



HAL
open science

Analyse multiomique des peptides d'extraits de levure et de leurs impacts fonctionnels sur *Streptococcus thermophilus*

Lucas Proust

► **To cite this version:**

Lucas Proust. Analyse multiomique des peptides d'extraits de levure et de leurs impacts fonctionnels sur *Streptococcus thermophilus*. Microbiologie et Parasitologie. Université Paris Saclay (COMUE), 2018. Français. <NNT : 2018SACLA032>. <tel-04416379>

HAL Id: tel-04416379

<https://theses.hal.science/tel-04416379v1>

Submitted on 25 Jan 2024

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

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



HAL Authorization

Analyse multiomique des peptides d'extraits de levure et de leurs impacts fonctionnels sur *Streptococcus thermophilus*

Thèse de doctorat de l'Université Paris-Saclay
préparée à AgroParisTech, l'Institut des sciences et industries du
vivant et de l'environnement

École doctorale n°581 : agriculture, alimentation, biologie,
environnement et santé (ABIES)
Spécialité de doctorat: Microbiologie

Thèse présentée et soutenue à Paris, le 5 décembre 2018, par

M. Lucas Proust

Thèse confidentielle

Composition du Jury :

Mme Catherine Béal Professeur, AgroParisTech, Grignon	Présidente
M. Ivan Mijakovic Professeur, Chalmers University of Technology, Göteborg	Rapporteur
M. Eric Guédon Directeur de Recherche, INRA, STLO, Rennes	Rapporteur
M. Eric Oriol Directeur Lesaffre Puri Center, Lesaffre, Loos	Examineur
M. Sébastien Nouaille Chargé de recherche, INRA, LISBP, Toulouse	Examineur
M. Vincent Juillard Directeur de Recherche, INRA, Micalis, Jouy-en-Josas	Directeur de thèse
M. Alain Sourabié Responsable Innovation, Procelys, Maisons-Alfort	Invité

Sommaire

Remerciements	i
Abbréviations	iii
Liste des Figures	v
Liste des Tableaux	ix
Introduction	1
1 Etude bibliographique : bactéries lactiques et peptides	7
1.1 Introduction	9
1.2 Peptidome extracellulaire en milieu de production	11
1.3 Principales fonctions assurées par les peptides	36
1.4 Conclusion	77
2 Séquençage et analyses génomiques de <i>Streptococcus thermophilus</i> N4L	79
2.1 Introduction	80
2.2 Complete Genome Sequence of the Industrial Fast-Acidifying Strain <i>Streptococcus thermophilus</i> N4L	80
2.3 Analyses génomiques complémentaires	83
2.4 Conclusion	89
3 Caractérisation des dynamiques d'utilisation de peptides d'extraits de levure par <i>Streptococcus thermophilus</i> N4L	91
3.1 Introduction	93
3.2 Approche quantitative : mesure de la consommation d'azote aminé par RP-HPLC	96
3.3 Approche qualitative : analyse peptidomique par spectrométrie de masse	110
3.4 Conclusion	173
4 Etude des variations d'expression génétique et de synthèse protéique en réponse à la composition des extraits de levure	175

Sommaire

4.1	Introduction	176
4.2	Matériel et méthodes	178
4.3	Résultats	184
4.4	Conclusion	205
	Discussion générale et perspectives	209
	Références	219
	Annexes	249
	Annexe A - Impacts différentiels des extraits de levure sur les fonctions et voies métaboliques de <i>S. thermophilus</i> N4L	250
	Annexe B - Résultats complémentaires de l'intégration multiomique <i>via</i> l'analyse DIABLO	270
	Annexe C - Valorisation du travail de thèse	273
	Annexe D - Formations	274

Remerciements

L'usage veut que les remerciements d'une thèse soient un espace de liberté où le presque-docteur peut enfin sortir sa plus belle plume, du moins ce qu'il en reste, et arroser copieusement ses collègues, amis et famille comme il lui plaît. Il est vrai que le ton et la forme varient d'une thèse à l'autre, allant du pur exercice diplomatique jusqu'aux débordements les plus fantasques. Néanmoins, malgré cette apparente liberté, on retrouvera quasi systématiquement dans cette prose une articulation en deux parties. La première, « sérieuse », visera à remercier encadrants, responsables hiérarchiques et jury car on ne saurait bien entendu s'y soustraire. Ce n'est que dans la deuxième partie que l'on s'occupera du reste des collègues et des proches à coup de traits d'esprit bien sentis, petits souvenirs touchants et autres bons sentiments (pour ne pas dire niaiseries). Tout ceci, je dois le reconnaître, m'ennuie. Or, j'ai cru percevoir ces dernières semaines quelques attentes dans mon entourage concernant mes propres remerciements. Pour ne point froisser les susceptibilités, ce n'est pas mon genre, tout en restant fidèle à mes valeurs, je vais proposer à la place un petit jeu qui mettra le lecteur à contribution s'il le souhaite. Ce n'est après tout que justice après le temps et l'énergie que mon directeur de thèse, l'excellent Vincent Juillard, et moi-même avons passé à rédiger et corriger ce document, sans compter les lectures attentives des membres de mon jury. J'ai donc caché dans la grille suivante les noms de toutes les personnes auxquelles je veux effectivement adresser mes plus sincères et chaleureux remerciements, ma reconnaissance, ou même, tout simplement, une petite pensée amicale. Vous qui m'avez côtoyé durant ma thèse, saurez-vous retrouver votre propre nom? Attention ce n'est pas facile, et je ne dévoilerai pas la solution! Si vous n'y arrivez pas, on va dire qu'il s'agit d'un malheureux oubli de ma part :)

Remerciements

K A K H T N S F Y I H B Z S N H B D X R R I L A G A M K H A
D B C Z Y U F Y O U N A R E Z E F O J C J T B S C Y K M I N
Z R W S G P M L J U P J A V B N F G L C L W H S Y F E M Z K
W V X R I R E X M E J S Q R D S E A J A K W B P K P J G U U
E P V U X B M D T N Y O U F O C E I N T Q Z O E Y X H Q R U
W D S I Y P S O Q I R G F J A N F Z T R I R C J O E P F O E
U F L V I L N I W T O J W W N E P N N S A V I I P I D W S R
N T M L Y O T Q R S Z K Y A Y W A D G E A B A L Y L E W H I
A N W D V T D I D I E R K B C P E E O U F B A V C A A E Q C
I N I P G C X E R R N D E W U R V F E C Y E E L S R P J Q T
J E G D P E R X C H N R Y R H C S G Y S S U N S C O C U R V
X Z P V G G N Y G C T R I L R T W W Q I I O X Q Q C C U O I
Z O Z O F G O P F R M G R H L Z I E L N C O Y W Y J G U T S
J U V O J R U D A X C E X I J I C E Q K F B C W S D W D V N
E F I J Y U X N D Y O N M C R V N K G Z Z M Y N Z U X W T X
F X L W U D D O J C T A W K I I E L I P U F L R A W B U E E
J W S H H Y E J K D S H U X R F F R G E L E Q O N R L O K N
J N A C I X C Y J M K P V E H I S E O O C M B Z O E F V M I
C Q C J S U X M A R I E H S V S I L R N U I Q S D D Y H A L
L K N R M F O R P B V T A V G L H E E E I B Y V X W D G X O
B R W J U E T A G G A S T S I C N M U Z M Q S Y L I L P B D
V D E K K I K M D C S N N M Z T E N I R A M U Y S P E Q M N
S U Q Z N E M E W R E O E K L L M S I S V N A E W F D A Z E
K Z O S T X E M P C M T H Y C Y O R N M Y L E N E A I N A W
G N E E P M N O N I Q O H G M R S Z A S P C Z S W X T F H G
O P J I A G A I S N A R D N A S I I M E Q N E A I X H L B Z
W S C A H I V F Y U K U D I T Q S Q A E V Z E L T X D V Q J
Y T Y B U P I Q E K G J E R O M E I M N S Q I B I D E N Y W
E Z S F K W O H F Y F Z F F W G Y F E X E V Q E Y N V L R F
Z Z S Z O G C S R G Z H E B Z J F M E K W G M Q Y F E Q A I

Abbreviations

ADN	Acide DésoxyriboNucléique
ARN	Acide RiboNucléique
ATP	Adénosine TriPhosphate
BCAA	Branched-Chain Amino Acid
BER	Balanced Error Rate
CDS	Coding DNA Sequence
CHO	Chinese Hamster Ovary
CID	Collision Induced Dissociation
DDBJ	DNA DataBank of Japan
DO	Densité Optique
ECF	Energy-Coupling Factor
EMBL	European Molecular Biology Laboratory
ENA	European Nucleotide Archive
EXL	Extrait De Levure
FC	Fold Change
FDR	False Discovery Rate
GAF	cGMP-stimulated phosphodiesterases, Adenylate cyclases, FhlA
GMP	Guanosine MonoPhosphate
GO	Gene Ontology
GTP	Guanosine TriPhosphate
HPLC	High Performance Liquid Chromatography
HSP	Heat Shock Protein
IMP	Inosine MonoPhosphate
KEGG	Kyoto Encyclopedia of Genes and Genomes
MCD	Milieu Chimiquement Défini
MFS	Major Facilitator Superfamily
MS	Mass Spectrometry
NBD	Nucleotide-Binding Domain
NCBI	National Center for Biotechnology Information
NRP	NonRibosomal Peptide

Abbréviations

ORF	Open Reading Frame
PDB	Protein Data Bank
PLS	Partial Least Squares/Projection to Latent Structure
PMF	Proton Motive Force
POT	Proton-dependent Oligopeptide Transporter
PTR	Peptide TRansport
QSAR	Quantitative StructureActivity Relationship
RMN	Résonance Magnétique Nucléaire
SAM	S-AdénosylMéthionine
SBD	Substrate-Binding Domain
SDS-PAGE	Sodium Dodecyl SulfatePolyAcrylamide Gel Electrophoresis
SHP	Short Hydrophobic Peptide
SPE	Solid Phase Extraction
TCS	Two Component System
TMD	TransMembrane Domain
TPR	TetratricoPeptide Repeat
TRAP	Tripartite ATP-independent Periplasmic
TTT	Tripartite Tricarboxylate Transporter
UFC	Unité Formant Colonie
wHTH	Winged Helix-Turn-Helix
XIC	eXtraction of Ion Currents

Liste des Figures

1	Stratégie scientifique générale du projet de thèse.	5
1.1	Procédé de fabrication des extraits de levure	13
1.2	Composition en acides aminés libres et engagés dans des liaisons peptidiques d'un extrait de levure	17
1.3	Exemples de conformations peptidiques possibles déterminées sous contraintes RMN	23
1.4	Principales enzymes protéolytiques caractérisées chez les bactéries lactiques et impliquées dans l'élaboration du peptidome bactérien extracellulaire	28
1.5	Recensement des peptides disponibles dans BactPepDB en fonction de leur taille	31
1.6	Représentation simplifiée des voies d'utilisation des acides aminés libres et peptides chez <i>L. lactis</i>	37
1.7	Structure canonique des protéines de transport de la superfamille MFS	40
1.8	Cycle de fonctionnement d'un transporteur de di-tripeptides de la famille PTR	41
1.9	Architecture modèle des ABC-transporteurs	43
1.10	Mécanismes d'import des ABC-transporteurs de type I	45
1.11	Organisation génétique des ABC-transporteurs de peptides de bactéries lac- tiques	47
1.12	Représentation schématique des interactions entre des résidus conservés de diverses SBPs structurellement caractérisées et les extrémités N- et C-terminales des peptides transportés	50
1.13	Illustration de la cavité de fixation d'OppA occupée par SLSQLSSQS et SLS- QSLSQS	51
1.14	Représentation schématique des flux d'utilisation des acides aminés libres et peptides chez <i>L. lactis</i> IL1403	56
1.15	Structure générale de CodY	61
1.16	Représentation schématique des deux voies de signalisation utilisés par les bactéries à Gram positif dans les mécanismes de <i>quorum sensing</i>	63
1.17	Processus de biosynthèse de la subtiline	67
1.18	Processus de maturation de cCF10 d' <i>E. faecalis</i>	68

1.19 Exemples de voies de signallement impliquant des phéromones associées aux régulateurs RRNPP	69
2.1 Arbre phylogénétique des 50 souches de <i>S. thermophilus</i> disponibles dans la base de données GenBank du NCBI	84
2.2 Répartition des principales catégories fonctionnelles telles qu'identifiées par l'annotation KEGG du génome de <i>S. thermophilus</i> N4L	86
2.3 Distribution des tailles de protéines et courts ORFs chez <i>S. thermophilus</i> N4L	88
3.1 Démarche expérimentale mise en œuvre afin d'effectuer un suivi du contenu peptidique lors de la croissance de <i>S. thermophilus</i> N4L	95
3.2 Cinétique de croissance <i>S. thermophilus</i> N4L mesurée par densité optique à 600 nm et par dénombrement sur boîtes pendant sa culture en bioréacteurs	100
3.3 Evolution au cours du temps du contenu en acides aminés libres dans les EXLs	101
3.4 Evolution de la consommation totale et par classe biochimique des acides aminés libres pendant la croissance de <i>S. thermophilus</i> N4L	102
3.5 Evolution de la composition en acides aminés des peptides de courte taille	103
3.6 Evolution de la consommation totale et par classe biochimique des acides aminés constituant les peptides de courte taille	104
3.7 Evolution de la composition en acides aminés des oligopeptides	105
3.8 Evolution de la consommation totale et par classe biochimique des acides aminés constituant les oligopeptides	106
3.9 Evolution du contenu d'azote aminé total et par origine pendant la croissance de <i>S. thermophilus</i> N4L	107
3.10 Protocole de peptidomique initial hérité de l'analyse de peptidomes bactériens	113
3.11 Principe de l'analyse de l'empreinte de masse par spectrométrie de masse en tandem	114
3.12 Protocole de peptidomique final	116
3.13 Principaux résultats obtenus lors l'analyse de répétabilité du protocole de peptidomique	118
3.14 Caractéristiques physico-chimiques des peptides identifiés lors de l'analyse de répétabilité	119
3.15 ACP sur l'ensemble des peptides caractérisés cinétiquement dans les deux EXLs	161
3.16 Stratégie déployée pour l'implémentation d'un modèle prédictif à partir des données de peptidomique	163
3.17 Elaboration du modèle sPLS-DA complet	166
3.18 Application du modèle sPLS-DA complet	167
3.19 Principaux résultats de l'analyse de répétabilité du protocole de peptidomique amélioré	171

3.20 Propriétés physico-chimiques des peptides identifiés durant l'analyse de répé- tabilité du protocole de peptidomique amélioré	172
4.1 Démarche expérimentale mise en œuvre lors des analyses transcriptomiques et protéomiques globales pendant de la croissance de <i>S. thermophilus</i> N4L . . .	177
4.2 Résultats généraux du séquençage du transcriptome de <i>S. thermophilus</i> N4L . .	184
4.3 Résultats généraux de l'analyse protéomique	185
4.4 ACP basées sur l'abondance des transcrits et protéines	186
4.5 <i>Heatmaps</i> représentant l'abondance des transcrits et protéines montrant des variations significatives entre les deux EXLs	188
4.6 Voies métaboliques et fonctions cellulaires surreprésentées parmi les gènes et protéines significativement affectés par la composition des EXLs	191
4.7 Gènes impliqués dans la biosynthèse d'acides aminés dont l'expression glo- bale est différenciellement affectée entre les EXLs	192
4.8 Protéines impliquées dans la biosynthèse d'acides aminés dont l'expression globale est différenciellement affectée entre les EXLs	193
4.9 Différences d'activité transcriptionnelle au niveau de la voie de biosynthèse des acides aminés branchés entre les deux EXLs	194
4.10 Différences d'abondance protéique au niveau de la voie de biosynthèse des acides aminés branchés entre les deux EXLs	195
4.11 Différences d'activité transcriptionnelle et d'abondance protéique au niveau des transporteurs d'acides aminés branchés et d'oligopeptides.	196
4.12 Expression de <i>codY</i> au cours du temps	196
4.13 Mécanismes de <i>quorum-sensing</i> gouvernant la production du streptide	197
4.14 Expression du locus streptide au cours du temps	199
4.15 Résultats de l'intégration multi-omique obtenue avec DIABLO	201
4.16 Nature des liens de corrélation entre gènes et protéines sélectionnés par l'ana- lyse DIABLO	203
4.17 Fonctions associées à la signature moléculaire des EXLs	204

Liste des Tableaux

1.1	Exemple de composition nutritionnelle d'un extrait de levure	12
1.2	Protéases vacuolaires de <i>S. cerevisiae</i>	15
1.3	Efficacité catalytique de quatre protéases commerciales pour la formation de peptides et d'acides aminés libres à partir de protéines de pomme de terre . . .	16
1.4	Principales caractéristiques d'hydrolysats de levure commerciaux	19
1.5	Nombre de peptides putatifs de 10 à 80 acides aminés recensés chez les principales bactéries lactiques	33
1.6	Classification structurale des SBPs et corrélation avec la nature du substrat transporté	44
1.7	Distribution des peptidases chez les bactéries lactiques	54
1.8	Systèmes de transports d'acides aminés libres identifiés chez les bactéries lactiques	55
1.9	Familles des peptides phéromones impliqués dans la voie de détection extracellulaire des systèmes de <i>quorum sensing</i> des bactéries à Gram positif . .	66
1.10	Familles des peptides phéromones impliqués dans la voie de détection intracellulaire des systèmes de <i>quorum sensing</i> des bactéries à Gram positif	72
2.1	Principales caractéristiques des annotations génomiques réalisées par la plateforme Agmial et l'outil PGAP du NCBI	83
2.2	Localisation cellulaire des protéines de <i>S. thermophilus</i> N4L prédite par Loca-teP v2	87
2.3	Phéromones associées à des régulateurs Rgg/ComR identifiées dans le génome de <i>S. thermophilus</i> N4L	89
3.1	Plan expérimental mis en œuvre pour la quantification du contenu azoté des EXLs	99
4.1	Nombre de transcrits et de protéines différentiellement exprimés	187

Introduction

Les bactéries lactiques constituent un groupe de microorganismes dont l'utilité technologique est connue depuis des millénaires. Traditionnellement associées à la fermentation des denrées alimentaires, leur usage se diversifie aujourd'hui et trouve de nombreuses applications dans l'industrie pharmaceutique, cosmétique et chimique. Toutefois, l'intérêt des bactéries lactiques traditionnelles utilisées en agroalimentaire reste considérable, en particulier dans le secteur laitier où les produits qui en sont dérivés (fromages, yaourts, laits fermentés) permettent de tirer la croissance d'un marché arrivé par ailleurs à maturité dans la plupart des pays développés (FAO, 2017).

La fabrication des ferments lactiques s'est largement industrialisée depuis la deuxième moitié du XX^{ème} siècle. Des souches sélectionnées sont cultivées à large échelle par fermentation puis concentrées, stabilisées et conditionnées avant d'être commercialisées. A l'issue d'un tel procédé industriel, les fabricants de ferments doivent être en mesure de fournir une biomasse microbienne viable en quantité suffisante et présentant un niveau métabolique optimal lui permettant d'assurer ses différentes fonctionnalités technologiques. Celles-ci sont nombreuses et dépendent de l'usage du ferment considéré, bien que le premier critère de choix reste avant tout l'activité acidifiante du ferment. Divers facteurs sont susceptibles d'influencer les rendements de production et les fonctionnalités du ferment durant l'ensemble du procédé de fabrication, allant du choix de la souche bactérienne jusqu'au type de conditionnement final. Les industriels cherchent donc à maîtriser chaque source de variabilité en standardisant leurs procédés afin de fournir de manière répétable un produit de qualité optimale. Dans cette optique, une attention toute particulière doit être portée à la composition du milieu de culture employé pour la croissance bactérienne. Celui-ci, en plus de satisfaire les exigences nutritionnelles du microorganisme cultivé, va également moduler son métabolisme et son état physiologique général. Cette incidence a été particulièrement étudiée dans un contexte de résistance aux stress subis lors des étapes dites *downstream* (concentration, stabilisation etc.) et de viabilité cellulaire en résultant, ce paramètre étant de première importance pour le maintien des performances technologiques du ferment.

Le choix du milieu de culture et des substrats le composant doivent donc être minutieusement raisonnés. La production industrielle des ferments lactiques utilise généralement

des milieux de culture dits semi-définis dans lesquels une partie des nutriments est apportée par des matières premières complexes dont l'exacte composition est inconnue. Les hydrolysats cellulaires et protéiques (peptones diverses, extraits de viande, extraits de levure etc.) sont des ingrédients particulièrement appréciés car ils permettent de réduire les coûts de production par rapport à des milieux chimiquement définis tout en assurant des rendements élevés grâce à leur richesse nutritionnelle. Ces substrats permettent en effet de fournir une large gamme de composés (sources carbonées, nucléotides, vitamines, minéraux), mais sont particulièrement utilisés comme sources d'azote aminé en raison de leur haute teneur en acides aminés libres et en peptides. Toutefois, leur nature complexe rend difficile leur caractérisation et la mise au point de milieux spécifiquement adaptés au type de microorganisme cultivé. Les bactéries lactiques étant nutritionnellement exigeantes, et, du fait que les milieux de culture sont susceptibles d'influencer les fonctionnalités du ferment produit, la fine caractérisation de ces substrats complexes ainsi que la compréhension des mécanismes par lesquels ils affectent la physiologie bactérienne constituent des enjeux majeurs pour les industriels.

Forts de ce constat, le fabricant de levure Lesaffre, *via* sa Business Unit Procelys dédiée à la commercialisation d'extraits de levure (EXLs), ainsi que le producteur de ferments lactiques Sacco ont décidé de s'associer dans un projet de recherche qui implique également l'UMR Microbiologie de l'Alimentation au service de la Santé (UMR Micalis) de l'INRA de Jouy-en-Josas. L'objectif de ce projet est d'étudier l'influence de la composition d'EXLs sur le comportement d'une bactérie lactique d'intérêt technologique majeur : *Streptococcus thermophilus*. Comme expliqué précédemment, la composition du milieu de culture est d'une grande importance pour les performances de production. De nombreuses observations empiriques à l'échelle industrielle ont ainsi confirmé que différents types d'EXLs ne conduisaient pas systématiquement aux mêmes résultats en termes de croissance bactérienne et de fonctionnalités technologiques, alors même qu'ils sont utilisés dans des milieux couvrant théoriquement l'ensemble des besoins nutritionnels de la souche cultivée.

Pour ce projet, il a été décidé d'aborder cette problématique en s'intéressant plus spécifiquement à la fraction peptidique des EXLs. Les peptides sont un des composés majoritaires de ces substrats, si ce n'est le plus abondant. Ils correspondent aux produits de dégradation des protéines de levure survenant lors de l'étape d'autolyse. L'hypothèse ayant motivé ce projet est que les différences de performances observées en production industrielle pourraient être en partie liées à cette fraction peptidique. Plus précisément, les peptides des EXLs pourraient en réalité avoir un double rôle en étant à la fois molécules nutritionnelles mais aussi régulatrices, c'est-à-dire ayant un effet ciblé sur le métabolisme bactérien. Cette hypothèse a été élaborée en raison du fait bien connu que les bactéries à Gram positif fabriquent elles-mêmes des peptides leur servant de molécules de signalisation, ou phéromones, coordonnant l'expression de fonctions cellulaires à l'échelle de la population à travers des méca-

nismes dits de *quorum sensing* (Monnet *et al.*, 2014). Etant donnée la quantité et la diversité des peptides présents dans les EXLs, il est envisageable que certains d'entre eux puissent soit avoir des effets similaires à des phéromones bactériennes, soit au contraire interférer avec les systèmes de communication de la bactérie. Ces gènes ainsi affectés, en fonction du ou des métabolismes auxquels ils participent, pourraient influencer la croissance et les fonctionnalités technologiques du ferment produit. En outre, il est également connu que les peptides participent indirectement à la régulation du métabolisme bactérien à travers certains de leurs acides aminés constitutifs qui, après hydrolyse intracellulaire, vont servir d'indicateurs de l'environnement nutritionnel pour la bactérie qui ajuste son métabolisme général en conséquence.

Ainsi, l'exploration de la diversité peptidique des EXLs représente un réel enjeu pour mieux comprendre les différentes dynamiques - tant nutritionnelles que régulatrices - pouvant avoir un impact sur la croissance et les aptitudes technologiques des ferments lactiques. Malheureusement, l'identification et la caractérisation des peptides en mélange complexe sont encore aujourd'hui techniquement difficiles. Les EXLs sont ainsi considérés comme étant principalement un mélange mal défini de divers peptides inconnus. De nombreuses questions restent ainsi encore en suspens concernant la fraction peptidique des EXLs (liste non exhaustive) :

- Quels sont les peptides en mélange ?
- Quelles sont leurs concentrations respectives ?
- Quelle est leur valeur nutritionnelle globale par rapport aux acides aminés libres ?
- La bactérie utilise-t-elle indifféremment l'ensemble des peptides de l'EXL, ou en assimile-t-elle préférentiellement certains pendant sa croissance ?
- Si oui, lesquels et sur quel(s) critère(s) ?
- Assimile-t-elle les mêmes types de peptides dans différents EXLs ?
- Les peptides assimilés sont-ils ceux satisfaisant le mieux les besoins nutritionnels de la souche cultivée ?
- Différents EXLs, utilisés par ailleurs dans des conditions similaires, produisent-ils des effets marqués d'un point de vue métabolique (expression génétique, synthèse protéique, production de métabolites) ?
- Si oui, quel est le rôle des peptides dans ces différences par rapport aux autres composés des EXLs ?
- Ces différences peuvent-elles être à l'origine de variations de performances technologiques entre EXLs ?

Les réponses à ces questions portent divers intérêts. En premier lieu, d'un point de vue académique, si la culture des bactéries lactiques est particulièrement bien décrite dans le lait, elle a été significativement moins étudiée dans des milieux complexes à base d'hydrolysats. Ces milieux peuvent pourtant être considérés comme des substrats naturels du fait de leur

large utilisation, tant en laboratoire qu'en industrie. Le rôle et les fonctions des peptides en lien avec les bactéries lactiques constituent également un champ disciplinaire à part entière dont il reste de nombreuses zones à explorer. De plus, on peut également relever le fait que la plupart des travaux de recherche sur les bactéries lactiques ont été réalisés avec l'espèce modèle *Lactococcus lactis*. Si celle-ci constitue bien le premier ferment lactique d'un point de vue commercial, *S. thermophilus* arrive néanmoins en deuxième position (Hols *et al.*, 2005). Or, il y a comparativement près de trois fois moins de publications associées à cette dernière bactérie. De nombreuses connaissances restent donc à valider ou acquérir chez elle. Enfin, sur un plan industriel, la meilleure caractérisation des peptides contenus dans les EXLs, la connaissance de leur utilisation par *S. thermophilus*, leurs effets sur son métabolisme et ses fonctionnalités technologiques représentent des informations importantes permettant potentiellement de proposer des pistes d'améliorations pour l'ensemble du procédé industriel, de la conception de l'EXL jusqu'à l'utilisation du ferment. A titre d'exemple, l'élaboration d'un EXL dont la fraction peptidique serait simplifiée constituerait un objectif souhaitable. Cette fraction ne contiendrait que des peptides à haute valeur nutritionnelle et permettrait également d'orienter à façon le métabolisme de la souche vers des performances technologiques élevées. Toutefois, avant d'arriver à ce résultat, de nombreuses connaissances restent à acquérir. C'est donc dans ce contexte particulier qu'intervient ce projet de thèse.

Pour tenter de répondre à un maximum de ces questions, une démarche scientifique relativement simple a été élaborée (Figure 1). Celle-ci se veut largement exploratoire et est basée sur la combinaison de différentes approches globales. Deux EXLs commerciaux différents fournis par Procelys ont été utilisés pour cultiver une souche industrielle de *S. thermophilus* fournie par l'entreprise Sacco. Ces cultures ont été conduites en bioréacteurs dans des conditions semi-industrielles et durant lesquelles des prélèvements réguliers des moûts de fermentation ont été effectués. A partir de ces échantillons, après séparation du surnageant de la biomasse, quatre types d'analyses globales ont été réalisées sur certains points d'intérêt. Tout d'abord, les surnageants recueillis ont servi à caractériser qualitativement le peptidome extracellulaire ainsi que son évolution au cours de la croissance bactérienne. Le peptidome correspond aux peptides en mélange, provenant ici des EXLs utilisés. Leur caractérisation a nécessité l'emploi d'une approche analytique originale basée sur la spectrométrie de masse et intégrant des traitements bioinformatiques et statistiques spécifiquement développés au cours de ce projet pour identifier les dynamiques d'utilisation des peptides. Toutefois, cette approche étant seulement semi-quantitative, une approche complémentaire basée sur l'analyse par chromatographie liquide haute performance (HPLC) a été employée afin de mesurer quantitativement les dynamiques d'utilisation des peptides mais aussi des acides aminés libres. Enfin, des approches de transcriptomique à l'échelle du génome (RNAseq) et de protéomique ont également été mises en œuvre afin d'évaluer l'impact métabolique et physiologique des EXLs sur les cellules en croissance. L'ensemble de ce dispositif expérimental

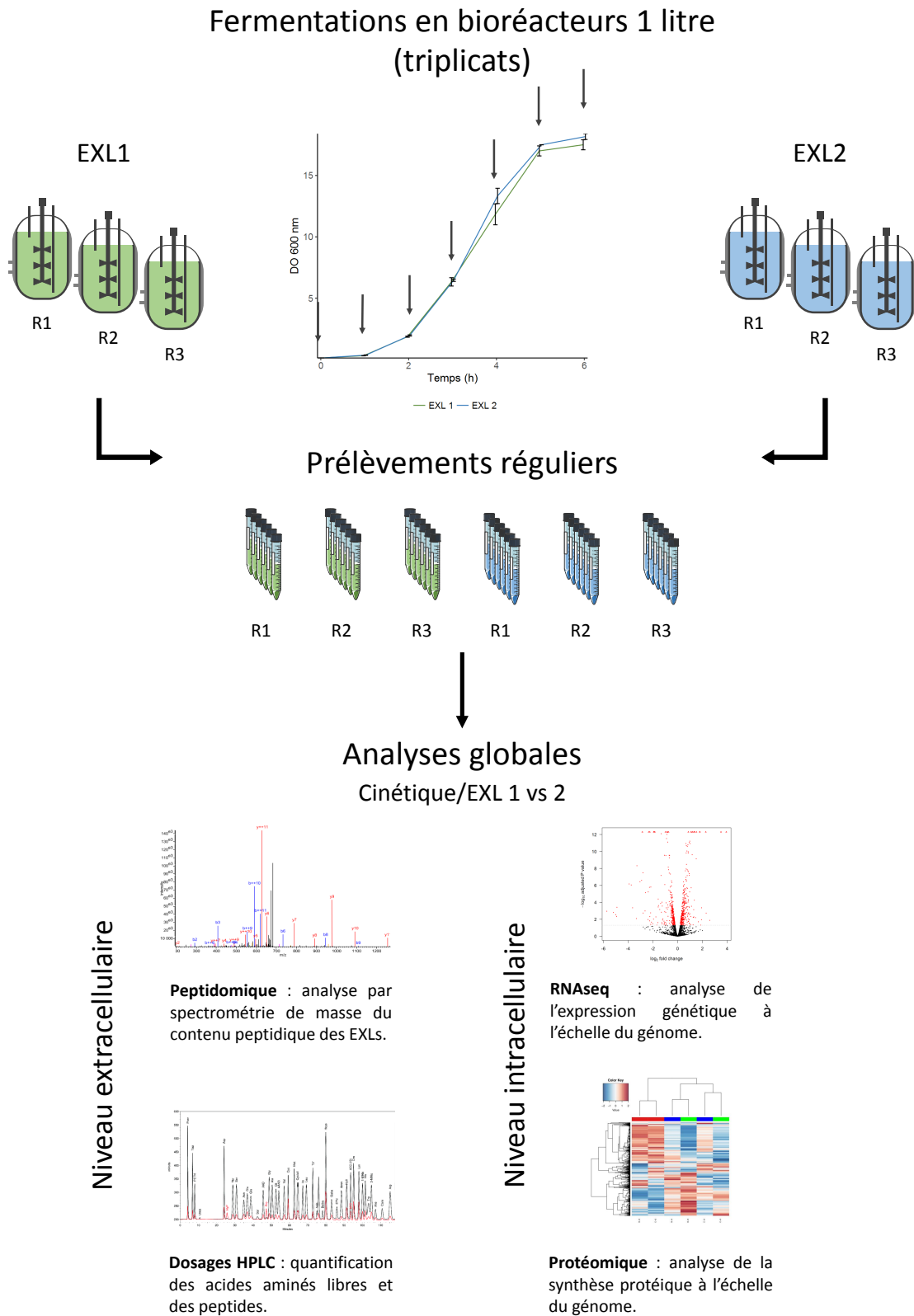


FIGURE 1 : Stratégie scientifique générale du projet de thèse.

tal permet ainsi d'appliquer ces différentes approches à la fois sur un plan cinétique mais aussi comparatif entre les deux EXLs. A côté de cette stratégie ayant articulé l'ensemble de la thèse, d'autres techniques ont également été déployées telles que la création de souches mutantes pour répondre à des questions plus spécifiques et valider des hypothèses.

Ce mémoire a donc pour rôle d'effectuer la synthèse de l'ensemble du travail effectué durant ce projet de thèse. Il est articulé en cinq parties : une étude bibliographique préalable pour contextualiser le sujet de thèse suivie de trois chapitres présentant les résultats obtenus et enfin une discussion générale. L'étude bibliographique a pour objet l'étude de l'utilisation des peptides par les bactéries lactiques dans un contexte de culture en milieu complexe de type EXL. L'accent sera mis en particulier sur les deux fonctions exercées par les peptides qui constituent le cœur de cette thèse : la nutrition et la régulation. La présentation des résultats commencera ensuite par un court chapitre présentant le travail de séquençage et d'analyse génomique réalisé sur la souche de travail et ayant fait l'objet d'un article publié. Le chapitre suivant s'intéressera à l'échelle extracellulaire précédemment mentionnée (Figure 1). Il aura pour objectif la présentation de la caractérisation quantitative (dosages HPLC) et qualitative (peptidomique) de la fraction peptidique des EXLs utilisés et du suivi de son évolution pendant la croissance bactérienne. Le dernier chapitre de résultats aura pour thème la caractérisation des impacts de deux EXLs de travail sur l'expression génétique et la synthèse protéique de la bactérie. Cette partie s'appuiera notamment sur les résultats précédemment obtenus concernant l'utilisation des peptides mesurée par peptidomique afin de formuler des hypothèses quant aux impacts mesurés. Enfin, la discussion soulignera les apports de ce travail tant d'un point de vue académique qu'industriel, et proposera des pistes de travail envisageables pour approfondir cette étude et valider certaines hypothèses émises.

Chapitre 1

Etude bibliographique : bactéries lactiques et peptides

Sommaire

1.1 Introduction	9
1.2 Peptidome extracellulaire en milieu de production	11
1.2.1 Peptides non bactériens issus d'extraits de levure	11
1.2.1.1 Procédé de fabrication des extraits de levure	12
1.2.1.2 Influence de l'étape de lyse cellulaire sur la nature des peptides	14
1.2.1.3 Contenu peptidique des extraits de levure	16
1.2.1.3.1 Caractérisation générale	16
1.2.1.3.2 Stratégies de fractionnement	21
1.2.1.3.3 Analyse peptidomique par spectrométrie de masse	21
1.2.1.3.4 Caractérisation structurale <i>in silico</i>	22
1.2.2 Peptides résultant d'une protéolyse d'origine bactérienne	24
1.2.2.1 Les protéases de paroi, acteurs principaux de la protéolyse nutritionnelle	24
1.2.2.2 Protéases impliquées dans la maturation protéique et la réponse aux stress	25
1.2.2.3 Autres activités protéolytiques	26
1.2.2.4 Lien entre systèmes protéolytiques et peptidome bactérien extracellulaire	26
1.2.3 Peptides résultant d'une biosynthèse bactérienne	28
1.2.3.1 Peptides à synthèse ribosomique	28
1.2.3.1.1 Détection <i>in silico</i> des courtes séquences codantes	29
1.2.3.1.2 Approches complémentaires pour l'identification de nouveaux peptides bactériens	31

1.2.3.1.3	Peptides encodés dans les génomes des bactéries lactiques	32
1.2.3.2	Peptides à synthèse non ribosomique	33
1.2.4	Peptidome extracellulaire : bilan	34
1.3	Principales fonctions assurées par les peptides	36
1.3.1	Nutrition	36
1.3.1.1	Transport des peptides	38
1.3.1.1.1	Transporteurs PTR	38
1.3.1.1.2	ABC-transporteurs	42
1.3.1.1.3	Transport des peptides : bilan	52
1.3.1.2	Clivage intracellulaire	53
1.3.1.3	Devenir des acides aminés libérés	55
1.3.1.4	Peptides nutritionnels : bilan	57
1.3.2	Régulation	58
1.3.2.1	Régulation indirecte : l'exemple de CodY	58
1.3.2.1.1	Mécanismes d'action de CodY	59
1.3.2.1.2	Rôle physiologique de CodY chez les bactéries lactiques	60
1.3.2.2	Régulation directe : mécanismes de <i>quorum sensing</i>	62
1.3.2.2.1	Peptides impliqués dans la détection extracellulaire	64
1.3.2.2.2	Peptides impliqués dans la détection intracellulaire	68
1.3.2.2.3	Influence des peptides nutritionnels sur les systèmes de <i>quorum sensing</i> des bactéries lactiques	73
1.3.2.3	Peptides régulateurs : bilan	75
1.4	Conclusion	77

1.1 Introduction

Lors de leur croissance en milieu riche, comme c'est notamment le cas en production industrielle, les bactéries lactiques sont en étroit contact avec une formidable diversité de peptides environnants. Ces peptides proviennent à la fois du milieu de culture, mais également de la propre production de la bactérie à travers son activité protéolytique et de biosynthèse. Tout au long de la croissance, bactéries et peptides interagissent en permanence, et ces derniers vont assurer diverses fonctions. Dans cette étude bibliographique, deux fonctions distinctes seront explorées : la nutrition et la régulation.

La nutrition constitue en effet le débouché majeur des peptides des milieux de production industriels qui contiennent à cet effet un ou plusieurs hydrolysats protéiques ou cellulaires en mélange et sources d'azote aminé (Ummadi et Curic-Bawden, 2008). Comme leur nom l'indique, les hydrolysats cellulaires sont issus de cellules entières tandis que les hydrolysats protéiques sont obtenus à partir de fractions protéiques, ces derniers étant plus connus sous le nom de peptones. Les hydrolysats cellulaires et protéiques ont des origines variées (animale, végétale, microbienne) ainsi que des modes de fabrication divers (autolyse, lyse acide, alcaline, enzymatique ou fermentaire; Pasupuleti *et al.*, 2008). Néanmoins, leur principal point commun est leur richesse en azote aminé, en particulier sous forme peptidique. Cette richesse en peptides constitue une différence notable par rapport au contexte laitier où les peptides natifs sont peu abondants et où les bactéries s'appuient principalement sur leur système protéolytique pour leur approvisionnement azoté. Cette richesse est ainsi susceptible de créer des dynamiques nutritionnelles distinctes de celles se produisant en lait, et il paraît donc légitime de les explorer.

La fonction de régulation par les peptides est quant à elle notamment connue à travers certains peptides fabriqués par les bactéries qui sont impliqués dans des mécanismes de *quorum sensing* où ils assurent la régulation de l'expression de gènes cibles à l'échelle de la population (Monnet *et al.*, 2014). Ces systèmes attirent grandement l'attention depuis plusieurs années, mais jusqu'à présent ils n'ont pour ainsi dire jamais été explorés dans un contexte de production industrielle. Il est alors légitime de se demander si, parmi les systèmes connus, certains ne sont pas susceptibles d'influencer l'issue de la production, mais on peut également s'interroger sur le lien entre peptides nutritionnels et régulateurs. Quels sont les éléments qui permettent de distinguer un peptide nutritionnel d'un peptide régulateur? Les uns peuvent-ils influencer les mécanismes par lesquels les autres sont pris en charge? Les réponses à ces questions permettraient d'apporter un nouvel éclairage sur la façon dont les peptides d'hydrolysats cellulaires et protéiques influencent les performances de production tant d'un point de vue quantitatif que qualitatif.

Cette étude bibliographique sera articulée en deux parties. La première partie aura pour

objectif d'explorer la diversité peptidique rencontrée par une bactérie lactique lors de sa croissance en milieu riche. L'EXL sera utilisé comme modèle de substrat complexe pourvoyeur de peptides, et seront également abordées les diverses voies par lesquelles les bactéries lactiques participent à la production de nouveaux peptides pendant leur croissance. La deuxième partie présentera ensuite les deux principales fonctions mentionnées précédemment : la nutrition et la régulation. Concernant la nutrition, étant donné le fait qu'il s'agit d'un domaine déjà très bien documenté, un accent particulier sera plutôt mis sur les systèmes de transport de peptides des bactéries lactiques. Ceux-ci représentent en effet les acteurs déterminants pour l'approvisionnement azoté en milieu riche, et les connaissances à leur sujet ont de plus récemment évolué. Enfin, pour ce qui est de la régulation, sera tout d'abord abordé un mode de régulation indirect où ce sont les acides aminés constitutifs des peptides qui participent à une voie de signalisation. Puis, les systèmes de *quorum sensing* seront finalement abordés. Etant donné qu'il s'agit là aussi d'un immense champ disciplinaire, cette partie ne cherchera pas non plus l'exhaustivité mais sera centrée sur quelques peptides phéromones connus à ce jour et sur le lien entre leur structure et leur fonction.

1.2 Peptidome extracellulaire en milieu de production

Dans un contexte de culture d'une bactérie lactique en milieu de production riche, il est possible de différencier les peptides constituant le milieu extracellulaire selon trois origines différentes :

- Peptides exogènes issus directement de l'hydrolysats entrant dans la composition du milieu de culture. Les EXLs serviront dans cette partie de modèle d'hydrolysats.
- Peptides résultant d'une hydrolyse d'origine bactérienne. Il peut s'agir soit d'une hydrolyse de peptides extracellulaires en fragments de plus courte taille, soit de l'hydrolyse de protéines bactériennes (recyclage protéique).
- Peptides synthétisés par la bactérie, par voie ribosomique ou non-ribosomique.

Cette partie vise à décrire, pour chacune de ces trois catégories, les procédés technologiques ou biologiques à l'origine de ces peptides.

1.2.1 Peptides non bactériens issus d'extraits de levure

Les EXLs ont pour première fonction la fourniture d'azote aminé sous forme d'acides aminés libres et de peptides directement utilisables par la bactérie cultivée (Ummadi et Curic-Bawden, 2008). Ils correspondent à la fraction soluble obtenue après autolyse de cellules de levures. Ainsi, en comparaison avec des hydrolysats protéiques classiques - appelés également peptones - qui résultent de l'hydrolyse de fractions protéiques strictes, les EXLs sont composés d'une plus grande variété de nutriments (Tableau 1.1) issus du contenu intracellulaire de la levure (Chae *et al.*, 2001). De plus, contrairement aux hydrolysats d'origine végétale, ils présentent l'avantage de ne pas être soumis à des fluctuations d'ordre climatique, géographique ou saisonnier. Ils constituent donc des produits largement utilisés en fermentation industrielle.

La fraction d'azote aminé (acides aminés libres, peptides, éventuellement protéines) des EXLs représente généralement la part majoritaire de ces substrats. Toutefois la connaissance précise de sa composition reste encore limitée, notamment au niveau peptidique. Les EXLs sont ainsi largement considérés comme étant principalement un mélange mal défini de divers peptides. Cette mauvaise connaissance de la nature des peptides fournis a deux principales origines. Tout d'abord, les peptides étant issus de la dégradation des protéines de levure, leur nature et quantité sont fortement dépendantes du procédé de fabrication, et plus précisément de l'étape de lyse. La deuxième limite est d'ordre analytique car il n'existe à ce jour aucune technique permettant d'identifier et de quantifier avec précision l'ensemble des peptides en mélange. Cette partie présentera donc d'une part le procédé de fabrication des extraits de levure et la façon dont sa conduite est susceptible de moduler la composition pep-

TABLEAU 1.1 : Exemple de composition nutritionnelle d'un extrait de levure. Adapté d'Ogoshi *et al.* (2014).

Composant	Pourcentage de matière sèche (%)
Protéines	50,00
Lipides	0,20
Glucides	22,20
Acides nucléiques	5,40
Fibres	0,40
Cendres	8,20
Humidité	6,00
Macrominéraux ^a	5,51
Microminéraux ^b (ppm)	666
Vitamines ^c (ppm)	209

^a Soufre, sodium, phosphore, potassium, magnésium, calcium

^b Fer, cuivre, zinc, manganèse, chlorure

^c Niacine, biotine, acide pantothénique, thiamine, riboflavine, pyridoxine, cyanocobalamine, niacine

tidique du produit final. Puis, un point sera fait sur les principales techniques analytiques mises en œuvre jusqu'à présent pour caractériser les mélanges complexes de peptides.

1.2.1.1 Procédé de fabrication des extraits de levure

Le procédé général de fabrication des EXLs est présenté dans la Figure 1.1. Historiquement, les EXLs ont d'abord été utilisés en tant qu'exhausteurs de goût dans l'industrie agroalimentaire. Cette orientation transparait encore fortement dans la fabrication des EXLs à destination des milieux de fermentation. Ainsi, les souches de levure utilisées sont principalement des souches à application agroalimentaire appartenant préférentiellement aux genres *saccharomyces* (*S. cerevisiae*, *S. rouxii*, *S. uvarum*), *kluveromyces* (*K. lactis*, *K. fragilis*) ou *candida* (*C. utilis*; Kanegae *et al.*, 1989; Potman et Wesdorp, 1994). En Europe, la culture de *S. cerevisiae* est généralement préférée et s'effectue en milieu à base de molasses de betterave. La biomasse obtenue en fin de fermentation, qui est également appelée crème de levure après récolte et lavages, est hydrolysée par un processus d'autolyse durant lequel les enzymes endogènes vont progressivement dégrader les macromolécules. Cette autolyse est en général induite par un stress thermique modéré (50-70 °C) durant lequel les cellules sont tuées et perméabilisées (plasmolyse) mais permettant de maintenir l'activité des enzymes endogènes (Kanegae *et al.*, 1989). La plasmolyse peut également être facilitée par l'adjonction de sels (NaCl) ou de solvants organiques (toluène, chloroforme, acétate d'éthyle, formaldéhyde etc.), ces derniers possédant en outre une fonction antimicrobienne (Breddam et Beenfeldt, 1991). L'autolyse se fait ensuite dans des conditions contrôlées de pH et de température per-

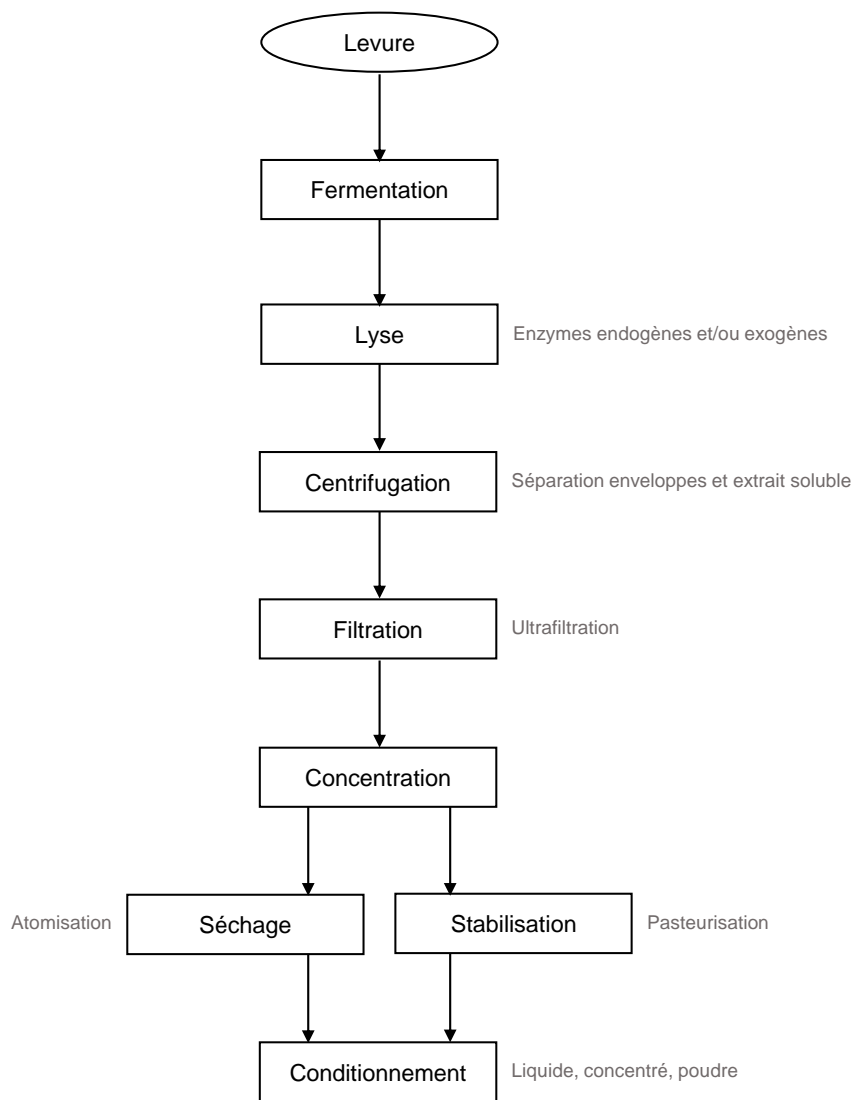


FIGURE 1.1 : Procédé de fabrication des extraits de levure

mettant aux différentes enzymes endogènes de dégrader les macromolécules. Les protéases vont ainsi libérer acides aminés et peptides, les acides nucléiques sont hydrolysés par des nucléases, le glycogène et tréhalose par des glycosidases, les lipides par des lipases et les glucanes de la paroi par des glucanases. Ces produits d'hydrolyse vont, avec les vitamines et minéraux, diffuser à travers les enveloppes endommagées dans le milieu extracellulaire. Il est également courant pour les industriels d'ajouter des enzymes exogènes afin d'orienter à façon l'hydrolyse de la levure. La papaïne, une endoprotéase à cystéine à large spectre, est régulièrement utilisée du fait de son action synergique avec les protéases endogènes de la levure (Chao *et al.*, 1980). La subtilisine, une endoprotéase à sérine commercialisée sous le nom d'Alcalase, est aussi largement employée du fait de sa faible spécificité de clivage (Kalum, 2009). Deux autres enzymes sont également couramment utilisées et permettent l'enrichissement en nucléotides spécifiques : la 5'-phosphodiesterase et l'adénosine monophosphate (AMP)-déaminase, qui permettent d'enrichir les EXLs respectivement en 5'-GMP (5'-Guanosine MonoPhosphate) et en 5'-IMP (5'-Inosine MonoPhosphate; Chae *et al.*, 2001). Dans certaines fabrications, il est également possible de conduire l'hydrolyse de la levure exclusivement *via* des enzymes exogènes. Ceci nécessite l'inactivation préalable des enzymes endogènes par choc thermique (température > 70 °C). Une fois l'hydrolyse terminée, les enveloppes sont séparées de l'extrait soluble par centrifugation. Le pH de ce dernier est ensuite généralement ajusté de façon à augmenter la teneur en insolubles par cristallisation des sels en solution. Ceux-ci sont ensuite retirés par ultrafiltration. Finalement, l'EXL fabriqué peut être soit séché par atomisation si vendu sous forme de poudre, soit concentré par évaporation et pasteurisé si vendu sous forme liquide.

1.2.1.2 Influence de l'étape de lyse cellulaire sur la nature des peptides

Le contenu en peptides des EXLs est ainsi sous dépendance directe de l'étape de lyse cellulaire qui peut donc se faire selon trois modes de conduite : autolyse stricte, autolyse couplée à des activités enzymatiques exogènes, ou hydrolyse entièrement exogène. Le processus d'autolyse de la levure a été largement documenté (Charpentier et Feuillat, 1993), et il repose en grande partie sur le système protéolytique de la levure lui aussi bien caractérisé depuis plusieurs années (Achstetter et Wolf, 1985b, 1985a; Jones, 1991). Les enzymes protéolytiques à l'œuvre durant l'autolyse sont principalement d'origine vacuolaire (Bhalová et Beran, 1979). Sept protéases vacuolaires ont été à ce jour caractérisées (Hecht *et al.*, 2014). Elles incluent trois métalloprotéases, trois protéases à sérine et une protéase aspartique. Elles présentent au total deux activités endoprotéasiques, trois aminoprotéases et deux carboxyprotéases (Tableau 1.2). Du fait la diversité des modes de clivage et des spécificités de ces enzymes, la dégradation des structures protéiques durant l'autolyse est susceptible de libérer une grande variété de peptides de taille et de composition diverses. Sommairement, les exoprotéases vont permettre l'enrichissement en acides aminés libres et en courts peptides,

tandis que les endoprotéases sont responsables de la libération d'oligopeptides de poids moléculaire intermédiaire. Le niveau d'hydrolyse général est quant à lui premièrement fonction du temps d'autolyse (Charpentier et Feuillat, 1993). L'optimisation des paramètres d'autolyse et la connaissance de leur impact sur le profil peptidique du produit fini font naturellement partie du savoir-faire propre à chaque fabricant. Les informations disponibles à ce sujet sont donc très limitées dans la littérature publique.

TABLEAU 1.2 : Protéases vacuolaires de *S. cerevisiae*. Adapté de Hecht *et al.* (2014).

Protease	Gène	Activité	Fonction	Acides aminés identifiés en position P1
Protéinase A	PEP4	Endoprotéase aspartique	Initiation de la cascade d'activation des protéases; dégradation des protéines	Phe, Leu, Tyr, Trp, Thr, Asn, Gln, Glu, Lys, Ala, Ile
Protéinase B	PRB1	Endoprotéase à sérine	Activation des protéases; dégradation des protéases	Leu, Arg, Phe, Tyr, Gln, Lys
Carboxypeptidase Y	PRC1	Carboxypeptidase à sérine	Dégradation des peptides	Ala, Gly, Val, Leu, Ile, Met, Phe
Carboxypeptidase S	CPS1	Métalloprotéase à zinc	Dégradation des peptides	Gly, Phe
Aminopeptidase I	APE1	Métalloprotéase à zinc	Dégradation du glutathion	Leu, Cys/Gly
Aminopeptidase Y	APE3	Métalloprotéase	Inconnue	Pro, Ala, Leu, Met, Phe, Tyr, Ser, Lys, Arg
Dipeptidylaminopeptidase B	DAP2	Dipeptidase à sérine	Inconnue	Xaa-Ala, Xaa-Pro

L'influence d'enzymes exogènes sur l'hydrolyse protéique est en revanche mieux documentée. On peut notamment relever une étude récente réalisée à partir d'hydrolysats de protéines de pomme de terre qui a détaillé les effets de quatre solutions enzymatiques commerciales : la papaïne et l'Alcalase précédemment citées, Novo Pro-D, une protéase issue de *Bacillus sp.*, et Flavourzyme, un mélange de protéases provenant d'*Aspergillus oryzae* (Waglay et Karboune, 2016). L'efficacité de ces quatre protéases a ainsi été caractérisée à travers leur activité protéolytique, le degré d'hydrolyse obtenu dans des conditions contrôlées, et *via* les proportions relatives de peptides et d'acides aminés obtenues (Tableau 1.3). Les auteurs ont ainsi montré que le plus faible taux d'hydrolyse était atteint avec Novo Pro-D et ce malgré une haute activité protéolytique. Cette faible hydrolyse est probablement attribuable à la spécificité de clivage étroite de cette enzyme qui hydrolyse préférentiellement les liaisons peptidiques au niveau des résidus aromatiques et soufrés. Les trois autres protéases ont quant à elles des spécificités plus larges permettant une meilleure digestion du substrat protéique. On peut également remarquer que le cocktail enzymatique Flavourzyme a conduit à une plus forte libération d'acides aminés libres, ce qui est dû à la présence d'exoprotéases dans sa composition.

TABLEAU 1.3 : Efficacité catalytique de quatre protéases commerciales pour la formation de peptides et d'acides aminés libres à partir de protéines de pomme de terre. D'après Waglay et Karboune (2016).

Enzyme ^a	Activité protéolytique (U/ml) ^b	Degré d'hydrolyse (%)	Proportion relative de peptides (%)	Proportion relative d'acides aminés (%)
Novo Pro-D	6468	4,5-10	32-56	44-68
Alcalase	9061	> 93	83	17
Flavourzyme	153,2	> 98	15	85
Papaïne	139,0	> 83	86	14

^a Les digestions ont été réalisées avec 2 U d'activité protéolytique par ml de mélange réactionnel à 50 °C pendant 16h.

^b L'activité protéolytique a été déterminée avec de l'azocaséine selon une méthode adaptée de Trengove *et al.* (1999). Une unité correspond à la quantité d'enzyme nécessaire pour augmenter d'une unité l'absorbance à 440 nm/min.

Outre cette caractérisation, un autre intérêt de cette étude provient de l'utilisation de la spectrométrie de masse pour identifier qualitativement les peptides libérés lors des hydrolyses réalisées. Les auteurs ont ainsi pu identifier les protéines parentales desquelles proviennent les peptides de même que les spécificités de clivage enzymatique associées. Cette stratégie de caractérisation par spectrométrie de masse est ainsi transférable à tout type d'hydrolysats - dont les EXLs - et sera plus explicitement détaillée dans la partie 1.2.1.3.3.

1.2.1.3 Contenu peptidique des extraits de levure

1.2.1.3.1 Caractérisation générale

Pour une crème de levure donnée, il est ainsi possible d'aboutir à des produits finis aux profils peptidiques très différents en modulant à façon les activités enzymatiques gouvernant l'étape de lyse. Le contenu en azote aminé des EXLs commerciaux est couramment caractérisé par différents indicateurs :

Le degré d'hydrolyse (DH), correspondant au pourcentage de liaisons peptidiques rompues durant l'autolyse. Il est le plus souvent estimé par le ratio AN/TN. L'azote aminé (AN) représente le nombre de groupements amine libres d'un EXL et inclut donc les groupements des acides aminés libres, la chaîne latérale des lysines et les extrémités N-terminales des peptides. Il est généralement calculé par dosage au formaldéhyde ou à l'acide trinitrobenzène-sulfonique (TNBS; Pasupuleti et Braun, 2008). L'azote total (TN) inclut à la fois l'azote aminé et les sels d'ammonium quaternaires. Il est titré par la méthode de Kjeldahl.

La distribution de masses moléculaires, permettant d'évaluer la quantité de peptides en fonction de leur taille. Cette distribution est classiquement mesurée par chromatographie d'exclusion stérique couplée à une détection de l'absorbance à 220 nm (Visser *et al.*, 1992).

Dosage des acides aminés libres et acides aminés totaux après hydrolyse acide. Fournies par certains fabricants, ces mesures notamment permettent d'évaluer avec plus de préci-

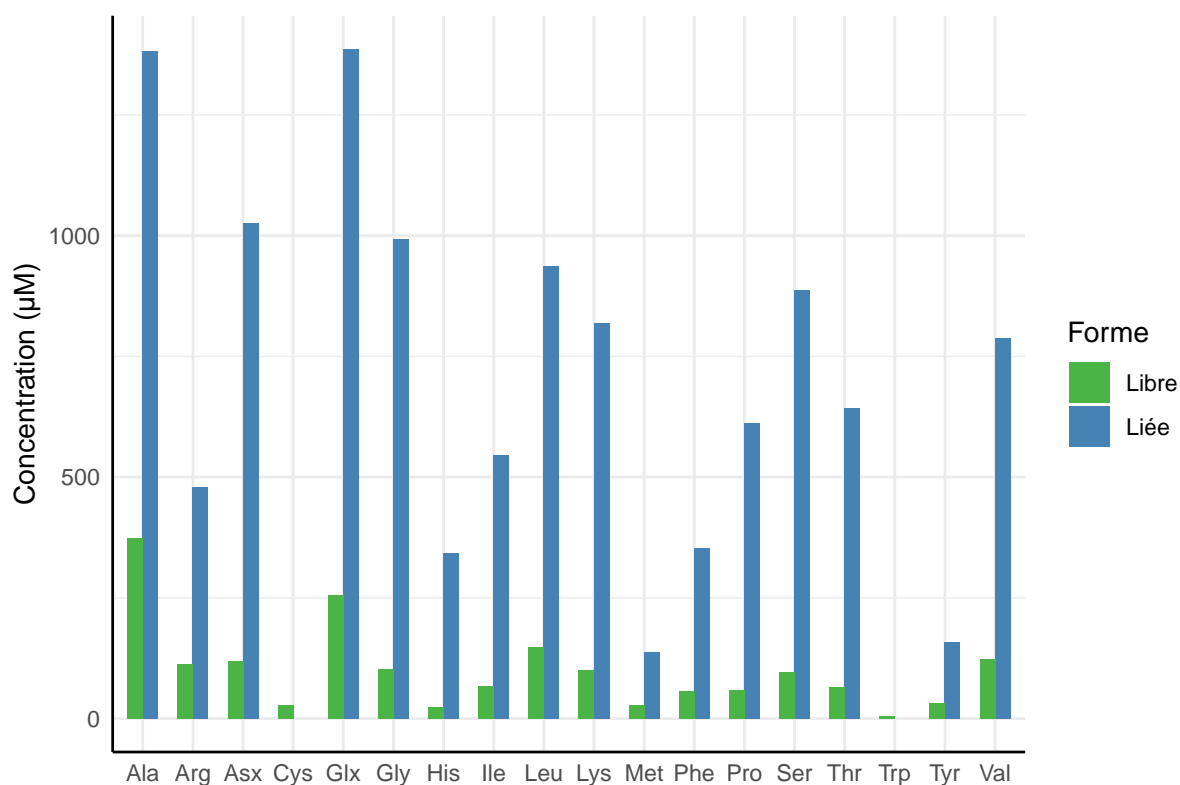


FIGURE 1.2 : Composition en acides aminés libres et engagés dans des liaisons peptidiques d'un extrait de levure en solution à 2 g/l. La cystéine et tryptophane étant dégradés durant l'hydrolyse acide, leur valeur sous forme liée n'est pas disponible. Asx = Asp + Asn, Glx = Glu + Gln. D'après Kevvai *et al.*, 2014.

sion la composition en acides aminés engagés dans des liaisons peptidiques (différence entre acides aminés libres et totaux). L'une des méthodes d'analyse couramment utilisée est la chromatographie liquide haute performance en phase inverse (RP-HPLC; Molnár-Perl, 2011). Celle-ci est classiquement couplée à une dérivation pré-colonne des acides aminés à l'*o*-phtalaldéhyde (OPA, aminés primaires) et au fluorénylméthoxycarbonyl (FMOC, amines secondaires) permettant une détection et quantification simultanément en ultraviolet et fluorescence. Une détection en spectrométrie de masse est également possible pour une quantification plus fine et une identification certaine des acides aminés. Cette approche a notamment été développée par Kevvai *et al.* (2014) (Figure 1.2). Il s'agit dans ce travail d'un EXL particulier car fabriqué à façon et marqué isotopiquement à l'azote 15 afin d'effectuer un suivi des flux d'azote durant la croissance de *L. lactis* (les conclusions de cette étude seront développées dans la partie 1.3.1.3). Les valeurs mesurées révèlent une surabondance des peptides par rapport aux acides aminés libres dans cet EXL, avec des concentrations totales respectives de 11,5 mM et 1,80 mM. Ces valeurs correspondent à une concentration finale d'EXL de 2 g/l, ce qui est dans la fourchette basse des gammes de concentration généralement employées dans les milieux de culture. L'alanine et le couple glutamate/glutamine sont les acides aminés les plus abondants, tant sous forme libre que liée. Les acides aminés soufrés ainsi que le tryptophane et la tyrosine sont à l'inverse minoritaires. Cet EXL ayant été obtenu en laboratoire, il serait intéressant de croiser ces valeurs avec celles d'EXLs commerciaux.

De telles données sont disponibles dans le Tableau 1.4 qui référence les caractéristiques générales de quelques hydrolysats de levure (EXLs et peptones) disponibles sur le marché. Ces données permettent notamment de constater une certaine variabilité du degré d'hydrolyse (AN/TN) qui est relativement bien corrélé avec la distribution des masses moléculaires des différents produits. Le produit le moins hydrolysé (Hy-Yest 101, AN/TN = 13,9 %) présente ainsi des composés de plus haut poids (près de 20 % de masses supérieures à 2 kDa) tandis que le plus hydrolysé (Yeast Extract, BD Biosciences, AN/TN = 60 %) est enrichi en produits de faible masse inférieure à 250 Da. La distribution des acides aminés totaux, en proportions relatives, présente en revanche une certaine homogénéité entre produits. L'acide glutamique est systématiquement le composé majoritaire, conformément aux données de la Figure 1.2, suivi de l'alanine ou de l'acide aspartique en fonction des produits. On retrouve également les acides aminés soufrés (cystéine et méthionine), la tyrosine ainsi que l'histidine comme étant les acides aminés minoritaires. La prévalence de l'acide glutamique est caractéristique de l'aptitude de *S. cerevisiae* à accumuler fortement ce composé durant sa croissance (Watson, 1976; Singh et Neelakantan, 1989). Cela reflète également l'orientation initiale des EXLs vers le marché agroalimentaire, le glutamate étant responsable de la saveur umami en conjonction avec le 5'-GMP et 5'-IMP (Kurihara, 2015).

TABLEAU 1.4 : Principales caractéristiques d'hydrolysats de levure commerciaux. Informations compilées à partir des fiches techniques mises à disposition par les différents fabricants.

Entreprise	Produit	AN/TN (%)	Distribution masses moléculaires (%)						Acides aminés totaux ^a (% azote aminé total)				
			>10 kDa	5-10 kDa	2-5 kDa	1-2 kDa	1k-500 Da	< 500 Da	Alanine	Arginine	Acide aspartique	Cystine	Acide glutamique
Kerry	Hy-Yest 101	13,9	4	6,5	9	7,9	10,1	62,5	7,9	5,2	10,2	0,5	18,4
	Hy-Yest 412	44,6	0,2	0,1	3,6	14,4	17,2	64,4	8,6	5,4	11,0	0,5	17,8
	Hy-Yest 413	41,3	0	0	3,2	15,1	18,4	63,3	6,7	5,7	11,0	0	16,9
	Hy-Yest 444	42,9	0	0,1	3,5	13,1	16,3	67,1	7,6	4,9	10,8	0	23,4
	Hy-Yest 502	30	0	0,7	6,3	13,8	18,1	61,1	8,5	5,1	11,5	0,6	21,3
	Hy-Yest 503	31,4	0,3	0,5	6,2	14,3	19,1	59,6	7,9	5,2	11,4	1	22,0
	Hy-Yest 504	39,1	0,3	0,6	6,2	14	18	60,9	7,9	5,2	10,8	0,2	20,0
BD Biosciences			>10 kDa	5-10 kDa	2-5 kDa	500-2k Da	250-500 Da	< 250 Da					
	Bacto Yeast Extract	55	0	< 1	6	22	10	62	10,7	5,0	10,2	ND	18,0
	Yeast Extract	60	0	< 1	3	18	9	69	9,7	4,7	9,2	ND	17,2
	Bacto Yeast Extract, Technical	54	0	< 1	5	21	9	64	7,8	7,3	10,4	ND	18,4
	Difco Yeast Extract, UF	56	0	< 1	5	22	9	63	9,8	4,7	9,8	ND	18,1
Organotechnie			>10 kDa	1-10 kDa	500-1k Da	300-500 Da	< 300 Da						
	Yeast Extract 19512	41-60	0	10,7	13	17,2	59,1		7,6	5,2	11,2	1,1	20,5
Bio Springer	Springer 0207	49-63	ND	ND	ND	ND	ND		7,6	4,2	10,8	2,2	18,2
	Springer 0251	38-58	ND	ND	ND	ND	ND		8,6	4,2	4,5	0,8	19,6
	Springer 2506	26-49	ND	ND	ND	ND	ND		8,7	4,8	9,5	0,6	23,4
	Springer HYP-A	13-32	ND	ND	ND	ND	ND		6,8	5,9	11,3	0,3	14,0

Entreprise	Produit	Acides aminés totaux ^a (% azote aminé total)											
		Glycine	Histidine	Isoleucine	Leucine	Lysine	Méthionine	Phénylalanine	Proline	Sérine	Thréonine	Tyrosine	Valine
Kerry	Hy-Yest 101	4,5	2,0	8,8	7,0	7,3	1,4	4,1	3,9	5,2	5,2	3,2	5,2
	Hy-Yest 412	4,7	2,1	5,1	7,5	8,0	1,6	4,4	3,8	4,9	4,9	3,7	6,1
	Hy-Yest 413	3,6	1,9	4,8	7,1	9,5	1,2	4,3	8,6	4,5	4,5	4,3	5,5
	Hy-Yest 444	4,6	2,0	4,9	7,1	7,5	1,5	4,1	3,9	4,4	4,6	3,1	5,6
	Hy-Yest 502	5,3	2,1	4,3	6,8	7,9	1,5	3,8	4,0	4,3	4,5	3,2	5,3
	Hy-Yest 503	5,2	1,9	4,2	6,6	7,9	1,0	3,9	3,9	5,0	5,2	2,7	5,0
	Hy-Yest 504	4,0	1,9	4,2	6,4	8,7	1,3	4,2	4,0	4,6	4,4	6,7	5,2
BD Biosciences	Bacto Yeast Extract	5,7	2,5	5,7	7,9	8,8	1,5	5,0	3,8	3,1	3,1	2,3	6,7
	Yeast Extract	5,1	2,2	7,3	9,7	7,6	1,7	6,9	3,6	3,0	2,8	1,9	7,5
	Bacto Yeast Extract, Technical	4,8	2,2	5,0	8,0	8,2	2,2	4,8	3,7	5,0	3,9	2,2	6,5
	Difco Yeast Extract, UF	5,3	2,2	6,9	8,5	8,3	1,5	6,5	3,4	3,1	2,9	1,5	7,4
Organotechnie	Yeast Extract 19512	4,7	2,1	4,9	6,9	7,8	1,5	4,7	4,7	5,0	4,9	1,5	5,8
Bio Springer	Springer 0207	4,6	2,2	5,3	8,0	7,3	1,9	4,3	4,2	5,0	4,9	2,0	6,0
	Springer 0251	5,4	2,4	4,6	8,5	7,1	2,2	4,9	5,7	5,8	6,8	2,2	6,7
	Springer 2506	5,4	2,2	4,2	6,4	7,6	1,1	3,8	4,3	5,0	5,5	2,1	5,5
	Springer HYP-A	4,8	2,6	4,5	7,9	8,2	2,0	4,8	4,9	5,8	6,1	4,2	5,9

^a Les valeurs de concentrations relatives des acides aminés ont été obtenues après hydrolyse acide complète du contenu protéique. Elles correspondent à la somme des acides aminés libres et engagés dans des liaisons peptidiques. Durant l'étape d'hydrolyse, le tryptophane est entièrement détruit, la cystéine partiellement dégradée et dimerisée en cystine, et la glutamine et asparagine converties en acide glutamique et aspartique, respectivement.

ND : donnée non disponible

1.2.1.3.2 Stratégies de fractionnement

Ces différentes méthodes, du fait de leur relative simplicité, représentent la principale stratégie de caractérisation des EXLs. Il est toutefois possible d'affiner cette caractérisation en séparant et en quantifiant certaines molécules ou groupes de molécules sur la base de leur propriétés physico-chimiques (taille, charge et hydrophobicité) par le biais de fractionnements chromatographiques (exclusion stérique, échange d'ions et phase inverse, respectivement) ou de filtrations tangentielles. Cette stratégie permet en outre d'évaluer la bioactivité des différentes fractions obtenues. Mosser *et al.* (2012) ont ainsi montré qu'une fraction chromatographique d'un EXL contenant principalement des peptides cationiques et hydrophiles permettait de supporter le mieux la croissance d'une ligne cellulaire issue d'ovaires de hamster (CHO) avec des performances comparables à celles de l'EXL complet. Les mêmes auteurs ont également suggéré que les courts peptides - di et tripeptides - présents dans un perméat de nanofiltration étaient les sources d'azote supportant la croissance de la même lignée cellulaire (Mosser *et al.*, 2015).

1.2.1.3.3 Analyse peptidomique par spectrométrie de masse

Diverses techniques analytiques permettent donc d'évaluer la composition peptidique générale des EXLs. Néanmoins, celles-ci ne sont pas capables pas de lever entièrement le voile sur la complexité de ces substrats. Si les stratégies de fractionnement permettent de réduire significativement cette complexité, il n'est toujours pas possible de séparer chaque constituant individuellement. Les mesures d'acides aminés libres et totaux ont quant à elles l'avantage d'être quantitatives et de mesurer la composition globale des peptides en acides aminés. Toutefois, elles ne renseignent pas sur les séquences propres des peptides.

Pour pallier ces limitations, une solution complémentaire pourrait venir de la spectrométrie de masse. Si celle-ci est employée avec succès depuis de nombreuses années pour séquencer des peptides dans un contexte d'analyse protéomique *bottom-up* (Link *et al.*, 1999; Wolters *et al.*, 2001; Yates, 2004), son utilisation pour l'analyse de mélanges complexes de peptides natifs - appelée également peptidomique (Schulz-Knappe *et al.*, 2001) - est plus limitée et ne s'est réellement développée que récemment (Bingeman *et al.*, 2017). La majorité des études portent sur des échantillons eucaryotes : venin (Tashima *et al.*, 2012), lait et denrées alimentaires (Boutrou *et al.*, 2013; Lahrichi *et al.*, 2013), salive (Zhang *et al.*, 2012; Tian *et al.*, 2017), végétaux (Okamoto *et al.*, 2015, 2016) ou encore culture de lignées cellulaires *in vitro* (Tsuchiya *et al.*, 2015). Cette technologie a également été utilisée en microbiologie pour décoder le peptidome intracellulaire de *Salmonella typhimurium* (Manes *et al.*, 2007) et extracellulaire de *L. lactis* (Guillot *et al.*, 2016), les deux étant issus d'une activité protéolytique endogène. Cette aspect sera abordé plus en détail dans la partie 1.2.2.4. Dans le domaine des hydrolysats protéiques, on peut trouver plusieurs exemples portant sur des

substrats variés et visant le plus souvent à identifier de nouveaux peptides bioactifs : hydrolysats de protéines de laits (Norris *et al.*, 2014; Nongonierma *et al.*, 2018), de pommes de terre (Rajendran *et al.*, 2016; Waglay et Karboune, 2016), de carpe (Borawska *et al.*, 2016) ou encore de légumineuses (Zinchenko *et al.*, 2018). Toutefois, à ce jour, aucune étude portant sur l'analyse peptidomique des EXLs n'a encore été publiée.

Il convient néanmoins de préciser que l'analyse peptidomique par spectrométrie de masse n'est pas encore arrivée pleinement à maturité. Elle emprunte divers outils et méthodes à la métabolomique et à la protéomique et souffre encore d'un certain nombre de limitations. Non exhaustivement, ces limitations incluent de faibles rendements d'extraction et de purification des peptides, ainsi qu'une identification limitée voire impossible de certaines gammes de masses et de modifications post-traductionnelles (Dallas *et al.*, 2015; Bingeman *et al.*, 2017). En particulier, l'identification de peptides de courte taille - di à pentapeptides - est techniquement impossible par une analyse de *mass matching* classiquement utilisée avec les oligopeptides. Leur analyse nécessite une approche de séquençage *de novo* incluant une optimisation préalable des conditions de fragmentation ainsi qu'une vérification manuelle des spectres générés (O'Keeffe et FitzGerald, 2015). A ce jour, il n'est ainsi pas possible d'obtenir une caractérisation entièrement exhaustive des mélanges complexes de peptides en une seule analyse. Finalement, la quantification relative sans marquage de type *spectral counting* ou extraction des courants d'ions (XIC), bien qu'aisée est assez peu précise. Néanmoins, malgré toutes ces limitations, l'analyse peptidomique constitue à ce jour le seul outil permettant le séquençage de peptides en mélange complexe, et pourrait donc être mise à profit pour une étude plus approfondie des EXLs.

1.2.1.3.4 Caractérisation structurale *in silico*

La spectrométrie de masse constitue donc un outil de choix pour la production de larges listes de séquences peptidiques. Néanmoins, ces données ne renseignent que sur la structure primaire des peptides, c'est-à-dire l'enchaînement des acides aminés les constituant. La connaissance de leur structure tridimensionnelle représente un niveau supplémentaire de caractérisation permettant de mieux comprendre le rôle de certains résidus dans une éventuelle activité fonctionnelle ou impliqués dans une interaction avec une autre molécule. La structure topologique d'un peptide peut classiquement être résolue grâce à des approches expérimentales telles que la cristallographie aux rayons X (Spencer et Nowick, 2015), la spectroscopie à résonance magnétique nucléaire (RMN; Gomathi et Subramanian, 1996) ou le dichroïsme circulaire vibrationnel (Keiderling, 2002). Bien qu'efficaces, ces approches sont avant tout ciblées et ne permettent pas une analyse haut débit de mélanges complexes de peptides. De plus, malgré les progrès réalisés dans ces domaines analytiques, en particulier la spectrométrie RMN, le nombre d'entrées dans la Protein Data Bank (PDB; Berman *et al.*, 2000) pour les protéines à chaîne unique de moins de 50 résidus reste relativement peu

élevé (3693 au 8 août 2018) par rapport aux nombres de structures uniques de toutes tailles référencées (56249).

La meilleure solution pour l'analyse à haut-débit de structures peptidiques reste donc à l'heure actuelle la modélisation *in silico*. Pour les protéines, les techniques privilégiées reposent sur de la modélisation par homologie (Martí-Renom *et al.*, 2000). Il est en effet admis que des protéines montrant une identité de séquence élevée vont également avoir tendance à adopter une conformation similaire. Or, le faible nombre de structures peptidiques expérimentalement résolues limite fortement la pertinence de ces approches pour les peptides (Thévenet *et al.*, 2015). De plus, les peptides sont globalement des molécules plus flexibles que les protéines du fait d'un plus faible nombre d'interactions stabilisatrices et peuvent ainsi adopter différentes conformations (Figure 1.3). La substitution d'un seul acide aminé dans la chaîne peptidique peut drastiquement affecter la structure globale, ce qui limite également la pertinence de la modélisation par homologie.

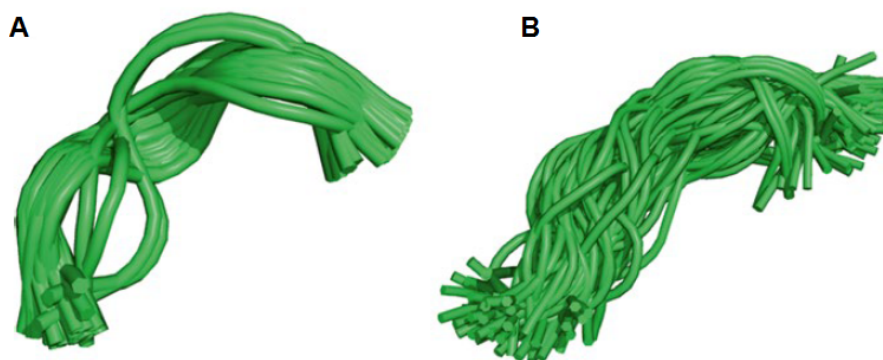


FIGURE 1.3 : Exemples de conformations peptidiques possibles déterminées sous contraintes RMN. (A) Met-enképhaline (YGGFM); (B) domaine peptidique de la glycoprotéine MUC2 (XPTTTPLKX). Thévenet *et al.*, 2015.

La modélisation *de novo* est donc privilégiée pour les peptides. A l'heure actuelle, les approches disponibles peuvent être divisées en deux catégories (Thévenet *et al.*, 2015). La première catégorie fait appel à des simulations de dynamique moléculaire visant à décrire la position et le mouvement de chaque atome d'une molécule par unité de temps. Cette approche à l'avantage d'être non biaisée et applicable à tout type de molécule mais nécessite d'importantes capacités de calcul. Pour la caractérisation haut-débit de peptides, on privilégie plutôt des modèles élaborés spécifiquement pour les peptides permettant une caractérisation plus rapide. Différentes approches ont ainsi été mises au point et certaines sont disponibles sous forme d'outils accessibles en ligne (se référer à Thévenet *et al.*, 2015 pour une revue récente des principaux outils disponibles). A ce jour, on ne trouve pas encore d'exemple d'application de tels outils sur des données de séquençage d'hydrolysats protéiques. Il pourrait néanmoins s'agir d'une perspective intéressante, en particulier dans

un contexte de recherche de peptides bioactifs basée sur le principe de relation quantitative structure à activité (Perkins *et al.*, 2003).

1.2.2 Peptides résultant d'une protéolyse d'origine bactérienne

Les peptides issus de l'EXL - et plus généralement de tout hydrolysats cellulaire ou protéique - forment la première catégorie de peptides présents dans le milieu de culture en contact avec une bactérie lactique lors de sa croissance. La deuxième catégorie concerne les peptides formés par l'activité protéolytique de la bactérie. Deux systèmes protéolytiques distincts sont particulièrement bien connus chez les bactéries lactiques : un système affilié à la nutrition et un autre impliqué dans la maturation des protéines et les réponses aux stress. Ces deux systèmes ont été très bien étudiés et plusieurs revues les ont couverts en profondeur (système protéolytique nutritionnel : Kunji *et al.*, 1996a; Savijoki *et al.*, 2006; Liu *et al.*, 2010; système protéolytique lié aux stress : van de Guchte *et al.*, 2002; Papadimitriou *et al.*, 2016). Les principaux acteurs de ces systèmes seront décrits dans la suite.

1.2.2.1 Les protéases de paroi, acteurs principaux de la protéolyse nutritionnelle

Le système protéolytique nutritionnel repose en premier lieu sur une classe particulière de protéines : les protéases de paroi (Prt). Celles-ci regroupent l'ensemble des endoprotéases ancrées à la paroi bactérienne et identifiées chez de nombreuses bactéries lactiques : lactocoques (PrtP), lactobacilles (PrtB, PrtR, PrtH) et *S. thermophilus* (PrtS). Une synthèse des différentes protéases de paroi identifiées et de leurs principales caractéristiques (taille, spécificité de substrat, classe biochimique, température et pH d'activité optimale) est disponible dans Atlan *et al.* (2008), p350-351. Ces protéases ont été principalement étudiées à travers leur capacité à hydrolyser les caséines du lait en oligopeptides qui servent de substrats indispensables à la croissance en lait (Juillard *et al.*, 1995). Les peptides ainsi libérés vont ensuite être transloqués intracellulairement par divers transporteurs dédiés et vont progressivement subir une hydrolyse complète sous l'action de plusieurs peptidases. Les notions de transport de peptides et d'hydrolyse intracellulaire seront développées dans la partie 1.3.1.

On peut noter qu'à ce jour, il n'existe aucune étude portant sur l'activité éventuelle des protéases de paroi en milieu complexe basé sur des hydrolysats protéiques et donc dépourvu de protéines exogènes. Cette absence provient du fait que i) le lait est le milieu d'étude privilégié des bactéries lactiques, et ii) d'une moindre pertinence d'une telle étude du fait d'une probable répression de CodY sur l'expression des protéases de paroi en milieu complexe par rapport au lait en raison d'une plus forte abondance nutritionnelle (voir partie 1.3.2.1). Néanmoins, même si leur synthèse est fortement réprimée, les protéases sont malgré tout

susceptibles d'intervenir, et l'on peut se demander si elles sont effectivement capables d'hydrolyser les peptides environnants. Il est généralement admis que toutes les activités peptidolytiques des bactéries lactiques ont une origine exclusivement intracellulaire (Christensen *et al.*, 1999; Savijoki *et al.*, 2006). Les bactéries lactiques possèdent en effet un nombre important de peptidases à localisation cytoplasmique possédant des spécificités complémentaires et capables de libérer entièrement les acides aminés engagés dans les peptides internalisés. Toutefois, cette assertion néglige certainement l'activité des Prt qui, sans offrir un tel degré d'hydrolyse, sont probablement capables de dégrader des oligopeptides en fragments de plus courte taille. Cette aptitude n'a jamais été explorée directement mais est implicitement démontrée dans plusieurs études où des fragments de caséine (e.g. fragment 1-23 de la caséine α_{s1}) ainsi que de courts peptides chromogènes ont été utilisés pour caractériser diverses protéases de paroi telles que PrtS de *S. thermophilus* (Fernandez-Esplá *et al.*, 2000). Il est donc tout à fait probable que les protéases de paroi participent à l'élaboration du peptidome extracellulaire par un clivage additionnel de certains peptides en fragments de plus courte taille.

1.2.2.2 Protéases impliquées dans la maturation protéique et la réponse aux stress

Les bactéries lactiques possèdent plusieurs protéines appartenant à la famille des protéines de choc thermique (HSP, heat shock protein) qui ont une double fonction chaperon et protéase. Ces protéines sont impliquées dans le repliement des protéines naissantes, dans le recyclage protéique, et elles sont particulièrement actives en conditions de stress où elles assurent le repliement et la dégradation des protéines endommagées (van de Guchte *et al.*, 2002; Papadimitriou *et al.*, 2016). HtrA est une protéase à sérine à activité trypsine qui a été identifiée chez *Escherichia coli* comme jouant le rôle de chaperon à basse température et dont l'activité protéolytique prédominait à des températures normales et hautes (Spiess *et al.*, 1999). Il s'agit d'une protéine attachée à la membrane dont le site catalytique est exposé à la surface externe de la cellule (Poquet *et al.*, 2000). Son activité concerne ainsi les protéines de surface, c'est-à-dire liées à la membrane, à la paroi, sécrétées dans le milieu extérieur ou encore les lipoprotéines. Des homologues ont été caractérisés chez *L. lactis* (Poquet *et al.*, 2000) et *Lb. helveticus* (Smeds *et al.*, 1998). En outre, il a été montré que son expression était induite chez *S. thermophilus* entre la phase de croissance exponentielle précoce et tardive de la bactérie lorsque cocultivée avec *Lb. delbrueckii* subsp. *bulgaricus* (Herve-Jimenez *et al.*, 2009). De même, sa production est stimulée chez *Lb. casei* en présence de composés phénoliques (Rivas-Sendra *et al.*, 2011).

Bien que cela semble contre-intuitif, il a été suggéré que les protéases de stress à localisation intracellulaire pouvaient également intervenir dans la production de peptides bactériens se retrouvant dans le milieu extracellulaire (voir partie 1.2.2.4 pour plus de détails). Les bac-

téries lactiques en possèdent plusieurs. La première, FtsH, est une métalloprotéase zinc-dépendente appartenant également à la famille des HSPs et rattachée à la membrane. Son site catalytique, contrairement à HtrA, est exposé côté cytosolique et a donc pour cibles principales les protéines cytoplasmiques et membranaires (Ito et Akiyama, 2005). Initialement identifiée chez *E. coli*, elle possède également une double activité chaperon-protéase (Santos et De Almeida, 1975). Elle a depuis été caractérisée chez *L. lactis* (Nilsson *et al.*, 1994), *O. oeni* (Bourdineaud *et al.*, 2003) ou encore *Lb. plantarum* (Fiocco *et al.*, 2009). On peut également noter la présence dans le génome de nombreuses bactéries lactiques d'un homologue de l'endoprotéase zinc-dépendente HtpX de *E. coli* impliquée dans le recyclage des protéines membranaires en conjonction avec FtsH (Sakoh *et al.*, 2005). Cette protéine n'a toutefois pas été étudiée chez les bactéries lactiques, mais des données d'expression sont disponibles chez *S. thermophilus* (Herve-Jimenez *et al.*, 2009).

Finalement, des complexes protéiques appartenant à la famille des protéines Clp assurent le recyclage et la réponse aux stress dans le cytoplasme de la cellule : ClpX, ClpE, ClpC, ClpB, ClpL et ClpP (Papadimitriou *et al.*, 2016). Ces protéines, à l'exception de ClpP, sont des Clp ATPases et ont une fonction de chaperon. ClpX, ClpE et ClpC sont en outre capables de dimériser avec ClpP, une protéase à sérine, formant ainsi un complexe protéique assurant la dégradation des protéines largement endommagées et dont le repliement est impossible. ClpL et ClpB en revanche ne peuvent s'associer avec ClpP et n'assurent qu'une seule fonction de chaperon.

1.2.2.3 Autres activités protéolytiques

Les bactéries lactiques possèdent d'autres activités protéolytiques plus marginales et moins bien documentées en comparaison avec les deux précédents systèmes. On peut tout d'abord mentionner les deux peptidases, DacA et DacB, associées à la paroi bactérienne et impliquées dans la maturation du peptidoglycane (Courtin *et al.*, 2006; Roces *et al.*, 2012). SipL et LspA sont deux peptidases à localisation membranaire respectivement responsables du clivage des séquences signal des protéines sécrétées et des lipoprotéines (Venema *et al.*, 2003; Morello *et al.*, 2008). Enfin, la protéase Eep joue également un rôle dans la maturation des lipoprotéines chez *L. lactis* et *S. thermophilus* (Juillard *et al.*, 2016).

1.2.2.4 Lien entre systèmes protéolytiques et peptidome bactérien extracellulaire

Les systèmes précédemment décrits constituent ainsi les principales activités protéolytiques connues des bactéries lactiques. En raison du relatif faible nombre de protéases impliquées, en particulier en surface des cellules, il a longtemps été pensé que les bactéries lactiques étaient faiblement protéolytiques (Morello *et al.*, 2008). Toutefois, une étude ré-

cente, sans remettre totalement en cause cette affirmation, a permis de mettre en évidence une protéolyse étonnamment étendue chez *L. lactis* (Guillot *et al.*, 2016). Dans cette étude, *L. lactis* IL1403 - une souche PrtP-négative - a été cultivée dans un milieu chimiquement défini ne contenant initialement pas de peptides. Le contenu en peptides bactériens non modifiés du surnageant, prélevé en phase de croissance exponentielle ($DO_{600nm} = 1,0$), a été examiné en spectrométrie de masse et a révélé un nombre important de peptides accumulés (> 1800). En retraçant l'origine des peptides sur les protéines parentales et en déterminant les spécificités de clivage à l'origine de leur libération, les auteurs ont soulevé plusieurs faits marquants. Ils ont tout d'abord confirmé l'implication de HtrA dans la formation de ce peptidome extracellulaire et plus précisément de son action spécifique sur les protéines de surface. De plus, il a été montré qu'il existait toujours une protéolyse de surface chez un mutant $\Delta htrA$, suggérant ainsi la présence d'un ou plusieurs acteurs encore inconnus dans la protéolyse de surface, contrairement aux conclusions d'une précédente étude (Poquet *et al.*, 2000). En outre, un travail en tout point similaire réalisé chez *S. thermophilus* LMD-9 - une souche PrtS-positif - a également montré l'implication de PrtS dans le recyclage des protéines de surface (Boulay *et al.*, résultats non publiés), montrant que ses fonctions sont plus étendues que la seule dégradation des caséines. La double inactivation des gènes *prtS* et *htrA* chez *S. thermophilus* a de même mis en évidence une protéolyse de surface résiduelle d'origine inconnue. Cette conclusion fait particulièrement écho à deux études récentes menées sur la même souche où quatre activités protéolytiques de surface ont été identifiées dans le double mutant $\Delta prtS \Delta htrA$: activités aminopeptidase, X-prolyl dipeptidyl peptidase, peptidyl dipeptidase et carboxypeptidase (Hafeez *et al.*, 2013, 2015). Ces résultats nécessitent maintenant une validation formelle en identifiant le ou les effecteurs de cette protéolyse de surface inconnue.

L'analyse du peptidome extracellulaire de *L. lactis* a également révélé une forte présence de peptides issus de l'hydrolyse de protéines à localisation strictement cytoplasmique, suggérant un phénomène d'export de ces peptides dont le mécanisme est là aussi encore inconnu. Une hypothèse avancée est la capacité récemment découverte des bactéries à Gram positif à sécréter des vésicules extracellulaires (Brown *et al.*, 2015; Kim *et al.*, 2015). Des peptides issus de protéines transmembranaires ont également été détectés dans le peptidome extracellulaire, bien que dans des proportions significativement moindres en comparaison avec les peptides de surface et cytoplasmiques. Enfin, plus d'une centaine de peptides provenant du clivage de la séquence signal de lipoprotéines et de protéines sécrétées ont été identifiés. De nombreuses activités protéolytiques sont donc à l'œuvre durant la croissance bactérienne et sont responsables de la constitution d'un large peptidome extracellulaire. Sachant qu'une analyse *in silico* a révélé la présence de 40 protéases dans le génome de *L. lactis* IL1403 dont moins de la moitié a été caractérisée (Juillard *et al.*, 2016), il est probable que de nombreuses activités protéolytiques restent encore à élucider chez les bactéries lactiques. Il est toutefois

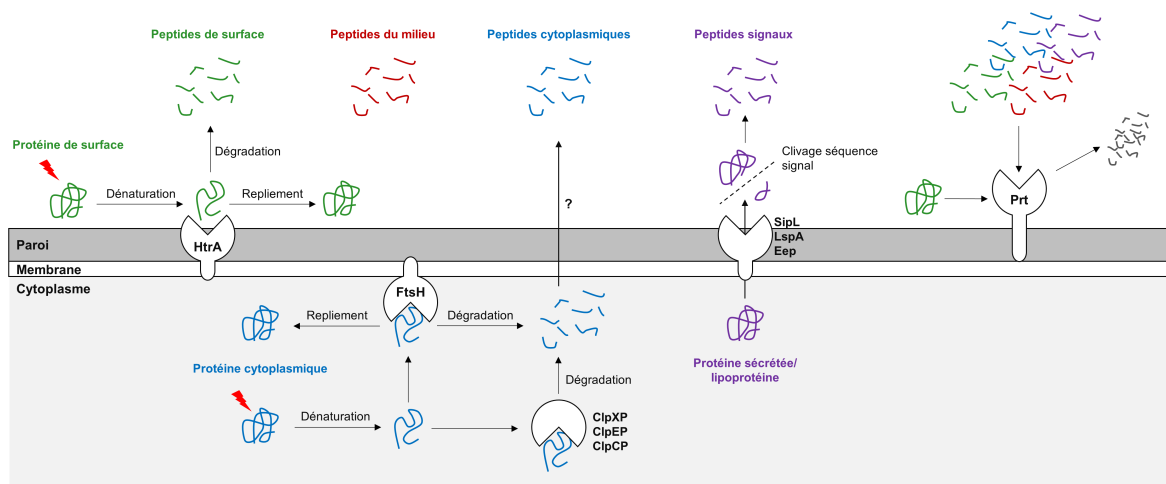


FIGURE 1.4 : Principales enzymes protéolytiques caractérisées chez les bactéries lactiques et impliquées dans l'élaboration du peptidome bactérien extracellulaire. Le contexte est celui d'un milieu riche de type EXL dépourvu de protéines exogènes. Les peptides provenant du recyclage des protéines membranaires ont été omis du fait d'une connaissance limitée des protéases impliquées dans leur libération et de leur faible identification relative dans le peptidome de *L. lactis*. Les peptides cytoplasmiques peuvent être exportés par un mécanisme encore inconnu, où bien être recyclés par des peptidases intracellulaires (non indiqué). Il ne peut être exclu qu'une partie des fragments en cours de dégradation par ces peptidases soient eux-mêmes exportés hors de la cellule. Adapté de Savijoki *et al.* (2006), Guillot *et al.* (2016), Juillard *et al.* (2016), Papadimitriou *et al.* (2016).

possible de proposer un schéma résumant les différentes voies connues par lesquelles le peptidome extracellulaire des bactéries lactiques est élaboré (Figure 1.4).

1.2.3 Peptides résultant d'une biosynthèse bactérienne

Les deux parties 1.2.1 et 1.2.2 précédentes ont permis de mettre en exergue l'étendue du peptidome extracellulaire d'un milieu riche dans un contexte de croissance d'une bactérie lactique. Celui-ci comprend une vaste diversité de peptides résultant d'une hydrolyse protéique, qu'elle soit d'origine biotechnologique (peptides du milieu) ou strictement biologique (peptides de protéolyse bactérienne). Néanmoins, ce panorama ne serait pas tout à fait complet s'il n'était pas fait mention d'une dernière catégorie de peptides composant ce peptidome : les peptides biosynthétisés par la bactérie. On distinguera ici les peptides à synthèse ribosomique d'une part, et les peptides à synthèse non ribosomique d'autre part.

1.2.3.1 Peptides à synthèse ribosomique

Comme leur nom l'indique, les peptides à synthèse ribosomique sont élaborés par les bactéries en suivant les voies classiques de la synthèse protéique, à savoir transcription d'une

séquence codante en ARN messager (ARNm) et traduction de cet ARNm en peptide. La traduction donne naissance à un précurseur qui va ensuite subir, dans la plupart des cas, une hydrolyse afin de libérer le peptide mature. Ce dernier peut ensuite être éventuellement modifié post-traductionnellement par des systèmes enzymatiques spécialisés (Ortega et van der Donk, 2016). Les peptides concernés dans cette partie proviennent donc de séquences codantes de courte taille (sCDSs, small Coding DNA Sequences). La taille maximale des peptides encodés par ces sCDS est arbitraire et varie entre 50 et 100 acides aminés selon les sources, tandis que la taille minimale généralement acceptée est de l'ordre de 10 résidus. Des peptides de plus courte taille sont en effet rarement rencontrés dans la nature. A titre d'exemple, parmi les peptides antimicrobiens répertoriés, seulement 0.6 % d'entre eux ont une taille inférieure à 10 acides aminés (Wang *et al.*, 2009). Une exception notable dans un autre domaine concerne le peptide phéromone contrôlant l'expression du facteur de virulence SpeB de *Streptococcus pyogenes* : SIP (SpeB-Inducing Peptide). Celui-ci est un peptide de 8 acides aminés (MWLLLLFL) synthétisé directement sous forme mature (Do *et al.*, 2017).

La régulation fait partie des différentes fonctions assurées par les peptides et sera développée dans la partie 1.3.2. Il est ainsi largement reconnu que les génomes de nombreuses bactéries, tant à Gram négatif que positif, contiennent de nombreuses sCDSs à l'origine de peptides présentant des fonctions biologiques variées (Hobbs *et al.*, 2011). La caractérisation de nouveaux peptides bactériens repose le plus souvent sur des approches expérimentales ciblées (Rey *et al.*, 2014). En effet, les approches globales dans ce domaine se heurtent encore à des limites computationnelles et expérimentales. Néanmoins, les techniques et outils s'améliorant depuis quelques années, l'étude systématique et à large échelle des peptides bactériens constitue désormais un champ disciplinaire émergent. Cette partie a donc pour but de présenter quelles sont les approches globales disponibles, leurs principales limites et de faire un point sur les données disponibles concernant les peptides encodés dans les génomes des bactéries lactiques.

1.2.3.1.1 Détection *in silico* des courtes séquences codantes

L'étude systématique des peptides bactériens est principalement limitée par la difficulté à détecter *in silico* les sCDSs dans les génomes bactériens. Si les cadres ouverts de lecture de courte taille (sORFs, small Open Reading Frames : séquences génétiques situées entre un codon d'initiation de la traduction et un codon d'arrêt) sont distribués dans les génomes de l'ensemble des êtres vivants et aisément identifiables (Andrews et Rothnagel, 2014), il est en revanche beaucoup plus difficile pour les outils d'annotation de déterminer si ces séquences sont réellement codantes ou non. La base de données génomiques RefSeq (Pruitt *et al.*, 2007) du NCBI (National Center for Biotechnology Information) contient ainsi très peu de gènes codant des protéines d'une taille inférieure à 50 résidus (Warren *et al.*, 2010).

L'annotation des génomes a recours à des approches dites intrinsèques et extrinsèques (Borodovsky *et al.*, 1994). Les approches intrinsèques sont principalement utilisées pour la première phase de l'annotation (annotation structurale) consistant à détecter les séquences codantes (protéines, ARN de transfert, ARN ribosomiaux etc.) et les signaux génétiques associés (sites de fixation des ribosomes, promoteurs, terminateurs etc.; Bryson *et al.*, 2006). Elles font notamment appel à des modèles mathématiques tels que les modèles de Markov cachés qui évaluent le potentiel codant de séquences génétiques en fonction de leur composition nucléotidique (Nicolas *et al.*, 2002). Les approches extrinsèques quant à elles vont surtout être utilisées pour l'annotation fonctionnelle des séquences codantes précédemment détectées. Elles reposent sur la comparaison de séquences homologues par des alignements multiples (BLAST), sur des programmes de détections de motifs et domaines protéiques caractéristiques ou encore sur la conservation de synténie chez des espèces apparentées. Si l'annotation structurale est généralement efficace avec les protéines de taille normale (Bryson *et al.*, 2006), la discrimination entre sORFs codants et non codants est rendue plus délicate par leur courte taille du fait d'une information de séquence plus limitée (Ibrahim *et al.*, 2007b). Pour les mêmes raisons, les approches extrinsèques souffrent à leur tour d'un manque de sensibilité pour l'annotation fonctionnelle des sORFs potentiellement codants. Elles sont en effet limitées par le peu d'information disponible dans les différentes bases de données sur cette classe de molécules. En résumé, il est à la fois difficile de prédire de nouvelles sCDSs à l'échelle du génome et de leur attribuer une fonction biologique. De plus, la validation expérimentale des prédictions est rendue difficile par le fait que la plupart des peptides sont maturés, leur forme active étant plus courte que celle contenue dans les sCDS.

Néanmoins, depuis quelques années, les tentatives se multiplient pour remédier à ces problèmes. Grâce à l'important volume de données de séquençage désormais disponibles, une approche de génomique comparée constitue probablement la solution la plus directe pour identifier de nouveaux peptides dans les génomes bactériens. En effet, la conservation d'un sORF donné dans le génome de différentes espèces est en général un bon indicateur d'une éventuelle fonction biologique, quand bien même son identification à l'échelle individuelle reste assortie d'une importante incertitude (Rey *et al.*, 2014). Warren *et al.* (2010) ont ainsi utilisé une stratégie basée sur des alignements de séquences multiples pour explorer les génomes et plasmides procaryotes de la base de données RefSeq et estimer le nombre de sORFs potentiellement codant mais non-annotés dans la base de données. Plus d'un millier de sORFs correspondent à cette description. Rey *et al.* (2014) ont réalisé une approche analogue sur les données de RefSeq non pas basée sur des alignements multiples mais sur BactgeneSHOW, un programme de détection des CDS utilisant des modèles de Markov cachés pour la détection des sites de fixation des ribosomes et l'analyse de la composition nucléotidique des séquences génétiques (Nicolas *et al.*, 2002). Ce travail a permis la créa-

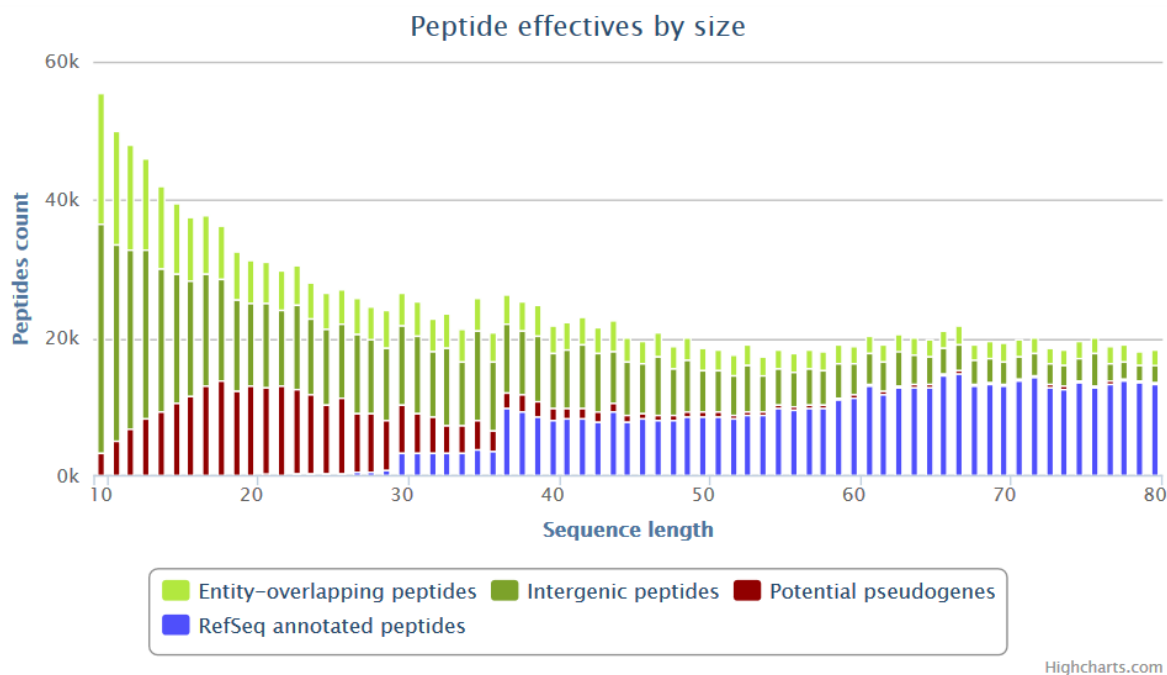


FIGURE 1.5 : Recensement des peptides disponibles dans BactPepDB en fonction de leur taille. Source : BactPepDB, <http://bactpepdb.rpbs.univ-paris-diderot.fr>, mise à jour du 8 juin 2018

tion de la base de données BactPepDB recensant de façon exhaustive tous les peptides de 10 à 80 acides aminés potentiellement encodés dans l'ensemble des génomes bactériens actuellement disponibles dans RefSeq. Ces peptides sont classés selon différents critères (déjà annotés dans RefSeq, pseudogènes potentiels, localisation intergénique ou se superposant à d'autres régions codantes) et sont assortis d'informations biologiques complémentaires (informations structurales, présence dans des espèces proches). Au 8 juin 2018, BactPepDB comprenait ainsi 1 747 413 peptides putatifs provenant de 557 genres bactériens, 1 226 espèces et 2 240 souches différents (Figure 1.5). Bien qu'incluant certainement de nombreux faux-positifs, grâce aux informations et filtres proposés, cette base permet d'explorer et éventuellement cibler de potentiels nouveaux peptides bactériens. Un important critère de sélection consiste notamment en la conservation d'un sORF intra- voire inter- espèce.

1.2.3.1.2 Approches complémentaires pour l'identification de nouveaux peptides bactériens

Afin de réduire l'incertitude liée à l'identification à large échelle de nouveaux peptides, il est possible de coupler les outils de détection *in silico* précédemment cités avec des approches omiques complémentaires. La protéogénomique fait ainsi appel à des stratégies de protéomique et peptidomique pour réannoter les génomes à partir de données de spectrométrie de masse (Hemm *et al.*, 2010; Tinoco et Saghatelian, 2011; Müller *et al.*, 2013).

Finalement, la biologie des ARNs offre également des perspectives intéressantes pour la découverte de nouveaux peptides. Les progrès récents en matière de transcriptomique (puces, RNAseq) ont mis en évidence une activité transcriptionnelle jusqu'alors insoupçonnée chez les bactéries (Wade et Grainger, 2014). De nombreux ARNs de petite taille ont ainsi été découverts et initialement déclarés comme non-codants car non annotés dans les génomes bactériens. Parmi ces ARNs non-codants ont en particulier été caractérisés des ARNs à fonction régulatrice et n'étant effectivement pas traduits en protéines (Waters et Storz, 2009; Storz *et al.*, 2011). Néanmoins, il s'avère qu'une autre fraction de ces ARNs pourrait en réalité être composée de précurseurs de petites protéines ou peptides et seraient donc en réalité codants. Friedman *et al.* (2017) ont ainsi développé un outil (DiSCO-Bac : Discovery of sRNA Coding ORFs in Bacteria) permettant d'explorer l'univers des petits ARNs bactériens à la recherche de ceux supportés par des ORFs codants et susceptibles d'être donc traduits en peptides. Il convient finalement de noter que la séparation entre ARNs régulateurs et précurseurs peptidiques n'est pas stricte car il existe en effet plusieurs exemples d'ARNs possédant cette double fonction (Gimpel et Brantl, 2017). L'exemple le mieux connu est celui de l'ARNIII de *Staphylococcus aureus*. Celui-ci est en effet capable d'une part de réguler l'expression d'au moins 12 autres gènes au niveau traductionnel par appariement avec leur ARNm associé, provoquant selon les cas une activation ou inhibition de la traduction. D'autre part, cet ARNIII possède également un sCDS codant un peptide cytotoxique (Verdon *et al.*, 2009).

1.2.3.1.3 Peptides encodés dans les génomes des bactéries lactiques

Bien que les peptides occupent d'importantes fonctions chez les bactéries lactiques, il existe encore très peu d'études disponibles à l'heure actuelle ayant visé à identifier de manière systématique les peptides encodés dans leurs génomes. Il est toutefois possible de citer un travail précurseur basé sur la génomique comparée de 19 souches de streptocoques et une souche d'*Enterococcus faecalis* qui a permis d'identifier un total de 789 petits ORFs putatifs dans les génomes analysés dont respectivement 35 et 39 chez *S. thermophilus* CNRZ1066 et LMG18311 (Ibrahim *et al.*, 2007b). Cette liste de petits ORFs a en particulier permis de découvrir une nouvelle famille de gènes encodant des peptides de petite taille et hydrophobes appelés SHP (Short Hydrophobic Peptides) et largement distribués dans les génomes des streptocoques. Les SHP ont fait l'objet d'études ultérieures ayant démontré leur implication dans des régulations de type *quorum-sensing* qui seront détaillées plus en détail dans la partie 1.3.2.

Plus récemment, van der Meulen *et al.* (2016) ont utilisé des approches de RNAseq différentielles afin de revisiter le transcriptome de *L. lactis* MG1363. Outre de nouveaux petits ARNs régulateurs, un total de 134 ORFs putatifs possédant des signaux génétiques associés à la traduction ont aussi pu être identifiés. Toutefois, leur traduction effective en peptides n'a pas été validée expérimentalement.

Finalement, à ce jour, les deux études précédemment citées constituent les uniques tentatives menées pour identifier de façon systématique et globale de nouveaux peptides chez les bactéries lactiques. Or, les données fournies par la base de données BactPepDB, bien qu'incluant vraisemblablement une grande part de faux positifs, montrent qu'il existe un grand nombre de sORFs encore inexplorés chez ces bactéries (Tableau 1.5). Le couplage de ces données génomiques avec des approches globales de type RNAseq ou protéomique/peptidomique constitue certainement une stratégie efficace pour identifier de nouveaux peptides bactériens.

TABLEAU 1.5 : Nombre de peptides putatifs de 10 à 80 acides aminés recensés chez les principales bactéries lactiques. Source : BactPepDB.

	<i>Streptococcus thermophilus</i> ^a	<i>Lactococcus lactis</i> ^b	<i>Lactobacillus bulgaricus</i> ^c	<i>Lactobacillus casei</i> ^d	<i>Oenococcus oeni</i> ^e
Total	4941	2165	2690	4185	481
Annoté dans RefSeq	1771	774	803	2347	153
Intergénique	1880	872	1141	1009	188
Chevauchement d'une région codante	574	243	275	550	70
Partiellement dupliqué	716	276	471	279	70
Possible séquence signal	84	15	42	56	4
Possible segments transmembranaires	652	291	295	699	68
Possibles ponts disulfures	596	321	337	384	32
Homologie de structure dans PDB	1177	458	684	885	92

Les résultats correspondent au total de sORFs identifiés à partir des souches bactériennes suivantes :

^a CNRZ1066, JIM8232, LMD-9, LMG18311, MN-ZLW-002, ND03

^b CV56, IL1403, IO-1, KF147

^c 2038, ATCC11842, ATCC BAA-365, ND02

^d ATCC334, BD-II, BL23, LC2W, LOCK919, W56, Zhang

^e PSU-1

1.2.3.2 Peptides à synthèse non ribosomique

En parallèle des peptides à synthèse ribosomique, les bactéries sont également capables produire des peptides par une voie indépendante de l'ARN messager. Le mécanisme de production de ces peptides non ribosomiques (NRPs, NonRibosomal Peptides) a été décrit pour la première fois en 1971 à travers l'étude de deux antibiotiques : la gramicidine S et la tyrocidine (Lipmann *et al.*, 1971). Les NRPs sont fabriqués par le biais de complexes multienzymatiques que l'on retrouve principalement chez les bactéries (*Proteobacteria*, *Actinobacteria*, *Firmicutes*, *Cyanobacteria*) et champignons (*Ascomycota*) encodés dans des clusters génétiques (Wang *et al.*, 2014). Les NRPs ont la particularité de présenter une grande diversité structurale. Ils incorporent en effet des acides aminés non protéinogènes (Walsh *et al.*, 2013), des résidus non protéiques et peuvent présenter une architecture linéaire, cyclique et/ou ramifiée (Kopp et Marahiel, 2007). Ces structures hautement spécifiques leur confèrent diverses propriétés fonctionnelles dont certaines sont largement exploitées à des fins industrielles. On retrouve ainsi parmi ces molécules des surfactants, antibiotiques, toxines, sidérophores ou encore pigments (Martínez-Núñez et López, 2016).

Peu de NRPs ont été à ce jour identifiés et caractérisés chez les bactéries lactiques. L'analyse *in silico* du génome de *Lb. plantarum* WCFS1 a révélé le premier exemple de la présence d'un complexe multienzyme dédié à la synthèse d'un NRP dans ce groupe de bactéries (Kleebezem *et al.*, 2003). Depuis, des systèmes impliqués dans la résistance au stress oxydatif ont été découverts chez *Streptococcus mutans* et *L. lactis* (Wu *et al.*, 2010; Golomb *et al.*, 2017) ainsi qu'un antibiotique, la reutericycline, chez *Lactobacillus reuteri* (Lin *et al.*, 2015). Enfin, une analyse globale de milliers de génomes représentatifs des différents domaines du vivant a détecté des systèmes putatifs dans les génomes de *Lactobacillus brevis* et *Lactobacillus helveticus* (Wang *et al.*, 2014). Pour une vision générale des différents NRPs identifiés et de leurs fonctions, la base de données Norine est spécialisée dans le référencement des NRPs et met à disposition différents outils pour leur analyse (<http://bioinfo.lifl.fr/norine>, Caboche *et al.*, 2008).

1.2.4 Peptidome extracellulaire : bilan

Cette première partie a permis de démontrer que l'environnement immédiat d'une bactérie lactique cultivée en milieu riche était densément peuplé par une grande diversité de peptides organisés en un peptidome complexe. Idéalement, ce peptidome devrait être caractérisé qualitativement (nombre de peptides présents et séquences) et quantitativement (concentration absolue de chaque peptide). Dans les faits, les informations disponibles à ce sujet sont très limitées, mais des conjectures sont toutefois possibles.

Ainsi, dans cet univers peptidique, les peptides provenant du milieu de culture sont vraisemblablement les plus abondants en raison des concentrations auxquels sont généralement employés les hydrolysats cellulaires et protéiques (jusqu'à plusieurs dizaines de g/l). Des concentrations en peptides de l'ordre du millimolaire ont notamment été observées chez un EXL (Figure 1.2). En revanche, la connaissance de la nature exacte de ces peptides et l'étendue de leur diversité reste encore à ce jour limitée. Toutefois, l'analyse récente du peptidome extracellulaire de *L. lactis* (ainsi que *S. thermophilus*, *E. faecalis* et *Listeria monocytogenes*; Guillot *et al.*, 2016) a révélé une activité protéolytique bien plus intense qu'initialement imaginée et responsable de la libération de milliers de peptides dans le milieu extracellulaire. Les peptides du milieu résultant également d'une protéolyse, on peut émettre l'hypothèse que leur diversité est du même ordre de grandeur que celle du peptidome bactérien protéolytique. En comparaison, le nombre de peptides issus d'une biosynthèse bactérienne est très probablement inférieur de plusieurs ordres de grandeur, bien que de nombreux systèmes restent encore certainement à découvrir chez les bactéries lactiques. Enfin, ces rapports d'abondance ne sont évidemment pas figés mais évoluent dynamiquement au cours de la croissance bactérienne. Les peptides du milieu sont initialement les seuls en contact avec la bactérie. Du fait de leur utilisation, leur nombre décroît progressivement tandis que

s'accumulent en parallèle tout ou partie des peptides produits par la bactérie. Ce mélange dynamique et complexe de peptides va assurer diverses fonctions qui seront abordées dans la partie suivante.

1.3 Principales fonctions assurées par les peptides

Dans la partie précédente, les peptides ont été séparés arbitrairement en trois groupes en fonction de leur origine : peptides du milieu, peptides résultant de l'activité protéolytique bactérienne, et peptides biosynthétisés. Toutefois, la séparation entre peptides du milieu et peptides issus de la protéolyse bactérienne est finalement assez artificielle puisque les deux groupes sont issus d'un même processus de dégradation protéolytique intense. Les peptides ainsi formés se retrouvent dans le milieu extérieur au sein d'une nébuleuse peptidique relativement mal caractérisée. De tels peptides seront désormais qualifiés de « peptides de dégradation ». Par opposition, se retrouvent également dans cette nébuleuse des peptides synthétisés et maturés à façon par les bactéries qui sont, dans l'ensemble, mieux décrits d'un point de vue structural et fonctionnel. Conformément au principe de relation quantitative structure à activité (QSAR, Quantitative Structure–Activity Relationship; Perkins *et al.*, 2003), l'activité fonctionnelle d'un peptide découle premièrement de sa structure. A cet égard, on peut considérer que les peptides synthétisés par les bactéries et présentant des modifications structurales spécifiques sont des molécules hautement spécialisées. Ceci est particulièrement bien illustré avec les NRPs (voir partie précédente), et les peptides à synthèse ribosomique ne sont pas en reste car certains d'entre eux subissent d'importantes modifications post-traductionnelles leur conférant des propriétés uniques. Pour un panorama complet de la diversité biochimique de ces peptides modifiés, les lecteurs pourront se référer aux revues de Arnison *et al.* (2013) et Ortega et van der Donk (2016).

Cette partie s'intéressera aux fonctions de nutrition et de régulation qui, à quelques exceptions près dans le cas de la régulation, sont assurées par des peptides « simples », c'est-à-dire ne présentant pas de modifications structurales spécifiques. Ceux-ci constituent en effet la majorité des peptides de dégradation et donc de l'environnement peptidique autour de la bactérie. Il est donc pertinent d'explorer comment la simple structure primaire d'un peptide peut déterminer à la fois l'efficacité de son utilisation nutritionnelle, de son action régulatrice, mais aussi brouiller les cartes entre ces deux fonctions.

1.3.1 Nutrition

La nutrition constitue la fonction de base des peptides de dégradation. Dans le lait, les oligopeptides libérés par l'action des protéases de paroi Prt sur les caséines constituent la principale source d'azote aminé pour les bactéries lactiques (Juillard *et al.*, 1995). Les acides aminés libres et peptides naturellement présents sont en effet limitants pour assurer une croissance bactérienne complète. Un milieu utilisant des hydrolysats cellulaires ou protéiques contient en revanche un riche panel d'acides aminés libres et de peptides permettant un approvisionnement immédiat et suffisant en azote aminé. De fait, l'intérêt des protéases de

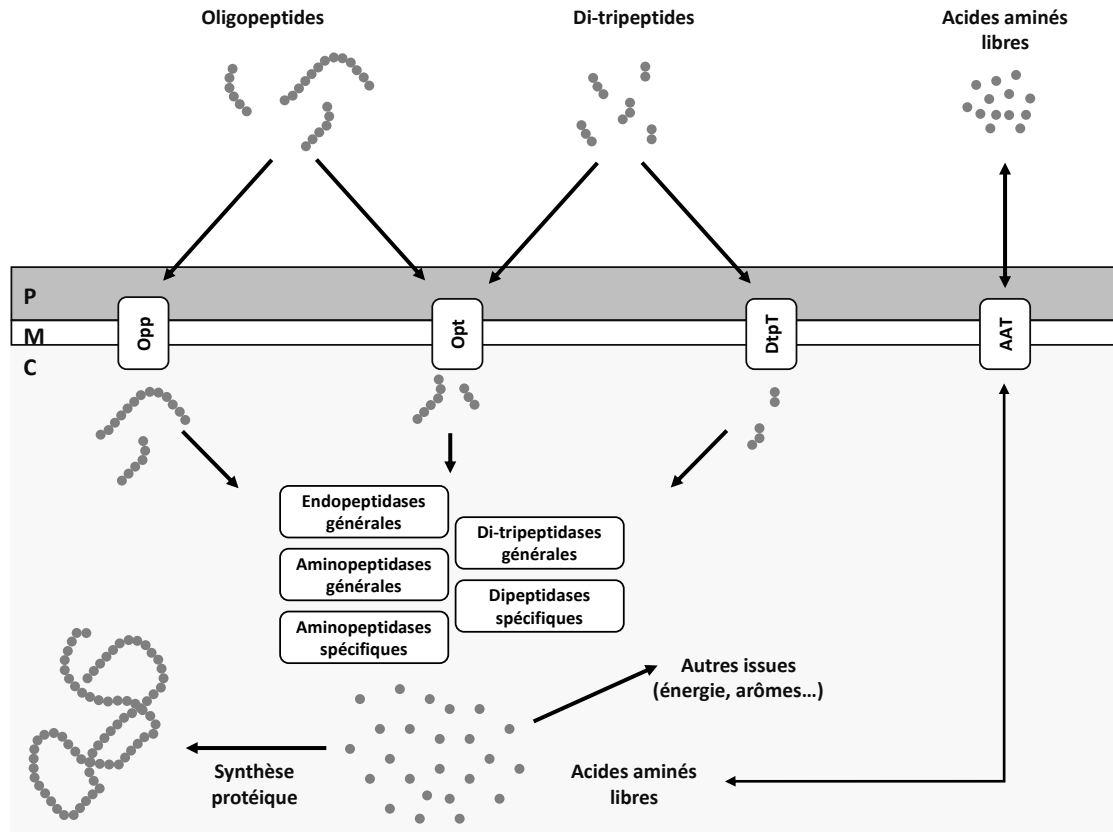


FIGURE 1.6 : Représentation simplifiée des voies d'utilisation des acides aminés libres et peptides chez *L. lactis*. *Opp* ABC transporteur des oligopeptides, *Opt* ABC transporteur des di-tripeptides et oligopeptides, *DtpT* transporteur des di-tripeptides, *AAT* ensemble des transporteurs des acides aminés libres. D'après Kunji *et al.* (1996), Savijoki *et al.* (2006).

paroi pour la croissance dans de tels milieux est limité. A titre d'exemple, un mutant $\Delta prtS$ de la souche *S. thermophilus* CNRZ385 ne voit pas sa croissance être affectée par rapport à la souche sauvage $PrtS^+$ en milieu M17 et dans du lait complétement avec de l'EXL (Courtin *et al.*, 2002). Cette similitude de phénotype peut soit provenir d'une absence complète d'hydrolyse de peptides extracellulaires en fragments de plus courte taille, ce qui est assez peu probable (voir partie 1.2.2.1). Soit les peptides libérés n'apportent aucun gain nutritionnel, hypothèse autrement plus pertinente compte tenue de la surabondance des peptides du milieu.

Il est aussi généralement admis que les peptides issus de la protéolyse bactérienne sont recyclés rapidement pour la synthèse de nouvelles protéines (Visick et Clarke, 1995 ; Becker et Darwin, 2017). Ceci concerne toutefois les peptides hydrolysés intracellulairement, aucune étude n'ayant entrepris de vérifier que leurs homologues extracellulaires étaient également métabolisés. Une seule exception à cela : le cas des mucopeptides issus de la dégradation du peptidoglycane. Il a en effet été montré chez *E. coli* que ceux-ci pouvaient être réutilisés

directement ou après clivage additionnel pour l'addition de nouveaux brins de glycanes au peptidoglycane existant (Goodell, 1985).

Quoiqu'il en soit, les peptides à vocation nutritionnelle, quelle que soit leur origine, vont être successivement transportés, hydrolysés et métabolisés (Figure 1.6). Ces différentes étapes seront revues dans la suite. Une attention particulière sera portée sur l'étape de transport du fait de son rôle dans l'utilisation sélective de certains peptides en lien avec leur séquence propre, et susceptible de créer des dynamiques d'approvisionnement particulières.

1.3.1.1 Transport des peptides

Les bactéries lactiques ont à leur disposition différents systèmes de transport actifs permettant la translocation des peptides à travers la membrane cytoplasmique. On classe généralement les transporteurs de peptides en fonction de la source d'énergie utilisée. On distingue ainsi d'une part les transporteurs actifs primaires appelés ABC-transporteurs (ATP-Binding Cassette) car alimentés par l'hydrolyse de l'ATP, et d'autre part les transporteurs actifs secondaires qui couplent le transport des peptides à la force proton motrice (PMF, Proton Motive Force).

1.3.1.1.1 Transporteurs PTR

1.3.1.1.1.1 Organisation générale des transporteurs PTR

Les cotransporteurs de peptides/protons appartiennent à la famille des transporteurs PTR (Peptide Transport; Steiner *et al.*, 1995) également connue sous le nom de POT (Proton-dependent Oligopeptide Transporter). Cette famille fait elle-même partie d'un groupe plus large appelé MFS (Major Facilitator Superfamily) qui inclut divers transporteurs membranaires dédiés à l'entrée de petites molécules dans la cellule à l'aide de gradients chimiosmotiques (Pao *et al.*, 1998). Les transporteurs de peptides PTRs sont ubiquitaires chez les bactéries et eucaryotes (mais absents chez les archées) où ils prennent en charge principalement des di- et tripeptides (Daniel *et al.*, 2006) et également certains médicaments chez les mammifères (Smith *et al.*, 2013). Ils ont en revanche divergé chez les plantes supérieures où ils prennent en charge le transport de nitrate, d'hormones et de molécules de défense (Newstead, 2017).

La séquence des protéines de la famille MFS est étonnamment bien conservée tant chez les prokaryotes que chez les eucaryotes (Daniel *et al.*, 2006). Les transporteurs PTR en adoptent donc la structure canonique : ils sont organisés en deux domaines protéiques (domaines N et C) chacun constitué de six hélices transmembranaires formant un « V » autour du site

de fixation au substrat (Yan, 2015; Quistgaard *et al.*, 2016). Ces deux domaines sont eux-mêmes sous-divisés en deux répétitions inversées de trois hélices (Figure 1.7), et le site de fixation central est composé de résidus provenant à la fois des domaine N et C. Les transporteurs PTR bactériens possèdent en outre deux hélices additionnelles entre les deux domaines (Newstead, 2017). La stochiométrie exacte régissant le transport des peptides chez les protéines PTR bactériennes est encore mal connue. Entre 1 et 6 protons pourraient en effet être utilisés pour le transport d'un peptide, et ce nombre est susceptible d'être modifié par l'état de charge du peptide (Newstead, 2015, 2017). A ce jour, la structure tridimensionnelle de six transporteurs PTR bactériens a été résolue chez cinq espèces différentes : deux transporteurs chez *Shewanella oneidensis* (Newstead *et al.*, 2011; Guettou *et al.*, 2013), et un transporteur chez *Geobacillus kaustophilus* (Doki *et al.*, 2013), *S. thermophilus* (Solcan *et al.*, 2012), *E. coli* (Zhao *et al.*, 2014) et *Yersinia enterocolitica* (Boggavarapu *et al.*, 2015). Ces travaux ont en particulier permis d'élucider les mécanismes de translocation des peptides de cette famille de transporteurs et les résidus impliqués dans la fixation des substrats (Figure 1.8). A noter que le fonctionnement de ces transporteurs est réversible et dépend de la différence de pH et de potentiel de membrane $\Delta\Psi$ entre la cellule et le milieu extérieur. Il est toutefois probable que dans des conditions physiologiques usuelles, ce transporteur fonctionne principalement dans le sens de l'import.

1.3.1.1.1.2 DtpT, la perméase PMF-dépendante des di-tripeptides chez les bactéries lactiques

Les bactéries lactiques possèdent un unique représentant de la famille des transporteurs PTR appelé DtpT (Di-tri-peptide Transporter). Ubiquitaire, cette protéine a notamment constitué le premier exemple connu des transporteurs PTR lors de son identification et caractérisation initiale chez *L. lactis* (Hagting *et al.*, 1994, 1997; Fang *et al.*, 2000). Un homologue a également été caractérisé chez *Lb. helveticus* peu de temps après (Nakajima *et al.*, 1997). Ces différents travaux réalisés *in vitro* ont permis de mettre en avant la plus forte affinité de transport de DtpT envers les dipeptides, et sa préférence pour les résidus hydrophobes par rapport aux chargés. Les récents travaux de caractérisation structurale, en particulier chez *S. thermophilus*, ont finalement élucidé les mécanismes à l'origine de ces préférences (Solcan *et al.*, 2012; Martinez Molledo *et al.*, 2018). Premièrement, la préférence envers les dipeptides a une origine thermodynamique. DtpT transporte en effet les dipeptides et tripeptides selon deux modes de fixation différents (Parker *et al.*, 2014). Les dipeptides sont cotransportés en compagnie de 5 à 6 protons contre 3 pour les tripeptides, ce qui favoriserait leur accumulation intracellulaire. Ensuite, la spécificité biochimique du transporteur trouve son origine dans la topologie propre de son site de fixation et la façon dont celui-ci interagit avec le peptide transporté. Brièvement, des interactions se réalisant majoritairement avec la chaîne principale et universelle du peptide - en particulier des

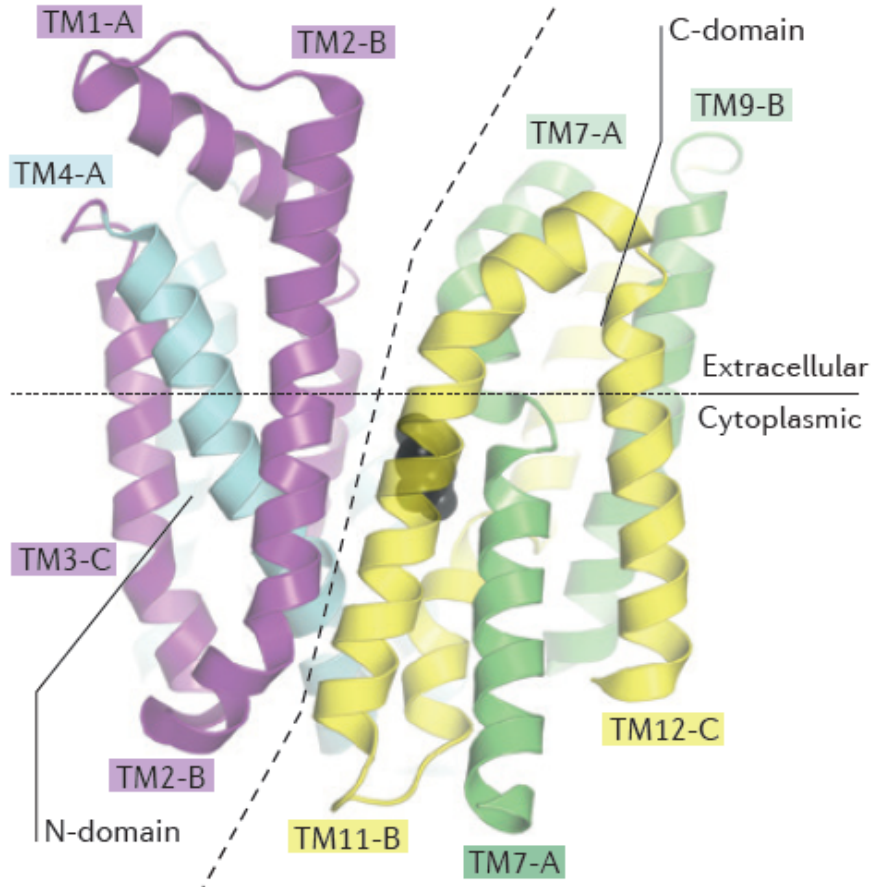


FIGURE 1.7 : Structure canonique des protéines de transport de la superfamille MFS à travers l'exemple du transporteur humain de glucose 3 (GLUT3, code PDB : [4ZW9](#)) vu de côté à une résolution de 1.5 Å. Les sphères noires centrales représentent le substrat (glucose). Les deux domaines N et C sont de façon identique composés de deux triplet d'hélices. Le premier triplet est coloré en violet, le second en bleu, le troisième en jaune et le quatrième en vert. Les hélices transmembranaires (TM) sont numérotées et labellisées A, B ou C selon qu'elles représentent la première, deuxième ou troisième hélice d'un triplet. Certains éléments éléments cytoplasmiques ont été omis pour des raisons de clarté. Quistgaard *et al.* (2016).

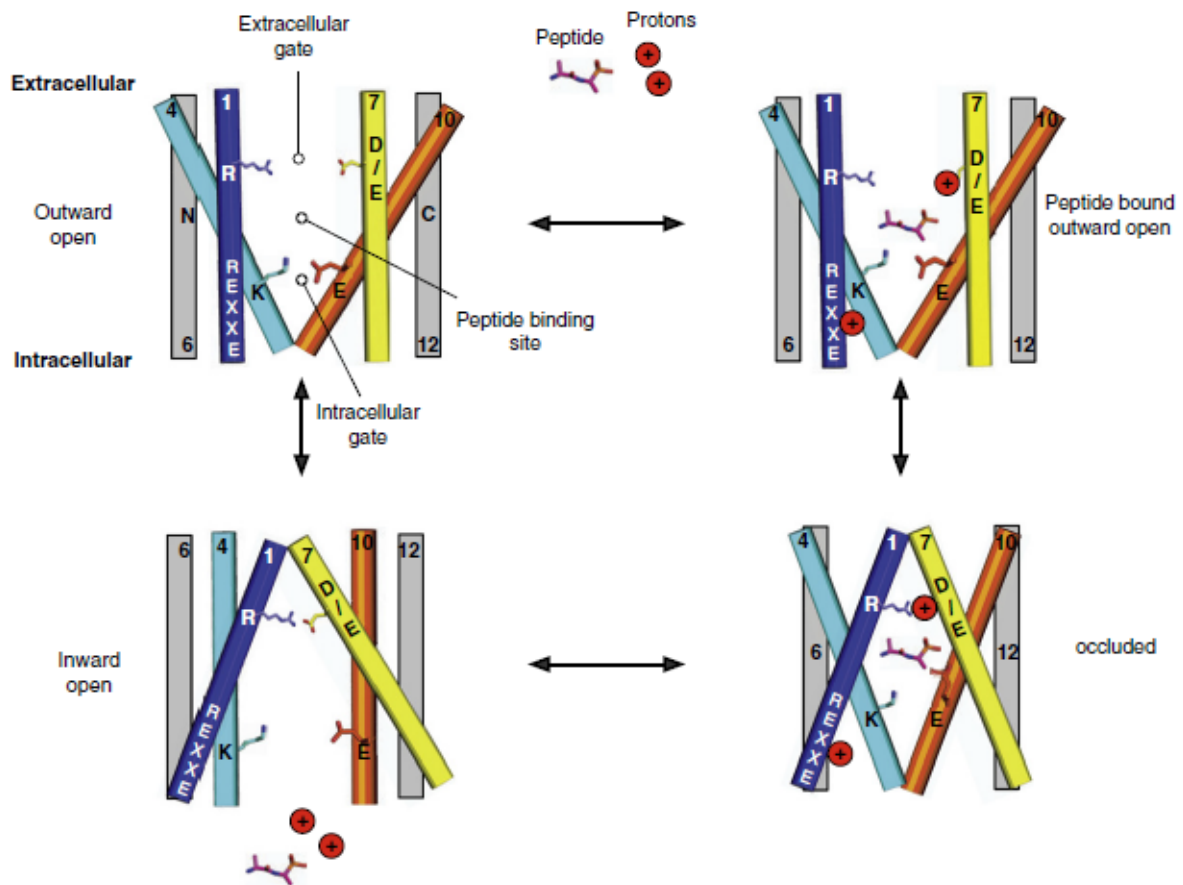


FIGURE 1.8 : Cycle de fonctionnement d'un transporteur de di-tripeptides de la famille PTR. Le transport du peptide et des protons est réalisé à travers une succession de changements conformationnels alternant entre un état ouvert vers l'extérieur et un état ouvert vers l'intérieur permettant respectivement la capture et la libération des substrats. Entre ces états le transporteur se trouve dans un état intermédiaire dit obstrué durant lequel le site de fixation est bloqué et non accessible. L'état ouvert vers l'extérieur est caractérisé par un resserrement des hélices 4 et 5 avec les hélices 10 et 11 qui forment une porte intracellulaire stabilisée par des ponts salins entre les résidus K126 et E400. Le transport est ensuite initié par la fixation de protons sur le motif ExxERFxYY de l'hélice 1 et sur des résidus acides de l'hélice 7. L'entrée du peptide provoque ensuite un changement conformationnel conduisant à l'obstruction complète du site de fixation grâce à la formation d'une porte extracellulaire stabilisée par des ponts salins entre R53-E312 et R33-E300. Un mouvement des protons au sein de cette structure couplé à la déstabilisation de l'interaction K126-E400 provoque ensuite l'ouverture du système du côté intracellulaire permettant la libération du peptide et des protons. Le cycle est ensuite réinitialisé par la reformation de la porte intracellulaire. Solcan *et al.* (2012), Newstead (2017).

interactions de type ponts salins (combinaison d'une liaison hydrogène et ionique) avec les extrémités amino et carboxy-terminales - vont faire pencher la nature du transport vers la multispécificité, voire l'aspécificité. A l'inverse, des interactions préférentielles avec certaines chaînes latérales vont au contraire augmenter la spécificité du transporteur. Se référer à la Figure 1.8 pour les mécanismes de transport généraux. En ce qui concerne DtpT de *S. thermophilus*, il est intéressant de noter la présence de quatre sous-cavités - ou poches - au sein du site de fixation de la perméase dans lesquelles se logent les chaînes latérales du peptide transporté. La deuxième de ces cavités, qui interagit avec le deuxième résidu de la chaîne peptidique, est particulièrement riche en résidus aromatiques (tyrosines) et offre donc un environnement favorisant les interactions hydrophobes (Martinez Molledo *et al.*, 2018), ce qui est concordant avec les préférences de transport observées *in vitro* chez plusieurs homologues de cette protéine.

Ainsi, si les mécanismes généraux gouvernant le fonctionnement des transporteurs PTR sont bien conservés chez les eucaryotes et procaryotes, la spécificité biochimique et la thermodynamique du transport de peptides sont variables en fonction des espèces, en particulier chez les bactéries. Le degré de spécificité envers une classe biochimique particulière de peptides résulte de la topologie propre du site de fixation. La relation structure-affinité de transport permet ainsi d'explicitier les préférences observées de la perméase DtpT des bactéries lactiques pour les dipeptides hydrophobes. Cette notion de spécificité de transport se retrouve également au sein de l'autre classe de transporteurs de peptides présents chez les bactéries lactiques : les-ABC transporteurs.

1.3.1.1.2 ABC-transporteurs

1.3.1.1.2.1 Organisation générale des ABC-transporteurs

Les ABC-transporteurs forment l'une des plus importantes familles de protéines connues et sont répandus dans les trois domaines du vivant où ils prennent en charge une grande variété de substrats (Higgins, 1992). Les ABC-transporteurs utilisent l'énergie fournie par l'hydrolyse de l'ATP et sont généralement classés selon leur sens de fonctionnement : importeurs ou exporteurs (Biemans-Oldehinkel *et al.*, 2006). Ces structures, contrairement aux transporteurs PTR, ne fonctionnent en effet que de façon unidirectionnelle. Les deux groupes sont basés sur une architecture centrale commune : deux domaines transmembranaires (TMDs, TransMembrane Domains) formant le canal de translocation, ainsi que deux domaines de fixation de l'ATP (NBDs, Nucleotide-Binding Domains) reliés aux TMDs côté cytoplasmique (Figure 1.9; Biemans-Oldehinkel *et al.*, 2006; Locher, 2016). Du point de vue de leur séquence, les ABC-transporteurs sont caractérisés par des motifs hautement conservés au niveau de leurs NBDs, tandis que les TMBs présentent des structures plus

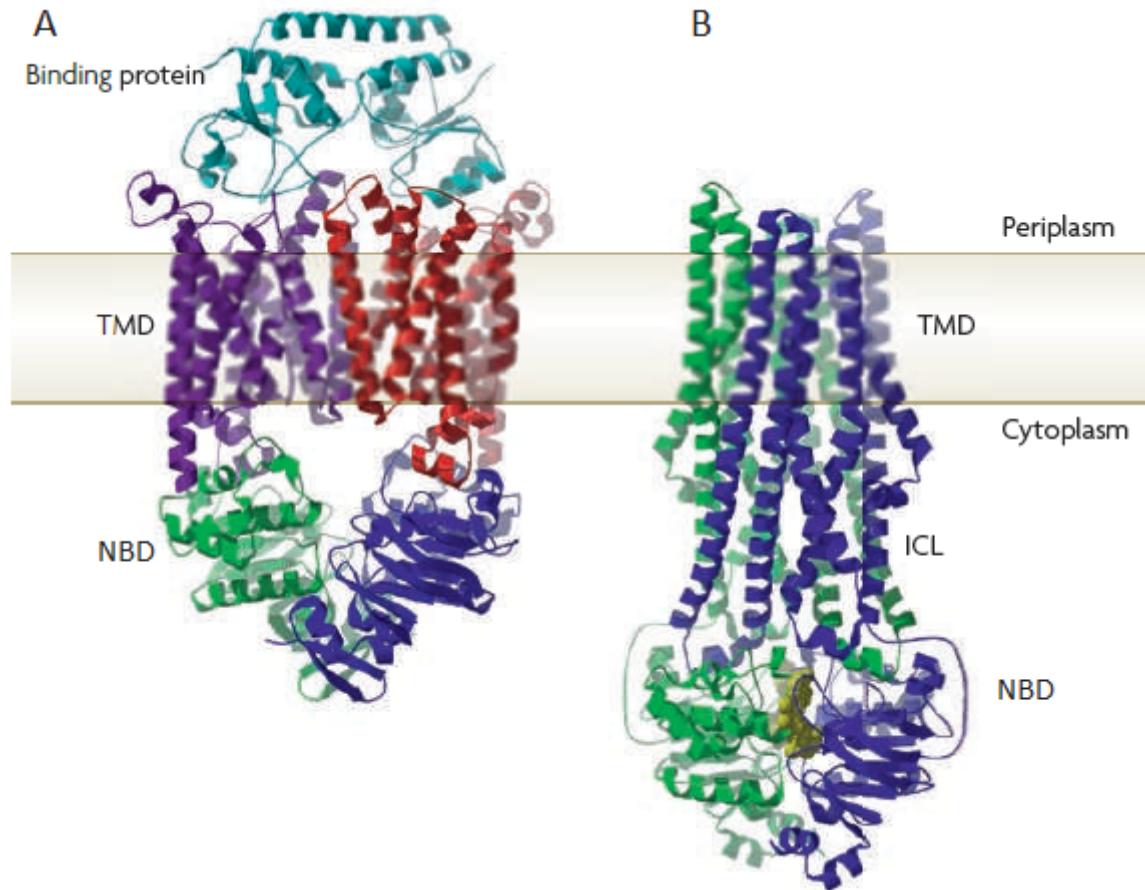


FIGURE 1.9 : Architecture modèle des ABC-transporteurs. (A) Importeur de la vitamine B12 d'*E. coli* (BtuCDF22, code PDB : [2QI9](#)). Le complexe protéique représenté est constitué des deux TMDs (BtuC, violet et rouge), des deux NBDs (BtuD, vert et bleu), ainsi que d'une copie de BtuE, la SBP à localisation périplasmique (cyan). (B) Protéine d'export multimédicamenteuse (multidrug exporter) de *Staphylococcus aureus* (Sav1866, code PDB : [2ONJ](#)) contenant des domaines TMD et NBD fusionnés. Les nucléotides fixés dans cette structure sont représentés en jaune. ICL boucle intracellulaire. Rees *et al.* (2009).

variables et adaptées au substrat transporté (Rees *et al.*, 2009). Les NBDs présentent ainsi au niveau de leur cœur catalytique un motif A de Walker ou *P-loop* (GXXGXGK[S/T]), un motif B de Walker ($\phi\phi\phi\phi$ D, où ϕ est un résidu aromatique), ou encore la signature caractéristique des ABC-transporteurs LSGGQ au niveau de leur domaine hélice- α .

Les ABC-importeurs sont une spécificité des prokaryotes. Ils sont divisés en trois sous-groupes : ABC-importeurs de type I, II ou III (Rice *et al.*, 2014). Cette distinction est basée sur des différences d'architecture générale et des variations dans leurs mécanismes de transport. En particulier, les ABC-transporteurs de type I et II utilisent tous les deux un cinquième domaine protéique : la protéine de fixation au substrat (SBD, Substrate-Binding Domain, Figure 1.9A). Les SBPs sont soit présentes sous forme libre dans l'espace périplasmique des bactéries à Gram négatif, soit rattachées à la membrane des bactéries à Gram positif et archées par un résidu N-terminal lipidique, ou plus marginalement fusionnées aux

TABLEAU 1.6 : Classification structurale des SBPs et corrélation avec la nature du substrat transporté. Scheepers *et al.* (2016).

Cluster	Sous-cluster	Nature du ligand	Associé avec	Total (2016)
A	A-I	Ions métalliques	ABC-transporteur de type II	55
	A-II	Sidérophores, hèmes, vitamine B12, molybdate, tungstate	ABC-transporteur de type II	
B	B-I	Sucres, polyols, autoinducteur 2	ABC-transporteur de type I, TCS, facteur de transcription	129
	B-II	Acides aminés	ABC-transporteur de type I	
	B-III	Acides aromatiques	ABC-transporteur de type I	
	B-IV	Composés natriurétiques	Récepteur de peptide natriurétique	
	B-V	Glutamate	Récepteur couplé à une protéine G	
C	C	Di-tripeptides, oligopeptides, cellobiose, opine, nickel	ABC-transporteur de type I	30
D	D-I	Sucres, polyols, spermidine/putrescine	ABC-transporteur de type I	108
	D-II	Thiamine, spermidine/putrescine	ABC-transporteur de type I, facteur de transcription	
	D-IIIa	Phosphate, molybdate, tungstate	ABC-transporteur de type I	
	D-IIIb	Molybdate, tungstate	ABC-transporteur de type I	
	D-IV	Ions ferreux	ABC-transporteur de type I	
E	E-I	Acides organiques, alcools aminés, dipeptides, glycérol-3-phosphate	Transporteur TRAP	70
	E-II	Acides aminés	Transporteur TTT	
F	F-I	Thiamine, pyrimidines, sulfonates, bicarbonate	ABC-transporteur de type I	106
	F-II	Méthionine	ABC-transporteur de type I	
	F-III	Solutés compatibles	ABC-transporteur de type I	
	F-IV	Acides aminés	ABC-transporteur de type I, récepteur ionotrope, récepteur couplé à une protéine G	

TMDs (van der Heide et Poolman, 2002). Elles ont pour rôle la reconnaissance, la fixation et la fourniture du substrat au translocon. Les transporteurs de type I et II divergent en revanche au niveau de l'organisation de leur TMDs. Les importeurs de type I contiennent en effet un plus faible nombre d'hélices transmembranaires (5 ou 6 hélices par domaine) par rapport au type II (10 hélices par domaine; Locher, 2009). Les importeurs de type III, également connus sous le nom de transporteurs ECF (Energy-Coupling Factor), constituent une addition récente à cette classification (Rodionov *et al.*, 2009). Ils ont la particularité de ne pas utiliser de SBP mais un composant S, une protéine membranaire intégrale, se substituant à un TMD et se chargeant de la capture du substrat (Rice *et al.*, 2014).

Les transporteurs de type I et II constituent néanmoins la majorité des exemples connus d'ABC-importeurs. La nature des substrats transportés est ainsi dictée par les SBPs associées à ces systèmes. Bien que les ABC-transporteurs constituent les partenaires majoritaires des SBPs, ces protéines peuvent également être associées avec d'autres structures protéiques telles que les transporteurs procaryotes secondaires de type TRAP (Tripartite ATP-independent Periplasmic transporter) ou TTT (Tripartite Tricarboxylate Transporter; Rosa *et al.*, 2018), certains systèmes à deux composants procaryotes (TCS, Two Component Systems; Neiditch *et al.*, 2006) ou encore les récepteur couplés à une protéine G et récepteurs ionotropes des eucaryotes (Armstrong et Gouaux, 2000). Les SBPs reposent sur une architecture de base commune. Elles sont formées de deux domaines α/β , chacun constitué d'un feuillet β central comprenant cinq brins β et entouré d'hélices- α (Berntsson *et al.*, 2010). Ces deux domaines forment des lobes connectés par une région charnière accordant, en

l'absence de substrat, une flexibilité au système avec les deux domaines pivotant autour de la charnière. Lorsqu'une molécule de ligand est capturée, les deux domaines se referment et l'enferment dans une cavité - le site de fixation - à l'interface entre les deux domaines, selon un processus nommé « *Venus fly trap* » (Mao *et al.*, 1982). Toutefois, en dépit de cette architecture commune et fonctionnement similaire, les SBPs présentent des différences structurales reflétant leur adaptation pour la capture d'une grande variété de substrats. Scheepers *et al.* (2016) ont ainsi récemment proposé une classification des différentes SBPs répertoriées dans la Protein Data Bank en fonction de leur homologie structurale (Tableau 1.6).

Les ABC-importeurs de peptides font ainsi partie des transporteurs de type I précédemment décrits et sont associés avec des SBPs appartenant au cluster C. Celles-ci se démarquent des autres SBPs par un domaine supplémentaire entre les deux lobes canoniques et ont une taille moyenne significativement supérieure (Berntsson *et al.*, 2010; Scheepers *et al.*, 2016). Ces ensembles protéiques vont permettre l'approvisionnement en peptides selon un mécanisme général décrit dans la Figure 1.10.

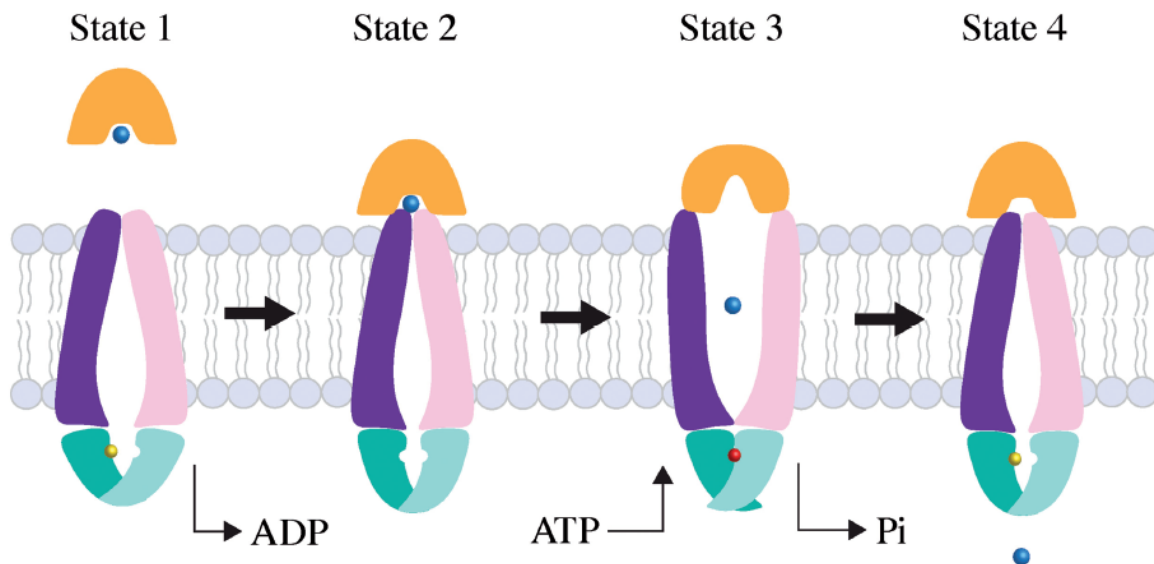


FIGURE 1.10 : Mécanismes d'import des ABC-transporteurs de type I. Les deux TMDs sont représentés en violet et rose, les deux NBDs en cyan et magenta, la protéine de fixation en orange, et le substrat par une sphère bleue. L'ATP et l'ADP sont respectivement représentés par une sphère jaune et rouge. Dans l'état 1, le transporteur se trouve dans un état ouvert vers l'intérieur. La fixation du couple SBP/substrat induit une fermeture partielle des deux NBDs et la libération d'ADP résiduel (état 2). La fixation de l'ATP déclenche ensuite la fermeture complète des deux NBDs, la réorientation des TMDs et l'ouverture de la SBD qui relargue son substrat dans le translocon (état 3). L'hydrolyse de l'ATP repositionne ensuite le transporteur dans un état ouvert vers l'intérieur, libérant le substrat dans le cytoplasme (état 4). Adapté de Rice *et al.* (2014).

1.3.1.1.2.2 ABC-transporteurs de peptides des bactéries lactiques

Les bactéries lactiques possèdent différents ABC-transporteurs dédiés à l'import de peptides. A ce jour, quatre systèmes ont été identifiés : Opp, Opt, Ami et Ots. Le système Opp est de loin le système le mieux connu et le plus répandu parmi de nombreuses espèces : *L. lactis* (Tynkkynen *et al.*, 1993), *Lb. bulgaricus* (Peltoniemi *et al.*, 2002), *Leuconostoc mesenteroides* (Germain-Alpettaz et Foucaud-Scheunemann, 2002) ou encore *Lb. helveticus* (Cremonesi *et al.*, 2013). Son ubiquité est telle qu'on le retrouve chez des espèces plus éloignées telles que *E. coli* (Klepsch *et al.*, 2011) ou encore *Salmonella typhimurium* (Tame *et al.*, 1994). Opt est un système identifié chez *L. lactis* et pour lequel règne une certaine confusion dans la littérature. Il a en effet été initialement appelé DtpP (Foucaud *et al.*, 1995), puis renommé Dpp (DiPeptide transporter) et Opt (OligoPeptide Transporter) par deux équipes différentes (Sanz *et al.*, 2001, 2003; Lamarque *et al.*, 2004, 2011). Les deux noms sont encore employés simultanément de nos jours dans la littérature en fonction des sources. Par mesure de simplification, Opt sera le nom présentement utilisé car il rend compte avec plus de justesse des modes de transport de ce système. Ami (ou Ali chez certains espèces), est le pendant d'Opp des streptococci et se retrouve donc chez *S. thermophilus* (Garault *et al.*, 2002) ou *Streptococcus pneumoniae* (Kerr *et al.*, 2004). Enfin, Ots est un système récemment identifié chez *S. thermophilus* (Goh *et al.*, 2011) et facultativement présent chez certaines souches uniquement (Jameh *et al.*, 2016).

Ces systèmes diffèrent au niveau de leur organisation génétique ainsi qu'au niveau du nombre de SBPs associées à chaque transporteur (Figure 1.11). Ils montrent également des capacités de transport variables en termes de tailles de peptides pris en charge. Opp de *L. lactis* et Ami de *S. thermophilus* sont en effet capables de transporter des peptides sur une gamme de tailles très étendue : de 4 à plus de 30 acides aminés pour Opp (limite de 35 acides aminés observée *in vitro*; Doeven *et al.*, 2004), et de 3 à plus de 20 pour Ami de *S. thermophilus* (limite de 24 acides aminés observée *in vivo*; Garault *et al.*, 2002). Cette capacité de transport étendue tranche radicalement avec les aptitudes plus limitées observées avec les systèmes Opp d'*E. coli* (2 à 5 acides aminés; Guyer *et al.*, 1986), de *S. typhimurium* (2 à 5 acides aminés; Tame *et al.*, 1994) ou encore le système App de *B. subtilis* (nonapeptides; Levnikov *et al.*, 2005). De même, les récents travaux de caractérisation du système Ots suggèrent son implication dans le transport de peptides de courte taille, éventuellement des di-tripeptides (Jameh *et al.*, 2016). Enfin, le transporteur Opt de *L. lactis* constitue un cas intéressant du fait des spécificités complémentaires des deux SBPs qui lui sont associées (Figure 1.11). En effet, contrairement par exemple au système Ami de *S. thermophilus* où les trois SBPs (AmiA1, AmiA2, AmiA3) présentent des spécificités proches, les deux SBPs OptS et OptA prennent respectivement en charge des dipeptides et des oligopeptides (de 3 à 6 résidus; Lamarque *et al.*, 2011). Cette dualité de transport est d'ailleurs à l'origine des désaccords concernant le nom de ce transporteur. Celui-ci a en

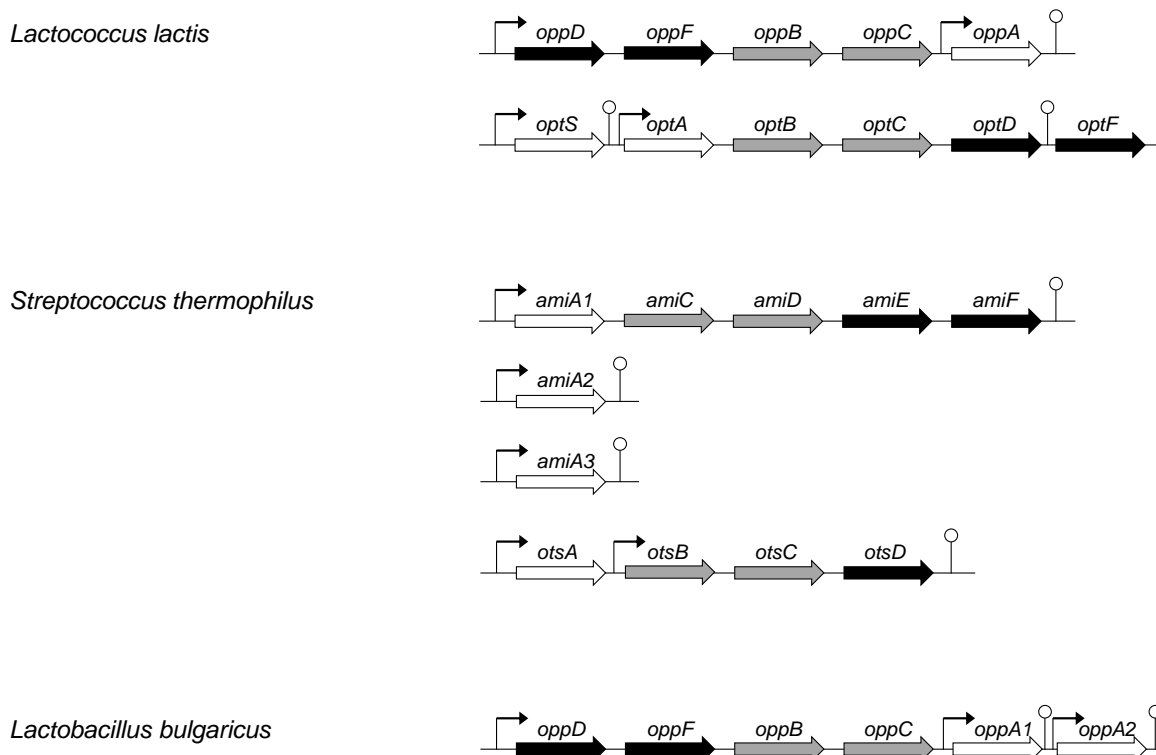


FIGURE 1.11 : Organisation génétique des ABC-transporteurs de peptides de bactéries lactiques. Les gènes des domaines de fixation de l'ATP (NBDs), des domaines transmembranaires (TMDs) et des protéines de fixation (SBPs) sont respectivement représentés en noir, gris et blanc. Adapté de Doeven *et al.* (2005), Atlan *et al.* (2008), Jameh *et al.* (2016).

effet été étudié en premier chez la souche MG1363 dans laquelle le gène *optA* n'est plus fonctionnel en raison d'une mutation. En raison de ce contexte génétique particulier, ce transport a ainsi été déclaré à tort comme strictement un transporteur de dipeptides, d'où le nom Dpp.

1.3.1.1.2.3 Relation structure-affinité de transport : le cas de OppA

Si la taille des peptides transportés représente le premier critère de distinction entre différents ABC-transporteurs, d'autres facteurs sont susceptibles d'intervenir dans les spécificités de transport d'un système donné. Similairement à ce qui a été décrit précédemment avec les transporteurs PTR, ces spécificités sont le reflet d'une organisation structurale favorisant le passage de peptides pourvus de propriétés physico-chimiques particulières découlant de leur structure primaire. L'originalité des ABC-importeurs vient du fait que ce sont les SBPs associées à ces systèmes qui dictent principalement les spécificités de transport (Doeven *et al.*, 2004). A cet égard, le système Opp de *L. lactis* et sa SBP associée, OppA, est probablement l'un des ABC-transporteurs les mieux décrits toutes espèces bactériennes confondues et constitue un cas d'étude intéressant.

L'activité de ce transporteur a en effet été caractérisée à la fois *in vivo* (Tynkkynen *et al.*, 1993; Juillard *et al.*, 1995, 1998; Kunji *et al.*, 1995; Detmers *et al.*, 1998; Picon *et al.*, 2000; Charbonnel *et al.*, 2003; Helinck *et al.*, 2003), *in vitro* (Lanfermeijer *et al.*, 1999, 2000; Detmers *et al.*, 2000; Doeven *et al.*, 2004), et la structure tridimensionnelle de la SBP OppA a également été résolue (Berntsson *et al.*, 2009, 2011). Des positions contradictoires ont longtemps été tenues concernant les spécificités de transport de ce système. Si toutes les études s'accordent pour reconnaître sa faculté à prendre en charge une large gamme de tailles - la taille maximale de 35 acides aminés ayant été observée *in vitro* (Doeven *et al.*, 2004) - divers travaux ont abouti à des conclusions opposées concernant d'autres spécificités de transport. Ainsi, Juillard *et al.* (1998) ont très tôt proposé un modèle de transport dépendant non seulement de la taille, mais également de l'hydrophobicité et de la charge nette du peptide. Il a été en effet observé *in vivo* un transport favorisé pour des peptides basiques, hydrophobes et d'une taille comprise entre 600 et 1100 Da, et un transport aboli pour les longs peptides anioniques. Une étude ultérieure du même laboratoire a accentué le rôle déterminant d'une charge positive pour le transport, et a également suggéré un modèle de compétition de transport dans lequel des peptides négativement chargés seraient fixés par la SBP OppA mais non transportés à travers le translocon OppDFBC (Helinck *et al.*, 2003). En parallèle de ces travaux, un autre laboratoire a au contraire longtemps soutenu l'absence totale d'influence de la composition en acides aminés d'un peptide sur son transport *via* le système Opp, tant chez *L. lactis* que chez d'autres espèces bactériennes (Doeven *et al.*, 2004). Ce modèle, hérité des travaux de caractérisation biochimique et structurale d'OppA de *S. typhimurium* (Tame *et al.*, 1994, 1996; Sleight *et al.*, 1997, 1999) voulait que les seules interactions possibles entre une SBP et un peptide se situent exclusivement au niveau de la chaîne principale de celui-ci, les chaînes latérales étant confinées indifféremment dans des poches hydratées à l'intérieur du site de fixation. Seule la taille était considérée comme déterminante. Néanmoins, l'étude de la structure de OppA par cristallographie aux rayons X a initié la révision de ces conclusions en révélant une situation plus complexe (Berntsson *et al.*, 2009).

Avant toute chose, l'étude de la topologie du site de fixation d'OppA a d'abord permis d'élucider la façon dont cette protéine pouvait accommoder une gamme aussi large de tailles. Les limitations de taille du système Opp de *S. typhimurium* (Tame *et al.*, 1994, 1995), App de *B. subtilis* (Levdikov *et al.*, 2005) ou encore Dpp (transporteur de dipeptides) d'*E. coli* (Dunten et Mowbray, 1995) sont en effet le résultat d'interactions non-covalentes de type ponts salins entre les extrémités amino- et carboxyterminales du peptide et des résidus conservés du site de fixation (Figure 1.12). Ces liaisons permettent notamment une interaction à haute affinité entre la SBP et le peptide transporté et ancrent ce dernier dans des positions fixes au sein de la protéine. Or, les résidus responsables de telles interactions sont absents chez OppA de *L. lactis*, ce qui permet à des peptides de tailles variées d'interagir avec la protéine

de fixation. De même, la taille maximale de transport est très bien corrélée avec le volume de la cavité interne dans laquelle les peptides sont enfermés. Sans surprise, OppA de *L. lactis* dispose de la cavité la plus volumineuse identifiée jusqu'à présent ($\approx 4900 \text{ \AA}^3$; Berntsson *et al.*, 2009), tandis que DppA d'*E. coli* est l'une des plus étroites ($\approx 700 \text{ \AA}^3$; Dunten et Mowbray, 1995). L'absence de ponts salins est néanmoins préjudiciable pour un transport de peptides à haute affinité. Cette absence est compensée chez OppA de *L. lactis* par la présence d'une poche hydrophobe au sein de la cavité interne qui permet la fixation préférentielle et à haute affinité de résidus à chaîne latérale hydrophobe, en particulier les acides aminés branchés (Figure 1.13). Cette stratégie n'est d'ailleurs pas sans rappeler le cas analogue de DtpT (voir partie 1.3.1.1.2). Enfin, cette étude a également révélé une haute affinité d'OppA pour des peptides riches en proline.

Ces travaux de caractérisation structurale ont donc permis de mettre en évidence une sélectivité dans le transport d'oligopeptides de *L. lactis* du fait d'une plus forte affinité du système Opp pour des peptides hydrophobes riches en proline. Si le facteur hydrophobicité concorde bien avec le modèle de transport observé *in vivo* par Juillard *et al.* (1998), l'influence de résidus proline n'avait été en revanche jamais observée tant *in vivo* qu'*in vitro*. De même, l'influence positive de résidus basiques mentionnée par ces derniers n'a pas trouvé d'explications dans la structure d'OppA de *L. lactis*. En revanche, il est important de préciser que chez *E. coli*, OppA démontre effectivement une préférence pour les peptides positivement chargés (Klepsch *et al.*, 2011). Cette spécificité vient du fait que la surface de son site de fixation est tapissée de charges négatives qui permettent d'attirer les peptides basiques. Etonnamment, il s'avère que cette découverte s'applique également à OppA de *S. typhimurium*, contrairement à tout ce qui avait été auparavant décrit à son sujet. En réalité, les précédents travaux, qui avaient tous conclu que le transport dans cette protéine était indépendant de la composition en acides aminés (Tame *et al.*, 1996; Sleight *et al.*, 1997, 1999), avaient quasi systématiquement fait usage de peptides positivement chargés (présence de lysines) pour étudier les spécificités de transport, biaisant de fait les résultats obtenus. S'il est tentant de vouloir transposer ce modèle de transport de peptides basiques à OppA de *L. lactis* afin de réconcilier définitivement les différents points de vue, il faut néanmoins rappeler que i) aucune donnée de caractérisation structurale ne supporte ce modèle chez cette bactérie, et ii) les OppA respectives d'*E. coli* et *S. typhimurium* partagent entre elles une forte identité de séquence (85.7 %; Klepsch *et al.*, 2011) alors que l'homologue de *L. lactis* ne présente qu'une identité très modérée (< 25 %). Toutefois, il convient finalement de signaler une dernière étude réalisée *in vivo* avec *S. thermophilus* (Juille *et al.*, 2005) selon une méthodologie très proche de celle adoptée par Juillard *et al.* (1998) avec *L. lactis*. Ces deux travaux ont en effet abouti à des conclusions très proches chez les deux espèces, en particulier concernant la préférence pour des peptides basiques, et ce malgré une identité de séquence limitée entre les trois AmiA de *S. thermophilus* et OppA de *L. lactis* ($\approx 20 \%$). Cette

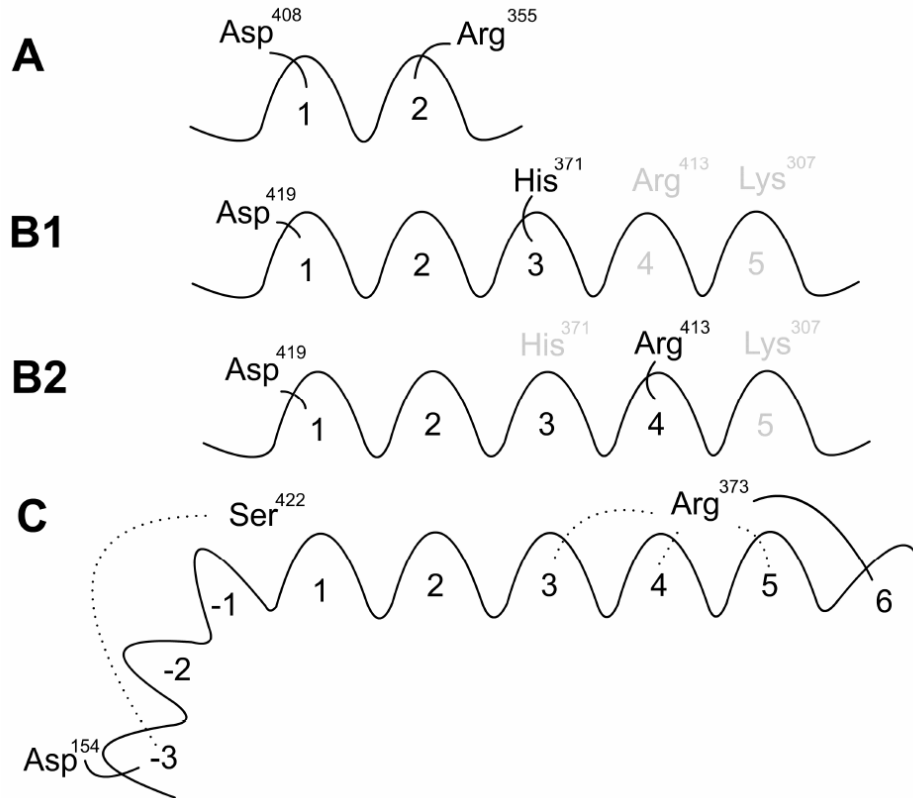


FIGURE 1.12 : Représentation schématique des interactions entre des résidus conservés de diverses SBPs structurellement caractérisées et les extrémités N- et C-terminales des peptides transportés. Les chaînes latérales des acides aminés sont positionnées dans des sous-cavités des sites de fixation numérotées ici à partir des extrémités N-terminales des peptides. Les extrémités chargées sont ancrées par des ponts salins avec des résidus conservés (traits pleins) , ce qui détermine les différentes tailles prises en charge. L'aspartate interagissant avec les extrémités N-terminales (A et B1/B2) est conservé dans de nombreuses protéines de fixation d'ABC-transporteurs de peptides. En outre, de nombreuses liaisons hydrogènes sont formées entre les sites de fixation et les chaînes principales des peptides. Ne sont représentées ici que les liaisons hydrogène intervenant au niveau des extrémités (C, traits pointillés). (A) Dipeptide lié à DppA d'*E. coli*. (B1) et (B2) tri- et térapeptide lié à OppA de *S. typhimurium*. Il est également possible qu'un pentapeptide soit par fixé par Lys³⁰⁷ à extrémité C-terminale. (C) Nonapeptide pris en charge par AppA de *B. subtilis*. Le résidu aspartate habituel est là remplacé par une sérine qui est également présente dans la séquence d'OppA de *L. lactis*. L'extrémité N-terminale est prise en charge à une autre position par un aspartate non conservé dans les autres SBPs. Doeven *et al.* (2006)

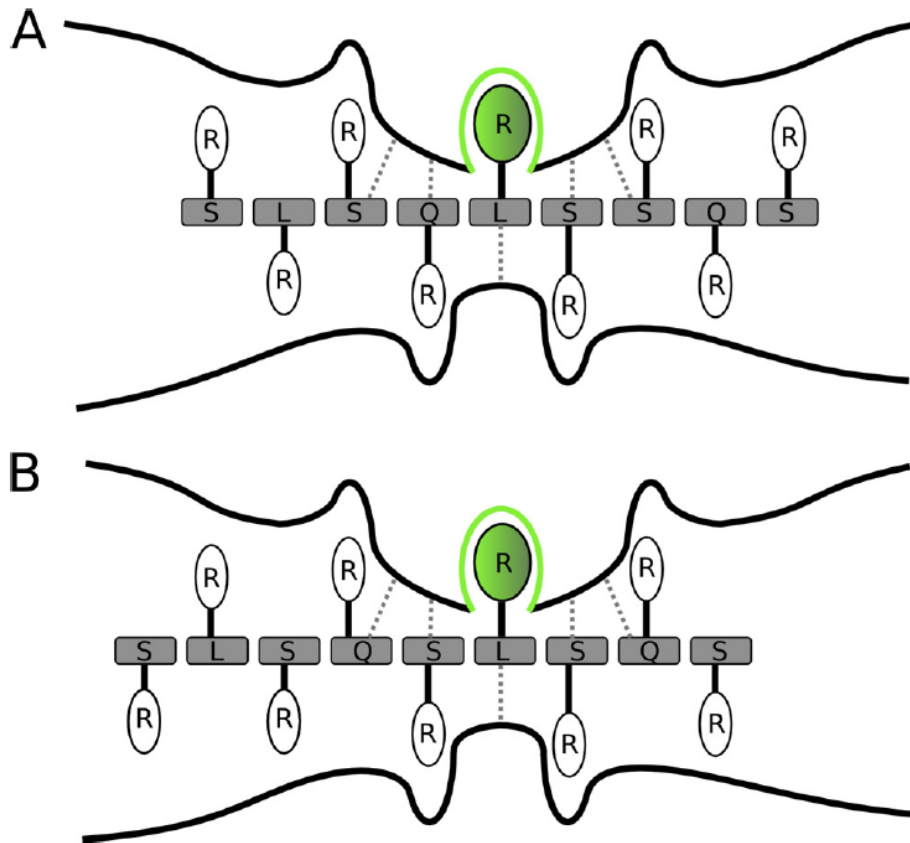


FIGURE 1.13 : Illustration de la cavité de fixation d'OppA occupée par SLSQLSSQS (A) et SLSQSLSSQS (B). La chaîne latérale hydrophobe de la leucine est représentée en vert et détermine la position d'ancrage du peptide à la poche hydrophobe. La partie centrale et étroite de la cavité interagit également avec la chaîne principale *via* cinq liaisons hydrogène. La cavité s'élargit ensuite de part et d'autre, permettant d'accomoder de longs peptides dont les extrémités restent libres. Berntsson *et al.* (2011)

différence de résultats entre études structurales/*in vitro* d'un côté et *in vivo* de l'autre peut provenir du fait que, dans le premier cas, seule la spécificité de la SBP est étudiée, tandis que dans le deuxième les préférences observées résultent de l'action du transporteur dans son ensemble (éventuelle influence du translocon dans la sélection des peptides?). Quoiqu'il en soit, la question du rôle d'une charge positive dans le transport d'oligopeptides *via* le système Opp de *L. lactis* reste encore ouverte.

1.3.1.1.3 Transport des peptides : bilan

Le transport des peptides par les bactéries constitue assurément une thématique complexe au sein de laquelle subsistent encore d'apparentes contradictions dans la littérature. En particulier, chez les bactéries lactiques, si les différentes études du transporteur DtpT sont dans l'ensemble cohérentes, tel n'est pas le cas pour les ABC-transporteurs Opt et surtout Opp de *L. lactis*. Bien que cela n'a pas toujours été le cas, il est maintenant unanimement reconnu que le transport d'oligopeptides *via* ce dernier système est fonction de la nature physico-chimique des peptides et de leur composition en acides aminés. En revanche, les préférences réelles de transport restent encore à élucider chez cette bactérie, et pourraient également être approfondies chez *S. thermophilus*, et explorées chez les autres bactéries lactiques.

En effet, si la mise en place de telles préférences de transport chez les bactéries semble avant tout répondre à une question utilitaire (« comment capturer et transporter efficacement des peptides exogènes? »), il est également légitime de mettre en regard ces préférences avec les fonctions physiologiques des transporteurs de peptides, et plus particulièrement vis-à-vis de leur rôle dans la nutrition azotée. *L. lactis* est en effet une espèce relativement exigeante et ses auxotrophies concernent les acides aminés branchés (leucine, isoleucine, valine), l'histidine, la méthionine, l'arginine et la glutamine, bien que ces besoins varient d'une souche à l'autre (Foucaud *et al.*, 2001). Les spécificités du transporteur Opp ne concordent donc qu'avec une partie des auxotrophies de la bactérie. De plus, en admettant l'hypothèse qu'Opp de *L. lactis* et Ami de *S. thermophilus* possèdent effectivement des préférences similaires, cette dernière espèce est en revanche moins exigeante d'un point de vue nutritionnel. Seules la leucine et la valine sont requises pour sa croissance, ainsi que les acides aminés soufrés chez certaines souches (méthionine, cystéine; Letort et Juillard, 2001). De fait, l'hypothèse d'une sélection des peptides par la SBP en fonction des auxotrophies propres de la bactérie apparaît comme assez peu probable. Cette absence de lien entre transport de peptides et besoins nutritionnels a d'ailleurs été mise en avant par Kevvai *et al.* (2014) lors de leur travail sur les flux d'azote aminé entre un EXL et *L. lactis* IL1403 (voir partie 1.3.1.3). Cette conclusion est d'autant plus intéressante qu'elle s'applique dans ce cas là au système Opt et à sa SBP associée OptA, le transporteur Opp n'étant pas fonctionnel chez IL1403. Enfin, la question de la redondance entre systèmes de transport chez certaines es-

pèces (e.g. Opt/Opp) ou encore de l'éventuel intérêt de systèmes comportant de multiples SBPs (Ami) reste ouverte et n'est à ce jour pas entièrement résolue.

Finalement, dans un contexte de croissance en milieu riche où les peptides de dégradation sont abondants et variés, il apparaît désormais que leur valeur nutritionnelle est non seulement fonction des acides aminés essentiels qu'ils contiennent, mais aussi des acides aminés « laissez-passer » qui vont permettre leur prise en charge préférentielle par un transporteur dédié et leur translocation dans la cellule.

1.3.1.2 Clivage intracellulaire

Une fois internalisés, il est généralement admis que les peptides à rôle nutritionnel vont être rapidement dégradés grâce à l'action concertée d'un large éventail de peptidases intracellulaires, conduisant à la libération complète des acides aminés les constituant. Les bactéries lactiques possèdent toutes un jeu complet de peptidases, de l'ordre d'une douzaine par espèce (Atlan *et al.*, 2008). Ces systèmes ont été dans l'ensemble abondamment caractérisés et documentés (Kunji *et al.*, 1996a; Christensen *et al.*, 1999; Savijoki *et al.*, 2006; Atlan *et al.*, 2008; Liu *et al.*, 2010; Juillard *et al.*, 2016). Ils incluent à la fois des endopeptidases, di-tripeptidases et aminopeptidases, et la majorité d'entre elles sont communes aux différentes espèces de bactéries lactiques (Tableau 1.7). Il est également généralement admis que les bactéries lactiques ne possèdent pas d'activité carboxypeptidase intracellulaire, l'ensemble des exopeptidases identifiées libèrent en effet entre un et deux résidus de l'extrémité N-terminale (Savijoki *et al.*, 2006).

Ces différentes peptidases ont également été classées en fonction de leur spécificité de substrat qui peut être large (clivage indifférencié) ou au contraire spécifique de quelques acides aminés. Le grand nombre d'aminopeptidases spécifiques de la proline (PepX, PepQ, PepR, PepI, PepP) est particulièrement frappant, mais trouve son explication dans la composition des caséines qui sont des protéines riches en proline (respectivement 7.5 %, 15 % et 17 % des résidus des caséines α , β et κ ; Gordon *et al.*, 1950). Certaines peptidases sont également plus spécifiques de certaines espèces telles que PepL, PepG, PepI et PepR qui se retrouvent principalement chez les lactobacilles et sont absentes chez les lactocoques et streptocoques. De même, PepS, identifiée uniquement chez *S. thermophilus*, présente une spécificité pour l'arginine et les acides aminés aromatiques (Fernandez-Espla et Rul, 1999).

TABLEAU 1.7 : Distribution des peptidases chez les bactéries lactiques. Adapté de Liu *et al.* (2010).

Classification	Peptidase	Spécificité de substrat	Souches bactériennes																					
			LAC	LJO	LGA	LDB	LBU	LHE	LPL	LBE	LRF	LSK	LSL	LCA	LRH	PPE	OOE	LME	LLX	LLA	LLM	STH	STU	STM
Endopeptidase	PepO	(X)n (X)n	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
	PepF	(X)n (X)n	P	P	P	P	P	P	P	P	A	P	P	P	P	P	P	P	P	P	P	P	P	P
	PepE/PepG	(X)n (X)n	P	P	P	P	P	P	P	P	P	A	P	P	A	P	A	A	A	A	A	A	A	A
Aminopeptidase générale	PepC	X (X)n	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
	PepN	X (X)n	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Aminopeptidase spécifique	PepX	X-Pro (X)n	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
	PepI	Pro X-(X)n	P	A	A	P	P	P	P	A	A	A	A	P	A	A	A	A	A	A	A	A	A	A
	PepR	Pro X	P	P	P	P	P	P	P	P	P	A	P	P	P	A	A	A	A	A	A	A	A	A
	PepL	Leu (X)n	P	A	P	A	A	A	P	A	A	A	A	P	A	P	A	A	A	A	A	A	A	A
	PepP	X Pro-(X)n	P	P	P	P	P	P	P	P	P	A	P	P	P	A	P	P	P	P	P	P	P	P
	PepQ	X Pro	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
	PepS	Arg/Aro (X)n	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	P	P
Aminopeptidase unique	PepM	Met (X)n	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
	PepA	Glu/Asp (X)n	P	P	P	P	P	P	A	A	A	A	A	A	P	A	A	P	P	P	P	P	P	P
	Pcp	pyroGlu (X)n	P	P	P	P	P	P	A	P	A	A	A	P	P	A	A	A	A	P	P	A	A	A
Dipeptidase	PepD	X X	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	A	P	P	P	A	A	A
	PepV	X X	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Tripeptidase	PepT	X X-X	P	P	P	P	P	P	P	P	P	P	A	P	P	P	P	P	P	P	P	P	P	P

P (vert) : présence d'au moins un gène dans le génome de la bactérie. **A** (rouge) : absence de gène. Nom des souches : **LAC** *Lb. acidophilus* NCFM, **LJO** *Lb. johnsonii* NCC533, **LGA** *Lb. gasseri* ATCC33323, **LDB** *Lb. bulgaricus* ATCC11842, **LBU** *Lb. bulgaricus* ATCCBAA365, **LHE** *Lb. helveticus* DPC4571, **LPL** *Lb. plantarum* WCFS1, **LBE** *Lb. brevis* ATCC367, **LRF** *Lb. reuteri* F275, **LSK** *Lb. sakei* 23K, **LSL** *Lb. salivarius* UCC118, **LCA** *Lb. casei* ATCC334, **LRH** *Lb. rhamnosus* GG, **PPE** *P. pentosaceus* ATCC25745, **OOE** *O. oeni* PSU1, **LME** *L. mesenteroides* ATCC8293, **LLX** *L. lactis* subsp. *lactis* IL1403, **LLA** *L. lactis* subsp. *cremoris* SK11, **LLM** *L. lactis* subsp. *cremoris* MG1363, **STH** *S. thermophilus* CNRZ1066, **STU** *S. thermophilus* LMG18311, **STM** *S. thermophilus* LMD-9.

TABLEAU 1.8 : Systèmes de transports d'acides aminés libres identifiés chez les bactéries lactiques. Atlan *et al.* (2008).

Transporteur	Acides aminés transportés
Transporteur PTR	Alanine, glycine
	Thréonine, Sérine
	Isoleucine, Leucine, Valine
	Tyrosine, Phénylalanine
	Méthionine
	Histidine
ABC-transporteur	Lysine
	Glutamine, glutamate
	Asparagine
Antiport	Proline, Glycine-bétaïne
	Arginine/ornithine
	Glutamate/gamma-aminobutyrate

Lorsqu'un peptide est internalisé, il va être séquentiellement hydrolysé (Kunji *et al.*, 1996a). Il est tout d'abord pris en charge par des peptidases dites *primaires* (PepN, PepA, PepI, PepN et PepC) qui vont directement libérer des acides aminés à partir d'oligopeptides. Puis, des peptidases *secondaires* (PepO, PepF, PepX, PepR, PepQ, PepT et PepV) vont compléter et terminer l'hydrolyse. Enfin, certaines peptidases ne sont pas impliquées dans la nutrition azotée. C'est notamment le cas de PepM qui participe à la maturation des protéines nouvellement synthétisées en retirant la méthionine N-terminale (Miller *et al.*, 1989).

1.3.1.3 Devenir des acides aminés libérés

L'hydrolyse des peptides conduit donc à la libération complète de leurs acides aminés constitutifs. Ceux-ci, en compagnie des acides libres du milieu transportés par leurs systèmes dédiés (Tableau 1.8), vont avoir différents débouchés : synthèse protéique, export hors de la cellule, interconversion, ou catabolisme pour la production d'énergie ou de molécules aromatiques. Seules les deux premières voies seront approfondies ici en raison de leur importance quantitative. Les travaux récents de Kevvai *et al.* (2014) ont permis de faire la lumière sur ces dynamiques dans un milieu à base d'EXL.

Les auteurs ont en effet cherché à étudier les dynamiques de consommation d'acides aminés libres et de peptides *via* leurs acides aminés constitutifs - qui seront dorénavant nommés « acides aminés peptidiques » par opposition aux acides aminés libres - lors d'une culture de *L. lactis* IL1403. Pour cela, un EXL isotopique marqué à l'azote 15 a été utilisé pour compléter un milieu synthétique ne contenant que des acides libres non marqués dans lequel a été cultivée la bactérie. Le suivi cinétique par chromatographie

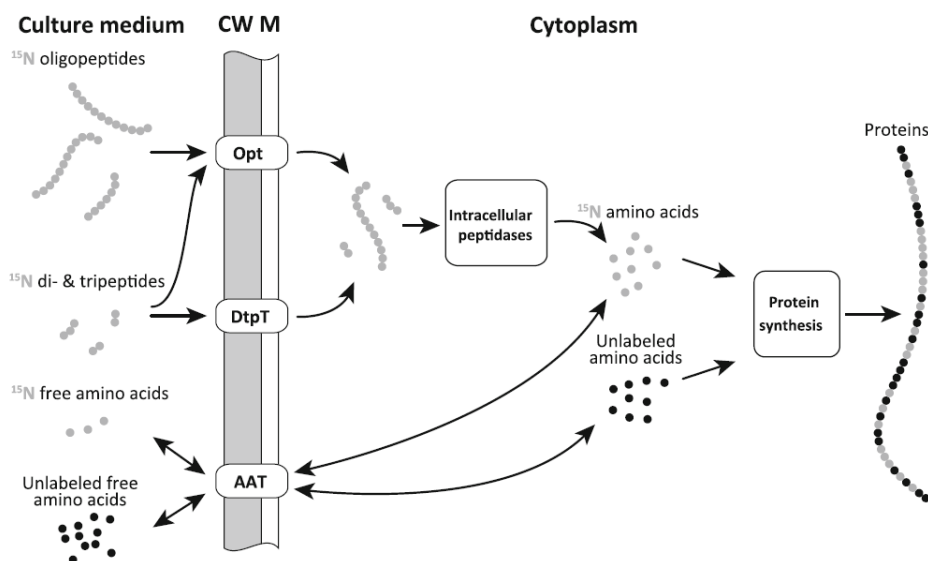


FIGURE 1.14 : Représentation schématique des flux d'utilisation des acides aminés libres et peptides chez *L. lactis* IL1403. CW paroi, M membrane, *Opt* ABC-transporteur des di-tripeptides et oligopeptides, *DtpT* transporteur PTR des di-tripeptides, *AAT* ensemble des transporteurs des acides aminés libres. Kevvai *et al.* (2014).

liquide couplée à la spectrométrie de masse des acides aminés libres (marqués et non marqués) et peptidiques (= différence entre acides aminés libres et totaux après hydrolyse acide) combiné à une modélisation mathématique a permis d'estimer les flux entrants et sortants de ces différentes formes d'azote aminé ainsi que leur intégration respective dans la biomasse totale (Figure 1.14).

Les auteurs ont tout d'abord noté une utilisation simultanée d'acides aminés libres et de peptides. En particulier, l'asparagine, la glutamine et l'arginine constituent les acides aminés les plus intensément consommés sous forme libre, l'arginine ayant quasiment disparue en fin de fermentation. De même, tous les acides aminés peptidiques ont été progressivement consommés durant l'expérience. Cette co-consommation d'acides aminés, tant sous forme libre que peptidique, est indépendante des auxotrophies de la souche, indiquant que l'assimilation d'un acide aminé donné dépend avant tout de sa biodisponibilité.

Néanmoins, cette consommation apparente résulte en réalité d'un bilan entre l'import d'acides aminés libres et de peptides, et le rejet d'acides aminés hors de la cellule. A cet égard, à l'exception du glutamate, les auteurs ont noté une augmentation de la proportion d'acides aminés libres isotopiques dans le milieu extérieur, indiquant une large excrétion des acides aminés peptidiques suite à leur hydrolyse intracellulaire. Ils ont ainsi estimé que la bactérie rejette au total autant d'acides aminés dans le milieu qu'elle n'en utilise pour la biosynthèse protéique.

En ce qui concerne la contribution respective des acides aminés libres ou peptidiques pour

la production de biomasse, la situation est variable selon l'acide aminé considéré et le stade de croissance. En début de culture, les peptides contribuent à hauteur de 30 % à 60 % de la biomasse produite en fonction de l'acide aminé considéré. Ces chiffres diminuent ensuite en fin de culture, traduisant une disparition des peptides les plus facilement assimilables. Il est ainsi tentant de dresser un parallèle entre ce phénomène et les spécificités des systèmes de transport décrits dans la partie précédente : les peptides ayant une plus forte affinité de transport sont certainement ceux disparaissant le plus vite. Dans l'éventualité où les acides aminés peptidiques dictant les spécificités de transport seraient les constituants majoritaires des peptides consommés en raison d'une préférence relativement stricte du transporteur (e.g. la SBP ne fixe que des peptides très majoritairement hydrophobes), on devrait en toute logique s'attendre à une consommation accrue de tels acides aminés. Or, ici, tous les acides aminés peptidiques voient leur concentration décroître au cours du temps de façon significative et relativement uniforme. Il semble donc que les spécificités du transporteur ne sont *a priori* pas suffisamment strictes pour prévenir l'assimilation d'un quelconque acide aminé peptidique (les peptides transportés incluent l'ensemble des différents acides aminés). En outre, si la souche IL1403 possède bien la perméase DtpT, le système Opp n'est en revanche pas fonctionnel et elle repose donc exclusivement sur Opt pour son approvisionnement en oligopeptides. A ce jour, les spécificités de ce système n'ont pas été étudiées et demeurent inconnues.

Pour résumer, à travers cette étude, il apparaît que dans un milieu riche de type EXL, les spécificités des transporteurs de peptides sont susceptibles de moduler les dynamiques d'approvisionnement en peptides au cours du temps, tout en permettant une fourniture large en acides aminés. Approximativement la moitié des acides aminés totaux transportés, tant sous forme libre que peptidique, va ensuite être utilisée pour la biomasse tandis que l'autre moitié sera excrétée hors de la cellule. Cet efflux considérable pourrait notamment contribuer à la force proton-motrice et être donc énergiquement favorable pour la bactérie (Konings, 2002). En revanche, des zones d'ombre subsistent encore concernant les effecteurs de cet export pour les acides aminés qui transitent par des systèmes non réversibles tels que les ABC-transporteurs.

1.3.1.4 Peptides nutritionnels : bilan

La nutrition représente donc le débouché majoritaire des peptides exogènes. Etant donnée leur abondance, ce flux nutritionnel est principalement occupé par les peptides issus du milieu de culture. Il est toutefois très vraisemblable que les peptides provenant de l'activité protéolytique bactérienne soit également recyclés par les mêmes processus. Si les protéases de paroi constituent les acteurs clés pour la nutrition dans le lait du fait de la faible abondance en acides aminés libres et peptides natifs, en milieu riche ce sont véritablement les

transporteurs qui tiennent ce rôle. Il est donc intéressant de noter que ceux-ci possèdent des spécificités propres qui vont influencer sur la nature des peptides entrants, tout en assurant une couverture complète de l'ensemble des acides aminés protéinogènes. Il est ensuite généralement reconnu qu'une fois transportés, les peptides sont immédiatement pris en charge par les peptidases intracellulaires et très rapidement dégradés. Or, à cet égard, la présence dans le milieu extérieur de peptides protéolytiques issus de protéines cytoplasmiques et exportés par un mécanisme inconnu (Guillot *et al.*, 2016) est surprenante. En effet, on pourrait s'attendre à ce que les peptides libérés par le recyclage protéique intracellulaire soient eux aussi entièrement dégradés, ce qui n'est *a priori* pas le cas pour une partie d'entre eux. Ce phénomène interroge donc sur la cinétique et la spécificité de substrat des peptidases. De plus, considérant certaines fonctions physiologiques connues de peptides bactériens, notamment les phéromones intracellulaires qui interagissent avec des régulateurs transcriptionnels, il est légitime de se demander si certains peptides transportés considérés *a priori* comme nutritionnels ne pourraient pas échapper à une hydrolyse immédiate et exercer d'autres fonctions chez la bactérie.

1.3.2 Régulation

En se basant sur les fonctions documentées que certains peptides produits par les bactéries exercent sur la physiologie cellulaire, il est possible d'émettre l'hypothèse que des peptides de dégradation pourraient participer à des mécanismes de régulation. Ce concept est d'ores et déjà supporté par le cas des sexe-phéromones d'*E. faecalis* impliquées dans l'induction de la conjugaison et qui sont produites par hydrolyse de la séquence signal de lipoprotéines par la protéase Eep (An *et al.*, 1999; Clewell *et al.*, 2000). Dans cette partie, la notion de régulation par des peptides sera explorée en raison de son implication connue dans de nombreuses fonctions physiologiques chez les bactéries lactiques. Ce contexte correspond à une régulation dite « directe » où un peptide intact va moduler l'action d'un régulateur transcriptionnel. Mais avant cela, il sera tout d'abord fait mention d'une régulation « indirecte » où ce sont les acides aminés constitutifs du peptide et libérés après hydrolyse intracellulaire qui sont impliqués dans une voie de signalisation.

1.3.2.1 Régulation indirecte : l'exemple de CodY

Un exemple bien connu de régulation indirecte par les peptides chez les bactéries lactiques concerne le régulateur transcriptionnel CodY. Il s'agit d'un régulateur global hautement conservé chez les bactéries à Gram positif et à faible pourcentage de G-C. Il a été initialement découvert chez *B. subtilis* en tant que facteur de répression de la perméase des dipeptides (Dpp) durant la phase de croissance exponentielle (Slack *et al.*, 1995). Depuis, son

rôle de senseur nutritionnel permettant l'adaptation à des changements environnementaux a été clairement établi. Il a été montré que CodY contrôlait chez *B. subtilis* l'expression de plus de 100 gènes et opérons (Brinsmade *et al.*, 2014) impliqués dans de nombreuses fonctions telles que la nutrition azotée, l'induction de la compétence, la sporulation, la réponse aux stress, la biosynthèse d'acides aminés branchés ou encore la synthèse de transporteurs d'acides aminés et de sucres (Serror et Sonenshein, 1996; Molle *et al.*, 2003). En outre, CodY intervient également dans la régulation de la virulence chez des espèces pathogènes telles que *S. pneumoniae* (Hendriksen *et al.*, 2008; Caymaris *et al.*, 2010), *L. monocytogenes* (Lobel *et al.*, 2012) ou *S. aureus* (Majerczyk *et al.*, 2008; Pohl *et al.*, 2009). En ce qui concerne les bactéries lactiques, des homologues de CodY de *B. subtilis* sont présents dans les génomes des genres *lactococcus* et *streptococcus*, mais pas *lactobacillus* (Groot Kormelink *et al.*, 2012). Ce régulateur a été particulièrement bien étudié depuis les années 2000 chez *L. lactis* où il est en particulier connu pour réguler le système protéolytique de la bactérie (Guédon *et al.*, 2001a, 2005). Plus récemment, le régulon du CodY de *S. thermophilus* a également été caractérisé (Lu *et al.*, 2015), ouvrant d'intéressantes perspectives sur d'éventuelles variations de régulation entre différentes espèces de bactéries lactiques.

1.3.2.1.1 Mécanismes d'action de CodY

CodY est ici donné en exemple de régulation indirecte médiée par les peptides en raison de son mode d'action particulier. En effet, l'activité de CodY est contrôlée par le contenu intracellulaire en acides aminés branchés (BCAAs, Branched-Chain Amino Acids) qui servent de signaux indiquant l'état nutritionnel de la bactérie (Guédon *et al.*, 2001b; Shivers et Sonenshein, 2004), ainsi que dans certains cas par le niveau de GTP (Guanosine TriPhosphate), indicateur du statut énergétique (Brinsmade et Sonenshein, 2011). Il a toutefois été montré que CodY de *L. lactis* n'était actif qu'en présence des BCAAs seulement et non de GTP (Petranovic *et al.*, 2004). Le contenu intracellulaire des BCAAs découle de la biosynthèse *de novo* de ces acides aminés, mais aussi de l'approvisionnement azoté de la bactérie, tant sous forme d'acides aminés libres que de peptides. Il est d'ailleurs notable que l'activité de CodY a initialement été détectée dans des milieux riches en peptides tant chez *B. subtilis* que *L. lactis* (Slack *et al.*, 1993, 1995; Guédon *et al.*, 2001a, 2001b).

Ainsi, en conditions de richesse nutritionnelle et énergétique, la fixation de BCAAs et/ou de GTP favorise la liaison de CodY à l'ADN sur un motif consensus de 15 nucléotides (AATTTTCWGAAAATT; den Hengst *et al.*, 2005; Guédon *et al.*, 2005) situé dans la région promotrice des gènes cibles. Dans la plupart des cas, CodY exerce une action répressive sur l'expression des gènes soit en entrant en compétition avec l'ARN polymérase ou un régulateur positif (Belitsky, 2011), soit en provoquant l'arrêt de la transcription par un mécanisme dit de « *roadblock* » (blocage de l'ARN polymérase; Belitsky et Sonenshein, 2011). En conditions de carence nutritionnelle, les concentrations intracellulaires de BCAAs

et GTP diminuent, entraînant une dissociation de CodY des régions promotrices ce qui lève la répression des gènes nécessaires pour l'adaptation à la carence nutritionnelle. La régulation positive par CodY est également possible mais n'a été observée jusqu'à présent que dans trois cas différents avec le gène de l'acétate kinase (*ackA*; Shivers *et al.*, 2006) et un petit ARN régulateur (BsrF; Preis *et al.*, 2009) de *B. subtilis*, ainsi que le gène *botA* codant la toxine botulique de *Clostridium botulinum* (Zhang *et al.*, 2014). Il a également été montré récemment que l'affinité de CodY variait en fonction du degré de conservation des sites de fixation (Belitsky et Sonenshein, 2013). Ceci entraîne de fait un contrôle hiérarchique de l'expression de son régulon et permet ainsi une adaptation fine aux variations environnementales (Brinsmade *et al.*, 2014). Ces différents exemples correspondent à un mode de régulation dit direct où CodY interagit physiquement avec ses gènes cibles. Néanmoins, du fait de son action pléiotrope et de son interaction avec des facteurs transcriptionnels, CodY intervient dans des réseaux de régulation où il module de façon indirecte l'expression de nombreux autres gènes. Sous ce mode d'action, CodY peut avoir un effet tant répresseur qu'activateur.

La structure tridimensionnelle de CodY a été résolue en deux temps. Deux fragments de la protéine ont d'abord été cristallisés (Levdikov *et al.*, 2006), mais ce n'est que récemment que sa structure entière a été caractérisée (Levdikov *et al.*, 2017). Dans des conditions physiologiques, CodY est organisé en dimère, chaque protomère comprenant un domaine GAF (cGMP-stimulated phosphodiesterases, Adenylate cyclases, FhlA) et un domaine wHTH (Winged Helix-Turn-Helix), les deux étant connectés par une hélice- α (Figure 1.15A). Les domaines GAF portent les sites de fixation aux substrats tandis que les domaines wHTH sont responsables de la liaison à l'ADN. La fixation de ligand induit un changement conformationnel caractérisé par un écartement des hélices- α entraînant l'éloignement des deux domaines wHTH (Figure 1.15B). Cet écartement est associé à une plus forte affinité des domaines wHTH pour l'ADN. Il a été en outre observé que la fixation d'un dimère CodY sur une molécule d'ADN favorisait la liaison d'un deuxième dimère selon un mode de coopération allostérique (Levdikov *et al.*, 2017).

1.3.2.1.2 Rôle physiologique de CodY chez les bactéries lactiques

La caractérisation du régulon de CodY, et plus largement de tout régulateur transcriptionnel, peut se faire à l'aide d'un mutant de délétion pour ce régulateur étudié à travers des approches transcriptomiques globales de type puces à ADN ou RNAseq. La grande majorité des informations concernant l'action physiologique de CodY chez les bactéries lactiques proviennent de *L. lactis* (Guédon *et al.*, 2001a, 2001b, 2005; Chambellon et Yvon, 2003; Petranovic *et al.*, 2004; den Hengst *et al.*, 2006). Dans cet organisme, CodY assure un contrôle strict du système protéolytique nutritionnel à travers une répression de la protéolyse extracellulaire de PrtP, du transport de peptides et de leur dégradation intracellulaire (Guédon

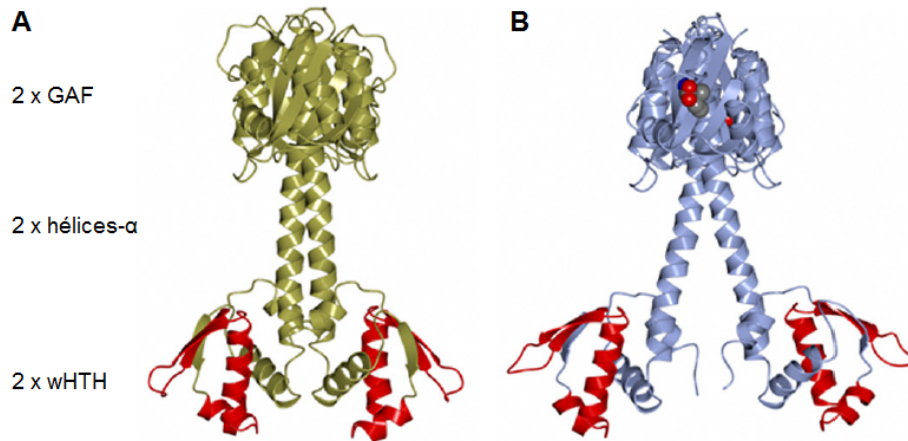


FIGURE 1.15 : Structure générale de CodY en l'absence de ligand (A) et en présence d'isoleucine (B). L'écartement des hélices- α et des domaines wHTH entraînés par la fixation d'isoleucine par les domaines GAF est visible. Levdikov *et al.* (2017).

et al., 2005). De plus, près de 45 % des gènes appartenant au régulon de CodY de *L. lactis* sont impliqués dans la biosynthèse d'acides aminés, incluant notamment les gènes responsables de la synthèse des BCAAs pour lesquels la répression est particulièrement intense. Cette information, couplée au fait que la perméase des BCAAs, BcaP, est également sous contrôle de CodY (den Hengst *et al.*, 2006), indique la présence d'une boucle de régulation négative dans les voies d'approvisionnement de ses propres effecteurs. De même, un motif de fixation CodY est présent à proximité de la boîte -35 de son propre promoteur et ce dans de nombreux genres bactériens, indiquant là aussi une probable boucle de régulation négative de CodY pour sa propre expression (Guédon *et al.*, 2005).

La caractérisation récente du régulon de CodY de *S. thermophilus* a révélé d'intéressantes particularités (Lu *et al.*, 2015). En comparaison avec *L. lactis*, le contrôle du système protéolytique par CodY y apparaît bien plus ténu car ni l'expression des gènes des transporteurs de peptides ni celle des peptidases ne sont affectées par l'inactivation de CodY. Son activité semble beaucoup plus orientée vers le transport et la biosynthèse d'acides aminés. Les auteurs attribuent cette caractéristique à l'absence de protéase de paroi chez la souche utilisée et donc à un probable contexte génétique davantage tourné vers la biosynthèse *de novo* d'acides aminés que vers un approvisionnement par les peptides exogènes. Il serait néanmoins intéressant de confronter ces résultats à ceux obtenus dans des conditions similaires avec une souche protéolytique. Enfin, la délétion de *codY* chez *S. thermophilus* a également conduit à la diminution de l'expression d'un certain nombre de gènes n'appartenant pas au régulon de CodY de *L. lactis*. Ces gènes incluent notamment les opérons *purCLFMNH* et *purDEK* responsables de la biosynthèse de l'IMP. En raison de l'absence de motif de fixation clairement identifiable dans les régions promotrices de ces opérons, il est davantage probable que cette régulation positive se fasse de façon indirecte à travers des réseaux de ré-

gulation. Elle suggère néanmoins un lien possible entre CodY et le métabolisme énergétique chez *S. thermophilus*, l'IMP étant le précurseur du GTP et de l'ATP.

On peut finalement relever que si les lactobacilles ne possèdent pas d'homologue de CodY dans leur génome, un régulateur « *CodY-like* » vient récemment d'être caractérisé chez *Lb. delbrueckii* subsp. *lactis* : YebC (Brown *et al.*, 2017). De même que lors des expériences antérieures réalisées avec *B. subtilis* et *L. lactis*, les auteurs ont noté une répression du système protéolytique dans une culture complémentée en peptides. Par des expériences de retard sur gel, ils ont également observé la liaison de la protéine YebC purifiée sur les régions promotrices des gènes du système protéolytique. Si sa fonction régulatrice doit encore être clairement établie en particulier *via* la construction d'un mutant de délétion, YebC constitue néanmoins un excellent candidat potentiel pour tenir un rôle équivalent à celui de CodY chez les lactobacilles.

Ainsi, à quelques variantes près, CodY tient un rôle majeur dans le contrôle du métabolisme azoté des bactéries lactiques. Il est frappant de constater l'utilisation des BCAAs en tant que molécules de signalement de l'environnement nutritionnel, et de croiser cela avec les spécificités connues des deux principaux systèmes de transport des bactéries lactiques, DtpT et Opp, qui favorisent l'import de peptides hydrophobes, contenant en particulier des BCAAs. Il semble ainsi émerger de toutes ces informations une stratégie globale et coordonnée pour la nutrition azotée où les BCAAs jouent un rôle central. A travers eux, les peptides d'un milieu riche participent à la régulation du métabolisme bactérien de façon indirecte, c'est-à-dire en raison du fait que leur composition intervient mais non leur séquence propre. La partie suivante va en revanche s'intéresser aux cas où les peptides sont directement des molécules régulatrices chez les bactéries lactiques en raison de leur séquence propre.

1.3.2.2 Régulation directe : mécanismes de *quorum sensing*

La régulation de fonctions physiologiques par les peptides est principalement connue à travers leur rôle dans la communication chimique inter-cellules, également appelée *quorum sensing*. Ce phénomène a été initialement observé il y a une quarantaine d'années chez les espèces marines luminescentes *Vibrio fischeri* et *Vibrio harveyi* qui coordonnent l'émission de lumière à l'échelle de la population grâce à des messagers chimiques (Nealson et Hastings, 1979). Chez les bactéries à Gram négatif, ces messagers - que l'on appelle également auto-inducteurs ou phéromones - sont principalement des acyl-homosérine lactones ou autres molécules synthétisées à partir de S-adénosylméthionine (SAM) capables de diffuser à travers la membrane bactérienne (Papenfort et Bassler, 2016). Chez les bactéries à Gram positif, ce sont principalement des oligopeptides qui remplissent cette fonction (Miller et Bassler, 2001). Ceux-ci sont produits en réponse à des conditions environnementales particulières sous forme de précurseurs peptidiques soit par synthèse ribosomale, soit, dans

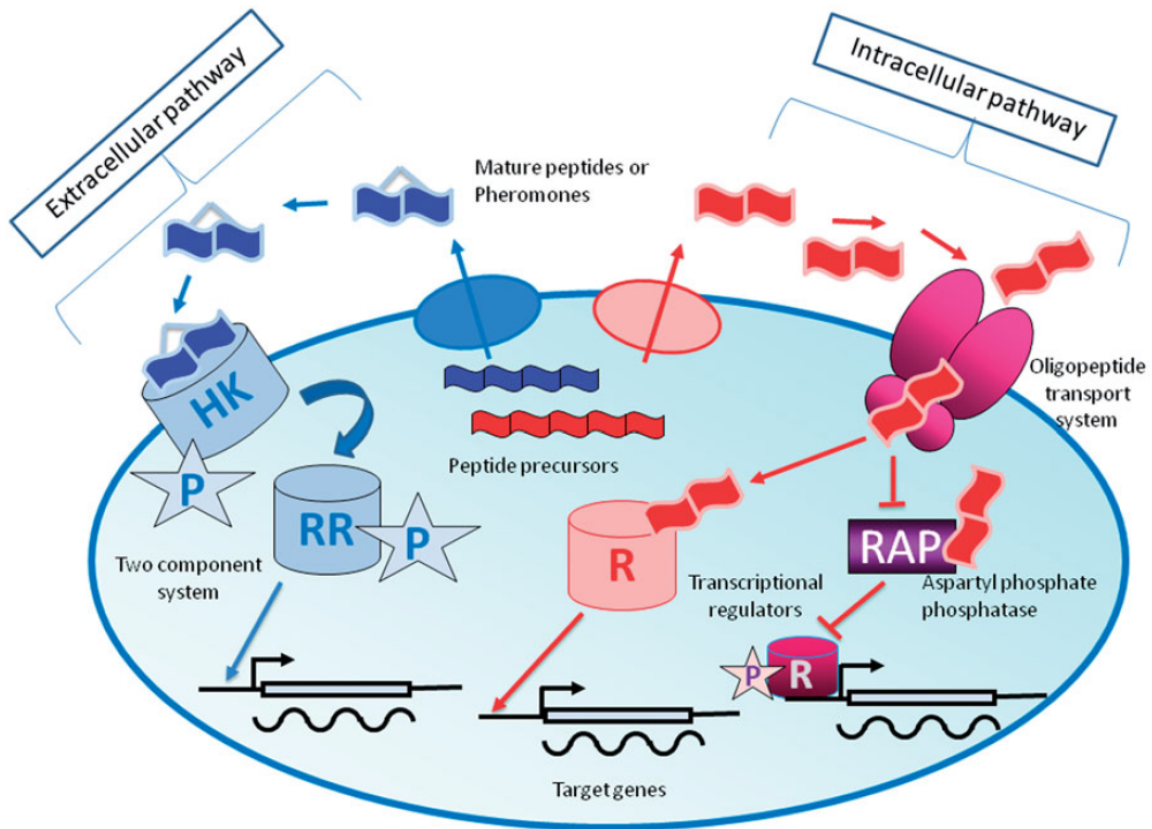


FIGURE 1.16 : Représentation schématique des deux voies de communication utilisées par les bactéries à Gram positif reposant sur des oligopeptides en tant que molécules d'auto-induction. A gauche (bleu) : mécanisme extracellulaire dans lequel la détection du signal s'opère à la surface de la cellule par le biais du senseur histidine kinase (HK) d'un système à deux composants et associé à un régulateur transcriptionnel cytosolique (RR, régulateur réponse). Les peptides phéromones sont généralement modifiés post-traductionnellement (cyclisation). A droite (rouge) : mécanisme basé sur la réimportation des phéromones matures par un ABC-transporteur et sur une détection intracellulaire directement par un régulateur transcriptionnel (R) ou par une aspartylphosphate phosphatase (RAP). Monnet *et al.* (2014)

quelques cas, par dégradation protéolytique de la séquence signal de lipoprotéines. Ils sont ensuite maturés (hydrolyse du précurseur) afin de libérer la forme active du peptide, puis éventuellement modifiés, et enfin activement exportés dans le milieu extracellulaire. Là, ils vont être détectés à l'échelle de la population à travers deux mécanismes de signalisation distincts (Figure 1.16). Le premier mécanisme, dit extracellulaire, est caractérisé par une détection du signal peptidique à la surface des cellules ainsi que par un découplage entre détection et réponse. A l'inverse, le deuxième mécanisme, dit intracellulaire, couple la détection et la réponse au sein d'un même effecteur protéique à localisation cytoplasmique. La réponse induite est propre à chaque système et consiste en la modulation coordonnée de l'expression de gènes à l'origine d'une fonction physiologique particulière (compétence, virulence, sporulation etc. ; Ng et Bassler, 2009).

Aujourd'hui, les connaissances accumulées sur la régulation de type *quorum sensing* chez les bactéries à Gram positif sont nombreuses et particulièrement bien décrites dans la littérature (Miller et Bassler, 2001; Kleerebezem, 2004; Lyon et Novick, 2004; Waters et Bassler, 2005; Ng et Bassler, 2009; Cook et Federle, 2014; Jimenez et Federle, 2014; Monnet *et al.*, 2014; Monnet et Gardan, 2015; Hawver *et al.*, 2016; Perez-Pascual *et al.*, 2016; Monedero *et al.*, 2017; Neiditch *et al.*, 2017). L'objectif de cette partie est de répertorier les familles de peptides connues intervenant dans les deux mécanismes de détection précédemment cités, d'en préciser les spécificités de séquence et structurales, et, ce faisant, d'apporter des éléments de réponse à la question suivante : des peptides à vocation nutritionnelle sont-ils susceptibles de participer à des mécanismes de *quorum sensing* chez les bactéries lactiques ? La notion de participation étant envisagée à la fois sous un angle positif (activité « *pheromone-like* ») mais aussi négatif (activité d'interférence ou d'inhibition).

1.3.2.2.1 Peptides impliqués dans la détection extracellulaire

La détection extracellulaire repose sur l'utilisation d'un système à deux composants (Figure 1.16). Ceux-ci sont composés d'une protéine transmembranaire, l'histidine kinase, dotée d'un domaine senseur extracellulaire. Celui-ci, suite à la perception d'une molécule signal, va induire l'autophosphorylation de la protéine au niveau d'un résidu histidine conservé. Le groupement phosphoryle est ensuite transféré à un résidu aspartate du domaine receveur d'un régulateur de réponse cytoplasmique, entraînant l'activation de son domaine effecteur (Gao et Stock, 2009; Casino *et al.*, 2010). Dans 87 % des cas connus, ce domaine effecteur est un domaine de fixation à l'ADN de type HTH (Ulrich *et al.*, 2005) permettant la régulation de l'expression de gènes cibles.

Ces systèmes de détection à deux composants sont les mieux connus et les plus documentés dans la littérature en comparaison avec leurs équivalents à un composant. Le prototype de ces systèmes est le système Agr (Tableau 1.9) initialement décrit chez *S. aureus* mais répandu dans de nombreuses espèces de *streptococci*, *clostridia* et *bacilli* (Wuster et Babu, 2008). Chez *S. aureus*, ce système est à l'origine de la régulation d'un large régulon comprenant plus d'une trentaine de gènes impliqués dans la virulence (Thoendel *et al.*, 2011). Il existe également des systèmes « *Agr-like* » tels que Lam de *Lb. plantarum* ou Fsr d'*E. faecalis*. Tous ces systèmes ont la particularité d'utiliser des peptides cyclisés (cycles thiolactones ou lactones) pour phéromones dont la maturation est couplée à l'export. Deux autres grandes familles de phéromones présentent également d'importantes modifications post-traductionnelles. La première est composée des peptides de compétence ComX dont les exemples représentatifs sont issus de l'espèce modèle *B. subtilis*. Ces peptides présentent un résidu tryptophane conservé, indispensable à la transduction du signal, et présentant des greffages de groupes isoprénoïdes. La deuxième famille de phéromones modifiées est celle des bactériocines de classe I. Il s'agit de peptides antibiotiques polycycliques contenant plu-

sieurs acides aminés inhabituels et qui ont la particularité d'être également des molécules autoinductrices élicitrices de leur propre biosynthèse (Kleerebezem, 2004). La subtiline de *B. subtilis* en est un exemple bien connu. Les étapes gouvernant sa biosynthèse sont représentés dans la Figure 1.17 et illustrent le fonctionnement des systèmes de *quorum sensing* à détection extracellulaire.

TABLEAU 1.9 : Familles des peptides phéromones impliqués dans la voie de détection extracellulaire des systèmes de *quorum sensing* des bactéries à Gram positif. Adapté de Monnet *et al.* (2014) et de la base de données Quorumpeps, <http://quorumpeps.ugent.be> (Wynendaele *et al.*, 2013).

Famille	Espèce	Nom	Séquence ^a	Modification post-traductionnelle ^b	Taille du propeptide (aa)	Origine du peptide	Fonction contrôlée	Récepteur (HK, RR)	Référence
AgrD , Accessory Gene Regulator	<i>Staphylococcus aureus</i>	AIP-I AIP-II AIP-III AIP-IV	YSTCDFIM GVNACSSLF INCDFLL YSTCYFIM	Cycle thiolactone entre C4 et M8 Cycle thiolactone entre C5 et F9 Cycle thiolactone entre C3 et L7 Cycle thiolactone entre C4 et M8	45-48	Petit ORF	Virulence	AgrC, ArgA	Novick (2003), Lyon et Novick (2004)
	<i>Staphylococcus epidermis</i>	AIP-I AIP-II AIP-III	DSVCASYF NASKYNPCSNYL NAKYNPCASYL	Cycle thiolactone entre C4 et F8 Cycle thiolactone entre C8 et L12 Cycle thiolactone entre C8 et L12	46	Petit ORF	Virulence	AgrC, ArgA	Otto <i>et al.</i> (1998), Olson <i>et al.</i> (2014)
	<i>Listeria monocytogenes</i>	AIP	CFMFV	Cycle thiolactone entre C1 et V5	53	Petit ORF	Virulence	AgrC, ArgA	Zetzmann <i>et al.</i> (2016)
LamD , Lactobacillus Agr-like Module	<i>Lactobacillus plantarum</i>	LamD	CVGIW	Cycle thiolactone entre C1 et W5	40	Petit ORF	Adhérence	LamC, LamA	Sturme <i>et al.</i> (2007)
GBAP , Gelatinase biosynthesis-activating pheromone	<i>Enterococcus faecalis</i>	GBAP	QNSPNIFGQWM	Cycle lactone entre S3 et M11	53	Petit ORF	Biofilm, virulence	FsrC, FsrA	Nakayama <i>et al.</i> (2001, 2006), Nishiguchi <i>et al.</i> (2009)
ComX	<i>Bacillus subtilis</i>	ComX RS-B-1 ComX RO-E-2 ComX natto ComX 168	MMDWHY GIFWEQ KWPPIE ADPIRQWGD	Géranyl-géranylation en W4 Géranyl-géranylation en W3 Farnésylation en W2 Farnésylation en W8	53-71	Petit ORF	Compétence, biofilm	ComP, ComA	Tortosa <i>et al.</i> (2001), Okada <i>et al.</i> (2007, 2015)
CSP , Competence Stimulating Peptide	<i>Streptococcus pneumoniae</i>	CSP-I CSP-II	EMRLSKFFRDFILQRKK EMRISRILDFLFLRKK	Non	41	Petit ORF	Compétence	ComD, ComE	Håvarstein <i>et al.</i> (1995, 1997), Pozzi <i>et al.</i> (1996)
Pit	<i>Lactobacillus plantarum</i>	PitA	EQLSFTSIGILQLLTIGTRSCWFFYCRY	Non	58	Petit ORF	Inconnue	PitK, PitR	Sturme <i>et al.</i> (2007)
Bactériocines de classe I (lantibiotiques), autoinductrices	<i>Bacillus subtilis</i>	Mersacidine	A(Abu)F(Abu)LPGGGGVA(Abu)L(Abu)(Dha)EAI	Ponts thioéthers entre (Ala1-S-Abu2), (Abu4-S-Ala12), (Abu13-S-Ala18) et (Ile19-NH-CH=CH-S-Abu15)	68	Petit ORF	Autoproduction, activité antimicrobienne	MrsK2, MrsR1R2	Hsu <i>et al.</i> (2003), Schmitz <i>et al.</i> (2006)
		Subtiline	WKAE(Dha)LA(Abu)PGAV(Abu)GALQ(Dhb)AFLQ(Abu)L(Abu)ANAKI(Dha)K	Ponts thioéthers entre (Ala3-S-Ala7), (Abu8-S-Ala11), (Abu13-S-Ala19), (Abu23-S-Ala26) et (Abu25-S-Ala28)	56	Petit ORF	Autoproduction, activité antimicrobienne	SpaK, SpaR	Kleerebezem (2004)
	<i>Lactococcus lactis</i>	Nisine	I(Dhb)AI(Dha)LA(Abu)PGAK(Abu)GALMGA NMK(Abu)A(Abu)AHASIHV(Dha)K	Ponts thioéthers entre (Ala3-S-Ala7), (Abu8-S-Ala11), (Abu13-S-Ala19), (Abu23-S-Ala26) et (Abu25-S-Ala28)	56	Petit ORF	Autoproduction, activité antimicrobienne	NisK, NisR	Kuipers <i>et al.</i> (1995)
	<i>Streptococcus pyogenes</i>	Streptine 1	VGARYLA(Abu)PGAAWKLVAF(Dhb)(Dhb)(Dhb)VK	Possibles ponts thioéthers entre (Ala3-S-Ala7), (Abu8-S-Ala12) et (Ala11-S-Ala17)	46	Petit ORF	Autoproduction, activité antimicrobienne	SrtK, SrtR	Wescombe and Tagg (2003)
Phéromones de bactériocines de classe II	<i>Carnobacterium maltaromaticum</i>	CbaX	SINSQIGKATSSISKCVSFFFKC	Non	41	Petit ORF	Carnobactériocine A	CbaK, CbaR	van Belkum <i>et al.</i> (2007)
	<i>Enterococcus faecium</i>	EntF	AGTKPQKGPASNIVECVSFLFKCN	Non	41	Petit ORF	Entéroisine A et B	EntK, EntR	Nilsen <i>et al.</i> (1998), O'Keefe <i>et al.</i> (1999)
	<i>Lactobacillus sakei</i>	IP-TX	TPGGFDIISGGPHVAQDVLNAIKDFFK	Non	45	Petit ORF	Sakacine T et X	StxK, StxR	Vaughan <i>et al.</i> (2003)
	<i>Streptococcus thermophilus</i>	BlpC-1 BlpC-2	SGWMDYINGFLKGGGQRTLPKDYNIQA SGWMDYINGFLKGGGQRT	Non	53	Petit ORF	Blp, Bactériocine-Like Peptide	BlpH, BlpR	Fontaine <i>et al.</i> (2007)

^a Exemples non exhaustifs

^b Abu, acide 2-aminobutyrique; Dha, 2,3-didésydroalanine; Dhb, acide 2,3-didésydroaminobutyrique

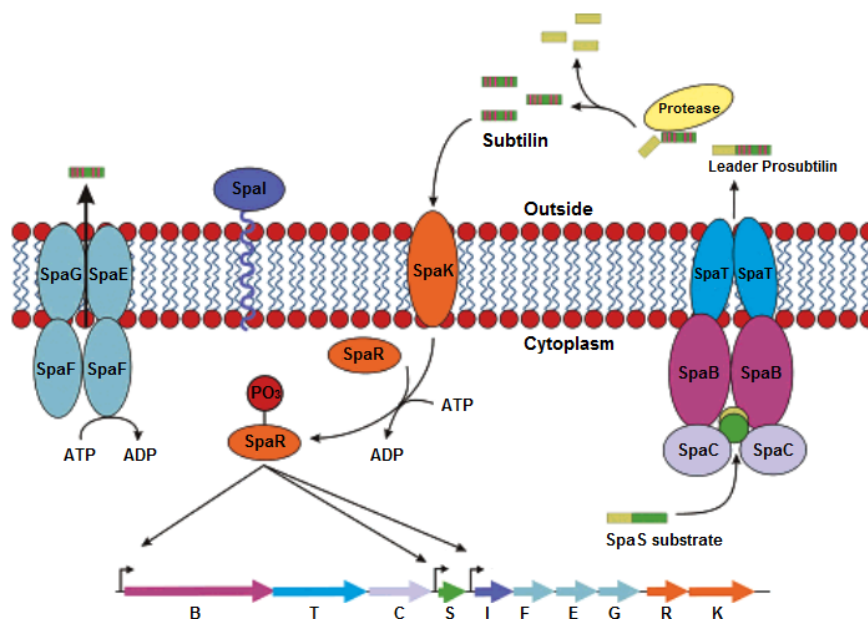


FIGURE 1.17 : Processus de biosynthèse de la subtiline. La subtiline mature fait office de signal permettant d'enclencher l'autophosphorylation de l'histidine kinase SpaK. Celle-ci transfère ensuite le groupement phosphate au régulateur de réponse SpaR. Une fois ainsi activé, SpaR régule l'expression de trois unités transcriptionnelles (*spaS*, *spaBTC* et *spaI-FEG*) impliquées dans la biosynthèse et l'auto-immunité à la subtiline. Le précurseur SpaS est modifié par un complexe multi-enzymatique associé à la membrane (SpaBC) puis transporté par l'ABC-transporteur SpaT. Une fois dans le milieu extérieur, le peptide est clivé afin de libérer sa forme mature. SpaEFG, un ABC-transporteur, ainsi que SpaI seraient impliqués dans l'auto-protection contre la subtiline. Adapté de Chatterjee *et al.* (2005)

On compte également quelques phéromones non modifiées dans les systèmes de détection extracellulaire. En particulier, les phéromones de compétence du système ComCDE des *streptococci* (CSP) ainsi que diverses phéromones régulant la production de bactériocines de classe II présentent toutes une structure linéaire. Bien qu'impliquées dans des fonctions différentes, ces deux classes de phéromones ont la particularité de correspondre à la partie N-terminale du précurseur peptidique clivé après un motif Gly-Gly conservé qui est également caractéristique des précurseurs des bactériocines de classe II elles-mêmes (Håvarstein *et al.*, 1994). Par comparaison avec les phéromones hautement modifiées précédemment citées, se pose la question de l'efficacité d'une telle structure non modifiée en regard de la stabilité et de la durée de vie des phéromones dans un contexte où l'activité protéolytique de surface de la bactérie peut être intense. Il est en effet vraisemblable que les peptides modifiés échappent plus efficacement à la dégradation protéolytique et véhiculent ainsi plus efficacement le signal moléculaire. Néanmoins, malgré cette meilleure efficacité pré-supposée, de nombreux autres systèmes également utilisent des phéromones non modifiées, notamment les systèmes à détection intracellulaire.

1.3.2.2 Peptides impliqués dans la détection intracellulaire

Dans la voie de signalisation intracellulaire, les peptides phéromonones sont transportés dans la cellule par un ABC-transporteur d'oligopeptides, généralement Opp ou son homologue Ami chez les *streptococci* (Cook et Federle, 2014). Une fois internalisé, le peptide va soit activer un régulateur transcriptionnel, soit inhiber l'action d'une aspartylphosphatase phosphatase RAP engagée dans une cascade de phosphorylation (Figure 1.16; Monnet *et al.*, 2014). Ces différents récepteurs intracellulaires, tant régulateurs transcriptionnels que phosphatases RAP, font partie d'un large groupe structural de régulateurs appelé RRNPP (Monnet et Gardan, 2015; Perez-Pascual *et al.*, 2016), nommé d'après les différentes familles de régulateurs qui le composent (Rgg, Rap, NprR, PlcR, PrgX). Ces régulateurs sont basés sur une architecture commune caractérisée par des motifs TPRs (Tetratricopeptide Repeats) responsables des interactions protéine-protéine, et par un domaine N-terminal HTH (Neiditch *et al.*, 2017). Ils agissent de concert avec des familles de peptides phéromones qui leur sont spécifiques (Tableau 1.10). En comparaison avec les systèmes de détection extracellulaire, il existe encore de nombreuses zones d'ombre concernant les différents effecteurs impliqués dans la détection intracellulaire, en particulier autour des mécanismes de maturation et de sécrétion des phéromones. Trois revues récentes (Monnet *et al.*, 2014; Perez-Pascual *et al.*, 2016; Neiditch *et al.*, 2017) dressent un panorama complet des connaissances actuelles sur les systèmes de *quorum sensing* basés sur les RRNPP. Elles détaillent en particulier l'organisation génétique de ces systèmes, leur mode d'action, les fonctions cellulaires contrôlées, ainsi que les principaux points de blocage dans la compréhension de chacune des étapes mises en œuvre et pour l'identification de nouveaux systèmes.



FIGURE 1.18 : Processus de maturation de cCF10 d'*E. faecalis*. La signal peptidase II (SPII) commence par détacher la séquence signal de la prolipoprotéine CcfA. La séquence signal est ensuite maturée par Eep. Puis, les 3 résidus C-terminaux du décapeptide cCF10 sont retirés par une activité carboxypeptidase (C.P.) libérant ainsi la forme mature de la phéromone. Les sites de clivage sont indiqués par des triangles. Chandler et Dunny, 2008

D'un point de vue structural, contrairement aux phéromones impliquées dans la détection extracellulaire, les peptides associés aux régulateurs RRNPP ne présentent pas de modification post-traductionnelle et sont directement le produit de l'hydrolyse d'un précurseur peptidique encodé par un sORF. Il existe néanmoins des exceptions notables à cette dernière règle : les sexe-phéromones d'*E. faecalis*, qui sont libérées à partir de la séquence signal de lipoprotéines (Antiporta et Dunny, 2002). Ce clivage fait intervenir la peptidase signal II, la protéase Eep ainsi qu'une ou plusieurs carboxypeptidases encore non déterminées (Figure

1.18; Chandler et Dunny, 2008). Ce mode de production particulier a également été observé chez *S. pyogenes* pour une phéromone intervenant dans la régulation de l'expression d'un facteur de virulence (Shelburne *et al.*, 2011). La Figure 1.19A présente les mécanismes généraux régissant la régulation de la conjugaison chez *E. faecalis*. On peut notamment relever le fait que la sexe-phéromone cCF10 (LVTLVfV) ainsi que son inhibiteur associé iCF10 (AITLIFI) présentent la particularité de transiter par une protéine de fixation dédiée au moment de leur réimportation : PrgZ (Leonard *et al.*, 1996). Celle-ci fonctionne en association avec le translocon OppBCDF mais, contrairement à son homologue OppA, présente une stricte spécificité pour ces deux phéromones hydrophobes. Cette spécificité lui permet de les sélectionner dans le milieu extérieur à de très faibles concentrations, de l'ordre d'une dizaine de picomoles dans des conditions physiologiques (Mori *et al.*, 1988; Dunny, 2007), alors que la concentration habituelle des phéromones peptidiques est estimée à quelques nanomoles (Monnet *et al.*, 2014). Le transport de cCF10 et iCF10 peut néanmoins être initié par OppA mais nécessite des concentrations plus élevées non conformes avec les concentrations physiologiques observées *in vivo* (Fixen *et al.*, 2007). La caractérisation de la structure tridimensionnelle de PrgZ a révélé une organisation de son site de fixation hautement adaptée aux chaînes latérales des deux phéromones (Berntsson *et al.*, 2012). En particulier, quatre poches hydrophobes accommodent les chaînes latérales des résidus 2, 4, 6, et 7, et deux poches hydrophiles sont alignées avec les résidus 3 et 5, le troisième résidu des deux phéromones (Thr) jouant un rôle important dans la spécificité d'interaction avec la protéine (Fixen *et al.*, 2007). PrgZ constitue néanmoins le seul exemple connu de SBP spécifiquement dédiée à l'import de peptides phéromones. Toutes les autres phéromones connues transitent en effet par OppA ou par les différentes AmiA/Alia des *streptococci* vraisemblablement de façon moins efficace en raison de la spécificité plus large de ces protéines.

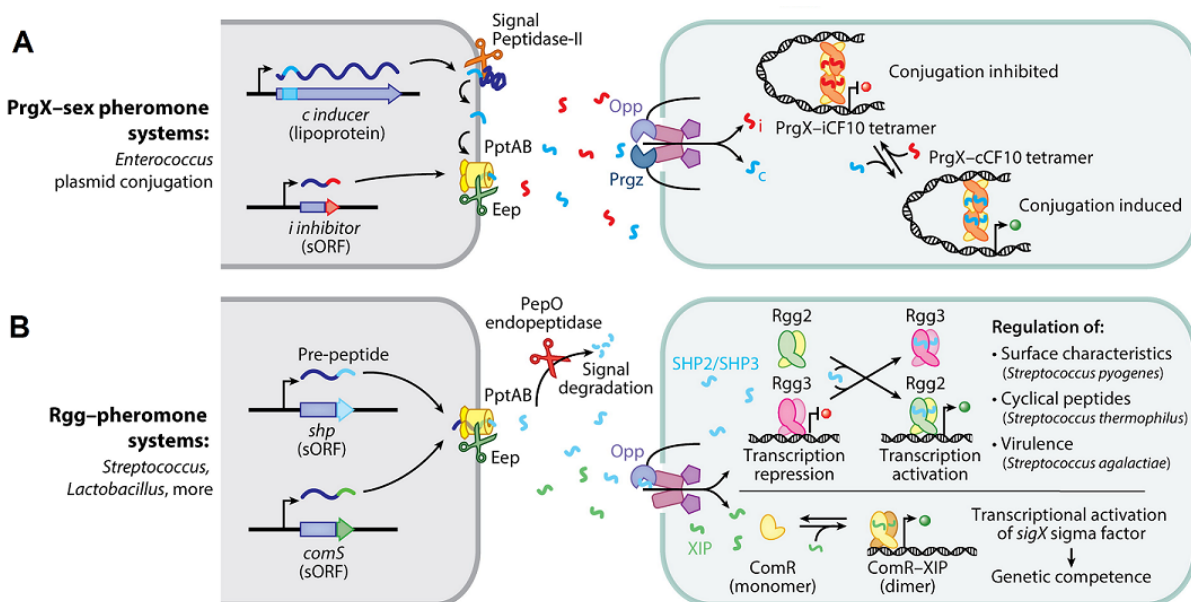


FIGURE 1.19 : (Légende page suivante)

FIGURE 1.19 : Exemples de voies de signalisation impliquant des phéromones associées aux régulateurs RRNPP. (A) Sexe-phéromones associées au régulateur PrgX. Les phéromones inductrices de la conjugaison chez les entérocoques (bleu) sont libérées chez les cellules donneuses et receveuses à partir de la séquence signal de lipoprotéines par la signal peptidase II et la protéase Eep, puis sont exportées par un ABC-transporteur nommé PptAB. Les phéromones inhibitrices de la conjugaison (rouge) sont produites à partir de sORFs uniquement chez les cellules donneuses afin de prévenir l'auto-induction de la conjugaison et sont également exportées par PptAB. PrgZ, un homologue d'OppA, améliore significativement la réception du signal en raison de sa haute spécificité pour le couple cCF10/iCF10. PrgX régule le transfert conjugatif du plasmide pCF10 en réprimant l'expression des gènes responsables de la conjugaison. La répression est favorisée lorsque iCF10 est fixée à PrgX, formant un complexe tétramérique qui se lie à deux sites de l'ADN afin de former une boucle perturbant l'activité de l'ARN polymérase. Cette répression est levée par le tétramère cCF10-PrgX qui altère la conformation du complexe et permet à l'ARN polymérase de transcrire les gènes responsables de la conjugaison. L'équilibre entre répression/activation est dicté par le ratio cCF10/iCF10 dans l'environnement immédiat de la cellule. Une cellule donneuse seule ou en présence d'autres cellules donneuses verra son système conjugatif réprimé en raison d'une plus forte abondance d'iCF10. En contact avec une cellule receveuse, ce ratio est modifié en faveur de cCF10, permettant d'enclencher le transfert entre les deux cellules. (B) Phéromones associées aux régulateurs Rgg. Les phéromones des Rgg (SHP et ComS) sont encodées par des sORFs et suivent un processus de maturation et d'export similaire à celui des sexe-phéromones des entérocoques impliquant également le couple Eep/PptAB. Dans certaines conditions, les SHPs sont dégradés par la peptidase PepO (modulation du signal), ce qui constitue le premier exemple documenté d'une enzyme impliquée dans la dégradation de phéromones appartenant aux systèmes RRNPP. Deux phéromones de *S. pyogenes* (SHP2, SHP3) sont fonctionnellement équivalentes et peuvent se lier indifféremment à Rgg2 et Rgg3. Rgg3, sans sa phéromone associée, agit comme un répresseur de l'expression de gènes impliqués dans des caractéristiques de surface de *S. pyogenes*. Lorsque SHP2 ou SHP3 se lient à leurs récepteurs, Rgg3 est déplacé de son site de fixation sur l'ADN, faisant place pour Rgg2 qui agit comme activateur de l'expression génétique. Dans d'autres espèces, des orthologues à Rgg2 et Rgg3 sont impliqués dans la régulation de systèmes variés, incluant la production d'un peptide cyclique chez *S. thermophilus* ou la virulence chez *S. agalactiae*. ComRS constitue un autre système Rgg-phéromone. ComS est encodé par un sORF et se nomme XIP sous sa forme mature. Lorsque fixé à ComR, il induit l'expression du facteur de transcription SigX (ou ComX) qui contrôle l'expression des gènes de la compétence. D'après Neiditch *et al.*, 2017.

De même que pour les sexe-phéromones d'*E. faecalis*, l'hydrophobicité est aussi la principale caractéristique des phéromones associées aux régulateurs Rgg en raison de leur teneur en BCAAs (Tableau 1.10). Ces phéromones sont spécifiques aux *streptococci* et ont été divisées en deux sous-groupes : les SHP d'une part, et les peptides de compétence ComS, également appelés XIP dans leur forme mature (Figure 1.19B). Les SHP ont été initialement identifiés chez *S. thermophilus* (Ibrahim *et al.*, 2007b) et présentent une séquence hautement caractéristique. Les peptides ComS, s'ils sont également associés à des régulateurs Rgg appelés ComR, s'éloignent de la séquence canonique des SHP en incluant notamment des résidus aromatiques. Les autres familles de phéromones - Phr, PapR et NprX - incluent

en revanche davantage de résidus polaires et chargés. Finalement, l'ensemble des peptides à détection intracellulaire a pour principale particularité une taille significativement plus courte (entre 5 et 10 acides aminés) que leurs homologues non modifiés détectés extracellulairement (ComX, Plt et phéromones de bactériocines de classe II, plus de 17 résidus).

Enfin, sur un niveau plus général et pour conclure cette partie, on peut relever le fait qu'il existe un décalage entre le nombre restreint de systèmes actuellement identifiés et caractérisés en comparaison avec l'important potentiel des bactéries en sORFs (voir partie 1.2.3.1.1), et ce tant pour les systèmes à détection extracellulaire qu'intracellulaire. De plus, se pose également la question du nombre minimal de résidus à considérer pour la conservation des sORFs dont la limite est aujourd'hui fixée à 10 résidus. L'exemple récent du SIP de *S. pyogenes* (MWLLLLFL; Do *et al.*, 2017) synthétisé directement sous une forme mature de 8 résidus laisse entrevoir un immense champ de possibilités pour l'identification de nouveaux systèmes basés sur des peptides de courte taille.

TABLEAU 1.10 : Familles des peptides phéromones impliqués dans la voie de détection intracellulaire des systèmes de *quorum sensing* des bactéries à Gram positif. Adapté de Monnet *et al.* (2014), Neiditch *et al.* (2017) et de la base de données Quorumpeps.

Famille	Espèce	Nom	Séquence ^a	Modification post-traductionnelle	Taille du propeptide (aa)	Origine du peptide	Fonction contrôlée	Récepteur	Référence
PhR , Phosphatase Regulator	<i>Bacillus anthracis</i>	BXA0205phr	GHTGG	Non	46	Petit ORF	Sporulation	Rap	Bongiorni <i>et al.</i> (2006)
	<i>Bacillus subtilis</i>	PhrA PhrC PhrE PhrF PhrG PhrH PhrI PhrK Phr60	ARNQT ERGMT SRNVT QRGMI EKMIG TDRNTT DRVGA/ADRVGA ERPVG SRNAT/ASRNAT	Non	40-45	Petit ORF	Compétence, sporulation, production d'enzymes de dégradation extracellulaires	RapA RapB, RapC RapE RapF RapG RapH RapI RapK Rap60	Lazizzera <i>et al.</i> (1997), Perego (1997), Jiang <i>et al.</i> (2000), Perego et Brannigan (2001), Ogura <i>et al.</i> (2003), Auchtung <i>et al.</i> (2005), Lanigan-Gerdes <i>et al.</i> (2008), Mirouze <i>et al.</i> (2011), Boguslawski <i>et al.</i> (2015)
PapR	<i>Bacillus cereus</i>	PapR-I PapR-II PapR-III PapR-IV	ADLPFEF SDMPFEF NEVPFEF SDLPFEH	Non	48	Petit ORF	Virulence	PlcR-I PlcR-II PlcR-III PlcR-IV	Slamti et Lereclus (2005), Bouillaud <i>et al.</i> (2008)
	<i>Streptococcus pneumoniae</i>	PhrA ^b	DVGKAD	Non	56	Petit ORF	Lantibiotique	TrpA ^b	Hoover <i>et al.</i> (2015)
NprX	<i>Bacillus cereus</i> , <i>Bacillus thuringiensis</i>	NprX-I	SKPDIYG	Non	43	Petit ORF	Nécrotrophisme	NprR-I NprR-II NprR-III NprR-IV NprR-V NprR-VI NprR-VII	Perchat <i>et al.</i> (2011), Dubois <i>et al.</i> (2013)
		NprX-II	SKPDIYG						
		NprX-III	SNPDIYG						
		NprX-IV	SRPDVLT						
		NprX-V	WTSDIYG						
		NprX-VI	WKPD[N/V/T][Y/L]G						
		NprX-VII	WRPDMSI						
Sexe-phéromones	<i>Enterococcus faecalis</i>	cAD1	LFSLVLAG	Non	143 166 275 272 234	Séquence signal de lipoprotéines	Conjugaison	TraA TraA PrgX - TraA	Clewell et Weaver (1989), Mori <i>et al.</i> (1988), Clewell <i>et al.</i> (2000), Flannagan et Clewell (2002)
		cAM373	AIFILAS						
		cCF10	LVTLVAV						
		cOB1	VAVLVLAG						
		cPD1	FLVMFLSG						
Inhibiteurs de sexe-phéromones	<i>Enterococcus faecalis</i>	iAD1	LFVVTIVG	Non	21-23	Petit ORF	Inhibition de la conjugaison	TraA TraA PrgX - TraA	Clewell et Weaver (1989), Mori <i>et al.</i> (1988), Clewell <i>et al.</i> (2000), Flannagan et Clewell (2002)
		iAM373	SIFTLVA						
		iCF10	AITLIFI						
		iOB1	SLTLILSA						
		iPD1	ALILTLVS						
SHP , Short Hydrophobic Peptide	<i>Streptococcus agalactiae</i>	SHP	DILIVGG	Non	22	Petit ORF	Virulence	Rgg1555	Fleuchot <i>et al.</i> (2013)
	<i>Streptococcus pyogenes</i>	SHP2	DILIVGG	Non	22	Petit ORF	Biofilm	Rgg2, Rgg3 Rgg2, Rgg3	Chang <i>et al.</i> (2011), Aggarwal <i>et al.</i> (2014)
		SHP3	DIIIVGG	23					
<i>Streptococcus thermophilus</i>	SHP1358 SHP1299	EGIIVVVG DIIIFPPFG	Non	23 26	Petit ORF	Streptide (peptide cyclique) Inconnue	Rgg1358 Rgg1299	Fleuchot <i>et al.</i> (2011, 2013)	
ComS/XIP^c , sigX-Inducing Peptide	<i>Streptococcus mutans</i>	ComS-II	GLDWWSL	Non	17	Petit ORF	Compétence	ComR ^d	Mashburn-Warren <i>et al.</i> (2010)
	<i>Streptococcus pyogenes</i>	ComS-II	SAVDWWRL	Non	31	Petit ORF	Compétence	ComR ^d	Mashburn-Warren <i>et al.</i> (2012)
		ComS-II	EFDWWNLG	32					
	<i>Streptococcus salivarius</i>	ComS-I	PYFTGCL	Non	24	Petit ORF	Compétence	ComR ^d	Fontaine <i>et al.</i> (2010), Gardan <i>et al.</i> (2013)
	<i>Streptococcus suis</i>	ComS-III ComS-III	WGTWVEE WGKWTDG	Non	21	Petit ORF	Compétence	ComR ^d	Fontaine <i>et al.</i> (2015)
<i>Streptococcus thermophilus</i>	ComS-I	LPYFAGCL	Non	24	Petit ORF	Compétence	ComR ^d	Fontaine <i>et al.</i> (2010), Gardan <i>et al.</i> (2013)	

^a Exemples non exhaustifs

^b Le système Tpr/Phr de *S. pneumoniae* est une addition récente à cette classification orthologue au système PlcR/PapR de *B. cereus* (Hoover *et al.*, 2015).

^c ComS, nom général; XIP, nom de la forme mature du peptide.

^d Les régulateurs ComR appartiennent à la famille des Rgg.

1.3.2.2.3 Influence des peptides nutritionnels sur les systèmes de *quorum sensing* des bactéries lactiques

A travers les quelques exemples répertoriés dans les Tableaux 1.9 et 1.10, il apparaît que divers systèmes de régulation *quorum sensing* sont présents dans plusieurs espèces de bactéries lactiques. Les systèmes de production de bactériocines régulés par un mécanisme de détection extracellulaire en sont des exemples particulièrement bien connus. De plus, depuis quelques années, les régulateurs Rgg associés à de courts peptides hydrophobes (SHP/ComS) des *streptococci* attirent également l'attention. En particulier, chez *S. thermophilus*, le système ComRS régulant la compétence est dans l'ensemble bien décrit. En revanche, la fonction physiologique de ses systèmes Rgg-SHP reste pour l'instant inconnue.

En se replaçant dans un contexte de croissance d'une bactérie lactique en contact avec de multiples peptides de dégradation - c'est-à-dire provenant soit du milieu nutritif, soit de sa propre activité protéolytique - la question de l'influence de ces peptides sur les systèmes de régulation *quorum sensing* de la bactérie mérite d'être abordée. Seront envisagés ici uniquement les peptides dont la fonction première est supposée d'ordre nutritionnel. La question d'une éventuelle influence de tels peptides sur les systèmes de *quorum sensing* peut s'envisager selon deux modalités : influence négative (interférence, inhibition) ou positive. Ce dernier cas de figure correspondant à la situation où un peptide nutritionnel présente une activité similaire à une phéromone bactérienne. Les peptides nutritionnels sont métabolisés par la bactérie en suivant une voie d'utilisation précise (voir partie 1.3.1) qui n'est que partiellement distincte de celle des phéromones. Il est en effet notable qu'à la fois les peptides nutritionnels et la majorité des phéromones à détection intracellulaires sont transportés par un même système (Opp/Ami). De ce transport partagé résulte un phénomène de compétition bien connu entre peptides nutritionnels et phéromones qui, en fonction de la quantité des peptides nutritionnels environnants, peut aboutir à une inhibition totale de la communication inter-cellules. Ainsi, il a été montré que l'expression de *comX*, le gène cible du système ComRS régulant la compétence de *S. thermophilus*, était réduite d'un facteur 200 en milieu M17 par rapport à un milieu chimiquement défini (MCD) ne contenant pas de peptides, empêchant ainsi toute possibilité de transformation génétique (Gardan *et al.*, 2009). De même, un MCD complétement avec un digestat tryptique de caséines (tryptone) entraîne une répression quasi-totale de l'expression de *comX* à partir de 0.5 g/l de tryptone (Gardan *et al.*, 2013). En revanche, une complémentation à 10 g/l d'un MCD avec des casaminoacides, un digestat de caséines contenant principalement des acides aminés libres, n'aboutit qu'à une répression partielle (28 %) de l'expression de *comX* par rapport à un MCD non complétement. La stratégie déployée par *E. faecalis* à travers sa SBP PrgZ spécifique à un couple de phéromones apparaît donc comme une solution efficace pour prévenir toute inhibition de communication par compétition de transport. PrgZ constitue néanmoins une exception dans le monde bactérien. Pourquoi une telle stratégie n'est-elle pas davantage

répandue? Il n'existe probablement pas de réponse universelle à cette question qui mériterait d'être traitée au cas par cas selon les systèmes de communication et les espèces voire souches bactériennes. En ce qui concerne l'exemple de *S. thermophilus* qui est une bactérie adaptée à la croissance en lait, la teneur en peptides y est de l'ordre de 50 mg/l (Juillard *et al.*, 1998), ce qui est compatible avec le fonctionnement de son module de compétence naturelle (Gardan *et al.*, 2013). L'éventuel avantage sélectif conféré par une protéine de fixation dédiée aux phéromones dans ce contexte là n'apparaît donc pas comme évident.

L'interférence par compétition de transport constitue donc la première et vraisemblablement la plus répandue forme d'influence des peptides nutritionnels sur les systèmes de *quorum sensing*. Il est néanmoins possible d'envisager une deuxième forme d'influence dans laquelle les peptides nutritionnels interagiraient directement avec les récepteurs des phéromones, à savoir les domaines senseurs extracellulaires des histidines kinases ou les régulateurs RRNPP intracellulaires. Cette interaction hypothétique repose sur la combinaison de trois conditions : i) une séquence et une structure des peptides nutritionnels suffisamment proche de celles des phéromones, ii) une permissivité du récepteur pour des peptides hétérologues, et iii) une absence de dégradation intracellulaire complète dans le cas d'une détection intracellulaire. La première de ces trois conditions rend en pratique hautement improbable la possibilité d'une interaction entre un peptide nutritionnel et une histidine kinase d'un système à détection extracellulaire. Les phéromones associées à ces systèmes sont en effet soit dotées de modifications post-traductionnelles spécifiques, soit d'une taille telle que la probabilité qu'un peptide nutritionnel de taille égale ait une séquence proche soit quasi-nulle. En revanche, les phéromones des systèmes à détection intracellulaire étant non modifiées et de plus courte taille, la probabilité qu'un peptide nutritionnel ait une identité de séquence ne serait-ce qu'approchante est plus importante.

Dans l'éventualité de la présence d'un tel peptide nutritionnel à séquence proche d'une phéromone, son effet sur un récepteur RRNPP dépend ensuite du degré de permissivité ou de spécificité de la protéine en question. Les systèmes de communication cellulaire bactériens sont généralement spécifiques au niveau de l'espèce voire de la souche. Il existe en effet une diversité allélique au sein d'une même famille régulateur-phéromone qui a amené à la définition des phérotypes qui correspondent à des groupes de spécificité. Une histidine kinase ou un régulateur RRNPP d'un phérotypé donné est ainsi capable de détecter des phéromones appartenant au même phérotypé, mais pas celles d'un autre phérotypé, et pourra même être dans certains cas inhibé par des variants de sa phéromone associée. De tels phérotypes ont été observés au sein des systèmes Agr de *S. aureus* (Ji *et al.*, 1997), ComCDE de *S. pneumoniae* (Håvarstein *et al.*, 1997), ou encore Phr et PapR des *bacilli* (Perego et Brannigan, 2001; Slamti et Lereclus, 2005). La spécificité de ces systèmes est avant tout dépendante de la séquence primaire des phéromones, une mutation d'un seul acide aminé étant susceptible d'entraîner d'importantes perturbations dans la transduction du signal (Perego, 1997;

MDowell *et al.*, 2001). Il existe néanmoins plusieurs exceptions à cette règle, en particulier au sein des systèmes Rgg-SHP et ComRS des *streptococci*. Il a en effet été montré que les systèmes Rgg-SHP de *S. agalactiae* et *S. mutans* pouvaient être activés par des SHP provenant d'une autre espèce (Fleuchot *et al.*, 2013). Plus surprenant, Fontaine *et al.* (2013) ont non seulement montré que le régulateur ComR de *S. thermophilus* pouvait être activé par des peptides ComS hétérologues, mais également par des peptides dérivant de caséines de lait imitant grossièrement la séquence de ComS de *S. thermophilus*, en particulier le peptide GAWYYVPL. Cette expérience constitue à ce jour le seul exemple répertorié d'activation d'un système de *quorum sensing* par des peptides exogènes nutritionnels, qui plus est chez une bactérie lactique. Il convient néanmoins de préciser que les peptides de caséines utilisés ont été identifiés sur la base d'une recherche d'une séquence ComS consensus dans les séquences des caséines et produits synthétiquement. Rien ne permet d'affirmer que de tels peptides sont libérés *in vivo* et ce dans des concentrations permettant leur détection. Il s'agit néanmoins d'une preuve de concept intéressante soutenant en partie l'étude des peptides nutritionnels en tant que potentielles molécules régulatrices.

1.3.2.3 Peptides régulateurs : bilan

La régulation de fonctions physiologiques par les peptides peut donc se réaliser selon deux modalités chez les bactéries lactiques : i) régulation indirecte, où ce sont les acides aminés constitutifs des peptides qui vont interagir avec un régulateur transcriptionnel, et ii) régulation directe, les peptides eux-mêmes activant un effecteur de la régulation de l'expression génétique. Cette partie a ainsi permis de souligner l'ambivalence des peptides considérés comme nutritionnels du fait de leur intervention connue en regard de la régulation indirecte (exemple de CodY), mais aussi de la régulation directe en raison de leur interférence possible au niveau des systèmes de communication bactériens (compétition de transport), voire leur éventuelle activité « *pheromone-like* ».

Dans ce dernier cas de figure, comment se fait-il qu'un tel peptide nutritionnel à activité régulatrice ne soit pas immédiatement dégradé par les peptidases intracellulaires après son transport, conformément au dogme de la nutrition azotée, et puisse au contraire interagir avec un régulateur RRNPP? Cette question, ainsi formulée, pré-suppose de manière simpliste que les peptides nutritionnels sont exclusivement pris en charge par les peptidases, et, réciproquement, que les phéromones n'interagissent qu'avec leurs récepteurs cibles. Or, la réalité est sans doute plus complexe. Il a notamment été récemment démontré chez *S. pyogenes* que l'endopeptidase PepO intervenait dans l'extinction du signal d'un système Rgg-SHP en dégradant la phéromone en réponse à des stress environnementaux, ces stress ayant pour effet d'augmenter spécifiquement l'expression de *pepO* (Wilkening *et al.*, 2016). Le sort d'un peptide tout juste internalisé est en réalité probablement déterminé par une combinai-

son dynamique de plusieurs facteurs, le premier d'entre eux étant l'affinité propre du peptide envers les peptidases d'un côté, et les récepteurs RRNPP de l'autre. Il serait ainsi intéressant d'évaluer expérimentalement la contribution des différentes peptidases dans la modulation du signal des systèmes de *quorum sensing* à détection intracellulaire, par exemple en utilisant des combinaisons de mutants simples et multiples de peptidases comme cela a été fait avec *L. lactis* pour suivre le devenir des peptides issus de la dégradation des caséines (Kunji *et al.*, 1996b).

1.4 Conclusion

L'objectif sous-jacent de cette étude bibliographique était de montrer que les peptides « simples », qu'ils soient issus d'un milieu de fermentation où du recyclage protéique bactérien, étaient considérés à tort comme étant uniquement des fragments de protéines pourvoyeurs d'acides aminés pour la production de biomasse. Bien que cela représente effectivement leur débouché principal, il apparaît que, pour un peptide donné, la nature de sa séquence peut avoir une influence considérable sur sa destinée, et par voie de conséquence sur la physiologie bactérienne. La structure primaire d'un peptide va en effet conditionner son efficacité de transport en premier lieu, ce qui donne lieu à des dynamiques d'approvisionnement, mais également son devenir intracellulaire. Dans l'immense majorité des cas, il sera dégradé par l'activité peptidolytique de la bactérie qui utilisera certains acides aminés particuliers, les BCAAs, comme indicateurs de richesse nutritionnelle de son environnement et adaptera son métabolisme en conséquence. Dans certains cas, il est également possible qu'un peptide nutritionnel, du fait d'une particularité de séquence, interagisse avec des facteurs de transcription et entraîne une modification de l'expression génétique de gènes cibles particuliers. Enfin, bien que cela n'ait pas été étudié ici, on peut également envisager que certains peptides présentent une bioactivité vis-à-vis de la bactérie sans que celle-ci soit nécessairement le produit d'une interaction avec un régulateur transcriptionnel. Les hydrolysats protéiques sont en effet une riche source de peptides bioactifs (Daliri *et al.*, 2017; Sánchez et Vázquez, 2017). Bien que ceux-ci aient été principalement étudiés dans un contexte de santé humaine ou animale, il est possible que certains peptides, du fait d'une physico-chimie particulière leur conférant certaines propriétés (activité antioxydante, chélation de métaux etc.), puissent également avoir un impact sur la physiologie bactérienne.

Les peptides sont donc des molécules dotées de nombreuses propriétés fonctionnelles. Quels sont les impacts de ces propriétés dans un contexte de production industrielle de ferments? La fabrication de levains lactiques est avant tout dictée par des impératifs de performances qui se déclinent sur deux niveaux : i) performances de production *stricto sensu*, où l'on va chercher à maximiser la productivité qui se définit comme le rapport entre la biomasse maximale obtenue en fin de fermentation et le temps d'obtention de cette biomasse, et ii) performances d'utilisation, qui se rapportent à la qualité des levains produits. Ces derniers doivent en effet présenter des propriétés fonctionnelles optimales lors de leur utilisation, la première de ces propriétés étant l'activité acidifiante. Ces deux niveaux de performances sont en interdépendance, car, au moment de leur récolte et conservation, la configuration métabolique des bactéries, qui résulte elle-même de l'historique de la culture, va grandement influencer la reprise d'activité du ferment et se répercuter sur les performances d'utilisation. Cette notion est notamment supportée dans la littérature par

le concept de « mémoire phénotypique », qui consiste tout simplement en la transmission de protéines dont la durée de vie est supérieure au temps de génération de la bactérie (Lambert et Kussell, 2014). Ainsi, les deux niveaux de performances peuvent être unifiés par la notion de « *fitness* ». Celle-ci se définit à l'échelle du métabolisme cellulaire comme étant la différence entre le bénéfice conféré par l'ensemble de la production protéique de la bactérie, et le coût nécessaire à cette production et à la maintenance protéique (Dekel et Alon, 2005). La maximisation du *fitness* dans un environnement donné passe donc par l'ajustement optimal des niveaux d'expression génétique et de synthèse protéique. Cette définition établie, il est aisé de voir comment les peptides, molécules capables d'influencer divers systèmes de régulation des bactéries lactiques, peuvent potentiellement avoir un impact tant lors de la production des ferments qu'*a posteriori* au moment de leur utilisation. L'extinction de systèmes coûteux et inutiles pendant la production ainsi que l'activation préalable de gènes conférant un avantage métabolique au moment de l'utilisation des ferments sont autant de stratégies souhaitables. Ceci peut s'envisager en modulant à façon le contenu en peptides du milieu de culture, ce qui passe par l'ajustement des conditions de leur obtention pendant l'étape de lyse de l'hydrolysats employé.

Toutefois, avant de mettre en œuvre une telle stratégie, certes innovante mais encore très hypothétique, de nombreux verrous restent à lever. Le premier d'entre eux est la caractérisation qualitative des mélanges complexes mal définis de peptides que sont les hydrolysats cellulaires et protéiques afin d'accéder à la connaissance de leur séquence. De même, il est nécessaire d'évaluer quel est l'impact réel de ces peptides nutritionnels sur le métabolisme bactérien, car, si leur potentiel régulateur théorique est établi, il faut également vérifier en pratique l'ampleur réelle des impacts métaboliques et leur nature. Ces deux axes constituent le cœur du travail de thèse réalisé qui sera présenté dans les chapitres suivants.

Chapitre 2

Séquençage et analyses génomiques de *Streptococcus thermophilus* N4L

Sommaire

2.1 Introduction	80
2.2 Complete Genome Sequence of the Industrial Fast-Acidifying Strain <i>Streptococcus thermophilus</i> N4L	80
2.3 Analyses génomiques complémentaires	83
2.3.1 Annotation complémentaire par l'outil PGAP du NCBI	83
2.3.2 Assignation d'information fonctionnelle de numéros de l'orthologie KEGG	85
2.3.3 Prédiction de la localisation cellulaire des protéines bactériennes	86
2.3.4 Identification de courts ORFs	87
2.4 Conclusion	89

2.1 Introduction

La souche bactérienne *S. thermophilus* N4L sur laquelle a été réalisée l'intégralité du travail de thèse a été fournie par le groupe Sacco. Il s'agit de l'un des ferments lactiques produits et commercialisés par l'entreprise et principalement utilisé pour sa haute capacité acidifiante. Son génome a été séquençé spécifiquement pour ce projet à l'aide d'un système PacBio qui utilise une technologie de séquençage haut-débit de dernière génération (*Single Molecule, Real-Time Sequencing*) produisant des *reads* de longue taille (>10000 pb). Couplée à une importante profondeur de séquençage (889X), cette technologie a permis d'assembler entièrement et de circulariser le génome de la bactérie qui a ensuite été annoté par l'intermédiaire de la plateforme Agmial (Bryson *et al.*, 2006). Ce travail a fait l'objet d'un article dans *Microbiology Resource Announcements* présenté dans la partie suivante. Diverses autres analyses ont ensuite été réalisées afin d'approfondir la caractérisation du génome bactérien et servir de support aux approches omiques développées dans le projet de thèse. Les éléments clés de ces analyses seront présentés en dernière partie de ce chapitre.

2.2 Complete Genome Sequence of the Industrial Fast-Acidifying Strain *Streptococcus thermophilus* N4L

Article publié dans *Microbiology Resource Announcements*, Volume 7, le 30 août 2018.



Complete Genome Sequence of the Industrial Fast-Acidifying Strain *Streptococcus thermophilus* N4L

Lucas Proust,^{a,c} Valentin Loux,^b Véronique Martin,^b Cristian Magnabosco,^d Martin Pedersen,^d Véronique Monnet,^a Vincent Juillard^a

^aMicalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

^bMaIAGE, INRA, Université Paris-Saclay, Jouy-en-Josas, France

^cProcelys, Lesaffre Group, Maisons-Alfort, France

^dSacco S.r.l., Cadorago, Italy

ABSTRACT *Streptococcus thermophilus* is one of the most used dairy starters for the production of yogurt and cheese. We report here the complete genome sequence of the industrial strain *S. thermophilus* N4L, which is used in dairy technology for its fast-acidifying phenotype.

Streptococcus thermophilus is a well-known lactic acid bacterium (LAB) widely used for the production of dairy products in association with other LAB species, such as *Lactococcus lactis* and *Lactobacillus delbrueckii* subsp. *bulgaricus*, for cheese and yogurt production, respectively. It belongs to the salivarius group and recently evolved from a commensal ancestor (1). Nowadays, it can be naturally found all around the world in dairy production facilities, whether artisanal or industrial (2). For industrial use, strains are carefully selected for their ability to ferment milk and confer desirable properties to the final product. In particular, due to the high prevalence of the cell wall proteinase *prtS* gene (1) in industrial isolates of *S. thermophilus*, they are often associated with higher productivity in milk when compared to their wild or traditional counterparts. The proteinase PrtS is responsible for the breakdown of caseins into peptides, which are used as sources of amino acids, thus allowing faster growth and acidification than do PrtS-negative strains. These released peptides are then internalized inside the bacterial cell primarily by the Ami oligopeptide transport system, which includes one or several peptide-binding proteins depending on the strain (3).

S. thermophilus N4L is an industrial strain used precisely for its fast-acidifying phenotype. A stationary-phase culture (M17 with 2% lactose) was subjected to genomic DNA (gDNA) extraction using the GenElute bacterial gDNA kit (Sigma-Aldrich, USA) following the protocol indicated by the manufacturer. Its gDNA was then sequenced (Macrogen, Inc., South Korea) on a PacBio RS II system (Pacific Biosciences, USA) using single-molecule real-time (SMRT) sequencing technology (4). A total of 150,292 polymerase reads were obtained, displaying an average length of 11,243 bp. These reads, after quality filtering and adapter trimming, were converted into 213,700 subreads with an average length of 7,626 bp. They were subsequently used for *de novo* genome assembly. For that purpose, Canu version 1.6 was employed with all default settings for subread correction, trimming, and assembly (5). The resulting genome corresponded to a unique circular chromosome of 1,831,756 bp with an 889× coverage depth and a G+C content of 39.05%.

Annotation was performed with the AGMIAL annotation platform (6). It revealed the presence of 2,300 open reading frames (ORFs), including 67 tRNAs and 18 rRNAs. This sequencing work also confirmed the presence of the genomic island encoding the *prtS* gene, as well as two copies of the oligopeptide-binding proteins AmiA1 and AmiA3, which explains the high acidification rates in milk observed with this strain. All of the

Received 20 July 2018 Accepted 27 July 2018
Published 30 August 2018

Citation Proust L, Loux V, Martin V, Magnabosco C, Pedersen M, Monnet V, Juillard V. 2018. Complete genome sequence of the industrial fast-acidifying strain *Streptococcus thermophilus* N4L. *Microbiol Resour Announc* 7:e01029-18. <https://doi.org/10.1128/MRA.01029-18>.

Editor Julie C. Dunning Hotopp, University of Maryland School of Medicine

Copyright © 2018 Proust et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Vincent Juillard, vincentjuillard@inra.fr.

genes required for natural competence are also present in the genome (7, 8), suggesting that the strain could theoretically be naturally transformable.

Data availability. The complete genome sequence of *S. thermophilus* strain N4L has been deposited in GenBank under the accession no. [LS974444](https://doi.org/10.26434/chemrxiv-2018-05-01) (BioProject no. [PRJEB27286](https://doi.org/10.26434/chemrxiv-2018-05-01)).

ACKNOWLEDGMENTS

The bacterial strain was provided by Sacco S.r.l., which also supported the sequencing costs. This work was performed in the framework of a project funded by the Association Nationale de la Recherche et de la Technologie (ANRT).

REFERENCES

1. Delorme C, Bartholini C, Bolotine A, Ehrlich SD, Renault P. 2010. Emergence of a cell wall protease in the *Streptococcus thermophilus* population. *Appl Environ Microbiol* 76:451–460. <https://doi.org/10.1128/AEM.01018-09>.
2. Delorme C, Legravet N, Jamet E, Hoarau C, Alexandre B, El-Sharoud WM, Darwish MS, Renault P. 2017. Study of *Streptococcus thermophilus* population on a world-wide and historical collection by a new MLST scheme. *Int J Food Microbiol* 242:70–81. <https://doi.org/10.1016/j.ijfoodmicro.2016.11.016>.
3. Garault P, Le Bars D, Besset C, Monnet V. 2002. Three oligopeptide-binding proteins are involved in the oligopeptide transport of *Streptococcus thermophilus*. *J Biol Chem* 277:32–39. <https://doi.org/10.1074/jbc.M107002200>.
4. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomanev A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korf J, Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138. <https://doi.org/10.1126/science.1162986>.
5. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
6. Bryson K, Loux V, Bossy R, Nicolas P, Chaillou S, van de Guchte M, Penaud S, Maguin E, Hoebeke M, Bessières P, Gibrat J-F. 2006. AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system. *Nucleic Acids Res* 34:3533–3545. <https://doi.org/10.1093/nar/gkl471>.
7. Fontaine L, Wahl A, Flécharde M, Mignolet J, Hols P. 2015. Regulation of competence for natural transformation in streptococci. *Infect Genet Evol* 33:343–360. <https://doi.org/10.1016/j.meegid.2014.09.010>.
8. Gardan R, Besset C, Gitton C, Guillot A, Fontaine L, Hols P, Monnet V. 2013. Extracellular life cycle of ComS, the competence-stimulating peptide of *Streptococcus thermophilus*. *J Bacteriol* 195:1845–1855. <https://doi.org/10.1128/JB.02196-12>.

2.3 Analyses génomiques complémentaires

2.3.1 Annotation complémentaire par l'outil PGAP du NCBI

Le génome annoté de *S. thermophilus* N4L a fait l'objet d'un dépôt initial dans l'*European Nucleotide Archive* (ENA) qui a ensuite automatiquement transmis ces informations à la *DNA DataBank of Japan* (DDBJ) et GenBank du NCBI. Avant sa mise en ligne sur le site du NCBI, le génome a été entièrement ré-annoté automatiquement par l'outil interne de l'institut *Prokaryotic Genome Annotation Pipeline* (PGAP, version 4.6) (Tatusova *et al.*, 2016) afin d'assurer une annotation homogène avec les autres génomes bactériens déjà hébergés. Les deux annotations fournies par la plateforme Agmial et l'outil PGAP sont largement convergentes. Un total de 1744 ORFs ont en effet été détectés et annotés de façon identique, ce qui représente respectivement 75 % et 89 % des ORFs totaux identifiés dans les deux annotations (Tableau 2.1). La principale différence entre les deux outils se situe au niveau de la prédiction des pseudogènes qui n'est pas fournie par Agmial. Ainsi, environ 13 % des CDSs annotés par PGAP correspondent à des pseudogènes, ce qui est conforme aux nombreux événements de pertes de fonctions ayant façonné le génome de *S. thermophilus* (Hols *et al.*, 2005). La différente gestion des pseudogènes entre les deux outils explique également le nombre plus élevé de CDSs détectées par Agmial qui annote comme de multiples éléments les différents fragments résultant d'un pseudogène. A l'inverse, PGAP a tendance à annoter en un seul pseudogène ces différents fragments, d'où le nombre peu élevé de pseudogènes qu'ont en commun les deux annotations (92). On retrouve donc en majorité parmi les CDSs spécifiques de l'une ou l'autre annotation soit des fragments protéiques non fonctionnels (Agmial), soit directement des pseudogènes entiers (PGAP). Enfin, PGAP intègre une annotation d'éléments structuraux tels que les petits ARNs non codants (4) ou encore les loci CRISPR (3).

TABLEAU 2.1 : Principales caractéristiques des annotations génomiques réalisées par la plateforme Agmial et l'outil PGAP du NCBI.

	Agmial	NCBI	Partagé
Total ORFs	2300	1950	1744
Total CDSs	2215	1861	1659
dont pseudogènes	ND	247	92
Total ARNs	85	89	85
dont ARNt	67	67	67
ARNr	18	18	18
ARNnc	ND	4	0
Loci CRISPR	ND	3	0

L'intégration de *S. thermophilus* N4L au sein de GenBank a également permis de positionner

la souche dans l'arbre phylogénétique de l'espèce (Figure 2.1). La bactérie est située sur une branche isolée d'un clade portant son plus proche parent, la souche KLDS SM, avec qui elle présente environ 95,26 % d'identité de séquence. La souche N4L présente également une identité de séquence importante (93,89 %) avec la souche industrielle de référence LMD-9 située dans un clade voisin et pour laquelle une littérature conséquente est disponible. La souche la plus éloignée est la CNCM I-1630 (81,04 %), et 42 souches ont plus de 90 % d'identité avec la N4L, démontrant l'homogénéité génétique du groupe.

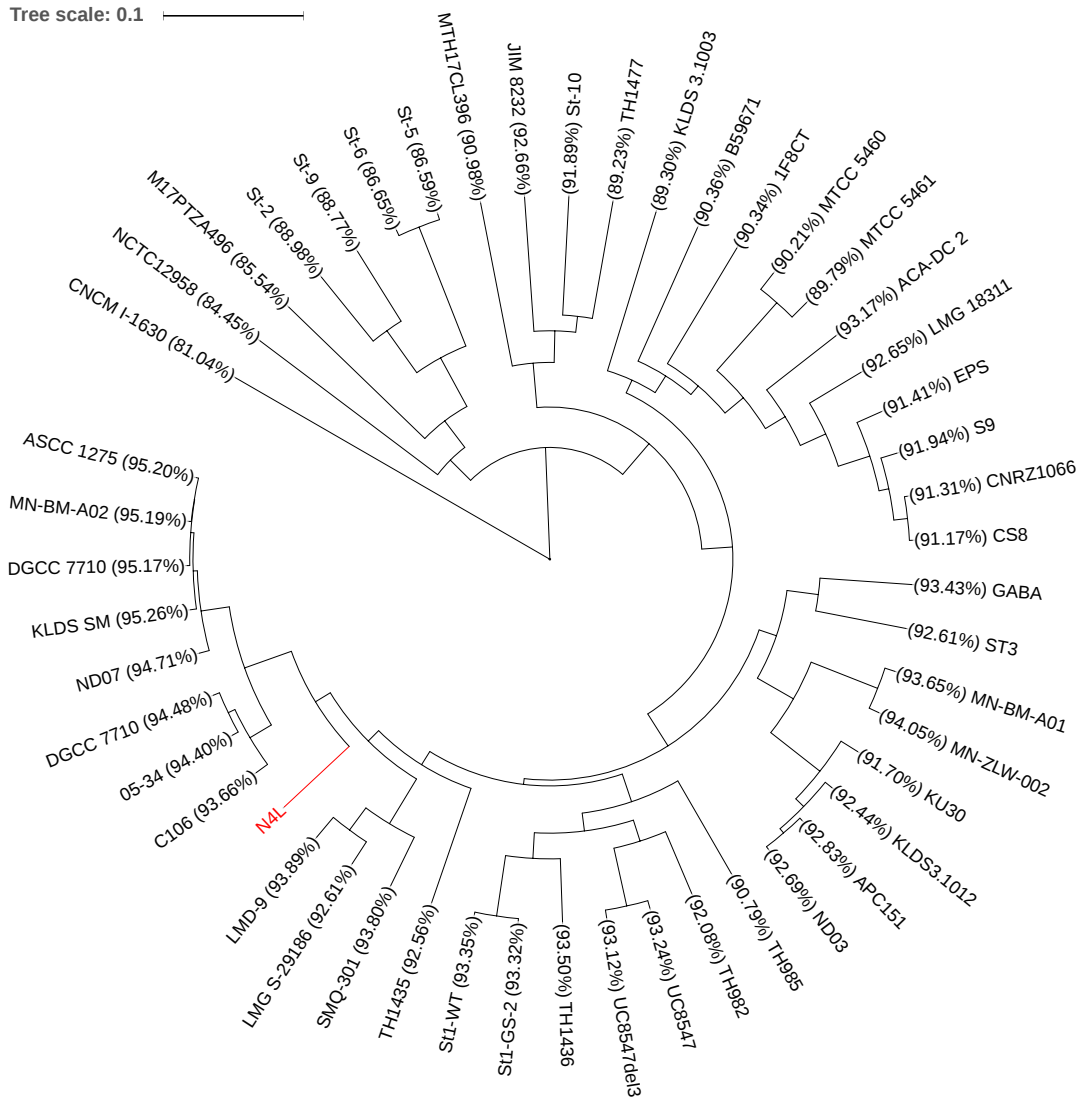


FIGURE 2.1 : Arbre phylogénétique des 50 souches de *S. thermophilus* disponibles dans la base de données GenBank du NCBI. Les souches ont été regroupées en fonction de leur similarité de séquence génomique déterminée par BLAST. Entre parenthèses est indiqué le pourcentage d'identité entre la séquence de la souche considérée et celle de la N4L, d'après les informations fournies par le NCBI. Arbre créé à l'aide de l'outil iTOL (Letunic et Bork, 2016).

L'annotation réalisée par le NCBI permet ainsi d'apporter des informations complémentaires par rapport à l'annotation initiale d'Agmial et tire profit de l'importante collection de données génomiques de l'institut pour positionner la souche au sein de son taxon. Néanmoins, en raison de son antériorité et par souci d'homogénéité, l'annotation d'Agmial est celle qui sera utilisée dans l'ensemble des résultats présentés dans la suite de ce manuscrit. Celle-ci a été enrichie fonctionnellement et structurellement afin d'approfondir la caractérisation du génome de *S. thermophilus* N4L.

2.3.2 Assignment d'information fonctionnelle de numéros de l'orthologie KEGG

Dans le but de mieux interpréter les résultats des analyses transcriptomiques et protéomiques qui seront présentés dans le Chapitre 4, il a été nécessaire de préalablement compléter l'annotation d'Agmial par une annotation fonctionnelle visant à assigner aux différents gènes prédits un vocabulaire hiérarchique contrôlé et standardisé. Différentes ontologies ont été développées dans ce but, la plus répandue étant la *Gene Ontology* (GO) qui permet de décrire un gène selon sa fonction moléculaire, son rôle biologique et la localisation cellulaire de son produit (Ashburner *et al.*, 2000). Bien que très performante, cette classification ne fait cependant pas de lien direct avec les différentes voies métaboliques (biosynthèse, signalisation, régulation etc.) ce qui est préjudiciable dans le contexte de cette étude.

L'ontologie proposée par la *Kyoto Encyclopedia of Genes and Genomes* (KEGG) permet précisément de pallier cette limitation (Mao *et al.*, 2005) et a donc été choisie pour décrire fonctionnellement le génome de *S. thermophilus* N4L. A cet effet, l'outil BlastKOALA (Kanehisa *et al.*, 2016) a été utilisé afin d'assigner des numéros « KO » (*KEGG Orthology*) aux différents gènes prédits par Agmial, permettant ainsi de les rattacher aux *pathways* et hiérarchies fonctionnelles décrits dans l'orthologie de KEGG. Brièvement, comme son nom l'indique, BlastKOALA est basé sur l'emploi de recherches BLAST permettant d'aligner les gènes fournis par l'utilisateur contre des groupes de gènes orthologues non-redondants appartenant à une même catégorie fonctionnelle. Lorsque l'homologie entre le gène et le groupe d'orthologues est détectée comme significative et qu'un certain score statistique est dépassé, le gène reçoit le numéro KO du groupe. Ici, ces groupes de gènes ont été limités à l'échelle du genre *streptococcus* uniquement, afin de réduire l'assignation potentielle de faux positifs.

Un total de 1023 gènes ont ainsi reçu un unique numéro KO, ce qui correspond à une annotation fonctionnelle de 46,3 % des CDSs prédites du génome (les gènes codant des ARNs n'ont pas été analysés). La répartition des principales catégories assignées est donnée dans la Figure 2.2. A titre d'illustration, le gène *pfk* (STN4L_00380) codant la 6-phosphofructokinase a reçu le numéro K00850 ce qui permet de l'affilier théoriquement jusqu'à 14 *pathways* dif-

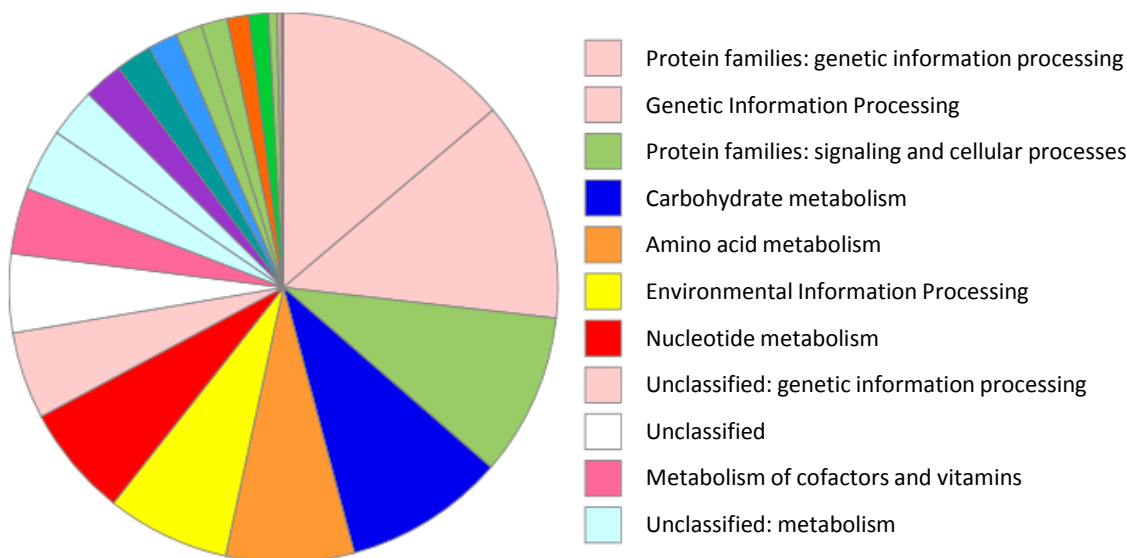


FIGURE 2.2 : Répartition des principales catégories fonctionnelles telles qu'identifiées par l'annotation KEGG du génome de *S. thermophilus* N4L.

férents, le premier étant évidemment la glycolyse. Ce type d'information sera utile ultérieurement lors de l'exploitation des résultats des approches omiques, en particulier afin de réaliser des analyses d'enrichissement sur les liste de gènes et protéines différemment exprimés entre les deux EXLs (voir Chapitre 4).

2.3.3 Prédiction de la localisation cellulaire des protéines bactériennes

De façon similaire, dans l'optique de mieux interpréter les résultats de protéomique, l'outil LocateP v2 (Zhou *et al.*, 2008) a été utilisé afin de prédire la localisation cellulaire possible des protéines identifiées par Agmial (Tableau 2.2). En l'absence de peptide signal dans les 60 premiers résidus côté N-terminal, ce qui représente ici plus de 76 % des cas (1686), les protéines sont considérées comme ayant une localisation cytoplasmique. Quelques protéines (16) possèdent un possible signal d'export peu après le 60ème résidu. Elles sont prédites comme étant également cytoplasmiques mais pourraient être exportées. Environ 20% des protéines sont localisées au niveau de la membrane, à travers soit de multiples domaines transmembranaires, soit par un ancrage de l'extrémité N-terminale. Enfin, 2% des protéines sont prédites comme étant sécrétées dans le milieu extracellulaire, et 6 sont rattachées à la paroi. Afin de conforter ces résultats, LocateP v2 a également été utilisé sur la version du génome annotée par le NCBI, ainsi que sur celui de la souche LMD-9. Dans les deux cas, des profils similaires ont été obtenus ne montrant que quelques variations marginales.

TABLEAU 2.2 : Localisation cellulaire des protéines de *S. thermophilus* N4L prédite par Loca-teP v2 (Zhou *et al.*, 2008).

Destination	Localisation subcellulaire	Nombre de protéines
Cytoplasmique	Intracellulaire	1686
	Intracellulaire/peptide signal après 60ème résidu	16
Membranaire	Transmembranaire	312
	Transmembranaire, extrémité N-terminale modifiée (lipidation)	6
	Ancrage N-terminal (pas de site de clivage)	125
	Ancrage N-terminal (avec site de clivage)	13
Pariétale	Ancrée à la paroi (motif LPxTG)	6
Extracellulaire	Ancrage lipidique	24
	Sécrétée (avec site de clivage)	21

2.3.4 Identification de courts ORFs

Le dernier travail de caractérisation du génome de la souche N4L a consisté à enrichir son annotation en courts ORFs (sORFs). Comme expliqué dans la Partie 1.2.3.1.1 de l'introduction bibliographique, les outils d'annotation actuels ont encore du mal à identifier efficacement les courtes séquences codantes dans les génomes bactériens. Ainsi, dans l'annotation du NCBI, les séquences codant des protéines de moins de 50 acides aminés ne représentent qu'environ 2 % du total identifié (39 CDSs). Dans l'annotation d'Agmial, ce chiffre augmente artificiellement à 13 % (292 CDSs) du fait de l'annotation des pseudogènes en de multiples fragments (Figure 2.3A). Mais dans les deux cas, l'annotation en courtes séquences codant des peptides d'intérêt physiologique reste limitée.

Afin de remédier à cela, la stratégie utilisée ici a consisté à identifier de façon exhaustive tous les ORFs putatifs du génome dont la taille est comprise entre 15 et 60 acides aminés et de confronter leur caractère codant par le biais des approches omiques au niveau transcriptionnel (RNAseq) et traductionnel (peptidomique). L'objectif final étant ici double. Il s'agit d'une part de tirer profit des approches globales pour éventuellement détecter de nouveaux systèmes de *quorum sensing* jusqu'alors non caractérisés, et d'autre part d'identifier des systèmes dont l'activité est différenciellement modulée (inhibition ou activation) par le contenu en peptides des deux EXLs du projet. Ces aspects seront développés ultérieurement dans les deux chapitres suivants. Afin d'identifier ces sORFs, le programme BactgeneSHOW (Nicolas *et al.*, 2002) a été utilisé en fixant le seuil d'exclusion à une valeur minimale (0.01) afin de retenir tous les ORFs situés dans la gamme de taille précisée. En effet, les valeurs de probabilité statistique habituellement utilisées dans l'annotation des gènes classiques n'ont qu'une portée limitée dans un contexte d'identification de sORFs. Des peptides prédits avec de faibles scores de probabilité statistique ont ainsi été expérimentalement détectés dans des surnageants de culture (Vincent Juillard, communication personnelle). Au total, 1016

sORFs ont été identifiés dans le génome de *S. thermophilus* N4L. Après élimination des séquences déjà identifiées par Agmial (320 cas) ou correspondant aux extrémités de gènes de plus longue taille (17), ce nombre a été finalement réduit à 679 sORFs (Figure 2.3B).

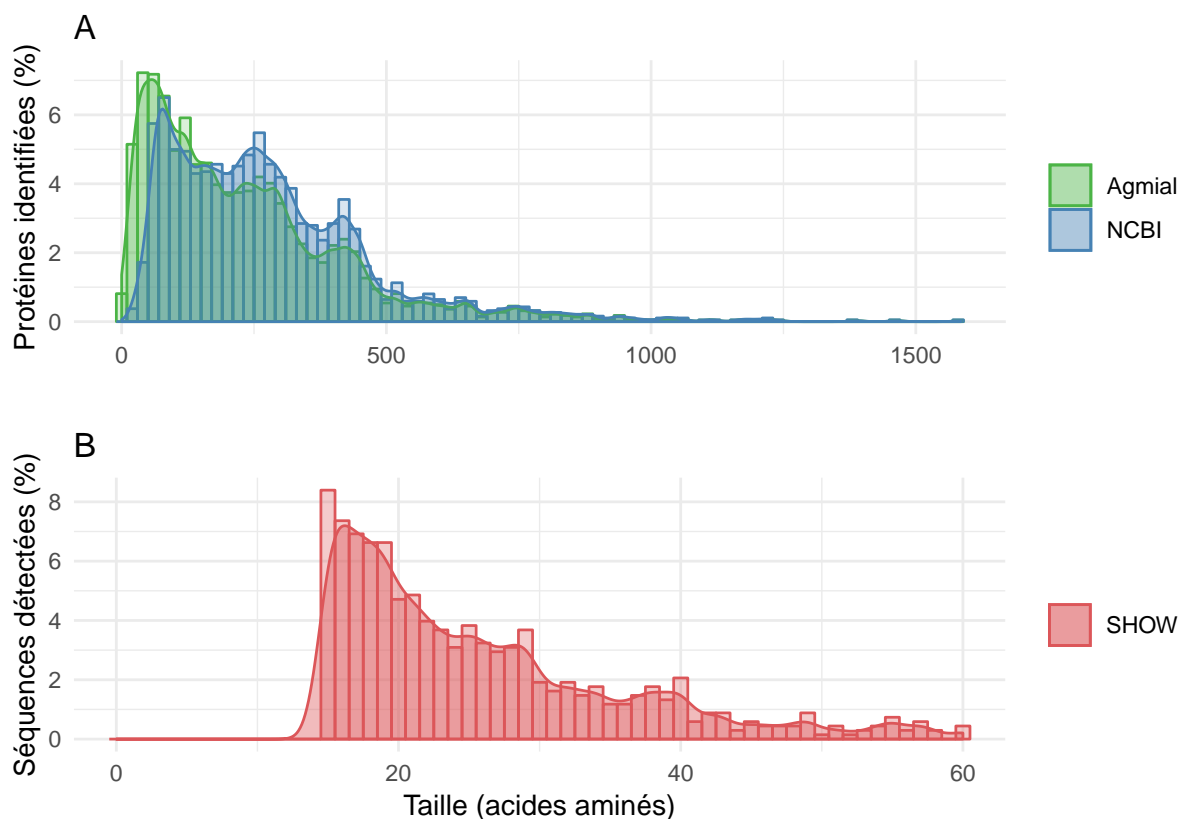


FIGURE 2.3 : Fréquence de distribution des tailles de protéines (A) et courts ORFs de 15 à 60 acides aminés (B) identifiés chez *S. thermophilus* N4L.

Du simple fait de sa taille importante, cette liste de sORFs contient naturellement de nombreux faux positifs non codants. Elle inclut également des fragments de pseudo-gènes et des séquences chevauchant de plus longs gènes. Toutefois, ce travail a permis d'identifier plusieurs phéromones impliquées dans les systèmes SHP/Rgg et ComRS dont deux qui n'avaient pas été retenues par Agmial et l'outil PGAP du NCBI (Tableau 2.3). Le couple SHP/Rgg (identifiants STN4L_short_044/STN4L_00156) ainsi identifié est homologue à celui du locus 1358 découvert dans la souche LMD-9 (Ibrahim *et al.*, 2007a; Fleuchot *et al.*, 2013) et assurant la régulation d'un gène cible codant un peptide modifié post-traductionnellement (couramment nommé « streptide ») dont la fonction biologique demeure encore aujourd'hui inconnue. La forme mature et sécrétée de ce peptide identifiée chez la LMD-9 est AKGDGWKVM. Celui-ci présente une structure cyclique en raison d'une liaison covalente entre les carbones des chaînes latérales de Lys2 et Try6 (Schramma *et al.*, 2015). Chez la souche N4L, le peptide équivalent (STN4L_00157) présente une mutation induisant la substitution de Lys7 par Glu suivi d'un codon stop (AKGDGWE). La question

de l'effet de cette mutation sur la maturation et la fonction de ce peptide mériterait d'être explorée.

TABLEAU 2.3 : Phéromones associées à des régulateurs Rgg/ComR identifiées dans le génome de *S. thermophilus* N4L.

Phéromone	Identifiant	Séquence	Position	Brin	Détection	Régulateur associé	Identifiant	Position	Brin
SHP	STN4L_short_044	MKKQILLLLLLVFEGHIVVVG	141562:141633	-	SHOW	Rgg	STN4L_00156	141724:142584	+
	STN4L_00649	MNKESFLAIIIIIFESIIIVAVG	526751:526822	+	Agmial		STN4L_00648	525807:526664	-
	STN4L_01808	MNKKALFSLLFVILEGHIVGVG	1456690:1456761	+	Agmial		STN4L_01807 ^a	1455739:1456437	-
	STN4L_02256	MKKQKLLLLLVVLCVCEGHIVLVG	1795496:1795567	+	Agmial		STN4L_02255	1794546:1795409	-
	STN4L_02276 ^b	MNISIKRFLMILLEIHHIAL	1810482:1810547	+	Agmial		STN4L_02275	1809489:1810361	-
ComS	STN4L_short_398	LKTLKIFVLFSLIPLPYFAGCL	1119094:1119168	-	SHOW	ComR	STN4L_01358	1119256:1120155	-

Les possibles formes matures des phéromones sont indiquées en rouge au sein des séquences des précurseurs.

^a Pseudogène.

^b Potentiel SHP non canonique.

La phéromone ComS également identifiée lors de ce travail de prédiction de sORFs appartient au système ComRS assurant la régulation de la compétence chez *S. thermophilus* qui est maintenant bien décrit dans la littérature (Fontaine *et al.*, 2010, 2015; Gardan *et al.*, 2013). D'autres systèmes SHP/Rgg potentiels ont également été identifiés (Tableau 2.3), mais n'ont pour l'instant pas été caractérisés.

Ces deux types de systèmes régulateurs SHP/Rgg et ComRS présentent la particularité d'associer au sein d'un même locus le régulateur transcriptionnel, sa phéromone en position directement adjacente, ainsi que leurs divers gènes cibles (cibles proximales). Au total, 80 protéines ont été annotées comme étant de potentiels régulateurs transcriptionnels dans le génome de la bactérie. Sur ce nombre, 39 présentent directement en position amont ou aval un ou plusieurs sORFs (dont les phéromones présentées dans le Tableau 2.3).

2.4 Conclusion

Ce travail de caractérisation du génome de *S. thermophilus* N4L a donc constitué un préalable nécessaire dans le cadre de ce projet de thèse faisant appel à des approches omiques. Globalement, cette souche possède des attributs dans la norme de son espèce. Elle présente en particulier une certaine proximité avec la souche LMD-9 qui est elle aussi une souche industrielle démontrant une haute capacité d'acidification en raison de la présence d'une protéase de paroi. En revanche, il convient de noter que la LMD-9 possède un système additionnel de transport de peptides, le système Ots (Jameh *et al.*, 2016), qui est absent chez la souche N4L. Celle-ci ne possède que les systèmes DtpT et Ami (possédant deux SBPs : AmiA1 et AmiA3) qui assurent son approvisionnement en peptides, thématique qui constitue le cœur du chapitre suivant.

Chapitre 3

Caractérisation des dynamiques d'utilisation de peptides d'extraits de levure par *Streptococcus thermophilus* N4L

Sommaire

3.1 Introduction	93
3.2 Approche quantitative : mesure de la consommation d'azote aminé par RP-HPLC	96
3.2.1 Introduction	96
3.2.2 Matériel et méthodes	96
3.2.2.1 Conditions de culture	96
3.2.2.2 Préparation des échantillons	97
3.2.2.3 Conditions d'analyse en RP-HPLC	98
3.2.2.4 Quantification des acides aminés	99
3.2.3 Résultats	100
3.2.3.1 Croissance bactérienne	100
3.2.3.2 Suivi des acides aminés libres	101
3.2.3.3 Suivi des peptides de faible poids moléculaire	103
3.2.3.4 Suivi des oligopeptides	104
3.2.3.5 Bilan de la consommation d'azote aminé	106
3.2.4 Discussion	107
3.3 Approche qualitative : analyse peptidomique par spectrométrie de masse .	110
3.3.1 Introduction	110

3.3.2	Evolution méthodologique pour la mise au point d'un procédé d'analyse peptidomique adapté aux EXLs	111
3.3.2.1	Elaboration du protocole de préparation des échantillons . . .	112
3.3.2.2	Estimation de la répétabilité technique	117
3.3.2.3	Mise en place des conditions de fermentation de <i>S. thermophilus</i> N4L	121
3.3.3	Manuscrit	121
3.3.4	Bilan	159
3.3.5	Résultats complémentaires	160
3.3.5.1	Modélisation sPLS-DA	160
3.3.5.1.1	Méthode	162
3.3.5.1.2	Résultats	164
3.3.5.1.3	Discussion	165
3.3.5.2	Amélioration du protocole expérimental pour une analyse peptidomique haut-débit	168
3.3.5.2.1	Stratégie d'amélioration du protocole	168
3.3.5.2.2	Estimation de la répétabilité technique du nouveau protocole	170
3.3.5.2.3	Bilan	172
3.4	Conclusion	173

3.1 Introduction

Cette partie a pour objectif de présenter le suivi de l'évolution du contenu en azote aminé - avec un accent marqué sur les peptides - réalisé lors des cultures de la souche *S. thermophilus* N4L en bioréacteurs dans des milieux basés sur deux EXLs industriels fournis par Procelys qui seront nommés ici EXL1 et EXL2. Ces EXLs ont été choisis au démarrage du projet de thèse sur la base i) d'un travail de *screening* de divers hydrolysats de levure ayant permis de valider l'excellente adéquation de *S. thermophilus* N4L avec ces deux substrats (données non présentées), et ii) de connaissances acquises à l'échelle industrielle où ont été observées des différences de performances technologiques (acidification dans le lait) vraisemblablement liées à une différence de composition entre ces deux EXLs. De plus, bien que tous les deux issus de l'hydrolyse de la levure *S. cerevisiae*, les EXLs choisis résultent de deux processus de fabrication très différents. Des écarts et spécificités en termes de composition peptidique sont ainsi attendus.

L'objectif de cette partie, qui a occupé la majeure partie de la thèse, est double. Il s'agit en effet de répondre aux deux questions suivantes : i) quels sont les peptides des deux EXLs? (vision statique), et ii) quels sont les peptides préférentiellement utilisés par la bactérie pendant sa croissance? (vision dynamique). La démarche expérimentale générale mise en œuvre pour répondre à ces questions est présentée dans la Figure 3.1. Celle-ci est organisée en trois parties. La première a consisté à traiter les surnageants de culture issus des moûts prélevés durant les fermentations afin d'en récolter les peptides. Pour cela, ces surnageants ont tout d'abord été soumis à différentes filtrations et ultrafiltrations successives afin de les débarrasser des composés de poids moléculaire supérieur à 3 kDa. Les peptides contenus dans le dernier ultrafiltrat ont ensuite été isolés par le biais d'une extraction en phase solide (SPE, Solid Phase Extraction). La phase polymérique utilisée permettant de retirer les composés hydrophiles de faible poids moléculaire (informations du fournisseur), la fraction éluée (FE) obtenue est ainsi majoritairement composée d'oligopeptides, tandis que les acides aminés et libres et courts peptides se retrouvent essentiellement dans la fraction non retenue (FNR) en compagnie des sucres, vitamines et autres petits composés hydrophiles (les composés fortement hydrophobes restent fixés à la phase solide).

A partir de ce processus de purification, deux approches analytiques complémentaires ont ensuite été déployées. La première approche, qui constitue véritablement le cœur du projet de thèse, a consisté à analyser par spectrométrie de masse le contenu peptidique des FEs récoltées au cours des fermentations. L'analyse en spectrométrie de masse de mélanges complexes de peptides est une technique relativement récente pour laquelle il n'existe pas encore d'outils « clé en main ». Elle a donc nécessité un important développement méthodologique, tant du point de vue expérimental que bioinformatique, qui sera développé dans ce chapitre. Néanmoins, cette approche souffre d'un certain nombre de limitations : elle n'est

en effet que semi-quantitative et ne permet pas d'analyser les peptides de faible poids moléculaire - en particulier les di- et tripeptides - qui peuvent pourtant constituer une fraction non négligeable des EXLs. Pour remédier à cela, une deuxième approche, complémentaire, a été déployée en cours de thèse. Celle-ci est basée sur le dosage des acides aminés libres et totaux par chromatographie liquide haute performance en phase inverse (RP-HPLC). Appliquée sur différents échantillons prélevés au cours du processus de purification avant et après hydrolyse acide, cette approche a permis de mesurer quantitativement l'évolution du contenu en acides aminés libres et acides aminés peptidiques en cours de fermentation et dans les deux EXLs. Bien que non qualitative (l'information sur la séquence des peptides est perdue), elle permet de réaliser un véritable suivi quantitatif de l'ensemble du contenu en azote aminé pendant la fermentation.

Ce chapitre détaillera donc les résultats obtenus à l'issue de ces deux approches expérimentales. Le dosage du contenu azoté par RP-HPLC sera tout d'abord abordé du fait de son périmètre plus large. Cela permettra ainsi de mieux introduire les analyses peptidomiques réalisées qui seront détaillées en deuxième partie.

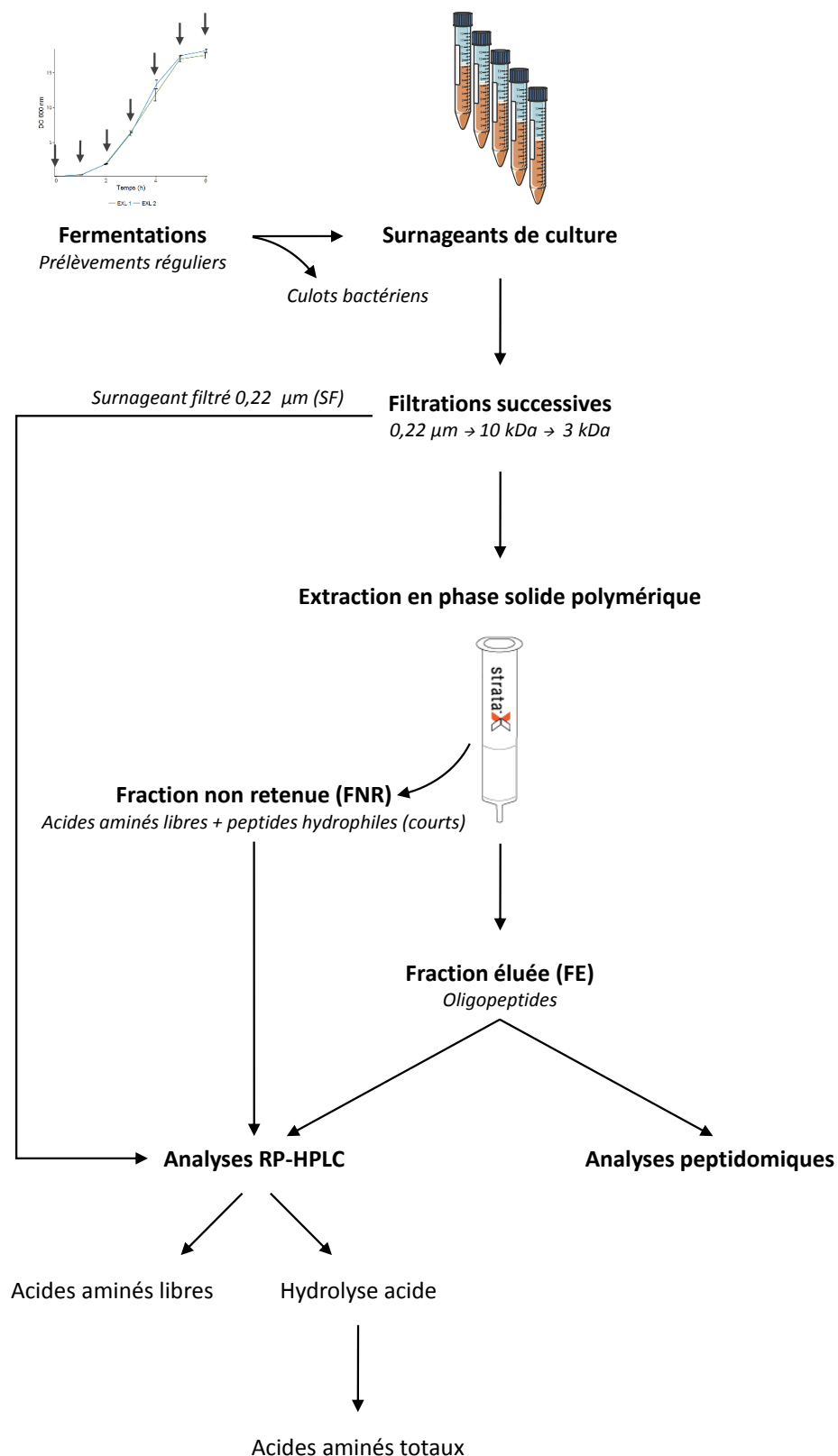


FIGURE 3.1 : Démarche expérimentale mise en œuvre afin d'effectuer un suivi du contenu peptidique lors de la croissance de *S. thermophilus* N4L.

3.2 Approche quantitative : mesure de la consommation d'azote aminé par RP-HPLC

3.2.1 Introduction

L'objectif du travail présenté dans cette partie est fournir une estimation des concentrations respectives des différentes formes d'azote aminé présentes dans les EXLs (acides aminés libres, courts peptides, oligopeptides) ainsi que leur évolution au cours de la croissance bactérienne. La question biologique sous-jacente est de déterminer leur importance nutritionnelle respective pour la croissance bactérienne. Pour cela, des échantillons ont été prélevés sur trois différents niveaux du processus de purification : surnageant filtré à 0,22 μm (SF) contenant l'ensemble des trois formes d'azote précitées, fraction non retenue (FNR) qui inclut la majorité des courts peptides et acides aminés libres, et fraction éluée (FE) qui contient l'essentiel des oligopeptides d'une masse moléculaire inférieure à 3 kDa. Le contenu en acides aminés a été mesuré avant et après hydrolyse acide, permettant de reconstituer par simple jeu de différences la teneur en acides aminés des différentes formes d'azote étudiées.

3.2.2 Matériel et méthodes

3.2.2.1 Conditions de culture

La souche *S. thermophilus* N4L, conservée à $-80\text{ }^{\circ}\text{C}$ en milieu M17 (Terzaghi et Sandine, 1975) supplémenté avec 1 % de lactose (m/v) et 20 % de glycérol (v/v), a été successivement pré-cultivée deux fois à $42\text{ }^{\circ}\text{C}$ en M17 contenant 5 % de lactose avant d'ensemencer (2 %, v/v) des cultures en bioréacteurs. Celles-ci ont été effectuées dans des fermenteurs d'un litre modèle BIOSTAT Qplus (Sartorius Stedim Biotech, Allemagne). Les milieux de culture étaient composés à 6 % de lactose (m/v), 0,01 % de CaCl_2 , 2 % d'EXL1 ou EXL2 (gamme NuCel, Procelys, France) et contenaient un mélange de vitamines tel que défini par Letort et Juillard (2001). Ces vitamines ont été ajoutées afin d'assurer des performances de croissance optimales et répétables, similaires à celles obtenues dans un milieu équivalent contenant du Tween 80 (données non présentées). Le Tween 80 est un facteur de croissance largement utilisé pour la culture des bactéries lactiques (Williams *et al.*, 1947). Cependant, il s'agit également d'un composé hautement ionisable connu pour perturber les analyses en spectrométrie de masse (Jäpelt *et al.*, 2016). Il n'a donc pas été utilisé durant l'ensemble du projet. Les EXLs ont été stérilisés séparément du reste du milieu directement dans les bioréacteurs pendant 20 min à $120\text{ }^{\circ}\text{C}$. Les autres constituants du milieu ont été stérilisés séparément par filtration à 0,22 μm à l'aide d'une membrane de polyéthersulfone (Millipore, Guyancourt, France), puis mélangés stérilement avec les EXLs dans les bioréacteurs. Le pH initial a été

ajusté avant inoculation à 6,6 par ajout de NaOH 2 M. Les fermentations ont été conduites durant 6 h à 40 °C sous une agitation fixée à 50 rpm et une régulation du pH à 6,0 (NaOH 2 M). Une pseudo anaérobiose a été établie par un barbotage préalable avant inoculation de diazote à 0,2 l/min pendant 1 h, puis durant la fermentation dans l'espace de tête au même débit. La croissance a été mesurée par densité optique (DO) à 600 nm et dénombrement sur boîtes en M17 + lactose 1 %. Des échantillons ont été prélevés chaque heure, en commençant par le milieu avant ensemencement ($t = 0$ h). Contrairement aux analyses peptidomiques qui ont été réalisées à partir de cultures conduites en triplicats, ici, le dosage des acides aminés libres et totaux n'a été effectué que sur une seule culture supplémentaire.

3.2.2.2 Préparation des échantillons

Les échantillons récoltés en cours de fermentation ont tout d'abord été centrifugés (3000 g, 10 min, 4 °C) afin de retirer la biomasse bactérienne. Les surnageants obtenus ont ensuite été mélangés avec de l'acide trifluoroacétique (TFA) et de l'acétonitrile (ACN) à des concentrations finales respectives de 0,1 % et 5 % (v/v). Ces mélanges ont ensuite été filtrés à 0,22 μ m par une membrane PVDF à faible rétention protéique (Millipore, Guyancourt, France). A cette étape, des aliquots (SF) ont été prélevés. Le reste de ces surnageants a ensuite été successivement ultrafiltré à 10 kDa et 3 kDa par des systèmes Amicon Ultra-15 *ad hoc* (Millipore, Guyancourt, France). Le dernier perméat a finalement subi une SPE basée sur l'emploi d'une cartouche StrataX de 200 mg (Phenomenex, Le Pecq, France). Celle-ci, après avoir été préalablement activée et équilibrée respectivement par du méthanol et du TFA à 0,1 % (v/v), a été chargée avec 4 ml d'échantillon. A cette étape la FNR a été récupérée et aliquotée. La cartouche a ensuite été lavée par 12 ml de TFA 0,1 % et ACN 5 % (v/v). Les peptides retenus ont finalement été élués par 1,5 ml de TFA 0,1 % et ACN 50 %. La FE a de même été récupérée et aliquotée. A noter que des détails concernant l'élaboration de ce procédé de préparation des échantillons seront fournis dans la partie 3.3 suivante.

En vue de la quantification de leur contenu en acides aminés libres, les trois échantillons obtenus (SF, FNR, FE) ont été préparés de la façon suivante. Les protéines et peptides ont tout d'abord été précipités par ajout de 50 μ l d'acide sulfosalicylique dans 450 μ l d'échantillon. Le mélange a été stocké durant une nuit à 4 °C et centrifugé (10 000 rpm, 10 min, 4 °C) le lendemain. Le surnageant a ensuite été dilué (1/50) dans un solution tampon constituée d'acide borique 0,2 M et de NaOH 0,2 M en ratio 1 :1 (v/v) afin d'ajuster le pH à environ 9,9. Ce pH basique est nécessaire pour une dérivation optimale des acides aminés (voir plus bas). Le dosage des acides aminés totaux a quant à lui nécessité une hydrolyse préalable. Pour cela, 100 μ l d'échantillon (SF, FNR, FE) ont été séchés toute une nuit par évaporation sous vide dans un système Speed-Vac (Savant, Thermo Scientific, US). Les échantillons séchés ont ensuite été hydrolysés sous vide en présence de HCl 6 N pendant 16 h à 110 °C. Passée cette

étape, les échantillons ont de nouveau été séchés pendant 2 h par évaporation sous vide afin de retirer toute trace d'acide, puis resuspendus dans 1 ml d'une solution neutralisante à 100 g/l de NaOH, 74 g/l de KCl et 61 g/l d'acide borique. Enfin, ils ont été dilués (1/5) dans la solution tampon décrite précédemment.

3.2.2.3 Conditions d'analyse en RP-HPLC

Une méthode de quantification par RP-HPLC a été spécifiquement développée afin d'analyser le contenu en acides aminés des différents échantillons préparés. Celle-ci est basée sur la dérivation pré-colonne des acides aminés à l'*o*-phtalaldéhyde (OPA) pour le dosage de l'ensemble des résidus hormis la proline et la cystéine qui ont, eux, été quantifiés *via* une dérivation au 9-fluorénylméthoxycarbonyle (FMOC). La procédure de quantification par l'OPA a été mise en œuvre de la façon suivante. Les échantillons ont tout d'abord été mis en contact avec le réactif pendant 7 min (ratio 1 :1, v/v). Les acides aminés ainsi dérivés ont ensuite été séparés sur une colonne Kinetex C18 Core-Shell (granulométrie : 2,6 μm ; taille de pore : 100 Å; dimensions : 150 x 4,6 mm; Phenomenex, Le Pecq, France). Un gradient de méthanol multiphasique a été appliqué afin de séparer efficacement les différents acides aminés. Le solvant A était constitué de tétrahydrofurane 3 % (THF, v/v) en solution dans de l'acétate de sodium 50 mM (pH 6,4). Le solvant B était composé de THF dilué à 5 % (v/v) dans du méthanol. Le profil du gradient était le suivant :

- 0-10 min : 100 % A
- 10-18 min : 0 à 35 % B
- 18-28 min : 35 à 38 % B
- 28-50 min : 38 à 60 % B
- 50-55 min : 60 à 100 % B
- 55-60 min : 100 % B
- 60-65 min : 100 % B à 100 % A
- 65-80 min : 100 % A

Les acides aminés ont été détectés durant ce gradient par une double détection UV (338 et 280 nm) couplée à une mesure fluorimétrique ($\lambda_{\text{Ex}} = 340 \text{ nm}$, $\lambda_{\text{Em}} = 425 \text{ nm}$).

La quantification de la proline et cystéine par dérivation au FMOC a tout d'abord nécessité une mise en contact entre réactif et échantillon pendant 7 min avec un ratio échantillon/FMOC de 2 :1 (v/v). La séparation chromatographique a été réalisée dans des conditions similaires à celles décrites précédemment mais avec un gradient simplifié :

- 0-5 min : 30 % B
- 5-50 min : 30 à 75 % B

Seule la double détection UV à 262 et 380 nm a été utilisée afin de détecter les acides aminés, la détection fluorimétrique étant hors gamme.

3.2.2.4 Quantification des acides aminés

Deux méthodes de quantification différentes ont été appliquées suivant la nature de l'échantillon analysé (Tableau 3.1). Le dosage par dérivation à l'OPA des acides aminés libres (sans hydrolyse) de l'EXL1 a requis une méthodologie complexe mêlant standardisation interne et externe. Un total de 19 injections a ainsi été nécessaire par point de prélèvement afin d'analyser l'échantillon pur puis complétement tour à tour avec un standard interne de chaque acide aminé (à l'exception de Pro et Cys). En parallèle, une standardisation externe a été réalisée en injectant des quantités connues de standards purs des différents acides aminés. Cette double quantification a permis d'une part de s'assurer de la correcte identification des temps de rétention des acides aminés dans les échantillons, mais surtout de valider la méthode par standardisation externe. Celle-ci a ensuite été appliquée seule aux échantillons SF dérivés au FMOC de l'EXL1 ainsi qu'à tous les échantillons de l'EXL2 quelle que soit la méthode de dérivation. En revanche, malgré de nombreuses tentatives, il n'a pas été possible de quantifier correctement la proline et cystéine par dérivation au FMOC dans l'ensemble des échantillons hydrolysés (détection impossible). Ces deux acides aminés n'ont donc été dosés que sous forme libre. De même, les acides aminés totaux issus de l'hydrolyse des différentes fractions éluées n'ont pu être quantifiés, tant avec la dérivation à l'OPA qu'au FMOC. L'origine la plus probable de ces problèmes techniques provient certainement de la difficulté à ajuster correctement le pH après hydrolyse jusqu'à une valeur proche de 9, ce qui limite fortement l'efficacité de la dérivation. Y parvenir aurait nécessité une dilution plus importante de l'échantillon, dont la concentration en acides aminés serait alors devenue trop faible pour être détectable.

TABLEAU 3.1 : Plan expérimental mis en œuvre pour la quantification du contenu azoté des EXLs.

Echantillon	EXL1				EXL2			
	AA libres		AA totaux		AA libres		AA totaux	
	OPA	FMOC	OPA	FMOC	OPA	FMOC	OPA	FMOC
SF	Int/Ext	Ext	Int/Ext	X	Ext	Ext	Ext	X
FNR	Int/Ext	ND	Int/Ext	X	Ext	ND	Ext	X
FE	Int/Ext	ND	X	X	Ext	ND	X	X

Int/Ext : standardisation interne et externe; **Ext** : standardisation externe uniquement; **ND** : non déterminé; **X** : quantification impossible

En dépit de ces limitations, les différentes analyses réalisées ont permis d'estimer la composition en acides aminés libres, courts peptides et oligopeptides de chaque échantillon. Ces valeurs ont été calculées de la façon suivante :

$$\text{Acides aminés libres} = \text{SF}_{\text{AA libres}}$$

$$\text{Courts peptides} = \text{FNR}_{\text{AA totaux}} - \text{FNR}_{\text{AA libres}}$$

$$\text{Oligopeptides} = \text{SF}_{\text{AA totaux}} - \text{FNR}_{\text{AA totaux}}$$

3.2.3 Résultats

3.2.3.1 Croissance bactérienne

Le suivi de la biomasse bactérienne en cours de fermentation a été réalisé par mesure de la $\text{DO}_{600\text{nm}}$ et par dénombrement sur boîtes (Figure 3.2). Les résultats des mesures de DO montrent des niveaux de performances très similaires entre les deux EXLs (Figure 3.2A). Les taux de croissance calculés sur la partie linéaire (entre 1 h et 3 h) de la courbe $\ln(\text{DO}) = f(t)$ atteignent respectivement $1,5 \text{ h}^{-1}$ pour l'EXL1 et $1,4 \text{ h}^{-1}$ pour l'EXL2, ce qui correspond à des temps de génération de 28 min et 30 min. Les résultats de suivi du dénombrement sont également très proches, bien qu'un léger gain de croissance semble s'opérer en faveur de la culture dans l'EXL2 où la biomasse y est quasiment deux fois plus abondante à 3h que dans l'EXL1 (environ $3,0 \cdot 10^9$ UFC/ml contre $1,7 \cdot 10^9$ UFC/ml, Figure 3.2B). Ceci se traduit par des taux de croissance, calculés sur la même période de temps que pour la DO, d'environ $1,7 \text{ h}^{-1}$ pour l'EXL1 et $1,9 \text{ h}^{-1}$ pour l'EXL2, soit des temps génération de 22 min et 25 min. Ces écarts sont par ailleurs reproductibles car également observés de façon systématique lors des cultures ultérieures réalisées en triplicats.

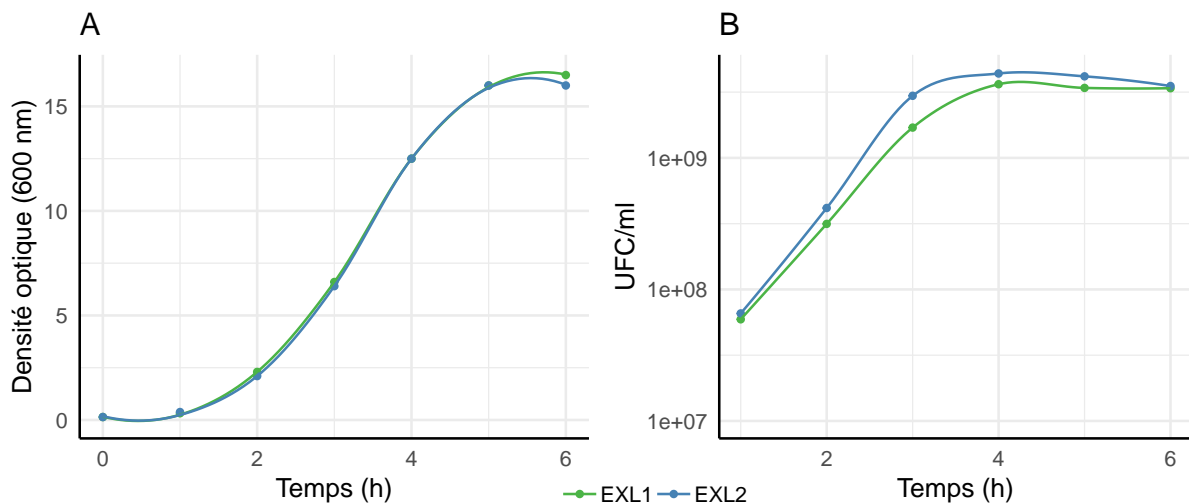


FIGURE 3.2 : Cinétique de croissance de *S. thermophilus* N4L mesurée par densité optique à 600 nm (A) et dénombrement sur boîtes (B) pendant sa culture en bioréacteurs.

3.2.3.2 Suivi des acides aminés libres

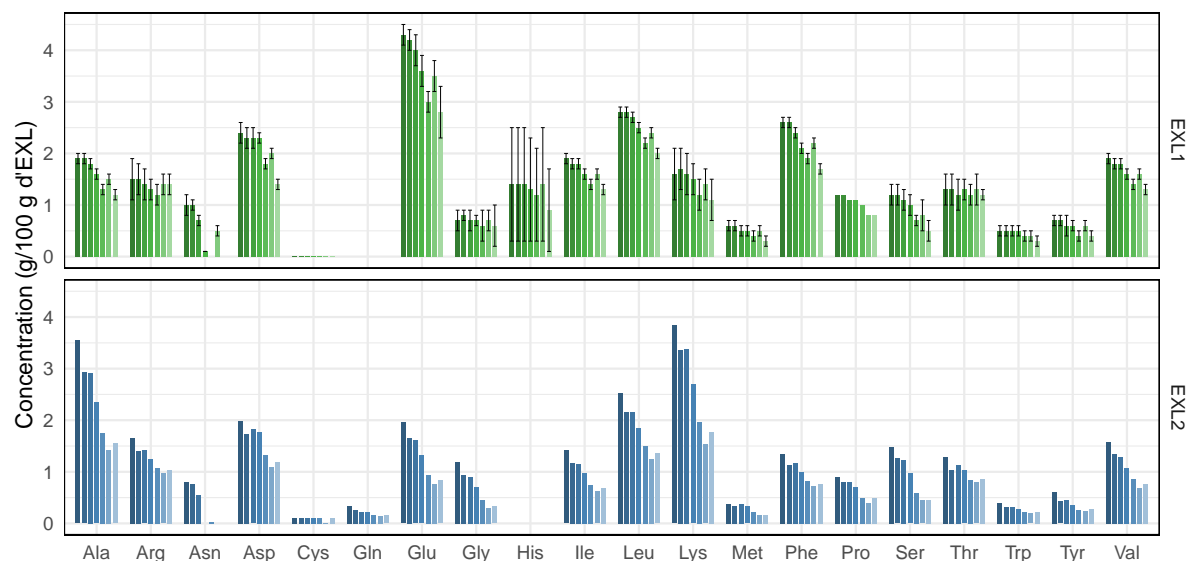


FIGURE 3.3 : Evolution au cours du temps du contenu en acides aminés libres dans les EXLs pendant la croissance de *S. thermophilus* N4L. Chaque colonne représente la concentration d'un acide aminé donnée mesurée toutes les heures, de 0 h (milieu avant inoculation) jusqu'à 6 h de culture. Les valeurs représentées correspondent à la détection par UV à 338 nm, la détection par fluorescence ayant donné des résultats très proches, et l'UV à 280 nm ayant principalement servi à confirmer la bonne assignation des pics de Trp et Tyr.

Le détail du contenu en acides aminés libres dans les deux EXLs est présenté dans la Figure 3.3. Les deux substrats présentent une quantité totale d'acides aminés libres relativement proche, avec respectivement 29,5 g d'acides aminés pour 100 g d'EXL1 et 27,3 g pour 100 g d'EXL2. Le glutamate est l'acide aminé le plus abondant de l'EXL1 (4,3 g/100 g) alors que la lysine est majoritaire (3,8 g/100 g) dans l'EXL2 suivie de près par l'alanine (3,5 g/100 g). Dans les deux substrats, les acides aminés soufrés (cystéine, méthionine), la glutamine, le tryptophane et la tyrosine sont minoritaires, la cystéine et glutamine n'ayant d'ailleurs pu être détectées dans l'EXL1. De façon plus suprenante, bien que soumise à une importante variabilité de mesure, l'histidine est présente dans des quantités appréciables dans l'EXL1 (1,4 g/100 g) alors qu'elle n'a pas été détectée dans l'EXL2. La principale différence de composition en acides aminés libres entre les deux EXLs réside ainsi dans leur composition en composés chargés. Les acides aminés basiques (lysine, arginine, histidine) représentent environ 15 % du total dans l'EXL1 contre 20 % de l'EXL2, et, de façon similaire, les acides (glutamate, aspartate) sont présents à hauteur de 22 % et 14 % du total. De même, en raison d'une plus forte abondance d'alanine dans l'EXL2, le contenu en acides aminés hydrophobes s'élève à 33 % du total dans ce substrat contre 28 % dans l'EXL1, les acides aminés branchés (leucine, isoleucine, valine) étant présents dans des concentrations similaires.

Une diminution de la concentration de la plupart des acides aminés est visible dans la Figure 3.3. Celle-ci correspond en réalité au bilan entre acides aminés importés d'un côté et exportés de l'autre par la bactérie à un temps donné. La méthodologie employée ici ne permettant pas de discriminer ces deux flux, on parlera donc de « consommation apparente ». La Figure 3.4A révèle que celle-ci est plus intense dans l'EXL2 que dans l'EXL1. La concentration totale d'acides aminés libres présents dans le milieu extérieur a en effet diminué d'environ 35 % dans l'EXL1 alors qu'elle a chuté de près de 50 % dans l'EXL2 entre le début et la fin de la fermentation. Ces valeurs se traduisent en des consommations apparentes d'acides aminés libres d'environ 1,7 g/h dans l'EXL1 et 2,7 g/h dans l'EXL2. La Figure 3.4B montre que la bactérie ne consomme également pas toutes les classes d'acides aminés de la même façon. Les acides aminés chargés positivement sont en effet consommés plus rapidement dans l'EXL2 que dans l'EXL1 (0,5 g/h contre 0,2 g/h). Le constat est similaire pour les acides aminés hydrophobes (0,8 g/h contre 0,5 g/h) et les acides aminés « autres » (0,7 g/h contre 0,4 g/h). Seuls les acides aminés chargés négativement et les aromatiques sont utilisés à des taux similaires (respectivement 0,4 g/h et 0,2 g/h).

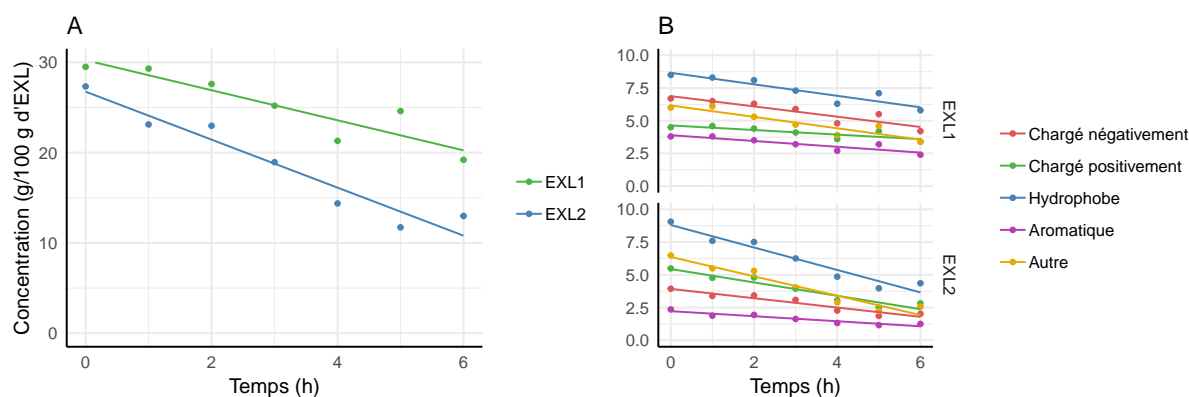


FIGURE 3.4 : Evolution de la consommation totale (A) et par classe biochimique (B) des acides aminés libres pendant la croissance de *S. thermophilus* N4L. Chargé négativement : Asp, Glu; chargé positivement : Arg, Lys, His; hydrophobe : Ala, Ile, Leu, Val; aromatique : Phe, Trp, Tyr; autre : Asn, Cys, Gln, Gly, Met, Pro, Ser, Thr.

Enfin, le dosage des acides aminés libres dans les SF et les FNR a permis d'évaluer l'efficacité du protocole de préparation des échantillons par SPE. En moyenne, plus de 90 % du total des acides aminés libres quantifiés dans les SF de l'EXL1 se retrouvent dans les FNR associées, alors qu'aucun acide aminé libre n'a pas été détecté dans les FE. Cette même valeur est d'environ 75 % dans l'EXL2, indiquant des pertes légèrement supérieures, bien que ce nombre se situe dans la marge d'erreur des mesures. Quoiqu'il en soit, ces données confirment que les acides aminés libres ne sont majoritairement pas retenus par la SPE, et auraient probablement été retrouvés dans les lavages, si cette fraction avait été analysée.

3.2.3.3 Suivi des peptides de faible poids moléculaire

Le contenu total en courts peptides est relativement similaire entre les deux EXLs. Il représente environ 14 g/100 g dans l'EXL1 et 12 g/100 g dans l'EXL2. Ici aussi des disparités sont à noter lorsque les acides aminés constitutifs des courts peptides sont regardés individuellement (Figure 3.5). Si dans les deux substrats le glutamate (incluant la glutamine convertie en glutamate durant le procédé d'hydrolyse) est majoritaire, la lysine est le deuxième acide aminé le plus abondant de l'EXL1 (2 g/100 g) alors qu'aucun résidu basique (à l'exception de l'histidine) n'a pu être détecté dans l'EXL2. A l'exception de la valine, les acides aminés hydrophobes ainsi que les aromatiques sont peu voire pas représentés dans les deux EXLs.

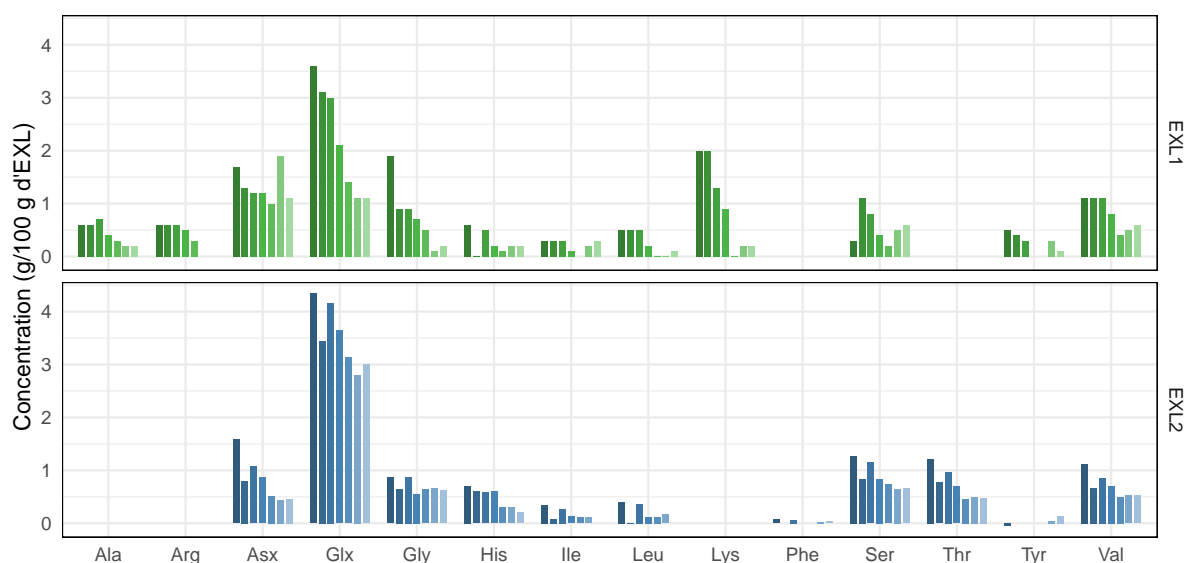


FIGURE 3.5 : Evolution de la composition en acides aminés des peptides de courte taille. Les valeurs correspondent à la différence de concentration entre acides aminés totaux après hydrolyse acide et acides aminés libres de la FNR. Asx = Asp + Asn, Glx = Glu + Gln. Cys et Trp étant entièrement dégradés durant l'hydrolyse acide et Met partiellement détruite, ces composés n'ont pas pu être mesurés. De même, Pro n'a pu être quantifiée en raison de la mauvaise qualité de la dérivation au FMOC dans les échantillons hydrolysés.

La Figure 3.6A indique que les petits peptides sont davantage consommés dans l'EXL1, avec une vitesse moyenne de disparition estimée à environ 1,7 g/h contre 0,9 g/h pour l'EXL2. Cette différence est principalement le résultat d'une plus forte consommation de courts peptides contenant des résidus basiques (Figure 3.6B) qui sont consommés à hauteur de 0,5 g/h dans l'EXL1 contre moins de 0,1 g/h dans l'EXL2. De même, les résidus hydrophobes disparaissent deux fois plus vite dans l'EXL1 que dans l'EXL2. Les résidus acides et « autres » disparaissent à des vitesses proches. Enfin, en raison de leur faible abondance, aucune consommation n'est observable au niveau des résidus aromatiques.

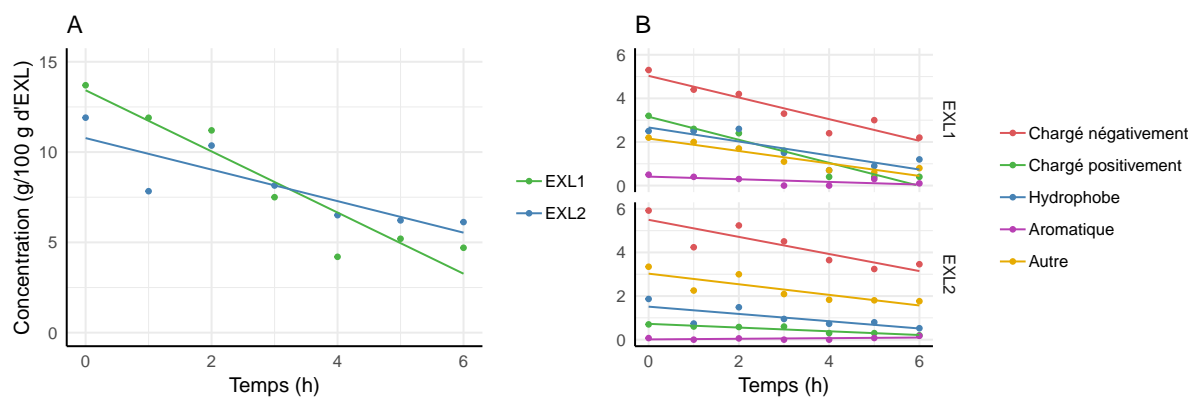


FIGURE 3.6 : Evolution de la consommation totale (A) et par classe biochimique (B) des acides aminés constituant les peptides de courte taille. Chargé négativement : Asx, Glx; chargé positivement : Arg, Lys, His; hydrophobe : Ala, Ile, Leu, Val; aromatique : Phe, Tyr; autre : Gly, Ser, Thr.

3.2.3.4 Suivi des oligopeptides

Si la composition globale en acides aminés libres et courts peptides est relativement proche entre les deux EXLs, ce n'est pas le cas de la fraction oligopeptidique (Figure 3.7). L'EXL2 contient en effet approximativement deux fois plus d'oligopeptides que l'EXL1 (31 g/100 g contre 14 g/100 g). Cette plus forte valeur est homogène sur l'ensemble des classes biochimiques qui montrent une répartition similaire entre les deux EXLs. Les acides aminés acides représentent en effet respectivement 37 % et 32 % du total de l'EXL1 et EXL2, les basiques 18 % dans les deux substrats, les hydrophobes 27 % et 24 %, les aromatiques 6 % et 8%, et enfin les autres résidus 11 % et 18 %. De façon similaires aux courts peptides, les oligopeptides sont majoritairement composés de glutamate (+ glutamine). Enfin, l'arginine, l'histidine et la tyrosine sont les résidus minoritaires.

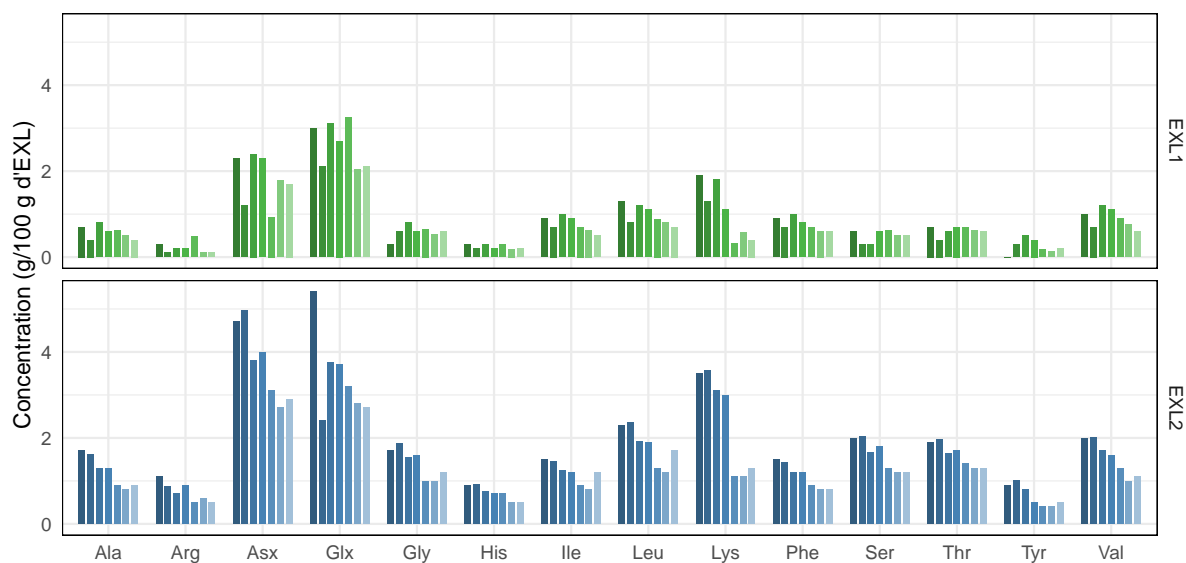


FIGURE 3.7 : Evolution de la composition en acides aminés des oligopeptides. Les valeurs correspondent à la différence de concentration entre acides aminés totaux après hydrolyse acide des échantillons SF et FNR.

Les profils de consommation des oligopeptides sont en revanche nettement différents entre les deux EXLs (Figure 3.8A). La vitesse de disparition des oligopeptides dans l'EXL2 est en effet près de 4 fois supérieure à celle observée dans l'EXL1 (2,6 g/h contre 0,7 g/h). Il est en particulier frappant de constater que les acides aminés aromatiques et « autres » ne diminuent pas au cours du temps (vitesses nulles) dans l'EXL1, alors qu'ils sont significativement consommés dans l'EXL2 (vitesses respectives de 0.2 g/h et 0.4 g/h). En revanche, dans les deux EXLs, les acides aminés chargés et hydrophobes sont ceux qui sont consommés le plus intensément. Toutefois, on peut noter que si les acides aminés chargés basiques et acides sont consommés à des vitesses égales dans l'EXL2 (0,7 g/h), dans l'EXL1, les basiques sont consommés près de deux fois plus vite que les chargés négativement (0,3 g/h contre 0,17 g/h respectivement). Les chargés positivement sont d'ailleurs quasiment épuisés dans l'EXL1 à 6 h de culture (0,7 g/100 g).

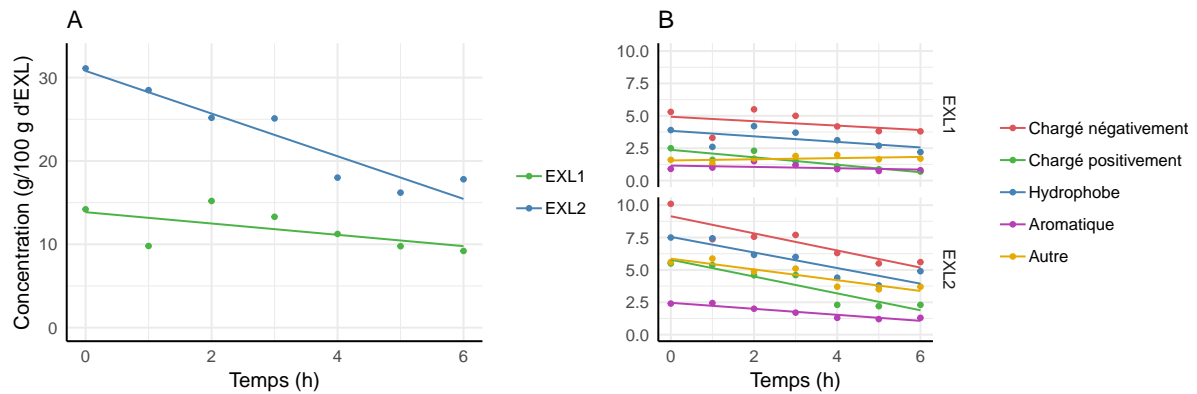


FIGURE 3.8 : Evolution de la consommation totale (A) et par classe biochimique (B) des acides aminés constituant les oligopeptides. Chargé négativement : Asx, Glx; chargé positivement : Arg, Lys, His; hydrophobe : Ala, Ile, Leu, Val; aromatique : Phe, Tyr; autre : Gly, Ser, Thr.

3.2.3.5 Bilan de la consommation d'azote aminé

La consommation globale d'azote aminé sous ses différentes formes est présentée dans la Figure 3.9, qui résume les différentes informations présentées précédemment. L'EXL2 contient ainsi environ 23 % d'azote aminé en plus par rapport à l'EXL1 (70,3 g/100 g contre 57,3 g/100 g). Cette différence est le résultat d'une surabondance des oligopeptides dans ce substrat, les acides aminés libres et courts peptides étant présents dans des quantités similaires. En fin de culture, la quantité résiduelle d'azote est équivalente dans les deux milieux, avec 33 g/100 g d'azote total dans l'EXL1 et 37 g/100 g dans l'EXL2. Les dynamiques ayant abouti à ces concentrations finales sont en revanche très différentes entre les deux EXLs.

La consommation d'azote est en effet plus importante et rapide dans l'EXL2 où la souche consomme environ 6 g d'azote total par heure, alors que cette vitesse n'est que de 4 g/h dans l'EXL1. Dans cet EXL, les acides aminés libres et courts peptides sont consommés à égale vitesse (1,7 g/h) alors que les oligopeptides ne disparaissent qu'à un rythme de 0,7 g/h. La situation est différente dans l'EXL2 où ce sont les oligopeptides et acides aminés libres qui sont consommés sensiblement de la même façon (respectivement 2,6 et 2,7 g/h) alors que les courts peptides sont moins sollicités (0,9 g/h).

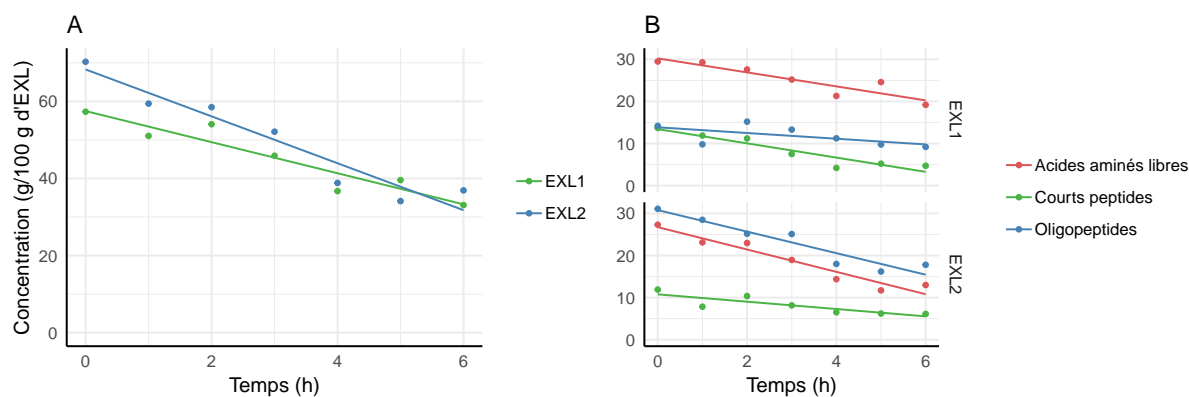


FIGURE 3.9 : Evolution du contenu d'azote aminé total (A) et par origine (B) pendant la croissance de *S. thermophilus* N4L.

3.2.4 Discussion

Cette démarche de dosage des acides aminés libres et constitutifs des peptides a donc permis i) de caractériser en profondeur le contenu global en azote aminé des deux EXLs du projet, et ii) d'identifier les dynamiques de consommation propres à chaque catégorie d'azote par *S. thermophilus* N4L durant sa culture en bioréacteurs. Afin de s'assurer de la validité des résultats, un soin particulier a été apporté sur la mise en place des procédures expérimentales employées. Une méthodologie spécifique et adaptée aux substrats de travail a ainsi été mise en place. La quantification des acides aminés libres dans les différents échantillons non hydrolysés n'a pas posé de problème majeur et la validité des résultats présentés a été confirmée par de multiples procédures analytiques complémentaires (standardisation interne et externe, couplage de la détection en UV et fluorimétrie). Le dosage des acides aminés totaux après hydrolyse en revanche s'est révélé être plus délicat principalement en raison de la difficulté à correctement ajuster le pH des échantillons après leur hydrolyse, condition indispensable pour une dérivation optimale. Ainsi la détection fiable des acides aminés totaux s'est révélée être impossible dans les FE contenant les oligopeptides inférieurs à 3 kDa. Ces fractions étant celles sur lesquelles ont été réalisées les analyses peptidomiques (voir partie 3.3), il était initialement prévu d'y quantifier également les oligopeptides afin d'avoir une exacte correspondance entre les deux approches. Malheureusement, cela n'a pas été possible, et les oligopeptides ont été quantifiés indirectement à partir des hydrolysats des SF et FNR. Cette procédure offre toutefois l'avantage d'inclure les oligopeptides supérieurs à 3 kDa et délivre ainsi une vision plus juste du réel contenu oligopeptidique des EXLs. Il convient également de rappeler que l'ensemble de cette approche ne constitue en aucun cas un réel bilan matière en raison des limitations techniques précitées, mais également du fait que la SPE n'opère pas seulement une séparation de taille, l'hydrophobicité des peptides (ainsi que d'autres interactions non spécifiques) intervenant

également au cours du procédé.

Malgré cela, des différences flagrantes de composition azotée et de dynamiques nutritionnelles ont été relevées entre les deux EXLs pendant la croissance de la souche N4L. L'EXL2 fournit en effet davantage d'azote aminé à la bactérie en raison d'une plus grande quantité d'oligopeptides, et ces derniers y sont consommés près de quatre fois plus vite que dans l'EXL1 dans lequel la croissance de la bactérie repose davantage sur l'utilisation des acides aminés libres et courts peptides. Cette importante disparition des oligopeptides dans l'EXL2 interroge quant à son origine et son impact sur la croissance bactérienne. Il est en effet possible d'envisager qu'une partie de cette diminution puisse non seulement provenir du transport des peptides dans la cellule, mais également d'une activité peptidolytique dans le milieu de culture qui les transformerait en courts peptides (activité endopeptidase) voire acides aminés libres (activité exopeptidase). Les acteurs les plus probables pouvant expliquer ces activités sont d'une part les protéases de surface (PrtS, HtrA...), et d'autre part les peptidases intracellulaires potentiellement libérées dans le milieu en raison d'une lyse cellulaire. Néanmoins, en l'absence de validation supplémentaire (e.g. : croissance d'une souche mutante $\Delta prtS$, mesure de la lyse cellulaire), il n'est pas possible à ce stade de fournir davantage d'explication concernant l'origine de ce phénomène. De plus, même si les oligopeptides de l'EXL2 transitent effectivement sous une forme intermédiaire, ils sont *in fine* assimilés par la bactérie qui utilise davantage d'azote aminé dans ce milieu. A ce titre, il est intéressant de relever qu'une croissance légèrement plus importante a été observée dans l'EXL2 sur la base des mesures de numération sur boîtes, ce qui semble bien se corréliser avec l'importante consommation azotée visible dans ce substrat. En revanche, là aussi en l'absence de validations complémentaires (e.g. dosages des autres constituants du milieu), il n'est pas possible d'établir un lien de causalité direct.

En ce qui concerne la consommation globale de peptides dans les deux EXLs en lien avec leur classe biochimique, il n'a pas été possible d'établir un lien clair entre les diminutions observées et les besoins nutritionnels de la bactérie - *S. thermophilus* n'ayant que peu si ce n'est pas d'auxotrophie stricte - ou les spécificités des deux systèmes de transport de la souche, à savoir DtpT et Ami. Il semble donc que l'assimilation d'un peptide donné soit avant tout dictée par sa biodisponibilité du moment, conclusion déjà énoncée par Kevvai *et al.* (2014) durant leur étude analogue réalisée chez *L. lactis* IL 1403. Enfin, il est intéressant de constater que, quelle que soit la classe d'azote aminé considérée, les diminutions observées sont globalement linéaires, indiquant un découplage partiel entre croissance bactérienne et consommation d'azote aminé. Ce découplage peut s'expliquer dans le cadre des acides aminés libres en raison de l'intense export des acides aminés en cours de culture venant compenser en partie la diminution liée à la consommation (Kevvai *et al.*, 2014). En revanche, la consommation des peptides - toutes tailles confondues - étant elle-même linéaire, il serait tentant d'y voir un possible reflet de l'intense activité protéolytique de sur-

face de la bactérie mise à jour par Guillot *et al.* (2016) potentiellement libératrice d'une importante quantité de nouveaux peptides.

Cette étude a quoiqu'il en soit révélé d'intéressantes différences en termes de composition mais aussi d'utilisation des peptides entre les deux EXLs. Ceci confirme notamment l'impact significatif du procédé industriel sur le contenu peptidique des hydrolysats protéiques et cellulaires. La partie suivante s'intéressera plus spécifiquement aux oligopeptides des deux EXLs dans une optique plus qualitative grâce à l'emploi de la spectrométrie de masse comme outil de caractérisation des mélanges complexes de peptides.

3.3 Approche qualitative : analyse peptidomique par spectrométrie de masse

3.3.1 Introduction

L'analyse par spectrométrie de masse du contenu peptidique des EXLs constitue la base sur laquelle a été élaboré ce projet de thèse. La connaissance précise des peptides contenus dans ces hydrolysats constitue en effet une information importante qui manque encore aux industriels fabriquant et utilisant ces substrats. La spectrométrie de masse permet ainsi de détecter les peptides en mélange au niveau individuel et de révéler leur séquence. De façon similaire au travail de suivi des acides aminés présenté dans la partie précédente, l'objectif de la présente étude est à la fois de caractériser les EXLs purs mais également d'étudier les dynamiques de variation de leur contenu peptidique pendant la croissance de *S. thermophilus* N4L. En outre, sur un plan plus académique, le contenu des milieux de culture en peptides bactériens, qu'ils soient issus d'une activité protéolytique (recyclage protéique) ou de biosynthèse, a également été analysé.

L'analyse peptidomique emprunte de nombreux éléments à l'approche protéomique *bottom-up* durant laquelle les protéines d'un échantillon sont identifiées et semi-quantifiées (quantification relative) sur la base de leurs peptides constitutifs produits par hydrolyse trypsique. Néanmoins, trois différences majeures séparent les deux approches et augmentent significativement la difficulté de l'analyse peptidomique. Premièrement, sur le plan expérimental, il est relativement plus simple d'isoler et purifier des protéines que des peptides en mélange complexe. Les protocoles actuels, principalement basés sur l'électrophorèse mono-dimensionnelle sur gel de polyacrylamide en présence de dodécylsulfate de sodium (SDS-PAGE), ciblent en effet les protéines intactes et ne permettent pas de récupérer efficacement les peptides de plus faible poids moléculaire. Deuxièmement, la digestion contrôlée des protéines repose dans la majorité des cas sur l'emploi de la trypsine, une protéase à sérine hydrolysant les liaisons peptidiques du côté C-terminal des résidus lysine et arginine. Les peptides ainsi libérés sont hautement compatibles avec les analyses en masse du fait i) d'une distribution statistique de taille en phase avec la gamme de mesure des appareils en raison de la fréquence de ces deux acides aminés dans les séquences protéiques, et ii) d'une structure spécifique (résidu basique en position C-terminale) facilitant leur ionisation et donc leur identification. A l'inverse, les peptides contenus dans des mélanges complexes tels que les hydrolysats protéiques ou cellulaires présentent une plus grande diversité de taille et de propriétés physico-chimiques, ce qui complique leur détection et identification de façon exhaustive en une seule analyse. En particulier, les courts peptides ainsi que les monochargés sont généralement exclus des analyses. Enfin, la troisième différence réside dans les méthodes de traitement des

données de masse. En protéomique *bottom-up*, les peptides ne constituent qu'un niveau intermédiaire d'information à partir duquel sont inférées les protéines de départ. Les données d'une protéine particulière agrègent ainsi les données de ses peptides constitutifs identifiés, ce qui permet de réduire l'incertitude de leur identification et quantification selon le principe de la loi des grands nombres. Par opposition, en peptidomique, le peptide constitue l'objet final de l'analyse. Sa propre information est donc directement soumise aux aléas et faiblesses propres à la spectrométrie de masse (variation de la sensibilité d'analyse au cours du temps, compétition pour l'ionisation etc.).

En raison de ces difficultés inhérentes à l'approche mais également du fait de la complexité de composition des EXLs, un important travail préalable de mise au point méthodologique a été entrepris afin d'adapter le protocole d'analyse peptidomique originellement employé dans le laboratoire au contexte de la présente étude. Cette partie va donc dans un premier temps présenter un récapitulatif de cette phase de mise au point. A l'issue de celle-ci, le protocole établi a été appliqué aux échantillons récoltés lors des cultures en bioréacteurs précédemment décrites. Les données qui en sont issues ont servi de support à la rédaction d'un manuscrit soumis pour publication qui sera présenté dans un deuxième temps. Le protocole de peptidomique a néanmoins continué d'être travaillé et amélioré après ces analyses dans le but de le simplifier et de le rendre notamment applicable à un contexte industriel. Ce dernier travail, ainsi que d'autres résultats non publiés, seront présentés dans la partie de résultats complémentaires. Enfin, les apports, les limites et les différentes perspectives ouvertes par l'ensemble de cette démarche seront discutés.

3.3.2 Evolution méthodologique pour la mise au point d'un procédé d'analyse peptidomique adapté aux EXLs

La mise au point du protocole d'analyse peptidomique adapté aux EXLs s'est déroulé en trois grandes phases :

- Phase 1 : mise au point du protocole « amont » pour la préparation des échantillons
- Phase 2 : estimation de la variabilité technique et reproductibilité du protocole
- Phase 3 : mise en place des conditions de fermentation

Ces phases vont être successivement décrites dans la présente partie avec pour double objectif de présenter d'une part le matériel et méthodes employés, et d'autre part d'explicitier les choix méthodologiques effectués et leur impact sur les résultats obtenus.

3.3.2.1 Elaboration du protocole de préparation des échantillons

Avant le démarrage du projet de thèse, des analyses préliminaires de différents EXLs fournis par Procelys ont été effectuées afin d'évaluer la faisabilité et la pertinence d'une telle approche sur les EXLs. Très brièvement, la méthodologie employée reposait sur une purification préalable des peptides des EXLs en solution par SPE et ultrafiltration suivie d'une injection directe en LC-MS/MS. Par cette méthode, le nombre de peptides uniques identifiés par EXL était de l'ordre de 100 à 200, avec une certaine variation suivant la nature de l'EXL analysé. Pour le projet de thèse, il a été décidé d'utiliser un protocole plus complexe directement transposé de l'analyse de peptidomes bactériens (analyse du surnageant de cultures réalisées en milieu chimiquement défini initialement sans peptide) dont le laboratoire a l'expertise (Guillot *et al.*, 2016). Le protocole utilisé pour les premières analyses des EXLs du projet est décrit dans la Figure 3.10.

La différence majeure entre ce protocole et les analyses préliminaires réalisées avant le démarrage du projet est la présence d'un fractionnement chromatographique (HPLC 1) préalablement à l'analyse LC-MS/MS. Chaque échantillon est ainsi divisé en plusieurs dizaines de fractions (initialement 35) qui sont ensuite individuellement analysées en spectrométrie de masse. Les peptides, au moment de leur entrée dans la source du spectromètre de masse, ont donc subi une séparation chromatographique bidimensionnelle dans deux conditions de pH différentes, améliorant grandement la profondeur d'analyse et le nombre de peptides identifiés. Ainsi, suivant ce protocole, les deux EXLs ont subi une première analyse MS selon le principe de l'empreinte de masse par spectrométrie de masse en tandem (MS/MS) décrit dans la Figure 3.11A. Au niveau bioinformatique, les peptides ont été identifiés selon la méthodologie habituellement employée en analyse protéomique (voir Figure 3.11B), mais certains paramètres ont néanmoins été adaptés. La principale modification indispensable est de préciser une absence de spécificité de clivage lors de l'interrogation de la banque protéique. En effet, en protéomique, dans le cas où la trypsine est employée pour l'hydrolyse, la présence d'une lysine ou arginine doit être indiquée en extrémité C-terminale. Ce clivage particulier permet ainsi de grandement réduire le champ d'investigation et limite le nombre de faux positifs. Afin de compenser cette absence d'hydrolyse spécifique lors de l'analyse peptidomique et maintenir un FDR (*False Discovery Rate*) bas, certains paramètres ont été fixés à des valeurs relativement stringentes. Ceci concerne en particulier la tolérance de masse sur la mesure des précurseurs (ici ± 5 ppm), ainsi que l'*E-value* maximale en dessous de laquelle un peptide est conservé (0.01). Enfin, en protéomique, il est également courant de ne conserver que les protéines pour lesquelles au moins deux peptides uniques ont été assignés afin de limiter le taux de fausses identifications. Ici, le peptide constituant l'objet final de l'analyse, toutes les identifications respectant les contraintes précédentes (tolérance de masse et *E-value*) ont été retenues.

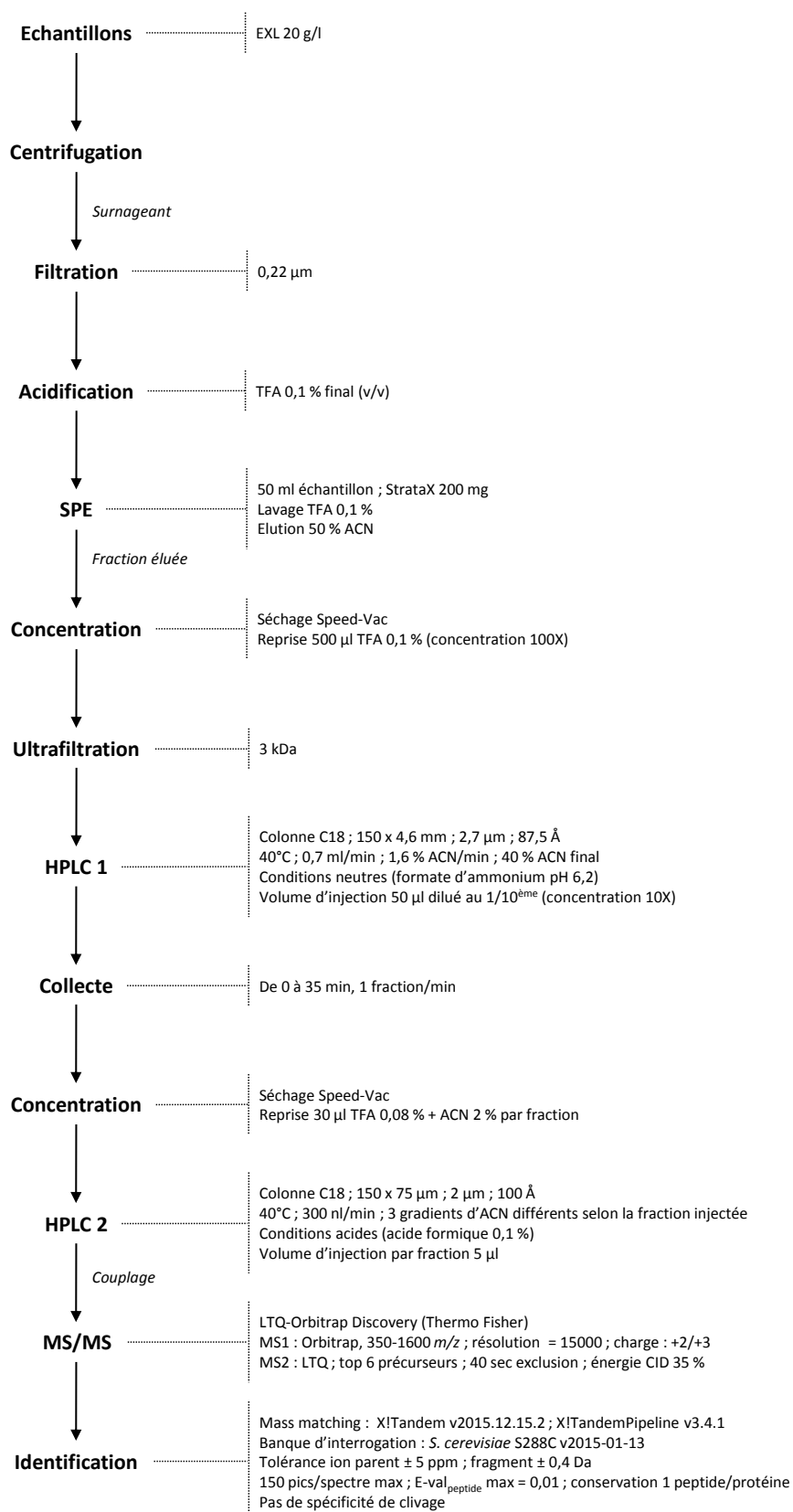
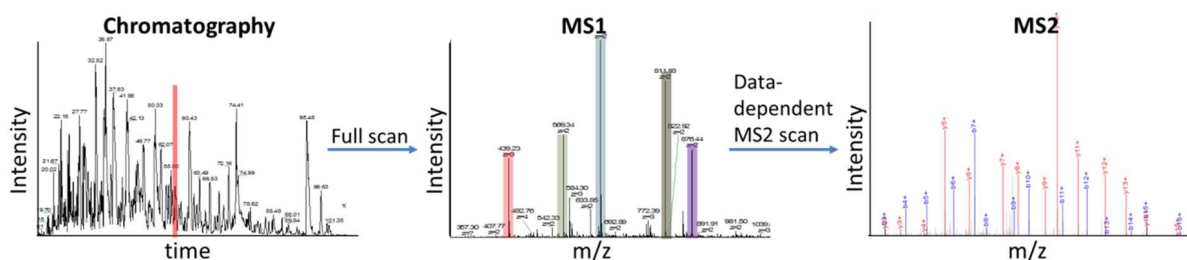


FIGURE 3.10 : Protocole de peptidomique initial hérité de l'analyse de peptidomes bactériens.

A: Data acquisition



B: Data analysis

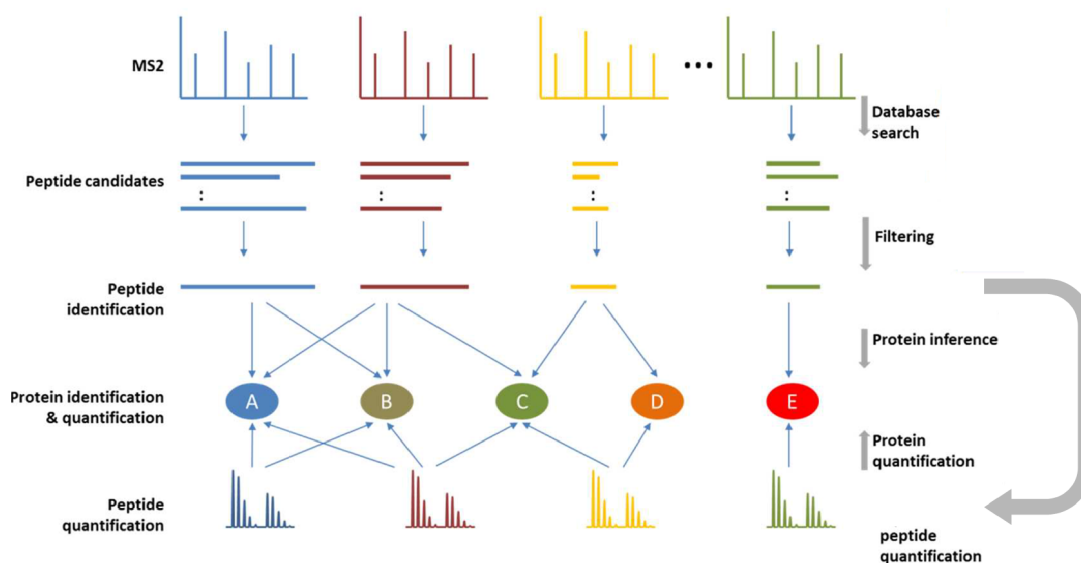


FIGURE 3.11 : (Légende page suivante)

L'ensemble de cette démarche a ainsi permis d'obtenir les résultats suivants :

- EXL1 : 6554 spectres de fragmentation assignés, 4036 peptides uniques identifiés, provenant de 1103 protéines de levure.
- EXL2 : 3555 spectres de fragmentation assignés, 2603 peptides uniques identifiés, provenant de 921 protéines de levure.

En comparaison avec les analyses réalisées avant le démarrage du projet, le fractionnement chromatographique employé ici a permis d'augmenter de plus d'un ordre de grandeur le nombre de peptides identifiés dans les deux EXLs. Il est d'ores et déjà possible de remarquer une importante différence dans le nombre de peptides totaux identifiés entre les deux EXLs (-36 % de peptides uniques dans l'EXL2 par rapport à l'EXL1). Cette information, combinée avec la connaissance de la quantité totale d'oligopeptides mesurée par HPLC (partie 3.2), indique que si les oligopeptides sont, en concentration, plus abondants dans l'EXL2, ils semblent en revanche y être moins diversifiés.

FIGURE 3.11 : Principe de l'analyse de l'empreinte de masse par spectrométrie de masse en tandem. (A) Un système LC-MS/MS réalise en continu l'acquisition du chromatogramme (courant d'ions total) issu de la séparation HPLC préalable des peptides. En parallèle, le spectromètre de masse itère en boucle un cycle constitué des événements suivants : mesure des masses entières (MS1), sélection des n masses les plus intenses (précurseurs), fragmentation et acquisition des spectres de fragmentation (MS2). La lecture MS1 s'effectue sur l'ensemble de la gamme de mesure de l'appareil, généralement comprise entre 300 et 2000 m/z . Dans un appareil de type LTQ-Orbitrap, la lecture MS1 est réalisée à haute résolution dans l'analyseur Orbitrap. La fragmentation se fait quant à elle dans le piège à ions linéaire (ou *linear ion trap*) par *collision induced dissociation* (CID) à l'aide d'un gaz inerte. La lecture MS2 est également réalisée dans le piège à ions à basse résolution. Une fois ce cycle terminé, le système applique un temps d'exclusion de l'ordre de 30 secondes (largeur moyenne d'un pic chromatographique) aux n masses analysées durant lequel elles ne seront plus fragmentées. Il recommence ensuite de nouveau ce cycle sur les autres masses entrantes. (B) Les données acquises durant l'ensemble du gradient chromatographique sont ensuite traitées par un processus bioinformatique. Des séquences peptidiques sont établies sur la base des masses entières (MS1) et des spectres de fragmentation (MS2) mesurés. Ces séquences sont ensuite comparées à des cartes de masses et spectres MS2 théoriques produits *in silico* à partir d'une banque protéique de référence. Lorsqu'il y a correspondance entre un spectre observé et spectre théorique et que celle-ci dépasse un seuil statistique choisi, on considère que le spectre est assigné et le peptide correspondant identifié. Si le temps d'éluion d'un peptide dépasse le temps d'exclusion fixé, il peut être fragmenté plusieurs fois, d'où le fait que le nombre de spectres assignés est toujours supérieur ou égal au nombre de peptides uniques identifiés. Le nombre de spectres associé à un peptide donné est donc un indicateur relatif de son abondance (*spectral counting*). Les peptides identifiés sont ensuite alignés sur les séquences protéiques de la banque d'interrogation afin d'inférer les protéines parentales. Adapté de Zhang *et al.*, 2013.

Ce protocole initial a fourni des résultats très satisfaisants concernant la profondeur d'analyse atteinte. Néanmoins, il souffre de deux limitations majeures : i) lenteur d'analyse, le fractionnement HPLC démultipliant le temps d'analyse en masse (ici, un seul échantillon initial équivaut à plus de 26 h d'analyse, sans compter les blancs et standards intermédiaires, ni la préparation préalable des échantillons), et ii) faible reproductibilité, liée en particulier à un probable encrassement du système LC-MS/MS. En effet, une deuxième série d'injections réalisées à partir des mêmes échantillons a montré une importante baisse du nombre de peptides identifiés de l'ordre de 40 %. En tenant compte des facteurs de concentrations et des volumes injectés en HPLC (voir Figure 3.10), on peut estimer la quantité finale d'EXL injectée en masse dans l'ensemble des 35 fractions comme étant de l'ordre de 1 mg, ce est probablement trop et responsable de la baisse de performances observée.

Un travail a donc été effectué afin de remanier ce protocole et trouver une juste balance entre profondeur d'analyse et maintien des performances du spectromètre de masse *via* un allègement de la quantité de matière injectée et une intensification de la purification des échantillons. Ce travail a abouti au protocole présenté dans la Figure 3.12.

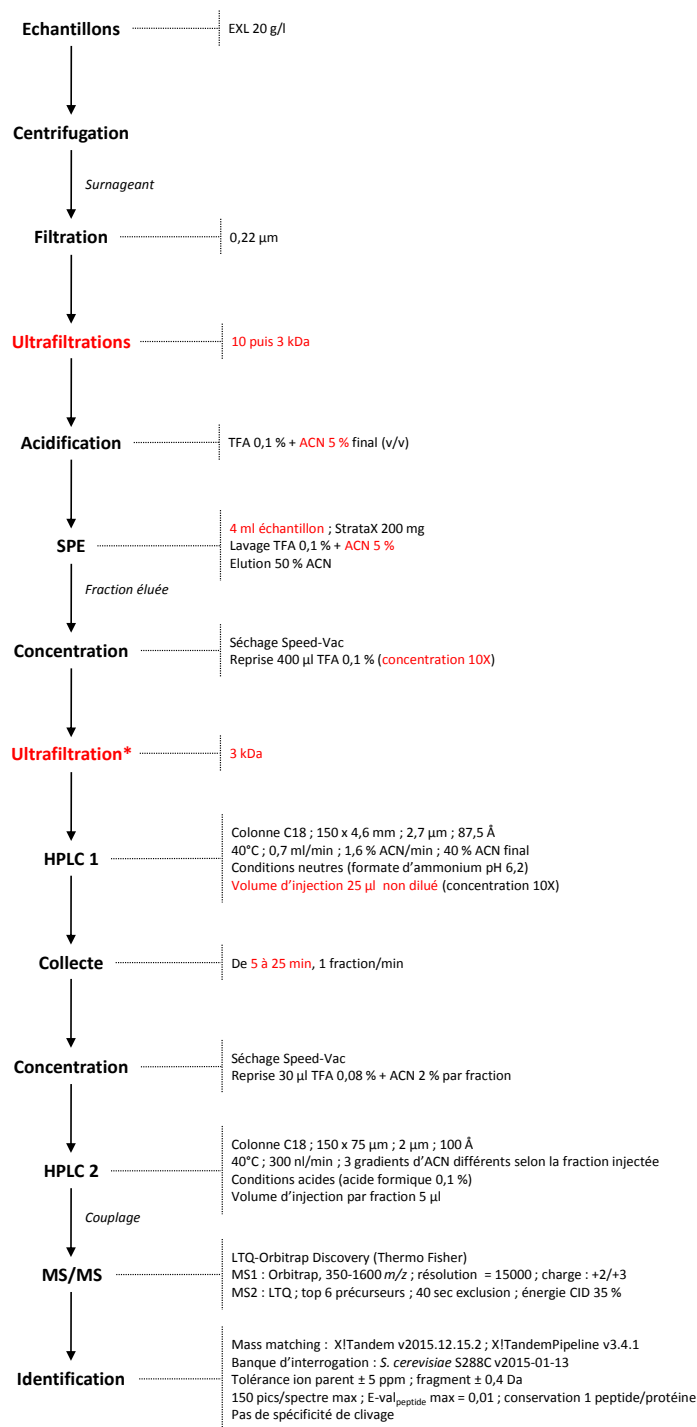


FIGURE 3.12 : Protocole de peptidomique final. En rouge sont indiquées les modifications opérées par rapport au protocole initial. Astérisque : étape d'ultrafiltration 3 kDa post-SPE employée uniquement avec les fermentations finales et non durant la deuxième phase de mise au point.

Succinctement, le premier axe de travail a consisté à réduire la charge peptidique supportée par la cartouche de SPE. Pour cela, le volume d'échantillon traité a d'une part été réduit à 4 ml - diminution par ailleurs nécessaire pour limiter le volume des futurs prélèvements en

bioréacteur et permettant un gain de temps important durant l'étape de SPE (chargement de l'échantillon sur la colonne) - et l'ultrafiltration à 3 kDa a été déplacée avant la SPE et a été réalisée en deux temps en ajoutant une ultrafiltration à 10 kDa intermédiaire. Ensuite, la SPE a été réalisée en présence de 5 % d'ACN afin d'éliminer une forte quantité de composés hydrophiles apparaissant en début de gradient à 214 nm lors de l'HPLC1. La quantité d'échantillon injectée durant cette chromatographie a par ailleurs été réduite de moitié. Enfin, le deuxième axe de travail a consisté à étudier la charge peptidique contenue dans les différentes fractions chromatographiques collectées afin de ne sélectionner que celles contenant la majeure partie des peptides identifiés. Les fractions collectées entre la 5^{ème} et 25^{ème} min du gradient ont ainsi été retenues car contenant plus de 90 % des peptides identifiés dans les deux EXLs. Les paramètres d'analyse LC-MS/MS n'ont pour leur part pas été modifiés. Une fois ces différents paramètres fixés, la variabilité technique de ce protocole a ensuite été évaluée.

3.3.2.2 Estimation de la répétabilité technique

Pour estimer la fidélité du protocole final élaboré, six répétitions techniques par EXL ont été effectuées incluant l'intégralité du processus analytique, de la préparation des échantillons jusqu'aux analyses en masse. Les EXLs n'ont pas été analysés directement purs comme précédemment mais ont été préalablement stérilisés (120 °C, 20 min) et mélangés avec les autres composants du milieu de culture (m/v) : lactose 6 %, CaCl₂ 0,01 % et acide ascorbique 0,05 %. Ceci devait permettre de se placer dans les conditions de fermentation et tenir compte d'un éventuel effet matrice. A noter que la composition de ce milieu a par la suite évolué, des détails seront donnés dans la partie 3.3.2.3 suivante.

Les résultats de la qualification du protocole sont présentés dans la Figure 3.13. En moyenne, 1403 peptides différents ont été identifiés dans l'EXL1 *via* 1911 spectres de fragmentation, et 885 peptides (1162 spectres) dans l'EXL2 (Figure 3.13A). En comparaison avec la première analyse réalisée avec le protocole initial, cela représente une diminution du nombre d'identifications par échantillon de plus de 60 %. Cette baisse, bien que considérable, a néanmoins permis de réduire la variabilité d'une analyse à l'autre dans des valeurs acceptables, le coefficient de variation ($CV = \sigma/\mu$) moyen étant d'environ 10 % dans les deux EXLs. Toutefois, l'analyse répétée d'un même échantillon permet de compenser en partie cette perte de profondeur d'identification. Ainsi, en combinant l'ensemble des six répétitions, ce sont 3610 peptides uniques (11465 spectres) qui ont été identifiés dans l'EXL1, 2440 peptides (6972 spectres) dans l'EXL2, et 755 peptides qui ont été identifiés dans les deux substrats (Figure 3.13B). On remarque à nouveau un nombre de peptides présents dans l'EXL2 moindre que dans l'EXL1 (-32 %), du même ordre que l'écart mesuré lors de l'analyse précédente (-36 %).

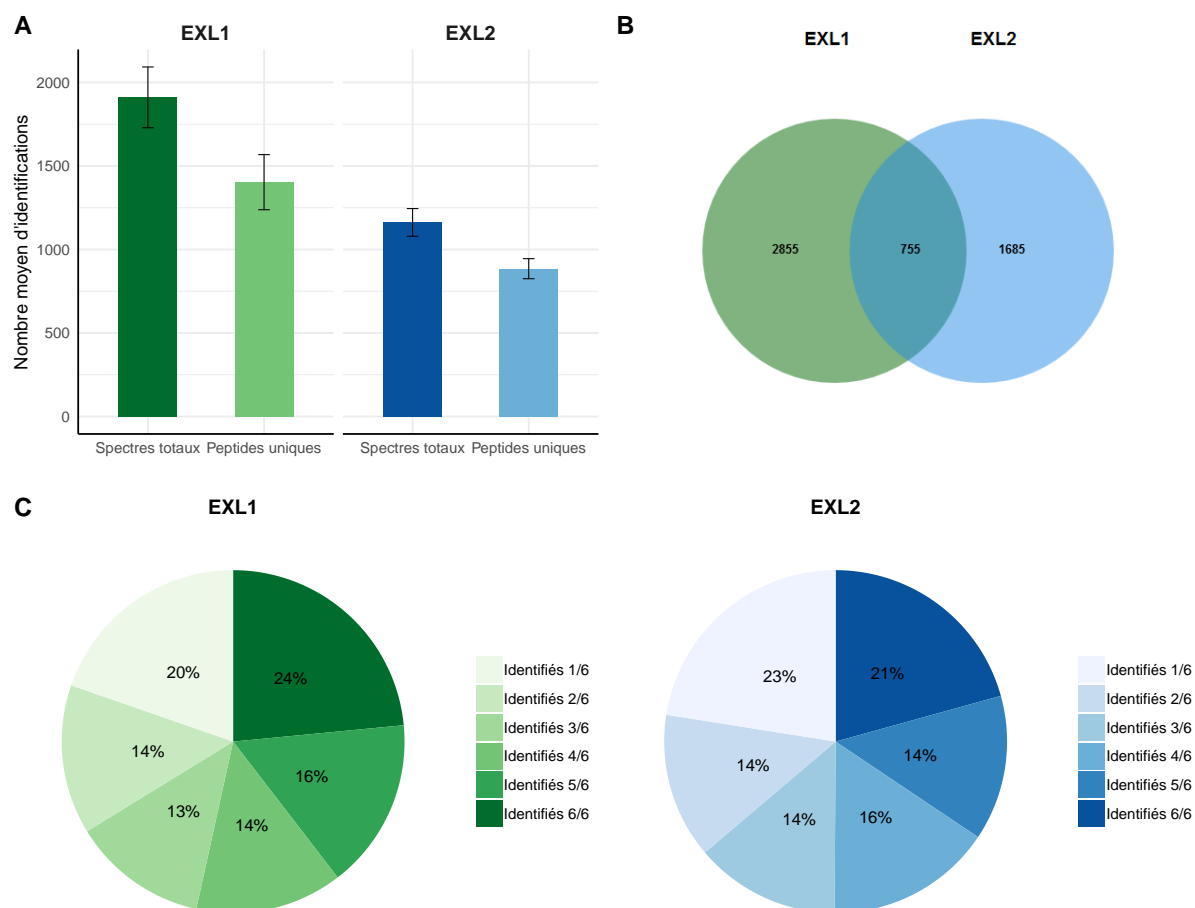


FIGURE 3.13 : Principaux résultats obtenus lors l'analyse de répétabilité du protocole de peptidomique. (A) Nombre moyen de spectres totaux assignés et peptides uniques identifiés par analyse (B) Diagramme de Venn montrant le total de peptides uniques communs et spécifiques aux EXLs détectés dans l'ensemble des six répétitions. (C) Parts moyennes de peptides à l'échelle d'une analyse détectés systématiquement dans l'ensemble des six répétitions (6/6), ponctuellement (de 5/6 à 2/6), et spécifiques à une seule répétition (1/6).

Si la répétabilité d'ensemble se révèle donc satisfaisante, la situation est plus mitigée à l'échelle du peptide individuel. La Figure 3.13C dresse un bilan de la répétabilité d'identification des peptides. Les peptides identifiés systématiquement lors des 6 répétitions représentent un peu moins de 25 % du nombre moyen de peptides identifiés par analyse dans les deux EXLs. Les peptides identifiés ponctuellement, entre 2 et 5 répétitions, représentent plus de la moitié des cas, et enfin les peptides spécifiques à une seule répétition comptent pour un peu moins d'un quart également. En d'autres termes, et pour résumer, si le nombre total de peptides identifiés est globalement constant, les peptides détectés ne sont pas systématiquement les mêmes. Cette variabilité est elle-même le résultat de la variabilité liée à la purification des peptides, mais aussi à leur détection dans le spectromètre de masse. Des phénomènes de compétition pour l'ionisation doivent en particulier intervenir au niveau de la source (électro-spray) rendant l'identification de certains d'entre

eux aléatoire.

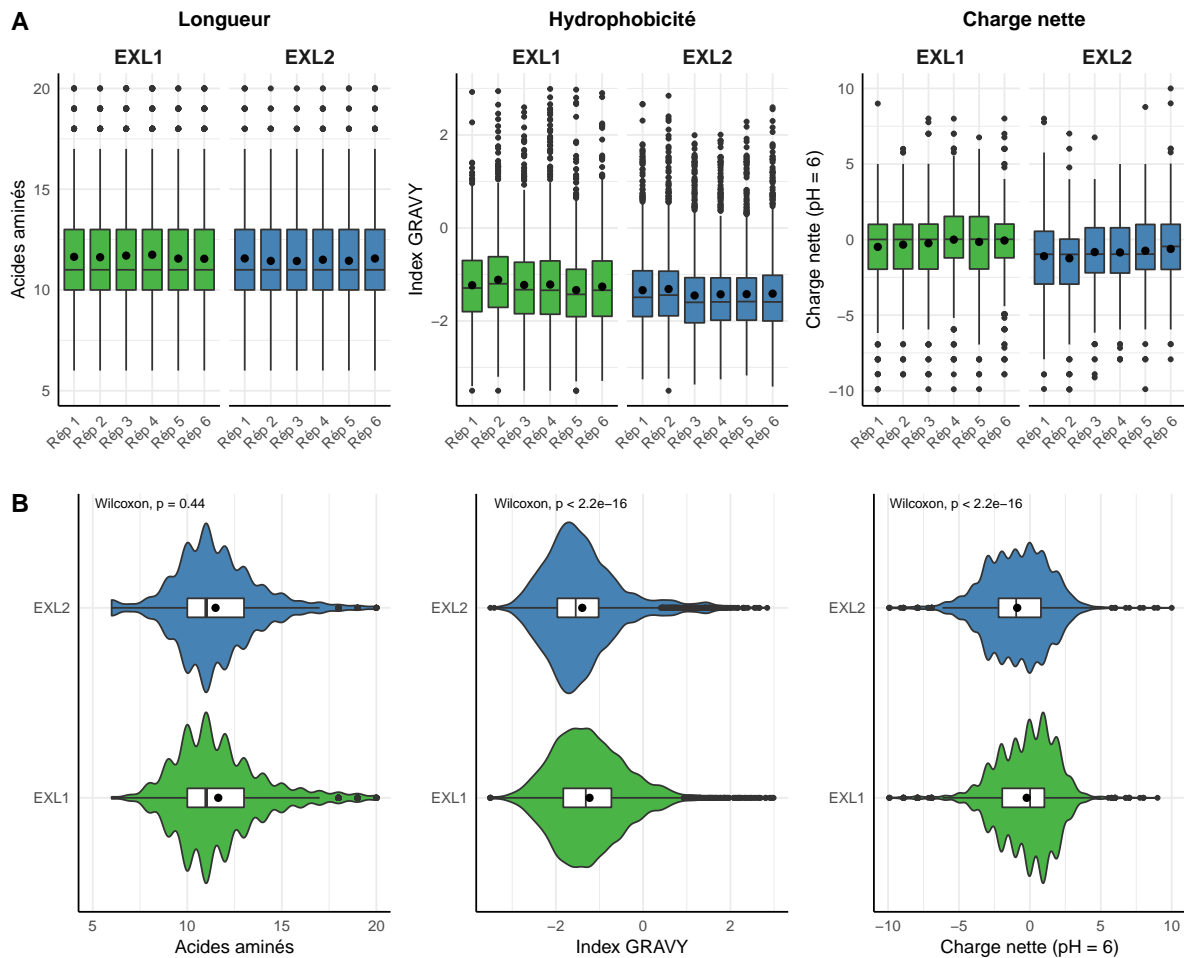


FIGURE 3.14 : Caractéristiques physico-chimiques des peptides identifiés lors de l'analyse de répétabilité. (A) Boxplots représentant la distribution par répétition de la taille, hydrophobicité et charge nette des peptides. Le rectangle central inclut 50 % des effectifs, la médiane de la distribution est représentée par une ligne horizontale et la moyenne par un point. (B) Violin plots et boxplots indiquant le détail de la distribution des caractéristiques physico-chimiques pour l'ensemble des répétitions combinées. Tous les graphes ont été réalisés à partir des spectres totaux assignés afin de tenir compte de la redondance d'identification.

Les peptides identifiés ont également été caractérisés du point de vue de leur taille, hydrophobicité (index GRAVY; Kyte et Doolittle, 1982) et charge nette calculée à pH = 6 (Moore, 1985). Ces propriétés sont représentées dans la Figure 3.14. La comparaison des propriétés d'ensemble entre répétitions (Figure 3.14A) montre tout d'abord une très bonne répétabilité concernant la taille des peptides identifiés. En revanche, des variations ont été détectées (ANOVA de Kruskal-Wallis, $p\text{-val} \leq 0,01$; Kruskal et Wallis, 1952) entre répétitions concernant l'hydrophobicité et la charge nette des peptides dans les deux EXLs. Ces variations, bien que significatives, sont néanmoins relativement minimales et ne dévient que faiblement autour de la valeur moyenne du critère physico-chimique considéré dans l'ensemble des répétitions du même EXL. Cette valeur moyenne ainsi que la distribution globale des trois propriétés est

présentée dans la Figure 3.14B. Des tests non-paramétriques de Wilcoxon-Mann-Whitney (Wilcoxon, 1945; Mann et Whitney, 1947) ont été effectués afin de détecter une différence de distribution entre les deux EXLs pour chacun des trois critères. Aucune différence significative n'a été observée au niveau de la taille moyenne des peptides entre les deux EXLs. Visuellement, on peut toutefois souligner une particularité de l'EXL2 qui contient légèrement plus de peptides de courte taille (6 résidus) que l'EXL1, bien que cette différence n'influe pas sur la moyenne générale et la médiane (Figure 3.14B). En revanche, des différences significatives ont été détectées au niveau de l'hydrophobicité et de la charge des peptides entre les deux EXLs. Plus précisément, davantage de peptides hydrophiles et chargés négativement ont été détectés dans l'EXL2 que dans l'EXL1.

Il ressort donc de ce travail préliminaire que la variabilité de l'analyse peptidomique est avant tout fonction de l'échelle à laquelle sont analysées les données. Lorsque celles-ci sont considérées sur une échelle globale, c'est-à-dire en regardant la masse peptidique dans son ensemble, le protocole d'analyse apparaît relativement robuste, à la fois en termes d'identifications totales mais aussi de conservation des propriétés physico-chimiques entre répétitions. En revanche, lorsque l'on se rapproche de l'échelle du peptide individuel, une importante variabilité apparaît résultant principalement du caractère partiellement aléatoire de l'identification des peptides. Ceci concerne en particulier les peptides dont l'abondance relative est la plus faible. Celle-ci peut être facilement estimée par *spectral counting*, c'est-à-dire en comptant le nombre de spectres (nombre de fragmentations) par peptide (Liu *et al.*, 2004; Colinge *et al.*, 2005). Un peptide abondant dispose en général d'un large pic chromatographique et peut être fragmenté plusieurs fois. De plus, en raison du dispositif expérimental employé ici (double fractionnement chromatographique), un peptide abondant peut également se retrouver dans plusieurs fractions différentes, augmentant ainsi ses chances de fragmentation pendant l'analyse en masse. À l'inverse, un peptide peu abondant n'est que peu fragmenté en raison d'un pic plus fin et du temps d'exclusion appliqué (voir Figure 3.11 pour plus d'explications). Ainsi, dans les deux EXLs analysés et en combinant les données de l'ensemble des 6 répétitions, plus de 80 % des peptides identifiés n'ont été fragmentés qu'une seule fois. À l'inverse, une faible minorité (environ 1 %) a été identifiée *via* un nombre de spectres plus important (≥ 4). Ces données d'abondance relative se corrélaient très bien avec la reproductibilité d'identification présentée dans la Figure 3.13C, les peptides les plus abondants se retrouvant dans le groupe des peptides systématiquement (6/6) ou fréquemment identifiés (4/6 et 5/6), et inversement pour les peu abondants. Dans l'objectif d'interpréter les données issues des cultures de *S. thermophilus* N4L, une stratégie d'analyse originale a donc dû être développée. Néanmoins, avant de réaliser ces cultures finales, un travail de mise en place des conditions de fermentation a également dû être effectué.

3.3.2.3 Mise en place des conditions de fermentation de *S. thermophilus* N4L

Avant de procéder aux fermentations ayant permis de produire les différentes données omiques, il a été nécessaire d'entreprendre un important travail de mise au point des conditions de fermentation. En effet, si les premières fermentations réalisées en début de thèse n'ont pas posé de problème majeur, celles-ci utilisaient du Tween 80 (0,1 %, v/v) dans les milieux de culture. Or, ce composé, un émulsifiant dérivé de l'acide oléique et de sorbitane, est hautement incompatible avec les analyses en masse (Jäpelt *et al.*, 2016). Des cultures réalisées dans des milieux dépourvus de ce composé ont révélé qu'il était un important facteur de croissance pour la bactérie, en particulier en bioréacteur. Dans des cultures à plus petite échelle et à pH non régulé, la simple addition d'acide ascorbique et de tampon phosphate (K_2HPO_4/KH_2PO_4) a suffi à rétablir des performances de croissance comparables à celles observées dans les mêmes conditions avec du Tween 80. En revanche, cela n'a pas été le cas dans les fermenteurs. Afin de retrouver des performances de croissance satisfaisantes et répétables, il a été nécessaire d'ajouter le mélange complet de vitamines utilisées dans le milieu chimiquement défini développé pour *S. thermophilus* (Letort et Juillard, 2001). Le milieu définitif utilisé s'éloigne donc des milieux simples employés en particulier à l'échelle industrielle, mais il s'agit d'un compromis indispensable pour la réalisation des analyses en spectrométrie de masse.

Enfin, une fois ces conditions de culture fixées et avant de réaliser les fermentations finales, une dernière modification du protocole de préparation des échantillons a été effectuée : ajout d'une ultrafiltration à 3 kDa supplémentaire après la SPE, voir Figure 3.12. Ceci encore une fois afin de limiter la charge de matière supportée par le système LC-MS/MS ainsi que la présence de matière insoluble lié au facteur de concentration de l'EXL après SPE (voir Figure 3.12).

3.3.3 Manuscrit

Manuscrit soumis le 4 octobre 2018 au journal Applied and Environmental Microbiology.

Insights into the complexity of yeast extract peptides and their utilization by *Streptococcus thermophilus*

Lucas Proust^{a,c}, Alain Sourabié^b, Martin Pedersen^c, Iris Besançon^b, Véronique Monnet^a and Vincent Juillard^{a*}

^aMicalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France;

^bProcelys, Lesaffre Group, Maisons-Alfort, France;

^cSacco S.r.l., Cadorago, Italy

*Corresponding author:

Tel: (+33) 134 652 068

Fax: (+33) 134 652 163

E-mail: vincent.juillard@inra.fr

Keywords

Streptococcus thermophilus, yeast extract, peptidomics, oligopeptide utilization, transport specificity

1 **Abstract**

2 *Streptococcus thermophilus*, an extensively used lactic starter, is generally produced in yeast extract-
3 based media containing a complex mixture of peptides whose exact composition remains elusive. In this
4 work, we aimed at investigating the peptide content of two different yeast extracts (YEs) and identifying
5 dynamics of peptide utilization during the growth of the industrial proteinase-positive *S. thermophilus*
6 N4L strain, cultivated in 1-liter bioreactors under pH-regulation. We set up a complete analytical
7 workflow based on mass spectrometry (peptidomics) that allowed the identification of about 4,000
8 different oligopeptides in both analyzed YEs during the time-course of the experiment. These peptides
9 ranged from 6 to more than 30 amino acids in length. The subsequent study of relative peptide abundance
10 over the course of the fermentation first revealed the activity of the oligopeptide carrier Ami in both
11 YEs. Specifically, we unambiguously showed that the physicochemical characteristics of consumed
12 peptides matched the known affinities of the Ami system. Moreover, by analyzing such a large number
13 of peptides, we were able to establish that peptide net charge is the major factor for oligopeptide
14 transport in *S. thermophilus* N4L. Finally, we also highlighted another peptide utilization dynamics
15 (diminution of anionic peptides) whose origin is intriguing and that was only specific to one YE.

16

17 **Importance**

18 Protein hydrolysates play an important role in industrial microbiology as substrates for fermentation
19 media. However, their peptide composition remains uncharacterized. This can be problematic for the
20 production of starter cultures for which the availability of nitrogen sources is essential. This shortfall
21 mostly results from a lack of appropriate analytical tools. Using yeast extract as a protein hydrolysate
22 model, we have developed a mass spectrometry-based analytical pipeline that makes it possible to
23 acutely analyze and monitor its peptide composition evolution during fermentation. Furthermore, this
24 method also constitutes an efficient tool to characterize in depth bacterial transport preferences *in vivo*,
25 as evidenced here with *Streptococcus thermophilus*. This approach could be therefore extended to other
26 complex substrates and other microorganisms to deepen our knowledge of peptide-related dynamics.

27 **Introduction**

28 Lactic acid bacteria (LAB) are a group of microorganisms displaying a wide range of properties making
29 them suitable for various applications in fields such as health (1), chemistry (2, 3) or cosmetics (4).
30 However, they have been historically used for food production (5), which still remains their main
31 outcome nowadays, in particular for dairy starters. Therefore, detailed information is available regarding
32 the growth of LAB in milk, especially regarding the proteolytic system on which they rely for their
33 amino nitrogen nutrition (6, 7). When considering the whole lifetime of industrial dairy starters, it
34 appears that the culture medium in which they are industrially produced also represents an important
35 substrate for these bacteria. Yet, in contrast to milk culture, less academic work has been performed on
36 this particular step, although it is of great importance as it impacts production yields and technological
37 functionalities of the starter (8). Production media usually rely on complex substrates such as cell or
38 protein hydrolysates that notably include yeast extracts (YEs), which are widely used for LAB industrial
39 production. YEs correspond to the soluble fraction of molecules released after either yeast autolysis or
40 controlled enzymatic lysis (9). They contain several classes of nutrients among which peptides that can
41 represent more than 50% of the total YE mass, although this proportion depends on their own
42 manufacturing process. They are thus mainly employed as amino nitrogen sources. If the importance of
43 peptides in milk cultures has been extensively covered - especially with the LAB model *Lactococcus*
44 *lactis* (10–13) less is known about their utilization in YE-based media.

45 On one hand, peptide utilization in YE naturally depends on the composition of the peptide
46 fraction, which precisely constitutes the main hurdle to this study due to the abundance and diversity of
47 peptides composing these substrates. To our knowledge, the YE peptide fraction has only been
48 characterized up to now *via* filtration or chromatographic fractionation strategies (14–17), or by
49 analyzing the composition in peptide-bound amino acids (18, 19). Nonetheless, the exact nature of the
50 peptides composing YEs still remains elusive, these compounds are therefore commonly considered as
51 black boxes containing a complex mixture of undefined peptides.

52 On the other hand, peptide utilization is also dependent on the transport machinery of the
53 cultivated LAB. Indeed, the only known LAB peptidases able to convert peptides to free amino acids
54 are intracellularly located (6, 20). Several peptide transport systems are available in LAB among which
55 two are particularly well documented and present in various species. The first one is the di- and tri-
56 peptide transporter DtpT, a secondary transporter belonging to the PTR family (proton motive-force
57 dependent carrier). It is a protein encoded by a single gene and constituted of 12 transmembrane domains
58 that couples peptide transport with the proton gradient (21). It has been identified in several LAB, among
59 which *L. lactis* (22), *Lactobacillus helveticus* (23) or *Streptococcus thermophilus* (24, 25). The second
60 well-known system for peptide uptake is an ABC (ATP-binding cassette) transporter dedicated to
61 oligopeptides (3 residues and more). It is best known as Opp in many LAB species, in particular in *L.*
62 *lactis* where it has been first characterized (26). It was equally found in *S. thermophilus* where it was
63 named Ami due to its high sequence homology with other streptococcal transporters (27). Both Opp and

64 Ami share a similar overall genetic organization. They consist of five conserved proteins organized in a
65 single operon: OppDFBCA and AmiA1CDEF, respectively. OppA and AmiA1 are lipoproteins
66 anchored to the cell membrane and devoted to peptide binding and delivery to the translocon formed by
67 OppBC/AmiCD. Peptide internalization is enabled by the cytoplasmic membrane-bound ATPases
68 OppDF/AmiEF that energize the whole system upon ATP binding and hydrolysis. However, the Ami
69 system possesses supplementary AmiA proteins, in opposition to OppA present in only one copy, which
70 is a characteristics of streptococci. Extra AmiA proteins are located in other parts of the genome, and
71 their number is strain-dependent.

72 These two combined systems, DtpT and Opp/Ami, thus allow for a large supply of various
73 peptides to the bacterial cells. However, peptide length is not the sole factor upon which these carriers
74 operate, as all peptides are not indiscriminately transported even when belonging to the adequate size
75 range. In the case of DtpT, it has been evidenced in several species that it had a higher affinity for
76 dipeptides over tripeptides, preferred hydrophobic substrates and worked less efficiently with cationic
77 peptides (23, 28–30). Concerning the oligopeptide carrier specificity, extensive information is available
78 for *L. lactis* Opp. Specifically, *in vivo*, *in vitro*, and structural characterization data are available (10–
79 12, 26, 31–39). In comparison, apart from its initial genetic identification and characterization (27), only
80 one *in vivo* study has been performed on *S. thermophilus* Ami transporter to determine its substrate
81 preferences (40). Complementary approaches based on the use of mixtures of milk peptides showed that
82 oligopeptide transport seemed to be favored by hydrophobic and positively charged peptides, whereas
83 long anionic ones were never taken up. However, this study was based on a limited number of peptides
84 notably resulting from the tryptic digestion of α_{s2} -casein and therefore presenting biochemical
85 similarities such as a positively charged C-terminal residue (Lys or Arg). One can therefore question
86 whether the trends shown by the consumption of such limited and specific peptide mixtures are
87 representative of the Ami oligopeptide transporter specificities, and furthermore whether they can be
88 extended to a YE-based medium where oligopeptides are abundant and available in the form of a large
89 complex mix.

90 In this study, we specifically aimed at qualitatively characterizing the oligopeptide fraction of
91 YE-based media and monitoring the dynamics of oligopeptide utilization occurring during the growth
92 of an industrial *S. thermophilus* strain. For that purpose, we developed a specific peptidomics-based
93 analytical pipeline combined with a dedicated data analysis workflow. This whole approach, applied to
94 cultures based on two different commercial YEs, revealed the complexity of YE peptide fractions as
95 well as peptide utilization dynamics in both YEs that notably reflected the activity and the specificity of
96 the oligopeptide transporter Ami.

97

98 **Material and methods**

99

100 Strain and preculture conditions

101 This work used *S. thermophilus* N4L (41), a PrtS-positive, AmiA2-negative and AmiA3-positive
102 industrial starter. This strain contains as well the gene encoding the DtpT transporter, and does not
103 possess the putative Ots peptide transport system facultatively present in other strains (24, 42).
104 Therefore, DtpT and Ami are the only known peptide transport systems identified in this strain. It was
105 stored at -80°C in M17 broth (43) containing 1% (w/v) lactose and supplemented with 20% (v/v)
106 glycerol. The strain was routinely pre-cultured in M17 broth containing 5% lactose at 42°C.

107

108 Bioreactor culture conditions

109 Bioreactor cultures were performed in a 1-liter bioreactor system BIOSTAT® Qplus (Sartorius Stedim
110 Biotech, Germany). Two successive precultures were performed prior to inoculating at 2% (v/v) the
111 fermenters. The culture medium contained (w/v) 6% lactose, 0.01% calcium chloride and 2% of YE1 or
112 YE2, two commercial YEs provided by Procelys (Maisons-Alfort, France) belonging to the Nucell®
113 range notably dedicated to dairy starters. A pool of vitamins as used in *S. thermophilus* chemically
114 defined medium (44) was also added in the culture media in order to ensure repeatable and optimal
115 growth performances similar to those obtained with equivalent Tween 80-containing media (data not
116 shown). Tween 80 is a widely used growth factor for lactic acid bacteria cultures (45). However, it is a
117 highly ionizable compound known to disrupt mass spectrometry analyses (46) and thus was not useable
118 in this study. The YE fraction of the media was sterilized by heat treatment along with the bioreactors
119 for 20 min at 120°C. The other components were sterilized by a 0.22 µm pore-sized polyethersulfone
120 membrane (Millipore, Guyancourt, France). The initial pH was adjusted beforehand to 6.6 with sodium
121 hydroxide 2 M. The batch fermentations were carried out during 6 hours at 40°C, the stirring was fixed
122 at 50 rpm and the pH regulated at 6.0 with sodium hydroxide 2 M. Pseudo anaerobic conditions were
123 set up by sparging nitrogen at 0.2 liter/min in the growth medium for 1 hour before inoculation, and in
124 the headspace at the same flow rate thereafter. Growth was followed by optical density (600 nm)
125 measurement and by online monitoring of the volume of base added.

126 Three independent cultures were carried out during which samples were collected each hour for
127 the first repetition. In order to maximize reproducibility, samples from the two subsequent repetitions
128 were taken when the added volumes of base reached the corresponding levels of the first repetition.

129

130 Yeast peptide identification

131 YE peptides in the culture media were identified using an approach adapted from a previous work (47).
132 Bacterial cells were first discarded from the fermentation samples by centrifugation (3000 g, 10 min,
133 4°C). The peptide-containing supernatants were filtered using a 0.22 µm pore-sized PVDF membrane
134 with low protein binding properties (Millipore, Guyancourt, France). Trifluoroacetic acid (TFA) and
135 acetonitrile (ACN) were then added at final concentrations of respectively 0.1% and 5% (v/v). The mixes
136 were centrifuged (3000 g, 10 min, 4°C) after a one-night storage at 4°C, then ultrafiltered successively
137 through 10 kDa and 3 kDa cut-off Amicon® Ultra-15 devices (Ultracel®-10K and Ultracel®-3K

138 membranes, resp., Millipore). The final permeates went through a solid phase extraction step using a
139 200 mg StrataX® cartridge according to manufacturer procedure (Phenomenex, Le Pecq, France).
140 Briefly, the activated cartridge was loaded with 4 ml of sample, washed with 5% ACN and 0.1% TFA,
141 and peptides were finally eluted with 1.5 ml of 50% ACN and 0.1% TFA in MilliQ water (Waters, St-
142 Quentin-en-Yvelines, France). The eluted fractions obtained were then dried overnight in a Speed-Vac
143 system (Savant, Thermo Scientific) and resolubilized in 400 µl of 0.1% TFA. Finally, the concentrates
144 were ultrafiltered once again through a 3 kDa cut-off Amicon® Ultra-4 device prior to a double
145 chromatographic separation (Figure 1A).

146 The first separation was performed on a Nucleoshell RP 18plus reversed-phase column (150 x
147 4.6 mm, 2.7 µm, 87.5 Å, Macherey-Nagel, Germany) at 40°C with an injection volume of 25 µl
148 corresponding to 250 µl of the initial supernatant. A linear gradient of acetonitrile (1.6% per min) in
149 ammonium formate (20 mM, pH 6.2) was applied with a 0.7 ml per min flow rate and fractions were
150 collected every minute. Preliminary analyses showed that the initial and last fractions collected
151 respectively before 5 min and after 25 min of the chromatographic run contained less than 5% of the
152 identified peptides (data not shown) and thus were discarded. The remaining 21 fractions were
153 subsequently dried overnight in a SpeedVac system and resuspended in 30 µl of 0.1% TFA and 2%
154 acetonitrile in MilliQ water. Aliquots of 5 µl were analyzed by a data-dependent tandem mass
155 spectrometry approach on the PAPPISO platform (INRA, Jouy-en-Josas). All the *modus operandi* of the
156 second chromatographic separation and peptide m/z detection - i.e. equipment and settings - were the
157 same as those previously described (47). Peptide identification was performed with X!Tandem version
158 2015.12.15.2 (Vengeance) and X!TandemPipeline (C++) version 0.2.16 (48) on the protein sequence of
159 *Saccharomyces cerevisiae* S288C (version 2015-01-13, www.yeastgenome.org). The main peptide
160 identification parameters were the following: no cleavage specificity, variable methionine oxidation
161 state and mass tolerance for parent and fragment ions of ± 10 ppm and ± 0.4 Da, respectively. Peptides
162 were conserved when showing an *E*-value ≤ 0.05, and only one peptide per parental protein was
163 considered as sufficient to enable identification. Contaminant peptides were discarded using a classical
164 proteomic contaminant database, and the False Discovery Rate was estimated using the reversed protein
165 database.

166

167 Peptide physicochemical characterization and class assignment

168 Peptides were characterized by nine different physicochemical properties listed in Table 1. At the
169 exception of the proportions of aromatic residues and proline which were manually calculated, all other
170 properties were computed using the aminoAcidProperties function of the R package 'alakazam' version
171 0.2.8 (49). Default settings were kept for scaling and normalization procedures.

172 Prior to differential analysis, peptides were grouped together in classes based on their
173 physicochemical closeness (Figure 1B). For that purpose, a specific procedure was performed in order
174 to assign every identified peptide to a unique physicochemical class. The range of each descriptor was

175 divided into three intervals (Table 1 and Figure 3). For each physicochemical criterion, a given peptide
176 can belong to only one interval. The physicochemical class or "barcode" of the peptide is then defined
177 as the combination of the respective intervals of the nine descriptors. As an example, the peptide
178 KGSIDEQHPRYGG belongs to the class "321223322". It is 13-residues long, therefore it belongs to
179 the interval 3 of length, its GRAVY index value is -1.73: interval 2, and so on.

180

181 Statistical analyses

182 Wilcoxon-Mann-Whitney (50, 51) and Kruskal-Wallis (52) non-parametric tests were used in order to
183 detect statistical differences in the peptide property distributions between 2 or more than 2 groups,
184 respectively. For that purpose, the *ad hoc* functions of the R package 'stats' version 3.4.3 were
185 employed.

186 Differential analysis over time was performed on the abundance of peptide classes with the R
187 script MassChroqR (version 0.3.8) of the MassChroQ pipeline (53). A contingency table was generated
188 beforehand that contained the total spectral counts of each physicochemical class - i.e. the sum of the
189 spectral counts of their constitutive peptides - in each analyzed sample. Classes showing low abundance
190 (< 5 spectra in all samples) and little variations (< 50% variation between the minimal and maximal
191 average abundance observed in the different samples) were discarded. Finally, spectral count data of
192 remaining classes were modeled *via* a generalized linear model with a Poisson distribution. An analysis
193 of variance (ANOVA) was then performed using a Chi² test to detect significant variations in peptide
194 class abundances with time considered as factor of analysis. The generated *p*-values were adjusted by a
195 FDR procedure (54). Classes showing an adjusted *p*-value ≤ 0.01 were considered as varying
196 significantly over time (Figure 1C). A kinetic profiling was then manually performed in order to separate
197 classes showing either a decrease, increase or fluctuating variations over-time. The physicochemical
198 properties of peptides belonging to the different kinetic profiles determines were then discriminated by
199 a principal component analysis (R package 'FactoMineR' version 1.41).

200 In order to assess the validity of this class-based procedure, data were also assign to random
201 classes, i.e. classes whose peptides were grouped randomly and not according to their physicochemical
202 characteristics. The total number of random classes and their general composition (number of peptides
203 per class) were fixed accordingly to those of the real classification procedure. They were then analyzed
204 as described above.

205

206 **Results**

207

208 *S. thermophilus* N4L was cultivated in two YE-based media in 1-liter bioreactors. The peptide content
209 of the culture supernatants was monitored during growth using mass spectrometry. Peptide identification
210 was performed on the initial medium before inoculation (t = 0 h) and then each hour from 3 to 6 hours
211 of growth, corresponding to the exponential and the early stationary growth phases (Figure 2). For

212 simplification reasons, the results will be presented sequentially. First, they will be detailed on only one
213 YE (YE1), and the results associated to YE2 will be presented more succinctly thereafter.

214

215 YE1 contains a large number of peptides with different levels of abundance

216 Between about 1,300 and 1,700 distinct peptides were identified in YE1 per analyzed time point from
217 approximately 1,900 to 2,600 fragmentation spectra, depending on the point considered (Table 2). These
218 values were consistent within biological repetitions as indicated by the low coefficients of variation
219 (average variation around the mean of 7% both at peptide and spectra levels), showing a good
220 reproducibility in terms of number of identified peptides. Nevertheless, the qualitative identification of
221 peptides was not as effective, as, for a given time, only an average of 55% of the peptides was identified
222 in all three repetitions. This confirms that non-tryptic peptide identification in complex mixtures is still
223 technically challenging, as already discussed ((47)). Combining all the identifications from all time
224 points and repetitions resulted in a total of 4,598 distinct peptides identified (FDR < 1%) from 32,920
225 fragmentation spectra during the course of the growth.

226 To estimate peptide relative abundance, spectral counting is considered as the simplest method
227 in a labeling-free approach (55, 56). Table 3 shows that the majority of peptides (close to 80%) were
228 actually scarcely identified with only one spectrum per peptide. Even though label-free mass
229 spectrometry only allows relative semi-quantification, these peptides are likely to be either the less
230 abundant ones, or to have poor yields of detection. Nevertheless, some peptides generated larger
231 numbers of spectra. In particular, the top ones (more than 10 spectra per peptide) represented less than
232 1% of the identified peptides in each sample but 5% to 10% of the total number of spectra. Thus, they
233 are likely to be the most abundant in the medium. All these data provide evidence that YE1 contains a
234 high peptide diversity with few of them being over-abundant.

235

236 Peptide physicochemical properties

237 Each identified peptide was characterized by a combination of nine physicochemical properties. These
238 properties are summarized in Table 1 and have been chosen to comprehensively describe YE peptide
239 diversity. The internal distributions of each property calculated from the initial medium peptidome
240 (before inoculation, $t = 0$ h) are represented in Figure 3. No significant difference could be detected (p -
241 value ≤ 0.01) between the three repetitions, and the other time points of analysis showed good
242 reproducibility as well (Figure S1).

243 Most of the nine physicochemical properties of the peptides initially identified in YE1 were
244 distributed within a relatively large range, reflecting a large physicochemical diversity. The detected
245 peptides were mostly hydrophilic, had a median net charge slightly superior to zero, and their average
246 bulkiness was close to 14 \AA^2 which is moderately inferior to that of the mean of the 20 standard amino
247 acids, ca. 15.4 \AA^2 (57) . This last finding suggests a slight over-representation in the yeast-derived
248 peptide sequences of relatively small residues. Finally, these peptides showed an average length of 10

249 residues, although this observation has to be tempered by the limits of the analytical pipeline. Indeed,
250 the upper length limit was driven by the purification process employed and more specifically by the 3
251 kDa ultrafiltration steps, while the lower limit - no detection of peptides shorter than 6 residues - was a
252 direct consequence of the chosen mass spectrometry range detection (350-1,600 m/z).

253

254 Identification of peptide kinetic dynamics

255 YE1 is composed of a large number of peptides displaying various physicochemical properties. This
256 inherent diversity made it suitable to study peptide utilization dynamics during the strain growth.
257 However, as previously described, only a very small number of identified peptides showed high levels
258 of relative abundance, as determined by their spectral counting data (Table 3). These few abundant
259 peptides were therefore the only ones on which the study of abundance variations during growth was
260 achievable. However, their number being too low, it was not possible to extrapolate clear general rules
261 regarding potential peptide utilization patterns. Likewise, the global evolution of peptide
262 physicochemical properties during growth did not reveal any clear trend (Figure S1). Therefore, in order
263 to identify which peptides were utilized and on which physicochemical basis, a specific analytical
264 workflow was developed. The underlying idea was to pool peptides showing close physicochemical
265 properties into groups in order to combine their spectral counts and therefore perform the study not with
266 individual peptides but on a larger scale (see the Material and methods section for explanations about
267 the grouping procedure). The limits of each interval chosen during the grouping were fixed according
268 to the internal distribution of each physicochemical property established from the initial peptidome ($t =$
269 0 h) before inoculation (see Figure 3 for the representation of these intervals and the general rules
270 regarding their constitution). Theoretically, there are $3^9 = 19,683$ different possible classes or
271 “barcodes”. Practically, not all are physically possible or exist biologically and experimentally only
272 1,308 were identified here. Of this total, 612 classes (47%) contained only one peptide, whereas the top
273 three most abundant classes pooled 41, 45 (two *ex aequo* classes) and 53 peptides, respectively.

274 After determining their relative abundance by summing the spectral counts of their constitutive
275 peptides, these classes were submitted to statistical analyses classically used with regular proteomic data
276 (Figure 1C). A first filter was applied to remove all classes showing low levels of abundance (less than
277 5 spectra per class) as their quantification over time would not be relevant. At this step, 1,040 classes
278 were discarded, i.e. 80% of the total. Then, a second filter was used on the 268 remaining abundant
279 classes in order to remove those considered largely constant, i.e. showing less than 50% of variation
280 between their minimal and maximal average abundance values in the different samples. A total of 45
281 classes (3%) matched this description and were not included in the following differential analysis. An
282 analysis of variance was finally performed on the 223 remaining abundant classes that showed sufficient
283 variation amplitude.

284 A total of 49 classes were declared as varying significantly over time at the threshold of an
285 adjusted p -value ≤ 0.01 . Their kinetic profiles are depicted in Figure 4. These profiles can be roughly

286 classified into 3 different groups: classes showing an unambiguous decrease (36 classes) or increase (2
287 classes) over time, and classes whose time-course evolution was fluctuating (11 classes). However, for
288 understanding bacterial physiology, it is of interest to characterize not only peptides that are utilized by
289 the strain but also those which are not and that are left aside in the external medium. On this basis, the
290 former 45 constant classes that were discarded during the second filtering procedure were therefore
291 reintegrated with the 49 others for the last part of the analysis. In total, 94 classes were selected and can
292 thus be classified into 4 different profiles whose main characteristics are given in Table 4. They enclosed
293 a total of 1,060 different peptides, which represents about 23% of the total identified peptides (4,598),
294 but enclosed about a third (10,412) of the total spectra (32,920) assigned during the whole experiment.
295 The remaining classes corresponded to the combination of low abundance classes and abundant ones
296 whose variations were not detected as significant.

297

298 Linking kinetic profiles to peptide composition

299 In an attempt to correlate those different kinetic dynamics with the physicochemical properties used to
300 design the former selected classes, a principal component analysis (PCA) was performed (Figure 5A).
301 The two first components explained more than 60% of the total inertia. At the exception of peptide
302 bulkiness and proline content, all other peptide properties were well represented with correlation
303 coefficients $> |0.6|$ on these axes. In order of importance, the first axis separated the peptides according
304 to their polarity, acidic content, charge, hydrophobicity and, to a lesser extent, peptide bulkiness. The
305 second axis mostly wore the information of the basic and aromatic content as well as peptide length.
306 Proline frequency was equally supported by both axes. Important and natural correlation links can be
307 noticed, for instance between peptide bulkiness and GRAVY index. Indeed, the most hydrophobic
308 residues (I, L and V) are also among the bulkiest (57). Similarly, polarity was anti-correlated with
309 GRAVY index and positively linked with peptide acidic content. Finally, longer peptides were
310 associated with a higher content in acidic residues, and *vice-versa*. This representation allowed the
311 segregation of the four kinetic profiles previously identified (increasing, decreasing, constant,
312 fluctuating).

313 Peptides belonging to both constant and fluctuating profiles were spread in wide and partially
314 overlapping areas, reflecting a common high physicochemical diversity. In contrast, decreasing and
315 increasing profiles were located in two specific and distinct zones. Strikingly, decreasing profile
316 corresponded to exclusively positively charged peptides (Figure 5B) that were also significantly shorter
317 ($p\text{-val} \leq 0.01$, median length of 8 amino acids). The positive charge was the result of both a higher
318 amount of basic residues and a lower fraction of acidic ones compared to that of other profiles.
319 Consistently, these peptides were less polar, and they also contained a higher proportion of hydrophobic
320 residues. Their aromatic and proline content was more variable and did not seem to constitute a good
321 discriminative factor. In contrast, increasing profile was made up of exclusively negatively charged
322 peptides that were also significantly longer (median length of 11 amino acids) and contained a higher

323 amount of proline (median content = 18%). Finally, peptides enclosed in constant and fluctuating
324 profiles displayed intermediate distributions regarding their length and net charge. Their overall
325 hydrophobicity was not significantly different to that of increasing profile, and both displayed low and
326 similar proline content.

327 The validity of this physicochemical class-based analysis was also assessed by the creation of
328 random classes, as described in the Material and Methods section. A 10-step validation process was
329 performed, i.e. 10 different sets of random classes were created and analyzed as detailed above. The
330 detection of kinetic variations was severely impaired in each iteration as a consequence of randomizing
331 peptide grouping (see Figure S2 for more details). This demonstrates that the dynamics identified
332 through our approach were not merely accidental but can be linked to peptide physicochemical
333 specificities.

334

335 Application to YE2

336 The exact same procedure was applied to YE2 to evaluate the robustness of our conclusions on peptide
337 utilization by *S. thermophilus*. Briefly, YE2 contrasts from YE1 by a lower peptide diversity. On
338 average, about 1,200 peptides (1,500 spectra) were detected by time point (see Table S1 for more
339 information), which is about 20% inferior to the number of identifications in YE1. Similarly, over the
340 whole time-course of the experiment, a total of 3,778 peptides (22,900 spectra) were identified (FDR <
341 3%), from which 32% (1215 peptides) were also detected in YE1. The physicochemical properties of
342 the peptides, although close to those of YE1, were not exactly identical (Figure S3). In particular, YE2
343 peptides displayed a higher average amount of acidic residues ($p\text{-val} \leq 0.01$), resulting in an inferior
344 global net charge and higher polarity. They were also bulkier, more hydrophilic and contained less
345 proline than YE1 peptides. YE2 peptides were barcoded as previously described and pooled into 1125
346 classes. On this total and after filtering procedures, 22 classes were detected as varying significantly
347 over time (20 decreasing, 1 increasing, 1 fluctuating, Figure S4) to which were then added 43 constant
348 classes. In total, these 65 classes enclosed 952 peptides, representing 25% of the total peptides (32% of
349 total spectra) identified during the experiment time-course, which is similar to the equivalent analyzed
350 fraction in YE1.

351 The projection in the PCA plane of these selected peptides revealed both consistent but also
352 specific dynamics compared to that of YE1 (Figure 6). More specifically, peptides belonging to the
353 decreasing profile were clustered in two distinct parts of the graph. The first group (dark green, 15
354 classes, 222 peptides) is the counterpart to the decreasing profile observed in YE1 as their constitutive
355 peptides display very similar physicochemical properties, in particular a strict positive net charge
356 (Figure 5B and 6B). The sole noticeable difference between them is the higher presence of short
357 hydrophobic peptides in YE2 (bottom right part of the PCA, Figure 6A). However, peptides of
358 completely different physicochemical profile were also determined as decreasing in YE2 (light green, 5
359 classes, 93 peptides). These peptides were significantly longer, more hydrophilic, and were exclusively

360 negatively charged. Finally, as in YE1, constant profile enclosed a large number of peptides displaying
361 various physicochemical properties. In contrast, fluctuating and increasing profiles only held 9 and 3
362 peptides, respectively, which is largely inferior to the equivalent profiles in YE1 (84 and 50 peptides).

363

364 **Discussion**

365

366 By using a mass spectrometry-based approach coupled with appropriate statistical tools, we were able
367 to shed light on the peptide content of yeast extract-based fermentation media, but also identify on a
368 large scale distinct patterns of peptide abundance variations during the growth of the widely used lactic
369 acid bacterium *Streptococcus thermophilus*. Both YEs used in this study displayed a high peptide
370 diversity with about 4,000 distinct peptides identified in both media. They possibly contain even more
371 peptides as the identification of some of them is still technically challenging (47, 58). Furthermore, our
372 approach allowed the detection of significant differences in terms of peptide composition between
373 analyzed YEs. The number of identified peptides was also large enough with an appropriate
374 physicochemical diversity to enable a robust analysis of peptide utilization by *S. thermophilus* N4L. By
375 pooling peptides into physicochemical classes, we were able i) to identify consistent kinetic profiles,
376 and ii) to compensate partially for the overall low relative abundance levels of individual peptides. As
377 this grouping procedure was based on peptide physicochemical properties, which are known to be
378 leading factors for bacterial peptide use, the temporal evolutions observed in selected classes can
379 reasonably be considered to mainly reflect peptide utilization dynamics of the strain N4L.

380 Four relevant kinetic profiles of peptide utilization have been observed: stagnation, decrease,
381 increase and fluctuation of spectral counts over time. These patterns might come from two plausible
382 origins: transport inside the cells, and/or peptide cleavage mediated by an extracellular hydrolytic
383 activity. This last case is especially suggested by the increasing profile observed. Indeed, as the
384 fermentation was performed in batch mode, the most likely explanation is that some peptides must have
385 been gradually hydrolyzed by the strain into smaller fragments. These fragments can share the same
386 barcodes as other peptides of the initial medium. Some of them are likely to be used by the strain, and
387 do not accumulate in large amounts in the medium. Some others are not, and progressively accumulate
388 in the external medium during the growth. This hypothesis is supported by the fact that, in YE1,
389 increasing classes are constituted of numerous scarce peptides, many of which were only detected in the
390 latter stages of fermentation. The cell-envelope located proteinase PrtS is the most plausible effector of
391 this increase, but it may not be the only one. The membrane-anchored protease HtrA, or another cell
392 surface activity that has not been yet characterized (47, 59, 60) could also be responsible for this
393 increase. Therefore, the observed peptide dynamics can be explained as follows: i) decrease: transport
394 and/or cleavage of initially present peptides; ii) increase: accumulation of cleavage products at a higher
395 rate than their transport (if any transport); iii) stagnation: neither transport/cleavage nor accumulation,
396 or both at similar rates; iv) fluctuating profile: combination of transport/cleavage and accumulation

397 within the same physicochemical class at various changing rates over time, or artefactual noise (peptide
398 identification variability).

399 In that respect, the consistent presence in both YEs of a large group of decreasing basic peptides
400 is noteworthy, and it is sensible to assume that this decrease is predominantly the consequence of
401 transport. As a first reason, the decrease depended on the global physicochemical property of the
402 peptides, and not on their amino acid sequence and order as is the case for the activity of serine-proteases
403 such as PrtS and HtrA (61, 62). Then, the main conclusion of our study is that this profile is firstly linked
404 to a systematic presence of a global positive net charge combined with a significantly shorter length and
405 a higher proportion of hydrophobic residues. This description perfectly matches that of the previously
406 mentioned study performed with a protease-negative strain on a small number of milk-derived peptides
407 (40). Our work, by relying on a vastly higher number of peptides in two different substrates, consolidates
408 these former results but also suggests a dominant role of a positive net charge for peptide transport. It is
409 thus reasonable to assume that most, if not all, of these decreasing peptides were actually preferentially
410 consumed by the strain and transported inside the cells.

411 This transport was mediated by the Ami system which is the only oligopeptide carrier identified
412 in the strain. It has been demonstrated in *L. lactis* that the oligopeptide-binding protein (OppA)
413 determined the overall peptide specificity of its cognate transporter (36). Similarly, in the case of *S.*
414 *thermophilus* N4L, the consumption of positively charged peptides is likely to be dictated by its own
415 oligopeptide-binding proteins, namely AmiA1 and AmiA3. Moreover, it has been formerly established
416 *in vivo* that peptide transport in both species displayed very similar specificities (12, 40). Indeed, it was
417 shown that *L. lactis* also preferentially used hydrophobic basic peptides ranging between 600 and 1,100
418 Da, although this bacterium can accommodate surprisingly long peptides up to 35 residues (37), which
419 is even longer to the maximal size (24 residues) observed with *S. thermophilus* (27). The ability of *L.*
420 *lactis* to carry various sizes of peptides and its preference for peptides containing branched-chain amino
421 acids - in particular isoleucine - were explained later on thanks to the crystallization of OppA (31, 32).
422 However, an apparent discrepancy remains in the literature between *in vivo* studies and structural
423 characterization concerning the role of peptide net charge in *L. lactis* OppA-based selection. If this factor
424 was determined *in vivo* as a major feature for transport (12), it has not been identified as coming into
425 play regarding binding mechanisms in structural data. In that perspective, it noteworthy that the crystal
426 structure of unliganded *E. coli* OppA binding site had revealed a negatively charged surface responsible
427 for the preferential binding of basic peptides (63). This finding was subsequently found to apply as well
428 for *S. typhimurium* OppA. Therefore, further work is needed to elucidate the role of peptide charge both
429 in *L. lactis* and *S. thermophilus*, as corroborating evidence seems to indicate that this factor may be a
430 widespread requisite feature for peptide transport.

431 Finally, it is striking that all these conclusions are in complete opposition with the
432 physicochemical properties of the second group of decreasing peptides specifically identified in YE2,
433 as these were longer hydrophilic and anionic peptides. Therefore, the hypothesis of a decrease due to

434 peptide transport appears unlikely. A more convincing explanation would be a decrease due to an
435 extracellular hydrolysis by the cell-wall anchored proteinase PrtS or by another actor. As this dynamics
436 was only observable in YE2, this might reflect a difference of concentration between both substrates in
437 acidic peptides. In any case, regardless of its origin, this second decrease profile shows that some
438 different peptide dynamics were at work between both YEs. We can thus question the biological impact
439 of this extra class of decreasing peptides and whether it constitutes an additional supply of nutritional
440 peptides for the strain in YE2, either directly through peptide transport or indirectly by the release of
441 peptide fragments of higher nutritional value. Although the strain showed very close growth
442 performance in both media (Figure 2), a slightly higher growth rate is noticeable in YE2, and
443 enumeration on Petri dishes notably revealed an average two-fold difference of final biomass in favors
444 of YE2 (data not shown). Nonetheless, as things stand, the impact of this supplementary peptide
445 dynamics remains an open question.

446 In conclusion, the identification of complex mixtures of peptides by mass spectrometry,
447 although still technically challenging, is progressively gaining attention (58) and is proving to be an
448 excellent exploratory approach to unravel the peptide content of complex media as well as study the
449 diverse oligopeptide utilization patterns of a bacterial species during its growth. Combined with
450 complementary approaches, it opens avenues for further characterization and optimization of protein
451 hydrolysate-based culture media and could also be used to deepen our knowledge of bacterial
452 physiology. In our context, the confirmation of the suspected systems responsible of the dynamics
453 identified (Ami, PrtS etc.) could be achieved *via* the construction of mutant strains that would be then
454 submitted to the same peptidomic analyses. Likewise, it would be interesting to explore whether, by
455 modulating the peptide composition of both YEs (enrichment or depletion of certain families of
456 peptides), we can modify the dynamics formerly observed as well as the growth yields.

457

458 **Acknowledgments**

459

460 We thank the INRA PAPPSO proteomics platform (<http://pappso.inra.fr>) for providing mass
461 spectrometry facilities, and Eloi Haudebourg and Sophie Liuu for performing the mass spec
462 experiments. We also thank Mylène Boulay for her support and technical expertise. Finally, we also
463 thank Simon Poirier, Rozenn Gardan and Françoise Rul for their critical reading.

464

465 **Funding information**

466

467 This project was funded by Association Nationale de la Recherche et de la Technologie (ANRT).

468 **References**

469

- 470 1. Hill D, Sugrue I, Arendt E, Hill C, Stanton C, Ross RP. 2017. Recent advances in microbial
471 fermentation for dairy and health. *F1000Res* 6.
- 472 2. Sauer M, Russmayer H, Grabherr R, Peterbauer CK, Marx H. 2017. The Efficient Clade: Lactic
473 Acid Bacteria for Industrial Chemical Production. *Trends Biotechnol* 35:756–769.
- 474 3. Othman M, Ariff AB, Rios-Solis L, Halim M. 2017. Extractive Fermentation of Lactic Acid in
475 Lactic Acid Bacteria Cultivation: A Review. *Front Microbiol* 8.
- 476 4. Izawa N, Sone T. 2014. Cosmetic Ingredients Fermented by Lactic Acid Bacteria, p. 233–242. *In*
477 *Microbial Production*. Springer, Tokyo.
- 478 5. Salque M, Bogucki PI, Pyzel J, Sobkowiak-Tabaka I, Grygiel R, Szmyt M, Evershed RP. 2013.
479 Earliest evidence for cheese making in the sixth millennium BC in northern Europe. *Nature*
480 493:522–525.
- 481 6. Christensen JE, Dudley EG, Pederson JA, Steele JL. 1999. Peptidases and amino acid catabolism
482 in lactic acid bacteria. *Antonie Van Leeuwenhoek* 76:217–246.
- 483 7. Kunji ER, Mierau I, Hagting A, Poolman B, Konings WN. 1996. The proteolytic systems of
484 lactic acid bacteria. *Antonie Van Leeuwenhoek* 70:187–221.
- 485 8. Ummadi M (Soni), Curic-Bawden M. 2008. Use of Protein Hydrolysates in Industrial Starter
486 Culture Fermentations, p. 91–114. *In* Pasupuleti, VK, Demain, AL (eds.), *Protein Hydrolysates*
487 *in Biotechnology*. Springer Netherlands.
- 488 9. Pasupuleti VK, Braun S. 2008. State of the Art Manufacturing of Protein Hydrolysates, p. 11–
489 32. *In* Pasupuleti, VK, Demain, AL (eds.), *Protein Hydrolysates in Biotechnology*. Springer
490 Netherlands.
- 491 10. Helinck S, Charbonnel P, Foucaud-Scheunemann C, Piard J-C, Juillard V. 2003. Charged
492 casein-derived oligopeptides competitively inhibit the transport of a reporter oligopeptide by
493 *Lactococcus lactis*. *J Appl Microbiol* 94:900–907.
- 494 11. Juillard V, Le Bars D, Kunji ER, Konings WN, Gripon JC, Richard J. 1995. Oligopeptides are
495 the main source of nitrogen for *Lactococcus lactis* during growth in milk. *Appl Environ*
496 *Microbiol* 61:3024–3030.

- 497 12. Juillard V, Guillot A, Le Bars D, Gripon JC. 1998. Specificity of milk peptide utilization by
498 *Lactococcus lactis*. *Appl Environ Microbiol* 64:1230–1236.
- 499 13. Kunji ER, Hagting A, De Vries CJ, Juillard V, Haandrikman AJ, Poolman B, Konings WN.
500 1995. Transport of beta-casein-derived peptides by the oligopeptide transport system is a crucial
501 step in the proteolytic pathway of *Lactococcus lactis*. *J Biol Chem* 270:1569–1574.
- 502 14. Mosser M, Kapel R, Aymes A, Bonanno L-M, Olmos E, Besançon I, Druaux D, Chevalot I,
503 Marc I, Marc A. 2011. Characterization of chromatographic yeast extract fractions promoting
504 CHO cell growth. *BMC Proceedings* 5:P99.
- 505 15. Mosser M, Kapel R, Chevalot I, Olmos E, Marc I, Marc A, Oriol E. 2015. Fractionation of yeast
506 extract by nanofiltration process to assess key compounds involved in CHO cell culture
507 improvement. *Biotechnol Prog* 31:875–882.
- 508 16. Spearman M, Lodewyks C, Richmond M, Butler M. 2014. The bioactivity and fractionation of
509 peptide hydrolysates in cultures of CHO cells. *Biotechnol Prog* 30:584–593.
- 510 17. Spearman M, Chan S, Jung V, Kowbel V, Mendoza M, Miranda V, Butler M. 2016.
511 Components of yeast (*Saccharomyces cerevisiae*) extract as defined media additives that support
512 the growth and productivity of CHO cells. *J Biotechnol* 233:129–142.
- 513 18. Kevvai K, Kütt M-L, Nisamedtinov I, Paalme T. 2014. Utilization of (15)N-labelled yeast
514 hydrolysate in *Lactococcus lactis* IL1403 culture indicates co-consumption of peptide-bound and
515 free amino acids with simultaneous efflux of free amino acids. *Antonie Van Leeuwenhoek*
516 105:511–522.
- 517 19. Kevvai K, Kütt M-L, Nisamedtinov I, Paalme T. 2016. Simultaneous utilization of ammonia,
518 free amino acids and peptides during fermentative growth of *Saccharomyces cerevisiae*. *Journal*
519 *of the Institute of Brewing* 122:110–115.
- 520 20. Savijoki K, Ingmer H, Varmanen P. 2006. Proteolytic systems of lactic acid bacteria. *Appl*
521 *Microbiol Biotechnol* 71:394–406.
- 522 21. Hagting A, vd Velde J, Poolman B, Konings WN. 1997. Membrane topology of the di- and
523 tripeptide transport protein of *Lactococcus lactis*. *Biochemistry* 36:6777–6785.
- 524 22. Hagting A, Kunji ER, Leenhouts KJ, Poolman B, Konings WN. 1994. The di- and tripeptide
525 transport protein of *Lactococcus lactis*. A new type of bacterial peptide transporter. *J Biol Chem*
526 269:11391–11399.

- 527 23. Nakajima H, Hagting A, Kunji ER, Poolman B, Konings WN. 1997. Cloning and functional
528 expression in *Escherichia coli* of the gene encoding the di- and tripeptide transport protein of
529 *Lactobacillus helveticus*. *Appl Environ Microbiol* 63:2213–2217.
- 530 24. Goh YJ, Goin C, O’Flaherty S, Altermann E, Hutkins R. 2011. Specialized adaptation of a lactic
531 acid bacterium to the milk environment: the comparative genomics of *Streptococcus*
532 *thermophilus* LMD-9. *Microb Cell Fact* 10 Suppl 1:S22.
- 533 25. Hols P, Hancy F, Fontaine L, Grossiord B, Prozzi D, Leblond-Bourget N, Decaris B, Bolotin A,
534 Delorme C, Dusko Ehrlich S, Guédon E, Monnet V, Renault P, Kleerebezem M. 2005. New
535 insights in the molecular biology and physiology of *Streptococcus thermophilus* revealed by
536 comparative genomics. *FEMS Microbiology Reviews* 29:435–463.
- 537 26. Tynkkynen S, Buist G, Kunji E, Kok J, Poolman B, Venema G, Haandrikman A. 1993. Genetic
538 and biochemical characterization of the oligopeptide transport system of *Lactococcus lactis*. *J*
539 *Bacteriol* 175:7523–7532.
- 540 27. Garault P, Le Bars D, Besset C, Monnet V. 2002. Three oligopeptide-binding proteins are
541 involved in the oligopeptide transport of *Streptococcus thermophilus*. *J Biol Chem* 277:32–39.
- 542 28. Fang G, Konings WN, Poolman B. 2000. Kinetics and substrate specificity of membrane-
543 reconstituted peptide transporter DtpT of *Lactococcus lactis*. *J Bacteriol* 182:2530–2535.
- 544 29. Martinez Molledo M, Quistgaard EM, Flayhan A, Pieprzyk J, Löw C. 2018. Multispecific
545 Substrate Recognition in a Proton-Dependent Oligopeptide Transporter. *Structure* 26:467–
546 476.e4.
- 547 30. Solcan N, Kwok J, Fowler PW, Cameron AD, Drew D, Iwata S, Newstead S. 2012. Alternating
548 access mechanism in the POT family of oligopeptide transporters. *The EMBO Journal* 31:3411–
549 3421.
- 550 31. Berntsson RP-A, Doeven MK, Fusetti F, Duurkens RH, Sengupta D, Marrink S-J, Thunnissen
551 A-MWH, Poolman B, Slotboom D-J. 2009. The structural basis for peptide selection by the
552 transport receptor OppA. *EMBO J* 28:1332–1340.
- 553 32. Berntsson RP-A, Thunnissen A-MWH, Poolman B, Slotboom D-J. 2011. Importance of a
554 hydrophobic pocket for peptide binding in lactococcal OppA. *J Bacteriol* 193:4254–4256.

- 555 33. Charbonnel P, Lamarque M, Piard J-C, Gilbert C, Juillard V, Atlan D. 2003. Diversity of
556 oligopeptide transport specificity in *Lactococcus lactis* species. A tool to unravel the role of
557 OppA in uptake specificity. *J Biol Chem* 278:14832–14840.
- 558 34. Detmers FJ, Kunji ER, Lanfermeijer FC, Poolman B, Konings WN. 1998. Kinetics and
559 specificity of peptide uptake by the oligopeptide transport system of *Lactococcus lactis*.
560 *Biochemistry* 37:16671–16679.
- 561 35. Detmers FJ, Lanfermeijer FC, Abele R, Jack RW, Tampe R, Konings WN, Poolman B. 2000.
562 Combinatorial peptide libraries reveal the ligand-binding mechanism of the oligopeptide receptor
563 OppA of *Lactococcus lactis*. *Proc Natl Acad Sci USA* 97:12487–12492.
- 564 36. Doeven MK, Abele R, Tampé R, Poolman B. 2004. The binding specificity of OppA determines
565 the selectivity of the oligopeptide ATP-binding cassette transporter. *J Biol Chem* 279:32301–
566 32307.
- 567 37. Doeven MK, Kok J, Poolman B. 2005. Specificity and selectivity determinants of peptide
568 transport in *Lactococcus lactis* and other microorganisms. *Mol Microbiol* 57:640–649.
- 569 38. Kunji ER, Fang G, Jeronimus-Stratingh CM, Bruins AP, Poolman B, Konings WN. 1998.
570 Reconstruction of the proteolytic pathway for use of beta-casein by *Lactococcus lactis*. *Mol*
571 *Microbiol* 27:1107–1118.
- 572 39. Lanfermeijer FC, Detmers FJ, Konings WN, Poolman B. 2000. On the binding mechanism of the
573 peptide receptor of the oligopeptide transport system of *Lactococcus lactis*. *EMBO J* 19:3649–
574 3656.
- 575 40. Juille O, Le Bars D, Juillard V. 2005. The specificity of oligopeptide transport by *Streptococcus*
576 *thermophilus* resembles that of *Lactococcus lactis* and not that of pathogenic streptococci.
577 *Microbiology (Reading, Engl)* 151:1987–1994.
- 578 41. Proust L, Loux V, Martin V, Magnabosco C, Pedersen M, Monnet V, Juillard V. 2018. Complete
579 Genome Sequence of the Industrial Fast-Acidifying Strain *Streptococcus thermophilus* N4L.
580 *Microbiol Res Announc* 7:e01029-18.
- 581 42. Jameh N, Galia W, Awussi AA, Roux E, Genay M, Perrin C, Dary-Mourot A. 2016.
582 Characterization of a new peptide transport system in *Streptococcus thermophilus*. *Food*
583 *Research International* 86:34–45.

- 584 43. Terzaghi BE, Sandine WE. 1975. Improved Medium for Lactic Streptococci and Their
585 Bacteriophages. *Appl Microbiol* 29:807–813.
- 586 44. Letort C, Juillard V. 2001. Development of a minimal chemically-defined medium for the
587 exponential growth of *Streptococcus thermophilus*. *Journal of Applied Microbiology* 91:1023–
588 1029.
- 589 45. Williams WL, Broquist HP, Snell EE. 1947. Oleic acid and related compounds as growth factors
590 for lactic acid bacteria. *Journal of Biological Chemistry* 170:619–630.
- 591 46. Jäpelt KB, Johnsen LG, Christensen JH. 2016. Removal of Polysorbate 80 by complexation prior
592 to LC-MS analysis. *Anal Bioanal Chem* 408:2303–2307.
- 593 47. Guillot A, Boulay M, Chambellon É, Gitton C, Monnet V, Juillard V. 2016. Mass Spectrometry
594 Analysis of the Extracellular Peptidome of *Lactococcus lactis*: Lines of Evidence for the
595 Coexistence of Extracellular Protein Hydrolysis and Intracellular Peptide Excretion. *J Proteome*
596 *Res* 15:3214–3224.
- 597 48. Langella O, Valot B, Balliau T, Blein-Nicolas M, Bonhomme L, Zivy M. 2016.
598 XTandemPipeline: A Tool to Manage Sequence Redundancy for Protein Inference and
599 Phosphosite Identification. *J Proteome Res*.
- 600 49. Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH. 2015.
601 Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data.
602 *Bioinformatics* 31:3356–3358.
- 603 50. Mann HB, Whitney DR. 1947. On a Test of Whether one of Two Random Variables is
604 Stochastically Larger than the Other. *Ann Math Statist* 18:50–60.
- 605 51. Wilcoxon F. 1945. Individual Comparisons by Ranking Methods. *Biometrics Bulletin* 1:80–83.
- 606 52. Kruskal WH, Wallis WA. 1952. Use of Ranks in One-Criterion Variance Analysis. *Journal of*
607 *the American Statistical Association* 47:583–621.
- 608 53. Valot B, Langella O, Nano E, Zivy M. 2011. MassChroQ: a versatile tool for mass spectrometry
609 quantification. *Proteomics* 11:3572–3577.
- 610 54. Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and
611 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B*
612 (Methodological) 57:289–300.

- 613 55. Colinge J, Chiappe D, Lagache S, Moniatte M, Bougueleret L. 2005. Differential proteomics via
614 probabilistic peptide identification scores. *Anal Chem* 77:596–606.
- 615 56. Liu H, Sadygov RG, Yates JR. 2004. A model for random sampling and estimation of relative
616 protein abundance in shotgun proteomics. *Anal Chem* 76:4193–4201.
- 617 57. Zimmerman JM, Eliezer N, Simha R. 1968. The characterization of amino acid sequences in
618 proteins by statistical methods. *J Theor Biol* 21:170–201.
- 619 58. Bingeman TS, Perlman DH, Storey DG, Lewis IA. 2017. Digestomics: an emerging strategy for
620 comprehensive analysis of protein catabolism. *Current Opinion in Biotechnology* 43:134–140.
- 621 59. Hafeez Z, Cakir-Kiefer C, Girardet J-M, Jardin J, Perrin C, Dary A, Miclo L. 2013. Hydrolysis
622 of milk-derived bioactive peptides by cell-associated extracellular peptidases of *Streptococcus*
623 *thermophilus*. *Appl Microbiol Biotechnol* 97:9787–9799.
- 624 60. Hafeez Z, Cakir-Kiefer C, Girardet J-M, Lecomte X, Paris C, Galia W, Dary A, Miclo L. 2015.
625 New Insights into the Proteolytic System of *Streptococcus thermophilus*: Use of Isradicin To
626 Characterize Cell-Associated Extracellular Peptidase Activities. *J Agric Food Chem* 63:7522–
627 7531.
- 628 61. Perona JJ, Craik CS. 1995. Structural basis of substrate specificity in the serine proteases.
629 *Protein Sci* 4:337–360.
- 630 62. Siezen RJ, Leunissen JA. 1997. Subtilases: the superfamily of subtilisin-like serine proteases.
631 *Protein Sci* 6:501–523.
- 632 63. Klepsch MM, Kovermann M, Löw C, Balbach J, Permentier HP, Fusetti F, de Gier JW,
633 Slotboom DJ, Berntsson RP-A. 2011. *Escherichia coli* peptide binding protein OppA has a
634 preference for positively charged peptides. *J Mol Biol* 414:75–85.
- 635 64. Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a
636 protein. *J Mol Biol* 157:105–132.
- 637 65. Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science*
638 185:862–864.
- 639 66. Moore DS. 1985. Amino acid and peptide net charges: A simple calculational procedure.
640 *Biochem Educ* 13:10–11.

642 Table 1: Peptide physicochemical properties and interval division used for peptide barcoding
 643

Peptide property	Details	Interval 1	Interval 2	Interval 3
Length	Number of amino acids	≤ 9	[10; 12]	≥ 13
GRAVY	Grand average of hydrophathy index (64)	≤ -2] -2 ; -0.5]	> -0.5
Bulkiness	Average section of amino acid side chains, in Å ² (57)	≤ 12.6] 12.6 ; 14.7]	> 14.7
Polarity	Average polarity of amino acids (65)	≤ 8.7] 8.7 ; 10.1]	> 10.1
Net charge	Normalized net charge calculated at pH = 6.0 (66)	< 0	[0 ; 1]	> 1
Basic residues	Proportion of lysine, arginine and histidine	= 0] 0 ; 0.12]	> 0.12
Acidic residues	Proportion of aspartic and glutamic acids	= 0] 0 ; 0.12]	> 0.12
Aromatic residues	Proportion of phenylalanine, tryptophan and tyrosine	= 0] 0 ; 0.12]	> 0.12
Proline	Proportion of proline	= 0] 0 ; 0.12]	> 0.12

644
 645
 646
 647
 648

Table 2: Summary of peptide identification in YE1 during growth of *S. thermophilus* N4L.

	t = 0 h	t = 3 h	t = 4 h	t = 5 h	t = 6 h
Peptides ^a	1560 ± 69	1742 ± 63	1514 ± 123	1402 ± 124	1290 ± 107
Total spectra ^b	2294 ± 86	2596 ± 119	2211 ± 157	2010 ± 198	1863 ± 190

649 Means of 3 independent repetitions ± standard deviation

650 ^a Number of distinct peptide sequences identified.

651 ^b Number of peptide spectrum matches. The values are higher than above as some peptides may be
 652 identified repeatedly.

653 Table 3: Average number of peptides identified as a function of their relative abundance
 654

Number of spectra per peptides	t = 0 h	t = 3 h	t = 4 h	t = 5 h	t = 6 h
More than 10	11 ± 0	14 ± 9	9 ± 1	11 ± 2	10 ± 4
From 5 to 9	40 ± 5	44 ± 8	37 ± 4	25 ± 5	27 ± 1
From 2 to 4	272 ± 18	316 ± 5	273 ± 22	252 ± 37	230 ± 40
1	1236 ± 55	1369 ± 60	1195 ± 102	1113 ± 86	1023 ± 67

655 Means of 3 independent repetitions ± standard deviation

656

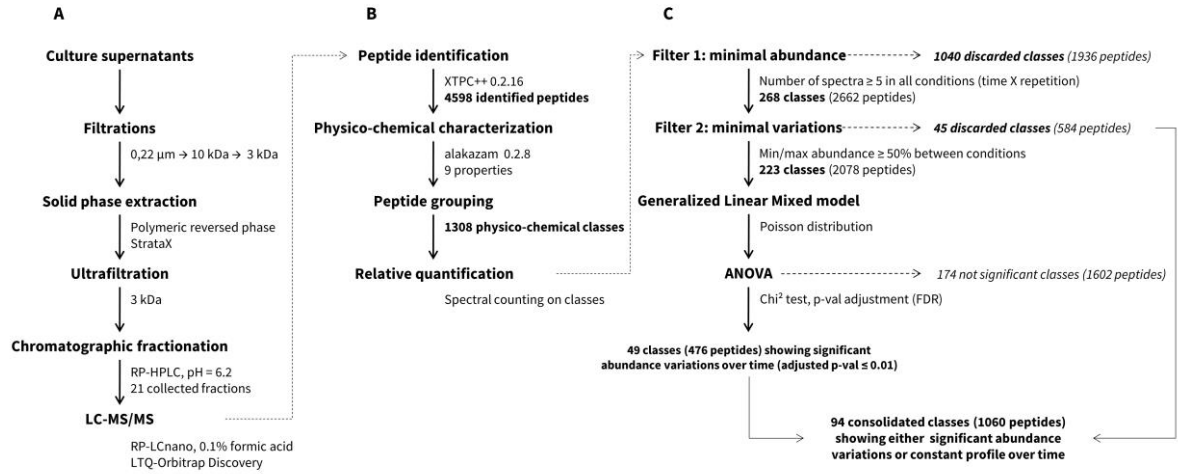
657

658 Table 4: Characteristics of the different kinetic profiles

659

	Decreasing	Increasing	Constant	Fluctuating
Number of classes	36	2	45	11
Total number of peptides	342	50	584	84
Total number of spectra	2785	617	6374	636

660



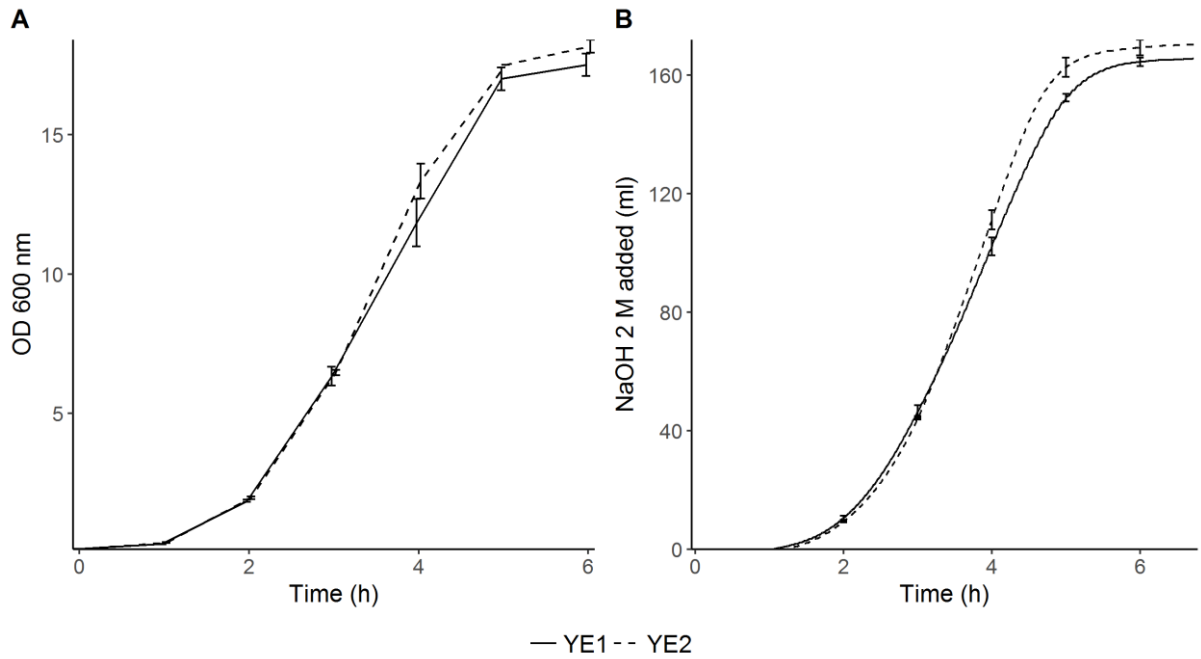
661

662

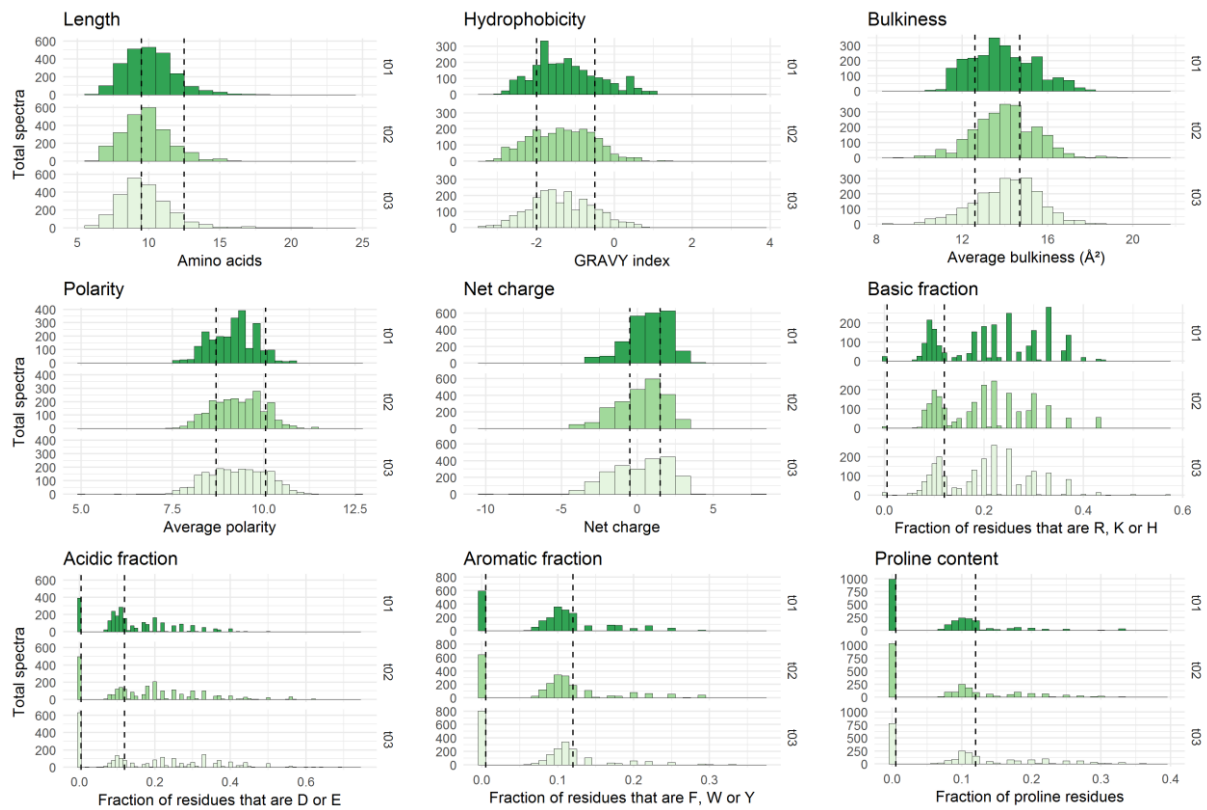
663 Figure 1. General workflow of the peptidomics analytical pipeline, starting from the experimental part

664 (A) to the bioinformatic (B) and statistical analyses (C). Results indicated are from YE1 data.

665 XTPC++: X!Tandem Pipeline (C++)

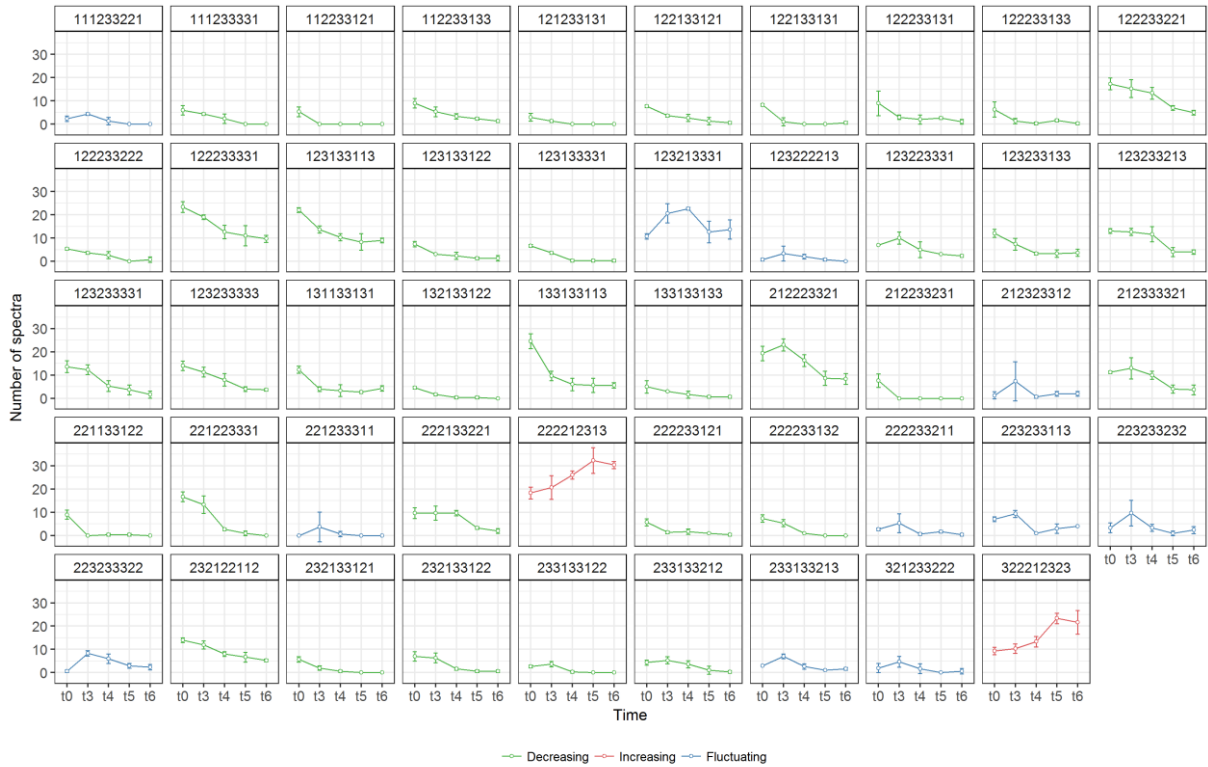


666
 667 Figure 2. Growth of *S. thermophilus* in YE-based medium throughout fermentations performed at pH 6.
 668 (A) Bacterial biomass evaluated by optical density measured each hour at 600 nm. (B) Online
 669 monitoring of NaOH (2 M) addition to the fermentation medium as a consequence of pH regulation.
 670 The values correspond to the means of three independent experiments; errors bars represent the standard
 671 deviations.



672
673

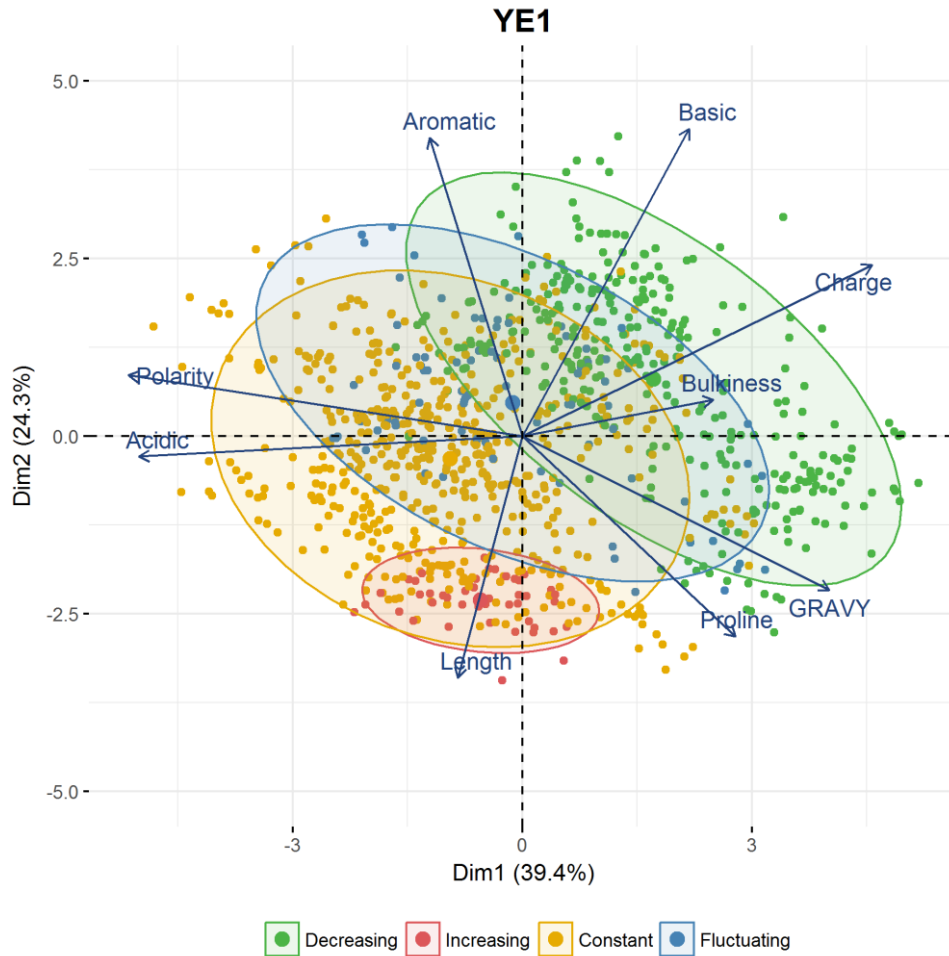
674 Figure 3: Physicochemical properties of the peptides initially present in YE1 before inoculation ($t = 0$
675 h). The three repetitions are represented separately. The x-axes correspond to the scales specific to each
676 physicochemical property. The y-axes represent the number of spectra observed. Vertical dashed lines
677 indicate the limits chosen for the definition of the intervals used for peptide barcoding (Table 1). As
678 peptide length, hydrophobicity, bulkiness, polarity and net charge show a bell-shaped distribution, the
679 limits were chosen so that they framed the profile peak. The remaining properties correspond to
680 frequencies of specific types of residues in the peptide amino acid composition. The first limit defines
681 the "zero sub-class": absence of the considered residues in the peptide composition, and the second limit
682 was arbitrarily fixed at 12%.



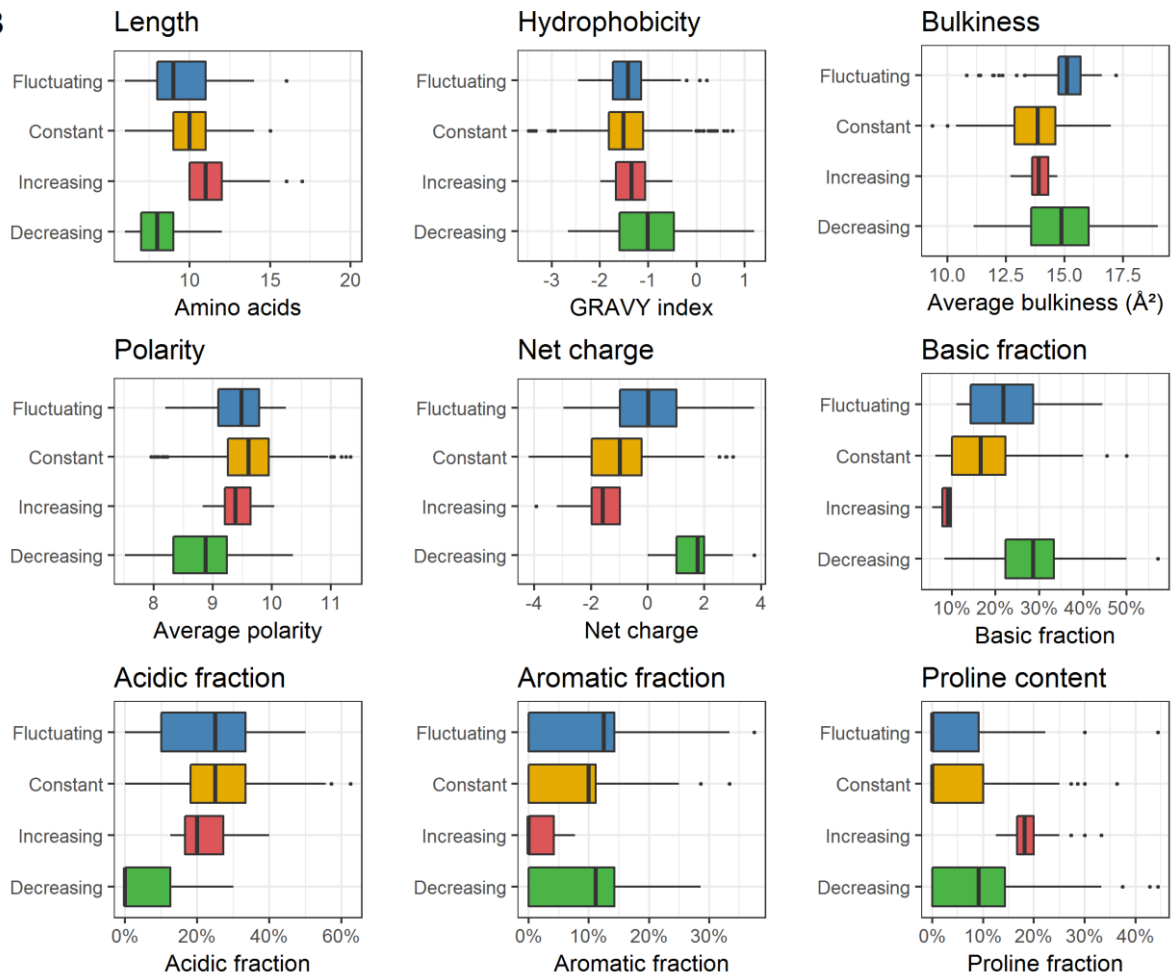
683
 684
 685
 686
 687
 688
 689
 690

Figure 4: Classes of peptides whose population is significantly modified (adjusted p -val ≤ 0.01) in YE1 during growth of *S. thermophilus*. The values correspond to the means of the spectra used to identify the peptides constituting the classes. The error bars represent the standard deviations. In green are classes showing a clear decrease over time, in red an increase over time, and in blue classes whose abundance evolution is fluctuating over time because of successive increase, decrease or stagnation (and all the combinations).

A



B



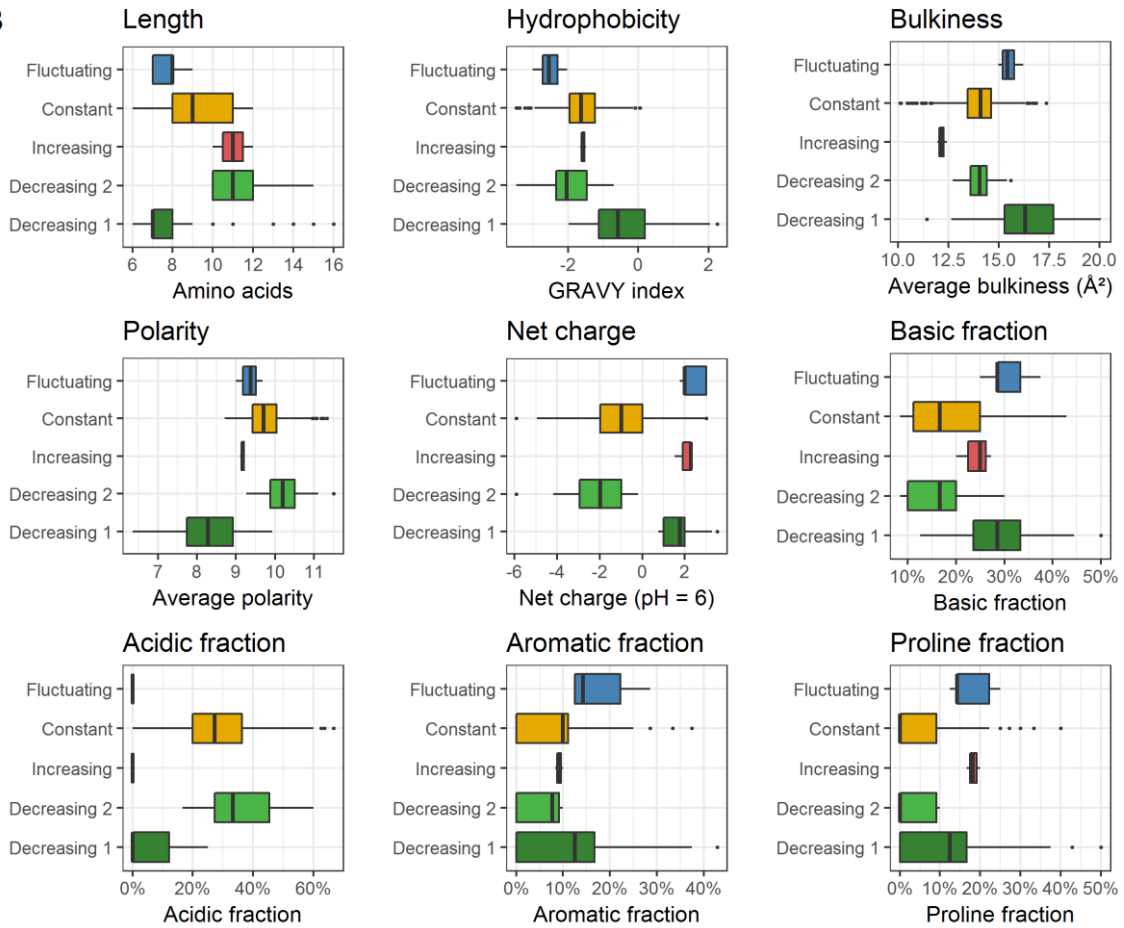
692

693 Figure 5: (A) PCA representation of peptides enclosed within the different kinetic profiles observed in
694 YE1. Each point represents a unique peptide. (B) Physicochemical properties of the peptides constituting
695 the four kinetic profiles identified. Each box plot displays the distribution of the values taken by all
696 unique peptides included in each kinetic profile. The boxes display the range between the first and third
697 quartile of each distribution, and the central bold lines their median value. Individual points are peptides
698 considered as outliers (more than 1.5 times the length of the box away from the box closest extremity).
699

A



B



701 Figure 6: (A) PCA representation of peptides enclosed within the different kinetic profiles observed in
702 YE2. (B) Physicochemical properties of the peptides constituting the four kinetic profiles identified.

Insights into the complexity of yeast extract peptides and their utilization by *Streptococcus thermophilus*

Lucas Proust^{a,c}, Alain Sourabié^b, Martin Pedersen^c, Iris Besançon^b, Véronique Monnet^a and Vincent Juillard^{a*}

^aMicalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France;

^bProcelys, Lesaffre Group, Maisons-Alfort, France;

^cSacco S.r.l., Cadorago, Italy

Supporting information

1 Table S1: Summary of peptide identification in YE2 during growth of *S. thermophilus* N4L.

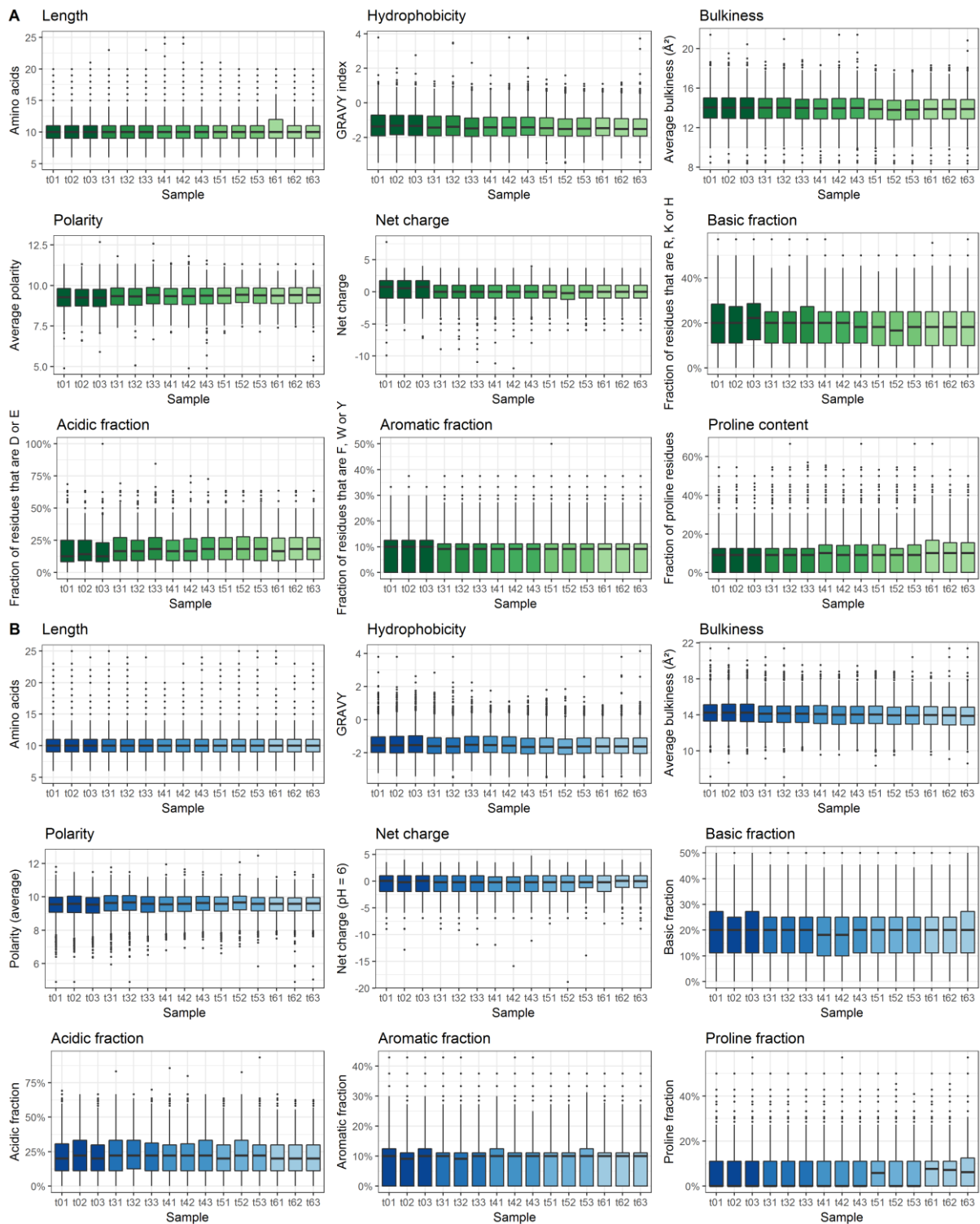
2

	t = 0 h	t = 3 h	t = 4 h	t = 5 h	t = 6 h
Peptides ^a	1384 ± 41	1272 ± 85	1072 ± 16	1050 ± 174	1042 ± 75
Total spectra ^b	1811 ± 40	1622 ± 130	1398 ± 27	1397 ± 243	1406 ± 95

3 Means of 3 independent repetitions ± standard deviation

4 ^a Number of distinct peptide sequences identified.

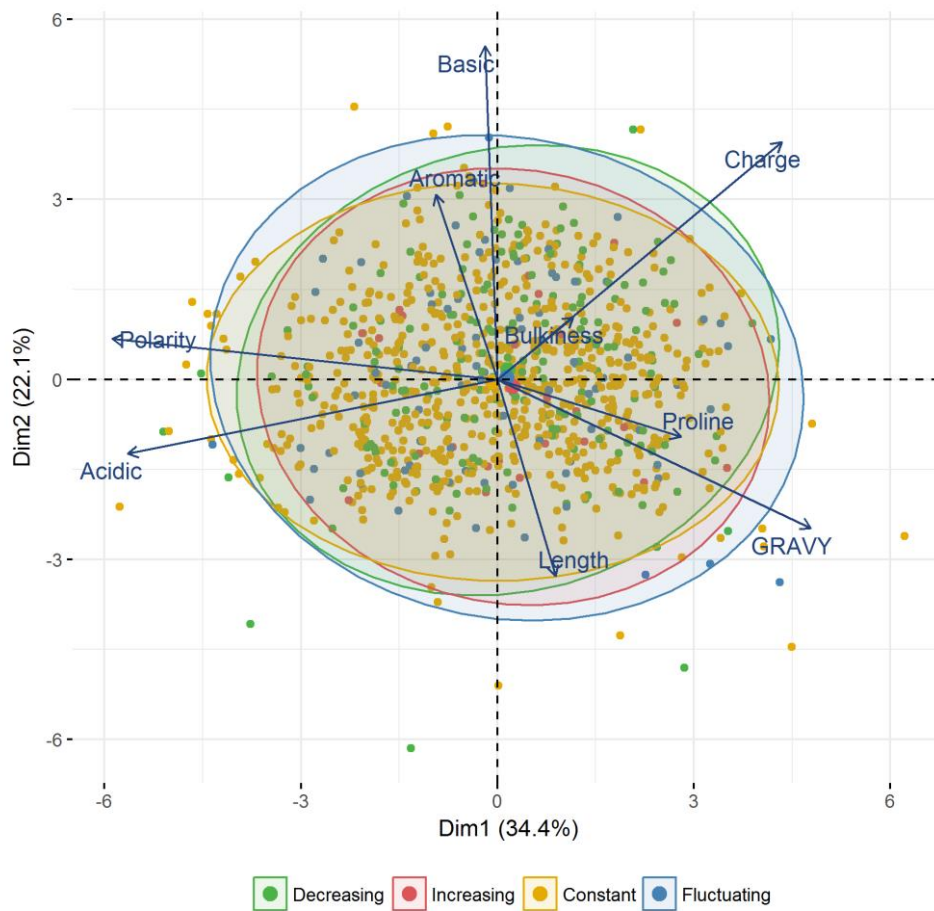
5 ^b Number of peptide spectrum matches. The values are higher than above as some peptides may be
6 identified repeatedly.



7
8

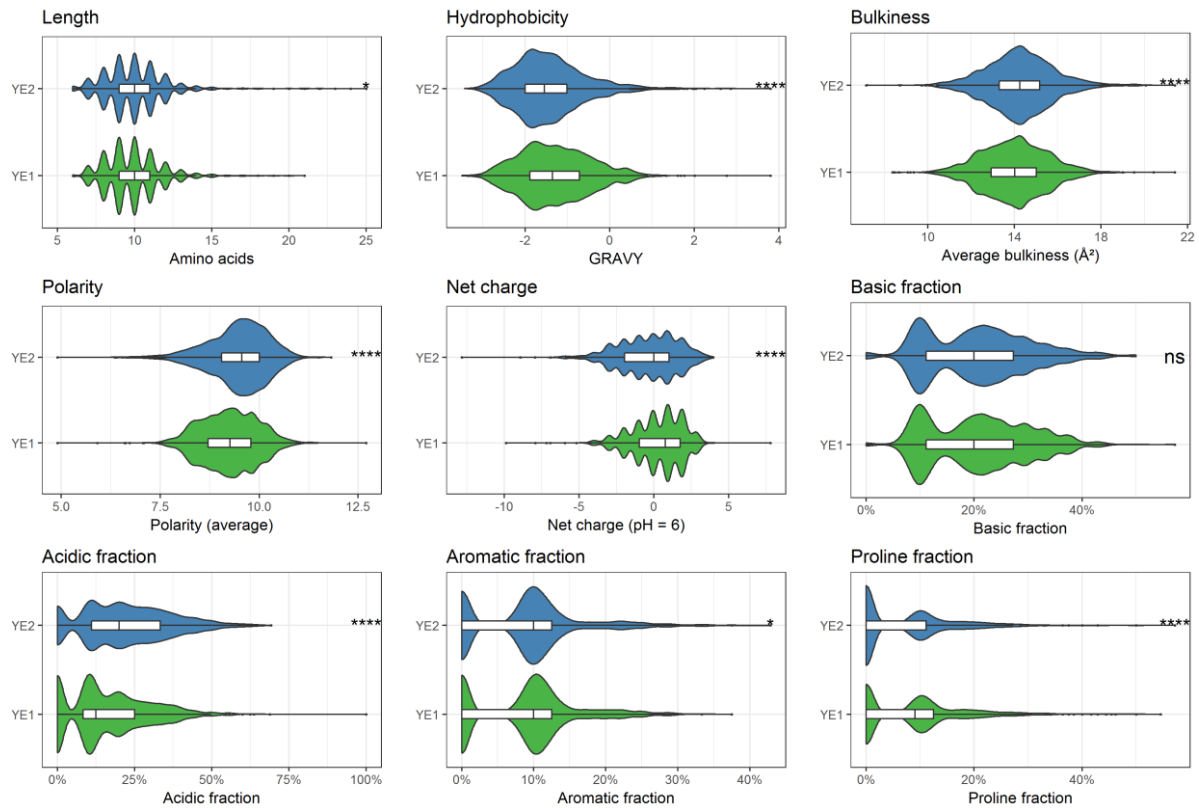
9 Figure S1: Physicochemical properties of the identified peptides in YE1 (A) and YE2 (B), as a function
 10 of the growth course. Each box plot represents the distribution of the total spectra identified in one
 11 repetition, each shade of green or blue one sampling time. From left to right: initial time (0 h), 3 h, 4 h
 12 5 h and 6 h of bacterial growth. All repetitions are displayed as indicated in the x-axes (e.g.: t01
 13 corresponds to the first repetition at t = 0 h). The boxes display the range between the first and third
 14 quartile of each distribution, and the central bold lines their median value. Individual points are peptide

15 spectra considered as outliers (more than 1.5 times the length of the box away from the box closest
16 extremity).



17

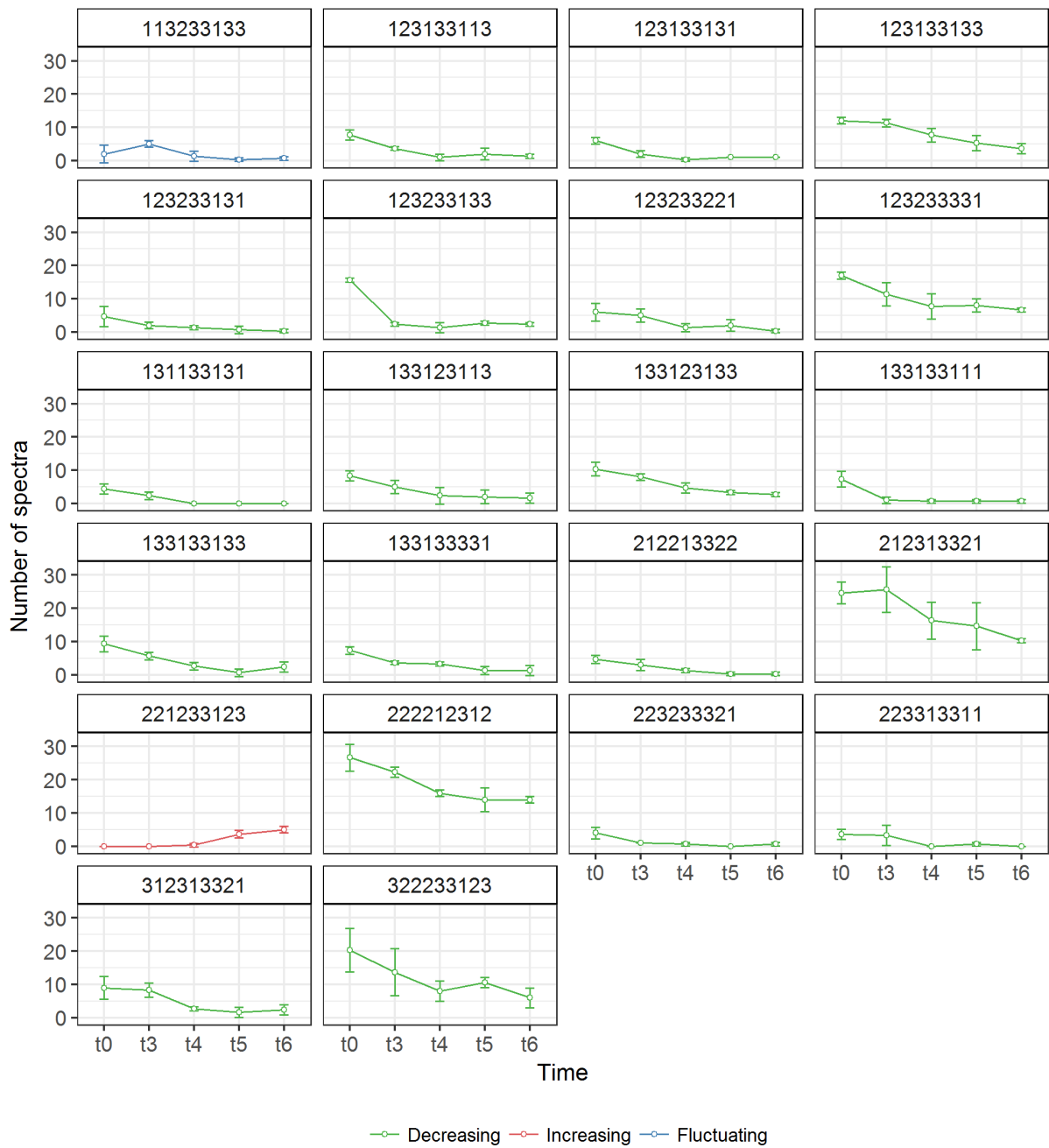
18 Figure S2: PCA representation of one (out of 10) distribution of kinetic profiles determined after
 19 randomly grouping peptides in YE1. On average in the 10 data sets, the number of random classes
 20 declared as significantly varying over time was about $30.7 (\pm 3.7)$, which is around 37% inferior to that
 21 of the grouping based on the physicochemical properties of peptides (49). The average number of total
 22 peptides they enclosed was also reduced by a two-fold factor. This reduction was mostly the result of a
 23 diminution of peptides belonging to the decreasing profiles ($-53\% \pm 20\%$) while, in contrast, the number
 24 of the subsequently added constant peptides increased by about $16\% (\pm 10\%)$.



25

26

27 Figure S3: Comparison of the physicochemical properties of the peptides initially present in YE1 and
 28 YE2 before inoculation ($t = 0$ h). The violin plots and boxplots combine spectra assigned in the three
 29 repetitions altogether. The x-axes correspond to the scales specific to each physicochemical property.
 30 Mean comparison results from Wilcoxon-Mann-Whitney tests are represented according to the
 31 following symbols: ns ($p\text{-val} > 0.05$), * ($0.01 < p\text{-val} \leq 0.05$), **** ($p\text{-val} \leq 0.0001$).



32

33 Figure S4: Classes of peptides whose population is significantly modified (adjusted p -val ≤ 0.01) in YE2
 34 during growth of *S. thermophilus*. The values correspond to the means of the spectra used to identify the
 35 peptides constituting the classes. The error bars represent the standard deviations. In green are classes
 36 showing a clear decrease over time, in red an increase over time, and in blue a class whose abundance
 37 evolution is fluctuating over time because of successive increase, decrease and stagnation.

3.3.4 Bilan

Le travail présenté dans ce manuscrit a permis d'explorer en profondeur le contenu peptidique des deux EXLs, d'identifier par une approche de regroupement par classes physico-chimiques les différentes dynamiques d'utilisation des peptides au cours du temps et, ce faisant, de révéler l'activité de transport d'oligopeptides de la bactérie. Comme indiqué dans la discussion de ce manuscrit, la présence d'un deuxième groupe de peptides dont l'abondance décroît au cours du temps dans l'EXL2 est surprenante. Sur la seule base de ces données, il n'est pas possible d'identifier l'origine de cette dynamique, mais on peut émettre l'hypothèse qu'elle est liée à la différence de concentration en oligopeptides entre les deux EXLs précédemment mesurée sur la base des analyses d'acides aminés (partie 3.2). De même, ces dynamiques (transport + décroissance) identifiées en peptidomique sont dans l'ensemble bien corrélées avec les vitesses de disparition des oligopeptides précédemment mesurées. Ainsi, dans l'EXL1, les acides aminés oligopeptidiques basiques disparaissent le plus vite (0,3 g/h), suivis ensuite des hydrophobes (0,2 g/h) et enfin des anioniques (un peu moins de 0,2 g/h). Dans l'EXL2, ces trois catégories disparaissent sensiblement à la même vitesse ($\approx 0,6$ g/h), reflétant la plus grande diversité physico-chimique des peptides décroissant au fil du temps observée en peptidomique. A noter que la disparition de résidus chargés négativement semble contradictoire, en particulier pour l'EXL1, avec les conclusions du manuscrit (transport = peptides basiques). Il convient toutefois de rappeler que i) la fraction oligopeptidique mesurée n'est pas strictement la même entre les deux analyses (absence d'ultrafiltration à 3 kDa lors de l'analyse des acides aminés, limites liées à la gamme de détection en peptidomique), et ii) les peptides identifiés comme transportés en peptidomique portent également des charges négatives, elles sont seulement en moins grand nombre que les charges positives (voir Figure 5B du manuscrit). Tout ceci montre que les deux approches sont réellement complémentaires, chacune apportant la qualité d'information manquant à l'autre. On soulignera également la présence dans l'EXL1 d'un groupe de peptides négativement chargés dont l'abondance augmente au cours du temps. Ce même groupe n'a pas été identifié dans l'EXL2, et l'origine de cette dynamique reste encore inconnue, bien que l'on puisse soupçonner l'activité protéolytique de la bactérie.

D'un point de vue méthodologique, on précisera que quelques modifications ont été opérées dans le protocole du manuscrit par rapport à celui précédemment présenté dans la Figure 3.12. Celles-ci concernent exclusivement les paramètres d'identification des peptides *in silico* par le logiciel X!TandemPipeline. Brièvement, la tolérance de masse sur les ions précurseurs a été augmentée à ± 10 ppm, ainsi que la *E*-value maximale tolérée par peptide qui est passée de 0,01 à 0,05. Ceci a notamment permis d'augmenter d'environ un facteur 2 le nombre de peptides identifiés sur les mêmes données avec les précédents paramètres. En contrepartie, le nombre maximal de pics par spectre a été réduit de 150 à 50. Il s'est en effet avéré que ce paramètre était responsable de l'identification de nombreux faux positifs en

raison de l'assignation douteuse de spectres de mauvaise qualité (bruit et faible intensité) produits en fin de gradient.

Enfin, on mentionnera ici que ces mêmes données de masse ont aussi servi à interroger la banque protéique de *S. thermophilus* N4L dans le but d'identifier des peptides bactériens produits en cours de croissance. Malheureusement, en dépit de nombreuses tentatives qui faisaient varier les paramètres d'identification, il n'a jamais été possible d'identifier de façon sûre ne serait-ce qu'un seul peptide bactérien. Ceci provient très vraisemblablement de la surabondance des peptides de levure dont le signal a « écrasé » celui de peptides produits par la souche.

3.3.5 Résultats complémentaires

Un certain nombre de travaux complémentaires ont été effectués suite à ces analyses peptidomiques. Deux résultats seront présentés dans cette partie. Le premier concerne l'élaboration d'un modèle statistique à partir des données du manuscrit visant i) à sélectionner parmi les neuf propriétés étudiées celles expliquant le mieux les trajectoires cinétiques observées et valider ainsi les précédentes interprétations, et ii) à prédire pour n'importe quel peptide donné si celui-ci est susceptible d'être transporté ou non. La deuxième partie de résultats complémentaires concerne l'élaboration et la mise en œuvre d'un protocole de peptidomique simplifié permettant une analyse d'échantillons « haut-débit ».

3.3.5.1 Modélisation sPLS-DA

Les analyses réalisées ont permis d'identifier le statut cinétique de 1060 peptides dans l'EXL1 et 952 dans l'EXL2, soit environ un quart du total identifié dans les deux cas. Dans chacun des deux jeux de données, un large sous-groupe de ces peptides a été identifié comme disparaissant graduellement au fil du temps et dont les caractéristiques physico-chimiques permettent, sur la base de travaux antérieurs, de rattacher cette diminution à un phénomène de transport. Il est ainsi possible de simplifier à nouveau ces données en distinguant les peptides transportés et non transportés (Figure 3.15). A noter que, en raison de l'absence de validation supplémentaire et par mesure de simplification, le groupe de peptides anioniques décroissant au cours du temps et spécifiques à l'EXL2 ont été également classés comme non transportés, l'hypothèse d'une décroissance liée à leur hydrolyse restant privilégiée.

Cette séparation des peptides en deux classes qualitatives (transporté/non transporté) a servi de base à l'élaboration d'un modèle statistique basé sur la sPLS-DA (*sparse Partial Least Squares Discriminant Analysis*). Brièvement, il s'agit d'une extension de la régression PLS, également appelée *Projection to Latent Structure*, permettant de séparer des groupes



FIGURE 3.15 : ACP sur l'ensemble des peptides dont le profil cinétique a été caractérisé dans les deux EXLs.

d'échantillons (les peptides) en fonction de variables quantitatives (les propriétés physico-chimiques) tout en maximisant la covariance avec une variable de réponse catégorielle, ici le statut de transport des peptides (Lê Cao *et al.*, 2011). La version parcimonieuse (*sparse*) de la PLS-DA permet en outre d'opérer une sélection des variables les plus informatives. Cette analyse a été réalisée en plusieurs étapes (Figure 3.16). Deux modèles intermédiaires ont tout d'abord été élaborés à partir des données des deux EXLs traitées séparément. Le but était de vérifier la faisabilité et la pertinence d'une telle approche sur ces données de peptidomique. Ces deux sous-modèles ont été validés *via* leur capacité à prédire l'aptitude au transport des peptides analysés de l'autre EXL. Un modèle complet a finalement été élaboré à partir des données combinées des deux EXLs. Il a notamment permis de prédire l'aptitude au transport des peptides qui n'avaient pas été analysés car appartenant à des classes physico-chimiques de faible abondance ou ne montrant pas de variations significatives au cours du temps. Cette partie présentera donc les résultats issus de ce dernier modèle complet.

A noter que dans cette partie, conformément à l'usage, le terme de "classe" utilisé fera désormais référence aux modalités de la variable de réponse (ici deux classes : transporté/non transporté), à ne pas confondre avec les classes physico-chimiques utilisées dans le manuscrit pour identifier les dynamiques cinétiques dans les données de peptidomique.

3.3.5.1.1 Méthode

L'analyse sPLS-DA a été réalisée par l'intermédiaire du package R « mixOmics » version 6.3.2 (Lê Cao *et al.*, 2011; Rohart *et al.*, 2017). Les peptides caractérisés dans les deux EXLs constituaient autant d'échantillons analysés. Les variables explicatives (matrice d'entrée X) étaient les propriétés physico-chimiques des peptides, et la variable de réponse (matrice de sortie Y) la classe des différents peptides telle que déterminée précédemment dans les résultats du manuscrit (transporté ou non). L'élaboration du modèle s'est faite en plusieurs étapes dont l'issue a été évaluée à chaque fois par une validation croisée et par le calcul d'un score de performances. Brièvement, la validation croisée consiste à sous-diviser l'échantillon de départ en k sous-ensembles (ou *folds*), à construire un modèle dit d'apprentissage à partir de $k - 1$ ensembles et de calculer à partir du sous-ensemble k restant (ensemble de validation) le score de performances du modèle d'apprentissage. L'opération est ensuite répétée en sélectionnant un autre ensemble de validation parmi les $k - 1$ ensembles n'ayant pas été utilisés pour la validation du modèle, et ainsi de suite jusqu'à ce que tous les ensembles aient été utilisés une fois comme ensemble de validation. Tout ce processus est ensuite répété n fois, au choix de l'utilisateur, afin de procéder à de nouveaux échantillonnages. Le score de performances utilisé ici est le BER (*Balanced Error Rate*) que l'on cherche donc à minimiser. Le BER représente la moyenne des taux d'erreur par classe. Ici, $BER = \frac{1}{2} \times (\frac{a}{n_1} + \frac{b}{n_2})$, où a est le nombre de peptides transportés faussement prédits comme non transportés, n_1 le nombre



FIGURE 3.16 : Stratégie déployée pour l'implémentation d'un modèle prédictif à partir des données de peptidomique.

total de peptides transportés, b le nombre de peptides non transportés faussement prédits comme transportés, et n_2 le nombre total de peptides non transportés.

La première étape a consisté en l'élaboration d'un modèle intermédiaire PLS-DA (sans sélection de variables) pour réaliser une pré-sélection des paramètres qui seront utilisés dans le modèle final, à savoir le nombre de composantes et le type de distance de prédiction (trois distances sont proposées dans mixOmics : distance maximale, de Mahalanobis, ou des centroïdes, voir documentation). Pour cela, la fonction *perf* a été utilisée afin de réaliser une validation croisée (nombre de sous-ensembles $k = 20$, nombre d'itérations $n = 10$) de ce modèle et sélectionner les paramètres permettant de minimiser le BER. La fonction *tune* a ensuite été utilisée afin d'affiner ce choix sous un modèle sPLS-DA et de sélectionner les variables les plus explicatives par composante. Cette fonction procède également par validation croisée ($k = 20$, $n = 100$). Une fois les paramètres définitivement arrêtés, le modèle final a été construit. Ses performances ont tout d'abord été évaluées par validation croisée (fonction *perf*, $k = 20$, $n = 100$), puis en utilisant la fonction *predict* afin de prédire la classe de l'ensemble des peptides ayant servi à son élaboration. Cette dernière approche, contrairement à la validation croisée, ne procède pas itérativement par prédiction sur un sous-ensemble de validation mais utilise le modèle complet pour prédire en une seule fois la classe de chacun des échantillons.

3.3.5.1.2 Résultats

Le premier modèle PLS-DA a été ajusté sur 5 composantes afin d'effectuer une pré-sélection des variables nécessaires pour le modèle final. Il a été déterminé que les deux premières composantes permettaient de minimiser le BER ($\approx 10\%$) en utilisant la distance des centroïdes. Ces paramètres ont été ensuite renseignés lors de l'étape de réglage des paramètres du modèle sPLS-DA. Celle-ci a confirmé que deux composantes étaient suffisantes pour optimiser le modèle. On peut toutefois noter que les peptides sont principalement discriminés par la première composante, bien que leur séparation ne soit pas totale (Figure 3.17A). Cette première composante a été construite à partir de 7 variables, alors que la deuxième inclut l'information des 9 variables initiales (Figure 3.17B). De la première composante ont ainsi été retirées les variables teneur en proline et en résidus aromatiques. La Figure 3.17C montre que la charge nette est le premier facteur pour discriminer la classe (transport) des peptides, ce qui conforte l'interprétation des résultats du manuscrit précédemment présenté. Par corollaire, contribuent fortement les variables corrélées à la charge nette du peptide (teneur en résidus acides, polarité et teneur en résidus basiques). Suivent ensuite l'hydrophobicité du peptide et son encombrement stérique, variables qui sont toutes deux également corrélées. Enfin, la taille du peptide n'intervient qu'en dernier dans la contribution à la première composante. Etant données les diverses corrélations entre ces variables, il semble ici raisonnable de résumer les facteurs comme participant au transport des peptides de la façon suivante

(par ordre d'importance) : charge nette, hydrophobicité et taille.

L'évaluation des performances de ce modèle, par validation croisée, a abouti à un BER légèrement inférieur à 10 %, indiquant qu'en moyenne pour chaque classe, le modèle a émis une assignation correcte dans 90 % des cas. La fonction *predict* a ensuite été utilisée successivement sur les peptides caractérisés des deux EXLs qui ont servi à construire le modèle. Les résultats de cette démarche sont indiqués dans la Figure 3.17D. Ceux-ci permettent de constater que les performances ne sont pas égales selon l'EXL, le modèle ayant donné de meilleures prédictions pour les peptides de l'EXL2 (BER = 6 %) par rapport à ceux de l'EXL1 (BER = 12 %). Cette différence résulte plus précisément de l'excellente capacité du modèle à identifier les peptides transportés dans l'EXL2, puisqu'un seul peptide a été mal assigné sur les 222 que compte ce groupe. Les autres groupes montrent des taux moyens d'erreurs proches variant entre 10 % et 13 %. La Figure 3.18A permet de précisément localiser où se situent les fausses prédictions sur un plan factoriel. Globalement, pour les deux EXLs, celles-ci sont logiquement situées au niveau de la zone où se recouvrent les deux classes de peptides transportés/non transportés (Figure 3.15). Ceci indique que la sPLS-DA n'a pas réussi à mieux séparer les peptides qu'une simple ACP.

Néanmoins, ses scores de prédiction étant relativement satisfaisants, ce modèle a finalement été utilisé sur l'ensemble des 75 % restants de peptides qui n'avaient pas pu être caractérisés dans les deux EXLs en raison soit d'une trop faible abondance, soit de variations détectées comme non significatives (Figure 3.18B). En moyenne dans les deux EXLs, environ un tiers des peptides non caractérisés disposent de propriétés physico-chimiques associées au transport. Si l'incertitude de ces prédictions est plus forte concernant les peptides situés à la frontière entre les deux groupes, elles sont en revanche beaucoup plus fiables pour les peptides les plus périphériques.

3.3.5.1.3 Discussion

Le premier objectif de ce modèle était de conforter l'interprétation des résultats précédemment présentés dans le manuscrit concernant les facteurs intervenant dans le transport des peptides. Cet objectif est donc rempli et la charge nette constitue effectivement le facteur déterminant pour le transport des peptides. En revanche, il n'a pas été possible d'améliorer la séparation des deux classes de peptides (transportés/non transportés) *via* la sPLS-DA en comparaison avec une simple ACP qui est une analyse non-supervisée. Ceci provient vraisemblablement du faible nombre de variables réellement indépendantes utilisées et suggère la présence d'un ou plusieurs autres facteurs non inclus dans l'étude et intervenant également dans le transport des peptides. La concentration des peptides en est logiquement un, car des phénomènes de compétition pour le transport doivent vraisemblablement intervenir dans de tels milieux complexes. Ainsi, un peptide présentant une haute affinité pour son

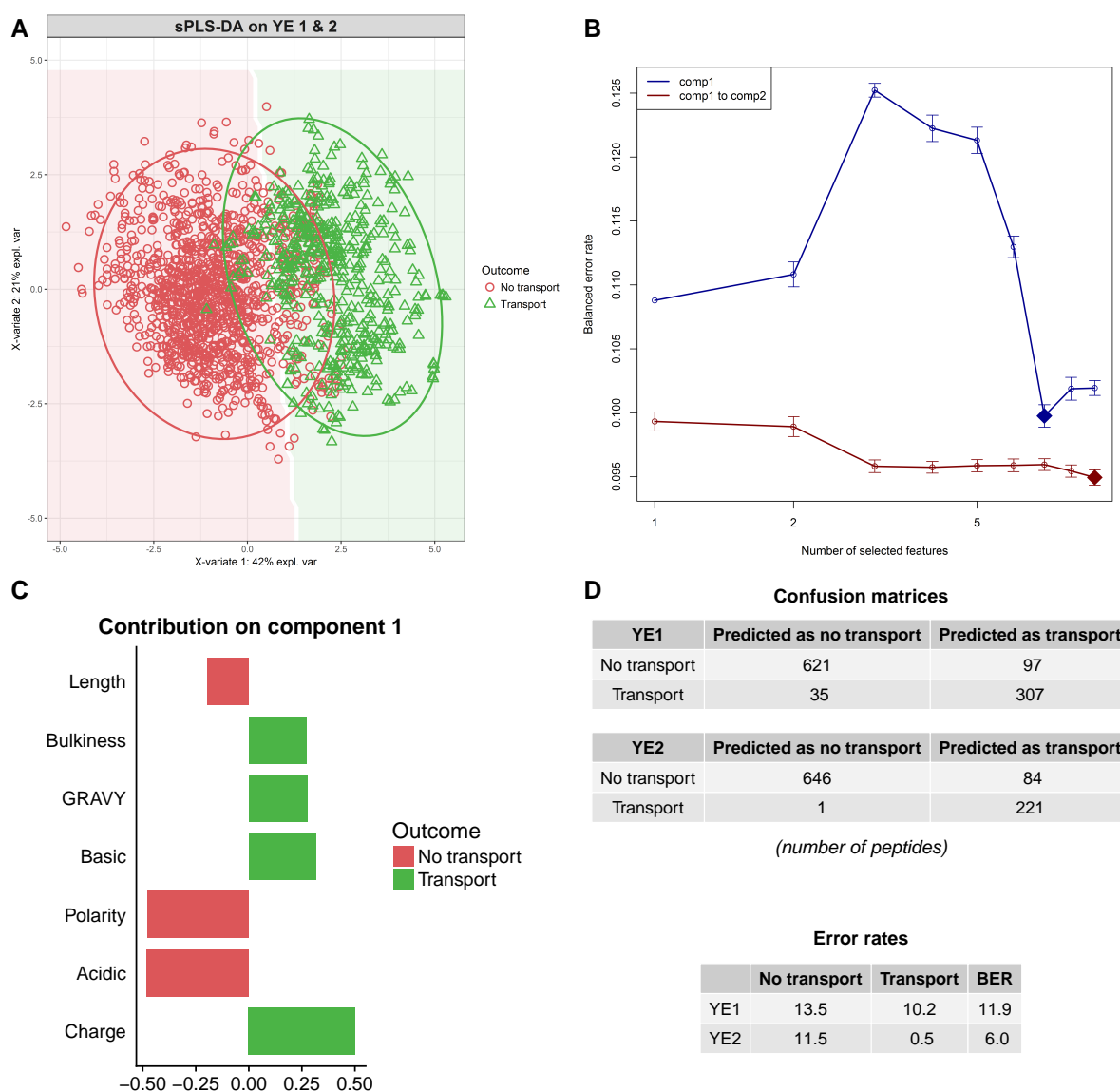


FIGURE 3.17 : Elaboration du modèle sPLS-DA complet à partir de l'ensemble des données peptidomiques caractérisées cinétiquement. (A) Graphe des échantillons (peptides) obtenu par sPLS-DA, les surfaces colorées représentent les aires de prédiction associées aux deux classes ayant servi à séparer les peptides (transporté/non transporté). (B) Résultats de l'étape de réglage des paramètres du modèle sPLS-DA utilisant les deux composantes par validation croisée. Le graphe représente l'évolution du BER (axe y) en fonction du nombre de variables sélectionnées (axe x). Les losanges représentent le nombre optimal de variables à sélectionner pour chacune des deux composantes. (C) Graphe des *loadings* (coordonnées standardisées) indiquant la classe pour laquelle la contribution des variables sélectionnées sur la composante 1 est maximale. (D) Matrices de confusion (en haut) représentant la répartition des prédictions par classe et par nature de prédiction (correcte ou erronée). Ces mêmes informations sont synthétisées dans le tableau des taux d'erreur (en bas), indiquant le pourcentage d'erreurs par classe ainsi que le BER pour chaque EXL.

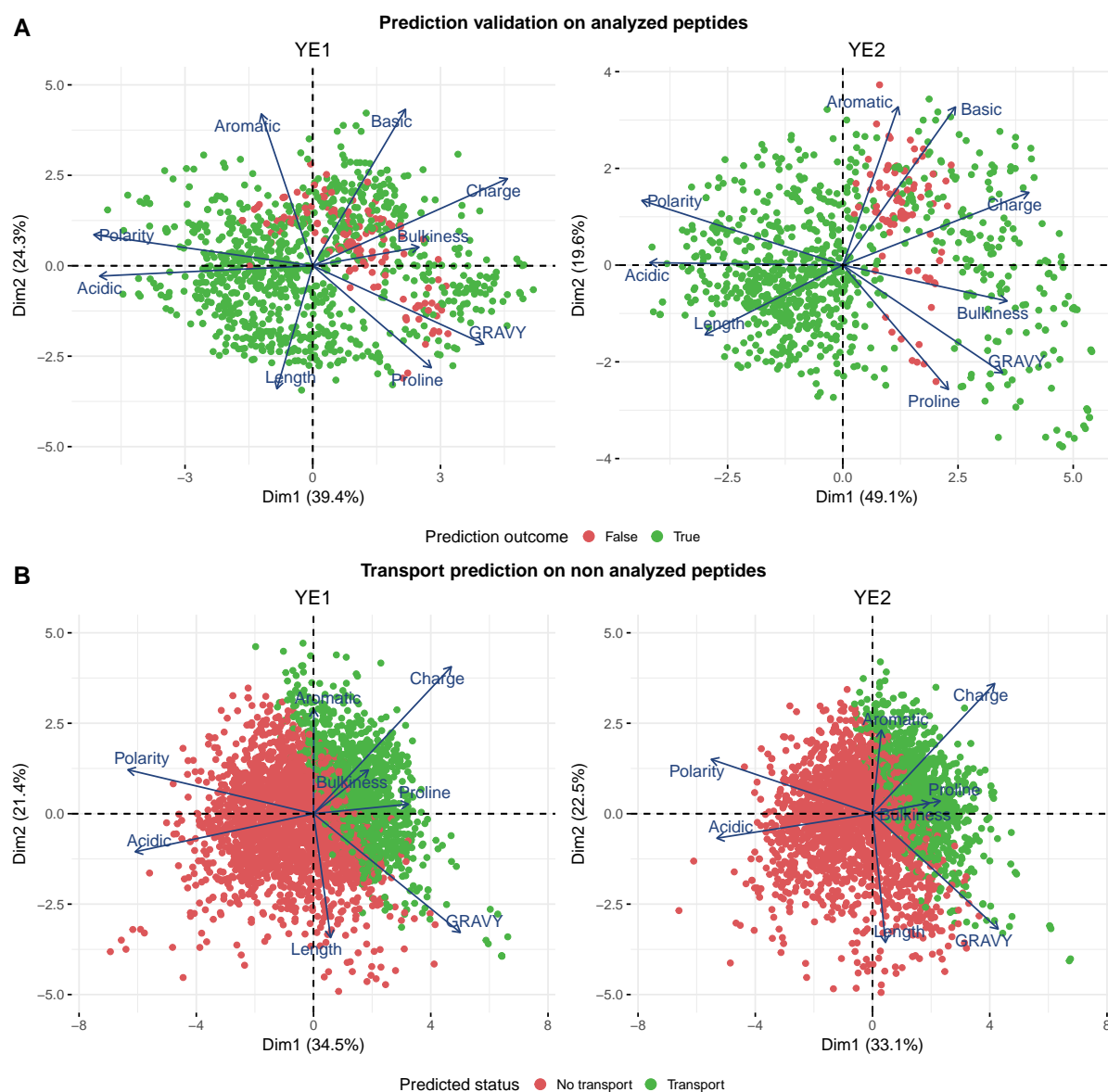


FIGURE 3.18 : Application du modèle sPLS-DA complet. (A) Visualisation de la véracité des prédictions sur les peptides cinétiqumement caractérisés des deux EXLs. (B) Prédiction de l'aptitude au transport des peptides non caractérisés des deux EXLs.

transporteur pourra ne pas être pris en charge si son abondance est trop faible par rapport à celle de compétiteurs moins spécifiques. Il est également possible d'émettre l'hypothèse que, outre sa composition générale en acides aminés, la position de ces acides aminés au sein du peptide ainsi qu'une éventuelle structure tridimensionnelle spécifique puissent intervenir dans les mécanismes de fixation avec la SBP du transporteur. Ainsi, deux peptides de propriétés physico-chimiques proches pourraient ne pas être transportés avec la même efficacité en raison d'un résidu placé différemment ou d'un repliement tridimensionnel particulier qui favoriseraient ou désavantageraient le transport de l'un des peptides par rapport à l'autre. Il serait intéressant d'explorer cette hypothèse en utilisant les outils de modélisation des structures de peptides disponibles tels que ceux présentés par Thévenet *et al.* (2015). Combinés avec la caractérisation structurale des diverses SBPs AmiA de *S. thermophilus*, ces outils pourraient permettre de modéliser les interactions peptides/SBPs et ainsi identifier avec beaucoup plus de certitude les peptides les plus susceptibles d'être transportés par la bactérie.

Néanmoins, même en l'absence de facteurs supplémentaires, le modèle actuel permet d'ores et déjà de prédire avec une assez bonne justesse quels sont les peptides présentant une composition compatible ou non avec leur transport par *S. thermophilus*. Afin de gagner en robustesse, il devrait être enrichi par l'ajout de données de peptidomique supplémentaires obtenues dans des conditions similaires à celles de la présente étude mais avec d'autres types d'EXLs voire d'autres hydrolysats protéiques. A terme, l'objectif serait d'élaborer un outil suffisamment robuste permettant d'explorer les potentialités de nouveaux substrats protéiques complexes sans qu'il soit nécessaire de réaliser une cinétique de croissance bactérienne complète. En effet, une des principales limites de l'approche peptidomique employée ici est sa lourdeur. En raison du fractionnement chromatographique préalable, le traitement d'un seul surnageant de culture nécessite plus 16 h d'analyse en spectrométrie de masse, sans compter le temps nécessaire à sa préparation et à son fractionnement. Afin de pouvoir déployer plus facilement cette approche, une simplification du procédé est nécessaire, ce qui a été entrepris et est présenté dans la partie suivante.

3.3.5.2 Amélioration du protocole expérimental pour une analyse peptidomique haut-débit

3.3.5.2.1 Stratégie d'amélioration du protocole

En se basant sur les résultats acquis, des travaux ont été menés afin d'alléger le protocole analytique et le rendre compatible avec une utilisation « routinière » tout en conservant de bonnes performances d'ensemble. Le goulot d'étranglement du protocole réside dans la première dimension chromatographique, responsable de la démultiplication du nombre

d'échantillons à analyser. Sa suppression est donc indispensable pour cet objectif de simplification. Ce travail a été réalisé en plusieurs étapes. Brièvement, un échantillon issu des manipulations précédentes - fraction éluée de l'EXL1 à $t = 0$ h (milieu initial), première répétition, dilution au 20^{ème} - a tout d'abord été analysé directement en LC-MS/MS dans des conditions similaires à celles utilisées précédemment. Par rapport au même échantillon préalablement fractionné, le nombre de spectres totaux et de peptides uniques identifiés ont été réduits respectivement de 74 % et 67 %. Pour compenser ces pertes, la première démarche entreprise a été d'améliorer la séparation chromatographique. Concrètement, cela a consisté à utiliser une colonne de 50 cm (contre 15 cm précédemment) et à rallonger le gradient chromatographique en amont du spectromètre de masse afin de mieux séparer les peptides et augmenter la profondeur d'analyse. Parmi les différents gradients testés, un simple gradient long de 3 h (contre 40 min précédemment) sur cette nouvelle colonne a permis de réduire les pertes d'identification, celles-ci étant désormais de 66 % (spectres) et 57 % (peptides) par rapport à l'échantillon fractionné. Sur la base de ces nouvelles conditions chromatographiques, la deuxième étape s'est focalisée sur la modification des paramètres de fragmentation. Ont tout d'abord été modifiés l'énergie de collision ainsi que le filtre sur l'état de charge des ions précurseurs (prise en compte des mono-chargés). Ces deux paramètres n'ont néanmoins pas permis d'augmenter significativement le nombre d'assignations. Enfin, la stratégie de sélection et d'exclusion des peptides précurseurs a été ajustée, passant d'une stratégie « top 6 » à « top 8 » (sélection des 8 précurseurs les plus intenses), et surtout, leur temps d'exclusion a été réduit de 40 sec à 10 sec. L'objectif ici était d'augmenter l'abondance spectrale des peptides identifiés afin d'améliorer la qualité de leur quantification relative en *spectral counting*. Cette modification a abouti à des résultats très satisfaisants, un important gain de spectres totaux ayant été observé ainsi qu'une légère augmentation du nombre de peptides uniques identifiés. Ainsi, suite à cette étape, les pertes par rapport à l'échantillon fractionné étaient de 49 % (spectres) et 54 % (peptides).

Tout ce travail a été effectué sur le même système LC-MS/MS que lors des précédentes analyses avec fractionnement, à savoir une nano-HPLC couplée à un LTQ-Orbitrap. La dernière phase d'amélioration a consisté à transposer cette analyse sur un spectromètre de masse de dernière génération : un Orbitrap Fusion Lumos Tribrid (Thermo Scientific), également couplé à une nano-HPLC. Très brièvement, cet appareil offre des conditions d'analyse grandement améliorées par rapport au précédent spectromètre, en particulier une fragmentation en HCD (*Higher energy Collisional Dissociation*), une stratégie « top N » (fragmentation de tous les précurseurs durant une fenêtre de temps ici de 3 sec), ou encore l'analyse des ions parents et fragments à haute résolution. Etant donnée la rapidité de l'appareil, le gradient chromatographique a été réduit à 105 min, mais la réduction du temps d'exclusion (10 sec) a été conservée. Tous les précurseurs présentant un état de charge de +1 à +4 ont été inclus. Enfin, en raison de l'analyse haute résolution des ions fragments, le nombre maximal toléré

de pics par spectre a été augmenté de 50 à 100. En effet, dans l'ancien système LC-MS/MS où l'acquisition des spectres MS2 se fait à basse résolution, ce paramètre lorsqu'il est augmenté est responsable de nombreux faux positifs et augmente considérablement le FDR (assignation de spectres de mauvaise qualité). L'analyse en haute résolution des ions fragments sur le nouveau système permet de compenser ce problème.

Sous ces conditions, des résultats d'identifications supérieurs ont été obtenus en comparaison avec ceux du même échantillon fractionné et analysé *via* le premier spectromètre de masse. Le nombre de spectres assignés a ainsi été amélioré de 72 %, et le nombre de peptides identifiés a augmenté de 44 %. En résumé, grâce à ce changement de technologie, une telle analyse peptidomique « *shotgun* » montre des résultats proches de ceux d'un triplicat d'analyses réalisées dans les anciennes conditions.

3.3.5.2.2 Estimation de la répétabilité technique du nouveau protocole

De façon similaire à ce qui avait été précédemment réalisé (partie 3.3.2.2), la reproductibilité technique du nouveau protocole a ensuite été évaluée. Pour cela, le même échantillon a été analysé 8 fois. Les principaux résultats de cette étude sont résumés dans la Figure 3.19. En moyenne, 2686 peptides distincts ont été identifiés par répétition par l'intermédiaire de 4650 spectres de fragmentation (Figure 3.19A). Ceci représente un ratio d'environ 8,6 spectres par peptides, ce même rapport étant d'environ 1,5 dans les données de l'échantillon équivalent présentées dans le manuscrit soumis pour publication. L'abondance spectrale moyenne a donc été grandement améliorée. Cette augmentation concerne précisément les peptides ayant été identifiés par un total de 2 à 4 spectres (Figure 3.19B). Cette catégorie représente ici environ 31 % des peptides identifiés par analyse, contre seulement 17 % dans les données précédentes. Logiquement, la proportion de peptides identifiés par un seul spectre a diminué, elle est ainsi passée de 79 % à 65 %. En revanche, la part de peptides identifiés avec plus de 5 spectres n'a pas évolué. La Figure 3.19C montre que les peptides ont été identifiés très majoritairement sous forme doublement chargée, ce qui était déjà le cas auparavant. Finalement, sur les huit répétitions effectuées, 1519 peptides ont été systématiquement identifiés, ce qui représente en moyenne 57 % des peptides identifiés par analyse (Figure 3.19D). Là aussi, on constate une nette amélioration de la répétabilité d'identification par rapport aux précédentes conditions, découlant de la plus forte abondance spectrale des peptides.

L'ensemble des peptides identifiés au cours des 8 analyses ont ensuite été caractérisés selon les 9 propriétés physico-chimiques décrites dans le manuscrit. Les résultats de cette caractérisation sont présentés dans la Figure 3.20. Les propriétés sont remarquablement conservées, aucune différence significative n'ayant pu être détectée entre répétitions, ce qui n'était pas le cas précédemment. Il faut toutefois rappeler qu'il s'agit ici de l'injection répétée du

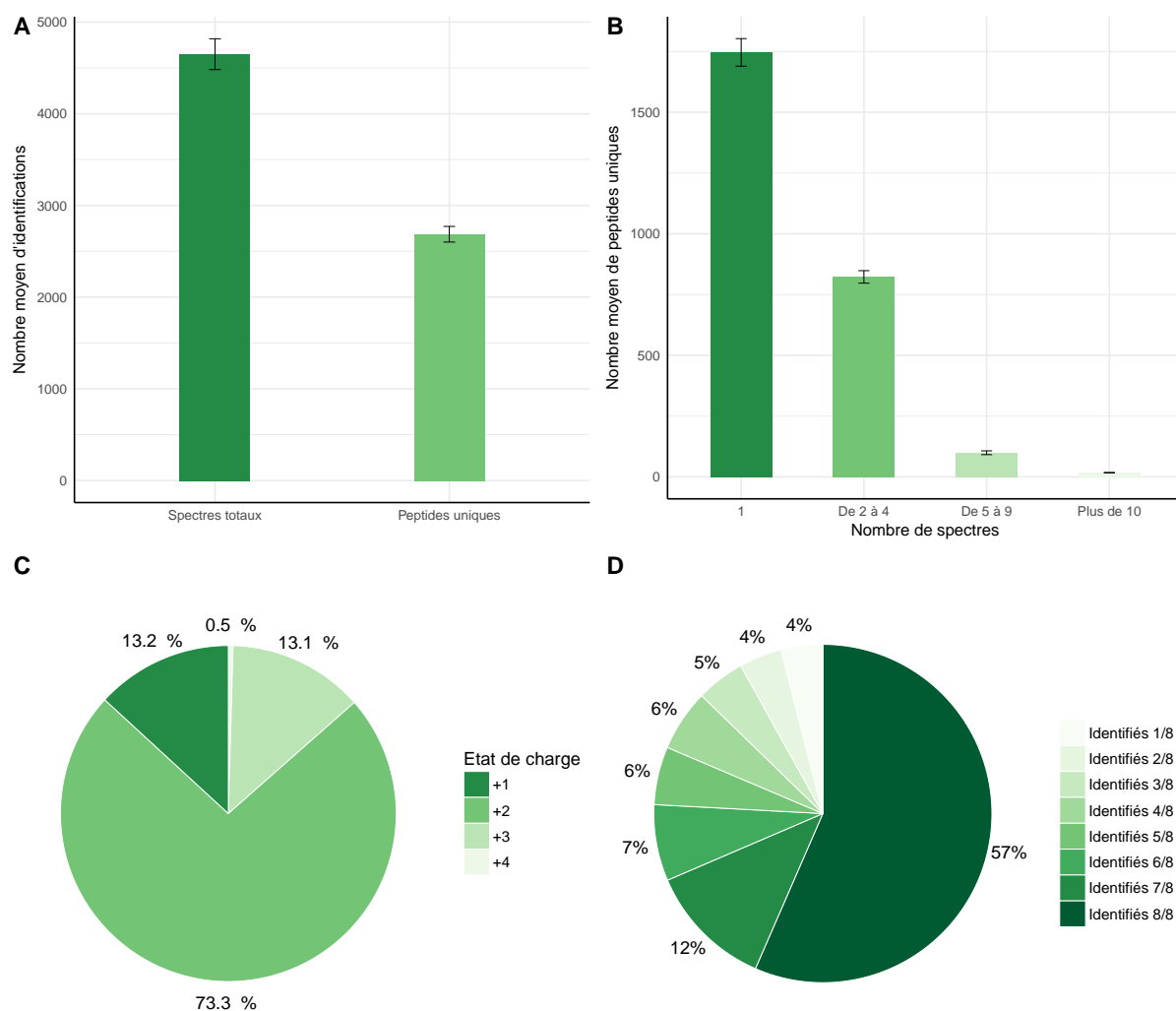


FIGURE 3.19 : Principaux résultats de l'analyse de répétabilité du protocole de peptidomique amélioré. (A) Nombre moyen de spectres totaux et peptides uniques identifiés par analyse. (B) Nombre moyen de peptides identifiés en fonction de leur abondance spectrale. (C) Distribution moyenne de l'état de charge des peptides identifiés. (D) Distribution moyenne de la fréquence d'identification des peptides sur l'ensemble des huit répétitions.

Chapitre 3 : Caractérisation des dynamiques d'utilisation de peptides d'extraits de levure par *Streptococcus thermophilus* N4L

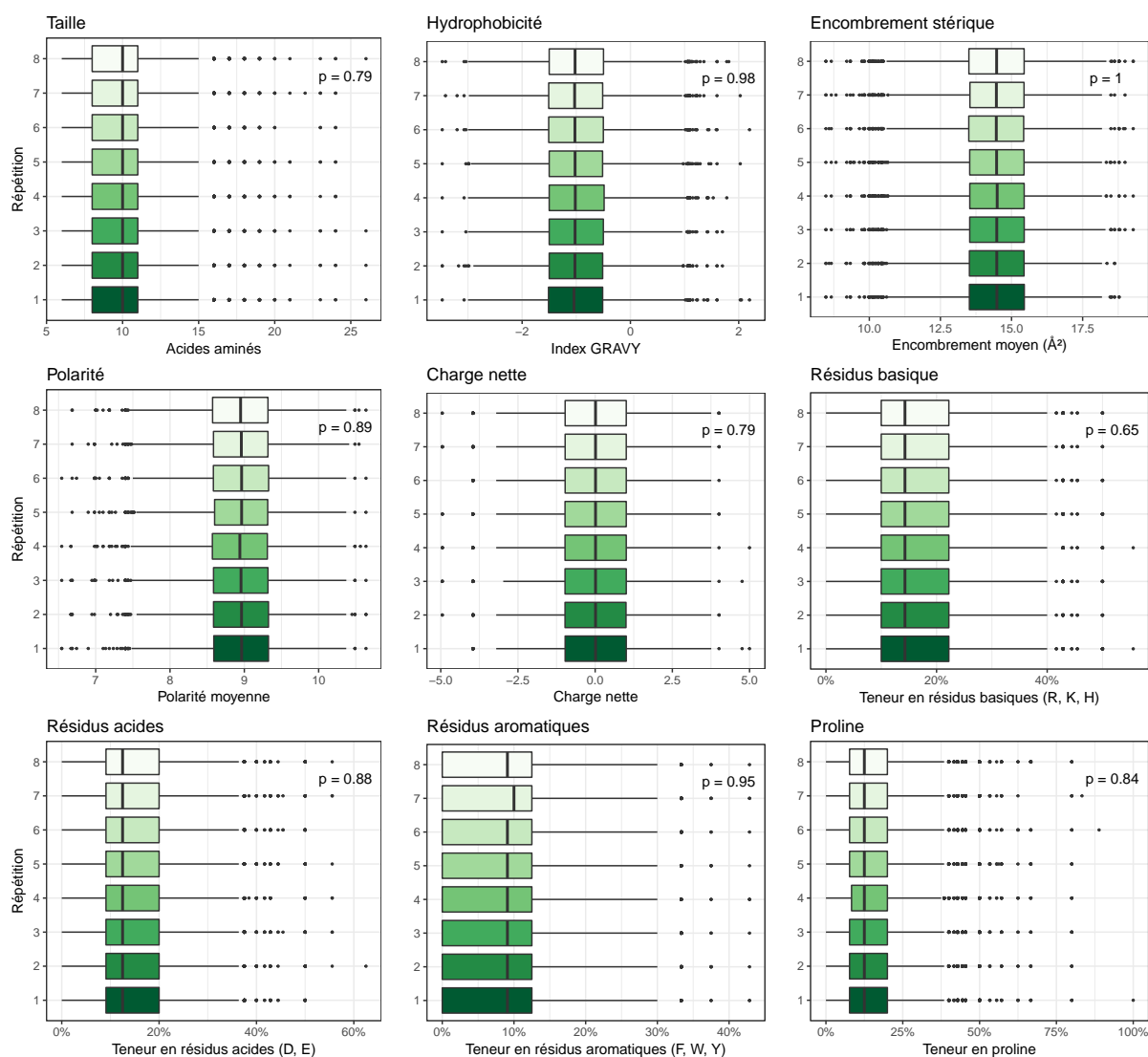


FIGURE 3.20 : Propriétés physico-chimiques des peptides identifiés durant l'analyse de répétabilité du protocole de peptidomique amélioré. Les boxplots ont été réalisés à partir des spectres totaux assignés afin de tenir compte de la redondance d'identification. Les valeurs de p-values indiquées correspondent aux résultats de tests de Kruskal-Wallis.

même échantillon (répétition technique au niveau du système LC-MS/MS). Dans les précédents résultats, la variabilité technique avait été évaluée sur l'ensemble du protocole, en incluant la phase de purification des échantillons. Néanmoins, il est aussi probable que la meilleure répétabilité observée ici bénéficie à la fois de la meilleure profondeur d'analyse, mais aussi de la suppression du fractionnement chromatographique qui devait contribuer aux variations précédemment observées.

3.3.5.2.3 Bilan

Tous ces résultats reflètent ainsi l'amélioration à la fois quantitative mais aussi qualitative obtenue grâce à ce nouveau protocole analytique. Celui-ci permet donc une analyse pepti-

domique beaucoup plus rapide, simple à déployer, mais également plus fiable et répétable. Ce degré de performances nécessite évidemment l'utilisation d'un spectromètre moderne que toutes les plateformes analytiques ne possèdent pas. Néanmoins, des résultats tout à fait corrects sont accessibles sur des machines moins performantes en ajustant convenablement les conditions chromatographique et de fragmentation. Etant donnée l'absence de fractionnement, il est également possible de compenser des résultats moins performants obtenus avec des systèmes d'ancienne génération par un nombre de répétitions plus important.

3.4 Conclusion

Les résultats présentés dans ce chapitre ont révélé les dynamiques d'utilisation des peptides par *S. thermophilus* N4L dans deux EXLs différents. La quantification du contenu azoté, par dosage des acides aminés libres, a en particulier montré que les deux substrats n'offraient pas des quantités égales d'oligopeptides à la bactérie, et que celle-ci présentait des cinétiques de consommation des peptides différentes en fonction du milieu. L'approche peptidomique a pour sa part mis à jour la complexité de la fraction oligopeptidique des hydrolysats, ceux-ci étant composés de plusieurs milliers de peptides en mélange présentant une vaste diversité de propriétés physico-chimiques. L'étude dynamique par spectrométrie de masse de l'utilisation des peptides a nécessité la mise au point d'un processus d'analyse des données basé sur le regroupement des peptides par classes physico-chimiques, en raison d'une abondance spectrale en moyenne trop faible à l'échelle du peptide individuel. Cette approche a permis de mettre à jour des tendances communes entre les deux EXLs, en particulier concernant la disparition d'un large groupe de peptides basiques dont l'origine peut très vraisemblablement être attribuée au transporteur Ami. En revanche, conformément à ce qui avait été observé lors des dosages d'acides aminés, des dynamiques spécifiques à chaque EXL ont également été caractérisées, telle que l'atteste l'identification d'un deuxième groupe de peptides disparaissant graduellement dans l'EXL2, et inversement d'un groupe de peptides dans l'EXL1 dont l'abondance augmente progressivement.

Pour identifier précisément les acteurs responsables de ces diverses dynamiques, des analyses complémentaires sont nécessaires. En particulier, il conviendrait de soumettre les souches mutantes Δami ou $\Delta prtS$ - les acteurs les plus probables de ces dynamiques - à une analyse peptidomique durant leur culture dans des conditions similaires. Ces deux souches mutantes ont été construites durant le projet de thèse. Néanmoins, en raison de la lourdeur du protocole d'analyse peptidomique, il a été nécessaire de travailler sur sa simplification afin de pouvoir le déployer plus facilement. Cette étape étant maintenant achevée, il devient plus réaliste d'analyser les deux mutants, ce qui constitue une perspective immédiate au

projet de thèse. Ensuite, il serait intéressant d'étendre cette approche à d'autres modèles bactériens et autres substrats, car cette démarche constitue probablement à ce jour l'outil le plus performant pour caractériser les peptides d'un hydrolysate protéique et analyser *in vivo* les préférences d'utilisation des peptides par les bactéries.

Enfin, sur un plan à nouveau plus méthodologique, grâce à cette simplification du protocole, il est désormais envisageable de quantifier les peptides identifiés non plus sur la base de leur abondance spectrale, mais par l'extraction de leurs courants d'ions (XIC). En raison du fractionnement chromatographique, il n'était pas possible de procéder à une telle quantification. Celle-ci nécessite en effet une étape préalable d'alignement des chromatogrammes afin d'homogénéiser les temps de rétention des divers pics identifiés dans les différents échantillons. Or cet alignement nécessite un nombre minimum (de l'ordre d'une centaine) de peptides identifiés en commun dans des fractions de niveau équivalent, ce qui n'était pas toujours le cas. En analyse *shotgun*, les peptides communs entre les différents échantillons sont suffisamment abondants, rendant possible une telle quantification. Celle-ci, bien que toujours relative, est plus précise que du simple *spectral counting*, et permettrait éventuellement de travailler à l'échelle du peptide individuel sans avoir besoin de recourir à un regroupement par classe physico-chimique. Néanmoins, les outils de protéomique actuels permettant de réaliser une quantification par XIC sont conçus pour quantifier les protéines et ne s'arrêtent pas au stade peptide. Un développement méthodologique est donc encore probablement à prévoir pour appliquer de façon automatisée cette méthode à la peptidomique. Finalement, de façon plus générale encore, un travail de fond en développement bioinformatique serait aussi nécessaire afin d'améliorer les rendements d'identification. A l'heure actuelle, ceux-ci sont en effet bien plus faibles que ceux observés en protéomique. Sur l'ensemble des spectres produits au cours d'une analyse peptidomique, seuls 5 % à 10 % d'entre eux sont assignés, contre 30 % à 40 % en protéomique. Il y a donc là aussi très certainement une importante voie d'amélioration à explorer.

Chapitre 4

Etude des variations d'expression génétique et de synthèse protéique en réponse à la composition des extraits de levure

Sommaire

4.1 Introduction	176
4.2 Matériel et méthodes	178
4.2.1 Conditions de culture et de prélèvements	178
4.2.2 Séquençage RNAseq et analyses transcriptomiques	178
4.2.3 Identification du protéome	179
4.2.4 Quantification des protéines et analyse différentielle	181
4.2.5 Analyses génomiques complémentaires	182
4.2.6 Test de surreprésentation et visualisation des résultats	182
4.2.7 Intégration multiomique par analyse multivariée	183
4.3 Résultats	184
4.3.1 Analyses générales	184
4.3.2 Analyse différentielle entre EXL1 et EXL2	187
4.3.3 Impact fonctionnel et métabolique des EXLs sur <i>S. thermophilus</i> N4L .	189
4.3.4 Impact des EXLs sur la régulation par les peptides : le cas du streptide	197
4.3.5 Signature moléculaire des EXLs par intégration multiomique	200
4.4 Conclusion	205

4.1 Introduction

Le chapitre précédent a fait état d'importantes différences en termes de composition de la fraction oligopeptidique des EXLs analysés, mais aussi concernant l'utilisation qu'en fait la souche *S. thermophilus* N4L. Ces différences d'utilisation sont le reflet d'une probable adaptation différentielle de la souche aux deux environnements qui lui ont été proposés, et, réciproquement, témoignent de l'impact des EXLs sur la bactérie. Ce chapitre va ainsi s'intéresser spécifiquement à cette question de l'impact des EXLs sur *S. thermophilus* en se focalisant sur le métabolisme bactérien.

En se replaçant dans le contexte général du projet de thèse, l'objectif principal auquel doit répondre cette partie est, dans un premier temps, de comprendre comment les deux EXLs choisis affectent différentiellement le métabolisme de la bactérie. Ensuite, en fonction des réponses apportées à cette question, il s'agira d'évaluer si certaines fonctions cellulaires affectées sont susceptibles d'avoir un rôle dans l'élaboration des performances et fonctionnalités technologiques du ferment. Ces notions, très larges, incluent d'une part les performances de croissance de la souche lors de sa production industrielle en milieu riche de type EXL, cet aspect étant notamment motivé ici par la différence de croissance observée en bioreacteurs. D'autre part, on s'intéressera également à un possible effet de conditionnement ou programmation cellulaire que pourraient avoir les EXLs sur les fonctionnalités technologiques futures de la bactérie. A titre d'exemple simple, si l'un des deux EXLs provoque une plus forte induction des systèmes de résistance aux stress de la bactérie, on pourra émettre l'hypothèse qu'une population bactérienne cultivée dans ces conditions aura une meilleure capacité de résistance aux procédés de stabilisation et conservation (congélation, lyophilisation), et disposera donc d'une bioactivité supérieure lors de son utilisation en tant que ferment. Enfin, la dernière étape de ce programme sera d'évaluer l'origine possible des différences métaboliques détectées, et en particulier une origine peptidique, afin si possible de corrélérer les résultats de peptidomique précédents avec des dynamiques métaboliques. L'objectif sous-jacent étant d'explorer la question de la régulation de fonctions cellulaires par les peptides des EXLs, et de son pilotage.

Pour répondre à ces différentes questions, deux approches globales ont été mises en œuvre afin de réaliser un suivi du métabolisme bactérien à l'échelle du transcriptome et du protéome. Le processus analytique général employé est décrit dans la Figure 4.1. Brièvement, celui-ci utilise sur les mêmes prises d'essai réalisées lors de l'analyse peptidomique (prélèvements entre 3 h et 5 h de culture). Chaque échantillon a *in fine* produit trois jeux de données différents : deux en protéomique et un en transcriptomique. Les jeux de données de protéomique correspondent à deux fractions : une fraction de protéines majoritairement cytoplasmiques (PC), et une autre fraction enrichie en protéines liées aux enveloppes (PE). Les données de transcriptomique correspondent pour leur part à la fusion de deux banques

différentes (se référer au Matériel et Méthodes suivant pour plus de détails). L'analyse de ces données s'est faite selon deux approches différentes : une analyse différentielle classique, afin de détecter les transcrits et protéines dont l'abondance a été significativement affectée par les EXLs, et ensuite *via* une intégration multi-omique afin de dégager les gènes et protéines variant de façon coordonnée dans l'ensemble des échantillons. Cette dernière démarche équivaut à identifier la signature moléculaire des EXLs.

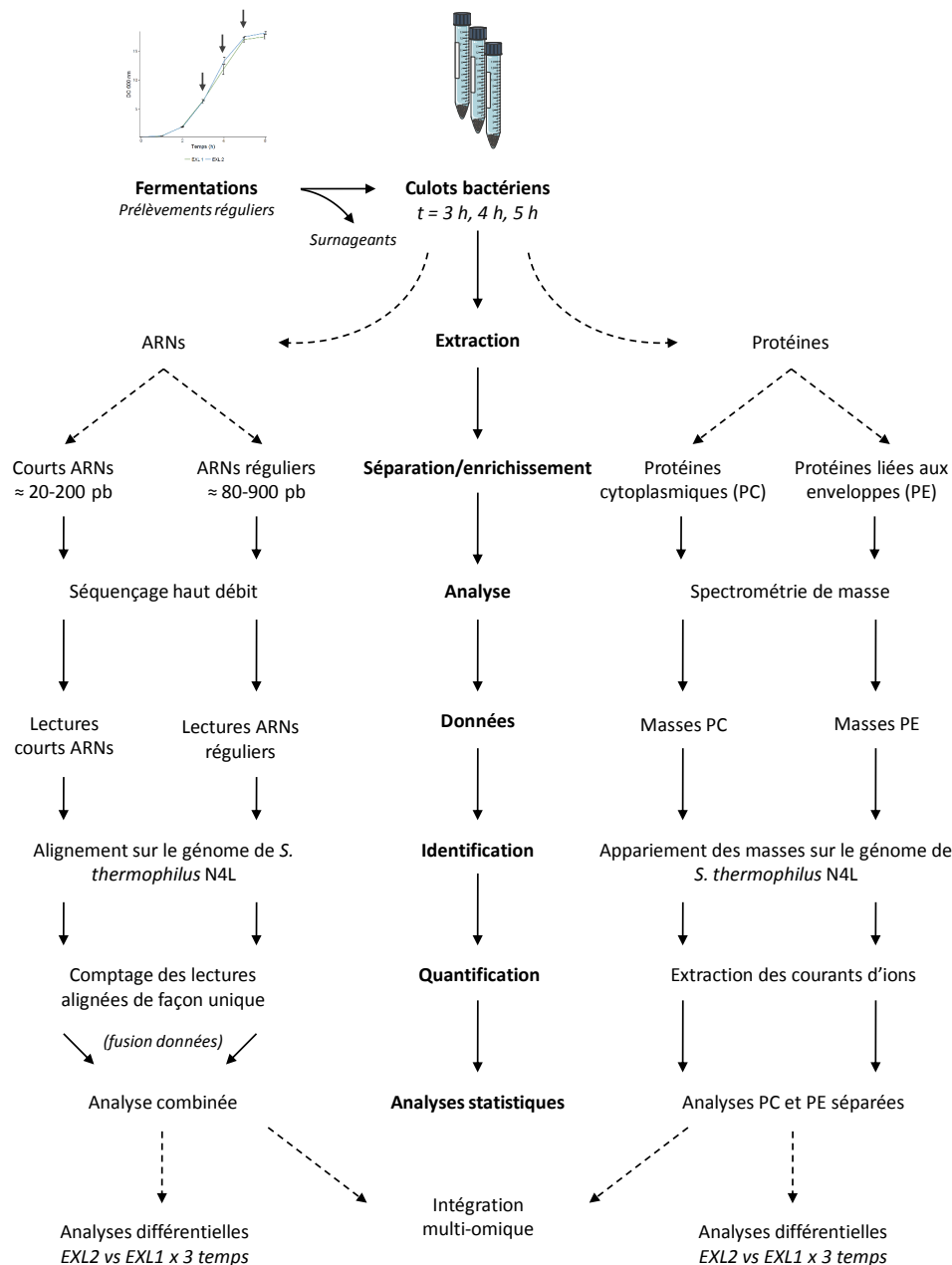


FIGURE 4.1 : Démarche expérimentale mise en œuvre lors des analyses transcriptomiques et protéomiques globales pendant de la croissance de *S. thermophilus* N4L.

4.2 Matériel et méthodes

4.2.1 Conditions de culture et de prélèvements

Les échantillons ayant servi de base à cette partie étant les mêmes que ceux utilisés pour l'analyse peptidomique, le détail des conditions expérimentales est indiqué dans le Matériel et Méthodes du manuscrit soumis pour publication (partie 3.3.3).

4.2.2 Séquençage RNAseq et analyses transcriptomiques

Les aliquots de culture prélevés à 3 h, 4 h et 5 h de croissance et destinés au séquençage des ARNs ont été rapidement centrifugés (18000 g, 15 sec), et les culots bactériens obtenus ont été plongés dans de l'azote liquide et stockés à -80 °C. Au moment de leur utilisation, ces cellules ont été ensuite lysées avec un FastPrep FT120 (Thermo Savant, USA). Pour ce faire, les aliquots congelés ont été tout d'abord repris dans un mélange de Tris 10 mM pH 7 et d'EDTA pH 8. Les suspensions cellulaires ont ensuite été mélangées avec du phénol-chloroforme 5 :1 (Sigma-Aldrich, USA) et des billes de lyse (Sigma-Aldrich, USA). Deux cycles de désintégration cellulaires ont été utilisés à 6,5 m/sec pendant 45 sec, et les débris cellulaires ont été culottés par centrifugation (16000 g, 10 min, 4 °C) de façon à pouvoir récupérer la phase aqueuse contenant les ARNs. L'extraction des ARNs totaux a ensuite été réalisée par l'intermédiaire du kit Direct-ZolTM RNA MiniPrep (Zymo Research, USA) suivant les instructions du fabricant. L'ADN génomique a été retiré grâce au kit DNA-freeTM Removal (Invitrogen, USA). Une sélection de taille sur gel d'agarose a ensuite été réalisée afin de séparer les courts transcrits (\approx 20-200 bp) des transcrits totaux. Ces derniers ont ensuite déplétés des ARNs ribosomiques par le kit Ribo-Zero rRNA removal (Epicentre, USA). Cette déplétion n'a en revanche pas été effectuée pour les courts transcrits. Enfin, la quantité finale et la qualité des deux préparations ont été évaluées respectivement par un fluorimètre Qubit (ThermoFisher Scientific, USA) et un système Bioanalyzer (Agilent, USA).

A partir de ces échantillons, deux banques de séquençage directionnelles ont été construites. La banque correspondant aux petits ARNs a été obtenue *via* le kit TrueSeq Preparation (Illumina, USA), et celle contenant les transcrits standards grâce au kit ScriptSeq Preparation (Illumina, USA). Le séquençage a ensuite été effectué sur un système Illumina NextSeq 500 (Illumina, USA). La taille des lectures choisie était de 75 pb et le mode de lecture en « *single-end* ». Les lectures brutes obtenues ont ensuite été traitées de la façon suivante. Elles ont été premièrement démultiplexées et compilées sous un format fastq *via* l'outil bcl2fastq2 (version 2.15.0). Puis, elles ont été débarassées de leurs adaptateurs avec Cutadapt (version 1.9.1). La qualité des lectures a été mesurée avant et après l'élimination des adaptateurs par Fastqc (version 0.11.5). BWA (version 0.6.2-r126) a ensuite été utilisé afin d'aligner les

lectures sur le génome de *S. thermophilus* N4L. Enfin, l'abondance relative des transcrits a été estimée par comptage des lectures s'alignant de façon unique sur chaque CDS prédits par l'outil Agmial (2215 CDS) ainsi que sur les sORFs identifiés par SHOW (679 sORFs) *via* l'option featureCounts du paquet Subread (version 1.5.2). Seules les lectures présentant un score de qualité supérieur à 10 ont été comptabilisées.

Après cette étape, les tables d'abondance issues des données des transcrits courts et réguliers ont été fusionnées en une seule table. Le seuil de taille choisi était de 180 nt (60 acides aminés). Plus explicitement, *in fine*, les données d'abondance issues de la banque des courts transcrits ont été utilisées pour la quantification de 363 CDS d'Agmial et les 679 sORFs de SHOW dont les produits ont une taille inférieure à ce seuil, et inversement pour la banque des transcrits réguliers, utilisée pour quantifier les 1852 CDS d'Agmial restants. Ces données fusionnées ont ensuite été importées dans l'environnement R de la suite SARTools (version, 1.6.3; Varet *et al.*, 2016) afin de procéder à l'analyse différentielle par l'intermédiaire de DESeq2 (version 1.18.1; Love *et al.*, 2014). L'expression génétique de la souche a été comparée entre les deux EXLs de façon séparée pour chacun des trois points de prélèvement. Les données d'abondance ont été normalisées (paramètre de normalisation « *median* », par défaut) dans chaque échantillon avant l'analyse différentielle proprement dite. Les *p*-values produites ont ensuite été ajustées par une procédure FDR (Benjamini et Hochberg, 1995). Les gènes montrant une *p*-value ajustée ≤ 0.05 ont été considérés comme différentiellement exprimés entre les deux EXLs.

4.2.3 Identification du protéome

Les aliquots des prélèvements précédemment décrits et destinés à l'analyse protéomique ont tout d'abord été centrifugés (3000 g, 10 min, 4 °C) afin de récupérer la biomasse bactérienne. Celle-ci a été lavée deux fois dans une solution de NaPO₄ 200 mM pH 6,8, puis congelée dans de l'azote liquide et conservée à -80 °C. Au moment de leur utilisation, les cellules ont été décongelées dans un tampon phosphate (Na₂HPO₄ 200 mM + NaH₂PO₄, pH 6.4) contenant un cocktail d'inhibiteurs de protéases utilisé suivant les préconisations du fournisseur (Sigma-Aldrich, P8465, USA). Elles ont ensuite été lysées dans un désintégrateur de cellules BAZIC Z (CellID, Warwickshire, U.K.) à 2,5 kBar, et les débris ont été retirés par centrifugation (21000 g, 20 min, 4 °C). Les lysats ont ensuite été ultracentrifugés (200000 g, 30 min, 4 °C) de façon à « culotter » une fraction enrichie en protéines liées aux enveloppes (PE). Ces culots enrichis en PE ont été séparés du surnageant contenant majoritairement des protéines cytoplasmiques (PC), et ont été resuspendus dans le tampon phosphate contenant les inhibiteurs de protéases précédemment employé. La concentration en protéines dans les surnageants PC et fraction PE a été évaluée par le kit Bradford Protein Assay (ThermoFisher Scientific, USA). Les protéines ont ensuite été mélangées avec 2 % de SDS, 6 % de glycérol, 25

mM de DTT, 50 mM de Tris pH 6,8 et 0,1 % de bleu de bromophénol (concentrations finales). Un total de cinq microgrammes de protéines ont été chargés sur des gels NuPage 4-12 % Bis-Tris (Novex, Invitrogen, USA) dans du tampon NuPage MES SDS (ThermoFisher Scientific, USA). Un courant de 200 V, 110 mA a été appliqué le temps d'une migration courte (environ 1 cm) des protéines dans le gel. Les gels ont ensuite été colorés avec du SimplyBlue™ SafeStain (ThermoFisher Scientific, USA) pendant 60 min puis lavés dans de l'eau Milli-Q. Les bandes de protéines ont été excisées des gels puis découpées en cubes de 1 mm. Ces cubes ont été lavés deux fois afin de retirer la coloration, tout d'abord avec un mélange d'acide acétique 10 % et éthanol 40 % (v/v), puis dans un mélange 50 :50 de NH₄HCO₃ 50 mM et d'acétonitrile (ACN). Puis, ils ont finalement été séchés avec de l'ACN 100 %. La moitié des cubes (2,5 µg de protéines par échantillon) a été utilisée pour la digestion enzymatique. Les protéines ont tout d'abord été réduites et alkylées avec du DTT 10 mM de l'iodoacétamide 55 mM selon le protocole décrit par Gundry *et al.* (2009). La digestion a ensuite été réalisée durant une nuit complète à 37 °C en présence d'une solution de trypsine d'origine porcine dissoute à 10 ng/µl (Promega, USA) dans un tampon NH₄HCO₃ 50 mM. La digestion complétée, les cubes de gel ont été lavés dans le même tampon NH₄HCO₃ 50 mM, et les peptides tryptiques ont ensuite été extraits en les incubant successivement dans 30 µl d'une solution d'acide trifluoroacétique (TFA) 0,5 % (v/v) et ACN 50 %, puis dans 30 µl d'ACN 100 %. Les fractions liquides obtenues lors des deux extractions ont ensuite été mélangées, séchées par évaporation sous vide dans un système SpeedVac (Savant, ThermoFisher Scientific, USA), puis finalement resuspendues dans 30 µl d'une solution de TFA 0,1 % et ACN 2 %.

Cinq µl de cette solution contenant les peptides tryptiques ont ensuite été concentrés sur une pré-colonne Pepmap™ 300 C18 LC (5 mm x 300 µm, granulométrie 5 µm, taille des pores 300 Å, ThermoFisher Scientific, USA) par un système Ultimate 3000 RSLCnano (ThermoFisher Scientific, USA) sous un débit de 20 µL/min de TFA 0,08 % et ACN 2 % pendant 3 min. Après ce temps, les peptides pré-chargés ont ensuite été transférés sur une colonne Pepmap™ 100 C18 (500 mm x 75 µm, granulométrie 3 µm, taille des pores 100 Å, ThermoFisher Scientific, USA) sous un gradient d'ACN à 0,24 %/min pendant 97 min dans de l'acide formique 0,1 % afin d'y être séparés. En sortie de colonne, les peptides ont été ionisés par électro-ébulisation sous un courant de 1,3 kV appliqué à un fin capillaire (10 µm diamètre interne, New Objective, USA). Un spectromètre de masse LTQ-Orbitrap Discovery, couplé à l'électro-ébulisation, a ensuite réalisé l'analyse des peptides sous un mode d'acquisition *data dependent*. Les masses complètes des peptides ont d'abord été mesurées à haute résolution (m/z 300-1400, résolution = 15000) dans l'analyseur Orbitrap. Les 8 ions doublement et triplement chargés les plus intenses ont été sélectionnés pour être fragmentés dans la trappe à ions linéaire par CID, où la masse des ions fragments produits a été mesurée à basse résolution. Les masses des 8 ions précurseurs ont ensuite été exclues pendant 30 secondes.

L'identification des peptides tryptiques puis des protéines parentales a finalement été réali-

sée avec le programme X!Tandem (version 2015.12.15.2, Vengeance) intégré dans le logiciel X!TandemPipeline (version 3.4.3; Langella *et al.*, 2016). La recherche des peptides a été réalisée en *mass matching* sur la séquence protéique de *S. thermophilus* N4L en incluant les contraintes d'identification suivantes : spécificité de clivage tryptique (Arg ou Lys en position C-terminale), carbamidométhylation de la cystéine, état d'oxydation variable de la méthionine, tolérance de masse de ± 10 ppm et $\pm 0,5$ Da respectivement pour les ions parents et fragments. Un seul site de clivage manqué a été autorisé, et un minimum de 4 ions fragments ont été requis pour qu'un peptide soit évalué. Une procédure d'affinage a ensuite été appliquée afin de réaliser une recherche plus approfondie sur les candidats présentant une E-value $\leq 0,01$ afin de rechercher la présence d'une acétylation en position N-terminale. Finalement, tous les peptides présentant une E-value $\leq 0,05$ ont été conservés, et un minimum de 2 peptides par protéine a été considéré comme nécessaire pour permettre l'identification de la protéine parentale. Le taux de faux positifs a été évalué par une procédure FDR opérée sur la banque protéique inverse.

4.2.4 Quantification des protéines et analyse différentielle

Les protéines identifiées ont été quantifiées sur la base de l'aire sous la courbe représentant l'intensité des courants d'ions extraits (XIC) de leurs peptides constitutifs. Pour cela, le logiciel MassChroQ a été employé (version 2.2.1; Valot *et al.*, 2011). Les chromatogrammes des différents échantillons ont tout d'abord été alignés sur la base des identifications communes. Les pics artéfactuels ont été ensuite filtrés, les pics des peptides ont été détectés par la méthode Zivy implémentée dans MassChroQ, et les pics des peptides non fragmentés ont été assignés sur la base de leur couple temps de rétention/valeur de *m/z*. Les valeurs brutes d'intensité des peptides ont ensuite été importées dans le paquet R MassChroQR (version 0.3.8) et transformées logarithmiquement. Une première procédure de filtrage a été appliquée sur la base des données chromatographiques : élimination des peptides dont la variation du temps de rétention entre échantillons était supérieure à 20 sec, élimination des peptides associés à un large pic chromatographique dont le temps d'élution était supérieur à 200 sec, et élimination de toutes les données obtenues après 105 min (fin de l'étape de séparation des peptides, début de l'étape de lavage à l'ACN). Les intensités des peptides ont ensuite été normalisées par le biais d'une méthode adaptée de Lyutvinskiy *et al.* (2013). A cette stade, une deuxième procédure de filtrage a été entreprise : élimination des peptides s'alignant sur de multiples protéines (peptides non spécifiques), et élimination des peptides montrant plus de 5 % de données manquantes. Suite à ces filtres, les protéines associées à un seul peptide restant ont également été retirées en vertu de la règle communément appliquée stipulant au minimum deux peptides par protéine identifiée. Finalement, en prévision de l'analyse statistique différentielle qui n'admet que des données complètes (pas de zéro ou de valeurs

vides), les données d'intensité manquantes restant dans le jeu de données ont été imputées (méthode d'extrapolation basée sur les valeurs prises dans d'autres répétitions), permettant ainsi de calculer les valeurs finales d'abondance protéique.

Préalablement à l'analyse différentielle, toutes les protéines montrant moins de 50 % de variation d'abondance entre les différents échantillons ont été retirées. Une ANOVA a finalement été effectuée sur les protéines restantes. Les p -values obtenues ont été ajustées par FDR (Benjamini et Hochberg, 1995), et toute protéine présentant une p -value ajustée $\leq 0,05$ a été considérée comme variant significativement.

4.2.5 Analyses génomiques complémentaires

Différentes analyses génomiques ont été réalisées : annotation fonctionnelle de numéros KEGG, prédiction de la localisation cellulaire des protéines, enrichissement en sORFs. Les procédures utilisées ainsi que les résultats associés sont disponibles dans le Chapitre 2.

4.2.6 Test de surreprésentation et visualisation des résultats

Un test de surreprésentation a été effectué sur l'annotation fonctionnelle des gènes et protéines déclarés comme différentiellement exprimés entre les deux EXLs. Ce test a été réalisé soit de façon séparée sur les listes de gènes et protéines associés à chaque temps d'analyse, soit en combinant l'ensemble des résultats sur les trois temps analysés. Pour ce faire, la fonction *enrichKEGG* du paquet R ClusterProfile (version 3.6.0; Yu *et al.*, 2012) a été utilisée. Cette fonction est basée sur la stratégie initialement employée par Boyle *et al.* (2004) pour la *Gene ontology*. Brièvement, un test hypergéométrique est utilisé afin de déterminer si la fréquence d'un terme d'ontologie mesurée dans un jeu de données particulier est significativement plus importante que dans un contexte (*background*) plus large. Les p -values ainsi produites pour chacun des termes identifiés ont ensuite été ajustées par FDR. Une fonction moléculaire est considérée comme significativement surreprésentée lorsque la p -value ajustée est inférieure ou égale à 0,05. Ici, le contexte utilisé pour évaluer la fréquence de distribution des termes KEGG est celui de l'ensemble de l'orthologie KEGG.

La visualisation des voies et fonctions surreprésentées dans les listes de gènes et protéines différentiellement exprimés a été réalisée par le paquet R pathview ainsi que par l'interface web du même outil (Luo et Brouwer, 2013; Luo *et al.*, 2017).

4.2.7 Intégration multiomique par analyse multivariée

En complément de l'analyse différentielle, une approche multivariée a été utilisée afin d'exploiter les résultats des analyses « omiques » de façon intégrée. Pour cela, le module DIABLO (*Data Integration Analysis for Biomarker discovery using Latent cOmponents*) du paquet R *mixOmics* (version 6.3.2) a été utilisé (Singh *et al.*, 2018). Cette approche d'intégration multiomique a été développée ici sur les trois jeux de données (RNAseq, fractions PC et PE de l'analyse protéomique) en combinant l'ensemble des tables d'abondance provenant des différents temps d'analyse et répétitions. Les jeux de données utilisés correspondent aux résultats normalisés obtenus respectivement *via* DESeq2 et MassChroqR. Les données de RNAseq ont été préalablement filtrées afin d'en retirer les sORFs identifiés par l'outil SHOW, ainsi que tous les gènes restants dont la valeur d'abondance minimale était strictement inférieure à 100. Les données de protéomique n'ont en revanche pas été filtrées.

L'analyse DIABLO reposant notamment sur la régression PLS et l'élaboration d'un modèle sPLS-DA, la procédure suivie est donc très proche de celle présentée dans la partie 3.3.5.1. Succinctement, une phase préliminaire est tout d'abord nécessaire afin de sélectionner le type de distance ainsi que le nombre optimal de composantes permettant de maximiser les performances (i.e. minimiser le BER) d'un modèle complet sans sélection de variables (fonction *perf.diablo* : validation croisée, *folds* $k = 6$, répétitions $n = 10$). En revanche, à la différence d'une simple analyse sPLS-DA univariée, une matrice dite de *design* doit être renseignée afin de régler le niveau de corrélation souhaité entre les différents jeux de données. Ici, une corrélation maximale a été recherchée entre les trois jeux de données *via* la matrice suivante :

$$\mathbf{Design} = \begin{array}{ccc} & \text{RNAseq} & \text{PC} & \text{PE} \\ \left[\begin{array}{ccc} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{array} \right] & \text{RNAseq} & \text{PC} & \text{PE} \end{array}$$

Une fois ces paramètres fixés, une deuxième étape intermédiaire doit avoir lieu afin de choisir la combinaison optimale de variables pour chacune des composantes sélectionnées. Ce dernier paramètre déterminé, le modèle final a été élaboré. Son niveau de performance a été évalué par validation croisée ($k = 6$, $n = 20$) et différentes sorties graphiques ont été produites par l'intermédiaire des fonctions *ad hoc* fournies par le paquet.

4.3 Résultats

4.3.1 Analyses générales

Le séquençage du transcriptome de *S. thermophilus* N4L pendant sa croissance en bioréacteurs a permis de produire environ 28 et 23 millions de lectures par échantillon respectivement issues de la banque des transcrits réguliers et celle des courts ARNs (Figure 4.2). Dans la première banque, environ 84 % de ces lectures ont été alignées de façon non ambiguë sur le génome de la bactérie. Ce ratio tombe à environ 7 % pour les lectures de la banque des courts transcrits en raison de l'absence de déplétion des ARN ribosomiques durant la préparation de cette banque. Après comptage des lectures alignées, les données d'abondance issues des deux banques ont été fusionnées sur la base de la taille des produits des différents gènes et ORFs putatifs (voir Matériel et Méthodes). Dans le jeu de données final, qui inclut donc les valeurs de quantification de l'ensemble des CDS et sORFs à travers les différents échantillons, 4 % du total (128 éléments) ne disposaient d'aucune valeur d'abondance (aucune lecture alignée), et 38 % (1084) ont montré une abondance strictement inférieure à 100 quel que soit l'échantillon considéré.

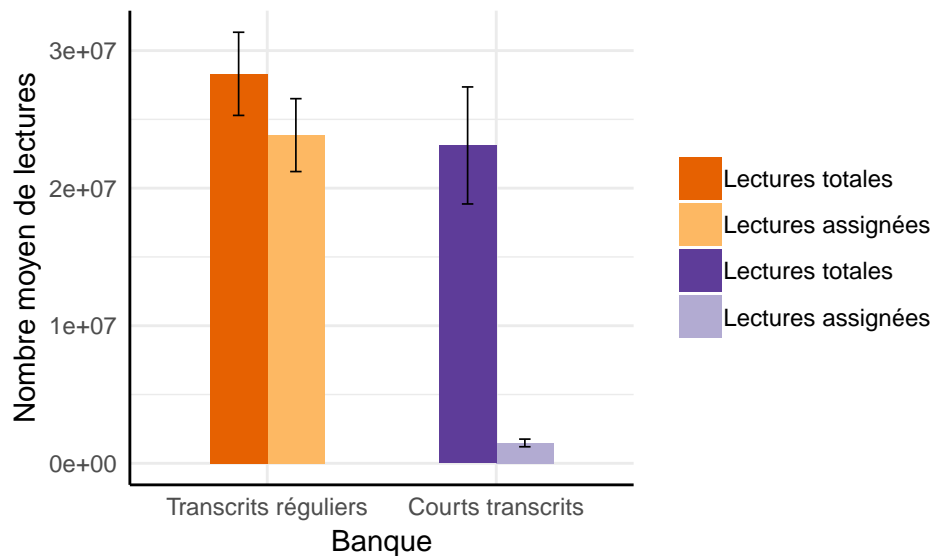


FIGURE 4.2 : Résultats généraux du séquençage du transcriptome de *S. thermophilus* N4L. Nombre moyen par échantillon de lectures séquençées et alignées sur le génome de *S. thermophilus* N4L en fonction de la banque de départ.

En ce qui concerne l'analyse du protéome de la bactérie, 530 et 714 protéines ont été identifiées et quantifiées respectivement dans les fractions PC et PE et pour lesquelles des valeurs d'abondance sont disponibles pour l'ensemble des échantillons. Un total de 478 protéines ont été identifiées dans l'une et l'autre des deux fractions, 52 protéines sont spécifiques de la fraction PC, et 236 de la fraction PE (Figure 4.3). Sur ce dernier nombre, plus de la moitié des

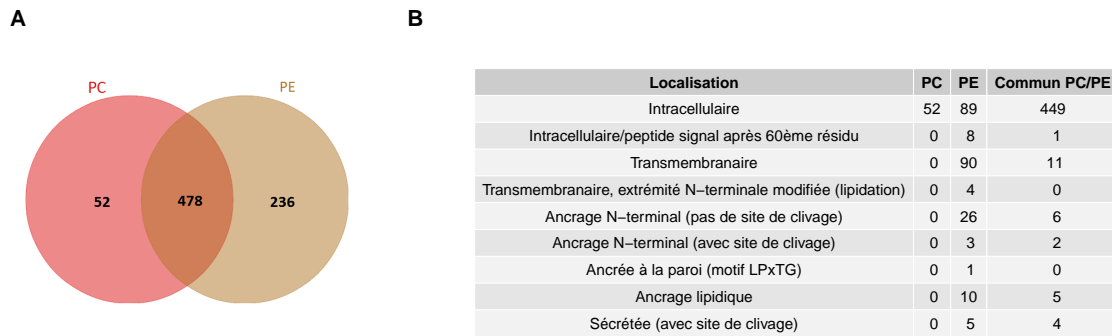


FIGURE 4.3 : Résultats généraux de l'analyse protéomique (A) Diagramme de Venn représentant le nombre total de protéines quantifiées communes ou spécifiques aux deux fractions analysées. (B) Nombre de protéines quantifiées en fonction de leur localisation cellulaire et de leur fraction d'origine. **PC** Protéines cytoplasmiques, **PE** Protéines liées aux enveloppes.

protéines (139) sont prédites par LocateP (Zhou *et al.*, 2008) comme n'étant pas localisées dans le cytoplasme. Ceci montre que la fraction PE est belle est bien enrichie en protéines liées aux enveloppes. Au total, ce sont donc 766 protéines distinctes qui ont été quantifiées durant ces analyses protéomiques. Ce chiffre représente 35 % des CDS identifiés par Agmial dans le génome de la bactérie (41 % des CDS de l'annotation du NCBI, 47 % si l'on retire les prédictions de pseudogènes). Un total de 599 protéines à probable localisation cytoplasmique ont été quantifiées, soit 78 % de l'ensemble. Ce ratio est du même ordre que celui du protéome complet prédit (Tableau 2.2), et il en est de même pour les protéines associées aux autres localisations cellulaires.

Des ACP ont été réalisées à partir des tables d'abondance du transcriptome et des protéomes identifiés dans les fractions PC et PE (Figure 4.4). Les graphes des individus - ici les différents échantillons - montrent un effet prépondérant du temps de prélèvement sur la structure des données d'abondance tant à l'échelle transcriptomique que protéomique. La première composante des différentes ACP porte en effet systématiquement l'information du facteur temps et résume entre 30 % et plus de 40 % de la variabilité totale (Figures 4.4 colonne 1). En particulier, à l'échelle transcriptomique, les échantillons prélevés à 5 h de culture sont très clairement séparés de ceux récoltés à 3 h et 4 h, reflétant très vraisemblablement le passage en phase stationnaire de la bactérie. L'information concernant l'effet des EXLs est portée principalement par la composante 2 au niveau transcriptomique (10,5 % de la variabilité totale), et plus marginalement par la composante 3 (Figure 4.4A2). A l'échelle protéomique, cette information est portée conjointement par les composantes 2 et 3 (Figures 4.4B2 et C2) qui résument à elles deux environ 29 % de la variabilité totale. Tout ceci montre que l'impact de la différence de composition des deux EXLs sur le métabolisme bactérien est plus modéré que celui lié au stade de croissance.

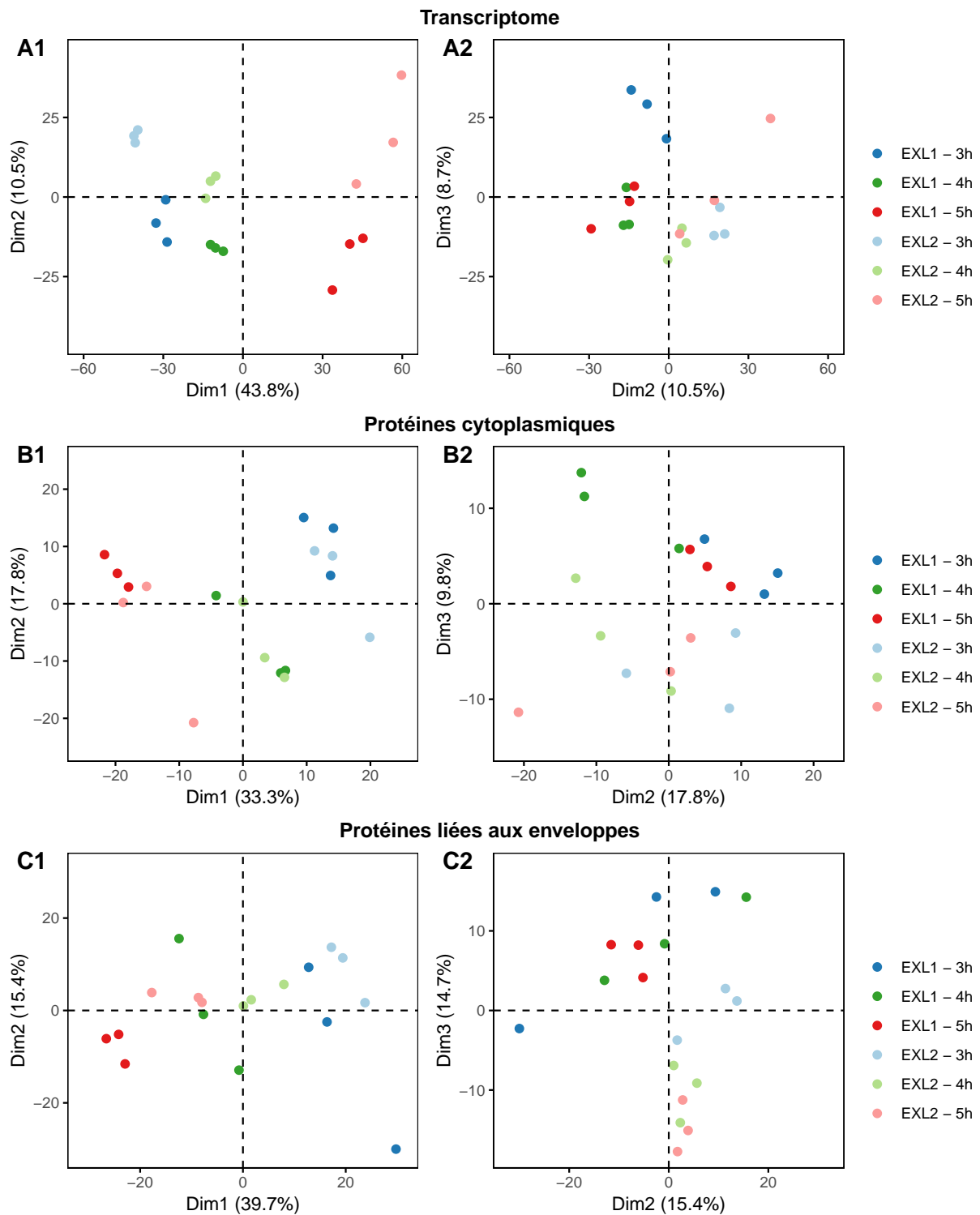


FIGURE 4.4 : ACP basées sur l'abondance des transcrits (A) et protéines (B et C) quantifiés respectivement par comptage des lectures et XIC. Graphes des individus représentés selon les composantes 1 et 2 (colonne 1), et 2 et 3 (colonne 2).

4.3.2 Analyse différentielle entre EXL1 et EXL2

Afin de détecter les transcrits et protéines dont l'abondance est spécifiquement affectée par la nature de l'EXL utilisé, une analyse différentielle (EXL2 *vs* EXL1) a été réalisée pour chacun des trois temps de prélèvement. Les résultats de ces différentes analyses sont regroupés dans le Tableau 4.1. A noter que les résultats des analyses protéomiques correspondent à la combinaison des résultats obtenus pour les fractions PC et PE. Seulement 4 protéines ont été doublement identifiées comme variant significativement entre les deux EXLs dans les deux fractions PC et PE à $t = 3$ h. Etant prédites comme cytoplasmiques, seuls les résultats de la fraction PC ont été conservés pour ces protéines (les résultats équivalents de la fraction PE sont par ailleurs tout à fait similaires). Au total, à l'échelle transcriptomique, 1034 gènes ou sORFs distincts ont été détectés comme variant significativement dans au moins l'un des temps d'analyse au seuil p -value ajustée $\leq 0,05$. Ce total inclut notamment 84 sORFs détectés par l'outil SHOW. Les *fold changes* (FC) estimés par DESeq2 varient entre -61 et +16, mais seulement 139 gènes montrent un $|FC| \geq 2$, et ce chiffre descend à 51 pour $|FC| \geq 4$. Ce dernier total inclut encore 34 sORFs. Cette surreprésentation provient du fait qu'en raison de leur courte taille, relativement peu de lectures sont alignées sur les sORFs par rapport aux transcrits réguliers. Ceci provoque une augmentation artificielle des FC estimés et ce malgré l'opération de compensation (*shrinkage*) réalisée par DESeq2 (Love *et al.*, 2014). Se pose donc la question de la justesse des résultats concernant ces sORFs. Des filtres supplémentaires basés sur la taille et l'abondance moyenne de ces sORFs semblent en tout état de cause nécessaires.

TABLEAU 4.1 : Nombre de transcrits et de protéines différentiellement exprimés (comparaison EXL2 vs EXL1).

Temps	Transcriptome			Protéome ^a		
	3 h	4 h	5 h	3 h	4 h	5 h
Surexprimés	355	155	233	19	14	19
Sous-exprimés	343	137	201	11	20	32
Total	698	292	434	30	34	51
Distincts	1034			91		

^a Combinaison des résultats issus des fractions PC et PE

Au niveau protéomique, ce sont 91 protéines dont l'abondance a été détectée comme significativement différente entre les deux EXLs dans un moins un des temps d'analyse. Sur ce nombre, 79 protéines sont issues des données de la fraction PE, 12 de la fraction PC, et un total de 54 protéines ont également été détectées comme différentiellement exprimées au niveau transcriptomique (recouvrement entre les deux analyses). Les FC estimés varient entre

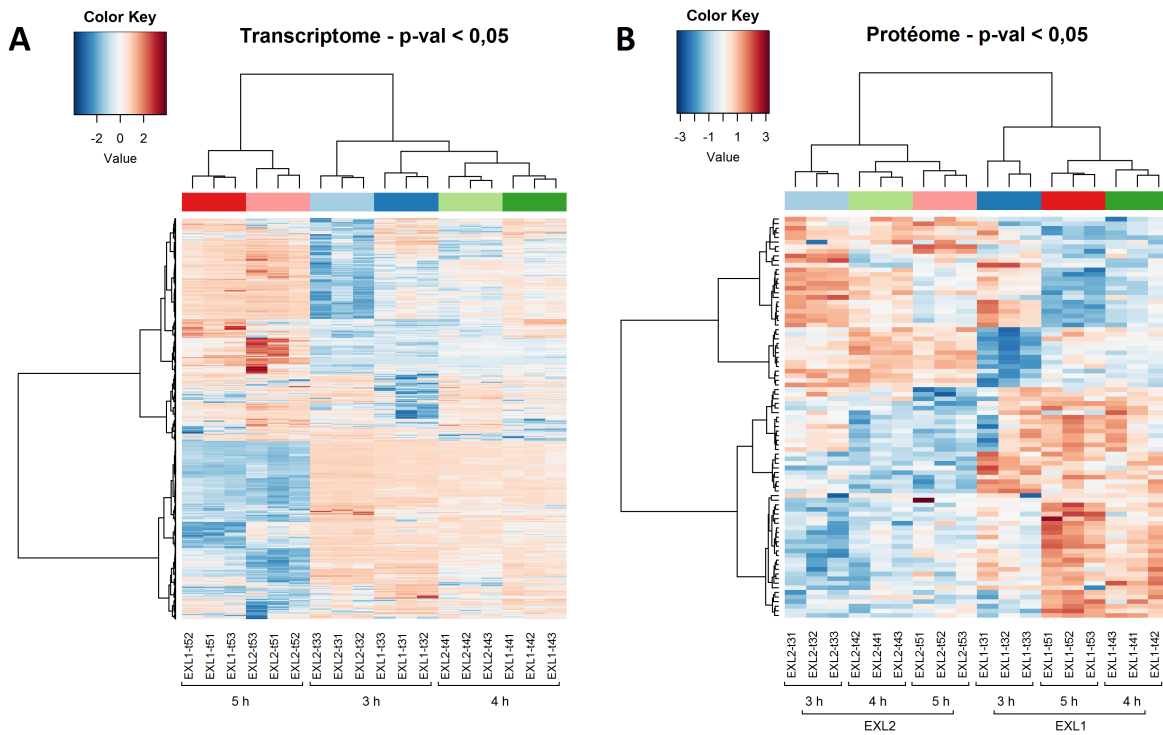


FIGURE 4.5 : *Heatmaps* représentant l'abondance normalisée des transcrits (A) et protéines (B) montrant des variations significatives entre les deux EXLs. Les données ont été préalablement centrées et réduites, les données de transcriptomique ont en outre été transformées afin de les rendre homoscédastiques (fonction *varianceStabilizingTransformation*, package R DESeq2 version 1.18.1). Les dendrogrammes ont été construits par *clustering* hiérarchique (méthode de Ward) basé sur des matrices de distances euclidiennes. Les échantillons sont représentés en colonne et reprennent le code couleur de la Figure 4.4 : bleu = 3 h, vert = 4 h, rouge = 5 h, foncé = EXL1, clair = EXL2.

-4,1 et 2,6, et seulement 19 protéines montrent un $|FC| \geq 2$. Bien que cette gamme de FC soit plus restreinte que celle des données de transcriptomique, il n'est toutefois pas possible de les comparer directement, étant donnée la différence de nature des données d'abondance (données de comptage contre intensité de courants d'ions) et de méthode de calcul des FC.

Les résultats de cette analyse différentielle ont été représentés sous forme de *heatmap* (Figure 4.5). Bien que sélectionnés sur une différence d'expression entre les deux EXLs, les gènes considérés comme significatifs sont encore résolus dans deux clusters structurés sur une base temporelle (Figure 4.5A). De façon similaire aux ACP précédemment réalisées (Figure 4.4), on retrouve ici aussi les échantillons prélevés à 5 h de culture dans un groupe séparé de celui des échantillons récoltés à 3 h et 4 h. Ceci indique que même parmi les gènes significativement affectés par la composition des EXLs, leur expression est encore davantage soumise au stade de croissance de la bactérie. Ceci contraste avec la situation à l'échelle protéomique où l'abondance des protéines détectées par l'analyse différentielle est premièrement influencée par la nature de l'EXL (Figure 4.5B). Ces représentations mettent ainsi en évidence que les analyses transcriptomiques et protéomiques ne reflètent pas exactement

de la même façon l'adaptation de l'activité métabolique de la cellule, ce qui a déjà été mentionné à plusieurs reprises dans la littérature (Perco *et al.*, 2010; Haider et Pal, 2013; Freiberg *et al.*, 2016).

4.3.3 Impact fonctionnel et métabolique des EXLs sur *S. thermophilus* N4L

Les listes de gènes et protéines issues de l'analyse différentielle ont été soumises à un profilage fonctionnel afin mettre en évidence les principales voies métaboliques et fonctions cellulaires dont l'activité est affectée par la composition des EXLs. Pour ce faire, les numéros Kegg (KO) assignés aux gènes et protéines différentiellement exprimés ont été soumis à un test de surreprésentation. Au total, ce sont 516 gènes distincts et 77 protéines qui ont été analysés de la sorte *via* leurs numéros KEGG, les CDS annotés par Agmial dans le génome de la bactérie n'ayant pas tous reçus un numéros KEGG.

Les résultats de cette analyse sont représentés Figure 4.6. Au niveau transcriptionnel, la principale fonction affectée par la composition des EXLs est systématiquement la biosynthèse des acides aminés, quel que soit le temps d'analyse considéré (Figure 4.6A). En combinant l'ensemble des résultats des trois points de prélèvement, ce sont ainsi plus de 60 gènes participant à la biosynthèse des acides aminés qui ont vu leur expression être significativement modifiée par la différence de composition des EXLs. Cet effet est également visible à l'échelle protéomique, bien que de façon plus modérée (total de 9 enzymes dont l'abondance a été modifiée, Figure 4.6B). Cet écart entre transcriptome et protéome est néanmoins cohérent avec la même différence observée entre le total de gènes et protéines différentiellement exprimés (1034 vs 91). Les Figures 4.7 et 4.8 permettent respectivement de localiser les gènes et protéines concernés sur l'ensemble des voies de biosynthèse des acides aminés. Sont ainsi affectés des gènes et protéines participant à la production des acides aminés branchés, de l'histidine ou encore des acides aminés soufrés. Toutefois, deux autres fonctions sont également surreprésentées au niveau protéomique à des niveau comparables à celui de la biosynthèse des acides aminés : le métabolisme des purines (9 protéines), et les protéines ribosomiques (10 protéines). Ces fonctions sont également surreprésentées au niveau transcriptomique, avec 39 gènes appartenant au métabolisme des purines et 28 gènes codant des protéines ribosomiques. On peut de plus relever la présence de gènes participant au métabolisme des pyrimidines surreprésentés à 3 h (16 gènes) et 5 h (11 gènes), ainsi que des gènes codant des aminoacyl-ARNt synthétases à 5 h (9 gènes). Ainsi, de façon plus synthétique, cette analyse a permis de faire ressortir le fait que l'activité de trois grandes fonctions cellulaires a été différentiellement modulée de façon coordonnée tant à l'échelle transcriptomique que protéomique entre les deux EXLs : la biosynthèse des acides aminés, le métabolisme des acides nucléiques et l'activité traductionnelle.

Il faut toutefois préciser que ces différences d'expression et d'abondance protéique entre EXLs restent relativement modérées. Les valeurs de FC mesurées varient en effet principalement entre 1,5 et un peu moins de 4 (valeurs absolues). Ces différences contrastent avec les valeurs largement plus élevées de FC mesurées, par exemple, lors d'expériences de comparaison entre un mutant de délétion et une souche sauvage. Ici, les résultats reflètent plutôt l'adaptation générale de la bactérie aux deux substrats et l'ajustement fin de ses niveaux d'expression en réponse à la différence de composition de son environnement.

Les résultats de ces tests de surreprésentation ne permettent pas en revanche de savoir dans quelle condition l'abondance d'un transcrite ou d'une protéine est augmentée ou réduite, ni son évolution au cours du temps. Pour cela, une étude au cas par cas est nécessaire mais ne sera pas détaillée ici de façon exhaustive en raison de sa lourdeur. A titre d'exemple, seuls les résultats correspondant à la voie de biosynthèse des acides aminés branchés sont donnés dans les Figures 4.9 et 4.10 (se référer à l'Annexe A pour les autres voies d'intérêt). Dans l'ensemble, la biosynthèse des acides aminés branchés semble être stimulée dans l'EXL2 par rapport à l'EXL1. Sont ainsi concernés les gènes *ilvC* (1.1.1.86), *ilvD* (4.2.1.9), ou encore *ilvE/bcaT* (2.6.1.42). Encore une fois, les valeurs de FC restent relativement modestes, le FC maximal ayant été détecté pour la protéine IlvE à 4 h et 5 h (FC = 1,9) dans la fraction PE. Ces résultats sont toutefois intéressants à double titre car i) ils sont globalement cohérents dans les trois jeux de données (ARN, PC, PE), et ii) ils font également écho aux valeurs d'abondance différentielle d'un ABC transporteur annoté comme transporteur des acides aminés branchés (Liv) et le système Ami prenant en charge les oligopeptides (Figure 4.11). Là aussi bien que modérée (FC_{max} = 1,6), la différence d'abondance générale est également en faveur de l'EXL2. Tout ceci (ainsi que d'autres résultats non détaillés ici) semble donc converger vers un mécanisme de régulation commun qui, d'après la littérature, pourrait venir du régulateur pléiotrope CodY.

Si l'abondance de ce régulateur a bien été quantifiée en protéomique, la différence entre les deux EXLs n'a pas été détectée comme significative. Elle l'a en revanche été au niveau transcriptomique, et la cinétique d'expression du gène est montrée dans la Figure 4.12. Celle-ci révèle deux profils d'expression très différents entre les deux EXLs. A 3 h, *codY* est légèrement plus exprimé dans l'EXL1 (FC = -1,3, significatif) mais son niveau reste relativement stable durant la suite de la cinétique. En revanche, dans l'EXL2, l'expression du gène est croissante au cours du temps, ce qui lui permet de dépasser le niveau d'expression dans l'EXL1 dès 4 h (FC = +1,2, non significatif) pour finir à un niveau d'expression double à 5 h (FC = +2,1, significatif). Ces résultats reflètent donc très vraisemblablement l'adaptation de la bactérie en réponse au contenu en azote aminé de son environnement. Tout ceci sera commenté plus en détail ultérieurement.

Chapitre 4 : Etude des variations d'expression génétique et de synthèse protéique en réponse à la composition des extraits de levure

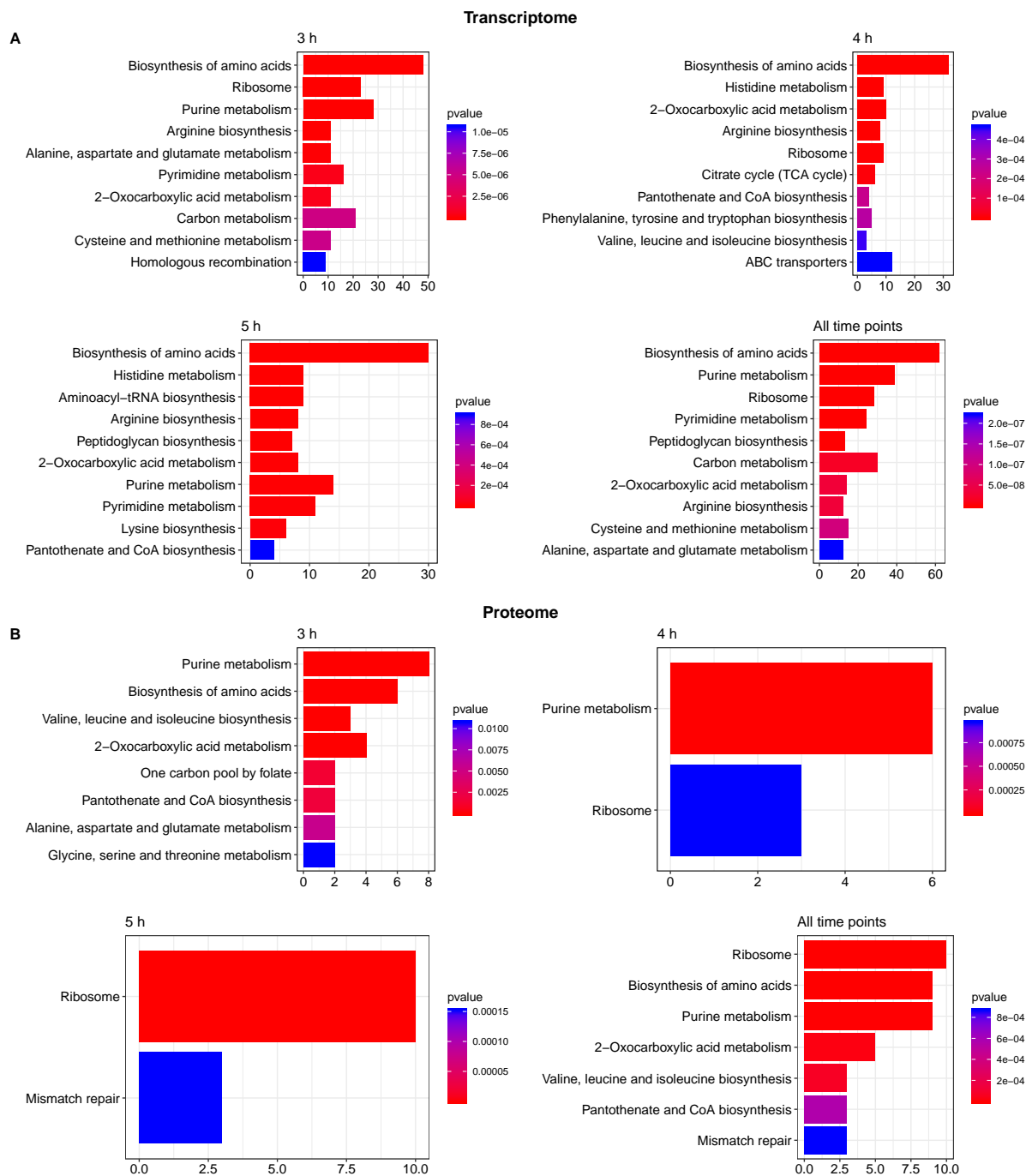


FIGURE 4.6 : Voies métaboliques et fonctions cellulaires surreprésentées parmi les gènes et protéines significativement affectés par la composition des EXLs. Analyse réalisée à partir de l'annotation fonctionnelle du génome par assignation de numéros KEGG. (A) Fonctions surreprésentées au niveau transcriptomique, seuls les 10 premiers résultats les plus significatifs sont affichés. (B) Echelle protéomique, tous les résultats du test de surreprésentation sont ici affichés. Les résultats sont tout d'abord présentés par temps de prélèvement séparément, puis le dernier graphe correspond aux résultats produits par la combinaison de l'ensemble des gènes ou protéines significativement affectés par les EXLs. Le nombre de gènes ou protéines affiliés aux différentes fonctions est porté par l'axe des abscisses.

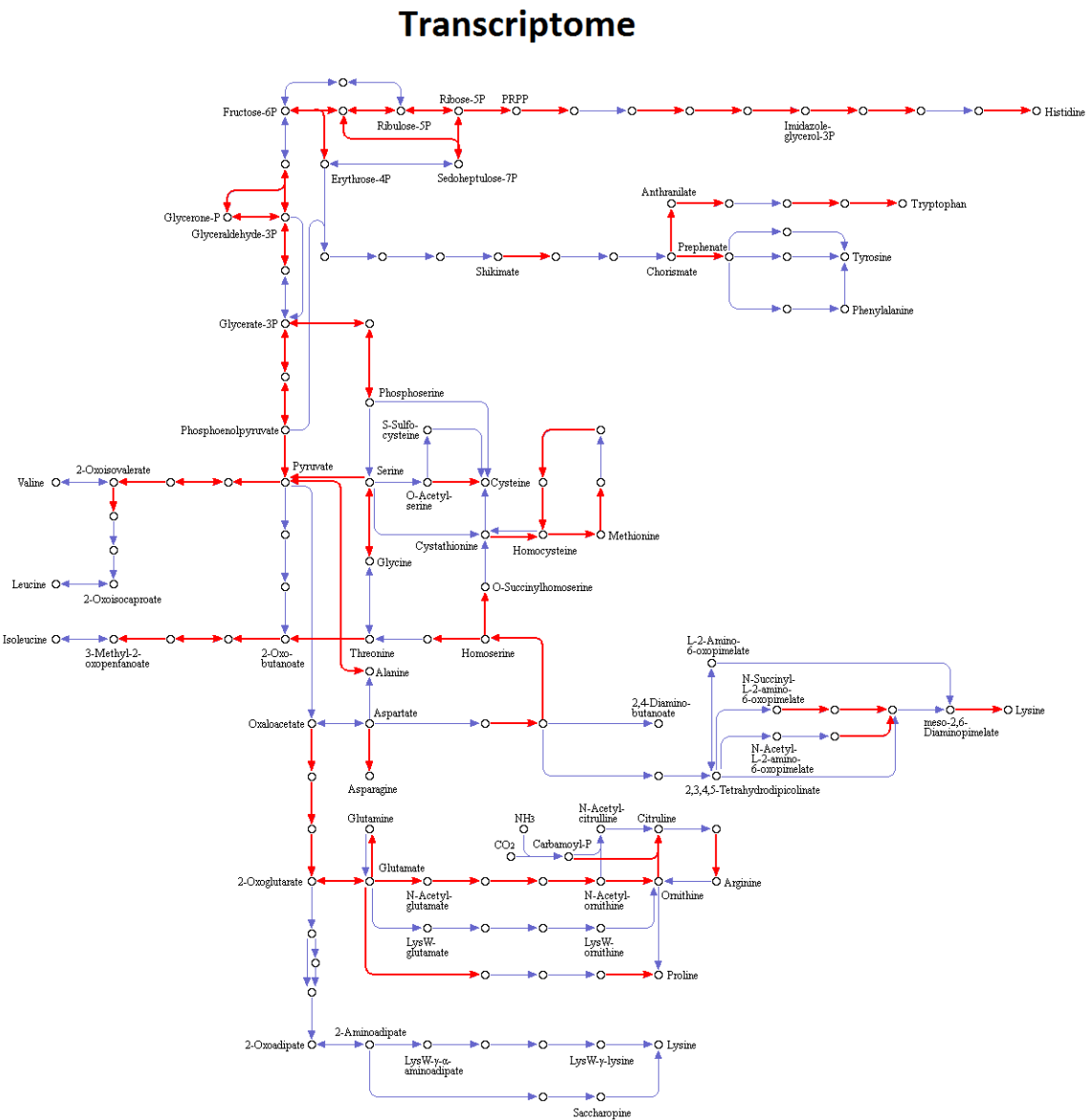


FIGURE 4.7 : Gènes impliqués dans la biosynthèse d'acides aminés dont l'expression globale est différenciellement affectée entre les EXLs. Représentation selon le *pathway* KEGG KO01230, les flèches rouges indiquent les gènes dont l'expression a été détectée comme significativement différente entre les deux EXLs dans au moins l'un des temps de prélèvement.

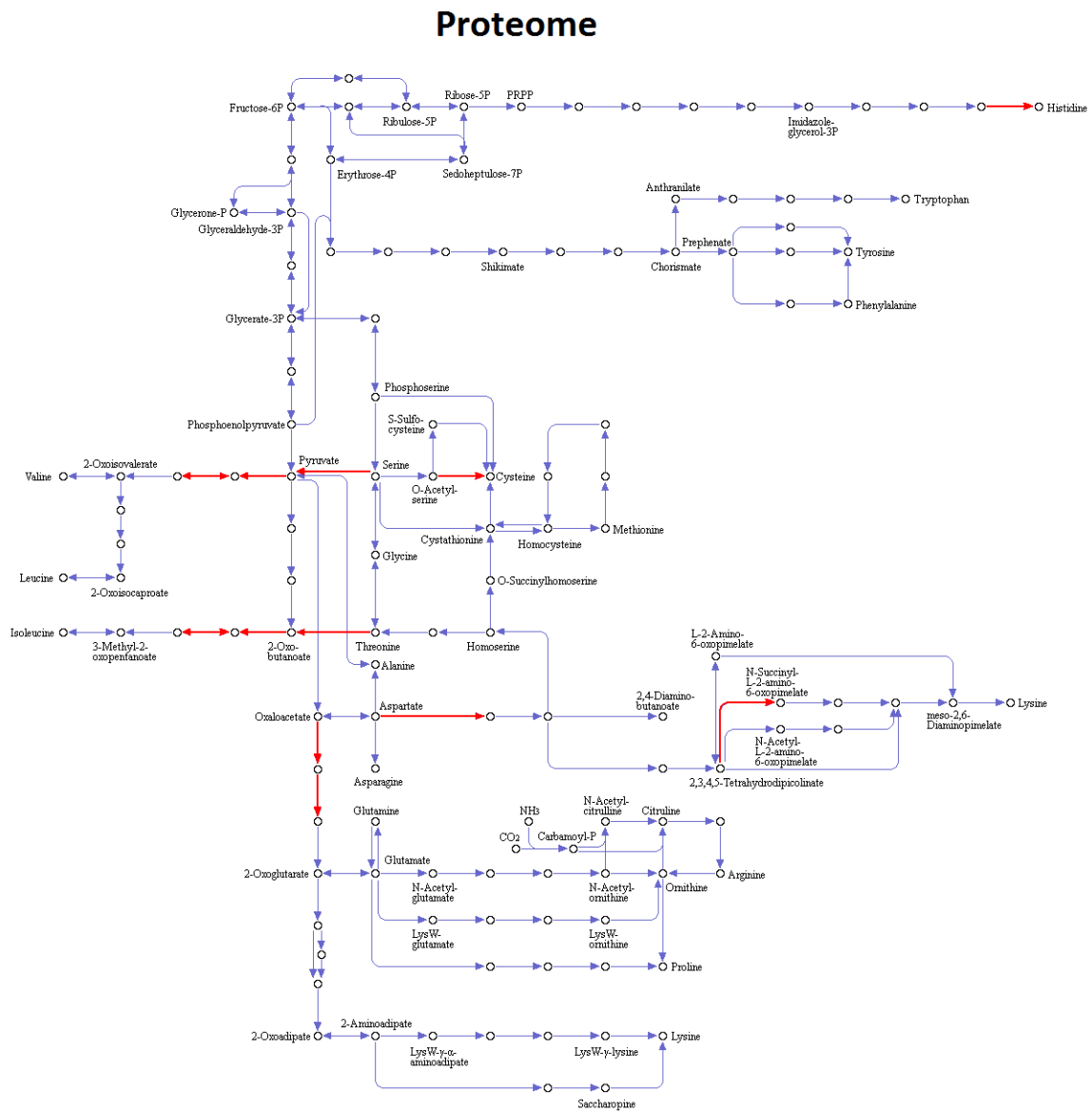


FIGURE 4.8 : Protéines impliquées dans la biosynthèse d'acides aminés dont l'abondance globale est différenciellement affectée entre les EXLs. Les caractéristiques de la figure prennent celles de la Figure 4.7.

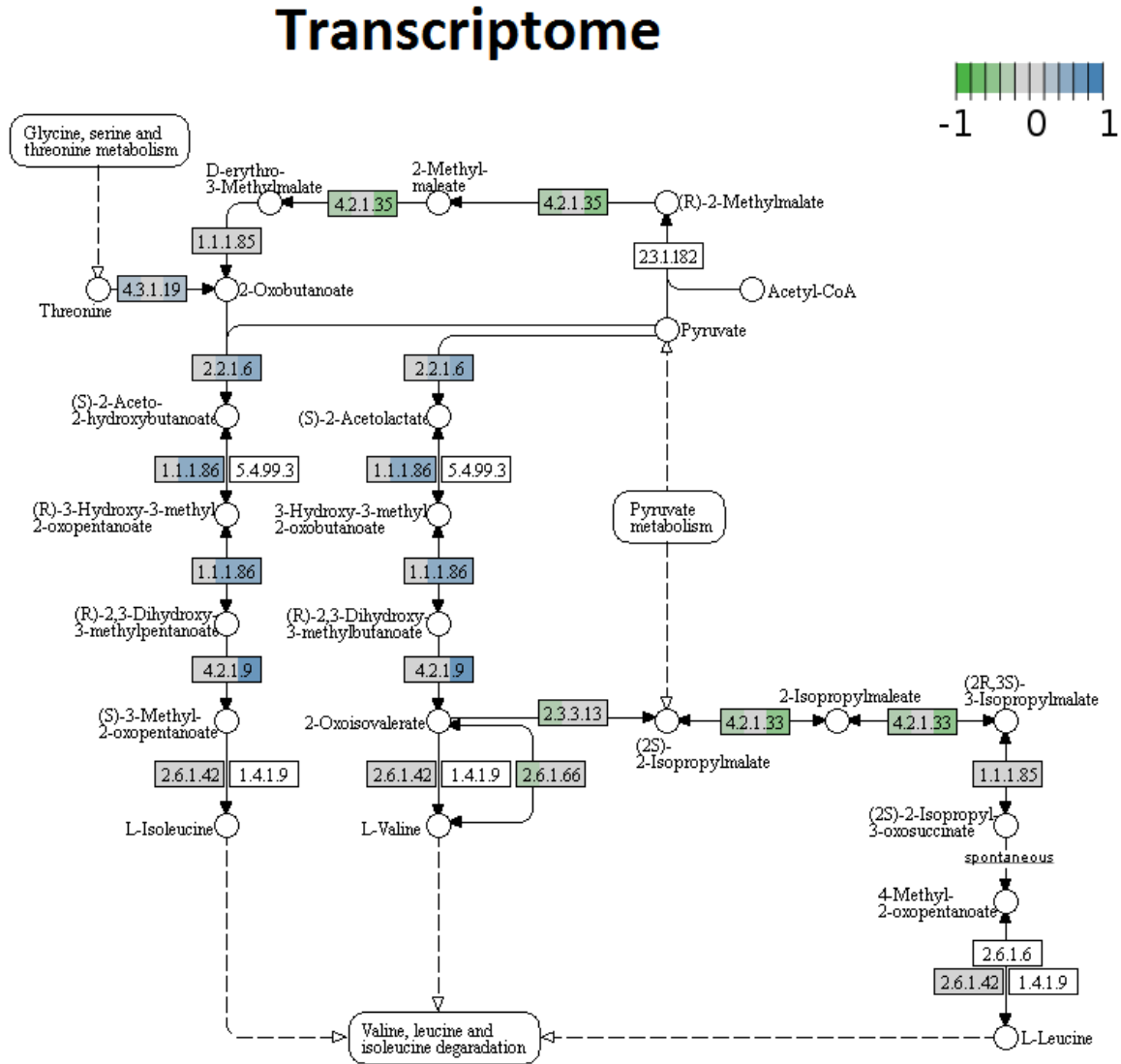


FIGURE 4.9 : Différences d'activité transcriptionnelle au niveau de la voie de biosynthèse des acides aminés branchés entre les deux EXLs. Les enzymes sont référencées par leur numéro E.C. dans un rectangle. Les rectangles colorés indiquent les gène ou les protéines détectés lors des analyses globales. Le code couleur représente le niveau relatif de différence d'abondance entre les deux conditions. Vert : EXL1 > EXL2; gris : EXL1 = EXL2; bleu EXL2 > EXL1; blanc : non détecté. Chaque rectangle est sous-divisé en trois parties, représentant les résultats propres à chaque temps d'analyse, respectivement 3 h, 4 h et 5 h de gauche à droite.

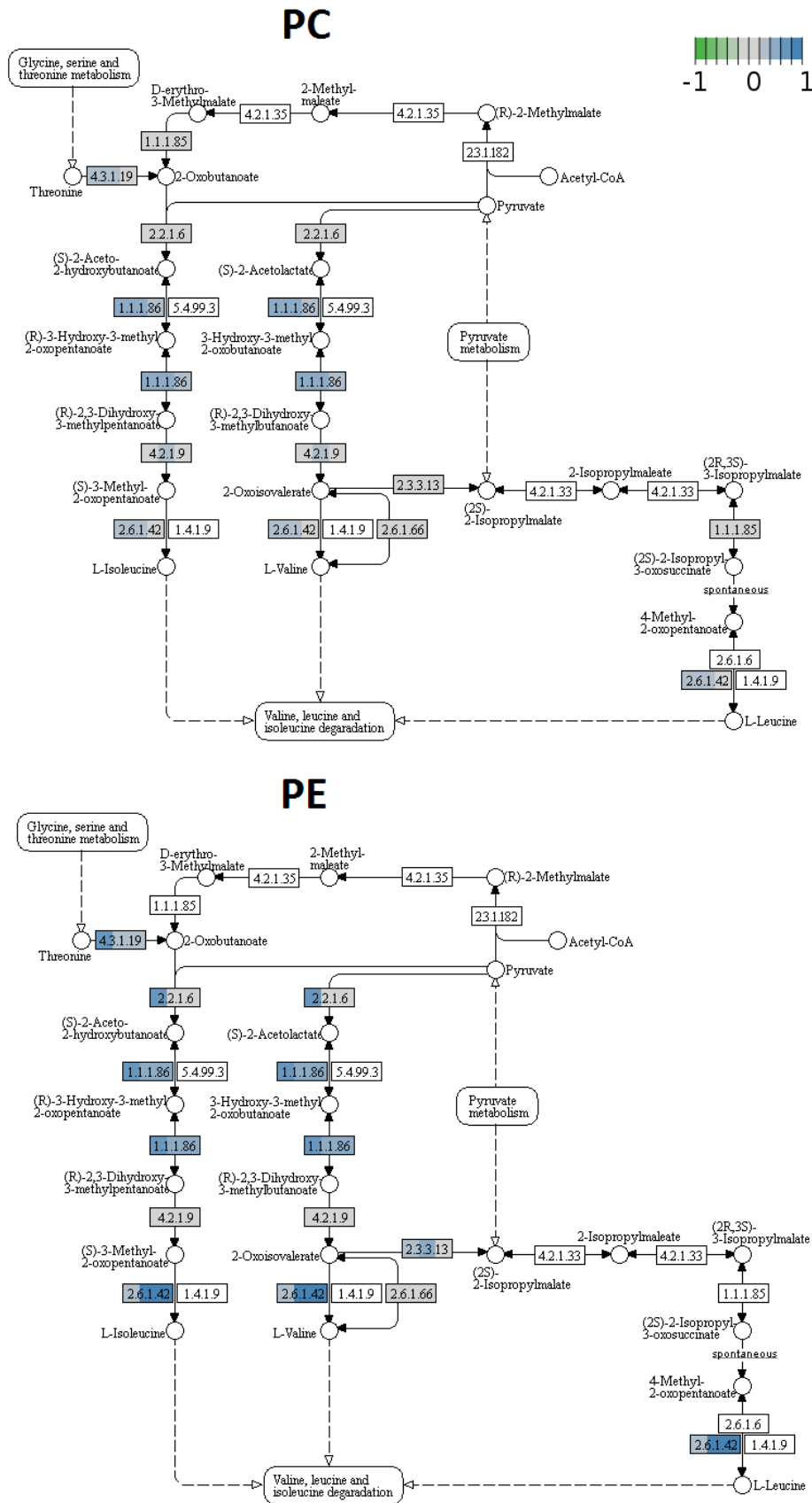


FIGURE 4.10 : Différences d'abondance protéique au niveau de la voie de biosynthèse des acides aminés branchés entre les deux EXLs. Les caractéristiques de la figure reprennent celles de la Figure 4.9.

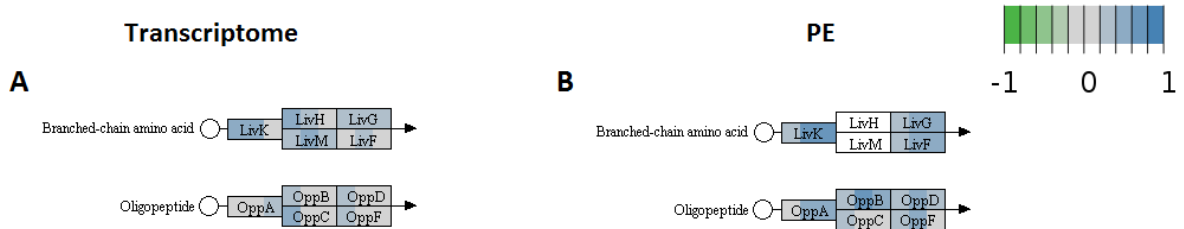


FIGURE 4.11 : Différences d'activité transcriptionnelle (A) et d'abondance protéique (B) au niveau des transporteurs d'acides aminés branchés (Liv) et d'oligopeptides (Ami). Etant donné la localisation membranaire de ces protéines, seuls les résultats de la fraction PE sont présentés ici. Les caractéristiques de la figure reprennent celles de la Figure 4.9.

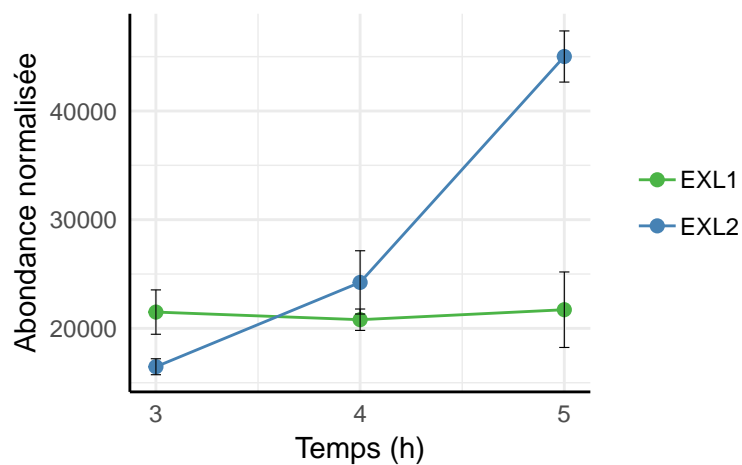


FIGURE 4.12 : Expression de *codY* au cours du temps.

4.3.4 Impact des EXLs sur la régulation par les peptides : le cas du streptide

Les résultats de la partie précédente constituent un bon exemple où un large ensemble de gènes et protéines apparentés varient de façon relativement modérée mais de manière coordonnée. Toutefois, l'analyse différentielle réalisée a également révélé un cas inverse où un nombre plus restreint de gènes ont vu leur expression être profondément modifiée par la composition des EXLs. Plus précisément, le gène ayant montré la valeur de FC la plus extrême durant l'ensemble de la cinétique est le gène *strA* (STN4L_00157) à 5 h avec un FC = -61. Il est de plus assorti d'une intense expression relative (nombre moyen de lectures alignées à 5 h = 71 000 dans l'EXL1). Ce gène, qui est donc largement surexprimé dans l'EXL1 en fin de croissance, est doublement intéressant car il s'agit d'un court ORF codant un peptide modifié post-traductionnellement et son expression est elle-même sous le contrôle d'un système de *quorum sensing*. La Figure 4.13 résume la voie de biosynthèse de ce peptide.

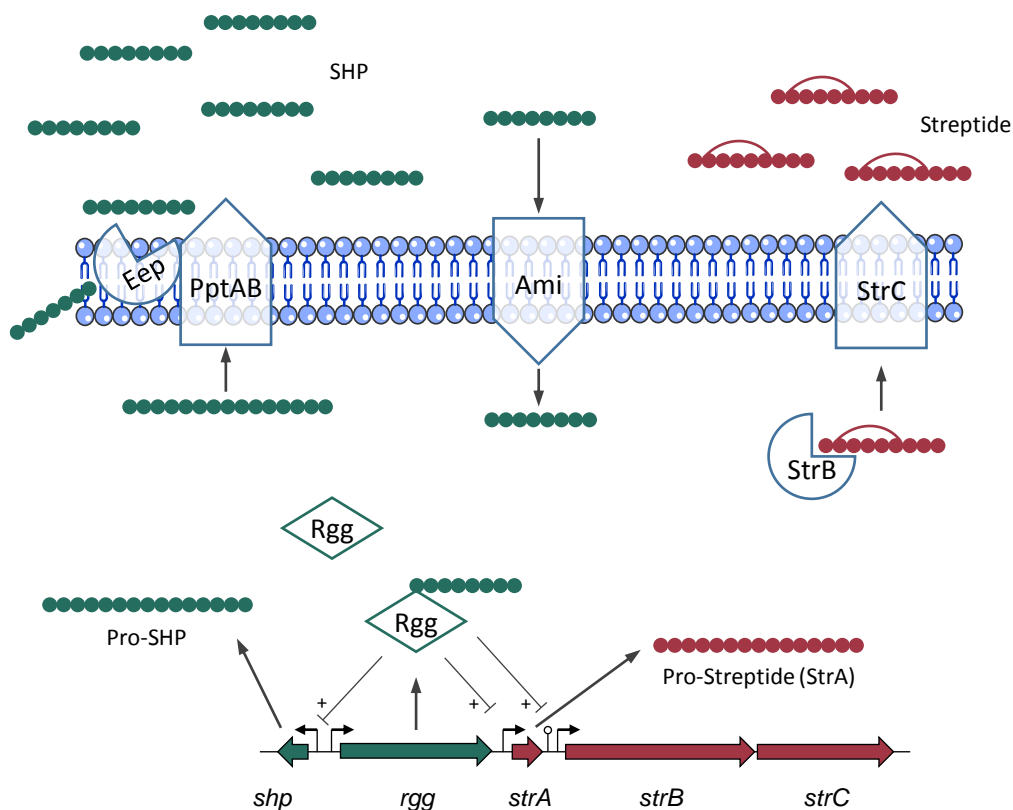


FIGURE 4.13 : (Légende page suivante)

Les données d'expression des différents gènes du locus sont représentées dans la Figure 4.14. A l'exception du régulateur Rgg, les gènes de ce locus suivent des profils similaires, reflétant le mécanisme commun de régulation. Ce locus est ainsi quasiment éteint dans l'EXL2 en

FIGURE 4.13 : Mécanismes de *quorum sensing* gouvernant la production du streptide, un peptide cyclique modifié post-traductionnellement. Le locus de production inclut (en vert) des gènes codant un système de régulation Rgg/SHP, ainsi que leurs gènes cibles (en rouge) : le gène du streptide (*strA*), d'une enzyme à radical SAM (*strB*) responsable de la cyclisation du streptide, ainsi que d'un ABC-exporteur (*strC*) se chargeant de son export. Le SHP est d'abord synthétisé sous la forme d'un pro-peptide, puis exporté et mûri *via* respectivement l'ABC-transporteur PptAB et la protéase Eep. Il s'accumule ensuite dans le milieu extracellulaire jusqu'à atteindre une concentration seuil permettant sa réimportation par le transporteur Ami. Dans la cellule, le SHP mature et le régulateur Rgg s'associent en un complexe capable d'interagir avec les régions promotrices des gènes cibles pour en activer l'expression. Le complexe Rgg-SHP interagit également avec le promoteur du SHP, formant une boucle de régulation positive. Adapté de Ibrahim *et al.* (2007a) et Schramma *et al.* (2015).

comparaison avec l'EXL1 où son expression globale croit fortement au cours du temps. Tous ces gènes (hormis celui du Rgg) font ainsi partie des 5 % des gènes détectés comme étant les plus différenciellement exprimés entre les deux EXLs, celui du streptide arrivant donc en tête. Ces différences n'ont toutefois pas pu être vérifiées au niveau protéomique, aucune des protéines correspondantes n'ayant pu être quantifiée. Ceci est attendu concernant le SHP et le streptide qui ne peuvent être détectés par cette approche en raison de leur taille (trop courte pour être retenue dans un gel SDS-PAGE) et de leur localisation principalement extracellulaire sous forme mature (sans compter la cyclisation du streptide requérant des filtres d'identification modifiés). En revanche, l'absence de quantification de la maturase à radical SAM et de l'exporteur pourrait provenir soit d'une absence complète de détection dans l'ensemble des points d'analyse (abondance systématiquement trop faible par rapport aux autres protéines bactériennes), soit d'une détection partielle dans certains échantillons uniquement. Dans ce dernier cas de figure, le modèle statistique utilisé en quantification par XIC ne tolérant pas les données manquantes, ces protéines ont pu être éliminées du jeu de données durant les diverses étapes de filtrage (voir partie Matériel et Méthodes). On précisera également que la présence du SHP et du streptide a été investiguée dans les données de peptidomique, tant de façon automatique *via* le logiciel X!TandemPipeline que manuellement en recherchant les spectres MS1 correspondant aux masses théoriques des précurseurs. Néanmoins là aussi cette recherche s'est révélée infructueuse, vraisemblablement en raison de la surabondance des peptides de levure dans le milieu. Concernant le SHP, il est également possible qu'il ait été perdu au cours des différentes étapes de préparation du protocole expérimental du fait de sa nature fortement hydrophobe.

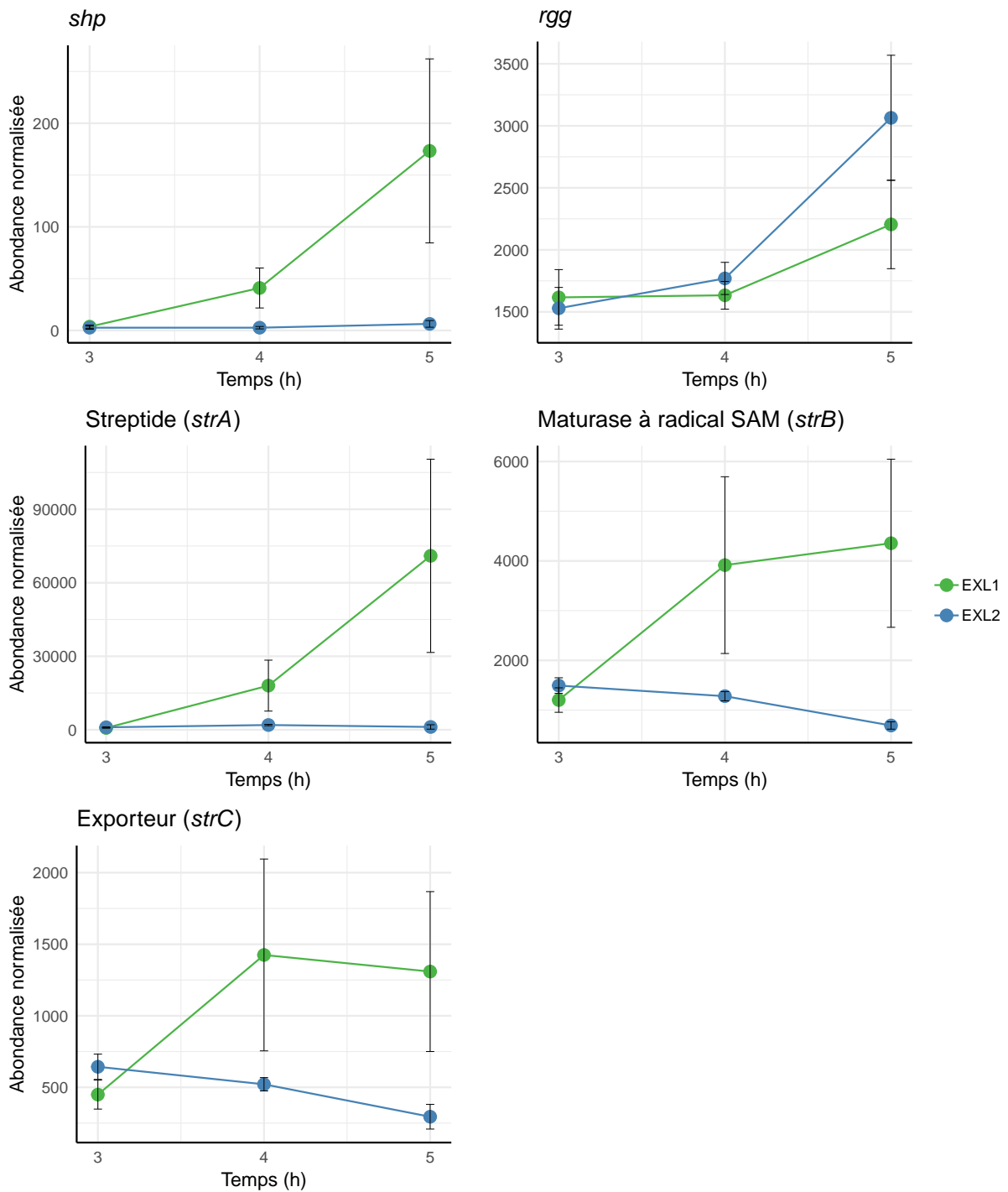


FIGURE 4.14 : Expression du locus streptide au cours du temps.

4.3.5 Signature moléculaire des EXLs par intégration multiomique

Les résultats présentés dans les deux parties précédentes reposent sur une approche par analyse différentielle. Ils sont donc en réalité le produit d'une analyse univariée itérée de multiples fois au travers des différents temps de prélèvement et des différents jeux de données. Ce type de démarche permet ainsi d'identifier de façon très efficace les dynamiques « intra-omiques » qui se produisent donc au sein d'un seul et même niveau moléculaire. Cette approche fait en revanche abstraction des possibles relations « inter-omiques », c'est-à-dire des dynamiques se produisant de façon coordonnée sur plusieurs niveaux moléculaires. En d'autres termes, les analyses univariées ne permettent pas d'identifier les variations corrélées entre, par exemple, le transcriptome et le protéome. Pourtant, d'un point de vue biologique, ce type d'information est très important car il reflète la façon dont un système est coordonné et modulé sur différentes échelles moléculaires. Le passage au haut-débit de nombreuses technologies d'analyse (séquençage, spectrométrie de masse etc.) ayant favorisé les approches multi-omiques, divers outils et méthodes ont logiquement été élaborés afin de répondre à la question de l'analyse intégrée des données produites. Le paquet R *mixOmics* (Rohart *et al.*, 2017) fait partie des outils les plus aboutis à l'heure actuelle et propose de nombreuses méthodes d'analyse multivariée adaptées à différentes configurations. Dans le présent contexte, le module DIABLO a été utilisé car il permet de réaliser une « *N*-intégration », c'est-à-dire une analyse de différents jeux de données omiques produits à partir des mêmes échantillons initiaux. Brièvement, DIABLO permet de sélectionner des variables, ici les gènes et protéines, dont les niveaux d'abondance sont hautement corrélés à travers les différents jeux de données tout en étant fortement associés à un phénotype d'intérêt, ici l'effet des EXLs sur l'activité métabolique de la bactérie. Cette démarche s'attache donc à la caractérisation de la « signature moléculaire » multiomique des EXLs.

Un modèle préliminaire sans sélection de variable a tout d'abord été établi et a permis de vérifier que deux composantes associées à une distance de prédiction de type maximale (« *max.dist* ») permettaient de minimiser le taux moyen d'erreur du modèle. Une grille de combinaisons a ensuite été utilisée afin de sélectionner le nombre optimal de gènes et protéines associés à chaque composante. La grille suivante a été utilisée pour chacun des trois jeux de données : 5, 10, 15, 20, ..., 50. Soit un total de $10 \times 10 \times 10 = 1000$ modèles différents qui ont été ajustés sur chacune des composantes, le tout dans une procédure itérative de validation croisée. A l'issue de cette démarche, un total de 20 gènes (5 sur la composante 1, 15 sur la composante 2), 20 protéines de la fraction PC (10 et 10) et 20 protéines de la fraction PE (15 et 5) ont été déterminés comme étant la combinaison permettant de maximiser les performances du modèle. Celui-ci a ainsi montré de très bonnes capacités de classification avec un BER d'environ 1,5 % seulement (taux d'erreur moyen de 2,2 % pour l'EXL1, 0,7 % pour l'EXL2).

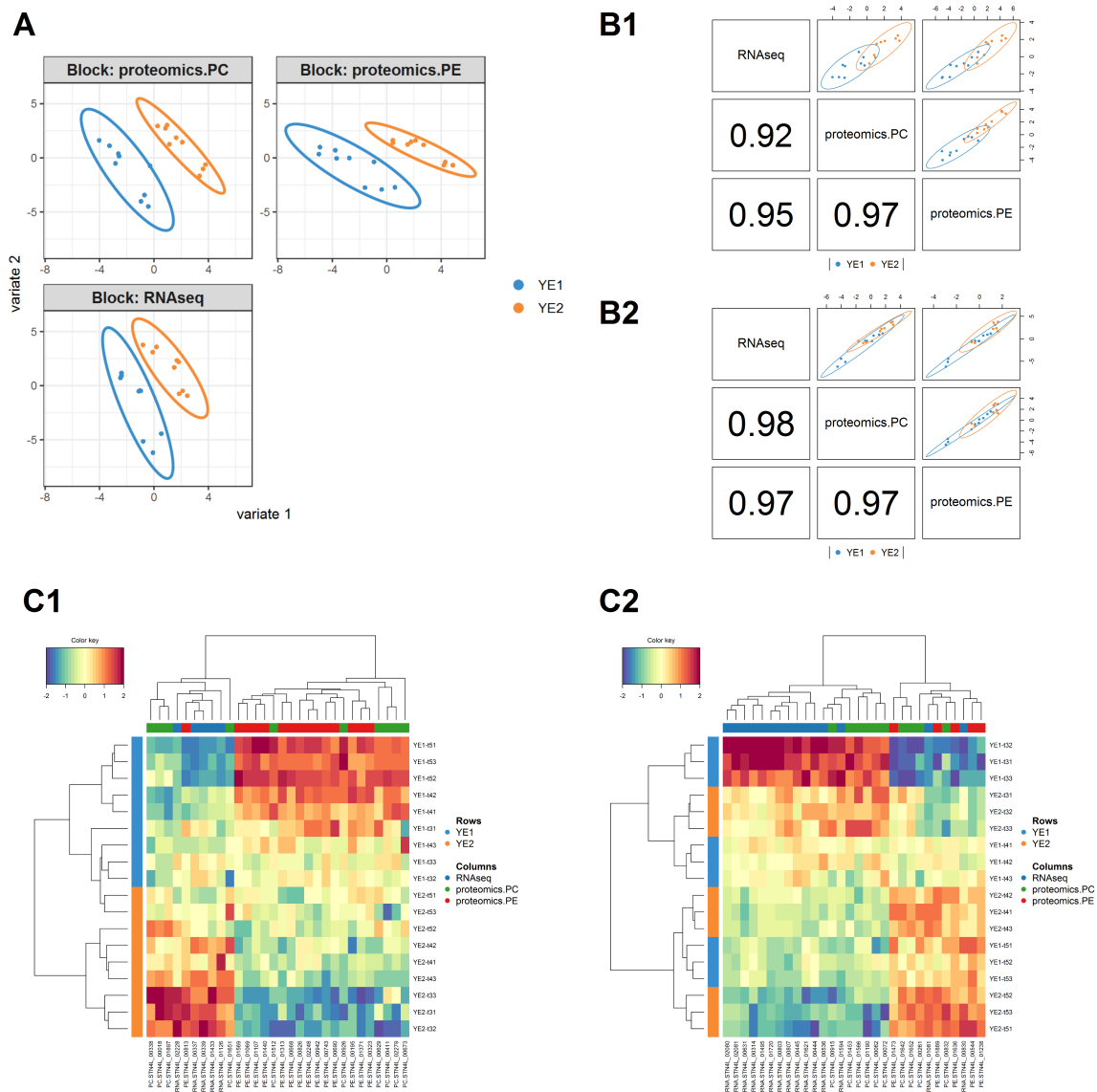


FIGURE 4.15 : Résultats de l'intégration multi-omique obtenue avec DIABLO. (A) Graphes des individus (échantillons) représentés selon leurs coordonnées sur les deux composantes du modèle dans chaque jeu de données séparé (bloc omique). (B) Graphes des individus (diagonale supérieure droite) représentés selon leurs coordonnées sur la première composante (B1) et deuxième composante (B2) et par paire de jeux de données. La lecture s'effectue de façon similaire à une matrice de corrélation, l'intersection d'une ligne et d'une colonne correspondant à la combinaison des blocs associés respectivement à ladite ligne et colonne. Les coefficients de corrélation de Pearson entre chaque paire de jeux de données sont indiqués par symétrie dans la diagonale inférieure gauche. (C) *Heatmaps* (distance euclidienne, méthode du lien complet) représentant l'abondance par échantillon (lignes) des gènes et protéines (colonnes, identifiant Agmial) sélectionnés sur la première composante (C1) et deuxième composante (C2).

Les principaux résultats de l'analyse DIABLO sont résumés dans la Figure 4.15. On peut premièrement relever le fait que ce travail a abouti à une très bonne séparation des échan-

tillons en fonction de l'EXL utilisé (Figure 4.15A). Ceci indique que les éléments moléculaires sélectionnés (gènes et protéines) ont vu leur abondance être affectée par la nature de l'EXL. On notera toutefois que cette séparation est principalement opérée par la première composante du modèle, ce qui est particulièrement visible en comparant la Figure 4.15B1 et la Figure 4.15B2. Cette dernière figure montre en particulier un fort recouvrement entre échantillons sur la base des éléments associés à la composante 2. Tout ceci indique que les deux groupes d'éléments sélectionnés sur les composantes 1 et 2 n'ont vraisemblablement pas été affectés par les EXLs avec la même intensité. En revanche, le niveau de corrélation observé entre les différents « blocs » omiques est très satisfaisant, quelle que soit la composante considérée (coefficients de Pearson $\geq 0,9$). Ceci indique que, au sein de chaque groupe d'éléments, les niveaux d'expression tant au niveau transcriptomique que protéomique (fractions PC et PE) sont très fortement corrélés (ou anti-corrélés) entre eux.

Les Figures 4.15C1 et C2 révèlent plus en profondeur la structure sous-jacente des éléments sélectionnés. En particulier, la Figure 4.15C1 montre bien que les éléments moléculaires associés à la première composante sont, à deux échantillons près (EXL2, t = 5 h, répétitions 1 et 3), résolus dans deux clusters structurés par la nature de l'EXL employé. A l'inverse, on retrouve dans les données des éléments propres à la deuxième composante (Figure 4.15C2) une résolution sur une base temporelle.

Enfin, la Figure 4.16 est davantage axée sur la nature des corrélations entre les éléments sélectionnés. Le graphe Circos (Figure 4.16A) indique une très bonne homogénéité et une forte densité de corrélation entre l'ensemble des éléments sélectionnés, ce qui est tout à fait cohérent au vu des coefficients de corrélation globaux observés entre « blocs » omiques (Figure 4.15B). Ce graphe révèle en outre que sur le total des 20 gènes sélectionnés, 13 ont une plus forte expression dans l'EXL1 et 7 dans l'EXL2. De façon similaire, 12 protéines (fraction PC) sont plus abondantes dans l'EXL1 (8 dans l'EXL2), et 14 autres protéines de la fraction PE sont retrouvées majoritairement dans l'EXL1 contre 6 dans l'EXL2. Le détail des niveaux de contribution de chacun de ces gènes et protéines est disponible dans l'Annexe B1. La Figure 4.16B permet toutefois d'affiner ces observations et révèle à la fois l'excellente corrélation intra-groupe (éléments d'une même composante), mais également une absence complète de corrélation inter-groupes (angle de 90° entre sous-éléments des deux composantes). Ceci est illustré de façon plus visuelle par le réseau de co-expression (Figure 4.16C) qui montre bien l'organisation en deux clusters séparés lorsque ces différents gènes et protéines sont organisés en fonction du degré de corrélation de leurs niveaux d'expression respectifs.

En résumé, l'analyse DIABLO a permis de faire émerger de l'ensemble des données omiques deux groupes de gènes et protéines constituant la signature moléculaire des deux EXLs utilisés dans cette étude. Le premier groupe, associé à la première composante, est constitué d'éléments dont les niveaux relatifs d'expression sont fortement similaires (fort degré de corrélation) et ce tant au niveau transcriptomique que protéomique. Ces éléments sont de plus

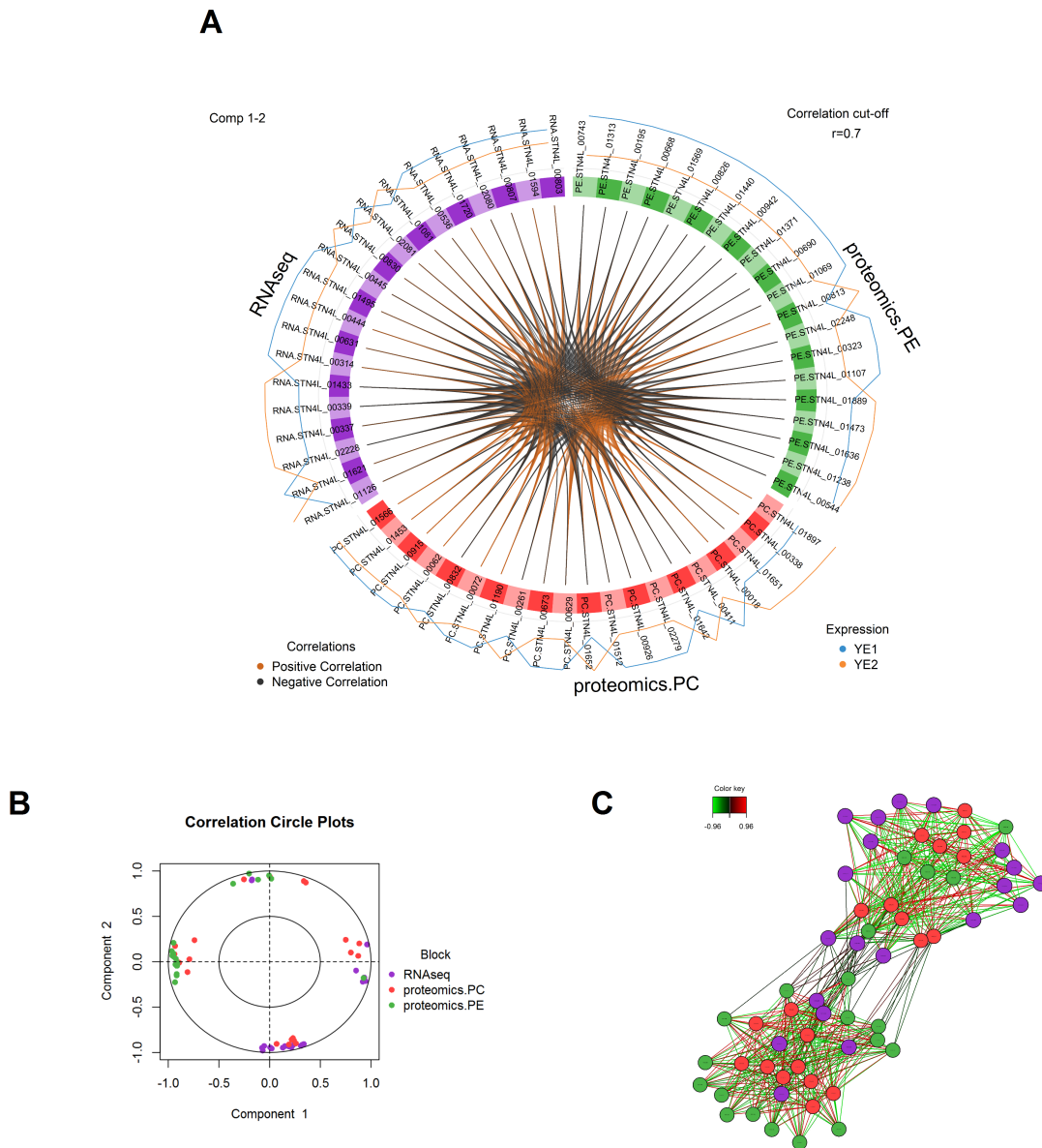


FIGURE 4.16 : Nature des liens de corrélation entre gènes et protéines sélectionnés par l'analyse DIABLO. (A) Graphe Circos montrant la corrélation (marron) ou anti-corrélation (noir, $r > 0,7$) entre l'ensemble des éléments sélectionnés sur les deux composantes. Les valeurs d'expression relative sont indiquées par des lignes au-dessus du cercle. L'identifiant des gènes et protéines est indiqué au sein du cercle. (B) Cercle de corrélation des variables colorées selon leur jeu de données d'origine. (C) Réseau de co-expression des gènes et protéines sélectionnés (composante 1 et 2).

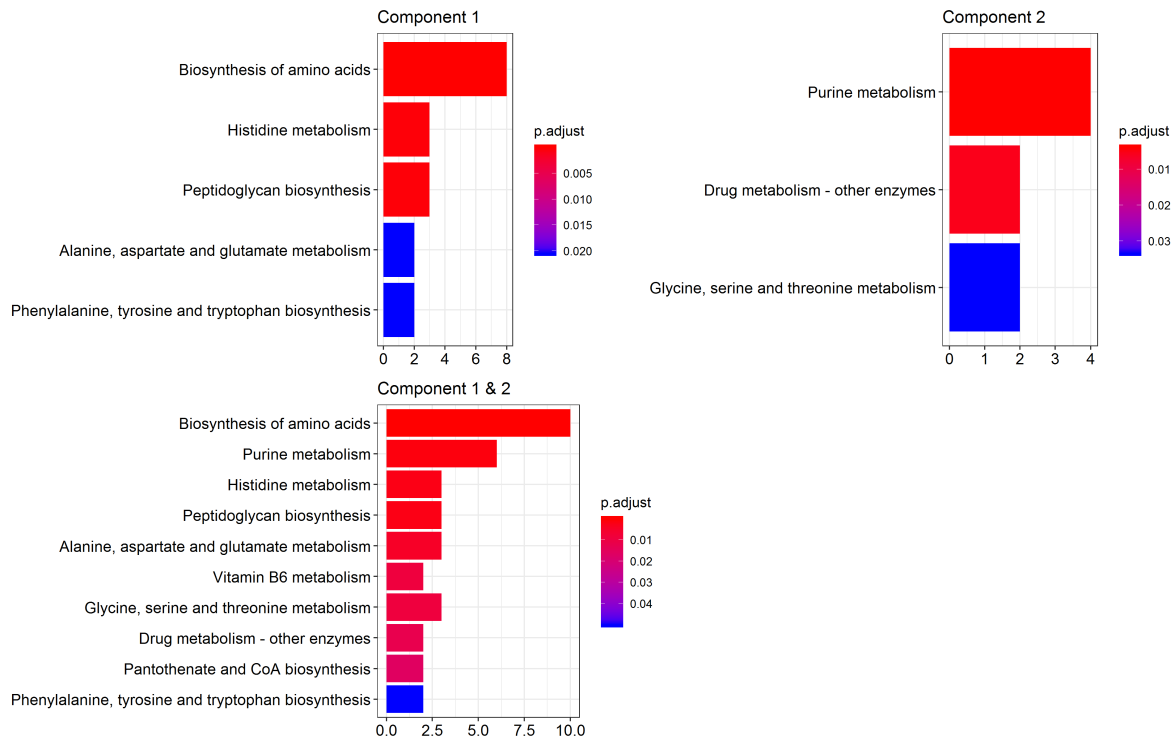


FIGURE 4.17 : Fonctions associées à la signature moléculaire des EXLs

principalement affectés par la nature de l'EXL utilisé. Le deuxième groupe pour sa part est également constitué de gènes et protéines corrélés et dont l'abondance a également été modulée par les EXLs mais de façon beaucoup plus modérée, le stade de croissance de la bactérie étant le facteur prévalent. Ce deuxième groupe suit en outre une trajectoire d'expression indépendante de celle du premier, suggérant des mécanismes d'interaction différents des EXLs sur les deux groupes. Pour autant, l'information portée par l'ensemble des éléments des deux groupes reste nécessaire pour optimiser les performances de classification du modèle. Le BER basé uniquement sur la première composante est en effet de 14 % mais passe à 1,5 % lorsque la deuxième composante est ajoutée.

Afin d'extraire l'information biologique contenue dans les résultats fournis par l'analyse DIABLO, les gènes et protéines associés aux deux groupes identifiés ont été soumis au test d'enrichissement précédemment décrit (partie 4.3.3) afin d'y détecter les fonctions cellulaires surreprésentées. Les résultats de cette analyse sont présentés dans la Figure 4.17. La biosynthèse des acides aminés et le métabolisme des purines sont les deux fonctions surreprésentées dans la signature moléculaire des EXLs. En ce qui concerne la biosynthèse des acides aminés, sont particulièrement concernés l'histidine, l'alanine, les acides aminés chargés négativement, ainsi que la glycine, sérine et thréonine. D'autres fonctions ont également été détectées comme surreprésentées bien que de façon plus modérée. Cela concerne notamment la voie de biosynthèse du peptidoglycane et de diverses vitamines. Il y a donc un excellent recouvrement des fonctions obtenues avec cette approche intégrative et l'analyse

différentielle précédemment décrite. La représentation visuelle des données d'expression sur ces voies métaboliques est disponible dans l'Annexe A, et la liste détaillée des gènes et protéines identifiés par l'analyse DIABLO est donnée dans l'Annexe B2. On relèvera finalement que l'information concernant la biosynthèse des acides aminés est portée majoritairement par la première composante, tandis que la deuxième est principalement associée au métabolisme des purines. L'expression des éléments associés à ces deux composantes n'étant pas corrélée, ceci suggère que le contrôle de ces fonctions fait intervenir soit des acteurs (ex : facteurs de transcription) différents, soit de mêmes acteurs produisant des effets très différents sur leurs systèmes cibles. A ce titre, on notera dans le groupe de protéines associées à la première composante la présence de deux régulateurs transcriptionnels : un régulateur appartenant à la famille GntR (STN4L_00673), et un régulateur annoté comme étant probablement YebC (STN4L_01440), le *CodY-like* des lactobacilles (Brown *et al.*, 2017). Ces deux protéines ont été détectées comme étant plus abondantes dans l'EXL1. Néanmoins, il semble que la composition de l'EXL ait eu un effet plus marqué sur la synthèse de YebC qui montre une plus forte contribution à la construction de la première composante de son propre bloc (fraction PE) en comparaison avec GntR (fraction PC, se référer à l'Annexe B1).

4.4 Conclusion

Le suivi de l'expression génétique et de la synthèse protéique de *S. thermophilus* N4L pendant sa croissance dans les deux EXLs a permis de mettre en évidence deux faits majeurs. Le premier d'entre eux concerne un système de régulation basé sur un mécanisme de *quorum sensing* dont l'expression, comparativement au reste du génome de la bactérie, a été la plus fortement affectée par la différence de composition entre EXLs. Ce système est dédié à la production d'un peptide modifié post-traductionnellement identifié il y a maintenant une dizaine d'années (Ibrahim *et al.*, 2007a) et faisant intervenir un régulateur transcriptionnel Rgg associé avec un peptide phéromone de type SHP. A ce stade, il convient d'ores et déjà de préciser que la fonction biologique de ce peptide modifié reste pour l'instant inconnue, et l'élucidation de cette fonction constitue un des axes de travail du laboratoire. Dans le contexte de la présente étude, il est en revanche plus pertinent de s'intéresser à l'origine de cette différence d'expression observée entre les deux EXLs. Deux hypothèses sont envisageables : i) la surexpression du système dans l'EXL1 par rapport à l'EXL2 pourrait provenir de la présence d'un ou plusieurs peptides de levure « *SHP-like* » spécifiques à cet EXL capables de stimuler l'expression du système par interférence positive, ou ii) l'extinction du système dans l'EXL2 serait au contraire le résultat d'une interférence négative où les peptides de l'EXL2 aurait fortement perturbé le signal moléculaire porté par le SHP, le système s'exprimant normalement dans l'EXL1. Etant donnée la cinétique d'expression croissante du système, la première hypothèse suppose que le peptide *SHP-like* serait de mieux

en mieux perçu au fur et à mesure de la croissance. Il pourrait donc s'agir soit d'un peptide résultant de l'activité peptidolytique de la bactérie s'accumulant progressivement, soit d'un peptide initialement présent mais dont le transport serait perturbé en début de croissance en raison d'une forte compétition au niveau du transporteur. Cette compétition serait ensuite progressivement levée du fait de la consommation globale des peptides compétiteurs. Bien que cette première hypothèse ne puisse être immédiatement exclue, la deuxième (inhibition par compétition dans l'EXL2) paraît néanmoins la plus vraisemblable, en particulier étant donnée la surabondance des oligopeptides mesurée par le dosage de leurs acides aminés constitutifs (partie 3.2). Il est en effet probable que cette quantité accrue de peptides a provoqué une compétition pour le transport plus intense au niveau du système Ami de la bactérie et ce au détriment du SHP. On peut de la même façon envisager une hypothèse annexe où des peptides de levure interféreraient non seulement au niveau du transporteur mais également au niveau du régulateur Rgg, atténuant ainsi son action élicitrice sur l'expression génétique. Néanmoins, quelle que soit l'origine précise de ce phénomène, elle est selon toute vraisemblance directement imputable au contenu peptidique des EXLs. A ce titre, l'activation de ce système dans l'EXL1 a constitué une réelle surprise au cours de ces travaux car, jusqu'à présent, les divers systèmes connus de *quorum sensing* portés par *S. thermophilus* n'ont été observés actifs que dans des milieux à faible teneur en peptides (ex : lait) voire dépourvus de peptides (MCD). Pourtant, chez l'espèce voisine *S. salivarius*, des systèmes de *quorum sensing* comme celui de la compétence sont bien fonctionnels dans des milieux complexes tels que le M17 (Christine Delorme, communication personnelle). Il conviendrait donc de vérifier si l'activité du système streptide dans l'EXL1 est comparable à celle observée en MCD ou si elle est malgré tout atténuée. Le phénotype associé à ce système restant pour l'instant inconnu, une façon d'en mesurer l'activité consisterait à doser par LC-MS/MS la concentration du streptide et/ou du SHP. Or la forte présence de peptides de levure dans l'EXL rend cette tâche techniquement difficile et nécessiterait un important développement méthodologique. Il conviendrait donc de privilégier une approche mettant en œuvre des fusions transcriptionnelles entre le promoteur du gène *shp* et un gène rapporteur (ex : luciférase).

Le deuxième fait marquant relevé au cours de ces travaux concerne un effet global des EXLs sur plusieurs fonctions cellulaires, en particulier la biosynthèse des acides aminés et le métabolisme des purines. Cet effet contraste dans sa forme à ce qui a été précédemment décrit concernant le locus streptide. En effet, ici, les différences d'abondance relevées sont beaucoup plus modérées, mais le fait qu'elles se produisent de façon coordonnée dans diverses voies métaboliques et sur les deux niveaux moléculaires étudiés (transcriptomique et protéomique) leur donne un poids tout aussi important. De plus, ces dynamiques ont été identifiées par deux approches analytiques indépendantes, d'une part par analyse différentielle, et d'autre part *via* une intégration multiomique. Bien que ces deux approches aient abouti à

des conclusions très similaires, l'information qu'elles apportent n'est pas strictement identique et les deux sont en réalité complémentaires. D'un côté, l'analyse différentielle repose sur des méthodes statistiques univariées qui permettent de produire des indicateurs statistiques (ex : p -values) très utiles pour évaluer la significativité et le degré de confiance d'un résultat. De tels indicateurs sont en revanche absents des approches multivariées telles que proposées par le paquet mixOmics. En contrepartie, l'intégration multiomique permet de mieux faire ressortir les dynamiques multi-échelles, résume plus efficacement l'information d'un système complexe, et permet d'identifier la signature moléculaire (les biomarqueurs) associée à un phénotype d'intérêt. Dans le contexte de cette étude, la significativité de l'effet des EXLs sur la biosynthèse des acides aminés et le métabolisme des purines est donc attesté par l'analyse différentielle. L'analyse DIABLO a pour sa part montré que ces dynamiques se produisaient dans deux groupes de gènes et protéines distincts suivant deux trajectoires d'expression différentes.

Là encore, il est possible d'émettre des hypothèses concernant l'origine de ces dynamiques. L'hypothèse la plus immédiate est que ces variations d'expression seraient le résultat de différences de concentration en divers acides aminés et bases nucléiques entre les deux EXLs. Ces différences seraient perçues par des facteurs de transcription locaux qui moduleraient en conséquence la voie de biosynthèse dont ils ont la charge. A titre d'illustration, des travaux réalisés chez *B. subtilis* et *S. pneumoniae* (Ebbole et Zalkin, 1989; Ng *et al.*, 2003) ont montré que l'activité de PurR, le répresseur de l'opéron *pur* responsable de la formation d'IMP, dépendait en partie de la concentration extracellulaire en purines qui elle-même influence la quantité de PRPP intracellulaire, le ligand de PurR. Un mécanisme similaire a par la suite été proposé chez *S. thermophilus* (Herve-Jimenez *et al.*, 2009). Ce type de schéma a également été identifié dans les voies de biosynthèse de divers acides aminés. Pour autant, on sait aussi que de nombreuses voies de biosynthèse, en plus de leurs régulateurs propres, sont également sous le contrôle direct ou indirect de régulateurs centraux. Ainsi, le même opéron *pur* a été identifié chez *S. thermophilus* comme faisant partie du régulon de CodY (Lu *et al.*, 2015). Dans la présente étude, l'intervention de régulateurs centraux est particulièrement suggérée par le fort degré de corrélation observé au sein des biomarqueurs dégagés par l'analyse DIABLO. CodY constitue un candidat tout à fait plausible, d'une part en raison de sa fonction physiologique connue et sa forte implication dans la régulation de la biosynthèse d'acides aminés chez *S. thermophilus* (Lu *et al.*, 2015), mais également du fait des données transcriptomiques qui ont montré une expression significativement supérieure dans l'EXL2 en fin de croissance. Ce gain d'expression est de plus cohérent avec la richesse peptidique mesurée dans l'EXL2, ce qui là aussi suggère une implication des peptides comme molécules responsables des dynamiques métaboliques observées. Plus précisément, seraient concernés les peptides riches en acides aminés branchés, les ligands de CodY.

Néanmoins, sous cette « hypothèse CodY », certaines voies de biosynthèse ne sont pas affectées de la façon attendue. En particulier, les gènes codant la biosynthèse et le transport des acides aminés branchés sont connus pour être très fortement réprimés par CodY. Or, ici, on observe plutôt une expression de ces différents gènes et protéines allant globalement dans le même sens que celle de CodY dans l'EXL2 (Figures 4.9, 4.10 et 4.11). On peut donc se demander si un autre acteur ne serait pas à l'œuvre dans la régulation centrale de ces diverses voies métaboliques. L'analyse DIABLO a montré que l'abondance du régulateur putatif YebC était plus importante dans l'EXL1 et était de plus fortement corrélée à la biosynthèse de plusieurs acides aminés. Ce régulateur n'a pas été étudié chez *S. thermophilus* mais il lui a été montré une activité *Cody-like* chez l'homologue de *Lb. lactis* (Brown *et al.*, 2017). Les protéines de ces deux espèces montrent une identité de séquence d'environ 44 %, ce qui n'est pas incompatible avec une similarité de fonction biologique. Tout ceci interroge donc sur l'éventuel rôle de cette protéine dans la régulation du métabolisme azoté de *S. thermophilus*, et sur son lien fonctionnel avec CodY.

Discussion générale et perspectives

La question de la relation entre peptides et bactéries constitue une thématique large et multifacette. Ce projet de thèse a apporté un certain nombre d'éclairages nouveaux à ce thème en se plaçant dans un contexte particulier, celui de la bactérie lactique d'intérêt industriel *Streptococcus thermophilus* cultivée dans des milieux riches à base d'extraits de levure. Les principaux apports de ce travail se situent à deux niveaux :

- i) Apports méthodologiques, avec d'une part l'élaboration d'une méthode de caractérisation haut-débit de mélanges complexes de peptides par spectrométrie de masse, le développement d'une méthode analytique permettant de mettre en évidence *in vivo* les spécificités de transport de peptides de la bactérie, et la conception d'un outil prédictif permettant de qualifier l'aptitude au transport des peptides. D'autre part, il s'agit de la première étude de microbiologie appliquée visant à combiner différentes approches omiques analysées de façon intégrée dans le but de mieux cerner la relation existant entre peptides et bactéries lactiques.
- ii) Apports de connaissances, concernant tout d'abord l'établissement de la charge nette comme facteur clé dictant l'assimilation d'un peptide par le transporteur Ami, mais aussi la mise en évidence de phénotypes spécifiques découlant de la composition propre des deux EXLs du projet. Ces phénotypes concernent d'une part la façon dont l'azote aminé est utilisé par la bactérie durant sa croissance (transport, hydrolyse), et d'autre part l'adaptation de son métabolisme *via* des réseaux de régulation complexes faisant intervenir des facteurs de transcription locaux (Rgg) et très vraisemblablement centraux (CodY etc.).

En se replaçant dans le contexte appliqué de la production industrielle d'hydrolysats protéiques et de ferments lactiques, la finalité de tout le travail initié durant cette thèse était de pouvoir corrélérer la composition d'un hydrolysate protéique avec ses performances vis-à-vis d'un microorganisme donné. Ici, cette notion de performances dépasse le simple cadre nutritionnel (fourniture des divers nutriments permettant d'assurer une croissance rapide), et englobe tout ce qui peut ensuite moduler les fonctionnalités technologiques du ferment. Idéalement, l'élaboration d'une méthode capable de prédire les performances d'un hydrolysate sur la base de sa simple composition procurerait un avantage compétitif certain pour

les fabricants d'hydrolysats et de ferments. Pour répondre à cet objectif ambitieux, il est nécessaire de travailler sur deux niveaux. Le premier concerne la caractérisation physico-chimique exhaustive des hydrolysats, et le deuxième s'intéresse à la compréhension fine des mécanismes moléculaires par lesquels les hydrolysats affectent le métabolisme et la physiologie bactérienne afin de faire le lien entre composition et performances. Avant d'atteindre cet objectif général, de nombreuses étapes doivent encore être franchies et des limitations levées. Néanmoins, le travail réalisé durant ce projet de thèse a permis d'apporter d'importantes contributions à cette problématique en se focalisant spécifiquement sur la fraction peptidique de deux EXLs.

L'objectif de cette partie conclusive sera donc de récapituler les différents apports du projet et les perspectives désormais ouvertes vis-à-vis de cet objectif général « corrélation composition-performances » en lien avec les deux niveaux d'étude précédemment mentionnés. Les contributions d'ordre plus académique seront naturellement discutées, ainsi que certains points particuliers qui n'ont pas été abordés dans les chapitres précédents.

Analyse peptidomique : un nouvel outil pour le laboratoire et l'industrie?

Les travaux réalisés durant cette thèse ont permis de faire grandement progresser l'efficacité de l'analyse de mélanges complexes de peptides. Pour rappel, les toutes premières analyses réalisées avant même le démarrage de la thèse avaient identifié tout au plus quelques centaines de peptides par EXL. Par la suite, ces performances ont été améliorées de plus d'un ordre de grandeur grâce à l'emploi d'une stratégie de fractionnement chromatographique. Néanmoins, ce gain de performances s'est fait au détriment du temps d'analyse nécessaire pour traiter un échantillon, qui a été démultiplié. La caractérisation des EXLs et autres hydrolysats protéiques étant l'une des priorités des partenaires industriels, il a ainsi été décidé de consacrer une partie de la fin de la thèse à la mise en place d'un processus analytique simplifié, qui soit en pratique applicable dans toute plateforme disposant d'un spectromètre de masse compatible avec l'analyse *bottom-up*. Cet objectif a été atteint, et il désormais possible de réaliser des analyses de mélanges complexes de peptides à haut-débit.

Cet travail ouvre ainsi de nombreuses perspectives tant du point de vue industriel qu'académique. En particulier, il a d'ores et déjà été montré au cours de ce travail que l'analyse peptidomique représentait un outil très puissant pour analyser *in vivo* les préférences d'utilisation des oligopeptides par une bactérie et ce grâce à l'approche originale d'analyse des données développée. La partie expérimentale de l'outil peptidomique étant désormais simplifiée, il est maintenant beaucoup plus facile d'envisager son utilisation afin de confirmer les hypothèses expliquant les cinétiques observées (transport *vs* hydrolyse). En particulier, il serait souhaitable d'explorer l'origine des augmentations de certaines classes de peptides détectées durant la croissance de *S. thermophilus* N4L, ainsi que la diminution spécifique à l'EXL2 faisant intervenir des peptides anioniques. Des mutants de délétion du transporteur

Ami ainsi que de différentes protéases permettraient d'apporter des éléments d'information à ce sujet. Dans ce dernier cas de figure, il serait particulièrement intéressant de savoir si certaines dynamiques observées sont à imputer à la protéase de paroi PrtS, ou bien d'autres protéases qui, dans un contexte laitier, n'interviennent pas dans la nutrition azotée basée sur les caséines (ex : HtrA, Eep). Enfin, de façon plus générale, la simplification de l'outil peptidomique facilite désormais l'étude d'autres souches et espèces possédant différents systèmes de transport et protéolytiques. A nouveau, cet outil se combinerait particulièrement bien avec l'utilisation de mutants de délétion afin d'étudier spécifiquement les préférences d'un seul transporteur dans l'éventualité où plusieurs systèmes seraient présents. De plus, dans le cas particulier des streptocoques qui possèdent plusieurs protéines de fixation (SBPs) associées au système Ami, des mutants permettraient également d'étudier finement les préférences respectives des différentes SBPs du système.

En outre, l'approche peptidomique découlant de la protéomique, elle permet non seulement d'identifier des peptides en mélange, mais également i) d'inférer les protéines parentales d'où ils proviennent, et ii) de déterminer les activités endopeptidasiques responsables de leur libération par alignement de séquences et identification des sites primaires de clivage. Un tel travail a déjà été réalisé avec succès par le laboratoire pour décoder le peptidome extracellulaire de *L. lactis* (Guillot *et al.*, 2016). Ce type d'information pourrait ainsi s'avérer précieux pour les fabricants d'hydrolysats protéiques. Cela leur permettrait en effet de retracer les protéines et activités enzymatiques responsables de la formation de peptides aptes ou non au transport et d'établir ainsi un lien direct avec le procédé de fabrication. L'utilisation d'un modèle prédictif comme celui décrit dans cette thèse contribuerait très positivement à un tel travail. Pour cela, un tel modèle devrait être préalablement établi pour un microorganisme donné sur la base de quelques fermentations dans différents hydrolysats. Si besoin, ces fermentations seraient à réaliser en comparant des mutants de transport et de protéase(s) à la souche sauvage afin d'identifier sans ambiguïté les dynamiques liées au transport. Cette étape terminée, le modèle obtenu pourrait être décliné sur les données de peptidomique de divers hydrolysats bruts (identification des peptides constitutifs initialement présents) sans avoir besoin de recourir à un suivi cinétique. Ceci permettrait de caractériser l'aptitude au transport de ces peptides, mais aussi d'identifier et de comparer les activités enzymatiques à l'œuvre durant leur production ayant favorisé l'apparition de peptides transportables ou non.

En l'état actuel des choses, cette démarche est encore très théorique et son intérêt pratique limité. Mais à terme, dans le contexte de l'objectif général de corrélation composition-performances, ce modèle constitue une ébauche de ce que pourrait être un véritable outil d'identification de « biomarqueurs de performances ». Cette notion sera explorée plus en détail dans la dernière partie de la discussion. A ce stade, l'outil d'analyse peptidomique mis au point durant ce projet constitue un important pas en avant dans la caractérisation

des hydrolysats protéiques et contribue donc au premier niveau d'information nécessaire à l'objectif corrélation composition-performance. Néanmoins, ici, le voile n'a pas été encore entièrement levé sur la fraction peptidique des EXLs. Il manque en effet toute la dimension quantitative à l'analyse par spectrométrie de masse, et la technologie a encore de nombreux progrès à faire dans ce domaine. A l'heure actuelle, le couplage avec une deuxième approche quantitative, tel que le dosage des acides aminés libres et totaux réalisé durant ce projet, reste la seule stratégie possible.

Pour autant, dans l'éventualité où l'entière composition peptidique d'un hydrolysats serait connue, cette masse d'information ne saurait suffire par elle-même à prédire les futures performances attendues de cet hydrolysats vis-à-vis d'un microorganisme donné. Il est bien entendu possible de procéder à des expériences de type plan d'expérience factoriel afin de tenter d'identifier certains déterminants de performances (ex : plan de Plackett-Burman; Plackett et Burman, 1946). Néanmoins, ces dispositifs impliquent d'être capable de faire varier de façon ciblée et contrôlée la concentration de quelques composés, ce qui est évidemment impossible dans le cas de substrats aussi complexes comprenant des milliers de peptides produits de façon semi-dirigée. Pour identifier des biomarqueurs de performances dans un tel contexte, la seule solution véritablement pertinente consiste à mesurer et comprendre les effets des hydrolysats sur le métabolisme et la physiologie du microorganisme.

De l'intérêt de la double analyse transcripto-protéomique

Durant l'élaboration de ce projet, il a été fait le choix de s'intéresser à l'effet spécifique des deux EXLs du projet sur le métabolisme de la bactérie. Pour rappel, ces deux EXLs avaient été choisis sur la base de leur bonne adéquation avec *S. thermophilus* N4L en termes de croissance, ainsi que sur des connaissances acquises à l'échelle industrielle concernant des différences de performances technologiques liées aux spécificités de composition de ces deux EXLs. Afin d'investiguer de façon approfondie l'effet des EXLs sur la souche modèle du projet, deux approches différentes ont été utilisées : transcriptomique et protéomique. Ce choix revêt à la fois un certain nombre d'avantages et d'inconvénients. Tout d'abord, le protocole expérimental suivi a été conçu de façon à ce que les différences métaboliques observées entre EXLs soient le reflet de la différence de composition globale entre les deux EXLs, incluant à la fois des différences de contenu azoté, mais également nucléique, carboné, de vitamines etc. Ceci a l'avantage de révéler la vraie nature des impacts susceptibles de se produire lors de l'utilisation de deux produits différents. Néanmoins dans le contexte de cette thèse davantage focalisée sur la fraction peptidique des EXLs, cela complexifie l'interprétation des résultats, et il est difficile de démêler les effets provenant des peptides de ceux provenant d'autres composés (sans compter les possibles interactions entre différentes classes de molécules). Pour autant, les effets principaux observés ici sur le locus streptide et la biosynthèse des acides aminés peuvent vraisemblablement être imputés à la différence de contenu azoté, voire directement peptidique, entre les EXLs. Concernant le locus streptide, l'origine

peptidique est la plus probable en raison des mécanismes de régulation de ce système reposant sur un peptide phéromone (SHP). Les peptides peuvent également avoir participé à la modulation de la biosynthèse des acides aminés par l'intermédiaire d'un régulateur central tel que CodY. Néanmoins, ces pistes sont encore préliminaires et demandent donc validation. Pour ce faire, une solution envisageable serait d'effectuer un suivi de l'expression de certains gènes cibles par RT-qPCR ou fusion transcriptionnelle dans un EXL donné dans lequel serait modifié uniquement sa fraction peptidique. Cette modification peut passer soit par des stratégies de fractionnement (déplétion ou enrichissement de certaines classes de peptides), soit par une complémentation avec des peptides synthétiques produits à façon. A titre d'exemple, il serait intéressant d'explorer l'hypothèse concernant l'origine de l'activité de CodY en modulant de la sorte la fraction en peptides branchés et en suivant l'expression du gène même de CodY ainsi que quelques gènes cibles choisis appartenant à son régulon. Néanmoins, manipuler seulement la fraction peptidique d'un tel substrat complexe reste plus difficile à mettre en œuvre qu'à concevoir. Une stratégie de fractionnement ne saurait en effet affecter exclusivement les peptides, et l'addition de peptides synthétiques, bien qu'aisée, ne sera peut-être pas suffisamment représentative de la réalité, et son effet sera de plus fortement dépendant de la séquence propre du ou des quelques peptides ajoutés. Une stratégie alternative, qui se combinerait d'ailleurs très bien avec la précédente, consisterait à utiliser des souches mutantes. Il serait notamment intéressant d'observer si les différences métaboliques provoquées par les compositions respectives des deux EXLs sont maintenue, atténuées voire abolies dans un mutant $\Delta codY$ cultivé dans les mêmes conditions que celles de la présente étude.

Il est également possible d'émettre quelques commentaires *a posteriori* concernant le choix d'une double analyse transcriptomique/protéomique pour mesurer l'effet des EXLs sur *S. thermophilus*. Conformément à ce qui a déjà été observé dans d'autres travaux, ces deux types d'analyse n'ont pas apporté exactement le même type d'information. En particulier, il s'est avéré que les données de RNAseq étaient beaucoup plus influencées par l'effet temps que celles de protéomique, et ces dernières se sont révélées beaucoup plus discriminantes pour le contexte de cette étude. Ainsi, si l'on se penche un peu plus en détail sur les résultats de l'intégration par mixOmics des différents jeux de données, le premier groupe d'éléments les plus fortement influencés par la nature des EXLs est composé majoritairement de données de protéomique (83 %). A l'inverse, le groupe associé à la deuxième composante du modèle, moins informatif, ne compte que 50 % de protéines. Pour autant, peut-on en conclure que les données de transcriptomique sont réellement moins informatives que celles de protéomique? Rien n'est moins sûr, d'importantes variations significatives ayant été observées pour de nombreux gènes. En l'état, la seule conclusion qu'il est possible de formuler est que, dans les conditions expérimentales du projet, les effets des EXLs sur le transcriptome de la bactérie sont moins marqués que ceux induits par le passage de la phase

de croissance exponentielle à stationnaire. Cette conclusion aurait peut-être été différente pour des EXLs engendrant des différences de croissance plus marquées. Quoiqu'il en soit, pourquoi de tels effets ne sont-ils pas aussi prégnants au niveau du protéome, et plus généralement, pourquoi un tel manque de corrélation entre transcriptomique et protéomique? Les deux principales raisons connues sont d'une part les divers mécanismes de régulation post-transcriptionnelle se produisant entre l'ARNm et la protéine finale, et d'autre part la différence de durée de vie entre ces deux types de molécules (Haider et Pal, 2013). Différentes stratégies seraient ainsi envisageables pour réaliser ces mêmes analyses en s'affranchissant de l'effet temps. Par exemple, une fermentation en mode continu (chemostat) permettrait de placer la bactérie en régime permanent, c'est-à-dire en croissance constante et contrôlée. Sur le plan scientifique, cette solution est sans doute la plus judicieuse, mais, en raison du contexte appliqué de l'étude, elle n'est pas pertinente. Les ferments lactiques classiques tels que *S. thermophilus* sont en effet rarement produits à l'échelle industrielle *via* un tel procédé (Béal *et al.*, 2008). Or, un des pré-requis de cette thèse était de se rapprocher le plus possible de conditions réelles de fermentation.

L'utilisation de méthodes d'analyse supervisées telle l'intégration multiomique par l'outil DIABLO prend alors tout son sens, car elle permet précisément de faire émerger des dynamiques sous-jacentes à partir de données où se confondent différents effets. Ici, cette analyse intégrée a été utilisée avec succès sur les deux niveaux d'analyse, transcriptomique et protéomique, et a permis de dégager une signature moléculaire propre à chacun des deux EXLs. En particulier, l'identification du régulateur putatif YebC est particulièrement intéressante car, bien qu'encore très préliminaire, cette piste ouvre d'importantes perspectives concernant les mécanismes de régulation du métabolisme azoté chez les bactéries lactiques. YebC a en effet été identifié chez *L. delbrueckii* subsp. *lactis* comme étant un substitut fonctionnel de CodY intervenant en particulier dans la répression du système protéolytique (Brown *et al.*, 2017). Il serait maintenant intéressant d'explorer le rôle de cette protéine chez des espèces qui possèdent un CodY fonctionnel, telles que *S. thermophilus* ou *L. lactis*, ainsi que leur interaction fonctionnelle éventuelle.

Pour autant, le potentiel des méthodes intégratives pour l'analyse de données multiomiques est loin d'avoir été entièrement exploité dans le contexte de cette étude. Certaines questions, qui restent à l'heure actuelle encore ouvertes, pourraient tout particulièrement bénéficier de cette approche. La première de ces questions est la suivante : les peptides des EXLs peuvent-ils réellement avoir une activité régulatrice de type phéromone sur des systèmes bactériens ciblés? Pour répondre à cette question, il est ainsi possible d'envisager l'emploi des données peptidomiques dans une approche intégrative afin de corrélérer des variations d'abondance de certains peptides ou familles de peptides au cours du temps à des variations d'expression de gènes ou d'abondance protéique. Cette stratégie nécessiterait l'emploi d'une autre méthode de quantification des peptides que celle utilisée ici (*spectral counting*) qui est as-

sez limitée. La quantification par extraction des courants d'ions (XIC) serait probablement plus adaptée car plus précise, et la simplification du protocole analytique, en particulier la suppression de la première dimension chromatographique, permet d'envisager une telle quantification. Une deuxième question, d'ordre plus fondamental, bénéficierait également de l'intégration multiomique : existe-t-il des systèmes de *quorum sensing* où le peptide phéromone et son régulateur associé ne se situeraient pas sur le même locus ? En effet, à l'heure actuelle, tous les systèmes connus associent en position proximale peptide phéromone, régulateur, et bien souvent gènes cibles. Mais pour autant, rien n'empêche de supposer que des systèmes reposant sur des gènes en position distale ne puissent exister. En particulier, l'enrichissement du génome de *S. thermophilus* en sORFs *via* l'outil SHOW a permis de révéler plusieurs séquences génétiquement isolées qui étaient fortement exprimées lors des fermentations, ce qui tend à démontrer leur caractère codant. Ces résultats n'ont pas été présentés dans ce manuscrit car ils s'éloignent de la thématique première de la thèse, mais ils constituent néanmoins une base de travail possible pour identifier de nouveaux systèmes de régulation basés sur les peptides. Ainsi, en corrélant les données d'expression de ces petits gènes avec celles du reste du génome (voire du protéome), il serait théoriquement possible de faire émerger de nouveaux candidats placés sous le contrôle d'un même système de régulation (s'il en existe).

Impact métabolique des extraits de levure : quelles conséquences pour le ferment ?

Les dynamiques identifiées à l'issue des analyses transcriptomique et protéomique semblent ainsi être au moins pour partie le fait de différences de contenu peptidique entre les deux EXLs, ce qui conforte l'hypothèse de départ, basée sur l'expérience (empirique) des partenaires industriels, voulant que les peptides n'aient pas qu'une simple fonction nutritionnelle lors d'une culture en milieu riche. Néanmoins, il est aussi très vraisemblable que les autres constituants du milieu ont également participé à l'élaboration de ces dynamiques. En particulier, la sollicitation de la voie de biosynthèse de l'IMP (précurseur des purines) dans l'EXL2 pourrait résulter d'une moindre abondance de certaines bases nucléiques dans ce produit. Ce même raisonnement pourrait s'appliquer aux voies de biosynthèse des acides aminés. Pourtant, en dépit du fait d'avoir réalisé des dosages des différentes formes d'azote aminé des EXLs (acides aminés libres, courts peptides et oligopeptides), il est assez difficile d'établir une corrélation évidente entre les résultats de dosage d'un acide aminé donné, toute forme confondue, dans les deux EXLs et les niveaux d'expression de sa voie de biosynthèse correspondante. Outre l'imprécision des mesures de dosage, ceci peut s'expliquer par l'intervention des régulateurs précédemment mentionnés et le chevauchement de divers réseaux de régulation. Quoiqu'il en soit, si la question de l'origine de ces dynamiques est intéressante, elle ne prend vraiment de sens que si l'on s'intéresse aux effets concrets de ces dynamiques sur les performances du ferment.

En d'autres termes, la question est de savoir s'il est possible de corrélérer les différences mé-

taboliques observées à des différences de croissance mais également de fonctionnalités technologiques : résistance aux procédés de stabilisation, au stockage, reprise d'activité, vitesse d'acidification etc. Or, lors de ce projet, la question des performances a été peu développée, principalement en raison de limitations de temps, et s'est arrêtée au niveau de la simple croissance en bioréacteur. Il est néanmoins possible de formuler plusieurs hypothèses quant à l'impact technologique des différences métaboliques observées entre les deux EXLs. Ainsi, en ce qui concerne le locus streptide, en supposant que ce système ne procure aucun avantage technologique, ce qui paraît plausible, on peut logiquement supposer que son expression induit un coût inutile affectant négativement le *fitness* bactérien général. Ceci pourrait d'ailleurs expliquer le léger gain de croissance observé dans l'EXL2 où ce système est éteint. Sous cette hypothèse, il serait alors judicieux de chercher à désactiver systématiquement ce circuit de régulation très répandu chez *S. thermophilus* par exemple en privilégiant un hydrolysate avec des peptides compétiteurs. À l'inverse, dans l'éventualité où l'on découvrirait un système de *quorum sensing* contrôlant ou affectant une fonction technologique importante (résistance au stress, production d'un composé aromatique etc.), il serait intéressant d'enrichir à façon un hydrolysate avec un ou des peptides capables d'activer ce système.

Ces hypothèses et pistes d'amélioration concernent toutefois des fonctions très ciblées. Or, les analyses omiques réalisées suggèrent également l'implication de régulateurs centraux dans les différences métaboliques observées. Dans le cas de CodY, en supposant son implication démontrée, il est possible que son action ait un fort impact sur les fonctionnalités technologiques générales du ferment, en particulier concernant son utilisation en matrice laitière. CodY contrôle en effet des systèmes primordiaux pour la croissance de la bactérie en lait, en particulier le système protéolytique. On peut ainsi émettre l'hypothèse que ce gène participe à la préparation du ferment à sa future utilisation en contrôlant la quantité de protéines protéolytiques synthétisées et disponibles au moment de sa récolte. Lorsque le ferment est ensuite utilisé pour inoculer du lait, son expression génétique est très rapidement réinitialisée par rapport à son état précédent (Martin Pedersen, communication personnelle). Or, ce n'est pas le cas du *pool* protéique hérité des générations antérieures en raison de la mémoire phénotypique des bactéries (transmission de protéines dont la durée de vie est supérieure au temps de génération de la bactérie; Lambert et Kussell, 2014). On peut ainsi voir la façon dont CodY, ou tout autre régulateur transcriptionnel, serait capable d'influencer à distance les performances du levain par un effet de programmation cellulaire. Tout cela est évidemment encore très hypothétique et demande bien entendu d'importantes validations expérimentales probablement complexes. Néanmoins, si cette hypothèse s'avère juste, elle fournirait un levier d'action pour programmer à façon, par le contenu en peptides de l'EXL, l'activité des facteurs de transcription impliqués dans le contrôle à distance des fonctions technologiques du ferment. Dans le cas précis de CodY, les effecteurs de

cette programmation pourraient être les peptides à fort contenu en acides aminés branchés. Ces exemples permettent ainsi de mieux discerner l'intérêt de l'approche peptidomique couplée à un modèle prédictif afin de déterminer si dans un hydrolysats donné, son contenu peptidique est susceptible d'affecter positivement ou négativement les propriétés fonctionnelles futures d'un ferment. L'intérêt d'un industriel serait alors d'ajuster le procédé de fabrication de ce produit en conséquence afin de l'enrichir en « bons » peptides associés à de meilleures performances, et de diminuer la teneur d'autres peptides interférant avec des systèmes technologiquement importants, tout ceci afin de vendre un produit hautement actif et à forte valeur ajoutée. L'analyse peptidomique montre son importance dans ce travail en raison de sa capacité à faire le lien entre composition peptidique et procédé de production (identification des protéines parentales et des activités enzymatiques).

Vers la biotechnologie prédictive

Tout cela ramène à la notion de biomarqueurs de performances précédemment évoquée. Dans sa définition la plus large possible, un biomarqueur serait ici toute caractéristique intrinsèque et mesurable d'un substrat complexe influençant ou permettant de prédire les performances obtenues lors de son utilisation dans un procédé biotechnologique donné. Cette définition englobe donc tout type de substrat, de nutriment, de paramètre physico-chimique (pH, potentiel redox etc.), de modèle microbien (bactérie, levure, cellule de mammifère, de végétal etc.) mais également de type de procédé fermentaire (*batch*, *fed-batch*, continu etc.) ainsi que sa finalité (production de biomasse, d'une molécule d'intérêt etc.). Ces biomarqueurs seraient ainsi à la fois les indicateurs et les leviers d'action permettant à un industriel d'optimiser à façon un produit pour une application donnée.

Le travail engagé lors de cette thèse ne constitue qu'une ébauche de toute la démarche qu'il serait nécessaire de déployer pour parvenir à ce résultat. Il faudrait en particulier « généraliser » cette approche, c'est-à-dire étendre progressivement le travail réalisé à d'autres modèles microbiens et des hydrolysats d'une autre nature que les EXLs. Il s'agit d'un long processus dont l'application pratique est loin d'être certaine, mais qui pourrait s'avérer extrêmement fructueux à long terme. Cette démarche a de plus le mérite de s'inscrire dans une méthode raisonnée et approfondie d'acquisition de connaissances sur des systèmes complexes. Si l'industrie des biotechnologies microbiennes procède encore beaucoup par essais et erreurs (« *trial and error* ») pour améliorer ses procédés de fermentation, la situation concurrentielle et les enjeux économiques du secteur imposent désormais d'acquérir une expertise technique et scientifique de plus en plus poussée. Le travail réalisé dans cette thèse est donc le reflet de la volonté des groupes Lesaffre et Sacco, en association avec l'INRA, de rester à la pointe de l'innovation dans leurs métiers respectifs.

Références

- Achstetter T, Wolf DH. Hormone Processing and Membrane-Bound Proteinases in Yeast. *The EMBO journal* 1985a; 4 : 173177.
- Achstetter T, Wolf DH. Proteinases, Proteolysis and Biological Control in the Yeast *Saccharomyces Cerevisiae*. *Yeast (Chichester, England)* 1985b; 1 : 139157.
- Aggarwal C, Jimenez JC, Nanavati D, Federle MJ. Multiple Length Peptide-Pheromone Variants Produced by *Streptococcus Pyogenes* Directly Bind Rgg Proteins to Confer Transcriptional Regulation. *The Journal of Biological Chemistry* 2014; 289 : 2242722436.
- An FY, Sulavik MC, Clewell DB. Identification and Characterization of a Determinant (Eep) on the *Enterococcus Faecalis* Chromosome That Is Involved in Production of the Peptide Sex Pheromone cAD1. *Journal of Bacteriology* 1999; 181 : 59155921.
- Andrews SJ, Rothnagel JA. Emerging Evidence for Functional Peptides Encoded by Short Open Reading Frames. *Nature Reviews Genetics* 2014; 15 : 193204.
- Antiporta MH, Dunny GM. ccfA, the Genetic Determinant for the cCF10 Peptide Pheromone in *Enterococcus Faecalis* OG1RF. *Journal of Bacteriology* 2002; 184 : 11551162.
- Armstrong N, Gouaux E. Mechanisms for Activation and Antagonism of an AMPA-Sensitive Glutamate Receptor : Crystal Structures of the GluR2 Ligand Binding Core. *Neuron* 2000; 28 : 165181.
- Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, et al. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products : Overview and Recommendations for a Universal Nomenclature. *Natural Product Reports* 2013; 30 : 108160.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology : Tool for the Unification of Biology. *The Gene Ontology Consortium. Nature Genetics* 2000; 25 : 2529.
- Atlan D, Béal C, Champomier-Vergès M-C, Chapot-Chartier M-P, Chouayekh H, Coccagn-Bousquet M, et al. Métabolisme et Ingénierie Métabolique. In : *Bactéries Lactiques, de La*

- Génétique Aux Ferments. Paris : Lavoisier; 2008. p. 271509.
- Auchtung JM, Lee CA, Monson RE, Lehman AP, Grossman AD. Regulation of a *Bacillus Subtilis* Mobile Genetic Element by Intercellular Signaling and the Global DNA Damage Response. *Proceedings of the National Academy of Sciences of the United States of America* 2005; 102 : 1255412559.
- Becker SH, Darwin KH. Bacterial Proteasomes : Mechanistic and Functional Insights. *Microbiology and Molecular Biology Reviews* 2017; 81 : e0003616.
- Belitsky BR. Indirect Repression by *Bacillus Subtilis* CodY via Displacement of the Activator of the Proline Utilization Operon. *Journal of Molecular Biology* 2011; 413 : 321336.
- Belitsky BR, Sonenshein AL. Roadblock Repression of Transcription by *Bacillus Subtilis* CodY. *Journal of Molecular Biology* 2011; 411 : 729743.
- Belitsky BR, Sonenshein AL. Genome-Wide Identification of *Bacillus Subtilis* CodY-Binding Sites at Single-Nucleotide Resolution. *Proceedings of the National Academy of Sciences of the United States of America* 2013; 110 : 70267031.
- Benjamini Y, Hochberg Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 1995; 57 : 289300.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Research* 2000; 28 : 235242.
- Berntsson RP-A, Doeven MK, Fusetti F, Duurkens RH, Sengupta D, Marrink S-J, et al. The Structural Basis for Peptide Selection by the Transport Receptor OppA. *The EMBO journal* 2009; 28 : 13321340.
- Berntsson RP-A, Schuurman-Wolters GK, Dunny G, Slotboom D-J, Poolman B. Structure and Mode of Peptide Binding of Pheromone Receptor PrgZ. *Journal of Biological Chemistry* 2012; 287 : 3716537170.
- Berntsson RP-A, Smits SHJ, Schmitt L, Slotboom D-J, Poolman B. A Structural Classification of Substrate-Binding Proteins. *FEBS letters* 2010; 584 : 26062617.
- Berntsson RP-A, Thunnissen A-MWH, Poolman B, Slotboom D-J. Importance of a Hydrophobic Pocket for Peptide Binding in Lactococcal OppA. *Journal of Bacteriology* 2011; 193 : 42544256.
- Béal C, Marin M, Fontaine E, Fonseca F, Obert J-P. Production et Conservation Des Ferments Lactiques et Probiotiques. In : *Bactéries Lactiques, de La Génétique Aux Ferments*. Paris : Lavoisier; 2008. p. 661785.

- Bhalová B, Beran K. Activation of Proteolytic Enzymes during Autolysis of Disintegrated Baker's Yeast. *Folia Microbiologica* 1979; 24 : 455461.
- Biemans-Oldehinkel E, Doeven MK, Poolman B. ABC Transporter Architecture and Regulatory Roles of Accessory Domains. *FEBS Letters* 2006; 580 : 10231035.
- Bingeman TS, Perlman DH, Storey DG, Lewis IA. Digestomics : An Emerging Strategy for Comprehensive Analysis of Protein Catabolism. *Current Opinion in Biotechnology* 2017; 43 : 134140.
- Boggavarapu R, Jeckelmann J-M, Harder D, Ucurum Z, Fotiadis D. Role of Electrostatic Interactions for Ligand Recognition and Specificity of Peptide Transporters. *BMC Biology* 2015; 13 : 58.
- Boguslawski KM, Hill PA, Griffith KL. Novel Mechanisms of Controlling the Activities of the Transcription Factors Spo0A and ComA by the Plasmid-Encoded Quorum Sensing Regulators Rap60-Phr60 in *Bacillus Subtilis*. *Molecular Microbiology* 2015; 96 : 325348.
- Bongiorni C, Stoessel R, Shoemaker D, Perego M. Rap Phosphatase of Virulence Plasmid pXO1 Inhibits *Bacillus Anthracis* Sporulation. *Journal of Bacteriology* 2006; 188 : 487498.
- Borawska J, Darewicz M, Vegarud GE, Minkiewicz P. Antioxidant Properties of Carp (*Cyprinus Carpio* L.) Protein Ex Vivo and in Vitro Hydrolysates. *Food Chemistry* 2016; 194 : 770779.
- Borodovsky M, Rudd KE, Koonin EV. Intrinsic and Extrinsic Approaches for Detecting Genes in a Bacterial Genome. *Nucleic Acids Research* 1994; 22 : 47564767.
- Bouillaut L, Perchat S, Arold S, Zorrilla S, Slamti L, Henry C, et al. Molecular Basis for Group-Specific Activation of the Virulence Regulator PlcR by PapR Heptapeptides. *Nucleic Acids Research* 2008; 36 : 37913801.
- Bourdineaud J-P, Nehmé B, Tesse S, Lonvaud-Funel A. The ftsH Gene of the Wine Bacterium *Oenococcus Oeni* Is Involved in Protection against Environmental Stress. *Applied and Environmental Microbiology* 2003; 69 : 25122520.
- Boutrou R, Gaudichon C, Dupont D, Jardin J, Airinei G, Marsset-Baglieri A, et al. Sequential Release of Milk Protein-Derived Bioactive Peptides in the Jejunum in Healthy Humans. *The American Journal of Clinical Nutrition* 2013; 97 : 13141323.
- Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, et al. GO ::TermFinder –Open Source Software for Accessing Gene Ontology Information and Finding Significantly Enriched Gene Ontology Terms Associated with a List of Genes. *Bioinformatics (Oxford, England)* 2004; 20 : 37103715.

- Breddam K, Beenfeldt T. Acceleration of Yeast Autolysis by Chemical Methods for Production of Intracellular Enzymes. *Applied Microbiology and Biotechnology* 1991 ; 35 : 323329.
- Brinsmade SR, Alexander EL, Livny J, Stettner AI, Segrè D, Rhee KY, et al. Hierarchical Expression of Genes Controlled by the *Bacillus Subtilis* Global Regulatory Protein CodY. *Proceedings of the National Academy of Sciences of the United States of America* 2014 ; 111 : 82278232.
- Brinsmade SR, Sonenshein AL. Dissecting Complex Metabolic Integration Provides Direct Genetic Evidence for CodY Activation by Guanine Nucleotides. *Journal of Bacteriology* 2011 ; 193 : 56375648.
- Brown L, Villegas JM, Elean M, Fadda S, Mozzi F, Saavedra L, et al. YebC, a Putative Transcriptional Factor Involved in the Regulation of the Proteolytic System of *Lactobacillus*. *Scientific Reports* 2017 ; 7 : 8579.
- Brown L, Wolf JM, Prados-Rosales R, Casadevall A. Through the Wall : Extracellular Vesicles in Gram-Positive Bacteria, Mycobacteria and Fungi. *Nature Reviews. Microbiology* 2015 ; 13 : 620630.
- Bryson K, Loux V, Bossy R, Nicolas P, Chaillou S, van de Guchte M, et al. AGMIAL : Implementing an Annotation Strategy for Prokaryote Genomes as a Distributed System. *Nucleic Acids Research* 2006 ; 34 : 35333545.
- Caboche S, Pupin M, Leclère V, Fontaine A, Jacques P, Kucherov G. NORINE : A Database of Nonribosomal Peptides. *Nucleic Acids Research* 2008 ; 36 : D326331.
- Casino P, Rubio V, Marina A. The Mechanism of Signal Transduction by Two-Component Systems. *Current Opinion in Structural Biology* 2010 ; 20 : 763771.
- Caymaris S, Bootsma HJ, Martin B, Hermans PWM, Prudhomme M, Claverys J-P. The Global Nutritional Regulator CodY Is an Essential Protein in the Human Pathogen *Streptococcus Pneumoniae*. *Molecular Microbiology* 2010 ; 78 : 344360.
- Chae HJ, Joo H, In MJ. Utilization of Brewer's Yeast Cells for the Production of Food-Grade Yeast Extract. Part 1 : Effects of Different Enzymatic Treatments on Solid and Protein Recovery and Flavor Characteristics. *Bioresource Technology* 2001 ; 76 : 253258.
- Chambellon E, Yvon M. CodY-Regulated Aminotransferases AraT and BcaT Play a Major Role in the Growth of *Lactococcus Lactis* in Milk by Regulating the Intracellular Pool of Amino Acids. *Applied and Environmental Microbiology* 2003 ; 69 : 30613068.
- Chandler JR, Dunny GM. Characterization of the Sequence Specificity Determinants Required for Processing and Control of Sex Pheromone by the Intramembrane Protease Eep and the Plasmid-Encoded Protein PrgY. *Journal of Bacteriology* 2008 ; 190 : 11721183.

- Chang JC, LaSarre B, Jimenez JC, Aggarwal C, Federle MJ. Two Group A Streptococcal Peptide Pheromones Act through Opposing Rgg Regulators to Control Biofilm Development. *PLOS Pathogens* 2011 ; 7 : e1002190.
- Chao KC, McCarthy EF, McConaghy GA. Yeast Autolysis Process. Patent US4218481A. 1980
- Charbonnel P, Lamarque M, Piard J-C, Gilbert C, Juillard V, Atlan D. Diversity of Oligopeptide Transport Specificity in *Lactococcus Lactis* Species. A Tool to Unravel the Role of OppA in Uptake Specificity. *The Journal of Biological Chemistry* 2003 ; 278 : 1483214840.
- Charpentier C, Feuillat M. Yeast Autolysis. In : *Wine Microbiology and Biotechnology*. Taylor & Francis ; 1993. p. 225242.
- Chatterjee C, Paul M, Xie L, van der Donk WA. Biosynthesis and Mode of Action of Lantibiotics. *Chemical Reviews* 2005 ; 105 : 633684.
- Christensen JE, Dudley EG, Pederson JA, Steele JL. Peptidases and Amino Acid Catabolism in Lactic Acid Bacteria. *Antonie Van Leeuwenhoek* 1999 ; 76 : 217246.
- Clewell DB, An FY, Flannagan SE, Antiporta M, Dunny GM. Enterococcal Sex Pheromone Precursors Are Part of Signal Sequences for Surface Lipoproteins. *Molecular Microbiology* 2000 ; 35 : 246247.
- Clewell DB, Weaver KE. Sex Pheromones and Plasmid Transfer in *Enterococcus Faecalis*. *Plasmid* 1989 ; 21 : 175184.
- Colinge J, Chiappe D, Lagache S, Moniatte M, Bougueleret L. Differential Proteomics via Probabilistic Peptide Identification Scores. *Analytical Chemistry* 2005 ; 77 : 596606.
- Cook LC, Federle MJ. Peptide Pheromone Signaling in *Streptococcus* and *Enterococcus*. *FEMS microbiology reviews* 2014 ; 38 : 473492.
- Courtin P, Miranda G, Guillot A, Wessner F, Mézange C, Domakova E, et al. Peptidoglycan Structure Analysis of *Lactococcus Lactis* Reveals the Presence of an l,d-Carboxypeptidase Involved in Peptidoglycan Maturation. *Journal of Bacteriology* 2006 ; 188 : 52935298.
- Courtin P, Monnet V, Rul F. Cell-Wall Proteinases PrtS and PrtB Have a Different Role in *Streptococcus Thermophilus/Lactobacillus Bulgaricus* Mixed Cultures in Milk. *Microbiology (Reading, England)* 2002 ; 148 : 34133421.
- Cremonesi P, Chessa S, Castiglioni B. Genome Sequence and Analysis of *Lactobacillus Helveticus*. *Frontiers in Microbiology* 2013 ; 3
- Daliri EB-M, Oh DH, Lee BH. Bioactive Peptides. *Foods* 2017 ; 6
- Dallas DC, Guerrero A, Parker EA, Robinson RC, Gan J, German JB, et al. Current Peptidomics : Applications, Purification, Identification, Quantification, and Functional Analysis.

- Proteomics 2015; 15 : 10261038.
- Daniel H, Spanier B, Kottra G, Weitz D. From Bacteria to Man : Archaic Proton-Dependent Peptide Transporters at Work. *Physiology* 2006; 21 : 93102.
- Dekel E, Alon U. Optimality and Evolutionary Tuning of the Expression Level of a Protein. *Nature* 2005; 436 : 588592.
- den Hengst CD, Groeneveld M, Kuipers OP, Kok J. Identification and Functional Characterization of the *Lactococcus Lactis* CodY-Regulated Branched-Chain Amino Acid Permease BcaP (CtrA). *Journal of Bacteriology* 2006; 188 : 32803289.
- den Hengst CD, van Hijum SAFT, Geurts JMW, Nauta A, Kok J, Kuipers OP. The *Lactococcus Lactis* CodY Regulon : Identification of a Conserved Cis-Regulatory Element. *The Journal of Biological Chemistry* 2005; 280 : 3433234342.
- Detmers FJ, Kunji ER, Lanfermeijer FC, Poolman B, Konings WN. Kinetics and Specificity of Peptide Uptake by the Oligopeptide Transport System of *Lactococcus Lactis*. *Biochemistry* 1998; 37 : 1667116679.
- Detmers FJ, Lanfermeijer FC, Abele R, Jack RW, Tampe R, Konings WN, et al. Combinatorial Peptide Libraries Reveal the Ligand-Binding Mechanism of the Oligopeptide Receptor OppA of *Lactococcus Lactis*. *Proceedings of the National Academy of Sciences of the United States of America* 2000; 97 : 1248712492.
- Do H, Makthal N, VanderWal AR, Rettel M, Savitski MM, Peschek N, et al. Leaderless Secreted Peptide Signaling Molecule Alters Global Gene Expression and Increases Virulence of a Human Bacterial Pathogen. *Proceedings of the National Academy of Sciences* 2017; 114 : E8498E8507.
- Doeven MK, Abele R, Tampé R, Poolman B. The Binding Specificity of OppA Determines the Selectivity of the Oligopeptide ATP-Binding Cassette Transporter. *The Journal of Biological Chemistry* 2004; 279 : 3230132307.
- Doeven MK, Kok J, Poolman B. Specificity and Selectivity Determinants of Peptide Transport in *Lactococcus Lactis* and Other Microorganisms. *Molecular Microbiology* 2005; 57 : 640649.
- Doeven M, Slotboom D, W. H. Thunnissen A-M, Poolman B. Hydrodynamic Properties and Crystallization of the Oligopeptide-Binding Protein OppA from *Lactococcus Lactis*. Unpublished results 2006
- Doki S, Kato HE, Solcan N, Iwaki M, Koyama M, Hattori M, et al. Structural Basis for Dynamic Mechanism of Proton-Coupled Symport by the Peptide Transporter POT. *Proceedings of the National Academy of Sciences* 2013; 110 : 1134311348.

- Dubois T, Perchat S, Verplaetse E, Gominet M, Lemy C, Aumont-Nicaise M, et al. Activity of the *Bacillus Thuringiensis* NprR-NprX Cell-Cell Communication System Is Co-Ordinated to the Physiological Stage through a Complex Transcriptional Regulation. *Molecular Microbiology* 2013; 88 : 4863.
- Dunny GM. The Peptide Pheromone-Inducible Conjugation System of *Enterococcus Faecalis* Plasmid pCF10 : Cellcell Signalling, Gene Transfer, Complexity and Evolution. *Philosophical Transactions of the Royal Society B : Biological Sciences* 2007; 362 : 11851193.
- Dunten P, Mowbray SL. Crystal Structure of the Dipeptide Binding Protein from *Escherichia Coli* Involved in Active Transport and Chemotaxis. *Protein Science : A Publication of the Protein Society* 1995; 4 : 23272334.
- Ebbole DJ, Zalkin H. Interaction of a Putative Repressor Protein with an Extended Control Region of the *Bacillus Subtilis* Pur Operon. *The Journal of Biological Chemistry* 1989; 264 : 35533561.
- Fang G, Konings WN, Poolman B. Kinetics and Substrate Specificity of Membrane-Reconstituted Peptide Transporter DtpT of *Lactococcus Lactis*. *Journal of Bacteriology* 2000; 182 : 25302535.
- FAO. Perspectives Agricoles de l'OCDE et de la FAO 2017-2026. Chapitre spécial : Asie du Sud-Est. Rome : FAO/OECD; 2017.
- Fernandez-Espla MD, Garault P, Monnet V, Rul F. *Streptococcus Thermophilus* Cell Wall-Anchored Proteinase : Release, Purification, and Biochemical and Genetic Characterization. *Applied and Environmental Microbiology* 2000; 66 : 47724778.
- Fernandez-Espla MD, Rul F. PepS from *Streptococcus Thermophilus*. A New Member of the Aminopeptidase T Family of Thermophilic Bacteria. *European journal of biochemistry / FEBS* 1999; 263 : 502510.
- Fiocco D, Collins M, Muscariello L, Hols P, Kleerebezem M, Msadek T, et al. The *Lactobacillus Plantarum* ftsH Gene Is a Novel Member of the CtsR Stress Response Regulon. *Journal of Bacteriology* 2009; 191 : 16881694.
- Fixen KR, Chandler JR, Le T, Kozlowicz BK, Manias DA, Dunny GM. Analysis of the Amino Acid Sequence Specificity Determinants of the Enterococcal cCF10 Sex Pheromone in Interactions with the Pheromone-Sensing Machinery. *Journal of Bacteriology* 2007; 189 : 13991406.
- Flannagan SE, Clewell DB. Identification and Characterization of Genes Encoding Sex Pheromone cAM373 Activity in *Enterococcus Faecalis* and *Staphylococcus Aureus*. *Molecular Microbiology* 2002; 44 : 803817.

- Fleuchot B, Gitton C, Guillot A, Vidic J, Nicolas P, Besset C, et al. Rgg Proteins Associated with Internalized Small Hydrophobic Peptides : A New Quorum-Sensing Mechanism in Streptococci. *Molecular Microbiology* 2011 ; 80 : 11021119.
- Fleuchot B, Guillot A, Mézange C, Besset C, Chambellon E, Monnet V, et al. Rgg-Associated SHP Signaling Peptides Mediate Cross-Talk in Streptococci. *PloS One* 2013 ; 8 : e66042.
- Fontaine L, Boutry C, de Frahan MH, Delplace B, Fremaux C, Horvath P, et al. A Novel Pheromone Quorum-Sensing System Controls the Development of Natural Competence in *Streptococcus Thermophilus* and *Streptococcus Salivarius*. *Journal of Bacteriology* 2010 ; 192 : 14441454.
- Fontaine L, Goffin P, Dubout H, Delplace B, Baulard A, Lecat-Guillet N, et al. Mechanism of Competence Activation by the ComRS Signalling System in Streptococci. *Molecular Microbiology* 2013 ; 87 : 11131132.
- Fontaine L, Wahl A, Flécharde M, Mignolet J, Hols P. Regulation of Competence for Natural Transformation in Streptococci. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* 2015 ; 33 : 343360.
- Foucaud C, Hemme D, Desmazeaud M. Peptide Utilization by *Lactococcus Lactis* and *Leuconostoc Mesenteroides*. *Letters in Applied Microbiology* 2001 ; 32 : 2025.
- Foucaud C, Kunji ER, Hagting A, Richard J, Konings WN, Desmazeaud M, et al. Specificity of Peptide Transport Systems in *Lactococcus Lactis* : Evidence for a Third System Which Transports Hydrophobic Di- and Tripeptides. *Journal of Bacteriology* 1995 ; 177 : 46524657.
- Freiberg JA, Le Breton Y, Tran BQ, Scott AJ, Harro JM, Ernst RK, et al. Global Analysis and Comparison of the Transcriptomes and Proteomes of Group A *Streptococcus* Biofilms. *mSystems* 2016 ; 1
- Friedman RC, Kalkhof S, Doppelt-Azeroual O, Mueller SA, Chovancová M, von Bergen M, et al. Common and Phylogenetically Widespread Coding for Peptides by Bacterial Small RNAs. *BMC genomics* 2017 ; 18 : 553.
- Gao R, Stock AM. Biological Insights from Structures of Two-Component Proteins. *Annual Review of Microbiology* 2009 ; 63 : 133154.
- Garault P, Le Bars D, Besset C, Monnet V. Three Oligopeptide-Binding Proteins Are Involved in the Oligopeptide Transport of *Streptococcus Thermophilus*. *The Journal of Biological Chemistry* 2002 ; 277 : 3239.
- Gardan R, Besset C, Gitton C, Guillot A, Fontaine L, Hols P, et al. Extracellular Life Cycle of ComS, the Competence-Stimulating Peptide of *Streptococcus Thermophilus*. *Journal of*

- Bacteriology 2013; 195 : 18451855.
- Gardan R, Besset C, Guillot A, Gitton C, Monnet V. The Oligopeptide Transport System Is Essential for the Development of Natural Competence in *Streptococcus Thermophilus* Strain LMD-9. *Journal of Bacteriology* 2009; 191 : 46474655.
- Germain-Alpettaz V, Foucaud-Scheunemann C. Identification and Characterization of an Oligopeptide Transport System in *Leuconostoc Mesenteroides* Subsp. *Mesenteroides* CNRZ 1463. *Letters in Applied Microbiology* 2002; 35 : 6873.
- Gimpel M, Brantl S. Dual-Function Small Regulatory RNAs in Bacteria. *Molecular Microbiology* 2017; 103 : 387397.
- Goh YJ, Goin C, O'Flaherty S, Altermann E, Hutkins R. Specialized Adaptation of a Lactic Acid Bacterium to the Milk Environment : The Comparative Genomics of *Streptococcus Thermophilus* LMD-9. *Microbial Cell Factories* 2011; 10 Suppl 1 : S22.
- Golomb BL, Yu AO, Coates LC, Marco ML. The *Lactococcus Lactis* KF147 Nonribosomal Peptide Synthetase/Polyketide Synthase System Confers Resistance to Oxidative Stress during Growth on Plant Leaf Tissue Lysate. *MicrobiologyOpen* 2017; 7
- Gomathi L, Subramanian S. Elucidation of Secondary Structures of Peptides Using High Resolution NMR. *Current Science* 1996; 71 : 553567.
- Goodell EW. Recycling of Murein by *Escherichia Coli*. *Journal of Bacteriology* 1985; 163 : 305310.
- Gordon WG, Semmett WF, Bender M. Alanine, Glycine and Proline Contents of Casein and Its Components¹. *Journal of the American Chemical Society* 1950; 72 : 42824282.
- Groot Kormelink T, Koenders E, Hagemeijer Y, Overmars L, Siezen RJ, de Vos WM, et al. Comparative Genome Analysis of Central Nitrogen Metabolism and Its Control by GlnR in the Class Bacilli. *BMC genomics* 2012; 13 : 191.
- Guettou F, Quistgaard EM, Trésaugues L, Moberg P, Jegerschöld C, Zhu L, et al. Structural Insights into Substrate Recognition in Proton-dependent Oligopeptide Transporters. *EMBO reports* 2013; 14 : 804810.
- Guédon E, Renault P, Ehrlich SD, Delorme C. Transcriptional Pattern of Genes Coding for the Proteolytic System of *Lactococcus Lactis* and Evidence for Coordinated Regulation of Key Enzymes by Peptide Supply. *Journal of Bacteriology* 2001a; 183 : 36143622.
- Guédon E, Serror P, Ehrlich SD, Renault P, Delorme C. Pleiotropic Transcriptional Repressor CodY Senses the Intracellular Pool of Branched-Chain Amino Acids in *Lactococcus Lactis*. *Molecular Microbiology* 2001b; 40 : 12271239.

- Guédon E, Sperandio B, Pons N, Ehrlich SD, Renault P. Overall Control of Nitrogen Metabolism in *Lactococcus Lactis* by CodY, and Possible Models for CodY Regulation in Firmicutes. *Microbiology (Reading, England)* 2005; 151 : 38953909.
- Guillot A, Boulay M, Chambellon É, Gitton C, Monnet V, Juillard V. Mass Spectrometry Analysis of the Extracellular Peptidome of *Lactococcus Lactis* : Lines of Evidence for the Coexistence of Extracellular Protein Hydrolysis and Intracellular Peptide Excretion. *Journal of Proteome Research* 2016; 15 : 32143224.
- Gundry RL, White MY, Murray CI, Kane LA, Fu Q, Stanley BA, et al. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* 2009; CHAPTER : Unit10.25.
- Guyer CA, Morgan DG, Staros JV. Binding Specificity of the Periplasmic Oligopeptide-Binding Protein from *Escherichia Coli*. *Journal of Bacteriology* 1986; 168 : 775779.
- Hafeez Z, Cakir-Kiefer C, Girardet J-M, Jardin J, Perrin C, Dary A, et al. Hydrolysis of Milk-Derived Bioactive Peptides by Cell-Associated Extracellular Peptidases of *Streptococcus Thermophilus*. *Applied Microbiology and Biotechnology* 2013; 97 : 97879799.
- Hafeez Z, Cakir-Kiefer C, Girardet J-M, Lecomte X, Paris C, Galia W, et al. New Insights into the Proteolytic System of *Streptococcus Thermophilus* : Use of Isracidin To Characterize Cell-Associated Extracellular Peptidase Activities. *Journal of Agricultural and Food Chemistry* 2015; 63 : 75227531.
- Hagting A, Knol J, Hasemeier B, Streutker MR, Fang G, Poolman B, et al. Amplified Expression, Purification and Functional Reconstitution of the Dipeptide and Tripeptide Transport Protein of *Lactococcus Lactis*. *European Journal of Biochemistry* 1997; 247 : 581587.
- Hagting A, Kunji ER, Leenhouts KJ, Poolman B, Konings WN. The Di- and Tripeptide Transport Protein of *Lactococcus Lactis*. A New Type of Bacterial Peptide Transporter. *The Journal of Biological Chemistry* 1994; 269 : 1139111399.
- Haider S, Pal R. Integrated Analysis of Transcriptomic and Proteomic Data. *Current Genomics* 2013; 14 : 91110.
- Hawver LA, Jung SA, Ng W-L. Specificity and Complexity in Bacterial Quorum-Sensing Systems. *FEMS microbiology reviews* 2016; 40 : 738752.
- Håvarstein LS, Coomaraswamy G, Morrison DA. An Unmodified Heptadecapeptide Pheromone Induces Competence for Genetic Transformation in *Streptococcus Pneumoniae*. *Proceedings of the National Academy of Sciences of the United States of America* 1995; 92 : 1114011144.

- Håvarstein LS, Hakenbeck R, Gaustad P. Natural Competence in the Genus *Streptococcus* : Evidence That *Streptococci* Can Change Pherotype by Interspecies Recombinational Exchanges. *Journal of Bacteriology* 1997; 179 : 65896594.
- Håvarstein LS, Holo H, Nes IF. The Leader Peptide of Colicin V Shares Consensus Sequences with Leader Peptides That Are Common among Peptide Bacteriocins Produced by Gram-Positive Bacteria. *Microbiology* 1994; 140 : 23832389.
- Hecht KA, O'Donnell AF, Brodsky JL. The Proteolytic Landscape of the Yeast Vacuole. *Cellular Logistics* 2014; 4 : e28023.
- Helinck S, Charbonnel P, Foucaud-Scheunemann C, Piard J-C, Juillard V. Charged Casein-Derived Oligopeptides Competitively Inhibit the Transport of a Reporter Oligopeptide by *Lactococcus Lactis*. *Journal of Applied Microbiology* 2003; 94 : 900907.
- Hemm MR, Paul BJ, Miranda-Ríos J, Zhang A, Soltanzad N, Storz G. Small Stress Response Proteins in *Escherichia Coli* : Proteins Missed by Classical Proteomic Studies. *Journal of Bacteriology* 2010; 192 : 4658.
- Hendriksen WT, Bootsma HJ, Estevão S, Hoogenboezem T, de Jong A, de Groot R, et al. CodY of *Streptococcus Pneumoniae* : Link between Nutritional Gene Regulation and Colonization. *Journal of Bacteriology* 2008; 190 : 590601.
- Herve-Jimenez L, Guillouard I, Guedon E, Boudebouze S, Hols P, Monnet V, et al. Postgenomic Analysis of *Streptococcus Thermophilus* Cocultivated in Milk with *Lactobacillus Delbrueckii* Subsp. *Bulgaricus* : Involvement of Nitrogen, Purine, and Iron Metabolism. *Applied and Environmental Microbiology* 2009; 75 : 20622073.
- Higgins CF. ABC Transporters : From Microorganisms to Man. *Annual Review of Cell Biology* 1992; 8 : 67113.
- Hobbs EC, Fontaine F, Yin X, Storz G. An Expanding Universe of Small Proteins. *Current Opinion in Microbiology* 2011; 14 : 167173.
- Hols P, Hancy F, Fontaine L, Grossiord B, Prozzi D, Leblond-Bourget N, et al. New Insights in the Molecular Biology and Physiology of *Streptococcus Thermophilus* Revealed by Comparative Genomics. *FEMS Microbiology Reviews* 2005; 29 : 435463.
- Hoover SE, Perez AJ, Tsui H-CT, Sinha D, Smiley DL, DiMarchi RD, et al. A New Quorum-Sensing System (TprA/PhrA) for *Streptococcus Pneumoniae* That Regulates a Lantibiotic Biosynthesis Gene Cluster. *Molecular Microbiology* 2015; 97 : 229243.
- Hsu S-TD, Breukink E, Bierbaum G, Sahl H-G, de Kruijff B, Kaptein R, et al. NMR Study of Mersacidin and Lipid II Interaction in Dodecylphosphocholine Micelles. Conformational Changes Are a Key to Antimicrobial Activity. *The Journal of Biological Chemistry* 2003;

- 278 : 1311013117.
- Ibrahim M, Guillot A, Wessner F, Algaron F, Besset C, Courtin P, et al. Control of the Transcription of a Short Gene Encoding a Cyclic Peptide in *Streptococcus Thermophilus* : A New Quorum-Sensing System? *Journal of Bacteriology* 2007a; 189 : 88448854.
- Ibrahim M, Nicolas P, Bessières P, Bolotin A, Monnet V, Gardan R. A Genome-Wide Survey of Short Coding Sequences in Streptococci. *Microbiology (Reading, England)* 2007b; 153 : 36313644.
- Ito K, Akiyama Y. Cellular Functions, Mechanism of Action, and Regulation of Ftsh Protease. *Annual Review of Microbiology* 2005; 59 : 211231.
- Jameh N, Galia W, Awussi AA, Roux E, Genay M, Perrin C, et al. Characterization of a New Peptide Transport System in *Streptococcus Thermophilus*. *Food Research International* 2016; 86 : 3445.
- Jäpelt KB, Johnsen LG, Christensen JH. Removal of Polysorbate 80 by Complexation Prior to LC-MS Analysis. *Analytical and Bioanalytical Chemistry* 2016; 408 : 23032307.
- Ji G, Beavis R, Novick RP. Bacterial Interference Caused by Autoinducing Peptide Variants. *Science (New York, N.Y.)* 1997; 276 : 20272030.
- Jiang M, Grau R, Perego M. Differential Processing of Propeptide Inhibitors of Rap Phosphatases in *Bacillus Subtilis*. *Journal of Bacteriology* 2000; 182 : 303310.
- Jimenez JC, Federle MJ. Quorum Sensing in Group A *Streptococcus*. *Frontiers in Cellular and Infection Microbiology* 2014; 4 : 127.
- Jones EW. Three Proteolytic Systems in the Yeast *Saccharomyces Cerevisiae*. *The Journal of Biological Chemistry* 1991; 266 : 79637966.
- Juillard V, Guillot A, Le Bars D, Gripon JC. Specificity of Milk Peptide Utilization by *Lactococcus Lactis*. *Applied and Environmental Microbiology* 1998; 64 : 12301236.
- Juillard V, Le Bars D, Kunji ER, Konings WN, Gripon JC, Richard J. Oligopeptides Are the Main Source of Nitrogen for *Lactococcus Lactis* during Growth in Milk. *Applied and Environmental Microbiology* 1995; 61 : 30243030.
- Juillard V, Lopez-Kleine L, Monnet V. Proteolytic Systems of Lactic Acid Bacteria (LAB). In : *Reference Module in Food Science*. Elsevier; 2016.
- Juille O, Le Bars D, Juillard V. The Specificity of Oligopeptide Transport by *Streptococcus Thermophilus* Resembles That of *Lactococcus Lactis* and Not That of Pathogenic Streptococci. *Microbiology (Reading, England)* 2005; 151 : 19871994.
- Kalum L. Method for Producing a Yeast Extract. Patent EP2097510A1. 2009

- Kanegae Y, Sugiyama Y, Minami K. Method for Producing Yeast Extract. Patent US4810509A. 1989
- Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA : KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology* 2016; 428 : 726731.
- Keiderling TA. Protein and Peptide Secondary Structure and Conformational Determination with Vibrational Circular Dichroism. *Current Opinion in Chemical Biology* 2002; 6 : 682688.
- Kerr AR, Adrian PV, Estevão S, de Groot R, Alloing G, Claverys J-P, et al. The Ami-AliA/AliB Permease of *Streptococcus Pneumoniae* Is Involved in Nasopharyngeal Colonization but Not in Invasive Disease. *Infection and Immunity* 2004; 72 : 39023906.
- Kevvai K, Kütt M-L, Nisamedtinov I, Paalme T. Utilization of (15)N-Labelled Yeast Hydrolysate in *Lactococcus Lactis* IL1403 Culture Indicates Co-Consumption of Peptide-Bound and Free Amino Acids with Simultaneous Efflux of Free Amino Acids. *Antonie Van Leeuwenhoek* 2014; 105 : 511522.
- Kim JH, Lee J, Park J, Gho YS. Gram-Negative and Gram-Positive Bacterial Extracellular Vesicles. *Seminars in Cell & Developmental Biology* 2015; 40 : 97104.
- Kleerebezem M. Quorum Sensing Control of Lantibiotic Production; Nisin and Subtilin Autoregulate Their Own Biosynthesis. *Peptides* 2004; 25 : 14051414.
- Kleerebezem M, Boekhorst J, Kranenburg R van, Molenaar D, Kuipers OP, Leer R, et al. Complete Genome Sequence of *Lactobacillus Plantarum* WCFS1. *Proceedings of the National Academy of Sciences* 2003; 100 : 19901995.
- Klepsch MM, Kovermann M, Löw C, Balbach J, Permentier HP, Fusetti F, et al. *Escherichia Coli* Peptide Binding Protein OppA Has a Preference for Positively Charged Peptides. *Journal of Molecular Biology* 2011; 414 : 7585.
- Konings WN. The Cell Membrane and the Struggle for Life of Lactic Acid Bacteria. *Antonie van Leeuwenhoek* 2002; 82 : 327.
- Kopp E, Marahiel MA. Where Chemistry Meets Biology : The Chemoenzymatic Synthesis of Nonribosomal Peptides and Polyketides. *Current Opinion in Biotechnology* 2007; 18 : 513520.
- Kruskal WH, Wallis WA. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association* 1952; 47 : 583621.
- Kuipers OP, Beerthuyzen MM, de Ruyter PG, Luesink EJ, de Vos WM. Autoregulation of Ni-

- sin Biosynthesis in *Lactococcus Lactis* by Signal Transduction. *The Journal of Biological Chemistry* 1995; 270 : 2729927304.
- Kunji ER, Hagting A, De Vries CJ, Juillard V, Haandrikman AJ, Poolman B, et al. Transport of Beta-Casein-Derived Peptides by the Oligopeptide Transport System Is a Crucial Step in the Proteolytic Pathway of *Lactococcus Lactis*. *The Journal of Biological Chemistry* 1995; 270 : 15691574.
- Kunji ER, Mierau I, Hagting A, Poolman B, Konings WN. The Proteolytic Systems of Lactic Acid Bacteria. *Antonie Van Leeuwenhoek* 1996a; 70 : 187221.
- Kunji ERS, Mierau I, Poolman B, Konings WN, Venema G, Kok J. Fate of Peptides in Peptidase Mutants of *Lactococcus Lactis*. *Molecular Microbiology* 1996b; 21 : 123131.
- Kurihara K. Umami the Fifth Basic Taste : History of Studies on Receptor Mechanisms and Role as a Food Flavor. *BioMed Research International* 2015; 2015
- Kyte J, Doolittle RF. A Simple Method for Displaying the Hydrophobic Character of a Protein. *Journal of Molecular Biology* 1982; 157 : 105132.
- Lahrichi SL, Affolter M, Zolezzi IS, Panchaud A. Food Peptidomics : Large Scale Analysis of Small Bioactive Peptides—a Pilot Study. *Journal of Proteomics* 2013; 88 : 8391.
- Lamarque M, Aubel D, Piard J-C, Gilbert C, Juillard V, Atlan D. The Peptide Transport System Opt Is Involved in Both Nutrition and Environmental Sensing during Growth of *Lactococcus Lactis* in Milk. *Microbiology (Reading, England)* 2011; 157 : 16121619.
- Lamarque M, Charbonnel P, Aubel D, Piard J-C, Atlan D, Juillard V. A Multifunction ABC Transporter (Opt) Contributes to Diversity of Peptide Uptake Specificity within the Genus *Lactococcus*. *Journal of Bacteriology* 2004; 186 : 64926500.
- Lambert G, Kussell E. Memory and Fitness Optimization of Bacteria under Fluctuating Environments. *PLoS Genetics* 2014; 10
- Lanfermeijer FC, Detmers FJ, Konings WN, Poolman B. On the Binding Mechanism of the Peptide Receptor of the Oligopeptide Transport System of *Lactococcus Lactis*. *The EMBO journal* 2000; 19 : 36493656.
- Lanfermeijer FC, Picon A, Konings WN, Poolman B. Kinetics and Consequences of Binding of Nona- and Dodecapeptides to the Oligopeptide Binding Protein (OppA) of *Lactococcus Lactis*. *Biochemistry* 1999; 38 : 1444014450.
- Langella O, Valot B, Balliau T, Blein-Nicolas M, Bonhomme L, Zivy M. XTandemPipeline : A Tool to Manage Sequence Redundancy for Protein Inference and Phosphosite Identification. *Journal of Proteome Research* 2016

- Lanigan-Gerdes S, Briceno G, Dooley AN, Faull KF, Lazazzera BA. Identification of Residues Important for Cleavage of the Extracellular Signaling Peptide CSF of *Bacillus Subtilis* from Its Precursor Protein. *Journal of Bacteriology* 2008; 190 : 66686675.
- Lazazzera BA, Solomon JM, Grossman AD. An Exported Peptide Functions Intracellularly to Contribute to Cell Density Signaling in *B. Subtilis*. *Cell* 1997; 89 : 917925.
- Leonard BA, Podbielski A, Hedberg PJ, Dunny GM. Enterococcus Faecalis Pheromone Binding Protein, PrgZ, Recruits a Chromosomal Oligopeptide Permease System to Import Sex Pheromone cCF10 for Induction of Conjugation. *Proceedings of the National Academy of Sciences of the United States of America* 1996; 93 : 260264.
- Letort C, Juillard V. Development of a Minimal Chemically-Defined Medium for the Exponential Growth of *Streptococcus Thermophilus*. *Journal of Applied Microbiology* 2001; 91 : 10231029.
- Letunic I, Bork P. Interactive Tree of Life (iTOL) v3 : An Online Tool for the Display and Annotation of Phylogenetic and Other Trees. *Nucleic Acids Research*; 44 : W242245.
- Levdikov VM, Blagova E, Joseph P, Sonenshein AL, Wilkinson AJ. The Structure of CodY, a GTP- and Isoleucine-Responsive Regulator of Stationary Phase and Virulence in Gram-Positive Bacteria. *Journal of Biological Chemistry* 2006; 281 : 1136611373.
- Levdikov VM, Blagova EV, Brannigan JA, Wright L, Vagin AA, Wilkinson AJ. The Structure of the Oligopeptide-Binding Protein, AppA, from *Bacillus Subtilis* in Complex with a Nonapeptide. *Journal of Molecular Biology* 2005; 345 : 879892.
- Levdikov VM, Blagova E, Young VL, Belitsky BR, Lebedev A, Sonenshein AL, et al. Structure of the Branched-Chain Amino Acid and GTP-Sensing Global Regulator, CodY, from *Bacillus Subtilis*. *The Journal of Biological Chemistry* 2017; 292 : 27142728.
- Lê Cao K-A, Boitard S, Besse P. Sparse PLS Discriminant Analysis : Biologically Relevant Feature Selection and Graphical Displays for Multiclass Problems. *BMC Bioinformatics* 2011; 12 : 253.
- Lin XB, Lohans CT, Duar R, Zheng J, Vederas JC, Walter J, et al. Genetic Determinants of Reutericyclin Biosynthesis in *Lactobacillus Reuteri*. *Applied and Environmental Microbiology* 2015; 81 : 20322041.
- Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, et al. Direct Analysis of Protein Complexes Using Mass Spectrometry. *Nature Biotechnology* 1999; 17 : 676682.
- Lipmann F, Gevers W, Kleinkauf H, Roskoski R. Polypeptide Synthesis on Protein Templates : The Enzymatic Synthesis of Gramicidin S and Tyrocidine. *Advances in Enzymology and Related Areas of Molecular Biology* 1971; 35 : 134.

- Liu H, Sadygov RG, Yates JR. A Model for Random Sampling and Estimation of Relative Protein Abundance in Shotgun Proteomics. *Analytical Chemistry* 2004; 76 : 41934201.
- Liu M, Bayjanov JR, Renckens B, Nauta A, Siezen RJ. The Proteolytic System of Lactic Acid Bacteria Revisited : A Genomic Comparison. *BMC Genomics* 2010; 11 : 36.
- Lobel L, Sigal N, Borovok I, Ruppin E, Herskovits AA. Integrative Genomic Analysis Identifies Isoleucine and CodY as Regulators of *Listeria Monocytogenes* Virulence. *PLoS genetics* 2012; 8 : e1002887.
- Locher KP. Review. Structure and Mechanism of ATP-Binding Cassette Transporters. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 2009; 364 : 239245.
- Locher KP. Mechanistic Diversity in ATP-Binding Cassette (ABC) Transporters. *Nature Structural & Molecular Biology* 2016; 23 : 487493.
- Love MI, Huber W, Anders S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biology* 2014; 15 : 550.
- Lu WW, Wang Y, Wang T, Kong J. The Global Regulator CodY in *Streptococcus Thermophilus* Controls the Metabolic Network for Escalating Growth in the Milk Environment. *Applied and Environmental Microbiology* 2015; 81 : 23492358.
- Luo W, Brouwer C. Pathview : An R/Bioconductor Package for Pathway-Based Data Integration and Visualization. *Bioinformatics* 2013; 29 : 18301831.
- Luo W, Pant G, Bhavnasi YK, Blanchard SG, Brouwer C. Pathview Web : User Friendly Pathway Visualization and Data Integration. *Nucleic Acids Research* 2017; 45 : W501W508.
- Lyon GJ, Novick RP. Peptide Signaling in *Staphylococcus Aureus* and Other Gram-Positive Bacteria. *Peptides* 2004; 25 : 13891403.
- Lyutvinskiy Y, Yang H, Rutishauser D, Zubarev RA. In Silico Instrumental Response Correction Improves Precision of Label-Free Proteomics and Accuracy of Proteomics-Based Predictive Models. *Molecular & cellular proteomics : MCP* 2013; 12 : 23242331.
- Majerczyk CD, Sadykov MR, Luong TT, Lee C, Somerville GA, Sonenshein AL. *Staphylococcus Aureus* CodY Negatively Regulates Virulence Gene Expression. *Journal of Bacteriology* 2008; 190 : 22572265.
- Manes NP, Gustin JK, Rue J, Mottaz HM, Purvine SO, Norbeck AD, et al. Targeted Protein Degradation by *Salmonella* under Phagosome-Mimicking Culture Conditions Investigated Using Comparative Peptidomics. *Molecular & cellular proteomics : MCP* 2007; 6 : 717727.
- Mann HB, Whitney DR. On a Test of Whether One of Two Random Variables Is Stochastically

- Larger than the Other. *The Annals of Mathematical Statistics* 1947; 18 : 5060.
- Mao B, Pear MR, McCammon JA, Quioco FA. Hinge-Bending in L-Arabinose-Binding Protein. The "Venus's-Flytrap" Model. *The Journal of Biological Chemistry* 1982; 257 : 11311133.
- Mao X, Cai T, Olyarchuk JG, Wei L. Automated Genome Annotation and Pathway Identification Using the KEGG Orthology (KO) as a Controlled Vocabulary. *Bioinformatics (Oxford, England)* 2005; 21 : 37873793.
- Martinez Molledo M, Quistgaard EM, Flayhan A, Pieprzyk J, Löw C. Multispecific Substrate Recognition in a Proton-Dependent Oligopeptide Transporter. *Structure* 2018; 26 : 467476.e4.
- Martínez-Núñez MA, López VEL y. Nonribosomal Peptides Synthetases and Their Applications in Industry. *Sustainable Chemical Processes* 2016; 4 : 13.
- Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A. Comparative Protein Structure Modeling of Genes and Genomes. *Annual Review of Biophysics and Biomolecular Structure* 2000; 29 : 291325.
- Mashburn-Warren L, Morrison DA, Federle MJ. A Novel Double-Tryptophan Peptide Pheromone Controls Competence in *Streptococcus* Spp. via an Rgg Regulator. *Molecular Microbiology* 2010; 78 : 589606.
- Mashburn-Warren L, Morrison DA, Federle MJ. The Cryptic Competence Pathway in *Streptococcus Pyogenes* Is Controlled by a Peptide Pheromone. *Journal of Bacteriology* 2012; 194 : 45894600.
- MDowell P, Affas Z, Reynolds C, Holden MT, Wood SJ, Saint S, et al. Structure, Activity and Evolution of the Group I Thiolactone Peptide Quorum-Sensing System of *Staphylococcus Aureus*. *Molecular Microbiology* 2001; 41 : 503512.
- Miller CG, Kukral AM, Miller JL, Movva NR. *pepM* Is an Essential Gene in *Salmonella Typhimurium*. *Journal of Bacteriology* 1989; 171 : 52155217.
- Miller MB, Bassler BL. Quorum Sensing in Bacteria. *Annual Review of Microbiology* 2001; 55 : 165199.
- Mirouze N, Parashar V, Baker MD, Dubnau DA, Neiditch MB. An Atypical Phr Peptide Regulates the Developmental Switch Protein RapH ∇ . *Journal of Bacteriology* 2011; 193 : 61976206.
- Molle V, Nakaura Y, Shivers RP, Yamaguchi H, Losick R, Fujita Y, et al. Additional Targets of the *Bacillus Subtilis* Global Regulator CodY Identified by Chromatin Immunoprecipitation

- and Genome-Wide Transcript Analysis. *Journal of Bacteriology* 2003; 185 : 19111922.
- Molnár-Perl I. Advancement in the Derivatizations of the Amino Groups with the O-Phthaldehyde-Thiol and with the 9-Fluorenylmethyloxycarbonyl Chloride Reagents. *Journal of Chromatography B* 2011; 879 : 12411269.
- Monedero V, Revilla-Guarinos A, Zúñiga M. Physiological Role of Two-Component Signal Transduction Systems in Food-Associated Lactic Acid Bacteria. *Advances in Applied Microbiology* 2017; 99 : 151.
- Monnet V, Gardan R. Quorum-Sensing Regulators in Gram-Positive Bacteria : 'Cherchez Le Peptide'. *Molecular Microbiology* 2015; 97 : 181184.
- Monnet V, Juillard V, Gardan R. Peptide Conversations in Gram-Positive Bacteria. *Critical Reviews in Microbiology* 2014 : 113.
- Moore DS. Amino Acid and Peptide Net Charges : A Simple Calculational Procedure. *Biochemical Education* 1985; 13 : 1011.
- Morello E, Bermúdez-Humarán LG, Llull D, Solé V, Miraglio N, Langella P, et al. *Lactococcus Lactis*, an Efficient Cell Factory for Recombinant Protein Production and Secretion. *Journal of Molecular Microbiology and Biotechnology* 2008; 14 : 4858.
- Mori M, Sakagami Y, Ishii Y, Isogai A, Kitada C, Fujino M, et al. Structure of cCF10, a Peptide Sex Pheromone Which Induces Conjugative Transfer of the *Streptococcus Faecalis* Tetracycline Resistance Plasmid, pCF10. *The Journal of Biological Chemistry* 1988; 263 : 1457414578.
- Mosser M, Kapel R, Aymes A, Bonanno L-M, Olmos E, Chevalot I, et al. Chromatographic Fractionation of Yeast Extract : A Strategy to Identify Physicochemical Properties of Compounds Promoting CHO Cell Culture. *Process Biochemistry* 2012; 47 : 11781185.
- Mosser M, Kapel R, Chevalot I, Olmos E, Marc I, Marc A, et al. Fractionation of Yeast Extract by Nanofiltration Process to Assess Key Compounds Involved in CHO Cell Culture Improvement. *Biotechnology Progress* 2015; 31 : 875882.
- Müller SA, FindeiSS S, Pernitzsch SR, Wissenbach DK, Stadler PF, Hofacker IL, et al. Identification of New Protein Coding Sequences and Signal Peptidase Cleavage Sites of *Helicobacter Pylori* Strain 26695 by Proteogenomics. *Journal of Proteomics* 2013; 86 : 2742.
- Nakajima H, Hagting A, Kunji ER, Poolman B, Konings WN. Cloning and Functional Expression in *Escherichia Coli* of the Gene Encoding the Di- and Tripeptide Transport Protein of *Lactobacillus Helveticus*. *Applied and Environmental Microbiology* 1997; 63 : 22132217.
- Nakayama J, Cao Y, Horii T, Sakuda S, Akkermans AD, de Vos WM, et al. Gelatinase

- Biosynthesis-Activating Pheromone : A Peptide Lactone That Mediates a Quorum Sensing in *Enterococcus Faecalis*. *Molecular Microbiology* 2001 ; 41 : 145154.
- Nakayama J, Chen S, Oyama N, Nishiguchi K, Azab EA, Tanaka E, et al. Revised Model for *Enterococcus Faecalis* Fsr Quorum-Sensing System : The Small Open Reading Frame *fsrD* Encodes the Gelatinase Biosynthesis-Activating Pheromone Propeptide Corresponding to *Staphylococcal AgrD*. *Journal of Bacteriology* 2006 ; 188 : 83218326.
- Nealson KH, Hastings JW. Bacterial Bioluminescence : Its Control and Ecological Significance. *Microbiological Reviews* 1979 ; 43 : 496518.
- Neiditch MB, Capodagli GC, Prehna G, Federle MJ. Genetic and Structural Analyses of RRNPP Intercellular Peptide Signaling of Gram-Positive Bacteria. *Annual Review of Genetics* 2017 ; 51 : 311333.
- Neiditch MB, Federle MJ, Pompeani AJ, Kelly RC, Swem DL, Jeffrey PD, et al. Ligand-Induced Asymmetry in Histidine Sensor Kinase Complex Regulates Quorum Sensing. *Cell* 2006 ; 126 : 10951108.
- Newstead S. Molecular Insights into Proton Coupled Peptide Transport in the PTR Family of Oligopeptide Transporters. *Biochimica Et Biophysica Acta* 2015 ; 1850 : 488499.
- Newstead S. Recent Advances in Understanding Proton Coupled Peptide Transport via the POT Family. *Current Opinion in Structural Biology* 2017 ; 45 : 1724.
- Newstead S, Drew D, Cameron AD, Postis VLG, Xia X, Fowler PW, et al. Crystal Structure of a Prokaryotic Homologue of the Mammalian Oligopeptideproton Symporters, PepT1 and PepT2. *The EMBO Journal* 2011 ; 30 : 417426.
- Ng W-L, Bassler BL. Bacterial Quorum-Sensing Network Architectures. *Annual Review of Genetics* 2009 ; 43 : 197222.
- Ng W-L, Kazmierczak KM, Robertson GT, Gilmour R, Winkler ME. Transcriptional Regulation and Signature Patterns Revealed by Microarray Analyses of *Streptococcus Pneumoniae* R6 Challenged with Sublethal Concentrations of Translation Inhibitors. *Journal of Bacteriology* 2003 ; 185 : 359370.
- Nicolas P, Bize L, Muri F, Hoebeke M, Rodolphe F, Ehrlich SD, et al. Mining *Bacillus Subtilis* Chromosome Heterogeneities Using Hidden Markov Models. *Nucleic Acids Research* 2002 ; 30 : 14181426.
- Nilsson D, Lauridsen AA, Tomoyasu T, Ogura T. A *Lactococcus Lactis* Gene Encodes a Membrane Protein with Putative ATPase Activity That Is Homologous to the Essential *Escherichia Coli* *ftsH* Gene Product. *Microbiology (Reading, England)* 1994 ; 140 (Pt 10) : 26012610.

- Nishiguchi K, Nagata K, Tanokura M, Sonomoto K, Nakayama J. Structure-Activity Relationship of Gelatinase Biosynthesis-Activating Pheromone of *Enterococcus Faecalis*. *Journal of Bacteriology* 2009; 191 : 641650.
- Nongonierma AB, Paoletta S, Mudgil P, Maqsood S, FitzGerald RJ. Identification of Novel Di-peptidyl Peptidase IV (DPP-IV) Inhibitory Peptides in Camel Milk Protein Hydrolysates. *Food Chemistry* 2018; 244 : 340348.
- Norris R, Poyarkov A, O’Keeffe MB, FitzGerald RJ. Characterisation of the Hydrolytic Specificity of *Aspergillus Niger* Derived Prolyl Endoproteinase on Bovine β -Casein and Determination of ACE Inhibitory Activity. *Food Chemistry* 2014; 156 : 2936.
- Novick RP. Autoinduction and Signal Transduction in the Regulation of Staphylococcal Virulence. *Molecular Microbiology* 2003; 48 : 14291449.
- Ogoshi RCS, Zangeronimo MG, Reis JSD, França J, Santos JPF, Pires CP, et al. Acidifying and Yeast Extract in Diets for Adults Cats. *Animal Science Journal* 2014; 85 : 555561.
- Ogura M, Shimane K, Asai K, Ogasawara N, Tanaka T. Binding of Response Regulator DegU to the aprE Promoter Is Inhibited by RapG, Which Is Counteracted by Extracellular PhrG in *Bacillus Subtilis*. *Molecular Microbiology* 2003; 49 : 16851697.
- Okada M, Nakamura Y, Hayashi S, Ozaki K, Usami S. Chemical Structure and Biological Activity of a Quorum Sensing Pheromone from *Bacillus Subtilis* Subsp. *Natto*. *Bioorganic & Medicinal Chemistry Letters* 2015; 25 : 42934296.
- Okada M, Yamaguchi H, Sato I, Cho SJ, Dubnau D, Sakagami Y. Structure-Activity-Relationship Studies on Quorum Sensing ComXRO-E-2 Pheromone. *Bioorganic & medicinal chemistry letters* 2007; 17 : 17051707.
- Okamoto S, Suzuki T, Kawaguchi M, Higashiyama T, Matsubayashi Y. A Comprehensive Strategy for Identifying Long-Distance Mobile Peptides in Xylem Sap. *The Plant Journal : For Cell and Molecular Biology* 2015; 84 : 611620.
- Okamoto S, Tabata R, Matsubayashi Y. Long-Distance Peptide Signaling Essential for Nutrient Homeostasis in Plants. *Current Opinion in Plant Biology* 2016; 34 : 3540.
- O’Keeffe MB, FitzGerald RJ. Identification of Short Peptide Sequences in Complex Milk Protein Hydrolysates. *Food Chemistry* 2015; 184 : 140146.
- Olson ME, Todd DA, Schaeffer CR, Paharik AE, Van Dyke MJ, Büttner H, et al. *Staphylococcus Epidermidis* Agr Quorum-Sensing System : Signal Identification, Cross Talk, and Importance in Colonization. *Journal of Bacteriology* 2014; 196 : 34823493.
- Ortega MA, van der Donk WA. New Insights into the Biosynthetic Logic of Ribosomally Syn-

- thesized and Post-Translationally Modified Peptide Natural Products. *Cell Chemical Biology* 2016; 23 : 3144.
- Otto M, Süßmuth R, Jung G, Götz F. Structure of the Pheromone Peptide of the *Staphylococcus Epidermidis* Agr System. *FEBS letters* 1998; 424 : 8994.
- Pao SS, Paulsen IT, Saier MH. Major Facilitator Superfamily. *Microbiology and molecular biology reviews* : *MMBR* 1998; 62 : 134.
- Papadimitriou K, Alegría Á, Bron PA, de Angelis M, Gobbetti M, Kleerebezem M, et al. Stress Physiology of Lactic Acid Bacteria. *Microbiology and molecular biology reviews* : *MMBR* 2016; 80 : 837890.
- Papenfort K, Bassler BL. Quorum Sensing Signalresponse Systems in Gram-Negative Bacteria. *Nature Reviews Microbiology* 2016; 14 : 576588.
- Parker JL, Mindell JA, Newstead S. Thermodynamic Evidence for a Dual Transport Mechanism in a POT Peptide Transporter. *eLife* 2014; 3
- Pasupuleti VK, Braun S. State of the Art Manufacturing of Protein Hydrolysates. In : Pasupuleti VK, Demain AL, editor(s). *Protein Hydrolysates in Biotechnology*. Springer Netherlands; 2008. p. 1132.
- Pasupuleti VK, Holmes C, Demain AL. Applications of Protein Hydrolysates in Biotechnology. In : Pasupuleti VK, Demain AL, editor(s). *Protein Hydrolysates in Biotechnology*. Springer Netherlands; 2008. p. 19.
- Peltoniemi K, Vesanto E, Palva A. Genetic Characterization of an Oligopeptide Transport System from *Lactobacillus Delbrueckii* Subsp. *Bulgaricus*. *Archives of Microbiology* 2002; 177 : 457467.
- Perchat S, Dubois T, Zouhir S, Gominet M, Poncet S, Lemy C, et al. A Cell-Cell Communication System Regulates Protease Production during Sporulation in Bacteria of the *Bacillus Cereus* Group. *Molecular Microbiology* 2011; 82 : 619633.
- Perco P, Mühlberger I, Mayer G, Oberbauer R, Lukas A, Mayer B. Linking Transcriptomic and Proteomic Data on the Level of Protein Interaction Networks. *ELECTROPHORESIS* 2010; 31 : 17801789.
- Perego M. A Peptide Export-Import Control Circuit Modulating Bacterial Development Regulates Protein Phosphatases of the Phosphorelay. *Proceedings of the National Academy of Sciences of the United States of America* 1997; 94 : 86128617.
- Perego M, Brannigan JA. Pentapeptide Regulation of Aspartyl-Phosphate Phosphatases. *Peptides* 2001; 22 : 15411547.

- Perez-Pascual D, Monnet V, Gardan R. Bacterial Cell-Cell Communication in the Host via RRNPP Peptide-Binding Regulators. *Frontiers in Microbiology* 2016; 7 : 706.
- Perkins R, Fang H, Tong W, Welsh WJ. Quantitative Structure-Activity Relationship Methods : Perspectives on Drug Discovery and Toxicology. *Environmental Toxicology and Chemistry* 2003; 22 : 16661679.
- Petranovic D, Guédon E, Sperandio B, Delorme C, Ehrlich D, Renault P. Intracellular Effectors Regulating the Activity of the *Lactococcus Lactis* CodY Pleiotropic Transcription Regulator. *Molecular Microbiology* 2004; 53 : 613621.
- Picon A, Kunji ER, Lanfermeijer FC, Konings WN, Poolman B. Specificity Mutants of the Binding Protein of the Oligopeptide Transport System of *Lactococcus Lactis*. *Journal of Bacteriology* 2000; 182 : 16001608.
- Plackett RL, Burman JP. The Design of Optimum Multifactorial Experiments. *Biometrika* 1946; 33 : 305325.
- Pohl K, Francois P, Stenz L, Schlink F, Geiger T, Herbert S, et al. CodY in *Staphylococcus Aureus* : A Regulatory Link between Metabolism and Virulence Gene Expression. *Journal of Bacteriology* 2009; 191 : 29532963.
- Poquet I, Saint V, Seznec E, Simoes N, Bolotin A, Gruss A. HtrA Is the Unique Surface Housekeeping Protease in *Lactococcus Lactis* and Is Required for Natural Protein Processing. *Molecular Microbiology* 2000; 35 : 10421051.
- Potman RP, Wesdorp J. Method for the Preparation of a Yeast Extract. Patent US5288509A. 1994
- Pozzi G, Masala L, Iannelli F, Manganelli R, Havarstein LS, Piccoli L, et al. Competence for Genetic Transformation in Encapsulated Strains of *Streptococcus Pneumoniae* : Two Allelic Variants of the Peptide Pheromone. *Journal of Bacteriology* 1996; 178 : 60876090.
- Preis H, Eckart RA, Gudipati RK, Heidrich N, Brantl S. CodY Activates Transcription of a Small RNA in *Bacillus Subtilis*. *Journal of Bacteriology* 2009; 191 : 54465457.
- Proust L, Loux V, Martin V, Magnabosco C, Pedersen M, Monnet V, et al. Complete Genome Sequence of the Industrial Fast-Acidifying Strain *Streptococcus Thermophilus* N4L. *Microbiol Res Announc* 2018; 7 : e0102918.
- Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequences (RefSeq) : A Curated Non-Redundant Sequence Database of Genomes, Transcripts and Proteins. *Nucleic Acids Research* 2007; 35 : D61D65.
- Quistgaard EM, Löw C, Guettou F, Nordlund P. Understanding Transport by the Major Facili-

- tator Superfamily (MFS) : Structures Pave the Way. *Nature Reviews Molecular Cell Biology* 2016; 17 : 123132.
- Rajendran SRCK, Mason B, Udenigwe CC. Peptidomics of Peptic Digest of Selected Potato Tuber Proteins : Post-Translational Modifications and Limited Cleavage Specificity. *Journal of Agricultural and Food Chemistry* 2016; 64 : 24322437.
- Rees DC, Johnson E, Lewinson O. ABC Transporters : The Power to Change. *Nature Reviews. Molecular Cell Biology* 2009; 10 : 218227.
- Rey J, Deschavanne P, Tuffery P. BactPepDB : A Database of Predicted Peptides from a Exhaustive Survey of Complete Prokaryote Genomes. *Database : The Journal of Biological Databases and Curation* 2014; 2014
- Rice AJ, Park A, Pinkett HW. Diversity in ABC Transporters : Type I, II and III Importers. *Critical Reviews in Biochemistry and Molecular Biology* 2014; 49 : 426437.
- Rivas-Sendra A, Landete JM, Alcántara C, Zúñiga M. Response of *Lactobacillus Casei* BL23 to Phenolic Compounds. *Journal of Applied Microbiology* 2011; 111 : 14731481.
- Roces C, Courtin P, Kulakauskas S, Rodríguez A, Chapot-Chartier M-P, Martínez B. Isolation of *Lactococcus Lactis* Mutants Simultaneously Resistant to the Cell Wall-Active Bacteriocin Lcn972, Lysozyme, Nisin, and Bacteriophage C2. *Applied and Environmental Microbiology* 2012; 78 : 41574163.
- Rodionov DA, Hebbeln P, Eudes A, ter Beek J, Rodionova IA, Erkens GB, et al. A Novel Class of Modular Transporters for Vitamins in Prokaryotes. *Journal of Bacteriology* 2009; 191 : 4251.
- Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics : An R Package for 'omics Feature Selection and Multiple Data Integration. *PLoS computational biology* 2017; 13 : e1005752.
- Rosa LT, Bianconi ME, Thomas GH, Kelly DJ. Tripartite ATP-Independent Periplasmic (TRAP) Transporters and Tripartite Tricarboxylate Transporters (TTT) : From Uptake to Pathogenicity. *Frontiers in Cellular and Infection Microbiology* 2018; 8 : 33.
- Sakoh M, Ito K, Akiyama Y. Proteolytic Activity of HtpX, a Membrane-Bound and Stress-Controlled Protease from *Escherichia Coli*. *Journal of Biological Chemistry* 2005; 280 : 3330533310.
- Santos D, De Almeida DE. Isolation and Characterization of a New Temperature-Sensitive Cell Division Mutant of *Escherichia Coli* K-12. *Journal of Bacteriology* 1975; 124 : 15021507.
- Sanz Y, Lanfermeijer FC, Renault P, Bolotin A, Konings WN, Poolman B. Genetic and Func-

- tional Characterization of Dpp Genes Encoding a Dipeptide Transport System in *Lactococcus Lactis*. *Archives of Microbiology* 2001; 175 : 334343.
- Sanz Y, Toldrá F, Renault P, Poolman B. Specificity of the Second Binding Protein of the Peptide ABC-Transporter (Dpp) of *Lactococcus Lactis* IL1403. *FEMS microbiology letters* 2003; 227 : 3338.
- Savijoki K, Ingmer H, Varmanen P. Proteolytic Systems of Lactic Acid Bacteria. *Applied Microbiology and Biotechnology* 2006; 71 : 394406.
- Sánchez A, Vázquez A. Bioactive Peptides : A Review. *Food Quality and Safety* 2017; 1 : 2946.
- Scheepers GH, Lycklama A Nijeholt JA, Poolman B. An Updated Structural Classification of Substrate-Binding Proteins. *FEBS letters* 2016; 590 : 43934401.
- Schmitz S, Hoffmann A, Szekat C, Rudd B, Bierbaum G. The Lantibiotic Mersacidin Is an Autoinducing Peptide. *Applied and Environmental Microbiology* 2006; 72 : 72707277.
- Schramma KR, Bushin LB, Seyedsayamdost MR. Structure and Biosynthesis of a Macrocyclic Peptide Containing an Unprecedented Lysine-to-Tryptophan Crosslink. *Nature Chemistry* 2015; 7 : 431437.
- Schulz-Knappe P, Zucht HD, Heine C, Jurgens M, Hess R, Schrader M. Peptidomics : The Comprehensive Analysis of Peptides in Complex Biological Mixtures. *Combinatorial Chemistry & High Throughput Screening* 2001; 4 : 207217.
- Serror P, Sonenshein AL. CodY Is Required for Nutritional Repression of *Bacillus Subtilis* Genetic Competence. *Journal of Bacteriology* 1996; 178 : 59105915.
- Shelburne SA, Olsen RJ, Makthal N, Brown NG, Sahasrabhojane P, Watkins EM, et al. An Amino-Terminal Signal Peptide of Vfr Protein Negatively Influences RopB-Dependent SpeB Expression and Attenuates Virulence in *Streptococcus Pyogenes*. *Molecular Microbiology* 2011; 82 : 14811495.
- Shivers RP, Dineen SS, Sonenshein AL. Positive Regulation of *Bacillus Subtilis* ackA by CodY and CcpA : Establishing a Potential Hierarchy in Carbon Flow. *Molecular Microbiology* 2006; 62 : 811822.
- Shivers RP, Sonenshein AL. Activation of the *Bacillus Subtilis* Global Regulator CodY by Direct Interaction with Branched-Chain Amino Acids. *Molecular Microbiology* 2004; 53 : 599611.
- Singh A, Gautier B, Shannon CP, Rohart F, Vacher M, Tebutt SJ, et al. DIABLO : From Multi-Omics Assays to Biomarker Discovery, an Integrative Approach. *bioRxiv* 2018 : 067611.
- Singh K, Neelakantan S. Amino Acid Composition of Yeast Single Cell Protein Grown on Pa-

- neer Whey. *Journal of Dairy Research* 1989; 56 : 813815.
- Slack FJ, Mueller JP, Sonenshein AL. Mutations That Relieve Nutritional Repression of the *Bacillus Subtilis* Dipeptide Permease Operon. *Journal of Bacteriology* 1993; 175 : 46054614.
- Slack FJ, Serror P, Joyce E, Sonenshein AL. A Gene Required for Nutritional Repression of the *Bacillus Subtilis* Dipeptide Permease Operon. *Molecular Microbiology* 1995; 15 : 689702.
- Slamti L, Lereclus D. Specificity and Polymorphism of the PlcR-PapR Quorum-Sensing System in the *Bacillus Cereus* Group. *Journal of Bacteriology* 2005; 187 : 11821187.
- Sleigh SH, Seavers PR, Wilkinson AJ, Ladbury JE, Tame JR. Crystallographic and Calorimetric Analysis of Peptide Binding to OppA Protein. *Journal of Molecular Biology* 1999; 291 : 393415.
- Sleigh SH, Tame JRH, Dodson EJ, Wilkinson AJ. Peptide Binding in OppA, the Crystal Structures of the Periplasmic Oligopeptide Binding Protein in the Unliganded Form and in Complex with Lysyllysine, *Biochemistry* 1997; 36 : 97479758.
- Smeds A, Varmanen P, Palva A. Molecular Characterization of a Stress-Inducible Gene from *Lactobacillus Helveticus*. *Journal of Bacteriology* 1998; 180 : 61486153.
- Smith DE, Cléménçon B, Hediger MA. Proton-Coupled Oligopeptide Transporter Family SLC15 : Physiological, Pharmacological and Pathological Implications. *Molecular Aspects of Medicine* 2013; 34 : 323336.
- Solcan N, Kwok J, Fowler PW, Cameron AD, Drew D, Iwata S, et al. Alternating Access Mechanism in the POT Family of Oligopeptide Transporters. *The EMBO Journal* 2012; 31 : 34113421.
- Spencer RK, Nowick JS. A Newcomer's Guide to Peptide Crystallography. *Israel journal of chemistry* 2015; 55 : 698710.
- Spiess C, Beil A, Ehrmann M. A Temperature-Dependent Switch from Chaperone to Protease in a Widely Conserved Heat Shock Protein. *Cell* 1999; 97 : 339347.
- Steiner H-Y, Naider F, Becker JM. The PTR Family : A New Group of Peptide Transporters. *Molecular Microbiology* 1995; 16 : 825834.
- Storz G, Vogel J, Wassarman KM. Regulation by Small RNAs in Bacteria : Expanding Frontiers. *Molecular Cell* 2011; 43 : 880891.
- Sturme MHJ, Francke C, Siezen RJ, de Vos WM, Kleerebezem M. Making Sense of Quorum Sensing in Lactobacilli : A Special Focus on *Lactobacillus Plantarum* WCFS1. *Microbiology (Reading, England)* 2007; 153 : 39393947.
- Tame JR, Dodson EJ, Murshudov G, Higgins CE, Wilkinson AJ. The Crystal Structures of

- the Oligopeptide-Binding Protein OppA Complexed with Tripeptide and Tetrapeptide Ligands. *Structure* (London, England : 1993) 1995; 3 : 13951406.
- Tame JR, Murshudov GN, Dodson EJ, Neil TK, Dodson GG, Higgins CF, et al. The Structural Basis of Sequence-Independent Peptide Binding by OppA Protein. *Science* (New York, N.Y.) 1994; 264 : 15781581.
- Tame JR, Sleight SH, Wilkinson AJ, Ladbury JE. The Role of Water in Sequence-Independent Ligand Binding by an Oligopeptide Transporter Protein. *Nature Structural Biology* 1996; 3 : 9981001.
- Tashima AK, Zelanis A, Kitano ES, Ianzer D, Melo RL, Rioli V, et al. Peptidomics of Three Bothrops Snake Venoms : Insights into the Molecular Diversification of Proteomes and Peptidomes. *Molecular & cellular proteomics : MCP* 2012; 11 : 12451262.
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Research* 2016; 44 : 66146624.
- Terzaghi BE, Sandine WE. Improved Medium for Lactic Streptococci and Their Bacteriophages. *Applied Microbiology* 1975; 29 : 807813.
- Thévenet P, Rey J, Moroy G, Tuffery P. De Novo Peptide Structure Prediction : An Overview. *Methods in Molecular Biology* (Clifton, N.J.) 2015; 1268 : 113.
- Thoendel M, Kavanaugh JS, Flack CE, Horswill AR. Peptide Signaling in the Staphylococci. *Chemical Reviews* 2011; 111 : 117151.
- Tian C, Sun X, Liu X, Huang X, Chen F, Zheng S. Salivary Peptidome Profiling Analysis for Occurrence of New Carious Lesions in Patients with Severe Early Childhood Caries. *PloS One* 2017; 12 : e0182712.
- Tinoco AD, Saghatelian A. Investigating Endogenous Peptides and Peptidases Using Peptidomics. *Biochemistry* 2011; 50 : 74477461.
- Tortosa P, Logsdon L, Kraigher B, Itoh Y, Mandic-Mulec I, Dubnau D. Specificity and Genetic Polymorphism of the Bacillus Competence Quorum-Sensing System. *Journal of Bacteriology* 2001; 183 : 451460.
- Trengove NJ, Stacey MC, MacAuley S, Bennett N, Gibson J, Burslem F, et al. Analysis of the Acute and Chronic Wound Environments : The Role of Proteases and Their Inhibitors. *Wound Repair and Regeneration : Official Publication of the Wound Healing Society [and] the European Tissue Repair Society*; 7 : 442452.
- Tsuchiya T, Osaki T, Minamino N, Sasaki K. Peptidomics for Studying Limited Proteolysis. *Journal of Proteome Research* 2015; 14 : 49214931.

- Tynkkynen S, Buist G, Kunji E, Kok J, Poolman B, Venema G, et al. Genetic and Biochemical Characterization of the Oligopeptide Transport System of *Lactococcus Lactis*. *Journal of Bacteriology* 1993; 175 : 75237532.
- Ulrich LE, Koonin EV, Zhulin IB. One-Component Systems Dominate Signal Transduction in Prokaryotes. *Trends in Microbiology* 2005; 13 : 5256.
- Ummadi M, Curic-Bawden M. Use of Protein Hydrolysates in Industrial Starter Culture Fermentations. In : Pasupuleti VK, Demain AL, editor(s). *Protein Hydrolysates in Biotechnology*. Springer Netherlands; 2008. p. 91114.
- Valot B, Langella O, Nano E, Zivy M. MassChroQ : A Versatile Tool for Mass Spectrometry Quantification. *Proteomics* 2011; 11 : 35723577.
- van de Guchte M, Serror P, Chervaux C, Smokvina T, Ehrlich SD, Maguin E. Stress Responses in Lactic Acid Bacteria. *Antonie Van Leeuwenhoek* 2002; 82 : 187216.
- van der Heide T, Poolman B. ABC Transporters : One, Two or Four Extracytoplasmic Substrate-Binding Sites? *EMBO reports* 2002; 3 : 938943.
- van der Meulen SB, de Jong A, Kok J. Transcriptome Landscape of *Lactococcus Lactis* Reveals Many Novel RNAs Including a Small Regulatory RNA Involved in Carbon Uptake and Metabolism. *RNA Biology* 2016; 13 : 353366.
- Varet H, Brillet-Guéguen L, Coppée J-Y, Dillies M-A. SARTools : A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLOS ONE* 2016; 11 : e0157022.
- Venema R, Tjalsma H, van Dijl JM, de Jong A, Leenhouts K, Buist G, et al. Active Lipoprotein Precursors in the Gram-Positive Eubacterium *Lactococcus Lactis*. *The Journal of Biological Chemistry* 2003; 278 : 1473914746.
- Verdon J, Girardin N, Lacombe C, Berjeaud J-M, Héchard Y. Delta-Hemolysin, an Update on a Membrane-Interacting Peptide. *Peptides* 2009; 30 : 817823.
- Visick JE, Clarke S. Repair, Refold, Recycle : How Bacteria Can Deal with Spontaneous and Environmental Damage to Proteins. *Molecular Microbiology* 1995; 16 : 835845.
- Visser S, Slangen CJ, Robben AJPM. Determination of Molecular Mass Distributions of Whey Protein Hydrolysates by High-Performance Size-Exclusion Chromatography. *Journal of Chromatography A* 1992; 599 : 205209.
- Wade JT, Grainger DC. Pervasive Transcription : Illuminating the Dark Matter of Bacterial Transcriptomes. *Nature Reviews Microbiology* 2014; 12 : 647653.
- Waglay A, Karboune S. Enzymatic Generation of Peptides from Potato Proteins by Selected

- Proteases and Characterization of Their Structural Properties. *Biotechnology Progress* 2016; 32 : 420429.
- Walsh CT, O'Brien RV, Khosla C. Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. *Angewandte Chemie (International ed. in English)* 2013; 52 : 70987124.
- Wang G, Li X, Wang Z. APD2 : The Updated Antimicrobial Peptide Database and Its Application in Peptide Design. *Nucleic Acids Research* 2009; 37 : D933D937.
- Wang H, Fewer DP, Holm L, Rouhiainen L, Sivonen K. Atlas of Nonribosomal Peptide and Polyketide Biosynthetic Pathways Reveals Common Occurrence of Nonmodular Enzymes. *Proceedings of the National Academy of Sciences of the United States of America* 2014; 111 : 92599264.
- Warren AS, Archuleta J, Feng W-c, Setubal JC. Missing Genes in the Annotation of Prokaryotic Genomes. *BMC Bioinformatics* 2010; 11 : 131.
- Waters CM, Bassler BL. Quorum Sensing : Cell-to-Cell Communication in Bacteria. *Annual Review of Cell and Developmental Biology* 2005; 21 : 319346.
- Waters LS, Storz G. Regulatory RNAs in Bacteria. *Cell* 2009; 136 : 615628.
- Watson TG. Amino-Acid Pool Composition of *Saccharomyces Cerevisiae* as a Function of Growth Rate and Amino-Acid Nitrogen Source. *Journal of General Microbiology* 1976; 96 : 263268.
- Wescombe PA, Tagg JR. Purification and Characterization of Streptin, a Type A1 Lantibiotic Produced by *Streptococcus Pyogenes*. *Applied and Environmental Microbiology* 2003; 69 : 27372747.
- Wilcoxon F. Individual Comparisons by Ranking Methods. *Biometrics Bulletin* 1945; 1 : 8083.
- Wilkening RV, Chang JC, Federle MJ. PepO, a CovRS-Controlled Endopeptidase, Disrupts *Streptococcus Pyogenes* Quorum Sensing. *Molecular Microbiology* 2016; 99 : 7187.
- Williams WL, Broquist HP, Snell EE. Oleic acid and related compounds as growth factors for lactic acid bacteria. *Journal of Biological Chemistry* 1947; 170 : 619630.
- Wolters DA, Washburn MP, Yates JR. An Automated Multidimensional Protein Identification Technology for Shotgun Proteomics. *Analytical Chemistry* 2001; 73 : 56835690.
- Wu C, Cichewicz R, Li Y, Liu J, Roe B, Ferretti J, et al. Genomic Island TnSmu2 of *Streptococcus Mutans* Harbors a Nonribosomal Peptide Synthetase-Polyketide Synthase Gene Cluster Responsible for the Biosynthesis of Pigments Involved in Oxygen and H₂O₂ Tolerance. *Applied and Environmental Microbiology* 2010; 76 : 58155826.

- Wuster A, Babu MM. Conservation and Evolutionary Dynamics of the Agr Cell-to-Cell Communication System across Firmicutes. *Journal of Bacteriology* 2008; 190 : 743746.
- Wynendaele E, Bronselaer A, Nielandt J, D'Hondt M, Stalmans S, Bracke N, et al. Quorum-peps Database : Chemical Space, Microbial Origin and Functionality of Quorum Sensing Peptides. *Nucleic Acids Research* 2013; 41 : D655D659.
- Yan N. Structural Biology of the Major Facilitator Superfamily Transporters. *Annual Review of Biophysics* 2015; 44 : 257283.
- Yates JR. Mass Spectral Analysis in Proteomics. *Annual Review of Biophysics and Biomolecular Structure* 2004; 33 : 297316.
- Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler : An R Package for Comparing Biological Themes among Gene Clusters. *Omics : A Journal of Integrative Biology* 2012; 16 : 284287.
- Zetzmann M, Sánchez-Kopper A, Waidmann MS, Blombach B, Riedel CU. Identification of the Agr Peptide of *Listeria Monocytogenes*. *Frontiers in Microbiology* 2016; 7 : 989.
- Zhang J, Zhou S, Li R, Cao T, Zheng H, Wang X, et al. Magnetic Bead-Based Salivary Peptidome Profiling for Periodontal-Orthodontic Treatment. *Proteome Science* 2012; 10 : 63.
- Zhang Y, Fonslow BR, Shan B, Baik M-C, Yates JR. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews* 2013; 113 : 23432394.
- Zhang Z, Dahlsten E, Korkeala H, Lindström M. Positive Regulation of Botulinum Neurotoxin Gene Expression by CodY in *Clostridium Botulinum* ATCC 3502. *Applied and Environmental Microbiology* 2014; 80 : 76517658.
- Zhao Y, Mao G, Liu M, Zhang L, Wang X, Zhang XC. Crystal Structure of the E. Coli Peptide Transporter YbgH. *Structure* 2014; 22 : 11521160.
- Zhou M, Boekhorst J, Francke C, Siezen RJ. LocateP : Genome-Scale Subcellular-Location Predictor for Bacterial Proteins. *BMC bioinformatics* 2008; 9 : 173.
- Zinchenko DV, Muranova TA, Melanyina LA, Belova NA, Miroshnikov AI. Soy and Rapeseed Protein Hydrolysis by the Enzyme Preparation Protosubtilin. *Applied Biochemistry and Microbiology* 2018; 54 : 294300.

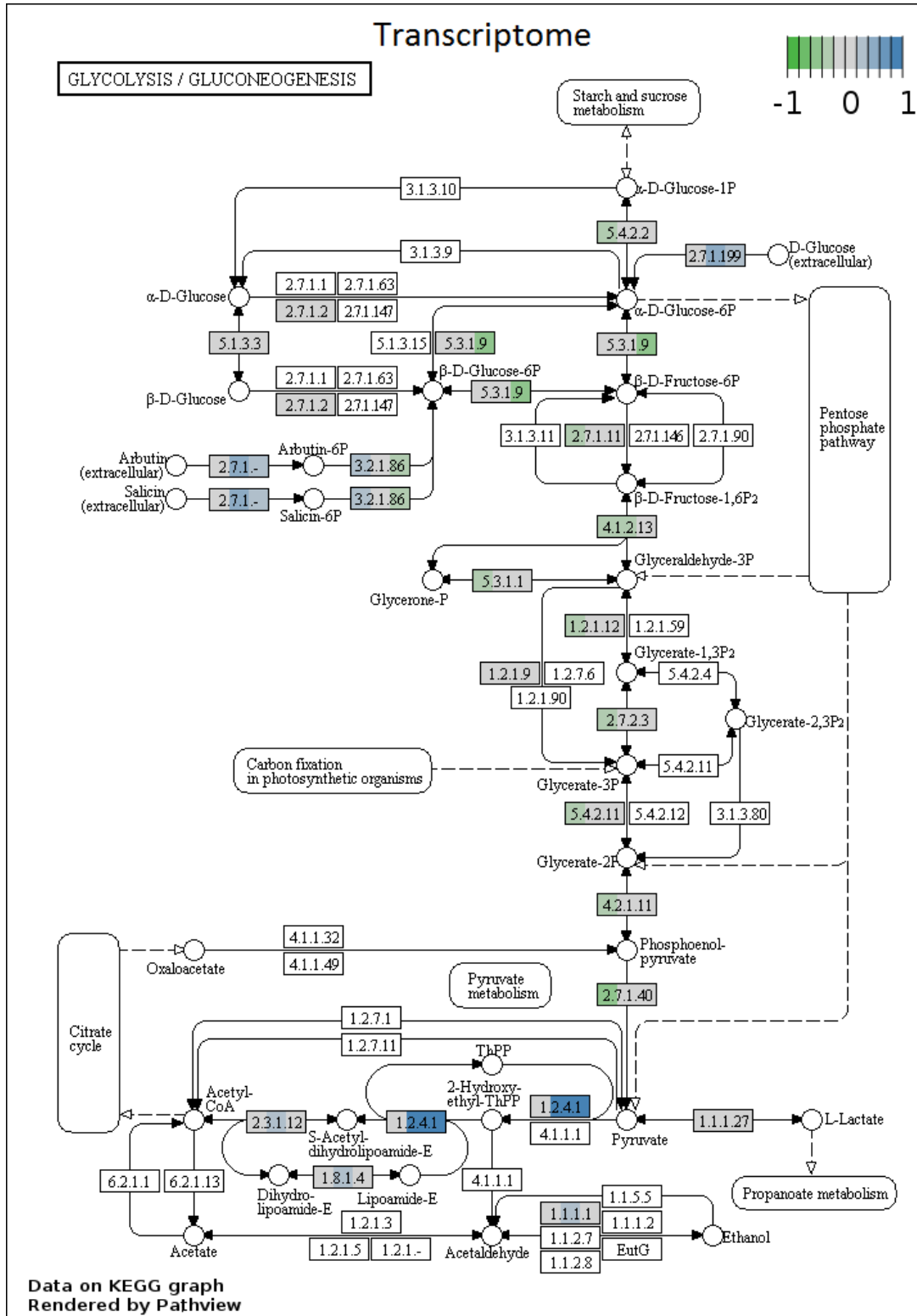
Annexes

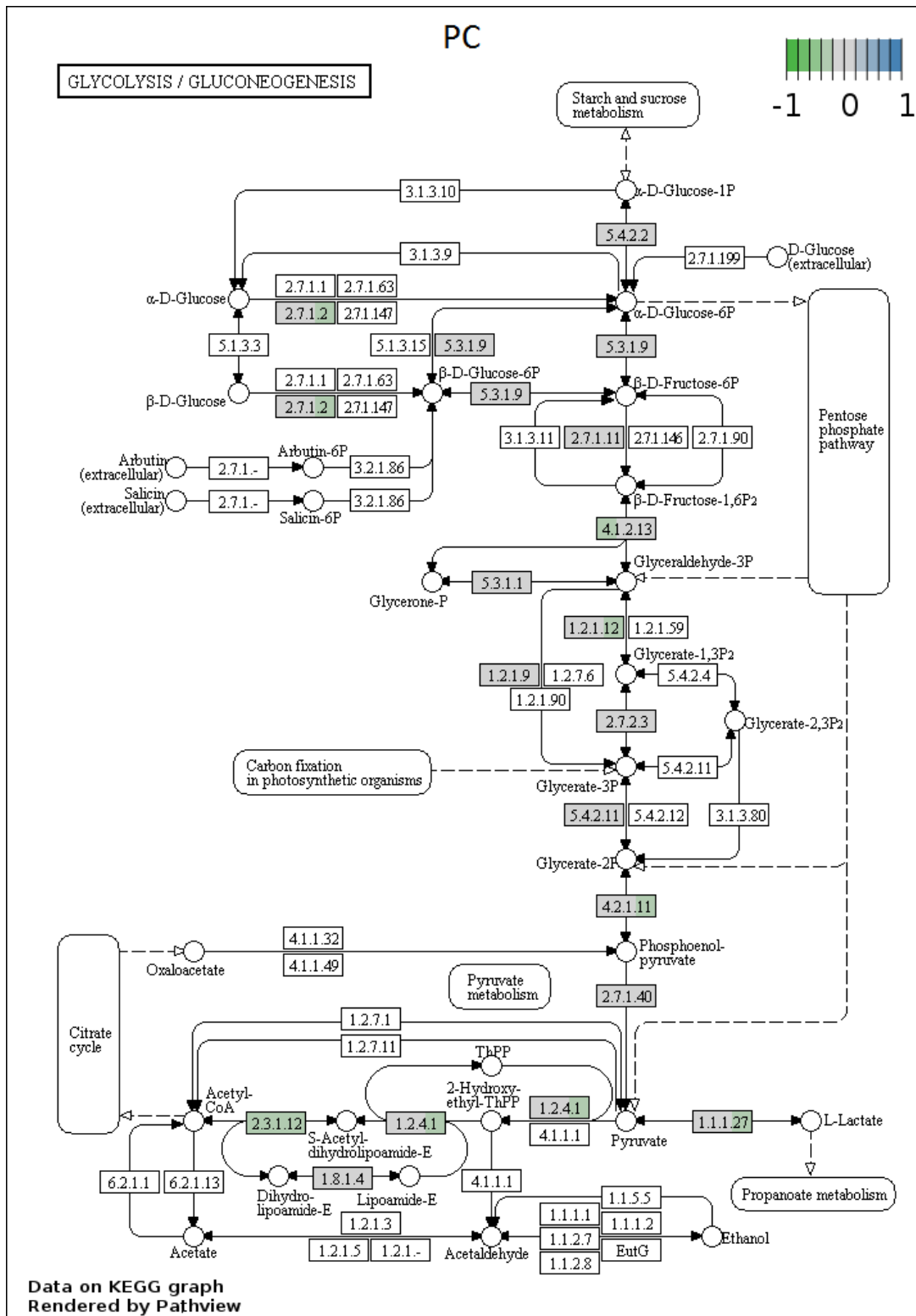
Sommaire

Annexe A - Impacts différentiels des extraits de levure sur les fonctions et voies métaboliques de <i>S. thermophilus</i> N4L	250
Annexe A1 - Glycolyse	250
Annexe A2 - Cycle de Krebs	252
Annexe A3 - Biosynthèse de l'arginine	254
Annexe A4 - Métabolisme de l'alanine, aspartate et glutamate	256
Annexe A5 - Biosynthèse de la lysine	258
Annexe A6 - Métabolisme de l'histidine	260
Annexe A7 - Biosynthèse des acides aminés aromatiques	262
Annexe A8 - Métabolisme des purines	264
Annexe A9 - Métabolisme des pyrimidines	266
Annexe A10 - Protéines ribosomiques	268
Annexe B - Résultats complémentaires de l'intégration multiomique <i>via</i> l'analyse DIABLO	270
Annexe B1 - Contribution des gènes et protéines sélectionnés à l'issue de l'analyse DIABLO	270
Annexe B2 - Liste des gènes et protéines sélectionnés	271
Annexe C - Valorisation du travail de thèse	273
Annexe D - Formations	274

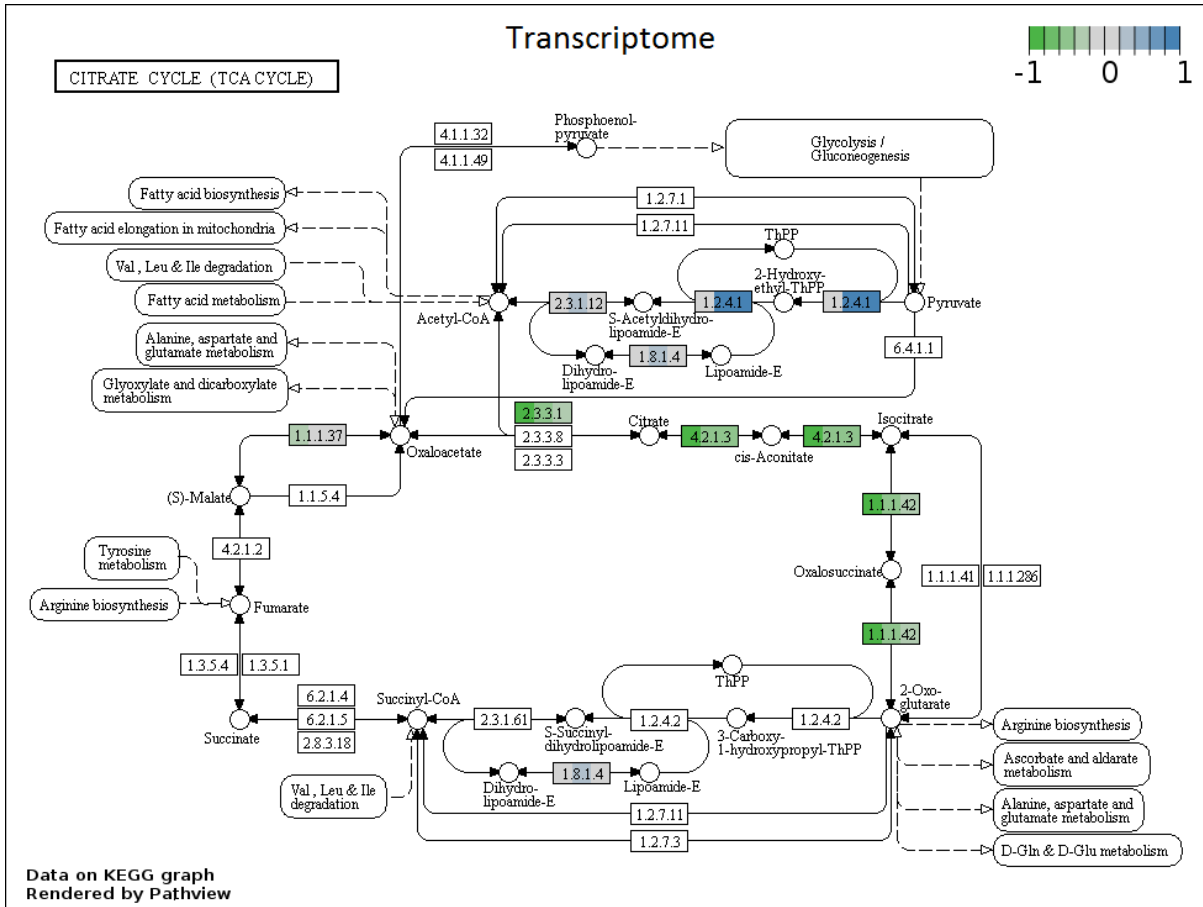
Annexe A - Impacts différentiels des extraits de levure sur les fonctions et voies métaboliques de *S. thermophilus* N4L

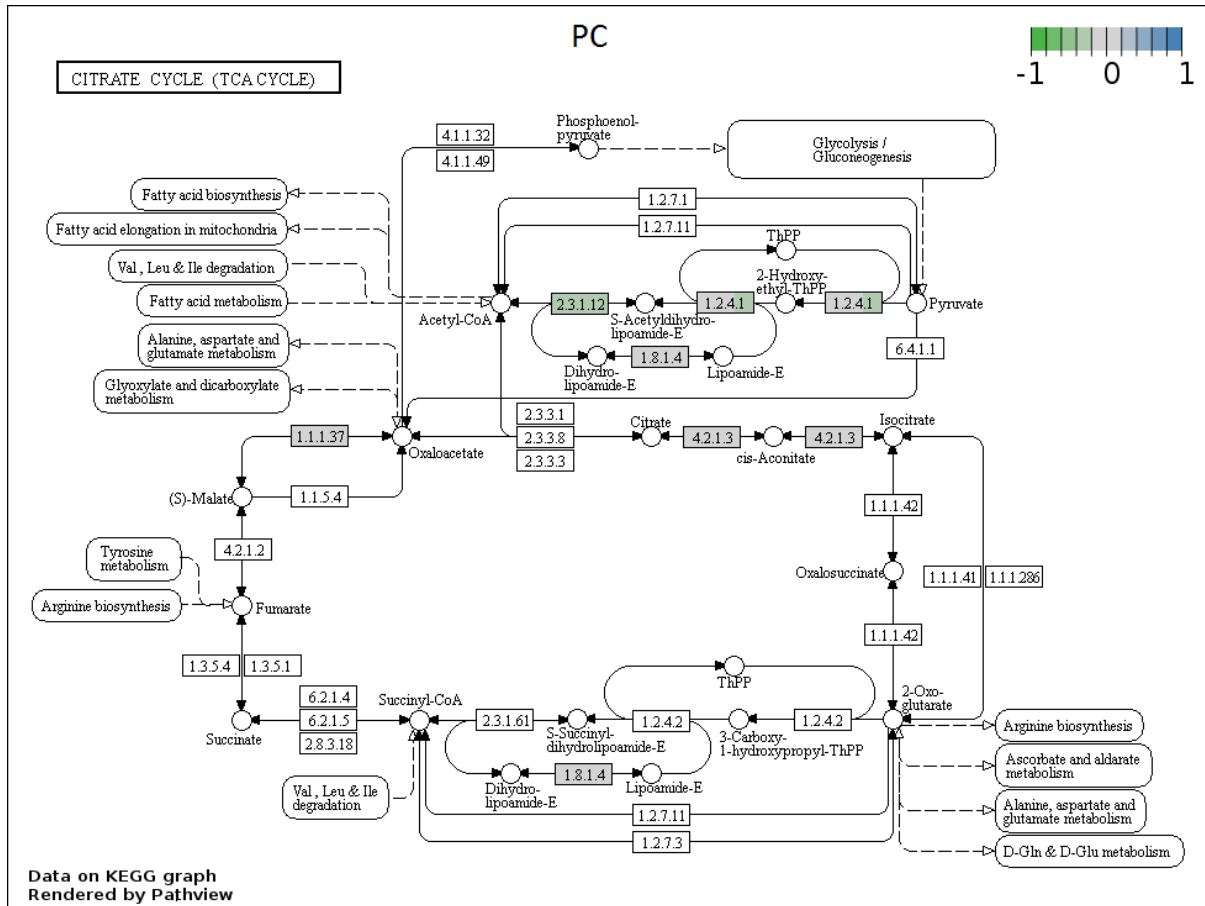
Annexe A1 - Glycolyse



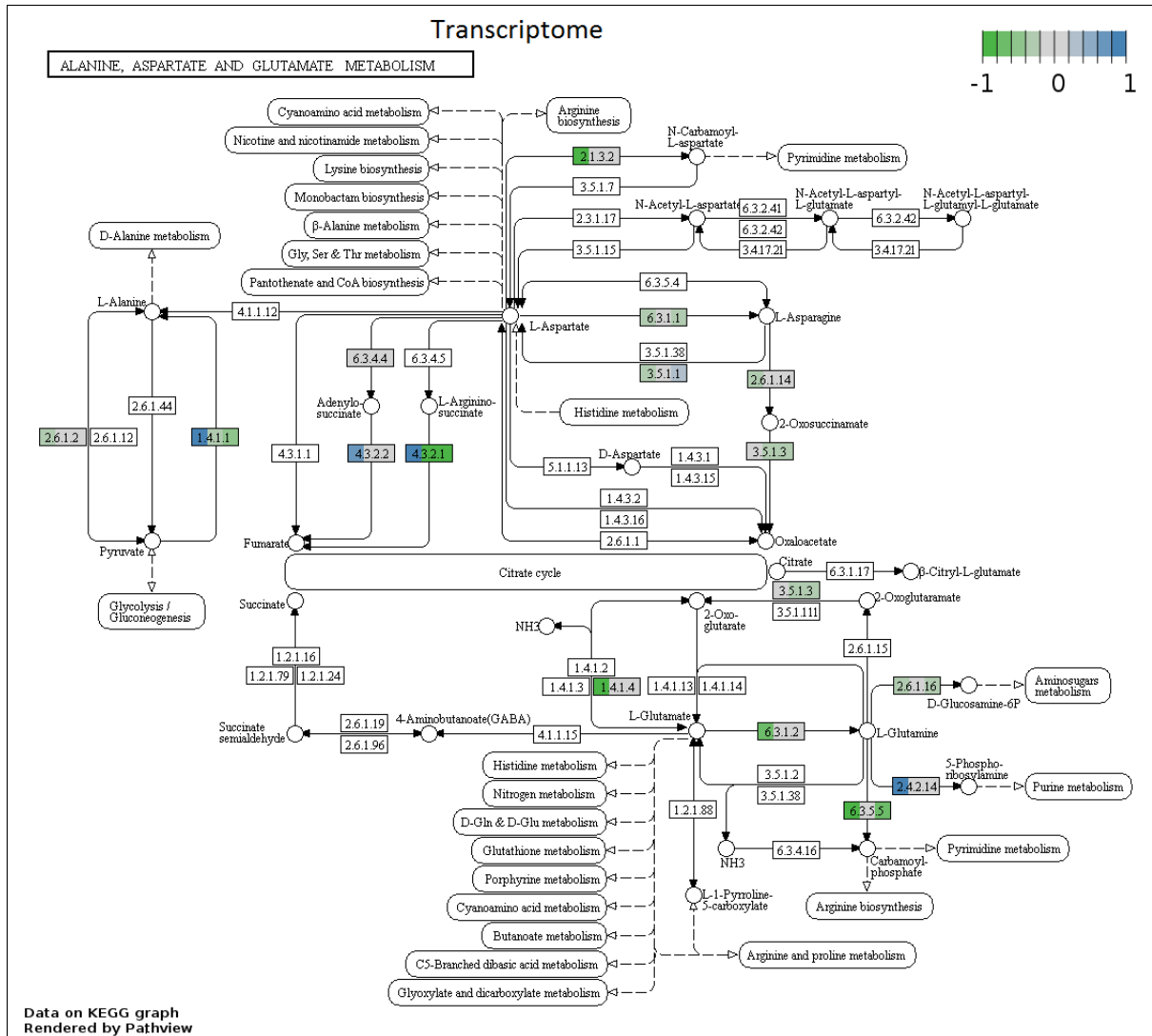


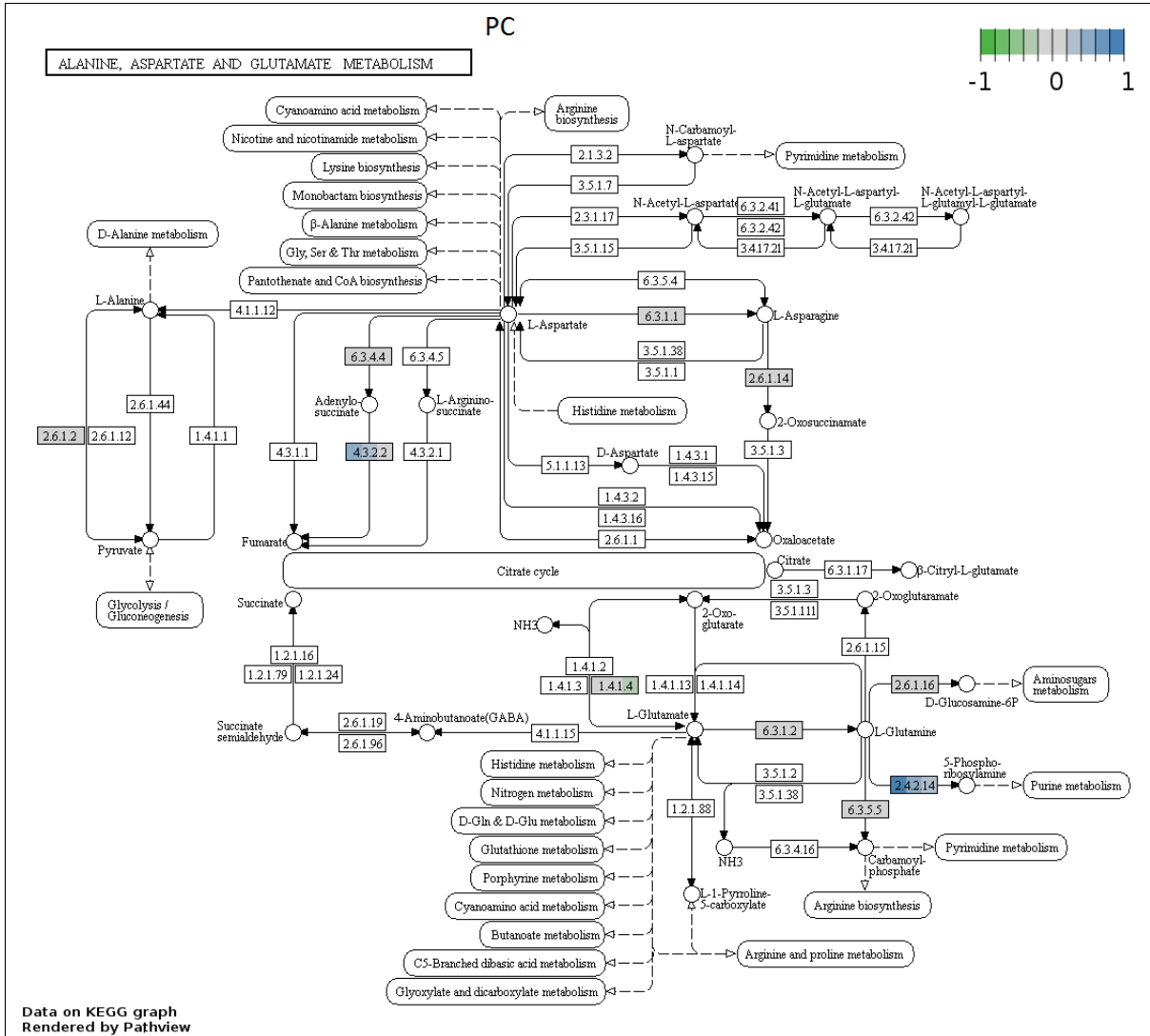
Annexe A2 - Cycle de Krebs



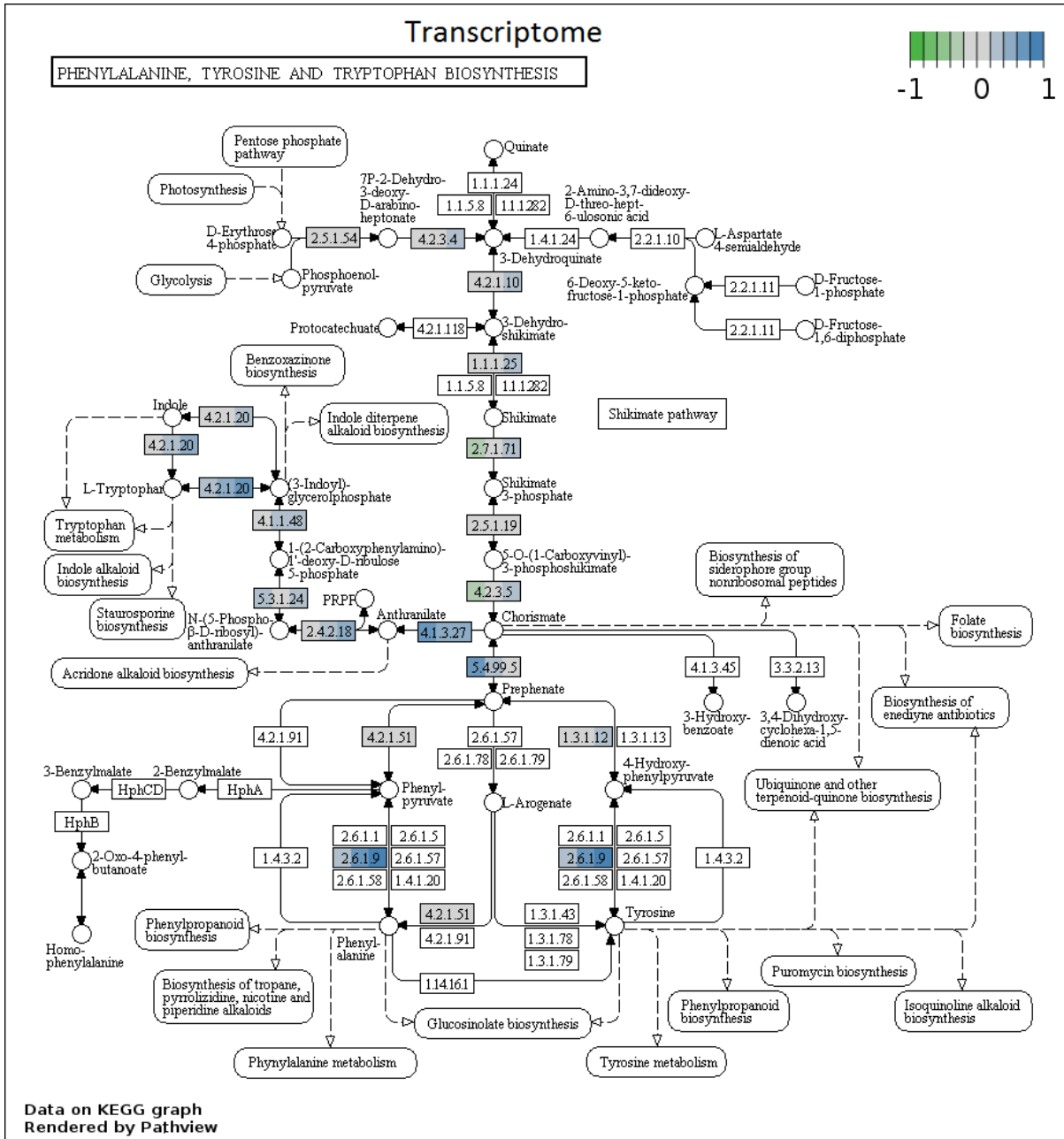


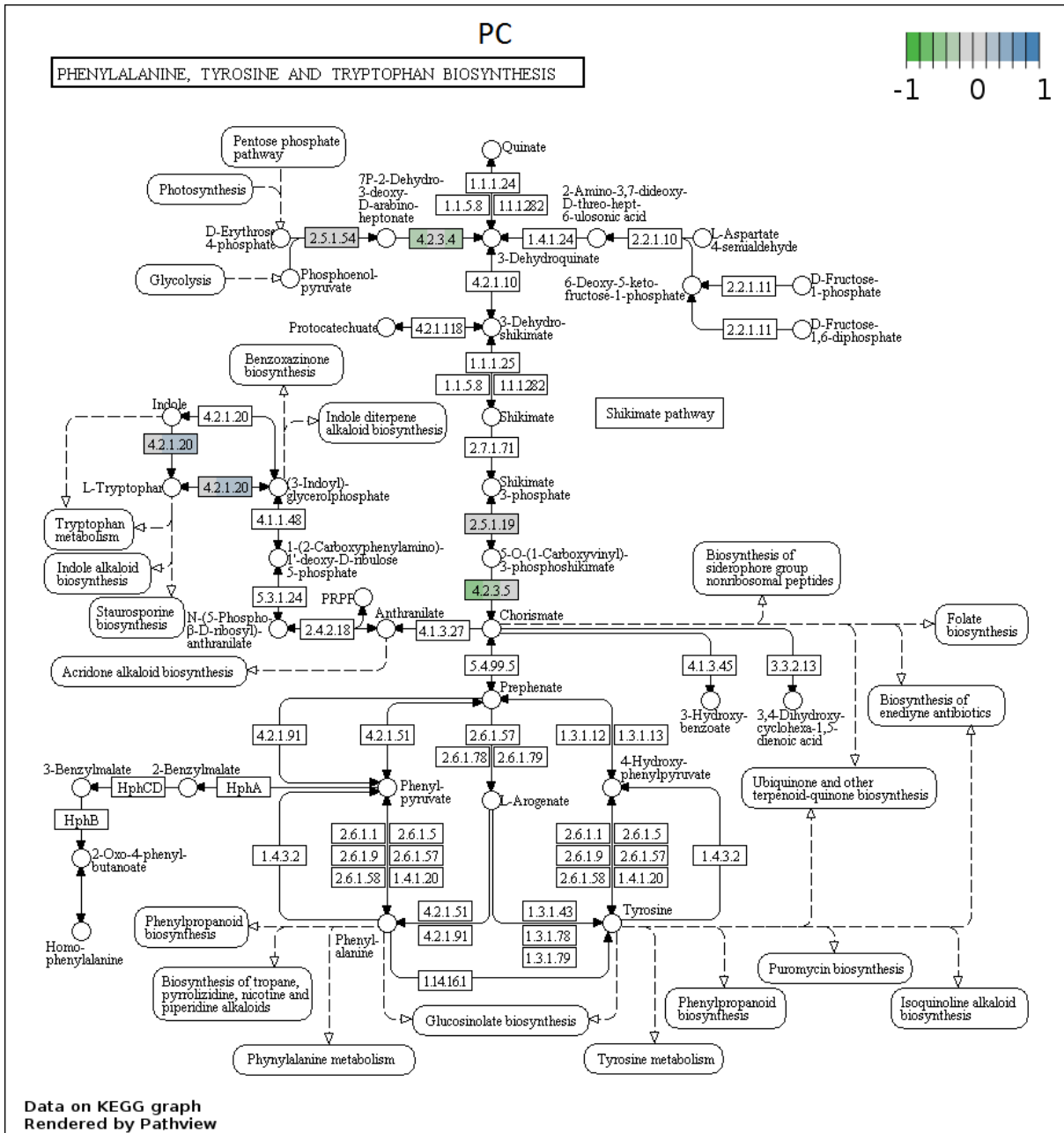
Annexe A4 - Métabolisme de l'alanine, aspartate et glutamate



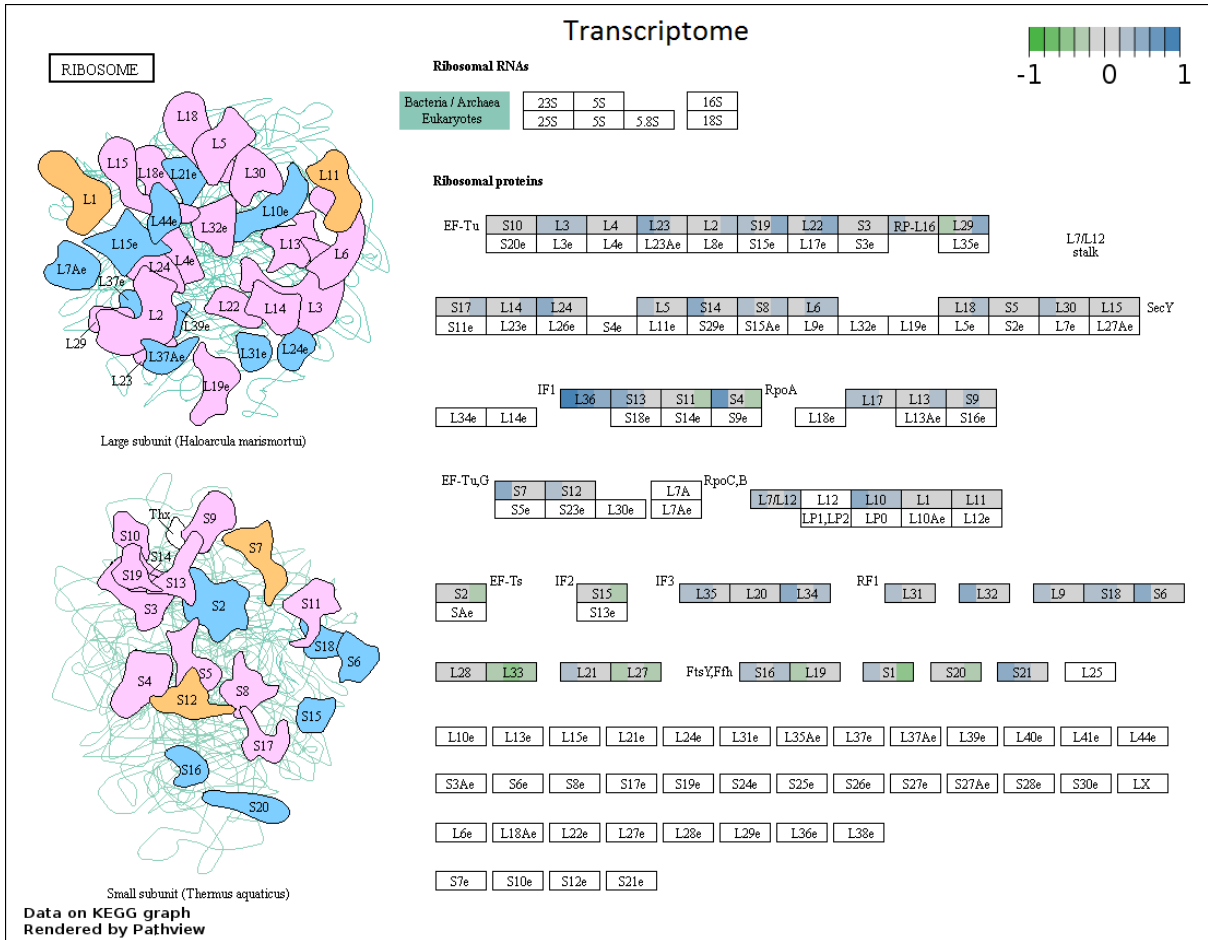


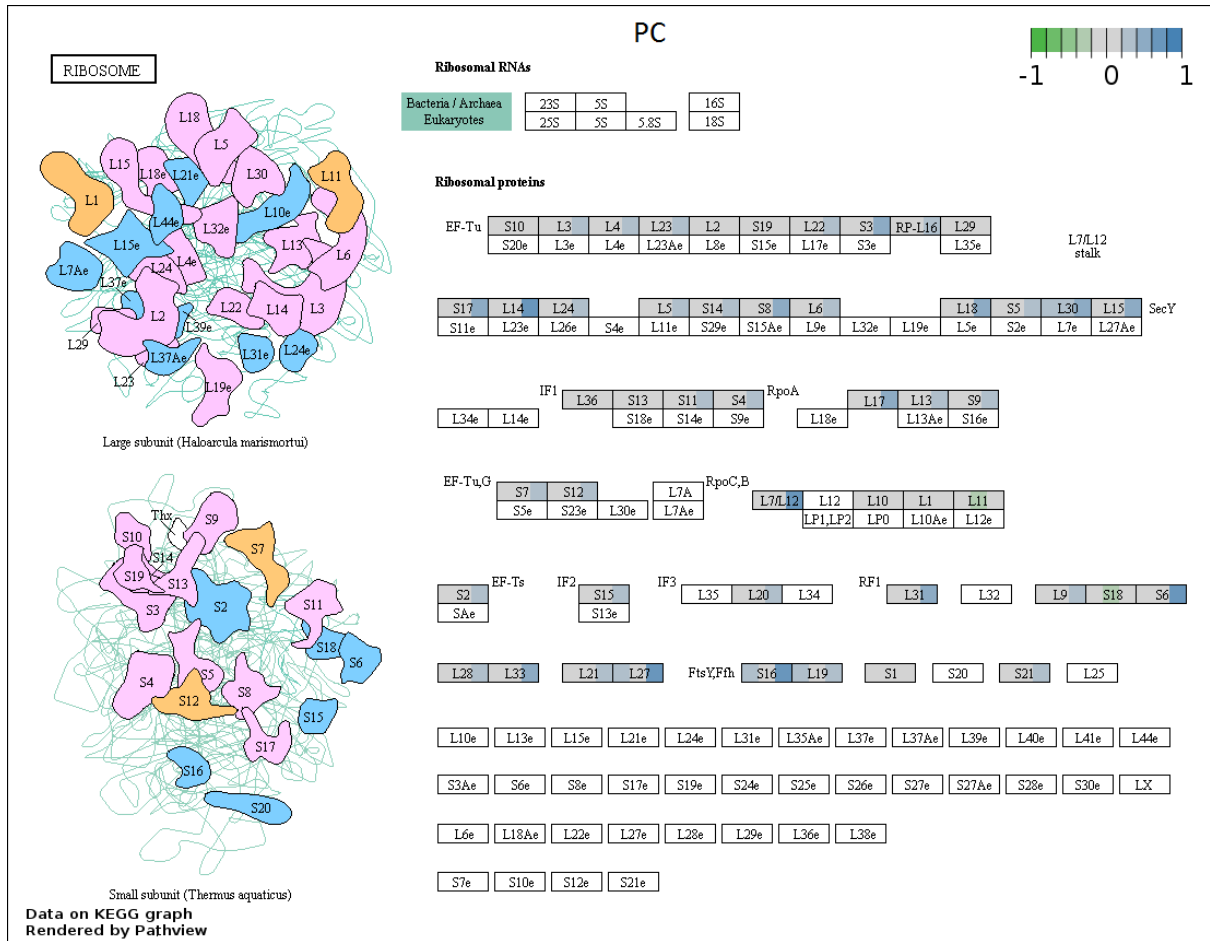
Annexe A7 - Biosynthèse des acides aminés aromatiques





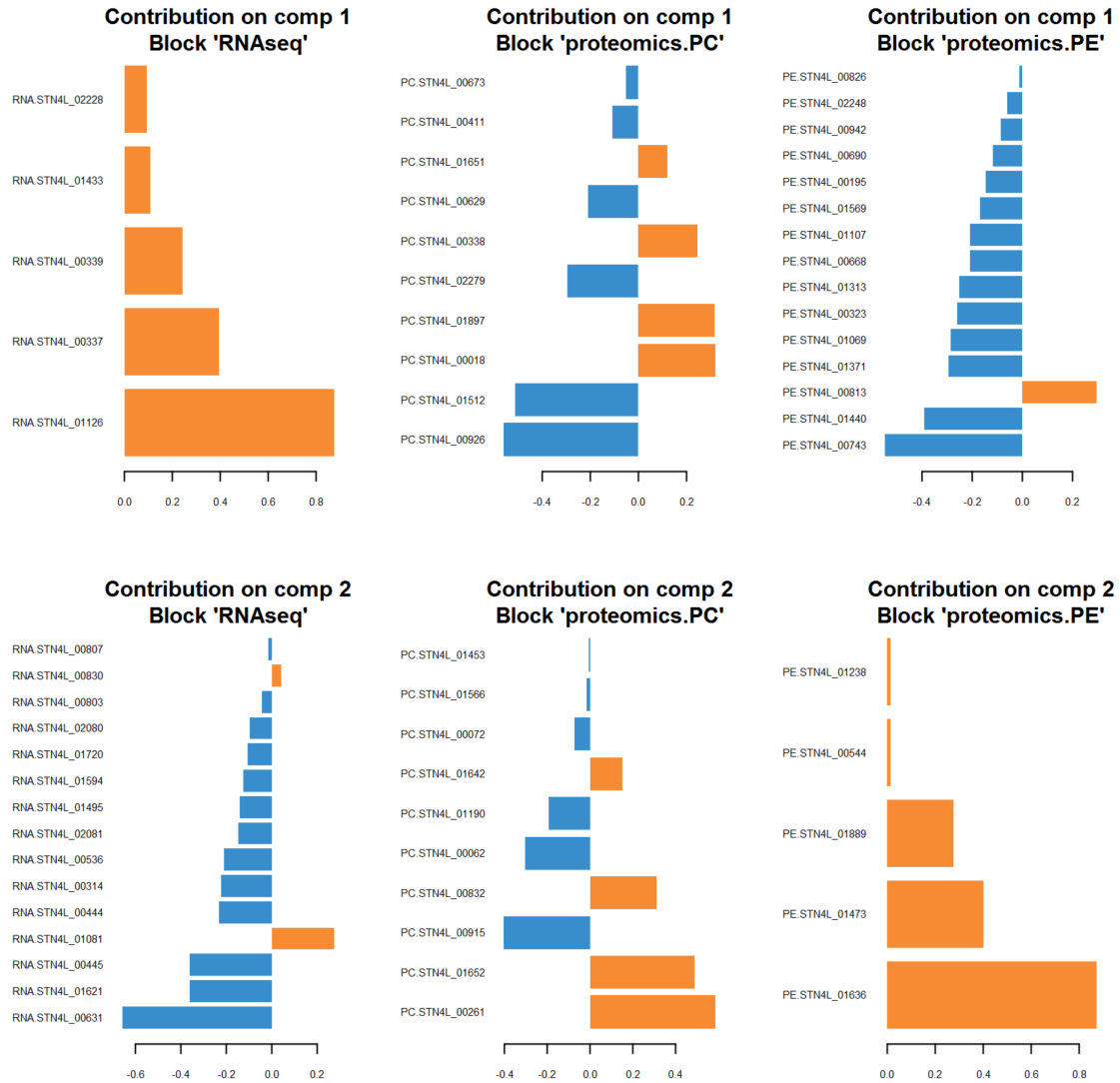
Annexe A10 - Protéines ribosomiques





Annexe B - Résultats complémentaires de l'intégration multi-omique *via* l'analyse DIABLO

Annexe B1 - Contribution des gènes et protéines sélectionnés à l'issue de l'analyse DIABLO



Annexe B2 - Liste des gènes et protéines sélectionnés

Composante	Origine	Identifiant	Gène	Produit	Contribution	
Composante 1	ARN	STN4L_01126	STN4L_01126	Uncharacterized protein	EXL2	
		STN4L_00337	hisG	ATP phosphoribosyltransferase	EXL2	
		STN4L_00339	hisB	Imidazoleglycerol-phosphate dehydratase	EXL2	
		STN4L_01433	STN4L_01433	ABC transporter permease	EXL2	
		STN4L_02228	trpE	Anthranilate synthase component I	EXL2	
	PC	STN4L_00926	aroB	3-dehydroquinate synthase	EXL1	
		STN4L_01512	cfa	Cyclopropane-fatty-acyl-phospholipid synthase	EXL1	
		STN4L_00018	serC	Phosphoserine aminotransferase	EXL2	
		STN4L_01897	ilvC	Ketol-acid reductoisomerase	EXL2	
		STN4L_02279	murZ	UDP-N-acetylglucosamine 1- carboxyvinyltransferase	EXL1	
		STN4L_00338	hisD	Histidinol dehydrogenase	EXL2	
		STN4L_00629	adhB	Alcohol dehydrogenase, zinc-containing	EXL1	
		STN4L_01651	purL	Phosphoribosylformylglycinamide synthase II (FGAM synthetase)	EXL2	
	STN4L_00411	metK	S-adenosylmethionine synthase	EXL1		
	STN4L_00673	gntR	Transcriptional regulator	EXL1		
	PE	STN4L_00743	aspC1	Aspartate aminotransferase	EXL1	
		STN4L_01440	yebC	Probable transcriptional regulatory protein yebC	EXL1	
		STN4L_00813	STN4L_00813	A0A249DLE6 (Uncharacterized protein)	EXL2	
		STN4L_01371	rdgB	Non-canonical purine NTP pyrophosphatase	EXL1	
		STN4L_01069	rluD	Pseudouridine synthase	EXL1	
		STN4L_00323	pho2	4-nitrophenylphosphatase (PNPPase)	EXL1	
		STN4L_01313	gidB	Ribosomal RNA small subunit methyltransferase G	EXL1	
		STN4L_00668	umuD	UV resistance protein UmuD	EXL1	
		STN4L_01107	tuf	Elongation factor Tu	EXL1	
		STN4L_01569	degV	Fatty acid-binding protein, DegV family	EXL1	
STN4L_00195		nrd	Nitroreductase	EXL1		
STN4L_00690		glmS	Glutamine-fructose-6-phosphate aminotransferase [isomerizing]	EXL1		
STN4L_00942		pplB	Peptidyl-prolyl cis-trans isomerase	EXL1		
STN4L_02248		murF	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D- alanine ligase	EXL1		
STN4L_00826		murD	UDP-N-acetylmuramoylalanine-D- glutamate ligase	EXL1		
Composante 2	ARN	STN4L_00631	czcD	Cation efflux system protein	EXL1	
		STN4L_01621	spxA	Arsenate reductase	EXL1	
		STN4L_00445	STN4L_00445	Uncharacterized protein	EXL1	
		STN4L_01081	STN4L_01081	Uncharacterized protein	EXL2	
		STN4L_00444	STN4L_00444	Membrane protein	EXL1	
		STN4L_00314	rmlC	DTDP-4-keto-L-rhamnose reductase	EXL1	
		STN4L_00536	STN4L_00536	Lipoprotein	EXL1	
		STN4L_02081	rexB	ATP-dependent helicase/deoxyribonuclease subunit B	EXL1	
		STN4L_01495	dtd	D-aminoacyl-tRNA deacylase	EXL1	
		STN4L_01594	STN4L_01594	A0A0P6VAY6 (Muramoyltetrapeptide carboxypeptidase)	EXL1	
		STN4L_01720	guaB	IMP dehydrogenase	EXL1	
		STN4L_02080	STN4L_02080	Putative uncharacterized conserved secreted protein	EXL1	
		STN4L_00803	glyA	Serine hydroxymethyltransferase	EXL1	
		STN4L_00830	ORF 2	Hypothetical 8.3 kDa protein	EXL2	
		STN4L_00807	tdk	Thymidine kinase	EXL1	
		PC	STN4L_00261	rncS	Ribonuclease 3	EXL2
			STN4L_01652	purC	Phosphoribosylaminoimidazole-succinocarboxamide synthase	EXL2
STN4L_00915	mip		Alkyl hydroperoxide reductase AhpD	EXL1		
STN4L_00832	dpr		Peroxide resistance protein,non-heme iron- containing ferritin	EXL2		
STN4L_00062	rpoD		RNA polymerase sigma factor SigA	EXL1		
STN4L_01190	thiI		Probable tRNA sulfurtransferase	EXL1		
STN4L_01642	purD		Phosphoribosylamine-glycine ligase	EXL2		
STN4L_00072	rgpD		Polysaccharide ABC exporter ATP-binding protein	EXL1		
STN4L_01566	panE		2-dehydropantoate 2-reductase	EXL1		
STN4L_01453	hslO		33 kDa chaperonin	EXL1		

Annexes

	STN4L_01636	purB1	Adenylosuccinate lyase	EXL2
	STN4L_01473	sufD	SufD	EXL2
PE	STN4L_01889	thrC	Threonine synthase	EXL2
	STN4L_00544	malQ	4-alpha-glucanotransferase	EXL2
	STN4L_01238	ppaC	Probable manganese-dependent inorganic pyrophosphatase	EXL2

Annexe C - Valorisation du travail de thèse

Publications scientifiques

Proust L., Loux V., Martin V., Magnabosco C., Pedersen M., Monnet V., Juillard V., 2018. Complete Genome Sequence of the Industrial Fast-Acidifying Strain *Streptococcus thermophilus* N4L. *Microbiol Res Announc* 7 :e01029-18

Proust L., Sourabié A., Pedersen M., Besançon I., Monnet V., Juillard V., 2018. Insights into the complexity of yeast extract peptides and their utilization by *Streptococcus thermophilus*. (Soumis à *Applied and Environmental Microbiology* en Octobre 2018)

Posters

Contribution of peptidomics to the study of *Streptococcus thermophilus* peptide utilization in complex media. Proust L., Liuu S., Haudebourg E., Sourabié A., Pedersen M., Besançon I., Monnet V., Juillard V. *21^{ème} colloque du Club des Bactéries Lactiques*. Lille. June 2017.

Insights into the nutritional and regulatory role of yeast extract components on *Streptococcus thermophilus*. Proust L., Haudebourg E., Sourabié A., Pedersen M., Besançon I., Monnet V., Juillard V. *1st International Conference on Microbial Food and Feed Ingredients*. Copenhagen, Denmark. May 2018.

Communications orales

Contribution of peptidomics to the study of *Streptococcus thermophilus* peptide utilization in complex media. Proust L., Liuu S., Haudebourg E., Sourabié A., Pedersen M., Besançon I., Monnet V., Juillard V. *Ecole de printemps du Club Jeune de la Société Française de Spectrométrie de Masse*. Trélon. March 2017.

Contribution of peptidomics to the study of *Streptococcus thermophilus* peptide utilization in complex media. Proust L., Liuu S., Haudebourg E., Sourabié A., Pedersen M., Besançon I., Monnet V., Juillard V. *Journée des doctorants de l'Institut Micalis (Doc Micalis)*. Jouy-en-Josas. June 2017.

Médiation scientifique

Organisation et participation aux Journées Portes Ouvertes pour les 70 ans de l'INRA de Jouy-en-Josas. Octobre 2016.

Annexe D - Formations

Formation Ecole Doctorale ABIES

Presentation skills in English - getting your message across. Mars 2016, Paris, 14 h.

Formations INRA

Initiation à Galaxy / NGS sous Galaxy. Mars 2016, Jouy-en-Josas. 14 h.

Annotation de génomes. Juin 2016, Jouy-en-Josas. 21 h.

RNAseq : bioinformatique sous Galaxy et statistiques. Mars 2017, Jouy-en-Josas. 21 h.

Analyse des données de protéomique à l'aide des outils bioinformatiques développés par la plateforme PAPPSO. Novembre 2016, Gif-sur-Yvette. 20 h.

Formation CNRS

Caractérisation des protéines par spectrométrie de masse dans le contexte de la protéomique. Novembre 2016, Paris. 35 h.

Autres formations

Les fondamentaux du management de projet, Lesaffre/CEPI management. Juillet 2016, Marcq-en-Barœul. 14 h.

Séminaire : Ecole de printemps du Club Jeune de la Société Française de Spectrométrie de Masse. Mars 2017, Trélon. 74 h.

MixOmics advanced workshop. Juin 2018, Orsay. 14 h.

Titre : Analyse multiomique des peptides d'extraits de levure et de leurs impacts fonctionnels sur *Streptococcus thermophilus*

Mots clés : *Streptococcus thermophilus*, extrait de levure, peptide, nutrition, régulation, analyse multiomique

Résumé :

Les bactéries lactiques sont largement utilisées en tant que ferments dans l'industrie laitière. Leur production s'effectue généralement dans des milieux semi-définis ou complexes dans lesquels certains nutriments peuvent être apportés par des extraits de levure (EXLs). Ce projet de thèse s'est focalisé sur l'effet des peptides de deux EXLs (EXL1 et EXL2) sur une souche industrielle de *Streptococcus thermophilus*, un levain lactique d'intérêt économique majeur. L'hypothèse sous-jacente était que ces peptides pourraient avoir un double rôle de nutrition et de régulation de fonctions cellulaires pouvant présenter un intérêt technologique. Afin d'explorer cette question, une stratégie expérimentale à deux niveaux a été élaborée : i) caractérisation et suivi cinétique de la fraction peptidique des deux EXLs par spectrométrie de masse (peptidomique) durant la fermentation de *S. thermophilus* en bioréacteurs, et ii) suivi cinétique parallèle du transcriptome et du protéome de la bactérie. L'objectif final était de croiser ces deux niveaux d'information afin de corréler des différences de contenu peptidique avec des différences d'activation de systèmes participant aux performances globales du levain.

La caractérisation et le suivi du peptidome des EXLs en cours de fermentation a nécessité un important travail de développement méthodologique ayant abouti *in fine* à l'élaboration d'un outil analytique complet, combinant analyse peptidomique à haut-débit des échantillons et traitement bioinformatique et statistique des données. Cet outil a permis d'identifier environ 4000 peptides différents composant les deux EXLs. Le suivi cinétique a notamment permis de préciser la spécificité du transporteur d'oligopeptides de la bactérie (Ami). En particulier, il s'est

avéré qu'une charge nette positive était le facteur prévalent pour le transport des peptides chez *S. thermophilus*. En complément de cette approche semi-quantitative, des analyses quantitatives ont été réalisées sur des fractions peptidiques des EXLs (dosages différentiels par HPLC des acides aminés avant et après hydrolyse). Elles ont notamment permis de révéler d'importantes différences de teneurs en oligopeptides entre les deux EXLs.

En parallèle, le suivi transcriptomique et protéomique réalisé durant la croissance de la bactérie a révélé deux faits marquants. Le premier fait a trait à la surexpression dans l'EXL1 d'un locus génétique régulé par un mécanisme de *quorum sensing* utilisant un peptide phéromone comme signal moléculaire. Le deuxième fait marquant concerne diverses voies de biosynthèse (acides aminés et purines) différenciellement affectées par les deux EXLs. L'origine de ces dynamiques pourrait être au moins pour partie le fait de différences de contenu peptidique entre les deux substrats. Notamment, certaines voies de biosynthèse pourraient avoir été modulées différenciellement sous l'action de régulateurs centraux tels que CodY, dont l'activité est corrélée au contenu peptidique du milieu, ou encore YebC, un régulateur *CodY-like* dont le lien fonctionnel avec CodY reste encore inconnu chez *S. thermophilus*. Tous ces résultats ouvrent d'intéressantes perspectives pour mieux explorer le lien entre peptides et métabolisme bactérien. A terme, cette démarche pourrait se traduire par l'identification de biomarqueurs de performances dans les EXLs, et l'élaboration à façon de produits permettant de maximiser le potentiel technologique des ferments lactiques.

Title: Multiomic analysis of yeast extract peptides and their functional impacts on *Streptococcus thermophilus*

Keywords: *Streptococcus thermophilus*, yeast extract, peptide, nutrition, regulation, multiomic analysis

Abstract:

Lactic acid bacteria are widely used as starters in dairy industry. They are generally produced in complex fermentation media containing a wide array of nutrients that can be provided by yeast extracts (YEs). The main goal of this thesis project was to investigate the effect of the peptide fraction of two YEs (YE1 and YE2) on an industrial *Streptococcus thermophilus* strain, a major lactic acid starter. The underlying hypothesis of this whole project was that YE peptides could have a role in nutrition but also regulate cellular functions of technological relevance. In order to explore this question, a two-step strategy was elaborated: i) mass spectrometry characterization (peptidomics) of both YE peptide fractions and time course analysis of their relative abundance during the growth in bioreactors of *S. thermophilus*, and ii) parallel time course analysis of the strain transcriptome and proteome. The final objective was to cross these two levels of information in order to correlate differences of peptide content with differentially activated systems related to technological performances.

YE peptidome characterization and kinetic analysis first required an important methodological development. It eventually resulted in a complete analytical tool that combines high throughput peptidomic analysis as well as bioinformatic and statistical data processing. This powerful tool was able to identify around 4,000 different peptides in both YEs. Then, the time course analysis also clarified the *in vivo* substrate specificities of the oligopeptide transport system of the bacterium (Ami).

A peptide positive net charge notably turned out to be the leading factor governing peptide transport. In addition to this semi-quantitative approach, quantitative analyses were carried out on YE peptide fractions (differential HPLC analyses of amino acids before and after sample hydrolysis). They notably revealed significant differences in oligopeptides content between both YEs.

Meanwhile, genome scale transcriptomic and proteomic analyses performed during the strain growth highlighted two significant events. The first one concerns the overexpression in YE1 of a *quorum sensing*-based genetic locus that uses a pheromone peptide as molecular signal. The second event relates to several biosynthesis pathways (amino acids and purines) that were differentially affected by both YEs. These dynamics could result from differences in peptide content between both substrates. In particular, some pathways could have been differentially modulated by central regulators such as CodY, whose activity is correlated to the medium peptide richness, or YebC, a CodY-like regulator whose functional link with CodY is still unknown in *S. thermophilus*. All these results open avenues for a better understanding of the interplay between peptides and bacterial metabolism. In the future, this whole approach could lead to the identification of performance biomarkers in YEs, which in turns may eventually translate into the conception of new customized products granting high technological performances to dairy starters.

