

Table des matières

INTRODUCTION	1
CHAPITRE 1 ETUDE BIBLIOGRAPHIQUE.....	7
I. La lipase lip2 de <i>Yarrowia lipolytica</i>	8
1.1. Définition des lipases	8
1.2. Réactions catalysées	8
1.3. Sélectivité de substrat	10
1.4. Mécanisme catalytique	12
1.5. Structure générale	13
1.5.1. Le repliement α/β	13
1.5.2. La triade catalytique	14
1.5.3. Le trou oxyanion	15
1.5.4. Le volet amphiphile ou « lid »	17
1.5.5. Topologie du site actif	20
1.6. Notre enzyme modèle : Lip2 de <i>Yarrowia lipolytica</i>	21
1.6.1. Découverte de Lip2	21
1.6.2. Structure de Lip2	22
1.6.3. Propriétés enzymatiques de Lip2	24
1.6.4. Intérêt Biotechnologique	25
1.6.5. Etude de la dynamique de Lip2	27
II. La protéine OmpA de <i>Klebsiella pneumoniae</i>	28
2.1. <i>Klebsiella pneumoniae</i>	28
2.2. L'enveloppe bactérienne	29
2.2.1. Les composants de l'enveloppe bactérienne	30
2.2.2. Le lipopolysaccharide	32
2.3. Les protéines de la membrane externe	33
2.3.1. La structure générale	33
2.3.2. L'insertion et l'interaction avec les lipides	35
2.3.3. Les fonctions des OMP	38
2.4. Biogenèse des protéines de la membrane externe	43
2.4.1. Le passage de la membrane interne	43
2.4.2. Les voies des chaperonnes	43
2.4.3. La machinerie BAM	45
2.5. La protéine « Outer membrane protein A »	46
III. Dynamique des protéines	50
3.1. Méthodes d'analyse de la dynamique fonctionnelle des protéines	50
3.2. RMN et dynamique	51
3.2.1. Le protéasome	54
3.2.2. La protéine CAP et la mesure de l'entropie conformationnelle	58
3.2.3. Le lysozyme du phage T4, et la caractérisation d'états « invisibles »	60
3.2.4. La rhodopsine d' <i>Anabaena</i>	62

IV. Références	68
CHAPITRE 2 EXPRESSION DE LA LIPASE LIP2 DE YARROWIA LIPOLYTICA EN VUE D'UNE ETUDE PAR RMN EN SOLUTION.....	87
I. Introduction	90
II. Material and methods	92
2.1. Chemicals	92
2.2. Strains and culture conditions.....	92
2.3. Vector construction	93
2.3.1. <i>Escherichia coli</i> vector.....	93
2.3.2. Cell-free vector	93
2.3.3. <i>Pichia pastoris</i> vector	94
2.4. Protein expression	94
2.4.1. Expression in <i>Escherichia coli</i>	94
2.4.2. Cell-free expression	95
2.4.3. Expression in <i>Pichia pastoris</i>	95
2.4.4. Expression in <i>Yarrowia lipolytica</i>	96
2.5. Characterization of the expressed lipase.....	97
2.5.1. Lipase activity.....	97
2.5.2. Protein quantification.....	97
2.5.3. SDS-PAGE analysis.....	97
2.5.4. Western Blot.....	97
2.6. Purification	98
2.7. NMR spectroscopy.....	98
III. Results.....	98
3.1. Expression in <i>Escherichia coli</i>	98
3.2. Cell free expression	99
3.3. Expression in <i>Pichia pastoris</i>	99
3.4. Expression in <i>Yarrowia lipolytica</i>	100
IV. Discussion	101
V. Conclusion	103
VI. Acknowledgments.....	104
VII. Supplementary information.....	105
VIII. References.....	110
CHAPITRE 3 STRUCTURE CRISTALLOGRAPHIQUE ET DYNAMIQUE PAR RMN EN SOLUTION DU DOMAINE C-TERMINAL DE OMPA DE KLEBSIELLA PNEUMONIAE.....	113
I. Introduction	116
II. Material and Methods	117
2.1. Cloning, expression and protein purification.....	117
2.2. Crystallization.....	118
2.3. Structure determination.....	118

2.4. SEC-MALLS.....	119
2.5. NMR experiments.....	119
III. Results and Discussion	120
3.1. Tridimensional structure of C-terminal KpOmpA.	120
3.2. Binding site and sequence conservation.	124
3.3. Putative role of conserved D325 – R326 – R327.	125
3.4. Dimerization interface.	128
3.5. Dynamics in solution.	131
IV. Conclusion:.....	135
V. Supplementary information.....	136
5.1. Supplementary Figures.....	136
5.2. Supplementary Tables.....	143
VI. References	144
 CHAPITRE 4 ETUDE DE LA DYNAMIQUE DU DOMAINE N-TERMINAL DE OMPA DE KLEBSIELLA PNEUMONIAE PAR RMN DU SOLIDE ET PROTEOLYSE MENAGEE.....	 149
I. Introduction.....	152
II. Results and Discussion	153
2.1. Ultrafast solid state NMR:.....	153
2.2. Proteolysis experiments:.....	158
III. Supplementary information.....	163
3.1. Construction of the KpOmpA recombinant proteins	163
3.2. Protein expression and purification.....	163
3.3. Proteoliposomes preparation.....	164
3.4. Proteolysis experiments.....	164
3.5. Mass spectrometry.....	165
3.6. NMR spectroscopy.....	165
3.6.1. NMR experiments for KpOmpA assignment	166
3.6.2. Relaxation and REDOR NMR experiments	167
3.7. Supplementary Figures.....	169
3.8. Supplementary Tables.....	172
IV. References	179
 CONCLUSIONS ET PERSPECTIVES	 183
I. Expression de la lipase Lip2 de <i>Yarrowia lipolytica</i> pour la RMN.	184
II. Interaction de C-KpOmpA avec le DAP.	188
III. Expression de F-KpOmpA dans la membrane externe d' <i>E. coli</i>	193
IV. Références	197

Abréviations

AFM	Atomic Force Microscopy
AcN	Acetonitrile
ASR	<i>Anabaena</i> Sensory Rhodopsin
ASRT	<i>Anabaena</i> Sensory Rhodopsin Transducer
BAM	β -Barrel Assembly Machinery
BCL	<i>Burkholderia cepacia</i> Lipase
BMRB	Biological Magnetic Resonance data Bank
cAMP	Cyclic Adenosine Monophosphate
cGMP	Cyclic Guanosine Monophosphate
CHCA	α -Cyano-4-Hydroxycinnamic Acid
C-KpOmpA	C-terminal domain of KpOmpA
CMC	Critical Micelle Concentration
CSM	Chemical Shift Mapping
CP	Cross Polarization
CPMAS	Cross Polarization Magic Angle Spinning
CPMG	Carr – Purcell – Meiboom – Gill
CSI	Chemical Shift Index
DARR	Dipolar-Assisted Rotational Resonance
DAP	Diaminopimelic acid
DEER	Double Electron – Electron Resonance
DHA	cis-4,7,10,13,16,19-Docosahexanoic Acid
DHB	2,5-Dihydroxybenzoic Acid
DHPC	Di-Hexanoyl-Phosphatidyl-Choline (1,2-Dicaproyl-sn-glycero-3-Phosphocholine)
DNP	Dynamic Nuclear Polarization
DPC	Dodecyl-Phospho-Choline
EC	Enzyme Classification
EcOmpA	Outer Membrane Protein A from <i>Escherichia coli</i>
EPA	cis-5,8,11,14,17-Eicosapentanoic Acid
F-KpOmpA	Full-length polypeptide chain of KpOmpA
Hep	L-glycero-D-mannoheptose
HSQC	Heteronuclear Single Quantum Correlation
HPL	Human Pancreatic Lipase
IPTG	Isopropyl β -D-1-Thiogalactopyranoside
ITC	Isothermal Titration Calorimetry

Kdo	3-deoxy- α -D-mannooctulosonic acid
KpOmpA	Outer Membrane Protein A from <i>Klebsiella pneumoniae</i>
LPR	Lipid to Protein Ratio
LPS	Lipopolysaccharide
Lys-C	Lysine-C endoprotease
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionization - Time-Of-Flight mass spectrometry
MW	Molecular Weight
NAG	N-Acetylglucosamine
NAM	N-Acetylmuramic Acid
NaTDC	Sodium Taurodeoxycholate
N-KpOmpA	N-terminal domain of KpOmpA
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
OD	Optical Density
OM	Outer Membrane
OMP	Outer Membrane Protein
OmpA	Outer Membrane Protein A
PAL	Peptidoglycan Associated Lipoprotein
PLE	Polar Lipid Extract
pNPB	para-Nitrophenolbutyrate
ppm	parts per million
PUFA	Poly Unsaturated Fatty Acid
PDB	Protein Data Bank
PDSD	Proton-Driven Spin Diffusion
POTRA	Polypeptide Translocation Associated Domain
PRE	Paramagnetic Relaxation Enhancement
RMN	Résonance Magnétique Nucléaire
RPE	Résonance Paramagnétique Electronique
R ₁	Longitudinal (spin-lattice) relaxation rate
R _{1p}	R ₁ of spin-locked magnetization in rotating frame
R ₂	Transverse (spin-spin) relaxation rate
RX	Diffraction aux Rayons X
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate polyacrylamid Gel Electrophoresis
SMFS	Single Molecule Force Spectroscopy
SPA	Sinapinic Acid (3,5-dimethoxy-4-hydroxycinnamic acid)

STD	Saturation Transfert Difference
T_1	Longitudinal (spin-lattice) relaxation time
$T_{1\rho}$	T_1 of spin-locked magnetization in rotating frame
T_2	Transverse (spin-spin) relaxation time
TALOS	Torsion Angle Likelihood Obtained from Shift and sequence similarity
τ_c	Rotational Correlation Time
TOCSY	Total Correlation Spectroscopy
TRACT	TROSY for rotational correlation time
TROSY	Transverse Relaxation Optimized Spectroscopy
TFA	2,2,2-Trifluoroacetic Acid
XFEL	X-ray Free Electron Laser
Zw3-14	Zwittergent 3-14 (n-tetradecyl-N,N-dimethyl-3-ammino-1-propane sulfonate)

Introduction

Les protéines sont des polymères complexes qui interviennent dans l'ensemble des processus cellulaires grâce à la diversité de leurs fonctions, allant de la structure cellulaire à la catalyse en passant par la signalisation. L'explication de cette diversité de fonction des protéines a été une des motivations majeures de la biologie structurale. La cristallographie a fourni les premières structures atomiques de protéine (Kendrew *et al.* 1960), suivie par la résonance magnétique nucléaire (RMN) (Williamson *et al.* 1985). Ces structures ont alors permis l'émergence du concept de relation structure fonction. Depuis, le nombre de structures disponibles n'a cessé de croître suscitant la création de la Protein Data Bank dès 1971 (Berman 2008). Les techniques classiques de cristallographie et de résonance magnétique nucléaire utilisées pour résoudre les structures des protéines fournissent une structure moyenne avec une grande précision contribuant ainsi à une vision statique des protéines. Pourtant, la structure seule ne parvient pas à expliquer toutes les propriétés des protéines et la description de l'aspect dynamique des protéines a rapidement pris une importance croissante (McCammon *et al.* 1977, Henzler-Wildman *et al.* 2007, van den Bedem *et al.* 2015). Le travail effectué lors de cette thèse s'inscrit dans ce cadre où une description à la fois de la structure et de la dynamique est nécessaire à une compréhension des mécanismes moléculaires dont découlent les propriétés et par conséquent les fonctions des protéines. Cette importance de la dynamique est particulièrement vraie pour les enzymes qui catalysent les réactions chimiques. La remarquable augmentation de la vitesse des réactions catalysées résulte d'une diminution de la barrière d'activation par le catalyseur. Cependant, de nombreuses étapes précédant ou suivant l'étape chimique interviennent dans la catalyse enzymatique (Palmer 2015). L'échange entre différents états conformationnels peut ainsi avoir une influence sur la régulation allostérique, l'inhibition, la reconnaissance du substrat, la présence de réactions secondaires ou encore la libération du produit.

Le chapitre 1 de la thèse permettra au lecteur de se familiariser avec les protéines étudiées au cours de ce travail, la lipase Lip2 de *Yarrowia lipolytica*, puis la protéine OmpA de *Klebsiella pneumoniae*. Des exemples récents d'étude de la dynamique des protéines par RMN sont également exposés.

La structure cristallographique de la lipase Lip2 de *Yarrowia lipolytica* a été résolue en 2010 par les équipes de Catalyse et Ingénierie Moléculaire Enzymatiques (CIMES) du Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP) et de Biophysique Structurale de l'Institut de Pharmacologie et de Biologie Structurale (IPBS) (Bordes *et al.* 2010). La structure correspond à la conformation fermée de la lipase où le site actif est recouvert par un volet amphiphile, couramment dénommé lid, et est de fait inaccessible au substrat (PDB ID : 3O0D). Malgré plusieurs essais, il n'a pas été possible d'obtenir de structure de la conformation ouverte de Lip2. La RMN est une

technique qui a été appliquée avec succès à l'étude de la dynamique des protéines. Nous nous sommes donc tournés vers cette approche pour étudier les changements conformationnels de Lip2 en solution et l'influence de l'environnement sur ceux-ci.

Afin de mener une étude par RMN sur Lip2, une première étape est l'expression marquée aux isotopes ^{15}N , ^{13}C de cette protéine de 301 acides aminés. Dans cet objectif, nous avons exploré la production de Lip2 dans plusieurs systèmes d'expression et avec plusieurs stratégies de marquage. Nous avons réussi à exprimer Lip2 uniformément marquée ^{15}N chez la levure *Yarrowia lipolytica* et à enregistrer le premier spectre HSQC ^{15}N - ^1H de cette protéine. Ce travail fait l'objet du chapitre 2 de la thèse et a donné lieu à une publication dans la revue « Protein Expression and Purification » publiée en 2014 (Volume 101, Pages 14-20). Toutefois la nécessité d'avoir recours à un bioréacteur pour contrôler l'oxygénation et le pH lors de l'expression a considérablement alourdi la production d'échantillons. Par conséquent, une étude des conditions de culture est nécessaire pour notamment ajuster les quantités de glucose pour envisager un marquage ^{13}C telle que cela a été réalisé chez *Kluyveromyces lactis* (Sugiki *et al.* 2008).

Nous nous sommes donc tournés vers la protéine OmpA de *Klebsiella pneumoniae* (KpOmpA). Cette protéine a été largement étudiée dans l'équipe RMN Biologique Intégrative de l'IPBS. OmpA est une protéine de la paroi des bactéries gram-négatives. Le domaine N-terminal de KpOmpA est un tonneau β transmembranaire de la membrane externe. La structure de ce domaine solubilisé en micelles détergent a été résolue en 2009 par RMN en solution (PDB ID : 2K0L) (Renault *et al.* 2009). Le domaine C-terminal est périplasmique et des structures de domaines homologues sont disponibles. Nous avons pu résoudre par cristallographie les structures des conformations Apo et en complexe avec un ligand, le citrate, du domaine C-terminal de KpOmpA. Nous avons complété cette caractérisation structurale par une caractérisation de la dynamique de ce domaine par RMN en solution. Les résultats de cette étude sont décrits dans le chapitre 3 de la thèse.

Klebsiella pneumoniae est responsable d'infections nosocomiales respiratoires et urinaires chez les patients affaiblis ou immunodéprimés. KpOmpA intervient dans la modulation de l'immunité lors de l'interaction avec l'hôte (March *et al.* 2011). Les boucles du tonneau β de KpOmpA sont exposées à la surface extracellulaire de la membrane externe. La dynamique de ces boucles a été étudiée en solution par RMN liquide et en bicouche lipidique par RMN du solide (Iordanov *et al.* 2012). Ces études ont montré la présence d'un gradient de mobilité allant du centre du tonneau β aux extrémités des boucles extracellulaires. L'étude de la mobilité du domaine N-terminal de KpOmpA a été complétée par RMN du solide avec rotation à l'angle magique à très haute vitesse (60 kHz), qui permet des stratégies fondées sur la détection des protons. Les spectres des corrélations ^{15}N - ^1H de KpOmpA triplement marquée ^{15}N , ^{13}C , ^2H en protéoliposome ont pu être enregistrés et

l'attribution de la majeure partie du tonneau β pu être été réalisée. Toutefois, les signaux des boucles et des jonctions avec le tonneau β n'ont pas été observés que ce soit dans des expériences de transfert de polarisation basé sur les couplages dipolaires ou les couplages scalaires, révélant la présence de mouvements à des échelles de temps intermédiaires (μs – s). La dynamique particulière des boucles a été étudiée également par des expériences de protéolyse à la trypsine. Sur les 7 sites de clivage présents au niveau des boucles extracellulaires, seul celui présent dans la boucle L3 est accessible à la protéolyse témoignant de la dynamique particulière qui anime ces régions de KpOmpA. Le travail effectué sur le tonneau β de KpOmpA est détaillé dans le chapitre 4 de la thèse.

Enfin, le dernier chapitre inclue des résultats complémentaires et clos le manuscrit en discutant des perspectives qui peuvent être envisagées.

Berman H M. *The Protein Data Bank: a historical perspective*. Acta Crystallogr A, **2008**, 64(Pt 1): 88-95.

Bordes F, Barbe S, Escalier P, Mourey L, Andre I, Marty A and Tranier S. *Exploring the conformational states and rearrangements of Yarrowia lipolytica Lipase*. Biophys J, **2010**, 99(7): 2225-2234.

Henzler-Wildman K and Kern D. *Dynamic personalities of proteins*. Nature, **2007**, 450(7172): 964-972.

Iordanov I, Renault M, Reat V, Bosshart P D, Engel A, Saurel O and Milon A. *Dynamics of Klebsiella pneumoniae OmpA transmembrane domain: the four extracellular loops display restricted motion behavior in micelles and in lipid bilayers*. Biochim Biophys Acta, **2012**, 1818(9): 2344-2353.

Kendrew J C, Dickerson R E, Strandberg B E, Hart R G, Davies D R, Phillips D C and Shore V C. *Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å Resolution*. Nature, **1960**, 185(4711): 422-427.

March C, Moranta D, Regueiro V, Llobet E, Tomas A, Garmendia J and Bengoechea J A. *Klebsiella pneumoniae Outer Membrane Protein A Is Required to Prevent the Activation of Airway Epithelial Cells*. Journal of Biological Chemistry, **2011**, 286(12): 9956-9967.

McCammon J A, Gelin B R and Karplus M. *Dynamics of folded proteins*. Nature, **1977**, 267(5612): 585-590.

Palmer A G, 3rd. *Enzyme dynamics from NMR spectroscopy*. Acc Chem Res, **2015**, 48(2): 457-465.

Renault M, Saurel O, Czaplicki J, Demange P, Gervais V, Lohr F, Reat V, Piotto M and Milon A. *Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications*. J Mol Biol, **2009**, 385(1): 117-130.

Sugiki T, Shimada I and Takahashi H. *Stable isotope labeling of protein by Kluyveromyces lactis for NMR study*. Journal of biomolecular NMR, **2008**, 42(3): 159-162.

van den Bedem H and Fraser J S. *Integrative, dynamic structural biology at atomic resolution--it's about time*. Nat Methods, **2015**, 12(4): 307-318.

Williamson M P, Havel T F and Wuthrich K. *Solution Conformation of Proteinase Inhibitor-Ia from Bull Seminal Plasma by H-1 Nuclear Magnetic-Resonance and Distance Geometry*. Journal of Molecular Biology, **1985**, 182(2): 295-315.

Chapitre 1

Etude bibliographique.

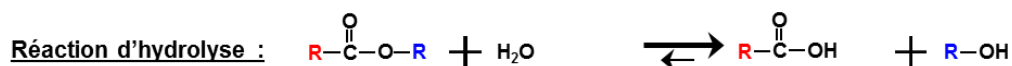
I. La lipase lip2 de *Yarrowia lipolytica*.

1.1. Définition des lipases.

Les lipases sont des enzymes de la famille des sérine-hydrolases définies comme des triacylglycerol-acylhydrolases (EC 3.1.1.3). Ces enzymes ubiquitaires sont largement distribuées chez les organismes procaryotes et eucaryotes. Elles interviennent dans de nombreux processus physiologiques cruciaux en particulier au niveau du métabolisme énergétique où elles catalysent la libération d'acides gras par saponification des triglycérides (Berg *et al.* 2002). Elles sont également très utilisées dans de nombreux procédés industriels en raison de leurs intéressantes propriétés de biocatalyseurs (Houde *et al.* 2004).

1.2. Réactions catalysées.

Les lipases catalysent naturellement l'hydrolyse de la liaison ester entre le glycérol et les acides gras des mono-, di- ou tri- glycérides. Toutefois, les lipases sont également capables d'hydrolyser de nombreux autres esters. De plus, placées dans des conditions thermodynamiques favorables, les lipases sont capables de catalyser des réactions de synthèse telles que des estérifications, des acidolyses, des alcoolyses ou des transestérifications (Figure 1). Leur capacité à accepter une large gamme de substrats leur permet de catalyser également des réactions d'amidation ou de thioestérification (Figure 1). Enfin, les lipases sont capables de présenter également des activités catalytiques annexes telles que des activités phospholipases, cholestérol estérases ou encore amidases (Svendsen 2000).



Réaction de synthèse :

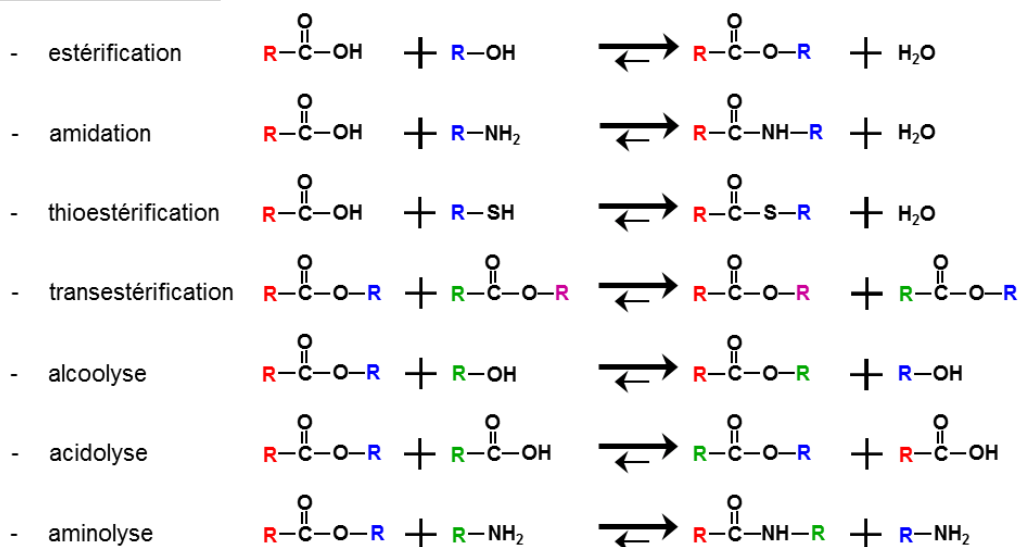


Figure 1. Exemples de réactions catalysées par les lipases

Une particularité des réactions catalysées par les lipases est qu'elles se déroulent en milieu biphasique, à l'interface entre les substrats hydrophobes, ou un solvant organique et un milieu aqueux (Reis *et al.* 2009). Ainsi, une propriété importante qui différencie les lipases des estérases, est le phénomène d'activation interfaciale (Figure 2) (Verger 1976). Ce phénomène se traduit par une activité accrue quand un substrat passe au-dessus de sa limite de solubilité et forme une émulsion, induisant la formation d'une interface lipide-eau. Néanmoins, les bases moléculaires à l'origine de ce phénomène demeurent encore mal comprises et certaines lipases telle que la lipase de *Pseudomonas aeruginosa* ne présentent pas ce phénomène d'activation interfaciale (Jaeger *et al.* 1993).

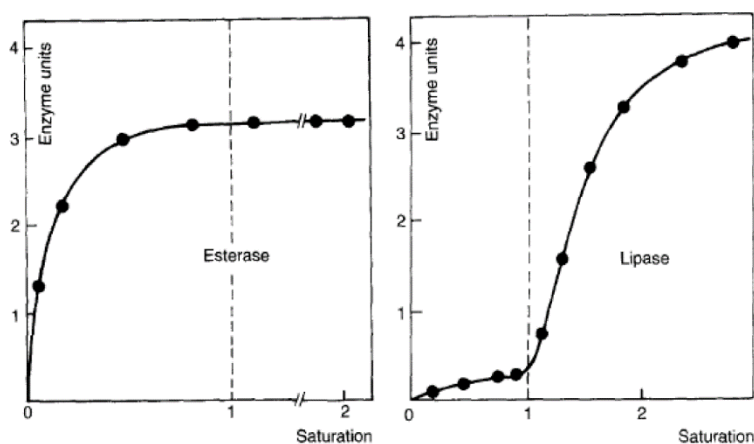


Figure 2. Activité enzymatique en fonction de la concentration d'un ester, les lignes verticales représentant la limite de solubilité de l'ester dans le solvant (Verger 1976).

1.3. Sélectivité de substrat.

Même si les lipases acceptent une large gamme de substrats, leur intérêt réside essentiellement dans leur sélectivité. Ainsi, on distingue différents types de sélectivité des lipases vis-à-vis de leurs substrats :

- La sélectivité vis-à-vis des acides gras, en particulier la longueur de la chaîne carbonée et la présence ainsi que la position de substitutions ou d'insaturations, appelée *typosélectivité*. Les lipases catalysent préférentiellement l'hydrolyse d'esters à longues chaînes carbonées ($C > 10$, tels que l'acide oléique C18 :1) alors que les estérases ont une activité plus importante sur les esters à chaînes courtes ($C < 10$, tels que la tributyrine C4). La présence d'insaturations sur la chaîne carbonée ainsi que leur position va influencer les conformations que le substrat pourra adopter et donc la forme présentée à l'enzyme. Cette sélectivité pour la longueur des groupements acyle ainsi que leur forme dépend fortement de la topologie et de la plasticité du site actif de la lipase. La typosélectivité inclut également la préférence pour les mono-, di- ou triglycérides.
- La sélectivité vis-à-vis de la position hydrolysée préférentiellement dans les triacylglycérols ou encore appelée *régiosélectivité*. En effet, les lipases peuvent présenter une spécificité différente en fonction de la position de la liaison ester sur le squelette du glycérol. On distingue les positions sn1 et sn3 correspondant à des fonctions alcool primaire de la position sn2 correspondant à un alcool secondaire. Si la plupart des lipases microbiennes ont une spécificité sn1 – sn3, seules quelques-unes ont une spécificité sn2 telle que la lipase de *Staphylococcus aureus* (Horchani *et al.* 2010). Ici encore, la taille et l'architecture du site actif accueillant le substrat jouent un rôle déterminant dans la régiosélectivité. La structure de la lipase de *Burkholderia cepacia* a été résolue par cristallographie en présence d'un analogue de triglycéride permettant ainsi d'identifier trois poches pouvant accommoder les trois acides gras en position sn1, sn2 et sn3 (Figure 3) (PDB ID : 5LIP) (Lang *et al.* 1998, Jaeger *et al.* 1999).

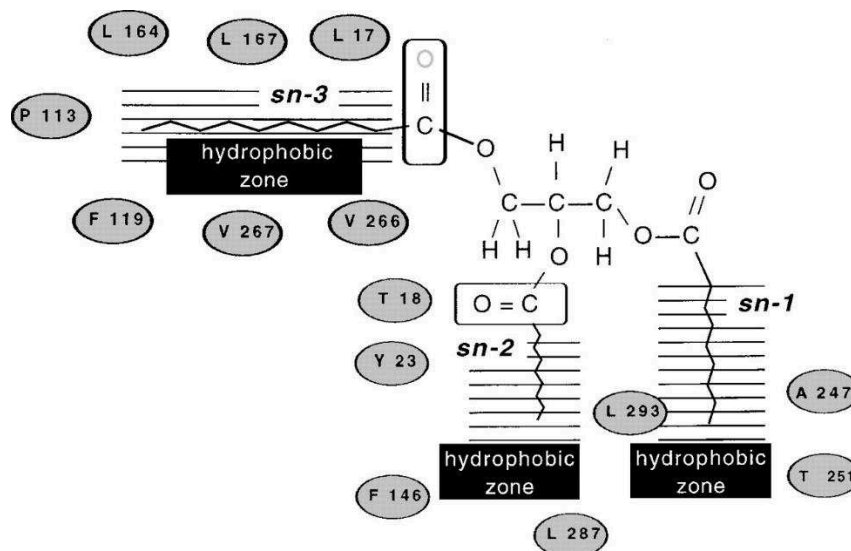


Figure 3. Représentation du site actif de la lipase de *Burkholderia cepacia* avec un triglycéride comme substrat. Les trois poches accommodant les chaînes carbonées sn1, sn2 et sn3 sont indiquées ainsi que les résidus bordant ces poches (Jaeger *et al.* 1999).

- Enfin, la sélectivité vis-à-vis d'un couple d'énantiomères, appelée *énantiosélectivité*. Les énantiomères sont des isomères images l'un de l'autre dans un miroir mais non superposables. Cette propriété particulière appelée chiralité est souvent due à la présence d'un carbone asymétrique dans la molécule, c'est-à-dire un carbone tétraédrique ayant quatre substituants différents. Si les énantiomères présentent des propriétés physico-chimiques proches, voire identiques, telles que la solubilité, le point de fusion ou le point d'ébullition, il n'en est pas de même pour les propriétés biologiques. Ainsi, si un énantiomère peut présenter une activité biologique recherchée, le second énantiomère peut être dépourvu de cette activité biologique, voire même avoir d'autres activités biologiques complètement différentes. Ce phénomène est à la base de nombreux effets indésirables des molécules thérapeutiques (Soykova Pachnerova 1963). Par conséquent, la résolution de mélange racémique présente un grand intérêt pour l'industrie pharmaceutique. Toutefois, l'explication de la spécificité d'une lipase pour un énantiomère reste difficile à mettre en évidence et plusieurs approches complémentaires telles que la cristallographie (Colton *et al.* 2011), le docking moléculaire (Berglund *et al.* 1999), la dynamique moléculaire (Guieysse *et al.* 2003), la modélisation d'intermédiaires réactionnels (Norin *et al.* 1994) ou l'étude de la géométrie du site actif et des trajectoires du substrat (Cortes *et al.* 2005) sont souvent nécessaires pour élucider les facteurs moléculaires jouant un rôle clé.

1.4. Mécanisme catalytique.

Le mécanisme catalytique des lipases fait intervenir une triade catalytique, constituée de trois acides aminés, un aspartate, une histidine et une sérine ainsi qu'un trou oxyanion constitué par les groupes amides NH de la chaîne principale de deux résidus (Figure 4) (Cygler *et al.* 1994). Le mécanisme catalytique peut être divisé en deux étapes successives. La première étape d'acylation débute par un transfert de proton entre les trois résidus de la triade catalytique de la sérine vers l'aspartate par l'intermédiaire de l'histidine. L'oxygène activé de la chaîne latérale de la sérine va alors attaquer le groupe carbonyle de l'ester pour former un intermédiaire tétraédrique appelé oxyanion. Cet intermédiaire est stabilisé par un réseau de liaisons hydrogènes incluant les résidus du trou oxyanion qui équilibrent les charges négatives de l'oxygène de l'intermédiaire tétraédrique. Un réarrangement électronique au niveau de l'intermédiaire réactionnel va libérer le groupement alcoolique de l'ester tandis que le carbone du groupement carboxylique va rester lié de façon covalente à l'oxygène de la sérine formant ainsi l'acyl-enzyme.

La seconde étape de déacylation débute par une attaque nucléophile sur le groupement carbonyle pour former un second intermédiaire réactionnel stabilisé par un réseau de liaisons hydrogènes de la manière identique à la première étape. Le réarrangement électronique de l'intermédiaire réactionnel permet la régénération de la fonction hydroxyle de la sérine catalytique et la libération de la partie acide de l'ester. En fonction de la nature du groupement effectuant l'attaque nucléophile lors de la seconde étape, la réaction peut être une hydrolyse s'il s'agit d'une molécule d'eau ou d'une alcoolise dans le cas d'un alcool.

Dans des conditions thermodynamiques favorables tel qu'un milieu présentant une faible activité de l'eau, le mécanisme inverse peut être utilisé pour synthétiser un ester à partir d'un alcool et d'un acide en libérant une molécule d'eau. Ce mécanisme catalytique ne permet toutefois pas de rendre compte de toutes les réactions catalysées par les lipases telles que typiquement les réactions d'acidolyse.

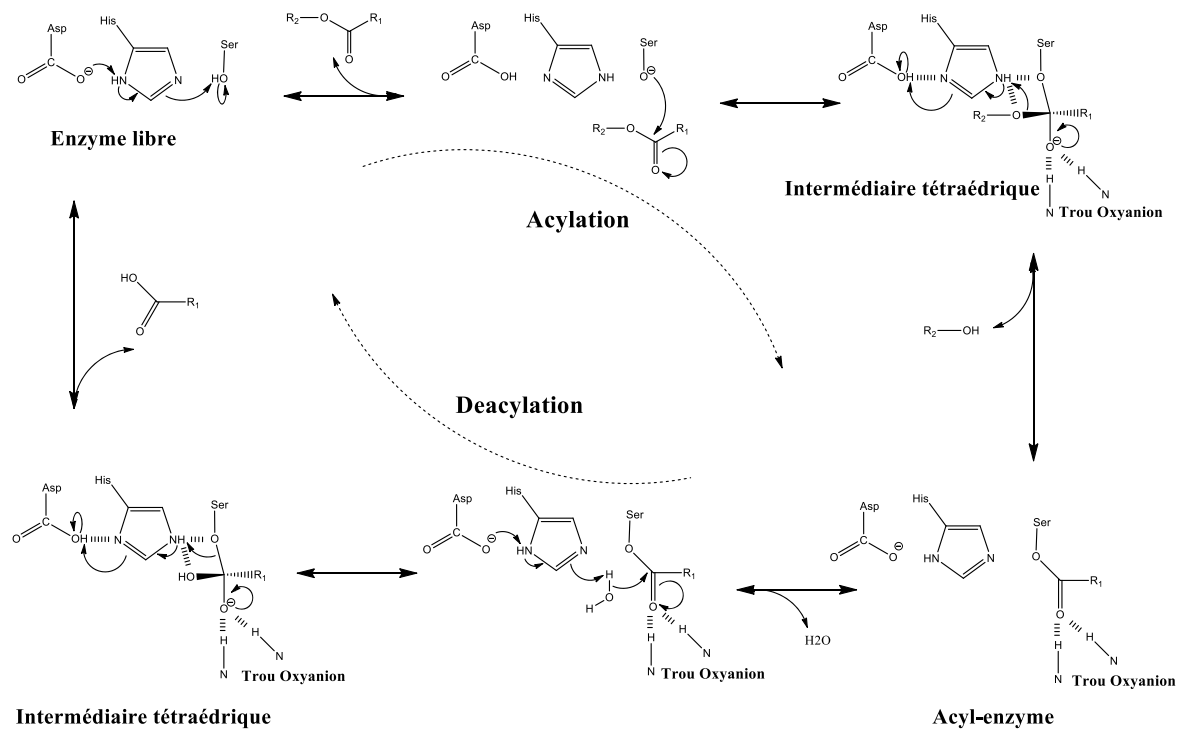


Figure 4. Mécanisme catalytique des lipases.

1.5. Structure générale.

De nombreuses lipases ont été caractérisées structuralement principalement par cristallographie. En comparant ces différentes structures cristallisées avec ou sans ligand, plusieurs éléments structuraux communs ont pu être mis en évidence.

1.5.1. Le repliement α/β

Les éléments de structures secondaires des lipases s'agencent pour adopter un motif de structure tertiaire appelé le repliement α/β . Ce motif n'est pas propre aux lipases et est partagé par de nombreuses autres hydrolases telles que des sérine-protéases, des cystéine-protéases ou les subtilisines (Ollis *et al.* 1992). Ce motif se compose d'un feuillet β central mixte, formé par 7 brins β parallèles et 1 brin β antiparallèle, et de 6 hélices α positionnées de part et d'autre du feuillet β central connectant les brins β 1 à β 8 (Figure 5) (Jaeger *et al.* 1999).

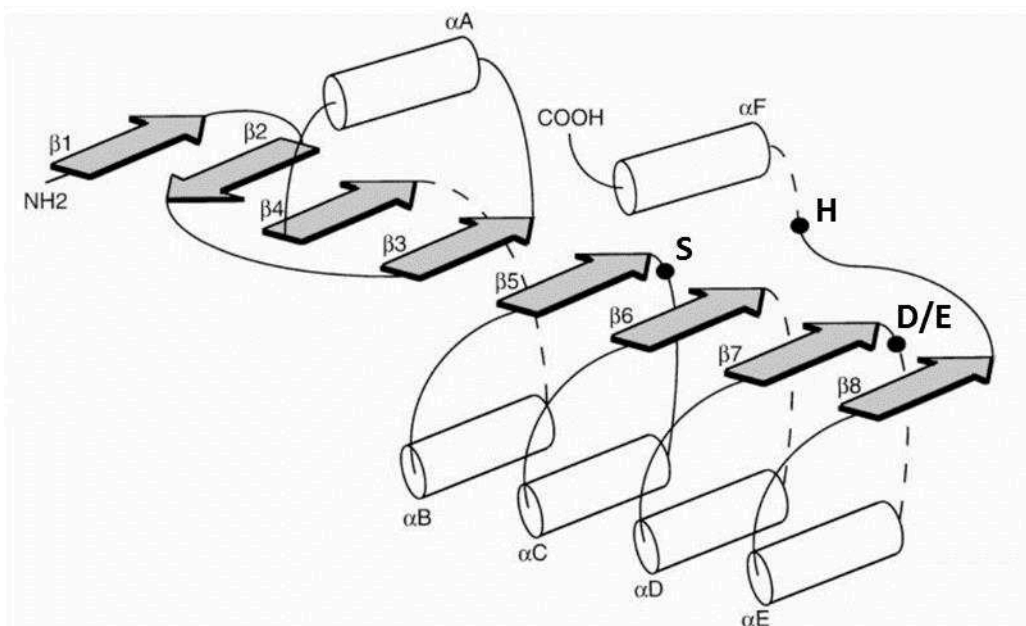


Figure 5. Repliement α/β . Représentation canonique du repliement α/β . Les brins β du feuillet central sont représentés par des flèches tandis que les hélices α sont représentées par des cylindres. Le positionnement des résidus de la triade catalytique est représenté par trois points noirs (Jaeger *et al.* 1999).

De nombreuses variations de ce repliement α/β sont retrouvées parmi les structures des différentes familles de lipases, principalement concernant le nombre d'hélices α et de brins β , la longueur des boucles ou encore l'architecture du site actif liant le substrat (Ollis *et al.* 1992, Schrag *et al.* 1997a, Pleiss *et al.* 1998).

1.5.2. La triade catalytique.

Les acides aminés de la triade catalytique catalysent la réaction et constituent par conséquent un élément très conservé dans les structures des lipases. Cette triade est constituée d'une sérine, un acide aspartique ou glutamique et une histidine (Brady *et al.* 1990) comme dans les sérine-protéase avec toutefois un ordre différent dans la séquence primaire qui reflète la différence de repliement avec les sérine-protéases du clan PA de la classification MEROPS (Rawlings *et al.* 2014). Les résidus de la triade catalytique sont placés à proximité immédiate dans l'espace par le repliement α/β . La sérine et l'aspartate / glutamate sont localisés en fin des brins $\beta 5$ et $\beta 7$, respectivement, alors que l'histidine est localisée dans la boucle suivant le brin $\beta 8$. La sérine catalytique fait partie d'une séquence consensus G – X – S – X – G située dans un « γ -turn » appelé coude nucléophile (Ollis *et al.* 1992). Cette séquence est remplacée par la séquence G – D – S – L dans une sous-famille de lipases identifiée en 2004 (Akoh *et al.* 2004).

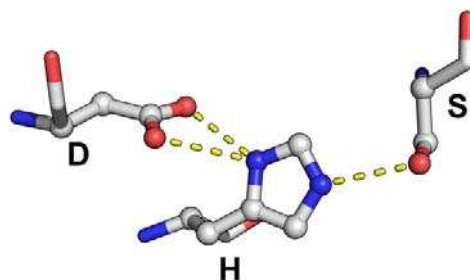


Figure 6. Triade catalytique de la lipase de *Rhizomucor miehei* (PDB ID : 4TGL)
(D, aspartate; H, histidine et S, sérine)

L'acide glutamique va jouer le rôle de résidu acide et l'histidine, le rôle de résidu basique alors que la sérine sera le résidu nucléophile. Ainsi l'aspartate va former une liaison hydrogène avec le groupe imidazole de l'histidine permettant à celle-ci d'adopter une orientation favorable en direction de l'oxygène γ de la sérine (Figure 6). De plus, l'interaction avec l'aspartate va modifier le pK_a du groupement imidazole de l'histidine. Durant l'étape de catalyse, le groupement imidazole de l'histidine rend l'oxygène γ de la sérine plus nucléophile et permet l'attaque du carbonyle du substrat. La conservation de l'aspartate / glutamate souligne l'importance de l'interaction avec l'histidine pour que cette dernière puisse jouer son rôle d'accepteur de proton (Dodson *et al.* 1998).

1.5.3. Le trou oxyanion

L'intermédiaire tétraédrique formé au cours des étapes d'acylation et de déacylation intervenant dans la catalyse se retrouve stabilisé par les résidus du trou oxyanion. La charge négative portée par l'oxygène de l'intermédiaire réactionnel est plus particulièrement stabilisée par deux liaisons hydrogènes impliquant des groupements amides du squelette peptidique.

Le premier résidu du trou oxyanion est localisé dans la boucle suivant le brin $\beta 3$. Deux types de trous oxyanion ont été identifiés chez les lipases en fonction de la topologie de ces résidus. Dans le premier type de trou oxyanion, appelé type **GX**, la stabilisation du résidu du trou oxyanion se fait par une interaction entre sa chaîne latérale et un résidu d'amarrage. Par contre, dans le deuxième type, appelé type **GGGX**, la stabilisation du résidu du trou oxyanion se fait par l'interaction entre la chaîne latérale du résidu X et la chaîne latérale d'une alanine (Figure 7) (Pleiss *et al.* 2000). Les lipases de type fongiques telles que celle de *Yarrowia lipolytica* ont un trou oxyanion de type **GX** avec une sérine ou

une thréonine pour le résidu X. Le type de trou oxyanion semble être corrélé avec la spécificité pour le substrat. Ainsi les lipases avec un trou oxyanion de type **GX** hydrolysent préférentiellement les substrats avec des longueurs de chaînes carbonées moyennes à longues alors que le trou oxyanion de type **GGGX** est retrouvé dans des lipases ayant une spécificité pour des substrats avec des longueurs de chaînes carbonées courtes ou des estérases. Un troisième type de trou oxyanion, dénommé type Y, a été identifié mais ce type regroupe essentiellement des peptidases et des estérases (Widmann *et al.* 2010). Dans ce type Y, le trou oxyanion n'est pas formé par un groupement amide du squelette peptidique mais par le groupement hydroxyle de la chaîne latérale d'une tyrosine (Fischer *et al.* 2006).

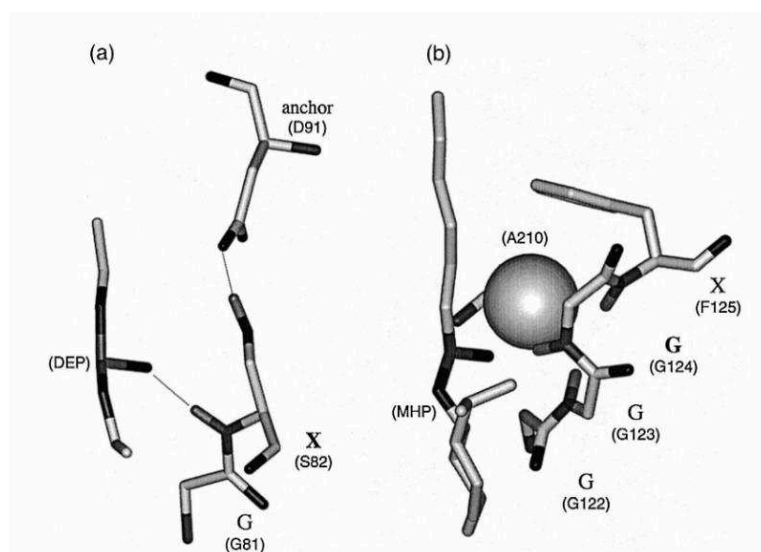


Figure 7. Trou oxyanion retrouvé chez les lipases. (a) Trou oxyanion de type **GX** retrouvé chez *Rhizomucor miehei* (PDB ID : 4TGL). Une liaison hydrogène entre le résidu S82 et un résidu d'ancrage, D91, permet une orientation adéquate du résidu du trou oxyanion G81 qui forme une liaison hydrogène par sa fonction amide avec le diéthylphosphonate (DEP). (b) Trou oxyanion de type **GGGX** retrouvé chez *Candida rugosa* (PDB ID : 1LPM). La fonction amide du résidu du trou oxyanion G124 est orientée vers l'inhibiteur, le (1R)-menthyl hexyl phosphonate (MHP), par l'interaction entre la chaîne latérale de l'alanine A210 et de la phénylalanine F125 (Pleiss *et al.* 2000)

Le second résidu du trou oxyanion suit directement la sérine catalytique dans la séquence consensus **G – X – S – X – G** et est par conséquent localisé dans le coude nucléophile, après le brin $\beta 5$.

Le trou oxyanion peut être préformé dans la conformation inactive dite « fermée » de la lipase comme dans le cas de la lipase de *Candida rugosa* (Grochulski *et al.* 1994b), ou ne se former que lorsque la lipase adopte une conformation active dite « ouverte » suite à un changement conformationnel d'un volet amphiphile présent dans le site actif comme dans le cas de la lipase de *Burkholderia cepacia* (Schrag *et al.* 1997b).

1.5.4. Le volet amphiphile ou « lid »

La détermination des premières structures de lipases par cristallographie, a révélé la présence d'un volet amphiphile fréquemment rencontré chez les lipases et constitué d'une ou plusieurs hélices α présentant une longueur et une complexité variables selon l'enzyme considérée. La résolution de structure de lipases cristallisées en présence d'inhibiteurs mimant le substrat dans le site actif a permis de mettre en évidence plusieurs conformations intermédiaires de ce volet, aussi appelé « lid » (Figure 8) (Brzozowski *et al.* 1991, Derewenda *et al.* 1992b, Grochulski *et al.* 1993, Grochulski *et al.* 1994a). En conformation qualifiée de « ouverte », la surface hydrophobe exposée est augmentée et le site catalytique devient accessible au substrat (Brady *et al.* 1990, Winkler *et al.* 1990, Derewenda *et al.* 1992a).

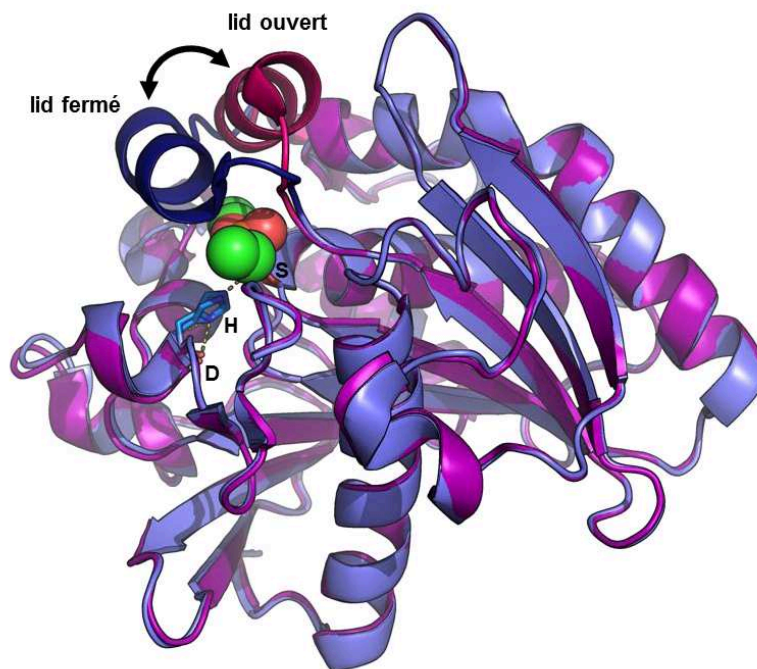


Figure 8. Structure de la lipase de *Rhizomucor miehei* avec le lid en conformation fermée en bleu (PDB ID : 3GTL) (Brzozowski *et al.* 1992) et en conformation ouverte en violet (PDB ID : 4GTL) (Derewenda *et al.* 1992a). L'inhibiteur suicide, le diéthylphosphonate, est représenté en sphères colorées en vert, il est présent uniquement dans la conformation ouverte. Les acides aminés catalytiques représentés en sticks sont indiqués, D pour l'aspartate 203, H pour l'histidine 257 et S pour la sérine 144.

Le passage d'une conformation fermée à une conformation ouverte a fourni une possible explication mécanistique au phénomène d'activation interfaciale mentionné précédemment. Ainsi, en l'absence d'interface hydrophobe, la conformation fermée est prédominante expliquant la faible activité catalytique observée. Lorsque la concentration en substrat est augmentée, et passe au-dessus de la concentration micellaire critique, la

formation d'une interface hydrophobe par le substrat insoluble permet une transition de la lipase vers une conformation ouverte ayant une activité catalytique accrue.

Néanmoins ce mécanisme reste mal compris. Ainsi la découverte de lipases présentant un volet amphiphile mais ne présentant pas ce phénomène d'activation interfaciale telle que la lipase de *Pseudomonas aeruginosa* (Jaeger *et al.* 1993) ou encore d'enzymes lipolytiques autres que les lipases ne possédant pas de volet amphiphile mais présentant un phénomène d'activation interfaciale telle que la phospholipase A2 pancréatique, (Pieterse *et al.* 1974), suggère un lien complexe entre l'activité et les états conformationnels du lid.

Par ailleurs, certaines lipases ont pu être cristallisées en l'absence d'inhibiteur suicide sous forme ouverte, telle que la lipase de *Burkholderia cepacia* (PDB ID : 2LIP, 3LIP) (Schrage *et al.* 1997b), ou partiellement ouverte telle que la lipase de *Rhizopus delemar* (PDB ID : 1TIC) (Derewenda *et al.* 1994) et celle de *Thermomyces lanuginosa* (PDB ID : 1DT3, 1DT5, 1DTE) (Brzozowski *et al.* 2000). La stabilisation de structures ouvertes ou partiellement ouverte du lid en l'absence de ligand suggère donc une dynamique plus complexe et il est probable que la nature des acides aminés constituant le lid ainsi que sa longueur jouent un rôle clef dans sa mobilité conformationnelle et influent donc grandement sur l'activité et la spécificité vis-à-vis du substrat (Brocca *et al.* 2003, Skjot *et al.* 2009).

Plusieurs techniques ont été utilisées pour explorer la dynamique et les changements conformationnels des lipases. On peut ainsi citer la fluorescence (Luthi-Peng *et al.* 1992, Yapoudjian *et al.* 2002), la diffraction des neutrons (Hermoso *et al.* 1997), la diffusion des neutrons aux petits angles (Pignol *et al.* 2000), la spectroscopie infrarouge à transformée de Fourier (Noinville *et al.* 2002) ou la résonance paramagnétique électronique (RPE) (Belle *et al.* 2007). Cette dernière technique a permis d'obtenir des informations particulièrement intéressantes sur la dynamique du lid de la lipase pancréatique humaine (HPL). L'introduction d'une cystéine par mutation d'un acide aminé du lid (D249C) a permis de fixer une sonde paramagnétique sur le lid de la HPL. De façon intéressante, le passage d'une conformation fermée à une conformation ouverte s'est accompagné d'une restriction de la mobilité de la sonde. Cela a permis de discriminer entre les états ouvert et fermé et d'étudier l'influence de la présence de sels biliaires, le sodium taurodeoxycholate (NaTDC), et de la colipase sur l'équilibre conformationnel du lid de la HPL entre les états fermé et ouvert. La présence de NaTDC au-delà de 1 mM déplace l'équilibre en faveur de la conformation ouverte, cet effet étant accentué par la présence de la colipase. Toutefois, la colipase seule ne modifie pas l'équilibre conformationnel. La colipase aurait donc un effet stabilisateur sur l'ouverture du lid par le NaTDC. De plus, une dilution de la concentration en NaTDC a entraîné un retour de l'équilibre en faveur de la conformation fermée malgré la présence de la colipase, retour qui a été inhibé par l'ajout d'un inhibiteur suicide se liant à la sérine

catalytique et immobilisant les résidus du trou oxyanion. Ainsi, cette étude a démontré le rôle déterminant joué par les sels biliaires dans l'ouverture du lid de la HPL.

La RPE a également permis de mesurer les distances entre la cystéine 249, située dans le lid du variant D249C, et la cystéine C181, située dans une boucle en face du lid, pour les conformations fermée et ouverte en solution par « Double Electron – Electron Resonance » (DEER) (Ranaldi *et al.* 2010). Ainsi à l'aide de deux sondes paramagnétiques, des distances de 19 Å et 42 Å ont été estimées pour les conformations fermée et ouverte, respectivement. Ces distances mesurées en solution sont identiques à celles retrouvées dans les structures cristallographiques de HPL en conformations fermée (PDB ID : 1N8S) et ouverte (PDB ID : 1IPB), supportant l'hypothèse que ces conformations déterminées par cristallographie existent en solution.

En l'absence de techniques biophysiques permettant d'appréhender les échelles de temps auxquelles se produisent les changements conformationnels des domaines protéiques, la dynamique moléculaire est une bonne alternative puisqu'elle permet de simuler les réarrangements conformationnels se produisant au cours du temps et d'accéder à une description atomistique. Un exemple en est donné pour la lipase de *Burkholderia cepacia* (BCL). Seule la structure cristallographique ouverte de cette lipase est disponible, que ce soit avec ou sans inhibiteur / substrat (Kim *et al.* 1997, Schrag *et al.* 1997b, Lang *et al.* 1998, Luic *et al.* 2001, Mezzetti *et al.* 2005, Luic *et al.* 2008). Le lid de BCL est constitué des résidus 118 – 159 et est formé par les hélices $\alpha 4$ et $\alpha 5$ reliées par une boucle. A partir de la forme ouverte, une simulation de dynamique moléculaire de 20 ns en solvant aqueux a permis de passer à une forme fermée alors que la même simulation en solvant apolaire, l'octane, ou en présence d'une interface hydrophobe, eau / octane, n'a pas permis d'atteindre cette conformation fermée (Barbe *et al.* 2009). Une seconde simulation à partir de la conformation fermée placée en solvant apolaire, l'octane, a permis de repasser à une conformation ouverte du lid de BCL. En plus de souligner le rôle crucial de l'environnement sur la dynamique du lid de BCL, cette simulation a également mis à jour les détails du mécanisme de fermeture et d'ouverture du lid. Ainsi des simulations de dynamique moléculaire restreinte ont montré l'importance des résidus 214 – 261. Ces résidus forment notamment l'hélice $\alpha 9$ qui fait face au lid. Alors que des restrictions des mouvements du squelette peptidique et des chaînes latérales de ces résidus bloquent le lid dans une conformation semi-fermée, la mobilité des chaînes latérales seules suffit pour passer à une conformation pleinement fermée. Le même phénomène est observé lors de l'ouverture du lid lors de simulations de dynamique moléculaire restreinte à partir de la forme fermée en solvant apolaire. L'importance de la nature des résidus constituant le lid a aussi été mise en évidence en remplaçant un acide aminé apolaire, la valine V138, par un résidu apolaire de faible taille, une alanine, ou un résidu polaire, une sérine. Ces changements ont conduit

à une impossibilité du lid à repasser en conformation fermée en solvant aqueux. Un phénomène analogue a été observé pour la mutation de la phénylalanine F142 en sérine. Cette mutation F142S empêche la fermeture du lid en le stabilisant dans une conformation intermédiaire. Cette étude de BCL par dynamique moléculaire illustre la complexité de la dynamique du lid en montrant l'implication du solvant, des interactions inter-domaines et de la composition du lid dans cette dynamique. Une étude semblable par dynamique moléculaire du lid de BCL a permis de retrouver le rôle crucial joué par l'environnement sur la conformation du lid de BCL (Trodler *et al.* 2009).

1.5.5. Topologie du site actif

Le site actif se situe au sommet du feuillet β central. La topologie du site de fixation de la partie acyle du substrat varie entre les lipases et détermine la spécificité vis-à-vis de la longueur de chaîne du substrat, du nombre d'insaturations ou de leur position. Il est possible de classifier les lipases en trois catégories en fonction de la topologie du site de liaison au substrat (Figure 9) (Pleiss *et al.* 1998).

- site en forme de crevasse située près de la surface avec la sérine catalytique située au fond de la crevasse (exemple de la lipase de *Rhizomucor miehei*)
- site en forme d'entonnoir avec la sérine catalytique située au fond de l'entonnoir (exemples des lipases de *Candida antarctica* (lipase B) et *Burkholderia cepacia*)
- site en forme de tunnel avec la sérine catalytique située près de l'ouverture du tunnel (exemple de la lipase de *Candida rugosa*)

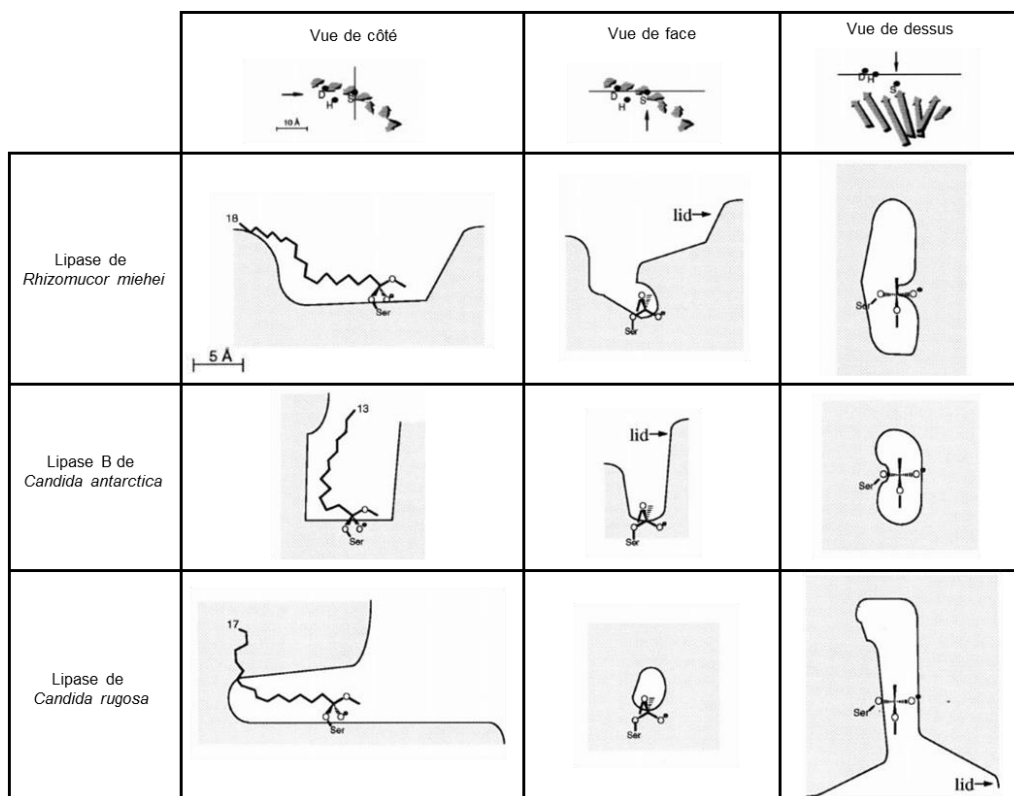


Figure 9. Site de liaison du substrat. Représentation des sites de liaison des lipases de *Rhizomucor miehei*, *Candida antarctica* (lipase B) et *Candida rugosa*. L'orientation de la vue est indiquée par rapport au feuillet β central et aux trois acides aminés catalytiques (D, aspartate ; H, histidine ; S, sérine). Sur chaque vue, la position de la sérine catalytique est indiquée ainsi que l'orientation du groupement acyle du substrat et la longueur maximale de la chaîne hydrocarbonée. La position du lid est indiquée le cas échéant. Adapté de (Pleiss *et al.* 1998).

1.6. Notre enzyme modèle : Lip2 de *Yarrowia lipolytica*.

1.6.1. Découverte de Lip2

La lipase lip2 de *Yarrowia lipolytica* a été découverte en 2000 (Pignede *et al.* 2000a, Pignede *et al.* 2000b). Lip2 est une lipase extracellulaire sécrétée et elle est responsable de la majorité de l'activité lipolytique de la levure *Yarrowia lipolytica*. La forme mature de Lip2 est une glycoprotéine de 38 kDa et de 301 acides aminés et est encodée par le gène *LIP2* (Pignede *et al.* 2000a). Ainsi, ce gène encode un polypeptide de 334 acides aminés dont les 33 premiers constituent un signal peptide appelé le PrePro permettant la sécrétion de l'enzyme dans le milieu extracellulaire. Ce signal peptide est composé d'une courte séquence « Pre » de 13 acides aminés suivis par quatre dipeptides X-Ala ou X-Pro et d'une séquence « Pro » de 10 acides aminés se terminant par un motif de clivage Lysine – Arginine. Ce motif est le substrat d'une endoprotéase KEX2 encodée par le gène *XPR6* (Enderlin *et al.* 1994).

Quatre différents isoformes de Lip2 ayant des poids moléculaires entre 36 et 38 kDa ont été découverts (Aloulou *et al.* 2007b). Deux glycosylation majeures, ont été identifiées pour l'isoforme principal de Lip2, $\text{Man}_8\text{GlcNAc}_2$ et $\text{Man}_9\text{GlcNAc}_2$ en position N134 et N113,

constituées de α -D-mannose (Man) et de N-acétyl- α -D-glucosamine (GlcNAc) (Jolivet *et al.* 2007). Toutefois ces glycosylations ne sont pas nécessaires à la sécrétion, au repliement ou à l'activité bien qu'il ne soit pas exclu qu'elles puissent moduler les phénomènes d'activation interfaciale ou d'adsorption de l'enzyme (Aloulou *et al.* 2007b, Jolivet *et al.* 2007, Yu *et al.* 2007, Aloulou *et al.* 2013).

1.6.2. Structure de Lip2

La structure de lip2 a été résolue en conformation fermée en 2010 (PDB ID : 3O0D) par cristallographie à une résolution de 1.7 Å (Bordes *et al.* 2010, Aloulou *et al.* 2013). Lip2 adopte le repliement en α/β avec 5 hélices α distribuées de part et d'autre d'un feuillet β central constitué de 9 brins β (Figure 10).

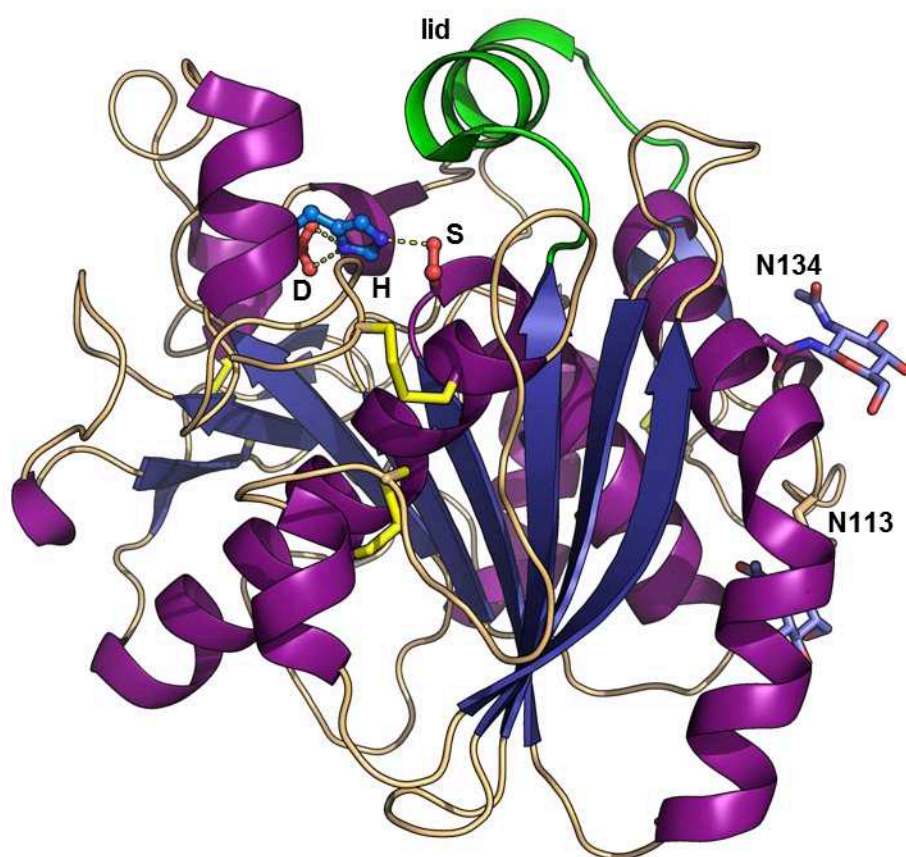


Figure 10. Structure du repliement α/β de Lip2 de *Yarrowia lipolytica* (PDB ID : 3O0D) (Bordes *et al.* 2010). Le feuillet β central est coloré en bleu alors que les hélices α sont en violet et le lid en vert respectivement. Les résidus catalytiques sont indiqués, D pour l'aspartate 230, H pour l'histidine 289, S pour la sérine 162. Les ponts disulfures sont colorés en jaune et les sites de glycosylations N134 et N113 sont indiqués avec une représentation des résidus NAG.

Par alignement de séquences homologues, les résidus D230 – H289 – S162 et L163 – T88 ont été identifiés comme constituant la triade catalytique et le trou oxyanion,

respectivement. Toutefois, lors de la superposition avec des structures de lipases homologues résolues en conformation fermée, si les résidus D230 – H289 – S162 et L163 se superposent correctement avec les acides aminés correspondant dans les autres structures, la position de T88 diffère considérablement de S82 pour *Rhizomucor miehei*, S83 pour *Thermomyces lanuginosa* et T83 pour *Rhizopus niveus* (Figure 11A) (PDB ID : 3TGL, 1DT3 et 1LGY, respectivement). De la même manière, le lid de lip2 a été identifié comme étant formé par les résidus T88 – L105. En comparaison des lipases homologues, le lid de lip2 apparaît plus long de 3 résidus, I98 – R99 – I100, ajoutant ainsi un tour supplémentaire en C-terminal de l'hélice α (Figure 11B). Egalement, l'allongement de deux résidus du brin β 3 chez lip2 perturbe fortement la conformation de la boucle 87 – 90 qui inclue le second résidu potentiel du trou oxyanion, T88.

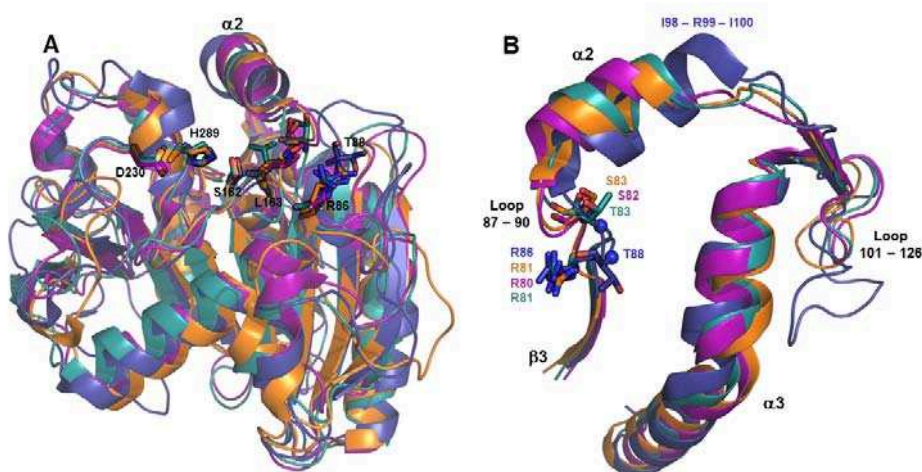


Figure 11. Superposition des structures fermées des lipases de *Yarrowia lipolytica* en bleu (PDB ID : 3O0D), *Thermomyces lanuginosa* en orange (PDB ID : 1DT3), *Rhizomucor miehei* en magenta (PDB ID : 3TGL) et de *Rhizopus niveus* en cyan (PDB ID : 1LGY). A. Structure complète. Les résidus catalytiques, D230 H289 et S162, ainsi que la leucine L163 formant le trou oxyanion sont représentés en sticks. Le second résidu putatif, la thréonine T88 est également représentée en stick ainsi que la chaîne latérale de l'arginine R86. L'hélice α 2 formant le lid est indiquée. La numérotation des résidus correspond à celle de Lip2 de *Yarrowia lipolytica*. B. Vue rapprochée au niveau du lid. Les résidus R86 et T88 ainsi que leurs homologues sont représentés en stick. L'azote de la thréonine 88 et des résidus équivalent est représenté en sphère. Les résidus 98 – 99 – 100 formant un tour d'hélice α supplémentaire dans le lid de Lip2 sont également indiqués.

Lip2 possède également neuf cystéines qui forment quatre ponts disulfures (Figure 12). Le pont disulfure entre les cystéines C30 – C299 lie ensemble les extrémités N- et C-terminales de la protéine. Les autres ponts disulfures, C43 – C47, C120 – C123 et C265 – C273 sont situés dans des boucles et jouent vraisemblablement un rôle dans leur flexibilité. Enfin, la cystéine C244 est libre et il a été montré qu'elle était impliquée dans les mécanismes de dénaturation thermique de la lipase en favorisant les phénomènes d'agrégation (Bordes *et al.* 2011).

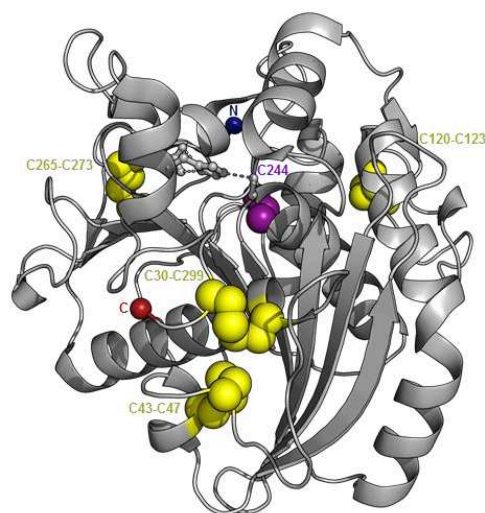


Figure 12. Ponts disulfures et cystéines de Lip2. Les cystéines sont représentées sous forme de sphère. Les cystéines formant des ponts disulfures sont colorées en jaune alors que la cystéine 244 est en violet. Les extrémités N- et C-terminales de la chaîne polypeptidique sont indiquées en bleu et en rouge respectivement.

1.6.3. Propriétés enzymatiques de Lip2

Les propriétés biochimiques de Lip2 ont été caractérisées par différentes études (Aloulou *et al.* 2007a, Aloulou *et al.* 2007b, Yu *et al.* 2007). Ainsi Lip2 est capable d'hydrolyser des triglycérides à chaîne carbonée courte (tributyryne), moyenne (trioctanoylglycerol) ou longue (huile d'olive en émulsion avec de la gomme arabique). L'activité spécifique maximale a été mesurée pour les triglycérides à chaîne longue. Par ailleurs, pour la tributyrine et le trioctanoylglycerol, l'activité n'est détectable qu'en présence de sodium taurodeoxycholate (NaTDC). Cette observation suggère que Lip2 est engagée dans une interaction complexe avec les interfaces hydrophobes et requiert la présence d'agents tensioactifs tels que le NaTDC ou la gomme arabique pour être active sur les triglycérides. De plus, contrairement à de nombreuses lipases, Lip2 n'est pas inhibée par les sels biliaires (Aloulou *et al.* 2007a). Enfin, Lip2 présente une régiosélectivité de type sn1,3 lors de l'hydrolyse des triglycérides comme en témoigne l'accumulation de 2-monoglycérides au cours de la réaction.

Le pH optimal de Lip2 a été trouvé à 6.0 et la lipase est stable après incubation à un pH entre 4.0 et 7.0. Par contre, une incubation à des pH inférieurs à 3.0 ou supérieurs à 8.0 entraîne une diminution importante et irréversible de l'activité (Aloulou *et al.* 2007b). De même, une température optimale de 40 °C a été mesurée avec une inactivation rapide pour des températures au-delà de 50 °C (Yu *et al.* 2007). Un variant thermostable avec une demi-vie accrue à 60 °C a été obtenu en remplaçant la cystéine 244 par une alanine (Bordes *et al.* 2011). La présence de solvant a également un effet sur l'activité de lip2, ainsi 10% d'acétonitrile ou 20% d'éthanol, d'isopropanol ou d'acétone entraînent une inactivation de l'enzyme. Une concentration de 10% de méthanol, éthanol, isopropanol, acétone ou DMSO

est sans effet sur l'activité. La présence d'ions divalents a aussi un effet modulateur sur l'activité catalytique de Lip2 et si les ions Ca^{2+} et Mg^{2+} l'augmente, les ions Zn^{2+} , Ni^{2+} ou Cu^{2+} la diminue (Yu *et al.* 2007).

1.6.4. Intérêt Biotechnologique.

Les propriétés de la lipase lip2 de *Yarrowia lipolytica* en font une enzyme avec un fort potentiel d'applications industrielles (Fickers *et al.* 2011).

En effet, lip2 a montré un intérêt certain dans la résolution de mélanges racémiques. Les acide 2-halogéno-carboxyliques sont d'importants intermédiaires dans la synthèse de molécules thérapeutiques (analgésiques, prostaglandines, prostacyclines, pénicillines semi-synthétique). Lip2 a montré une énantiosélectivité de 28 pour l'énantiomère S lors de la résolution d'un mélange racémique de 2-bromo-p-tolyl-acétique-éthyl-ester par transesterification (Guieysse *et al.* 2004). De même, lip2 a montré également son efficacité pour la résolution de 2-bromo-o-tolylacétique-éthyl-ester, un précurseur d'analgésiques et d'antagonistes au récepteur de l'angiotensine 2, avec une énantiosélectivité de 27. Par une approche semi-rationnelle, il a été possible d'améliorer l'énantiosélectivité de lip2 vis-à-vis des esters 2-bromo-arylacétique (Bordes *et al.* 2009). Sur la base d'un modèle tridimensionnel de lip2 généré par modélisation comparative, 5 positions de résidus du site de liaison au substrat ont été sélectionnés. L'importance de ces positions dans l'énantiosélectivité a été étudiée par mutagénèse dirigée des acides aminés sélectionnés. Les positions V232 et D97 se sont révélées particulièrement importantes et ont été étudiées plus en détail. Sur ces deux positions, les variants substitués par les 19 autres acides aminés possibles ont été générés. Le variant V232S s'est avéré avoir une énantiosélectivité supérieure à 200 pour la transformation de l'énantiomère S du 2-bromo-phényl-acétique-octyl-ester et du 2-bromo-o-tolyl-acétique-octyl-ester lors de réactions d'hydrolyse. De plus, cette amélioration de l'énantiosélectivité s'accompagne d'une amélioration de la vitesse de réaction vis-à-vis de l'énantiomère S pour les deux substrats. La substitution de la position 232 par un acide aminé ayant une chaîne latérale volumineuse, tel qu'une leucine, isoleucine, phénylalanine ou une tyrosine, entraîne un changement de l'énantiosélectivité en faveur de l'énantiomère R par diminution de l'activité catalytique sur l'énantiomère S. Le même phénomène est observé à la position 97 avec le remplacement de l'aspartate par une alanine. Enfin la combinaison de ces deux mutations dans le double variant D97A-V232F a permis non seulement de changer l'énantiosélectivité pour l'énantiomère R mais aussi d'augmenter l'activité vis-à-vis de cet énantiomère (Cambon *et al.* 2010). Ce changement d'énantiosélectivité est le fruit d'une diminution par un facteur 1000 de l'activité catalytique pour l'énantiomère S avec une augmentation par un facteur 20 de l'activité pour l'énantiomère R. Il est ainsi remarquable que la mutation de seulement deux positions a

permis de changer l'énantiosélectivité de Lip2. De plus, le changement de l'énantiosélectivité et de l'activité catalytique du double variant n'est pas seulement une simple cumulation des propriétés de chaque simple variant, puisqu'il a été constaté que ces deux mutations avaient un effet synergique sur les propriétés énantiosélectives de Lip2.

Lip2 s'est également révélée être une enzyme efficace pour la purification d'acide gras polyinsaturés (PUFA) de type oméga-3 (Casas-Godoy *et al.* 2014). Ainsi l'acide cis-4,7,10,13,16,19-docosahexanoïque (DHA) et l'acide cis-5,8,11,14,17-eicosapentanoïque (EPA) ont pu être purifiés par hydrolyse d'un mélange d'esters d'éthyle extrait d'huile de poisson. La purification enzymatique des acides gras polyinsaturés se fait par résolution cinétique. Les acides gras riches en insaturations sont souvent plus lentement catalysés que les acides gras saturés ou mono-insaturés. En effet, la présence de plusieurs insaturations augmente l'encombrement stérique de la chaîne carbonée en limitant sa mobilité au niveau du site actif. La comparaison des vitesses de réaction entre les différents esters avec le nombre et la position des insaturations a confirmé que l'activité de lip2 était plus affectée par la présence de doubles liaisons sur la chaîne carbonée que par la longueur de cette dernière. Cet effet est d'autant plus délétère que la première insaturation est proche du groupement carbonyle de l'ester. Ces propriétés ont permis dès lors d'atteindre une pureté de 90% en oméga-3 et de 77% en DHA avec un rendement de 77% et de 94%, respectivement, dans un réacteur ouvert permettant l'évaporation de l'éthanol produit au cours de l'hydrolyse.

L'insuffisance pancréatique exocrine se traduit par un défaut de la sécrétion exocrine du pancréas constituée principalement d'enzymes participant à la digestion. Cette insuffisance se traduit par une malabsorption des nutriments et une dénutrition. Si les causes peuvent être multiples, telles que des pancréatites chroniques ou aiguës, des tumeurs pancréatiques ou la mucoviscidose, le traitement reste identique et consiste en des suppléments par voie orale d'extrait pancréatique de porc contenant des protéases, des amylases et des lipases (Keller *et al.* 2003). Néanmoins, en raison de l'origine des enzymes utilisées, ce traitement présente un risque de transmission viral inter-espèce. *Yarrowia lipolytica* est une levure non-pathogénique pour l'homme et a été classifiée comme « generally recognized as safe » (GRAS) par la Food and Drug Administration (FDA, USA) (Madzak *et al.* 2004, Groenewald *et al.* 2014). Les caractéristiques de lip2, son activité sur les triglycérides à longue chaîne ainsi que sa stabilité à pH acide et en présence de sels biliaires, en font un bon candidat pour le traitement de l'insuffisance pancréatique exocrine (Aloulou *et al.* 2007a, Aloulou *et al.* 2007b, Turki *et al.* 2010).

1.6.5. Etude de la dynamique de Lip2

Seule la conformation dite fermée de lip2 a pu être obtenue par cristallographie malgré toutes les tentatives pour obtenir une conformation ouverte (Bordes *et al.* 2010). Pour pallier au manque d'information expérimentale, des simulations réalisées par dynamique moléculaire ont été réalisées pour étudier plus en détails le mécanisme d'ouverture du volet amphiphile et accéder à un modèle structural de la conformation ouverte de lip2. Ainsi des simulations de 20 ns ont été réalisées à partir de la structure de lip2 en conformation fermée en solvant explicite, soit de l'eau, soit dans un milieu biphasique eau/octane. Alors que dans l'eau, lip2 demeure dans une conformation fermée, lorsqu'elle est placée dans le mélange eau/octane, on observe de larges changements conformationnels au niveau du volet amphiphile. Lip2 adopte alors une conformation dite « semi-ouverte », qui n'évolue pas vers une conformation complètement ouverte même lorsque la simulation est poursuivie pendant 60 ns. Bien qu'il ne puisse être exclu qu'une simulation plus longue soit nécessaire pour passer à une conformation ouverte, il a été suggéré que la présence d'une interface seule n'était pas suffisante pour atteindre cette conformation ouverte.

L'ajout d'un substrat lié de façon covalente à la sérine catalytique dans le site actif a permis, en poursuivant la simulation, de passer de la conformation semi-ouverte à une conformation ouverte du lid. De plus, à la fin de la simulation, l'intermédiaire tétraédrique et les résidus catalytiques (S162, H289 et D230) établissent les liaisons hydrogènes classiquement décrites. Toutefois, la thréonine T88, second résidu potentiel du trou oxyanion, se retrouve dans une conformation qui ne lui permet pas de stabiliser l'intermédiaire réactionnel alors que la chaîne latérale de l'arginine R86 est réorientée vers le site catalytique. Le groupement guanidinium de l'arginine R86 devient alors susceptible de stabiliser la charge négative de l'intermédiaire tétraédrique. Toutefois, les expériences de mutagenèse dirigée réalisées sur les résidus R86 et T88 n'ont pas permis de préciser le rôle de chacun de ces acides aminés, les mutants à ces deux positions s'étant avérés inactifs (R86A, R86K, T88L, T88V et T88A) (Bordes *et al.* 2009).

La dynamique moléculaire a donc permis de suggérer un mécanisme d'ouverture du lid en deux étapes nécessitant à la fois la présence d'une interface hydrophobe et l'interaction avec le substrat pour atteindre une conformation ouverte pleinement active. Par ailleurs, cette approche a aussi souligné la particularité du trou oxyanion de lip2 due à la conformation particulière la boucle incluant la thréonine T88 et la possibilité d'une participation de l'arginine R86 dans la catalyse a été proposée.

Afin de mieux comprendre les relations entre les changements conformationnels suggérés par la dynamique moléculaire et l'activité de lip2, nous avons entrepris une investigation plus poussée basée sur l'étude de la dynamique par RMN. Ainsi, les objectifs

étaient de caractériser la dynamique de lip2 et de mettre en évidence les déterminants structuraux et environnementaux qui dirigent cette dynamique et son influence sur les propriétés de catalyseur de la lipase. La nécessité d'explorer des mouvements à différentes échelles de temps, mouvement de chaîne latérales, du squelette peptidique et de domaines nous a orientés vers une approche par RMN. En effet, cette technique permet d'explorer la dynamique sur une large gamme de temps et à une résolution atomique. Toutefois, la taille importante des lipases limite leur étude par spectroscopie RMN. Seule la lipase de *Pseudomonas mendocina* a été étudiée par RMN à ce jour pour examiner l'effet stabilisateur de deux mutations (Sibille *et al.* 2006). La taille importante de lip2, 301 acides aminés et 37 kDa après glycosylation, impose le recours à un marquage $^{15}\text{N}^{13}\text{C}$ et, idéalement à la perdeutériation de la protéine. Nous avons ainsi commencé par explorer les possibilités d'expression avec marquage uniforme de lip2 dans *Escherichia coli*, dans les levures *Pichia pastoris* et *Yarrowia lipolytica*, ainsi que l'expression en système acellulaire. Nous avons aussi effectué diverses tentatives de marquage acide – aminé spécifique (en particulier avec des leucines), dans le but de sonder spécifiquement la dynamique du lid.

Les résultats obtenus dans le cadre de cette étude constituent le premier chapitre de thèse et font l'objet d'une première publication dans PROTEIN EXPRESSION AND PURIFICATION Volume : 101 Pages : 14-20.

II. La protéine OmpA de *Klebsiella pneumoniae*.

2.1. *Klebsiella pneumoniae*

Klebsiella pneumoniae fait partie de la famille des *Enterobacteriaceae* qui dépend du phylum des *Proteobacteria* regroupant un grand nombre d'espèces de bactéries gram-négatives. Ces gram-négatives sont usuellement des bacilles de taille moyenne de 1 à 5 μm , oxydase négative et nitrate réductase positives, aéro-anaérobies facultatives et utilisant la fermentation du glucose, mobiles ou immobiles (Borman *et al.* 1944, Kauffmann 1956). Ces bactéries sont largement retrouvées dans l'environnement mais également au niveau du tractus intestinal de l'homme et des animaux où elles sont commensales. On distingue dans ce groupe les pathogènes strictes responsables d'infections tels que la fièvre typhoïde (*Salmonella typhi*), la dysenterie (*Shigella dysenteriae*) ou la peste (*Yersinia pestis*), des pathogènes opportunistes qui provoquent des infections chez les patients affaiblis ou immunodéprimés tels que *Escherichia coli*, *Klebsiella pneumoniae*, *Enterobacter* ou *Serratia*. Néanmoins, ces pathogènes opportunistes peuvent être responsables d'infections chez des individus sains par l'acquisition de facteurs de virulence

comme pour les souches d'*Escherichia coli* entérotoxiques (ETEC), entéropathogène (EPEC) ou entérohémorragique (EHEC).

Klebsiella pneumoniae est le représentant le plus important cliniquement du genre *Klebsiella*. Ce pathogène opportuniste est retrouvé au niveau du tractus gastro-intestinal mais aussi du tractus urinaire et des voies respiratoires supérieures (Podschun *et al.* 1998). En raison de sa capacité à former des biofilms, *Klebsiella pneumoniae* est aussi fréquemment retrouvée au niveau des dispositifs médicaux tels les cathéters ou les sondes d'intubation endotrachéales (Schroll *et al.* 2010). La capacité à former des biofilms *in vivo* (Jagnow *et al.* 2003), ainsi que des phénotypes de multirésistance dus à la présence fréquente de β -lactamases à spectre étendu ou de carbapénémases (Paterson *et al.* 2004, Munoz-Price *et al.* 2013), rendent les infections opportunistes causées par *Klebsiella pneumoniae* particulièrement délicates à traiter.

La capsule polysaccharidique constitue le principal facteur de virulence de *Klebsiella pneumoniae*. Cette capsule correspond à l'antigène K dont certains sérotypes, principalement K1 et K2, donnent un phénotype hypermucovisqueux associé à une virulence accrue (Lin *et al.* 2011, Shon *et al.* 2013). La capsule protège *Klebsiella pneumoniae* de la phagocytose en bloquant les phénomènes de reconnaissance et d'internalisation (Sahly *et al.* 2000, Cortes *et al.* 2002, Evrard *et al.* 2010, Pan *et al.* 2011). De plus, les formes virulentes K1 de *Klebsiella pneumoniae* peuvent survivre en intracellulaire après la phagocytose ce qui va permettre la formation de foyers d'infections secondaires (Wu *et al.* 2008). En plus de la capsule, le lipopolysaccharide responsable de l'antigène O, les fimbriae et les protéines de la membrane externe constituent des facteurs de virulence supplémentaires. Notamment OmpA de *Klebsiella pneumoniae* joue un rôle important dans la pathogénicité en prévenant l'activation des cellules épithéliales des voies respiratoires en agissant sur les voies de signalisation de NF- κ B, p38 et p44/42 (March *et al.* 2011) et en contribuant à la résistance à la phagocytose par les macrophages alvéolaires (March *et al.* 2013). OmpA est donc un facteur important de la virulence de *Klebsiella pneumoniae* qui permet à la bactérie d'échapper aux mécanismes de défense du système immunitaire. Une étude structurale et dynamique de OmpA permettra de mettre à jour les mécanismes moléculaires permettant à cette protéine de jouer un rôle dans la virulence de *Klebsiella pneumoniae*, tout particulièrement les boucles extracellulaires du domaine transmembranaire comme nous le verrons au chapitre 3.

2.2. L'enveloppe bactérienne.

L'enveloppe bactérienne est une structure qui sépare le cytoplasme de la cellule de l'environnement extérieur. La coloration de Gram permet de définir deux principales

catégories de bactéries en fonction de leurs propriétés à retenir le cristal violet. Cette propriété traduit une structuration différente de l'enveloppe bactérienne entre les bactéries Gram-positives et les bactéries Gram-négatives (Figure 13). Bien que les deux types de bactéries possèdent une membrane interne accompagnée d'un peptidoglycane, l'épaisseur de cette couche de peptidoglycane est substantiellement plus épaisse, de 20 à 80 nm contre 1 à 7 nm, chez les Gram-positives que chez les Gram-négatives (Cabeen *et al.* 2005). De plus, les bactéries Gram-négatives possèdent une membrane externe ou « Outer Membrane » (OM) dont sont dépourvues les bactéries Gram-positives. *Klebsiella pneumoniae* étant un bacille Gram-négatif, nous nous concentrerons sur la description de l'enveloppe des bactéries Gram-négatives.

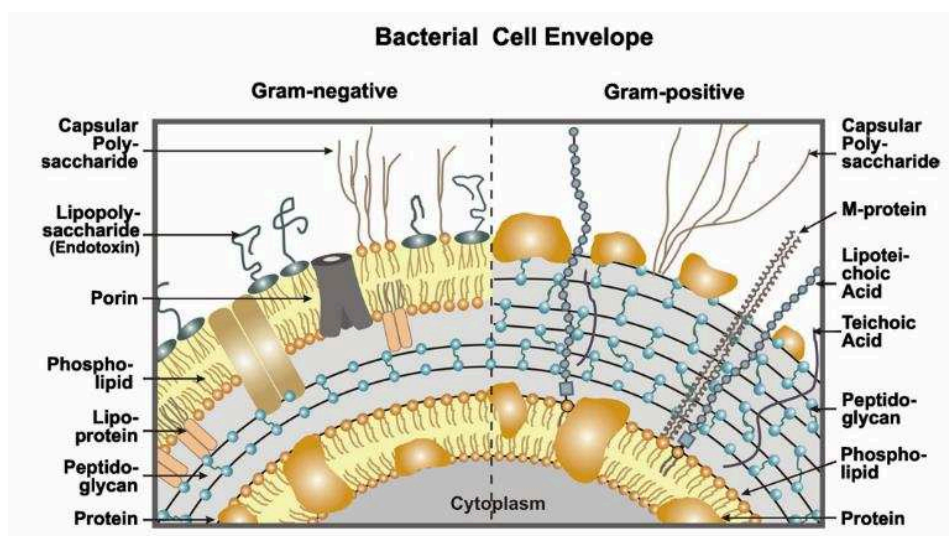


Figure 13. Représentation schématique de l'enveloppe bactérienne, Gram-négative à droite et Gram-positives à gauche. L'enveloppe des bactéries Gram-négatives est constituée d'une couche de peptidoglycane plus fine ainsi que d'une membrane externe absente chez les bactéries Gram-positives. Figure reprise de (Muller-Loennies *et al.* 2007).

2.2.1. Les composants de l'enveloppe bactérienne

La membrane cytoplasmique encore appelée membrane interne chez les Gram-négatives, délimite le cytoplasme et est constituée par une bicouche de phospholipides. Par exemple, la membrane cytoplasmique d'*Escherichia coli* est principalement composée de phosphatidyléthanolamine et de phosphatidylglycérol ainsi que de phosphatidylsérine et de cardiolipide dans une moindre mesure. Cette membrane assure de nombreuses fonctions qui sont effectuées au niveau des membranes des organites chez les eucaryotes supérieurs. De nombreuses protéines associées avec la production énergétique, la synthèse des lipides, la sécrétion des protéines ou encore le transport sont retrouvées à ce niveau.

La membrane interne est entourée d'une couche de peptidoglycane situés dans l'espace périplasmique. Le peptidoglycane est un polymère glycosidique constitué d'une répétition d'une unité di-saccharidique de N-acétylglucosamine (NAG) et d'acide N-acétylmuramique (NAM) liés en β 1 – 4. La longueur des brins saccharidiques est variable en fonction des espèces bactériennes (Vollmer *et al.* 2010). Le groupement lactoyl en position 3 du NAM permet le branchement d'un pentapeptide dont la séquence est L-Alanine – γ -D-Glutamate – acide diaminopimélique (DAP) – D-Alanine – D-Alanine chez la plupart des bactéries gram-négatives (Figure 14). Ces pentapeptides peuvent être ramifiés entre eux par des liaisons peptidiques impliquant le DAP en position 3 du pentapeptide et la D-Alanine en position 4 d'un peptide voisin. Néanmoins de nombreuses variations existent en fonction des espèces bactériennes. Notamment, la structure du peptidoglycane des bactéries Gram-positives diffère par la présence d'une Lysine à la place du DAP et la présence de plusieurs acides aminés supplémentaires assurant la liaison entre les parties peptidiques du peptidoglycane (Vollmer *et al.* 2008a).

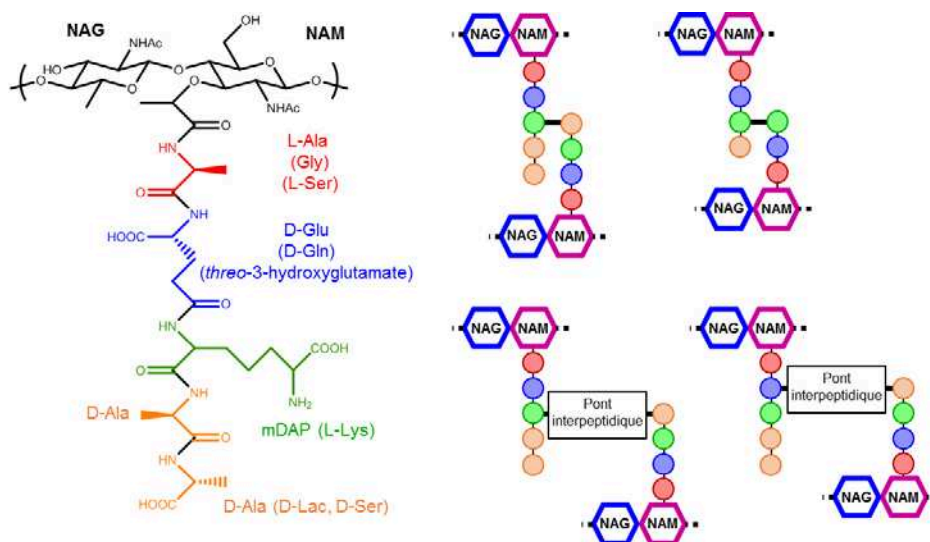


Figure 14. Composition du peptidoglycane. La partie saccharidique est une répétition d'un disaccharide de N-Acétylglucosamine (NAG) et d'acide N-Acétylmuramique (NAM). La partie peptidique est constituée du pentapeptide L-Alanine – D-Glutamate – mésoDAP – D-Alanine – D-Alanine chez la plupart des bactéries gram-négatives. Les variations possibles de séquence sont indiquées en parenthèses. La liaison interpeptidique se fait directement entre le résidu mDAP en position 3 et la D-Alanine en position 4 mais d'autres possibilités sont indiquées y compris faisant intervenir un pont interpeptidique (Vollmer 2012).

Ce pontage entre les brins glycosidiques permet un agencement du peptidoglycane sous la forme d'un polymère réticulé flexible, le degré de réticulation variant en fonction de l'espèce bactérienne mais aussi des conditions rencontrées par la bactérie. Le peptidoglycane détermine la forme de la bactérie et protège les cellules contre les chocs osmotiques tout en permettant le passage de molécules de haute masse moléculaire (Demchick *et al.* 1996). De nombreuses enzymes sont capables de remodeler

le peptidoglycane participant à sa plasticité (Vollmer *et al.* 2008b). Les enzymes participant à la synthèse et au remodelage du peptidoglycane sont ainsi la cible de nombreux antibiotiques dont les β -lactamines et les glycopeptides, soulignant l'importance de l'intégrité du peptidoglycane pour la survie des bactéries.

La membrane externe est formée par une bicouche lipidique asymétrique et délimite l'espace périplasmique avec la membrane interne. Elle est propre aux bactéries Gram-négatives. Alors que le feuillet interne de la membrane externe est composé de phospholipides, son feuillet externe est majoritairement composé par le lipopolysaccharide (LPS).

2.2.2. Le lipopolysaccharide

Le LPS est un glycolipide constitué par le lipide A, un cœur oligosaccharidique et une partie polysaccharidique qualifiée d'antigène O (Figure 15). Le lipide A est un disaccharide de NAG lié en β 1 – 6, acylé en position 2 et 3, phosphorylé en position 1 et 4'. Le lipide A possède 6 chaînes aliphatiques, C12 ou C14, β hydroxylées dont 4 sont directement liées au NAG sur les positions 2 et 3, et 2 sont liées aux fonctions β hydroxyles des précédentes chaînes acyles. Les lipides β hydroxylées sont caractéristiques du lipide A qui est considéré comme la partie la plus toxique du LPS (Rietschel *et al.* 1994). La structure du lipide A et notamment la longueur et le degré de saturation des chaînes acyles peut varier en fonction de l'espèce bactérienne ou des conditions de croissance (Muller-Loennies *et al.* 2007).

Le cœur oligosaccharidique est divisé en deux parties. La partie interne est composée d'acide 3-déoxy- α -D-mannooctulosonique (Kdo) et de L-glycéro-D-mannoheptose (Hep) avec divers substituants tels que le rhamnose, le galactose, la glucosamine, la NAG, des résidus Kdo ou des groupements phosphates (Firdich *et al.* 2005). La région du cœur interne est riche en charges négatives apportées par ces phosphorylations et les résidus Kdo. Ces charges négatives jouent un rôle important en permettant la capture d'ions divalents ou de polyamines qui assurent la liaison entre les molécules de LPS (Nikaido *et al.* 1985). Ces charges sont également d'importants médiateurs de l'interaction entre le LPS et les charges positives des protéines de la membrane externe (OMP) (de Cock *et al.* 1999, Hagge *et al.* 2002). Enfin, des mutants présentant un cœur tronqué de la région interne ont montré une perméabilité plus grande de la membrane externe et une plus grande sensibilité aux composés hydrophobes avec l'apparition de phospholipides dans le feuillet externe et une diminution du nombre des OMP (Ames *et al.* 1974, Koplów *et al.* 1974, Smit *et al.* 1975). Une caractéristique de *Klebsiella pneumoniae* par rapport aux autres entérobactériacées est le remplacement de ces phosphorylations par des acides galacturoniques. La structure du cœur externe est plus variable avec une composition d'hexoses tels que le D-glucose, le D-galactose ou la D-glucosamine. Des résidus Kdo et

Hep sont également retrouvés dans le cœur externe chez *Klebsiella pneumoniae* (Firdich *et al.* 2005).

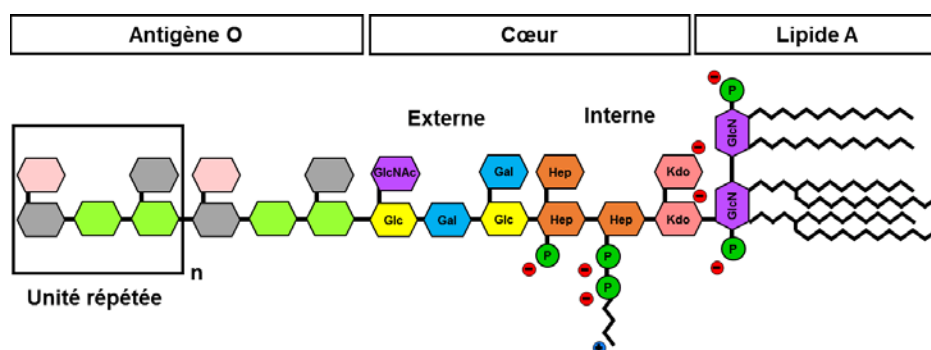


Figure 15. Structure générale du lipopolysaccharide (LPS) d'*Escherichia coli* constituant le feuillet externe de la membrane externe. Le lipide A forme la partie lipidique du LPS avec deux résidus phosphorylés de glucosamine (GlcN) possédant 6 chaînes acyle β hydroxylées. Le cœur oligosaccharidique comprend une partie interne riche en charges négatives composée d'acide 3-déoxy- α -D-mannooctulosonique (Kdo) et de L-glycero-D-mannoheptose (Hep) phosphorylés. La partie externe est composée d'hexoses de type glucoses (Glc), galactoses (Gal) ou N-acétylglucosamines (GlcNAc). La partie externe est appelée l'antigène O et est composée de motifs répétés variant en fonction de l'espèce et de la souche bactérienne (Muller-Loennies *et al.* 2007).

L'antigène O constitue la partie la plus externe du LPS. Il est formé par la répétition d'un motif oligosaccharidique de 2 à 8 monosaccharides (Stenutz *et al.* 2006). La chaîne polysaccharidique de l'antigène O peut contenir jusqu'à 50 unités répétées chez les entérobactériacées. L'antigène O constitue ainsi une véritable couche polysaccharidique autour de la bactérie qui la protège en empêchant l'accès à la membrane de nombreuses molécules. Ainsi, une perte de l'antigène O est associée à une diminution significative de la virulence due à une sensibilité accrue au complément. La composition de l'antigène O varie entre les espèces bactériennes et même au sein d'une même espèce en fonction de la souche. Cette variation de composition s'accompagne aussi de nombreuses modifications telles que des acétylations, des méthylations, des phosphorylations, des amidations ou des substitutions par divers acides aminés. Chaque type d'antigène O contient donc des groupements chimiques différents qui constituent autant d'épitopes reconnus par des anticorps spécifiques déterminant le sérotype d'une souche particulière (Muller-Loennies *et al.* 2007).

2.3. Les protéines de la membrane externe

2.3.1. La structure générale

La membrane externe inclue également de nombreuses protéines transmembranaires appelées « Outer Membrane Protein » (OMP) ayant diverses fonctionnalités. Alors que les protéines de la membrane interne sont majoritairement

constituées d'hélices α , les OMP sont formées par un nombre pair de brins β qui s'organisent en un tonneau β dont l'axe central est parallèle à la normale à la bicouche lipidique (ou légèrement incliné).

Les structures en tonneaux β sont retrouvées dans la membrane externe des bactéries Gram-négatives mais aussi dans les membranes externes des mitochondries et des chloroplastes (Galdiero *et al.* 2007). Une explication plausible de cette exclusivité des tonneaux β aux membranes externes est que ces protéines doivent passer une membrane interne avant de s'insérer dans la membrane externe. La structure en brin β évite la présence de domaines hydrophobes, caractéristiques des hélices α transmembranaires, au sein de la séquence primaire et qui pourraient bloquer la protéine dans la membrane interne (Koebnik *et al.* 2000). En effet, l'organisation en feuillet β implique une distribution alternative de résidus hydrophobes et hydrophiles dans la séquence primaire afin que tous les résidus d'un même type se retrouvent du même côté du feuillet (Figure 16) (Wimley 2003). Il est ainsi remarquable que la structure de tonneau β ne soit pas réservée aux protéines transmembranaires mais que des protéines solubles adoptent également cette structuration (Freer *et al.* 1970, Sawyer *et al.* 1978). Les protéines solubles organisées en tonneau β possèdent jusqu'à 8 brins β arrangés autour d'un cœur hydrophobe. Au contraire dans les OMP, le cœur de la protéine est hydrophile alors que les résidus hydrophobes sont exposés à l'extérieur du tonneau β . Les OMP comportent un nombre pair de brins β antiparallèles de 9 à 11 résidus légèrement inclinés de 20° à 40° par rapport à la normale de la bicouche lipidique (Sansom *et al.* 1995). Le repliement du feuillet β en tonneau β grâce à une interaction directe entre les deux brins β N- et C-terminal permet de s'assurer que tous les groupements polaires, amide ou carbonyle, du squelette peptidique sont engagés dans des liaisons hydrogènes y compris ceux enfouis au sein de la bicouche lipidique. Les extrémités N- et C-terminales de la chaîne peptidique sont périplasmiques et les connections entre les brins β sont des « turns » courts du côté périplasmique tandis que les boucles extracellulaires sont usuellement plus longues avec une plus grande variabilité de séquence. Le cœur de la protéine est polaire jusqu'à 12 brins β alors qu'une augmentation du nombre de brins β s'accompagne de l'apparition d'un canal (Galdiero *et al.* 2007).

Les acides aminés exposés aux lipides sont majoritairement apolaires tels que des phénylalanines, tyrosines, tryptophanes, valines et leucines (Ulmschneider *et al.* 2001). Les résidus aromatiques sont particulièrement présents au niveau de l'interface de la bicouche lipidique (Figure 16). L'intérieur de nombreuses OMP contient généralement un canal qui peut être obstrué par la présence d'une boucle ou d'un domaine formant un bouchon et dont les résidus sont principalement des acides aminés polaires de petite taille tels que des glycines, sérines, thréonines, asparagines ou glutamines.

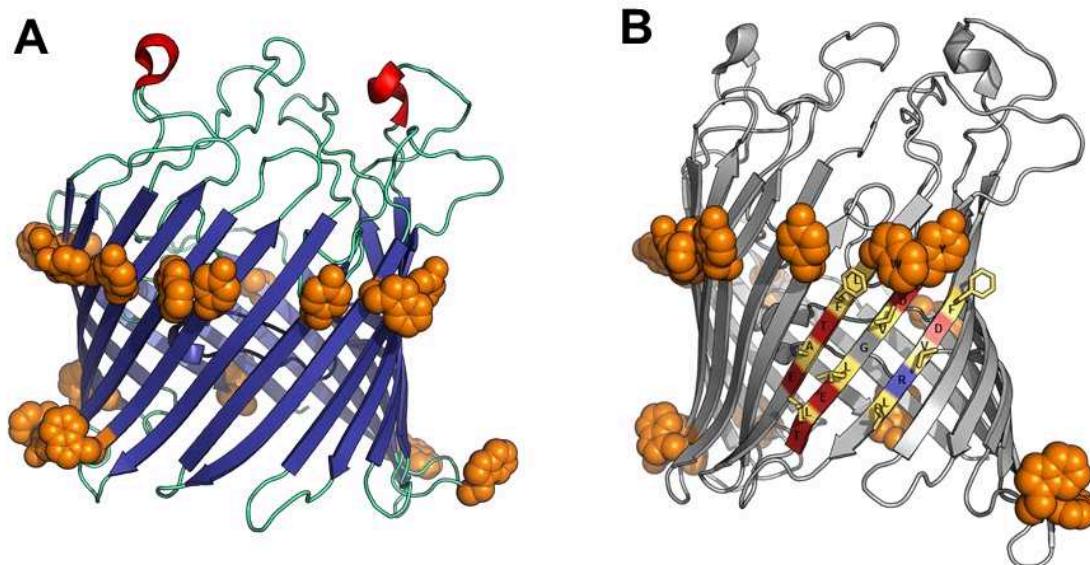


Figure 16. Structure de la porine LamB d'*Escherichia coli* (PDB ID : 1MAL). A. Les résidus aromatiques de type tryptophane et tyrosine, ici représentés en sphère orange, sont particulièrement présents au niveau de l'interface de la bicouche lipidique. B. Détail de l'organisation en dyade qui permet le feuillet β . Les résidus apolaires colorés en jaune (Phe, Leu, Val et Ala) sont orientés vers l'extérieur du tonneau β alors que les résidus polaires colorés en rouge (Glu, Asp, Thr) ou en bleu (Arg) sont orientés vers l'intérieur.

2.3.2. L'insertion et l'interaction avec les lipides

L'épaisseur de la partie hydrophobe du tonneau β des OMP est en moyenne de 24 Å, ce qui correspond à l'épaisseur du cœur hydrophobe de la membrane externe (Pogozheva *et al.* 2014). Cette partie hydrophobe est séparée des extrémités polaires de la protéine par deux ceintures de résidus tyrosine et tryptophane qui interagissent favorablement avec l'interface de la membrane (Killian *et al.* 2000). Alors que les groupements phénols des tyrosines se situent au niveau des groupes carbonyles des lipides de la membrane, les groupements indoles des tryptophanes sont enfouis 3 à 5 Å plus profondément vers le centre de la bicouche (Figure 17). Cette zone de la bicouche lipidique possède une polarité augmentée ainsi qu'une capacité en accepteur de liaisons hydrogène accrue comparativement au centre de la bicouche et est qualifiée de zone de « polarité moyenne » (Hanshaw *et al.* 2008). Contrairement à la membrane interne, au niveau de la membrane externe, les zones de polarité moyenne des feuillet interne et externe présentent une dissymétrie due à l'asymétrie dans la composition lipidique de ces feuillet et qui se reflète dans la distribution des résidus tyrosine et tryptophane (Pogozheva *et al.* 2014). Enfin, les surfaces polaires de la protéine incluent des résidus chargés négativement, aspartate et glutamate, ou positivement, lysine et arginine. La distribution des résidus chargés entre les faces polaires périplasmiques et extracellulaires varie avec la charge portée. Dans les protéines membranaires, les charges positives sont habituellement retrouvées du côté interne de la membrane plasmique, règle dite « positive

inside » (von Heijne 1992, Nilsson *et al.* 2005). Toutefois, dans les OMP, les résidus chargés positivement ont tendance à être retrouvés sur la face extracellulaire où il est possible qu'ils interagissent avec le LPS chargé négativement (Jackups *et al.* 2005). Au contraire, les résidus chargés négativement sont plus fréquents sur la face périplasmique (Pogozheva *et al.* 2013), où ils peuvent interagir avec les chaperonnes chargée positivement de l'espace périplasmique telles que Skp (Qu *et al.* 2009).

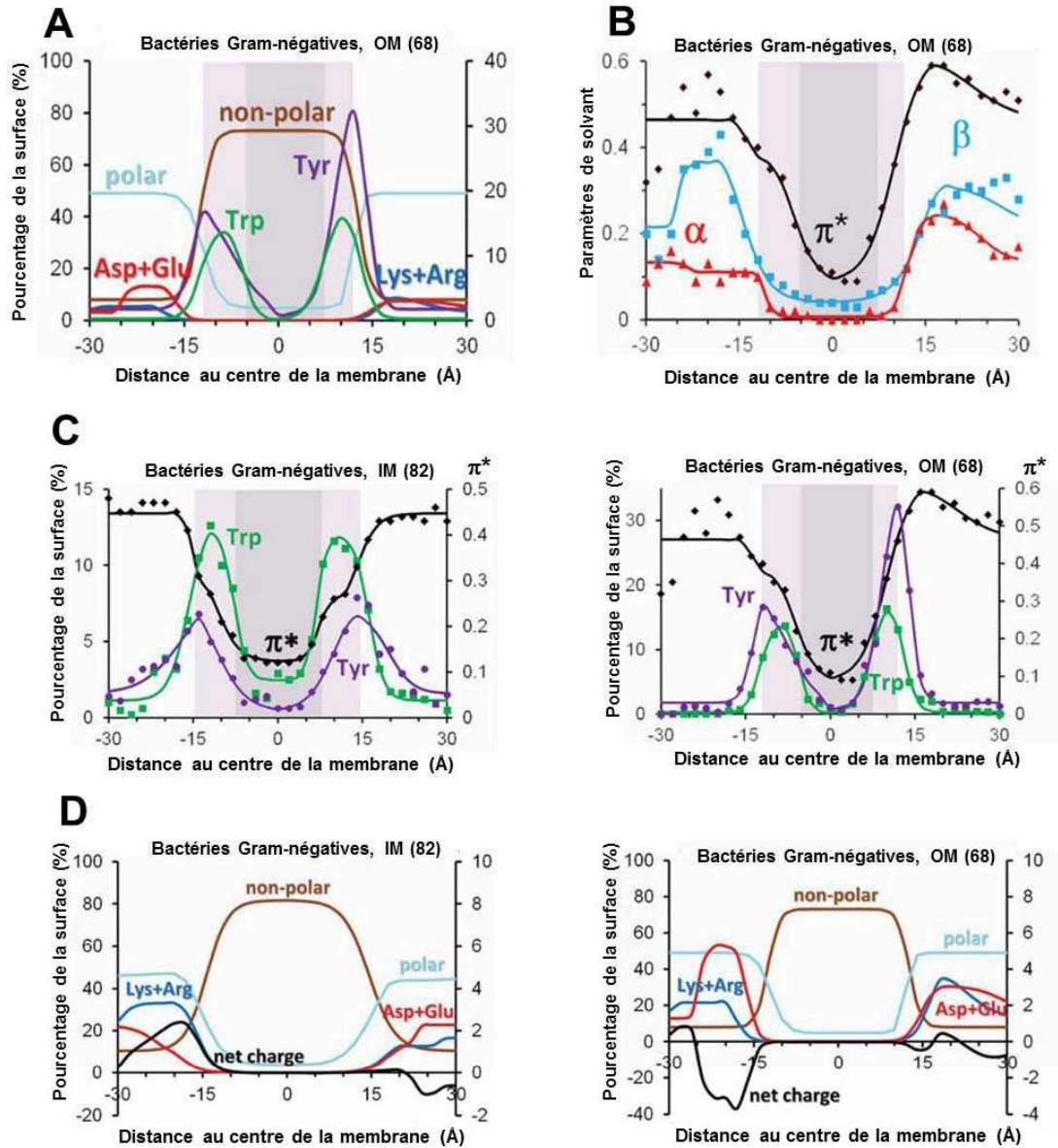


Figure 17. A. Distributions des atomes de 68 OMP en contact avec les lipides : atomes polaires (N et O de tous les acides aminés), atomes apolaires (C et S des résidus Val, Leu, Ile, Met, Cys, Phe, Tyr et Trp), groupes aromatiques (groupe benzène des Tyr et groupe indole des Trp) et groupes chargés négativement (Asp et Glu) ou positivement (Lys et Arg). B. Profil de paramètres de polarité calculés à partir des atomes en contact avec les lipides à la surface (donneur de liaisons hydrogènes, α ; accepteur de liaisons hydrogènes, β et polarisabilité, π^*) (Pogozheva *et al.* 2013). Comparaison des distributions entre 68 protéines membranaires de la membrane externe et 82 protéines membranaires de la membrane interne des groupes aromatiques en C. et des groupes chargés en D. Les régions de « polarité moyenne » colorées en gris clair ont été placées en suivant les zones d'accumulation des groupements indoles des tryptophanes. Le cœur hydrophobe est coloré en gris foncé. Adapté de (Pogozheva *et al.* 2014).

Un exemple des interactions entre les OMP et les lipides de la membrane externe est donné par la protéine FhuA dont la structure a été résolue par cristallographie en présence de LPS (Figure 18) (PDB ID : 1QFG). Un site de reconnaissance du LPS a été identifié. Les groupements phosphates du lipide A sont en interaction avec 8 résidus basiques (K280, K306, K351, R382, R384, K439, K441 et R474) et 3 résidus polaires (E304, Q353 et D386)

alors que les chaînes acyles interagissent avec 6 résidus aromatiques (F231, F235, Y284, F302, F355 et F380) (Ferguson *et al.* 2000, Adamian *et al.* 2011).

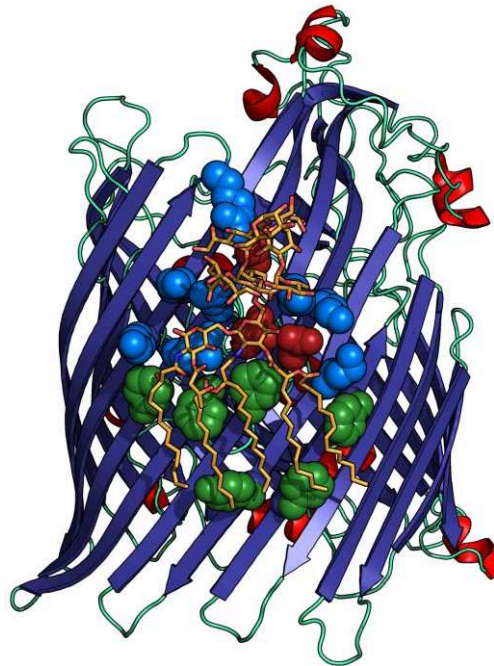


Figure 18. Structure de FhuA d'*Escherichia coli* avec une molécule de LPS, en orange (PDB ID : 1QFG) (Ferguson *et al.* 2000). Les résidus interagissant avec le LPS sont représentés sous forme de sphères. Les 8 résidus basiques (K280, K306, K351, R382, R384, K439, K441 et R474) et les 3 résidus polaires en interaction (E304, Q353 et D386) en interaction avec la partie polaire du LPS sont colorés en bleu et en rouge respectivement alors que les 6 résidus aromatiques (F231, F235, Y284, F302, F355 et F380) interagissant avec les parties acyles apolaires du LPS sont colorés en vert.

2.3.3. Les fonctions des OMP

Les OMP assurent un nombre important de fonctions au niveau de la membrane externe (Figure 19). Par leur exposition à la surface de la bactérie, les OMP jouent un rôle dans l'adhésion aux surfaces et l'interaction avec le système immunitaire. OmpX d'*Escherichia coli* joue un rôle important dans la virulence en neutralisant les mécanismes de défense de l'hôte (Heffernan *et al.* 1994). Composée de 8 brins β connectés par 3 turns périplasmiques et 4 courtes boucles extracellulaires, la principale caractéristique d'OmpX est l'allongement de 4 de ses brins β au-delà de l'interface de la membrane (Figure 20) (PDB ID : 1QJ9) (Vogt *et al.* 1999). Cette structure a été proposée comme support de la fonction d'adhésine et d'inhibition du système du complément. De façon intéressante, les boucles sont visibles dans la densité électronique, témoignant de la rigidité de la structure. L'importance de l'adhésion dans la pathogénicité est souligné par l'exemple d'OpcA de *Neisseria meningitidis* qui lie des protéoglycanes tels que l'héparine ou l'héparane sulfate et participe à l'internalisation du méningocoque dans les cellules épithéliales (PDB ID : 1K24) (Prince *et al.* 2002).

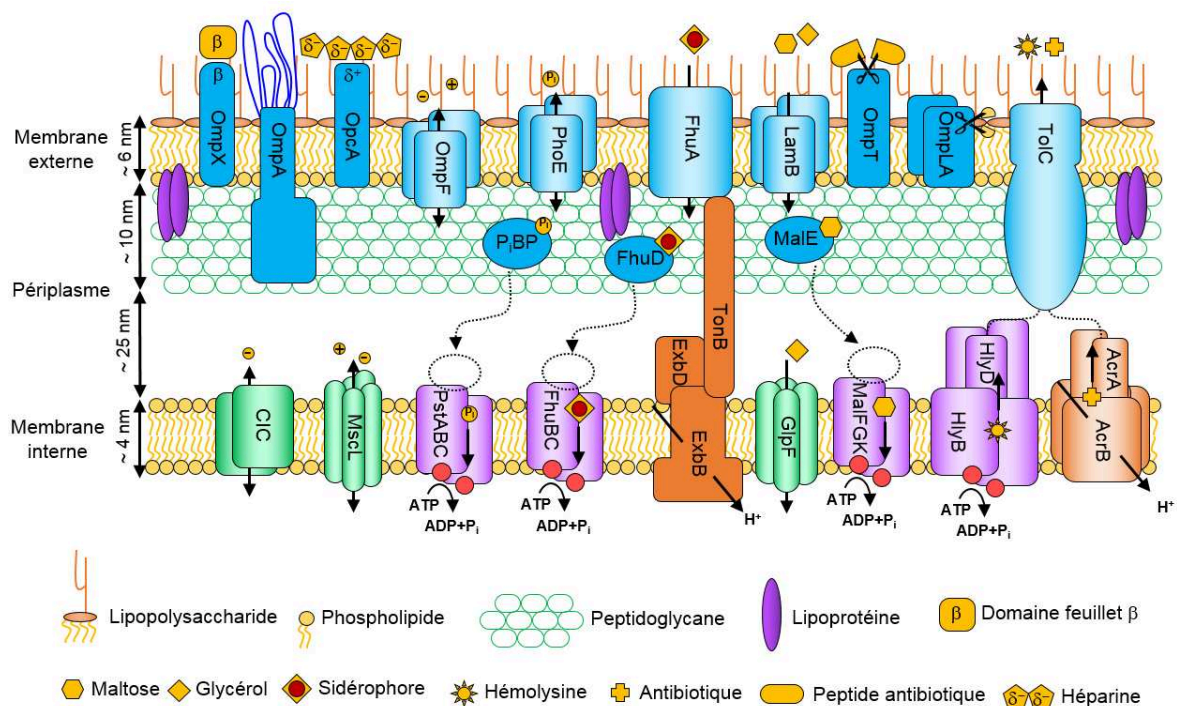
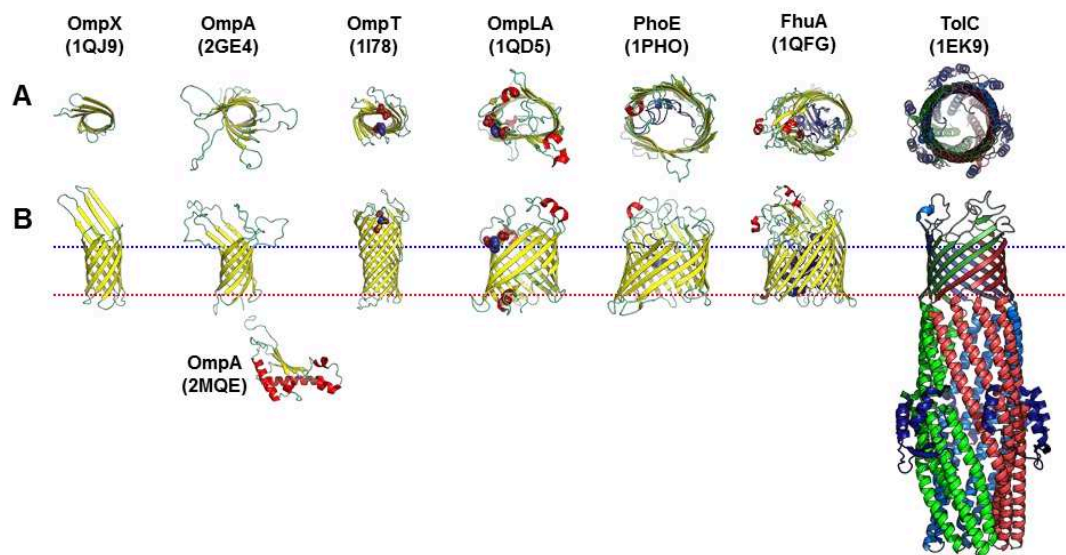


Figure 19. Fonctions biologiques d'OMP caractérisées structuellement. Les OMP permettent la diffusion des molécules à travers la membrane externe. Les ions diffusent passivement par des porines non-spécifiques telles qu'OmpF puis via des canaux ioniques au niveau de la membrane interne (CIC, MscL). Les sucres et les alcools passent par des porines spécifiques tels que LamB ou PhoE avant d'être liés à des protéines périplasmiques (MalE, PiBP) puis sont transférés dans le cytoplasme par des transporteurs ABC (MalFGK, PstABC) ou par des canaux (GlpF). Les sidérophores sont importés par des transporteurs actifs dépendant de TonB tels que FhuA avant de se lier à une protéine périplasmique (FhuD) et d'être importé dans le cytoplasme par un transporteur ABC (FhuBC). La sécrétion de toxines comme l'α-hémolysine ou de substances toxiques comme les antibiotiques est assurée par TolC couplée à des pompes d'efflux actives de la membrane interne (HlyDB, AcrAB). Les OMP assurent aussi des rôles d'adhésion aux cellules de l'hôte, OmpX et OmpA, de structure de l'enveloppe, OmpA ou encore des activités enzymatiques, OmpT et OmpLA. Adapté de (Faraldo-Gomez *et al.* 2003).

OmpA est une protéine majoritaire de la membrane externe et joue un rôle dans l'intégrité de l'enveloppe bactérienne. OmpA est constitué de deux domaines dont seul le domaine N-terminal est transmembranaire (Figure 20). Comme OmpX, OmpA d'*Escherichia coli* est composée de 8 brins β (PDB ID : 1QJP) (Pautsch *et al.* 2000). Toutefois les boucles extracellulaires sont hautement mobiles et n'ont pas de densité électronique visible dans la structure cristallographique. L'étude de la mobilité de ces boucles chez OmpA de *Klebsiella pneumoniae* a été au cœur de notre étude et sera décrite en détail au chapitre 4. Ces boucles pourraient jouer un rôle important dans l'interaction avec la réponse immunitaire (Confer *et al.* 2013). Enfin, le domaine C-terminal, dont la structure a été également résolue dans le cas de *Escherichia coli* (PDB ID : 2MQE) (Ishida *et al.* 2014), est localisé dans le périplasme où il joue un rôle dans l'interaction avec le peptidoglycane (Park *et al.* 2012).

Certaines OMP assurent des fonctions enzymatiques. Ainsi la protéase OmpT est un tonneau β constitué de 10 brins β et dont les boucles extracellulaires contiennent le site catalytique (Figure 20) (PDB ID : 1I78) (Vandeputte-Rutten *et al.* 2001). OmpLA possède également une activité enzymatique. Cette phospholipase est composée de 12 brins β formant un tonneau β dont le canal est obstrué par les extrémités C- et N-terminales ainsi que par les boucles L1, L4 et L6 (PDB ID : 1QD5) (Snijder *et al.* 1999). Son site actif est situé sur le côté du tonneau β , à hauteur des têtes polaires des lipides du feuillet externe de la membrane externe. Normalement monomérique, l'association en dimère permet la formation du site de liaison du substrat et d'un trou oxyanion fonctionnel. Elle pourrait être impliquée dans la sécrétion de colicine et de facteurs de virulence.

Les OMP ont aussi un rôle de porine permettant une diffusion passive de molécules hydrophiles à travers la membrane externe (Figure 19) (Nikaido 2003). Composés de 16 brins β , ces porines dites non spécifiques ont un canal qui permet à de petites molécules solubles et des ions de diffuser. Toutefois, la boucle L3 de ces porines n'est pas exposée à la surface extracellulaire mais va être repliée dans la lumière du canal où elle va former un goulet d'étranglement à mi-hauteur (Figure 20). La présence de résidus chargés à ce niveau forme un champ électrique transverse local qui constitue une barrière énergétique pour les solutés de faible polarité tels que de nombreux antibiotiques (Schulz 1994). En fonction de la charge portée par ces résidus, la porine va également avoir une sélectivité préférentielle pour les anions ou pour les cations. Ainsi la présence d'une lysine K125 dans la porine PhoE explique une sélectivité préférentielle pour les anions (PDB ID : 1PHO) (Cowan *et al.* 1992). Ces porines forment des homotrimères, chaque sous-unité ayant un canal transmembranaire. L'association sous forme de trimère induit la formation d'un cœur hydrophobe entre les 3 sous-unités (Gehring *et al.* 1989, Schulz 2002). La membrane externe comprend également des porines ayant une spécificité pour certains substrats. LamB est ainsi une porine spécifique des malto-oligosaccharides qui permet la diffusion passive du maltose (PDB ID : 1MAL) (Schirmer *et al.* 1995). Ces porines spécifiques forment également des homotrimères mais chaque monomère est constitué de 18 brins β au lieu de 16.



	OmpX	OmpA	OmpT	OmpLA	PhoE	LamB	FhuA	TolC
PDB	1QJ9	1QJP	1I78	1QD5	1PHO	1MAL	2FCP	1EK9
Brins β par monomère	8	8	10	12	16	18	22	4
Etat oligomérique	Monomère	Monomère	Monomère	Dimère	Trimère	Trimère	Monomère	Trimère
Nombre de canaux	-	-	-	-	3	3	1	1
Fonction principale	Adhésine	Structure	Enzyme	Enzyme	Porine	Transport passif	Transport actif	Efflux actif

Figure 20. Exemples de structures de OMP d'*Escherichia coli*. A. Vue de la face extracellulaire. B. Vue de la face latérale. Les limites approximatives de la membrane externe sont indiquées en pointillés, bleus pour la face extracellulaire et rouges pour la face périplasmique. Les brins β , les hélices α et les boucles sont colorés en jaune, rouge et vert respectivement à l'exception de TolC où les chaînes A, B et C du trimère ont été colorées en rouge, bleu et vert respectivement alors que les extrémités N- et C-terminales des 3 chaînes formant une ceinture autour du tonneau α sont colorées en bleu foncé. La boucle L3 de PhoE et le domaine N-terminal de LamB obstruant le canal ont également été colorés en bleu foncé. Pour OmpT et OmpLA, les acides aminés catalytiques ont été indiqués par des sphères bleues pour l'histidine et rouge pour l'aspartate/asparagine et la sérine. Le numéro PDB est indiqué pour chaque structure et toutes les structures sont des structures cristallographiques à l'exception de OmpA où les deux domaines sont des structures RMN étant donné que les boucles sont absentes des structures cristallographiques du domaine N-terminal (PDB ID :1QJP, 1BXW) et que le domaine C-terminal a été résolu uniquement par RMN. Dans le tableau, sont indiqués le numéro PDB de la structure cristallographique, le nombre de brins β par monomère, l'état oligomérique, le nombre de canaux et la fonction principale de chaque OMP.

Certains métabolites tel que le fer se trouvent en concentration nanomolaire dans les conditions physiologiques, à pH neutre et en présence d'oxygène, due à la rapide oxydation des ions Fe^{2+} en ions Fe^{3+} et à leur précipitation sous forme d'hydroxydes insolubles (Faraldo-Gomez *et al.* 2003). Des concentrations aussi faibles ne permettent pas un apport suffisant par simple diffusion passive. Certaines OMP sont ainsi capables d'assurer un transport actif à travers la membrane externe. C'est le cas de la protéine FhuA qui permet la capture de sidérophores, des molécules de faible poids moléculaire sécrétées par les bactéries qui chélatent le fer avec une haute affinité. Le transport actif contre un gradient de concentration nécessite de l'énergie, habituellement fournie par l'hydrolyse d'ATP ou par un gradient électrochimique transmembranaire. Toutefois ces sources

d'énergie ne sont pas disponibles au niveau de la membrane externe en raison de sa nature perméable. L'énergie nécessaire est fournie par le « complexe Ton » qui permet de coupler le transport actif à travers la membrane externe au gradient électrochimique de la membrane interne. Le complexe Ton est localisé au niveau de la membrane interne et est formé par la protéine membranaire ExbB et les protéines ExbD et TonB ancrées dans la membrane interne. FhuA est composé d'un domaine N-terminal qui adopte un repliement globulaire et d'un domaine C-terminal formant un tonneau β de 22 brins β (Figure 20) (PDB ID : 2FCP) (Ferguson *et al.* 1998). Le domaine N-terminal va obstruer le canal et former avec les boucles extracellulaires un site de liaison à haute affinité pour les sidérophores. Il a été montré chez BtuB, transporteur actif de la vitamine B12 (PDB ID : 1KMO, 1KMP), que la liaison du ligand entraînait un changement conformationnel du côté périplasmique d'une séquence « Ton box » qui permettait une interaction avec la protéine TonB (Merianos *et al.* 2000, Chimento *et al.* 2003, Postle *et al.* 2003). Il a été proposé que cette interaction avec TonB induise un changement conformationnel qui permettrait la translocation et la libération du ligand dans l'espace périplasmique (Faraldo-Gomez *et al.* 2003).

La membrane externe assure une protection des bactéries en empêchant de nombreuses substances tels que des enzymes, des antibiotiques ou des détergents, d'accéder à la membrane interne. Cette fonction de barrière est également un obstacle majeur pour la sécrétion et l'export de molécules par les bactéries Gram-négatives. La protéine TolC d'*Escherichia coli* permet l'export de nombreuses molécules hors de la cellule (Wandersman *et al.* 1990, Fralick 1996, Li *et al.* 2009). TolC est un trimère qui adopte la structure d'un tonneau β à 12 brins β antiparallèles au niveau de la membrane externe tandis que la partie périplasmique forme un tonneau constitué de 12 hélices α antiparallèles (Figure 20) (PDB ID : 1EK9) (Koronakis *et al.* 2000). Chaque monomère contribue à 4 structures secondaires, brins β ou hélice α , pour chacun des tonneaux. Les extrémités N- et C-terminales vont former une structure α/β qui va former une ceinture au niveau du tonneau α . TolC forme un canal avec un diamètre moyen de 19.8 Å et de 140 Å de long qui traverse la membrane externe et se projette dans l'espace périplasmique (Hinchliffe *et al.* 2013). Alors que le tonneau β est largement ouvert à la membrane externe, le canal de TolC est fermé à son extrémité périplasmique où 3 des 6 paires d'hélices α se tordent vers l'intérieur du canal. Les aspartates 374 vont alors former un détroit de 3.9 Å de diamètre qui ferme le canal de TolC aux substrats (Koronakis *et al.* 2000, Andersen *et al.* 2002). TolC interagit au niveau de la membrane interne avec un adaptateur périplasmique ancré dans la membrane interne et une protéine transmembranaire qui joue le rôle de transporteur actif tels que les couples HlyD-HlyB ou AcrA-AcrB (Thanabalu *et al.* 1998). Plusieurs homologues de TolC ont été retrouvés chez des pathogènes opportunistes tel que

Pseudomonas aeruginosa conférant à ces bactéries des capacités de tolérance accrue vis-à-vis d'un grand nombre de molécules (Hancock *et al.* 2002).

2.4. Biogenèse des protéines de la membrane externe.

2.4.1. Le passage de la membrane interne

Contrairement aux protéines transmembranaires à hélice α qui sont insérées dans la membrane au cours de la traduction, l'insertion dans la membrane externe est post-traductionnelle pour les OMP. Les OMP sont traduites dans le cytoplasme sous forme de précurseur avec une courte séquence signal en N-terminal. La chaperonne SecB maintient la chaîne peptidique sous forme non repliée (Figure 21) (Bechtluft *et al.* 2010). La chaîne peptidique est alors exportée à travers la membrane interne par l'ATPase SecA à travers le complexe transmembranaire SecYEG (de Keyser *et al.* 2003, Kusters *et al.* 2011).

2.4.2. Les voies des chaperonnes

Dans l'espace périplasmique, la chaîne peptidique est prise en charge par les chaperonnes Skp (Chen *et al.* 1996, Harms *et al.* 2001, Korndorfer *et al.* 2004, Burmann *et al.* 2013) et SurA, cette dernière possédant une activité peptidyl-prolyl isomérase et reconnaissant préférentiellement les séquences peptidiques riches en résidus aromatiques (Figure 21) (Lazar *et al.* 1996, Behrens *et al.* 2001, Bitto *et al.* 2002, Xu *et al.* 2007). Il a été montré que Skp assiste le repliement d'OmpA en présence de LPS (Schafer *et al.* 1999, Bulieris *et al.* 2003). De plus Skp est capable de lier le domaine N-terminal de OmpA non replié sans perturber le repliement correct du domaine C-terminale (Walton *et al.* 2009). SurA pourrait jouer un rôle prépondérant dans l'activité chaperonne au niveau du périplasme tandis que Skp aurait un rôle de voie de secours avec la protéine DegP. Ainsi les mutations SurA/Skp⁻ et SurA/DegP⁻ présentent un effet bactériostatique et bactéricide respectivement ce qui n'est pas le cas des mutations Skp/DegP⁻ (Rizzitello *et al.* 2001). Toutefois, l'importance de chacune des voies restent à déterminer ; ainsi il a été observé que les mutations Skp/DegP⁻ entraînaient un arrêt de la croissance dépendant de la température avec une accumulation de protéines mal repliées dans le périplasme (Schafer *et al.* 1999). Ainsi des interactions au sein d'un réseau plus complexe ou une action séquentielle de ces deux chaperonnes n'est pas exclue.

D'autres protéines sont susceptibles d'intervenir à ce stade telles que DsbA qui catalyse la formation de ponts disulfures ou la protéine DsbC qui réduit les cystéines ayant formé des ponts disulfures incorrects, les cystéines réduites par DsbC pouvant par la suite être oxydées par DsbA pour former les ponts disulfures corrects (Merdanovic *et al.* 2011). Ainsi, il a été démontré qu'une absence de DsbA s'accompagnait d'un défaut de formation

du pont disulfure présent dans OmpA et d'une cinétique de repliement plus longue (Bardwell *et al.* 1991).

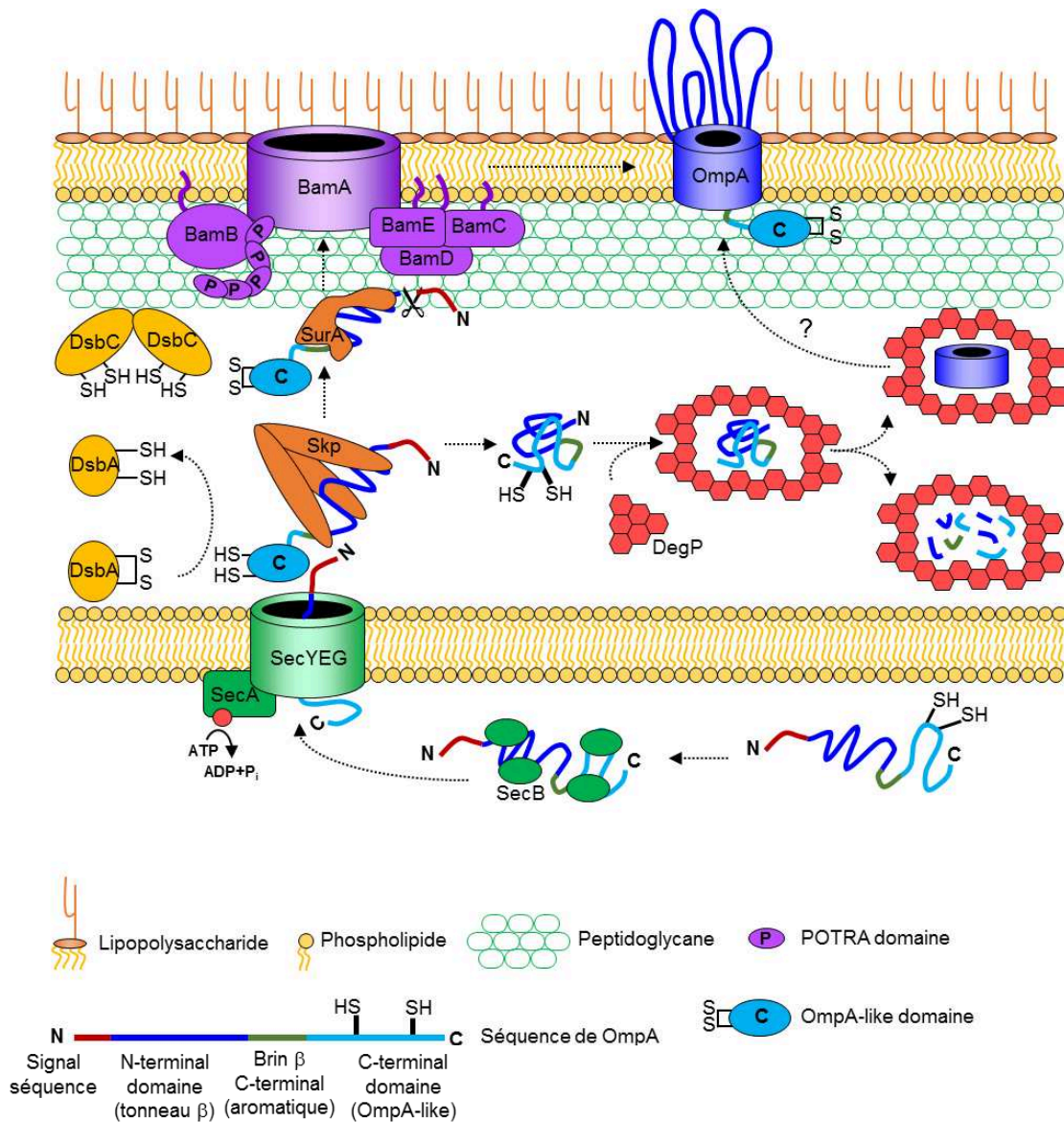


Figure 21. Modèle de la synthèse et de l'export des OMP dans la membrane externe. Au niveau cytoplasmique, les chaperonnes SecB prennent maintiennent la chaîne peptidique néo-synthétisée sous forme non repliée pour permettre son export par le système Sec dans le périplasme. Les chaperonnes périplasmiques SurA et Skp vont adresser la OMP à la machinerie BAM qui va assurer son insertion dans la membrane externe. Les protéines DsbA et DsbC catalysent la formation de ponts disulfures tandis que la protéine DegP prend en charge les OMP mal repliées présentes au niveau du périplasme.

Les OMP mal repliées présentes au niveau du périplasme sont prises en charge par la protéase DegP (Figure 21). Cette protéine possède une activité protéase et une activité chaperonne, le changement entre les deux fonctions étant contrôlé par la température (Spiess *et al.* 1999). Sous forme inactive, DegP forme un hexamère. La liaison d'un substrat mal replié entraîne la formation de 12-mer ou de 24-mer qui vont s'organiser

sous la forme d'une cage encapsulant le substrat dans sa cavité centrale (Jiang *et al.* 2008). La cavité centrale a ainsi une taille suffisante pour accueillir une OMP repliée (Krojer *et al.* 2008). De nombreuses OMP peuvent être substrat de DegP telles que OmpA, OmpC, OmpF ou LamB (Krojer *et al.* 2008). Toutefois, une fois correctement repliée, il n'est pas clair si la OMP va réintégrer la voie périplasmique des chaperonnes ou si elle va être insérée dans la membrane externe (Ricci *et al.* 2012). Ainsi, le rôle respectif des chaperonnes SurA, Skp et DegP dans les voies de prise en charge et de repliement des OMP au niveau du périplasme reste encore à définir (Sklar *et al.* 2007).

2.4.3. La machinerie BAM

L'insertion des OMP dans la membrane externe est assurée par la « β -Barrel Assembly Machinery » (BAM). Le complexe BAM est formé par BamA, une OMP composée d'un tonneau β transmembranaire et de 5 domaines « polypeptide translocation associated » (POTRA) périplasmiques placés en N-terminal avec lesquels vont venir interagir 4 lipoprotéines, BamB, BamC, BamD et BamE (Figure 21) (Wu *et al.* 2005, Kim *et al.* 2007, Ni *et al.* 2014). BamB et BamD interagissent directement avec BamA alors que BamC et BamE forment un sous-complexe avec BamD. Il a été montré que les lipoprotéines BamB et BamD étaient capables de lier BamA non repliée et de faciliter son repliement et son insertion dans la membrane *in vitro*. BamD lie également OmpA non repliée mais BamA est indispensable pour un repliement et une insertion correcte de OmpA (Hagan *et al.* 2013). Il a été également proposé que le dernier brin β du tonneau des OMP comportait un signal de reconnaissance et d'adressage à la membrane externe (de Cock *et al.* 1997). Ainsi, ce brin β se termine presque toujours par un résidu aromatique, une phénylalanine le plus souvent (de Cock *et al.* 1997). Alors que la mutation ou la délétion de ce résidu bloque l'assemblage *in vivo* et entraîne une accumulation des OMP au niveau du périplasme, ces variants se replient correctement dans une bicouche lipidique *in vitro* (Jansen *et al.* 2000, Misra *et al.* 2000).

La résolution des structures de BamA de *Neisseria gonorrhoea* et de *Haemophilus ducreyi* a mis en évidence la structure particulière du tonneau β transmembranaire (Noinaj *et al.* 2013). Le tonneau est constitué de 16 brins β antiparallèles formant un canal de 13000 Å³. L'intérieur de ce canal présente une surface fortement électro-négative. Du côté extracellulaire, le tonneau est fermé par les boucles extracellulaires L4, L6 et L7. Du côté périplasmique, alors que la structure chez *Haemophilus ducreyi* est pleinement ouverte sur le périplasme, le domaine POTRA 5 empêche tout accès dans la structure de *Neisseria gonorrhoea*. De plus, les brins β 1 et β 16 interagissent par 8 liaisons hydrogènes chez *Haemophilus ducreyi* alors qu'une seule liaison hydrogène est présente chez *Neisseria gonorrhoea* où le brin C-terminal se tord vers l'intérieur du tonneau β . Ces deux structures

pourraient refléter deux conformations de BamA où le domaine POTRA 5 et les brins $\beta 1$ et $\beta 16$ joueraient le rôle de porte avec le périplasma et la bicouche lipidique respectivement. La conformation retrouvée dans la structure de *Haemophilus ducreyi* serait alors la conformation ouverte sur le périplasma alors que la conformation retrouvée dans la structure de *Neisseria gonorrhoea* se rapprocherait plus d'une conformation fermée sur le périplasma et prête à s'ouvrir sur la bicouche lipidique. Des mécanismes similaires d'ouverture latérale sur la bicouche lipidique ont été décrits chez FadL, un transporteur de d'acide gras, ou PagP, une palmitoyl transférase, tous deux présents dans la membrane externe d'*Escherichia coli* (Hearn *et al.* 2009, Cuesta-Seijo *et al.* 2010).

2.5. La protéine « Outer membrane protein A ».

Comme mentionné précédemment, la protéine OmpA est constituée de deux domaines, un domaine N-terminal transmembranaire et un domaine C-terminal périplasmique.

La structure du domaine N-terminal d'*Escherichia coli* (N-EcOmpA) a été résolue par cristallographie (PDB ID : 1BXW et 1QJP) (Pautsch *et al.* 1998, Pautsch *et al.* 2000) et par RMN en solution (Figure 22) (PDB ID : 1G90 et 2GE4) (Arora *et al.* 2001, Cierpicki *et al.* 2006). Les 8 brins β antiparallèles du tonneau β forment un angle de 45° avec la normale de la bicouche lipidique et sont connectés par 3 courts turns du côté périplasmique et par 4 longues boucles du côté extracellulaire (Pautsch *et al.* 1998). La surface du tonneau est constituée de résidus apolaires et de résidus aromatiques, ces derniers étant retrouvés à la hauteur de l'interface de la bicouche lipidique. L'intérieur du tonneau est polaire et les chaînes latérales forment un réseau hydrogène dense. Malgré la présence de poches remplies par des molécules d'eau, aucun canal permettant une diffusion de molécules n'est visible. Un pont salin entre un glutamate (E52) et une arginine (R138) flanqué par une phénylalanine (F40) et une tyrosine (Y94) constitue la barrière principale à une éventuelle diffusion de molécules dans le tonneau. Le faible facteur B cristallographique témoigne de la rigidité des chaînes latérales participant au réseau d'interaction à l'intérieur du tonneau. Alors que le tonneau présente une forte structuration, les boucles présentent une mobilité élevée. Ainsi, les boucles L1, L2 et L4 n'ont pas de densité électronique dans la structure cristallographique (Pautsch *et al.* 1998).

La structure RMN de OmpA a permis de retrouver les principaux éléments mis à jour par la cristallographie (Figure 22) (Arora *et al.* 2001). Ainsi, OmpA solubilisée en dodécylphosphocholine (DPC) adopte le même repliement que celui trouvé dans la structure cristallographique en présence du n-octyltétraoxyéthylène. Les boucles extracellulaires présentent une grande variabilité due à l'absence de contraintes

structurales à ce niveau (Arora *et al.* 2001). Toutefois, les brins β de la structure RMN sont de 1 à 3 résidus plus courts que dans la structure cristallographique. De plus la RMN permet de mettre en évidence « un gradient de mobilité » se répartissant du centre du tonneau β jusqu'au niveau des boucles avec des mouvements μ s – ms au niveau des extrémités du tonneau β qui se traduisent par une diminution de l'intensité et un élargissement du signal à ce niveau (Arora *et al.* 2001).

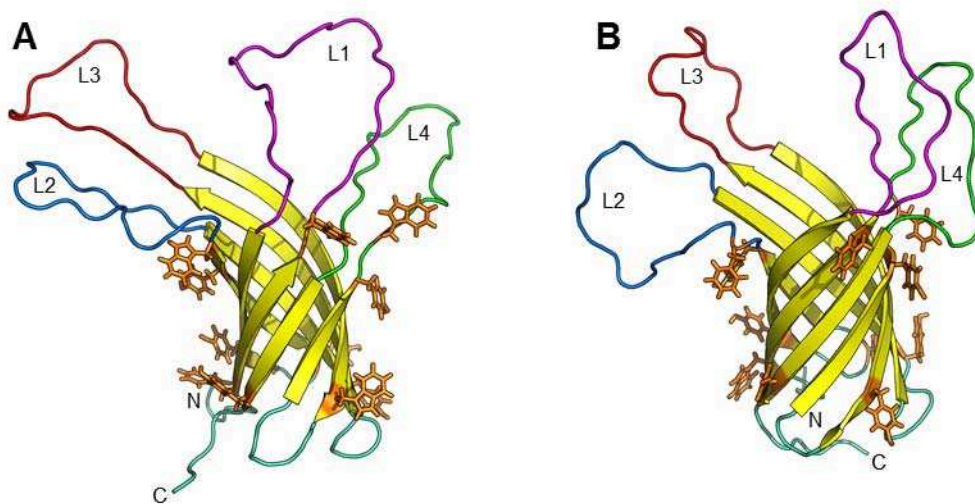


Figure 22. Structure par RMN de OmpA. A. OmpA de *Klebsiella pneumoniae* en micelle de DHPC (PDB ID : 2K0L) B. OmpA d'*Escherichia coli* en micelle de DPC (PDB ID : 2GE4). Les résidus aromatiques situés à hauteur de l'interface de la bicouche lipidique sont colorés en orange. Les quatre boucles extracellulaires ainsi que les extrémités périplasmiques N- et C-terminales sont indiquées.

La structure RMN du domaine N-terminal de *Klebsiella pneumoniae* (N-KpOmpA) a été résolue dans l'équipe de RMN de l'IPBS en 2009 (Figure 22) (PDB ID : 2K0L) (Renault *et al.* 2010). Cette structure a été résolue par RMN en présence de DHPC et avec la séquence native de la protéine alors que l'introduction de mutations avaient été nécessaires dans les études cristallographiques et RMN de EcOmpA. De plus, les boucles L1 et L3 sont plus longues chez *Klebsiella pneumoniae* que chez *Escherichia coli*. La structure et la dynamique en solution de KpOmpA sont très semblables à celles mise en évidence chez *Escherichia coli* (Arora *et al.* 2001, Renault *et al.* 2010). Notamment le gradient de mobilité a été retrouvé et caractérisé plus en détail chez KpOmpA. L'échelle de temps des mouvements de KpOmpA en DHPC va ainsi de la semaine (observation de l'échange de protons amides) au cœur du tonneau β à la nanoseconde au niveau des boucles en passant par la milliseconde au niveau de l'interface de la bicouche lipidique (observation d'élargissements de raies sur les spectres HSQC). Ce gradient de mobilité est aussi présent

quand KpOmpA est inséré dans une bicouche lipidique, un environnement plus proche de la membrane externe que les détergents (Iordanov *et al.* 2012). Après reconstitution en protéoliposome de KpOmpA, des mouvements intermédiaires de type μs – ms ont été mis en évidence au niveau de l'interface entre les boucles et le tonneau β par RMN du solide. De plus cette étude a montré que toutes les régions des boucles n'avaient pas la même mobilité (Renault *et al.* 2009, Iordanov *et al.* 2012). Les boucles L2, L4 et les parties N-terminales de L1 et L3 présentent également une mobilité restreinte avec des caractéristiques dynamiques typiques des mouvements intermédiaires. Seules les parties C-terminales de L1 et L3 présentent des caractéristiques dynamiques de boucle avec des mouvements rapides de type nanoseconde. Les boucles extracellulaires de KpOmpA jouent un rôle important dans les propriétés immunologiques de la protéine telle que la reconnaissance médiée par les récepteurs LOX-1 et SREC-1 ou encore la reconnaissance humorale médiée par PTX3 (Jeannin *et al.* 2005, Smith *et al.* 2007, Confer *et al.* 2013). Ayant un rôle dans la régulation de la réaction hôte-pathogène, la dynamique de ces boucles est susceptible de moduler cette réaction en intervenant lors de l'interaction avec différents ligands.

La présence d'un second domaine en C-terminal susceptible de se lier au peptidoglycane laisse penser que OmpA pourrait avoir un rôle structural dans la cohésion de l'enveloppe bactérienne, le domaine N-terminal ayant une fonction d'ancrage dans la membrane externe. La réponse du domaine transmembranaire à une tension mécanique a été étudiée dans l'équipe de RMN de l'IPBS en collaboration avec D.J. Muller, ETH Bâle, par microscopie à force atomique (AFM) et spectrométrie de forces sur molécule unique (SMFS) (Bosshart *et al.* 2012). Ainsi, sous l'effet d'une force de traction appliquée au domaine C-terminal, le domaine N-terminal de KpOmpA va se déplier par étapes successives. Les brins $\beta 8$ – $\beta 7$ – $\beta 6$ forment la première étape, suivis par les brins $\beta 5$ - $\beta 4$ lors de la deuxième étape, les brins $\beta 3$ – $\beta 2$ dans une troisième étape et le brin $\beta 1$ en dernière étape. De façon intéressante, la force nécessaire pour extraire le brin $\beta 1$ est deux à quatre fois plus importante que pour les étapes précédentes. Enfin si le brin $\beta 1$ n'est pas extrait et que la force appliquée est relâchée, N-KpOmpA est capable de se renaturer dans la membrane comme en témoigne le fait que les profils de dépliement obtenus après repliement sont identiques à ceux obtenus avant repliement. KpOmpA pourrait ainsi jouer un rôle d'absorbeur de stress mécanique et y répondre en se dépliant par étapes puis en se repliant une fois celui-ci passé.

Le domaine C-terminal de OmpA est commun à de nombreuses autres protéines périplasmiques des bactéries Gram-négatives. Ce domaine est désigné comme « OmpA domain » ou « OmpA-like domain » dans les bases de données PFAM et PROSITE

respectivement et suit un repliement de type $\beta/\alpha/\beta/\alpha/\beta$ (Figure 23) (Sigrist *et al.* 2013, Finn *et al.* 2014). Parmi les protéines comportant un domaine OmpA en C-terminal, on retrouve des porines telles que OmpA mais aussi des lipoprotéines associées au peptidoglycane (Pal) ou des protéines du moteur flagellaire telles que MotB. Chez les porines, les deux domaines sont reliés par une séquence riche en proline comprenant un motif formé par la répétition à quatre reprises du tandem AP chez *Klebsiella pneumoniae* OmpA, *Escherichia coli* OmpA et *Salmonella enterica* OmpA. Le domaine C-terminal soluble se situe au niveau périplasmique où il interagit avec le peptidoglycane (Figure 23) (Parsons *et al.* 2006, Park *et al.* 2012). Plusieurs structures sont disponibles dont celles de *Neisseria meningitidis* (PDB ID : 1R1M), *Escherichia coli* (PDB ID : 2MQE) ou *Salmonella enterica* (PDB ID : 4ERH) (Grizot *et al.* 2004, Ishida *et al.* 2014). Bien que présentant des identités de séquence de l'ordre de 20-30%, les domaines OmpA-like présentent une remarquable conservation de leur structure tertiaire.

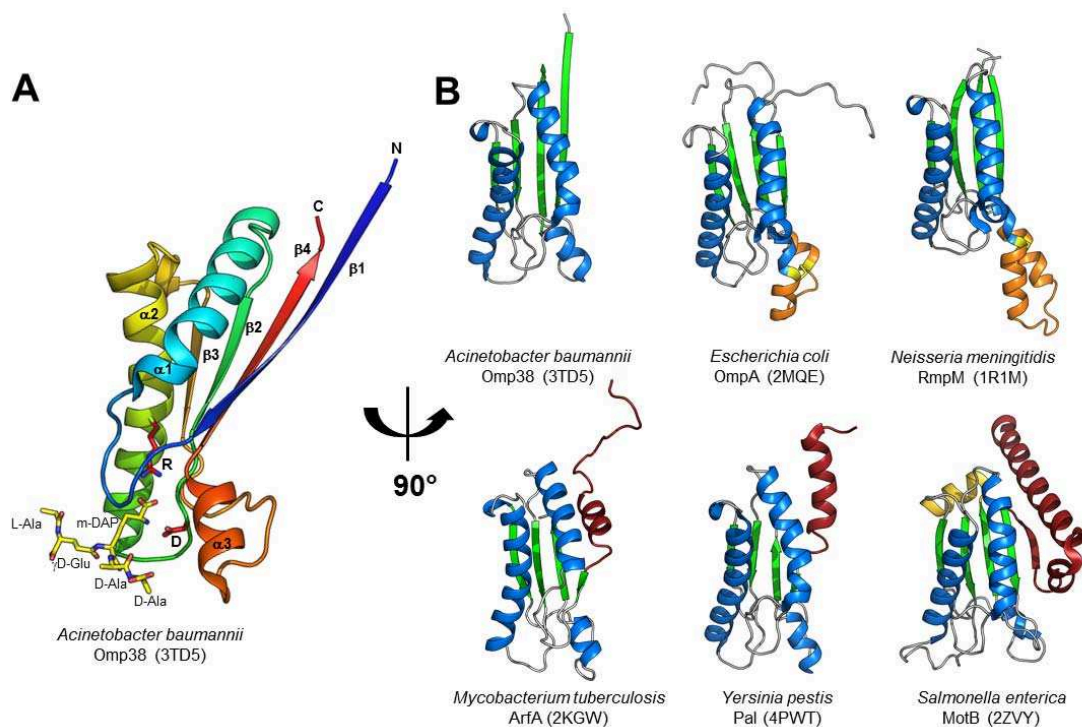


Figure 23. Domaine OmpA-like. A. Structure cristallographique du domaine C-terminal de Omp38 d'*Acinetobacter baumannii* en complexe avec un pentapeptide synthétique retrouvé dans le peptidoglycane. Les deux acides aminés importants pour l'interaction avec le DAP, une arginine et un aspartate, ont été représentés en rouge. B. Structures de différents domaines OmpA-like. Le cœur du domaine est coloré en vert, bleu et gris pour les brins β , les hélices α et les boucles respectivement. Les variations structurales et les structures supplémentaires en N-terminal sont colorées en rouge et en jaune pâle du côté C-terminal. L'élongation de la boucle $\beta 3 - \alpha 3$ et le pont disulfure retrouvés dans certains domaines sont colorés en orange et en jaune respectivement. Les identifiants PDB de chaque structure sont indiqués.

Les principales variations se situent au niveau des extrémités N- et C-terminales du domaine où des brins β et des hélices α supplémentaires peuvent être présentes comme dans MotB de *Helicobacter pylori* (PDB ID : 1CYQ) ou de *Salmonella enterica* (PDB ID : 2ZVY) (Kojima *et al.* 2008, Roujeinikova 2008). Le brin β 1 présente également une grande variabilité de structure et peut-être remplacé par une hélice α comme chez *Mycobacterium tuberculosis* (PDB ID : 2KGW) ou dans les Pal telle que celle de *Yersinia pestis* (PDB ID : 4PWT) (Yang *et al.* 2011). La partie la plus variable du cœur du domaine OmpA-like se situe entre le brin β 3 et l'hélice α 3 où la longueur de la boucle varie en fonction du domaine considéré. En général longue de 8 – 10 résidus, cette boucle est allongée à 15 résidus chez *Klebsiella pneumoniae*, *Escherichia coli* et *Salmonella enterica*, *Neisseria meningitidis* ayant la boucle la plus longue avec 26 résidus qui s'organisent pour former une hélice α supplémentaire (Figure 23) (Grizot *et al.* 2004). De façon intéressante, les domaines OmpA-like présentant une boucle β 3 – α 3 possèdent un pont disulfure absent dans les autres séquences homologues ce qui suggère que la structuration de cette partie du domaine est importante. Plusieurs travaux ont suggéré que ces domaines OmpA-like pouvaient former des dimères, *in vitro* pour *Escherichia coli* OmpA, ou *in vivo* pour *Salmonella enterica* MotB (Kojima *et al.* 2008, Kojima *et al.* 2009, Marcoux *et al.* 2014). En plus du rôle d'ancrage du peptidoglycane, il a été montré que ce domaine pouvait jouer un rôle de chaperonne lors du repliement de OmpA en réduisant les phénomènes d'agrégation dus au domaine N-terminal non replié (Danoff *et al.* 2011).

La structure et la dynamique du domaine C-terminal de OmpA de *Klebsiella pneumoniae* ont été caractérisés par cristallographie et par RMN dans le cadre de cette thèse.

III. Dynamique des protéines

3.1. Méthodes d'analyse de la dynamique fonctionnelle des protéines

Il existe une grande variété de méthodes biophysiques et spectroscopiques permettant d'appréhender la dynamique fonctionnelle de macromolécules biologiques. Nous en citerons quelques-unes, sans chercher nullement à être exhaustifs, mais simplement comme point d'entrée vers une littérature extrêmement riche. Au-delà de la RMN que nous développerons plus loin, et de manière souvent complémentaire avec elle, les méthodes de diffraction des rayons X (Putnam *et al.* 2007, Hennig *et al.* 2014) ou des neutrons (Reat *et al.* 1998, Gabel *et al.* 2002) ont depuis longtemps été utilisées pour cela. La cristallographie aux rayons X, bien que principalement utilisée pour fournir des structures 3D à haute résolution apporte des renseignements précieux sur les équilibres

conformationnels ne serait-ce qu'en fournissant plusieurs conformations différentes possibles. Un exemple majeur dans le domaine des protéines membranaires a consisté à la résolution de structure par l'équipe du prix Nobel de Chimie 2012, B. Kobilka, de la structure 3D du récepteur β_2 adrénergique dans son état « inactif », stabilisé par un antagoniste (Rasmussen *et al.* 2007), puis de la structure du même récepteur dans son état « actif », en présence d'un agoniste et d'une protéine G hétérotrimérique (Rasmussen *et al.* 2011). Les avancées récentes de la cristallographie, en particulier grâce au XFEL (sources laser femto seconde ultra puissantes) (Marx 2014), permettent d'envisager dans un avenir proche une cristallographie dynamique à température ambiante fournissant non plus des images mais bien des films décrivant un équilibre conformationnel ou un processus réactionnel (Feld *et al.* 2014). Une autre technique majeure de la biologie structurale moderne est bien entendu la microscopie électronique qui connaît ces dernières années des progrès fulgurants en termes de résolution (la résolution atomique devenant possible dans des cas favorables) ainsi que de diversité des complexes étudiables, sans limite de taille ou de complexité (Bai *et al.* 2015). Un point fort de cette approche est la possibilité de sélectionner des particules au sein d'un mélange hétérogène afin soit de le décrire dans sa diversité, soit d'améliorer la résolution en écartant les particules mal repliées. Dans un exemple issu des travaux de l'IPBS (équipe de biophysique structurale de L. Mourey), la combinaison d'approches cristallographiques et de cryo-EM a permis de mettre en évidence la dynamique de formation et la diversité d'états des γ Turc qui sont des complexes d'initiation de la formation de microtubules de tubuline (Guillet *et al.* 2011, Kollman *et al.* 2011, Kollman *et al.* 2015). Les progrès récents de la microscopie à force atomique, en particulier dans sa version à haute fréquence, permettent de mettre en évidence et de quantifier des événements dynamiques tels que mouvements de domaines ou les mécanismes de dénaturation de protéines (en utilisant la spectroscopie de force sur molécule unique, SMFS) (Borgia *et al.* 2008, Eghiaian *et al.* 2014). Enfin, sans les détailler, nous citerons de nombreuses autres méthodes spectroscopiques telles que la fluorescence, la résonance paramagnétique électronique, les spectroscopies Infra-Rouge, Raman, ou Mössbauer qui chacune contribue à l'élaboration d'une description dynamique des systèmes.

3.2. RMN et dynamique

Les paramètres RMN sensibles à des changements de structure de protéines d'une manière dépendante du temps (Figure 24) sont les phénomènes de relaxation de spin dans le repère du laboratoire (Palmer 2015, Torchia 2015) (en particulier la relaxation des paires ^{15}N - ^1H pour les liaisons peptidiques) (Palmer 2015), l'analyse quantitative fine des effets

Overhauser nucléaires (Vogeli 2014), la dispersion de relaxation (Palmer *et al.* 2006, Hansen *et al.* 2008, Ban *et al.* 2013) et l'analyse des formes de raies (Palmer *et al.* 2006, Palmer 2014), la relaxation paramagnétique ou PRE (Clore 2013), l'analyse des couplages dipolaires résiduels (Chen *et al.* 2012, Rezaei-Ghaleh *et al.* 2012) ou des échanges hydrogène deutérium (Krishna *et al.* 2004, Wildes *et al.* 2004). La combinaison de ces différentes techniques permet de caractériser les mouvements dans des échelles de temps allant de la pico seconde (10^{-12} s) à la seconde ou plus. Au-delà de la seconde, des expériences d'acquisition 2D rapides permettent un suivi en temps réel des phénomènes cinétiques (Gal *et al.* 2009). Il convient aussi de souligner que le champ d'application de la RMN va au-delà de l'étude des molécules purifiées en solution et / ou reconstituées en membranes, mais s'étend aujourd'hui en direction de la RMN *in cellulo* (Hansel *et al.* 2014) et de la RMN de protéines dans leur environnement natif (Renault *et al.* 2012a, Renault *et al.* 2012b, Schanda *et al.* 2014).

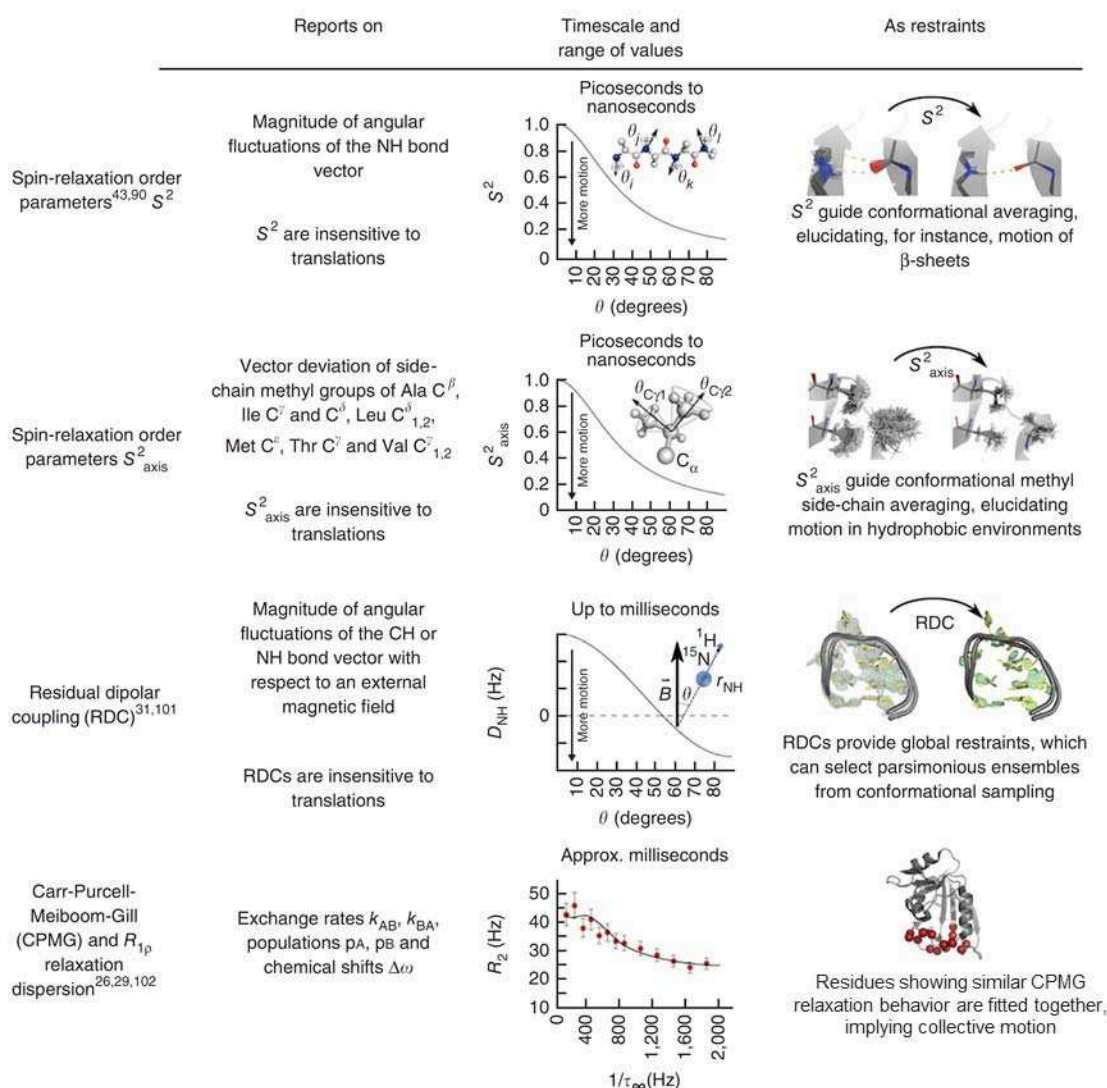


Figure 24. Présentation des principales méthodes de relaxation RMN utilisées et de la nature des mouvements qu'elles permettent d'appréhender. Adapté de (van den Bedem *et al.* 2015)

Le marquage uniforme ^2H , ^{15}N et ^{13}C des protéines (Goto *et al.* 1999, Lundstrom *et al.* 2009) permet ainsi d'établir des profils dynamiques tout le long de squelette peptidique pour des protéines de taille moyenne (200 à 300 acides aminés). Au-delà, il est possible de caractériser des complexes macromoléculaires de très haute masse moléculaire, jusqu'à 1 MDa, au moyen de marquages spécifiques en particulier des méthyls (Tugarinov *et al.* 2007, Ayala *et al.* 2009, Ruschak *et al.* 2010a, Kay 2011). Ceci a pu être réalisé par exemple sur le protéasome par l'équipe de L. Kay (Ruschak *et al.* 2012) ou sur des protéines membranaires (Wang *et al.* 2014b) telles que le récepteur $\beta 2$ -adrénergique solubilisé en micelle de détergents (Nygaard *et al.* 2013).

Dans ce qui suit, nous avons choisi d'illustrer de manière plus détaillée les apports de la RMN à l'étude de la dynamique de trois systèmes, le protéasome, le complexe CAP-

ADN et le lysozyme du phage T4. Nous avons choisi ces deux exemples car ils nous semblent particulièrement révélateurs des potentialités extraordinaires de la RMN moderne.

3.2.1. Le protéasome

Le protéasome est un assemblage supramoléculaire intracellulaire chargé de la dégradation des protéines ubiquitinylées. Son étude est un excellent exemple de ce que peut apporter la RMN lors de l'étude de complexes supramoléculaires (Rosenzweig *et al.* 2014). La particule 20S du protéasome est un complexe de 670 kDa formé par 4 anneaux heptamériques ayant une organisation $\alpha_7\beta_7\beta_7\alpha_7$ (Figure 25B). Les deux anneaux internes β_7 délimitent la chambre catalytique alors que les deux anneaux externes α_7 forment les entrées du protéasome. Deux chambres intermédiaires situées entre les anneaux α_7 et β_7 stockent les substrats avant leur dégradation dans la chambre interne. Les anneaux α_7 assurent également l'interaction entre la particule 20S, « cœur du protéasome », et d'autres particules régulatrices telles que la particule 19S reconnaissant les substrats portant une signature poly-ubiquitine ou la particule 11S impliquée dans la réponse immunitaire (Figure 25C) (Ruschak *et al.* 2011).

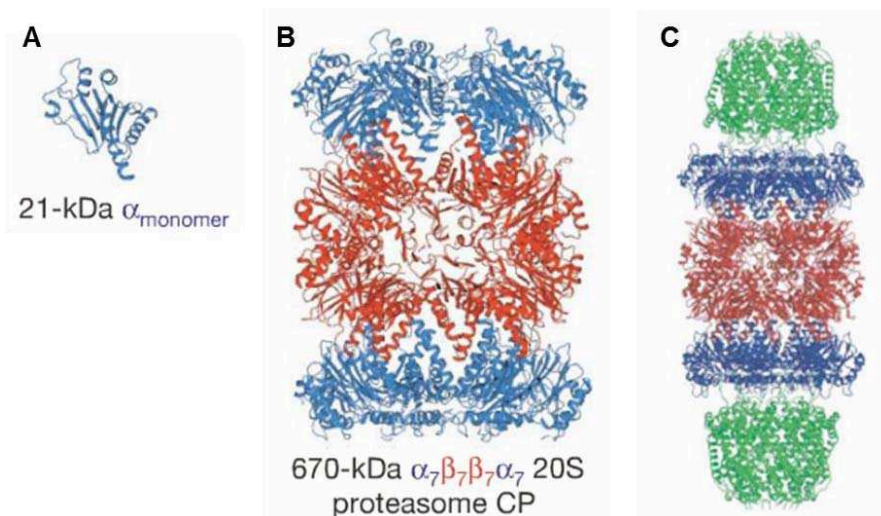


Figure 25. Structure du protéasome. A. Structure d'une sous-unité α . B. Assemblage du cœur du protéasome $\alpha_7\beta_7\beta_7\alpha_7$. Les sous-unités β formant la chambre catalytique sont colorées en rouge et les sous-unités α formant la porte et la chambre intermédiaire sont colorées en bleu. C. Structure du cœur du protéasome avec la particule régulatrice 11S colorée en vert. Adapté de (Sprangers *et al.* 2007).

Par un marquage spécifique des groupements méthyls, l'équipe de L. Kay a pu étudier la dynamique de ce complexe (Tugarinov *et al.* 2007). Les anneaux α_7 externe forment une entrée de 13 Å de diamètre par laquelle une chaîne polypeptidique non repliée accède à la chambre intermédiaire (Figure 26A et B). Il a pu être montré que les extrémités

N-terminales des sous-unités α présentaient des mouvements à l'échelle de la milliseconde et étaient capable de passer dans la chambre intermédiaire (Sprangers *et al.* 2007). L'utilisation de sonde paramagnétique a permis de calculer des contraintes de distance et d'établir que l'extrémité N-terminale pouvait adopter deux ensembles conformationnels, l'un où les 7 à 10 premiers acides aminés se situent à l'intérieur de la chambre intermédiaire, et l'autre où ces acides aminés sont situés à l'extérieur (Figure 26C) (Religa *et al.* 2010). De plus, sur les 7 extrémités N-terminales des sous-unités α de l'anneau $\alpha 7$, en moyenne 2 extrémités sont dans une conformation les plaçant à l'intérieur. Les deux états « interne » et « externe » sont en équilibre et s'échangent à l'échelle de la seconde avec une durée de vie de 2 s et 6.5 s respectivement à 45°C. Enfin, un déplacement de l'équilibre en faveur de la forme ouverte, par la mutation E25P ou par l'association avec la particule régulatrice 11S, s'est révélé être associé à une activité protéolytique accrue. L'étude des mouvements des extrémités N-terminales dans des milieux de viscosité variable a permis de mettre en évidence l'importance des collisions avec les molécules d'eau dans le passage entre les conformations « interne » et « externe » (Latham *et al.* 2014). Ce changement conformationnel s'effectue par étapes impliquant des segments inférieurs à 4 Å et l'équilibre conformationnel ainsi que sa cinétique ne sont pas influencés par la forte concentration en protéine présente dans un lysat cellulaire.

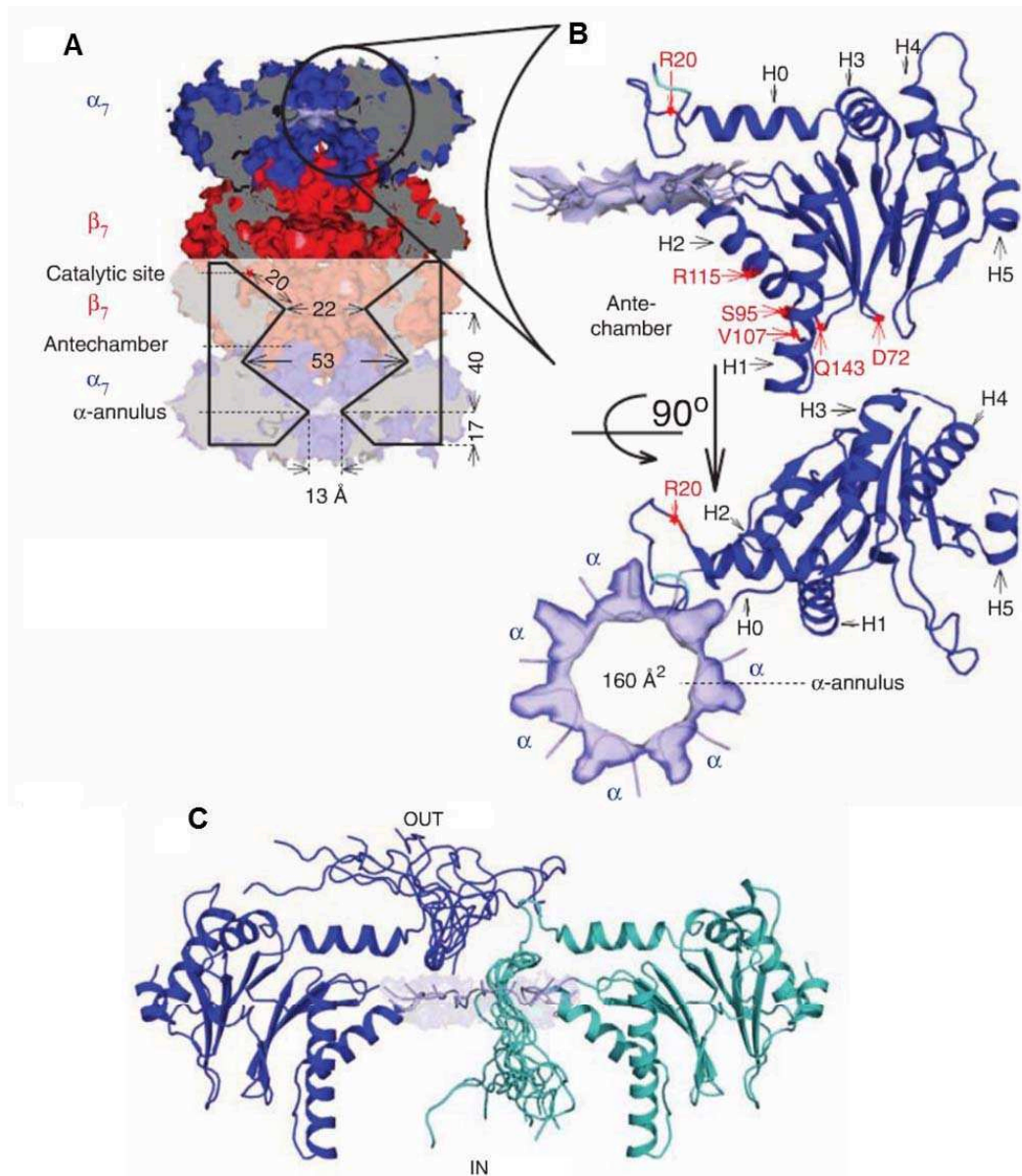


Figure 26. Conformation des résidus formant la porte du protéasome. A. Section du protéasome montrant la position de la chambre intermédiaire et de la chambre catalytique. B. Vue rapprochée d'une sous-unité α et de l'entrée de la chambre intermédiaire du protéasome. C. Ensemble de 10 structures montrant l'extrémité d'une sous-unité α en conformation externe (en bleu) ou interne (en cyan). Adapté de (Religa *et al.* 2010).

La RMN a aussi permis d'étudier l'état du substrat dans la chambre intermédiaire (Ruschak *et al.* 2010b). Ainsi en fixant par un linker les protéines EnHD, FynSH3 et Pin1WW à la sous-unité α , la chambre intermédiaire a pu être reconstituée avec un substrat dans sa cavité. Alors que ces protéines sont stables sous forme libre à 50°C, l'encapsulation dans la chambre intermédiaire s'est accompagnée d'une déstructuration du substrat qui adopte un ensemble de conformations non-repliées. De plus cette déstructuration s'est révélée être plus importante pour EnHD, constituée d'hélices α , que pour Pin1WW, constituée de brins β . Ainsi, alors que EnHD restait non-repliée à 10°C,

PinWW adoptait une conformation repliée vers 23 – 36°C alors que son T_m est normalement de 62°C. En plus d'abaisser le T_m , l'encapsulation de PinWW s'est aussi accompagnée d'un ralentissement de la cinétique de repliement et d'une augmentation de son τ_c . Cette dernière observation suggère que l'effet déstabilisateur de l'encapsulation passe par une interaction entre le substrat et la paroi de la chambre intermédiaire. Ainsi, le mutant PinWW2 qui est plus stable que PinWW, présente une conformation proche de sa forme repliée dans la chambre interne et un τ_c identique à sa forme libre. Enfin la protéolyse entre les différents substrats, EnHD, PinWW et PinWW2, s'est révélée d'autant plus efficace que le substrat était déstabilisé dans la chambre intermédiaire. La RMN a permis dans cette étude de mettre à jour l'effet de la chambre intermédiaire sur l'équilibre conformationnel du substrat.

Une régulation allostérique du protéasome a été mise en évidence par l'étude des déplacements chimiques des groupements méthyl des anneaux $\alpha 7$ et $\beta 7$ du cœur du protéasome (Ruschak *et al.* 2012). La particule 11S se lie à la surface externe de l'anneau $\alpha 7$ au niveau d'un site de liaison hydrophobe avec un K_d de 10 μM (Sprangers *et al.* 2007). Cette liaison entraîne un changement dans les déplacements chimiques d'un groupe de méthyls. Ces méthyls sont situés à l'interface entre les sous-unités α et β et forment un chemin connectant directement le site de liaison situé à la surface de $\alpha 7$ au site catalytique de $\beta 7$ situé à l'intérieur de la chambre catalytique (Figure 27). Ainsi, le déplacement chimique d'un résidu de la poche S3 du site catalytique, liant une chaîne latérale du substrat, est affecté par la liaison de la particule 11S et cette liaison est associée à un changement du profil des peptides générés par le protéasome. Cet effet est d'autant plus remarquable que la particule 11S joue un rôle important dans la réponse immunitaire en intervenant dans la genèse des peptides antigéniques qui seront présentés au système immunitaire (Niedermann *et al.* 1997, Shimbara *et al.* 1997). Enfin, la chloroquine, un inhibiteur du protéasome, ne se lie pas au site actif mais à 23 Å du site actif le plus proche et perturbe les déplacements chimiques des même méthyls que ceux perturbés par la liaison de la particule 11S et situés à l'interface entre les sous-unités α et β . Ainsi l'effet inhibiteur de la chloroquine est donc de nature allostérique en perturbant l'équilibre conformationnel du protéasome.

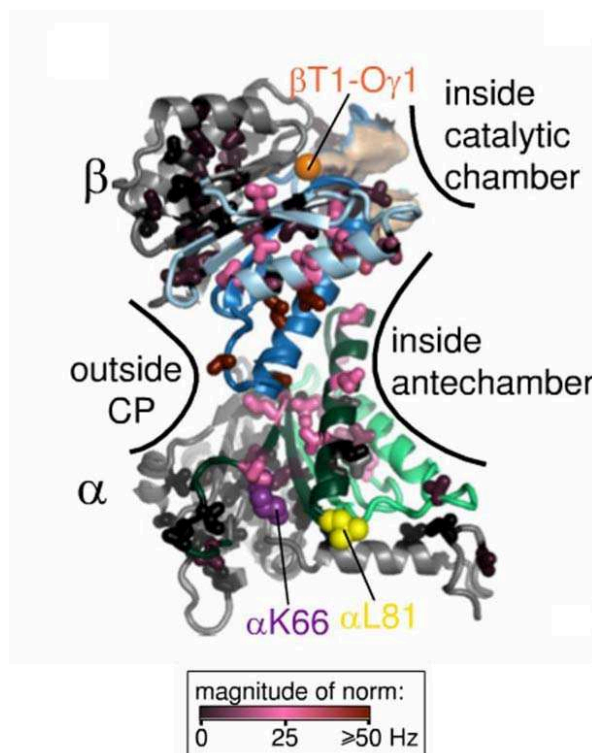


Figure 27. Régulation allostérique du protéasome. Les sous-unités α et β sont colorées en vert et en bleu, respectivement. Les résidus clef du site de liaison de la sous-unité α , K66 et L81, sont indiqués en mauve et en jaune. La thréonine catalytique T1 et la poche S3 de la sous-unité β sont indiquées par une sphère orange et par une surface, respectivement. Les méthyles sont colorés proportionnellement au déplacement chimique observé. Les méthyles montrant les déplacements chimiques les plus importants forment un chemin reliant le site de liaison de la sous-unité α au site catalytique de la sous-unité β . Adapté de (Ruschak *et al.* 2012).

3.2.2. La protéine CAP et la mesure de l'entropie conformationnelle

Dans une étude récente, la RMN a permis d'étudier la contribution de l'entropie conformationnelle à l'enthalpie libre de liaison lors de la formation du complexe ADN-protéine par la protéine CAP et les liens avec la régulation allostérique de CAP (Tzeng *et al.* 2012). La protéine CAP est constituée de deux domaines, un premier domaine liant l'ADN (DBP pour « DNA Binding Domain ») et un second domaine liant l'AMP cyclique (AMPc) (CBP pour « cAMP Binding Domain »). L'AMPc régule par allostérie la formation du complexe CAP-ADN. En effet, la fixation de l'AMPc sur le domaine CBP va entraîner une réorientation du domaine DBP qui va alors pouvoir reconnaître l'ADN (Figure 28A). Il existe un équilibre conformationnel régulé par la liaison à l'AMPc entre une conformation inactive et une conformation active (Popovych *et al.* 2009).

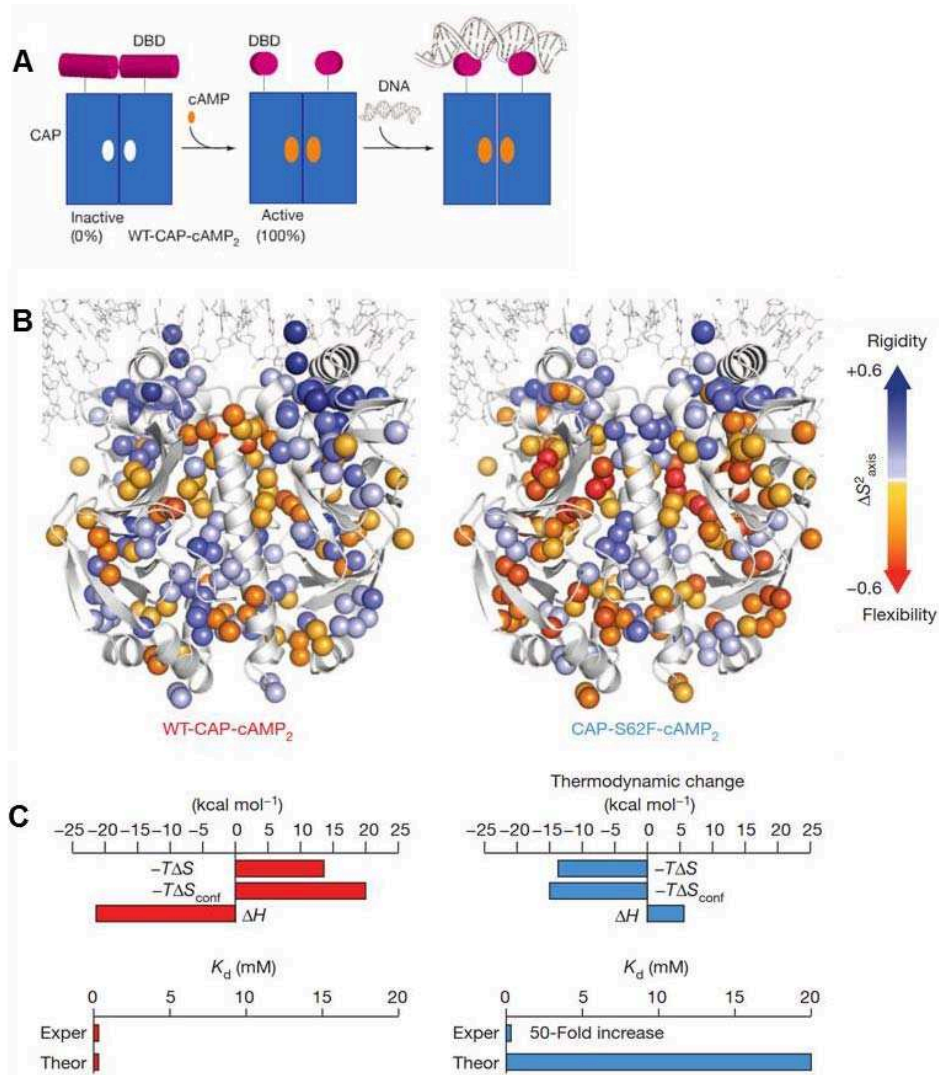


Figure 28. Liaison de la protéine CAP à l'ADN. A. L'activation de CAP par l'AMP cyclique permet la liaison à l'ADN. B. Effet de la liaison à l'ADN sur le paramètre d'ordre S^2_{axis} des méthyles. Une variation positive témoigne d'une augmentation de la rigidité. L'ADN est représenté en stick, la protéine en cartoon et les groupements méthyles des alanines, valines, leucines, isoleucines et méthionines sont représentés par des sphères. C. Entropie de liaison ($-T\Delta S$) mesurée par ITC, entropie conformationnelle ($-T\Delta S_{conf}$) déterminée par RMN et les constantes de dissociation (K_d) expérimentales (Exper) et théoriques (Theor). Adapté de (Tzeng *et al.* 2012).

Plusieurs variants de CAP montrent des équilibres conformationnels différents entre les formes actives et inactives. Toutefois, ces différences entre les populations actives et inactives ne permettent pas d'expliquer les différences d'affinité pour l'ADN observées entre les variants. De plus, l'étude par titration calorimétrique isotherme (ITC) a montré que même si les variants présentaient une enthalpie libre de liaison à l'ADN équivalente, la contribution de l'entropie et de l'enthalpie à cette enthalpie libre variait avec le variant considéré. Cette variation de l'entropie ne peut s'expliquer que par un changement de l'entropie conformationnelle de CAP. En effet, les mutations des variants sont situées à distance de la surface d'interaction entre CAP et l'ADN et ne perturbent donc pas celle-ci. Par conséquent, la surface enfouie lors de la formation du complexe est la même entre les

différents variants et la solvation contribue de manière identique pour tous les variants à l'énergie de liaison, comme en témoigne les paramètres thermodynamiques de la liaison mesurés par ITC dans l'eau et dans l'eau deutériée (Chervenak *et al.* 1994, Tzeng *et al.* 2012). En mesurant le paramètre d'ordre des groupements méthyls, qui reflète l'amplitude des mouvements à l'échelle pico – nanoseconde, il est possible d'évaluer l'entropie conformationnelle de CAP et de suivre sa variation suite à la formation du complexe avec l'ADN (Akke *et al.* 1993, Yang *et al.* 1996). Ainsi, lors de la liaison à l'ADN de la conformation active de CAP-cAMP, on observe une augmentation des paramètres d'ordre ce qui traduit une rigidification de CAP-cAMP et donc une diminution de l'entropie conformationnelle (Figure 28B). Le phénomène inverse est observé pour le variant CAP-S62F-cAMP où la liaison à l'ADN entraîne une augmentation des paramètres d'ordre, traduisant une flexibilité accrue et donc une entropie conformationnelle plus grande. Le K_d expérimental est 50 fois plus faible que ce que laissait présager l'équilibre conformationnel, la forme inactive de ce variant étant prédominante avec une forme active qui ne représente que 2% de sa population (Figure 28C). L'effet de l'entropie conformationnelle sur l'affinité pour l'ADN de CAP est aussi souligné dans le variant CAP-A144T. Ce variant présente des populations actives et inactives équivalentes en l'absence et en présence de cGMP. Néanmoins, CAP-A144T lie l'ADN avec une affinité 75 fois plus faible que CAP-A144T-cGMP. Ainsi CAP-A144T présente une diminution de son entropie conformationnelle lors de la liaison à l'ADN au contraire de CAP-A144T-cGMP où la liaison entraîne une augmentation de l'entropie conformationnelle, expliquant la différence d'affinité observée. Enfin plusieurs variants présentent des affinités similaires bien qu'ayant des entropies conformationnelles différentes, la variation d'enthalpie venant compenser ces différences entre les entropies conformationnelles. Ainsi l'étude RMN de la dynamique des groupements méthyls de CAP a permis d'expliquer les différences d'affinités entre les variants qui ne pouvaient être appréhendées à partir de la structure seule.

3.2.3. Le lysozyme du phage T4, et la caractérisation d'états « invisibles »

La RMN permet également d'étudier la structure de conformations minoritaires et par conséquent invisibles par des approches structurales usuelles. Ainsi la structure de l'état excité minoritaire du variant L99A du lysozyme du phage T4 a été déterminé par RMN. Le remplacement d'une leucine par une alanine crée une cavité au niveau du cœur hydrophobe de la protéine (Eriksson *et al.* 1992b). Cette cavité enfouie dans la protéine peut accommoder une molécule de benzène (Figure 29A et C) (Eriksson *et al.* 1992a).

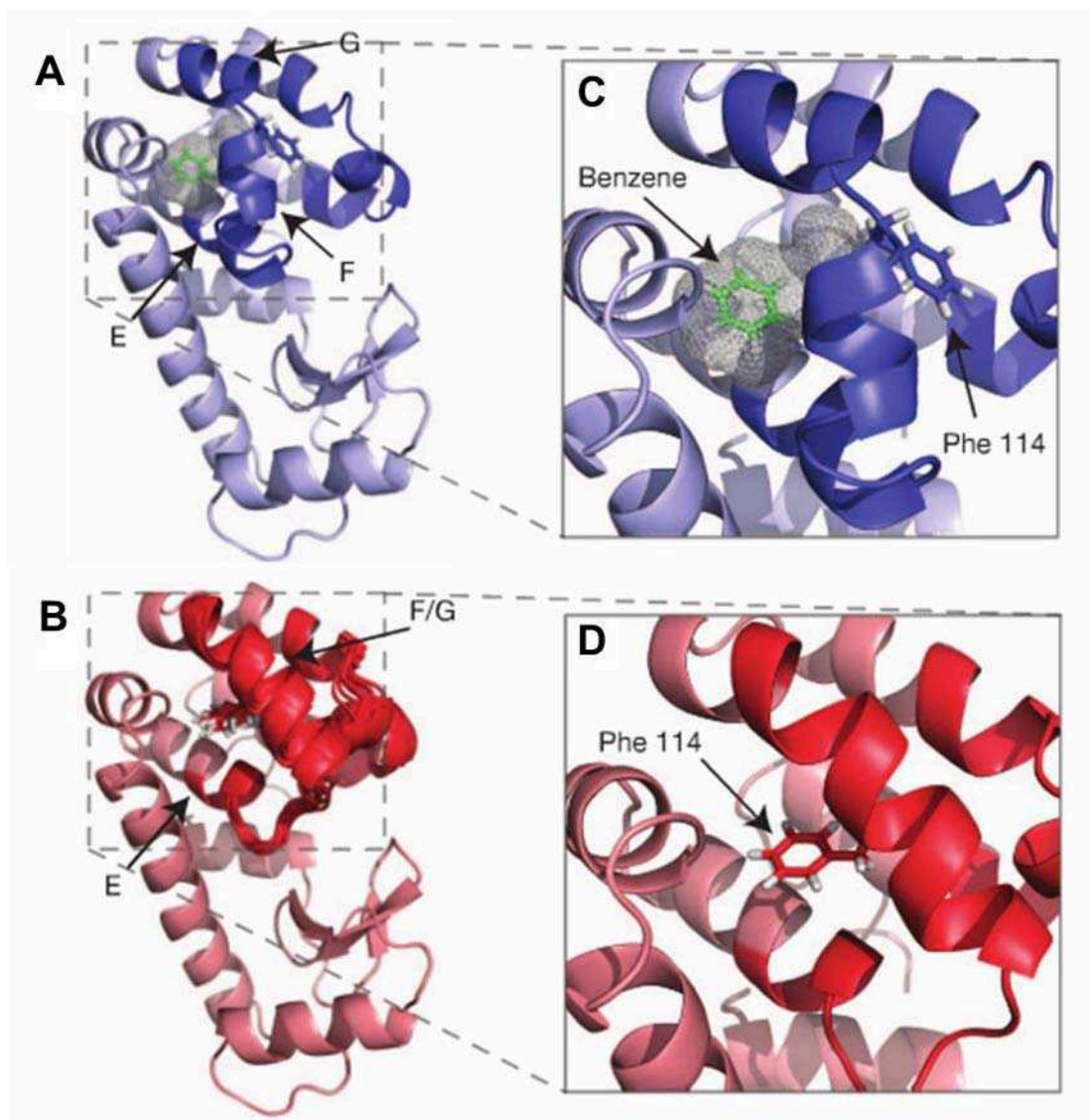


Figure 29. Structure de l'état « invisible » de L99A-T4L. A. Structure cristallographique de l'état de base de L99A-T4L avec une molécule de benzène, en vert, dans la cavité. Les positions des hélices E, F et G sont indiquées ainsi que la chaîne latérale de F114. B. Superposition des 10 structures de plus basse énergie de l'état excité de L99A-T4L. C. et D. Vue rapprochée de la chaîne latérale de F114 dans les structures de l'état basal (C.) et de l'état excité (D.) de L99A-T4L. Adapté de (Bouvignies *et al.* 2011).

Les structures cristallographiques du variants L99A et de la forme sauvage sont identiques mais le variant L99A présente un élargissement significatif de certaines résonances témoignant de la présence de mouvements à l'échelle de la micro – milliseconde. Des expériences de relaxation – dispersion de type Carr – Purcell – Meiboom – Gill (CPMG) ont pu mettre en évidence la présence de deux états, un état basal et un état excité, avec un équilibre en faveur de l'état basal, 95% de la population à 25°C, et s'échangeant à l'échelle de la seconde (Bouvignies *et al.* 2011). Les expériences de CPMG ont permis d'obtenir les déplacements chimiques ^1H , ^{15}N et ^{13}C de presque tout le squelette peptidique ainsi que des groupements méthyls des chaînes latérales. En utilisant le logiciel CS-ROSETTA, il a été possible de retrouver la structure de l'état excité à partir de la

structure de l'état basal et des déplacements chimiques mesurés pour l'état excité (Shen *et al.* 2008). Les différences structurales sont situées au voisinage de la cavité où deux courtes hélices α consécutives à l'état basal forment une hélice α continue à l'état excité (Figure 29B). Ce réarrangement entraîne le comblement de la cavité par la chaîne latérale de la phénylalanine F114 (Figure 29D). L'ajout de deux mutations supplémentaires, G113A et R119P, a permis de stabiliser l'état excité et d'enregistrer un spectre HSQC ^1H - ^{15}N de l'état excité qui était invisible pour le simple variant L99A. Ce triple variant L99A-G113A-R119P a permis de valider la structure déterminée à partir des déplacements chimiques mesurés par les expériences de relaxation – dispersion. De plus, des expériences de relaxation – dispersion sur le triple variant ont montré que celui-ci présentait une conformation minoritaire correspondant à l'état basal de L99A. Néanmoins la présence de la phénylalanine F114 dans la cavité rend l'état excité incapable de lier le benzène, ce qui suggère la présence d'une troisième conformation capable de lier le ligand. Cette étude a permis de déterminer la structure d'un état conformationnel minoritaire invisible et de trouver à partir de cette structure, deux mutations permettant de complètement renverser l'équilibre conformationnel entre l'état basal et l'état excité.

3.2.4. La rhodopsine d'*Anabaena*.

La « sensory rhodopsin » d'*Anabaena* (ASR) fait partie d'une famille de photorécepteurs utilisant le rétinal comme chromophore pour réagir à la lumière. Ces photorécepteurs peuvent ainsi utiliser l'énergie de la lumière pour assurer un transport actif d'ions à travers la membrane. C'est le cas de la bactériorhodopsine et de la halorhodopsine qui permettent un transport actif des protons ou des ions chlorures, respectivement (Oesterhelt *et al.* 1973, Schobert *et al.* 1982). La lumière peut également activer ces photorécepteurs, comme pour les « sensory rhodopsin » I et II. La transduction du signal passe alors via l'interaction avec une protéine partenaire transmembranaire, HtrI et HtrII, respectivement (Hoff *et al.* 1997). La « sensory rhodopsin » d'*Anabaena* (ASR) a également une fonction de récepteur à la lumière. Toutefois, dans le cas d'ASR, la transduction du signal passe par une protéine cytoplasmique soluble, l'« *Anabaena* Sensory Rhodopsin Transducer » (ASRT) (Figure 30A) (Jung *et al.* 2003, Kim *et al.* 2014). La structure de l'ASR a été résolue par cristallographie des rayons X (PDB ID : 1XIO) (Vogele *et al.* 2004). Elle adopte un repliement semblable à celui de la bactériorhodopsine (Henderson *et al.* 1990). L'ASR est constituée de 7 hélices α transmembranaires. Le rétinal est situé dans une poche formée entre les hélices α et est lié par une liaison covalente à la lysine 210 qui forme une base de Schiff. L'isomérisation du rétinal de la forme tout-*trans* à la forme 13-*cis* sous l'effet de la lumière, accompagné de

la déprotonation de la base de Schiff entraine la dissociation d'ASRT qui va directement réguler l'expression génique (Kondoh *et al.* 2011, Kim *et al.* 2014).

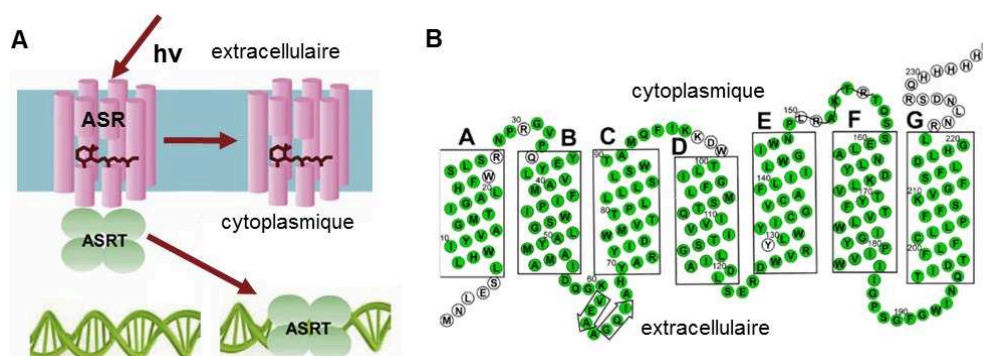


Figure 30. « *Anabaena* Sensory Rhodopsin » (ASR). A. Représentation de l'ASR avec le rétinale coloré en marron et un transducteur soluble tétramérique (ASRT) interagissant avec la face cytoplasmique de l'ASR. Sous l'effet de la lumière, l'isomérisation du rétinale va entraîner des changements conformationnels de l'ASR qui aboutiront à la dissociation de l'ASRT qui aura alors un effet régulateur sur l'expression génique. Adapté de (Wang *et al.* 2014a). B. Séquence de l'ASR, structures secondaires et attribution. Les hélices α transmembranaires sont indiquées par des rectangles. Les résidus attribués par RMN du solide sont colorés en vert. Adapté de (Good *et al.* 2014).

L'utilisation de la RMN du solide avec rotation à l'angle magique a permis d'étudier l'ASR reconstituée en bicouche lipidique, un environnement proche de la membrane cellulaire. 169 résidus sur 229 ont pu être attribués par des expériences de type CONCA, NCACX et NCOCX sur un échantillon uniformément marqué $^{15}\text{N}^{13}\text{C}$ (Shi *et al.* 2011). L'utilisation d'un marquage fractionnel en utilisant le $[1,3\text{-}^{13}\text{C}]$ -glycérol et le $[2\text{-}^{13}\text{C}]$ -glycérol comme source de carbone a permis d'obtenir des spectres avec une meilleure résolution, particulièrement en ce qui concerne les résidus aromatiques et les résidus à chaîne latérale longue (Wang *et al.* 2013b). 37 nouveaux résidus ont pu être attribués, particulièrement au niveau des boucles, portant le nombre de résidus attribués à 206 sur 229 (Figure 30B). L'analyse des déplacements chimiques ^{13}C confirme la présence des 7 hélices α transmembranaires, montrant que l'ASR adopte bien en protéoliposome le même repliement que celui de la structure cristallographique. Toutefois, deux brins β sont observés dans la boucle extracellulaire 1, entre les hélices B et C, alors que ces boucles ne sont pas visibles dans la structure cristallographique (Figure 30B) (PDB ID : 1XIO) (Vogelely *et al.* 2004). Enfin, le rétinale n'est retrouvé que dans un seul isomère d'après le spectre 1D- ^{15}N alors que deux isomères sont présents dans la structure cristallographique.

Plusieurs expériences de biochimie, notamment des expériences de « cross-linking », ont montrés que l'ASR oligomérise sous forme de trimères lors de la solubilisation en micelles de détergent et lors de la reconstitution en bicouche lipidique (Wang *et al.* 2012). Des expériences de relaxation paramagnétique ont permis de définir la surface d'interaction entre les sous-unités du trimère. Une sonde paramagnétique placée sur la cystéine 26

(mutant S26C), à l'extrémité cytoplasmique de l'hélice A, induit une réduction des signaux des résidus appartenant aux hélices A et B, ainsi qu'à la boucle A-B, tous ces résidus étant situés à moins de 15 Å de la cystéine 26. En plus des signaux de ces résidus, les signaux de résidus des hélices D et E sont également atténués, bien que situés à plus de 25 Å de la cystéine 26 portant la sonde paramagnétique (Figure 31A). Cette atténuation ne peut s'expliquer que par une proximité spatiale de l'hélice A d'un monomère avec les hélices D et E d'un autre monomère (Wang *et al.* 2012).

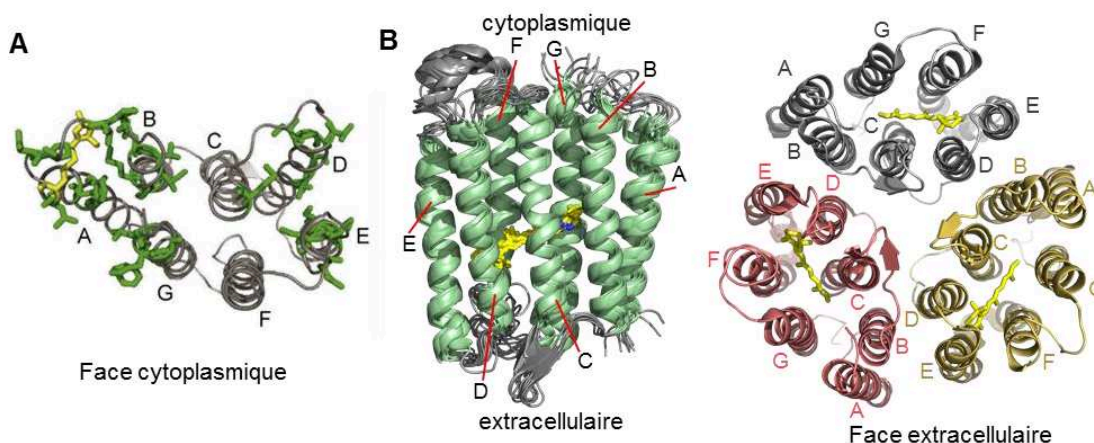


Figure 31. Structure du trimère de l'ASR par RMN du solide. A. Vue de la face cytoplasmique de l'ASR. La sonde paramagnétique est colorée en jaune et les chaînes latérales des résidus affectés sont colorées en vert. Adapté de (Ward *et al.* 2015a). B. Structure des 10 conformères de plus basse énergie déterminés par RMN du solide et agencement en trimère (PDB ID : 2M3G). Le rétinol est coloré en jaune. Adapté de (Wang *et al.* 2013a, Wang *et al.* 2014a).

L'attribution de la majorité des déplacements chimiques de l'ASR ainsi que la détermination des interfaces de trimérisation ont permis d'établir une structure tridimensionnelle de l'ASR en protéoliposomes par RMN du solide (Wang *et al.* 2013a). Les déplacements chimiques ont été utilisés pour identifier de façon non-ambiguë des pics de corrélation qui ont par la suite permis de générer une liste de contraintes spatiales. Couplée aux angles dièdres déduits à partir des déplacements chimiques des atomes du squelette peptidique par le logiciel TALOS+ (Shen *et al.* 2009), ces contraintes spatiales ont permis de calculer un modèle à basse résolution de l'ASR. Ce modèle à basse résolution a alors servi de point de départ à un protocole d'attribution des pics de corrélation ambiguës par le logiciel ARIA (Bardiaux *et al.* 2012). Ce protocole a permis d'identifier des contraintes spatiales entre des résidus des hélices B et des hélices D ou E, confirmant ainsi les surfaces d'oligomérisation identifiées précédemment par la PRE (Figure 31A) (Wang *et al.* 2012). Enfin, dans une dernière étape, l'ensemble des contraintes structurales a été utilisé pour calculer un ensemble de 10 conformères du trimère de l'ASR ayant un RMSD du squelette peptidique de 0.8 Å au niveau des hélices transmembranaire et de 1.8 Å pour tout le trimère

(PDB ID : 2M3G) (Wang *et al.* 2013a). La structure de l'ASR par RMN du solide retrouve le même repliement, sous la forme de 7 hélices α , que la structure cristallographique (Figure 31B) (PDB ID : 1XIO) (Vogelely *et al.* 2004). Toutefois alors qu'un dimère est présent dans la structure cristallographique, l'ASR forme un trimère dans la structure RMN (Figure 31B). Les sous-unités du trimère interagissent via l'hélice B d'une sous-unité avec les hélices D et E d'une autre sous-unité, principalement par les chaînes latérales de résidus aromatiques. La position rapprochée des surfaces d'interaction place les chaînes latérales en position pour former des liaisons hydrogènes et exclue une interaction médiée par les lipides. Les hélices A et E, qui constituent ou sont proches des surfaces d'interaction, sont inclinées de 10° par rapport à leur position dans la structure cristallographique. Par ailleurs, alors que les boucles de la face extracellulaire ne sont pas visibles dans la structure cristallographique, elles sont clairement ordonnées dans la structure RMN avec la boucle B-C qui adopte une structure en « β hairpin ». Au contraire, les boucles de la face cytoplasmique sont bien définies dans le cristal, notamment en raison de l'empilement cristallin, alors qu'elles sont plus mobiles dans la structure RMN, comme l'ont confirmé par la suite les expériences d'échange H_2O/D_2O .

L'étude de l'ASR insérée dans la membrane interne d'*Escherichia coli* a retrouvé une structuration de l'ASR similaire à celle adoptée après reconstitution en protéoliposome. Des résonances appartenant à 82 résidus sur 229 ont pu être attribuées en combinant des expériences de type 2D-NCA, 3D-CANCO et 3D-NCACB et en s'aidant de l'attribution réalisée précédemment en bicouche lipidique (Ward *et al.* 2015b). L'absence de perturbation des déplacements chimiques de la base de Schiff ainsi que de nombreux résidus formant la poche du rétinale indique que la conformation tout-*trans* du rétinale observée en protéoliposome n'est pas affectée par le changement d'environnement. La même remarque peut être faite pour les hélices α transmembranaires, la boucle extracellulaire B-C organisée en deux brins β antiparallèles, ainsi que les interfaces entre les sous-unités du trimère où les résonances identifiées ne présentent pas de perturbations significatives des déplacements chimiques par rapport à l'ASR en protéoliposome. Les perturbations les plus importantes des déplacements chimiques sont retrouvées pour des résidus appartenant aux hélices B ou D et orientés vers l'intérieur du monomère. Ces perturbations peuvent être une conséquence d'une différence de la pression latérale exercée sur l'ASR par la bicouche lipidique dans la membrane interne d'*E. coli* due à une composition différente de celle des protéoliposomes.

L'incubation dans le D_2O permet d'échanger les 1H labiles de la protéine, principalement les 1H amides des liaisons peptidiques, par le 2H du solvant. La cinétique de cet échange dépend de l'accessibilité au solvant ainsi que de l'engagement de ces 1H

labiles dans des liaisons hydrogène au niveau de structures secondaires et de la dynamique de ces structures secondaires. Le suivi de la cinétique d'échange des ^1H labiles permet ainsi d'accéder à la dynamique de la protéine et de définir les surfaces accessibles au solvant. L'incubation dans le D_2O de l'ASR, montre que les segments transmembranaires sont protégés du solvant alors que les résidus des faces extracellulaire et cytoplasmique sont en contact direct avec celui-ci comme en témoigne l'échange $^1\text{H}/^2\text{H}$ observé à ce niveau y compris dans la boucle B-C qui forme deux brins β (Figure 32A). De plus, la face cytoplasmique semble plus exposée au solvant que la face extracellulaire (Shi *et al.* 2011). La partie cytoplasmique de l'ASR présente un caractère hydrophile plus important que les autres rhodopsines microbiennes et un réseau de liaison hydrogène incluant 4 molécules d'eau relie le site du rétinale à la face cytoplasmique dans la structure cristallographique (PDB ID : 1XIO) (Vogelely *et al.* 2004). L'exposition à la lumière accroît le phénomène d'échange $^1\text{H}/^2\text{H}$, particulièrement à la face cytoplasmique où les extrémités des hélices B, C, G et F voient l'intensité de leurs signaux s'atténuer voire complètement disparaître (Figure 32A) (Wang *et al.* 2011). Cette augmentation de l'échange sous l'effet de la lumière traduit une dynamique plus importante de l'ASR dans ces conditions particulièrement au niveau de la face cytoplasmique où vient interagir l'ASRT. L'interaction entre l'ASR et l'ASRT est en effet directement régulée par la conformation du rétinale et à l'état de protonation de la base de Schiff sous l'effet de la lumière (Kondoh *et al.* 2011).

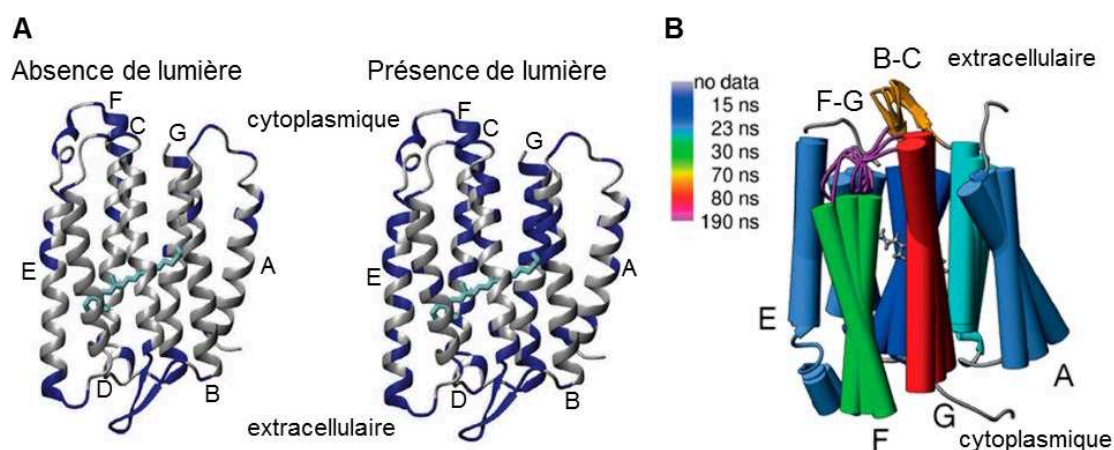


Figure 32. Dynamique de l'ASR. A. Echange $^1\text{H}/^2\text{H}$ en absence et en présence de lumière. Les résidus ayant un proton amide échangeable sont colorés en bleu. Adapté de (Wang *et al.* 2014a). B. Echelle de temps, amplitudes et directions des mouvements obtenus par l'analyse « 3D Gaussian Axial Fluctuation » de $R_{1\rho}$, S^2_{NH} et $S^2_{\text{CaH}\alpha}$. Adapté de (Good *et al.* 2014).

A côté de l'échange $^1\text{H}/^2\text{H}$ des ^1H labiles, la RMN du solide permet également de sonder directement la dynamique des protéines. La mesure des paramètres d'ordre à partir des interactions dipolaires permet d'accéder à l'amplitude des mouvements inférieurs à la centaine de microsecondes, alors que les temps de relaxation spin-réseau dans le

référentiel tournant ($R_{1\rho}$) donnent accès aux temps caractéristiques de ces mouvements (Good *et al.* 2014). L'analyse des couplages dipolaires ^{15}N - $^1\text{H}_\text{N}$, $^{13}\text{C}_\alpha$ - $^1\text{H}_\alpha$, $^{13}\text{C}'$ - ^{15}N et $^{13}\text{C}_\alpha$ - ^{15}N des résidus de l'ASR a révélé de faibles différences entre les parties transmembranaires et les boucles exposées au solvant, suggérant que les hélices et les boucles de l'ASR présentaient une mobilité réduite et des mouvements d'amplitude équivalente. Au contraire, la relaxation $R_{1\rho}$ affiche des différences significatives entre les hélices transmembranaires et les boucles. Les valeurs de $R_{1\rho}$ mesurées au niveau des hélices indiquent des mouvements rapides, autour de la dizaine de nanosecondes, alors que les valeurs plus élevées au niveau des boucles, particulièrement pour les boucles extracellulaires B-C et F-G, sont la marque de mouvements plus lents, autour de la centaine de nanosecondes. Pour tester la présence de mouvements collectifs de l'ASR, les mouvements des structures secondaires ont été modélisés par des mouvements de corps rigides ayant un temps et 3 amplitudes suivant 3 axes orthogonaux par une approche « 3D Gaussian Axial Fluctuation ». Les boucles B-C et F-G ont également été considérées comme des corps rigides, la boucle B-C consistant en deux brins β antiparallèles et des éléments de structures secondaire β étant également présents au niveau de la boucle F-G (Shi *et al.* 2011, Wang *et al.* 2013a). Cette approche retrouve des mouvements de 20 à 30 ns pour les 6 hélices A – F et des mouvements plus lents proches de 90 ns pour la septième hélice G, les boucles B-C et F-G étant animées de mouvements de l'ordre de 70 et 190 ns, respectivement (Figure 32B) (Good *et al.* 2014). Toutefois, le manque de données expérimentales n'a pas permis de modéliser les mouvements des boucles de la face cytoplasmique qui assurent l'interaction avec l'ASRT.

IV. Références

- Adamian L, Naveed H and Liang J. *Lipid-binding surfaces of membrane proteins: evidence from evolutionary and structural analysis*. Biochim Biophys Acta, **2011**, 1808(4): 1092-1102.
- Akke M, Brueschweiler R and Palmer A G. *NMR order parameters and free energy: an analytical approach and its application to cooperative calcium(2+) binding by calbindin D9k*. Journal of the American Chemical Society, **1993**, 115(21): 9832-9833.
- Akoh C C, Lee G C, Liaw Y C, Huang T H and Shaw J F. *GDSL family of serine esterases/lipases*. Prog Lipid Res, **2004**, 43(6): 534-552.
- Aloulou A, Benarouche A, Puccinelli D, Spinelli S, Cavalier J F, Cambillau C and Carriere F. *Biochemical and structural characterization of non-glycosylated Yarrowia lipolytica LIP2 lipase*. European Journal of Lipid Science and Technology, **2013**, 115(4): 429-441.
- Aloulou A, Puccinelli D, De Caro A, Leblond Y and Carriere F. *A comparative study on two fungal lipases from Thermomyces lanuginosus and Yarrowia lipolytica shows the combined effects of detergents and pH on lipase adsorption and activity*. Biochim Biophys Acta, **2007a**, 1771(12): 1446-1456.
- Aloulou A, Rodriguez J A, Puccinelli D, Mouz N, Leclaire J, Leblond Y and Carriere F. *Purification and biochemical characterization of the LIP2 lipase from Yarrowia lipolytica*. Biochim Biophys Acta, **2007b**, 1771(2): 228-237.
- Ames G F, Spudich E N and Nikaido H. *Protein composition of the outer membrane of Salmonella typhimurium: effect of lipopolysaccharide mutations*. J Bacteriol, **1974**, 117(2): 406-416.
- Andersen C, Koronakis E, Hughes C and Koronakis V. *An aspartate ring at the TolC tunnel entrance determines ion selectivity and presents a target for blocking by large cations*. Molecular Microbiology, **2002**, 44(5): 1131-1139.
- Arora A, Abildgaard F, Bushweller J H and Tamm L K. *Structure of outer membrane protein A transmembrane domain by NMR spectroscopy*. Nat Struct Biol, **2001**, 8(4): 334-338.
- Ayala I, Sounier R, Use N, Gans P and Boisdouvier J. *An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein*. Journal of Biomolecular Nmr, **2009**, 43(2): 111-119.
- Bai X C, McMullan G and Scheres S H W. *How cryo-EM is revolutionizing structural biology*. Trends in Biochemical Sciences, **2015**, 40(1): 49-57.
- Ban D, Sabo T M, Griesinger C and Lee D. *Measuring Dynamic and Kinetic Information in the Previously Inaccessible Supra-tau(c) Window of Nanoseconds to Microseconds by Solution NMR Spectroscopy*. Molecules, **2013**, 18(10): 11904-11937.
- Barbe S, Lafaquiere V, Guieysse D, Monsan P, Remaud-Simeon M and Andre I. *Insights into lid movements of Burkholderia cepacia lipase inferred from molecular dynamics simulations*. Proteins, **2009**, 77(3): 509-523.
- Bardiaux B, Malliavin T and Nilges M. *ARIA for solution and solid-state NMR*. Methods Mol Biol, **2012**, 831: 453-483.

Bardwell J C, McGovern K and Beckwith J. *Identification of a protein required for disulfide bond formation in vivo*. Cell, **1991**, 67(3): 581-589.

Bechtluft P, Nouwen N, Tans S J and Driessen A J. *SecB--a chaperone dedicated to protein translocation*. Mol Biosyst, **2010**, 6(4): 620-627.

Behrens S, Maier R, de Cock H, Schmid F X and Gross C A. *The SurA periplasmic PPlase lacking its parvulin domains functions in vivo and has chaperone activity*. Embo Journal, **2001**, 20(1-2): 285-294.

Belle V, Fournel A, Woudstra M, Ranaldi S, Prieri F, Thome V, Currault J, Verger R, Guigliarelli B and Carriere F. *Probing the opening of the pancreatic lipase lid using site-directed spin labeling and EPR spectroscopy*. Biochemistry, **2007**, 46(8): 2205-2214.

Berg J M, Tymoczko J L, Stryer L and Clarke N D. *Biochemistry*. **2002**, New York, Freeman and Company.

Berglund P, Vallikivi I, Fransson L, Dannacher H, Holmquist M, Martinelle M, Bjorkling F, Parve O and Hult K. *Switched enantiopreference of Humicola lipase for 2-phenoxyalkanoic acid ester homologs can be rationalized by different substrate binding modes*. Tetrahedron-Asymmetry, **1999**, 10(21): 4191-4202.

Bitto E and McKay D B. *Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins*. Structure, **2002**, 10(11): 1489-1498.

Bordes F, Barbe S, Escalier P, Mourey L, Andre I, Marty A and Tranier S. *Exploring the conformational states and rearrangements of Yarrowia lipolytica Lipase*. Biophys J, **2010**, 99(7): 2225-2234.

Bordes F, Cambon E, Dossat-Letisse V, Andre I, Croux C, Nicaud J M and Marty A. *Improvement of Yarrowia lipolytica lipase enantioselectivity by using mutagenesis targeted to the substrate binding site*. Chembiochem, **2009**, 10(10): 1705-1713.

Bordes F, Tarquis L, Nicaud J M and Marty A. *Isolation of a thermostable variant of Lip2 lipase from Yarrowia lipolytica by directed evolution and deeper insight into the denaturation mechanisms involved*. J Biotechnol, **2011**, 156(2): 117-124.

Borgia A, Williams P M and Clarke J. *Single-molecule studies of protein folding*. Annual Review of Biochemistry, **2008**, 77: 101-125.

Borman E K, Stuart C A and Wheeler K M. *Taxonomy of the Family Enterobacteriaceae*. J Bacteriol, **1944**, 48(3): 351-367.

Bosshart P D, Iordanov I, Garzon-Coral C, Demange P, Engel A, Milon A and Muller D J. *The transmembrane protein KpOmpA anchoring the outer membrane of Klebsiella pneumoniae unfolds and refolds in response to tensile load*. Structure, **2012**, 20(1): 121-127.

Bouvignies G, Vallurupalli P, Hansen D F, Correia B E, Lange O, Bah A, Vernon R M, Dahlquist F W, Baker D and Kay L E. *Solution structure of a minor and transiently formed state of a T4 lysozyme mutant*. Nature, **2011**, 477(7362): 111-114.

Brady L, Brzozowski A M, Derewenda Z S, Dodson E, Dodson G, Tolley S, Turkenburg J P, Christiansen L, Hage-Jensen B, Norskov L and et al. *A serine protease triad forms the catalytic centre of a triacylglycerol lipase*. Nature, **1990**, 343(6260): 767-770.

Brocca S, Secundo F, Ossola M, Alberghina L, Carrea G and Lotti M. *Sequence of the lid affects activity and specificity of Candida rugosa lipase isoenzymes*. Protein Sci, **2003**, 12(10): 2312-2319.

Brzozowski A M, Derewenda U, Derewenda Z S, Dodson G G, Lawson D M, Turkenburg J P, Bjorkling F, Huge-Jensen B, Patkar S A and Thim L. *A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex*. Nature, **1991**, 351(6326): 491-494.

Brzozowski A M, Derewenda Z S, Dodson E J, Dodson G G and Turkenburg J P. *Structure and Molecular-Model Refinement of Rhizomucor-Miehei Triacylglyceride Lipase - a Case-Study of the Use of Simulated Annealing in Partial Model Refinement*. Acta Crystallographica Section B-Structural Science, **1992**, 48: 307-319.

Brzozowski A M, Savage H, Verma C S, Turkenburg J P, Lawson D M, Svendsen A and Patkar S. *Structural origins of the interfacial activation in Thermomyces (Humicola) lanuginosa lipase*. Biochemistry, **2000**, 39(49): 15071-15082.

Bulieris P V, Behrens S, Holst O and Kleinschmidt J H. *Folding and insertion of the outer membrane protein OmpA is assisted by the chaperone Skp and by lipopolysaccharide*. J Biol Chem, **2003**, 278(11): 9092-9099.

Burmann B M, Wang C and Hiller S. *Conformation and dynamics of the periplasmic membrane-protein-chaperone complexes OmpX-Skp and tOmpA-Skp*. Nat Struct Mol Biol, **2013**, 20(11): 1265-1272.

Cabeen M T and Jacobs-Wagner C. *Bacterial cell shape*. Nat Rev Microbiol, **2005**, 3(8): 601-610.

Cambon E, Piamtongkam R, Bordes F, Duquesne S, Andre I and Marty A. *Rationally engineered double substituted variants of Yarrowia lipolytica lipase with enhanced activity coupled with highly inverted enantioselectivity towards 2-bromo phenyl acetic acid esters*. Biotechnol Bioeng, **2010**, 106(6): 852-859.

Casas-Godoy L, Meunchan M, Cot M, Duquesne S, Bordes F and Marty A. *Yarrowia lipolytica lipase Lip2: an efficient enzyme for the production of concentrates of docosahexaenoic acid ethyl ester*. J Biotechnol, **2014**, 180: 30-36.

Chen K and Tjandra N. *The Use of Residual Dipolar Coupling in Studying Proteins by NMR*. Nmr of Proteins and Small Biomolecules, **2012**, 326: 47-67.

Chen R and Henning U. *A periplasmic protein (Skp) of Escherichia coli selectively binds a class of outer membrane proteins*. Mol Microbiol, **1996**, 19(6): 1287-1294.

Chervenak M C and Toone E J. *A Direct Measure of the Contribution of Solvent Reorganization to the Enthalpy of Binding*. Journal of the American Chemical Society, **1994**, 116(23): 10533-10539.

Chimento D P, Mohanty A K, Kadner R J and Wiener M C. *Substrate-induced transmembrane signaling in the cobalamin transporter BtuB*. Nat Struct Biol, **2003**, 10(5): 394-401.

Cierpicki T, Liang B, Tamm L K and Bushweller J H. *Increasing the accuracy of solution NMR structures of membrane proteins by application of residual dipolar couplings. High-resolution structure of outer membrane protein A*. J Am Chem Soc, **2006**, 128(21): 6947-6951.

Clore G M. *Seeing the invisible by paramagnetic and diamagnetic NMR*. Biochemical Society Transactions, **2013**, 41: 1343-1354.

Colton I J, Yin D, Grochulski P and Kazlauskas R J. *Molecular Basis of Chiral Acid Recognition by Candida rugosa Lipase: X-Ray Structure of Transition State Analog and Modeling of the Hydrolysis of Methyl 2-Methoxy-2-phenylacetate*. Advanced Synthesis & Catalysis, **2011**, 353(13): 2529-2544.

Confer A W and Ayalew S. *The OmpA family of proteins: roles in bacterial pathogenesis and immunity*. Vet Microbiol, **2013**, 163(3-4): 207-222.

Cortes G, Borrell N, de Astorza B, Gomez C, Sauleda J and Alberti S. *Molecular analysis of the contribution of the capsular polysaccharide and the lipopolysaccharide O side chain to the virulence of Klebsiella pneumoniae in a murine model of pneumonia*. Infect Immun, **2002**, 70(5): 2583-2590.

Cortes J, Simeon T, de Angulo V R, Guieysse A D, Remaud-Simeon M and Tran V. *A path planning approach for computing large-amplitude motions of flexible molecules*. Bioinformatics, **2005**, 21: 1116-1125.

Cowan S W, Schirmer T, Rummel G, Steiert M, Ghosh R, Pauptit R A, Jansonius J N and Rosenbusch J P. *Crystal structures explain functional properties of two E. coli porins*. Nature, **1992**, 358(6389): 727-733.

Cuesta-Seijo J A, Neale C, Khan M A, Moktar J, Tran C D, Bishop R E, Pomes R and Prive G G. *PagP crystallized from SDS/cosolvent reveals the route for phospholipid access to the hydrocarbon ruler*. Structure, **2010**, 18(9): 1210-1219.

Cygler M, Grochulski P, Kazlauskas R J, Schrag J D, Bouthillier F, Rubin B, Serreji A N and Gupta A K. *A Structural Basis for the Chiral Preferences of Lipases*. Journal of the American Chemical Society, **1994**, 116(8): 3180-3186.

Danoff E J and Fleming K G. *The soluble, periplasmic domain of OmpA folds as an independent unit and displays chaperone activity by reducing the self-association propensity of the unfolded OmpA transmembrane beta-barrel*. Biophys Chem, **2011**, 159(1): 194-204.

de Cock H, Brandenburg K, Wiese A, Holst O and Seydel U. *Non-lamellar structure and negative charges of lipopolysaccharides required for efficient folding of outer membrane protein PhoE of Escherichia coli*. J Biol Chem, **1999**, 274(8): 5114-5119.

de Cock H, Struyve M, Kleerebezem M, van der Krift T and Tommassen J. *Role of the carboxy-terminal phenylalanine in the biogenesis of outer membrane protein PhoE of Escherichia coli K-12*. J Mol Biol, **1997**, 269(4): 473-478.

de Keyzer J, van der Does C and Driessen A J. *The bacterial translocase: a dynamic protein channel complex*. Cell Mol Life Sci, **2003**, 60(10): 2034-2052.

Demchick P and Koch A L. *The permeability of the wall fabric of Escherichia coli and Bacillus subtilis*. J Bacteriol, **1996**, 178(3): 768-773.

Derewenda U, Brzozowski A M, Lawson D M and Derewenda Z S. *Catalysis at the Interface - the Anatomy of a Conformational Change in a Triglyceride Lipase*. Biochemistry, **1992a**, 31(5): 1532-1541.

Derewenda U, Swenson L, Wei Y, Green R, Kobos P M, Joerger R, Haas M J and Derewenda Z S. *Conformational lability of lipases observed in the absence of an oil-water interface: crystallographic studies of enzymes from the fungi Humicola lanuginosa and Rhizopus delemar*. J Lipid Res, **1994**, 35(3): 524-534.

Derewenda Z S, Derewenda U and Dodson G G. *The crystal and molecular structure of the Rhizomucor miehei triacylglyceride lipase at 1.9 Å resolution*. J Mol Biol, **1992b**, 227(3): 818-839.

Dodson G and Wlodawer A. *Catalytic triads and their relatives*. Trends Biochem Sci, **1998**, 23(9): 347-352.

Eghiaian F, Rico F, Colom A, Casuso I and Scheuring S. *High-speed atomic force microscopy: Imaging and force spectroscopy*. FEBS letters, **2014**, 588(19): 3631-3638.

Enderlin C S and Ogrydziak D M. *Cloning, nucleotide sequence and functions of XPR6, which codes for a dibasic processing endoprotease from the yeast Yarrowia lipolytica*. Yeast, **1994**, 10(1): 67-79.

Eriksson A E, Baase W A, Wozniak J A and Matthews B W. *A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene*. Nature, **1992a**, 355(6358): 371-373.

Eriksson A E, Baase W A, Zhang X J, Heinz D W, Blaber M, Baldwin E P and Matthews B W. *Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect*. Science, **1992b**, 255(5041): 178-183.

Evrard B, Balestrino D, Dosgilbert A, Bouya-Gachancard J L J, Charbonnel N, Forestier C and Tridon A. *Roles of Capsule and Lipopolysaccharide O Antigen in Interactions of Human Monocyte-Derived Dendritic Cells and Klebsiella pneumoniae*. Infection and Immunity, **2010**, 78(1): 210-219.

Faraldo-Gomez J D and Sansom M S. *Acquisition of siderophores in gram-negative bacteria*. Nat Rev Mol Cell Biol, **2003**, 4(2): 105-116.

Feld G K and Frank M. *Enabling membrane protein structure and dynamics with X-ray free electron lasers*. Current opinion in structural biology, **2014**, 27: 69-78.

Ferguson A D, Hofmann E, Coulton J W, Diederichs K and Welte W. *Siderophore-mediated iron transport: crystal structure of FhuA with bound lipopolysaccharide*. Science, **1998**, 282(5397): 2215-2220.

Ferguson A D, Welte W, Hofmann E, Lindner B, Holst O, Coulton J W and Diederichs K. *A conserved structural motif for lipopolysaccharide recognition by procaryotic and eucaryotic proteins*. Structure, **2000**, 8(6): 585-592.

Fickers P, Marty A and Nicaud J M. *The lipases from Yarrowia lipolytica: genetics, production, regulation, biochemical characterization and biotechnological applications*. Biotechnol Adv, **2011**, 29(6): 632-644.

Finn R D, Bateman A, Clements J, Coghill P, Eberhardt R Y, Eddy S R, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer E L, Tate J and Punta M. *Pfam: the protein families database*. Nucleic Acids Res, **2014**, 42(Database issue): D222-230.

Fischer M, Thai Q K, Grieb M and Pleiss J. *DWARF--a data warehouse system for analyzing protein families*. BMC Bioinformatics, **2006**, 7: 495.

- Fralick J A. *Evidence that TolC is required for functioning of the Mar/AcrAB efflux pump of Escherichia coli*. Journal of Bacteriology, **1996**, 178(19): 5803-5805.
- Freer S T, Kraut J, Robertus J D, Wright H T and Xuong N H. *Chymotrypsinogen: 2.5-angstrom crystal structure, comparison with alpha-chymotrypsin, and implications for zymogen activation*. Biochemistry, **1970**, 9(9): 1997-2009.
- Firdich E and Whitfield C. *Lipopolysaccharide inner core oligosaccharide structure and outer membrane stability in human pathogens belonging to the Enterobacteriaceae*. J Endotoxin Res, **2005**, 11(3): 133-144.
- Gabel F, Bicout D, Lehnert U, Tehei M, Weik M and Zaccai G. *Protein dynamics studied by neutron scattering*. Quarterly Reviews of Biophysics, **2002**, 35(4): 327-367.
- Gal M, Kern T, Schanda P, Frydman L and Brutscher B. *An improved ultrafast 2D NMR experiment: Towards atom-resolved real-time studies of protein kinetics at multi-Hz rates*. Journal of Biomolecular Nmr, **2009**, 43(1): 1-10.
- Galdiero S, Galdiero M and Pedone C. *beta-Barrel membrane bacterial proteins: structure, function, assembly and interaction with lipids*. Curr Protein Pept Sci, **2007**, 8(1): 63-82.
- Gehring K B and Nikaido H. *Existence and purification of porin heterotrimers of Escherichia coli K12 OmpC, OmpF, and PhoE proteins*. J Biol Chem, **1989**, 264(5): 2810-2815.
- Good D B, Wang S, Ward M E, Struppe J, Brown L S, Lewandowski J R and Ladizhansky V. *Conformational dynamics of a seven transmembrane helical protein Anabaena Sensory Rhodopsin probed by solid-state NMR*. J Am Chem Soc, **2014**, 136(7): 2833-2842.
- Goto N K, Gardner K H, Mueller G A, Willis R C and Kay L E. *A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated N-15-, C-13-, H-2-labeled proteins*. Journal of Biomolecular Nmr, **1999**, 13(4): 369-374.
- Grizot S and Buchanan S K. *Structure of the OmpA-like domain of RmpM from Neisseria meningitidis*. Mol Microbiol, **2004**, 51(4): 1027-1037.
- Grochulski P, Bouthillier F, Kazlauskas R J, Serreqi A N, Schrag J D, Ziomek E and Cygler M. *Analogs of reaction intermediates identify a unique substrate binding site in Candida rugosa lipase*. Biochemistry, **1994a**, 33(12): 3494-3500.
- Grochulski P, Li Y, Schrag J D, Bouthillier F, Smith P, Harrison D, Rubin B and Cygler M. *Insights into interfacial activation from an open structure of Candida rugosa lipase*. J Biol Chem, **1993**, 268(17): 12843-12847.
- Grochulski P, Li Y, Schrag J D and Cygler M. *Two conformational states of Candida rugosa lipase*. Protein Sci, **1994b**, 3(1): 82-91.
- Groenewald M, Boekhout T, Neuveglise C, Gaillardin C, van Dijck P W and Wyss M. *Yarrowia lipolytica: safety assessment of an oleaginous yeast with a great industrial potential*. Crit Rev Microbiol, **2014**, 40(3): 187-206.
- Guieysse D, Salagnad C, Monsan P, Remaud-Simeon M and Tran V. *Towards a novel explanation of Pseudomonas cepacia lipase enantioselectivity via molecular modelling of the enantiomer trajectory into the active site*. Tetrahedron-Asymmetry, **2003**, 14(13): 1807-1817.

Guieysse D, Sandoval G, Faure L, Nicaud J M, Monsan P and Marty A. *New efficient lipase from Yarrowia lipolytica for the resolution of 2-bromo-arylacetic acid esters*. Tetrahedron-Asymmetry, **2004**, 15(22): 3539-3543.

Guillet V, Knibiehler M, Gregory-Pauron L, Remy M H, Chemin C, Raynaud-Messina B, Bon C, Kollman J M, Agard D A, Merdes A and Mourey L. *Crystal structure of gamma-tubulin complex protein GCP4 provides insight into microtubule nucleation*. Nature structural & molecular biology, **2011**, 18(8): 915-U1401.

Hagan C L, Westwood D B and Kahne D. *bam Lipoproteins Assemble BamA in vitro*. Biochemistry, **2013**, 52(35): 6108-6113.

Hagge S O, de Cock H, Gutschmann T, Beckers F, Seydel U and Wiese A. *Pore formation and function of phosphoporin PhoE of Escherichia coli are determined by the core sugar moiety of lipopolysaccharide*. J Biol Chem, **2002**, 277(37): 34247-34253.

Hancock R E and Brinkman F S. *Function of pseudomonas porins in uptake and efflux*. Annu Rev Microbiol, **2002**, 56: 17-38.

Hansel R, Luh L M, Corbeski I, Trantirek L and Dotsch V. *In-Cell NMR and EPR Spectroscopy of Biomacromolecules*. Angewandte Chemie-International Edition, **2014**, 53(39): 10300-10314.

Hansen D F, Vallurupalli P and Kay L E. *Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states*. Journal of Biomolecular Nmr, **2008**, 41(3): 113-120.

Hanshaw R G, Stahelin R V and Smith B D. *Noncovalent keystone interactions controlling biomembrane structure*. Chemistry, **2008**, 14(6): 1690-1697.

Harms N, Koningstein G, Dontje W, Muller M, Oudega B, Luirink J and de Cock H. *The early interaction of the outer membrane protein PhoE with the periplasmic chaperone Skp occurs at the cytoplasmic membrane*. Journal of Biological Chemistry, **2001**, 276(22): 18804-18811.

Hearn E M, Patel D R, Lepore B W, Indic M and van den Berg B. *Transmembrane passage of hydrophobic compounds through a protein channel wall*. Nature, **2009**, 458(7236): 367-370.

Heffernan E J, Wu L, Louie J, Okamoto S, Fierer J and Guiney D G. *Specificity of the complement resistance and cell association phenotypes encoded by the outer membrane protein genes rck from Salmonella typhimurium and ail from Yersinia enterocolitica*. Infect Immun, **1994**, 62(11): 5183-5186.

Henderson R, Baldwin J M, Ceska T A, Zemlin F, Beckmann E and Downing K H. *Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy*. J Mol Biol, **1990**, 213(4): 899-929.

Hennig J and Sattler M. *The dynamic duo: Combining NMR and small angle scattering in structural biology*. Protein Science, **2014**, 23(6): 669-682.

Hermoso J, Pignol D, Penel S, Roth M, Chapus C and Fontecilla-Camps J C. *Neutron crystallographic evidence of lipase-colipase complex activation by a micelle*. EMBO J, **1997**, 16(18): 5531-5536.

Hinchliffe P, Symmons M F, Hughes C and Koronakis V. *Structure and operation of bacterial tripartite pumps*. Annu Rev Microbiol, **2013**, 67: 221-242.

Hoff W D, Jung K H and Spudich J L. *Molecular mechanism of photosignaling by archaeal sensory rhodopsins*. Annu Rev Biophys Biomol Struct, **1997**, 26: 223-258.

Horchani H, Ben Salem N, Chaari A, Sayari A, Gargouri Y and Verger R. *Staphylococcal lipases stereoselectively hydrolyse the sn-2 position of monomolecular films of diglyceride analogs. Application to sn-2 hydrolysis of triolein*. J Colloid Interface Sci, **2010**, 347(2): 301-308.

Houde A, Kademi A and Leblanc D. *Lipases and their industrial applications: an overview*. Appl Biochem Biotechnol, **2004**, 118(1-3): 155-170.

Iordanov I, Renault M, Reat V, Bosshart P D, Engel A, Saurel O and Milon A. *Dynamics of Klebsiella pneumoniae OmpA transmembrane domain: the four extracellular loops display restricted motion behavior in micelles and in lipid bilayers*. Biochim Biophys Acta, **2012**, 1818(9): 2344-2353.

Ishida H, Garcia-Herrero A and Vogel H J. *The periplasmic domain of Escherichia coli outer membrane protein A can undergo a localized temperature dependent structural transition*. Biochim Biophys Acta, **2014**, 1838(12): 3014-3024.

Jackups R, Jr. and Liang J. *Interstrand pairing patterns in beta-barrel membrane proteins: the positive-outside rule, aromatic rescue, and strand registration prediction*. J Mol Biol, **2005**, 354(4): 979-993.

Jaeger K E, Dijkstra B W and Reetz M T. *Bacterial biocatalysts: molecular biology, three-dimensional structures, and biotechnological applications of lipases*. Annu Rev Microbiol, **1999**, 53: 315-351.

Jaeger K E, Ransac S, Koch H B, Ferrato F and Dijkstra B W. *Topological characterization and modeling of the 3D structure of lipase from Pseudomonas aeruginosa*. FEBS Lett, **1993**, 332(1-2): 143-149.

Jagnow J and Clegg S. *Klebsiella pneumoniae MrkD-mediated biofilm formation on extracellular matrix- and collagen-coated surfaces*. Microbiology, **2003**, 149(Pt 9): 2397-2405.

Jansen C, Heutink M, Tommassen J and de Cock H. *The assembly pathway of outer membrane protein PhoE of Escherichia coli*. Eur J Biochem, **2000**, 267(12): 3792-3800.

Jeannin P, Bottazzi B, Sironi M, Doni A, Rusnati M, Presta M, Maina V, Magistrelli G, Haeuw J F, Hoeffel G, Thieblemont N, Corvaia N, Garlanda C, Delneste Y and Mantovani A. *Complexity and complementarity of outer membrane protein A recognition by cellular and humoral innate immunity receptors*. Immunity, **2005**, 22(5): 551-560.

Jiang J, Zhang X, Chen Y, Wu Y, Zhou Z H, Chang Z and Sui S F. *Activation of DegP chaperone-protease via formation of large cage-like oligomers upon binding to substrate proteins*. Proc Natl Acad Sci U S A, **2008**, 105(33): 11939-11944.

Jolivet P, Bordes F, Fudalej F, Cancino M, Vignaud C, Dossat V, Burghoffer C, Marty A, Chardot T and Nicaud J M. *Analysis of Yarrowia lipolytica extracellular lipase Lip2p glycosylation*. FEMS Yeast Res, **2007**, 7(8): 1317-1327.

Jung K H, Trivedi V D and Spudich J L. *Demonstration of a sensory rhodopsin in eubacteria*. Mol Microbiol, **2003**, 47(6): 1513-1522.

Kauffmann F. *A simplified biochemical table of Enterobacteriaceae*. Acta Pathol Microbiol Scand, **1956**, 39(2): 103-106.

Kay L E. *Solution NMR spectroscopy of supra-molecular systems, why bother? A methyl-TROSY view*. Journal of magnetic resonance, **2011**, 210(2): 159-170.

Keller J and Layer P. *Pancreatic Enzyme Supplementation Therapy*. Curr Treat Options Gastroenterol, **2003**, 6(5): 369-374.

Killian J A and von Heijne G. *How proteins adapt to a membrane-water interface*. Trends in Biochemical Sciences, **2000**, 25(9): 429-434.

Kim K K, Song H K, Shin D H, Hwang K Y and Suh S W. *The crystal structure of a triacylglycerol lipase from Pseudomonas cepacia reveals a highly open conformation in the absence of a bound inhibitor*. Structure, **1997**, 5(2): 173-185.

Kim S, Malinverni J C, Sliz P, Silhavy T J, Harrison S C and Kahne D. *Structure and function of an essential component of the outer membrane protein assembly machine*. Science, **2007**, 317(5840): 961-964.

Kim S Y, Yoon S R, Han S, Yun Y and Jung K H. *A role of Anabaena sensory rhodopsin transducer (ASRT) in photosensory transduction*. Mol Microbiol, **2014**, 93(3): 403-414.

Koebnik R, Locher K P and Van Gelder P. *Structure and function of bacterial outer membrane proteins: barrels in a nutshell*. Mol Microbiol, **2000**, 37(2): 239-253.

Kojima S, Furukawa Y, Matsunami H, Minamino T and Namba K. *Characterization of the periplasmic domain of MotB and implications for its role in the stator assembly of the bacterial flagellar motor*. J Bacteriol, **2008**, 190(9): 3314-3322.

Kojima S, Imada K, Sakuma M, Sudo Y, Kojima C, Minamino T, Homma M and Namba K. *Stator assembly and activation mechanism of the flagellar motor by the periplasmic region of MotB*. Mol Microbiol, **2009**, 73(4): 710-718.

Kollman J M, Greenberg C H, Li S, Moritz M, Zelter A, Fong K K, Fernandez J J, Sali A, Kilmartin J, Davis T N and Agard D A. *Ring closure activates yeast gamma TuRC for species-specific microtubule nucleation*. Nature structural & molecular biology, **2015**, 22(2): 132-137.

Kollman J M, Merdes A, Mourey L and Agard D A. *Microtubule nucleation by gamma-tubulin complexes*. Nature Reviews Molecular Cell Biology, **2011**, 12(11): 709-721.

Kondoh M, Inoue K, Sasaki J, Spudich J L and Terazima M. *Transient dissociation of the transducer protein from anabaena sensory rhodopsin concomitant with formation of the M state produced upon photoactivation*. J Am Chem Soc, **2011**, 133(34): 13406-13412.

Koplow J and Goldfine H. *Alterations in the outer membrane of the cell envelope of heptose-deficient mutants of Escherichia coli*. J Bacteriol, **1974**, 117(2): 527-543.

Korndorfer I P, Dommel M K and Skerra A. *Structure of the periplasmic chaperone Skp suggests functional similarity with cytosolic chaperones despite differing architecture*. Nat Struct Mol Biol, **2004**, 11(10): 1015-1020.

Koronakis V, Sharff A, Koronakis E, Luisi B and Hughes C. *Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export*. Nature, **2000**, 405(6789): 914-919.

Krishna M M G, Hoang L, Lin Y and Englander S W. *Hydrogen exchange methods to study protein folding*. Methods, **2004**, 34(1): 51-64.

Krojer T, Sawa J, Schafer E, Saibil H R, Ehrmann M and Clausen T. *Structural basis for the regulated protease and chaperone function of DegP*. Nature, **2008**, 453(7197): 885-890.

Kusters I and Driessen A J. *SecA, a remarkable nanomachine*. Cell Mol Life Sci, **2011**, 68(12): 2053-2066.

Lang D A, Mannesse M L, de Haas G H, Verheij H M and Dijkstra B W. *Structural basis of the chiral selectivity of Pseudomonas cepacia lipase*. Eur J Biochem, **1998**, 254(2): 333-340.

Latham M P, Sekhar A and Kay L E. *Understanding the mechanism of proteasome 20S core particle gating*. Proc Natl Acad Sci U S A, **2014**, 111(15): 5532-5537.

Lazar S W and Kolter R. *SurA assists the folding of Escherichia coli outer membrane proteins*. J Bacteriol, **1996**, 178(6): 1770-1773.

Li X Z and Nikaido H. *Efflux-mediated drug resistance in bacteria: an update*. Drugs, **2009**, 69(12): 1555-1623.

Lin Y C, Lu M C, Tang H L, Liu H C, Chen C H, Liu K S, Lin C J, Chiou C S, Chiang M K, Chen C M and Lai Y C. *Assessment of hypermucoviscosity as a virulence factor for experimental Klebsiella pneumoniae infections: comparative virulence analysis with hypermucoviscosity-negative strain*. BMC Microbiology, **2011**, 11.

Luic M, Stefanic Z, Ceilinger I, Hodoscek M, Janezic D, Lenac T, Asler I L, Sepac D and Tomic S. *Combined X-ray diffraction and QM/MM study of the Burkholderia cepacia lipase-catalyzed secondary alcohol esterification*. J Phys Chem B, **2008**, 112(16): 4876-4883.

Luic M, Tomic S, Lescic I, Ljubovic E, Sepac D, Sunjic V, Vitale L, Saenger W and Kojic-Prodic B. *Complex of Burkholderia cepacia lipase with transition state analogue of 1-phenoxy-2-acetoxybutane: biocatalytic, structural and modelling study*. Eur J Biochem, **2001**, 268(14): 3964-3973.

Lundstrom P, Vallurupalli P, Hansen D F and Kay L E. *Isotope labeling methods for studies of excited protein states by relaxation dispersion NMR spectroscopy*. Nature Protocols, **2009**, 4(11): 1641-1648.

Luthi-Peng Q and Winkler F K. *Large spectral changes accompany the conformational transition of human pancreatic lipase induced by acylation with the inhibitor tetrahydrolipstatin*. Eur J Biochem, **1992**, 205(1): 383-390.

Madzak C, Gaillardin C and Beckerich J M. *Heterologous protein expression and secretion in the non-conventional yeast Yarrowia lipolytica: a review*. J Biotechnol, **2004**, 109(1-2): 63-81.

March C, Cano V, Moranta D, Llobet E, Perez-Gutierrez C, Tomas J M, Suarez T, Garmendia J and Bengoechea J A. *Role of bacterial surface structures on the interaction of Klebsiella pneumoniae with phagocytes*. PLoS One, **2013**, 8(2): e56847.

March C, Moranta D, Regueiro V, Llobet E, Tomas A, Garmendia J and Bengoechea J A. *Klebsiella pneumoniae* Outer Membrane Protein A Is Required to Prevent the Activation of Airway Epithelial Cells. *Journal of Biological Chemistry*, **2011**, 286(12): 9956-9967.

Marcoux J, Politis A, Rinehart D, Marshall D P, Wallace M I, Tamm L K and Robinson C V. *Mass spectrometry defines the C-terminal dimerization domain and enables modeling of the structure of full-length OmpA*. *Structure*, **2014**, 22(5): 781-790.

Marx V. *Structural biology: 'seeing' crystals the XFEL way*. *Nature Methods*, **2014**, 11(9): 903-+.

Merdanovic M, Clausen T, Kaiser M, Huber R and Ehrmann M. *Protein quality control in the bacterial periplasm*. *Annu Rev Microbiol*, **2011**, 65: 149-168.

Merianos H J, Cadieux N, Lin C H, Kadner R J and Cafiso D S. *Substrate-induced exposure of an energy-coupling motif of a membrane transporter*. *Nat Struct Biol*, **2000**, 7(3): 205-209.

Mezzetti A, Schrag J D, Cheong C S and Kazlauskas R J. *Mirror-image packing in enantiomer discrimination molecular basis for the enantioselectivity of B.cepacia lipase toward 2-methyl-3-phenyl-1-propanol*. *Chem Biol*, **2005**, 12(4): 427-437.

Misra R, CastilloKeller M and Deng M. *Overexpression of protease-deficient DegP(S210A) rescues the lethal phenotype of Escherichia coli OmpF assembly mutants in a degP background*. *J Bacteriol*, **2000**, 182(17): 4882-4888.

Muller-Loennies S, Brade L and Brade H. *Neutralizing and cross-reactive antibodies against enterobacterial lipopolysaccharide*. *Int J Med Microbiol*, **2007**, 297(5): 321-340.

Munoz-Price L S, Poirel L, Bonomo R A, Schwaber M J, Daikos G L, Cormican M, Cornaglia G, Garau J, Gniadkowski M, Hayden M K, Kumarasamy K, Livermore D M, Maya J J, Nordmann P, Patel J B, Paterson D L, Pitout J, Villegas M V, Wang H, Woodford N and Quinn J P. *Clinical epidemiology of the global expansion of Klebsiella pneumoniae carbapenemases*. *Lancet Infect Dis*, **2013**, 13(9): 785-796.

Ni D, Wang Y, Yang X, Zhou H, Hou X, Cao B, Lu Z, Zhao X, Yang K and Huang Y. *Structural and functional analysis of the beta-barrel domain of BamA from Escherichia coli*. *FASEB J*, **2014**, 28(6): 2677-2685.

Niedermann G, Grimm R, Geier E, Maurer M, Realini C, Gartmann C, Soll J, Omura S, Rechsteiner M C, Baumeister W and Eichmann K. *Potential immunocompetence of proteolytic fragments produced by proteasomes before evolution of the vertebrate immune system*. *J Exp Med*, **1997**, 186(2): 209-220.

Nikaido H. *Molecular basis of bacterial outer membrane permeability revisited*. *Microbiol Mol Biol Rev*, **2003**, 67(4): 593-656.

Nikaido H and Vaara M. *Molecular basis of bacterial outer membrane permeability*. *Microbiol Rev*, **1985**, 49(1): 1-32.

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

Noinaj N, Kuszak A J, Gumbart J C, Lukacik P, Chang H, Easley N C, Lithgow T and Buchanan S K. *Structural insight into the biogenesis of beta-barrel membrane proteins*. *Nature*, **2013**, 501(7467): 385-390.

Noinville S, Revault M, Baron M H, Tiss A, Yapoudjian S, Ivanova M and Verger R. *Conformational changes and orientation of Humicola lanuginosa lipase on a solid hydrophobic surface: an in situ interface Fourier transform infrared-attenuated total reflection study*. Biophys J, **2002**, 82(5): 2709-2719.

Norin M, Haeffner F, Achour A, Norin T and Hult K. *Computer modeling of substrate binding to lipases from Rhizomucor miehei, Humicola lanuginosa, and Candida rugosa*. Protein Sci, **1994**, 3(9): 1493-1503.

Nygaard R, Zou Y Z, Dror R O, Mildorf T J, Arlow D H, Manglik A, Pan A C, Liu C W, Fung J J, Bokoch M P, Thian F S, Kobilka T S, Shaw D E, Mueller L, Prosser R S and Kobilka B K. *The Dynamic Process of beta(2)-Adrenergic Receptor Activation*. Cell, **2013**, 152(3): 532-542.

Oesterhelt D and Stoeckenius W. *Functions of a new photoreceptor membrane*. Proc Natl Acad Sci U S A, **1973**, 70(10): 2853-2857.

Ollis D L, Cheah E, Cygler M, Dijkstra B, Frolov F, Franken S M, Harel M, Remington S J, Silman I, Schrag J and et al. *The alpha/beta hydrolase fold*. Protein Eng, **1992**, 5(3): 197-211.

Palmer A G. *Chemical exchange in biomacromolecules: Past, present, and future*. Journal of magnetic resonance, **2014**, 241: 3-17.

Palmer A G. *Enzyme Dynamics from NMR Spectroscopy*. Accounts of Chemical Research, **2015**, 48(2): 457-465.

Palmer A G and Massi F. *Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy*. Chemical reviews, **2006**, 106(5): 1700-1719.

Pan Y J, Lin T L, Hsu C R and Wang J T. *Use of a Dictyostelium Model for Isolation of Genetic Loci Associated with Phagocytosis and Virulence in Klebsiella pneumoniae*. Infection and Immunity, **2011**, 79(3): 997-1006.

Park J S, Lee W C, Yeo K J, Ryu K S, Kumarasiri M, Hesek D, Lee M, Mobashery S, Song J H, Kim S I, Lee J C, Cheong C, Jeon Y H and Kim H Y. *Mechanism of anchoring of OmpA protein to the cell wall peptidoglycan of the gram-negative bacterial outer membrane*. FASEB J, **2012**, 26(1): 219-228.

Parsons L M, Lin F and Orban J. *Peptidoglycan recognition by Pal, an outer membrane lipoprotein*. Biochemistry, **2006**, 45(7): 2122-2128.

Paterson D L, Ko W C, Von Gottberg A, Mohapatra S, Casellas J M, Goossens H, Mulazimoglu L, Trenholme G, Klugman K P, Bonomo R A, Rice L B, Wagener M M, McCormack J G and Yu V L. *International prospective study Klebsiella pneumoniae bacteremia: Implications of extended-spectrum beta-lactamase production in nosocomial infections*. Annals of Internal Medicine, **2004**, 140(1): 26-32.

Pautsch A and Schulz G E. *Structure of the outer membrane protein A transmembrane domain*. Nat Struct Biol, **1998**, 5(11): 1013-1017.

Pautsch A and Schulz G E. *High-resolution structure of the OmpA membrane domain*. Journal of Molecular Biology, **2000**, 298(2): 273-282.

Pieterse W A, Vidal J C, Volwerk J J and de Haas G H. *Zymogen-catalyzed hydrolysis of monomeric substrates and the presence of a recognition site for lipid-water interfaces in phospholipase A2*. Biochemistry, **1974**, 13(7): 1455-1460.

Pignede G, Wang H, Fudalej F, Gaillardin C, Seman M and Nicaud J M. *Characterization of an extracellular lipase encoded by LIP2 in Yarrowia lipolytica*. J Bacteriol, **2000a**, 182(10): 2802-2810.

Pignede G, Wang H J, Fudalej F, Seman M, Gaillardin C and Nicaud J M. *Autocloning and amplification of LIP2 in Yarrowia lipolytica*. Appl Environ Microbiol, **2000b**, 66(8): 3283-3289.

Pignol D, Ayvazian L, Kerfelec B, Timmins P, Crenon I, Hermoso J, Fontecilla-Camps J C and Chapus C. *Critical role of micelles in pancreatic lipase activation revealed by small angle neutron scattering*. J Biol Chem, **2000**, 275(6): 4220-4224.

Pleiss J, Fischer M, Peiker M, Thiele C and Schmid R D. *Lipase engineering database - Understanding and exploiting sequence-structure-function relationships*. Journal of Molecular Catalysis B-Enzymatic, **2000**, 10(5): 491-508.

Pleiss J, Fischer M and Schmid R D. *Anatomy of lipase binding sites: the scissile fatty acid binding site*. Chem Phys Lipids, **1998**, 93(1-2): 67-80.

Podschun R and Ullmann U. *Klebsiella spp. as nosocomial pathogens: Epidemiology, taxonomy, typing methods, and pathogenicity factors*. Clinical Microbiology Reviews, **1998**, 11(4): 589-+.

Pogozheva I D, Mosberg H I and Lomize A L. *Life at the border: Adaptation of proteins to anisotropic membrane environment*. Protein Science, **2014**, 23(9): 1165-1196.

Pogozheva I D, Tristram-Nagle S, Mosberg H I and Lomize A L. *Structural adaptations of proteins to different biological membranes*. Biochim Biophys Acta, **2013**, 1828(11): 2592-2608.

Popovych N, Tzeng S R, Tonelli M, Ebright R H and Kalodimos C G. *Structural basis for cAMP-mediated allosteric control of the catabolite activator protein*. Proc Natl Acad Sci U S A, **2009**, 106(17): 6927-6932.

Postle K and Kadner R J. *Touch and go: tying TonB to transport*. Mol Microbiol, **2003**, 49(4): 869-882.

Prince S M, Achtman M and Derrick J P. *Crystal structure of the OpcA integral membrane adhesin from Neisseria meningitidis*. Proc Natl Acad Sci U S A, **2002**, 99(6): 3417-3421.

Putnam C D, Hammel M, Hura G L and Tainer J A. *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution*. Quarterly Reviews of Biophysics, **2007**, 40(3): 191-285.

Qu J, Behrens-Kneip S, Holst O and Kleinschmidt J H. *Binding regions of outer membrane protein A in complexes with the periplasmic chaperone Skp. A site-directed fluorescence study*. Biochemistry, **2009**, 48(22): 4926-4936.

Ranaldi S, Belle V, Woudstra M, Bourgeas R, Guigliarelli B, Roche P, Vezin H, Carriere F and Fournel A. *Amplitude of pancreatic lipase lid opening in solution and identification of*

spin label conformational subensembles by combining continuous wave and pulsed EPR spectroscopy and molecular dynamics. Biochemistry, **2010**, 49(10): 2140-2149.

Rasmussen S G F, Choi H J, Rosenbaum D M, Kobilka T S, Thian F S, Edwards P C, Burghammer M, Ratnala V R P, Sanishvili R, Fischetti R F, Schertler G F X, Weis W I and Kobilka B K. *Crystal structure of the human beta(2) adrenergic G-protein-coupled receptor.* Nature, **2007**, 450(7168): 383-U384.

Rasmussen S G F, DeVree B T, Zou Y Z, Kruse A C, Chung K Y, Kobilka T S, Thian F S, Chae P S, Pardon E, Calinski D, Mathiesen J M, Shah S T A, Lyons J A, Caffrey M, Gellman S H, Steyaert J, Skinotis G, Weis W I, Sunahara R K and Kobilka B K. *Crystal structure of the beta(2) adrenergic receptor-Gs protein complex.* Nature, **2011**, 477(7366): 549-U311.

Rawlings N D, Waller M, Barrett A J and Bateman A. *MEROPS: the database of proteolytic enzymes, their substrates and inhibitors.* Nucleic Acids Res, **2014**, 42(Database issue): D503-509.

Reat V, Patzelt H, Ferrand M, Pfister C, Oesterhelt D and Zaccai G. *Dynamics of different functional parts of bacteriorhodopsin: H-H-2 labeling and neutron scattering.* Proceedings of the National Academy of Sciences of the United States of America, **1998**, 95(9): 4970-4975.

Reis P, Holmberg K, Watzke H, Leser M E and Miller R. *Lipases at interfaces: a review.* Adv Colloid Interface Sci, **2009**, 147-148: 237-250.

Religa T L, Sprangers R and Kay L E. *Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR.* Science, **2010**, 328(5974): 98-102.

Renault M, Pawsey S, Bos M P, Koers E J, Nand D, Tommassen-van Boxtel R, Rosay M, Tommassen J, Maas W E and Baldus M. *Solid-State NMR Spectroscopy on Cellular Preparations Enhanced by Dynamic Nuclear Polarization.* Angewandte Chemie-International Edition, **2012a**, 51(12): 2998-3001.

Renault M, Saurel O, Czaplicki J, Demange P, Gervais V, Lohr F, Reat V, Piotto M and Milon A. *Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications.* J Mol Biol, **2009**, 385(1): 117-130.

Renault M, Saurel O, Demange P, Reat V and Milon A. *Solution-state NMR spectroscopy of membrane proteins in detergent micelles: structure of the Klebsiella pneumoniae outer membrane protein A, KpOmpA.* Methods Mol Biol, **2010**, 654: 321-339.

Renault M, Tommassen-van Boxtel R, Bos M P, Post J A, Tommassen J and Baldus M. *Cellular solid-state nuclear magnetic resonance spectroscopy.* Proceedings of the National Academy of Sciences of the United States of America, **2012b**, 109(13): 4863-4868.

Rezaei-Ghaleh N, Blackledge M and Zweckstetter M. *Intrinsically Disordered Proteins: From Sequence and Conformational Properties toward Drug Discovery.* Chembiochem, **2012**, 13(7): 930-950.

Ricci D P and Silhavy T J. *The Bam machine: a molecular cooper.* Biochim Biophys Acta, **2012**, 1818(4): 1067-1084.

Rietschel E T, Kirikae T, Schade F U, Mamat U, Schmidt G, Loppnow H, Ulmer A J, Zahringer U, Seydel U, Di Padova F and et al. *Bacterial endotoxin: molecular relationships of structure to activity and function.* FASEB J, **1994**, 8(2): 217-225.

Rizzitello A E, Harper J R and Silhavy T J. *Genetic evidence for parallel pathways of chaperone activity in the periplasm of Escherichia coli*. J Bacteriol, **2001**, 183(23): 6794-6800.

Rosenzweig R and Kay L E. *Bringing dynamic molecular machines into focus by methyl-TROSY NMR*. Annu Rev Biochem, **2014**, 83: 291-315.

Roujeinikova A. *Crystal structure of the cell wall anchor domain of MotB, a stator component of the bacterial flagellar motor: implications for peptidoglycan recognition*. Proc Natl Acad Sci U S A, **2008**, 105(30): 10348-10353.

Ruschak A M and Kay L E. *Methyl groups as probes of supra-molecular structure, dynamics and function*. Journal of Biomolecular Nmr, **2010a**, 46(1): 75-87.

Ruschak A M and Kay L E. *Proteasome allostery as a population shift between interchanging conformers*. Proceedings of the National Academy of Sciences of the United States of America, **2012**, 109(50): E3454-3462.

Ruschak A M, Religa T L, Breuer S, Witt S and Kay L E. *The proteasome antechamber maintains substrates in an unfolded state*. Nature, **2010b**, 467(7317): 868-871.

Ruschak A M, Slassi M, Kay L E and Schimmer A D. *Novel proteasome inhibitors to overcome bortezomib resistance*. J Natl Cancer Inst, **2011**, 103(13): 1007-1017.

Sahly H, Podschun R, Oelschlaeger T A, Greiwe M, Parolis H, Hasty D, Kekow J, Ullmann U, Ofek I and Sela S. *Capsule impedes adhesion to and invasion of epithelial cells by Klebsiella pneumoniae*. Infection and Immunity, **2000**, 68(12): 6744-6749.

Sansom M S and Kerr I D. *Transbilayer pores formed by beta-barrels: molecular modeling of pore structures and properties*. Biophys J, **1995**, 69(4): 1334-1343.

Sawyer L, Shotton D M, Campbell J W, Wendell P L, Muirhead H and Watson H C. *The atomic structure of crystalline porcine pancreatic elastase at 2.5 Å resolution: comparisons with the structure of alpha-chymotrypsin*. J Mol Biol, **1978**, 118(2): 137-208.

Schafer U, Beck K and Muller M. *Skp, a molecular chaperone of gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins*. J Biol Chem, **1999**, 274(35): 24567-24574.

Schanda P, Triboulet S, Laguri C, Bougault C M, Ayala I, Callon M, Arthur M and Simorre J P. *Atomic Model of a Cell-Wall Cross-Linking Enzyme in Complex with an Intact Bacterial Peptidoglycan*. Journal of the American Chemical Society, **2014**, 136(51): 17852-17860.

Schirmer T, Keller T A, Wang Y F and Rosenbusch J P. *Structural Basis for Sugar Translocation through Maltoporin Channels at 3.1-Ångstrom Resolution*. Science, **1995**, 267(5197): 512-514.

Schobert B and Lanyi J K. *Halorhodopsin is a light-driven chloride pump*. J Biol Chem, **1982**, 257(17): 10306-10313.

Schrag J D and Cygler M. *Lipases and alpha/beta hydrolase fold*. Methods Enzymol, **1997a**, 284: 85-107.

Schrag J D, Li Y, Cygler M, Lang D, Burgdorf T, Hecht H J, Schmid R, Schomburg D, Rydel T J, Oliver J D, Strickland L C, Dunaway C M, Larson S B, Day J and McPherson A. *The open conformation of a Pseudomonas lipase*. Structure, **1997b**, 5(2): 187-202.

Schroll C, Barken K B, Krogfelt K A and Struve C. *Role of type 1 and type 3 fimbriae in Klebsiella pneumoniae biofilm formation*. BMC Microbiol, **2010**, 10: 179.

Schulz G E. *Chapter 15 Structure-function relationships in porins as derived from a 1.8 Å resolution crystal structure*. **1994**, *New Comprehensive Biochemistry*, Ghuysen J M and Hakenbeck R, Elsevier, Volume 27: 343-352.

Schulz G E. *The structure of bacterial outer membrane proteins*. Biochimica Et Biophysica Acta-Biomembranes, **2002**, 1565(2): 308-317.

Shen Y, Delaglio F, Cornilescu G and Bax A. *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts*. J Biomol NMR, **2009**, 44(4): 213-223.

Shen Y, Lange O, Delaglio F, Rossi P, Aramini J M, Liu G, Eletsky A, Wu Y, Singarapu K K, Lemak A, Ignatchenko A, Arrowsmith C H, Szyperski T, Montelione G T, Baker D and Bax A. *Consistent blind protein structure generation from NMR chemical shift data*. Proc Natl Acad Sci U S A, **2008**, 105(12): 4685-4690.

Shi L, Kawamura I, Jung K H, Brown L S and Ladizhansky V. *Conformation of a seven-helical transmembrane photosensor in the lipid environment*. Angew Chem Int Ed Engl, **2011**, 50(6): 1302-1305.

Shimbara N, Nakajima H, Tanahashi N, Ogawa K, Niwa S, Uenaka A, Nakayama E and Tanaka K. *Double-cleavage production of the CTL epitope by proteasomes and PA28: role of the flanking region*. Genes Cells, **1997**, 2(12): 785-800.

Shon A S, Bajwa R P and Russo T A. *Hypervirulent (hypermucoviscous) Klebsiella pneumoniae: a new and dangerous breed*. Virulence, **2013**, 4(2): 107-118.

Sibille N, Favier A, Azuaga A I, Ganshaw G, Bott R, Bonvin A M, Boelens R and van Nuland N A. *Comparative NMR study on the impact of point mutations on protein stability of Pseudomonas mendocina lipase*. Protein Sci, **2006**, 15(8): 1915-1927.

Sigrist C J, de Castro E, Cerutti L, Cuče B A, Hulo N, Bridge A, Bougueleret L and Xenarios I. *New and continuing developments at PROSITE*. Nucleic Acids Res, **2013**, 41(Database issue): D344-347.

Skjot M, De Maria L, Chatterjee R, Svendsen A, Patkar S A, Ostergaard P R and Brask J. *Understanding the Plasticity of the alpha/beta Hydrolase Fold: Lid Swapping on the Candida antarctica Lipase B Results in Chimeras with Interesting Biocatalytic Properties*. Chembiochem, **2009**, 10(3): 520-527.

Sklar J G, Wu T, Kahne D and Silhavy T J. *Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli*. Genes Dev, **2007**, 21(19): 2473-2484.

Smit J, Kamio Y and Nikaido H. *Outer membrane of Salmonella typhimurium: chemical analysis and freeze-fracture studies with lipopolysaccharide mutants*. J Bacteriol, **1975**, 124(2): 942-958.

Smith S G, Mahon V, Lambert M A and Fagan R P. *A molecular Swiss army knife: OmpA structure, function and expression*. FEMS Microbiol Lett, **2007**, 273(1): 1-11.

Snijder H J, Ubarretxena-Belandia I, Blaauw M, Kalk K H, Verheij H M, Egmond M R, Dekker N and Dijkstra B W. *Structural evidence for dimerization-regulated activation of an integral membrane phospholipase*. Nature, **1999**, 401(6754): 717-721.

- Soykova Pachnerova E. *Effect of Thalidomide on the Pathogenesis of Abnormalities in Newborn Infants*. Lek Veda Zahr, **1963**, 34: 162-166.
- Spiess C, Beil A and Ehrmann M. *A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein*. Cell, **1999**, 97(3): 339-347.
- Sprangers R and Kay L E. *Quantitative dynamics and binding studies of the 20S proteasome by NMR*. Nature, **2007**, 445(7128): 618-622.
- Stenutz R, Weintraub A and Widmalm G. *The structures of Escherichia coli O-polysaccharide antigens*. FEMS Microbiol Rev, **2006**, 30(3): 382-403.
- Svendsen A. *Lipase protein engineering*. Biochim Biophys Acta, **2000**, 1543(2): 223-238.
- Thanabalu T, Koronakis E, Hughes C and Koronakis V. *Substrate-induced assembly of a contiguous channel for protein export from E.coli: reversible bridging of an inner-membrane translocase to an outer membrane exit pore*. EMBO J, **1998**, 17(22): 6487-6496.
- Torchia D A. *NMR studies of dynamic biomolecular conformational ensembles*. Progress in Nuclear Magnetic Resonance Spectroscopy, **2015**, 84: 14-32.
- Trodler P, Schmid R D and Pleiss J. *Modeling of solvent-dependent conformational transitions in Burkholderia cepacia lipase*. BMC Struct Biol, **2009**, 9: 38.
- Tugarinov V, Sprangers R and Kay L E. *Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy*. Journal of the American Chemical Society, **2007**, 129(6): 1743-1750.
- Turki S, Jabloun Z, Mrabet G, Marouani A, Thonart P, Diouani M F, Ben Abdallah F, Amara A, Rejeb A and Kallel H. *Preliminary safety assessment of Yarrowia lipolytica extracellular lipase: results of acute and 28-day repeated dose oral toxicity studies in rats*. Food Chem Toxicol, **2010**, 48(8-9): 2393-2400.
- Tzeng S R and Kalodimos C G. *Protein activity regulation by conformational entropy*. Nature, **2012**, 488(7410): 236-240.
- Ulmschneider M B and Sansom M S P. *Amino acid distributions in integral membrane protein structures*. Biochimica Et Biophysica Acta-Biomembranes, **2001**, 1512(1): 1-14.
- van den Bedem H and Fraser J S. *Integrative, dynamic structural biology at atomic resolution-it's about time*. Nature Methods, **2015**, 12(4): 307-318.
- Vandeputte-Rutten L, Kramer R A, Kroon J, Dekker N, Egmond M R and Gros P. *Crystal structure of the outer membrane protease OmpT from Escherichia coli suggests a novel catalytic site*. Embo Journal, **2001**, 20(18): 5033-5039.
- Vergar R. *Interfacial enzyme kinetics of lipolysis*. Annu Rev Biophys Bioeng, **1976**, 5: 77-117.
- Vogeley L, Sineshchikov O A, Trivedi V D, Sasaki J, Spudich J L and Luecke H. *Anabaena sensory rhodopsin: a photochromic color sensor at 2.0 Å*. Science, **2004**, 306(5700): 1390-1393.
- Vogeli B. *The nuclear Overhauser effect from a quantitative perspective*. Progress in Nuclear Magnetic Resonance Spectroscopy, **2014**, 78: 1-46.

Vogt J and Schulz G E. *The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence*. Structure with Folding & Design, **1999**, 7(10): 1301-1309.

Vollmer W. *Bacterial outer membrane evolution via sporulation?* Nat Chem Biol, **2012**, 8(1): 14-18.

Vollmer W, Blanot D and de Pedro M A. *Peptidoglycan structure and architecture*. FEMS Microbiol Rev, **2008a**, 32(2): 149-167.

Vollmer W, Joris B, Charlier P and Foster S. *Bacterial peptidoglycan (murein) hydrolases*. FEMS Microbiol Rev, **2008b**, 32(2): 259-286.

Vollmer W and Seligman S J. *Architecture of peptidoglycan: more data and more models*. Trends Microbiol, **2010**, 18(2): 59-66.

von Heijne G. *Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule*. J Mol Biol, **1992**, 225(2): 487-494.

Walton T A, Sandoval C M, Fowler C A, Pardi A and Sousa M C. *The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains*. Proc Natl Acad Sci U S A, **2009**, 106(6): 1772-1777.

Wandersman C and Delepelaire P. *TolC, an Escherichia coli outer membrane protein required for hemolysin secretion*. Proc Natl Acad Sci U S A, **1990**, 87(12): 4776-4780.

Wang S and Ladizhansky V. *Recent advances in magic angle spinning solid state NMR of membrane proteins*. Prog Nucl Magn Reson Spectrosc, **2014a**, 82: 1-26.

Wang S, Munro R A, Kim S Y, Jung K H, Brown L S and Ladizhansky V. *Paramagnetic relaxation enhancement reveals oligomerization interface of a membrane protein*. J Am Chem Soc, **2012**, 134(41): 16995-16998.

Wang S, Munro R A, Shi L, Kawamura I, Okitsu T, Wada A, Kim S Y, Jung K H, Brown L S and Ladizhansky V. *Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein*. Nat Methods, **2013a**, 10(10): 1007-1012.

Wang S, Shi L, Kawamura I, Brown L S and Ladizhansky V. *Site-specific solid-state NMR detection of hydrogen-deuterium exchange reveals conformational changes in a 7-helical transmembrane protein*. Biophys J, **2011**, 101(3): L23-25.

Wang S, Shi L, Okitsu T, Wada A, Brown L S and Ladizhansky V. *Solid-state NMR (1)(3)C and (1)(5)N resonance assignments of a seven-transmembrane helical protein Anabaena Sensory Rhodopsin*. Biomol NMR Assign, **2013b**, 7(2): 253-256.

Wang S L and Ladizhansky V. *Recent advances in magic angle spinning solid state NMR of membrane proteins*. Progress in Nuclear Magnetic Resonance Spectroscopy, **2014b**, 82: 1-26.

Ward M E, Brown L S and Ladizhansky V. *Advanced solid-state NMR techniques for characterization of membrane protein structure and dynamics: application to Anabaena Sensory Rhodopsin*. J Magn Reson, **2015a**, 253: 119-128.

Ward M E, Wang S, Munro R, Ritz E, Hung I, Gor'kov P L, Jiang Y, Liang H, Brown L S and Ladizhansky V. *In situ structural studies of Anabaena sensory rhodopsin in the E. coli membrane*. Biophys J, **2015b**, 108(7): 1683-1696.

Widmann M, Juhl P B and Pleiss J. *Structural classification by the Lipase Engineering Database: a case study of Candida antarctica lipase A*. BMC Genomics, **2010**, 11: 123.

Wildes D and Marqusee S. *Hydrogen-exchange strategies applied to energetics of intermediate processes in protein folding*. Energetics of Biological Macromolecules, Pt E, **2004**, 380: 328-349.

Wimley W C. *The versatile beta-barrel membrane protein*. Curr Opin Struct Biol, **2003**, 13(4): 404-411.

Winkler F K, D'Arcy A and Hunziker W. *Structure of human pancreatic lipase*. Nature, **1990**, 343(6260): 771-774.

Wu J H, Wu A M, Tsai C G, Chang X Y, Tsai S F and Wu T S. *Contribution of fucose-containing capsules in Klebsiella pneumoniae to bacterial virulence in mice*. Exp Biol Med (Maywood), **2008**, 233(1): 64-70.

Wu T, Malinverni J, Ruiz N, Kim S, Silhavy T J and Kahne D. *Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli*. Cell, **2005**, 121(2): 235-245.

Xu X, Wang S, Hu Y X and McKay D B. *The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues*. J Mol Biol, **2007**, 373(2): 367-381.

Yang D and Kay L E. *Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding*. J Mol Biol, **1996**, 263(2): 369-382.

Yang Y, Auguin D, Delbecq S, Dumas E, Molle G, Molle V, Roumestand C and Saint N. *Structure of the Mycobacterium tuberculosis OmpATb protein: a model of an oligomeric channel in the mycobacterial cell wall*. Proteins, **2011**, 79(2): 645-661.

Yapoudjian S, Ivanova M G, Brzozowski A M, Patkar S A, Vind J, Svendsen A and Verger R. *Binding of Thermomyces (Humicola) lanuginosa lipase to the mixed micelles of cis-parinaric acid/NaTDC*. Eur J Biochem, **2002**, 269(6): 1613-1621.

Yu M R, Qin S W and Tan T W. *Purification and characterization of the extracellular lipase Lip2 from Yarrowia lipolytica*. Process Biochemistry, **2007**, 42(3): 384-391.

Chapitre 2

**Expression de la lipase lip2 de *Yarrowia lipolytica* en vue
d'une étude par RMN en solution.**

Résumé

Le marquage de la lipase Lip2 de *Yarrowia lipolytica* est une étape indispensable à son étude par RMN. Ici, l'expression uniformément marquée ^{15}N de Lip2 a été étudiée dans plusieurs systèmes d'expression. L'expression dans le système procaryote *Escherichia coli* n'a pas permis d'obtenir de forme active de Lip2, que ce soit par une expression classique *in cellulo* ou par une approche cell-free. Le passage dans des hôtes eucaryotes que sont les levures *Pichia pastoris* et *Yarrowia lipolytica* s'est révélé bénéfique pour obtenir une lipase à la fois soluble et active. Toutefois, la lipase exprimée dans *Pichia pastoris* a montré une plus faible activité spécifique que celle exprimée dans *Yarrowia lipolytica* ce qui s'est traduit par la présence de deux populations sur le spectre HSQC ^{15}N - ^1H , une majoritaire probablement mal repliée et une minoritaire correctement repliée. L'expression dans *Yarrowia lipolytica* a permis d'obtenir la plus haute activité spécifique, donnant le premier spectre HSQC ^{15}N - ^1H de Lip2 majoritairement repliée correctement. Ce travail a été publié sous : **Nars, G., Saurel, O., Bordes, F., Saves, I., Remaud-Simeon, M., Andre, I., Milon, A., and Marty, A.,** *Production of stable isotope labelled lipase Lip2 from Yarrowia lipolytica for NMR: investigation of several expression systems.* Protein Expr Purif, **2014**, 101: 14-20.

Production of stable isotope labelled lipase Lip2 from *Yarrowia lipolytica* for NMR: investigation of several expression systems

G. Nars¹⁻⁴, O. Saurel⁴, F. Bordes¹⁻³, I. Saves⁴, M. Remaud-Siméon¹⁻³, I. André¹⁻³, A. Milon⁴ and A. Marty¹⁻³*

¹ Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France

² INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse, France

³ CNRS, UMR5504, F-31400 Toulouse, France

⁴ CNRS, IPBS UMR 5089, Institut de Pharmacologie et de Biologie Structurale, 205 route de Narbonne, BP 64182, F-31077 Toulouse, Cedex 04, France

*Corresponding author at: Université de Toulouse, INSA, UPS, INP, LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, France. Tel.: +33 (0) 6 72 31 04 01. E-mail address: alain.marty@insa-toulouse.fr (A. Marty).

Abstract

Extracellular lipase Lip2 from *Yarrowia lipolytica* is a promising biocatalyst with unusual structural features, as indicated by x-ray crystallography. These features comprise a mobile domain called the lid that controls access to the catalytic site. Conformational rearrangements of the lid have been suggested to regulate lipase enzymatic activities. We used nuclear magnetic resonance to investigate the dynamics of Lip2 by exploring four expression systems, *Escherichia coli*, cell-free, *Pichia pastoris* and *Yarrowia lipolytica* to produce uniformly labelled enzyme. The expression of Lip2 was assessed by determining its specific activity and measuring ¹⁵N-¹H HSQC spectra. *Yarrowia lipolytica* turned out to be the most efficient expression system. Here, we report the first use of *Yarrowia lipolytica* as an expression host for the production of uniform stable isotopic labelled protein for further structural and dynamics studies using NMR.

Keywords

Yarrowia lipolytica lipase lip2

Cell-free

Yarrowia lipolytica

Isotope labelling

Yeast

I. Introduction

Lipases (triacylglycerol acyl hydrolase EC 3.1.1.3) are enzymes that naturally hydrolyze the ester bonds of triglycerides. In favorable thermodynamic conditions (i.e. low water activity), they are also able to catalyze synthesis reactions on various substrates with high selectivity. The versatility of these enzymes makes them very attractive for many biotechnological applications. Amongst lipases, *Yarrowia lipolytica* lipase Lip2, a protein of 37 kDa, exhibits the highest activity on long-chain triglycerides reported to date (Aloulou *et al.* 2007b). It is also a good candidate for treatment of pancreatic exocrine insufficiency, as Lip2 remains highly active at low pH and it is not inhibited by bile salts. Its selectivity towards many different substrates and in particular, its ability to resolve racemic mixtures of 2-bromo-arylacetic acid esters (Guieysse *et al.* 2004), an important class of pharmaceutical intermediates, encourage its use in chemistry. Investigation of the structural determinants involved in its selectivity has already enabled the development of highly selective catalysts (Bordes *et al.* 2009, Cambon *et al.* 2010). Recently, the three-dimensional structure of Lip2 was identified by X-ray crystallography (Bordes *et al.* 2010, Aloulou *et al.* 2013) (PDB: 3O0D). Lip2 adopts the typical α/β hydrolase fold observed in lipases belonging to the fungal lipase family. An important structural feature of many lipases is the mobile α helix, called lid. Depending on its conformation, the lid controls access to the catalytic site, which can be either closed or open. Unfortunately, all X-ray structures of Lip2 obtained to date showed the lid in the catalytically inactive closed conformation (Bordes *et al.* 2010, Aloulou *et al.* 2013), consequently limiting insight into the catalytic mechanism. Indeed, the activation mechanism by which the lid switches from a closed to an open conformation is still not fully understood. Molecular dynamics studies suggested a complex lid opening motion involving both interactions with the substrate and with the lipid/water interface (Bordes *et al.* 2010). A better understanding of this process is crucial for the design of optimized lipase catalysts.

Further biophysical studies are required to be able to characterize the lid molecular motion of Lip2 lipase in more detail. To gain insight into protein dynamics, NMR spectroscopy provides a broad range of approaches that make it possible to sample protein dynamics at ps to ms time scales (Mittermaier *et al.* 2006, Osawa *et al.* 2012). To do so, one relies heavily on 3D NMR and stable isotope labeling schemes, using ^2H , ^{13}C and ^{15}N , for determination of the structure and chemical shift assignment procedures, and for relaxation studies that enable access to spatially defined motion time scales and order parameters (Gardner *et al.* 1998).

For this purpose, well-defined minimal media are crucial for the production of recombinant proteins. Among the expression systems available for the production of labeled samples,

recombinant *Escherichia coli* is the most widely used thanks to the high-level of protein expression and moderate cost of the growth media. However, as a prokaryotic expression system, it suffers from some drawbacks. The overproduction of eukaryotic recombinant protein often leads to misfolded protein aggregates in insoluble inclusion bodies. One possible way to overcome this limitation is the co-expression of molecular chaperones to favor correct folding and hence the solubility of the recombinant protein (de Marco *et al.* 2007). Another problem is that the formation of the disulfide bond, which requires an oxidative environment, is not compatible with the bacterial low intracellular redox state. Even if some strains have been engineered by deleting the glutathione reductase (*gor*) and the thioredoxine reductase (*trxB*) leading to an altered cytoplasmic reducing potential (Samuelson 2011), it remains challenging and often leads to the formation of insoluble inclusion bodies. An alternative strategy for the production of proteins containing disulfide bonds is the use of an acellular transcription/translation system. The cell-free reaction allows the direct supplementation of the reaction mixture with additives like metal ions, detergents, cofactors and/or binding partners. The redox potential can be fine tuned during the reaction to favor formation of the disulfide bond (Michel *et al.* 2012).

An alternative option is the use of recombinant eukaryotic hosts, including mammalian cells (Dutta *et al.* 2012) or baculovirus-infected insect cells (Saxena *et al.* 2012) with the necessary post-translational mechanisms. However, this generally leads to a low protein yield and the production of isotopic labeled samples for NMR studies would be expensive, because of the high cost of a rich culture medium. Yeasts thus appear to be the most attractive host as they are able to grow in low-cost minimal medium using ammonium chloride as nitrogen source, and glucose, glycerol, or methanol as carbon source. Yeasts combine the ease of use of *E. coli* with higher protein expression yields. Moreover, protein secretion and crucial post-translational modifications such as the formation of disulfide bonds and glycosylation are effective. The non-conventional methylotrophic yeast *Pichia pastoris* has already been used to produce large quantities of uniform $^{15}\text{N}^{13}\text{C}$ double labeled proteins using ^{15}N ammonium chloride and ^{13}C methanol (Wood *et al.* 1999). *Kluyveromyces lactis* is another yeast that was recently used to produce uniform $^{15}\text{N}^{13}\text{C}$ double labeled proteins like in *Pichia pastoris* but by replacing the ^{13}C methanol carbon source by ^{13}C glucose (Sugiki *et al.* 2008).

Lip2 lipase is a glycosylated protein containing four disulfide bonds. Lip2 lipase is secreted by the non-conventional yeast *Yarrowia lipolytica*, which has not been used for the production of isotopically labeled proteins to date. In the present study, we compared the expression of Lip2 in four different expression systems and checked correct folding and stability using ^1H - ^{15}N HSQC, a prerequisite step for NMR structural studies and for the characterization of the lid motions upon activation of Lip2.

We tested and compared the four most promising strategies for protein expression and labelling: (i) homologous expression in *Y. lipolytica* (Pignede *et al.* 2000), (ii) heterologous expression in the yeast *P. pastoris* [1] and (iii) in *E. coli*, the most widely used expression system for labeled proteins (Ohki *et al.* 2008), and (iv) cell free expression, which has proven to be very efficient while allowing amino-acid specific labeling (Ozawa *et al.* 2005). For reasons we discuss below, homologous expression turned out to be the most efficient strategy for Lip2.

II. Material and methods

2.1. Chemicals

Restriction enzymes, phusion-HF polymerase, and T4-DNA-ligase were purchased from New England Biolabs (NEB) (Ipswich, MA, US); the primers came from Eurogentec (Liège, Belgium); tryptone, peptone, yeast extract, yeast nitrogen base (YNB) with or without ammonium sulfate and without amino acids were purchased from BD Difco (Franklin Lakes, NJ, US). Unless stated otherwise, other chemicals, were purchased from Sigma-Aldrich (St. Louis, MO, US). ¹⁵N labeled ammonium sulfate was purchased from Cambridge Isotope Laboratories (Tewksbury, MA, US).

2.2. Strains and culture conditions

Escherichia coli strain DH5 α (Life Technologies, Carlsbad, CA, US) was used for vector construction and amplification. Strains BL21(DE3) (Clontech, Mountain View, CA, US) and Shuffle T7 (NEB, Ipswich, MA, US) were used for expression. The BL21(Star) strain and Rosetta2 strains used for the preparation of cell-free extract were purchased respectively from Life Technology (Carlsbad, CA, US) and from Merck Millipore (Molsheim, France). *E. coli* strains were grown at 37 °C or 30 °C in LB medium (1% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl with 1.5% (w/v) solid agar) supplemented with the corresponding antibiotic, 40 mg/L of kanamycin, 25 mg/L of Zeocin, or 100 mg/L of Ampicillin.

Pichia pastoris strains X-33 and SMD1168H (Life Technologies, Carlsbad, CA, US) positive transformants were selected on YPDS solid medium (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose and 1 M sorbitol containing 1.5% (w/v) agar) supplemented with 100 mg/L of Zeocin. The transformants were screened to identify those producing the highest lipase activity using the following procedure: colonies were picked and cultured in 25 mL of BMGY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM K₂HPO₄/KH₂PO₄ pH 6.0, 1.34% (w/v) YNB with ammonium sulfate and without amino acids,

4 x 10⁻⁵% (w/v) biotin and 1% (v/v) glycerol) in a 250 mL baffled Erlenmeyer flask at 28 °C under stirring at 120 rpm. After the culture reached an OD₆₀₀ = 6, cells were harvested by centrifugation and resuspended in 25 mL of BMMY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM K₂HPO₄/KH₂PO₄ pH 6.0, 1.34% (w/v) YNB with ammonium sulfate and without amino acids, 4 x 10⁻⁵% (w/v) biotin and 1% (v/v) methanol) to trigger expression. One percent methanol (v/v) was added every day and OD₆₀₀, lipase activity and secretion were measured throughout the culture period.

The construction of the Yarrowia lipolytica strain JMY1212 – JMP61-pTEF-LIP2 used in this work is described in (Bordes et al. 2010).

2.3. Vector construction

Standard molecular genetic techniques were used (Sambrook et al. 1989). The DNA sequence encoding the *Yarrowia lipolytica* lipase Lip2 sequence (Uniprot: Q9P8F7) was amplified by polymerase chain reaction (PCR) using JMP61-LIP2 vector from (Bordes et al. 2010) as template. Primers were designed to amplify the 906 bp sequence encoding the mature extracellular lipase of the LIP2 gene without the 99 bp sequence encoding the signal peptide from *Y. lipolytica*. All primer sequences are listed in Supplementary Table S1 and maps of the different vectors constructed are given in Supplementary Figure S1. For all constructions, sequences were checked by DNA sequencing (GATC-Biotech, Konstanz, Germany) to ensure that no mutation was introduced during the PCR.

2.3.1. *Escherichia coli* vector

For expression in *Escherichia coli*, the forward primer aLip2_ECf and the reverse primer aLip2_ECr were used to introduce *Nde*I and *Not*I restriction sites, respectively (Supplementary Table S1). The PCR product was digested with these enzymes and ligated to the vector pET-26b(+) digested with the same restriction enzymes and subsequently gel purified. The construction pET26b-Lip2 introduced an additional N-terminal methionine in the mature Lip2 sequence (Table 1). The resulting pET26b-Lip2 construction was transformed into chemically competent *E. coli* strains BL21(DE3) and Shuffle T7 *E. coli* strains. Positive transformants were selected on LB agar plate supplemented with 40 mg/L of kanamycin.

2.3.2. Cell-free vector

Two constructions were tested for cell free expression. The forward primer, aHisLip2_CFF, was used as an alternative to the forward primer aLip2_ECf, which introduces the *Nde*I restriction site. The aHisLip2_CFF primer introduces a sequence coding for a histidine tag followed by the factor Xa cleavage site in addition to the *Nde*I restriction

site. The reverse primer aLip2_CFr was used to introduce the *Xma*I restriction site (Supplementary Table S1). The PCR products were digested with *Nde*I and *Xma*I and ligated into the vector PIVEX2.3d digested with the same restriction enzymes and subsequently gel purified. The resulting PIVEX-Lip2 construction codes for the mature Lip2 lipase with an additional methionine at the N-terminal end and the PIVEX-HisLip2 construction codes for the mature Lip2 lipase with an additional N-terminal sequence MKSGHHHHHSSGIEGR containing the his-tag and the factor Xa cleavage site (Table 1). The two vector constructs were transformed into *E. coli* DH5 α and selected on LB medium supplemented with 100 mg/L of ampicillin.

2.3.3. *Pichia pastoris* vector

For expression in *Pichia pastoris*, the wild type *LIP2* gene was amplified using the forward primer aLip2_PPf and the reverse primer aLip2_PPr to introduce *Eco*RI and *Xba*I restriction sites, respectively (Supplementary Table S1). The PCR product was digested with these enzymes and ligated to the vector PICZ α A digested by the same restriction enzymes and gel purified. The vector pPICZ α A-Lip2 encodes a fusion of the mature Lip2 with the N-terminal α -factor prepro signal sequence from *Saccharomyces cerevisiae* (Table 1). This construction adds two additional amino acids (glutamic acid and phenylalanine) derived from the *Eco*RI site in pPICZ α A vector (Yu *et al.* 2007) at the N-terminal extremity of the sequence of the mature Lip2 lipase. After amplification of the vector in DH5 α *E. coli* strain in LB supplemented with 25 mg/L of Zeocin, the pPICZ α A-Lip2 vector was linearized by *Pme*I and transformed by electroporation into *P. pastoris* strains X-33 or SMD1168H. Electro-competent cells were obtained following the condensed protocol described in (Lin-Cereghino *et al.* 2005).

2.4. Protein expression

2.4.1. Expression in *Escherichia coli*

One colony of transformed *E. coli* was resuspended in 10 mL LB medium containing 50 mg/L ampicillin and incubated with 160 rpm agitation at 30 °C or 37 °C for BL21(DE3) and 30 °C for Shuffle T7. When the OD₆₀₀ reached 0.6 – 0.8, expression of the lipase was induced by adding IPTG at a concentration of 0.1 or 1 mM for BL21(DE3) and 1 mM for Shuffle T7. After 4 h of induction, cells were collected by centrifugation at 10,000 x *g* for 30 min and the bacterial pellet was stored at -80 °C, overnight. The pellet was then thawed at room temperature and resuspended in 50 mM Tris-HCl pH 7.5, 300 mM NaCl and 0.25 mg/mL lysozyme for 5 min. Cells were lysed on ice by sonication and the soluble fraction was isolated by centrifugation 10,000 x *g* for 30 min.

2.4.2. Cell-free expression

The cell extract for cell-free expression was homemade. The protocols described in (Schneider *et al.* 2010) were used as a starting point to prepare the S30 extract and the cell-free expression reaction.

Three liters of YTPG medium (15 g/L yeast extract, 24 g/L tryptone, 7.5 g/L NaCl, 2.89 g/L KH_2PO_4 , 9.13 g/L K_2HPO_4 , 100 mM glucose) were inoculated at a $\text{OD}_{600} = 0.05$ with a culture of BL21(Star) transformed with the plasmid pAR1219 (Sigma, St. Louis, MO, US) to express the T7 RNA polymerase. The strain Rosetta2 co-transformed with the plasmid pAR1219 and pRDKJG (Caspers *et al.* 1994) was used for the S30 extract enriched with chaperones. The temperature was set at 37 °C and agitation at 500 rpm. When the OD_{600} reached 0.6, 1 mM IPTG was added to trigger the expression of the T7 RNA polymerase. The culture was stopped during the exponential growth stage, and placed on ice. Next, cells were harvested by centrifugation at 6000 x g for 15 min. The cell pellet was washed two times in 500 mL, and in 200 mL of S30-A buffer (10 mM Tris-Acetate pH 8.2, 14 mM Mg^{2+} acetate, 0.6 mM KCl, 6 mM β -mercaptoethanol) and was finally resuspended in 1 mL S30-B buffer (10 mM Tris-Acetate pH 8.2, 14 mM Mg^{2+} acetate, 0.6 mM KCl, 1 mM DTT, 0.1 mM PMSF, 1 mg/mL lysozyme, 0.05% tween 20) per gram of cells, and frozen by immersion in liquid nitrogen and stored overnight at -80 °C. The lysate was then thawed on ice and centrifuged at 30,000 x g for 30 min. The supernatant was collected and adjusted to a NaCl concentration of 400 mM by adding a solution of 5 M NaCl. The cleared lysate was then incubated at 42 °C for 45 min and dialyzed with a 12 – 14 kDa molecular weight cut-off (MWCO) dialysis membrane against 5 L of S30-C buffer (10 mM Tris-Acetate pH 8.2, 14 mM Mg^{2+} acetate, 0.6 mM K^+ acetate, 0.5 mM DTT) for 2 h then overnight. After centrifugation at 30,000 x g for 30 min, the supernatant was collected, fractionated in aliquots of 200 μL , frozen by immersion in liquid nitrogen and stored at -80 °C until further use.

A continuous exchange cell-free expression (CECF) (Schneider *et al.* 2010) system based on *E. coli* extracts was used for the expression of Lip2. Analytical reactions were performed in a Mini-CECF-Reactor with a reaction mixture (RM) volume of 60 μL . The RM was separated from the feeding mixture (FM) by a membrane with a 10 kDa MWCO. For all reactions, the RM:FM ratio was 1:14 and all reactions were carried out at 30 °C for 18 h with gentle shaking. The GFP was used as a test protein to determine optimal conditions and as a control.

2.4.3. Expression in *Pichia pastoris*

An overnight pre-culture at 120 rpm agitation at 28 °C in 10 mL YPD medium (1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v) dextrose) in a 100 mL Erlenmeyer flask was used

to inoculate 10 mL of ^{15}N labeled BMG medium (100 mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$ pH 6.0, 0.34% (w/v) YNB without ammonium sulfate and without amino acids, $4 \times 10^{-5}\%$ (w/v) biotin, 0.5% (w/v) $^{15}\text{NH}_4\text{Cl}$ and 1% (v/v) glycerol) in a 100 mL Erlenmeyer flask at $\text{OD}_{600} = 1$. After approximately 24 h of incubation at 28 °C at 120 rpm agitation, when OD_{600} reached 20, pH was checked and adjusted to 6.0 with 1 M NaOH and induction was done by adding 1% (v/v) methanol. After 24 h of induction, cells were harvested by centrifugation $10,000 \times g$ for 20 min and 100 mL of supernatant was collected and stored at 4 °C.

2.4.4. Expression in *Yarrowia lipolytica*

Cultivation was performed as described previously in the patent (Leblond *et al.* 2007) in a 10 L reactor (BIOSTAT C-plus Sartorius) containing 3 L basal salt medium (BSM) and 100 g/L of glucose. The original synthetic medium was optimized for expression of Lip2 using a constitutive promoter and ^{15}N labeled protein expression. The $(^{15}\text{NH}_4)_2\text{SO}_4$ of the BSM was used as ^{15}N labeled nitrogen source and the BSM composition did not include glutamate. The 14% ammonia solution used as base to regulate the pH at 6.0 was replaced by a solution of KOH 5 M. The temperature was set at 28 °C and the agitation rate was set at 300 rpm. The bioreactor culture was inoculated with a culture of 200 mL minimal medium (200 mL BSM with 100 mM Na/KPO₄, pH 6.8 and 30 g/L glucose) in a 3 L Erlenmeyer flask at 28 °C and at 120 rpm agitation for 24 h. Culture was performed in batch mode and cell growth was monitored by the OD_{600} and the decrease in the partial pressure in oxygen. The agitation rate was increased to 1000 rpm and $(^{15}\text{NH}_4)_2\text{SO}_4$ was added by pulses of 10 g/L in order to maintain cell growth throughout fermentation. At the end of fermentation, a total of 50 g $(^{15}\text{NH}_4)_2\text{SO}_4$ was added and the supernatant was collected by centrifugation at $10,000 \times g$ for 20 min.

The supernatant was kept for further purification and the cell pellet was used to prepare a solution of ^{15}N -yeast extract. Cells were resuspended in 3 L of water and autolysis was performed by heating the yeasts at 50 °C with stirring for 24 h. After centrifugation at $10,000 \times g$ for 30 min, the supernatant was collected and the pH was adjusted to 7.0. The solution was sterilized at 121 °C for 20 min and 10 mL of sample was lyophilized and weighed.

Cultivation in an Erlenmeyer flask was performed at 28 °C and at 120 rpm agitation. An overnight pre-culture of 25 mL YPD in a 250 mL baffled Erlenmeyer flask was used to inoculate 200 mL of ^{15}N -yeast extract dextrose medium (^{15}N -YD) (20 g/L ^{15}N -yeast extract solution with 100 mM Na/KPO₄, pH 6.8 and 30 g/L glucose) in a 2 L Erlenmeyer flask at an $\text{OD}_{600} = 1$. After 24 h, cells were harvested by centrifugation at $10,000 \times g$ for 20 min and the supernatant was collected.

2.5. Characterization of the expressed lipase

2.5.1. Lipase activity

Lipase activity was determined by monitoring the hydrolysis of para-nitrophenolbutyrate (pNPB) into butyrate and para-nitrophenol (Quinn *et al.* 1982). The optimized method is described in (Bordes *et al.* 2007). One unit of lipase activity was defined as the amount of enzyme releasing 1 μ mol of para-nitrophenol per min at 30 °C and pH 7.2.

2.5.2. Protein quantification

Protein concentration was determined using the Bio-Rad Protein Assay kit (Bio-Rad) according to the manufacturer's instructions using bovine serum albumin (BSA) as standard (Bradford 1976).

2.5.3. SDS-PAGE analysis

Samples were loaded onto Any kD Mini-PROTEAN TGX gel (Bio-Rad, Hercules, CA, USA). Precision Plus Protein™ All Blue Standards (Bio-Rad) was used as molecular weight marker. Migration was performed with Bio-Rad Mini-Protean Tetra Cell at 175V in running buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 0.1% (w/v) SDS). Gel was stained with PageBlue protein staining solution (Thermo Scientific, Waltham, MA, USA).

2.5.4. Western Blot

The SDS gel was transferred to a nitrocellulose membrane (0.45 μ m Protran® BA85 from GE Healthcare, Life Sciences – Whatman™, Uppsala Sweden) with the Bio-Rad Mini Trans-Blot Cell at 100 V for 1 hour in a transfer buffer (3.03 g/L Tris-HCl, 14.4 g/L glycine, 20% ethanol). After checking the transfer with Ponceau red staining, membrane was blocked overnight at 4 °C under gentle agitation with PBS1X (20 mM NaKPO₄ 150 mM NaCl pH 7.3), 5% (w/v) skim milk, 0.5% (v/v) Tween 20. Membrane was then incubated in the same solution with 1/5000 diluted primary monoclonal antibodies (mouse Anti-His antibodies, Life Technologies). After incubation for 1 hour at room temperature, the membrane was washed 3 times for 15 min with PBS1X, 1% (w/v) skim milk, 0.5% (v/v) Tween 20. The membrane was then incubated in the same solution supplemented with 1/5000 diluted alkaline phosphatase-conjugated secondary antibodies (goat Anti-Mouse IgG Alkaline Phosphatase, Sigma-Aldrich), for 1 hour at room temperature. Next, the membrane was washed 3 times for 15 min with PBS1X, 0.5% (v/v) Tween 20. Band detection was done by AP Chromogenic substrate BCIP/NBT (BCIP®/NBT Liquid Substrate System, Sigma-Aldrich, St. Louis, MO, US) incubation.

2.6. Purification

Lip2 expressed in *P. pastoris* or *Y. lipolytica* was purified by hydrophobic interaction chromatography and subsequent size exclusion chromatography. The culture supernatant was collected by centrifugation at 10,000 x g for 30 min and ammonium sulfate was added at a concentration of 1 M. It was loaded at 8 mL/min on Butyl-Sepharose 4 Fast Flow gel (GE Healthcare, Uppsala Sweden). The column was washed with 20 mM Tris-HCl pH 7.8 and Lip2 was eluted with a linear gradient of ammonium sulfate ranging from 1 M to 0 M in 20 mM Tris pH 7.8 on 125 mL. Active fractions were identified by checking pNPB hydrolysis and concentration on an Amicon apparatus with 10 kDa MWCO (Millipore, Billerica, MA, US). Lip2 was then loaded on a Superdex 200 16/60 Hi-Load (GE Healthcare, Uppsala, Sweden) at 1 mL/min previously equilibrated in 20 mM NaPO₄ pH 6.5 and 100 mM NaCl. Active fractions identified by pNPB hydrolysis were concentrated in an Amicon apparatus 10 kDa MWCO (Millipore, Billerica, MA, USA) at a concentration of 250 µM.

2.7. NMR spectroscopy

For NMR characterization, purified Lip2 from *P. pastoris* and *Y. lipolytica* were prepared at 300 µM respectively in 20 mM phosphate buffer pH 6.3 (90% H₂O/10% D₂O). NMR experiments were performed at 295 K using a Bruker Avance 600 MHz spectrometer (Bruker Biospin) equipped with a cryoprobe TCI. ¹H-¹⁵N HSQC experiments were acquired in the echo/anti-echo mode (256 data points) and watergate suppression was used to reduce the water peak. All data were acquired and processed using Topspin v2.1 software (Bruker Biospin).

Table 1. List of constructs used in this work.

Expression system	Construct	Expressed protein	Promoter	Inducer	Resistance marker	Reference
<i>E. coli</i>	pET26b-Lip2	Lip2	T7lac	IPTG	Kanamycin	Novagen
Cell-free	PIVEX2.3d-Lip2	Lip2	T7	-	Ampicillin	Roche
	PIVEX2.3d-HisLip2	HisLip2	T7	-	Ampicillin	Roche
<i>P. pastoris</i>	PICZαA-Lip2	Lip2	AOX1	Methanol	Zeocin	Life technology
<i>Y. lipolytica</i>	JMP61-LIP2	Lip2	TEF	-	Kanamycin, Uracil selection	(Bordes <i>et al.</i> 2010)

III. Results

3.1. Expression in *Escherichia coli*

The wild type sequence of the mature form of Lip2 with an additional N-terminal methionine was cloned in the plasmid pET26b(+). BL21(DE3) and Shuffle T7 strains were tested at 30 or 37 °C with IPTG induction at concentration of 0.1 or 1 mM. After lysis and separation of the soluble and insoluble fractions, none of these fractions displayed any

detectable activity and SDS-PAGE analysis (Supplementary Figure S2) revealed an intense band in the insoluble fraction, corresponding to high level expression of a protein compatible with the theoretical size of Lip2 (33 kDa). No such intense band was observed in the soluble fraction. Therefore, under the conditions tested, Lip2 lipase is expressed as an insoluble and inactive form in *E. coli*.

3.2. Cell free expression

Expression of Lip2, with and without a histidine tag, in cell-free systems was attempted using 2 homemade extracts (S30: BL21Star strain transformed with the pAR1219 plasmid or Chap-S30: Rosetta2 strain co-transformed with the pAR1219 plasmid and pRDKJG plasmid, both enriched in T7 RNA polymerase and the second in chaperones DnaK, DnaJ and GrpE). Lip2 and HisLip2 lipases were revealed in the insoluble fraction by SDS-PAGE analysis (Supplementary Figure S3A lanes 1-4 and 6-9) and confirmed by Western blot using Anti-His antibodies for HisLip2 (Supplementary Figure S3B lanes 1-2 and 4-5). No lipase activity was detected in either the soluble or insoluble fraction.

Optimization of cell-free reaction conditions by adding chaperone to the S30 extract (Supplementary Figure S3A lanes 10-14 and B lanes 6-7), controlling oxidizing conditions with GSSG:GSH, and reducing the reaction time enabled us to obtain a fully soluble protein (Supplementary Figure S3 C and D). However, the protein was not active and no activity was detected neither in the soluble nor in the insoluble fractions.

3.3. Expression in *Pichia pastoris*

The construction pPICZ α A-Lip2 brought the sequence of the mature Lip2 under the control of the methanol inducible AOX1 promoter and in frame with the α -factor prepro signal sequence from *Saccharomyces cerevisiae*. Two strains of *P. pastoris* were used, the wild-type strain X33 and the protease-deficient strain SMD1168H. In minimal ¹⁵N labeled BMG and BMM media, activity was detected after 12 hours of induction by methanol and a band appeared at 37 kDa on SDS-PAGE gel, as previously reported (Yu *et al.* 2010). Nevertheless, a decrease in activity was observed after 48 hours with the two strains used (X33 or SMD1168H). As a consequence, culture was stopped after 24 hours of induction to avoid degradation of the protein. An activity of 1.5 U/mL of supernatant was obtained and the presence of a single band at 37 kDa was checked by SDS-PAGE analysis (Figure 1A). To determine the specific activity, Lip2 was purified in two steps, first, by hydrophobic interaction chromatography followed by size-exclusion chromatography (Supplementary Table S2). After purification and concentration, a specific activity of 3.5 U/mg of protein was

detected and the ^{15}N - ^1H HSQC spectrum of the protein was recorded (Figure 1A). The specific activity and the ^{15}N - ^1H HSQC spectrum were identical in the two strains.

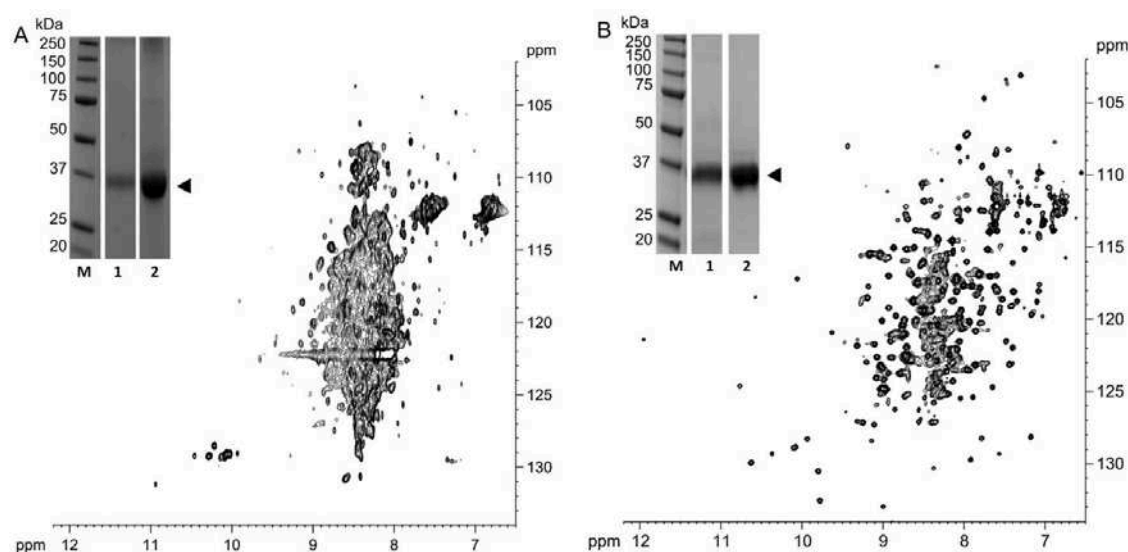


Figure 1. SDS-PAGE analysis and 2D [^{15}N - ^1H] HSQC spectrum of uniform ^{15}N labeled Lip2 expressed in *Pichia pastoris* (A) and in *Yarrowia lipolytica* (B). Lane M: molecular weight marker; Lane 1: supernatant; Lane 2: purified and concentrated sample; the expected size is indicated by a black arrow (37 kDa). Spectra were recorded using a Bruker Avance III HD spectrometer operating at a ^1H Larmor frequency of 600 MHz.

3.4. Expression in *Yarrowia lipolytica*

For expression of Lip2 in *Yarrowia lipolytica*, the strain JMY1212 – JMP61-pTEF-LIP2 was first cultivated in Erlenmeyer flasks in minimal medium for 24 hours. In these conditions, no activity was detected and no band was observed on SDS-PAGE gels. To obtain the expression of Lip2 in minimal ^{15}N labeled medium, a bioreactor culture was performed, as this allows better control of the oxygenation rate and pH during growth. Using this strategy, an activity of 8 U/mL was detected in the supernatant. On SDS-PAGE gel, the main band corresponding to Lip2 was observed at 37 kDa, as expected (Figure 1B). After purification by hydrophobic interaction chromatography followed by size-exclusion chromatography (Supplementary Table S2) and concentration, a specific activity of 15 U/mg was detected and the ^{15}N - ^1H HSQC spectrum was recorded (Figure 1B).

At the end of fermentation, the cells were recycled to produce a solution of 47 g/L of dry ^{15}N -yeast extract. This ^{15}N -yeast extract was used for culture in an Erlenmeyer flask. After 24 hours, an activity of 12 U/mL was found in the supernatant with a band at 37 kDa on SDS-PAGE gel. After purification and concentration, a specific activity of 15 U/mg was found and the ^{15}N - ^1H HSQC spectrum was recorded. Both the specific activity and the NMR spectrum were similar to those in Lip2 resulting from the fermentation in minimal medium.

IV. Discussion

The main purpose of this work was to produce uniformly ^{15}N labeled Lip2 for NMR analyses of the protein dynamic changes upon activation. We decided to avoid the use of fusion proteins, as reported in (Kumari *et al.* 2012), since Lip2 molecular weight is already quite high for NMR studies, and to avoid subsequent tag cleavage steps that may lead to further loss of protein activity. Lip2 has already been successfully expressed in *Y. lipolytica* and *P. pastoris* for structural studies under the control of TEF or GAP constitutive promoters (Bordes *et al.* 2010, Aloulou *et al.* 2013). However, in these cases, Lip2 was expressed in complex media, which is a serious obstacle to isotopic labeling.

Given our inability to express a functional lipase in *E. coli* and in cell-free, we turned to eukaryotic expression systems and tested two non-conventional yeasts, *Pichia pastoris* and *Yarrowia lipolytica*, in parallel. The methylotrophic yeast *P. pastoris* has already been shown to successfully express and secrete functional Lip2 lipase under the control of the constitutive GAP promoter (Aloulou *et al.* 2007a) and the methanol inducible AOX1 promoter (Yu *et al.* 2007). The use of an inducible promoter makes it possible to decouple protein expression from biomass production. The first step (biomass production) can therefore be performed using unlabeled media and the second step (protein expression) can be carried out using labeled carbon and a nitrogen source, thus reducing the cost of producing labeled protein. As our goal was to produce labeled Lip2, we chose to use the methanol inducible promoter AOX1.

The oleaginous yeast *Y. lipolytica* is not only the native organism to produce Lip2 lipase but it is an efficient non-conventional yeast for the production of heterologous protein (Madzak *et al.* 2004). Lip2 was already successfully using the oleic acid inducible POX2 promoter (Pignede *et al.* 2000, Jolivet *et al.* 2007), leading to a high specific activity (110 U/mg). However, oleic acid is known to lead to the production of lipid-protein complexes that are not suitable for subsequent structural studies (Aloulou *et al.* 2007b). Therefore, we opted to use the TEF constitutive promoter, even if for some unknown reason, the obtained specific activity was lower. The previously constructed strain JMY1212-Jmp61 pTEF-Lip2 (Bordes *et al.* 2010) was consequently used to express non-aggregated Lip2. Although no activity was obtained in minimal medium in shake flasks, cultivation in controlled conditions in a bioreactor enabled the production of ^{15}N uniformly labeled Lip2. The differences observed among the different cultures can be explained by the acidification of the medium due to culture in minimal medium. Acidification could thus favor the production of the well-described AXP protease secreted by *Y. lipolytica* in acid conditions (Glover *et al.* 1997).

When produced in *P. pastoris* and *Y. lipolytica*, the same apparent molecular weight of 37 kDa was found on SDS-PAGE analysis (Figure 1A and Figure 1B). This suggests that, as previously described (Aloulou *et al.* 2007b, Yu *et al.* 2007), Lip2 is similarly glycosylated in the two yeasts. However, specific activity after purification was 4 times higher when expressed in *Y. lipolytica* than in *P. pastoris* with 15 U/mg and 3.5 U/mg, respectively. Furthermore, the ^{15}N - ^1H HSQC spectra of Lip2 differed significantly depending on the host used. In *P. pastoris*, a ^{15}N - ^1H HSQC spectrum was produced with a narrow spectral dispersion of intense signals, mainly between 8 and 8.5 ppm for ^1H chemical shift, characteristic of an unfolded protein (Figure 1A). By decreasing the levels closer to the baseline, it was possible to reveal weak signals with increased dispersion. This suggests that two populations with different conformations coexist in this sample, a major unfolded population and a minor folded population of Lip2. This hypothesis is supported by the low specific activity, 3.5 U/mg, of the sample, and was confirmed by the strict superposition with the HSQC spectrum of Lip2 obtained from *Y. lipolytica*. When produced in *Y. lipolytica*, the well-dispersed chemical shift of the ^1H and ^{15}N nuclei in the ^{15}N - ^1H HSQC spectrum indicates that the ^{15}N -labeled Lip2 secreted by *Y. lipolytica* into the culture medium is correctly folded (Figure 1B). This was corroborated by the higher specific activity, 15 U/mg, of the sample.

It cannot be excluded that the two additional amino acids (glutamic acid and phenylalanine) introduced at the N-terminal extremity of the sequence during the construction (derived from the EcoRI site in pPICZ α A vector) might contribute to the misfolded protein and to the low specific activity. However, the group of R.D. Schmid (Yu *et al.* 2007) previously used successfully this strategy to express Lip2 lipase. A well-known problem of the expression in *P. pastoris* which could explain the lack of activity, is the proteolysis of recombinant proteins (Cregg 2007). *P. pastoris* cells have a vacuole that contains proteases, such as proteinase A. Even if secreted, recombinant proteins can be in contact with the protease in the culture medium due to cell lysis. However the use of the strain SMD1168H, deleted for the PEP4 gene, which encodes the proteinase A, did not alter the quality of the production in *P. pastoris* and in all the strains used, the protein produced appeared as a single unproteolysed band. Another possible explanation for the lack of activity of the secreted protein could be a different organization of the disulfide bridge in the lipase from that in the native arrangement, as many combinations between the 9 cysteines present in Lip2 could occur. Some examples in the literature show that, under the expression of AOX1 promoter, some heterologous protein production in *P. pastoris* could be impaired by the incorrect formation of disulfide bridges (Lange *et al.* 2001, Wu *et al.* 2012). The correct formation of disulfide bonds in protein folding is a critical process for heterologous protein production, and is responsible for oxidative stress of the host cells

(Tyo *et al.* 2012). In this case, this metabolic stress is added to a second environmental stress related to the use of methanol as carbon source for the inducing phase and which creates additional oxidative stress due to the generation of reactive oxygen species (Vanz *et al.* 2012). It can be assumed that this dual stress causes a significant disturbance in the redox state of the cells, leading to the formation of wrong disulfide bridges and a partially unfolded protein. This could also be in agreement with the recent successful crystallization of a properly folded Lip2 lipase using *P. pastoris* as host strain (Aloulou *et al.* 2013) for which the constitutive GAP promoter was used. Furthermore, constructs and the expression protocol differed slightly from previous attempts to express Lip2 in *Pichia pastoris* with the methanol inducible promoter AOX1, making direct comparison difficult. Expression of Lip2 in *Y. lipolytica* in minimal medium failed in uncontrolled pH culture conditions. The use of a bioreactor made it possible to control this parameter and resulted in a high rate of production of the ^{15}N labeled recombinant protein. However, the use of a bioreactor also increased the volume and hence the cost of labeled production. One possible solution is the use of inducible promoters like the POX promoter, but as already mentioned it is not suitable for subsequent structural studies. To circumvent the lack of other inducible promoters in *Y. lipolytica*, we were able to recycle the ^{15}N of the biomass resulting from the fermentation to produce a ^{15}N labeled rich medium. This homemade rich medium enabled us to produce uniform ^{15}N labeled Lip2 in a shake flask. After purification, the ^{15}N -Lip2 produced in the shake flask had the same specific activity and the same ^1H - ^{15}N HSQC spectrum as the ^{15}N -Lip2 produced in the bioreactor in minimal medium (Figure 1B). Recycling the labeled isotopes that were not incorporated in the recombinant protein enabled us to increase the production of ^{15}N labeled Lip2 from 200 mg (in the bioreactor) to 250 mg (in an Erlenmeyer flask) of purified protein per liter of culture. This can be combined with a simple solution consisting in using a small bioreactor (i.e. 250 mL). Taking into account the yield of expression, one should obtain about 50 mg of pure protein, which corresponds to approximately three NMR samples. Since we confirmed that the protein remains stable for several months at $-20\text{ }^{\circ}\text{C}$, this down scaling paves the way for elucidating protein dynamics by NMR at reasonable cost.

V. Conclusion

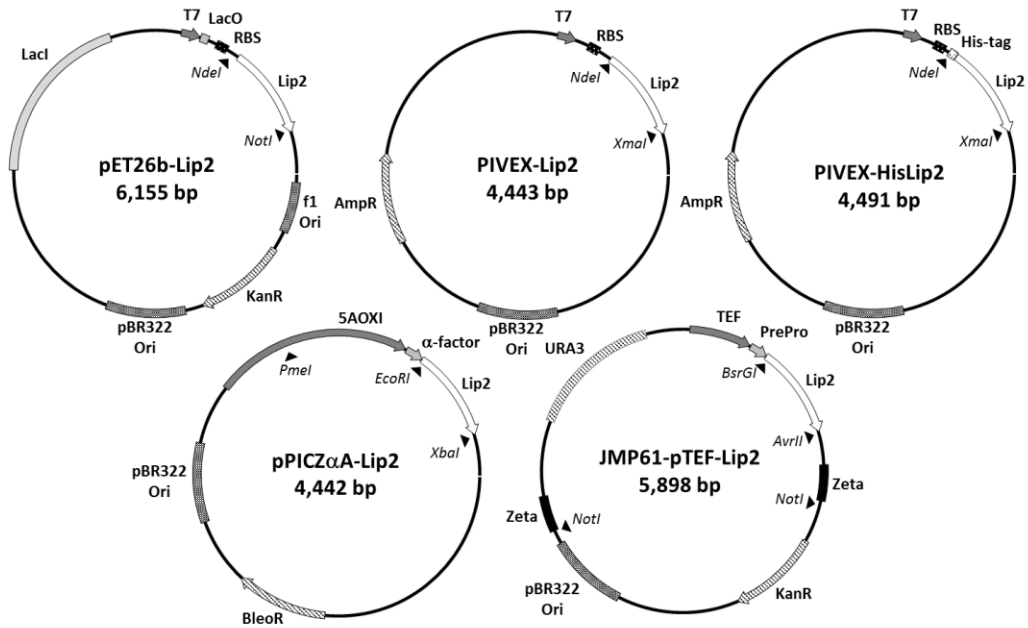
In this work, we studied the expression of Lip2 from *Yarrowia lipolytica* in four different prokaryote and eukaryote expression systems for the purpose of subsequent NMR studies. Among the systems tested, labeled Lip2 lipase was only successfully produced using yeasts, either *P. pastoris* or *Y. lipolytica*.

While the use of *P. pastoris* yeast combined with the inducible promoter AOX1 led to the production of two lipase populations with mainly a partially folded protein, the use of the *Y. lipolytica* expression system enabled production of uniform ^{15}N labeled Lip2, with good solubility and specific activity, and a homogeneous conformation as revealed by NMR. The first ^1H - ^{15}N HSQC spectrum of Lip2 from *Yarrowia lipolytica* was recorded and the good dispersion of ^1H and ^{15}N chemical shifts in the spectrum paves the way for further studies of the functional dynamics of this 37 kDa protein using NMR techniques.

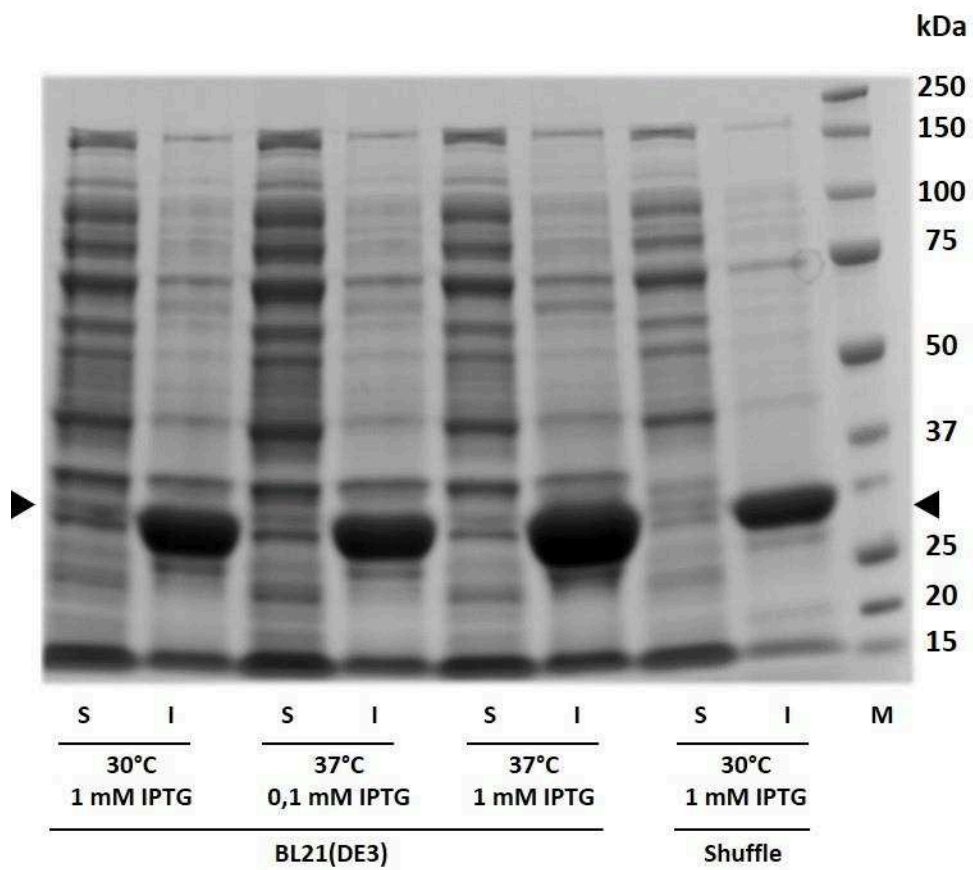
VI. Acknowledgments

The plasmid pRDKJG was a generous gift from Dr P. Genevaux. The NMR spectra were recorded on a Bruker Avance 600 MHz spectrometer (Bruker Biospin) which is part of the Integrated Screening Platform of Toulouse (PICT, IBISA). G. N. is supported by a Ph.D. grant from the Pôles de Recherche et d'Enseignement Supérieur de l'Université de Toulouse and the Région Midi-Pyrénées, France.

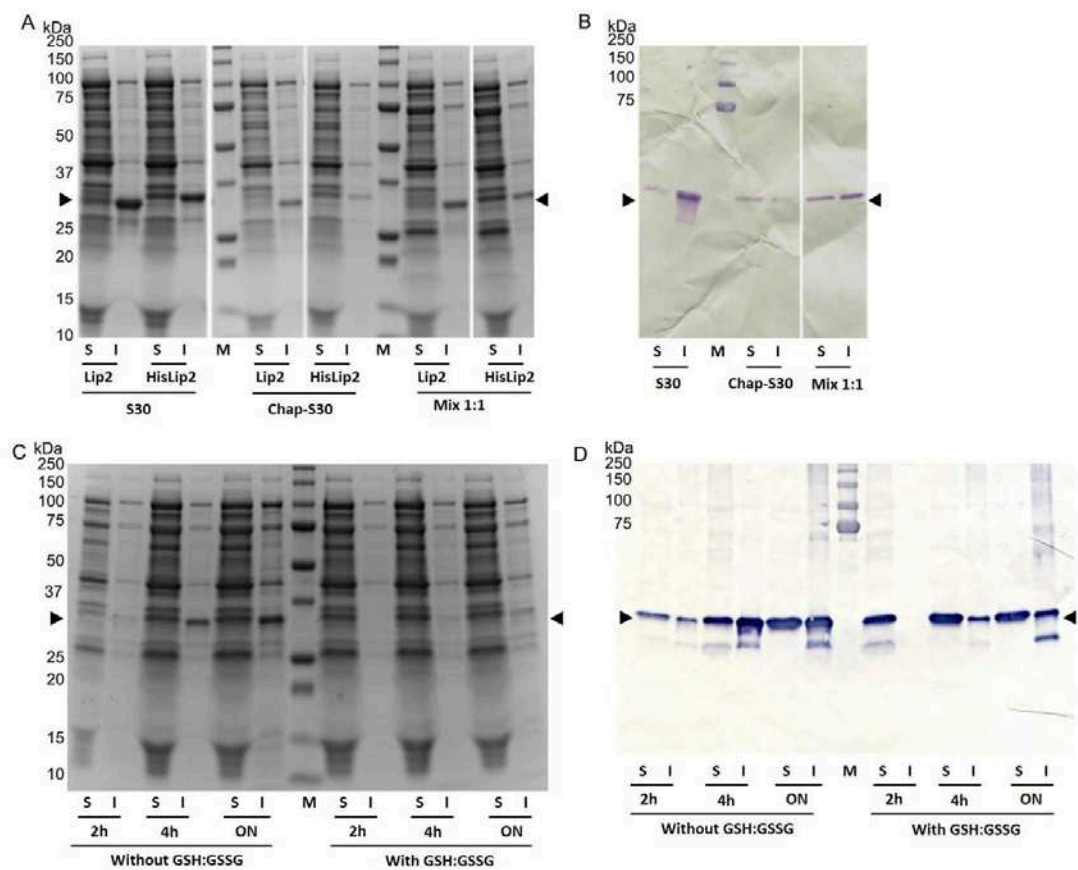
VII. Supplementary information



Supplementary Figure S1. Map of plasmids pET26b-Lip2, PIVEX-Lip2, PIVEX-HisLip2, pPICZ α A-Lip2 and JMP61-pTEF-Lip2. pBR322 Ori, pBR322 origin of replication; f1 Ori, f1 origin of replication; KanR, kanamycin resistance; AmpR, ampicillin resistance; BleoR, bleomycin resistance; URA3, uracil auxotrophy complementation; T7, T7 promoter; 5AOXI, alcohol oxidase 1 promoter; TEF, transcription elongation factor promoter; RBS, ribosome binding site; LacO, lac operator; LacI, lac inhibitor; α -factor, alpha factor signal sequence from *Saccharomyces cerevisiae*; PrePro, prepro Lip2 signal sequence from *Yarrowia lipolytica*; Lip2, mature sequence of Lip2 lipase from *Yarrowia lipolytica*; His-Tag, 6 histidine tag; Zeta, zeta sequence; NdeI, NotI, XmaI, PmeI, EcoRI, XbaI, BsrGI and AvrII restriction sites.



Supplementary Figure S2. SDS-PAGE of expression of Lip2 in *E. coli* testing the strain BL21(DE3) and the strain Shuffle at 30 °C or 37 °C with induction by 0.1 or 1 mM of IPTG. Lane S: soluble fraction; Lane I: insoluble fraction; Lane M: molecular weight marker; the expected size is indicated by a black arrow (33 kDa).



Supplementary Figure S3. Cell-free expression using S30 extract without chaperone (S30), with chaperone (Chap-S30) or a 1:1 mix of the two S30 extracts (Mix 1:1). SDS-PAGE (A) of Lip2 and HisLip2 and Western Blot (B) of HisLip2. Cell-free expression using a 1:1 mix of S30 extract without and with chaperone after 2 hours, 4 hours or overnight incubation in absence or presence of GSH:GSSG. SDS-PAGE (C) and Western Blot (D) of HisLip2. Lane S: soluble fraction; Lane I: insoluble fraction; Lane M: molecular weight marker; the expected size is indicated by a black arrow (33 kDa).

Supplementary Table S1. List of primers used in this work.

Primer	Expression system	Vector	Insertion	Sequence
aLip2_ECf	<i>E. coli</i>	pET26b(+)	NdeI	AATCTACGCATATGGTGTACACCTCTACCGAGACC
aLip2_ECr	<i>E. coli</i>	pET26b(+)	NotI	TTAGCGGGCCGCTTATTAGATACCACAGACACCCTC
aLip2_CFf	Cell-free	PIVEX2.3d	NdeI	AATCTACGCATATGGTGTACACCTCTACCGAGACC
aHisLip2_CFf	Cell-free	PIVEX2.3d	NdeI + His-tag	AATCTACGCATATGAAATCTGGTCAATCATCATCATCATCAITCTAGCG GCATCGAAGGCCGCGTGTACACCTCTACCGAGACC
aLip2_CFr	Cell-free	PIVEX2.3d	XmaI	ATTAGGCCCCGGGTTATTAGATACCACAGACACCCTCG
aLip2_PPF	<i>P. pastoris</i>	PICEZ α A	EcoRI	GACGCCGTACGGCGAATTCGTGTACACCTCTACCGAGACC
aLip2_PPr	<i>P. pastoris</i>	PICEZ α A	XbaI	GACGCCCTAGGTCTAGATCAGATACCACAGACACCCTCG

Supplementary Table S2. Purification steps of Lip2 expressed in *P. pastoris* and *Y. lipolytica*.

Purification step	Total activity (U)	Total Protein (mg)	Specific activity (U/mg)	Yield (%)	Purification (Fold)
<i>Pichia pastoris</i>					
Supernatant	150	97	1.5	100	1
Butyl Sepharose	90	46	1.9	60	1.3
Superdex 200	20	5.7	3.5	13	2.3
<i>Yarrowia lipolytica</i>					
Supernatant	24000	3600	6.7	100	1
Butyl Sepharose	17100	1596	10.7	71	1.6
Superdex 200	10455	688	15.2	44	2.3

VIII. References

- Aloulou A, Benarouche A, Puccinelli D, Spinelli S, Cavalier J F, Cambillau C and Carriere F. *Biochemical and structural characterization of non-glycosylated Yarrowia lipolytica LIP2 lipase*. European Journal of Lipid Science and Technology, **2013**, 115(4): 429-441.
- Aloulou A, Puccinelli D, De Caro A, Leblond Y and Carriere F. *A comparative study on two fungal lipases from Thermomyces lanuginosus and Yarrowia lipolytica shows the combined effects of detergents and pH on lipase adsorption and activity*. Biochim Biophys Acta, **2007a**, 1771(12): 1446-1456.
- Aloulou A, Rodriguez J A, Puccinelli D, Mouz N, Leclaire J, Leblond Y and Carriere F. *Purification and biochemical characterization of the LIP2 lipase from Yarrowia lipolytica*. Biochim Biophys Acta, **2007b**, 1771(2): 228-237.
- Bordes F, Barbe S, Escalier P, Mourey L, Andre I, Marty A and Tranier S. *Exploring the conformational states and rearrangements of Yarrowia lipolytica Lipase*. Biophys J, **2010**, 99(7): 2225-2234.
- Bordes F, Cambon E, Dossat-Letisse V, Andre I, Croux C, Nicaud J M and Marty A. *Improvement of Yarrowia lipolytica lipase enantioselectivity by using mutagenesis targeted to the substrate binding site*. Chembiochem, **2009**, 10(10): 1705-1713.
- Bordes F, Fudalej F, Dossat V, Nicaud J-M and Marty A. *A new recombinant protein expression system for high-throughput screening in the yeast Yarrowia lipolytica*. Journal of microbiological methods, **2007**, 70(3): 493-502.
- Bradford M M. *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding*. Analytical biochemistry, **1976**, 72: 248-254.
- Cambon E, Piamtongkam R, Bordes F, Duquesne S, Andre I and Marty A. *Rationally engineered double substituted variants of Yarrowia lipolytica lipase with enhanced activity coupled with highly inverted enantioselectivity towards 2-bromo phenyl acetic acid esters*. Biotechnol Bioeng, **2010**, 106(6): 852-859.
- Caspers P, Stieger M and Burn P. *Overproduction of Bacterial Chaperones Improves the Solubility of Recombinant Protein-Tyrosine Kinases in Escherichia coli*. Cellular and Molecular Biology, **1994**, 40(5): 635-644.
- Cregg J M. *Pichia Protocols*. **2007**, Humana Press.
- de Marco A, Deuerling E, Mogk A, Tomoyasu T and Bukau B. *Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli*. BMC biotechnology, **2007**, 7.
- Dutta A, Saxena K, Schwalbe H and Klein-Seetharaman J. *Isotope labeling in mammalian cells*. Methods in molecular biology (Clifton, N.J.), **2012**, 831: 55-69.
- Gardner K H and Kay L E. *The use of $2H$, ^{13}C , ^{15}N multidimensional NMR to study the structure and dynamics of proteins*. Annual review of biophysics and biomolecular structure, **1998**, 27: 357-406.

Glover D J, McEwen R K, Thomas C R and Young T W. *pH-regulated expression of the acid and alkaline extracellular proteases of Yarrowia lipolytica*. Microbiology, **1997**, 143(9): 3045-3054.

Guieysse D, Sandoval G, Faure L, Nicaud J M, Monsan P and Marty A. *New efficient lipase from Yarrowia lipolytica for the resolution of 2-bromo-arylacetic acid esters*. Tetrahedron-Asymmetry, **2004**, 15(22): 3539-3543.

Jolivet P, Bordes F, Fudalej F, Cancino M, Vignaud C, Dossat V, Burghoffer C, Marty A, Chardot T and Nicaud J M. *Analysis of Yarrowia lipolytica extracellular lipase Lip2p glycosylation*. FEMS Yeast Res, **2007**, 7(8): 1317-1327.

Kumari A, Verma V V and Gupta R. *Comparative biochemical characterization and in silico analysis of novel lipases Lip11 and Lip12 with Lip2 from Yarrowia lipolytica*. World J Microbiol Biotechnol, **2012**, 28(11): 3103-3111.

Lange S, Schmitt J and Schmid R D. *High-yield expression of the recombinant, atrazine-specific Fab fragment K411B by the methylotrophic yeast Pichia pastoris*. Journal of Immunological Methods, **2001**, 255(1–2): 103-114.

Leblond Y, Mouz N, Marty A and Uribe Larrea J-I. *Method for producing lipase, transformed Yarrowia lipolytica cell capable of producing said lipase and their uses*. **2007**, WO/2007/144475.

Lin-Cereghino J, Wong W W, Xiong S, Giang W, Luong L T, Vu J, Johnson S D and Lin-Cereghino G P. *Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris*. BioTechniques, **2005**, 38(1): 44,-46, 48.

Madzak C, Gaillardin C and Beckerich J M. *Heterologous protein expression and secretion in the non-conventional yeast Yarrowia lipolytica: a review*. J Biotechnol, **2004**, 109(1-2): 63-81.

Michel E and Wüthrich K. *Cell-free expression of disulfide-containing eukaryotic proteins for structural biology*. FEBS Journal, **2012**, 279(17): 3176-3184.

Mittermaier A and Kay L E. *New tools provide new insights in NMR studies of protein dynamics*. Science (New York, N.Y.), **2006**, 312(5771): 224-228.

Ohki S Y and Kainosho M. *Stable isotope labeling methods for protein NMR spectroscopy*. Progress in Nuclear Magnetic Resonance Spectroscopy, **2008**, 53(4): 208-226.

Osawa M, Takeuchi K, Ueda T, Nishida N and Shimada I. *Functional dynamics of proteins revealed by solution NMR*. Current Opinion in Structural Biology, **2012**, 22(5): 660-669.

Ozawa K, Dixon N E and Otting G. *Cell-free synthesis of N-15-labeled proteins for NMR studies*. J Biomol Life, **2005**, 57(9): 615-622.

Pignede G, Wang H, Fudalej F, Gaillardin C, Seman M and Nicaud J M. *Characterization of an extracellular lipase encoded by LIP2 in Yarrowia lipolytica*. J Bacteriol, **2000**, 182(10): 2802-2810.

Quinn D M, Shirai K, Jackson R L and Harmony J A K. *Lipoprotein lipase-catalyzed hydrolysis of water-soluble p-nitrophenyl esters. Inhibition by apolipoprotein C-II*. Biochemistry, **1982**, 21(26): 6872-6879.

Sambrook J, Fritsch E F and Maniatis T. *Molecular cloning*. **1989**, Cold spring harbor laboratory press New York.

Samuelson J C. *Recent Developments in Difficult Protein Expression: A Guide to E. coli Strains, Promoters, and Relevant Host Mutations*. **2011**, Heterologous Gene Expression in *E.coli*, Jr T C E and Xu M-Q, Humana Press: 195-209.

Saxena K, Dutta A, Klein-Seetharaman J and Schwalbe H. *Isotope labeling in insect cells*. Methods in molecular biology (Clifton, N.J.), **2012**, 831: 37-54.

Schneider B, Junge F, Shirokov V A, Durst F, Schwarz D, Dötsch V and Bernhard F. *Membrane protein expression in cell-free systems*. Methods in molecular biology (Clifton, N.J.), **2010**, 601: 165-186.

Sugiki T, Shimada I and Takahashi H. *Stable isotope labeling of protein by Kluyveromyces lactis for NMR study*. Journal of biomolecular NMR, **2008**, 42(3): 159-162.

Tyo K E J, Liu Z, Petranovic D and Nielsen J. *Imbalance of heterologous protein folding and disulfide bond formation rates yields runaway oxidative stress*. BMC Biology, **2012**, 10(1).

Vanz A L, Lünsdorf H, Adnan A, Nimtz M, Gurramkonda C, Khanna N and Rinas U. *Physiological response of Pichia pastoris GS115 to methanol-induced high level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes*. Microbial Cell Factories, **2012**, 11(1).

Wood M J and Komives E A. *Production of large quantities of isotopically labeled protein in Pichia pastoris by fermentation*. Journal of biomolecular NMR, **1999**, 13(2): 149-159.

Wu D, Chu J, Hao Y-Y, Wang Y-H, Zhuang Y-P and Zhang S-L. *Incomplete protein disulphide bond conformation and decreased protein expression result from high cell growth during heterologous protein expression in Pichia pastoris*. Journal of Biotechnology, **2012**, 157(1): 107-112.

Yu M, Wen S and Tan T. *Enhancing production of Yarrowia lipolytica lipase Lip2 in Pichia pastoris*. Engineering in Life Sciences, **2010**, 10(5): 458-464.

Yu M R, Lange S, Richter S, Tan T W and Schmid R D. *High-level expression of extracellular lipase Lip2 from Yarrowia lipolytica in Pichia pastoris and its purification and characterization*. Protein Expression and Purification, **2007**, 53(2): 255-263.

Chapitre 3

Structure cristallographique et dynamique par RMN en solution du domaine C-terminal de OmpA de *Klebsiella pneumoniae*.

Résumé

Ce chapitre décrit l'étude structurale conduite sur le domaine C-terminal de KpOmpA (C-KpOmpA). La structure de ce domaine a été résolue par cristallographie. C-KpOmpA adopte une structuration suivant le repliement $\beta\alpha\beta\alpha\beta\beta$ des domaines OmpA-like retrouvés dans de nombreuses protéines périplasmiques et susceptible de reconnaître le peptidoglycane. Deux structures de ce domaine ont pu être résolues, une forme Apo et une forme en complexe avec une molécule de citrate qui vient se placer dans une cavité décrite comme le site de reconnaissance de l'acide diaminopimélique, un acide aminé propre au peptidoglycane de nombreuses bactéries gram-négatives. La possibilité de formation de dimères a été explorée et un équilibre entre une forme monomérique majoritaire et une forme dimérique minoritaire a été observée. Enfin la dynamique de ce domaine a été étudiée par analyse des temps de relaxation ^{15}N révélant une structure stable du domaine en solution avec des mouvements de l'ordre de la pico – nanoseconde de faible amplitude et l'absence de mouvements sur l'échelle de la micro – milliseconde. Les deux structures cristallographiques ainsi que les déplacements chimiques sont en cours de dépôt dans la « Protein Data Bank » (PDB) et la « Biological Magnetic Resonance Data Bank » (BMRB), respectivement.

Structure and dynamics of the C-terminal domain of OmpA from *Klebsiella pneumoniae*.

Abstract

The OmpA-like domain is a structural motif found in the C-terminal part of numerous proteins constitutive of the cell-wall of gram-negative bacteria. In *Klebsiella pneumoniae*, the outer membrane protein A is a two domain membrane protein: the N-terminal β -barrel is anchored into the outer membrane and its 3D structure in DHPC detergent micelles has been previously solved by liquid state NMR. Here we used X-ray crystallography to determine the structure of its C-terminal domain in its Apo form and in complex with a citrate molecule into the peptidoglycan binding pocket. After complete assignment of its NMR spectrum in solution, ^{15}N relaxation NMR experiments were performed in order to characterize its dynamics. It appears to be well structured, including the loops, and to display limited motion in the ps-ns or μs -ms range.

Keywords

^{15}N relaxation

Crystallography

Dimerization

Peptidoglycan

I. Introduction

The cell-wall structure of bacteria is closely related to their pathogenicity. *Klebsiella pneumoniae* is a gram-negative rod-shaped bacteria and an opportunistic pathogen responsible of respiratory and urinary tract infections. In gram-negative bacteria, the cytoplasmic membrane is surrounded by a thin layer of peptidoglycan and an outer membrane. OmpA, a major protein of the outer membrane, plays diverse roles in the infection process. Important pathogenic roles include bacterial adhesion and invasion as well as evasion of hosts defenses (Confer *et al.* 2013). OmpA is also involved in the cell-wall structure maintenance. It has been shown that a deletion of OmpA increased outer membrane vesicle formation (Deatherage *et al.* 2009). Furthermore, a small non-coding RNA, VrrA, was found to repress OmpA translation in *Vibrio cholerae* which results in an increase of outer membrane vesicles (Song *et al.* 2008) and the sRNA mechanism to downregulate OmpA is conserved across gram-negative species including *Escherichia coli*, *Yersinia pestis* and *Klebsiella pneumoniae* (Argaman *et al.* 2001).

The OmpA protein is composed of two domains: while the N-terminal domain is a transmembrane β -barrel (Renault *et al.* 2009), the C-terminal domain is a soluble globular protein localized in the periplasm where it interacts with the peptidoglycan (Ishida *et al.* 2014). This domain is part of the OmpA-like domain superfamily which is a widespread structural motif found in integral transmembrane porins (OmpA), small lipid anchored proteins (Pal) or inner membrane proton channel (MotB). The C-terminal part of the protein RmpM from *Neisseria meningitidis* was the first solved structure of an OmpA-like domain (Grizot *et al.* 2004). The structure of RmpM showed the presence of a hydrophilic groove which could accommodate a glycan strand. The purification and the resolution by NMR of the peptidoglycan associated lipoprotein of *Haemophilus influenzae* revealed a direct interaction with a biosynthetic peptidic precursor from the peptidoglycan (Parsons *et al.* 2006). More details on the interaction were disclosed with the crystallographic structure of the OmpA C-terminal domain from *Acinetobacter baumannii* complexed with a synthetic peptidoglycan fragment (Park *et al.* 2012). The peptidoglycan is a polymer of repeating disaccharide units of N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM). The NAM moieties are branched with a pentapeptide which crosslinks neighboring peptidoglycan strands (Vollmer *et al.* 2008). In most of the gram-negative bacteria, the pentapeptide is L-Ala- γ -D-Glu-meso-DAP-D-Ala-D-Ala where the *meso*-DAP is the diaminopimelate amino acid. The recognition between the OmpA-like domain and the pentapeptide is mediated by the side chain of the DAP which interact with the sidechain of a strictly conserved arginine.

The peptidoglycan layer forms a sacculus around the bacteria which protects the cell from rupture as a result from osmotic pressure of the cytoplasm. Additionally, the peptidoglycan layer maintains the specific shape of the bacterial cell. While the C-terminal of OmpA interacts with the peptidoglycan, the transmembrane N-terminal β -barrel acts as an anchor in the outer-membrane. It has been shown that the N-terminal part of OmpA from *Klebsiella pneumoniae* (N-KpOmpA) undergoes a sequential unfolding in response to a tensile load applied on the C-terminal domain (C-KpOmpA) (Bosshart *et al.* 2012). While the structure of the N-terminal domain of KpOmpA has been solved by liquid state NMR (Renault *et al.* 2009) and its dynamics studied by solid state NMR (Iordanov *et al.* 2012), no structural data were available for its C-terminal domain. Here, we report the characterization of C-KpOmpA by combined biophysical approaches. We solved its structure by X-ray crystallography and assessed its oligomeric state by NMR and SEC-MALLS. Using solution state NMR on an ^{15}N - ^{13}C labelled domain, we could characterize its dynamics by ^{15}N relaxation experiments after complete assignment of its backbone resonances.

II. Material and Methods

2.1. Cloning, expression and protein purification

The gene encoding for the C-terminal domain of OmpA from *Klebsiella pneumoniae* (OmpA²⁰⁷⁻³⁴⁴) with a His₆-tag and a TEV cleavage site in N-terminal was chemically synthesized and inserted into the pJ411 vector (DNA2.0). The protein was expressed in *Escherichia coli* BL21(DE3) strain in M9 minimal medium with 30 mg/L of Kanamycin, containing 2 g/L $^{15}\text{NH}_4\text{Cl}$ (99%, ^{15}N ; Cambridge Isotope Laboratories) and 4 g/L ^{13}C -Glucose (99%, U- $^{13}\text{C}_6$; Cambridge Isotope Laboratories) for uniform ^{15}N / ^{13}C labelling, at 37°C and 190 rpm. Overexpression was performed by adding 1 mM IPTG when the OD₆₀₀ reached 0.8 and after 4 hours of induction, cell were harvested and the bacterial pellet was stored overnight at -80°C. The pellet was thawed at room temperature and resuspended in a lysis buffer (50 mM Na₂HPO₄, 150 mM NaCl and 0.2% Triton X-100 pH 7.4). Cells were lysed by sonication on ice and the soluble fraction was isolated by centrifugation 1 hour 20,000 x g at 4°C. The supernatant was loaded on a column filled with 5 mL of TALON Superflow resin (GE Healthcare), previously equilibrated in a phosphate buffer (50 mM Na₂HPO₄, 150 mM NaCl pH 7.4). After a washing step, elution was performed with 20% of elution buffer (50 mM Na₂HPO₄, 150 mM NaCl and 500 mM Imidazole pH 7.4). The N-terminal His₆-tag was cleaved using a His₆-tagged TEV protease during a dialysis step in the phosphate buffer for 4 hours at room temperature. The cleaved His-tag and the TEV protease were removed by a second step of affinity purification on TALON Superflow resin (GE Healthcare) where the

flow-through was collected. A final step of purification was performed by size-exclusion chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare) in a phosphate buffer (5 mM Na₂HPO₄, 50 mM NaCl pH 6.5). Purity was checked by SDS-PAGE with silver staining. Protein was concentrated using an Amicon membrane filter with a 10 kDa molecular weight cut-off. Protein concentration was measured by OD_{280nm} using a molecular extinction coefficient of 6640 M⁻¹cm⁻¹.

2.2. Crystallization

Crystallization screening of C-KpOmpA was performed using the sitting-drop vapor-diffusion method with NeXtal Tubes AmSO₄ Suites, NeXtal Tubes PEGs Suite, NeXtal Tubes JCSG Core Suite I, NeXtal Tubes JCSG Core Suite II and NeXtal Tubes JCSG+ Suite (Qiagen) in 96-well plates using a high-throughput crystallization robot, Innovadyne Nanodrop ExtY (Solve Scientific). Sitting drop were produced by mixing equal volume, 200 nL each, of protein solution at 10 mg/mL and reservoir solution. After 7 days at 285°K, crystals were obtained with HEPES 0.1 M pH 7.5 MPD 70% (v/v) or citric acid 0.1M pH 4.0 PEG 6000 20% (w/v) and belonged to space group C222₁ and P4₃22, respectively. Our attempts to obtain C-KpOmpA in complex with a ligand either by co-crystallization or by soaking C-KpOmpA crystals were realized by adding in solutions either DL-2,6-Diaminopimelic acid (Sigma-Aldrich) or a synthetic pentapeptide L-Ala-γD-Glu-DAP-D-Ala-D-Ala at a concentration of 10 mM.

2.3. Structure determination

X-ray experiments were carried out at 100°K. Crystals of C-KpOmpA were soaked in a cryo-protecting solution composed of 10% glycerol in mother solution for the P4₃22 crystals prior flash-cooling. Diffraction data sets were collected on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, or on the XALOC beamline at the ALBA Synchrotron, Cerdanyola del Vallès, Spain (Juanhuix *et al.* 2014). Diffraction intensities were integrated and scaled using XDS (Kabsch 2010). The crystallization form were obtained of the C-KpOmpA. The structures were solved by molecular replacement method using Phaser (McCoy *et al.* 2007) from the CCP4 (Potterton *et al.* 2003) or Phenix software (Adams *et al.* 2010) using a C-KpOmpA Robetta model (<http://rosetta.bakerlab.org/>) (Kim *et al.* 2004). Final structures of C-KpOmpA crystallized in space group C222₁ and P4₃22 were refined to final R/R free values of 0.19/0.22 and 0.16/0.19 respectively, using REFMAC5 (Murshudov *et al.* 2011). Models were built manually using COOT (Emsley *et al.* 2004). Water molecules were manually checked after automatic assignment. Ligands were assigned using LigandFit from Phenix

program (Adams *et al.* 2010). Refinement statistics are listed in Supplementary Table S1. The representations of the structures were made with PYMOL (DeLano, W.L., Copyright 2009 – 2010 Schrödinger LLC).

2.4. SEC-MALLS

Protein sample buffered in 5 mM Na₂HPO₄, 50 mM NaCl pH 6.5 was loaded on a Shodex 402.5 (300A) column (Wyatt Technology, France) using an Agilent 1260 Infinity LC chromatographic system (Agilent Technology) coupled to a multiangle laser light scattering (MALLS) detection system. The column was equilibrated with a 0.1- μ m filtered 5 mM Na₂HPO₄, 50 mM NaCl pH 6.5 buffer. Data were collected using a DAWN HELEOS 8+ (8-angle) and Optilab T-rEX refractive index detector (Wyatt Technology Corp., Toulouse France). Sample concentration was 12 mg/mL. 10 μ L of protein sample was loaded on the column and the separation was performed at a flow rate of 0.35 mL.min⁻¹ at 25°C. Results were analyzed using the ASTRA 6.0.2.9 software (Wyatt Technology Corp.).

2.5. NMR experiments

All NMR experiments were performed at 293°K using a Bruker Avance III 600 MHz spectrometer (Bruker Biospin) equipped with a cryoprobe TCI on a sample of 1 mM ¹⁵N, ¹³C-labelled C-KpOmpA in 5 mM sodium phosphate buffer pH 6.5, 150 mM NaCl, 5 μ M 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) and 10% D₂O. The effective rotational correlation time constant τ_c was determined using ¹H-¹⁵N TRACT NMR experiment (Lee *et al.* 2006). The following experiments were acquired for assignment purpose: two-dimensional ¹⁵N-HSQC and three-dimensional CBCA(CO)NH, CBCANH, HN(CO)CA, HNCA, (H)CC(CO)NH, H(CC)(CO)NH, ¹⁵N-HSQC-TOCSY and ¹⁵N-HSQC-NOESY. T₁ T₂ and ¹H-¹⁵N heteronuclear NOE measurements were taken using standard Bruker programs. T₁ delays were 10 (duplicated), 20, 50, 100, 200, 300, 500, 800 (duplicated), 1200, 1500, 2000, 2500 and 3500 (duplicated) ms while T₂ delays were 6 (duplicated), 10, 18, 34, 50, 66, 82, 98, 130, 146, 162 (duplicated), 194, 226 and 250 ms. The duplicate points were acquired for error estimations during the analysis. ¹H-¹⁵N NOE saturated and unsaturated spectral measurements were recorded in an interleaved manner. R₁ and R₂ relaxation rates were estimated by fitting the T₁ and T₂ peak intensities to a single exponential decay. ¹H-¹⁵N steady-state NOE values were determined as the ratio of the peak intensities in the spectra recorded with and without pre-saturation of ¹H. NOE errors were estimated from the noise standard deviation. All data were acquired and processed using Topspin v 3.2 software (Bruker Biospin) and analyzed using the CcpNMR analysis software (Vranken *et al.* 2005).

TENSORV2 software was used to determine overall and internal mobility parameters (Dosset *et al.* 2000).

III. Results and Discussion

3.1. Tridimensional structure of C-terminal KpOmpA.

The crystallographic structure of the C-terminal domain of *Klebsiella pneumoniae* OmpA was determined and refined at a resolution of 1.81 and 1.90 Å for the Apo and the protein in complex with the citrate, respectively. The structures were solved by molecular replacement using a model build using Robetta server (<http://robetta.bakerlab.org/>) (Simons *et al.* 1997, Bradley *et al.* 2005). In both case, the model encompasses residues 209 – 334 out of the expressed C-terminal domain (207 – 344). The data processing and refinement statistics are summarized in Supplementary Table S1. The first and the last 10 residues are not visible in the electron density map and therefore probably disordered. The overall structure of C-KpOmpA belongs to the OmpA-like domain which is constituted of a $\beta 1/\alpha 2/\beta 2/\alpha 3/\beta 3/\alpha 4/\beta 4$ motif described for the first time in *Neisseria meningitidis* RmpM by (Grizot *et al.* 2004) (Figure 1A). The core domain is composed of two long α helices ($\alpha 2$ and $\alpha 3$) and a mixed β sheet, formed by 3 parallel ($\beta 1$ to $\beta 3$) and 1 antiparallel ($\beta 4$) β strand. The segment connecting $\beta 1$ to $\alpha 2$ is formed by a short helix $\alpha 1$ followed by a type II β -turn. A supplementary helix $\alpha 4$ and a series of type I and II' β -turn form an extension aside from the core structure between strands $\beta 3$ and $\beta 4$.

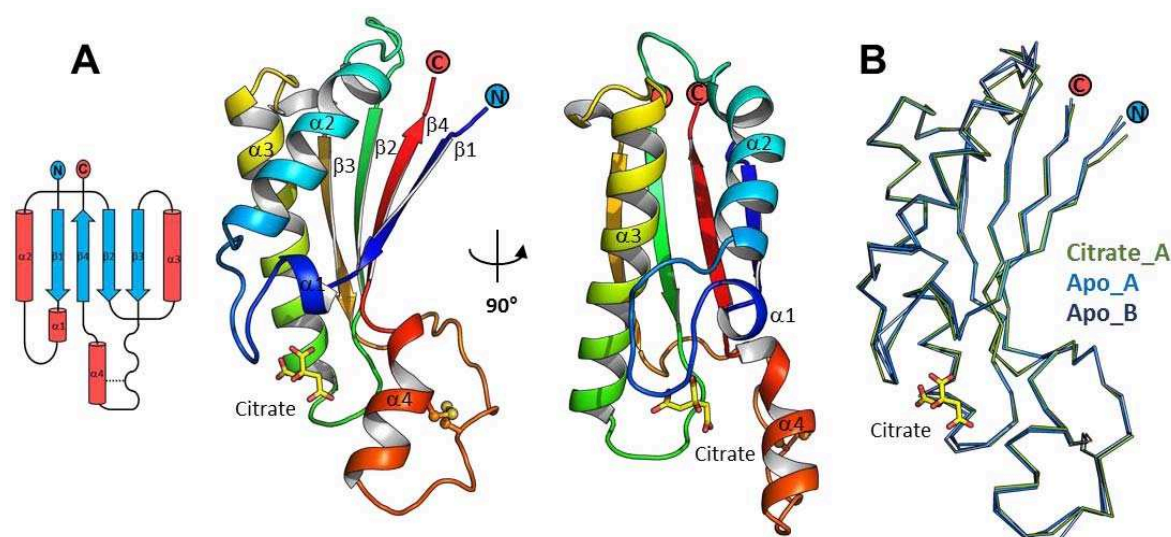


Figure 1. A. Cartoon representation of the C-KpOmpA structure color coded from blue to red from the N- to the C-terminus. The citrate is shown in stick mode and the disulfide bond between cysteine 309 and cysteine 321 is shown as ball and stick. B. Ribbon representation of the backbone superposition of the Holo (green) and the Apo (light blue for chain A and dark blue for chain B) form of C-KpOmpA. The citrate is shown in stick mode.

This bulge domain is specific of the 4 sequences of OmpA structures known. This extension corresponds to residues 300 to 327 on the alignment of sequences of OmpA-like domains extracted from the Protein Data Bank (Figure 2). For most homologous structures, this bulge is formed by a loop of 10 to 16 amino acids followed by a short α helix 4 ($\alpha 4$), composed of 11 amino acids. However, this bulge is 9 amino-acids longer for *Klebsiella pneumoniae*, *Salmonella enterica* (PDB ID: 4ERH) and *Escherichia coli* (PDB ID: 2MQE) (Ishida *et al.* 2014), and 20 amino-acids longer for *Neisseria meningitidis* (PDB ID: 1R1M) (Grizot *et al.* 2004). The identity between the sequences of the four OmpA structures varied from 35% to 95% (Supplementary Figure S1). Furthermore, they exhibit two cysteines forming a disulfide bond absent from the other sequences. This insertion in the sequence of the bulge is reflected structurally by an extension of the loop (300 – 308) to a series of β -turn for *Klebsiella pneumoniae* and *Salmonella enterica* or a α -helix for *Escherichia coli* and *Neisseria meningitidis*. Besides the helix $\alpha 4$ is also extended and the disulfide bond bridges the two sides of the bulge together.

β_1 $\xrightarrow{\alpha_1}$ β_2 $\xrightarrow{\alpha_2}$ β_3 $\xrightarrow{\alpha_3}$

Kp_OmpA	210	220	230	240	250	260	270	280
Kp_OmpA	APEVATKHF	TLKSDV	LFNFNKAT	LKPEGQQA	LDQLYTQ	LSNMDPKDGS	AVVLGY	TDRI
Kp_OmpA	APEVQTKHF	TLKSDV	LFNFNKAT	LKPEGQQA	LDQLYSQ	LSNLPKDGVS	VVLGCT	DRI
Ec_OmpA	APEVQTKHF	TLKSDV	LFNFNKST	LKPEGQQA	LDQLYSQ	LSNLPKDGVS	VVLGCT	DRI
Se_OmpA	APQYVDET	ISLSAKT	LFGFDKDS	LRAEAQDN	LKVLAQRLGQT	NIQSVRVEG	HTDFM	GSDKY
Nm_RmpM	APQYVDET	ISLSAKT	LFGFDKDS	LRAEAQDN	LKVLAQRLGQT	NIQSVRVEG	HTDFM	GSDKY
Ab_Pal	GVTEANKA	ALAKRV	HFDDYDSSD	LSTEDYQT	LQAHQAQFLMAN	ANSKVALT	CHTDER	GTTREY
Ab_Pal	GVTEANKA	ALAKRV	HFDDYDSSD	LSTEDYQT	LQAHQAQFLMAN	ANSKVALT	CHTDER	GTTREY
Bp_Pal	DPNLDPNS	PLAKRV	VFDFDYS	VQDQYQAL	LQHAQYLKSH	PSQRHILIQ	QNTDER	GTTSEY
Bp_Pal	DPNLDPNS	PLAKRV	VFDFDYS	VQDQYQAL	LQHAQYLKSH	PSQRHILIQ	QNTDER	GTTSEY
Ec_Pal	QARLQMQL	QONNI	VFDLDKYD	IGSDFAQM	LDAHAFNLRSN	PSDKVTV	EGHADER	GTTPEY
Ec_Pal	QARLQMQL	QONNI	VFDLDKYD	IGSDFAQM	LDAHAFNLRSN	PSDKVTV	EGHADER	GTTPEY
Yp_Pal	QARLQMQL	QONNI	VFDLDKYD	IGSDFAQM	LDAHAFNLRSN	PSDKVTV	EGHADER	GTTPEY
Yp_Pal	QARLQMQL	QONNI	VFDLDKYD	IGSDFAQM	LDAHAFNLRSN	PSDKVTV	EGHADER	GTTPEY
Hi_OmpP6	GGYSVADL	QQRYNT	VYFGFDKYD	ITGEYVQI	LDAHAYLNAT	PAAKVL	VEGNTDER	GTTPEY
Hi_OmpP6	GGYSVADL	QQRYNT	VYFGFDKYD	ITGEYVQI	LDAHAYLNAT	PAAKVL	VEGNTDER	GTTPEY
Ab_Omp38	QPQELTED	LNMEIR	VFFDTNKS	IKDQYKPE	IAKVAEKLSEY	PNATARI	EGHTDNT	GPRKL
Ab_Omp38	QPQELTED	LNMEIR	VFFDTNKS	IKDQYKPE	IAKVAEKLSEY	PNATARI	EGHTDNT	GPRKL
Ec_YiaD	TRSGDNI	IILNMPNN	VTFDSSAT	LKPAGANT	LTGVA	MLKEY	PKTAVNI	IGYTDST
Ec_YiaD	TRSGDNI	IILNMPNN	VTFDSSAT	LKPAGANT	LTGVA	MLKEY	PKTAVNI	IGYTDST
Mt_Arfa	ADLQSAIN	AVTGGP	IAFGNDGAS	LIPADYEI	ILNRVADKLKAC	PDARVT	INGYTDST	GSEGI
Mt_Arfa	ADLQSAIN	AVTGGP	IAFGNDGAS	LIPADYEI	ILNRVADKLKAC	PDARVT	INGYTDST	GSEGI
Ppa_T6SS	EDAVDRSV	VTIRGDEL	FFASAS	VRDEFQPL	LLRIADALRKV	KGQVLT	GHSDNR	PIATLR
Ppa_T6SS	EDAVDRSV	VTIRGDEL	FFASAS	VRDEFQPL	LLRIADALRKV	KGQVLT	GHSDNR	PIATLR
Hp_MotB	EQIDQGS	ILKLPSN	LFENATSDA	INQDMMLY	IERIAKIQKL	PKRVHIN	VRGFTDDT	PLVTRF
Hp_MotB	EQIDQGS	ILKLPSN	LFENATSDA	INQDMMLY	IERIAKIQKL	PKRVHIN	VRGFTDDT	PLVTRF
Se_MotB	QEGRLIQI	IDSQNR	PMFKTGS	VEPYMRDI	LRAIAPVLNGI	PNRISL	LAGHTDDFPY	ANGEKGS
Se_MotB	QEGRLIQI	IDSQNR	PMFKTGS	VEPYMRDI	LRAIAPVLNGI	PNRISL	LAGHTDDFPY	ANGEKGS
Dv_MotB	VFDEGVIT	LRVPS	SEVLFAPGAVE	LAPGADR	VLATL	KDLF	IRR	REQNIN
Dv_MotB	VFDEGVIT	LRVPS	SEVLFAPGAVE	LAPGADR	VLATL	KDLF	IRR	REQNIN

α_4 $\xrightarrow{\beta_4}$ α_5 $\xrightarrow{\beta_5}$

Kp_OmpA	290	300	310	320	330	340		
Kp_OmpA	GI	PAGK	I	SARGM	G	ESNP	VTGNTCD	
Kp_OmpA	GI	PAGK	I	SARGM	G	ESNP	VTGNTCD	
Ec_OmpA	GI	PADK	I	SARGM	G	ESNP	VTGNTCD	
Ec_OmpA	GI	PADK	I	SARGM	G	ESNP	VTGNTCD	
Se_OmpA	GI	PSDK	I	SARGM	G	ESNP	VTGNTCD	
Se_OmpA	GI	PSDK	I	SARGM	G	ESNP	VTGNTCD	
Nm_RmpM	GV	PVSR	I	SAVGL	G	ESQAQMTQVCE	AEVAKLGAKVSKAKKREALIAC	IEPDRRV
Nm_RmpM	GV	PVSR	I	SAVGL	G	ESQAQMTQVCE	AEVAKLGAKVSKAKKREALIAC	IEPDRRV
Ab_Pal	GV	NPQO	L	EA	VSYG	KEA	VPNPGHDE	
Ab_Pal	GV	NPQO	L	EA	VSYG	KEA	VPNPGHDE	
Bp_Pal	GV	SADQ	I	SIVSYG	G	KEKPA	VLGHDE	
Bp_Pal	GV	SADQ	I	SIVSYG	G	KEKPA	VLGHDE	
Ec_Pal	GV	SADQ	I	SIVSYG	G	KEKPA	VLGHDE	
Ec_Pal	GV	SADQ	I	SIVSYG	G	KEKPA	VLGHDE	
Yp_Pal	GV	SADQ	I	SIVSYG	G	KEKPA	VLGHDE	
Yp_Pal	GV	SADQ	I	SIVSYG	G	KEKPA	VLGHDE	
Hi_OmpP6	GV	DAGK	L	GT	VSYG	EEKPA	VLGHDE	
Hi_OmpP6	GV	DAGK	L	GT	VSYG	EEKPA	VLGHDE	
Ab_Omp38	NV	DASR	L	STQ	GFAWD	P	IADNKT	
Ab_Omp38	NV	DASR	L	STQ	GFAWD	P	IADNKT	
Ec_YiaD	GV	DASR	L	RTQGL	GPAN	P	IASNST	
Ec_YiaD	GV	DASR	L	RTQGL	GPAN	P	IASNST	
Mt_Arfa	GV	AGDH	I	ATVGL	G	SVN	P	IASNATP
Mt_Arfa	GV	AGDH	I	ATVGL	G	SVN	P	IASNATP
Ppa_T6SS	TG	DAGR	F	TA	EGRS	D	TE	P
Ppa_T6SS	TG	DAGR	F	TA	EGRS	D	TE	P
Hp_MotB	GV	PNQO	L	SF	SSY	G	STN	P
Hp_MotB	GV	PNQO	L	SF	SSY	G	STN	P
Se_MotB	GI	DNGK	L	RV	VGMA	AT	MR	L
Se_MotB	GI	DNGK	L	RV	VGMA	AT	MR	L
Dv_MotB	GI	EPAR	L	TAT	GL	ELD	P	LPNTSD
Dv_MotB	GI	EPAR	L	TAT	GL	ELD	P	LPNTSD

Figure 2. A. Multiple sequence alignment of OmpA-like domains present in the Protein Data Bank. The sequences were retrieved from Uniprot database and the C-terminal periplasmic domains were aligned by using the program MAFFT (Kato et al. 2002). Figure rendering was made by using the server ESPript (<http://esprict.ibcp.fr/ESPript/ESPript/>) (Robert et al. 2014). Residue numbers of C-KpOmpA and secondary structure elements are indicated above the alignment. Cysteines forming a disulfide bond are highlighted in green. Strictly conserved residues are in white with red background while residues well-conserved within a group are in red with yellow background. Residues important for the interaction with the citrate are highlighted below the sequence alignment with star when they interact by their side-chain and with triangle when they interact by their backbone amide. For each sequence, the Uniprot ID and the PDB ID are given. Kp_OmpA, *Klebsiella pneumoniae* OmpA [OMPA_KLEPN]; Ec_OmpA, *Escherichia coli* OmpA [OMPA_ECOLI – 2MQE]; Se_OmpA, *Salmonella enterica* OmpA [D0ZTJ5_SALT1 – 4ERH]; Nm_RmpM, *Neisseria meningitidis* RmpM [OMP4_NEIMA – 1R1M]; Ab_Pal, *Acinetobacter baumannii* Pal [D0CBL2_ACIBA – 4G4V]; Bp_Pal, *Burkholderia pseudomallei* Pal [Q63RA7_BURPS – 4B5C]; Ec_Pal, *Escherichia coli* Pal [PAL_ECOLI – 1OAP]; Yp_Pal, *Yersinia pestis* Pal [Q7CH55_YERPE – 4PWT]; Hi_OmpP6, *Haemophilus influenza* OmpP6 [PAL_HAEIN – 2AIZ]; Ab_Omp38, *Acinetobacter baumannii* Omp38 [OMP38_ACIBA – 3TD5]; Ec_YiaD, *Escherichia coli* YiaD [YIAD_ECOLI – 2K1S]; Mt_ArfA, *Mycobacterium tuberculosis* ArfA [ARFA_MYCTU – 2KGW]; Pa_T6SS, *Pseudomonas aeruginosa* T6SS [Q9I754_PSEAE – 4B62]; Hp_MotB, *Helicobacter pylori* MotB [MOTB_HELPY – 3CYQ]; Se_MotB, *Salmonella enterica* MotB [MOTB_SALTY – 2ZVY] and Dv_MotB, *Desulfovibrio vulgaris* MotB [Q729W9_DESVH – 3KHN].

The other OmpA-like domains are arranged in two supplementary group by sequence identity. The sequences of Pal have an identity between 44% and 90%, and include sequences of *Acinetobacter baumannii* Pal (PDB ID: 4G4V), *Burkholderia pseudomallei* Pal (PDB ID: 4B5C) (Gourlay et al. 2013), *Escherichia coli* Pal (PDB ID: 1OAP), *Yersinia pestis* Pal (PDB ID: 4PWT) and *Haemophilus influenza* OmpP6 (PDB ID: 2AIZ) (Parsons et al. 2006). A third group with sequence identity between 27% and 43% is composed of sequences of *Acinetobacter baumannii* Omp38 (PDB ID: 3TD5) (Park et al. 2012), *Escherichia coli* YiaD (PDB ID: 2K1S), *Mycobacterium tuberculosis* ArfA (PDB ID: 2KGW) (Yang et al. 2011) and *Pseudomonas aeruginosa* T6SS (PDB ID: 4B62).

Finally, OmpA-like domains are also retrieved in MotB protein as *Helicobacter pylori* MotB (PDB ID: 3CYP) (Roujeinikova 2008), *Salmonella enterica* MotB (PDB ID: 2ZVY) (Kojima et al. 2009) and *Desulfovibrio vulgaris* MotB (PDB ID: 3KHN). However these three MotB proteins have a low sequence identity between 16% and 24%.

OmpA-like sequences have low sequence identity within the entire group (under 30%), and it is thus remarkable that they share a conserved fold. For instance, using DaliLite server for structural alignment (http://ekhidna.biocenter.helsinki.fi/dali_lite/) (Hasegawa et al. 2009), C-KpOmpA have a backbone root mean square deviation (RMSD) varying from 0.8 to 3.0 Å with the afore-mentioned structure. The core is the most structurally conserved while variations are found mainly in the bulge and at the N and C-terminii. Indeed, the N-terminal β strand is significantly shorter in the structure of *Acinetobacter baumannii* Pal (PDB ID: 4G4V), *Burkholderia pseudomallei* Pal (PDB ID: 4B5C), *Escherichia coli* Pal (PDB ID: 1OAP) and *Haemophilus influenza* OmpP6 (PDB ID: 2AIZ). This short β strand is even partially or completely replaced by a α helix for *Yersinia pestis* Pal (PDB ID: 4PWT) and *Mycobacterium tuberculosis* ArfA (PDB ID: 2KGW), respectively. On the contrary, the N-

terminal β strand is completed by a supplementary β strand and a long α helix in the structure of *Pseudomonas aeruginosa* T6SS (PDB ID: 4B62), *Salmonella enterica* MotB (PDB ID: 2ZVY) and *Desulfovibrio vulgaris* MotB (PDB ID: 3KHN). In a same way, a supplementary α helix is found after the last C-terminal β strand in the structure of *Helicobacter pylori* MotB (PDB ID: 3CYP) and *Salmonella enterica* MotB (PDB ID: 2ZVY).

3.2. Binding site and sequence conservation.

Crystal obtained in presence of citric acid presented an electron density inside the peptidoglycan binding which has been identified in the homologous *Acinetobacter baumannii* OmpA (PDB ID: 3TD4) (Park *et al.* 2012). A molecule of citrate could be fitted inside the electron density (Figure 3A). The carboxyl group of citrate interacts with the guanidinium group of the conserved arginine 275 (2.8 Å) while a second carboxyl group binds the backbone amide of asparagine 224 (2.8 Å). This N224 is part of the β -turn N222 – K225 which orients the backbone amide group of N224 inward the binding pocket. This β -turn keeps also the carbonyl group of N222 and K225, residues *i* and *i*+3 of the β -turn respectively, directed to the N_ϵ (2.9 Å) and the N_{η_2} (3.1 Å) of the guanidinium group of the sidechain of R275. This orientation of the backbone carbonyl toward the R275 is maintained throughout the crystallographic structures of homologous OmpA-like domains. R275 is also in close contact with the apolar residues L271 and F221. F221 is strictly conserved while L271 is substituted by a valine only in the *Desulfovibrio vulgaris* MotB sequence (Figure 2). These two residues form a hydrophobic interaction network with R275 sidechain and completed by Y267 (Figure 3B). The importance of steric hindrance and hydrophobicity of F221 and L271 to maintain the R275 sidechain toward the binding site are outlined by the conservation of these residues among the OmpA-like sequences. The backbone RMSD between the citrate complexed form and the Apo form is 0.9 Å and 0.6 Å for chain A and chain B of the Apo form, respectively. Main differences are localized in the loop between helix α_2 and strand β_2 (Figure 1B) and to a lesser extend in the loop between helix α_4 and the series of β -turn as well as the β -turn between helix α_1 and α_2 (Supplementary Figure SC).

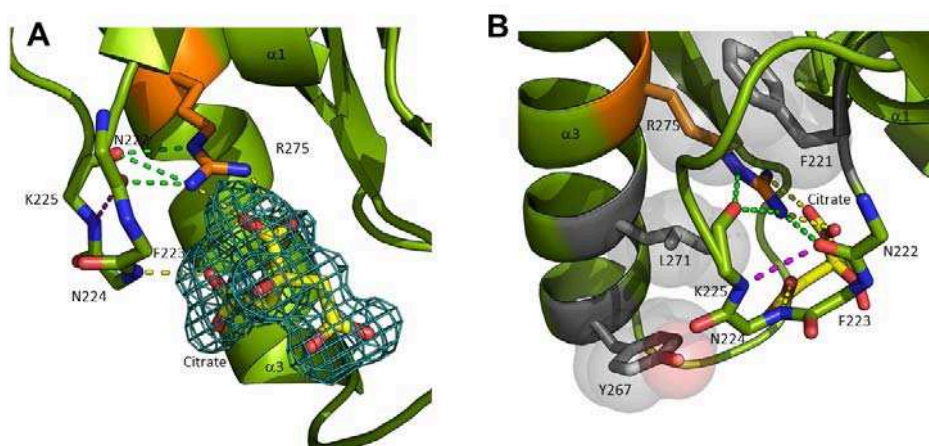


Figure 3. A. View of the citrate inside the binding site represented in stick mode and its electron density in the omit map contoured at 3 σ . Conserved arginine 275 is represented as orange stick while the backbone of β -turn α 1- β 2 is represented as green stick. Interactions between the citrate and C-KpOmpA are represented as yellow dashed lines and interactions between R275 sidechain and the backbone are shown as green dashed lines. Interaction between CO of N222 and NH of K225, residues i and $i+3$ of β -turn α 1- β 2 respectively, is represented as magenta dashed line. B. Side view of the citrate in the binding site. Conserved hydrophobic residues F221, L271 and Y267 are represented as sticks and transparent grey spheres.

3.3. Putative role of conserved D325 – R326 – R327.

Highly conserved residues on the multiple sequence alignment could be classified into three distinct groups. Most of the conserved residues are mainly apolar amino-acids (Figure 2), mapped on the structure of C-KpOmpA, they are part of the both helix α 2 and α 3 and constitute the hydrophobic core of the protein (Figure 3A). More interestingly polar amino-acids R275, D260 and N268 (Figure 3B), are located inside the binding site of C-KpOmpA and have been shown to be important for interaction with the peptidoglycan (Supplementary Figure SA) (Park *et al.* 2012). Finally, D325, R326 and R327 are localized between helix α 4 and strand β 4, at the hinge between the bulge and the core domain (Figure 3B). These three amino acid are strictly conserved for the two arginines or substituted only by an asparagine for the aspartic acid (Figure 2) and participate to the structure of the bulge.

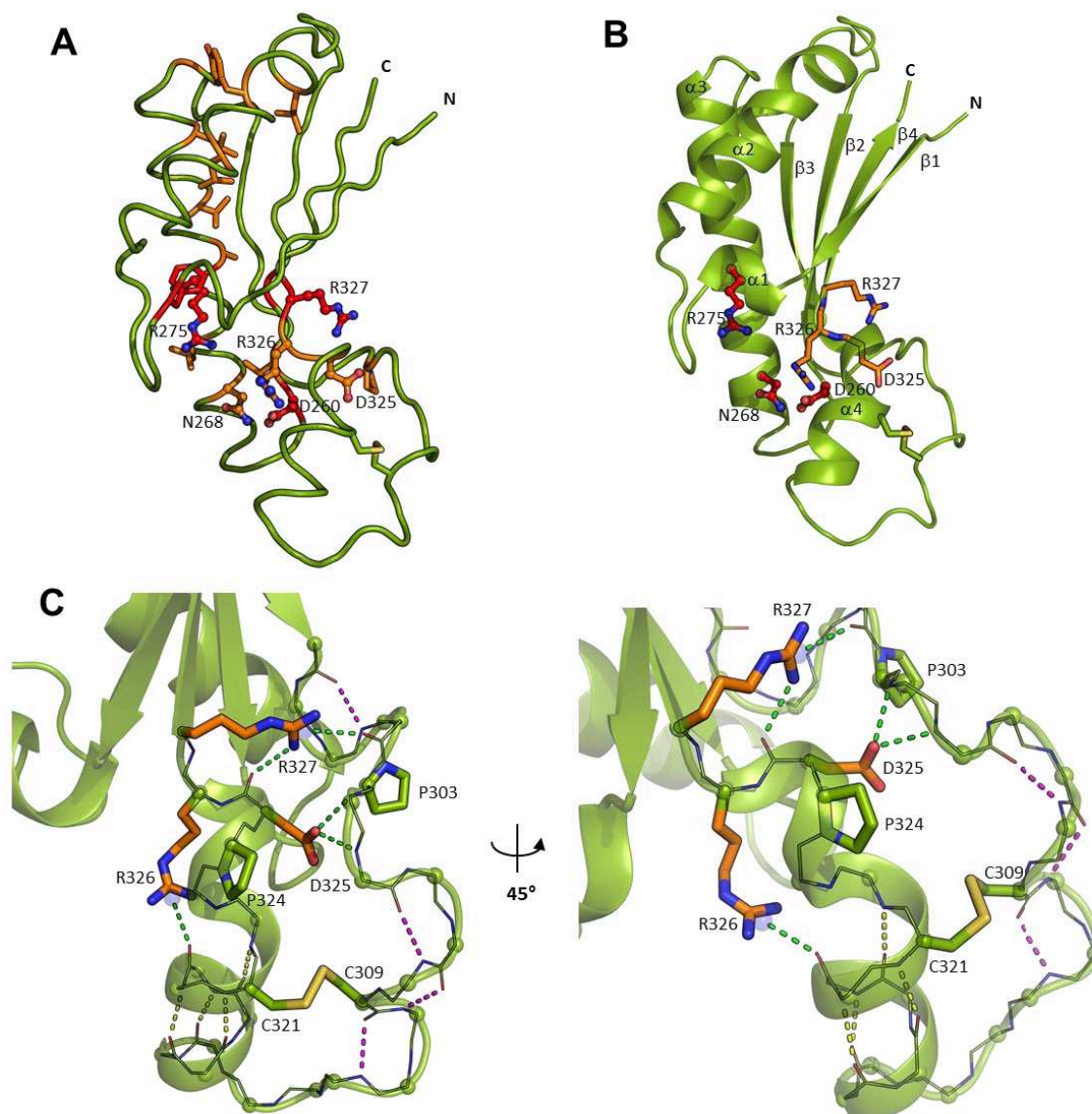


Figure 4. A. Structure of C-KpOmpA with conserved amino acid. Conserved amino acid are colored in red for strictly conserved amino acids and in orange for amino acids conserved in 75% of sequences. Apolar amino acids are represented as sticks while polar amino acid are represented as ball and sticks. B. Conserved polar amino acid represented as sticks on C-KpOmpA structure. Amino acid interacting with the citrate or the DAP in homologous OmpA-like domains are colored in red and represented as ball and sticks (Park *et al.* 2012). C. Structure of the bulge of C-KpOmpA. The backbone is represented as lines and spheres for C α atoms. Sidechains of proline, cysteine and polar conserved amino acids are shown as stick. Interaction between CO and NH are represented by dash lines in magenta for β -turn, yellow for α helix and green between sidechain and backbone.

The first strand of the bulge encompasses residue 298 – 312 and is lacking a regular α -helix structure as found in *Escherichia coli* OmpA or *Neisseria meningitidis* RmpM. Instead residue 305 – 312 form a series of 3 β -turns, T305 – T308, N307 – D310, C309 – V312, and are linked to the second strand of the bulge by the disulfide bond between C309 and C321 (Figure 3C). The upstream sequence includes β -turn M298 – S301 while the sidechain of R327 is directed toward the backbone carbonyl of N302 (3.1 Å). In a same way, D325 carboxylate point towards V304 and T305 backbone NH (2.7 Å and 2.8 Å) as

found in crystallographic structure of homologous OmpA-like domains (Figure 5). However, the interaction between R327 and the backbone carbonyl can be mediated by a molecule of water as in *Neisseria meningitidis* RmpM (PDB ID: 1R1M). With the disulfide bond, these two residues participate to the tight association of the both strands of the bulge.

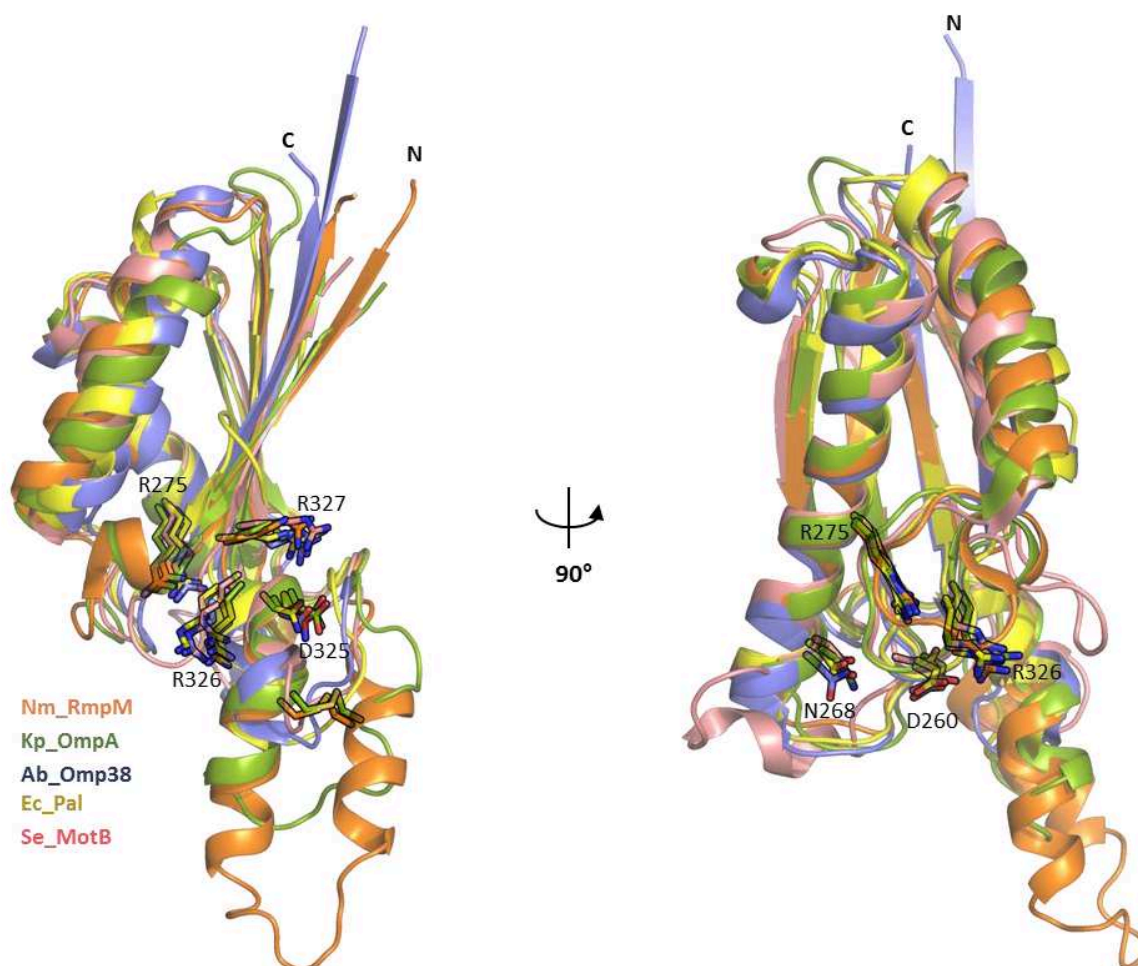


Figure 5. Superposition of homologous OmpA-like C-terminal domains. Structures were retrieved from Protein Data Bank and aligned with Pymol. Sidechain of cysteines and conserved polar residues are shown as stick. *Klebsiella pneumoniae* OmpA; Nm_RmpM, *Neisseria meningitidis* RmpM (PDB ID: 1R1M); Ab_Omp38, *Acinetobacter baumannii* Omp38 (PDB ID: 3TD5); Ec_Pal, *Escherichia coli* Pal (PDB ID: 1OAP) and Se_MotB, *Salmonella enterica* MotB (PDB ID: 2ZVY). For clarity, the first N-terminal α helix and β strand as well as the last C-terminal α helix of *Salmonella enterica* MotB are not shown.

The second strand includes residues 315 – 327 and is composed of one α helix, residues 315 – 322, followed by a 3.10 helix, residues 323 – 327. The two helices form an angle of 138° and the proline 324 support the transition between the two helices. The bend between the two helices allows sidechain of R326 to contact backbone carbonyl of I319 (2.7 Å). However this interaction is only retrieved in *Salmonella enterica* OmpA (PDB ID: 4ERH) and *Neisseria meningitidis* RmpM (PDB ID: 1R1M) crystallographic structure. In other homologous OmpA-like structure, R326 sidechain coordinates a molecule of water

which is close to the carbonyl backbone of residue 319 (less than 3.0 Å). Furthermore, R326 is close to the peptidoglycan binding site even if no direct interaction involved in the binding was retrieved in structures of *Acinetobacter baumannii* Pal (PDB ID: 4G4V), *Haemophilus influenza* OmpP6 (PDB ID: 2AIZ) (Parsons *et al.* 2006) or *Acinetobacter baumannii* Omp38 (PDB ID: 3TD4 and 3TD5) (Park *et al.* 2012). Therefore the role of this residue is not clear although its importance is emphasized by its conservation in the multiple sequence alignment (Figure 2).

3.4. Dimerization interface.

In the citrate bound C-KpOmpA crystal, there is one molecule per asymmetric unit while two molecules are present for the Apo form. It has been shown that OmpA-like domains are able to form dimers (Grizot *et al.* 2004, Kojima *et al.* 2009, Zheng *et al.* 2011, Marcoux *et al.* 2014). Therefore, we examined the possible interface between molecules present in the asymmetric unit or related by crystallographic symmetry. A first interface could be generated by a crystallographic twofold symmetry axis of the citrate bound C-KpOmpA. This interface of 801 Å² involves the loop $\alpha1 - \alpha2$, the loop $\beta2 - \alpha3$ and the helix $\alpha4$ and places the two binding sites in front of each other distant of 9 Å (Supplementary Figure S3A). A second interface of 628 Å², found by a twofold crystallographic symmetry of the citrate bound C-KpOmpA, places the two molecules side by side and is composed of helix $\alpha3$ and the strand $\beta3$ (Figure 8A). In this configuration, the two binding sites are 35 Å distant each other. Finally the third interface involves strand $\beta1$ and the series of β -turn from the bulge placing the two molecules side by side (Supplementary Figure S3B). This interface has an area of 482 Å² and separates the two binding sites by 32 Å. This configuration is found in the asymmetric unit of the Apo form of C-KpOmpA or by a threefold symmetry axis of the citrate complexed C-KpOmpA.

The first interface is unlikely as the binding site is completely buried between the two molecules and would not be accessible for interaction with the peptidoglycan (Supplementary Figure S3A). The third interface has an accessible binding site, however the dimerization is based on the interaction between the bulge and the strand $\beta1$ (Supplementary Figure S3B). As previously stated, the length of the bulge varies among the OmpA-like sequences. In a similar way, the sequence corresponding to the $\beta1$ strand adopts various structures in its N-terminal part.

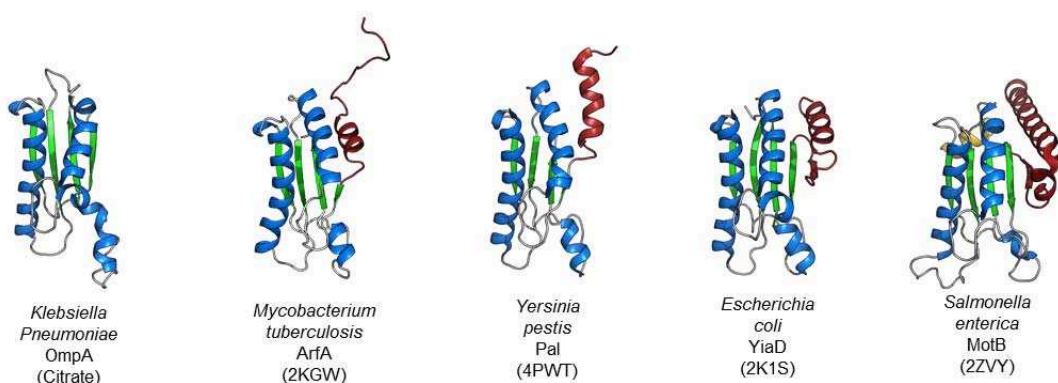


Figure 6. Structures of *Klebsiella pneumoniae* OmpA, *Mycobacterium tuberculosis* ArfA (PDB ID: 2KGW), *Yersinia pestis* Pal (PDB ID: 4PWT), *Escherichia coli* YiaD (PDB ID: 2K1S) and *Salmonella enterica* MotB (PDB ID: 2ZVY). Secondary structure elements of the core of OmpA-like domain are colored in blue and green. Supplementary elements in N-terminal or C-terminal are colored in red and yellow, respectively.

Indeed, this β strand is shorter and replaced or completed by an α helix in structures of *Mycobacterium tuberculosis* ArfA, 2KGW; *Escherichia coli* Pal, 1OAP; *Acinetobacter baumannii* Pal, 4G4V; *Burkholderia pseudomallei* Pal, 4B5C; *Yersinia pestis* Pal, 4PWT and *Haemophilus influenza* OmpP6, 2AIZ; while a supplementary β strand and α helix precede it in structures of *Escherichia coli* YiaD, 2K1S; *Pseudomonas aeruginosa* T6SS, 4B62; *Salmonella enterica* MotB, 2ZVY and *Desulfovibrio vulgaris* MotB, 3KHN (Figure 6). Therefore, the structural variability among OmpA-like domain for this interface is a serious impediment for a common dimerization process (Figure 7).

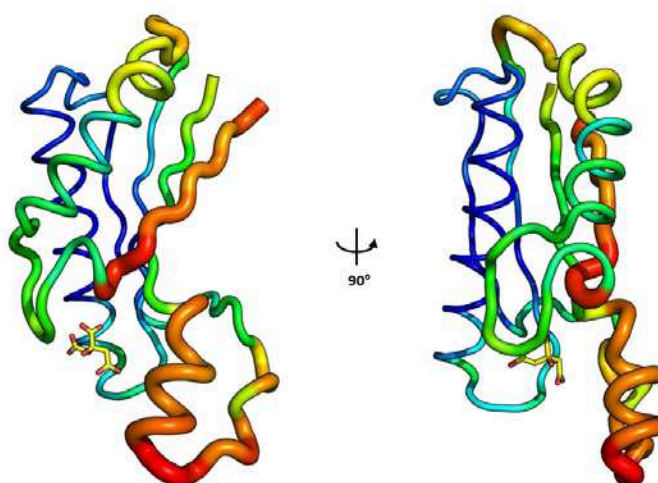


Figure 7. Structural variation among the OmpA-like domains. The structure of C-KpOmpA complexed with the citrate was submitted to the ENDscript server (<http://endscript.ibcp.fr>) (Robert *et al.* 2014). Homologous OmpA-like domains are retrieved from the Protein Data Bank and aligned using ProFit. Size of the tube is proportional to the r.m.s. deviation per residue between C α pairs. Following OmpA-like domain were used by ENDscript: *Neisseria meningitidis* RmpM (PDB ID: 1R1M), *Salmonella enterica* (PDB ID: 4ERH), *Escherichia coli* OmpA (PDB ID: 2MQE), *Acinetobacter baumannii* Pal (PDB ID: 4G4V), *Haemophilus influenzae* OmpP6 (PDB ID: 2AIZ), *Escherichia coli* Pal (PDB ID: 1OAP), *Yersinia pestis* Pal (PDB ID: 4PWT), *Escherichia coli* YiaD (PDB ID: 2K1S), *Mycobacterium tuberculosis* ArfA (PDB ID: 2KGW), *Burkholderia pseudomallei* Pal (PDB ID: 4B5C), *Pseudomonas aeruginosa* T6SS (PDB ID: 4B62) and *Desulfovibrio vulgaris* MotB (PDB ID: 3KHN).

The second interface formed by helix $\alpha 3$ and strand $\beta 3$ is the same as in *Salmonella enterica* MotB dimer (2ZVY). Indeed, this interface of dimerization was described in the crystallographic structure of *Salmonella enterica* MotB and confirmed in solution by size-exclusion chromatography and site-directed mutagenesis (Kojima *et al.* 2009). This interface places the two peptidoglycan binding sites on the same side of the dimer and 35 Å apart from each other (Figure 8A) while the two N-termini are located on the opposite side of the dimer and separated by 27 Å (Figure 8C).

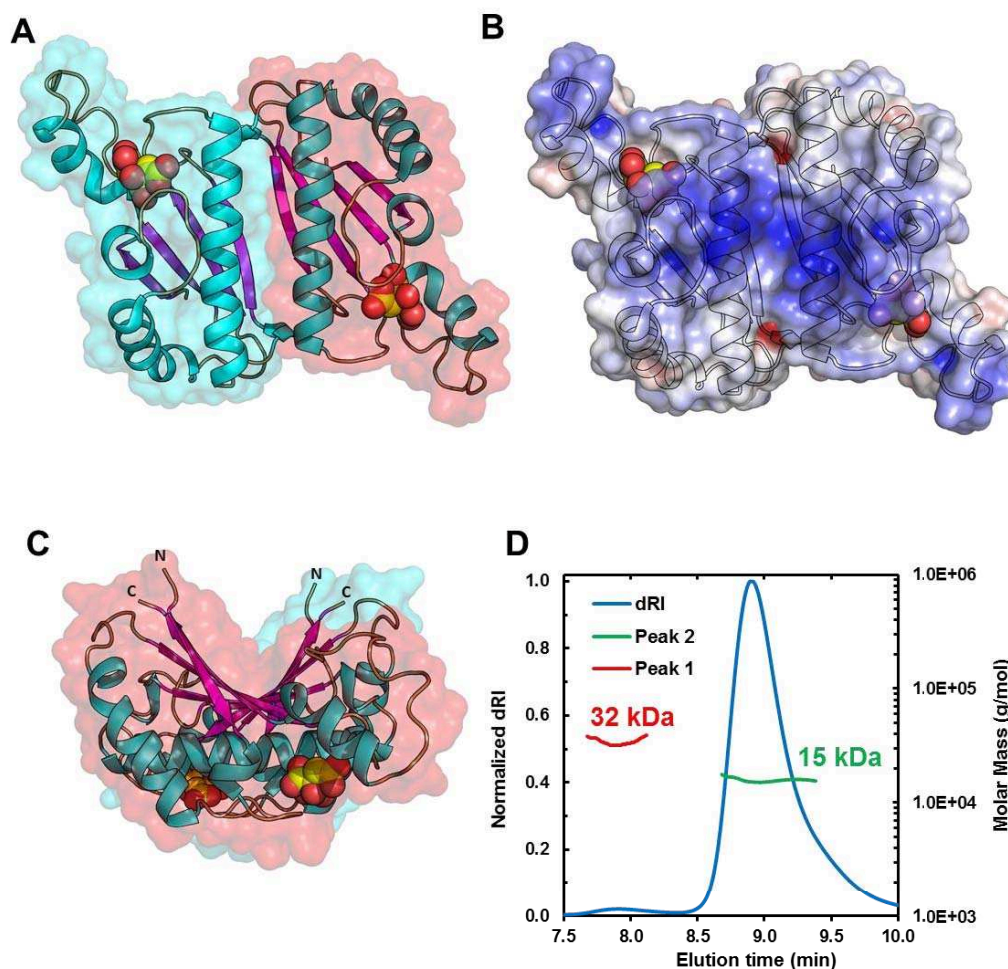


Figure 8. A. Cartoon representation of putative C-KpOmpA dimer with surface representation of each monomer colored in red and cyan. Citrate is shown as sphere. B. Electrostatic potential mapped on the surface of C-KpOmpA and colored from red for negatively charged region (-7 kTe^{-1}) to blue for positively charged region ($+7 \text{ kTe}^{-1}$). The electrostatic potential was calculated using all default values with the PDB2PQR server and the Adaptive Poisson-Boltzmann Solver (Baker *et al.* 2001, Dolinsky *et al.* 2004). C. Side view of the putative C-KpOmpA dimer represented as cartoon and surface with citrate shown as sphere. D. Elution profile from SEC-MALLS analysis of C-KpOmpA. Differential refractive index and calculated molecular mass are plotted against retention time.

Calculation of the electrostatic surface potential shows the presence of a positively charged groove crossing the surface between two binding site which could accommodate the negatively charged peptidoglycan (Figure 8B). To check the possibility of dimerization of C-KpOmpA in solution, we analyzed its elution profile by SEC-MALLS (Figure 8D). C-KpOmpA elutes as two peaks by size exclusion chromatography, at a 0.8 mM concentration. The major peak had a molecular mass of 15 kDa which corresponds to the monomer while the minor peak had a molecular mass of 32 kDa as expected for the dimer. In a 0.8 mM solution C-KpOmpA exists as a mixture of monomer and dimer, the monomer being dominant. Thus the affinity between the two C-terminal domains alone in solution is fairly weak and other interactions must contribute *in vivo* for this dimerization to be effective. A similar equilibrium between monomeric and dimeric form was observed for *Escherichia coli* OmpA (Ishida *et al.* 2014, Marcoux *et al.* 2014).

3.5. Dynamics in solution.

To further characterize KpOmpA-C, we investigated its dynamics in solution by NMR. All backbone amide resonances could be assigned in the ^1H - ^{15}N HSQC spectrum (Supplementary Figure S4) using conventional three-dimensional CBCA(CO)NH, CBCANH, HN(CO)CA and HNCA experiments with ^{15}N - ^{13}C labeled protein (Supplementary Figure S5). Resonance assignments will be deposited in the BMRB database. TALOS-N was used to predict secondary structures of C-KpOmpA in solution (Shen *et al.* 2013). Fold of C-KpOmpA in solution was assessed by comparing predicted secondary structure with secondary structures identified in C-KpOmpA crystallographic structure using the DSSP Software (Supplementary Figure S6) (Kabsch *et al.* 1983, Touw *et al.* 2015). Since a similar fold was observed, we did not perform an independent NMR structure determination but rather focused our efforts in studying the oligomerization state and the protein dynamics. Based on TRACT experiments (Supplementary Figure S7) (Lee *et al.* 2006), an effective rotational correlation time constant τ_c of 8.3 ns at 293K was measured which is coherent with a monomeric globular protein of 15 kDa (Rossi *et al.* 2010). This is consistent with the conclusion derived from SEC-MALLS experiments, in which the monomeric form was the major one.

To probe the motion at a picosecond – nanosecond timescale, we recorded ^{15}N relaxation parameters for the amide ^{15}N nuclei (Figure 9A). Consistently with a well-structured protein, the majority of the $^1\text{H} - ^{15}\text{N}$ heteronuclear NOE values were higher than 0.8. Residues M246 and S264 in the loops $\alpha 2 - \beta 2$ and $\beta 2 - \alpha 3$, respectively, as well as residues K313 – A314 – R315 in the loop between the series of β -turn and $\alpha 4$ helix exhibited R_2 values lower than the average, indicating rapid motion in the picosecond – nanosecond timescale. Significantly lower R_2 values were recorded for residues from the N and the C-termini indicating that these parts of the protein are fully disorganized (Figure 9A). ^{15}N relaxation parameters were used to perform standard Lipari-Szabo model-free analysis (Lipari *et al.* 1982a, Lipari *et al.* 1982b). The best fitting of the relaxation parameters were obtained using an axially symmetric tensor model for the molecular rotational diffusion tensor. No significant improvement was found using a fully asymmetric tensor model.

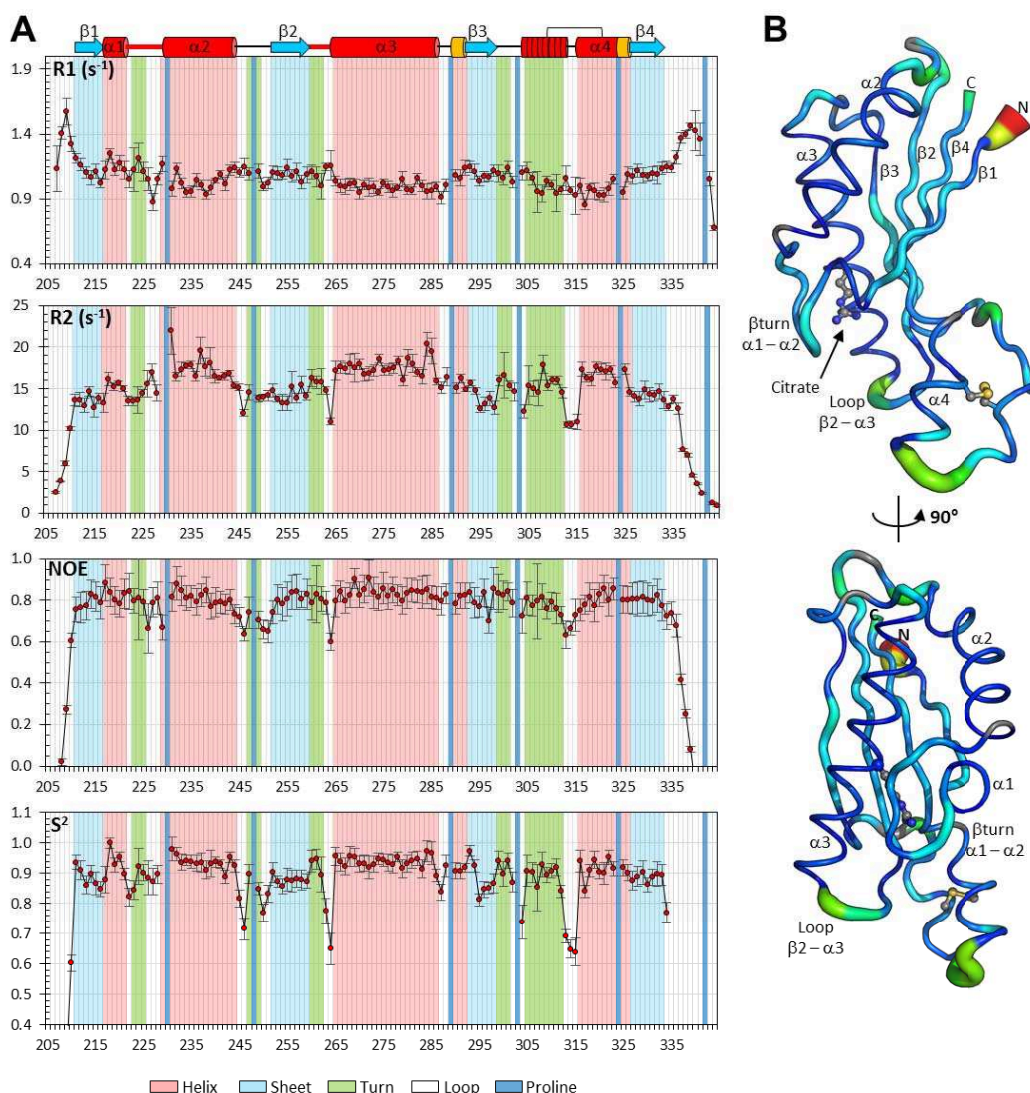


Figure 9. A. Backbone dynamics of C-KpOmpA. ^{15}N R1, ^{15}N R2, ^1H - ^{15}N steady-state heteronuclear NOEs and S^2 order parameter are plotted as a function of residue number. Secondary structures are shown on the top and by the background color. B. Sausage representation C-KpOmpA structure colored by S^2 order parameter. Radius of cartoon representation is proportional to the inverse of S^2 order parameter. Residues for which no data are available are shown in grey.

Most residues exhibited values of S^2 between 0.8 and 1.0, indicating the absence of fast motion in the N – H vector. With a mean S^2 of 0.94, α helices present a slightly higher order parameter than β strands which have a mean S^2 of 0.89. Of particular interest are the order parameters of residues 299 – 312 within the bulge. These residues are not structured in α helix or β strand, however their order parameter S^2 is particularly high with a mean value of 0.89 and a minimum of 0.74 for V304 (Figure 9A). This can be explained by the presence of stabilizing interactions described earlier. Sidechain – backbone interactions (R327 – N302, D325 – V304/T305), proline 303, disulfide bond (C309 – C321) and β -turn (M298 – S301, T305 – T308, N307 – D310, C309 – V312) participate to the structuration of these residues (Figure 4C).

Four loops present interesting order parameter values, the other junction between α helix and β strand being too short, two amino-acid or less, to have noticeable mobility. Excluding residues from the N-terminal, residues K313 – A314 and R315 constituting the loop between the series of β -turn and the α_4 helix have the lower order parameter with S264, 0.65, and are therefore the most flexible residues of the domain (Figure 9B). S264 is part of the loop between β_2 strand – α_3 helix which is part of the peptidoglycan binding site. While S263 and G264 have order parameter of 0.7 – 0.65, the beginning of this loop presents a lower mobility, 0.9, which may be related to the presence of an Asx turn between D260 and I262. The loop between α_1 helix – α_2 helix is also part of the binding site however residues 222 – 228 exhibit a S^2 between 0.82 and 0.92 which can be explained again by the presence of a β -turn (N222 – K225) and the sidechain – backbone (K225 – R275) interactions previously described (Figure 3A). Furthermore the presence of a phenylalanine in the loop, F223, could participate to the high order parameter (Goodman *et al.* 2000, Reddy *et al.* 2010). Finally the loop between α_2 helix – β_2 strand have a mean order parameter of 0.80 and the presence of an α -turn (D247 – G251) and proline 248 seems to significantly participate to increased order parameter.

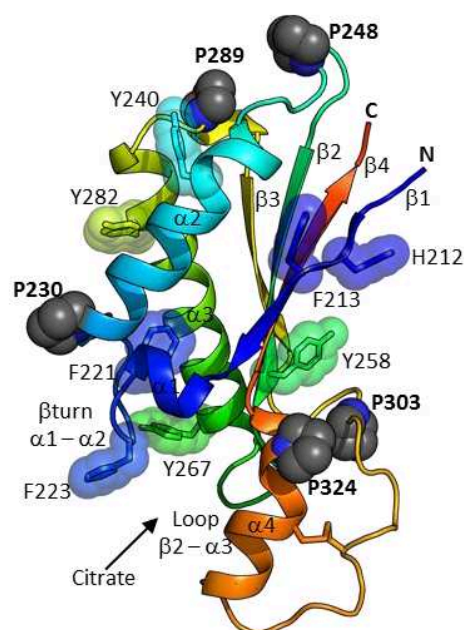


Figure 10. Cartoon representation of C-KpOmpA with proline sidechains shown as sphere and sidechains of aromatic residues shown as sticks.

The overall S^2 order parameter of the loops of the C-terminal domain is remarkably high. Furthermore, we could fit the ^{15}N relaxation parameter without any R2 exchange contribution, suggesting the absence of microsecond – millisecond timescale motion.

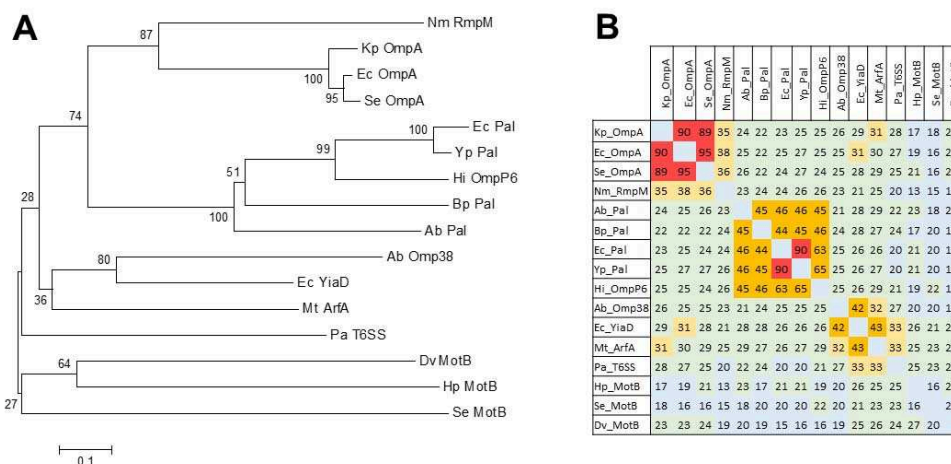
Therefore, the high level of structuration of the domain and the presence of small secondary structure or proline (Figure 10) in almost all the loops is a reasonable explanation.

IV. Conclusion:

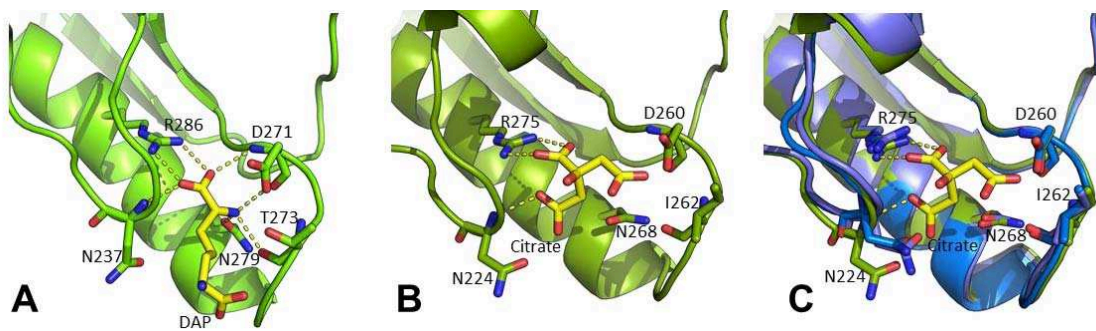
We have determined by crystallography and NMR spectroscopy the structure of the C-terminal domain of *Klebsiella pneumoniae* OmpA. The domain adopts a characteristic α/β fold of the OmpA-like domain (Grizot *et al.* 2004). Besides the Apo form, we solved the structure of the domain bound to a citrate molecule, mimicking the peptidoglycan. The resolution of the C-terminal domain in complex with the citrate confirmed the peptidoglycan binding site and the essential role of the R245. The binding site compare to similar structures of OmpA-like domain (Park *et al.* 2012) defined essential amino acids responsible for a specific binding of the peptidoglycan. Apo form of C-KpOmpA crystallized with two molecule in asymmetric unit. The properties of oligomerization of the domain were analyzed in solution by NMR and SEC-MALLS. We found that this domain is able to form dimer although the dimeric form remains marginal with a weak affinity higher than millimolar suggesting that dimerization is driven by other interactions to be effective *in vivo*. Finally, study of ^{15}N relaxation showed the absence of fast motion, characteristic of a well-structured protein, even for the extended sequence of the bulge where conserved polar amino-acids form interactions which contribute to the rigidity of the C-KpOmpA domain

V. Supplementary information

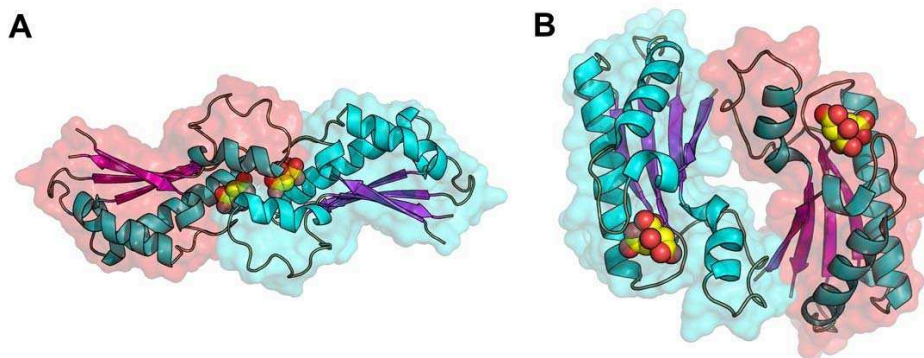
5.1. Supplementary Figures



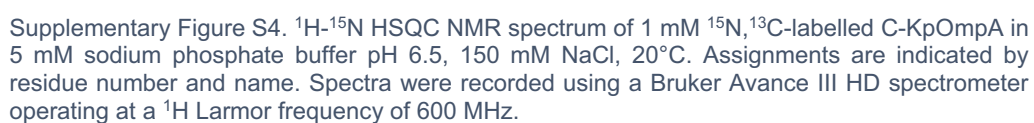
Supplementary Figure S1. A. Phylogenetic tree of OmpA-like domain. The evolutionary history was inferred using the Neighbor-Joining method (Saitou *et al.* 1987). The optimal tree with the sum of branch length = 7.24 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches (Felsenstein 1985). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method (Zuckerandl *et al.* 1965) and are in the units of the number of amino acid substitutions per site. The analysis involved 16 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 114 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura *et al.* 2013). B. Sequence identity in % and color coded: higher than 80% in red; higher than 40% in orange, higher than 20% in green and blue under 20%.

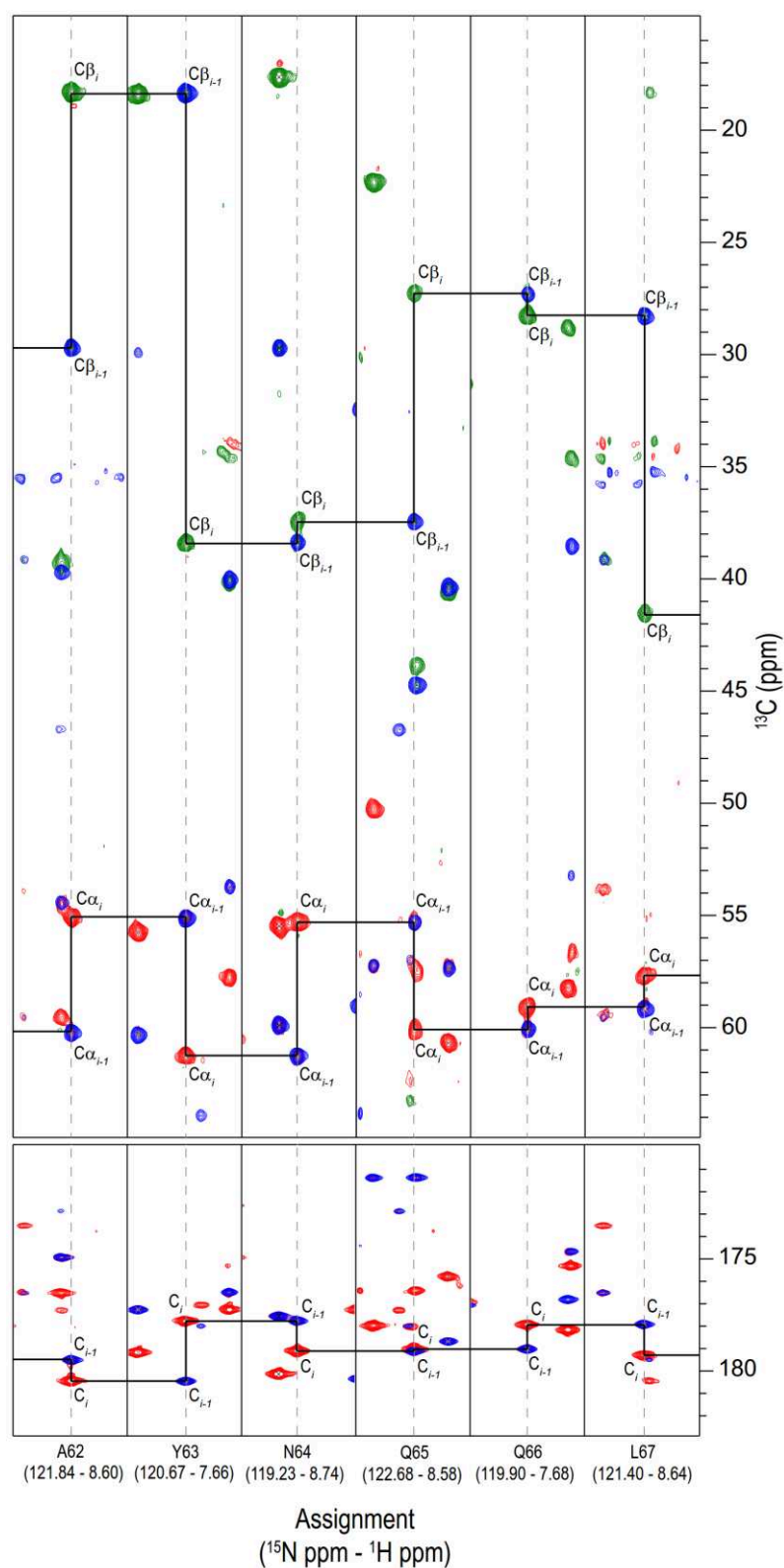


Supplementary Figure S2. A. DAP in the binding site of C-terminal domain of *Acinetobacter baumannii* OmpA (PDB ID: 3TD5) (Park *et al.* 2012). B. Citrate in the binding-site of C-terminal domain of *Klebsiella pneumoniae* OmpA. C. Superposition of the Holo and the Apo forms of C-terminal domain of *Klebsiella pneumoniae* OmpA, in green and in blue, respectively.

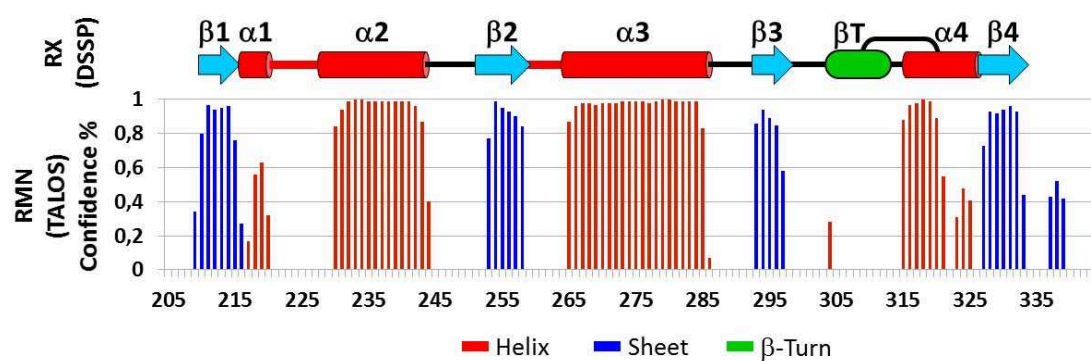


Supplementary Figure S3. A. Dimer formed by the first interface of dimerization involving the loop $\alpha1 - \alpha2$, loop $\beta2 - \alpha3$ and the helix $\alpha4$. B. Dimer formed by the third interface of dimerization and involving the strand $\beta1$ and the series of β -turn from the bulge.

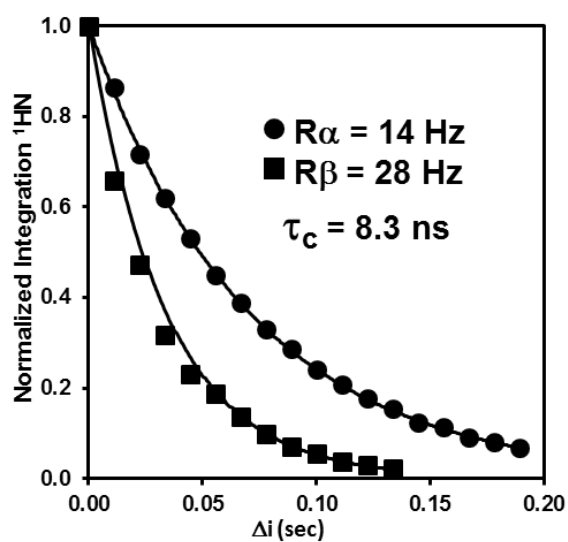




Supplementary Figure S5. Sequential residue connectivities using triple resonance NMR experiments CBCA(CO)NH/CBCANH and HNCO/HN(CA)CO. Spectra CBCA(CO)NH and HNCO are colored in blue while CBCANH and HN(CA)CO are colored in red and green for positive and negative intensity, respectively. Sequential ^{13}C connectivities for residues A62-L67 are indicated by a continuous line for $\text{C}\beta$, $\text{C}\alpha$ and C and labelled i or $i-1$ for intra and inter-residues correlation, respectively. Spectra were recorded using a Bruker Avance III HD spectrometer operating at a ^1H Larmor frequency of 600 MHz.



Supplementary Figure S6. Estimated secondary structure of C-KpOmpA in solution using TALOS-N to predict protein backbone torsion angle from NMR chemical shift (Shen *et al.* 2013) and comparison with secondary structure identified in C-KpOmpA crystallographic structure using the DSSP Software (Kabsch *et al.* 1983, Touw *et al.* 2015).



Supplementary Figure S7. ^1H - ^{15}N TRACT NMR experiment for the determination of the effective rotational correlation time constant τ_c (Lee *et al.* 2006). The one-dimensional proton signal intensity was integrated between 6.5 and 10 ppm. Exponential fit yielded $R_{\alpha} = 14 \text{ Hz}$ and $R_{\beta} = 28 \text{ Hz}$ resulting in a $\tau_c = 8.3 \text{ ns}$ which is coherent for a globular protein of about 15 kDa size.

5.2. Supplementary Tables

Supplementary Table S1. Data collection and refinement statistics of two crystals form of C-terminal domain of KpOmpA. (Statistics for the highest-resolution shell are shown in parentheses.)

	P4₃22	C222₁
Wavelength (Å...)	0.9794	0.9686
Resolution range (Å...)	39.65 – 1.9 (1.96 – 1.90)	76.63 – 1.81 (1.88 – 1.81)
Space group	P 43 2 2	C 2 2 21
Unit cell	a=73.4 Å, b=73.4 Å, c=61.5 Å $\alpha=\beta=\gamma=90^\circ$	a=98.3 Å, b=122.4 Å, c=75.6 Å $\alpha=\beta=\gamma=90^\circ$
Total reflections	112927 (11164)	170150 (14954)
Unique reflections	13918 (1356)	40795 (3818)
Multiplicity	8.1 (8.2)	4.2 (3.9)
Completeness (%)	99.89 (99.56)	98.27 (92.94)
Mean I / sigma (I)	22.15 (3.28)	13.21 (1.78)
Wilson B-factor	27.08	35.42
R-merge	0.063 (0.727)	0.056 (0.626)
R-meas	0.067	0.064
CC1/2	0.999 (0.916)	0.998 (0.859)
CC*	1 (0.978)	0.999 (0.961)
Reflections used for R-free		
R-work	0.167 (0.245)	0.196 (0.377)
R-free	0.189 (0.310)	0.228 (0.415)
CC(work)		
CC(free)		
Number of non-hydrogen atoms	1137	2192
macromolecules	969	1922
ligands	13	44
water	155	226
Protein residues	126	251
RMS(bonds)	0.011	0.024
RMS(angles)	1.34	2.11
Ramachandran favored (%)	99	98
Ramachandran allowed (%)	1	2
Ramachandran outliers (%)	0	0
Clashscore	2.55	2.77
Average B-factor	34.2	46.3
macromolecules	33.3	44.9
ligands	37.6	72.2
solvent	40.0	52.7

VI. References

- Adams P D, Afonine P V, Bunkoczi G, Chen V B, Davis I W, Echols N, Headd J J, Hung L W, Kapral G J, Grosse-Kunstleve R W, McCoy A J, Moriarty N W, Oeffner R, Read R J, Richardson D C, Richardson J S, Terwilliger T C and Zwart P H. *PHENIX: a comprehensive Python-based system for macromolecular structure solution*. Acta Crystallogr D Biol Crystallogr, **2010**, 66(Pt 2): 213-221.
- Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner E G, Margalit H and Altuvia S. *Novel small RNA-encoding genes in the intergenic regions of Escherichia coli*. Curr Biol, **2001**, 11(12): 941-950.
- Baker N A, Sept D, Joseph S, Holst M J and McCammon J A. *Electrostatics of nanosystems: application to microtubules and the ribosome*. Proc Natl Acad Sci U S A, **2001**, 98(18): 10037-10041.
- Bosshart P D, Iordanov I, Garzon-Coral C, Demange P, Engel A, Milon A and Muller D J. *The transmembrane protein KpOmpA anchoring the outer membrane of Klebsiella pneumoniae unfolds and refolds in response to tensile load*. Structure, **2012**, 20(1): 121-127.
- Bradley P, Misura K M and Baker D. *Toward high-resolution de novo structure prediction for small proteins*. Science, **2005**, 309(5742): 1868-1871.
- Confer A W and Ayalew S. *The OmpA family of proteins: roles in bacterial pathogenesis and immunity*. Vet Microbiol, **2013**, 163(3-4): 207-222.
- Deatherage B L, Lara J C, Bergsbaken T, Rassouljian Barrett S L, Lara S and Cookson B T. *Biogenesis of bacterial membrane vesicles*. Mol Microbiol, **2009**, 72(6): 1395-1407.
- Dolinsky T J, Nielsen J E, McCammon J A and Baker N A. *PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations*. Nucleic Acids Res, **2004**, 32(Web Server issue): W665-667.
- Dosset P, Hus J C, Blackledge M and Marion D. *Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data*. J Biomol NMR, **2000**, 16(1): 23-28.
- Emsley P and Cowtan K. *Coot: model-building tools for molecular graphics*. Acta Crystallogr D Biol Crystallogr, **2004**, 60(Pt 12 Pt 1): 2126-2132.
- Felsenstein J. *Confidence Limits on Phylogenies: An Approach Using the Bootstrap*. Evolution, **1985**, 39(4): 783-791.
- Goodman J L, Pagel M D and Stone M J. *Relationships between protein structure and dynamics from a database of NMR-derived backbone order parameters*. J Mol Biol, **2000**, 295(4): 963-978.
- Gourlay L J, Peri C, Ferrer-Navarro M, Conchillo-Sole O, Gori A, Rinchai D, Thomas R J, Champion O L, Michell S L, Kewcharoenwong C, Nithichanon A, Lassaux P, Perletti L, Longhi R, Lertmemongkolchai G, Titball R W, Daura X, Colombo G and Bolognesi M. *Exploiting the Burkholderia pseudomallei acute phase antigen BPSL2765 for structure-based epitope discovery/design in structural vaccinology*. Chem Biol, **2013**, 20(9): 1147-1156.

Grizot S and Buchanan S K. *Structure of the OmpA-like domain of RmpM from Neisseria meningitidis*. Mol Microbiol, **2004**, 51(4): 1027-1037.

Hasegawa H and Holm L. *Advances and pitfalls of protein structural alignment*. Curr Opin Struct Biol, **2009**, 19(3): 341-348.

Iordanov I, Renault M, Reat V, Bosshart P D, Engel A, Saurel O and Milon A. *Dynamics of Klebsiella pneumoniae OmpA transmembrane domain: the four extracellular loops display restricted motion behavior in micelles and in lipid bilayers*. Biochim Biophys Acta, **2012**, 1818(9): 2344-2353.

Ishida H, Garcia-Herrero A and Vogel H J. *The periplasmic domain of Escherichia coli outer membrane protein A can undergo a localized temperature dependent structural transition*. Biochim Biophys Acta, **2014**, 1838(12): 3014-3024.

Juanhuix J, Gil-Ortiz F, Cuni G, Colldelram C, Nicolas J, Lidon J, Boter E, Ruget C, Ferrer S and Benach J. *Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the Alba Synchrotron*. Journal of Synchrotron Radiation, **2014**, 21(4): 679-689.

Kabsch W. *Xds*. Acta Crystallogr D Biol Crystallogr, **2010**, 66(Pt 2): 125-132.

Kabsch W and Sander C. *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features*. Biopolymers, **1983**, 22(12): 2577-2637.

Katoh K, Misawa K, Kuma K and Miyata T. *MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform*. Nucleic Acids Res, **2002**, 30(14): 3059-3066.

Kim D E, Chivian D and Baker D. *Protein structure prediction and analysis using the Robetta server*. Nucleic Acids Res, **2004**, 32(Web Server issue): W526-531.

Kojima S, Imada K, Sakuma M, Sudo Y, Kojima C, Minamino T, Homma M and Namba K. *Stator assembly and activation mechanism of the flagellar motor by the periplasmic region of MotB*. Mol Microbiol, **2009**, 73(4): 710-718.

Lee D, Hilty C, Wider G and Wuthrich K. *Effective rotational correlation times of proteins from NMR relaxation interference*. J Magn Reson, **2006**, 178(1): 72-76.

Lipari G and Szabo A. *Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity*. Journal of the American Chemical Society, **1982a**, 104(17): 4546-4559.

Lipari G and Szabo A. *Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results*. Journal of the American Chemical Society, **1982b**, 104(17): 4559-4570.

Marcoux J, Politis A, Rinehart D, Marshall D P, Wallace M I, Tamm L K and Robinson C V. *Mass spectrometry defines the C-terminal dimerization domain and enables modeling of the structure of full-length OmpA*. Structure, **2014**, 22(5): 781-790.

McCoy A J, Grosse-Kunstleve R W, Adams P D, Winn M D, Storoni L C and Read R J. *Phaser crystallographic software*. J Appl Crystallogr, **2007**, 40(Pt 4): 658-674.

Murshudov G N, Skubak P, Lebedev A A, Pannu N S, Steiner R A, Nicholls R A, Winn M D, Long F and Vagin A A. *REFMAC5 for the refinement of macromolecular crystal structures*. Acta Crystallogr D Biol Crystallogr, **2011**, 67(Pt 4): 355-367.

Park J S, Lee W C, Yeo K J, Ryu K S, Kumarasiri M, Hesek D, Lee M, Mobashery S, Song J H, Kim S I, Lee J C, Cheong C, Jeon Y H and Kim H Y. *Mechanism of anchoring of OmpA protein to the cell wall peptidoglycan of the gram-negative bacterial outer membrane*. FASEB J, **2012**, 26(1): 219-228.

Parsons L M, Lin F and Orban J. *Peptidoglycan recognition by Pal, an outer membrane lipoprotein*. Biochemistry, **2006**, 45(7): 2122-2128.

Potterton E, Briggs P, Turkenburg M and Dodson E. *A graphical user interface to the CCP4 program suite*. Acta Crystallogr D Biol Crystallogr, **2003**, 59(Pt 7): 1131-1137.

Reddy T and Rainey J K. *Interpretation of biomolecular NMR spin relaxation parameters*. Biochem Cell Biol, **2010**, 88(2): 131-142.

Renault M, Saurel O, Czaplicki J, Demange P, Gervais V, Lohr F, Reat V, Piotto M and Milon A. *Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications*. J Mol Biol, **2009**, 385(1): 117-130.

Robert X and Gouet P. *Deciphering key features in protein structures with the new ENDscript server*. Nucleic Acids Res, **2014**, 42(Web Server issue): W320-324.

Rossi P, Swapna G V, Huang Y J, Aramini J M, Anklin C, Conover K, Hamilton K, Xiao R, Acton T B, Ertekin A, Everett J K and Montelione G T. *A microscale protein NMR sample screening pipeline*. J Biomol NMR, **2010**, 46(1): 11-22.

Roujeinikova A. *Crystal structure of the cell wall anchor domain of MotB, a stator component of the bacterial flagellar motor: implications for peptidoglycan recognition*. Proc Natl Acad Sci U S A, **2008**, 105(30): 10348-10353.

Saitou N and Nei M. *The neighbor-joining method: a new method for reconstructing phylogenetic trees*. Mol Biol Evol, **1987**, 4(4): 406-425.

Shen Y and Bax A. *Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks*. J Biomol NMR, **2013**, 56(3): 227-241.

Simons K T, Kooperberg C, Huang E and Baker D. *Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions*. J Mol Biol, **1997**, 268(1): 209-225.

Song T, Mika F, Lindmark B, Liu Z, Schild S, Bishop A, Zhu J, Camilli A, Johansson J, Vogel J and Wai S N. *A new Vibrio cholerae sRNA modulates colonization and affects release of outer membrane vesicles*. Mol Microbiol, **2008**, 70(1): 100-111.

Tamura K, Stecher G, Peterson D, Filipinski A and Kumar S. *MEGA6: Molecular Evolutionary Genetics Analysis version 6.0*. Mol Biol Evol, **2013**, 30(12): 2725-2729.

Touw W G, Baakman C, Black J, te Beek T A, Krieger E, Joosten R P and Vriend G. *A series of PDB-related databanks for everyday needs*. Nucleic Acids Res, **2015**, 43(Database issue): D364-368.

Vollmer W, Blanot D and de Pedro M A. *Peptidoglycan structure and architecture*. FEMS Microbiol Rev, **2008**, 32(2): 149-167.

Vranken W F, Boucher W, Stevens T J, Fogh R H, Pajon A, Llinas M, Ulrich E L, Markley J L, Ionides J and Laue E D. *The CCPN data model for NMR spectroscopy: development of a software pipeline*. Proteins, **2005**, 59(4): 687-696.

Yang Y, Auguin D, Delbecq S, Dumas E, Molle G, Molle V, Roumestand C and Saint N. *Structure of the Mycobacterium tuberculosis OmpATb protein: a model of an oligomeric channel in the mycobacterial cell wall*. Proteins, **2011**, 79(2): 645-661.

Zheng C, Yang L, Hoopmann M R, Eng J K, Tang X, Weisbrod C R and Bruce J E. *Cross-linking measurements of in vivo protein complex topologies*. Mol Cell Proteomics, **2011**, 10(10): M110 006841.

Zuckerkandl E and Pauling L. *Evolutionary divergence and convergence in proteins*. Evolving genes and proteins, **1965**, 97: 97-166.

Chapitre 4

**Etude de la dynamique du domaine N-terminal de OmpA
de *Klebsiella pneumoniae* par RMN du solide et
protéolyse ménagée.**

Résumé

La dynamique du domaine N-terminal de KpOmpA (N-KpOmpA) a été explorée en bicouche lipidique par une approche combinée de techniques biophysique et biochimique. Ce domaine adopte une structuration en tonneau β transmembranaire présentant 4 boucles sur son versant extracellulaire. L'étude par RMN du solide avec rotation à l'angle magique à très haute vitesse (60 kHz) a permis d'enregistrer une série de spectres 2D-3D avec détection des protons sur des protéoliposomes de protéine triplement marquée [$^1\text{H}^N, ^2\text{H}, ^{13}\text{C}, ^{15}\text{N}$]. 70% des résidus constituant le tonneau β ont ainsi pu être attribués. Toutefois, peu de signaux ont été observés au niveau des boucles, que ce soit dans des expériences de transfert de polarisation basées sur les couplages dipolaires (parties rigides) ou des couplages scalaires (parties très dynamiques « liquid like »). Ceci confirme à nouveau la présence de mouvements μs – s dans ces régions. L'étude par protéolyse à la trypsine a montré que les sites du tonneau β étaient comme attendu, bien protégés de la digestion lors de l'insertion dans bicouche lipidique. De façon plus remarquable, sur les 7 sites présents dans les boucles extracellulaires, un seul site présent dans la boucle L3 est sensible au clivage tryptique, et ce quelles que soient les conditions expérimentales (protéine reconstituée en micelles ou bicouches lipidiques, domaine N-terminal seul ou protéine entière, durée ou température de réaction). Ces résultats soulignent la dynamique particulière de la boucle L3. Cette boucle est par ailleurs plus longue que les 3 autres, et plus longue chez *Klebsiella pneumoniae* que dans d'autres entérobactéries non pathogènes. Ces résultats suggèrent que cette boucle flexible pourrait être un épitope privilégié impliqué dans la reconnaissance de KpOmpA (et donc de la bactérie) par les cellules du système immunitaire. Des expériences visant à confirmer cette hypothèse ont été engagées avec l'équipe du Professeur J. Bengoechea en Irlande.

Ultrafast MAS NMR and limited proteolysis of an integral membrane protein, KpOmpA, reveal distinct dynamics of its barrel and loop 3 domains.

Abstract

The Outer membrane proteins A of *enterobacteriaceae* play a pivotal role on their pathogenicity via their direct implication in cell-cell recognition (for adhesion and immune response mechanisms). This activity is intimately linked to the OmpA structure and dynamics. ¹H detected Ultrafast MAS NMR allowed partial assignment and site specific relaxation analysis of KpOmpA β-barrel membrane domain reconstituted in liposomes. Proteolysis experiments revealed the existence of a unique trypsin cleavage site within loop L3 over 14 potential Lys and Arg cleavable residues distributed throughout the N-terminal membrane domain. This was observed both for full-length KpOmpA and its membrane domain, in micelles as well as in lipid bilayers. By combining advanced NMR and biochemical approaches we found a unique feature of OmpA loop 3 dynamics and we propose that its length and accessibility to protein – protein interactions play a pivotal role in pathogenicity.

Keywords

Extracellular loop

Klebsiella pneumoniae

MALDI-TOF mass spectrometry

Order parameter

Relaxation times

I. Introduction

The outer membrane protein A from *Klebsiella pneumoniae* (KpOmpA) belongs to the OmpA family of proteins with homologues found in many Gram-negative bacteria (Confer *et al.* 2013). It is a monomeric, two-domain protein residing in the cell envelope. The transmembrane domain is a β -barrel structure (Renault *et al.* 2009) composed of 8 β strands, 4 long extracellular loops (L1-4) and 3 short periplasmic turns (T1-3). A flexible hinge region links the barrel with the soluble C-terminal domain of the molecule, which is responsible for binding to the peptidoglycan network in the bacterial periplasm. KpOmpA contributes to pathogenicity via its implication in cell-cell adherence, its pro-inflammatory role and its implication in innate and adaptive immunity. Surface exposure and high copy number make OmpA-like proteins good candidates for vaccine development (Jeannin *et al.* 2005, Kurupati *et al.* 2011).

Previously we have investigated the dynamic behavior of different N-KpOmpA regions in detergent micelles (Renault *et al.* 2009) and in lipid bilayers (Iordanov *et al.* 2012) and concluded that a major portion of the extracellular moiety of the protein exhibits motions at ms to μ s timescales. Interestingly, the extracellular loops L2 and L4 are well conserved between KpOmpA and *Escherichia coli* OmpA, while L1 and L3 are more variable and it is thus important to characterize the protein dynamics at a site-specific level.

Recent progresses in solid state NMR of protein, in particular using very high spinning frequency (~ 60 kHz) have opened the way for residue specific analyses of relaxation rates and thus residue specific analyses of protein motions. Here we applied these methodologies and could measure for the first time a residue specific order parameter profile of NH bonds within the entire β -barrel domain.

In an orthogonal approach we applied trypsin proteolysis and MALDI-TOF mass spectrometry to assess loop flexibility. Limited proteolysis is a widely used tool in structural biology of membrane proteins, to isolate protein domains suitable for X-ray crystallography or to facilitate the assignment of protein NMR spectra, particularly for β -barrel membrane proteins (Fox *et al.* 2013, Columbus 2015). It has long been recognized that loop mobility and accessibility is associated with efficient proteolysis (Bender *et al.* 1965, Fontana *et al.* 1986). A very elegant application of proteolysis to membrane proteins was used by D. Engelman and collaborators to establish the two stage folding model of bacteriorhodopsin (Huang *et al.* 1981, Popot *et al.* 1986). Although KpOmpA membrane domain contains 14 potential cleavage sites for trypsin, a very specific cleavage site within loop 3 could be observed, for the full length protein as well as for the N-terminal domain, in detergent micelles as well as in lipid bilayers.

II. Results and Discussion

The reconstitution and characterization of the transmembrane domain of KpOmpA into liposomes of *Escherichia coli* polar extract, at a lipid-to-protein ratio LPR = 0.5 w:w, was reported by our group in (Iordanov *et al.* 2012). Liposomes integrity and protein folding were controlled by electron microscopy, by SDS PAGE to check the “heat modifiability” properties of Omp proteins (Supplementary Figure S1A) and finally by ^{13}C solid state NMR, since ^{13}C chemical shift deviation of CO, C α and C β resonances are dependent of the secondary structure elements. Ser and Thr cross peaks on a 2D PDSD ^{13}C - ^{13}C correlation form two well separated groups, one for residues within the β -barrel and one for residues in the loops. By analyzing the ^{13}C relaxation properties of these cross peaks (T_1 , $T_{1\rho}$ and CP build up rates) we could show that the loops, particularly loops L1 and L3 were subject to μs to ms motion (Iordanov *et al.* 2012). The next challenge was to provide a description of the protein motion at a residue specific level.

2.1. Ultrafast solid state NMR:

Recent methodological developments enables proton detected ssNMR using fast spinning frequency. Due to the increased sensitivity, a complete set of 3D experiments was developed that allow sequential assignment and site specific relaxation analyses (Barbet-Massin *et al.* 2014). To date, most of solid state proton-detected studies were performed on microcrystalline proteins for sensitivity and resolution issues, therefore analyzing a membrane protein of more than 200 residues like KpOmpA is still challenging. The ^{15}N - ^1H correlation spectra of KpOmpA in *E. coli* lipid proteoliposomes are illustrated in Figure 1 using either scalar or dipolar based polarization transfer. The scalar-based ^{15}N - ^1H correlation displays only few peaks, mostly corresponding to the alanine residues of the highly mobile C-terminus. In contrast, the dipolar-based ^{15}N - ^1H correlation spectrum looks like the solution HSQC spectrum, the peaks are essentially spread from 8 to 10 ppm in the ^1H -detected dimension, a region that corresponds mainly to the β -barrel residues.

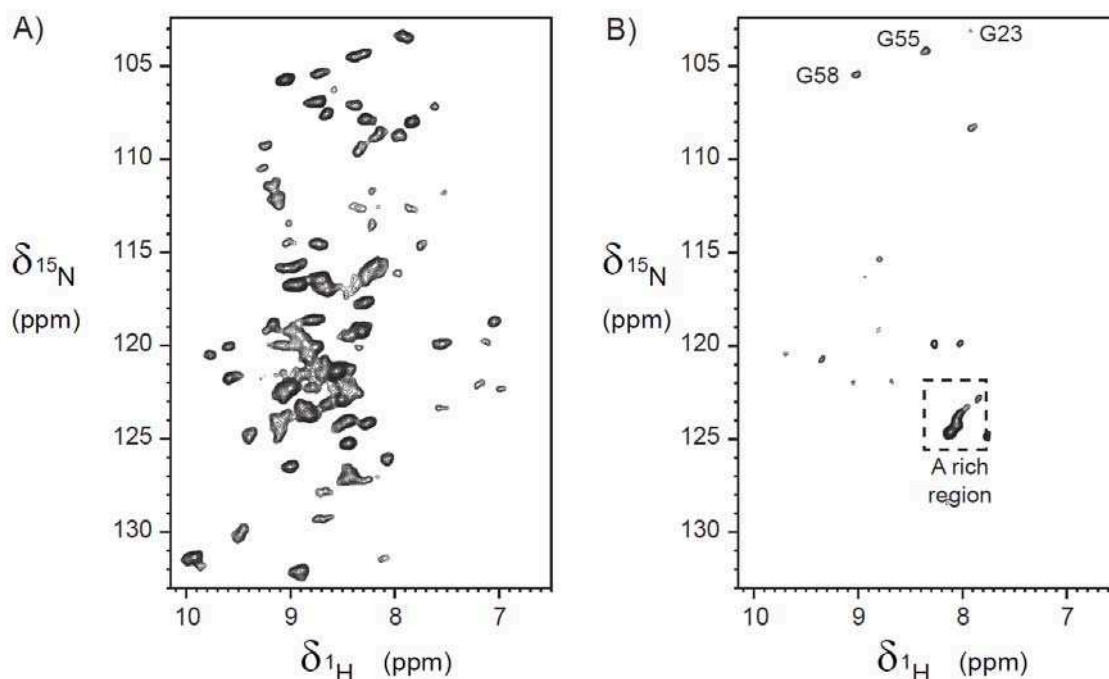


Figure 1. 2D proton-detected ^1H - ^{15}N correlation under fast-MAS of the transmembrane domain of KpOmpA reconstituted in liposome of *E. coli* polar extract (Spectra were recorded at 1 GHz spectrometer under 60 kHz MAS frequency and 25°C). A) CP-based experiment B) J-based polarization transfer experiment. Despite few weak peaks (including Gly 23, 55 and 58), the most intense peaks (dotted box) correspond to an alanine rich region and by comparison with solution state assignment can be assigned to the C-terminus which is a highly mobile and rich in alanine residues (about 30% of the residues). The proton linewidth on the CP-based ^1H - ^{15}N correlation is about 170 Hz (i.e. 0.17 ppm), determined on the 2D spectrum from five resolved peaks while the ^1H T2' determined by the integration of all amide peaks gave 0.9 ms and thus an average linewidth of 350 Hz. This is in line with the spectrum where intense peaks described above coexist with weaker and broader peaks.

In order to assign KpOmpA resonances, we performed the following set of 3D experiment, (H)CANH, (H)CONH and (H)(CA)CB(CACO)NH that connect $\text{CA}_i\text{N}_i\text{H}_i$, $\text{CO}_{i-1}\text{N}_i\text{H}_i$ and $\text{CB}_{i-1}\text{N}_i\text{H}_i$ nuclei respectively (Barbet-Massin *et al.* 2014). Despite its low sensitivity, the last experiment was particularly efficient for the assignment thanks to the intrinsic CB diversity compared to narrow the CO and CA spectral dispersions. Based on (CB_{i-1} , CO_{i-1} , N_i , H_i , CA_i) connectivities, we could assign unambiguously 49 residues by comparison with the known KpOmpA chemical shifts in detergent micelles of DHPC (Figure 2 and Supplementary Table S2) (Renault *et al.* 2009). The 48 assigned residues were distributed uniformly throughout the β -barrel core of the protein (Figure 3).

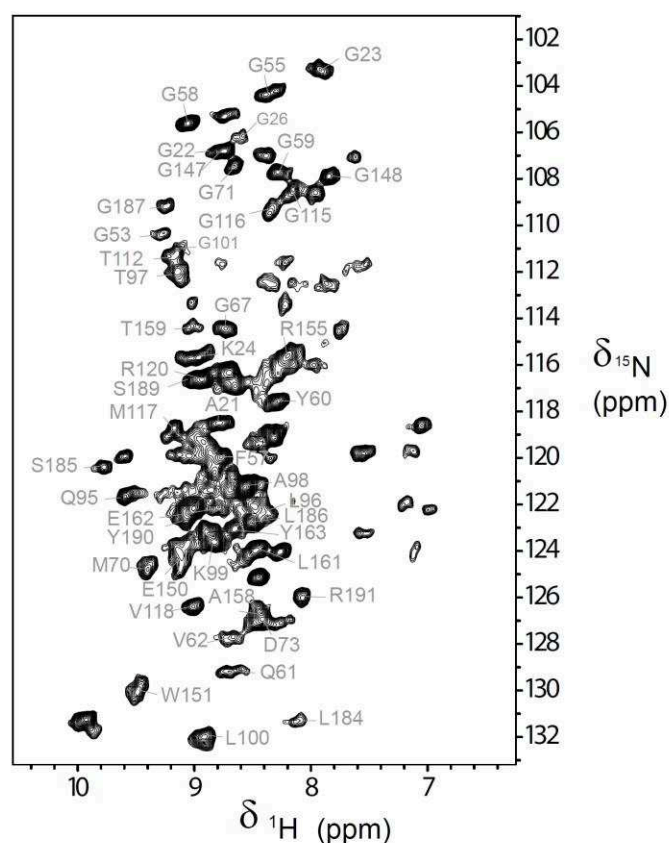


Figure 2. ^{15}N - ^1H correlation spectrum of $[\text{U-}^{15}\text{N}, ^2\text{H}, ^{13}\text{C}]$ -labelled KpOmpA reconstituted in *Escherichia coli* polar extract liposomes, recorded at 1 GHz spectrometer under 60 kHz MAS frequency and 25°C. Assigned residues are annotated.

This feature is in line with the NMR pulse sequence used that required efficient multiple cross polarization steps and therefore select the rigid elements. Accordingly, most of these residues are concentrated in the most rigid part of the β -barrel and only few of them are located in strands extremities known to be more mobile: this is clearly illustrated in the ^{15}N - ^1H spectrum where G26 and G53 are considerably weaker to G22 and G58 in the rigid core of their respective strand.

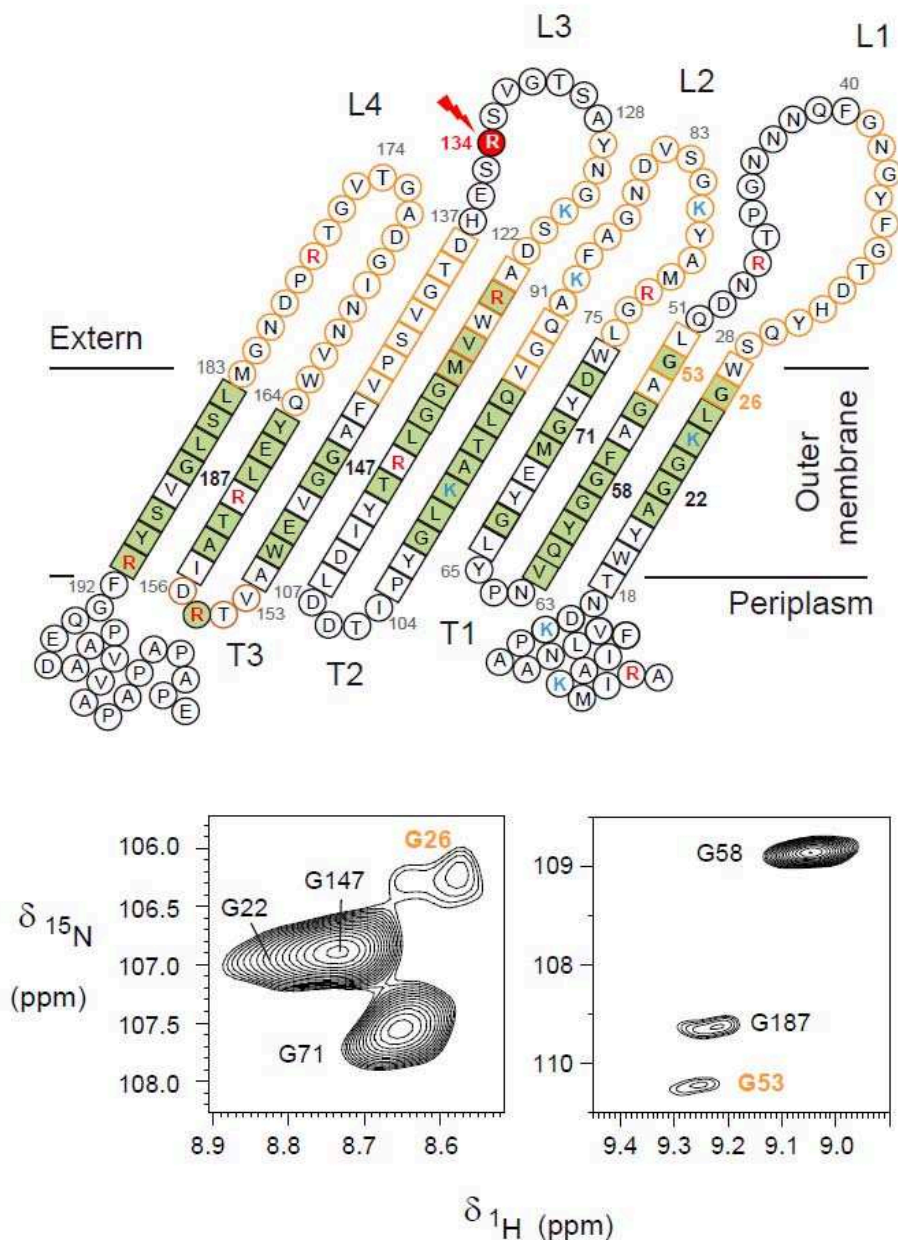


Figure 3. Topological representation of assigned residues in KpOmpA TM domain illustrating experimental evidence illustrating residue-specific mobility within distinct protein segments and Arg and Lys potential cleavage sites for trypsin (highlighted in red and blue respectively). Residues within β -sheet and random coil regions are represented by squares and circles, respectively. Color code: assigned residues are highlighted in green, residues that exhibit restricted mobility from previous studies (i.e., intermediate time scale motions) are colored in orange, red and blue color respectively for Arg and Lys residues and the unique cleavage site (i.e., R134) accessible to Trypsin is highlighted in red.

All the residues assigned belong to the β -barrel. Two glycine residues at the strand-to-loop transition (i.e., G26, G53) are weaker compared to G22 and G58 in the same respective strand, illustrating again that this region is prone to a specific dynamic with intermediate time scale motions, like chemical exchange. Most of the residues in the orange region could not be assigned since the magnetization vanished during the multiple magnetization transfer steps of the (H)(CA)CB(CACO)NH experiment, therefore most of the residues assigned belong to the rigid core of the β -barrel. The second halves of loops L1 and L3 could not be assigned, even with J-based polarization transfer experiments, illustrating that in our experimental conditions these regions are no longer in fast isotropic motional regime.

Note: The first three residues ARI result from the cloning strategy, so that the numbering used herein is +3 with respect to the original sequence.

Recent publications demonstrated the proton-detected solid state NMR under fast MAS spinning as an emerging methodology to give a quantitative description of the backbone dynamics (e.g, nano-seconds to microseconds scale motions) on microcrystalline proteins, leading to a site-specific order parameter of motion and associated time scale from various sources of relaxation: R_1 , $R_{1\rho}$, CSA-dipolar cross correlated relaxation and dipolar coupling (Schanda *et al.* 2010, Lewandowski *et al.* 2015). We thus recorded R_1 and $R_{1\rho}$ of ^{15}N relaxation rates and amide bond dipolar coupling on every assigned peak (Figure 4).

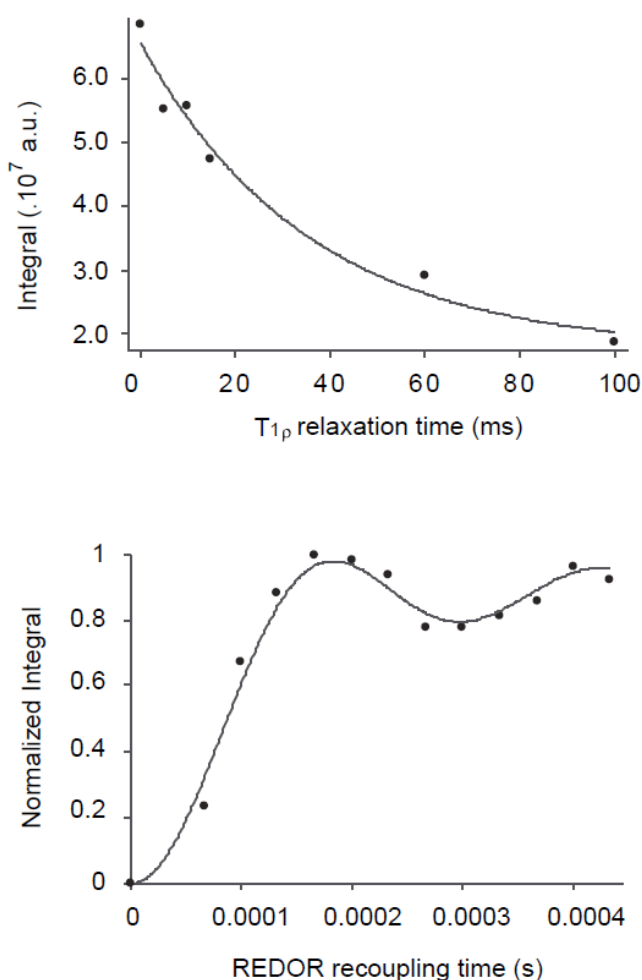


Figure 4. Illustration of the determination of $T_{1\rho}$ and dipolar coupling values for G23 residue. $T_{1\rho}$ value of 49.2 ± 11 ms and dipolar coupling value of 9155 ± 310 Hz were determined by fitting of the relaxation decay by a mono-exponential (upper panel) and REDOR dephasing curve (bottom panel) by equation 1, using a simulated annealing-based global optimization program, namely GOSA (Czaplicki *et al.* 2006).

Particularly informative is the order parameter profile along the β -barrel (Figure 5). It appears to be fairly homogeneous, every residues displaying values above 0.8 (average

value: 0.85 ± 0.05). Figure 5 also shows the expected order parameter profile under the assumption of fast axial diffusion along the bilayer normal. Due to different orientations of NH vector with respect to the diffusion axis, a $((3\cos^2\beta-1)/2)$ dependency should be observed in the profile, several NH residues giving very small order parameters. This result shows the absence of fast axial diffusion under our experimental conditions and explains the uniform linewidth observed in the spectra.

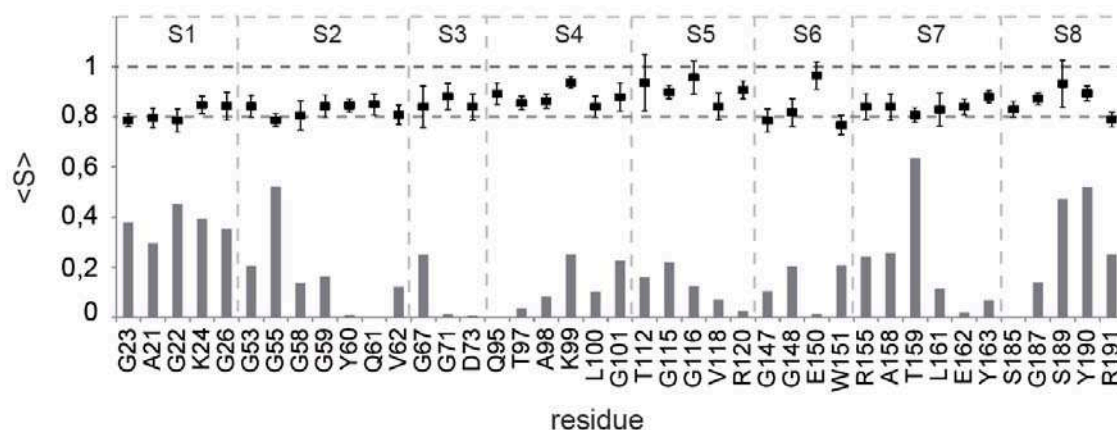


Figure 5. Histogram representation of amide bond order parameter (black box) per assigned residue (data in Supplementary Table S1). Grey bars represent the value of $|(\cos^2\beta-1)/2|$, where β is the angle the amide bond angle with respect to the long principal axis of the system (i.e. β -barrel), which is assumed to be closed to the principal axis of the rotational diffusion motion (β values were computed from the known 3D structure of KpOmpA, pdb 2K0L).

Time scale of motions could not be extracted at a site specific level since T_1 exhibited too large errors on individual peak, the average value over the barrel are $T_{1p} = 36 \pm 15$ ms and $T_1 = 38 \pm 12$ s (site specific T_{1p} values are given in Supplementary Table S1).

In conclusion, we could obtain H-N correlations within few minutes and record an (H)(CA)CB(CACO)NH 3D experiment that enabled to assign 70% of the barrel rigid core. Residues in the loops and part of the β -strands at the strand-loop transition did not show up both in dipolar and scalar-based polarization transfer. This tells us that these regions are neither rigid nor highly mobile but are rather animated by intermediate time scale (i.e., μ s to ms) motions. We could thus describe (for the first time for a membrane protein in its fluid lipid bilayer) an order parameter profile at a residue specific level, but not the complex dynamics present in the extracellular loops.

2.2. Proteolysis experiments:

In a complementary strategy to assess protein flexibility, we applied the long established proteolysis strategy (Bender *et al.* 1965, Fontana *et al.* 1986, Popot *et al.* 1986). KpOmpA contains 18 Lys (7 in N-term and 11 in C-term domains) and 15 Arg (9 in N-term

and 6 in C-term domains), thus potentially 33 cleavage sites for trypsin (16 in the N-term domain, Figure 3). Considering this very large number of cleavage sites we first used Lys-C, a protease specific for Lysine residues. To our surprise, the N-terminal domain proved to be fully resistant to proteolysis, both when solubilized in detergent micelles or reconstituted in lipid bilayers (Supplementary Figure S2 and Supplementary Figure S3), while the C-terminal domain was fully digested on only 50% of the proteins when reconstituted in liposomes due to the two possible orientations of the protein within a bilayer. The proteolysis resistance was observed for up to 12 h at 37°C and at an enzyme to substrate ratio of up to 1%, in Zw3-14 detergent micelles as well as in *E. coli* proteoliposomes at Lipid to protein molar ratio of 15, 60 and 150 and in nine different 20 mM Tris buffer conditions (pH 6, 7 and 8; [NaCl] = 0, 150 and 300 mM).

The three Lys cleavage site in the extracellular loops are located in loops L2 and the beginning of loop L3 (Figure 3), and thus do not report on L1 and L3 loops accessibility; we then used trypsin capable of cleavage at both Lys and Arg residues, thus covering positions in all four loops. Figure 6 shows that heat denaturation of the Trypsin-digested N-KpOmpA (in 0.1% Zw3-14 and in proteoliposomes) resulted in fragments separation from the (otherwise folded) β -barrel. MALDI-TOF analysis (Figure 6B) identified the cleavage products as fragments of sizes 14057 and 8814 Da (in the micellar sample), and 14083 and 8805 Da (in the PLE sample), closely matching the calculated sizes of the fragments generated upon cleavage at Arg134 on loop L3.

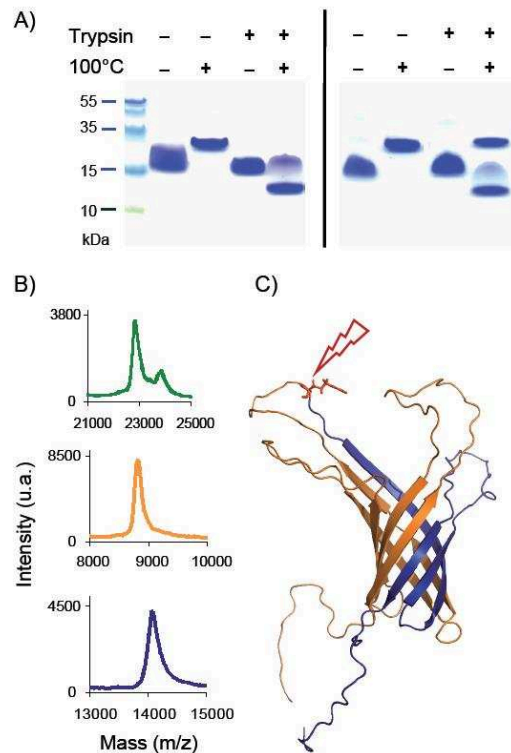


Figure 6. Complete digestion of N-KpOmpA with Trypsin: (A) SDS PAGE gel of digested N-KpOmpA in *left panel*: 0.1% Zw3-14 (SER = 200:1 w/w, 37°C for 2.5h) or in *right panel*: PLE liposomes (SER = 100:1 w/w, LPR = 60 mol/mol, 37°C for 4h). The presence or absence of Trypsin and of heat denaturation (100°C) of the aliquots are indicated respectively with '+' and '-' above each well. The unfolded intact protein chain and the larger cleavage product (~14 kDa) are indicated with solid and dashed arrows, respectively. (B) Normalized MALDI-TOF linear mode spectra of digested N-KpOmpA in in PLE liposomes, showing intact N-terminal domain at ~22.8 kDa (green) and the cleaved fragments at ~8.8 (blue)) and ~14.1 kDa (orange). (C) 3D structures of the N-terminal domains with a color coding corresponding to the above fragments.

Thus, over the 16 potential cleavage sites, a unique reaction occurred at Arg134, whatever the amount of protease or the incubation time tested. This was observed both on the N-Term domain and on the full length protein, in micelles and in lipid bilayers. The transmembrane domain (containing 3 Lys and 5 Arg) was found, as expected, to be protected from proteolysis. More sticking was the result concerning the extracellular loops which contain 7 potential cleavage sites, one Arg in L1, L2, L3 (at the end of it) and L4, 2 Lys in L2 and one Lys at the beginning of L3 (Figure 3).

Considering that proteolytic cleavage is limited by loops motion and accessibility (Bender *et al.* 1965, Fontana *et al.* 1986) this points to the very specific mobility of loop L3, particularly of its C-terminal half (since Lys 124 at the beginning of loop L3 is not cleaved). This loop may be involved in the adherence of the bacterial pathogen to the surface of the target cells. This is usually facilitated by specific interaction between a cell surface protein on the bacterium and the extracellular matrix (ECM) proteins in the target cell, among which fibronectin (Westerlund *et al.* 1993, Confer *et al.* 2013). In order to check whether this peculiar dynamics of the KpOmpA L3 loop has any relationship with pathogenicity we have

aligned several OmpA sequences from pathogenic and non-pathogenic strains of *enterobacteriaceae* (Table 1). *Klebsiella pneumoniae* has particularly long L1 and L3 loops. Most pathogenic strains (*Klebsiella*, *Salmonella*, *Shigella*, *Yersinia*, *Citrobacter*) possess a long L3 loop. Among *E. coli* strains, the non-pathogenic, commensal K12 strain possesses a short L3 loop, while pathogenic *E. coli* strains may have a short L3 loop (EHEC, ExPEC, UPEC, EPEC) or a long L3 loop (EAEC and ETEC) (Rasko *et al.* 2008).

Table 1. Sequence alignment of pathogenic and nonpathogenic *enterobacteriaceae*. The alignment was performed using the Uniprot server (<http://www.uniprot.org/>) A) list of the strains which have been aligned. B) The sequence alignment has been performed on the entire OmpA sequences. Here is presented a focus on the region surrounding loops L3. The positions of the β sheets 4-6, belonging to the transmembrane β -barrel of KpOmpA, are highlighted in yellow. In red is shown the unique Arg forming the trypsin cleavage site within loop L3.

A)

Uniprot entry name	Organism	L3 Loop
OMPA_ECOLI	<i>Escherichia coli</i> (strain K12)	short
D3QPF3_ECOCB	<i>Escherichia coli</i> O55:H7 (strain CB9615 / EPEC)	short
Q1RDQ7_ECOUT	<i>Escherichia coli</i> (strain UT189 / UPEC)	short
OMPA_ENTAE	<i>Enterobacter aerogenes</i>	short
OMPA_KLEPN	<i>Klebsiella pneumoniae</i>	Long
OMPA_SALTI	<i>Salmonella typhi</i>	Long
Q32HV0_SHIDS	<i>Shigella dysenteriae</i> serotype 1 (strain Sd197)	Long
OMPA_YERPE	<i>Yersinia pestis</i>	Long
D2TS87_CITRI	<i>Citrobacter rodentium</i> (strain ICC168)	Long
D3H0H9_ECO44	<i>Escherichia coli</i> O44:H18 (strain 042 / EAEC)	Long

B)

	$\beta 4$	IL2	$\beta 5$	EL3	$\beta 6$
SP P0A910 OMPA_ECOLI	SVENGAYKAQGVQLTAKLGYPITDDLDIYTRLGGMVWRADTKSNVY-G----				KNHDTGVS
TR D3QPF3 D3QPF3_ECOCB	SVENGAYKAQGVQLTAKLGYPITDDLDIYTRLGGMVWRADTKSNVY-G----				KNHDTGVS
TR Q1RDQ7 Q1RDQ7_ECOUT	SVENGAYKAQGVQLTAKLGYPITDDLDVYTRLGGMVWRADTKSNVY-G----				KNHDTGVS
SP P09146 OMPA_ENTAE	VKVNQAFSSQAVQLTAKLGYPITDDLDIYTRLGGMVWRADSSNSIA-G----				DNHDTGVS
SP P24017 OMPA_KLEPN	SVDNGAFKAQGVQLTAKLGYPITDDLDIYTRLGGMVWRADTKSNVY-G----				DNHDTGVS
SP Q8Z7S0 OMPA_SALTI	DNTNGAYKAQGVQLTAKLGYPITDDLDVYTRLGGMVWRADTKSNVP-GGASTKDHDGTGS				
TR Q32HV0 Q32HV0_SHIDS	SVENGAYKAQGVQLTAKLGYPITDDLDIYTRLGGMVWRADTKANVP-GGASFKDHDGTGS				
SP Q8ZG77 OMPA_YERPE	DINNGAFKAQGVQLAAKLSYPVQDLVYTRLGGLVWRADAKGSFDGGLDRASGHDGTGS				
TR D2TS87 D2TS87_CITRI	SVENGAYKAQGVQLTAKLGYPITDDLDIYTRLGGMVWRADTKSHNNLTGASTKDHDGTGS				
TR D3H0H9 D3H0H9_ECO44	DNINGAYKAQGVQLTAKLGYPITDDLDVYTRLGGMVWRADTKSNVP-GGVSTKDHDGTGS				

It has been shown that OmpA from *Mannheimia haemolytica* A1 (a bacteria associated with bovine pneumonic pasteurellosis) is involved in fibronectin binding, and that treatment of the bacteria with trypsin abolishes fibronectin binding (Lo *et al.* 2007). *M. haemolytica* possesses a long L3 loop, with two potential sites for trypsin (VQQHVDKDSRIKV compared to GNYASTGVSRSRSE for *K. pneumoniae* and the shorter NVYGKN for *E. coli*). It is thus tempting to speculate that loop L3 is highly flexible and well presented to the outer surface, contrary to the other extracellular loops, and thus accessible to trypsin. This makes it an important epitope for protein binding and cell adhesion. This hypothesis is currently under investigation in collaboration with Prof J.A. Bengoechea

(Belfast, Ireland) with various strains of *Klebsiella pneumoniae*, and various constructs of KpOmpA heterologously expressed in *E. coli*. We will look for the impact of loop3 on antimicrobial peptide resistance, inflammation (IL8 secretion and activation of signaling pathways) and the impact of loop3 in preventing phagocytosis, as was done in (Llobet *et al.* 2009, March *et al.* 2011).

In the present study, two very different and complementary strategies were applied to characterize the internal dynamics of an integral membrane protein. By applying state of the art ultra-fast MAS NMR methodologies, we could quantify the β -barrel order parameter profile at a residue specific level. It proved to be remarkably homogeneous within the entire barrel. Residues located within the extracellular loops escaped detection in these proton detected experiments, both using cross polarization and J based $^{15}\text{N} - ^1\text{H}$ transfers, confirming that the four loops are subject to motion at intermediate time scale. By contrast, proteolytic experiments revealed a specific feature of the long L3 loop, indicating that its motion is of sufficient amplitude for allowing productive contacts with trypsin, and by extension with other proteins required for cell-cell adhesion mechanisms.

III. Supplementary information

3.1. Construction of the KpOmpA recombinant proteins

The construction of the transmembrane domain of KpOmpA (N-KpOmpA, 210 a.a., ~23.4 kDa) as a separate expressible product in a pET21c vector was described previously (Renault *et al.* 2009).

The full-length polypeptide (F-KpOmpA, containing in addition the soluble C-terminal domain) was PCR-amplified from genomic DNA of *Klebsiella pneumoniae* strain 52145 (kindly provided by Dr. Tournibize, Institut Pasteur, Paris) using forward (GTGGCACTGCATATGAAAGCTACCGTAGCGCAGGCCGCTCCGAAAGATAAC) and reverse (GGTACCAGCTCGAGAGCCGCCGGCTGAGTTACAACCTTCT) primers, and Phusion Hot Start II High-Fidelity DNA Polymerase (Fynnzymes). The underlined sequences represent NdeI and XhoI restriction sites used for subcloning in pET26b vector. The cloned F-KpOmpA gene was sequence-verified (MilleGen).

N-KpOmpA contains a short modification at the N-terminus (ARIMKAIFVLNA) and a hexahistidine tag at the C-terminus, immediately after the flexible hinge region of the molecule that natively links its two domains. F-KpOmpA (359 a.a., ~38.7 kDa) contains a short modification at the N-terminus (MKATVAQA), the full-length polypeptide chain of KpOmpA and a thrombin cleavage site (LVPRGS) followed by a hexahistidine tag at the C-terminus. Both plasmids (pET21c-N-KpOmpA and pET26b-F-KpOmpA) were electrotransferred to *E. coli* BL21 (DE3) strain.

3.2. Protein expression and purification

The expression/purification procedure for the triple isotopically labelled N-terminal domain of the outer membrane protein from *Klebsiella pneumoniae*, [U-¹³C,¹⁵N,²H] N-KpOmpA, was described previously (Renault *et al.* 2009, Iordanov *et al.* 2012). Both the N- and F-KpOmpA recombinant proteins lack the native N-terminal signal peptide for outer membrane localization and therefore accumulate as inclusion bodies in the cytoplasm. The expression/purification procedure for N-KpOmpA was described previously (Renault *et al.* 2009) and the one used for F-KpOmpA did not vary significantly. Briefly, exponentially growing *E. coli* BL21 (DE3) cells harboring the pET26b-F-KpOmpA plasmid were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside for 4 hours at 37°C and then collected. Upon cell lysis, the released inclusion bodies were washed and dissolved in 6 M Guanidine/HCl buffer containing 25 mM Tris pH 8.5 and 5 mM EDTA. The denatured protein was refolded in 15x excess volume of 1% w/v Zwittergent 3-14 (n-tetradecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, Anatrace) buffer containing 25 mM Tris pH 8.5 and

150 mM NaCl. The detergent concentration was then reduced to 0.1% Zw3-14 via dialysis (12-14 kDa cutoff, Spectra/Por) and the protein was purified by affinity chromatography (Ni-NTA Superflow, QIAGEN) with elution step of 400 mM imidazole. The imidazole was removed by a second dialysis and the final protein stock (in 0.1% Zw3-14, 25 mM Tris pH 8.5, 150 mM NaCl) was stored at 4°C. The concentration of the purified proteins was determined spectrophotometrically by measuring the absorbance at 280 nm and using the theoretical molar extinction coefficients of 50880 M⁻¹cm⁻¹ and 56140 M⁻¹cm⁻¹ for N- and F-KpOmpA, respectively. The purity and successful refolding of the proteins was monitored with SDS PAGE (Nakamura *et al.* 1976b).

3.3. Proteoliposomes preparation

Proteoliposomes containing N-KpOmpA were prepared as described previously (Iordanov *et al.* 2012), using detergent dilution method adapted from procedures described by Rigaud *et al.* (Rigaud *et al.* 1988). Briefly, *E. coli* Polar Lipids Extract (PLE, Avanti Polar Lipids) was dissolved in chloroform, dried under nitrogen stream and then in a vacuum chamber overnight. This lipid film was dissolved at 10 mg/ml in 2% n-octyl-β-d-glucopyranoside (Carbosynth), 20 mM Tris pH 8.5, 100 mM NaCl and sonicated in a water bath several times. Purified N- or F-KpOmpA was bound to nickel-chelating resin, extensively washed with 2% n-octyl-β-d-glucopyranoside solution containing 20 mM Tris pH 8.5 and 100 mM NaCl and eluted with the same buffer supplemented with 400 mM imidazole. Mixed micelles were formed by adding the *E. coli* Polar Lipids Extract solution to the protein eluate at the desired lipid-to-protein ratio (LPR), ranging from 0.5 to 5 w/w (equivalent to ~15-150:1 mol/mol) in the different experiments. This mixture was incubated for 10 min at 4°C under gentle mixing and then dialyzed at 37°C against detergent-free buffer, until complete removal of the detergent and formation of large unilamellar vesicles. During the dialysis, the NaCl was gradually removed in 25 mM steps. The final proteoliposomes sample is thus found in only 20 mM Tris buffer pH 8.5.

3.4. Proteolysis experiments

The commercially available proteases ‘Endoproteinase Lysine-C sequencing grade’ (Lys-C, Roche) and ‘Sequencing Grade Modified Trypsin’ (Promega) were supplied as lyophilized powders and dissolved according to the manufacturer’s instructions, resulting in 0.1 mg/ml stocks for each enzyme.

Digestion experiments with Lys-C: for proteolysis experiments in ‘complete digestion’ conditions, 50-450 µg of N-KpOmpA (1-3 mg/ml in the 0.1% Zw3-14 stock or in the proteoliposomes in 20 mM Tris) were directly mixed with 10 µl protease stock (1 µg enzyme)

to the desired substrate-to-enzyme ratio (SER, ranging from 50:1 to 450:1 w/w). The volumes ratio of this mixture was at least 10:1 in favor of the KpOmpA-buffer, in order to ensure that the reaction conditions are determined by the KpOmpA-buffer. The reaction was incubated at 37°C for several hours or overnight and then aliquots were taken for SDS PAGE and MALDI-TOF. The aliquot for mass spectrometry was added to equal volume of 0.1% trifluoroacetic acid (TFA) and 50% acetonitrile (AcN) in 1:1 v/v ratio. If the aliquot material was considered too diluted, it was dried under vacuum (miVac Duo Concentrator, GeneVac) and resuspended in 10 µl of the same TFA/AcN solution. For monitoring the digestion kinetics of the protein, N- or F-KpOmpA (in micellar or bilayer environment) was treated in the same way, except that the reaction temperature was kept at 16°C for the micellar sample and at 37°C for the reconstituted protein sample. At certain time-points (0-120 min), aliquots from the reaction mixture were collected and immediately added either to the SDS PAGE loading dye and boiled for 5 min, or to equal volume of TFA/AcN solution and frozen at -20°C until analyzed by mass spectrometry.

3.5. Mass spectrometry

1 µl of each aliquot (1-2 µg KpOmpA) was deposited on a MALDI-plate and mixed with 1 µl of matrix solution (either alpha-cyano-4-hydroxycinnamic acid (CHCA, Fluka Analytical) or 3,5-dimethoxy-4-hydroxycinnamic acid (SPA), as indicated in text). Upon drying at room temperature, the plate was loaded into the vacuum chamber of the spectrometer. All mass spectrometry measurements were performed on a Voyager DE-STR (Applied Biosystems) MALDI-TOF spectrometer, using Voyager Instrument Control Panel software for data acquisition and Data Explorer software for data processing and analysis. The spectra were acquired either in linear (1-50 kDa range, 25 kV accelerating voltage, 300 ns delay time) or in reflectron (0.8-4 kDa range, 20 kV accelerating voltage, 240 ns delay time) mode, with positive ion detection and 1000 shots per spectrum in each case. Spectral calibration was done with commercial compounds of known sizes.

3.6. NMR spectroscopy

All experiments were carried out on Bruker Avance III 1000 MHz and 800 MHz standard bore spectrometers (operating at static fields of 23.4 and 18.7 T, respectively), both equipped with a 1.3 mm triple tuned (^1H , ^{13}C , ^{15}N) CP-MAS probe, at $\omega_r/2\pi = 60$ kHz sample spinning rate. Chemical shifts were referenced to DSS. The sample temperature was 25°C, as indicated by the chemical shift of supernatant water.

3.6.1. NMR experiments for KpOmpA assignment

Proton detected NMR experiments, 2D CP and J based HN correlation, 3D (H)CANH, (H)CONH and (H)(CA)CO(CACO)NH were reported in (Barbet-Massin *et al.* 2014). In all these experiments, WALTZ-16 decoupling (Shaka *et al.* 1983) was used for ^{15}N and ^{13}C during ^1H acquisition, with RF frequencies of 10 kHz on both channels (90° pulse-length of 25 μs). Swept low-power TPPM (sITPPM) decoupling was used during ^{15}N chemical shift evolution (Lewandowski *et al.* 2010). The ^1H RF frequency was 15 kHz according to flip pulse calibration (optimized on echo sequence), with a pulse-length ramped by 40% around a reference 180° duration of 33.3 μs , and a phase difference of 41° . Supercycled CW irradiation (Zhou *et al.* 2008) was employed to suppress water during proton acquisition. CW was applied on-resonance with the water signal during 4 blocks of 75 ms, with alternating X and Y phases, while sample magnetization was stored as longitudinal z polarization on ^{15}N . This step always followed the indirect acquisition of the ^{15}N dimension.

In the 2D HN correlation, the cross-polarization (CP) from ^1H to ^{15}N was performed using a constant RF frequency applied to ^{15}N of 35 kHz and a linearly ramped pulse from 65% to 135% of a maximum RF frequency of 30 kHz on ^1H , for 1500 μs . For CP from ^{15}N to ^1H , the same RF frequencies were used, but for a time of 600 μs and a ramped pulse from 135% to 35% of the ^1H RF frequency. The experiment was acquired with 1.5k and 168 complex points in the ^1H and ^{15}N dimensions respectively (acquisition times of 25 ms and 23 ms respectively in the ^1H and ^{15}N dimensions), with an interscan recycle delay of 1 s and with 64 transients per increment. The States-TPPI method was employed for quadrature detection in the indirect dimension (Marion *et al.* 1989).

In the (H)CONH, CP from ^1H to ^{13}C was performed using a constant RF frequency of 45 kHz on ^{13}C and a linearly ramped pulse from 95% to 105% of a maximum RF frequency of 100 kHz on ^1H , for 3000 μs . CP from ^{13}C to ^{15}N was performed using a constant RF frequency of 29 kHz on ^{13}C and a tangentially ramped pulse from 90% to 110% of a 35 kHz RF pulse on ^{15}N , (Hediger *et al.* 1994) for 12000 μs . The ^{13}C carrier was set at 176 ppm. Back CP from ^{15}N to ^1H was performed using a constant RF frequency of 35 kHz on ^{15}N and a linearly ramped pulse from 135% to 65% of a 30 kHz RF pulse on ^{15}N , for 600 μs . The experiment was acquired with 64 and 40 complex points respectively in the ^{13}C and ^{15}N indirect dimensions (indirect acquisition times of 9.6 ms and 5 ms in the ^{13}C and ^{15}N dimensions respectively, and a direct acquisition time of 14.9 ms in the ^1H dimension), with 24 transients per increment and with an interscan recycle delay of 1 s.

In the (H)CANH, CP from ^1H to ^{13}C was performed using a constant RF frequency of 45 kHz on ^{13}C and a linearly ramped pulse from 95% to 105% of a maximum RF frequency of 100 kHz on ^1H , for 3000 μs . CP from ^{13}C to ^{15}N was performed using a constant RF frequency of 30 kHz on ^{13}C and a tangentially ramped pulse from 90% to 110% of a 35 kHz

RF pulse on ^{15}N , (Hediger *et al.* 1994) for 12000 μs . The ^{13}C carrier was set at 55 ppm. Back CP from ^{15}N to ^1H was performed using a constant RF frequency of 35 kHz on ^{15}N and a linearly ramped pulse from 135% to 65% of a 30 kHz RF pulse on ^{15}N , for 600 μs . The experiment was acquired with 128 and 40 complex points respectively in the ^{13}C and ^{15}N indirect dimensions (indirect acquisition times of 6.4 ms and 5 ms in the ^{13}C and ^{15}N dimensions respectively, and a direct acquisition time of 14.9 ms in the ^1H dimension), with 16 transients per increment and with an interscan recycle delay of 1 s.

In the (H)(CA)CB(CACO)NH experiment reported in (Barbet-Massin *et al.* 2014), CP from ^1H to ^{13}C was performed using a constant RF frequency of 40 kHz on ^{13}C and a linearly ramped pulse from 97% to 102% of a maximum RF frequency of 100 kHz on ^1H , for 4000 μs . Homonuclear INEPT blocks (Knight *et al.* 2011) were inserted between ^1H - ^{13}C and ^{13}C - ^{15}N CP in order to transfer coherences between CA-CB spins and CA-CO spins. The CA to CB transfers were achieved with a τ delay of 7.2 ms and Caliph band selective π -pulses using a Gaussian cascade pulses with a length of 150 μs . The ^{13}C Caliph carrier was set at 40 ppm. The CA to CO transfer was achieved with a τ delay of 4 ms and with CA and CO band selective π -pulses with lengths of 600 μs and 350 μs respectively (Q3 Gaussian cascade pulse). The ^{13}C carriers were set for CA and CO at 55 ppm and 176 ppm respectively. CP from ^{13}C to ^{15}N was performed using a constant RF frequency of 31 kHz on ^{13}C and a tangentially ramped pulse from 90% to 110% of a 24 kHz RF pulse on ^{15}N for 10000 μs . Back CP from ^{15}N to ^1H was performed using a constant RF frequency of 10 kHz on ^{15}N and a linearly ramped pulse from 97% to 102% of a 100 kHz RF pulse on ^{15}N , for 450 μs . The experiment was acquired with 80 and 32 complex points respectively in the ^{13}C and ^{15}N indirect dimensions (indirect acquisition times of 2 ms and 4 ms in the ^{13}C and ^{15}N dimensions respectively, and a direct acquisition time of 14.9 ms in the ^1H dimension), with 128 transients per increment and with an interscan recycle delay of 1 s.

Spectra were processed with Topspin 3.1 and the assignment was performed using CARA (Keller 2005) based on the assignment previously performed by solution state NMR (BMRB Entry 15651) (Renault *et al.* 2009). The assignment of the transmembrane domain of KpOmpA reconstituted in liposome is reported in Supplementary Table S2.

3.6.2. Relaxation and REDOR NMR experiments

^{15}N $T_{1\rho}$ experiments were performed by the insertion of a ^{15}N spin-lock pulse immediately after the ^1H to ^{15}N CP with a RF frequency of 20 kHz and with pulse durations of 0.5, 10, 15, 20, 40, 60 and 100 ms. ^{15}N T_1 experiments were performed by the insertion of a T_1 relaxation delay between the 90° ^{15}N pulse and the 4 blocks while sample magnetization was stored as longitudinal z polarization on ^{15}N , T_1 relaxation delays were of 1, 2, 2.8, 5, 18 and 36 s. Averaged dipolar coupling were measured from a REDOR

experiment on proton-detected ^1H - ^{15}N correlation. A REDOR block (Schanda *et al.* 2010) was inserted after t_1 evolution, basically this block consists of ^1H π -pulses at integer of the rotor period, leading to a rapid dephasing due to heteronuclear dipolar recoupling. 12 spectra were acquired with the following dephasing times 67, 100, 133, 167, 200, 233, 267, 300, 333, 367, 400 and 433 μs . For $T_{1\rho}$, T_1 and REDOR experiments, the cross-polarization (CP) from ^1H to ^{15}N was performed using a constant RF frequency applied to ^{15}N of 40 kHz and a linearly ramped pulse from 65% to 135% of a maximum RF frequency of 23 kHz on ^1H , for 1000 μs . For CP from ^{15}N to ^1H , the same RF frequencies were used, but for a time of 600 μs and a ramped pulse from 135% to 35% of the ^1H RF frequency. The experiment was acquired with 6k and 168 complex points in the ^1H and ^{15}N dimensions respectively (acquisition times of 30 ms and 17.9 ms respectively in the ^1H and ^{15}N dimensions), with an interscan recycle delay of 1 s and with 96 transients per increment. Peak integrals were plotted as a function of the relaxation delay and fitted using a mono-exponential decay function to determine T_1 and $T_{1\rho}$ relaxation time.

For each assigned peak, normalized REDOR $\Delta S/S_0$ curve as function of dipolar evolution time τ ($\tau = N \cdot t_r$) were generated (see Figure 3 for G23 residue), and average dipolar coupling (d) expressed in Hz was fitted according to Equation 1 (Gullion *et al.* 1989, Gullion 2006),

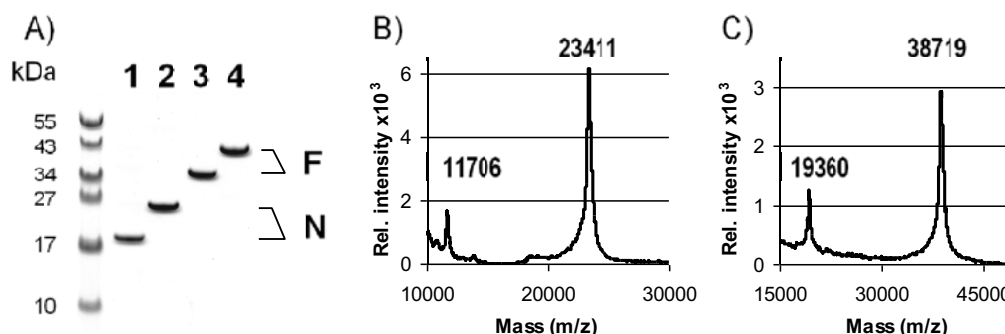
$$\frac{\Delta S}{S_0} = \frac{I_{0(Nt_r)} - I_{(Nt_r)}}{I_{0(0)}} = 1 - \frac{1}{2\pi} \int_0^{\pi/2} \int_0^{2\pi} \cos(2\sqrt{2}d(N \cdot t_r) \sin(2\beta) \sin(\gamma)) \sin\beta \, d\gamma d\beta$$

Equation 1

where N is the number of rotor period (t_r), $I_{0(Nt_r)}$ and $I_{(Nt_r)}$ are the peak integrals without and with dephasing ^1H π -pulses respectively, $I_{0(0)}$ is the peak integral at $N = 0$ and the orientation of the vector directed between ^{15}N and ^1H spins in the rotor frame is defined by the polar angle β and the azimuthal angle α . For each assigned residue, order parameters were calculated according to $\langle S \rangle = d_{\text{exp}}/d_{\text{rigid}}$, where the rigid-limit value was set to 11637 Hz, calculated with a amide bond length of 1.02 Å. Values of order parameters are reported in Supplementary Table S1.

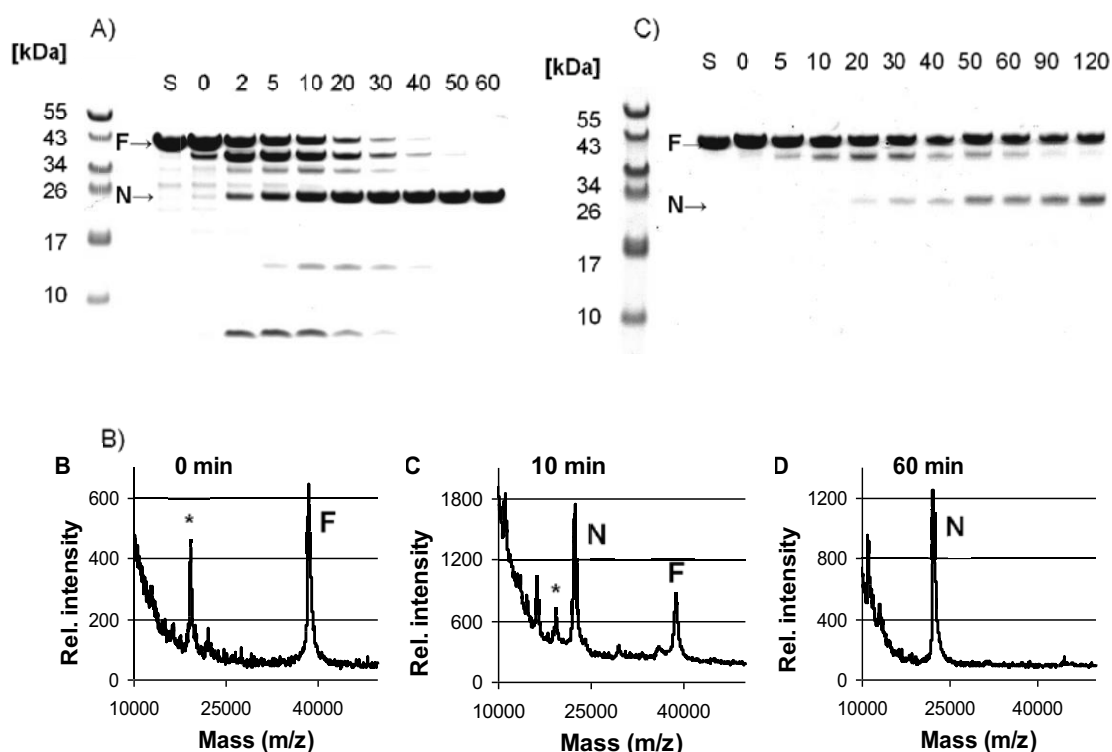
Spectra were processed and assigned peaks were integrated with Topspin 3.1. T_1 , $T_{1\rho}$ and dipolar coupling were fitted using GOSA (Czaplicki *et al.* 2006).

3.7. Supplementary Figures



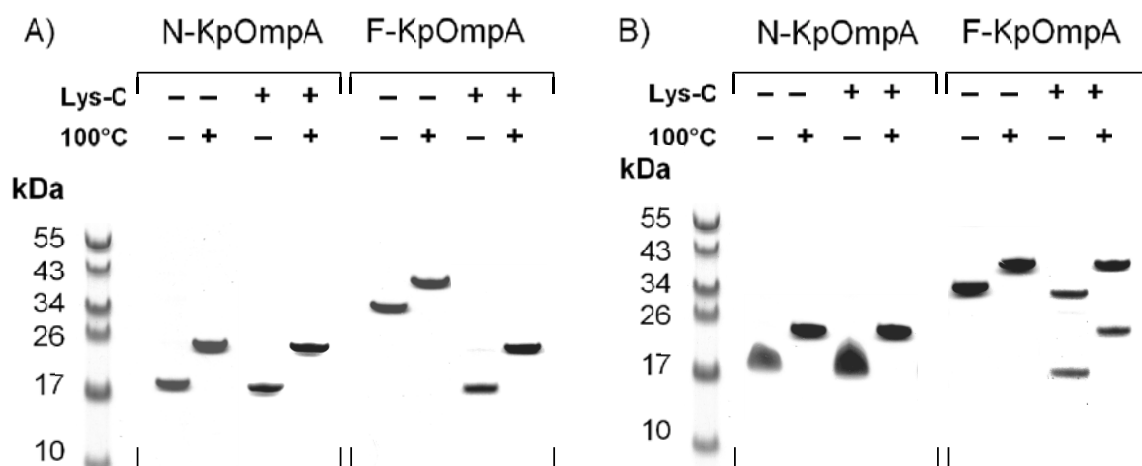
Supplementary Figure S1. Characterization of the protein samples: (A) SDS PAGE gel (Coomassie staining) of purified N-KpOmpA ('N', lanes 1 and 2) and F-KpOmpA ('F', lanes 3 and 4) in 0.1% Zw3-14 micelles. Lanes 2 and 4 are heat denatured (5 min at 100°C). Note the 'heat modifiability' effect on the protein band position. (B) and (C) MALDI-TOF linear mode spectra (matrix: SPA) of the purified N-KpOmpA (B) and F-KpOmpA (C) in 0.1% Zw3-14 micelles, exhibiting the expected sizes of ~23.4 and ~38.7 kDa, respectively, as well as 2⁺ ion peaks.

The protein purity and folding in the different sample preparations was monitored by SDS PAGE. KpOmpA (and the β -barrel proteins in general) exhibit the so called 'heat modifiability' (Nakamura *et al.* 1976a) which represents the intrinsic high stability of their structures. The barrel fold is preserved during electrophoresis in denaturing conditions unless the aliquot is boiled in advance, which reflects in different migration position of the protein band. Supplementary Figure S1A shows this effect for both N- and F-KpOmpA in micelles and was described previously for bilayer environment as well (Iordanov *et al.* 2012). Supplementary Figure S1B and C display linear mode MALDI-TOF spectra of the two protein constructs in Zw3-14 environment, exhibiting a major peak at the expected molecular weight and a 'double ion' peak with twice smaller mass-to-charge ratio (m/z). It is thus possible to use these techniques for monitoring the protease-induced degradation of KpOmpA and to map the cleavage sites. For instance, cleavage in the extracellular loops of the protein will remain 'masked' in the SDS PAGE gel due to the preserved barrel fold, unless the aliquot is heat denatured. On the other hand, the mass spectrometry is capable of identifying more precisely the exact cleavage site in the protein loops area, if any.



Supplementary Figure S2. Digestion kinetics of F-KpOmpA with Lys-C: (A) SDS PAGE gel of F-KpOmpA digested in 0.1% Zw3-14 micelles (SER = 450:1 w/w, 16°C), showing the complete transition of the protein from full-length chain (F) to the stable β -barrel (N). The time points (in minutes) are indicated above the respective gel wells. The 'zero time-point' (0 min) indicates that the aliquot was collected immediately after addition of the enzyme. 'S' denotes the protein stock solution before subjecting it to digestion. All aliquots are heat denatured (5 min, 100°C). (B) MALDI-TOF linear mode spectra (matrix: SPA) at time-points 0, 10 and 60 min of the reaction shown in (A). 'F' denotes the full-length protein (~38.7 kDa) and 'N' - the β -barrel alone (~22.4 kDa). The asterisk (*) indicates the 2⁺ ion of 'F'. (C) SDS PAGE gel of F-KpOmpA digested in PLE bilayers (LPR = 5 w/w, SER = 450:1 w/w, 37°C). All indications are as in (A).

The kinetics of F-KpOmpA digestion by Lys-C in Zw3-14 micelles Supplementary Figure S2A reveals the most vulnerable (i.e. quickly digested) K-sites of the molecule. These appear to be situated in the C-terminal domain of the protein. This gradual depletion of the F-KpOmpA pool and its conversion to N-KpOmpA is monitored precisely on the MALDI-TOF spectra, examples of which are given on Supplementary Figure S2B for 0, 10 and 60 min of reaction time. The experiment on F-KpOmpA reconstituted in PLE at LPR of 5 w/w (150:1 mol/mol) gave similar outcome, with the exception of preserving around 50% of the F-KpOmpA population at the end of the reaction, resulting in the two-band pattern seen on Supplementary Figure S2C. The reason for this is the random orientation of the protein in the lipid bilayer, leaving half of the C-terminal domain protected in the vesicle's lumen.



Supplementary Figure S3. Complete digestion of N- and F-KpOmpA with Lys-C visualized on SDS PAGE gels: (A) N- and F-KpOmpA were digested in 0.1% Zw3-14 micelles (SER = 100:1 w/w, 37°C overnight). The presence or absence of Lys-C and of heat denaturation (100°C) of the aliquots are indicated respectively with '+' and '-' above each well. (B) N- and F-KpOmpA were digested in PLE bilayers (LPR = 2 w/w, SER = 100:1 w/w, 37°C overnight). All indications are as in (A).

Supplementary Figure S3 shows the outcome of Lys-C complete digestion of N- and F-KpOmpA in Zw3-14 micelles (Supplementary Figure S3A) or PLE bilayers (Supplementary Figure S3B), indicating that the C-terminal domain (when present and exposed) is successfully targeted by the enzyme and completely absent at the end of the reaction, while the N-terminal transmembrane domain (either from F-KpOmpA or from N-KpOmpA) avoids digestion. The same result was observed in lipid bilayers with a LPR of 0.5, 2 and 5 w/w (i.e. 15, 60 and 150 lipids per protein), and in 9 different buffers, with three different NaCl concentrations (0, 150 and 300 mM) and three different pH values (6, 7, and 8).

3.8. Supplementary Tables

Supplementary Table S1. Order parameters and $T_{1\rho}$ values.

Residue	<S>	error	$T_{1\rho}$	error
G23	0,79	0,03	49,2	11,2
G22	0,79	0,05	33,3	23,9
G53	0,84	0,04	39,8	10,1
G55	0,79	0,03	56,1	21,7
G59	0,84	0,04	27,8	16,0
Y60	0,84	0,03	45,0	24,7
Q61	0,85	0,04	53,6	39,2
V62	0,81	0,04	24,3	12,5
D73	0,84	0,05	29,3	7,4
T97	0,85	0,03	90,0	46,5
A98	0,86	0,03	42,7	15,4
K99	0,94	0,02	51,0	31,3
L100	0,84	0,04	66,1	7,9
T112	0,94	0,11	29,3	6,5
G115	0,90	0,03	81,0	9,0
G116	0,96	0,06	16,0	9,9
R120	0,91	0,04	32,7	5,9
G147	0,79	0,05	33,3	23,9
G148	0,82	0,06	34,9	1,1
E150	0,96	0,06	38,5	17,1
W150	0,77	0,04	39,8	21,7
A158	0,84	0,05	29,3	7,4
T159	0,81	0,03	35,9	3,4
L161	0,83	0,07	49,3	4,4
E162	0,84	0,03	31,3	13,7
Y163	0,88	0,02	25,8	13,5
S185	0,83	0,03	23,3	7,9
S189	0,93	0,09	17,8	5,8
Y190	0,89	0,03	34,4	16,9
R191	0,79	0,03	47,3	28,3

Supplementary Table S2. Assigned resonances

Resid. Nb	Resid. type	Atom type	δ_{DSS} (ppm)
20	TYR	C	173.678
20	TYR	CB	40.613
21	ALA	C	175.391
21	ALA	CB	22.284
21	ALA	H	8.791
21	ALA	N	118.821
21	ALA	CA	50.746
22	GLY	C	171.478
22	GLY	CA	45.277
22	GLY	H	8.914
22	GLY	N	106.918
23	GLY	CA	43.968
23	GLY	C	170.671
23	GLY	H	7.904
23	GLY	N	103.423
24	LYS	C	173.167
24	LYS	CB	36.675
24	LYS	H	8.936
24	LYS	N	115.629
24	LYS	CA	53.566
25	LEU	C	175.699
25	LEU	CB	46.119
25	LEU	H	8.453
25	LEU	N	119.458
25	LEU	CA	52.751
26	GLY	CA	45.139
26	GLY	H	8.582
26	GLY	N	106.398
52	LEU	CB	43.221
53	GLY	H	9.227
53	GLY	N	110.374
54	ALA	CB	23.04
54	ALA	C	175.558
55	GLY	CA	46.56
55	GLY	H	8.378
55	GLY	N	104.278
56	ALA	C	172.667
56	ALA	CB	22.532
57	PHE	C	172.437
57	PHE	CB	41.714
57	PHE	H	8.769
57	PHE	N	119.984

57	PHE	CA	59.26
58	GLY	CA	44.592
58	GLY	C	171.029
58	GLY	H	9.053
58	GLY	N	105.659
59	GLY	CA	45.722
59	GLY	C	169.512
59	GLY	H	8.275
59	GLY	N	107.747
60	TYR	C	173.493
60	TYR	CB	41.506
60	TYR	H	8.287
60	TYR	N	117.605
60	TYR	CA	54.638
61	GLN	H	8.696
61	GLN	N	129.212
61	GLN	CA	53.418
62	VAL	C	173.69
62	VAL	CB	25.41
62	VAL	H	8.67
62	VAL	N	127.731
62	VAL	CA	64.068
66	LEU	C	173.974
66	LEU	CB	44.786
67	GLY	CA	43.508
67	GLY	H	8.774
67	GLY	N	114.482
69	GLU	CB	33.539
70	MET	CB	36.52
70	MET	H	9.415
70	MET	N	124.782
70	MET	CA	52.789
71	GLY	CA	45.209
71	GLY	H	8.677
71	GLY	N	107.514
72	TYR	C	173.41
72	TYR	CB	41.059
73	ASP	CA	52.935
73	ASP	H	8.472
73	ASP	N	127.299
94	VAL	C	173.998
94	VAL	CB	33.395
95	GLN	CA	51.643
95	GLN	CB	34.436
95	GLN	H	9.478

95	GLN	N	121.596
96	LEU	CA	53.377
96	LEU	C	175.243
96	LEU	CB	41.659
96	LEU	H	8.537
96	LEU	N	122.687
97	THR	C	173.394
97	THR	CB	72.913
97	THR	H	9.1
97	THR	CA	58.41
97	THR	N	112.3
98	ALA	CA	49.572
98	ALA	C	175.145
98	ALA	CB	19.192
98	ALA	H	8.557
98	ALA	N	121.287
99	LYS	C	174.609
99	LYS	CA	53.455
99	LYS	CB	32.502
99	LYS	H	8.852
99	LYS	N	123.65
100	LEU	CA	51.791
100	LEU	C	175.36
100	LEU	CB	43.664
100	LEU	H	8.911
100	LEU	N	131.995
101	GLY	CA	44.211
101	GLY	H	9.132
101	GLY	N	111.06
111	TYR	C	172.556
111	TYR	CB	41.691
112	THR	H	9.16
112	THR	CA	57.448
112	THR	N	111.578
113	ARG	C	175.157
113	ARG	CB	32.262
114	LEU	CA	52.567
114	LEU	C	176.044
114	LEU	CB	45.402
114	LEU	H	8.314
114	LEU	N	119.147
115	GLY	CA	45.977
115	GLY	C	171.509
115	GLY	H	8.1
115	GLY	N	108.369

116	GLY	CA	43.718
116	GLY	C	170.017
116	GLY	H	8.258
116	GLY	N	108.92
117	MET	CA	52.715
117	MET	C	174.846
117	MET	CB	37.921
117	MET	H	9.172
117	MET	N	118.75
118	VAL	H	9.007
118	VAL	N	126.403
118	VAL	CA	59.371
119	TRP	C	173.866
119	TRP	CB	34.153
120	ARG	CA	54.083
120	ARG	H	8.93
120	ARG	N	116.643
146	ALA	C	175.379
146	ALA	CB	22.244
147	GLY	CA	45.118
147	GLY	C	170.818
147	GLY	H	8.791
147	GLY	N	106.833
148	GLY	CA	45.556
148	GLY	H	7.938
148	GLY	N	108.604
149	VAL	C	174.754
149	VAL	CB	34.437
150	GLU	CA	49.572
150	GLU	CB	32.428
150	GLU	H	9.149
150	GLU	N	124.334
151	TRP	H	9.486
151	TRP	N	130.031
154	THR	C	174.235
154	THR	CB	71.085
155	ARG	H	8.198
155	ARG	N	115.65
155	ARG	CA	57.651
157	ILE	C	174.356
157	ILE	CB	38.209
158	ALA	C	176.784
158	ALA	CB	21.082
158	ALA	H	8.466
158	ALA	N	127.085

159	THR	H	9.052
159	THR	CA	58.225
159	THR	N	114.409
160	ARG	C	173.629
160	ARG	CB	32.556
161	LEU	CA	52.863
161	LEU	CB	44.333
161	LEU	N	123.972
162	GLU	C	173.37
162	GLU	CB	33.516
162	GLU	H	8.778
162	GLU	N	122.133
163	TYR	CA	56.117
163	TYR	H	8.753
163	TYR	N	123.241
183	MET	C	173.729
183	MET	CB	35.702
184	LEU	CA	52.678
184	LEU	C	175.416
184	LEU	CB	43.217
184	LEU	H	8.116
184	LEU	N	131.323
185	SER	H	9.781
185	SER	C	171.087
185	SER	CB	66.448
185	SER	CA	55.969
185	SER	N	120.737
186	LEU	CA	53.566
186	LEU	C	175.65
186	LEU	CB	44.715
186	LEU	H	8.511
186	LEU	N	122.793
187	GLY	CA	42.953
187	GLY	H	9.223
187	GLY	N	109.165
188	VAL	C	172.334
188	VAL	CB	34.466
189	SER	H	9.066
189	SER	C	171.915
189	SER	CB	66.347
189	SER	N	116.68
190	TYR	CA	54.49
190	TYR	CB	41.208
190	TYR	H	9.06
190	TYR	N	122.56

191	ARG	CA	53.446
191	ARG	H	8.065
191	ARG	N	125.999

IV. References

Barbet-Massin E, Pell A J, Retel J S, Andreas L B, Jaudzems K, Franks W T, Nieuwkoop A J, Hiller M, Higman V, Guerry P, Bertarello A, Knight M J, Felletti M, Le Marchand T, Kotelovica S, Akopjana I, Tars K, Stoppini M, Bellotti V, Bolognesi M, Ricagno S, Chou J J, Griffin R G, Oschkinat H, Lesage A, Emsley L, Herrmann T and Pintacuda G. *Rapid Proton-Detected NMR Assignment for Proteins with Fast Magic Angle Spinning*. Journal of the American Chemical Society, **2014**, 136(35): 12489-12497.

Bender M L and Kezdy J. *Mechanism of Action of Proteolytic Enzymes*. Annual review of biochemistry, **1965**, 34: 49-76.

Columbus L. *Post-expression strategies for structural investigations of membrane proteins*. Current opinion in structural biology, **2015**, 32: 131-138.

Confer A W and Ayalew S. *The OmpA family of proteins: Roles in bacterial pathogenesis and immunity*. Veterinary Microbiology, **2013**, 163(3-4): 207-222.

Czaplicki J, Cornélissen G and Halberg F. *GOSA, a simulated annealing-based program for global optimization of nonlinear problems, also reveals transyears*. Journal of applied biomedicine, **2006**, 4(2): 87-94.

Fontana A, Fassina G, Vita C, Dalzoppo D, Zamai M and Zambonin M. *Correlation between sites of limited proteolysis and segmental mobility in thermolysin*. Biochemistry, **1986**, 25(8): 1847-1851.

Fox D A and Columbus L. *Solution NMR resonance assignment strategies for beta-barrel membrane proteins*. Protein science : a publication of the Protein Society, **2013**, 22(8): 1133-1140.

Gullion T. *Rotational-Echo, Double-Resonance NMR*. **2006**, *Modern Magnetic Resonance*, Webb G, Springer Netherlands: 713-718.

Gullion T and Schaefer J. *Rotational-echo double-resonance NMR*. Journal of Magnetic Resonance (1969), **1989**, 81(1): 196-200.

Hediger S, Meier B H, Kurur N D, Bodenhausen G and Ernst R R. *NMR cross polarization by adiabatic passage through the Hartmann—Hahn condition (APHH)*. Chemical Physics Letters, **1994**, 223(4): 283-288.

Huang K S, Bayley H, Liao M J, London E and Khorana H G. *Refolding of an Integral Membrane-Protein - Denaturation, Renaturation, and Reconstitution of Intact Bacteriorhodopsin and 2 Proteolytic Fragments*. Journal of Biological Chemistry, **1981**, 256(8): 3802-3809.

Iordanov I, Renault M, Reat V, Bosshart P D, Engel A, Saurel O and Milon A. *Dynamics of Klebsiella pneumoniae OmpA transmembrane domain: the four extracellular loops display restricted motion behavior in micelles and in lipid bilayers*. Biochim Biophys Acta, **2012**, 1818(9): 2344-2353.

Jeannin P, Bottazzi B, Sironi M, Doni A, Rusnati M, Presta M, Maina V, Magistrelli G, Haeuw J F, Hoeffel G, Thieblemont N, Corvaia N, Garlanda C, Delneste Y and Mantovani A. *Complexity and complementarity of outer membrane protein a recognition by cellular and humoral innate immunity receptors*. Immunity, **2005**, 22(5): 551-560.

Keller R L J. *Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment*. **2005**, Diss., Naturwissenschaften, Eidgenössische Technische Hochschule ETH Zürich, Nr. 15947, 2005.

Knight M J, Webber A L, Pell A J, Guerry P, Barbet-Massin E, Bertini I, Felli I C, Gonnelli L, Pierattelli R, Emsley L, Lesage A, Herrmann T and Pintacuda G. *Fast Resonance Assignment and Fold Determination of Human Superoxide Dismutase by High-Resolution Proton-Detected Solid-State MAS NMR Spectroscopy*. *Angewandte Chemie International Edition*, **2011**, 50(49): 11697-11701.

Kurupati P, Ramachandran N P and Poh C L. *Protective Efficacy of DNA Vaccines Encoding Outer Membrane Protein A and OmpK36 of Klebsiella pneumoniae in Mice*. *Clinical and Vaccine Immunology*, **2011**, 18(1): 82-88.

Lewandowski J R, Halse M E, Blackledge M and Emsley L. *Protein dynamics. Direct observation of hierarchical protein dynamics*. *Science*, **2015**, 348(6234): 578-581.

Lewandowski J R, Sein J, Sass H J, Grzesiek S, Blackledge M and Emsley L. *Measurement of site-specific ¹³C spin-lattice relaxation in a crystalline protein*. *J Am Chem Soc*, **2010**, 132(24): 8252-8254.

Llobet E, March C, Gimenez P and Bengoechea J A. *Klebsiella pneumoniae OmpA confers resistance to antimicrobial peptides*. *Antimicrob Agents Chemother*, **2009**, 53(1): 298-302.

Lo R Y C and Sorensen L S. *The outer membrane protein OmpA of Mannheimia haemolytica A1 is involved in the binding of fibronectin*. *Fems Microbiology Letters*, **2007**, 274(2): 226-231.

March C, Moranta D, Regueiro V, Llobet E, Tomas A, Garmendia J and Bengoechea J A. *Klebsiella pneumoniae Outer Membrane Protein A Is Required to Prevent the Activation of Airway Epithelial Cells*. *Journal of Biological Chemistry*, **2011**, 286(12): 9956-9967.

Marion D, Ikura M and Bax A. *Improved solvent suppression in one- and two-dimensional NMR spectra by convolution of time-domain data*. *Journal of Magnetic Resonance (1969)*, **1989**, 84(2): 425-430.

Nakamura K and Mizushima S. *Effects of Heating in Dodecyl-Sulfate Solution on Conformation and Electrophoretic Mobility of Isolated Major Outer Membrane Proteins from Escherichia-Coli K-12*. *Journal of Biochemistry*, **1976a**, 80(6): 1411-1422.

Nakamura K and Mizushima S. *Effects of heating in dodecyl sulfate solution on the conformation and electrophoretic mobility of isolated major outer membrane proteins from Escherichia coli K-12*. *J Biochem*, **1976b**, 80(6): 1411-1422.

Popot J L, Trehwella J and Engelman D M. *Reformation of Crystalline Purple Membrane from Purified Bacteriorhodopsin Fragments*. *Embo Journal*, **1986**, 5(11): 3039-3044.

Rasko D A, Rosovitz M J, Myers G S, Mongodin E F, Fricke W F, Gajer P, Crabtree J, Sebahia M, Thomson N R, Chaudhuri R, Henderson I R, Sperandio V and Ravel J. *The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates*. *Journal of bacteriology*, **2008**, 190(20): 6881-6893.

Renault M, Saurel O, Czaplicki J, Demange P, Gervais V, Lohr F, Reat V, Piotto M and Milon A. *Solution state NMR structure and dynamics of KpOmpA, a 210 residue transmembrane domain possessing a high potential for immunological applications*. *J Mol Biol*, **2009**, 385(1): 117-130.

Rigaud J L, Paternostre M T and Bluzat A. *Mechanisms of Membrane-Protein Insertion into Liposomes during Reconstitution Procedures Involving the Use of Detergents .2. Incorporation of the Light-Driven Proton Pump Bacteriorhodopsin*. Biochemistry, **1988**, 27(8): 2677-2688.

Schanda P, Meier B H and Ernst M. *Quantitative analysis of protein backbone dynamics in microcrystalline ubiquitin by solid-state NMR spectroscopy*. J Am Chem Soc, **2010**, 132(45): 15957-15967.

Shaka A J, Keeler J and Freeman R. *Evaluation of a new broadband decoupling sequence: WALTZ-16*. Journal of Magnetic Resonance (1969), **1983**, 53(2): 313-340.

Westerlund B and Korhonen T K. *Bacterial proteins binding to the mammalian extracellular matrix*. Molecular microbiology, **1993**, 9(4): 687-694.

Zhou D H and Rienstra C M. *High-performance solvent suppression for proton detected solid-state NMR*. Journal of Magnetic Resonance, **2008**, 192(1): 167-172.

Conclusions et perspectives

Dans les précédents chapitres, nous avons présenté les principaux résultats obtenus au cours de ce travail sur les deux protéines modèles étudiées, la lipase Lip2 de la levure de *Yarrowia lipolytica* et la protéine OmpA de la bactérie *Klebsiella pneumoniae*. Nous allons à présent exposer différents prolongements en cours possibles ainsi que les résultats préliminaires obtenus et les perspectives envisageables à plus long terme.

I. Expression de la lipase Lip2 de *Yarrowia lipolytica* pour la RMN.

Parmi les systèmes d'expression étudiés dans le chapitre 2 pour produire Lip2 en vue de son étude par RMN, la levure *Yarrowia lipolytica* s'est révélée être le plus performant. Nous avons été en mesure de produire des échantillons de Lip2 uniformément marqué ^{15}N et d'enregistrer le spectre HSQC ^1H - ^{15}N de cette lipase (Figure 2A). Toutefois, une étude des conditions de culture afin de réduire les quantités de glucose ^{13}C est nécessaire pour envisager un marquage uniforme $^{15}\text{N}^{13}\text{C}$, comme cela a été réalisé chez *Kluyveromyces lactis* (Sugiki *et al.* 2008).

Afin de pouvoir facilement sonder la dynamique locale de Lip2 par des expériences de relaxation ^{15}N , nous avons aussi tenté une stratégie de marquage ^{15}N spécifique d'un type d'acide aminé, par exemple la leucine (Muchmore *et al.* 1989, Parker *et al.* 2004, Ozawa *et al.* 2006). Cette stratégie a été utilisée avec succès chez la levure *Pichia pastoris* pour obtenir un marquage ^{15}N spécifique des leucines, cystéines, méthionines et lysines (Chen *et al.* 2006).

```

1      11      21      31      41
VYTSTETSHI DQESYNFFEK YARLANIGYC VPGGTKIFKP FNCGLQCAHF

51      61      71      81      91
PNVELIEEFH DPRILFDVSG YLAVDHASKQ IYLVIRGTHSLEDVITDIRI

101     111     121     131     141
MQAPLTNFDL AANISSTATC DDCLVHNGFI QSYNNTYNQI GPKLDSVIEQ

151     161     171     181     191
YPDYQIAVTG HSLGGAAALL FGINLKVNGH DPLVVTLLGQP IVGNAGFANW

201     211     221     231     241
VDKLLFFGQEN PDVSKVSKDR KLYRITHRGD IVPQVPFWDG YQHCSGEVFI

251     261     271     281     291
DWPLIHPPLS NVVMCQGQSN KQCSAGNTLL QQVNVIGNHLQYFVTEGVCG

301
I

```

Figure 1. Séquence de Lip2 mature avec les leucines surlignées en vert. La séquence du lid, T88-L105, est surlignée en jaune et la séquence de l'hélice α faisant face au lid, N284-Q291, est surlignée en gris. Les acides aminés catalytiques, D230, H289 et S162, sont en rouges et les acides aminés du trou oxyanion T88, L163 et R86 sont en bleu.

La séquence mature de Lip2 contient 24 leucines (Figure 1). La leucine L91 est située sur le lid, alors que la leucine L105 est située en fin du lid, à la jonction avec un brin β . La leucine L290 est située sur une hélice α faisant face au lid. Enfin, la leucine L163 constitue le premier résidu du trou oxyanion (Figure 2B). L'identification des signaux correspondant aux leucines d'intérêt peut se faire rapidement en comparant le spectre HSQC ^1H - ^{15}N de la forme sauvage avec les spectres de mutants construits par mutagenèse dirigée.

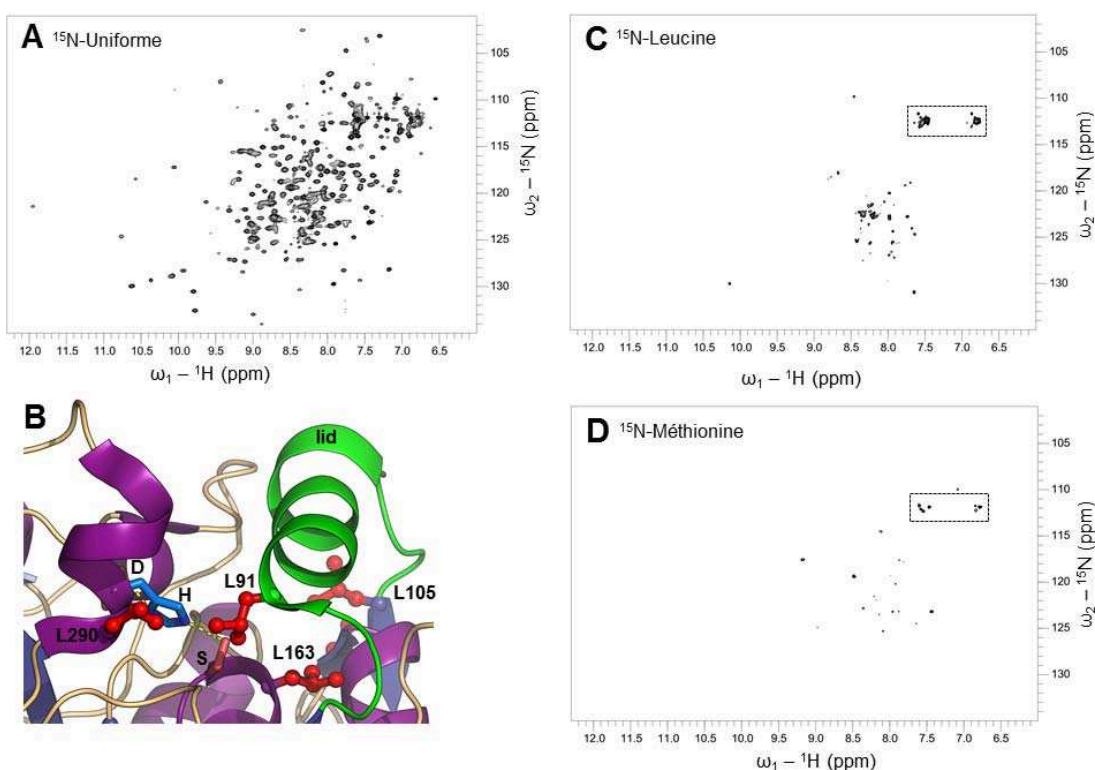


Figure 2. A. Spectre HSQC ^1H - ^{15}N de Lip2 uniformément marquée ^{15}N . B. Vue rapprochée du lid et du site actif de Lip2 en conformation fermée (PDB ID : 3O0D). Les leucines sont représentées en « ball and stick » et colorées en rouge. Le lid est coloré en vert et les acides aminés catalytiques sont représentés en « sticks ». C. et D. Spectre HSQC ^1H - ^{15}N de Lip2 spécifiquement marquée sur les leucines (C.) ou sur les méthionines (D.). Les résonances des chaînes latérales des asparagines et des glutamines sont encadrées. Les spectres ont été enregistrés à 293 K sur un spectromètre Bruker Avance III HD opérant à une fréquence de Larmor ^1H de 600 MHz.

Cependant, l'ajout de [^{15}N]-leucine lors de la production de Lip2 chez *Yarrowia lipolytica* n'a pas permis d'obtenir un marquage spécifique. En effet, le spectre HSQC ^1H - ^{15}N présente un nombre de corrélation supérieur aux 24 signaux attendus des leucines. La présence de signaux correspondant aux chaînes latérales des asparagines et des glutamines témoigne de phénomènes de transamination (Figure 2C). Enfin, la grande hétérogénéité des intensités des signaux ne permet pas de distinguer les corrélations provenant de leucines de celles provenant d'autres acides aminés suite à des réactions de transamination. Nous avons également essayé un marquage avec la [^{15}N]-méthionine, cet acide aminé n'induisant en principe aucun phénomène de transamination chez *Pichia pastoris* (Chen *et al.* 2006). La séquence mature de Lip2 inclue 2 méthionines (Figure 3), mais ici encore la présence de nombreuses corrélations sur l'HSQC ^1H - ^{15}N , dont certaines correspondent aux chaînes latérales des asparagines et glutamines (Figure 2D), indique l'existence de mélange isotopique.

Notre hypothèse est que, contrairement à l'expression dans *E. coli*, la durée d'induction de production de protéine chez *Yarrowia* est très longue (12 heures contre 4 à 6 heures pour une induction forte chez *E. coli*). Ceci laisse tout le temps nécessaire à un

ensemble de voies cataboliques et anaboliques pour redistribuer le marquage isotopique au sein de la biomasse et donc des protéines exprimées.

```

1      11      21      31      41
VYTSTETSHI DQESYNFFEK YARLANIGYC VGPGTKIFKP FNCGLQCAHF

51     61     71     81     91
PNVELIEEFH DPRLIFDVSG YLAVDHASKQ IYLVIRGTHS LEDVITDIRI

101    111    121    131    141
MQAPI TNFDL AANISSTATC DDCLVHNGFI QSYNNNTYNQI GPKLDSVIEQ

151    161    171    181    191
YPDYQIAVTG HSLGGAAALL FGINLKVNGH DPLVVTLGQP IVGNAGFANW

201    211    221    231    241
VDKLFQEN PDVSKVSKDR KLYRITHRGD IVPQVPFWDG YQHCSGEVFI

251    261    271    281    291
DWPLIHPPLS NVVMCQGQSN KQCSAGNTLL QQVNVIGNHL QYFVTEGVCG

301
I

```

Figure 3. Séquence de Lip2 mature avec les méthionines surlignées en vert. La séquence du lid, T88-L105, est surlignée en jaune et la séquence de l'hélice α faisant face au lid, N284-Q291, est surlignée en gris. Les acides aminés catalytiques, D230, H289 et S162, sont en rouges et les acides aminés du trou oxyanion, T88 et L163, ainsi que R86 sont en bleu.

D'autres stratégies de marquage spécifiques sont envisageables. Il est ainsi possible de marquer spécifiquement les groupements méthyls des valines, leucines et isoleucines à partir de précurseurs métaboliques marqués. Cette approche a ainsi permis de marquer sélectivement ^1H - ^{13}C les groupements méthyls de protéines deutérées afin d'en étudier la dynamique (Goto *et al.* 1999, Tugarinov *et al.* 2007). Ces protocoles ont récemment été adaptés pour marquer les groupements méthyls des isoleucines en milieu deutéré chez les levures *Kluyveromyces lactis* et *Pichia pastoris* (Miyazawa-Onami *et al.* 2013, Clark *et al.* 2015).

Une autre stratégie prometteuse est le couplage chimique d'une sonde RMN sur les chaînes latérales d'acides aminés par modification chimique en utilisant des réactifs chimiques marqués (Kosen 1989). Les résidus les plus couramment utilisés pour ces modifications chimiques sont les tyrosines, les cystéines et les lysines. L'introduction d'atome de fluor ^{19}F a ainsi été utilisée pour marquer des tyrosines (Hebel *et al.* 1990) ou des cystéines (Klein-Seetharaman *et al.* 1999, Liu *et al.* 2012). Les cystéines peuvent également être utilisées pour introduire des groupements méthyl ^{13}C formant des ^{13}C -S-méthylthiocystéines (Religa *et al.* 2011). Cette approche est très intéressante, combinant

l'avantage donné par l'observation de groupements méthyls (Tugarinov *et al.* 2003), avec la grande similarité entre les S-méthylthiocystéines et les méthionines, ce dernier aspect laissant présager que cette modification altèrera peu les propriétés structurales et dynamiques des protéines étudiées. Enfin, des groupements méthyls peuvent également être introduits sur les chaînes latérales des lysines (Means *et al.* 1968). La méthylation des lysines n'entraîne pas de modification structurale majeure, un ou deux groupements méthyls étant ajouté et la charge de la chaîne latérale étant maintenue (Jentoft *et al.* 1979, Rypniewski *et al.* 1993, Larda *et al.* 2012). Le groupe de B. Kobilka a ainsi pu utiliser cette technique pour étudier le récepteur β 2-adrénérique (Bokoch *et al.* 2010).

II. Interaction de C-KpOmpA avec le DAP.

Nous avons résolu par cristallographie la structure du domaine C-terminal de KpOmpA sous forme apo ou en complexe avec le citrate. La molécule de citrate est située dans une poche qui a été identifiée comme site de liaison à l'acide diaminopimélique (DAP) (Parsons *et al.* 2006, Park *et al.* 2012), un acide aminé retrouvé dans la partie peptidique du peptidoglycane (Vollmer *et al.* 2008). Nous avons alors utilisé des expériences RMN de différence de transfert de saturation (STD) pour rechercher une interaction entre le DAP et C-KpOmpA (Mayer *et al.* 2001).

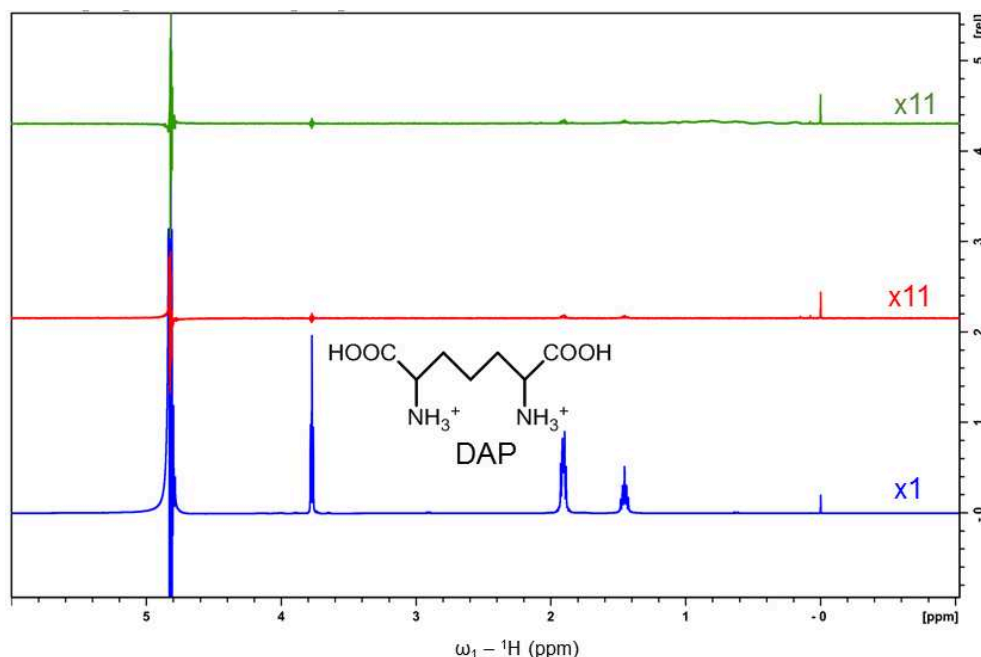


Figure 4. Spectre STD du DAP. Spectre de référence (irradiation à 18005 Hz) du DAP en bleu, spectre STD (irradiation à -190 Hz) du DAP sans C-KpOmpA, en rouge et avec C-KpOmpA, en vert. Le signal à 4.8 ppm correspond à l'eau et le signal à 0 ppm correspond au DSS utilisé pour la calibration des spectres. La structure du DAP est indiquée ainsi que les échelles relatives. Spectres enregistrés sur un spectromètre Bruker Avance III HD opérant à une fréquence de Larmor ^1H de 600 MHz à 293 K sur un échantillon de DAP à 5 mM dans un tampon 5 mM phosphate pH 6.5, 50 mM NaCl, 100 μM DSS, 10% D_2O avec ou sans 50 μM de C-KpOmpA.

Nous n'avons pas observé de signal STD du DAP en présence de C-KpOmpA (Figure 4). Nous avons également essayé différentes concentrations ainsi que différents tampons et pH mais ces expériences n'ont montré aucune interaction entre le DAP et le domaine C-KpOmpA dans les conditions testées.

Nous avons également enregistré des spectres HSQC ^1H - ^{15}N de C-KpOmpA en absence et en présence de DAP (Figure 5A). Les variations de déplacement chimique ont été calculées avec l'Équation 1, où δ_{H} et δ_{N} sont les déplacements chimiques ^1H et ^{15}N en ppm, respectivement, et α , un facteur d'échelle fixé à 0.15 (Williamson 2013).

$$d = \sqrt{\frac{1}{2}[\delta_{\text{H}}^2 + (\alpha \cdot \delta_{\text{N}}^2)]}$$

Équation 1. Calcul des variations de déplacement chimique

L'ajout de 5 mM de DAP à 0.8 mM de C-KpOmpA, soit un ratio protéine / ligand de 1/6, entraîne des shifts mineurs au niveau des corrélations des résidus G263 et E265, de 0.013 ppm et 0.005 ppm respectivement (Figure 5B).

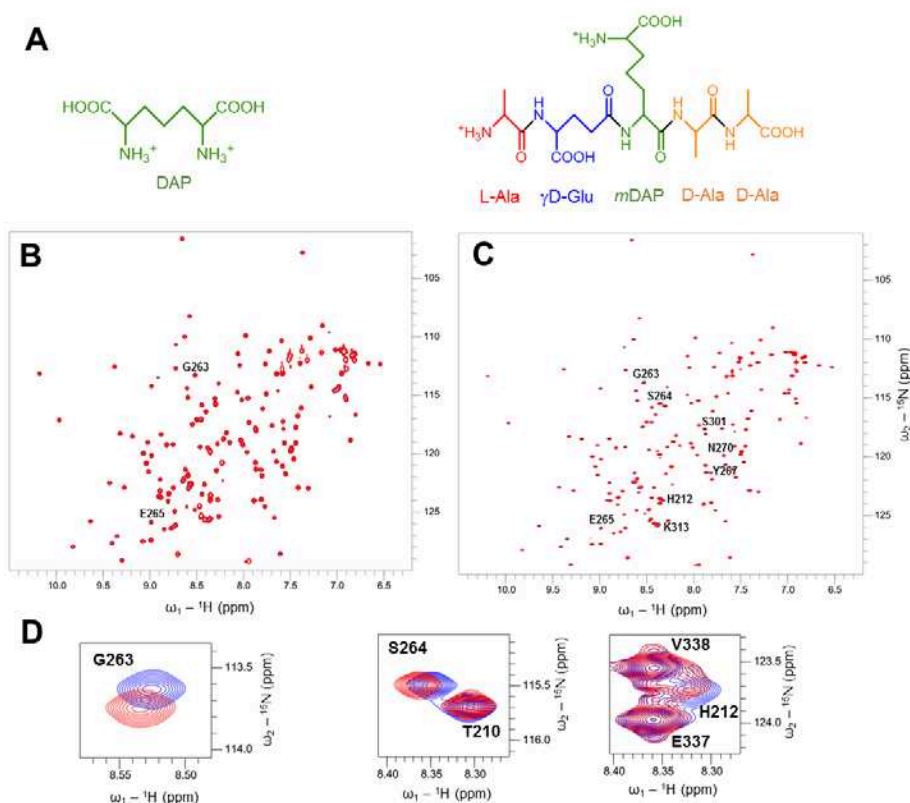


Figure 5. A. Structure du DAP et du pentapeptide L-Ala- γ D-Glu-*m*DAP-D-Ala-D-Ala utilisés dans les expériences de déplacement chimique. B. Spectre de C-KpOmpA avec les positions présentant un déplacement chimique en présence de DAP indiquées. C. Spectre de C-KpOmpA avec les positions présentant un déplacement chimique en présence du pentapeptide indiquées. D. Agrandissement des positions présentant les déplacements chimiques les plus importants en présence du pentapeptide. Spectre sans le pentapeptide en bleu et spectre avec le pentapeptide en rouge. Les spectres ont été enregistrés à 293 K sur un spectromètre Bruker Avance III HD opérant à une fréquence de Larmor ^1H de 600 MHz. Les expériences avec le DAP (B) ont été enregistrés sur un échantillon de C-KpOmpA 800 μM dans un tampon 5 mM phosphate pH 6.5, 50 mM NaCl, 100 μM DSS, 10% D_2O avec ou sans 5 mM de DAP. Les expériences avec le pentapeptide (C et D) ont été enregistrés sur un échantillon de C-KpOmpA 300 μM dans le même tampon avec ou sans 1.5 mM de pentapeptide.

Le remplacement du DAP par un pentapeptide synthétique (Figure 5A) incluant le DAP entraîne des déplacements chimiques plus nombreux pour un ratio protéine / ligand similaire de 1/5 (soit 1.5 mM de pentapeptide pour 0.3 mM de C-KpOmpA) (Figure 5C). Les déplacements chimiques les plus importants correspondent aux résidus H212, 0.017 ppm, G263, 0.020 ppm et S264, 0.015 ppm (Figure 5D).

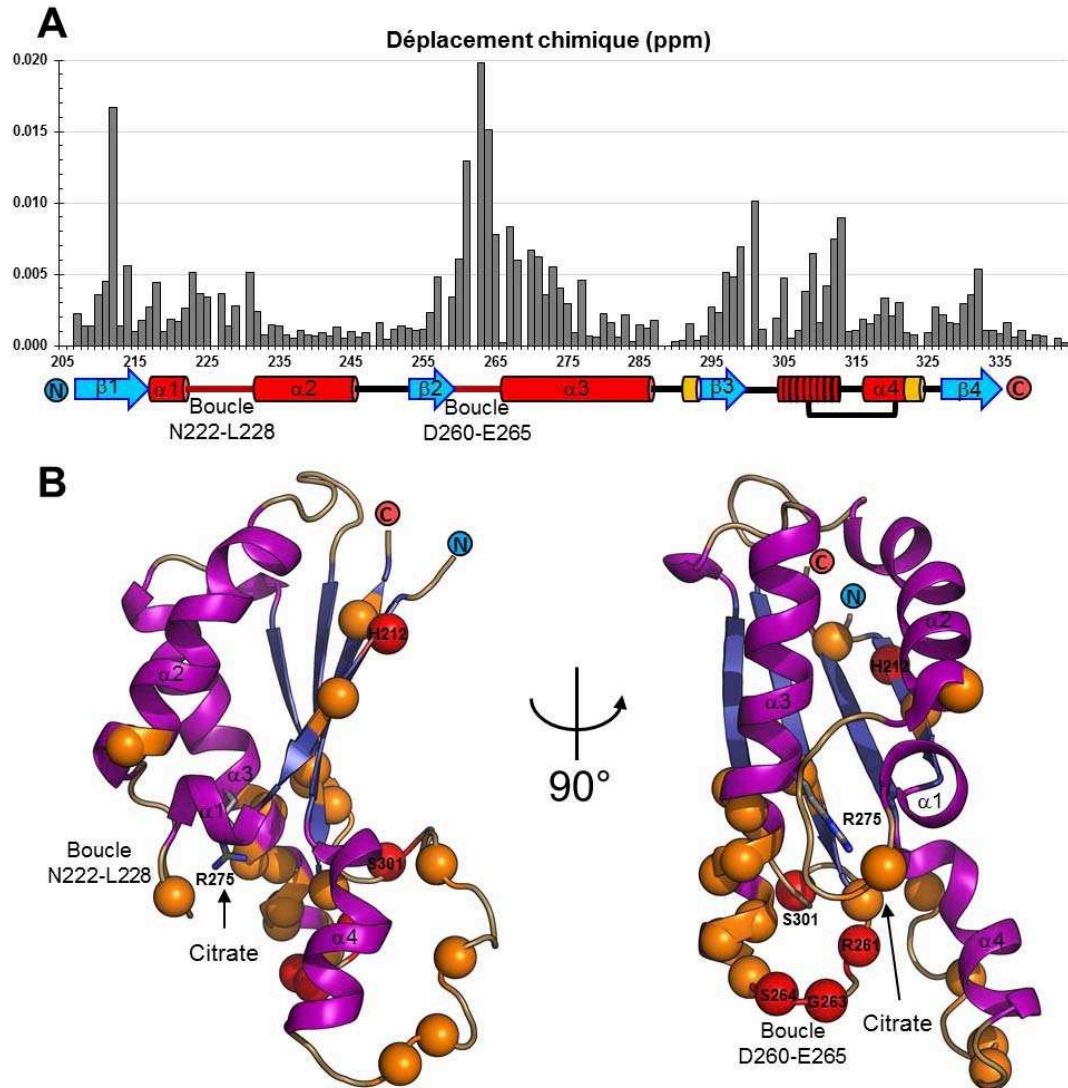


Figure 6. A. Déplacements chimiques induits par l'ajout du pentapeptide en fonction de la séquence de C-KpOmpA. Les structures secondaires de C-KpOmpA sont représentées et annotées en dessous de la séquence. B. Déplacements chimiques représentés sur la structure cristallographique de C-KpOmpA en complexe avec le citrate. Les azotes des groupements amides présentant un déplacement chimique supérieur à 0.005 ppm sont représentés par des sphères oranges et ceux présentant un déplacement chimique supérieur à 0.010 ppm sont représentés par des sphères rouges et sont annotés.

Les déplacements chimiques induits par le pentapeptide sont majoritairement situés au niveau de la boucles D260-E265 et de l'hélice α_3 (Figure 6A). Sur la structure, la boucle D260-E265 correspond à la face postérieure de la poche de liaison au citrate et donc au peptidoglycane via le DAP (Figure 6B). Toutefois, des déplacements chimiques supérieurs à 0.005 ppm sont également retrouvés au niveau du brin β_1 ainsi que du brin β_3 et de la série de β -turn qui précède l'hélice α_4 . De plus, la boucle N222-L228 n'est que modérément affectée par l'ajout du pentapeptide, avec des déplacements chimiques inférieurs à 0.005 ppm, alors qu'elle constitue la face antérieure de la poche de liaison. Enfin, lors d'expériences de titration de C-KpOmpA par le pentapeptide, les déplacements chimiques

observés ne suivent pas une variation linéaire avec la concentration en pentapeptide : ceci suggère soit la coexistence de plusieurs sites de liaison, soit que les variations sont trop faibles pour être significatives.

La faible amplitude des déplacements chimiques observés ainsi que l'impossibilité d'obtenir une courbe de titration nous ont amenés à nous demander si un ligand avec une plus forte affinité n'était pas déjà présent au niveau site de liaison, empêchant toute interaction avec le DAP ou le pentapeptide. Il a en effet été rapporté que lors de la purification de la protéine Pal d'*Haemophilus influenzae*, un précurseur biosynthétique du peptidoglycane restait lié à la protéine tout au long des étapes de purification (Parsons *et al.* 2006). Une dénaturation partielle de la protéine en chauffant celle-ci à 47°C a été nécessaire pour rompre l'interaction avec le précurseur. De même, le protocole de purification du domaine C-terminal de OmpA d'*Acinetobacter baumannii* comporte une étape de dénaturation – renaturation à l'urée 6 M, sans toutefois que les auteurs n'aient justifié cette étape (Park *et al.* 2012).

Nous avons enregistré des spectres HSQC ^1H - ^{15}N de C-KpOmpA avant dénaturation à l'urée 6 M (Figure 7A), pendant la dénaturation (Figure 7B) et après renaturation par dialyse sur la nuit à 4°C pour éliminer l'urée (Figure 7C). La structure de C-KpOmpA après renaturation est identique à celle avant dénaturation comme en témoigne l'excellente superposition des spectres HSQC ^1H - ^{15}N (Figure 7D). Quelques faibles variations de déplacements chimiques (entre 0.010 et 0.015 ppm) sont observées pour les corrélations correspondant aux résidus G263, S264, E265 et H212. Nous avons également repris sur C-KpOmpA renaturée les expériences de STD avec le DAP et de cartographie de déplacement chimique avec le DAP et le pentapeptide synthétique. Les résultats furent identiques à ceux obtenus précédemment avec C-KpOmpA sans étape de dénaturation – renaturation.

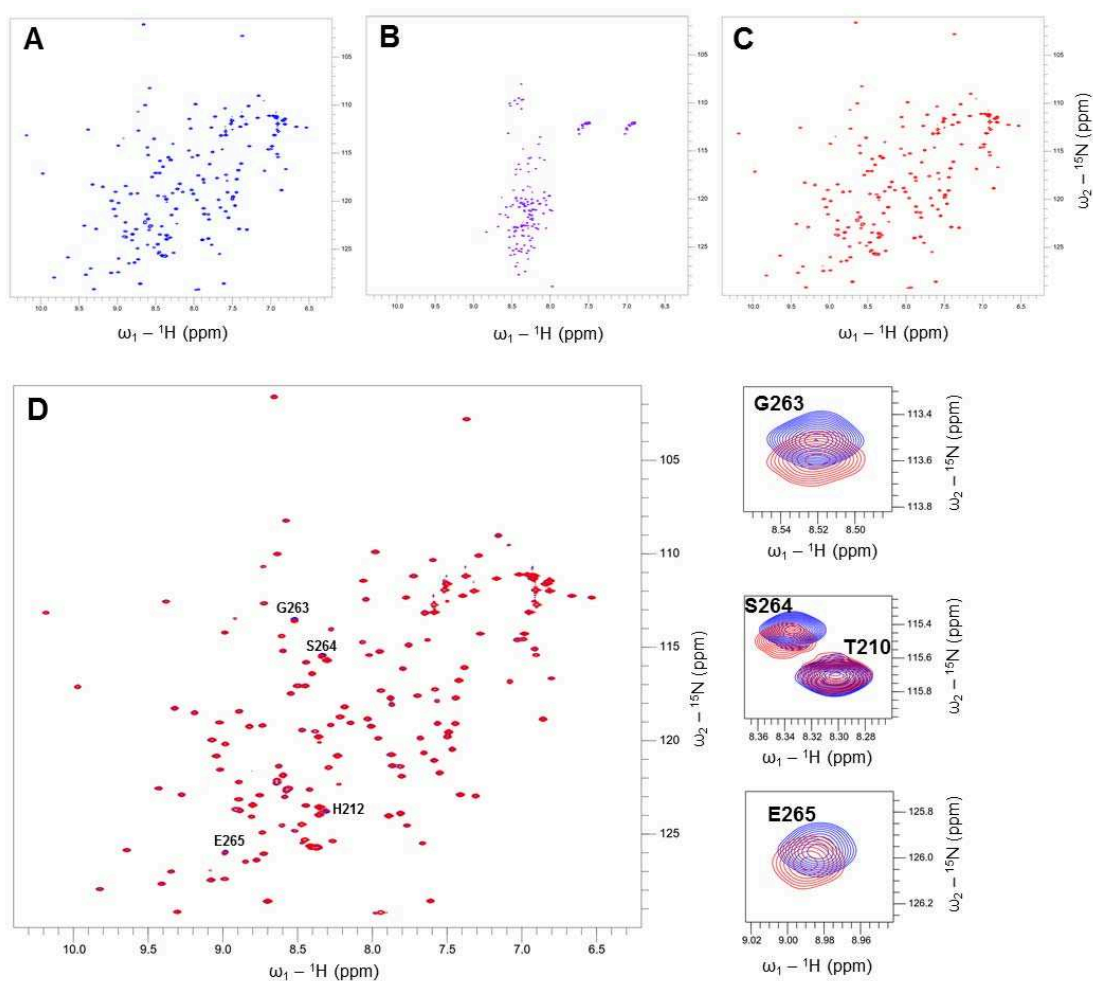


Figure 7. Spectres HSQC ^1H - ^{15}N de C-KpOmpA. A. Avant dénaturation. B. Dénaturé à l'urée 8 M. C. Après renaturation par dialyse. D. Superposition des spectres avant dénaturation (A.) et après renaturation (C.) avec agrandissement pour les positions G59, S60 et E61. Les spectres ont été enregistrés à 293 K sur un spectromètre Bruker Avance III HD opérant à une fréquence de Larmor ^1H de 600 MHz sur un échantillon de C-KpOmpA 300 μM dans un tampon 5 mM phosphate pH 6.5, 50 mM NaCl, 10 μM DSS, 10% D_2O .

En conclusion, il apparaît que l'interaction entre le DAP et le domaine C-terminal de KpOmpA *in vitro* est de très faible affinité ($K_d > \text{mM}$). Elle pourrait nécessiter un peptidoglycane complet pour être observée, comme le suggère l'augmentation du nombre de pics affectés lors de l'utilisation du pentapeptide synthétique à la place du DAP seul (Figure 5B et Figure 5C).

III. Expression de F-KpOmpA dans la membrane externe d'*E. coli*.

Afin d'étudier KpOmpA dans des conditions plus proches de son environnement natif, un projet a été entrepris dans l'équipe pour exprimer et adresser la séquence complète de KpOmpA (F-KpOmpA) à la membrane externe d'*Escherichia coli*. Cette étude fait suite aux résultats remarquables observés par M. Renault lors de son post doctorat dans l'équipe de M. Baldus (Utrecht, NL), et connus sous le terme « cellular solid state NMR » (Renault *et*

al. 2012). Ces études ont montré la possibilité de mesurer et d'analyser des spectres de RMN du solide de protéines de la paroi d'*E. coli*, sans les isoler de leur environnement naturel, sur bactérie entière ou sur préparation de paroi.

Les conditions d'expression et de marquage isotopique de F-KpOmpA recombinante ont été mises au point dans des souches d'*E. coli* déficientes pour les protéines OmpA et OmpF endogènes afin de favoriser l'adressage et l'insertion correcte de F-KpOmpA dans la membrane externe (Renault *et al.* 2012). Après cassage des bactéries et isolement de l'enveloppe cellulaire (EC) par ultracentrifugation du lysat clair (LC), les fractions cellulaires EC et LC obtenues à partir de cultures induites et non induites sont analysées par différentes méthodes biochimiques. L'analyse des fractions sur gel SDS-PAGE après révélation au bleu de Coomassie indique la présence de la protéine recombinante intacte après induction (Figure 8). La comparaison des fractions LC et EC montre un enrichissement important de F-KpOmpA dans l'enveloppe bactérienne. Des expériences complémentaires ont confirmé sa localisation et son insertion dans la membrane externe. Enfin, la différence de mobilité électrophorétique de F-KpOmpA au sein des préparations EC natives et dénaturées par chauffage (100°C, 10 min, 2% SDS) est caractéristique d'un repliement correct du tonneau β N-terminal de F-KpOmpA dans la membrane externe de *E. coli* (Nakamura *et al.* 1976).

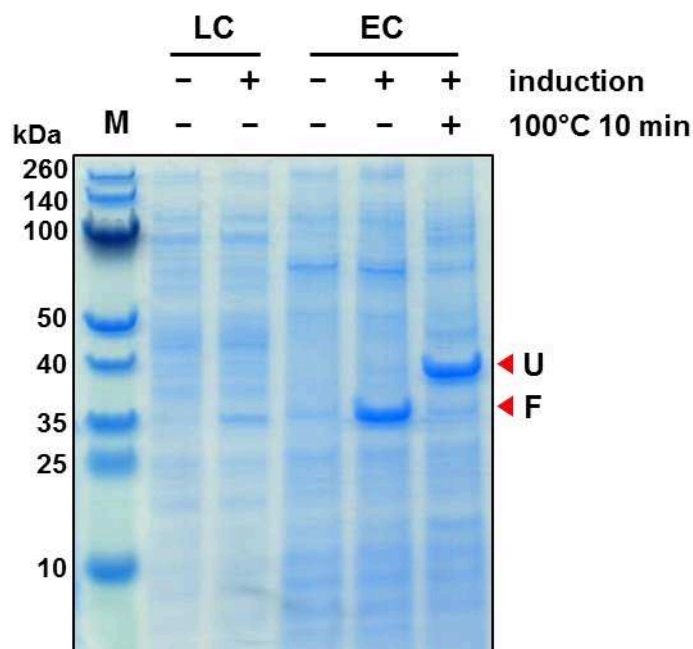


Figure 8. Expression de (U-¹³C, ¹⁵N)-KpOmpA recombinante dans la membrane externe d'*E. coli*. Analyse sur gel SDS-PAGE après révélation au bleu de Coomassie des protéines composants les fractions LC et EC avec (+) et sans (-) induction par l'IPTG, avec (+) et sans (-) denaturation des échantillons par chauffage. Les positions de KpOmpA correctement repliée (F) ou dénaturée par la chauffage (U) sont indiquées à droite du gel. M, marqueur de taille ; LC, lysat clair ; EC, enveloppe cellulaire.

Dans ce contexte, des préparations d'enveloppe cellulaire uniformément marquées ^{13}C et ^{15}N ont été produites afin d'essayer de caractériser la protéine F-KpOmpA recombinante *in situ*, par des approches de RMN du solide cellulaire (Renault *et al.* 2012). La Figure 9 montre un ensemble de spectres de corrélations ^{13}C - ^{13}C DARR (Takegoshi *et al.* 2001, Takegoshi *et al.* 2003) enregistrés sur des préparations EC contenant (induction IPTG) ou non (pas d'induction) la protéine recombinante F-KpOmpA. Ces expériences de RMN du solide mettent en jeu des transferts d'aimantation via les couplages dipolaires et des temps de mélange courts, permettant ainsi de détecter majoritairement les corrélations intra-résiduelles de segments protéiques rigides. Afin de garantir la stabilité des préparations cellulaires au cours de l'acquisition des données, les spectres ont été enregistrés à une température de +2°C. De manière générale, les spectres obtenus sont de bonne qualité et de nombreuses corrélations ^{13}C - ^{13}C isolées sont observées (largeur de raie moyenne à mi-hauteur estimée à 0.8 ppm pour les signaux ^{13}C). La superposition des spectres des échantillons non-induits (Figure 9A) et induits (Figure 9B) permet de visualiser de nombreuses corrélations supplémentaires sur le spectre de l'échantillon induit, montrant que les signaux détectés proviennent majoritairement de la protéine F-KpOmpA surexprimée (Figure 9C). Ces résultats préliminaires ouvrent de nouvelles perspectives pour la caractérisation structurale et dynamique de la protéine F-KpOmpA fonctionnelle, en interaction avec le peptidoglycane. L'étude RMN de préparations cellulaires nécessite la mise en place de stratégies spécifiques permettant de faire face à la grande complexité spectrale et la faible sensibilité spectroscopique de ces échantillons. Plusieurs approches sont en cours de développement telles que l'utilisation de nouvelles constructions protéiques, de stratégies de marquage spécifiques et de méthodes RMN du solide couplant la détection des protons et la rotation à très haute vitesse, ou l'utilisation de la polarisation nucléaire dynamique (DNP).

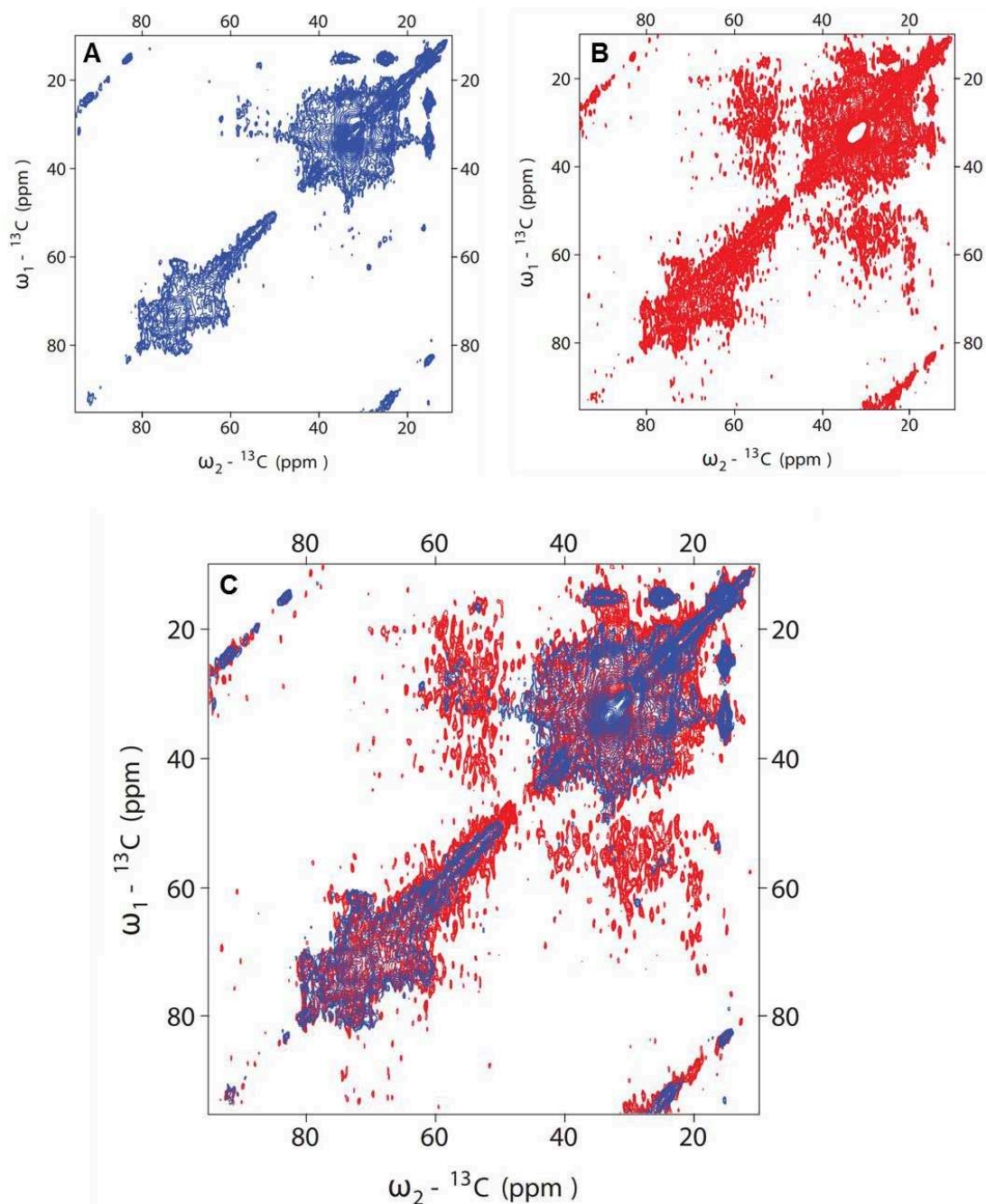


Figure 9. Spectres de corrélation 2D ^{13}C - ^{13}C DARR obtenus sur des préparations d'enveloppe cellulaires de ($\text{U-}^{13}\text{C}, ^{15}\text{N}$)- *E. coli* avec ou sans induction de KpOmpA. A. Echantillon non-induit. B. Echantillon Induit. C. Superposition des spectres des échantillons non-induit (A), en bleu, et induit (B), en rouge. L'ensemble des données RMN ont été obtenues sur un spectromètre Bruker Avance III HD opérant à une fréquence de Larmor ^1H de 700 MHz et équipé d'une sonde MAS E-Free 3.2 mm (Bruker BioSpin). Les expériences 2D ^{13}C - ^{13}C DARR ont été enregistrées à une fréquence de rotation de 12 KHz et à une température de $+2^\circ\text{C}$ en utilisant un temps de mélange de 50 ms et un temps d'acquisition de 38 heures.

IV. Références

- Bokoch M P, Zou Y, Rasmussen S G, Liu C W, Nygaard R, Rosenbaum D M, Fung J J, Choi H J, Thian F S, Kobilka T S, Puglisi J D, Weis W I, Pardo L, Prosser R S, Mueller L and Kobilka B K. *Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor*. Nature, **2010**, 463(7277): 108-112.
- Chen C Y, Cheng C H, Chen Y C, Lee J C, Chou S H, Huang W and Chuang W J. *Preparation of amino-acid-type selective isotope labeling of protein expressed in Pichia pastoris*. Proteins, **2006**, 62(1): 279-287.
- Clark L, Zahm J A, Ali R, Kukula M, Bian L, Patrie S M, Gardner K H, Rosen M K and Rosenbaum D M. *Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote Pichia pastoris*. J Biomol NMR, **2015**, 62(3): 239-245.
- Goto N K, Gardner K H, Mueller G A, Willis R C and Kay L E. *A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated N-15-, C-13-, H-2-labeled proteins*. Journal of Biomolecular Nmr, **1999**, 13(4): 369-374.
- Hebel D, Kirk K L, Cohen L A and Labroo V M. *First direct fluorination of tyrosine-containing biologically active peptides*. Tetrahedron Letters, **1990**, 31(5): 619-622.
- Jentoft J E, Jentoft N, Gerken T A and Dearborn D G. *¹³C NMR studies of ribonuclease A methylated with [¹³C]Formaldehyde*. J Biol Chem, **1979**, 254(11): 4366-4370.
- Klein-Seetharaman J, Getmanova E V, Loewen M C, Reeves P J and Khorana H G. *NMR spectroscopy in studies of light-induced structural changes in mammalian rhodopsin: applicability of solution (¹⁹F) NMR*. Proc Natl Acad Sci U S A, **1999**, 96(24): 13744-13749.
- Kosen P A. [5] *Spin labeling of proteins*. **1989**, *Methods in Enzymology*, Norman J O and Thomas L J, Academic Press, Volume 177: 86-121.
- Larda S T, Bokoch M P, Evanics F and Prosser R S. *Lysine methylation strategies for characterizing protein conformations by NMR*. J Biomol NMR, **2012**, 54(2): 199-209.
- Liu J J, Horst R, Katritch V, Stevens R C and Wuthrich K. *Biased signaling pathways in beta2-adrenergic receptor characterized by ¹⁹F-NMR*. Science, **2012**, 335(6072): 1106-1110.
- Mayer M and Meyer B. *Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor*. J Am Chem Soc, **2001**, 123(25): 6108-6117.
- Means G E and Feeney R E. *Reductive alkylation of amino groups in proteins*. Biochemistry, **1968**, 7(6): 2192-2201.
- Miyazawa-Onami M, Takeuchi K, Takano T, Sugiki T, Shimada I and Takahashi H. *Perdeuteration and methyl-selective H-1, C-13-labeling by using a Kluyveromyces lactis expression system*. Journal of Biomolecular Nmr, **2013**, 57(3): 297-304.
- Muchmore D C, McIntosh L P, Russell C B, Anderson D E and Dahlquist F W. *Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance*. Methods Enzymol, **1989**, 177: 44-73.

Nakamura K and Mizushima S. *Effects of heating in dodecyl sulfate solution on the conformation and electrophoretic mobility of isolated major outer membrane proteins from Escherichia coli K-12*. J Biochem, **1976**, 80(6): 1411-1422.

Ozawa K, Wu P S, Dixon N E and Otting G. *N-Labelled proteins by cell-free protein synthesis. Strategies for high-throughput NMR studies of proteins and protein-ligand complexes*. FEBS J, **2006**, 273(18): 4154-4159.

Park J S, Lee W C, Yeo K J, Ryu K S, Kumarasiri M, Hesek D, Lee M, Mobashery S, Song J H, Kim S I, Lee J C, Cheong C, Jeon Y H and Kim H Y. *Mechanism of anchoring of OmpA protein to the cell wall peptidoglycan of the gram-negative bacterial outer membrane*. FASEB J, **2012**, 26(1): 219-228.

Parker M J, Aulton-Jones M, Hounslow A M and Craven C J. *A combinatorial selective labeling method for the assignment of backbone amide NMR resonances*. J Am Chem Soc, **2004**, 126(16): 5020-5021.

Parsons L M, Lin F and Orban J. *Peptidoglycan recognition by Pal, an outer membrane lipoprotein*. Biochemistry, **2006**, 45(7): 2122-2128.

Religa T L, Ruschak A M, Rosenzweig R and Kay L E. *Site-directed methyl group labeling as an NMR probe of structure and dynamics in supramolecular protein systems: applications to the proteasome and to the ClpP protease*. J Am Chem Soc, **2011**, 133(23): 9063-9068.

Renault M, Tommassen-van Boxtel R, Bos M P, Post J A, Tommassen J and Baldus M. *Cellular solid-state nuclear magnetic resonance spectroscopy*. Proceedings of the National Academy of Sciences of the United States of America, **2012**, 109(13): 4863-4868.

Rypniewski W R, Holden H M and Rayment I. *Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: An x-ray analysis at 1.8-Å resolution*. Biochemistry, **1993**, 32(37): 9851-9858.

Sugiki T, Shimada I and Takahashi H. *Stable isotope labeling of protein by Kluyveromyces lactis for NMR study*. Journal of biomolecular NMR, **2008**, 42(3): 159-162.

Takegoshi K, Nakamura S and Terao T. *^{13}C – ^1H dipolar-assisted rotational resonance in magic-angle spinning NMR*. Chemical Physics Letters, **2001**, 344(5–6): 631-637.

Takegoshi K, Nakamura S and Terao T. *^{13}C – ^1H dipolar-driven ^{13}C – ^{13}C recoupling without ^{13}C rf irradiation in nuclear magnetic resonance of rotating solids*. The Journal of Chemical Physics, **2003**, 118(5): 2325-2341.

Tugarinov V, Hwang P M, Ollerenshaw J E and Kay L E. *Cross-correlated relaxation enhanced ^1H [bond] ^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes*. J Am Chem Soc, **2003**, 125(34): 10420-10428.

Tugarinov V, Sprangers R and Kay L E. *Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy*. Journal of the American Chemical Society, **2007**, 129(6): 1743-1750.

Vollmer W, Blanot D and de Pedro M A. *Peptidoglycan structure and architecture*. FEMS Microbiol Rev, **2008**, 32(2): 149-167.

Williamson M P. *Using chemical shift perturbation to characterise ligand binding*. Prog Nucl Magn Reson Spectrosc, **2013**, 73: 1-16.

Auteur: Guillaume NARS

Titre: Dynamique fonctionnelle des protéines: études d'une lipase et d'une protéine A de membrane externe de bactérie.

Directeurs de thèse: Alain MILON et Isabelle ANDRE

Laboratoires:

- IPBS, Université de Toulouse UPS – CNRS UMR 5089, 205 route de Narbonne, 31077Toulouse (France)
- LISBP, UMR INSA – CNRS 5504 – INRA 792, 135 avenue de Ranguel, 31077 Toulouse (France)

Discipline: Biologie structurale

Date et lieu de soutenance: 29 septembre 2015, Toulouse, France

Résumé:

La compréhension de la fonction des protéines et des systèmes biologiques passe par une connaissance fine des mécanismes moléculaires sous-jacents. La cristallographie et la résonance magnétique nucléaire permettent d'appréhender ces mécanismes au niveau atomique en fournissant des informations sur la structure et sur la dynamique des macromolécules biologiques. Nous nous sommes ainsi intéressés à deux protéines, la lipase lip2 de la levure *Yarrowia lipolytica* et la protéine membranaire OmpA de la bactérie *Klebsiella pneumoniae*. Nous avons recherché des conditions d'expression de la protéine lip2 marquée uniformément ou spécifiquement sur une boucle (appelée « lid ») afin d'en étudier la dynamique. Des conditions de marquage uniforme à l'azote 15 de lip2 recombinante dans *Yarrowia lipolytica* ont été mises au point, mais le marquage acide aminé spécifique n'a pu être réalisé à cause de phénomènes de dilution isotopique trop importants dans cette levure. Nous avons résolu par cristallographie aux rayons X la structure du domaine C-terminal de la protéine OmpA et étudié sa dynamique en solution par RMN (techniques de relaxation ¹⁵N). Nous avons caractérisé la dynamique de son domaine N-terminal membranaire reconstitué en liposomes par RMN du solide : en utilisant la rotation à l'angle magique à 60kHz et à la détection ¹H sur un spectromètre 1 GHz, nous avons pu attribuer une majorité des résonances du tonneau β et établir un profil de paramètre d'ordre des vecteurs NH. Des expériences de protéolyse ménagée ont révélé par ailleurs un site de coupure unique à la trypsine au sein de la boucle extracellulaire L3. Enfin, une première caractérisation de la protéine complète exprimée dans la membrane externe d'*Escherichia coli* a été entreprise par RMN du solide sur membranes externes natives.

Title: Protein functional dynamics: studies of a lipase and a bacterial outer membrane protein A

Abstract:

Understanding the function of proteins and biological systems requires an accurate knowledge of the underlying molecular mechanisms. Crystallography and nuclear magnetic resonance provide a detailed description of these mechanisms, with an atomic resolution, by providing data on both structures and motions. We investigated two proteins, the lip2 lipase from the yeast *Yarrowia lipolytica* and the membrane protein OmpA from the bacteria *Klebsiella pneumoniae*. We tried to produce lip2 with uniform and amino-acid specific stable isotope labelling on its functional loop (the lid) for NMR experiments. The homologous recombinant expression in *Yarrowia lipolytica* turned out to be the most efficient for uniform labelling but failed for specific labelling due to extensive isotope scrambling. We solved the structure of OmpA C-terminal domain by X-ray crystallography, and analyzed its dynamics in solution by NMR (¹⁵N relaxation techniques). We characterized its transmembrane N-terminal domain in proteoliposomes by solid state NMR: using state of the art ultra-fast MAS (60 kHz), ¹H detection and a 1 GHz spectrometer, we could assign most β-barrel resonances and establish a NH order parameter profile. In a complementary approach, we used proteolysis to reveal a unique trypsin cleavage site on the extracellular loop 3. Finally, a first characterization of the full-length protein expressed in the outer membrane of *Escherichia coli* was initiated by solid state NMR on intact outer membranes.

Keywords: *Yarrowia lipolytica* lipase, *Klebsiella pneumoniae* OmpA, stable isotope labelling of recombinant proteins, liquid and solid state NMR, crystallography, relaxation, proteolysis.