14*(9), 569–585. http://doi.org/10.1002/mats.200500010

Ode, H., Nakashima, M., Kitamura, S., Sugiura, W., Sato, H. (2012). Molecular dynamics simulation in virus research. *Virology*, *3*, 258. http://doi.org/10.3389/fmicb.2012.00258

Oijen, A. M. van, Ketelaars, M., Köhler, J., Aartsma, T. J., Schmidt, J. (1999). Unraveling the Electronic Structure of Individual Photosynthetic Pigment-Protein Complexes. *Science*, *285*(5426), 400–402. http://doi.org/10.1126/science.285.5426.400

Oliver, R. C., Lipfert, J., Fox, D. A., Lo, R. H., Doniach, S., Columbus, L. (2013). Dependence of micelle size and shape on detergent alkyl chain length and head group. *PloS One*, *8*(5), e62488. http://doi.org/10.1371/journal.pone.0062488

Olsen, J. D., Robert, B., Siebert, C. A., Bullough, P. A., Hunter, C. N. (2003). Role of the C-Terminal Extrinsic Region of the α Polypeptide of the Light-Harvesting 2 Complex of Rhodobacter sphaeroides: A Domain Swap Study†. *Biochemistry*, *42*(51), 15114–15123. http://doi.org/10.1021/bi035411h

Pan, J., Heberle, F. A., Tristram-Nagle, S., Szymanski, M., Koepfinger, M., Katsaras, J., Kučerka, N. (2012). Molecular structures of fluid phase phosphatidylglycerol bilayers as determined by small angle neutron and X-ray scattering. *Biochimica Et Biophysica Acta*, *1818*(9), 2135–2148. http://doi.org/10.1016/j.bbamem.2012.05.007

Papiz, M. Z., Prince, S. M., Howard, T., Cogdell, R. J., Isaacs, N. W. (2003). The Structure and Thermal Motion of the B800–850 LH2 Complex from Rps. acidophila at 2.0 Å Resolution and 100 K: New Structural Features and Functionally Relevant Motions. *Journal of Molecular Biology*, *326*(5), 1523–1538. http://doi.org/10.1016/S0022-2836(03)00024-X

Parrinello, M., Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, *52*(12), 7182–7190. http://doi.org/10.1063/1.328693

Pauli, W. (1925). Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Physik*, *31*(1), 765–783. http://doi.org/10.1007/BF02980631

Pérez, J., Koutsioubas, A. (2015). Memprot: a program to model the detergent corona around a membrane protein based on SEC-SAXS data. *Acta Crystallographica. Section D, Biological Crystallography*, *71*(Pt 1), 86–93. http://doi.org/10.1107/S1399004714016678

Pérez, J., Nishino, Y. (2012). Advances in X-ray scattering: from solution SAXS to achievements with coherent beams. *Current Opinion in Structural Biology*, *22*(5), 670–678. http://doi.org/10.1016/j.sbi.2012.07.014

Perilla, J. R., Goh, B. C., Cassidy, C. K., Liu, B., Bernardi, R. C., Rudack, T., Schulten, K. (2015). Molecular dynamics simulations of large macromolecular complexes. *Current Opinion in Structural Biology*, *31*, 64–74. http://doi.org/10.1016/j.sbi.2015.03.007

Peter, H. R., Johnson, G. B., Mason, K. A., Losos, J. B. (2011). *Biologie*. De Boeck Supérieur.

Poger, D., Mark, A. E. (2010). Turning the growth hormone receptor on: Evidence that hormone binding induces subunit rotation. *Proteins: Structure, Function, and Bioinformatics*, *78*(5), 1163–1174. http://doi.org/10.1002/prot.22636

Prince, S. M., Howard, T. D., Myles, D. A. A., Wilkinson, C., Papiz, M. Z., Freer, A. A., Isaacs, N. W. (2003). Detergent structure in crystals of the integral membrane light-harvesting complex LH2 from Rhodopseudomonas acidophila strain 10050. *Journal of Molecular Biology*, *326*(1), 307–315.

Prince, S. M., Papiz, M. Z., Freer, A. A., McDermott, G., Hawthornthwaite-Lawless, A. M., Cogdell, R. J., Isaacs, N. W. (1997). Apoprotein structure in the LH2 complex from Rhodopseudomonas acidophila strain 10050: modular assembly and protein pigment interactions. *Journal of Molecular Biology*, *268*(2), 412–423. http://doi.org/10.1006/jmbi.1997.0966

Pullerits, T., Sundström, V. (1996). Photosynthetic Light-Harvesting Pigment−Protein Complexes: Toward Understanding How and Why. *Accounts of Chemical Research*, *29*(8), 381–389. http://doi.org/10.1021/ar950110o

Rahman, A., Stillinger, F. H. (1971). Molecular Dynamics Study of Liquid Water. *The Journal of Chemical Physics*, *55*(7), 3336–3359. http://doi.org/10.1063/1.1676585

Ranck, J.-L., Ruiz, T., Péhau-Arnaudet, G., Arnoux, B., Reiss-Husson, F. (2001). Two-dimensional structure of the native light-harvesting complex LH2 from Rubrivivax gelatinosus and of a truncated form. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1506*(1), 67–78. http://doi.org/10.1016/S0005-2728(01)00185-2

Rancova, O., Sulskus, J., Abramavicius, D. (2012). Insight into the structure of photosynthetic LH2 aggregate from spectroscopy simulations. *The Journal of Physical Chemistry. B*, *116*(27), 7803–7814. http://doi.org/10.1021/jp302817p

R Development Core Team. (2008). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org

Reddy, T., Shorthouse, D., Parton, D. L., Jefferys, E., Fowler, P. W., Chavent, M., Sansom, M. S. P. (2015). Nothing to Sneeze At: A Dynamic and Integrative Computational Model of an Influenza A Virion. *Structure*, *23*(3), 584–597. http://doi.org/10.1016/j.str.2014.12.019

Robert, B., Andrianambinintsoa, S., Lutz, M. (1985). Structural Characterization of High 800 nm-Absorbing Light-Harvesting Complexes from Rhodospirillales from Their Resonance Raman Spectra. *Journal of Biochemistry*, *98*(2), 349–354.

Russell, N. J., Coleman, J. K., Howard, T. D., Johnston, E., Cogdell, R. J. (2002). Rhodopseudomonas acidophila strain 10050 contains photosynthetic LH2 antenna complexes that are not enriched with phosphatidylglycerol, and the phospholipids have a fatty acyl composition that is unusual for purple non-sulfur bacteria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1556*(2–3), 247–253. http://doi.org/10.1016/S0005-2728(02)00369-9

Rutkauskas, D., Novoderezhkin, V., Gall, A., Olsen, J., Cogdell, R. J., Hunter, C. N., van Grondelle, R. (2006). Spectral Trends in the Fluorescence of Single Bacterial Light-Harvesting Complexes: Experiments and Modified Redfield Simulations. *Biophysical Journal*, *90*(7), 2475–2485. http://doi.org/10.1529/biophysj.105.075903

Ryckaert, J.-P., Ciccotti, G., Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, *23*(3), 327–341. http://doi.org/10.1016/0021-9991(77)90098-5

Rycroft, C. H. (2009). VORO++: a three-dimensional voronoi cell library in C++. *Chaos (Woodbury, N.Y.)*, *19*(4), 041111. http://doi.org/10.1063/1.3215722

Sanders, S. A., Sammalkorpi, M., Panagiotopoulos, A. Z. (2012). Atomistic simulations of micellization of sodium hexyl, heptyl, octyl, and nonyl sulfates. *The Journal of Physical Chemistry. B*, *116*(8), 2430–2437. http://doi.org/10.1021/jp209207p

Sardet, C., Tardieu, A., Luzzati, V. (1976). Shape and size of bovine rhodopsin: A small-angle X-ray scattering study of a rhodopsin-detergent complex. *Journal of Molecular Biology*, *105*(3), 383–407. http://doi.org/10.1016/0022-2836(76)90100-5

Sauer, K., Cogdell, R. J., Prince, S. M., Freer, A., Isaacs, N. W., Scheer, H. (1996). Structure-Based Calculations of the Optical Spectra of the LH2 Bacteriochlorophyll-Protein Complex from Rhodopseudomonas acidophila. *Photochemistry and Photobiology*, *64*(3), 564–576. http://doi.org/10.1111/j.1751-1097.1996.tb03106.x

Scheuring, S., Busselez, J., Lévy, D. (2005). Structure of the Dimeric PufX-containing Core Complex of Rhodobacter blasticus by in Situ Atomic Force Microscopy. *Journal of Biological Chemistry*, *280*(2), 1426–1431. http://doi.org/10.1074/jbc.M411334200

Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., Gunsteren, W. F. van. (2011). Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal*, *40*(7), 843–856. http://doi.org/10.1007/s00249-011-0700-9

Scholes, G. D., Jordanides, X. J., Fleming, G. R. (2001). Adapting the Förster Theory of Energy Transfer for Modeling Dynamics in Aggregated Molecular Assemblies. *The Journal of Physical Chemistry B*, *105*(8), 1640–1651. http://doi.org/10.1021/jp003571m

Seddon, A. M., Curnow, P., Booth, P. J. (2004). Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1666*(1–2), 105–117. http://doi.org/10.1016/j.bbamem.2004.04.011

Silva, A. W. S. da, Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes*, *5*(1), 1–8. http://doi.org/10.1186/1756-0500-5-367

Skjevik, Å. A., Madej, B. D., Walker, R. C., Teigen, K. (2012). LIPID11: A Modular Framework for Lipid Simulations Using Amber. *The Journal of Physical Chemistry B*, *116*(36), 11124–11136. http://doi.org/10.1021/jp3059992

Skou, S., Gillilan, R. E., Ando, N. (2014). Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nature Protocols*, *9*(7), 1727–1739. http://doi.org/10.1038/nprot.2014.116

Smyth, C., Fassioli, F., Scholes, G.D., (2012). Measures and implications of electronic coherence in photosynthetic light-harvesting. *Philos. Trans. R. Soc. Lond. Math. Phys. Eng. Sci*. 370, 3728–3749. doi:10.1098/rsta.2011.0420

Sundström, V., Pullerits, T., van Grondelle, R. (1999). Photosynthetic Light-Harvesting: Reconciling Dynamics and Structure of Purple Bacterial LH2 Reveals Function of Photosynthetic Unit. *The Journal of Physical Chemistry B*, *103*(13), 2327–2346. http://doi.org/10.1021/jp983722+

Svergun, D., Barberato, C., Koch, M. H. J. (1995). CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography*, *28*(6), 768–773. http://doi.org/10.1107/S0021889895007047

Tanford, C. (1972). Micelle shape and size. *The Journal of Physical Chemistry*, *76*(21), 3020–3024. http://doi.org/10.1021/j100665a018

Taveneau, C., Blondeau, K., Bressanelli, S. (2015). Definition and expression in E. coli of large fragments from the human lipid kinase phosphatidylinositol 4-kinase type III alpha, and purification of a 1100-residue N-terminal module. *Protein Expression and Purification*, *114*, 121–127. http://doi.org/10.1016/j.pep.2015.06.018

Tessier, M. B., DeMarco, M. L., Yongye, A. B., Woods, R. J. (2008). Extension of the GLYCAM06 Biomolecular Force Field to Lipids, Lipid Bilayers and Glycolipids. *Molecular Simulation*, *34*(4), 349–363. http://doi.org/10.1080/08927020701710890

Tieleman, D. P., Berendsen, H. J. C. (1996). Molecular dynamics simulations of a fully hydrated dipalmitoylphosphatidylcholine bilayer with different macroscopic boundary conditions and parameters. *The Journal of Chemical Physics*, *105*(11), 4871–4880. http://doi.org/10.1063/1.472323

van Eerden, F. J., de Jong, D. H., de Vries, A. H., Wassenaar, T. A., Marrink, S. J. (2015). Characterization of thylakoid lipid membranes from cyanobacteria and higher plants by molecular dynamics simulations. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1848*(6), 1319–1330. http://doi.org/10.1016/j.bbamem.2015.02.025

van Oijen, A. M., Ketelaars, M., Köhler, J., Aartsma, T. J., Schmidt, J. (1998). Spectroscopy of Single Light-Harvesting Complexes from Purple Photosynthetic Bacteria at 1.2 K. *The Journal of Physical Chemistry B*, *102*(47), 9363–9366. http://doi.org/10.1021/jp9830629

van Oijen, A. M., Ketelaars, M., Köhler, J., Aartsma, T. J., Schmidt, J. (2000). Spectroscopy of Individual Light-Harvesting 2 Complexes of Rhodopseudomonas acidophila: Diagonal Disorder, Intercomplex Heterogeneity, Spectral Diffusion, and Energy Transfer in the B800 Band. *Biophysical Journal*, *78*(3), 1570–1577. http://doi.org/10.1016/S0006-3495(00)76709-9

Vanquelef, E., Simon, S., Marquant, G., Garcia, E., Klimerak, G., Delepine, J. C., Dupradeau, F.-Y. (2011). R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Research*, gkr288. http://doi.org/10.1093/nar/gkr288

Verlet, L. (1967). Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*, *159*(1), 98–103. http://doi.org/10.1103/PhysRev.159.98

Voronoi, G. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. *Journal für die reine und angewandte Mathematik*, *134*, 198–287.

Warshel, A., Levitt, M. (1976). Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, *103*(2), 227–249. http://doi.org/10.1016/0022-2836(76)90311-9

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology (Clifton, N.J.)*, *112*, 531–552.

Woods, R. J., Chappelle, R. (2000). Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates. *Journal of Molecular Structure: THEOCHEM*, *527*(1–3), 149–156. http://doi.org/10.1016/S0166-1280(00)00487-5

Wu, M., Strid, Å., Eriksson, L. A. (2013). Development of non-standard arginine residue parameters for use with the AMBER force fields. *Chemical Physics Letters*, *584*, 188–194. http://doi.org/10.1016/j.cplett.2013.08.071

Xu, D., Zhang, Y. (2012). Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins*, *80*(7), 1715–1735. http://doi.org/10.1002/prot.24065

Yoneda, Y., Noji, T., Katayama, T., Mizutani, N., Komori, D., Nango, M., Dewa, T. (2015). Extension of Light-Harvesting Ability of Photosynthetic Light-Harvesting Complex 2 (LH2) through Ultrafast Energy Transfer from Covalently Attached Artificial Chromophores. *Journal of the American Chemical Society*, *137*(40), 13121–13129. http://doi.org/10.1021/jacs.5b08508

Zhang, W. M., Meier, T., Chernyak, V., Mukamel, S. (1998). Exciton-migration and three-pulse femtosecond optical spectroscopies of photosynthetic antenna complexes. *The Journal of Chemical Physics*, *108*(18), 7763–7774. http://doi.org/10.1063/1.476212

Zuber, H. (1985). Structure and Function of Light-Harvesting Complexes and Their Polypeptides. *Photochemistry and Photobiology*, *42*(6), 821–844. http://doi.org/10.1111/j.1751-1097.1985.tb01655.x

Zuber, H., Brunisholz, R. (1991). Structure and function of antenna polypeptides and chlorophyll-protein complexes: Principles and variability. *In Chlorophylls (Scheer, H., Ed.)*, 627–703.

Titre : Simulations de dynamique moléculaire du complexe collecteur de lumière de type 2 d'une bactérie pourpre dans différents environnements micellaires et membranaire.

Mots clés : dynamique moléculaire, complexe LH2, POPC, LDAO, bOG, SAXS, désordre dynamique

Résumé : Les bactéries photosynthétiques pourpres comme *Rhodopseudomonas acidophila* strain 10050 disposent pour collecter la lumière d'un appareil photosynthétique constitué de complexes protéiques membranaires (LH1 et LH2) avec des pigments spécialisés. La lumière est principalement absorbée par ces pigments liés au complexe LH2 et l'énergie d'excitation résultante est ensuite transférée au complexe LH1 et, de là au centre réactionnel. Des expériences de spectroscopie de fluorescence sur des complexes LH2 uniques ont montré que l'intensité et la position de transition électronique du complexe pouvaient fortement fluctuer avec le temps décrivant un « désordre dynamique » en lien avec la fonction biologique du complexe LH2. Dans le but de mieux comprendre l'origine de ce désordre à l'échelle atomique, nous avons utilisé les approches de simulations de dynamique moléculaires classiques et quantiques. Nous avons d'abord modélisé le complexe LH2 dans différents environnements micellaires et membranaires constitués de détergents (dimethyldodecylamine-N-oxide, LDAO) et le β octyle glucoside, bOG)) et d'un modèle de membrane formé d'un phospholipide le 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, POPC). Des étapes préliminaires de paramétrisation des modèles pour les différents détergents et pigments ont d'abord été réalisées. Ensuite, pour tester et valider la construction des modèles, des expériences de SAXS avec les complexes LH2-LDAO et LH2-bOG ont été réalisées. Nous avons trouvé un bon accord avec nos modèles et les spectres de SAXS, notamment pour le complexe LH2-bOG. Les simulations moléculaires, nous ont permis d'étudié plus en détails la structure des complexes LH2-détergents et membrane ainsi que les interactions peptide-pigment, pigment-pigment en fonction de l'environnement. Nos résultats ont montré une influence significative de la dynamique du complexe et les interactions pigment-pigment et pigment-protéine en fonction de l'environnement micellaire ou membranaire. Enfin, afin de relier les variations des interactions entre les différents composants du complexe décrits dans nos simulations, aux variations d'absorption du complexe LH2 et au désordre dynamique, des calculs quantiques ont été réalisés à partir de structures atomiques représentatives de nos simulations. Les premiers résultats, encourageants, nous ont permis de modéliser les variations d'absorbance en relative accord avec les modèles théoriques.

Title: Molecular Dynamics simulations of the light-harvesting 2 complex of a purple bacterium in different micellar and membrane environments.

Keywords : molecular dynamics, LH2 complex, POPC, LDAO, bOG, membrane, SAXS, dynamic disorder

Abstract: Purple photosynthetic bacteria, such as *Rhodopseudomonas acidophila* strain 10050 have a synthetic apparatus which is composed by membrane protein complexes (LH1 and LH2) with specialized pigments to harvesting the light. This latter is mainly absorbed by photosynthetic pigments of the LH2 complex and the resulting excitation energy is transferred to LH1 complex, then to reaction center. Fluorescence spectroscopy experiments carried out with a single LH2 complexes have shown that the intensity and the position of the electronic transition of complexes can strongly fluctuate with the time describing a « dynamic disorder ». This disorder is strongly related to the biological function of the LH2 complex. The molecular description of the dynamic disorder of LH2 complex is great interest to get a precise comprehension about the ability to use of solar energy of these complexes. In order to obtain a precise picture of the dynamical disorder at the atomic level, we combined classical and quantum molecular dynamics simulations. We first model the LH2 complex in different biomimetic environments composed of detergents (dimethyldodecylamine-N-oxide, LDAO) and the β octyl glucoside, bOG)) or a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, POPC membrane. To validate the LH2-detergents complex models, small angle X-ray scattering (SAXS) experiments were performed. We find a good agreement of the SAXS results with simulation in particular with LH2-bOG models. In the second part of this thesis, we specifically studied with MD simulations the LH2 structure as well as the interactions between peptide and pigment and pigment-pigment in the different environments. Our results revealed a significant impact of the environment on the complex structure and dynamic of the interactions between pigments and pigment-protein depending on the environment. Finally to relate these variations with the molecular disorder we performed additional quantum calculation to model the absorbance variations changes by taken some representative snapshots of the simulations.