



HAL
open science

Caractérisation du locus glycosylasparaginase d'*enterococcus faecalis* et de son implication dans la virulence

Victor Combret

► **To cite this version:**

Victor Combret. Caractérisation du locus glycosylasparaginase d'*enterococcus faecalis* et de son implication dans la virulence. Bactériologie. Normandie Université, 2023. Français. NNT : 2023NORMC252 . tel-04431533

HAL Id: tel-04431533

<https://theses.hal.science/tel-04431533>

Submitted on 1 Feb 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.



THÈSE

Pour obtenir le diplôme de doctorat

Spécialité **PHYSIOLOGIE ET BIOLOGIE DES ORGANISMES - POPULATIONS - INTERACTIONS**

Préparée au sein de l'**Université de Caen Normandie**

Caractérisation du locus glycosylasparaginase d'*Enterococcus faecalis* et de son implication dans la virulence

Présentée et soutenue par
VICTOR COMBRET

Thèse soutenue le 05/12/2023
devant le jury composé de :

MME CRISTEL ARCHAMBAUD	Chargée de recherche - Institut National de Recherche pour l'Agriculture, l'alimentation et l'environnement	Rapporteur du jury
M. MATTHIEU JULES	Professeur des universités - AgroParisTech	Rapporteur du jury
MME SHAYNOOR DRAMSI	Directrice de recherche - Institut Pasteur	Membre du jury
M. ALAIN RINCE	Professeur des universités - Université de Caen Normandie	Président du jury
M. NICOLAS SAUVAGEOT	Maître de conférences - Université de Caen Normandie	Directeur de thèse

Thèse dirigée par **NICOLAS SAUVAGEOT** (COMMUNICATION BACTÉRIENNE ET STRATÉGIES ANTI-INFECTIEUSES)

Remerciements

Je tiens à exprimer toute ma gratitude aux membres de mon jury, le Docteur Cristel Archambaud, le Docteur Shaynoor Dramsi, le Professeur Matthieu Jules, le Professeur Alain Rince et le Docteur Nicolas Sauvageot pour me faire l'honneur d'examiner et juger ce travail.

Je remercie la région Normandie et le Ministère de l'Enseignement Supérieur et de la Recherche pour avoir financé mon doctorat.

Je tiens ensuite à remercier mon directeur de thèse, le Docteur Nicolas Sauvageot. Merci de m'avoir guidé et soutenu pendant ces trois années. Merci de toujours avoir pris le temps de répondre à mes questions et pour tous tes conseils. Merci d'avoir mis les mains dans le cambouis quand j'étais bloqué. Merci pour ta bonne humeur au quotidien, tes blagues, ton soutien, ta bienveillance, ton humanité. Merci pour ces bons moments au labo et ceux en dehors du labo. Merci de m'avoir fait découvrir Jean-Michel Jarre et le synthétiseur. Désolé de t'avoir cassé les oreilles avec mes musiques de sauvage. Merci à toi pour m'avoir fait grandir scientifiquement et humainement. Je suis fier d'avoir été ton padawan durant ces trois années. Du fond du cœur, merci à toi, le meilleur directeur de thèse de toute la galaxie.

Merci à tous les membres du laboratoire CBSA et ex-U2RM qui sont là et qui ont été là. Merci à vous, Axel, Abdellah, Alain, Nico V, Aurélie, Cécile, Florie, Eliette, Caroline, Sacha, Sonia, Isabelle, Sabrina, pour votre aide, votre gentillesse, pour votre altruisme et vos conseils. Travailler avec vous est un véritable plaisir. Merci pour cette ambiance incroyable, une ambiance que l'on souhaiterait retrouver partout. Un grand merci à mes acolytes, Diane, Ben, Solenn, Alex, Fab et Julie. Vous êtes totalement fous ! Il suffit de voir la décoration de notre bureau et de notre magnifique porte de la honte. Travailler avec vous est une vraie joie, vous n'êtes pas que des collègues, vous êtes des amis. Merci aux anciens doctorants Delphine et Valentin. Delphine, nouvelle gourou de la secte sportive de la branche rennaise, merci pour ces fous-rires et les craquages le vendredi après-midi. Tu n'es pas à l'abri que l'on redébarque tous ensemble comme cet été. J'ai hâte de te revoir. Valentin, je ne pensais pas dire ça, mais tes sifflements et râlements me manquent ! Merci d'avoir pris le temps lorsque j'en avais besoin. J'espère te revoir bientôt.

J'aimerai aussi remercier Mathilde Leverrier. Jamais je n'aurai pu rêver mieux comme stagiaire ! Je suis certain que tu vas décrocher ta thèse, tu en as les capacités. Je te souhaite le meilleur, mais aussi pleins de (bonnes) dingeries.

Merci à Benoît et Julien de la plateforme Proteogen. Venir vous voir a toujours été une bouffée d'air lorsque j'en avais ras le bol des expériences qui rataient. Merci à vous pour votre aide sur les manips, et surtout de m'avoir initié au mini tennis de table ! Il me reste encore un peu moins d'un an pour vous montrer qui est le champion ! À nos prochaines parties. Merci à Stéphane et Thomas de la plateforme Imogere pour votre accueil et explications sur comment dompter le Typhoon.

Merci à vous, Quentin, Jérem, Marion, Tiffanie, Jenn, Gauthier, Charlotte, Louis et Adrien, mes amis. Ce n'est pas toujours facile de se croiser depuis que j'ai quitté notre terre natale, mais c'est toujours un plaisir de vous retrouver lorsque je reviens en Auvergne ou lors de vos visites. J'ai tellement de choses à vous raconter !!! Merci pour votre soutien, merci de ces moments de détente, merci de faire partie de ma vie. Je vous aime.

Merci aux copains du master Microbio ! J'espère que l'on gardera longtemps ce lien que l'on a. Même si ça fait 3 ans maintenant (3 ANS ???!!! Mon dieu que le temps passe vite) que l'on a terminé le master, à chaque fois que l'on se retrouve, c'est comme si jamais on ne s'était quitté. Hâte de vous retrouver durant les fêtes de Noël !

Je tiens à remercier ma famille, Maman, Papa, Théo, Kaï, Mat, Diane, Papi, Mémé et Mamie pour m'avoir toujours soutenu et d'avoir cru en moi. Je vous aime. Mamie, j'espère que tu apprécieras cette thèse maintenant que tu es experte en microbiologie ! J'espère vous rendre fier le 5 Décembre.

Je souhaiterai terminer ces remerciements, et te remerciant ma Choupie. Ces 3 années à tes côtés ont été incroyables. J'ai également apprécié la vie caennaise. Mais je pense que j'apprécierai tous les endroits où j'irai du moment que tu es avec moi. Tu es ma maison, mon point de repère. Merci pour ton soutien sans faille durant ces 8 dernières années. Nous sommes des grands maintenant ! Je souhaite que notre voyage ensemble continue encore longtemps.

Table des matières

Remerciements.....	1
Table des matières.....	3
Liste des abréviations.....	7
Liste des figures.....	9
Liste des tableaux.....	10
Contexte.....	11
Introduction bibliographique.....	13
1 Le genre <i>Enterococcus</i>	14
1.1 Historique.....	14
1.2 Caractéristiques des entérocoques.....	14
1.2.1 Caractéristiques générales.....	14
1.2.2 Epidémiologie et antibiorésistance.....	16
1.2.3 Facteurs de virulence.....	17
1.3 La souche de référence : OG1RF.....	20
1.4 Régulation et métabolisme d' <i>E. faecalis</i>	20
1.4.1 Les régulateurs transcriptionnels HutC de la famille GntR.....	21
1.4.1.1 Domaine de fixation à l'ADN.....	22
1.4.1.2 Le domaine régulateur.....	24
1.4.1.3 Le domaine « linker ».....	24
1.4.2 Les systèmes phosphotransférases PEP dépendants (PTS).....	25
1.4.2.1 Fonctionnement général des systèmes PTS.....	26
1.4.2.2 Protéines EI et HPr.....	27
1.4.2.3 Les différentes classes de PTS.....	29
1.4.2.4 Protéines EII.....	33
1.4.2.5 Régulation du métabolisme par le PTS, et autres fonctions.....	35
1.5 Métabolisme du N-acétylglucosamine.....	39
2 Protéines glycosylées et enzymes déglycosylantes.....	41
2.1 Protéines glycosylées.....	42
2.1.1 N-Glycosylations chez les eucaryotes.....	42
2.1.2 O-Glycosylations chez les eucaryotes.....	45
2.1.3 La glycosylation chez les bactéries.....	46
2.2 Enzymes déglycosylantes.....	50
2.2.1 Endo- β -N-acétylglucosaminidases.....	51
2.2.2 Peptide N-glycanases.....	53

2.2.3	Glycosylasparaginase.....	55
3	Lien entre métabolisme et virulence chez <i>E. faecalis</i>	58
4	CRISPR-Cas.....	61
5	Objectifs	63
Matériels et méthodes	67
1	Matériels biologiques.....	68
1.1	Conditions de culture et souches bactériennes	68
1.2	Plasmides.....	70
1.2.1	pMAD.....	70
1.2.2	pMSP3535	70
1.2.3	pGEM-T easy.....	71
1.2.4	pQE30	71
1.2.5	pET29b+	71
1.2.6	pVPL3004.....	72
1.2.7	pVDM1001.....	72
2	Cellules compétentes et transformations par électroporation	73
2.1	<i>Escherichia coli</i>	73
2.2	<i>Enterococcus faecalis</i>	73
3	Cellules compétentes et transformations par choc thermique	74
3.1	<i>Escherichia coli</i>	74
4	Biologie moléculaire.....	75
4.1	Réaction de polymérisation en chaîne (PCR)	75
4.2	Electrophorèse en gel d'agarose.....	75
4.3	Restriction et ligature de fragments d'ADN	75
4.4	Purification des fragments d'ADN.....	76
4.5	Extraction de plasmides d' <i>Escherichia coli</i>	76
4.6	Extraction d'ADN chromosomique.....	76
4.7	Extraction d'ARN totaux.....	76
4.8	Réaction de transcription inverse	77
4.9	PCR quantitative.....	77
5	Constructions de mutants et complémentations	77
5.1	Mutant de délétion : Mutagenèse par recombinaison homologue.....	77
5.2	Mutant de délétion : CRISPR-Cas9	78
5.3	Complémentations des mutants	78
5.4	Séquençage du génome entier.....	79
6	Test d'inhibition à la streptozotocine.....	79

7	Méthodes biochimiques.....	79
7.1	Surexpression et purification des protéines présentant une étiquette poly-histidine.....	79
7.2	SDS-PAGE.....	80
7.3	Retard sur gel	81
7.4	Test enzymatique : Glycosylasparaginase (EgaG)	81
7.5	Test enzymatique : Peptidase (EgaP)	82
8	Survie chez <i>Galleria mellonella</i>	82
9	Biofilm	83
	Résultats	85
1	Régulation de l'opéron <i>ega</i> chez <i>E. faecalis</i>	86
1.1	Identification du locus glycosylasparaginase d' <i>E. faecalis</i> OG1RF	86
1.2	Le locus <i>ega</i> est organisé en deux structures opéroniques	92
1.3	Séquençage entier du génome de la souche Δ <i>egaG</i>	93
1.4	Suivi de croissance des différents mutants <i>ega</i>	93
1.5	Identification du PTS EIIA	95
1.6	Identification du produit de déglycosylation utilisable par le locus <i>ega</i>	96
1.7	EgaR est un répresseur transcriptionnel du locus <i>ega</i>	96
1.8	Répression catabolique du locus glycosylasparaginase	102
2	Opéron glycosylasparaginase et virulence	104
2.1	EgaG code une glycosylasparaginase	104
2.2	Test d'inhibition à la streptozotocine.....	106
2.3	EgaP est une dipeptidase intracellulaire	107
2.4	Le locus <i>ega</i> est impliqué dans la virulence	109
3	Mise en place de l'outil CRISPR-Cas9 chez <i>E. faecalis</i>	111
	Discussion et perspectives	115
1	Caractérisation du locus glycosylasparaginase chez <i>E. faecalis</i> OG1RF	116
2	L'outil de mutagenèse CRISPR-Cas9 chez <i>E. faecalis</i>	122
	Références bibliographiques	125
	Annexes	141
1	Tableaux supplémentaires	142
2	Publication scientifique : « Utilization of glycoprotein-derived N-acetylglucosamine-L-asparagine during <i>Enterococcus faecalis</i> infection depends on catabolic and transport enzymes of the glycosylasparaginase locus ».....	146

Liste des abréviations

ADN	: Acide désoxyribonucléique
ARN	: Acide ribonucléique
ARNr	: Acide ribonucléique ribosomique
cas	: CRISPR associated proteins
CBSA	: Communication Bactérienne et Stratégies Anti-infectieuses
CcpA	: Catabolite control protein A
cdM17	: carbon depleted M17
cre	: catabolite response element
CRISPR	: Clustered regularly interspaced short palindromic repeats
DBD	: DNA Binding Domain
D.O.	: Densité Optique
Dol	: Dolichol
Domaine E-b/O	: Domaine de liaison à l'effecteur et/ou domaine d'oligomérisation
EI	: Enzyme I
EMSA	: Retard sur gel
ENGase	: Endo- β -N-acétylglucosaminidases
ERAD	: Endoplasmic-reticulum-associated protein
FBP	: Fructose-1,6-bisphosphate
Fuc	: Fucose
Fru	: Fructose
Gal	: Galactose
GalNAc	: N-acétylgalactosamine
GlcNAc	: N-acétylglucosamine
Glu	: Glucose
HPr	: Histidine containing protein
Hyp	: Hydroxyproline
Hys	: Hydroxylysine
LB	: Lysogeny broth
LPSN	: List of prokaryotic names with standing in nomenclature
Man	: Mannose
MSCRAMM	: Microbial Surface Components Recognizing Adhesive Matrix Molecules
NaCl	: Chlorure de Sodium
OMS	: Organisation Mondiale de la Santé
OST	: Oligosaccharyltransférase
PCR	: Polymerase Chain Reaction
<i>pgl</i>	: protein glycosylation
PRD	: PTS Regulation Domain

PTS	: Phosphotransferase system (Système phosphotransférase)
QS	: Quorum sensing
RACE	: Rapid amplification of cDNA-ends
RC	: Répression catabolique
RT-qPCR	: Reverse transcriptase quantitative polymerase chain reaction
Sp	: Spectinomycine
TGI	: Tractus gastro-intestinal
wHTH	: winged Helix-Turn-Helix
WT	: Wild type (souche sauvage)
X-Gal	: 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

Liste des figures

Figure 1 : Image d' <i>E. faecalis</i> au microscope électronique à balayage.....	15
Figure 2 : Arbre phylogénétique correspondant aux protéines de la famille GntR.....	21
Figure 3 : Structure de YvoA, régulateur transcriptionnel de la famille GntR, sous-famille HutC.....	23
Figure 4 : Représentation structurale des domaines C-terminaux des sous-familles GntR.....	24
Figure 5 : Implication du domaine linker de la protéine YvoA dans la régulation.....	25
Figure 6 : Fonctionnement général des systèmes PTS.....	27
Figure 7 : Structure primaire de l'enzyme EI.....	27
Figure 8 : Structure tridimensionnelle de la protéine HPr d' <i>E. coli</i>	28
Figure 9 : Représentation des différentes classes de PTS chez <i>E. faecalis</i> OG1RF.....	30
Figure 10 : Exemple du fonctionnement du PTS glucose permettant l'import du maltose chez <i>E. faecalis</i> OG1RF.....	31
Figure 11 : Exemple du fonctionnement du PTS cellobiose permettant l'import de gentiobiose chez <i>E. faecalis</i> OG1RF.....	32
Figure 12 : Fonctionnement et implication de l'activité kinase et phosphorylase de l'enzyme HPrK/P.....	36
Figure 13 : Exclusion et expulsion de l'inducteur.....	37
Figure 14 : Structure des domaines PRD.....	38
Figure 15 : Métabolisme du N-acétylglucosamine chez <i>E. faecalis</i>	40
Figure 16 : Voie de biosynthèse des oligosaccharides liés aux lipides et transfert de l'oligosaccharide à la protéine au niveau de la membrane du réticulum endoplasmique.....	43
Figure 17 : Représentation de différentes chaînes N-glycanes.....	45
Figure 18 : Représentation des différents noyaux O-glycanes.....	46
Figure 19 : N-glycosylation chez <i>C. jejuni</i>	47
Figure 20 : Voie de biosynthèse Pse chez <i>C. jejuni</i> et <i>H. pylori</i>	49
Figure 21 : Hydrolyse de la chaîne glycanes par les enzymes endo- β -N-acétylglucosaminidases.....	51
Figure 22 : Hydrolyse de la chaîne glycanes par les enzymes PNGases.....	53
Figure 23 : Présentation schématique des structures primaires des PNGases de divers organismes.....	54
Figure 24 : Hydrolyse de la liaison glycosidique C1-N du GlcNAc-L-Asn par les glycosylasparaginases.....	56
Figure 25 : Modèle d'autoprotéolyse de la glycosylasparaginase.....	57
Figure 26 : Expression différentielle des gènes d' <i>E. faecalis</i> au cours de l'infection du péritoine de la souris, en fonction des catégories fonctionnelles.....	59
Figure 27 : Système CRISPR-Cas9.....	62
Figure 28 : Modèle initial du fonctionnement du locus comportant la glycosylasparaginase chez <i>E. faecalis</i>	64
Figure 29 : Test enzymatique de l'activité glycosylasparaginase.....	82
Figure 30 : Organisation génétique du locus <i>ega</i> d' <i>E. faecalis</i> OG1RF.....	86
Figure 31 : Séquence du gène <i>egaD</i> et structure des protéines EgaD.....	88
Figure 32 : Comparaison de l'organisation structurale du locus <i>ega</i>	91
Figure 33 : 5'RACE-PCR dans l'espace intergénique compris entre <i>egaP</i> et <i>egaG</i>	92
Figure 34 : Analyse par RT-PCR de la co-transcription des gènes de l'opéron <i>ega</i>	92
Figure 35 : Etude de la croissance de la souche sauvage et des différents mutants <i>ega</i>	94
Figure 36 : Expression relative des gènes <i>egaR</i> , <i>egaP</i> , <i>egaG</i> , <i>egaB</i> , <i>egaC</i> et <i>egaD1</i> dans la souche sauvage OG1RF.....	96
Figure 37 : Structure de la protéine EgaR.....	97
Figure 38 : Effet du GlcNAc-Asn sur l'expression des gènes <i>ega</i>	98

Figure 39 : Expériences de retard sur gel des régions promotrices P1 et P2 en présence d'EgaR	99
Figure 40 : Expériences de retard sur gel avec les motifs HutC mutés pour les régions promotrices P1 et P2 et la protéine EgaR.....	100
Figure 41 : EMSA des régions promotrices P1 et P2 en présence de GlcNAc-L-Asn, avec la protéine EgaR.....	101
Figure 42 : Expériences de retard sur gel de la région promotrice P1 en présence de GlcNAc ou Asparagine, avec la protéine EgaR	102
Figure 43 : Impact de la répression catabolique sur les gènes <i>egaG</i> , <i>egaB</i> , <i>egaC</i> , <i>egaD1</i>	103
Figure 44 : Purification de la protéine glycosylasparaginase (EgaG) et analyse par spectrométrie de masse.....	104
Figure 45 : Analyse cinétique de l'activité glycosylasparaginase d'EgaG.....	105
Figure 46 : Détermination du site de coupure par la glycosylasparaginase	106
Figure 47 : Test d'inhibition à la streptozotocine.	107
Figure 48 : Dégradation de la RNaseB par EgaP.....	108
Figure 49 : Analyse de la structure primaire de EgaP et analyse par spectrométrie de masse	109
Figure 50 : Implication du locus <i>ega</i> dans la virulence	109
Figure 51 : Représentation schématique des étapes réalisées pour obtenir les plasmides comportant l'outil CRISPR-Cas9 dirigé contre le gène <i>egaC</i>	111
Figure 52 : Présence du plasmide pCRISPR- <i>egaC</i> chez les souches mutées après transformation. ...	112
Figure 53 : Vérification de l'outil CRISPR-Cas9.....	113
Figure 54 : Efficacité de l'outil CRISPR-Cas9 après modification de la cassette de résistance	114
Figure 55 : Modèle proposé de la voie catabolique du GlcNAc-L-Asn chez <i>E. faecalis</i>	121

Liste des tableaux

Tableau 1 : Résistances antimicrobiennes chez les entérocoques	17
Tableau 2 : Principaux facteurs d'opportuniste ou de virulence décrits chez <i>E. faecalis</i>	19
Tableau 3 : Fusion des différents domaines PTS.....	34
Tableau 4 : Différents types de glycosylations bactériennes.....	50
Tableau 5 : Composition des ENGases les plus étudiées.....	52
Tableau 6 : Impact de différentes conditions de stress sur le métabolisme d' <i>E. faecalis</i>	60
Tableau 7 : Souches utilisées dans ce travail.	68
Tableau 8 : Plasmides utilisés dans ce travail.	72
Tableau 9 : Résultat de l'analyse de détection de mutation dans la souche Δ <i>egaG</i> . VNM : Variant nucléotidique multiple ; VNS : Variant nucléotidique simple.	93
Tableau S1 : Composition du milieu M17.	142
Tableau S2 : Composition du milieu cdM17.	142
Tableau S3 : Composition du milieu LB.....	142
Tableau S4 : Oligonucléotides utilisés pour la construction des mutants et des complémentations.	142
Tableau S5 : Oligonucléotides utilisés pour la construction des plasmides de surexpression.	144
Tableau S6 : Oligonucléotides utilisés pour les expériences de RT-qPCR.....	144
Tableau S7 : Oligonucléotides utilisés pour l'expérience de 5'Race-PCR.	144
Tableau S8 : Autres oligonucléotides.....	145

Contexte

Les entérocoques sont des bactéries opportunistes que l'on retrouve dans le tractus gastro-intestinal (TGI) [1]. Depuis le début des années 2000, les entérocoques, principalement *Enterococcus faecalis* et *Enterococcus faecium* sont considérés comme des pathogènes d'intérêt clinique provoquant diverses infections : intra-abdominales ou urinaires, des septicémies ou des endocardites. Ces microorganismes ont la capacité de s'adapter aux environnements hostiles et à résister à de nombreux antibiotiques, rendant leur traitement difficile [2,3]. Dans l'objectif de prévenir et traiter ces infections, l'étude du mécanisme infectieux chez ces bactéries est primordiale. Pour proliférer chez leur hôte, ces pathogènes recherchent activement une source d'énergie utilisable. Lors de différentes études, l'unité de recherche Communauté Bactérienne et Stratégies Anti-Infectieuses a montré que lors d'une infection, *E. faecalis* surexprimait de nombreux gènes codant pour des enzymes impliquées dans la dégradation des polysaccharides nommées glycosidases [4,5]. Ces enzymes sont capables de cliver la liaison entre 2 résidus osidiques (glucose, fructose, mannose, N-acétylglucosamine (GlcNAc), ...) composant le polysaccharide et sont spécifiques du type de liaison qui unit ces résidus (liaisons α ou β).

Une source potentielle de polysaccharides chez les eucaryotes sont les protéines glycosylées. Il existe deux principaux types de glycosylation : la N-glycosylation consistant à l'ajout de sucres sur le groupement amine (NH_2) du résidu asparagine, et la O-glycosylation consistant à l'ajout de sucre sur des résidus sérine/thréonine [6,7]. La structure et la composition des chaînes polysaccharidiques de ces protéines varient, mais dans le cas des protéines N-glycosylées, une structure commune pentasaccharidique $\text{GlcNAc}_2\text{Man}_3$ est systématiquement greffée à l'asparagine [6]. Trois principaux types de glycosidases capables de dégrader cette structure pentasaccharidique ont été décrits : d'une part les endo- β -N-acétylglucosaminidases [8] clivant la liaison entre les deux résidus GlcNAc, et d'autre part les peptidases N-glycanases [9] et les glycosylasparaginases [10] clivant la liaison entre un résidu asparagine et un résidu GlcNAc. Ces enzymes sont présentes chez les mammifères où elles participent à la dégradation des protéines, mais peu d'informations sont disponibles sur leur rôle chez les procaryotes.

Introduction bibliographique

1 Le genre Enterococcus

1.1 Historique

Thiercelin décrit en 1899 un microorganisme diplocoque à Gram positif résidant dans le TGI et capable de provoquer des infections. Compte tenu de sa forme et de sa localisation, il propose de nommer ce microorganisme « entérocoque » [11]. La même année, MacCallum et Hastings rapportent le cas d'une endocardite infectieuse fatale provoquée par la bactérie « *Micrococcus zymogenes* », maintenant reconnue comme étant *Enterococcus faecalis* [2]. En 1906, Andrewes et Horder identifient un microorganisme présentant des propriétés similaires à celles données par Thiercelin, et le nomment *Streptococcus faecalis* [12]. Ce n'est qu'en 1919 que Orla-Jensen apporte la première description d'un autre streptocoque : *Streptococcus faecium* mettant en avant les différences de fermentation avec *Streptococcus faecalis*. Jusqu'en 1984, les entérocoques étaient associés au groupe D des streptocoques d'après la classification de Lancefield [13]. La distinction entre streptocoques et entérocoques a enfin été proposée suite à l'apparition d'études basées sur l'hybridation ADN-ADN, analyses de séquences ARNr 16S, analyses immunologiques de certaines enzymes comme la superoxyde dismutase et enfin par des caractéristiques biochimiques telles que le test de tolérance au chlorure de sodium (NaCl) ou galerie API Rapid Strep panel [14–19]. Actuellement, 83 espèces appartenant au genre *Enterococcus* ont été décrites selon le site « List of prokaryotic names with standing in nomenclature » (LPSN) [20].

1.2 Caractéristiques des entérocoques

1.2.1 Caractéristiques générales

Les entérocoques appartiennent phylogénétiquement au phylum des *Bacillota* (anciennement *Firmicutes*). Ces bactéries sont Gram positives, non sporulantes, mésophiles, catalase et oxydase négatives, aéro-anaérobies facultatives, chimio-organotrophes et majoritairement immobiles [14,21]. Les entérocoques possèdent un génome ayant un faible pourcentage en bases GC (37 %-45 %) et riche en régions mobiles [17]. La cellule est de forme sphérique ou ovoïde, d'une taille comprise entre 0,6 µm et 2,5 µm, isolée ou associée en paires ou courtes chainettes (**Figure 1**) [22]. Les entérocoques se distinguent des streptocoques de par leur capacité à croître en milieu hostile. En effet, ces bactéries peuvent supporter des

conditions défavorables pour d'autres bactéries telles que des fortes concentrations en NaCl (6,5 %), en sels biliaires (40 %), des pH compris entre 4,1 et 9,6, et des températures allant de 10 °C à 45 °C alors que sa température optimale est de 35 °C. Ces bactéries résistent également à la congélation et à la pasteurisation basse (60 °C durant 30 minutes) [2,23–25].

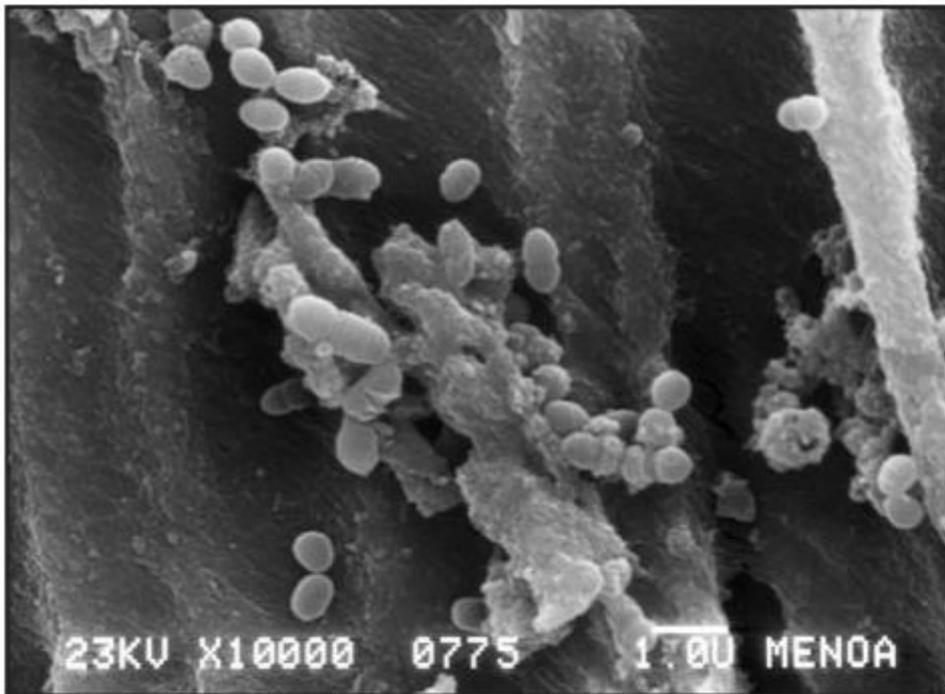


Figure 1 : Image d'*E. faecalis* au microscope électronique à balayage (Bulacio et al., 2015) [26].

Les entérocoques sont des bactéries ubiquitaires retrouvées dans l'environnement et le plus souvent dans le TGI des mammifères, dont l'Homme [1,27]. Leur présence dans l'environnement, comme dans les milieux aquatiques, peut refléter une contamination fécale du milieu [14,28,29]. *E. faecalis* et *E. faecium* sont les bactéries du genre *Enterococcus* les plus documentées car au sein de ce genre, elles sont les plus abondantes dans les fèces humaines, et sont les principales espèces provoquant des infections chez l'Homme. Elles représentent 0,1 % du microbiote chez l'Homme en bonne santé [2,23]. Ce pourcentage est susceptible de varier en fonction du compartiment intestinal, de l'alimentation ou encore de l'âge de l'individu [14]. Enfin, certaines espèces du genre ont été décrites comme étant capables d'assimiler le cholestérol, de produire des acides gras à courtes ou longues chaînes, et peuvent même être utilisées comme probiotiques [23,30–32]. Cependant, la Food and Drug Administration n'a jamais approuvé de probiotique contenant des entérocoques pour traiter,

guérir ou améliorer l'état d'un patient malade. De plus, les études menées montrant les effets bénéfiques sont controversées [33].

1.2.2 Epidémiologie et antibiorésistance

Bien que membre commensal du microbiote intestinal, certaines souches d'entérocoques sont considérées comme des pathogènes opportunistes pouvant être à l'origine d'infections s'ils traversent la barrière intestinale [34]. Ces infections peuvent aller des infections intra-abdominales, du tractus urinaire, des tissus mous, des valves cardiaques jusqu'aux méningites dans les cas les plus graves [2,3,14,35]. Le genre *Enterococcus* est responsable de 14 % et 10 % des infections associées aux soins respectivement aux Etats-Unis d'Amérique et en Europe, avec les pathogènes *E. faecalis* et *E. faecium* représentant 90 à 95 % des cas [2]. De façon plus anecdotique, des infections par les espèces *E. avium*, *E. casseliflavus*, *E. durans*, *E. gallinarum*, *E. hirae*, *E. mundtii* et *E. raddinosus* ont également été signalées [36,37]. En France, entre 2001 et 2017, le genre *Enterococcus* représentait la seconde cause des infections nosocomiales (9,4 %) derrière *Klebsiella pneumoniae* (11,2 %) selon le rapport de l'Institut National de Veille Sanitaire de 2018 [38]. Leur présence en milieu hospitalier est expliquée par plusieurs aspects qui sont que i) les entérocoques ont une forte tolérance aux conditions environnementales et une forte adaptabilité à leur environnement, ii) ces microorganismes ont la capacité de survivre sur des surfaces inertes sur de longues périodes [2,3] et iii) ils possèdent des résistances intrinsèques et acquises aux antibiotiques (**Tableau 1**). Des résistances intrinsèques aux aminoglycosides (gentamycine), β -lactamines (ampicilline), lincosamides (clindamycin), céphalosporines (cefepime) ou encore aux inhibiteurs de la voie métabolique des folates ont été en effet décrites dans la littérature [3,39]. De plus, des résistances acquises aux antibiotiques d'intérêt clinique ont été observées, telles que des résistances aux fluoroquinolones (ciprofloxacine), oxazolidinoles (linézolide), tétracycline (tigécycline) ou encore aux glycopeptides (vancomycine) [3]. Bien que les entérocoques ne soient pas des pathogènes hautement virulents [40], l'Organisation Mondiale de la Santé (OMS) a classé *E. faecium* en tant que pathogène de priorité élevée pour la recherche et le développement de nouveaux moyens de lutte. Cela est dû à ses nombreuses résistances aux antibiotiques, dont notamment la vancomycine, antibiotique utilisé en dernier recours [41,42].

Tableau 1 : Résistances antimicrobiennes chez les entérocoques (García-Solache et Rice., 2019) [3].

Famille d'antimicrobiens	Gène(s) de résistance / Opéron(s) représentatif(s)	Mécanisme de résistance
Aminoglycosides	<i>aac-2'-aph-2'-Ie, aph-3'-IIIa</i>	Modification de l'aminoglycoside
β-lactamines	<i>pbp4 (E. faecalis), pbp5 (E. faecium)</i>	Réduction de l'affinité pour l'antibiotique
Chloramphénicol	<i>cat</i>	Acétylation du chloramphénicol
Clindamycine	<i>Isa(A)</i>	Pompe à efflux putative
Daptomycine	<i>liaFSR</i>	Altération de la charge de la membrane et de sa fluidité
Erythromycine	<i>ermB</i>	Méthylation du ribosome
Fluoroquinolones	<i>gyrA, parC</i>	Modifications dans la région déterminant la résistance aux quinolones
Glycopeptides	<i>vanA, vanB, vanD, vanM</i>	Modification du précurseur du peptidoglycane se terminant par un D-lactate
	<i>vanC, vanE, vanG, vanL, vanN</i>	Modification du précurseur du peptidoglycane se terminant par une D-sérine
Oxazolidinones	rRNA genes	Mutations réduisant l'affinité
	<i>cfr</i>	Méthylation de l'ARN 23S
Rifampicine	<i>rpoB</i>	Mutations ponctuelles réduisant l'affinité
Streptomycine	<i>ant-6</i>	Modification de la streptomycine
Tétracycline	<i>tet(L)</i>	Efflux
	<i>tet(M)</i>	Protection ribosomique
Tigécycline	<i>tet(L), tet(M)</i>	Augmentation de l'expression

1.2.3 Facteurs de virulence

En plus des capacités d'adaptations et de résistances aux conditions environnementales ou aux antibiotiques, les entérocoques possèdent différents facteurs de virulence intrinsèques leur permettant d'initier le processus d'infection. Plusieurs facteurs ont été caractérisés chez *E. faecalis* (**Tableau 2**). La première étape d'infection est l'adhésion de la bactérie à la surface des cellules de l'hôte, impliquant chez *E. faecalis* différents types de protéines de surfaces dont les adhésines. Ces dernières possèdent en extrémité N-terminale un peptide signal et d'un domaine RGD constitué des 3 AA arginine, glycine, acide aspartique, et un motif LPxTG en extrémité C-terminale, leur permettant de s'ancrer dans la paroi bactérienne [43,44]. Trois adhésines principales ont été décrites chez les entérocoques : Asp1, Asc10 et Asa1, toutes produites par les plasmides conjugatifs pPD1, pCF10 et pAD1, respectivement [43,44]. Il a été montré que les adhésines Asa1 et Asc10 sont impliquées dans la liaison aux macrophages *via* leur domaine N-terminal RGD, favorisant ainsi l'internalisation des cellules bactériennes. Une fois internalisées dans les macrophages, les bactéries inhibent

l'attaque oxydative (respiratory burst en anglais), leur permettant de survivre et d'être transloquées [44]. Ces protéines permettent également l'agrégation entre bactéries donneuses et receveuses favorisant ainsi la conjugaison bactérienne [43,44]. D'autres adhésines, nommées MSCRAMM (Microbial Surface Components Recognizing Adhesive Matrix Molecules) permettent la fixation de la bactérie à des protéines de la matrice extracellulaire comme le collagène, la laminine, le fibrinogène, la fibronectine ou encore la thrombospondine. Ici encore, trois principales protéines MSCRAMM ont été décrites : Acm (Adhesion of collagen, chez *E. faecium*), Scm (Second collagen adhesion, chez *E. faecium*) et Ace (Adhesion of collagen, chez *E. faecalis*) [43,44]. Deux opérons (*ebp* et *bee*) dont les produits permettent la formation des pili, et la protéine HylA, sont impliqués dans le développement d'endocardites et la formation de biofilm [43–45]. Les biofilms sont des populations microbiennes adhérentes sur des surfaces et les unes aux autres, et englobées dans une matrice autoproduite [46]. Cette vie en communauté permet d'augmenter la résistance des cellules aux antibiotiques et à la phagocytose, et favorise également le transfert horizontal de gènes de résistance du fait de la proximité entre les cellules [2,47]. La production de biofilm est favorisée par le Quorum Sensing (QS). Le QS induit une modification transcriptionnelle chez les bactéries, ce qui conduit, entre autres, à l'expression de facteurs de virulences. Chez *E. faecalis*, le QS régule les gènes *fsr* contrôlant ainsi la formation de biofilm, mais aussi la sécrétion de peptidases comme GelE ou SprE. Ces peptidases sont impliquées dans i) la longueur des chaînettes d'entérocoques, ii) l'autolyse libérant l'ADN utilisé pour la matrice extracellulaire du biofilm, iii) la dégradation des protéines de surface mal repliées et iv) la dégradation des protéines de l'hôte, dont les éléments du complément [44,48–52]. La cytolysine est une protéine sécrétée chez plusieurs espèces d'entérocoques dont notamment *E. faecalis*. Cette bactériocine possède une activité hémolytique contre les cellules procaryotes et eucaryotes, notamment contre les macrophages et leucocytes polymorphonucléaires [53]. De plus, certaines souches d'*E. faecalis* dont la V583 possèdent un îlot de pathogénicité. Cet îlot d'environ 150 kb possède 129 cadres ouverts de lecture codant des facteurs de virulence comme la protéine Esp (Enterococcal surface protein), des substances d'aggrégations (Asa1, Asc10), cytolysine ou de résistance à la nisine. Plusieurs éléments impliqués dans le métabolisme sont retrouvés dans l'îlot de pathogénicité dont des protéines des systèmes transporteurs ABC et PTS, ce qui laisse suggérer que la recherche de

sources de carbone est une problématique importante pour la bactérie en condition infectieuse [25,54].

Tableau 2 : Principaux facteurs d'opportuniste ou de virulence décrits chez *E. faecalis* : rôles connus lors du processus d'infection et pathologie associée (Salze, 2019) [55].

Facteurs de virulence	Exemples	Rôles principaux	Pathologie associée
Facteurs de la surface cellulaire	Substance d'agrégation (Asc10, Asp1, Asa1)	Interaction et internalisation par les cellules de l'hôte, évasion du système immunitaire	Endocardite
	Protéine de surface (Esp)	Formation de biofilm, associée aux souches résistantes à l'ampicilline	Infection urinaire
	Adhésine au collagène (Ace)	Attachement aux tissus de l'hôte, bactériémie	Endocardite
	Protéine A liant la fibronectine (EfbA),	Attachement à la matrice extracellulaire	Endocardite
	Métalloprotéase de membrane (Eep),	X	Endocardite
	Glycolipides (BgsA, BgsB),	X	Endocardite
	Polysaccharides (CpsA, CpsB, Epa)	Capsule, échappement au système immunitaire, formation de biofilm	X
Facteurs sécrétés	Hémolysine, Cytolysine (Cyl)	Cytotoxicité	X
	Gélatinase (GelE),	X	Endocardite
	Sérine protéase (SprE)	Survie en infection intrapéritonéale (souris)	X
Protéines de stress	Protéine générale de stress Gls24	Résistance aux stress	Endocardite
	Peroxydases (Npr, Tpx)	Résistance dans les macrophages	X
Pili et fimbriae	EbpABC	Adhérence au collagène, formation de biofilm, interaction avec les plaquettes	Endocardite

Les croix (X) indiquent des éléments non déterminés.

Une étude menée par Vergis et ses collaborateurs en 2002 a mis en évidence qu'il n'y a pas de corrélation significative entre la mortalité et la présence d'un ou plusieurs des facteurs de virulence (cytolysine, Esp et gélatinase) [56]. Au contraire, certaines souches

dépourvues de ces facteurs de virulence sont malgré tout capables de provoquer des infections chez l'Homme [57].

1.3 La souche de référence : OG1RF

Le génome de la souche *E. faecalis* OG1RF a été intégralement séquencé en 2008 [32]. Cette souche dérive de la souche OG1, souche commensale pouvant provoquer des caries chez les rats [32,58]. Cette souche a été sélectionnée en laboratoire et mutée pour résister à la rifampycine (OG1-R) puis à l'acide fusidique (OG1RF) [32,58]. La souche OG1RF possède un génome de 2,7 Mpb qui contrairement à la souche clinique V583 est dépourvu d'îlot de pathogénicité, de plasmide, et présente peu d'éléments génétiques mobiles. Ceci peut s'expliquer par la présence chez OG1RF et l'absence chez V583 de loci CRISPR [59]. Bien que ne possédant pas d'îlot de pathogénicité, elle est capable d'infection et est couramment étudiée lors d'infection en modèle animal tel que le lapin [60], la souris [61], le rat [62] ou encore les nématodes [63]. La souche OG1RF en tant qu'objet d'étude a 2 avantages par rapport à la souche V583 : l'absence de plasmide permet d'utiliser plus d'outils moléculaires, et elle est plus facilement transformable. La souche OG1RF est donc devenue une souche de référence pour l'étude des entérocoques et a été utilisée lors de ce travail.

1.4 Régulation et métabolisme d'*E. faecalis*

Pour répondre aux changements environnementaux et intracellulaires, les microorganismes ont la capacité de percevoir le milieu dans lequel ils se trouvent. Le signal est capté puis transmis, la plupart du temps, à des protéines capables de se lier à l'ADN : les régulateurs transcriptionnels. En réponse à un stimulus, le régulateur transcriptionnel va se fixer à l'ADN et entraîner une modification de l'expression du/des gène(s) concerné(s) pour permettre au microorganisme de s'adapter aux changements et/ou de maintenir une balance précise de métabolites. Il est estimé que 10 % du génome des bactéries code des régulateurs transcriptionnels, dont ceux de la famille GntR étudiée lors de ce travail de thèse [64].

sept sous-familles en fonction des groupes de branches qui ressortent de l'arbre représenté. L'ellipse ombrée montre la nouvelle sous-sous-famille constituée par les homologues de *citO* appartenant à la sous-famille FadR (Blancato et al., 2008) [70].

Ces protéines sont retrouvées sous forme homodimérique, même s'il a déjà été rapporté une forme homotétramérique [67]. Chaque sous-famille reconnaît un motif de liaison à l'ADN lui étant propre, mais il est admis que l'ensemble des régulateurs transcriptionnels GntR reconnaissent le motif 5'- N_YGTN_XACN_Y-3' (X et Y étant le nombre de répétitions, et N n'importe quel nucléotide). Les nucléotides compris entre 5'- N_YGT et ACN_Y-3' ne sont presque exclusivement que des A et des T [64]. Ces régulateurs agissent majoritairement en tant que répresseurs même si des cas d'activateurs et activateurs/répresseurs de façon simultanée ont été décrits [66]. La fixation d'un métabolite sur le régulateur entraîne un changement de conformation, ce qui influence ses propriétés de liaison à l'ADN permettant ainsi de nuancer son activité. Bien que la sous-famille FadR soit la plus représentée au sein des GntR (40 % de l'ensemble des régulateurs GntR), cette étude s'intéressera seulement à la sous-famille HutC (30 % de l'ensemble des régulateurs GntR) dont le premier représentant découvert est un répresseur impliqué dans l'utilisation de l'histidine chez *Pseudomonas putida* [67]. Les régulateurs de cette famille sont impliqués dans la régulation de la production d'antibiotiques et ou peuvent être des senseurs de différents métabolites impliqués dans le métabolisme comme les acides gras, le GlcNac ou divers sucres [66,69]. Les régulateurs de la sous famille HutC possèdent 3 domaines qui seront détaillés ci-dessous.

1.4.1.1 Domaine de fixation à l'ADN

La partie N-terminale des régulateurs de la sous famille HutC contient un domaine de fixation à l'ADN de 69 AA. Malgré une composition en AA variable, la structure de ce domaine de fixation reste très conservée. Ce domaine de fixation possède une structure « winged Helix-Turn-Helix » (wHTH) [66,67], capable de se lier sous forme de dimère à des motifs ADN palindromiques de séquence 5' - N_YGTMTAKACN_Y - 3' espacés de 5 à 8 pb entre les sites [64]. La seconde hélice α appelée « hélice de reconnaissance » interagit avec le sillon principal de l'ADN *via* des interactions spécifiques [67]. Différents résidus impliqués dans l'interaction avec l'ADN sont conservés. Une étude cristallographique du régulateur GntR NagR, appartenant à

la sous-famille HutC, a montré que les arginines en position 38 et 48 ainsi que la glycine en position 69 sont impliquées dans l'interaction à l'ADN (**Figures 3A et 3B**) [71]. Les arginines, réalisant une liaison hydrogène bidentée avec la guanine du principal sillon de l'ADN, sont conservées chez différents régulateurs GntR, dont le régulateur YvoA appartenant à la sous-famille HutC [67].

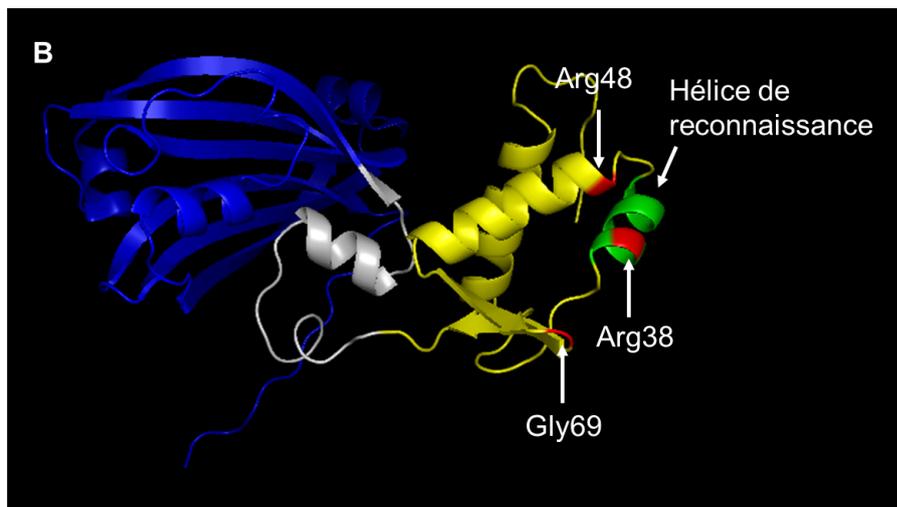
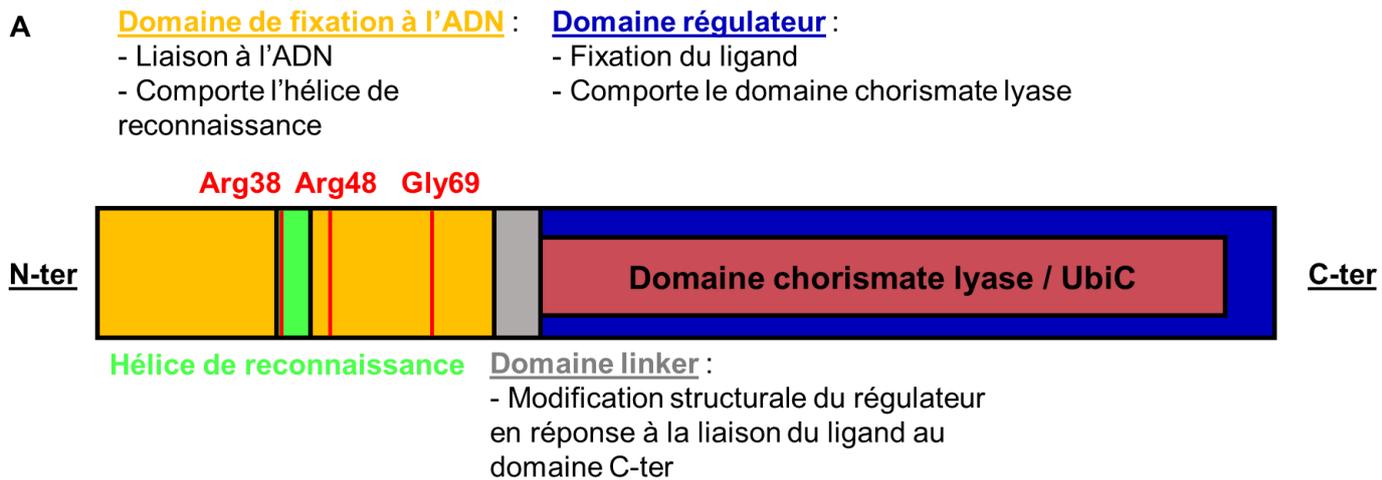


Figure 3 : Structure de YvoA, régulateur transcriptionnel de la famille GntR, sous-famille HutC. **A :** Schéma de la structure primaire. Les domaines sont représentés par des blocs de couleurs jaune, grise, bleue. Une description de ces 3 domaines est indiquée. Les 3 AA impliqués dans la liaison à l'ADN sont indiqués en rouge. **B :** Structure tridimensionnelle. Les domaines et acides aminés d'intérêt sont représentés avec le même code couleur que dans la figure 3A. Par soucis de clarté, le domaine chorismate lyase n'a pas été représenté ici. Un seul monomère est représenté bien que la protéine s'organise sous forme homodimérique. La structure tridimensionnelle de la protéine YvoA est disponible sur la banque de structure tridimensionnelle de macromolécules MMDB. ID : 79163 [72].

1.4.1.2 Le domaine régulateur

La partie C-terminale comporte le domaine de liaison à l'effecteur et/ou domaine d'oligomérisation (domaine E-b/O). Ce domaine de 170 AA est variable, distinguant ainsi les sous-familles GntR entre elles. Le domaine régulateur est constitué d'hélices α et de feuillets β , comme représenté en **Figure 4** [66]. L'étude de ce domaine met en évidence une homologie structurale avec la chorismate lyase (UbiC chez *E. coli*) (**Figures 3A et 3B**), qui assure le retrait du pyruvate du chorismate. Cependant, aucune activité chorismate lyase n'a été mesurée. Cette structure présente une grande cavité dans laquelle le ligand viendra se fixer [67,73,74].

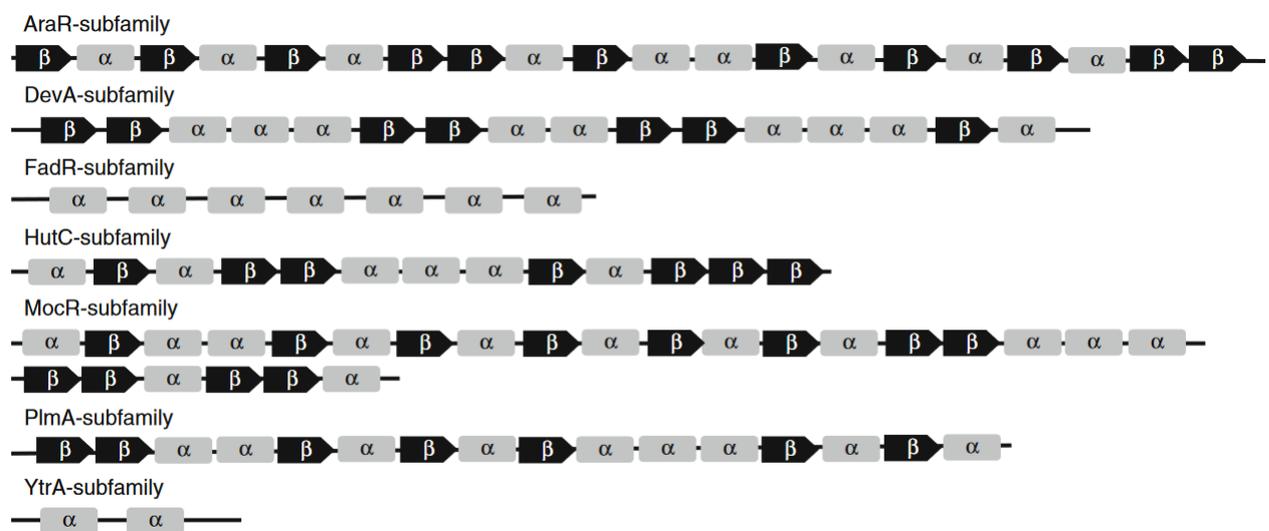


Figure 4 : Représentation structurale des domaines C-terminaux des sous-familles GntR. Les prédictions de structure secondaire ont été réalisées à l'aide du serveur web ProteinPredict et ont été comparées à la structure 3D de FadR. Les hélices α et les feuillets β ne sont pas à l'échelle et représentent simplement l'arrangement structural de ces domaines (Hoskisson et Rigali, 2009) [66].

1.4.1.3 Le domaine « linker »

Comme dit précédemment, le domaine « linker » permet de relier les domaines N-terminal et C-terminal. Chez les protéines de la famille GntR, ce domaine fait entre 1 et 60 AA, même si certains domaines linkers de 150 AA ont été identifiés [75]. En raison d'une grande hétérogénéité, aucun motif ni structure n'a pu être déterminé concernant les « linkers » GntR. Ce domaine est aussi impliqué dans différentes activités, dont la modification structurale du régulateur lorsqu'un ligand s'y lie [67,75]. Un exemple peut être donné avec la protéine YvoA. La fixation de l'effecteur GlcNAc-6P à YvoA modifie la structure du domaine « linker » (21 AA, 77-97) le faisant passer d'une conformation boucle vers hélice. Cette transition permet la

séparation des deux domaines DBD fixés à l'ADN entraînant le décrochage du régulateur à l'ADN (**Figure 5**) [67,74].

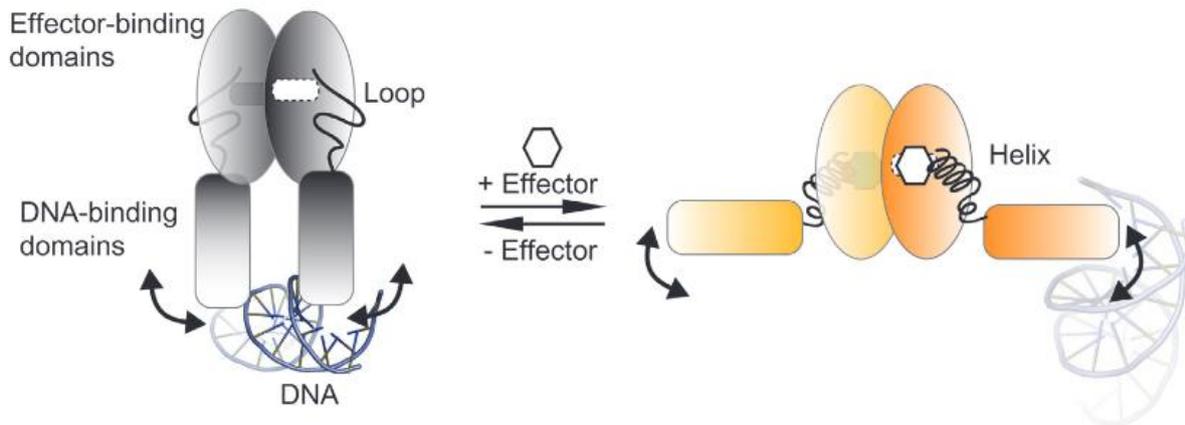


Figure 5 : Implication du domaine linker de la protéine YvoA dans la régulation. Dans la conformation de liaison à l'ADN (en gris), l'homodimère YvoA se lie à l'ADN avec ses hélices de reconnaissance de l'ADN appariées à l'interface du dimère et se liant dans les sillons majeurs successifs. Ce faisant, YvoA est capable de bloquer la transcription de tout gène en aval. Lors de la liaison de l'effecteur (GlcNAc-6-P), la région de liaison interdomaine subit une transition de boucle à hélice, forçant les domaines de liaison à l'ADN à se séparer dans un mouvement semblable à celui d'un "jumping jack" (orange). YvoA n'est alors plus en mesure de se lier au même duplex d'ADN avec les deux domaines de liaison à l'ADN simultanément, ce qui entraîne la dérégulation des gènes contrôlés par YvoA. (Resch et al., 2010) [74].

De par leur rôle, les régulateurs transcriptionnels permettent à la bactérie une rapide adaptation à son milieu. D'autres mécanismes d'adaptations sont mis en place par la bactérie, dont la régulation du métabolisme *via* les systèmes phosphotransférases (PTS).

1.4.2 Les systèmes phosphotransférases PEP dépendants (PTS)

Comme vu précédemment, les bactéries du genre *Enterococcus* possèdent une remarquable capacité d'adaptation à leur environnement. Une des problématiques importantes pour la persistance de la bactérie est la recherche d'une source de carbone métabolisable. *E. faecalis* est capable d'utiliser au moins 13 sources différentes qui seront métabolisées par 3 voies métaboliques principales : la glycolyse, la voie Entner-Doudoroff, et la voie des pentoses phosphates. Pour métaboliser ces composés, la bactérie doit, dans un premier temps, les internaliser. Pour cela, *E. faecalis* dispose de trois types de transporteurs : les perméases, les systèmes ABC (ATP Binding Cassette) et surtout les systèmes PTS.

Les transporteurs PTS sont des complexes enzymatiques permettant l'import d'un sucre-PTS dans le cytoplasme de la cellule via une cascade de phosphorylation aboutissant à la phosphorylation de manière concomitante du sucre internalisé. Ils ont été décrits pour la première fois chez *E. coli* en 1964 [76] avant d'être mis en évidence chez les *Bacillota*, entérobactéries et archées. [77–79]. À ce jour, 7 familles de système PTS ont été décrites ou identifiées *in silico* chez *E. faecalis* (glucose, fructose, lactose, mannose, galactitol, glucitol et L-ascorbate) [77] pour un total de 82 gènes selon la base de données KEGG [80]. Ce complexe enzymatique est constitué de deux groupes de protéines :

- Les protéines impliquées dans l'utilisation de l'ensemble des sucres, *i.e.* protéines EI (Enzyme I) et HPr (Histidine containing Protein).
- Les protéines spécifiques du sucre-PTS transporté et comprenant trois domaines : EIIA, EIIB, EIIC voire quatre dans le cas des systèmes PTS Mannose (EIIABCD).

1.4.2.1 Fonctionnement général des systèmes PTS

Le phosphoenolpyruvate (PEP) est le donneur de phosphate qui cède son groupement phosphate à la protéine EI au niveau de l'His191. Ce groupement phosphate est ensuite transféré à HPr au niveau de l'His15 puis transmis à une enzyme EIIA, puis à l'enzyme EIIB et finalement au sucre lors de son passage à travers le domaine transmembranaire formé par l'enzyme EIIC (**Figure 6**).

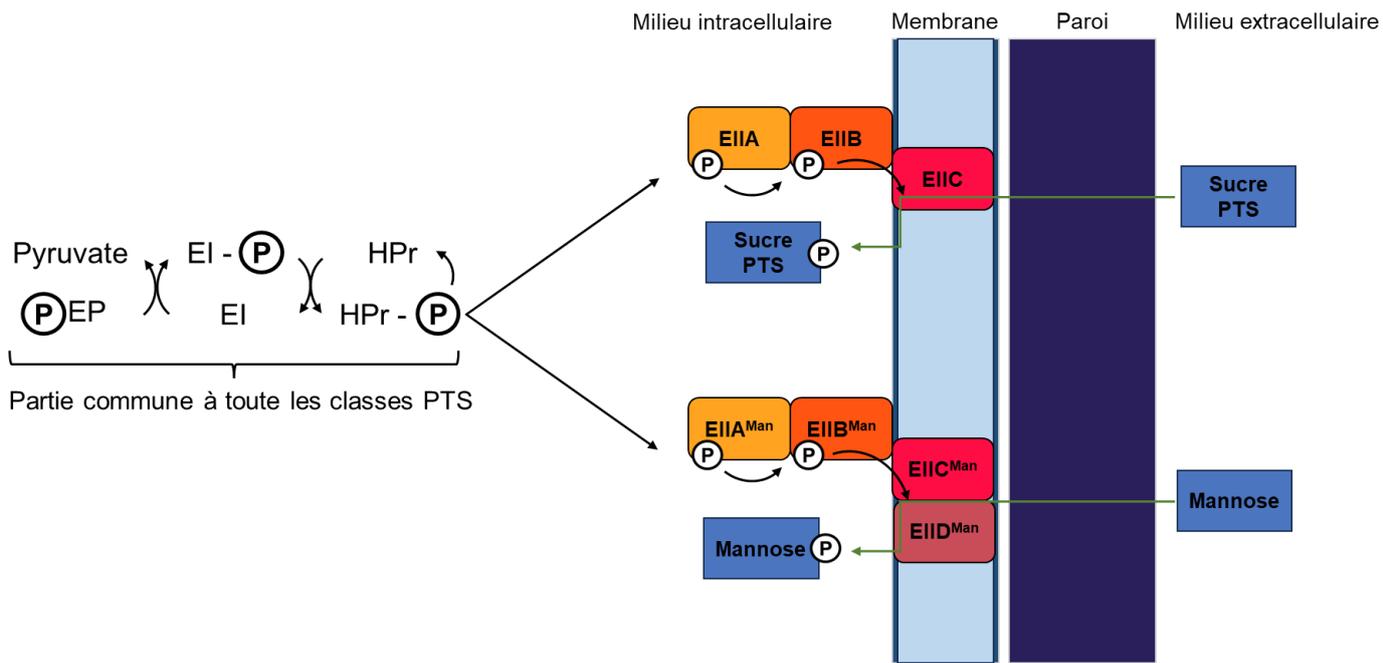


Figure 6 : Fonctionnement général des systèmes PTS. Les flèches noires indiquent un transfert du groupement phosphate. Les flèches vertes indiquent le transport du substrat.

1.4.2.2 Protéines EI et HPr

Les protéines EI et HPr sont codées par les gènes *ptsI* et *ptsH*, respectivement. Ces gènes ont été retrouvés chez différents *Bacillota* [81–84], et sont organisés en structure opéronique bicistronique [85]. Il existe toutefois des exceptions chez le genre *Bacillus* [86], où cet opéron contient en plus les gènes *ptsT* et *ptsG*. Ces gènes codent respectivement un anti-terminateur et une enzyme IICBA de la classe glucose.

La protéine EI d'*E. faecalis* a un poids moléculaire de 63,4 kDa et possède deux domaines fonctionnels (**Figure 7**).

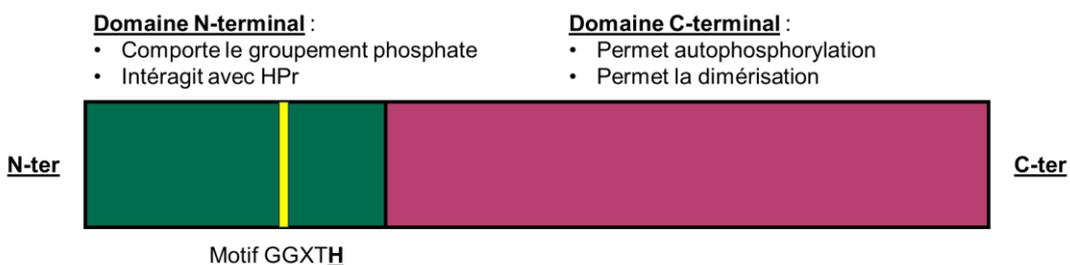


Figure 7 : Structure primaire de l'enzyme EI. Le domaine N-terminal est symbolisé en vert et le domaine C-terminal en violet. Le motif contenant l'histidine phosphorylée (His191 chez *E. faecalis*) est représenté en jaune.

Le domaine N-terminal est impliqué dans la phosphorylation de la protéine ainsi que dans l'interaction avec HPr. Le domaine C-terminal permet quant à lui l'autophosphorylation de EI lorsque la protéine est sous forme dimérique. La séquence d'EI est globalement conservée au sein d'une même espèce. Dans le domaine N-terminal, le motif GGXT[S,C]H[S,T,A]X[L,I,V,M]₂[A,S,T,G] est un motif retrouvé dans la séquence des protéines de la famille PPS-EI-PPDK, qui sont phosphorylées sur l'histidine de leur domaine central [87]. L'histidine de ce motif (His191 chez *E. faecalis*) est phosphorylable au niveau de son groupement azote N3 de l'anneau imidazole. L'autophosphorylation de EI a lieu lorsque la protéine se trouve sous forme dimérique [88,89] et en présence d'ion divalent (Mg^{2+} , Mn^{2+} , Zn^{2+} ou Co^{2+}) [76] et de PEP dans le milieu. L'enzyme EI sous sa forme homodimérique phosphorylée va céder le groupement phosphate à la protéine HPr.

HPr est une protéine monomérique d'un poids moléculaire de 9,3 kDa chez *E. faecalis*, qui est phosphorylée sur l'azote N1 de l'His15 transitoirement avant de transmettre le groupement phosphate à la protéine EIIA. La région autour de l'His15 et sa structure tridimensionnelle sont très conservées [90,91]. Cette protéine a une structure en sandwich β formée par 4 feuillets β antiparallèles et de 3 hélices α dont une plus courte que les deux autres, organisée sous la forme $\beta\alpha\beta\beta\alpha\beta\alpha$ (**Figure 8**).

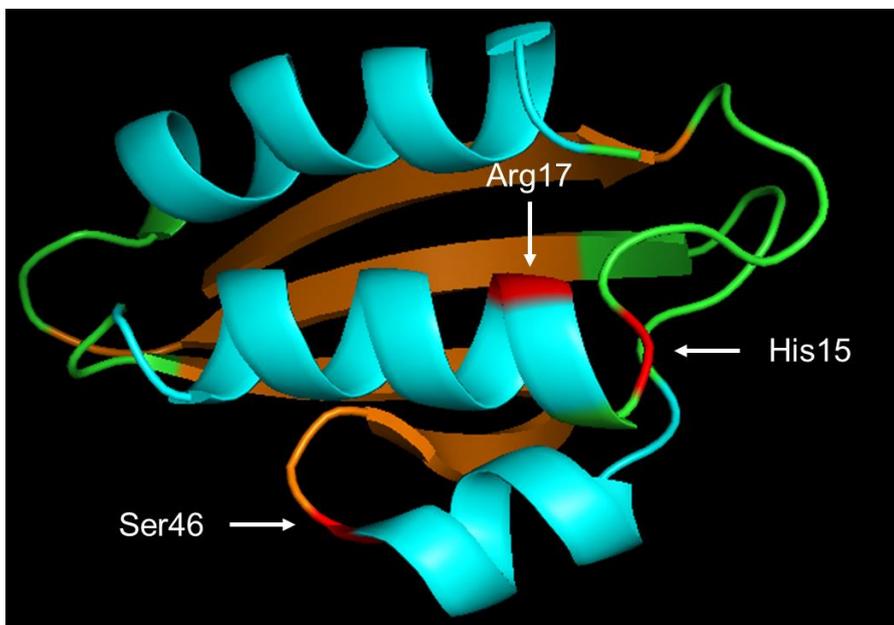


Figure 8 : Structure tridimensionnelle de la protéine HPr d'*E. coli*. Les hélices α sont représentées en bleu et les feuillets β en orange. Les AA d'intérêts sont indiqués en rouge. La structure tridimensionnelle de la protéine HPr est disponible sur la banque de structure tridimensionnelle de macromolécules MMDB. ID : 56311 [72].

La phosphorylation en His15 ne change que très peu la structure tridimensionnelle, mais permet à la protéine HPr d'interagir avec l'Arg17 du domaine EIIA pour transférer le groupement phosphate. Il a également été rapporté uniquement chez les *Bacillota* que HPr peut être phosphorylée sur son résidu Ser46 par l'enzyme HPrK/P qui est une kinase/phosphorylase [92]. Cette forme de phosphorylation sera décrite dans la partie « 1.4.2.5. Régulation du métabolisme par le PTS et autres fonctions ».

1.4.2.3 Les différentes classes de PTS

À ce jour, 7 classes de systèmes PTS ont été décrites ou identifiées *in silico* chez *E. faecalis* [77,93]. Ces familles PTS sont présentées en **Figure 9**, et seules les PTS des classes glucose et lactose seront discutés dans ce manuscrit.

PTS de la classe glucose : *E. faecalis* possède six systèmes PTS de la classe glucose, qui permettent l'import de différents β -Glucosides tels que le glucose, le GlcNac, le maltose, le sucrose et le tréhalose [93]. Un exemple pouvant être donné est celui de l'import de maltose (**Figure 10**). L'expression de l'opéron *malt-mapP* est régulée par le régulateur MalR dont le gène se trouve à proximité dans l'opéron *malPBMR*. En absence de maltose, MalR agit comme un répresseur. En présence de maltose, ce composé est internalisé et phosphorylé par la protéine PTS EII CBA MalT, puis déphosphorylé par l'enzyme MapP. Sous sa forme non phosphorylé ce sucre est clivé en glucose et glucose-1P par MalP. Ce dernier est converti en glucose-6P par MalB avant d'entrer dans la voie de la glycolyse [94]. En présence de glucose, MalR exerce sur ces deux structures opéroniques une répression catabolique indépendante de CcpA (Catabolite control protein A) en interagissant avec P-Ser-Hpr [94].

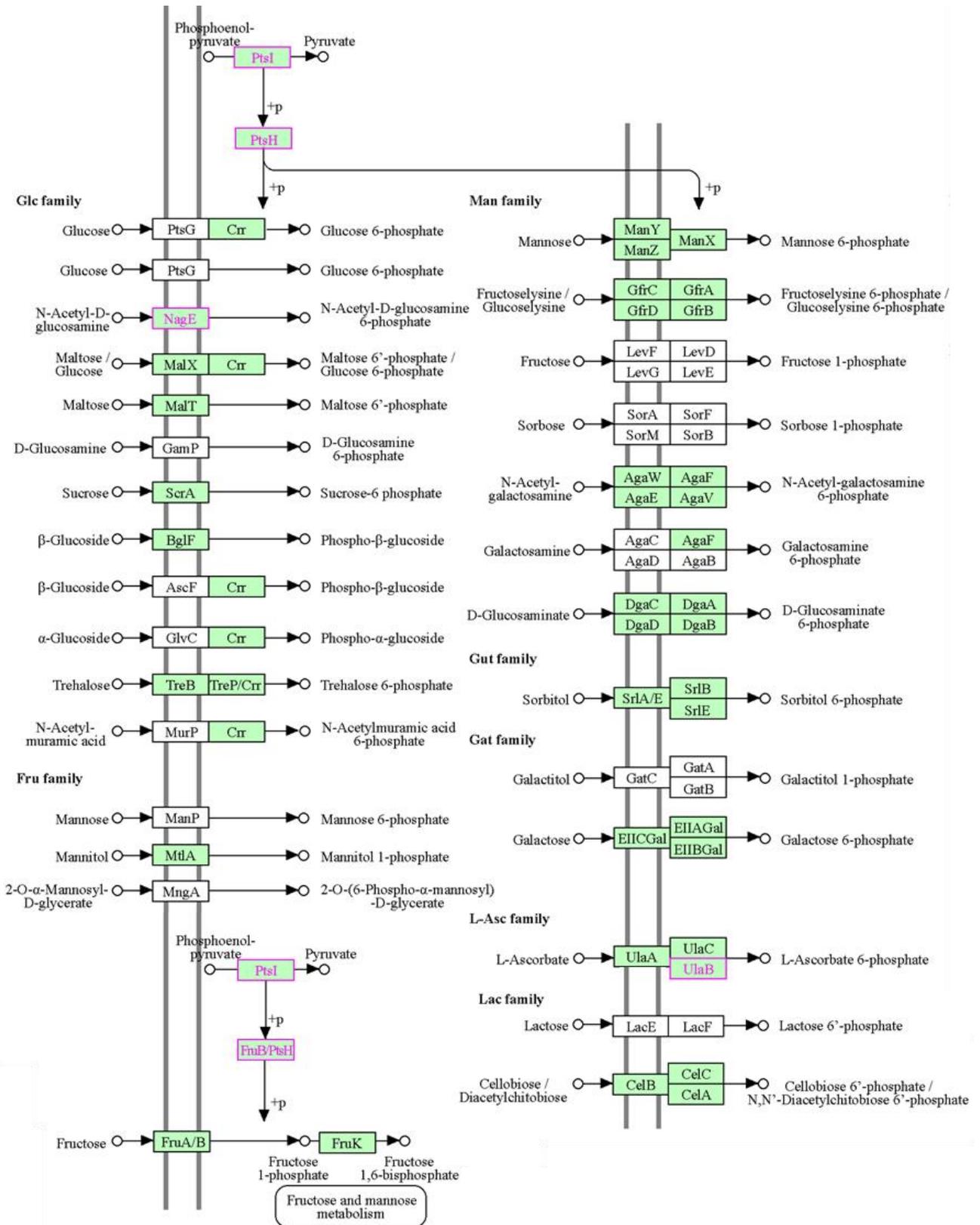


Figure 9 : Représentation des différentes classes de PTS chez *E. faecalis* OG1RF. Les cases vertes représentent les PTS identifiés chez *E. faecalis* OG1RF (Base de donnée KEGG PATHWAY) [93].

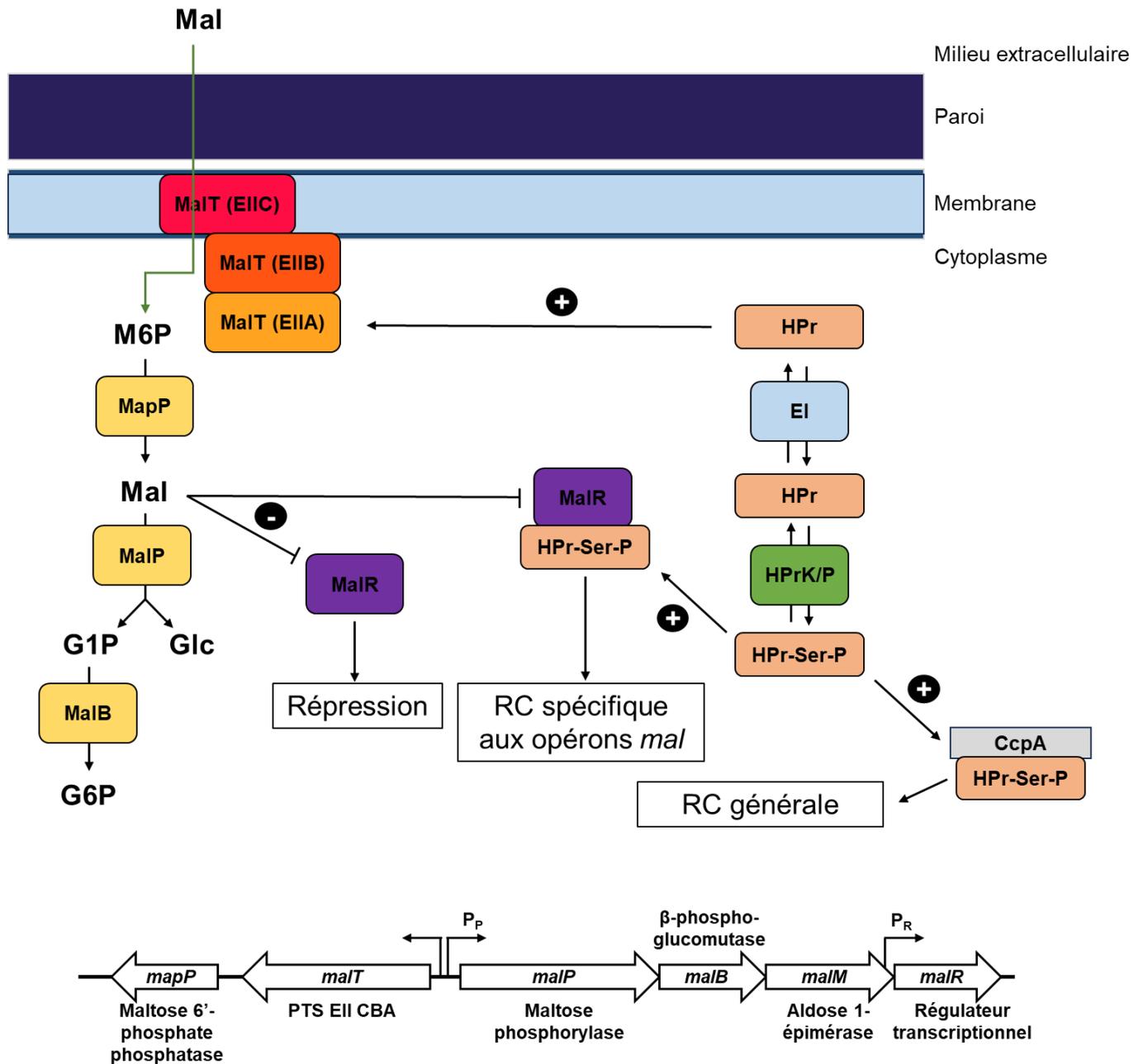


Figure 10 : Exemple du fonctionnement du PTS glucose permettant l'import du maltose chez *E. faecalis* OG1RF. Mal : Maltose ; M6P : Maltose-6P ; G1P : Glucose-1P ; G6P : Glucose-6P ; Glc : Glucose (adapté de Grand et al., 2020) [94].

PTS de la classe lactose : *E. faecalis* possède 13 gènes annotés comme faisant partie des PTS de la classe lactose/cellobiose [93]. Un exemple peut être donné avec le gène *genB* codant un PTS EIIC^{Cel} (**Figure 11**) [5].

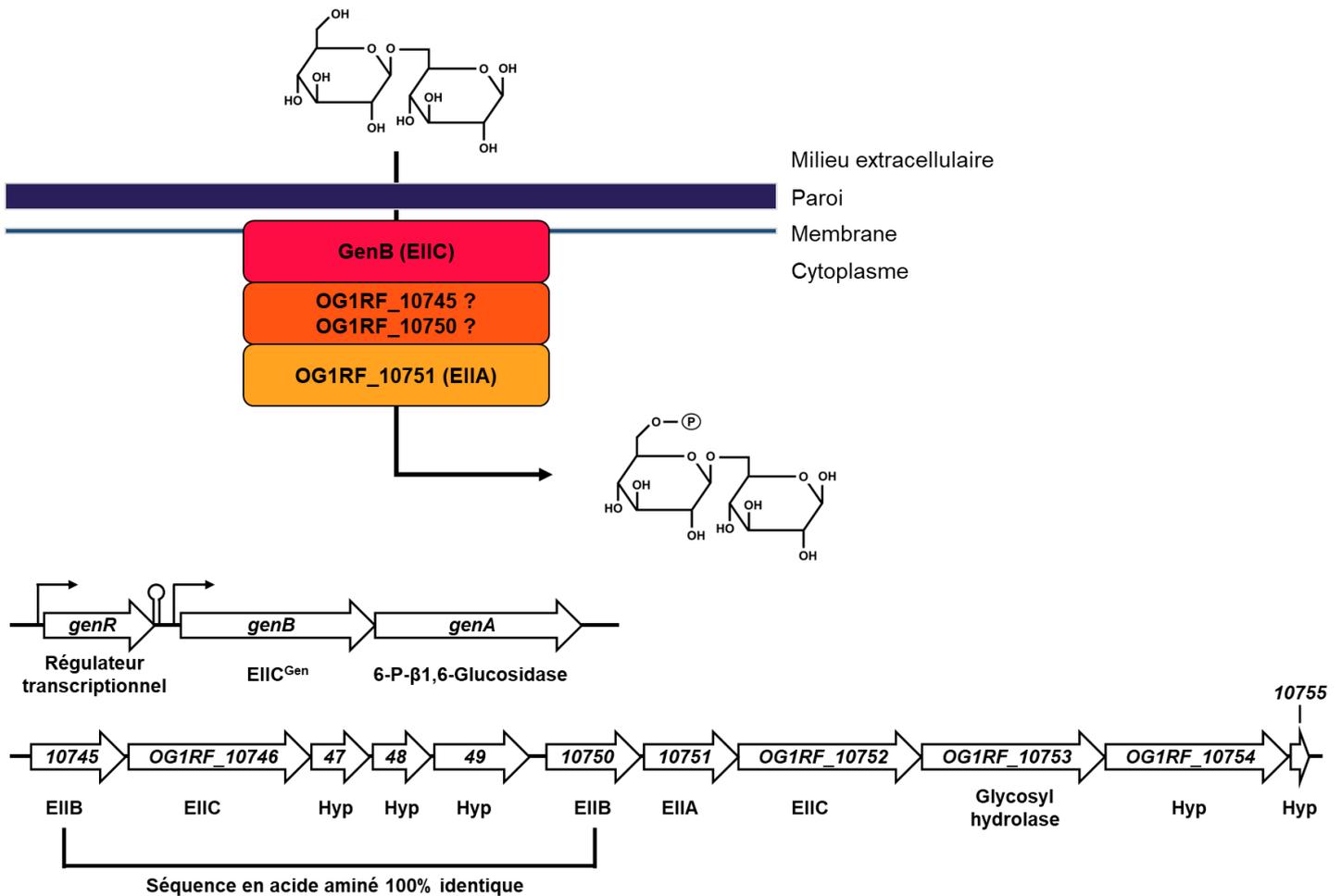


Figure 11 : Exemple du fonctionnement du PTS cellobiose permettant l'import de gentiobiose chez *E. faecalis* OG1RF (adapté de Grand et al., 2019) [5].

Ce gène se trouve au sein de l'opéron bicistronique *genBA*. En aval du gène *genB* codant pour la protéine PTS EIIC_{Cel}, se trouve le gène *genA* codant une 6-phospho- β -glucosidase. Un gène présent en amont code pour un activateur transcriptionnel GenR régulant l'expression de cet opéron. De plus, cette structure opéronique est également soumise à la répression catabolique médiée par le complexe P-Ser-HPR/CcpA. Ce transporteur PTS n'est pas le transporteur principal permettant l'utilisation du cellobiose, mais est impliqué dans l'utilisation du gentiobiose, un diholoside composé de deux unités de D-glucose reliées par une liaison osidique du type β (1 \rightarrow 6) et de l'amygdaline. Cet opéron et sa région

chromosomique voisine ne comportent pas de gènes codant des protéines EIIA et EIIB. Cependant pour être fonctionnel, GenB doit interagir avec un PTS EIIA et EIIB. L'analyse du génome d'*E. faecalis* a permis d'identifier une région (OG1RF_10745 à OG1RF_10755) comportant 2 EIIC, 2 EIIB et 1 EIIA de la classe cellobiose (**Figure 11**). Cet EIIA^{Cel} (OG1RF_10751) est le seul retrouvé chez *E. faecalis*. Une souche d'*E. faecalis* mutée pour cet EIIA perd sa capacité à utiliser le gentiobiose, l'amygdaline et le cellobiose comme source de carbone [5]. Ceci suggère que ce PTS EIIA est partagé entre différents complexes PTS de la classe cellobiose. Comme il sera vu plus tard, un PTS la classe Lactose/Cellobiose sera traitée dans la partie résultat ainsi que du rôle de la protéine OG1RF_10751 dans ce PTS.

1.4.2.4 Protéines EII

Les PTS nécessitent la présence d'enzymes spécifiques à l'import d'un substrat donné. Cette spécificité est portée par les protéines/domaines EII qui forment un complexe multi-enzymatique constitué d'une part des domaines EIIA et EIIB cytoplasmiques, et du domaine EIIC (et de EIID dans le cas du mannose) transmembranaire. Les domaines PTS EIIA, EIIB et EIIC sont essentiellement retrouvés de façon individuelle, ou bien fusionnés entre eux de sorte à former une protéine multi-domaine (**Tableau 3**). Bien que plus rare, il est également décrit dans la littérature l'association entre des domaines spécifiques (EII) et non spécifiques (EI et HPr) en une seule protéine à multi-domaine [77].

Tableau 3 : Fusion des différents domaines PTS (adapté de Barabote et Saier, 2005) [77].

Domaines fusionnés	Fusion des domaines des différentes classes PTS ^a .							
	Glc	Fru	Lac	Gut	Gat	Man	Asc	Dha
EI-IIA	-	1	-	-	-	-	-	-
HPr-IIA	-	3	-	-	-	-	-	-
IIA-HPr	-	10	-	-	-	-	-	2
HPr-EI-IIA	-	7	-	-	-	-	-	-
IIA-HPr-EI	8	10	-	-	-	-	-	3
IIA	54	101	76	20	25	81	47	26
IIB	15	26	81	-	25	56	45	42
IIC	7	30	98	16	31	83	49	47
IID	-	-	-	-	-	83	-	-
IIAB	-	-	-	-	-	28	4	-
IIBA	1	-	-	-	-	-	-	-
IIBC	75	37	-	16	-	-	-	-
IICB	80	21	8	-	-	-	7	-
IICD	-	-	-	-	-	2	-	-
IIABC	4	47	-	-	-	-	-	-
IIACB	-	3	-	-	-	-	-	-
IIBCA	56	4	-	-	-	-	-	-
IICBA	36	15	-	-	-	-	-	-

^a : Glc, Fru, Lac, Gut, Gat, Man, Asc et Dha indiquent respectivement les familles PTS glucose, fructose, lactose, glucitol, galactitol, mannose, ascorbate et dihydroxyacétone. Les nombres indiquent la proportion de chaque domaine trouvé dans les 106 espèces bactériennes de l'étude de Barabote et Saier, 2015) qui possèdent des homologues PTS. L'ordre des domaines est celui indiqué dans la première colonne.

Les systèmes PTS peuvent être classés selon les critères suivants : spécificité des substrats, homologie des séquences des complexes EII, et organisation des différents domaines. Les EII d'un même groupe partagent généralement un minimum de 25% d'identité de séquence [78]. Les gènes codants les EII pour une même classe de PTS sont le plus souvent organisés sous forme de structure opéronique comportant, en plus, un ou plusieurs gènes impliqués dans le métabolisme du sucre transporté, et un régulateur transcriptionnel. Le régulateur permet l'expression des gènes de l'opéron en présence du substrat qui est généralement le sucre en question. Il est possible que des domaines EII fonctionnant ensemble soient issus de gènes se trouvant sur des régions éloignées sur le chromosome [78,95]. Sur les 82 gènes annotés PTS chez *E. faecalis*, 70 codent des protéines EII dont 34 EIIC uniques ou fusionnés, démontrant l'importance de la recherche de substrat pour la bactérie. Toutefois, il est encore compliqué de prédire *in silico* la spécificité des substrats des domaines EIIC.

Le domaine EIIC, contrairement à HPr et aux domaines EIIB et EIIA, permet l'utilisation spécifique d'un substrat. C'est *via* la protéine EIIC que le sucre migre du milieu extracellulaire vers le cytoplasme bactérien. Lorsque le substrat est lié à la protéine ou au domaine EIIC, le groupement phosphate alors porté par le domaine EIIB est transféré au sucre, ce qui diminue l'affinité de ce dernier à EIIC et permet ainsi de libérer le substrat phosphorylé dans le cytoplasme bactérien. Sous sa forme phosphorylée, la polarité du substrat est augmentée, ce qui l'empêche de diffuser à travers la membrane [96]. Ce sucre phosphorylé sera ensuite métabolisé par la bactérie.

1.4.2.5 Régulation du métabolisme par le PTS, et autres fonctions

Bien que la fonction principale des PTS soit d'internaliser et de phosphoryler des sources de substrats, d'autres fonctions lui ont été décrites. En fonction de leur état de phosphorylation, les protéines des PTS peuvent exercer des fonctions régulatrices [95,97].

En présence de plusieurs sources de substrats métabolisables dans le milieu, la cellule va préférentiellement utiliser le substrat le plus rapidement métabolisable et réprimer l'expression des gènes impliqués dans l'utilisation d'une autre source de carbone. Ce phénomène se nomme Répression Catabolique (RC). Chez les *Bacillota*, la RC est médiée par les protéines HPr, HPrK/P, CcpA et les sites *cre* (catabolite responsive element). Nous avons décrit précédemment que la protéine HPr peut être phosphorylée sur l'His15, et que cette phosphorylation sera par la suite transférée au domaine EIIA. La protéine HPr possède un autre site de phosphorylation réversible sur sa sérine en position 46 (Ser46). Cette phosphorylation est réalisée par l'enzyme Kinase/Phosphorylase HPrK/P [98]. Chez *E. faecalis*, cette protéine est constituée de 311 AA et possède deux domaines [98] : i) le domaine N-terminal qui permet la fixation de deux ions phosphates [99], et ii) le domaine C-terminal qui comporte les activités kinases et phosphorylases. Un motif Walker A ((G/A)xxxxGK(S/T)) y est également retrouvé, et permet de former une P-loop qui interagira avec le phosphate de l'ATP [99]. Cette protéine HPrK/P possède une structure quaternaire hexamérique pouvant fixer jusqu'à 6 molécules HPr en simultanément [98]. La présence de fructose-1,6-bisphosphate (FBP) et d'ATP dans le milieu va stimuler l'activité kinase de HPrK/P sur son substrat HPr (**Figure 12**) [100,101].

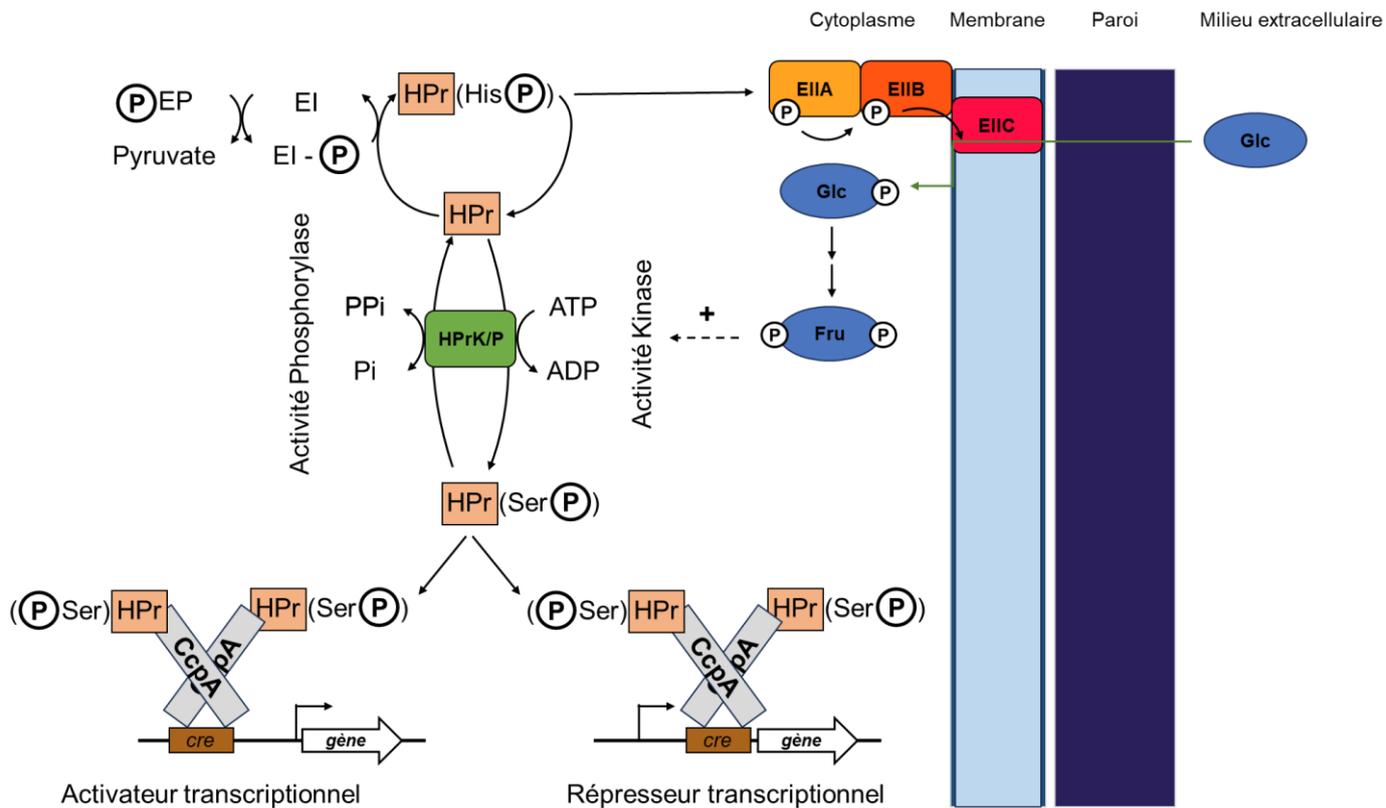


Figure 12 : Fonctionnement et implication de l'activité kinase et phosphorylase de l'enzyme HPrK/P. Les flèches pleines indiquent le déroulé de la réaction. La flèche en pointillée indique que le Fructose-1,6-bisphosphate stimule l'activité kinase de HPrK/P (adapté de Stülke et Schmalisch, 2004) [102].

Il est toutefois notable que le FBP n'est pas consommé dans la réaction, mais que son rôle se limite à stimuler HPrK/P. En présence de phosphate inorganique, l'activité phosphorylase sera exercée et catalysera le départ du groupement phosphate de la Ser46 et libérera un pyrophosphate. Ce dernier pourra être utilisé par HPrK/P pour phosphoryler la Ser46 si les conditions du milieu sont propices [99]. Il a été montré que la participation de HPr-Ser-P dans la cascade de phosphorylations des PTS est anecdotique [103]. La protéine HPr-Ser-P ne permet donc pas le transport de substrat, mais plutôt la régulation de l'activité et/ou de l'expression par l'intermédiaire de différents acteurs cellulaires, principalement la protéine CcpA. Cette protéine de 36,2 kDa chez *E. faecalis* est un régulateur transcriptionnel de la famille LacI/GaIR reconnaissant des séquences ADN appelées boîtes *cre*. Ces motifs ADN sont des palindromes imparfaits de 15 pb. Ils ont pour séquence consensus chez *E. faecalis* 5' – WTGWAARCGYWWWCW – 3' [104,105]. CcpA ne peut cependant pas se fixer seul sur les motifs *cre*, il doit obligatoirement former un complexe avec la protéine HPr-Ser-P. La

formation du complexe HPr-Ser-46 et CcpA permet la dimérisation de CcpA. Cette forme dimérique active est capable de se lier à l'ADN [106,107]. La localisation du complexe permet de déterminer s'il y aura une RC ou une activation catabolique. Le complexe peut être un activateur si la boîte *cre* se trouve en amont du promoteur, ou un répresseur si la boîte *cre* se trouve au sein du promoteur ou en aval de ce dernier [108]. L'affinité de ce complexe au motif *cre* peut être modulée : des études ont montré que les complexes HPr-Ser-P/CcpA auxquels sont liés les molécules FBP ou nicotinamide adénine dinucléotide phosphate (NADP+) ont une affinité à l'ADN 120x et 300x plus forte respectivement. Ces métabolites stimulent le contact entre HPr-Ser-P et CcpA, ce qui favorise la formation de complexes stables [109,110].

Les systèmes PTS peuvent également réguler le métabolisme par des mécanismes d'exclusion et d'expulsion de l'inducteur (**Figure 13**). Ce premier mécanisme est réalisé par la protéine HPr-Ser-P qui interagit avec des transporteurs ABC, inhibant ces derniers et empêchant l'entrée des substrats [97,111]. Le second mécanisme consiste à rejeter dans le milieu extracellulaire les sucres non métabolisables présents dans le cytoplasme de la bactérie lorsque celle-ci se trouve en présence d'un sucre facilement métabolisable comme le glucose. Une étape préalable de déphosphorylation des sucres peut avoir lieu [112–116].

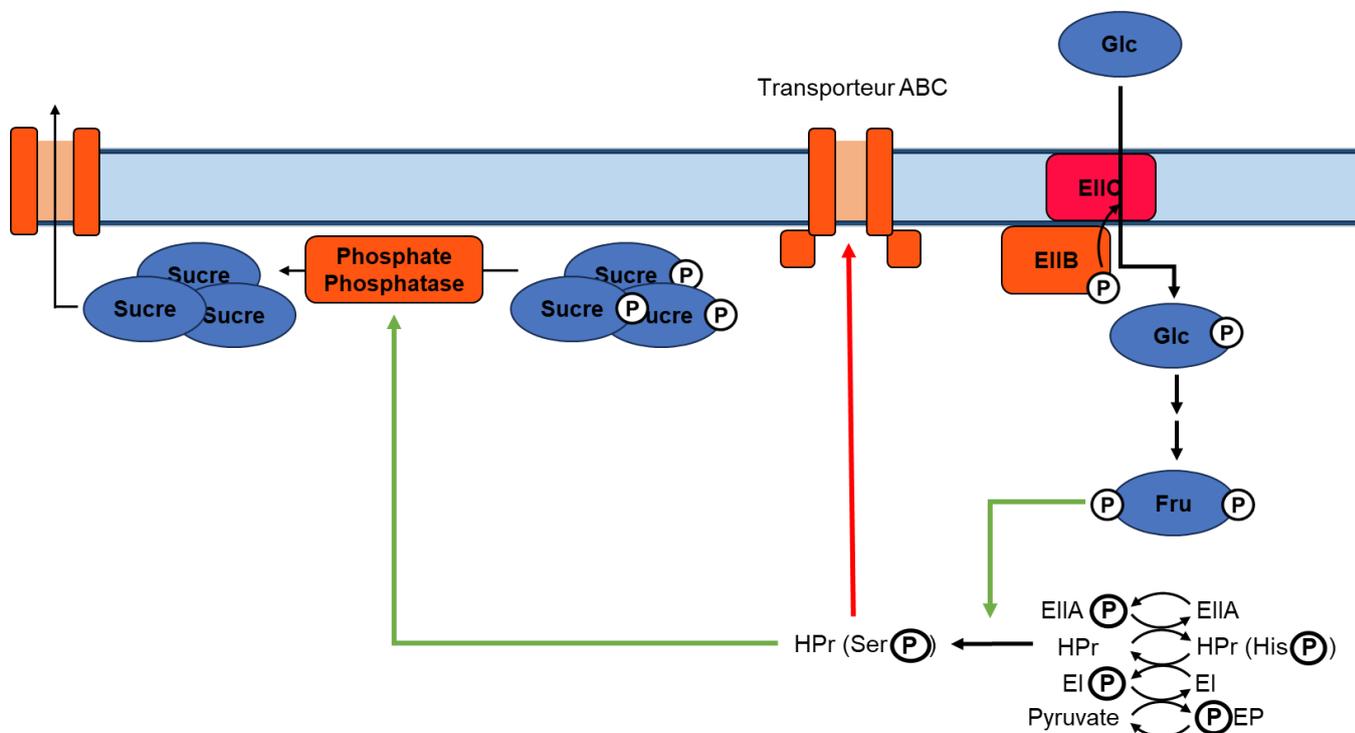


Figure 13 : Exclusion et expulsion de l'inducteur. Les flèches vertes indiquent la stimulation de l'activité. La flèche rouge indique une inhibition.

En plus de leur rôle dans l'import du substrat, les protéines HPr-His-P, EIIA-His-P et EIIB-His-P exercent un rôle de régulation *via* la phosphorylation de protéines non-PTS, qui sont majoritairement des régulateurs (**Figure 14**), des transporteurs ou des protéines de fonction encore non identifiée [77]. Parmi les régulateurs, on retrouve des activateurs transcriptionnels comme la protéine MltR chez *B. subtilis*, ou bien des anti-terminateurs comme la protéine LicT également retrouvée chez *B. subtilis* [117,118]. Ces protéines possèdent dans leurs régions N- ou C-terminales des domaines PRD (PTS Regulation Domain) de 170 AA, organisés sous forme de dimère, et possédant 1 à 2 histidines phosphorylables par les acteurs protéiques du PTS [96,119]. Ils sont très fréquemment retrouvés chez les *Bacillota*, et permettent une régulation ciblée de certains gènes du métabolisme, dont ceux des PTS [97]. En fonction de la phosphorylation des histidines des régulateurs à domaine PRD, les changements structuraux associés ont pour conséquence la stimulation ou l'inhibition de l'activité de la protéine. La phosphorylation d'un domaine PRD par HPr-His-P permet de stimuler son activité. Cependant, en présence d'un sucre facilement métabolisable comme le glucose, HPr-His-P transférera préférentiellement le groupement phosphate au EIIA du sucre correspondant et non plus au domaine PRD du régulateur, induisant ainsi son inactivation [120]. Lorsque le substrat spécifique des protéines EIIA et EIIB est absent, ces dernières peuvent transférer leur groupement phosphate aux domaines PRD des régulateurs. La phosphorylation par EIIA et EIIB inhibe l'activité de ces régulateurs [120].

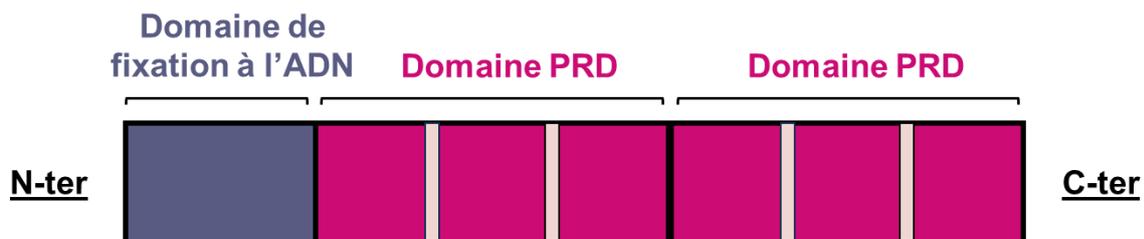


Figure 14 : Structure des domaines PRD. Les domaines d'intérêt sont représentés par des rectangles. Le domaine de liaison aux acides nucléiques à l'extrémité N-terminale est symbolisé en bleu, les domaines PRD sont représentés en violet. Les AA phosphorylables au sein des PRD sont indiqués par des rectangles clairs. A. Schéma de la structure de la protéine LicT de *B. subtilis*, un régulateur à domaines PRD fonctionnant comme un anti-terminateur (adapté de Grand, 2019) [121].

Enfin, la protéine HPr existe sous la forme doublement phosphorylée P-Ser-HPr-His-P [77,122]. Chez *Streptococcus*, la part de forme doublement phosphorylée varie entre 25%-75% selon la souche sur l'ensemble des 4 formes de phosphorylation possibles pour la

protéine HPr. Sous cette forme doublement phosphorylée, la protéine serait capable de transférer son groupement phosphate His15 aux protéines EIIA [122]. Cependant, la forme doublement phosphorylée chez *E. faecalis* n'est que très anecdotique car ne représente que $2 \pm 2\%$ de l'ensemble des formes retrouvées. Cela est dû au fait que la phosphorylation de l'His15 de HPr est diminuée d'un facteur 5000 si la Ser46 de la protéine est déjà phosphorylée [101].

1.5 Métabolisme du N-acétylglucosamine

Le N-acétylglucosamine (GlcNAc) est un monosaccharide dérivé du glucose. Ce composé peut se polymériser en établissant des liaisons β 1-4 et former des dimères (chitobioses) ou bien des chaînes linéaires (chitine). La chitine est le second polysaccharide naturel le plus abondant après la cellulose [123]. Le GlcNAc est retrouvé dans diverses structures comme les parois cellulaires des champignons, les exosquelettes des arthropodes, la muqueuse digestive, la muréine, l'acide hyaluronique ou encore les protéines glycosylées [123,124]. Les bactéries métabolisent ce composé en tant que source de carbone et d'azote [124]. Pour le métaboliser, *E. faecalis* fait appel à différents gènes et structures opéroniques, et procède en différentes étapes (**Figure 15**).

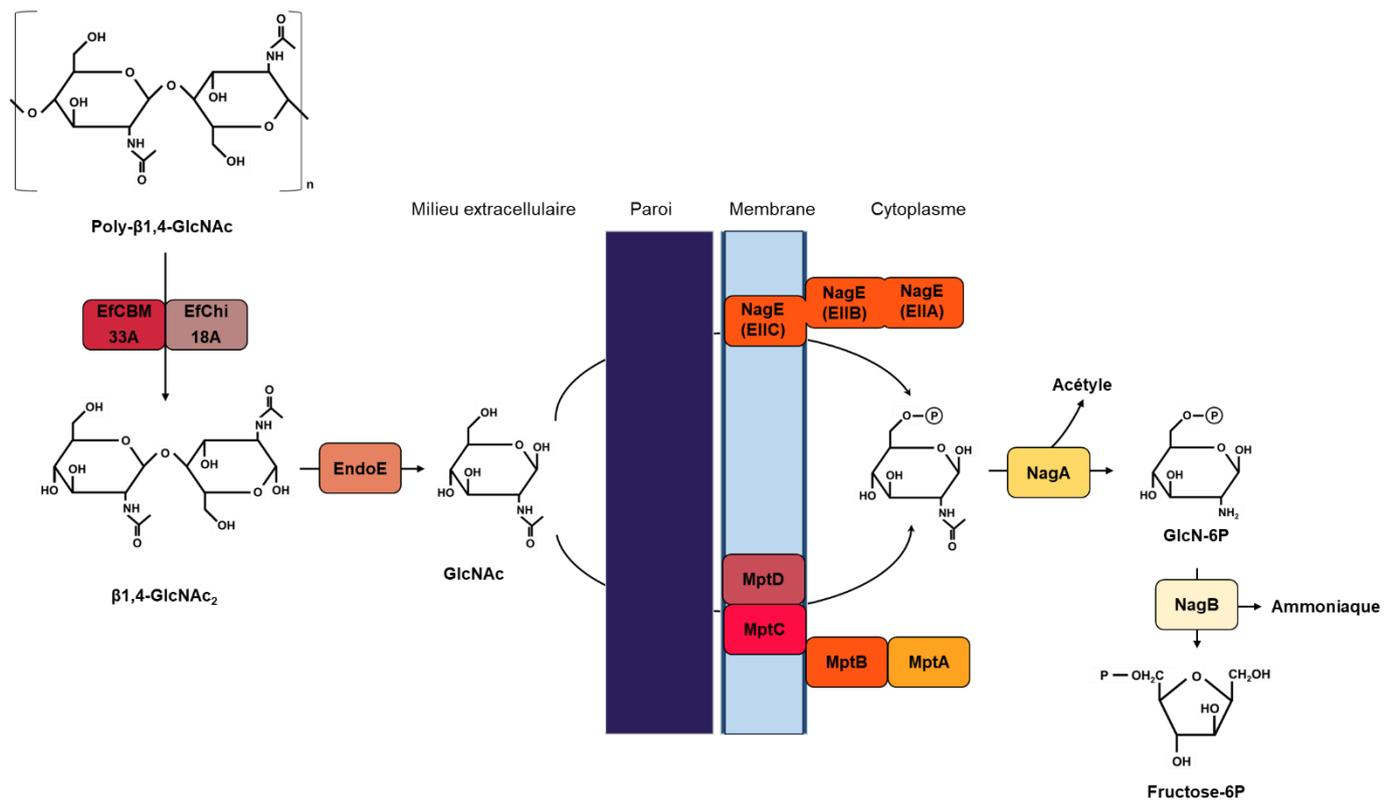


Figure 15 : Métabolisme de la N-acétylglucosamine chez *E. faecalis*.

Dans un premier temps, la bactérie va dégrader dans le milieu extracellulaire des composés comportant des résidus GlcNAc. L'activité chitinolytique d'*E. faecalis* a été caractérisée et permet de fournir un bon exemple du métabolisme du GlcNAc. *E. faecalis* possède un opéron comportant les gènes *ef0362* et *ef0361*, codant respectivement pour les enzymes EfcBM33A et EfChi18A sécrétées dans le milieu extracellulaire et étant impliquées dans la dégradation de la chitine. EfChi18A dégrade la chitine par hydrolyse alors qu'EfcBM33A dégrade la chitine par un clivage oxydatif nécessitant du cuivre comme co-facteur. Il a été montré qu'EfChi18A est capable d'hydrolyser des résidus chitines comportant 6 résidus GlcNAc, mais qu'elle utilise préférentiellement des substrats comportant 3 à 5 résidus GlcNAc [125]. Les deux enzymes agissent synergiquement en dépolymérisant la chitine en GlcNAc et GlcNAc₂. Le chitobiose est ensuite dégradé par l'enzyme EndoE en deux résidus GlcNAc [124]. Cette enzyme, qui sera décrite ultérieurement, comporte un domaine GH20 à l'extrémité C-terminale catalysant le clivage entre deux résidus GlcNAc adjacents.

Dans un second temps, le GlcNAc extracellulaire va être internalisé par *E. faecalis*. Chez *B. subtilis*, le GlcNAc est internalisé par le PTS NagP alors que chez *S. mutans* et *S. pneumoniae* il est pris en charge par le PTS glucose/mannose ManLMN [124]. Deux transporteurs PTS

présentant des similarités de séquences ont été identifiés chez *E. faecalis* : le PTS MptBACD et le PTS NagE. Il a été démontré chez *E. faecalis* que ces PTS permettent l'internalisation et la phosphorylation du GlcNAc extracellulaire [124]. Le gène *nagE* code un PTS EIICBA dans lequel les trois domaines sont fusionnés, et se trouve au sein d'une structure opéronique contenant un anti-terminateur transcriptionnel à domaine PRD, *nagY*. En présence de GlcNAc, les gènes de ce locus sont surexprimés, permettant l'expression du PTS NagE et de l'anti-terminateur. Des travaux actuellement menés au laboratoire CBSA par Soussan et collaborateurs ont montré que l'anti-terminateur NagY régule sa propre structure opéronique, mais également l'expression du gène *hylA* codant une hyaluronidase. Dans cette étude, les auteurs ont montré que cette hyaluronidase est capable de dégrader l'acide hyaluronique, un polysaccharide composé d'une répétition de motifs D-glucuronique β 1-3 N-acétylglucosamine liés par liaison β 1-4.

Une fois transporté et phosphorylé par les systèmes PTS, le GlcNAc-6P va subir des modifications avant de rejoindre la voie de la glycolyse. L'enzyme NagA va dans un premier temps soustraire le groupement acétyle du GlcNAc-6P, puis l'enzyme NagB va désaminer le GlcN-6P en Fructose-6P et ammoniacque (**Figure 15**) [124].

2 Protéines glycosylées et enzymes déglycosylantes

Les protéines nouvellement formées par le ribosome ne sont pas toujours totalement fonctionnelles. Pour le devenir totalement, elles peuvent subir une phase de maturation pendant et après la synthèse. La plupart des processus biologiques des cellules sont réalisés par des protéines ayant subi cette phase de maturation [126]. Ces modifications sont de natures variées : clivage, modification directe d'AA, ajout de groupements chimiques, de polypeptides, ou de molécules complexes [127,128]. Ces modifications permettent de réguler l'activité enzymatique, de modifier la conformation tridimensionnelle de la protéine, de la faire interagir avec des partenaires spécifiques, ou encore d'influer sur sa localisation cellulaire [128]. Bien que plus de 400 modifications aient été rapportées [126], nous nous focaliserons exclusivement sur la glycosylation des protéines dans ce manuscrit.

2.1 Protéines glycosylées

La modification la plus fréquemment retrouvée est la glycosylation des protéines. Il était admis que ces glycosylations n'étaient pas présentes chez les microorganismes, mais on sait maintenant qu'elles sont retrouvées chez les procaryotes et les eucaryotes unicellulaires et multicellulaires [129]. Il est estimé que 50% des protéines produites seraient glycosylées chez les eucaryotes [130]. Il est possible qu'une même protéine porte plusieurs glycosylations à des sites différents et qu'il peut y avoir différentes structures oligosaccharidiques pour un même site de glycosylation (glycoformes) [131]. Les principaux types de glycosylations sont la N-glycosylation et la O-glycosylation, où des résidus osidiques sont respectivement greffés à une asparagine, et à une sérine (Ser) ou une thréonine (Thr). Un troisième type de glycosylation plus anecdotique est la C-glycosylation, qui ne se sera pas développée dans ce manuscrit.

2.1.1 N-Glycosylations chez les eucaryotes

La N-glycosylation est une structure oligosaccharidique greffée à un résidu asparagine dans une séquence Asn-Xxx-Ser/Thr d'une protéine, avec Xxx pouvant être n'importe quel AA excepté la proline. Cette structure est constituée de Glc, Man et de GlcNAc. Cette modification est une modification co-translationnelle et post-translationnelle. Elle débute dans le réticulum endoplasmique (RE) et se poursuit dans l'appareil de Golgi. La synthèse de protéines N-glycosylées est un processus hautement conservé avec deux phases (**Figure 16**) [6].

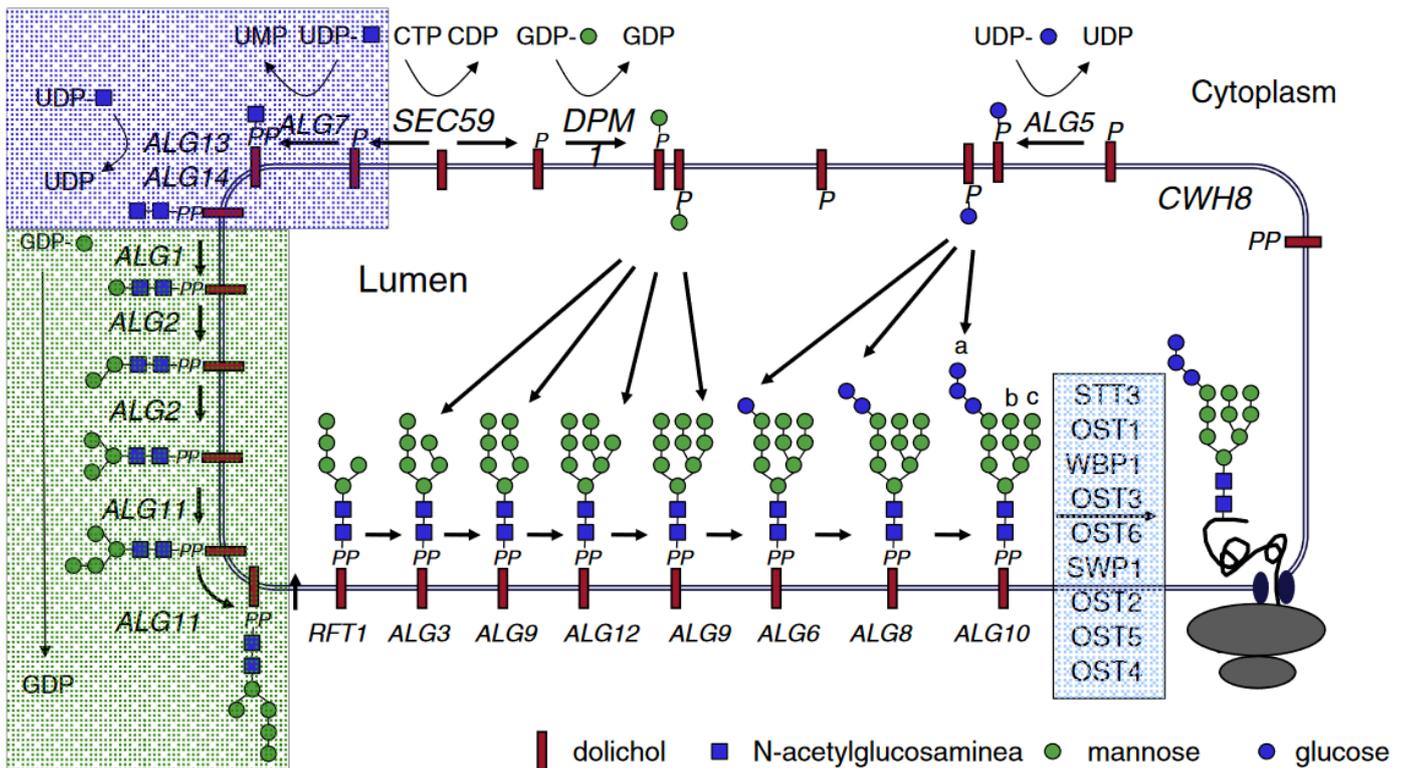


Figure 16 : Voie de biosynthèse des oligosaccharides liés aux lipides et transfert de l'oligosaccharide à la protéine au niveau de la membrane du réticulum endoplasmique. La topologie de la biosynthèse des oligosaccharides liés aux lipides catalysée par les glycosyltransférases codées par les différents loci ALG est présentée. Les glycosyltransférases actives du côté cytoplasmique de la membrane du RE qui utilisent l'UDP-GlcNAc et le GDP-Man, respectivement, forment deux complexes enzymatiques indiqués par un carré bleu et un rond vert. Après la translocation de l'oligosaccharide Man₅GlcNAc₂ lié aux lipides, une extension supplémentaire par des glycosyltransférases orientées vers le lumen produit l'oligosaccharide Glc₃Man₉GlcNAc₂ (a, b et c se réfèrent aux trois antennes de l'oligosaccharide) qui est transféré à des résidus Asn sélectionnés de polypeptides transloqués par le complexe d'oligosaccharyltransférases (Aebi, 2013) [6].

La première phase consiste en la synthèse d'un oligosaccharide greffé à un lipide. Des glycosyltransférases (ALG pour asparagin-linked glycosylation) interviennent tout au long de ce processus pour ajouter de nouveaux résidus à la glycosylation en cours de synthèse (**Figure 16**). Différentes glycosyltransférases ALG interviennent successivement pour, à terme, former une chaîne oligosaccharidique GlcNAc₂Man₅. Cet oligosaccharide sera formé sur du dolichol phosphorylé (Dol-P) retrouvé ancré dans la membrane du RE. L'enzyme RFT1, une flippase, permet de faire passer le glycolipide (Dol-P-P-GlcNAc₂Man₅) de la membrane extérieure vers la lumière du RE. Ce glycolipide est transformé en Dol-P-P-GlcNAc₂Man₉Glc₃ avant d'être transféré par le complexe oligosaccharyltransférase (OST) à la protéine naissante au niveau d'un résidu Asn-Xxx-Ser/Thr. Le transfert de l'oligosaccharide a lieu lorsque le résidu

asparagine de la protéine naissante traverse le translocon dans le RE. Cependant, un motif Asn-Xxx-Ser/Thr n'est pas forcément glycosylé. Il est possible que la glycosylation du motif soit ignorée s'il se trouve proche de la séquence N-terminale ou C-terminale, ou si deux motifs soient trop proches ou à proximité d'une cystéine [6,132].

Après le transfert de l'oligosaccharide à la protéine naissante, une étape de vérification de la conformation a lieu. Un faible nombre de défauts de repliement conduira la glycoprotéine dans un cycle de « reglycosylation/déglycosylation » alors qu'un fort nombre de défauts de repliement conduira la glycoprotéine à être dégradée par le système ERAD (endoplasmic reticulum associated protein degradation). Dans le cas où aucun défaut de repliement n'est détecté, la glycoprotéine sera transportée par des vésicules vers l'appareil de Golgi pour sa maturation. Cette dernière consiste au retrait de résidus Man et l'addition d'autres monosaccharides (GlcNAc, Gal, Fuc et acide sialique) par l'action de plusieurs glycosyltransférases et glycoside hydrolases [6,132]. Cette maturation conduit à la formation de 3 types de chaînes oligosaccharidiques (**Figure 17**) :

- Oligosaccharide riche en Man : seuls des résidus Man sont présents en plus de la structure de base.
- Oligosaccharide hybride : contient des résidus Fuc, Gal, Glu, Man et GlcNAc.
- Oligosaccharide complexe : semblable au type hybride, contient en plus des résidus N-acétylneuraminique (acide sialique).

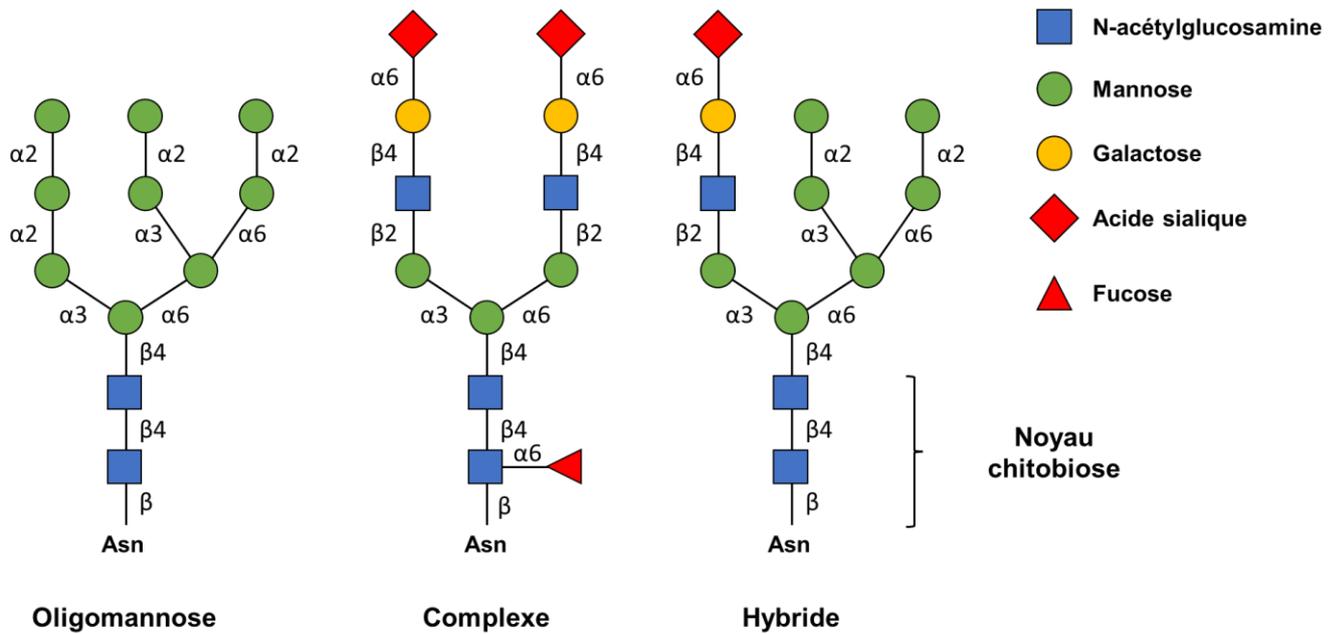


Figure 17 : Représentation de différentes chaînes N-glycanes. Les N-glycanes des séquences Asn-X-Ser/Thr dans les glycoprotéines des eucaryotes sont de trois types généraux : oligomannose, complexe et hybride. Chaque N-glycane contient le noyau commun $Man_3GlcNAc_2Asn$. Les N-glycanes complexes peuvent avoir jusqu'à six branches initiées par GlcNAc et chacune peut être allongée avec des répétitions $Gal\beta 1-4GlcNAc$ (LacNAc) (adapté de Stanley et al., 2015) [133].

2.1.2 O-Glycosylations chez les eucaryotes

Un autre type de modification des protéines fréquemment retrouvé est la liaison O-glycosidique de glycane sur la chaîne latérale d'une Ser ou Thr d'une protéine [134], ou d'une hydroxylysine (Hys) [135] et hydroxyproline (Hyp) [136] pour le collagène. L'élaboration de ces structures O-glycosidiques est réalisée par des glycosyltransférases, dont certaines communes avec la N-glycosylation. Cependant, contrairement à la N-glycosylation où un lipide précurseur se voyait greffer les monosaccharides avant de les transférer, ici les monosaccharides sont greffés directement et progressivement à la protéine. La O-glycosylation a lieu dans l'appareil de Golgi et sur des protéines ayant déjà acquis leur structure tridimensionnelle. Les motifs consensus reconnus par les glycosyltransférases ne sont pas bien connus mais il semblerait que la présence d'une proline (Pro) dans le voisinage d'un site d'O-glycosylation favoriserait l'ajout de monosaccharides [137].

Dû à la très grande variété d'O-glycosylations recensées, il n'est pas aussi aisé de les classer que les N-glycosylations. Des résidus Fuc, GlcNAc, Man, Glc et Xylose peuvent se greffer aux chaînes latérales des AA [137]. Cependant, le type de O-glycosylations le plus fréquemment retrouvé est l'ajout de GalNAc, formant ainsi une O-glycosylation de type mucine [138]. La structure des chaînes glycanes comporte trois parties : le noyau, le squelette linéaire ou ramifié, et le monosaccharide terminal. On recense 8 structures de noyaux fréquemment rencontrées (**Figure 18**) [7]. Les résidus terminaux sont le Fuc ou l'acide sialique.

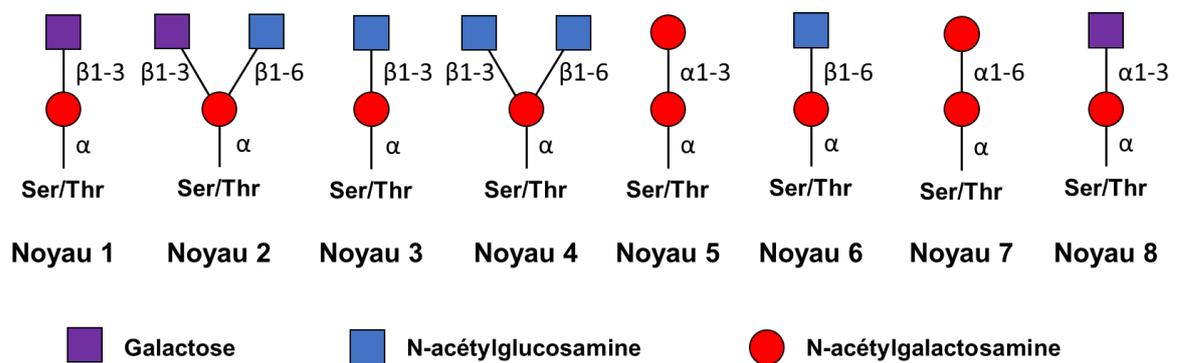


Figure 18 : Représentation des différents noyaux O-glycanes. Les noyaux 1 à 4 sont ceux les plus fréquents et retrouvés chez les glycoprotéines. (adapté de Hounsell et al., 1996) [7].

Les noyaux 1 à 4 sont retrouvés chez de nombreuses glycoprotéines, dont principalement dans les muqueuses. Les noyaux 5, 6, 7 et 8 sont moins caractérisés et sont localisés respectivement à la surface des adénocarcinomes, dans le méconium humain et la K-caséine bovine, dans les muqueuses sous-maxillaires bovines, et au niveau de la muqueuse respiratoire humaine [7].

2.1.3 La glycosylation chez les bactéries

Il est maintenant admis que les protéines glycosylées sont retrouvées dans les trois domaines du vivant [139]. En 1976, la première protéine glycosylée non eucaryote a été mise en évidence chez l'archée *Halobacterium salinarium* [140]. Depuis, des protéines N- et O-glycosylées ont été découvertes chez les bactéries, dont notamment *Campylobacter jejuni*, un pathogène associé aux muqueuses gastro-intestinales chez l'Homme [141], qui est devenu un modèle d'étude des glycosylations protéiques bactériennes (**Figure 19**) [142].

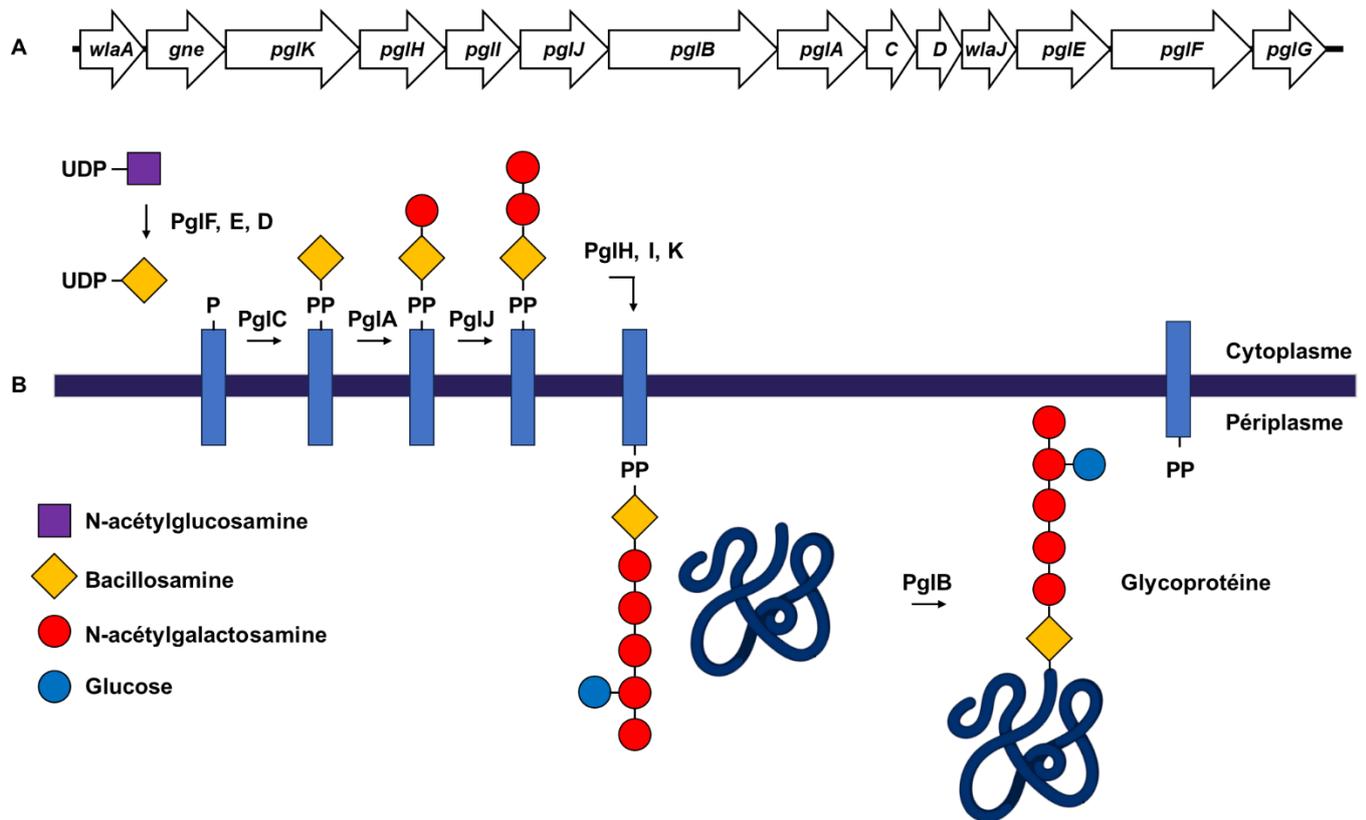


Figure 19 : N-glycosylation chez *C. jejuni*. **A.** Structure de l'opéron *pgl* chez *C. jejuni* JCM 2013. **B.** Voie de N-glycosylation chez *C. jejuni* (adapté de Chen et al., 2017 et de Srichaisupakit et al., 2014) [143,144].

C. jejuni est capable de produire des protéines N- et O-glycosylées. La N-glycosylation chez ce microorganisme est médiée par les gènes de l'opéron *pgl* (Protein **G**lycosylation) (Figures 19) et présente des similitudes avec la N-glycosylation des eucaryotes. Dans un premier temps, les protéines PglF (déshydratase), PglE (aminotransférase) et PglD (acétyltransférase) convertissent le UDP-HexNAc en bacillosamine dans le cytoplasme. Ce sucre aminé est alors greffé au support lipidique undécaprényl phosphate retrouvé au niveau de la face interne de la membrane *via* la phosphoglycosyltransférase PglC. Les glycosyltransférases PglA, PglJ, PglH et PglI permettraient l'ajout de 5 résidus GalNAc et d'un résidu glucose ramifié pour former un heptasaccharide (GalNAc- α 1,4GalNAc- α 1,4-[Glc β 1,3]-GalNAc- α 1,4-GalNAc- α 1,4-GalNAc- α 1,3diNAcBac- β 1) greffé à un lipide. La protéine PglK permet le passage du complexe lipide-heptasaccharide du cytoplasme vers le périplasma. L'heptasaccharide est alors transféré à l'asparagine de la protéine acceptrice au niveau du motif consensus D/E-X1-N-X2-S/T, où X peut être n'importe quel AA excepté la Pro [143]. Chez *C. jejuni*, 78 protéines N-glycosylées ont pu être recensées, toutes des protéines d'enveloppe

dont des protéines périplasmiques. Différentes études ont montré l'impact de la délétion des gènes du locus *pgl* sur le fitness et la virulence de cette bactérie. La délétion de ces gènes entraîne une diminution des capacités d'adhésion et d'invasion chez les cellules épithéliales humaines, une diminution de la formation de biofilm, une résistance aux antibiotiques altérée et une motilité diminuée [145].

Chez les procaryotes, deux mécanismes d'O-glycosylations ont été décrits impliquant ou non des oligosaccharyltransférases (OSTases). Le système d'O-glycosylation OSTase dépendant est semblable à la N-glycosylation. Une glycosyltransférase catalyse l'ajout d'un monosaccharide activé par un nucléotide à un lipide au niveau de la membrane. Puis des monosaccharides sont ajoutés successivement par différentes glycosylastransférases avant que le complexe lipide-glycane ne soit retourné du côté périplasmique de la membrane où sera transférée la chaîne glycane à la protéine acceptrice. Le système d'O-glycosylation OSTase indépendant est l'ajout de monosaccharides activés par un nucléotide directement à la protéine acceptrice par des glycosyltransférases cytoplasmiques lors du transfert de la protéine hors du cytoplasme [146]. Chez *C. jejuni*, des O-glycosylations ont été décrites au niveau des sous-unités FlaA et FlaB du flagelle *via* le locus de gènes *pse* et un système d'O-glycosylation OTase indépendant [146]. Le locus de gènes *pse* permet, à partir d'UDP-GlcNAc, de synthétiser de l'acide pseudaminique-CMP qui sera ajouté aux sous-unités FlaA et FlaB par l'action de la protéine PseE (**Figure 20**) [147]. La perte de O-glycosylation sur les flagelles impacte négativement la motilité [148,149].

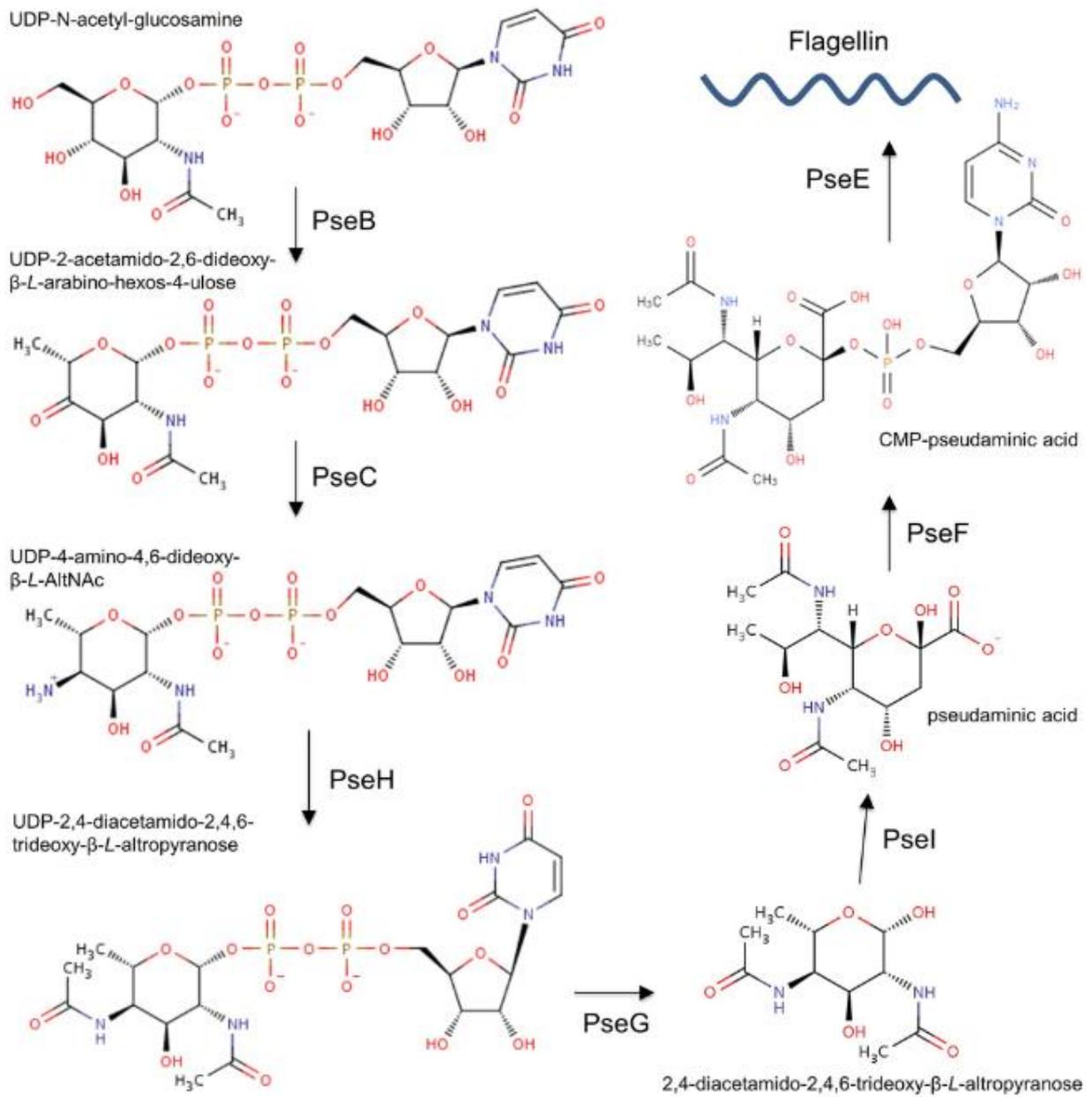


Figure 20 : Voie de biosynthèse Pse chez *C. jejuni* et *H. pylori* (Salah Ud-Din et Roujeinikova, 2017) [146].

Tableau 4 : Différents types de glycosylations bactériennes (Szymanski et al., 2005) [148]

Organisme	Protéine	Structure du glycane	Variabilité dans la structure	Lien	Sites	Sucres par site	Transfert de sucre	Effet biologique
<i>Campylobacter jejuni</i>	Multiple, sécrétées	GalNAc ₂ (Glc)GalNAc ₃ Bac	Non	N*	1-3	7	Bloc	Assemblage/Compétence du TS4, Colonisation, Adhérence/Invasion, Antigenicité des protéines
<i>Campylobacter coli</i> VC167	Flagelline FlaA	Acide pseudaminique, Acetaminido, Deoxypentose + Pse/PseAm	Oui	O	16	1-2	Séquentiel	Assemblage du flagelle, Perte de la motilité, Antigenicité des protéines
<i>Campylobacter jejuni</i> 81-176	Flagelline FlaA	Acide pseudaminique, Acetaminido, O-acetaminido, dihydroxypropionyl	Oui	O	19	1	Séquentiel	Assemblage du flagelle, Perte de la motilité, Antigenicité des protéines
<i>Helicobacter pylori</i>	FlaA, FlaB	Acide pseudaminique	Non	O	7,10	1	Séquentiel	Assemblage du flagelle, Perte de la motilité
<i>Neisseria meningitidis</i>	Piline	Gal ₁₋₂ DATDH	Oui	O	1	2-3	Séquentiel	Élimination via le système du complément
<i>Neisseria gonorrhoeae</i>	Piline	Gal ₁₋₂ DATDH	Oui	O	1	2-3	Séquentiel	Corrélation avec la maladie
<i>Pseudomonas aeruginosa</i> PAK	Flagelline	Jusqu'à 11 sucres liés par le rhamnose, y compris le 4-amino-4,6-dideoxyhexose viosamine	Oui	O	2	Various	Séquentiel	Inconnu
<i>Pseudomonas aeruginosa</i> JJ682	Flagelline	Rhamnose	-	O	2	1	Séquentiel	Inconnu
<i>Pseudomonas aeruginosa</i> 1244	Piline	Acide hydroxybutyryl-formyl-pseudaminique avec xylose et N-acétylfucosamine	-	O	1	3	Bloc	Inconnu
<i>Escherichia coli</i> H10407	Adhésine TibA	Heptoses	-	-	-	-	-	Adhérence/Invasion
<i>Escherichia coli</i> 2787	Adhésine AIDA-I	Heptoses	-	O	19	1	Séquentiel	Adhérence
<i>Haemophilus influenzae</i>	Adhésine HWM1	Gal, Glc, Man	-	-	-	-	-	Stabilité de la protéine / Adhérence

Lien : Type de liaison du glycane à la protéine ; *Sites* : Nombre de sites par protéine modifiée ; *Sucres par site* : Nombre de sucres par site modifié. *Bien qu'il s'agisse du seul système de glycosylation lié à l'azote qui ait été caractérisé, plusieurs systèmes Pgl homologues ont récemment été identifiés chez *C. coli*, *C. upsaliensis*, *C. lari*, *W. succinogenes* et *D. desulfuricans*. *AIDA-I* : adhésine impliquée dans l'adhérence diffuse ; *Bac* : bacillosamine ; *DADTH* : diacétamido-trideoxy hexose ; *Gal* : galactose ; *Glc* : glucose ; *HWM1* : glycoprotéine adhésine de masse moléculaire élevée ; *Man* : mannose ; *PseAm* : analogue 5-acétamido de l'acide pseudaminique ; *T4SS* : Système de sécrétion de type IV.

La glycosylation des flagelles a été mise en évidence chez de nombreux genres bactériens (**Tableau 4**) : *Campylobacter*, *Helicobacter*, *Aeromonas*, *Clostridium*, *Caulobacter* ou encore *Pseudomonas*. Dans la souche *E. coli* 2787, une adhésine auto-transporteuse impliquée dans l'adhésion et codée par le gène *aidA* a été montrée comme étant glycosylée par la protéine Aah. La perte de la glycosylation entraîne une perte de fonction de la protéine et ainsi une diminution de l'adhérence par la bactérie [148,150].

2.2 Enzymes déglycosylantes

Les procaryotes ont la capacité d'utiliser les glycosylations de leur hôte *via* l'intermédiaire d'enzymes déglycosylantes : les glycosylhydrolases dont les endo- β -N-acétylglucosaminidases et les asparaginases (ou peptides N-glycanases et glycosylasparaginases, respectivement).

2.2.1 Endo- β -N-acétylglucosaminidases

Les endo- β -N-acétylglucosaminidases (ENGases) sont des enzymes réalisant le clivage entre les deux résidus GlcNAc du noyau chitobiosyle des N-glycoprotéines par des domaines glycosyl hydrolases (GH) (**Figure 21**) [8].

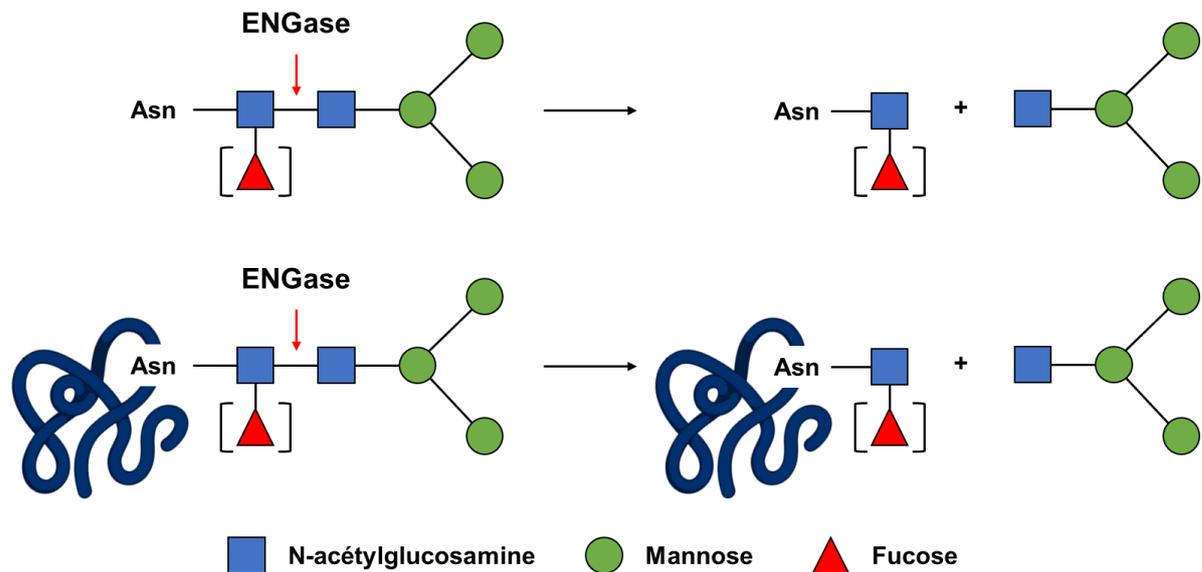


Figure 21 : Hydrolyse de la chaîne glycanne par les enzymes endo- β -N-acétylglucosaminidases. La coupure représentée en rouge a lieu au niveau de la liaison β 1-4 du noyau chitobiose. Certaines ENGases ont la capacité de cliver cette liaison lorsqu'un résidu Fucose se trouve sur un résidu GlcNAc du noyau chitobiosyle.

Les domaines GH sont classifiés en fonction de leur similarité de séquence et, à ce jour, la base de données CAZy (Carbohydrate Active enZymes) en recense 185 membres. Les ENGases sont retrouvées chez les eucaryotes et les bactéries, et possèdent principalement les domaines GH18 et GH85, bien qu'il soit possible de retrouver des ENGases comportant un domaine GH20 (**Tableau 5**) [8,151–158].

Tableau 5 : Composition des ENGases les plus étudiées (adapté de J. Fairbanks., 2017) [159].

Enzyme	Source	GH	Structure glycane hydrolysée	Hydrolyse les glycanes fucosylés
Endo A	<i>Arthrobacter protophormiae</i>	85	Riche en mannose / Hybride	Non
Endo D	<i>Streptococcus pneumoniae</i>	85	Structures tronquées	Oui
Endo E	<i>Enterococcus faecalis</i>	18/20	Complexe	Oui
Endo F ₁	<i>Elizabethkingia meningoseptica</i>	18	Riche en mannose / Hybride	Non
Endo F ₂	<i>Elizabethkingia meningoseptica</i>	18	Riche en mannose / Hybride	Oui
Endo F ₃	<i>Elizabethkingia meningoseptica</i>	18	Riche en mannose / Hybride	Oui
Endo H	<i>Streptomyces griseus</i>	18	Riche en mannose / Hybride	Oui
Endo M	<i>Mucor hiemalis</i>	85	Riche en mannose / Hybride / Complexe	Non
Endo S	<i>Streptococcus pyogenes</i>	18	Complexe	Oui
Endo S2	<i>Streptococcus pyogenes</i>	18	Riche en mannose / Hybride / Bi antenne complexe	Oui

La famille GH85 est retrouvée principalement chez les eucaryotes, bien que des exemples bactériens puissent être donnés (EndoA d'*Arthrobacter protophormiae* ou EndoD de *S. pneumoniae*). Le domaine catalytique est composé du motif NXEX₄₀Y [160]. Les domaines GH85 possèdent également une activité de transglycosylation. Il a été montré qu'EndoA peut former de nouveaux oligosaccharides en transférant une structure glycane riche en Man à des monosaccharides (GlcNAc ou Glc). EndoM possède également cette activité de transglycosylation et peut transférer les structures glycanes complexes [152].

Les domaines GH18 possèdent une structure tridimensionnelle conservée (β/α) 8 barrel folds et un site catalytique DXXDXDXE situé à l'extrémité C-terminale du brin β 4 [153,159,160]. En plus du domaine GH18, un domaine GH20 peut être associé comme chez l'une des ENGase d'*E. faecalis*, EndoE. Le domaine GH20 possède un site catalytique GGDE et agit comme une N-acetyl- β -hexosaminidase catalysant l'hydrolyse des liaisons glycosidiques N-acétyl- β -hexosamine [158]. EndoE est une enzyme de 88 kDa pouvant utiliser la RNaseB et des IgG (Immunoglobuline G) comme substrats. L'utilisation de la RNaseB est réalisée par le domaine GH18 alors que l'utilisation des IgG comme substrat se fait par le domaine GH20 [158]. Cependant, des résultats contradictoires ont été obtenus dans une étude récente s'intéressant au rôle du domaine GH18 dans l'utilisation des IgG comme substrat [161]. Il a également été montré qu'EndoE est capable d'agir sur la lactoferrine [157]. Un homologue à EndoE est retrouvé chez *Streptococcus*, EndoS. Il a été montré qu'EndoS est un facteur de virulence aggravant les infections à streptocoques dû à sa capacité à cliver les chaînes lourdes des IgG [162,163].

2.2.2 Peptide N-glycanases

Les peptides N-glycanases (PNGases, ou peptide-N⁴-[N-acetyl-β-glucosaminy] asparagine amidase) sont des enzymes appartenant à la classe des amidases. Ces enzymes catalysent le clivage des glycoprotéines et glycopeptides contenant des oligosaccharides complexes et riches en Man en hydrolysant la liaison glycosyl-asparagine (**Figure 22**).

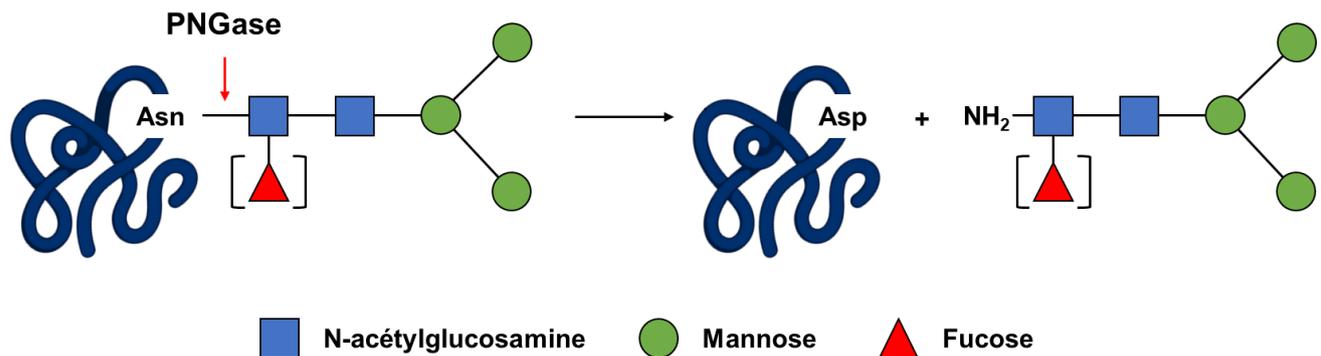


Figure 22 : Hydrolyse de la chaîne glycane par les enzymes PNGases. La coupure représentée en rouge a lieu au niveau de la liaison glycosyl-asparagine. Certaines PNGases ont la capacité de cliver cette liaison lorsqu'un résidu fucose se trouve sur un résidu GlcNAc du noyau chitobiosyle.

Cette réaction conduit à la libération d'aspartate et d'un résidu 1-amino-oligosaccharidique [9]. Il a été montré que les résidus 1-amino-saccharidiques ne sont pas stables en milieu aqueux et qu'une réaction de désamination non enzymatique spontanée a lieu, libérant ainsi l'oligosaccharide du groupement amine [164,165].

La première PNGase caractérisée a été identifiée chez l'amande [166]. Par la suite, d'autres PNGases ont été identifiées parfois chez une large variété d'organismes (**Figure 23**) : levures, bactéries, eucaryotes unicellulaires et pluricellulaires [8,151,167].

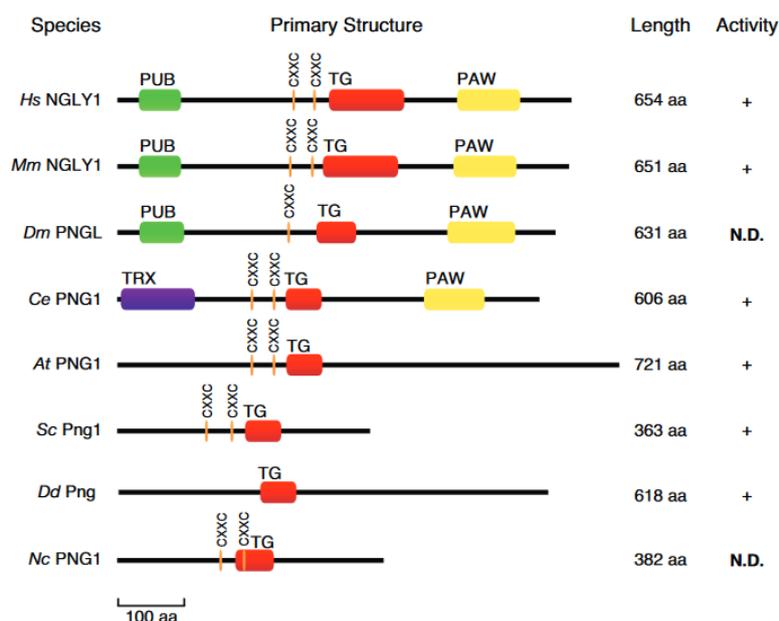


Figure 23 : Présentation schématique des structures primaires des PNGases de divers organismes. Les domaines fonctionnels ont été prédits à l'aide de l'outil web SMART (Simple Molecular Architecture Research Tool) (<http://smart.embl-heidelberg.de/>). HsNGLY1, H. sapiens N-glycanase1 (Q96IV0) ; MmNGLY1, M. musculus N-glycanase1 (Q9JI78) ; DmPNGL, D. melanogaster PNGase-like (Q7KRR5) ; CePNG1, C. elegans PNGase1 (Q9TW67) ; AtPNG1, A. thaliana PNGase1 (Q9FGY9) ; ScPng1, S. cerevisiae PNGase1 (Q02890), DdPng, D. discoideum PNGase (Q55FC8), NcPNG1 ; N. crassa PNGase1 (Q7SI01) ; PUB, domaine d'association peptide:N-glycanase ubiquitine (domaine d'interaction protéine-protéine) ; TG, domaine transglutaminase (domaine catalytique) ; PAW, domaine présent dans la PNGase et d'autres protéines de vers (domaine de liaison au mannose). CXXC est un motif putatif de liaison au zinc. N.D : non déterminé. (Hirayama et al., 2015) [167].

Même si elles catalysent la même réaction, une grande variabilité existe entre les différents organismes. Néanmoins, toutes les PNGases possèdent deux régions dans le domaine catalytique : Le transglutaminase-like core domain constitué des 3 AA cystéine, histidine et acide aspartique, essentiels pour l'activité de déglycosylation, et le domaine composé d'une paire de motif CXXC doigt de zinc. Ce dernier ne serait pas forcément nécessaire pour que la réaction ait lieu [167,168]. Enfin, la présence d'ion Zn^{2+} inhibe la réaction enzymatique des PNGases.

Le rôle et l'impact qu'ont les PNGases ne sont pas identiques suivant l'organisme. Chez les mammifères, il a été proposé que les PNGases joueraient un rôle important dans le contrôle de la biosynthèse des protéines N-glycosylées [151,169]. Dans le cas où les protéines

N-glycosylées adoptent une mauvaise conformation, ces dernières seraient dégradées dans la lumière du RE par le processus ERAD. Un défaut de déglycosylation, provoque une maladie de stockage lysosomal autosomique récessive nommée aspartylglucosaminurie. Cette accumulation affecte plus particulièrement les cellules neuronales, conduisant à une déficience intellectuelle et divers problèmes neurologiques et comportementaux [165,170]. Chez les plantes, les protéines mal repliées subissent également une dégradation, cependant le rôle de la PNGase cytosolique dans le mécanisme de déglycosylation dans les cellules végétales reste à élucider [8].

Enfin, des PNGases ont été identifiées chez différentes bactéries, mais seules celles de *Elizabethkingia meningoseptica* ont été caractérisées [171,172]. *E. meningoseptica* contient 2 PNGases nommées PNGaseF et PNGaseF-II de tailles différentes (354 AA et 567 AA respectivement). Ces deux enzymes possèdent à l'extrémité N-terminale un peptide signal de 30 AA, signifiant qu'elles sont sécrétées. Elles peuvent toutes les deux dégrader des chaînes glycanes complexes (IgG), riches en Man (RNaseB) ou hybrides (Ovalbumine). Cependant, l'ovalbumine et la RNaseB doivent être préalablement dénaturées avant d'être utilisées par la PNGaseF-II. Contrairement à PNGaseF, PNGaseF-II est capable de dégrader des résidus fucosylés [171]. Il a été montré que i) ces enzymes, PNGases et ENGases sont exprimées durant la phase stationnaire de croissance [153,158,173] lorsque les ressources en carbone dans le milieu ont été épuisées, ii) les produits de déglycosylation issus de ces enzymes pouvaient favoriser la croissance bactérienne [158], iii) ces enzymes ont un rôle dans la virulence de la bactérie, en clivant les chaînes lourdes des IgG, entraînant un défaut d'activation du complément [153,154,157,158]. Il a également été montré que la présence d'EndoS permet une résistance accrue de la bactérie à l'opsonophagocytose.

2.2.3 Glycosylasparaginase

Les enzymes déglycosylantes de type N4-(beta-N-acetylglucosaminy)-L-asparaginase (glycosylasparaginase) appartiennent à la famille des N-terminal nucléophile hydrolases (Ntm hydrolases). Ces enzymes catalysent la même réaction enzymatique que les PNGases, à la différence que le résidu asparagine glycosylé doit être au préalable séparé du reste de la partie protéique (**Figure 24**). Différents types de substrats comprenant un résidu asparagine sur une extrémité peuvent être utilisés par la glycosylasparaginase [174].

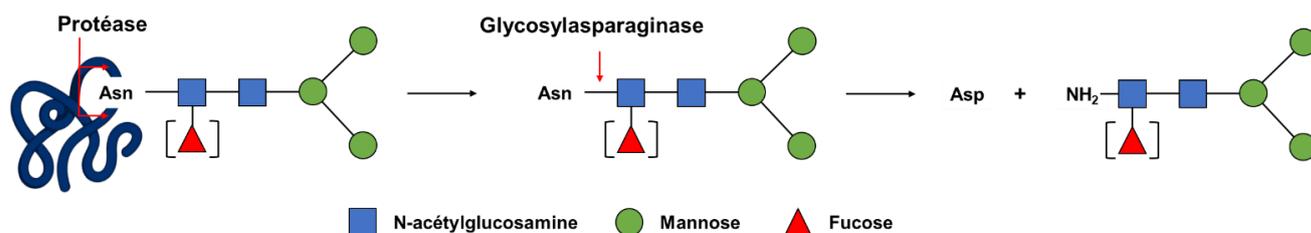


Figure 24 : *Hydrolyse de la liaison glycosidique C1-N du GlcNAc-L-Asn par les glycosylasparaginases. Les glycosylasparaginases ne peuvent cliver la liaison glycosidique que lorsqu'aucun autre AA n'est attaché au GlcNAc-L-Asn. Certaines glycosylasparaginases sont capables d'hydrolyser cette liaison lorsqu'un résidu fucose se trouve sur le GlcNAc.*

Les glycosylasparaginases ont été majoritairement étudiées chez les eucaryotes (insectes, rat, bovins, souris [175]) et plus particulièrement chez l'humain. Chez ce dernier, cette enzyme intervient dans la régulation de la glycosylation. Un défaut d'activité glycosylasparaginase provoque également des cas d'aspartylglucosaminurie. Bien que très peu étudié, les bactéries expriment également des glycosylasparaginases. L'analyse des génomes a permis d'identifier plusieurs microorganismes possédant des gènes codant des glycosylasparaginases putatives, mais seules les glycosylasparaginases de *Lactocaseibacillus casei* BL23 (anciennement *Lactobacillus casei*) et de *E. meningoseptica* ont été caractérisées [10,176].

La première glycosylasparaginase bactérienne a été identifiée en 1993 par Tarentino et Plummer chez *E. meningoseptica* [10]. Il a été montré que cette enzyme permet l'hydrolyse totale de GlcNAc-L-Asn en 10 minutes avec une activité spécifique de 25 $\mu\text{moles}/\text{min}/\text{mg}$ de protéine. Elle peut également utiliser des substrats glycosylés plus longs tel le Asn-(GlcNAc)₂(Man)₆. Cependant, l'enzyme utilisera préférentiellement les résidus glycosylés et permettra leur hydrolyse bien plus efficacement. L'utilisation d'un substrat par la glycosylasparaginase nécessite l'activation de cette dernière. Les glycosylasparaginases sont des hétérodimères constitués de sous-unités α et β générées depuis la proenzyme par autoprolyse [177]. Plusieurs études réalisées avec la glycosylasparaginase de *E. meningoseptica* ont permis d'identifier les AA impliqués dans le passage vers la forme active : histidine 150, asparagine 151, thréonine 152 [178–181]. Suite à la substitution de la thréonine 152 par d'autres AA, seuls les acides aminés ayant une chaîne latérale hydroxyle (T152S) ou thiol (T152C) ont pu permettre la formation d'une glycosylasparaginase active. Il a également été montré que l'ajout d'hydroxylamine en présence de la forme immature permet une

vitesse d'auto-protéolyse 144 fois plus rapide, suggérant la formation d'un intermédiaire thioester stable durant l'auto-protéolyse. À partir de ces données, un modèle a été proposé (Figure 25).

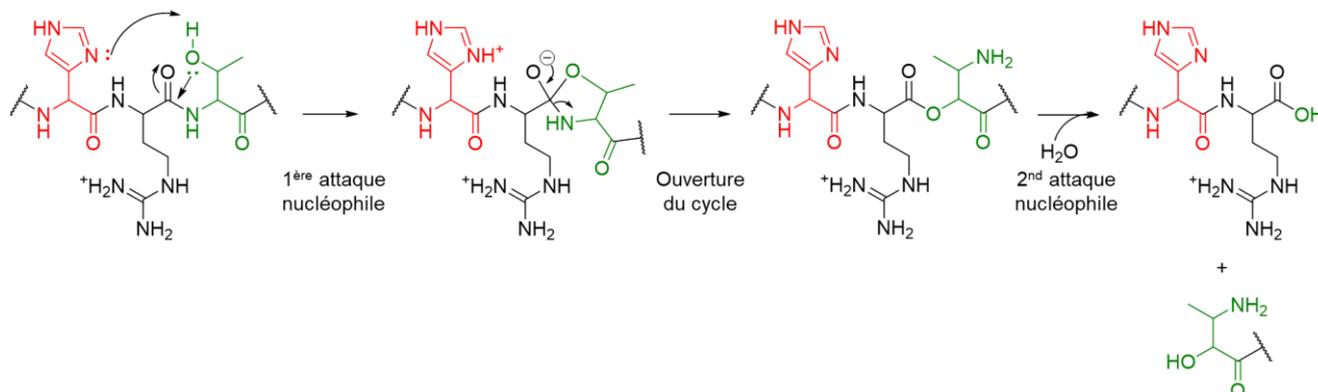


Figure 25 : *Modèle d'autoprotéolyse de la glycosylasparaginase*. Les AA sont représentés par différentes couleurs. L'histidine 150 en rouge, l'asparagine 151 en noire et la thréonine 152 en vert. Les flèches noires indiquent les réactions. Les doublets non liants sont représentés par deux points. (adapté de plusieurs sources) [178,182].

L'azote de la chaîne latérale de l'histidine 150 déprotone le groupement hydroxyle de la chaîne latérale de la thréonine 152. Ce dernier réagit alors sur le carbonyle de l'arginine 151 *via* une attaque nucléophile aboutissant à un intermédiaire cyclique. Puis l'ouverture du cycle permet d'obtenir le dérivé ester correspondant. Enfin, une molécule d'eau intervient comme nucléophile réagissant sur l'ester, ce qui provoque le clivage du précurseur de la glycosylasparaginase en deux sous-unités. En dehors de son implication dans l'auto-protéolyse, le résidu thréonine 152 est aussi impliqué dans l'activité enzymatique. Lorsque cet AA est remplacé par un autre et ne présente plus de résidu hydroxyle à l'extrémité N-terminale de la sous-unité β (correspondant à la position 152 de la pro-enzyme), la glycosylasparaginase perd sa capacité à cliver les résidus GlcNAc-L-Asn [178].

Une seconde glycosylasparaginase identifiée chez *L. casei* BL23 a été partiellement caractérisée [176]. Elle subit également un clivage auto-protéolytique et est capable de dégrader le GlcNAc-L-Asn avec une activité spécifique de $20.91 \mu\text{mol.mg protéine}^{-1}.\text{min}^{-1}$, valeur similaire à celle de *E. meningoseptica*. Une particularité notable est qu'elle est capable d'utiliser, moins efficacement cependant, des résidus GlcNAc-L-Asn sur lesquels un fucose est greffé sur le premier résidu GlcNAc. Cette particularité n'est pas partagée chez toutes les

glycosylasparaginases. Contrairement aux glycosylasparaginases de rat, celles de l'Homme ne peuvent pas hydrolyser ce type de composé [183].

En dehors de rares cas (p. ex. *C. jejuni*), les bactéries ne sont pas connues pour synthétiser des protéines glycosylées. La fonction des enzymes déglycosylantes chez ces dernières n'est donc pas similaire à celle des eucaryotes. Plusieurs faisceaux de preuves convergent vers l'hypothèse que ces enzymes déglycosylantes joueraient un rôle dans la recherche de source de carbone et dans la virulence. Aucune information sur la virulence ou l'utilisation des produits de déglycosylation n'est disponible concernant les glycosylasparaginases. En raison de leur activité enzymatique proche des PNGases et ENGases, nous pouvons supposer un rôle semblable des glycosylasparaginases. Ceci constitue l'objet d'étude de la partie expérimentale.

3 Lien entre métabolisme et virulence chez *E. faecalis*

Enterococcus faecalis est un pathogène ayant une forte capacité d'adaptation aux environnements hostiles. Cette capacité représente un avantage lors de l'étape de la colonisation de l'hôte. Cependant, *E. faecalis* est une bactérie hétérotrophe, ce qui implique qu'en condition infectieuse cette bactérie doit trouver une source de carbone et d'énergie chez l'hôte pour se multiplier et proliférer. L'importance de la recherche de nutriments en condition infectieuse a été mise en évidence dans une étude transcriptomique globale chez la souris [4], montrant la surexpression de 211 gènes et la répression de 157 gènes en condition d'infection (**Figure 26**).

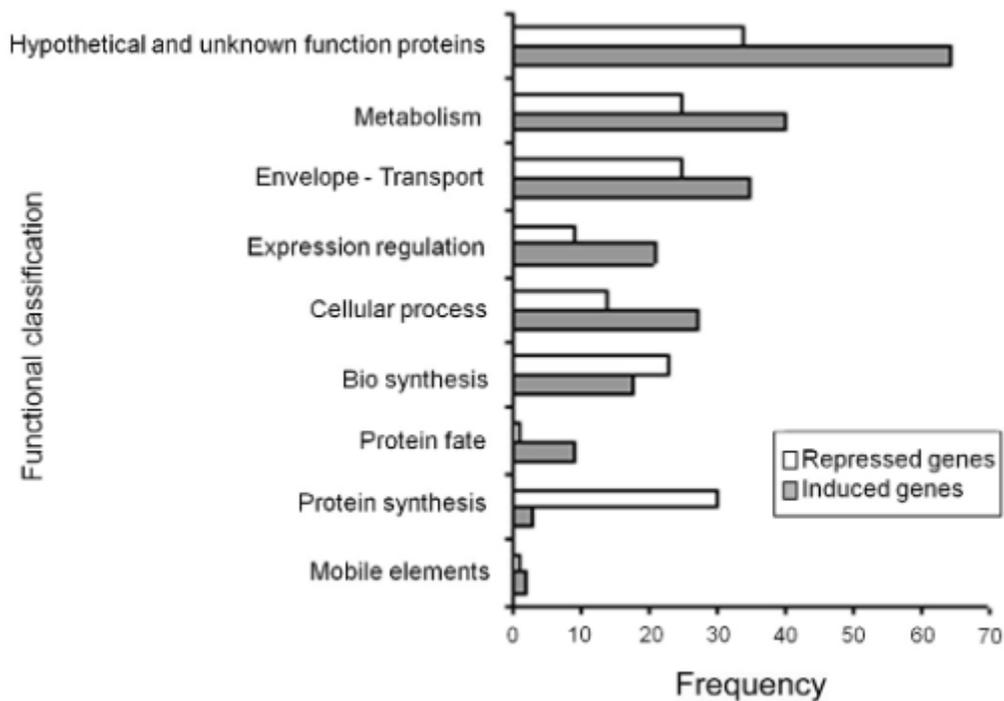


Figure 26 : Expression différentielle des gènes d'*E. faecalis* au cours de l'infection du péritoine de la souris, en fonction des catégories fonctionnelles. Les barres blanches et grises indiquent les gènes dont l'expression est respectivement réprimée et induite in vivo par rapport aux cellules cultivées jusqu'à la phase mi-log en BHI. (Muller et al., 2015) [4].

Dans la **Figure 26**, les 2^{ème} et 3^{ème} groupes de gènes surexprimés sont ceux liés au métabolisme (40) et au transport (34), montrant ainsi que les bactéries sont métaboliquement actives et cherchent une source d'énergie chez l'hôte. Parmi ces gènes, on retrouve ceux impliqués dans l'utilisation des α - et β -glucosides, ce qui tend à supposer que ces substrats sont importants pour *E. faecalis* en condition infectieuse. Les sucres sont ensuite préférentiellement métabolisés par la voie des pentoses phosphate plutôt que la glycolyse. Cette voie permet la synthèse de NADPH qui est un cofacteur impliqué dans de nombreuses réactions.

Le métabolisme du glycérol semble également jouer un rôle important en condition infectieuse. En modèle murin, la persistance dans les reins et le foie d'une souche doublement délétée des gènes *gldA* (glycerol déshydrogénase) et *gldK* (glycérol kinase) est négativement impactée [4]. Des résultats similaires ont été obtenus lors d'une étude transcriptomique dans différents milieux mimant des conditions de stress (Sels biliaires, Urée, Sérum sanguin, Acide, Stress oxydatif) retrouvées chez l'hôte [184]. Globalement, les gènes assimilés à l'utilisation du glycérol sont surexprimés dans chacune de ces conditions de stress. Néanmoins, chaque

stress induit des réponses spécifiques se traduisant par une modification spécifique de l'expression des gènes du métabolisme (**Tableau 6**).

Tableau 6 : Impact de différentes conditions de stress sur le métabolisme d'*E. faecalis* (Salze et al., 2020) [184].

Conditions environnementales mimées	Voies métaboliques positivement impactées
Sels biliaires	Maltose et Maltodextrine
Urine	Citrate, Ethanol, N-acétylglucosamine, Xylose, Galactose, Fer, Cuivre, Cadmium
Sérum sanguin	Citrate, Arginine, Acides aminés ramifiés.

Une étude transcriptomique réalisée chez *E. faecalis* suite à une culture dans un milieu contenant du gentiobiose a permis de mettre en évidence que certaines glycosylhydrolases, dont des Endo- β -N-acétylglucosaminidases, étaient surexprimées [5]. Parmi ces enzymes, EndoE, possède un rôle dans la virulence d'*E. faecalis*, comme expliqué précédemment. Des travaux réalisés par Soussan et collaborateurs [45] ont mis en évidence que le régulateur NagY impliqué le métabolisme du GlcNAc régulait également *hyla*, gène codant une hyaluronidase. En absence de HylA, une diminution significative de 3.2 fois de la formation de biofilm sur surfaces recouvertes d'acide hyaluronique est observée par rapport au témoin (souche sauvage). En plus de son rôle dans la formation de biofilm, son implication a également été mise en évidence dans la survie bactérienne lors d'expériences réalisées en modèle chenille *Galleria mellonella*. Les auteurs ont conclu que HylA permettait de fournir une source de nutriments à partir de glycosaminoglycanes procurant un avantage dans le processus d'infection à *E. faecalis*. Un dernier exemple du lien entre métabolisme et virulence peut être donné avec l'étude menée par Gao et collaborateurs. Cette étude met en évidence le lien entre CcpA, protéine de contrôle centrale du métabolisme impliquée dans la répression catabolique, et la protéine Ace, un facteur de virulence [185]. Lorsqu'*E. faecalis* est déléetée du gène *ccpA*, l'expression de la protéine Ace à la surface cellulaire est altérée. Cette altération se traduit par une adhésion au collagène significativement plus faible par rapport à la souche sauvage. Cependant, le mécanisme par lequel CcpA régule Ace n'a pas encore été mis en évidence.

4 CRISPR-Cas

En 1987, il a été rapporté chez *E. coli* la présence d'enchaînements de séquences répétées identiques séparées par des séquences uniques [186]. Des séquences similaires ont plus tard été mises en évidence chez d'autres bactéries et archées [187–189]. Ce n'est qu'à partir des années 2000, suite à des analyses *in silico*, que ces séquences ont été trouvées chez plus de 40 % des bactéries et 90 % des archées dont le génome a été séquencé [190]. Ces séquences répétées ont été nommées CRISPR pour « Clustered Regularly Interspaced Short Palindromic Repeats » et leur environnement génétique a été caractérisé [191]. Les séquences CRISPR sont souvent retrouvées à proximité des gènes codant des protéines conservées ayant des activités hélicases ou nucléases. Ces protéines conservées ont été nommées Cas pour « CRISPR associated proteins ». Plusieurs systèmes CRISPR-Cas ont été identifiés et classés en types et familles en fonction de l'association des différentes protéines Cas. Ce n'est qu'à partir de 2005 qu'émerge l'hypothèse selon laquelle CRISPR-Cas serait un système immunitaire adaptatif des microorganismes, impliqué dans la défense contre l'ADN exogène dont les éléments génétiques mobiles [192]. Cette hypothèse sera confirmée par de nombreux travaux montrant que CRISPR-Cas est capable de cliver les ADN étrangers tout en les gardant en mémoire.

Le locus comportant CRISPR-Cas peut être décomposé en 2 parties (**Figure 27A**). La première, appelée tracrRNA (trans-activating CRISPR RNA), comporte une ou plusieurs enzymes Cas. La seconde partie du locus CRISPR-Cas se nomme CRISPR array et est composée d'une succession de séquences répétées et identiques (crRNA) d'une taille comprise entre 21 et 47 pb, et de séquences variables nommées « spacers » d'une taille de 20 pb [190]. Par souci de clarté, seul l'exemple du système CRISPR-Cas9 sera décrit. La première étape est appelée interférence et fonctionne de la manière suivante (**Figure 27B**) : le locus CRISPR-Cas est transcrit permettant la production de la protéine Cas9, du tracrRNA et d'un ARN qui est clivé de sorte à former des séquences courtes d'ARN composées du crRNA et d'un spacer. Le transcrit tracrRNA s'hybride aux séquences ARN crRNA de sorte à former un duplex où viendra se fixer l'endonucléase Cas9. Ce complexe parcourt l'ADN et guide l'endonucléase Cas9 sur une portion d'ADN complémentaire aux nucléotides du spacer appelé protospacer. Le spacer s'apparie ensuite au protospacer, à condition que se trouve une séquence nucléotidique 5' – NGG – 3' en aval de celui-ci [193]. La deuxième étape est appelée étape de dégradation. Une

fois le spacer apparié, l'enzyme Cas9 provoque la cassure du double brin de l'ADN étranger. Celle-ci est reconnue par des nucléases bactériennes qui le dégrade entièrement [194,195].

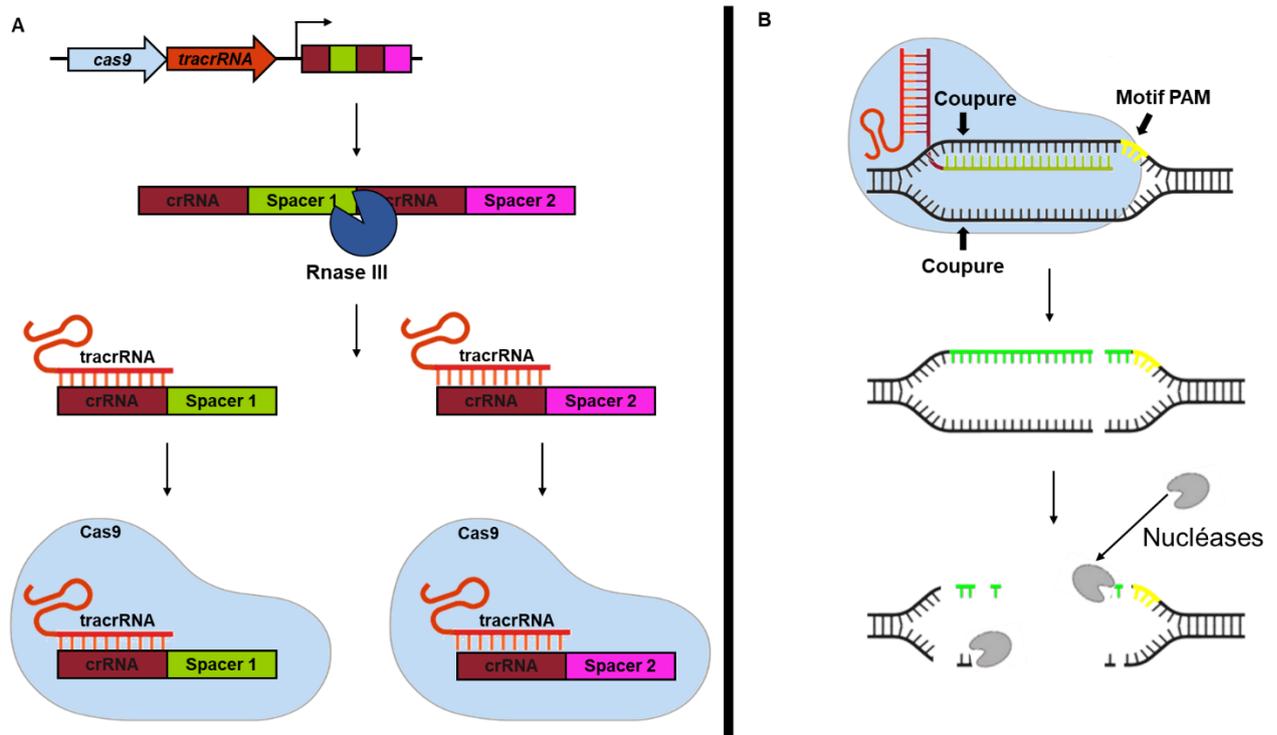


Figure 27 : Système CRISPR-Cas9. **A** : Mise en place du complexe CRISPR-Cas9. **B** : Interférence et dégradation (adapté de Hille et al., 2018) [196].

Suite à la découverte du mécanisme de fonctionnement, le système CRISPR-Cas a été détourné par Charpentier et ses collaborateurs pour en faire un outil de biologie moléculaire programmable pour cibler précisément un motif ADN et y induire une cassure double brin. Cet outil a été amélioré en combinant le crRNA, le tracrRNA et la séquence spacer en une seule molécule nommée ARN guide chimérique (ARNg) [197]. Chez les eucaryotes, il est utilisé pour l'édition de génomes. Chez les procaryotes, il est principalement utilisé pour resensibiliser les bactéries aux antibiotiques avec un ARNg conçu pour cibler les gènes de résistances [198]. Des outils CRISPR ont été développés et utilisés chez différents microorganismes tels que *Staphylococcus aureus* [198], *E. coli* [199] ou *Shewanella algae* [200] pour la génération de mutants. Cet outil permet de forcer la double recombinaison homologe avec un fragment d'ADN apporté à la bactérie suite à la coupure de l'ADN double brin.

5 Objectifs

La recherche d'une source de carbone en condition infectieuse est une problématique importante pour la survie de la bactérie. Les données de la littérature indiquent que les microorganismes possèdent un ensemble de glycosylases permettant la dégradation d'une large variété de polysaccharides complexes. Parmi ces glycosylases, les ENGases et PNGases utilisent comme substrat les protéines N-glycosylées retrouvées chez l'hôte. On pensait que ces enzymes étaient exclusives aux eucaryotes, mais des exemples ont été décrits chez les procaryotes. Certaines ont été caractérisées chez les bactéries, comme EndoE, mais globalement, peu d'information est disponible. L'analyse du génome d'*E. faecalis* OG1RF a mis en évidence un gène montrant des homologies de séquences avec les enzymes du type glycosylasparaginase. Ce gène (*OG1RF_10935* chez la souche OG1RF) est localisé au sein d'un locus de 7 gènes (*OG1RF_10933* à *OG1RF_10939*) organisés en une ou deux structures opéroniques. L'objectif de ce travail de thèse a été de caractériser ce locus et en particulier le gène codant pour la potentielle activité glycosylasparaginase chez *Enterococcus faecalis* OG1RF et de mettre en évidence son lien éventuel avec la virulence de la bactérie. La **Figure 28** présente le modèle hypothétique de fonctionnement du métabolisme des protéines glycosylées chez *E. faecalis*, développé à partir du rôle hypothétique des gènes présents dans le locus.

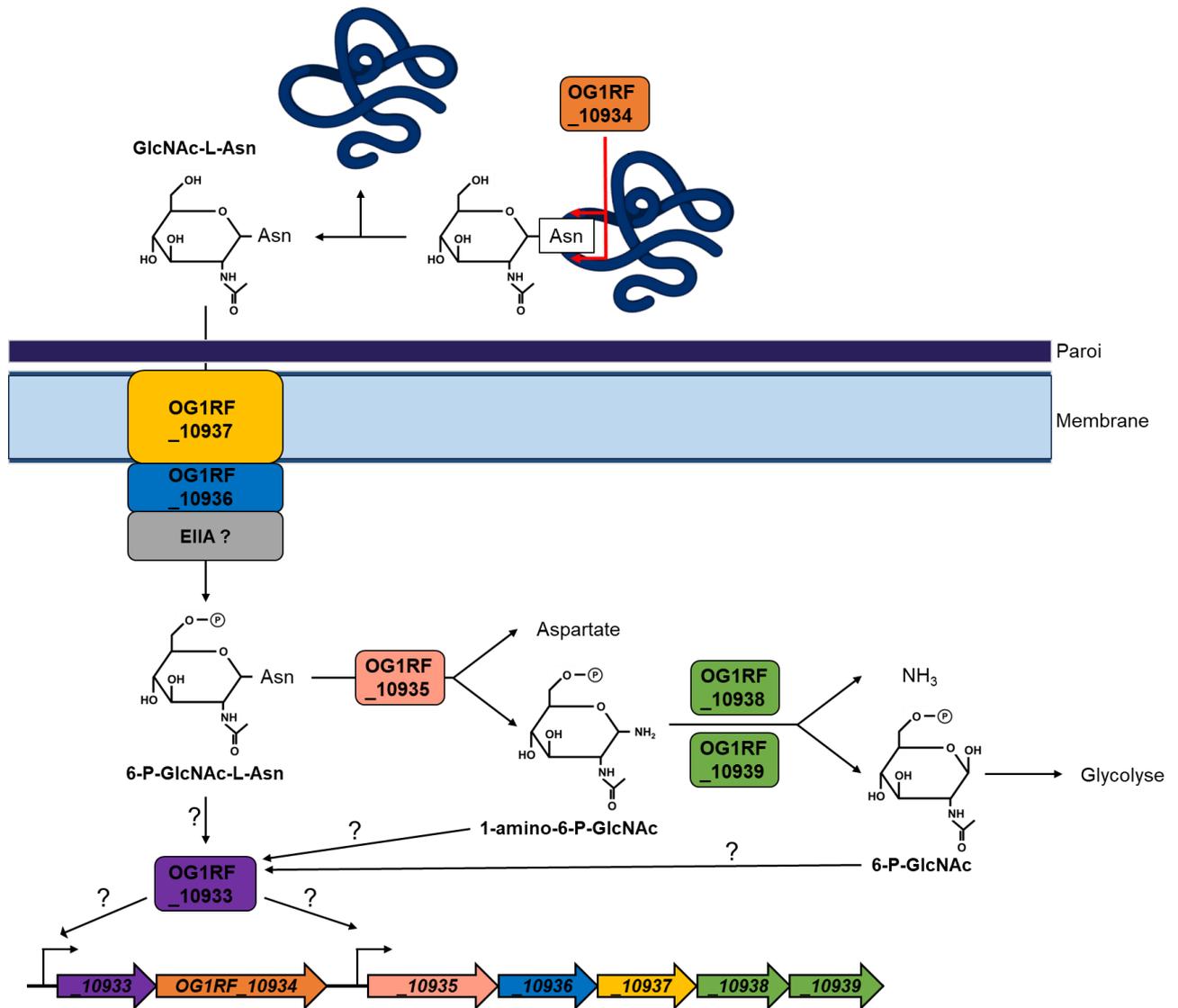


Figure 28 : Modèle initial du fonctionnement du locus comportant la glycosylasparaginase chez *E. faecalis*. Les flèches rouges indiquent où cliverait la peptidase. Le nom des molécules est indiqué en gras.

En effet, au sein de ce locus, et en amont du gène codant pour une glycosylasparaginase putative (*OG1RF_10935*), se trouve un gène codant pour un régulateur transcriptionnel et un gène codant une peptidase. Cette dernière permettrait de cliver les liaisons peptidiques de protéines glycosylées de part et d'autre de l'asparagine comportant la chaîne glycanique, générant ainsi le substrat de la glycosylasparaginase (GlcNAc-L-Asn). En aval de *OG1RF_10935* se trouvent deux gènes codant les sous-unités EIIB et EIIC d'un PTS annoté comme appartenant à la classe lactose/cellobiose. Ce PTS assurerait le transport et la phosphorylation du GlcNAc-L-Asn. La glycosylasparaginase d'*E. faecalis* utiliserait ce substrat

phosphorylé pour générer du N-acétylglucosaminylamine-6P et aspartate. Enfin les derniers gènes de l'opéron, *OG1RF_10938* et *OG1RF_10939*, codent pour des réductases qui permettraient le retrait du groupement amine, conduisant à la formation d'ammoniaque et de N-acétylglucosamine phosphorylé. Ce dernier rentrerait dans la voie de la glycolyse. Un des substrats générés pourrait interagir avec le régulateur (*OG1RF_10933*), ce qui modulerait la transcription des différents gènes de ce locus.

Une caractérisation de ce locus a donc été réalisée pour démontrer le modèle hypothétique à travers l'étude de l'activité glycosylasparaginase, ainsi que de la régulation et du transport de composés mimant les produits issus de la déglycosylation des protéines eucaryotes comme sources de carbone et d'énergie potentielle. Une partie des résultats présentés fait l'objet d'une publication soumise à un comité scientifique de relecture au sein du journal « Research in microbiology » et subit actuellement un processus de révision par les pairs. Cette publication intitulée « **Utilization of glycoprotein-derived N-acetylglucosamine-L-asparagine during *Enterococcus faecalis* infection depends on catabolic and transport enzymes of the glycosylasparaginase locus** » se trouve en annexe. Un projet annexe à la thèse a consisté à mettre en place un outil de mutagenèse utilisant la technologie CRISPR-Cas9 chez *E. faecalis*. Cet outil a déjà été réalisée chez *E. faecium* [201]. Cet article nous a servi de référence pour transposer l'outil CRISPR-Cas utilisable chez *E. faecalis* dans l'objectif de créer des mutants lors de ce travail de thèse.

Matériels et méthodes

1 Matériels biologiques

1.1 Conditions de culture et souches bactériennes

Les cultures d'*E. coli* ont été réalisées en routine en milieu Lysogeny Broth [202] à 37 °C avec agitation à 120 rpm. Lorsque nécessaire, des antibiotiques sont ajoutés au milieu aux concentrations finales suivantes : Ampicilline 100 µg/mL ; Kanamycine 25 µg/mL ; Spectinomycine 100 µg/mL et Chloramphénicol 25 µg/mL. Les plasmides d'intérêts ont été insérés dans des souches TOP10 (ThermoFischer Scientific), EC1000 [203] et NEB5α (New England Biolabs) avant de transformer *E. faecalis*. Les souches BL21(DE3) (New England Biolabs) et M15[pREP4] (Qiagen) ont permis la surexpression des protéines d'intérêts grâce aux plasmides pET29b+ (Novagen) et pQE30 (Qiagen).

Les cultures de routine d'*E. faecalis* ont été réalisées en milieu M17 (Biokar diagnostics) [204] supplémenté avec 0.5 % de glucose final (m/v) sans agitation à 37 °C. Lorsque nécessaire, des antibiotiques sont ajoutés au milieu aux concentrations finales suivantes : Erythromycine 75 µg/mL, Spectinomycine 200 à 900 µg/mL et Chloramphénicol 25 µg/mL. Lors des expérimentations de suivi de croissance, le milieu cdM17 [205] supplémenté avec la source de carbone désirée a été utilisé. La souche utilisée dans cette étude est la souche OG1RF [206].

La composition des différents milieux de culture utilisés lors de cette étude est indiquée en annexe. Les souches utilisées sont listées dans le **tableau 7**.

Tableau 7 : Souches utilisées dans ce travail.

Souches	Génotype	Source ou Référence
<i>Enterococcus faecalis</i>		
OG1RF	Souche sauvage dépourvue de plasmide. Fus ^R , Rif ^R .	[206]
OG1RF pMSP3535	Souche sauvage comportant le plasmide pMSP3535	Cette étude
Δ <i>egaR</i>	Souche mutante pour le gène <i>egaR</i> , dérivée de la souche parentale OG1RF.	Cette étude
Δ <i>egaR</i> pMSP3535:: <i>egaR</i>	Souche mutée pour le gène <i>egaR</i> , comportant le plasmide de complémentation pMSP3535:: <i>egaR</i>	Cette étude
Δ <i>egaR</i> pMSP3535	Souche mutée pour le gène <i>egaR</i> , comportant le plasmide vide pMSP3535	Cette étude
Δ <i>egaP</i>	Souche mutante pour le gène <i>egaP</i> , dérivée de la souche parentale OG1RF.	Cette étude

Δ egaP pMSP3535:: <i>egaP</i>	Souche mutée pour le gène <i>egaP</i> , comportant le plasmide de complémentation pMSP3535:: <i>egaP</i>	Cette étude
Δ egaP pMSP3535	Souche mutée pour le gène <i>egaP</i> , comportant le plasmide vide pMSP3535	Cette étude
Δ egaG	Souche mutante pour le gène <i>egaG</i> , dérivée de la souche parentale OG1RF.	Cette étude
Δ egaG pMSP3535:: <i>egaG</i>	Souche mutée pour le gène <i>egaG</i> , comportant le plasmide de complémentation pMSP3535:: <i>egaG</i>	Cette étude
Δ egaG pMSP3535	Souche mutée pour le gène <i>egaG</i> , comportant le plasmide vide pMSP3535	Cette étude
Δ egaB	Souche mutante pour le gène <i>egaB</i> , dérivée de la souche parentale OG1RF.	Cette étude
Δ egaB pMSP3535:: <i>egaB</i>	Souche mutée pour le gène <i>egaB</i> , comportant le plasmide de complémentation pMSP3535:: <i>egaB</i>	Cette étude
Δ egaB pMSP3535	Souche mutée pour le gène <i>egaB</i> , comportant le plasmide vide pMSP3535	Cette étude
Δ egaC	Souche mutante pour le gène <i>egaC</i> , dérivée de la souche parentale OG1RF.	Cette étude
Δ egaC pMSP3535:: <i>egaC</i>	Souche mutée pour le gène <i>egaC</i> , comportant le plasmide de complémentation pMSP3535:: <i>egaC</i>	Cette étude
Δ egaC pMSP3535	Souche mutée pour le gène <i>egaC</i> , comportant le plasmide vide pMSP3535	Cette étude
Δ egaD1D2	Souche mutante pour les gènes <i>egaD1D2</i> , dérivée de la souche parentale OG1RF.	Cette étude
Δ egaD1D2 pMSP3535:: <i>ega D1D2</i>	Souche mutée pour le gène <i>egaD1D2</i> , comportant le plasmide de complémentation pMSP3535:: <i>egaD1D2</i>	Cette étude
Δ egaD1D2 pMSP3535	Souche mutée pour leS gène <i>egaD1D2</i> , comportant le plasmide vide pMSP3535	Cette étude
Δ OG1RF_10751	Souche mutante pour le gène <i>OG1RF_10751</i> , dérivée de la souche parentale OG1RF.	[5]
Δ OG1RF_10751:: <i>OG1RF_10751</i>	Souche complémentée avec le gène <i>OG1RF_10751</i> provenant de la souche V583.	[5]
<u>Escherichia coli</u>		
TOP10	F- <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74 recA1 araD139</i> Δ (<i>ara-leu</i>) 7697 <i>galU galK rpsL</i> (Str ^R) <i>endA1 nupG</i> λ .	Thermo Fischer Scientific
EC1000	F- Δ (<i>araA-leu</i>)7697, [<i>araD139</i>]B/r, Δ (<i>codB-lacI</i>)3, <i>galK16, galE15</i> (<i>GalS</i>), λ -, <i>e14-</i> , <i>relA1, rpsL150</i> (strR), <i>spoT1, mcrB1</i>	[203]
NEB5 α	<i>fhuA2</i> Δ (<i>argF-lacZ</i>)U169 <i>phoA glnV44</i> Φ 80 Δ (<i>lacZ</i>)M15 <i>gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i> .	New England Biolabs

BL21(DE3)	F- <i>ompT hsdS_B (r_Bm_B-) gal dcm</i> (DE3).	New England Biolabs
BL21(DE3) pET29b+:: <i>egaR</i>	<i>E. coli</i> BL21 possédant le plasmide de surexpression pET29b+:: <i>egaR</i> .	Cette étude
M15pRep4	F-, Φ 80 Δ <i>lacM15</i> , <i>thi</i> , <i>lac</i> -, <i>mtl</i> -, <i>recA</i> +, Km ^R .	Qiagen
M15pRep4 pQE30:: <i>egaG</i>	<i>E. coli</i> M15pRep4 possédant le plasmide de surexpression pQE30:: <i>egaR</i> .	Cette étude

1.2 Plasmides

1.2.1 pMAD

Le plasmide pMAD a été utilisé pour la construction des différents mutants chez la souche *E. faecalis* OG1RF [207]. Ce plasmide est un vecteur navette de 9666 pb utilisable chez *E. coli* et *E. faecalis*. Il contient deux origines de réplication. L'origine de réplication pBR322 permet la réplication du plasmide à un fort nombre de copies chez les bactéries à Gram négatif, et la seconde pET194ts comporte un gène codant une réplicase thermosensible. Cette dernière fonctionne chez les bactéries à Gram positif à une température de 30 °C et devient inactive lorsque la température est au-delà de 42 °C. Ce plasmide possède également deux résistances aux antibiotiques ampicilline et érythromycine dues aux gènes *bla* et *ermC*. Ce vecteur permet également le « screening » bleu/blanc car il possède le gène *bgaB* codant une β -galactosidase. En présence de 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal), les souches possédant le plasmide apparaissent bleues sur le milieu de culture. Le site multiple de clonage se trouve en dehors des gènes cités, ce qui permet l'ajout de fragment ADN sans altérer le fonctionnement du plasmide. Les événements de recombinaison peuvent être contrôlés par les facteurs température et pression de sélection.

1.2.2 pMSP3535

Le plasmide pMSP3535 permet d'exprimer des protéines chez les bactéries à Gram positif [208]. Il a été utilisé pour compléter les souches mutées. Ce plasmide de 8353 pb contient deux origines de réplication. La première pBR322 permet la réplication du plasmide à un fort nombre de copie chez les bactéries à Gram négatif, et la seconde pAMB1 permet la réplication chez les bactéries à Gram positif. Ce plasmide contient le gène *ermB*, ce qui permet de discriminer les souches ayant ou non le plasmide sur un milieu comportant de l'érythromycine. On retrouve également le système à deux composantes *nisR/nisK* qui permet

en présence de nisine, l'expression des gènes se trouvant sous le contrôle du promoteur pNisA, où seront insérés les gènes codant pour les protéines d'intérêts.

1.2.3 pGEM-T easy

Le plasmide pGEM-T Easy (Promega) a été utilisé pour cloner les régions promotrices *pegaR* et *pegaG* du locus glycosylasparaginase, la région promotrice *pegaG* dont la séquence de fixation HutC a été mutée, ainsi qu'un fragment interne du gène *gusA*. Ce petit plasmide de 3015 pb est répliatif chez les bactéries à Gram négatif *via* une origine de réplication pUC ori, et possède également une origine de réplication phagique f1, qui permet sa réplication sous forme simple brin. Une sélection est réalisable avec une pression de sélection à l'ampicilline *via* le gène *ampR*. Le clonage se fait au niveau du site multiple de clonage du gène *lacZ*, ce qui permet de discriminer les clones ayant l'insert par un « screening » bleu/blanc. Ce vecteur est commercialisé sous forme linéaire avec des extrémités 3' simples brins possédant un nucléotide T pour permettre le clonage TA.

1.2.4 pQE30

Le plasmide pQE30 (Qiagen) a été utilisé pour surexprimer les protéines recombinantes EgaG et EgaP. Ce vecteur est répliatif chez les bactéries à Gram négatif et possède un gène codant une β -lactamase, ce qui lui confère une résistance à l'ampicilline. Le site multiple de clonage comporte une séquence permettant la production d'une protéine possédant en N-terminal une étiquette 6xHis, et placée sous le contrôle d'un promoteur inductible à l'IPTG. La protéine ainsi clonée présentera un tag 6xHis en N-terminal, ce qui permettra sa purification grâce à une chromatographie d'affinité sur colonne Ni-NTA ou ÄKTA start.

1.2.5 pET29b+

Le plasmide pET29b+ (Novagen) a été utilisé pour surexprimer les protéines recombinantes EgaR et MDH. Ce vecteur est répliatif chez les bactéries à Gram négatif par la présence de l'origine de réplication pRB322. Une sélection est réalisable avec une pression de sélection à la kanamycine grâce au gène *kanR*. Ce plasmide comporte un site multiple de clonage qui permet l'ajout d'une étiquette 6xHis en C-terminal de la protéine exprimée lorsque le gène d'intérêt est cloné dans le même cadre de lecture. La présence d'IPTG dans le

milieu inhibe l'action de LacI, exprimé constitutivement, permettant ainsi la surexpression de la protéine d'intérêt tagguée. La purification de la protéine est réalisée grâce à une chromatographie d'affinité sur colonne Ni-NTA ou ÄKTA start.

1.2.6 pVPL3004

Le plasmide pVPL3004 a été utilisé pour la mutagenèse avec CRISPR-Cas9 [201]. Ce plasmide de 9746 pb possède le gène *cas9* qui code pour l'endonucléase Cas9, et possède également la séquence tracrRNA qui permet la formation du double brin ARN avec le crRNA. pVPL3004 est maintenu dans la cellule avec une pression de sélection à l'érythromycine.

1.2.7 pVDM1001

Le plasmide pVDM1001 a été utilisé en mutagenèse avec CRISPR-Cas9 pour cloner le spacer et le fragment d'ADN contenant le gène d'intérêt délété à insérer dans le chromosome [201]. Ce plasmide de 4822 pb contient la séquence cDNA leader avec deux régions répétées entre lesquelles le spacer est intégré suite à une digestion par *BsaI*. Le fragment permettant la réparation du chromosome après clivage a été cloné en amont de la séquence cDNA leader. pVDM1001 est maintenu dans la cellule avec une pression de sélection à la spectinomycine.

Les plasmides utilisés et leurs caractéristiques sont listés dans le **tableau 8**.

Tableau 8 : Plasmides utilisés dans ce travail.

Plasmides	Caractéristiques	Référence ou source
pMAD	Origine de réplication pE194ts et pBR322, Em ^r , Amp ^r , <i>bgaB</i>	[207]
pMAD:: <i>ΔegaR</i>	Création du mutant <i>ΔegaR</i>	Cette étude
pMAD:: <i>ΔegaP</i>	Création du mutant <i>ΔegaP</i>	Cette étude
pMAD:: <i>ΔegaG</i>	Création du mutant <i>ΔegaG</i>	Cette étude
pMAD:: <i>ΔegaB</i>	Création du mutant <i>ΔegaB</i>	Cette étude
pMAD:: <i>ΔegaD1D2</i>	Création du mutant <i>ΔegaD1D2</i>	Cette étude
pMSP3535	Promoteur pNisA, Em ^r , <i>nisR</i> , <i>nisK</i>	[208]
pMSP3535:: <i>egaR</i>	Complémentation du mutant <i>ΔegaR</i>	Cette étude
pMSP3535:: <i>egaP</i>	Complémentation du mutant <i>ΔegaP</i>	Cette étude
pMSP3535:: <i>egaG</i>	Complémentation du mutant <i>ΔegaG</i>	Cette étude
pMSP3535:: <i>egaB</i>	Complémentation du mutant <i>ΔegaB</i>	Cette étude
pMSP3535:: <i>egaC</i>	Complémentation du mutant <i>ΔegaC</i>	Cette étude
pMSP3535:: <i>egaD1D2</i>	Complémentation du mutant <i>ΔegaD1D2</i>	Cette étude
pGEM-T Easy	Origines de réplication pUC et f1. Compatible avec le TA cloning. Amp ^R .	Promega
pGEMT- <i>pegaR</i>	Comporte la région promotrice du gène <i>egaR</i>	Cette étude

pGEMT- <i>pegaG</i>	Comporte la région promotrice du gène <i>egaG</i>	Cette étude
pGEMT-Mut <i>pegaG</i>	Comporte la région promotrice du gène <i>egaG</i> avec les motifs de fixation HutC mutés.	Cette étude
pGEMT <i>gus</i>	Comporte un fragment interne du gène <i>gusA</i>	[94]
pET29b+	Origine de réplication colE1. Comporte un tag 6xHis en C-terminal pour la purification de protéines. Km ^R .	Novagen
pET29b+:: <i>egaR</i>	Purification de la protéine 6xHis-EgaR	Cette étude
pQE30	Origine de réplication colE1. Comporte un tag 6xHis en N-terminal pour la purification de protéines. lacI ^q , Amp ^R , Cm ^R .	Qiagen
pQE30:: <i>egaG</i>	Purification de la protéine 6xHis-EgaG	Cette étude
pVDM1001	Comporte le gène <i>cas9</i> et le tracrRNA. Em ^r .	[201]
pCRISPR- <i>egaP</i>	Création du mutant Δ <i>egaP</i> via CRISPR-Cas9	Cette étude
pCRISPR- <i>egaC</i>	Création du mutant Δ <i>egaC</i> via CRISPR-Cas9	Cette étude
pVPL3004	Comporte le crRNA, spacer, et fragment ADN permettant la réparation du chromosome suite au clivage par la Cas9. Chi ^r .	[201]

2 Cellules compétentes et transformations par électroporation

2.1 *Escherichia coli*

Une préculture des souches est réalisée dans 10 mL de milieu LB pendant 18 h à 37 °C avant d'être utilisée pour ensemercer à 2 % 500 mL de milieu LB incubé à 37 °C avec agitation jusqu'à obtention d'une densité optique à 600nm (D.O. _{600 nm}) de 0,8. Après centrifugation (5.000 rpm ; 10 min ; 4 °C), les cellules sont lavées 3 fois dans de l'eau ultra-pure froide, puis lavées 3 fois dans du glycérol 10 % froid avant d'être reprises dans 500 µL de glycérol 10 % froid. Les cellules sont aliquotées et stockées à -80 °C jusqu'à utilisation. Les cellules compétentes et le plasmide d'intérêt à électroporer sont déposées dans des "GenePulser® electroporation cuvette 2 mm" (Biorad) refroidies, puis un choc électrique est délivré par un appareil GenePulser® (BioRad) selon les paramètres suivants : tension de 2,5 kV ; capacitance de 25 µF et résistance de 400 Ω. Les cellules sont alors resuspendues dans 1 mL de milieu LB puis sont laissées à incuber à 37 °C durant 45 min. Les cellules sont finalement ensemençées sur milieu LB comportant l'antibiotique adapté et incubées à 37 °C.

2.2 *Enterococcus faecalis*

Les cellules d'*E. faecalis* ont été préparées pour pouvoir être transformées par électroporation [209]. Un traitement contenant de la mutanolysine et du lysozyme est employé pour fragiliser la paroi des cellules d'*E. faecalis* avant transformation. Une préculture

des souches est réalisée dans 10 mL de milieu GM17 pendant 18 h à 37 °C, puis diluée au 100^{ème} dans 200 mL de milieu GM17 et incubée à nouveau à 37 °C jusqu'à obtention d'une D.O. _{600 nm} de 0,5 avant d'être laissée 20 min dans la glace. Les cellules sont alors séparées du milieu GM17 par centrifugation (6.000 rpm ; 10 min ; 4 °C) puis lavées dans du glycérol 10 % (v/v) froid. La paroi des cellules est fragilisée durant 15 min à 37 °C par un tampon (10 mM Tris pH8 ; 20 % (m/v) sucrose ; 10 mM EDTA) comportant de la mutanolysine (Sigma-Aldrich) et du lysozyme (Sigma-Aldrich) respectivement à 10 U/mL et 10 µg/mL. Après centrifugation (12.500 rpm ; 90 sec ; 4 °C), les cellules sont lavées 3 fois dans un tampon d'électroporation froid (0.5 M sucrose ; 10 % (v/v) glycérol) avant d'être aliquotées et stockées à -80 °C jusqu'à utilisation. Après ajout du plasmide, les cellules compétentes sont soumises à un choc électrique délivré par un appareil GenePulser® (BioRad) selon les paramètres suivants : tension de 2,25 kV ; capacitance de 25 µF et résistance de 400 Ω. Les cellules sont alors resuspendues dans 1 mL de milieu GM17 comportant du MgCl₂ (20 mM) et du CaCl₂ (2 mM), puis sont régénérées à 37 °C durant 1 h 30. Les cellules sont finalementensemencées sur milieu GM17 comportant l'antibiotique adapté et incubées à 30 °C ou 37 °C.

3 Cellules compétentes et transformations par choc thermique

3.1 *Escherichia coli*

Les cellules d'*E. coli* Top10, EC 1000, NEB5α et BL21 ont été préparées pour pouvoir être transformées par choc thermique [210]. Une préculture des souches est réalisée dans 10 mL de milieu LB pendant 18 h à 37 °C avec agitation, puis diluée au 100^{ème} dans 30 mL de milieu LB et incubée à nouveau à 37 °C jusqu'à obtention d'une D.O. _{600 nm} de 0,5 avant d'être aliquotée par 1 mL dans des microtubes. Les cellules sont alors centrifugées (4.000 rpm ; 10 min ; 4 °C) avant d'être resuspendues dans 90 µL de tampon TSS (1g PEG 8000 ; 30 mM MgCl₂ ; 5 % (v/v) DMSO ; QSP 50 mL LB) et stockées à -80 °C. Les cellules compétentes sont déposées 15 min dans la glace, puis 5 µL de tampon KCM 5X (0,5 M KCl ; 150 mM CaCl₂ ; 250 mM MgCl₂) et entre 1 et 5 µL de plasmide sont ajoutés. Le tube contenant les cellules et le plasmide est incubé 30 min dans la glace avant de subir un choc thermique de 42 °C pendant 90 sec, le tube est alors placé de nouveau 30 min dans la glace. Les cellules sont reprises dans 1 mL de milieu

LB et incubées pendant 1 h à 37 °C avant d'êtreensemencées sur milieu gélosé LB comportant l'antibiotique adapté.

4 Biologie moléculaire

4.1 Réaction de polymérisation en chaîne (PCR)

Les PCR ont été réalisées à l'aide de l'ADN polymérase GoTaq®G2 Flexi DNA Polymerase (Promega) en respectant les recommandations du fournisseur. Les oligonucléotides utilisés ont été synthétisés chimiquement par le groupe Eurofins. Les réactions PCR sont réalisées dans un thermocycleur "Biometra Tadvanced" (Analytik Jena).

Lorsqu'une amplification nécessitant une grande fidélité est nécessaire, la Q5® High-Fidelity DNA Polymerase (New England Biolabs) est utilisée en suivant les recommandations du fournisseur. Les différents oligonucléotides utilisés sont indiqués en annexe (**Tableaux S4 à S8**).

4.2 Electrophorèse en gel d'agarose

La vérification des produits PCR, digestions enzymatiques, et extraction d'ADN génomique ont été systématiquement effectués par électrophorèse en gel d'agarose 1 % (m/v). Les migrations ont été réalisées dans une cuve Sub-Cell® GT Cell (BioRad) sous un voltage constant de 90 V jusqu'à ce que le front de migration atteigne les 3/4 de la longueur du gel. Des marqueurs de masse moléculaire GeneRuler 1 kb DNA Ladder ou GeneRuler 100 pb DNA Ladder (ThermoFisher Scientific) ont été utilisés pour déterminer la taille des fragments d'ADN. Les acides nucléiques ont été révélés par ajout de MIDORI Green Advance (Dutscher) à 0,006 % final et détectés dans le lecteur à UV Axygen® Gel Documentation Systems (Axygen).

4.3 Restriction et ligature de fragments d'ADN

Les enzymes de restriction ont été utilisées dans leurs conditions optimales de fonctionnement conformément aux recommandations du fournisseur. Les réactions de ligatures entre fragments ADN ont été réalisées avec la T4 DNA Ligase (ThermoFischer Scientific) selon les recommandations du fournisseur.

4.4 Purification des fragments d'ADN

La purification des produits PCR et des fragments nucléotidique digérés par des enzymes de restriction ou issus de ligatures ont été purifiés à l'aide du kit " QIAquick PCR Purification Kit (Qiagen) " selon les recommandations du fournisseur.

4.5 Extraction de plasmides d'*Escherichia coli*

Les plasmides ont été obtenus à partir de souches d'*E. coli* possédant le plasmide d'intérêt à partir de 5 mL de milieu LB comportant l'antibiotique adapté pendant 18 h à 37 °C. L'extraction est réalisée à l'aide du kit « QIAprep Spin Miniprep Kit » (Qiagen) en suivant les recommandations du fournisseur. La qualité des plasmides a été vérifiée comme décrite précédemment et la concentration a été déterminée en utilisant le Nanodrop One™ (Thermofisher Scientific).

4.6 Extraction d'ADN chromosomique

L'ADN chromosomique a été extrait en utilisant le kit « Nucleospin Microbial DNA » (Macherey-Nagel) à partir de 5 mL d'une culture de nuit selon les recommandations du fournisseur. La qualité de l'ADN a été vérifiée par spectrophotométrie (Nanodrop) en mesurant l'absorbance à 260 nm. La pureté de l'ADN a été vérifiée en mesurant les absorbances à 230 et 280 nm.

4.7 Extraction d'ARN totaux

Les cellules sont, dans un premier temps, lysées au FastPrep-24™ 5G (MP Biomedicals™) par trois cycles de 30 sec à 6 mouvements par seconde espacés par 5 min de repos sur glace. Les ARN sont ensuite extraits en utilisant le kit « Direct-zol » (Zymo Research) selon les recommandations du fournisseur. La quantité d'ARN est mesurée par spectrophotométrie à 260 nm à l'aide d'un NanoDrop One™ (Thermofisher Scientific). La pureté des ARN a été vérifiée par la mesure des ratios A260/A280 et A260/A230 devant être tous deux supérieurs à 2. L'intégrité des ARN a été vérifiée par électrophorèse en gel d'agarose comprenant 0.06 % de chlore actif [211].

4.8 Réaction de transcription inverse

La réaction de transcription inverse a été réalisée à partir de 2 µg d'ARN à l'aide du kit « QuantiTect® Reverse Transcription » (Qiagen) selon les recommandations du fournisseur.

4.9 PCR quantitative

Les PCR quantitatives ont été réalisées dans des plaques 96 puits à l'aide du kit SYBR Green fluorescence (Promega). Les amplifications ont eu lieu dans des thermocycleurs CFX96TM Real-Time System (Bio-Rad) avec le programme suivant : 2 min à 98 °C ; 40 cycles de 5 sec à 98 °C et 20 sec à 61 °C avec mesure de la fluorescence à la fin de chaque amplification. A l'issue des cycles d'amplification, une augmentation progressive de la température de 65 °C à 95 °C avec acquisition de la fluorescence est réalisée pour générer les courbes de fusion. Le gène ADNr 5S est utilisé comme gène de référence.

5 Constructions de mutants et complémentations

5.1 Mutant de délétion : Mutagenèse par recombinaison homologue

Les mutants de l'opéron *ega* ont été obtenus suite à la délétion de la partie centrale du gène (75 %-90 %). Des PCR ont été réalisées de sorte à amplifier les régions amont et aval du gène d'intérêt. Les deux fragments obtenus présentent des extrémités compatibles entre eux, permettant de réaliser une 3^{ème} PCR de sorte à obtenir un amplicon d'environ 2000 pb contenant le gène tronqué à insérer dans le chromosome. Le plasmide pMAD a été linéarisé par PCR inverse, puis digéré par l'enzyme *DpnI* pour éliminer le plasmide méthylé ayant servi de matrice ADN. Le plasmide linéarisé ainsi que l'amplicon contenant le gène délété ont été transformés chez *E. coli* par choc thermique comme décrit précédemment. Le plasmide contenant l'insert est alors introduit chez *E. faecalis* par électroporation, puis intégré dans le chromosome par recombinaison homologue en réalisant une culture liquide à 42 °C avec de l'érythromycine comme pression de sélection. Le plasmide ne pouvant pas se répliquer est soit perdu, soit forcé de s'intégrer dans le chromosome. Le second évènement de recombinaison homologue est réalisé par repiquages successifs à 30 °C dans des milieux de culture sans antibiotique suivi d'un passage à 42 °C. Des vérifications à chaque étape sont

réalisées par PCR. Le mutant est finalement confirmé par un séquençage : soit des régions amont et aval de la partie délétée, soit par un séquençage complet du génome.

5.2 Mutant de délétion : CRISPR-Cas9

Le mutant *egaC* a été obtenu *via* CRISPR-Cas9 d'après le protocole décrit par de Maat et collaborateurs [201]. Des clonages successifs du spacer, puis de la région qui sert à la recombinaison ont été réalisés dans le plasmide pVDM1001. Le choix du motif ADN ciblé par le complexe CRISPR-Cas9 a été réalisé de sorte à ce qu'il fasse 30 pb de longueur et se trouve en amont d'une séquence NGG (Protospacer adjacent motif). Des oligonucléotides comportant cette séquence ont été choisis, avec les caractéristiques suivantes : le primer sens bordé en 5' d'un motif AAAC et en 3' d'un nucléotide G, et le primer antisens bordé en 5' d'un motif AAAAC. Les deux oligonucléotides sont mis en contact à une concentration respective de 20 µM, chauffés à 95 °C durant 5 min, et progressivement refroidis pour créer une séquence double brin correspondant au spacer bordé par des extrémités 5' simple brin compatibles avec celles générées lors de la digestion du plasmide pVDM1001 par *BsaI*. Le plasmide pVDM1001 est alors digéré par l'endonucléase de restriction *BsaI* avant d'être ligaturé avec ce spacer. Une fois le spacer cloné, la région permettant l'insertion de la mutation doit être insérée dans le plasmide. Pour cela, le produit de ligature contenant le spacer a été linéarisé par PCR, puis digéré par l'enzyme *DpnI* pour éliminer le plasmide sous sa forme circulaire. Trois PCR successives ont été réalisées comme décrit précédemment pour obtenir un amplicon comportant le gène *egaC* tronqué. Le plasmide linéarisé ainsi que l'amplicon contenant le gène délété ont été utilisés pour transformer *E. coli* par choc thermique. Le plasmide pVPL3004 ainsi que le plasmide comportant le spacer et la version du gène *egaC* muté ont été introduits chez *E. faecalis* par électroporation, et sélectionnés sur un milieu GM17 + érythromycine + spectinomycine. Des PCR sont alors réalisées pour vérifier l'insertion du fragment d'intérêt dans le chromosome suite au double crossing-over forcé par l'outil CRISPR-Cas9.

5.3 Complémentations des mutants

Le plasmide pMSP3535 a été utilisé pour effectuer des complémentations plasmidiques des mutants des gènes du locus *ega* en introduisant une copie sauvage du gène sur le plasmide dans la souche mutante correspondante. Les gènes sauvages ont été amplifiés

par PCR à partir de l'ADN chromosomique de la souche *E. faecalis* OG1RF avant d'être clonés au niveau du site multiple de clonage du plasmide pMSP3535. Pour s'assurer que les phénotypes observés n'étaient pas dû à la présence du plasmide, des contrôles ont été réalisés en utilisant le plasmide vide. Le plasmide est maintenu dans la cellule en ajoutant de l'érythromycine dans le milieu. L'expression des gènes clonés est induite par ajout de nisine à une concentration finale de 0.5 µg/mL.

5.4 Séquençage du génome entier

L'ADN chromosomique des souches OG1RF WT et OG1RF *ΔegaG* a été extrait à l'aide du Nucleospin Microbial DNA (Macherey-Nagel) selon les recommandations du fabricant. Les quantités d'ADN ont été déterminées à l'aide du spectrophotomètre NanoDrop One™ (ThermoFisher Scientific) et l'intégrité de l'ADN a été vérifiée par électrophorèse. Le séquençage Illumina a été sous-traité par la société Eurofins. L'analyse des séquences a été réalisée à l'aide du logiciel "CLC Genomics Workbench" (version 12.0.2) avec l'outil de détection des variants en comparant le génome de la souche *ΔegaG* à la souche WT. Tous les programmes ont été utilisés avec les paramètres par défaut.

6 Test d'inhibition à la streptozotocine

Pour chacune des souches utilisées, des cultures de nuit ont été réalisées en absence ou en présence de GlcNAc-L-Asn 0.3% et une autre n'en contenant pas. Les cultures de nuit sont lavées 3 fois par centrifugation (4.500 rpm ; 10 min) avant d'être ajustées à D.O._{600 nm} 0,12. Les cultures sont incubées 1 h à 37 °C avant d'être divisées en trois aliquots : un contenant de la streptozotocine à 50 µg/mL, un deuxième contenant 0.3 % de GlcNAc-L-Asn et 50 µg/mL de streptozotocine, et un dernier aliquot sans complément. Un suivi de croissance sur 24 h à D.O._{600 nm} est réalisée au « Infinite M nano » (Tecan).

7 Méthodes biochimiques

7.1 Surexpression et purification des protéines présentant une étiquette poly-histidine

Les protéines EgaR et Glutamate-Oxaloacétate Transaminase ont été surexprimées *via* le vecteur d'expression pET29b+. Les protéines EgaP et EgaG ont été surexprimées *via* le vecteur d'expression pQE30. Les cellules ont été cultivées dans 300 mL de milieu LB frais

comportant les antibiotiques adaptés, à 37 °C avec agitation 120 rpm. De l'IPTG à la concentration finale 0,5 mM est ajouté une fois la D.O._{600 nm} comprise entre 0,7 et 0,8 atteinte pour permettre l'expression des protéines d'intérêts. Après 2 h d'incubation à 37 °C avec agitation, les cellules sont lavées 3 fois par centrifugation (5.000 rpm ; 30 min ; 4°C) par un tampon de reprise (pH 8 ; NaCl 300 mM ; NaH₂PO₄ 50 mM). Après la dernière centrifugation, le culot bactérien est pesé, puis resuspendu dans un tampon contenant de l'imidazole (pH 8 ; NaCl 300 mM ; NaH₂PO₄ 50 mM ; Imidazole 10 mM). Les échantillons sont traités au lysozyme à une concentration finale de 1 mg/mL, durant 30 min. Les cellules sont ensuite lysées au FastPrep-24™ 5G (MP Biomedicals™) par 3 cycles de 30 sec à 6 mouvements par secondes, espacés par 5 min de repos sur glace. Une centrifugation (5.000 rpm ; 30 min ; 4 °C) est réalisée pour récupérer la fraction contenant les protéines solubles. Cette fraction est injectée dans l'ÄKTA Start (Cytiva) et passera au sein d'une colonne Ni-NTA préalablement équilibrée par un tampon (pH 8 ; NaCl 300 mM ; NaH₂PO₄ 50 mM ; Imidazole 10 mM). Plusieurs lavages sont réalisés en augmentant progressivement la concentration en imidazole jusqu'à 500 mM afin d'éluer la protéine d'intérêt. Les protéines sont ensuite dessalées dans une colonne PD10 Desalting Columns (GE Healthcare) dans un tampon de dessalage (Tris 50 mM pH 7,4 ; Glycérol 15 %). Le kit « Pierce™ BCA Protein Assay Kit » a été utilisé pour déterminer la concentration massique des protéines dessalées en utilisant une gamme BSA. Les protéines sont ensuite aliquotées et conservées à -80 °C.

7.2 SDS-PAGE

Après chaque purification de protéines, une électrophorèse sur gel polyacrylamide contenant du SDS (SDS-Page) est réalisée pour déterminer la qualité des extractions protéiques. La concentration totale en polyacrylamide est de 12,5 % (v/v), avec un ratio acrylamide/bis-acrylamide de 1 :29 (ITW Reagents). Les protéines sont dénaturées au préalable dans un tampon Laemmli (SDS 2 % ; Glycérol 10 % ; 2-mercaptoéthanol 5 % ; Bleu de bromophénol 0,002 % et Tris-HCl 0,0625M pH 6,8) durant 5 min à 95 °C. Les protéines sont ensuite déposées sur le gel avec un marqueur de poids moléculaire précoloré PageRuler Prestained Ladder (ThermoFischer Scientific) et séparées dans un tampon d'électrophorèse (Tris 25 mM pH 6,8 ; Glycine 200 mM ; SDS 0,1 % (v/v)) sous un ampérage constant de 25 mA dans un appareil Electrophoresis Power Supply – EPS 300 (Pharmacia Biotech). La visualisation

des bandes est réalisée en colorant le gel avec une solution de Bleu Brillant de Coomassie (Méthanol 30 % (v/v) ; Acide acétique 7,5 % (v/v) ; Bleu brillant 2,5 mg/mL).

7.3 Retard sur gel

Les fragments d'ADN d'intérêts ont été obtenus par PCR en utilisant une amorce marquée en 5' par une molécule fluorescente Cyanine 5 (Eurofins). Les protéines et les fragments d'ADN cibles sont incubés simultanément à 37 °C durant 15 min dans un tampon (20 mM Tris-HCl pH 7.4 ; 20 mM KCl ; 10 mM MgCl₂ ; 1 mM DTT ; 0.1 mM EDTA ; 50 % Glycérol ; 0.2 µg/mL Herring sperm). Les mélanges sont ensuite séparés par une électrophorèse en gel de polyacrylamide 5 % (v/v) à l'obscurité sous une tension continue de 140 V dans un tampon TBE 0,5X. La lecture des bandes ADN se fait par mesure de la fluorescence par un appareil Typhoon 5 (Cytiva) avec les paramètres suivants : Excitation 625 nm ; Emission : 680 nm ; Résolution 100 µm.

7.4 Test enzymatique : Glycosylasparaginase (EgaG)

L'activité glycosylasparaginase a été mise en évidence par une méthode enzymatique en trois étapes (**Figure 29**) [212]. L'hydrolyse du GlcNAc-L-Asn par la glycosylasparaginase libère du 1-amino-GlcNAc et de l'aspartate. La transamination de l'aspartate par la glutamate-oxaloacétate transaminase (GOT) en présence d' α -cétoglutarate génère du glutamate et de l'oxaloacétate. Finalement, la malate déshydrogénase (MDH) convertit l'oxaloacétate et le NADH en malate et NAD⁺. La consommation du NADH est suivie par mesure de l'absorbance à D.O. 340 nm dans un « Infinite M nano » (Tecan). Les réactions ont été réalisées dans un volume final de 150 µL contenant du sodium phosphate (20 mM ; pH 7,5) ; α -cétoglutarate (0,1 mM) ; NADH (0,2 mg) ; GOT (7,5 U) ; MDH (7,5 U) avec différentes concentrations de glycosylasparaginase et de GlcNAc-L-Asn.

L'activité spécifique a été déterminée de la manière suivante. La cinétique de consommation du NADH a été calculée lors de l'oxydation maximale du NADH (pente maximale de la courbe). Cette valeur a ensuite été divisée par la quantité de protéine (en mg) afin d'obtenir l'activité spécifique en µmol/mg de protéine/min.

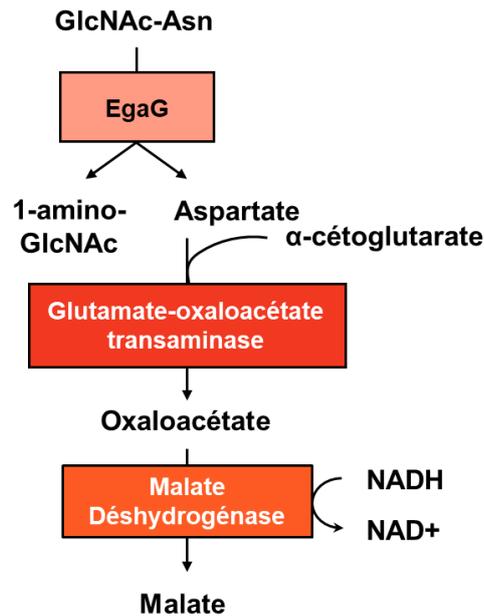


Figure 29 : Test enzymatique de l'activité glycosylasparaginase.

7.5 Test enzymatique : Peptidase (EgaP)

L'activité peptidase a été expérimentée en modifiant le test enzymatique de la glycosylasparaginase. Le même milieu réactionnel a été utilisé, à la différence de la substitution du GlcNAc-L-Asn par la peptidase EgaP (1,4 µg) et de RNase B (30 µg). Un second test a été effectué en dénaturant au préalable 30 µg RNase B (95 °C ; 1 h). Cette solution a ensuite été incubée avec l'enzyme EgaP (5,61 µg) 1 h à 37 °C avant de servir de substrat pour la glycosylasparaginase comme représenté ci-dessus.

8 Survie chez *Galleria mellonella*

Les souches utilisées ont été mises en culture comme décrit précédemment dans du GM17 à 37 °C en présence de nisine 0.5 µg/mL et érythromycine. Les cultures ont été centrifugées (5.000 rpm ; 5 min ; 4 °C) en milieu de phase exponentielle et lavées deux fois dans une solution saline (0,9 % NaCl). La concentration bactérienne a été ajustée pour injecter $9 \cdot 10^5$ UFC/larves dans le dernier segment des larves *G. mellonella* âgées de 7 semaines à l'aide d'une pompe à seringue (Thermofischer Scientific) et d'une aiguille 26G (Terumo). Vingt larves sont infectées par condition testées et incubées à 37 °C. Un dénombrement des larves vivantes a été réalisé toutes les 2 h durant 24 h après infection. Trois réplicas ont été réalisés pour un total de 60 larves par souche bactérienne. Les résultats sont représentés par des courbes de Kaplan-Meier.

9 **Biofilm**

Après une culture de nuit à 37 °C sans agitation dans un milieu GM17, les cellules ont été lavées et resuspendues dans un milieu GM17 frais contenant 2 % de glucose à une D.O._{600 nm} finale de 0,2. Pour chaque souche testée, 6 puits d'une microplaque 96 puits sont remplis avec 100 µL de la culture ajustée à D.O._{600nm} 0,2. La plaque est laissée à incuber 24 h à 37 °C sans agitation. Les puits sont lavés avec 300 µL d'une solution saline (0,9 % NaCl) puis incubés 15 min avec 125 µL de cristal violet 0,1 %. Deux lavages avec une solution saline sont réalisés, après séchage des puits, 125 µL d'acide acétique 30 % sont ajoutés. Après 15 min à température ambiante, l'absorbance est mesurée à D.O._{570nm} dans un « Infinite M nano» (Tecan).

Résultats

1 Régulation de l'opéron *ega* chez *E. faecalis*

1.1 Identification du locus glycosylasparaginase d'*E. faecalis* OG1RF

L'analyse du génome d'*E. faecalis* OG1RF a révélé la présence d'un gène (*OG1RF_10935*) annoté N4-(β -N-acetylglucosaminy)-L-asparaginase et codant une protéine de 337 AA appartenant à la famille des peptidases T2/asparaginases (**Figure 30A**). Il fait partie d'un locus de 7 gènes codant des activités qui pourraient être impliquées dans le métabolisme des protéines glycosylées chez *E. faecalis* OG1RF.

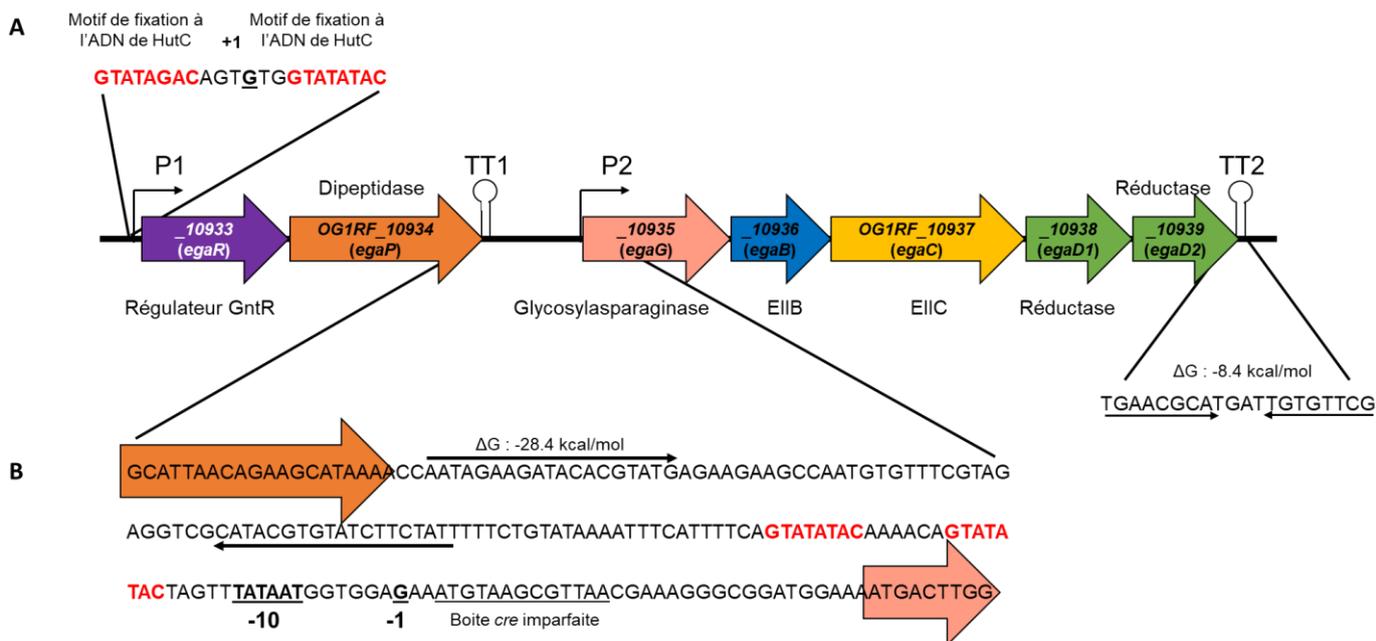


Figure 30 : Organisation génétique du locus *ega* d'*E. faecalis* OG1RF. **A** : Schéma du locus *ega*. Les flèches représentent les ORF et leur orientation indique la direction de la transcription. Les noms des gènes sont indiqués dans les flèches, les fonctions prédites sont indiquées au-dessus ou au-dessous. Les promoteurs sont symbolisés par de fines flèches coudées. Les structures en épingle à cheveux représentent des terminateurs transcriptionnels putatifs situés 2 nucléotides en aval du codon stop d'*egaP* (TT1), et 26 nucléotides en aval du codon stop d'*egaD2* (TT2). **B** : Région intergénique du locus *ega*. Les flèches noires indiquent les séquences inversées répétées permettant une possible structure en épingle à cheveux. Les séquences ADN rouges représentent le motif de fixation à l'ADN des régulateurs de la sous famille HutC.

Cette famille de protéines correspond au groupe des Ntn-hydrolases (nucléophile N-terminal (Ntn) -hydrolases) qui catalysent le clivage des liaisons amides dans une variété de macromolécules. Parmi ces protéines, on retrouve les peptidases à thréonine, les L-asparaginases et les glycosylasparaginases. Au vu de l'activité suspectée de cette protéine,

nous proposons donc de nommer ce locus *ega* (pour Enterococcal glycosylasparaginase) et le gène *OG1RF_10935 egaG*.

Les gènes situés en aval de *egaG*, nommés *egaB* (*OG1RF_10936*) et *egaC* (*OG1RF_10937*), codent respectivement les composants perméases EIIB et EIIC d'un système PTS incomplet de classe lactose/cellobiose. L'absence d'un gène codant pour le domaine EIIA dans le voisinage des éléments EIIB et EIIC suggère qu'un EIIA dont le gène est situé ailleurs sur le chromosome doit être recruté pour que ce système PTS soit complet et fonctionnel. Les distances entre ces trois gènes sont de 19 et 18 pb, cela suggère que ces gènes sont organisés en structure opéronique. Deux autres gènes, *OG1RF_10938* et *OG1RF_10939*, semblent être co-transcrits avec *egaG*, *egaB* et *egaC* et ont été nommés *egaD1* et *egaD2*. Ces gènes sont annotés comme codant des protéines hypothétiques. Les analyses des séquences génomiques d'autres souches d'*E. faecalis* (V583, JH2-2, EnGen0311) ont montré que, dans toutes les souches testées, *egaD1* et *egaD2* sont fusionnés en un seul gène, nommé *egaD*. Un examen plus approfondi de la séquence *OG1RF* a révélé que la division en deux ORF est le résultat d'un décalage du cadre ouvert de lecture et que l'ORF *egaD2* commence avec un codon de départ alternatif TTG situé à l'extrémité 3' d'*egaD1* (**Figures 30A et 31A**). Ceci suggère fortement que, dans la souche *OG1RF*, le gène *egaD* est naturellement muté. L'analyse Blast de la séquence en acides aminés d'*egaD* de la souche V583 a révélé une homologie significative (54% d'identité ; 100% de couverture) avec GrdB, qui fait partie du complexe glycine/sarcosine/bétaïne réductase retrouvé chez *Clostridium sp* (**Figure 31B**). Ce complexe est composé des trois protéines GrdA, GrdB et GrdC. GrdB est une sélénoprotéine qui métabolise la glycine, la sarcosine ou la bêtaïne en acétyl-phosphate et en ammoniacque [213]. La recherche des deux autres composants de ce complexe dans le génome d'*E. faecalis* par l'outil Blast avec GrdA et GrdC de *Clostridium sp*. comme requêtes n'a pas révélé de similarité significative avec les protéines d'*E. faecalis*. Cela suggère que le rôle joué par *EgaD* est très certainement différent de GrdB. Néanmoins, *EgaD* semble être fonctionnellement lié aux autres gènes présents dans ce locus (**Figure 30**).

A

1 5' – ATG AAA AAA ATC CTT TTA GTC TTA AAC CAC AGT TTA GCA GGA TTT GGC AGT GAT GAA CAA GCT CAG CTG AAA CCA 75

76 AGT GGT AAA AAA GCT CCC TTA GGC CCA GGG GTA ACA TTG CTA CCA TTT TTA CAG AAG CAA GGA GCG GAA ATT GTT 150

151 GCG ACA CTT TAT TGT GGT GAC CAA TAT TAT TTG GAA AAC GAG GAA GAG GTA GCA AAG AAA TTT ATT GGA GTT TCA TGC 225

226 CGA TGC GGT TCT TTG CGG ACC TGC CAT GCA TTA TCC TAA TTT TGG AGA AAT GGC CGC TCA TCT AGC TTG TAA GTT TAA 300

301 TGC TGC GAG CAT TCC TGC AAT TGC AGC GAT GGC AGA AGA AAA TCC AGC AGT TAG TCA CTA TTA TCA GCA AGT ACC GAT 375

376 TGT GAA GAT GCC CAA AAA GGG CGG CAT TGG TTT AAA TAA TTC GTT TAA ACA AAT GGC GCA ATT GGT CGT GGC AAA 450

451 AGC AAA TGG AAA AGA AAC CAA ACA GTT AGA AGA AAC AAG TTG TTT TAA – 3' 508

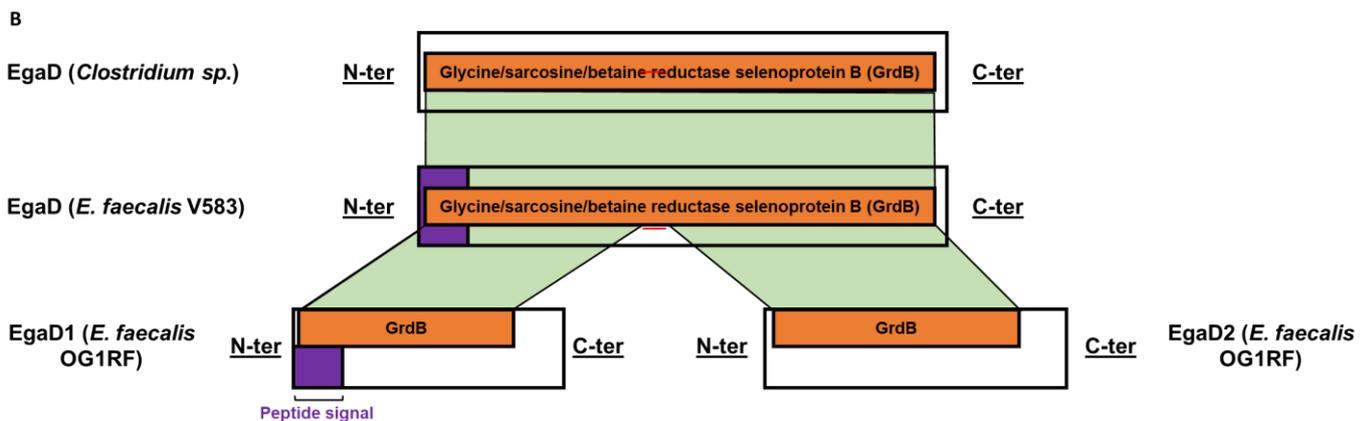
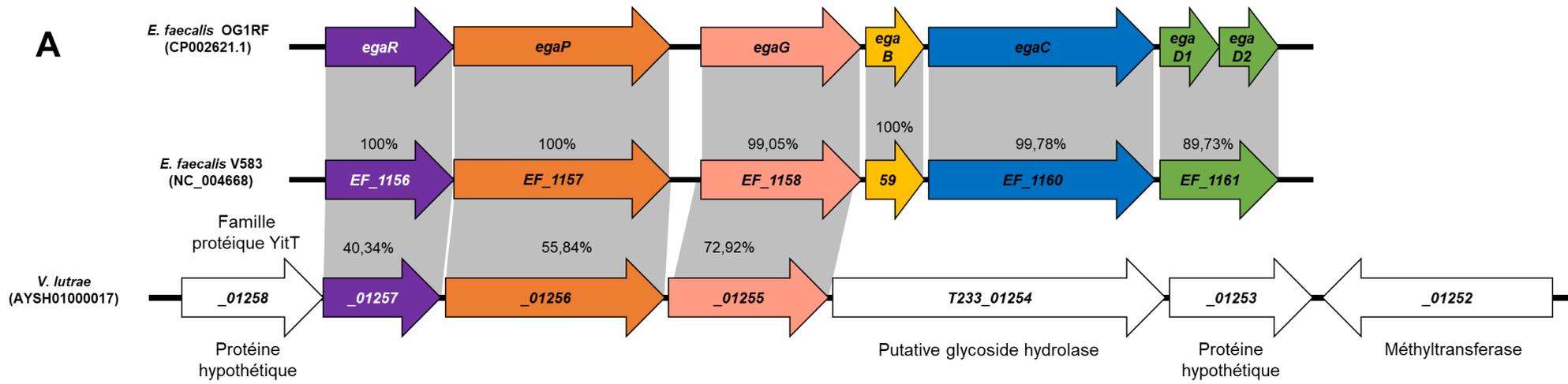


Figure 31 : Séquence du gène *egaD* et structure des protéines *EgaD*. **A** : Analyse de la séquence nucléotidique des gènes *egaD1* et *egaD2* chez *E. faecalis* OG1RF. Les codons starts sont indiqués en rouge. Les codons stops sont indiqués en vert. La séquence d'*egaD1* est soulignée en bleu. La séquence d'*egaD2* est soulignée en jaune. **B** : Comparaison de la structure primaire de la protéine *EgaD* entre *E. faecalis* OG1RF, V583 et *Clostridium* sp. Les domaines d'intérêts sont représentés par des rectangles. Le peptide signal est représenté en violet. Le domaine effecteur GrdB est indiqué en orange. Les parallélépipèdes verts représentent les zones conservées des domaines effecteurs.

Deux gènes sont présents en amont de *egaG*. Ces gènes nommés *egaR* (OG1RF_10933) et *egaP* (OG1RF_10934) codent pour des protéines annotées respectivement comme étant un régulateur transcriptionnel de la famille GntR (facteur de transcription), et une dipeptidase Mn²⁺ dépendante. Entre ces deux gènes et l'opéron contenant *egaG* se trouve une région intergénique de 169 pb (**Figure 30B**). L'analyse de cette région a permis d'identifier différents motifs d'ADN. Une séquence inversée répétée trouvée en aval de la peptidase semble indiquer la présence d'un terminateur de transcription putatif de 68 pb. Cependant, l'analyse de cette séquence par des sites de prédiction de terminateurs de transcription (ARNold, Genome2D) ne semble pas révéler la présence de ce genre de structure et il n'a pas été identifié non plus dans l'analyse transcriptomique menée chez *E. faecalis* V583 par Michaux et ses

collaborateurs [214]. Néanmoins, la présence d'un motif de type inversé-répété parfait d'une longueur de 19 nucléotides suivi par d'une répétition de T nous permet de calculer un ΔG de -28.4 kcal/mol et ce malgré une boucle de 30 nucléotides. Cette région intergénique semble aussi contenir une région promotrice. Une boîte -10 potentielle semble être présente 46 pb en amont du codon start de la glycosylasparaginase. Deux motifs ADN sont retrouvés à proximité de cette boîte : un motif de fixation à l'ADN appartenant à la sous-famille HutC et un motif *cre* imparfait.

Des homologues du locus *ega* d'*E. faecalis* OG1RF organisés de manière identique ou en partie identique ont été trouvés chez la souche V583, *Vagococcus lutrae* et *Lactocaseibacillus casei* BL23 (**Figures 32A et 32B**) suggérant que les produits des gènes *ega* pourraient être liés sur le plan fonctionnel.



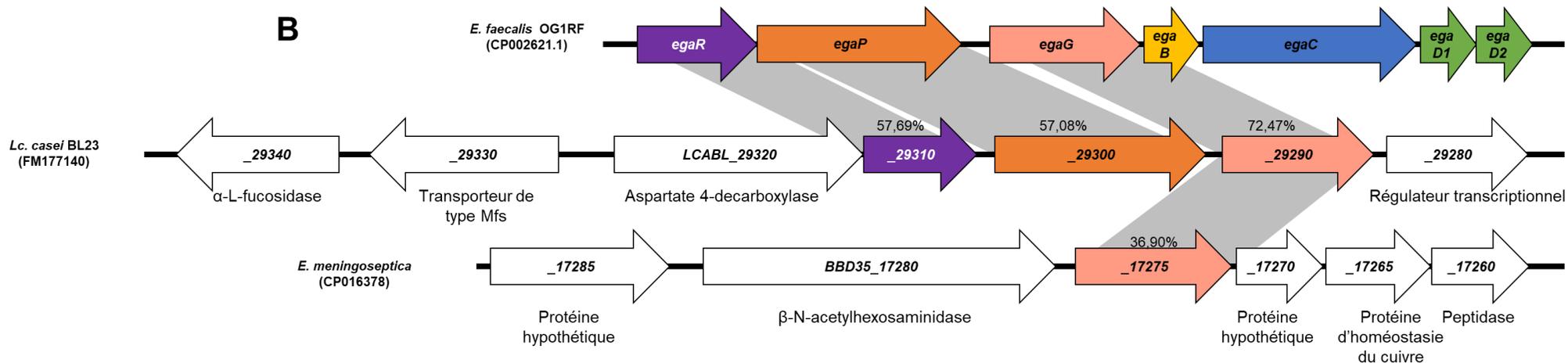


Figure 32 : Comparaison de l'organisation structurelle du locus *ega*. **A :** Organisation structurelle du locus *ega* chez la souche *E. faecalis* V583 et *Vagococcus lutrae*. **B :** Organisation structurelle du locus glycosylasparaginase chez *Lc. casei* BL23 et *E. meningoseptica*. Les flèches de couleurs représentent les gènes du locus *ega* ainsi que leurs homologues retrouvés chez différents microorganismes. Le pourcentage d'identité des séquences d'AA des protéines apparentées est indiqué au-dessus de l'homologue en question.

1.2 Le locus *ega* est organisé en deux structures opéroniques

Bien qu'un seul site d'initiation de la transcription (TSS) en amont d'*egaR* ait été identifié dans l'analyse globale des TSS d'*E. faecalis* V583 (Michaux et collaborateurs), l'organisation générale des gènes *ega* et la présence d'un terminateur transcriptionnel putatif entre les gènes *egaP* et *egaG*, suggèrent qu'ils sont probablement séparés en deux unités transcriptionnelles [214]. Nous avons confirmé expérimentalement cette organisation sous la forme de deux structures opéroniques indépendantes par 5'RACE-PCR (**Figure 33**) et RT-PCR (**Figure 34**). Les promoteurs P1 et P2 ont été cartographiés respectivement à 49 et 34 pb en amont des codons de départ d'*egaR* et d'*egaG* (**Figure 30**). Le TSS du locus *ega* identifié précédemment [214] est identique au TSS P1 chez OG1RF. Les résultats de l'analyse de RT-PCR montrent que ce locus est bien organisé en deux structures opéroniques avec *egaRP* d'une part et *egaGBCD1D2* d'autre part. En effet, aucune amplification n'a pu être mise en évidence entre *egaP* et *egaG* alors qu'une bande de taille identique au contrôle ADN génomique a pu être obtenue pour chacune des autres PCR.

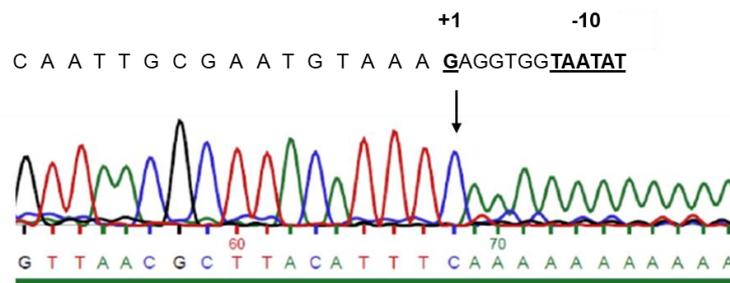


Figure 33 : 5'RACE-PCR dans l'espace intergénique compris entre *egaP* et *egaG*.

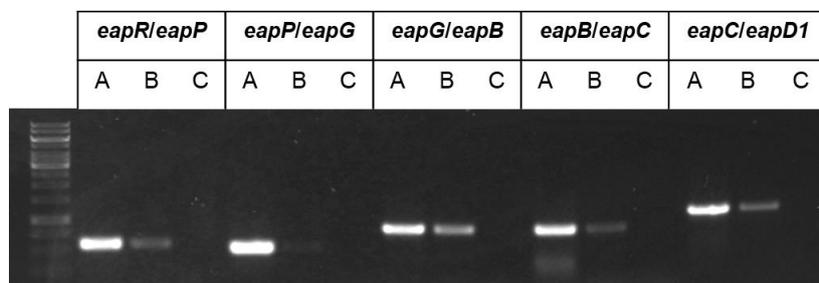


Figure 34 : Analyse par RT-PCR de la co-transcription des gènes de l'opéron *ega*. **A** : Contrôle positif (ADNg), **B** : ADNg, **C** : Contrôle négatif (ARN traité à la DNase et non rétro-transcrit).

1.3 Séquençage entier du génome de la souche *ΔegaG*

Afin de s'assurer que le phénotype observé pour la souche *ΔegaG* est bien dû à la délétion de ce gène, un séquençage intégral du chromosome a été effectué. Le séquençage a produit un résultat contenant 14.538.084 et 13.928.060 « paired-end reads sets » pour les souches WT et *ΔegaG* respectivement. Les différences entre les souches WT et *ΔegaG* sont répertoriées dans le **tableau 9**.

Tableau 9 : Résultat de l'analyse de détection de mutation dans la souche *ΔegaG*. VNM : Variant nucléotidique multiple ; VNS : Variant nucléotidique simple.

Position dans le génome	Annotation	Type de variant	Modification des nucléotides et acides aminés	Commentaire
978560 / 978561	<i>egaG</i> /OG1RF N4-(beta-N-acetylglucosaminy)-L-asparaginase	VNM	ACG en GAA Thr en Glu	Changement d'acide aminé ayant lieu dans la partie C-ter non tronquée du gène muté. N'impacte pas le cadre de lecture.
1460011	<i>apt</i> /OG1RF_11400 adenine phosphoribosyltransferase	VNS	TAC en TAT Tyr en Tyr	Mutation silencieuse

En dehors de la délétion désirée au sein du gène *egaG*, deux mutations sont présentes : une provoquant une mutation silencieuse au niveau du gène *apt*, et l'autre changeant un acide aminé dans la partie C-ter non tronquée du gène *egaG*. Etant donné que cette dernière mutation ne provoque pas de changement dans le cadre de lecture, nous pouvons conclure que les phénotypes qui seront observés pour la souche *ΔegaG* seront dû à l'effet de la délétion du gène *egaG*.

1.4 Suivi de croissance des différents mutants *ega*

Pour effectuer une analyse fonctionnelle des gènes *ega*, des mutants ont été construits par délétion dans le cadre ouvert de lecture de la partie centrale de chaque gène dans la souche OG1RF, à l'exception de *egaD1* et *egaD2*, qui sont le résultat d'une mutation naturelle. En raison du rôle attendu de la glycosylasparaginase, la souche OG1RF et les différents mutants ont été cultivés en milieu cdM17 supplémenté avec des molécules imitant les produits libérés lors de la déglycosylation des protéines (**Figure 35**).

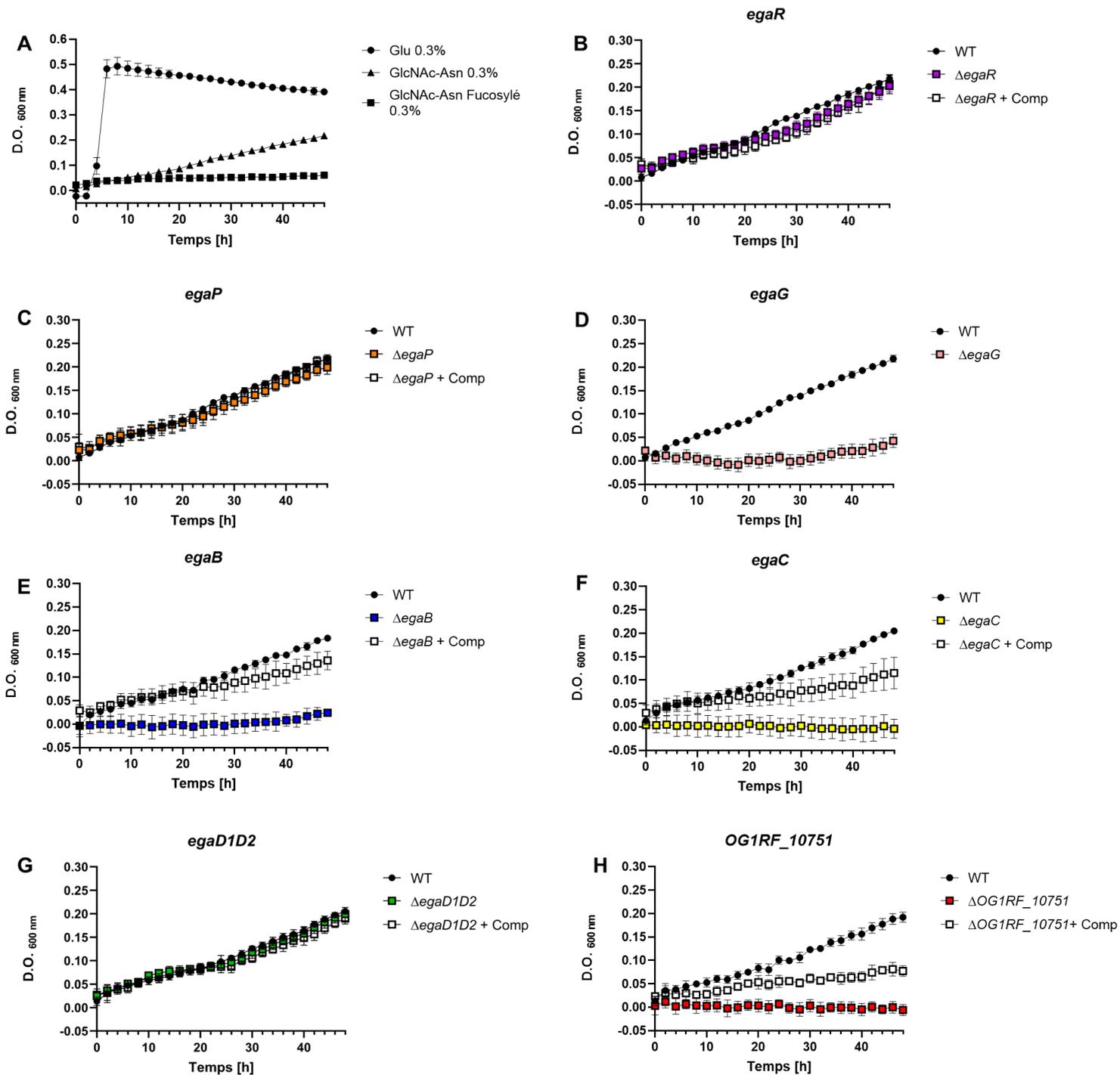


Figure 35 : Etude de la croissance de la souche sauvage et des différents mutants *ega*. **A** : Croissance de la souche *OG1RF* WT d'*E. faecalis* sur 0,3 % de glucose (cercle), 0,3 % de GlcNAc-L-Asn (triangle) ou 0,3 % de GlcNAc-L-Asn Fucosylé (carré). Croissance avec 0.3 % de GlcNAc-L-Asn de la souche Δ *egaR* (**B**), de la souche Δ *egaP* (**C**), de la souche Δ *egaG* (**D**), de la souche Δ *egaB* (**E**), de la souche Δ *egaC* (**F**), de la souche Δ *egaD1D2* (**G**), de la souche Δ *OG1RF_10751* (**H**). **B-H** : Les souches WT sont indiquées par un cercle, les souches mutées par un carré de couleur et les souches complémentées par un carré blanc.

Deux molécules ont été testées comme substrats, toutes deux contenant la structure centrale du N-glycane, GlcNAc, lié à une asparagine (2-acétamido-2-déoxy- β -D-glucopyranosyl L-asparagine, GlcNAc-L-Asn). La différence entre les deux substrats est la présence ou non d'un résidu fucose lié au GlcNAc. Comme le montre la **Figure 35A**, la souche OG1RF est capable de se développer uniquement en présence de GlcNAc-L-Asn. Aucune croissance n'a été observée avec le glycane fucosylé, indiquant que la bactérie ne peut pas internaliser ou métaboliser ce composé.

Les mutants Δ *egaG*, Δ *egaB* et Δ *egaC* cultivés en présence de GlcNAc-L-Asn présentent des défauts de croissance, suggérant que la glycosylasparaginase et les composants PTS EIIB et EIIC sont impliqués dans l'utilisation de ce composé (**Figures 35D, 35E et 35F**). L'inactivation des gènes *egaR*, *egaP* et *egaD1D2* n'a pas eu d'impact sur la croissance des mutants correspondants (**Figures 35B, 35C et 35G**) : ce résultat indique que les protéines EgaR, EgaP et EgaD1D2 ne semblent pas directement impliquées dans l'utilisation du composé GlcNAc-L-Asn et est cohérent avec le rôle présumé de EgaR comme répresseur transcriptionnel, celui de EgaP comme peptidase, et du rôle autre que réductase des protéines EgaD1D2 comme supposé initialement.

1.5 Identification du PTS EIIA

Les protéines EgaB et EgaC sont deux éléments d'un transporteur PTS appartenant à la classe du lactose/cellobiose. Nous avons montré lors des expériences de croissance que ces deux protéines étaient nécessaires au transport du GlcNAc-L-Asn. L'analyse de la séquence du génome a révélé un seul gène codant pour un composant EIIA, appartenant à la même classe que les protéines EgaBC. Cette protéine EIIA, OG1RF_10751, a déjà été montrée comme complémentaire des composants EIIB/EIIC catalysant l'absorption du gentiobiose, un 1,6- β -diglucoside [5]. La croissance en présence de GlcNAc-L-Asn du mutant Δ OG1RF_10751 et celle de la souche complémentée ont donc été analysées (**Figure 35H**). Ce mutant ne croit pas sur GlcNAc-L-Asn dans ces conditions, alors que la complémentation restaure partiellement la croissance. Ce résultat suggère fortement qu'avec EgaB et EgaC, la protéine EIIA^{Cel} OG1RF_10751 catalyse le transport et la phosphorylation de GlcNAc-L-Asn médiés par le PTS.

1.6 Identification du produit de déglycosylation utilisable par le locus *ega*

Afin de confirmer que le locus *ega* est requis pour l'utilisation de GlcNAc-L-Asn par *E. faecalis*, l'expression des six gènes (*egaR*, *egaP*, *egaG*, *egaB*, *egaC* et *egaD1*) a été évaluée par RT-qPCR avec des cellules cultivées en milieu cdM17 supplémenté avec 0,3% de glucose ou de GlcNAc-L-Asn. Il faut noter que l'expression de tous les gènes est très faible lorsque les cellules sont cultivées sur glucose seul. La croissance sur GlcNAc-L-Asn a fortement induit l'expression des gènes *ega* (**Figure 36**). Par rapport aux cellules cultivées sur glucose, l'expression des gènes *egaR* et *egaP* est multipliée respectivement par des valeurs comprises entre 20 et 30, et celle des gènes *egaG*, *egaB*, *egaC* et *egaD1* par des valeurs comprises entre 200 et 300. Cette différence de niveau d'expression entre *egaRP* et *egaGBCD1* confirme bien l'organisation générale en deux opérons montrée par les expériences de RT-PCR.

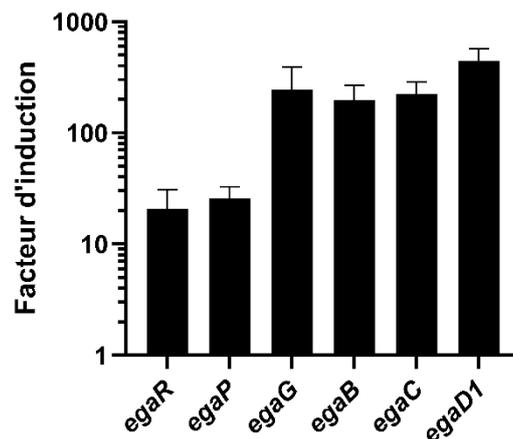


Figure 36 : Expression relative des gènes *egaR*, *egaP*, *egaG*, *egaB*, *egaC* et *egaD1* dans la souche sauvage OG1RF, déterminée par RT-qPCR. La souche sauvage a été cultivée dans un milieu cdM17 contenant 0,3 % de glucose ou 0,3 % de GlcNAc-L-Asn. Le facteur d'induction a été calculé en divisant le niveau d'expression déterminé pour un mutant pendant la croissance sur GlcNAc-L-Asn par le niveau d'expression observé pendant sa croissance sur glucose.

1.7 EgaR est un répresseur transcriptionnel du locus *ega*

La localisation constante du gène *egaR* à proximité d'*egaG* suggère que EgaR pourrait être le régulateur de l'expression de l'opéron *egaGBCD1D2*. EgaR est annoté comme appartenant aux régulateurs de transcription de la famille GntR. Ces régulateurs fonctionnent généralement comme des répresseurs et contrôlent l'expression de nombreux gènes impliqués dans les voies métaboliques [70,215–217]. L'analyse structurale de la protéine indique la présence de deux domaines en N-terminal et C-terminal (**Figure 37**).

- A**
- | | |
|--|---|
| <p>Domaine de fixation à l'ADN :</p> <ul style="list-style-type: none"> - Liaison à l'ADN - Comporte l'hélice de reconnaissance | <p>Domaine régulateur :</p> <ul style="list-style-type: none"> - Fixation du ligand ? - Comporte le domaine chorismate lyase |
|--|---|

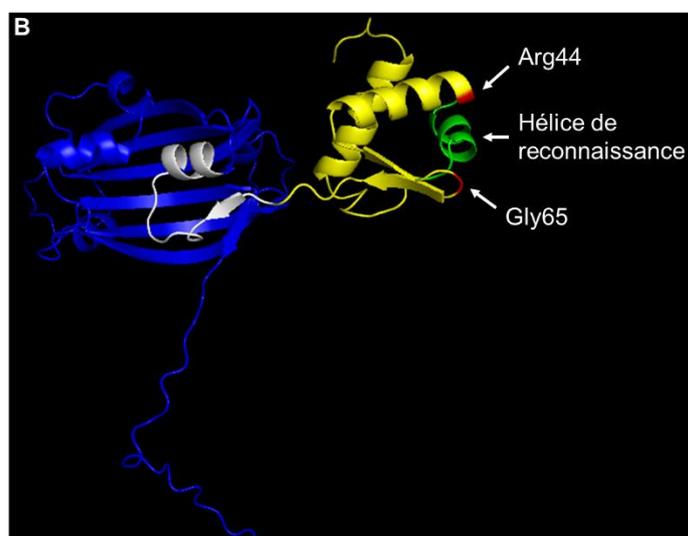
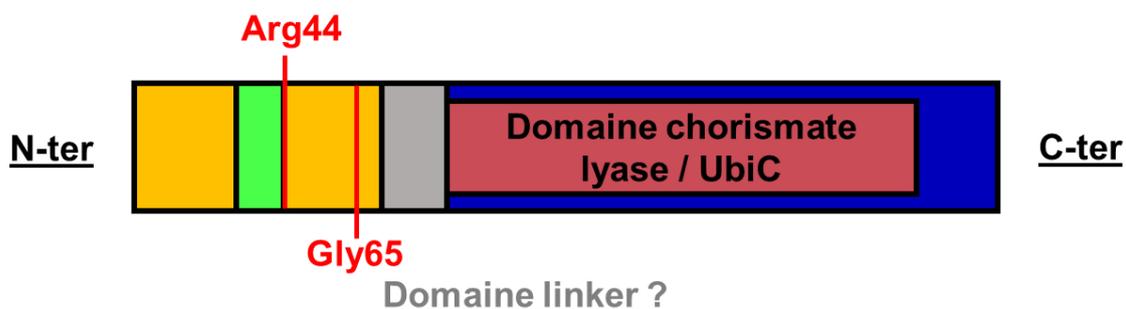


Figure 37 : Structure de la protéine EgaR. **A :** Schéma de la structure primaire. Les domaines sont représentés par des blocs de couleurs jaune, grise et bleue. Les AA indiqués en rouge correspondent à ceux de YvoA impliqués dans la liaison à l'ADN **B.** Structure tridimensionnelle. Les domaines et AA d'intérêts sont représentés selon le même code couleur que dans la figure 3A. Par soucis de clarté, le domaine chorismate lyase n'a pas été représenté ici.

La partie N-terminale contient un domaine de fixation à l'ADN de 68 AA. Les AA Arg44 et Gly65 correspondent aux AA Arg48 et Gly69 décrits précédemment comme étant impliqués dans la liaison à l'ADN pour la protéine YvoA. La position de l'hélice de reconnaissance est également similaire à celle retrouvée pour YvoA, suggérant un rôle analogue. Entre les parties N-terminale et C-terminale se trouve une région de 19 AA ne possédant aucun domaine fonctionnel. Au vu de sa position, et de sa longueur, nous supposons que cette séquence correspond au domaine linker. La partie C-terminale contient un domaine associé au régulateur de transcription UbiC (UTRA), un domaine de liaison au ligand conservé qui

modulerait l'activité des facteurs de transcription bactériens en réponse à la liaison de petites molécules telles que les sucres phosphorylés [218]. Ce domaine est caractéristique de la sous-famille HutC, indiquant donc que le régulateur EgaR appartient à cette sous-famille. Afin de confirmer expérimentalement le rôle d'EgaR dans la régulation de l'opéron *egaGBCD*, nous avons mesuré les niveaux d'expression des gènes du locus glycosylasparaginase par RT-qPCR dans le mutant de délétion *egaR* de la souche OG1RF (**Figure 38**) : l'expression de tous les gènes est significativement plus élevée chez le mutant comparé à la souche sauvage, en accord avec une levée de la répression. Cette surexpression est également observée dans le cas du gène *egaP*, suggérant que EgaR autorégule son expression. La complémentation avec l'allèle de type sauvage *egaR* rétablit la répression de tous les gènes (**Figure 38**).

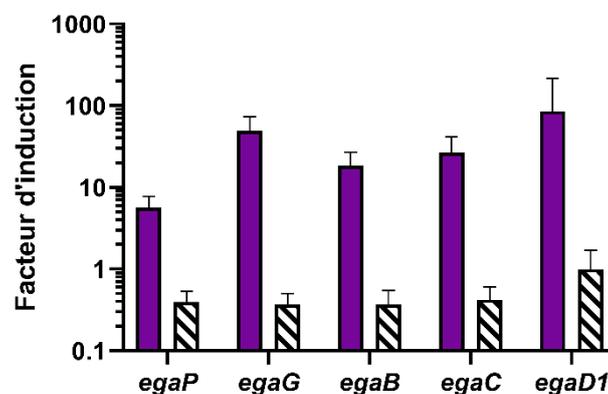


Figure 38 : Effet du GlcNAc-Asn sur l'expression des gènes *ega*. Expression des gènes *egaP*, *egaG*, *egaB*, *egaC* et *egaD1* dans la souche Δ *egaR* (histogrammes violets) ou dans la souche Δ *egaR* complétementé (histogrammes hachurées), cultivées avec du glucose, par rapport à leur expression dans la souche de type sauvage cultivée avec glucose. Le facteur d'induction représente le rapport de l'expression d'un gène dans Δ *egaR* ou Δ *egaR* complétementé divisé par son expression dans la souche de type sauvage. Les valeurs moyennes calculées à partir de trois expériences indépendantes sont présentées. Les écarts-types sont indiqués.

Pour montrer que la protéine EgaR interagit directement avec les deux régions promotrices du locus *ega*, la protéine EgaR purifiée a été utilisée pour des expériences de retard sur gel (EMSA). Trois fragments d'ADN marqués par fluorescence et contenant soit les régions promotrices d'*egaRP* (P1) ou d'*egaGBCD* (P2), soit une petite partie du gène *gus* (ADN de contrôle) ont été incubés en présence d'EgaR. Des retards de migration sont uniquement apparus pour les deux régions promotrices et pour des concentrations d'EgaR supérieures à 200 nM, suggérant une interaction spécifique directe entre EgaR et les régions d'ADN

promotrices (**Figure 39**). Ces interactions sont considérées comme spécifiques car la protéine n'interagit pas avec l'ADN contrôle, indiquant qu'EgaR régule spécifiquement l'expression des opérons *ega*.

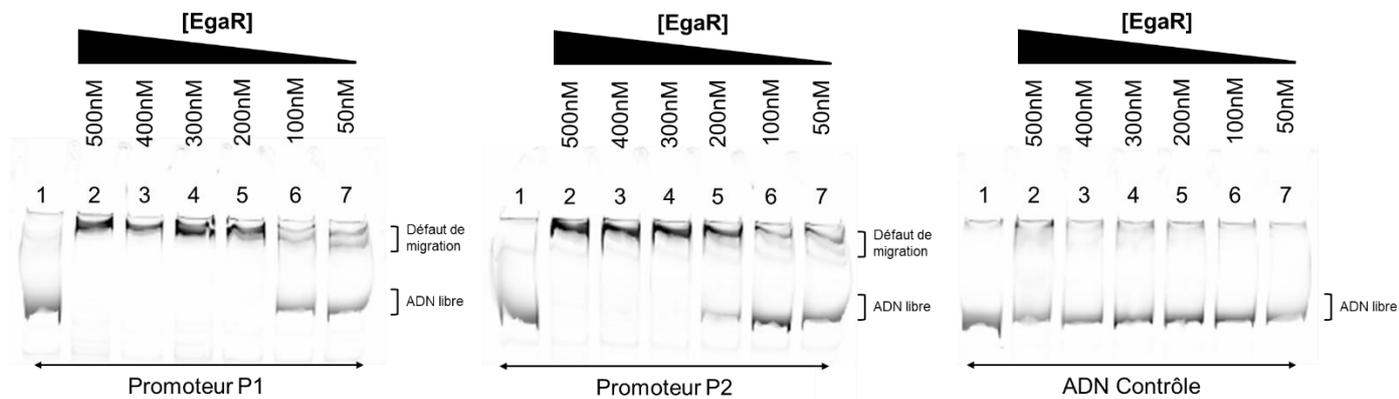


Figure 39 : *Expériences de retard sur gel des régions promotrices P1 et P2 en présence d'EgaR. Les EMSA ont été réalisées avec 5 nM d'ADN marqué et différentes concentrations d'EgaR allant de 500 nM à 50 nM (colonnes 2 à 7) comme indiqué au-dessus du gel. Un contrôle négatif a été effectué avec un fragment d'ADN endogène du gène gus. L'ADN sans protéine a été déposé sur la piste 1.*

Le motif de liaison à l'ADN de la sous-famille HutC/GntR a été déterminé par Suvorova et collaborateurs [64]. Il est composé d'un motif répété conservé de 8 pb, 5'– GTMTAKAC – 3', séparé par 5-6 pb [64]. De tels motifs conservés sont présents dans les régions promotrices du locus *ega* (**Figure 30**). Afin de mettre en évidence le rôle de ce motif dans l'interaction avec le régulateur EgaR, nous avons réalisé des versions modifiées des régions promotrices comportant les deux motifs HutC mutés (**Figures 40A et 40B**). Les nucléotides conservés ont été remplacés par des nucléotides non complémentaires (A et T -> C/G ; C et G -> A/T). Les essais de liaison avec EgaR en présence d'un motif répété HutC muté ne présentent aucun défaut de migration visible, ce qui indique que la protéine EgaR n'est plus capable de se fixer et semble se lier spécifiquement aux motifs 5'– GTMTAKAC – 3' situés à proximité de la boîte -10 du site d'initiation de la transcription (**Figure 40C**). Ceci est cohérent avec la fonction de répresseur de la protéine EgaR. Des expériences de DNA Footprinting ont été entreprises sans nous permettre d'obtenir de résultat.

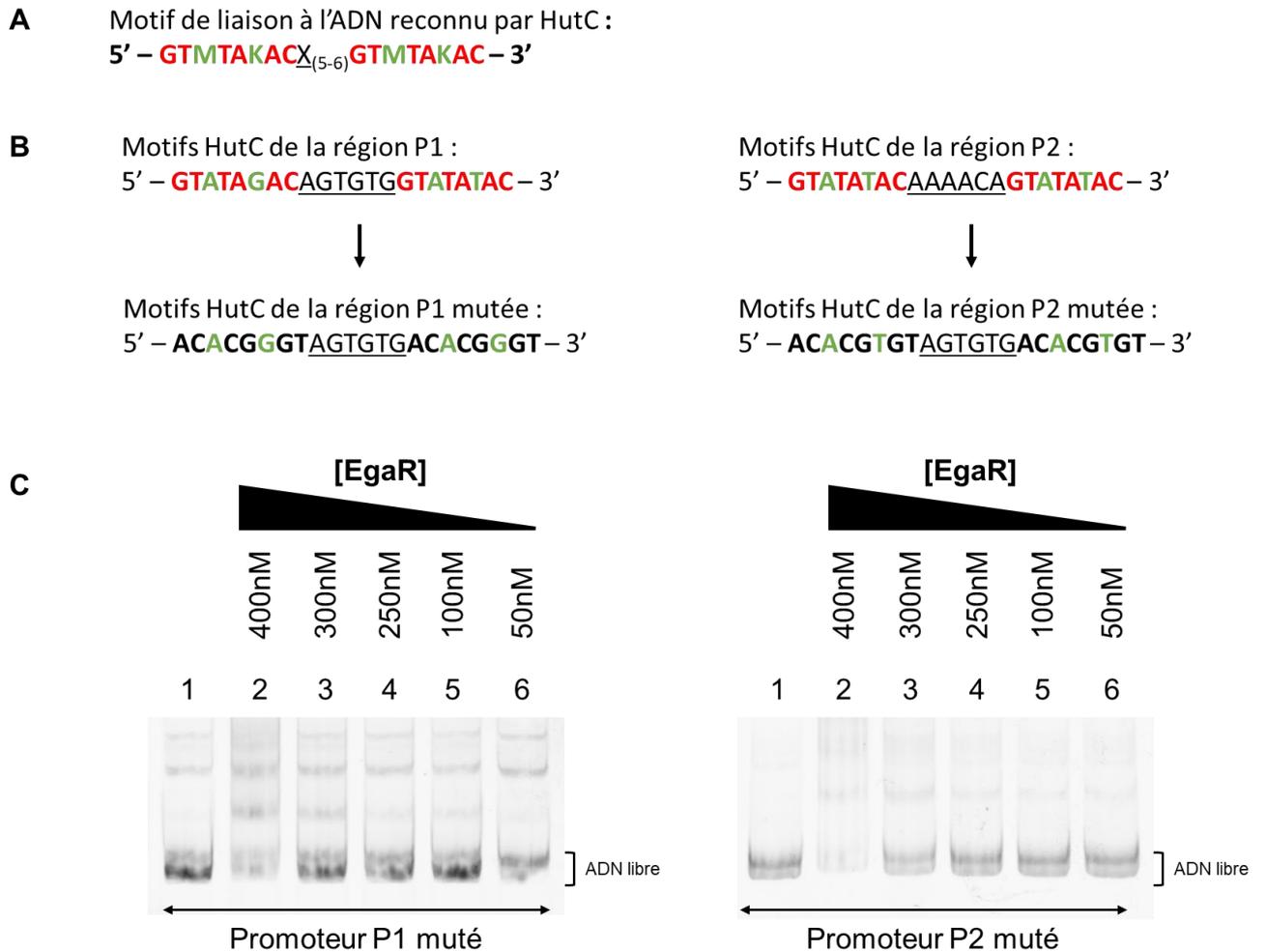


Figure 40 : Expériences de retard sur gel avec les motifs HutC mutés pour les régions promotrices P1 et P2 et la protéine EgaR. **A** : Motif de liaison à l'ADN reconnu par les régulateurs transcriptionnels de la sous-famille HutC. En rouge sont indiqués les nucléotides conservés. En vert sont indiqués les nucléotides pouvant varier. M indique la présence d'une adénine ou cytosine. K indique la présence d'une guanine ou thymine. En noir souligné est représentée la portion ADN séparant les deux motifs de liaison HutC. **B** : Représentation des régions promotrices comportant des motifs HutC non mutés ou mutés. Le code couleur est identique à A. **C** : Le retard sur gel a été réalisé avec 5 nM d'ADN marqué et une concentration de protéine EgaR allant de 400 nM à 50 nM (colonnes 2 à 6) comme indiqué au-dessus du gel. L'ADN sans protéine a été déposé sur la piste 1.

Dans le cas de la protéine EgaR, le GlcNAc-L-Asn phosphorylé pendant le transport catalysé par le PTS devrait se lier au répresseur et l'empêcher d'interagir avec son site opérateur. Cependant, le composé phosphorylé n'est pas disponible dans le commerce : le retard sur gel a donc été réalisé en présence de GlcNAc-L-Asn non phosphorylé (**Figure 41**).

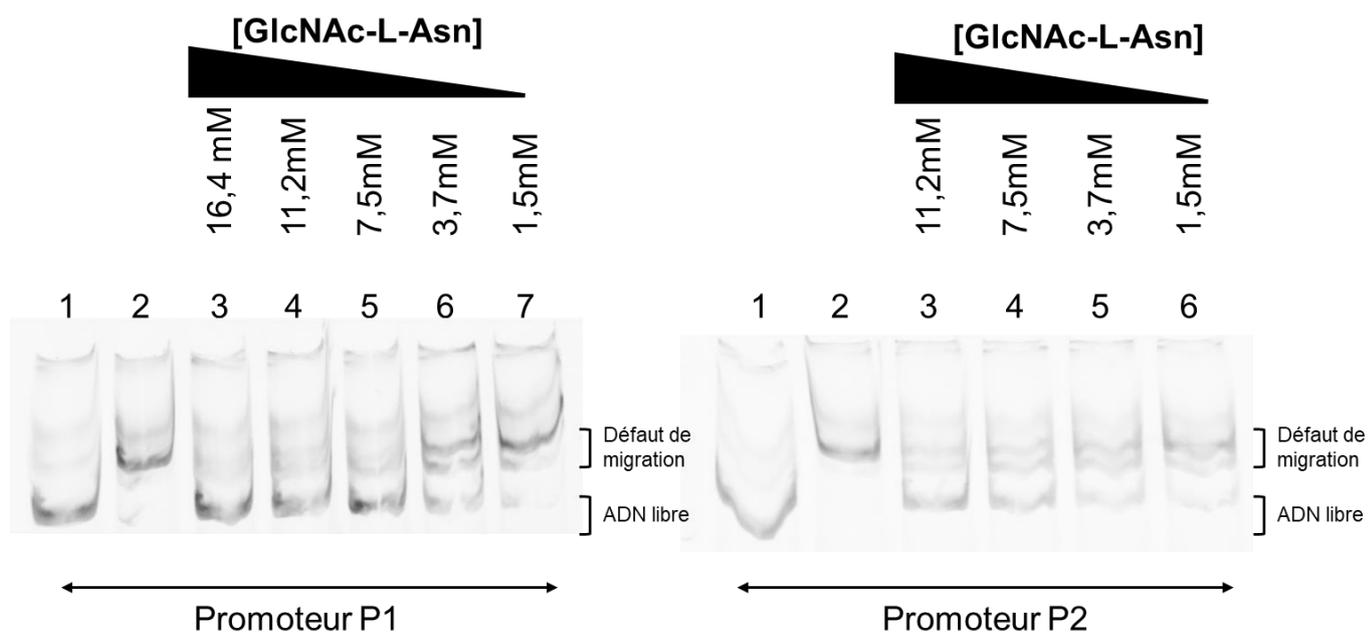


Figure 41 : EMSA des régions promotrices P1 et P2 en présence de GlcNAc-L-Asn, avec la protéine EgaR. Les EMSA ont été réalisées avec 5 nM d'ADN marqué, 250 nM d'EgaR et différentes concentrations de GlcNAc-L-Asn allant de 16,4 mM à 1,5 mM (colonnes 3 à 7) comme indiqué au-dessus du gel. L'ADN sans EgaR et sans GlcNAc-L-Asn a été déposé dans la colonne 1 et l'ADN avec 250 nM de la protéine EgaR mais sans GlcNAc-L-Asn sur la colonne 2.

En partant d'une condition (250 nM EgaR) où tout l'ADN présente un retard de migration (**Figure 41** piste 2), des quantités croissantes de GlcNAc-L-Asn ont été ajoutées. Lorsque la concentration de GlcNAc-L-Asn atteint 7,5 mM, l'ADN ne montre plus de défaut de migration. Cela confirme que le GlcNAc-L-Asn empêche EgaR de se lier à la région promotrice des deux opérons *ega*. Ce résultat est cohérent avec l'induction des gènes *ega* observée lors de la RT-qPCR pour les cellules cultivées en présence de GlcNAc-L-Asn (**Figure 36**). Il est intéressant de noter qu'alors que les hydrates de carbone transportés par le PTS ne fonctionnent généralement comme inducteurs que lorsqu'ils sont phosphorylés, le GlcNAc-L-Asn a empêché l'interaction EgaR/ADN lorsqu'il n'était pas phosphorylé. De plus, comme EgaR est un membre de la sous-famille HutC, connue pour lier principalement des AA, nous pouvons supposer que la liaison de GlcNAc-L-Asn à EgaR est principalement médiée par la partie asparagine. Pour tester cette possibilité, des EMSA ont été réalisées avec de l'asparagine ou du GlcNAc (**Figure 42**). Quel que soit la concentration en GlcNAc ou asparagine utilisée, un retard de migration est observé : cela indique que les deux parties de GlcNAc-L-Asn sont nécessaires à sa liaison à EgaR.

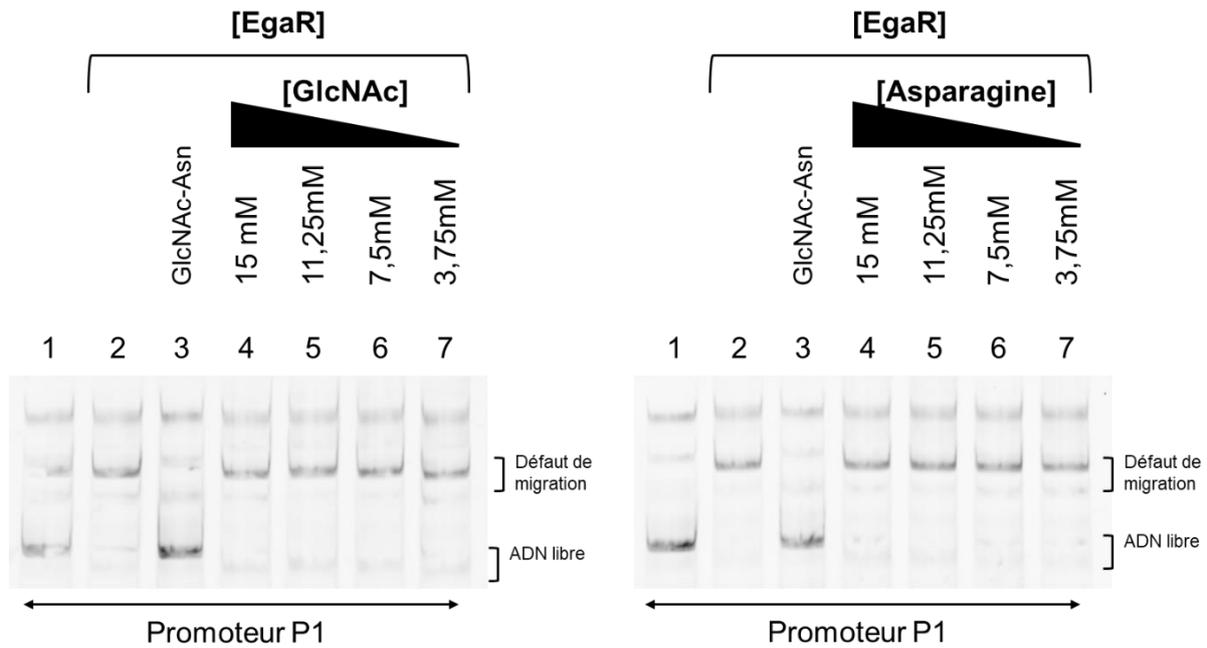


Figure 42 : Expériences de retard sur gel de la région promotrice P1 en présence de GlcNAc ou Asparagine, avec la protéine EgaR. Les EMSA ont été réalisées avec 5 nM d'ADN marqué, 250 nM de la protéine EgaR et différentes concentrations de GlcNAc ou Asparagine allant de 15 mM à 3,75 mM (colonnes 4 à 7) comme indiqué au-dessus du gel. L'ADN sans EgaR et sans GlcNAc-L-Asn a été déposé dans la piste 1. L'ADN avec 250 nM de la protéine EgaR sans GlcNAc-L-Asn a été déposé dans la piste 2. L'ADN avec 250 nM de la protéine EgaR et avec 7,5 mM de GlcNAc-L-Asn a été déposé dans la piste 3.

1.8 Répression catabolique du locus glycosylasparaginase

L'analyse de la région intergénique entre *egaRP* et *egaGBCD1D2* (**Figure 30B**) montre qu'elle contient une boîte *cre* imparfaite où le dernier nucléotide de la séquence 5' – WTGWAARCGYWWWCW – 3' n'est pas retrouvé. La présence de ce motif suggère que les gènes *egaGBCD1D2* sont soumis à la répression catabolique de manière CcpA-dépendant. Afin de le confirmer expérimentalement, nous avons mesuré les niveaux d'expression des gènes par RT-qPCR dans la souche OG1RF lors d'une croissance en milieu cdM17 contenant du glucose, glucose + GlcNAc-L-Asn ou GlcNAc-L-Asn seul (**Figure 43A**). L'expression des gènes *egaGBCD1D2* en condition glucose + GlcNAc-L-Asn a été augmentée entre 2,5 et 5 fois par rapport à la condition glucose. Cette surexpression est très inférieure à celle retrouvée sur milieu GlcNAc-L-Asn seul, indiquant ainsi que les gènes *egaGBCD1D2* sont soumis à la répression catabolique. L'expression de ces gènes a également été mesurée dans la souche OG1RF délétée du gène *ccpA* (**Figure 43B**). Dans cette souche, si la répression catabolique est exercée par CcpA, alors l'expression des gènes mesurée en présence de glucose + GlcNAc-L-

Asn devrait être similaire à celle obtenue sur milieu GlcNAc-L-Asn seule. L'expression de ces gènes dans les 3 conditions testées chez la souche $\Delta ccpA$ est plus forte que pour la souche OG1RF sauvage. Cependant, les niveaux de surexpressions des gènes de la condition glucose + GlcNAc-L-Asn ne sont pas équivalents à ceux obtenus en condition GlcNAc-L-Asn seul. Nous pouvons donc en conclure que les gènes *egaGBCD1D2* sont soumis à la répression catabolique médiée d'une part par la protéine CcpA, mais également à la répression catabolique non médiée par CcpA. Malgré la présence d'une boîte *cre* dans la région promotrice, des expériences de retard sur gel n'ont pas permis de montrer la fixation de CcpA.

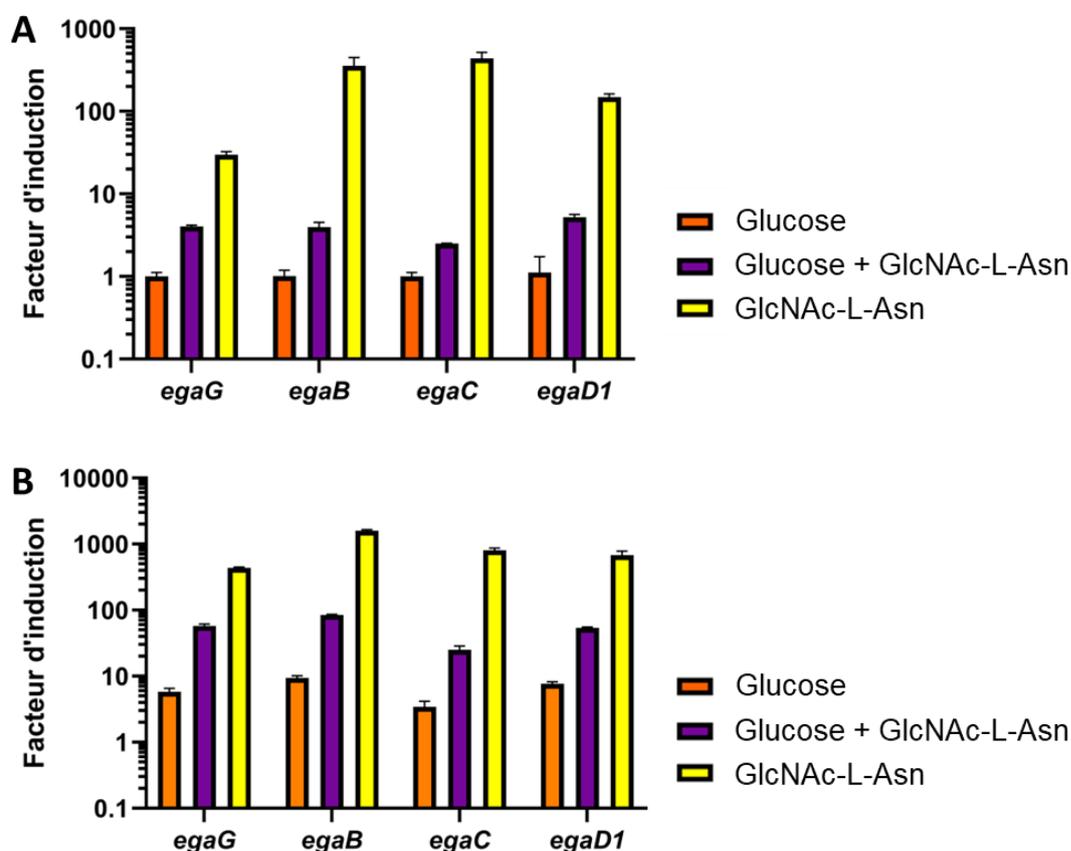


Figure 43 : Impact de la répression catabolique sur les gènes *egaG*, *egaB*, *egaC*, *egaD1*. **A** : La souche sauvage a été cultivée dans un milieu cdM17 contenant 0,3 % de glucose, ou 0,3 % de GlcNAc-L-Asn + 0,3 % glucose, ou seulement 0,3 % GlcNAc-Asn. Le facteur d'induction a été calculé en divisant le niveau d'expression déterminé d'une condition par le niveau d'expression observé pendant sa croissance sur glucose. **B** : La souche délétée du gène *ccpA* a été cultivée dans un milieu cdM17 contenant 0,3 % de glucose, ou 0,3 % de GlcNAc-L-Asn + 0,3 % glucose, ou seulement 0,3 % GlcNAc-Asn. Le facteur d'induction a été calculé en divisant le niveau d'expression déterminé pour une condition par le niveau d'expression observé dans la souche sauvage pendant sa croissance sur glucose.

2 Opéron glycosylasparaginase et virulence

2.1 EgaG code une glycosylasparaginase

Pour vérifier si EgaG présente l'activité glycosylasparaginase prédite, la pro-enzyme a été surexprimée dans *E. coli* et purifiée (**Figure 44A**). Après électrophorèse, deux bandes ont été observées sur gel SDS-PAGE avec des masses moléculaires estimées à 18 kDa et 16 kDa. L'analyse par spectrométrie de masse a permis d'identifier que chacune des bandes contient des peptides correspondant à la protéine EgaG, indiquant que ces bandes correspondent aux domaines N-terminal et C-terminal de l'enzyme active après auto-protéolyse (**Figure 44B**). Une recherche BLAST-P avec EgaG a mis en évidence une identité de séquence de 72,47 % et 38,40 % avec la glycosylasparaginase de *Lc. casei* (AsnA2) et *E. meningoseptica*, respectivement [28,29]. Les alignements de séquences ont révélé que les AA impliqués dans l'autoprotéolyse dans AsnA2 sont conservés dans EgaG (His149, Asp150 et Thr151). Ces résultats montrent bien que la glycosylasparaginase d'*E. faecalis* s'active selon un modèle identique à celui de *Lc casei* et *E. meningoseptica* et suggère que la protéine purifiée peut être fonctionnelle.

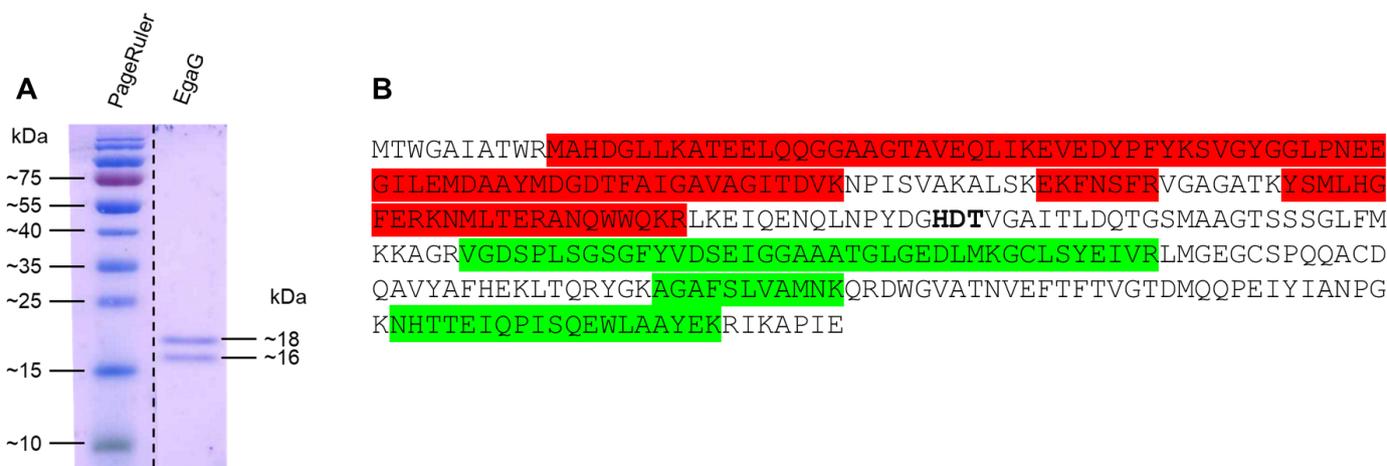


Figure 44 : Purification de la protéine glycosylasparaginase (EgaG) et analyse par spectrométrie de masse. **A** : Deux sous-unités sont obtenues après purification, indiquant que la protéine subit un processus d'auto-clivage. L'image a été obtenue à partir des résultats d'un même gel mais n'étant pas sur des puits continus ; la ligne verticale hachurée indique où l'image a été coupée. **B** : Séquence primaire de la protéine EgaG. Les séquences surlignées en vert correspondent aux fragments peptidiques retrouvés dans la bande de 18 kDa. Les séquences surlignées en rouge correspondent aux fragments peptidiques retrouvés dans la bande à 16 kDa. Les AA représentés en gras (HDT) sont conservés chez EgaG et sont ceux ayant été montrés comme impliqués dans l'auto-protéolyse.

Nous avons ensuite déterminé l'activité d'EgaG en mesurant la consommation de NADH dans un essai enzymatique couplé (voir Matériels et Méthodes) (**Figure 45**). Lorsque GlcNAc-L-Asn et EgaG sont présents dans le mélange, une forte diminution de l'absorbance est observée, indiquant que EgaG doit libérer de l'aspartate du substrat, qui est ensuite métabolisé en oxaloacétate puis en malate accompagné de la consommation de NADH (**Figure 45**).

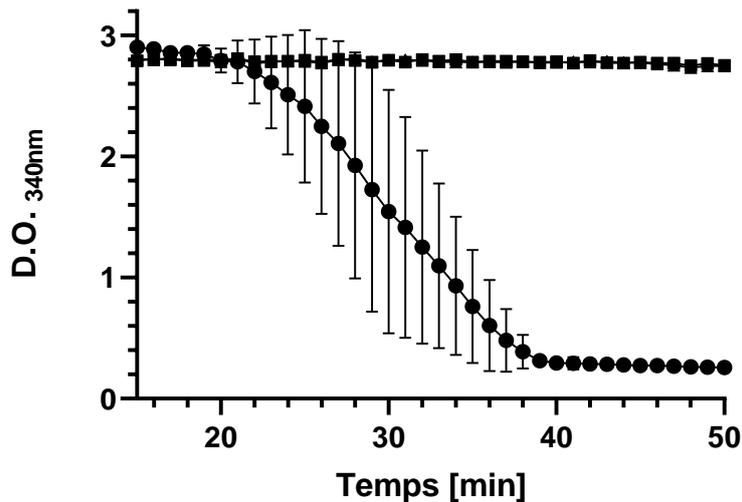


Figure 45 : Analyse cinétique de l'activité glycosylasparaginase d'EgaG. L'aspartate produit par la glycosylasparaginase est utilisé par les enzymes GOT et MDH. Cette dernière utilise le NADH comme cofacteur. La progression de la réaction est contrôlée en mesurant l'absorbance du NADH à D.O. $_{340\text{ nm}}$. L'analyse a été réalisée en présence (cercles noirs) et en l'absence de l'enzyme EgaG (carrés noirs) dans le milieu réactionnel. Le GlcNAc-L-Asn a été ajouté à une concentration finale de 6 mM. Les valeurs moyennes calculées à partir de trois expériences indépendantes sont présentées ainsi que les écarts-types.

Ce résultat indique que l'hydrolyse du GlcNAc-L-Asn doit se produire entre le NH_2 et le carbone γ de l'asparagine et libère ainsi de l'aspartate et du 1-amino-GlcNAc au lieu du GlcNAc et de l'asparagine. Pour confirmer la nature de la coupure par la glycosylasparaginase, nous avons réalisé un essai enzymatique comme indiqué en **Figure 46A**. Seul le test enzymatique réalisé avec l'aspartate comme substrat a permis d'obtenir une diminution de l'absorbance qui est semblable à celle obtenue lors du test enzymatique de la glycosylasparaginase (**Figure 46B**). Ce résultat confirme que l'hydrolyse du GlcNAc-L-Asn se produit entre le NH_2 et le carbone γ de l'asparagine et libère ainsi de l'aspartate et du 1-amino-GlcNAc au lieu du GlcNAc et de l'asparagine (**Figure 46C**). L'activité spécifique de la glycosylasparaginase d'*E. faecalis* a également pu être déterminée, et est de $19,35 \mu\text{mol}.\text{mg} \text{ protéine}^{-1}.\text{min}^{-1}$ pour le substrat GlcNAc-L-Asn.

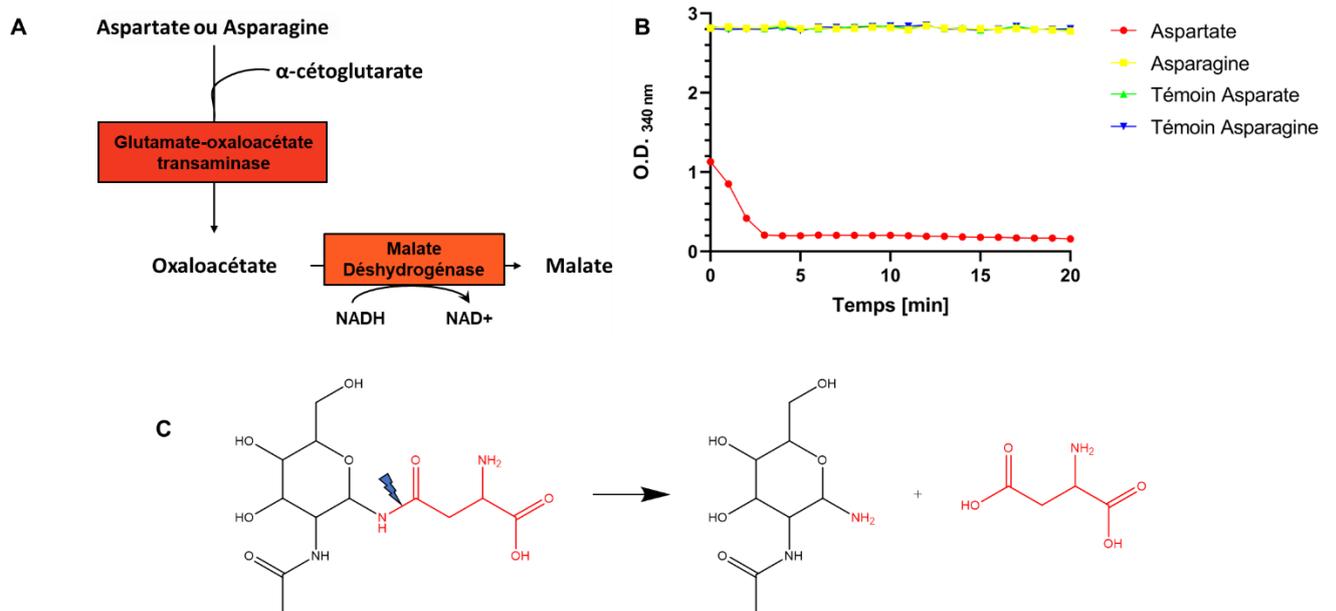


Figure 46 : Détermination du site de coupure par la glycosylasparaginase. **A** : Représentation schématique de l'essai enzymatique employé pour déterminer le site de coupure par la glycosylasparaginase. **B** : De manière similaire au test réalisé pour montrer l'activité glycosylasparaginase, un suivi d'absorbance est réalisé à D.O. $_{340\text{ nm}}$. Les témoins ont été réalisés dans un milieu réactionnel ne comportant pas la Glutamate-Oxaloacétate-Transaminase. **C** : Représentation schématique du site de coupure de la glycosylasparaginase sur le substrat GlcNAc-L-Asn. La partie noire du substrat représente le GlcNAc, et la partie rouge représente l'asparagine. La coupure, représentée par un éclair bleu, permet la libération de 1-amino-GlcNAc et aspartate.

2.2 Test d'inhibition à la streptozotocine

Dans l'objectif de caractériser l'internalisation du GlcNAc-Asn par EgaC, un test d'inhibition à la streptozotocine a été réalisé (**Figure 47**). La streptozotocine est un antibiotique ayant une structure analogue au GlcNAc. Cet antibiotique est internalisé dans la cellule bactérienne via les PTS^{Nag} et inhibe la croissance [219]. Nous avons montré précédemment que la souche ΔegaC n'était pas capable d'utiliser le GlcNAc-L-Asn. De ce fait, nous avons donc dû ajouter une source de glucose dans le milieu. Il a également été montré qu'une répression catabolique avait lieu. Cependant, en présence de glucose et GlcNAc-L-Asn, une faible surexpression des gènes du locus a été mesurée. Dans l'expérience réalisée, aucun retard de croissance n'est observé en présence ou absence de streptozotocine. Ce résultat semble indiquer que le PTS EIIC EgaC ne reconnaît pas la partie GlcNAc mais la partie asparagine du GlcNAc-L-Asn ou bien la molécule dans sa totalité pour internaliser cette molécule dans la cellule.

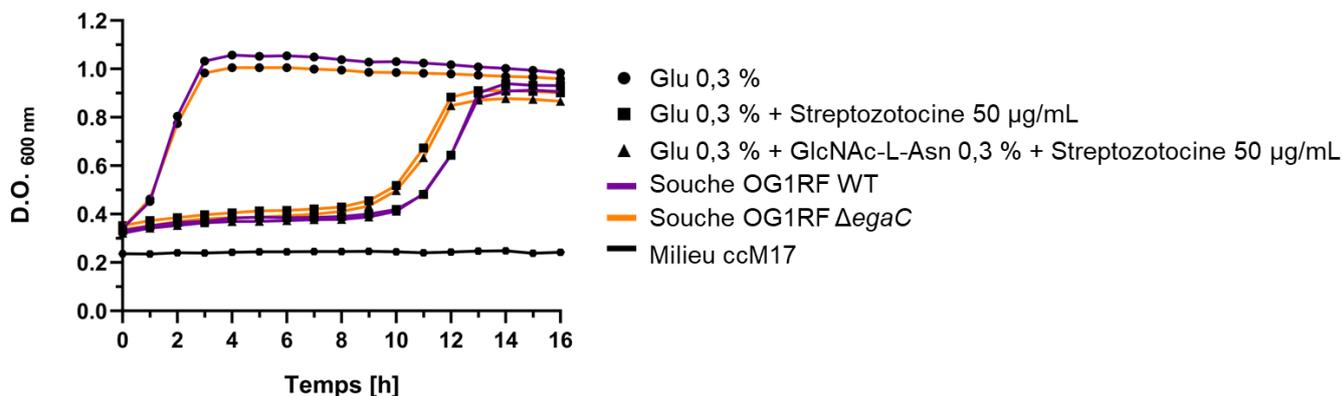


Figure 47 : Test d'inhibition à la streptozotocine. Culture des souches WT et Δ egaC en milieu ccM17 + Glu 0,3% comportant ou non de la streptozotocine 50 µg/mL et du GlcNAc-L-Asn 0,3%.

2.3 EgaP est une dipeptidase intracellulaire

Les études de croissance (**Figure 35C**) ont révélé que la délétion de la « dipeptidase-like » EgaP n'affecte pas l'utilisation du GlcNAc-L-Asn. Nous avons pour première hypothèse que cette protéine serait sécrétée dans le milieu extracellulaire et qu'elle permettrait la coupure des liaisons peptidiques de part et d'autre de l'asparagine des glycoprotéines de l'hôte pour libérer un substrat pour la glycosylasparaginase. Pour tester cette hypothèse, le test de mesure d'activité de la glycosylasparaginase a été modifié en ajoutant l'enzyme EgaP en plus de EgaG et en substituant le substrat GlcNAc-L-Asn par la glycoprotéine RNase B (**Figure 48**). Cette glycoprotéine est un substrat standard souvent utilisé pour mesurer l'activité des protéines déglycosylantes. En présence de ce composé, l'absorbance ne varie pas, indiquant qu'aucun substrat métabolisable par la glycosylasparaginase n'est libéré par EgaP. Nous avons réitéré l'expérience en dénaturant au préalable la RNase B par un traitement thermique, et en prétraitant cette dernière avec l'enzyme EgaP. Aucune diminution d'absorbance n'a été mesurée, indiquant que l'action de EgaP ne permet pas la libération de substrat utilisable dans le milieu par la glycosylasparaginase, ou en tout cas pas dans les conditions testées.

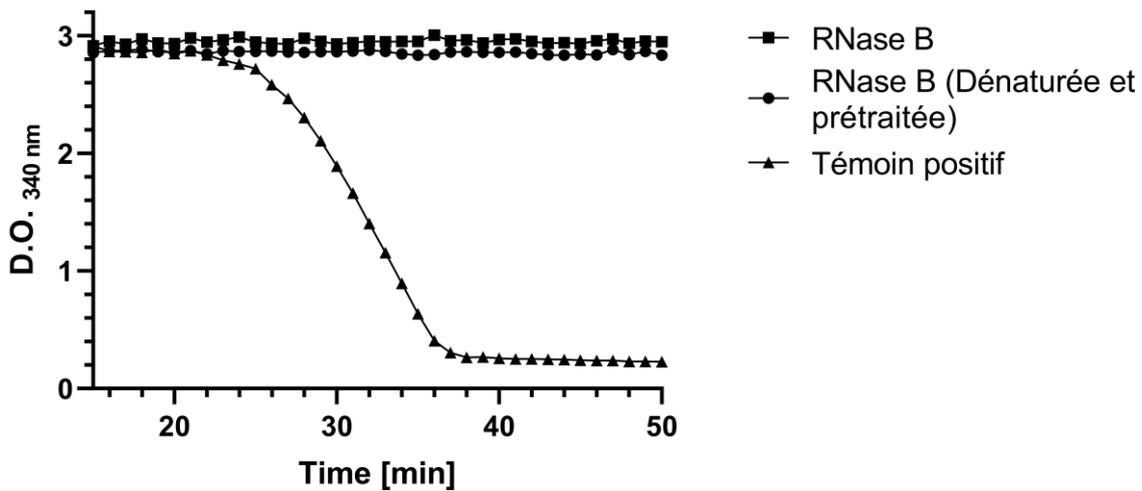
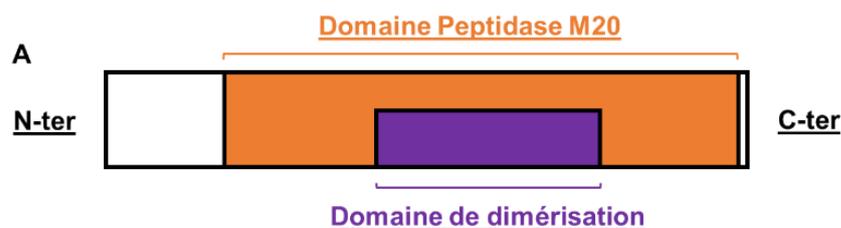


Figure 48 : Dégradation de la RNaseB par EgaP. Le test enzymatique utilisé ici est identique à celui utilisé pour mesurer l'activité de la glycosylasparaginase à la différence de la substitution du substrat GlcNAc-L-Asn par la RNase B et de l'ajout de la protéine EgaP. La mesure effectuée en présence de la RNase B et d'EgaP est représentée par les carrés noirs. La mesure effectuée en présence de la RNase B dénaturée et après pré-traitement par EgaP est représentée par les cercles noirs. Le témoin positif (triangles noirs) correspond à l'utilisation de GlcNAc-L-Asn (6 mM) comme substrat.

Au vu de ces résultats, nous nous sommes interrogés sur le rôle de cette peptidase et sa localisation. L'analyse INTERPROSCAN [220] de la protéine EgaP a révélé un domaine Peptidase M20A de 213 AA qui permettrait la dégradation de dipeptides hydrophobiques (**Figure 49A**). Quant à sa localisation, aucun peptide signal n'est détecté, laissant supposer une localisation intracellulaire. Des analyses de spectrométrie de masse du sécrétome et du protéome de la souche OG1RF WT et OG1RF Δ egaR ont mis en évidence une localisation intracellulaire de la protéine EgaP, remettant en cause l'hypothèse initiale avancée quant à son fonctionnement et sa localisation (**Figure 49B**).



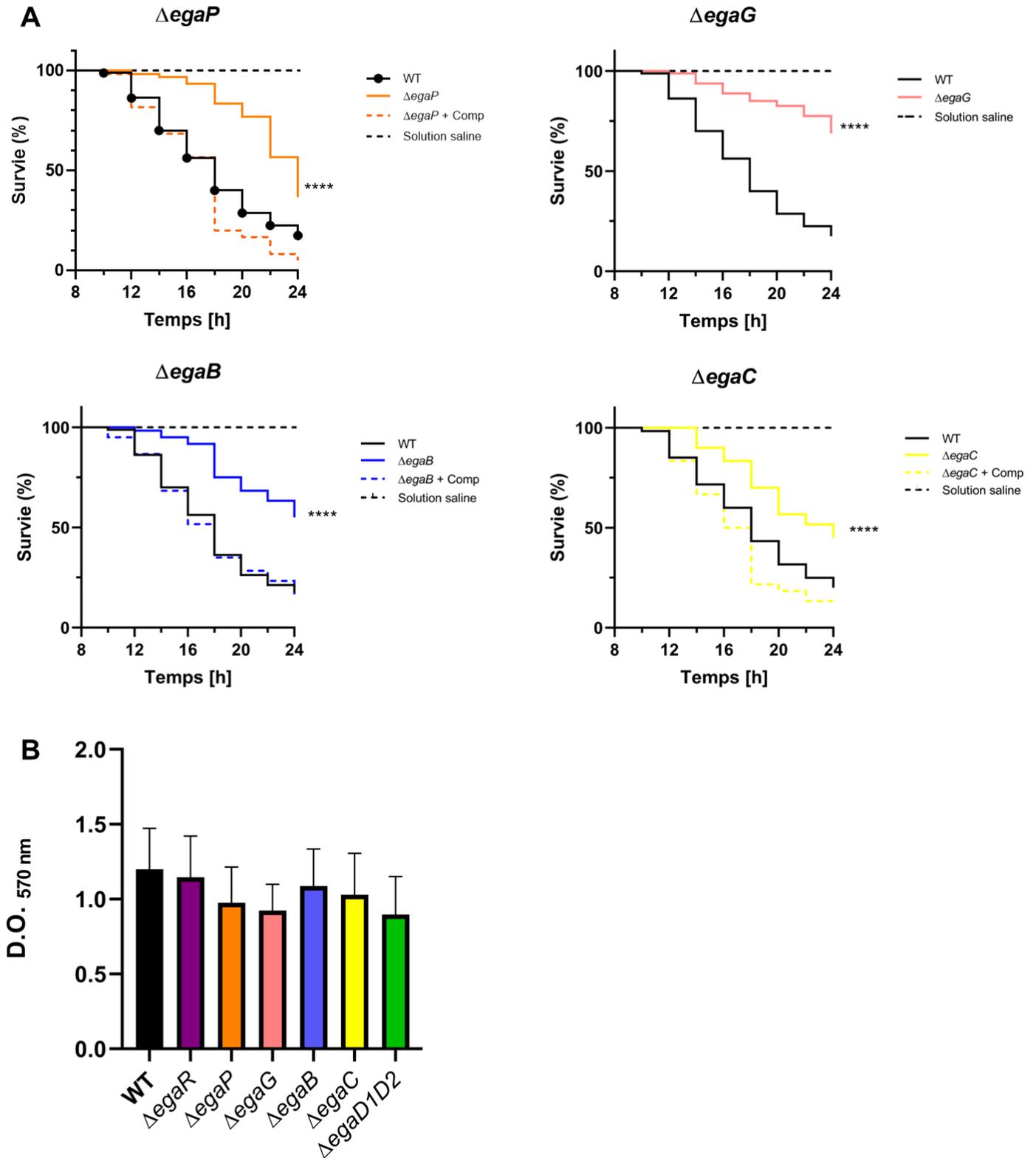
B

Protein.Group	Protein.Ids	Protein.Names	Genes	Δ egaR - Protéome	Δ egaR - Sécrétome	WT - Protéome	WT - Sécrétome
Q836F6	Q836F6	Q836F6_ENTFA	EgaP	22288,6		51422,2	
H7C6V7	H7C6V7	H7C6V7_ENTFA	EF_2476	9404,05		8260,27	
H7C6X1	H7C6X1	H7C6X1_ENTFA	EF_1565	2543,19		3094,52	
H7C6X2	H7C6X2	H7C6X2_ENTFA	pheA	6135,2		6517,18	
H7C6X6	H7C6X6	H7C6X6_ENTFA	EF_0394	51963,7	9,55E+06	58523,8	1,93E+07
H7C6X7	H7C6X7	H7C6X7_ENTFA	EF_2195	2568,67		3863,56	
H7C6X9	H7C6X9	H7C6X9_ENTFA	EF_1340	38184,4	210666	30862,3	

Figure 49 : Analyse de la structure primaire de EgaP et analyse par spectrométrie de masse. **A** : Les domaines d'intérêts sont représentés par des rectangles. Le domaine peptidase M20 est indiqué en orange. Le domaine de dimérisation est indiqué en violet. **B** : Analyse du protéome et sécrétome de la souche OG1RF WT et OG1RF Δ egaR par spectrométrie de masse.

2.4 Le locus *ega* est impliqué dans la virulence

L'implication du locus *ega* dans la virulence a été testée en réalisant des expériences de survie en modèle animal et de formation de biofilm. Les mutants de délétion des gènes *egaG*, *egaP*, *egaB* et *egaC* ont été testés dans le modèle animal *G. mellonella* (**Figure 50A**). Ce modèle, en plus d'être peu coûteux et sans contrainte éthique, possède un système immunitaire inné similaire à celui des mammifères [63,221]. Ces caractéristiques en font un modèle de choix pour l'étude de la virulence bactérienne. Par rapport à la souche de type sauvage, pour laquelle le taux de survie à 24 h après l'injection est de 25 %, plus de 50 % des insectes infectés par le mutant Δ egaG survivent après la même période de temps. Des résultats similaires ont été obtenus avec les mutants Δ egaP, Δ egaB et Δ egaC. Cela montre que l'utilisation de substrats GlcNAc-L-Asn-like dérivés de glycoprotéines est importante pour la virulence d'*E. faecalis*. Enfin nous avons observé que les mutations dans le locus *ega* n'ont pas d'impact sur la formation de biofilm (**Figure 50B**).



3 Mise en place de l'outil CRISPR-Cas9 chez *E. faecalis*

Afin d'étudier le locus *ega*, de nombreux mutants devaient être construits. Pour cela, un nouvel outil de mutagenèse basé sur le système CRISPR-Cas9 a été développé chez *E. faecalis*. Pour mettre en place l'outil, nous nous sommes basés sur l'article de Maat et collaborateurs [201]. Dans un premier temps, afin de déléter le gène *egaC*, nous avons déterminé la séquence spacer qui permettra de guider le complexe CRISPR-Cas9 sur son gène cible: 5' - TGAATTTCTACAAACGCCCTTATTGAGTCT - 3'. Ce spacer a été cloné entre les régions crRNA du plasmide pVDM1001 (pVDM1001-*egaC*). Un fragment ADN comportant le gène *egaC* tronqué a été cloné dans le plasmide pVDM1001-*egaC* pour permettre la double recombinaison homologue suite à la coupure de l'ADN double brin. Ces modifications ont conduit à l'élaboration du plasmide pCRISPR-*egaC* (Figure 51).

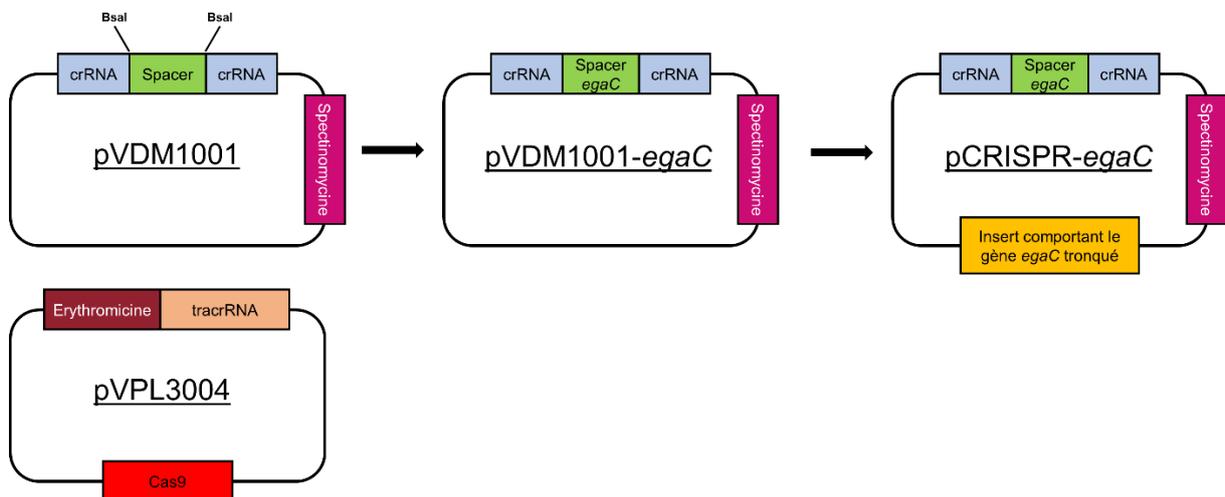


Figure 51 : Représentation schématique des étapes réalisées pour obtenir les plasmides comportant l'outil CRISPR-Cas9 dirigé contre le gène *egaC*. Le plasmide pVDM1001-*egaC* a été développé en insérant le spacer (vert) dirigé contre *egaC* entre les deux crRNA. Le plasmide pCRISPR-*egaC* a été développé en insérant le fragment d'ADN permettant la double recombinaison homologue (jaune) après coupure du chromosome.

Les plasmides pVPL3004 et pCRISPR-*egaC* ont été introduits dans la souche OG1RF. Une cinquantaine de colonies ayant poussé sur milieu GM17 contenant de l'érythromycine et spectinomycine ont été testées. Seulement 4% de ces colonies présentaient la mutation attendue dans leur chromosome. Cette faible efficacité nous a conduit à nous demander si la pression de sélection exercée par la spectinomycine était suffisante pour maintenir le plasmide pCRISPR-*egaC* dans la cellule, ou si l'outil CRISPR-Cas9 ne fonctionnait pas chez *E. faecalis*. Pour tester la 1^{ère} hypothèse, des PCR ont été réalisées pour montrer la présence

ou l'absence du plasmide pCRISPR-*egaC* dans les clones présentant la délétion d'intérêt. Aucune des 3 souches Δ *egaC* obtenues ne possèdent le plasmide pCRISPR-*egaC* (**Figure 52**). Au vu de ce résultat, nous avons réitéré l'expérience de cassure du double brin d'ADN en ensemençant après transformation les bactéries sur des milieux comportant différentes concentrations de spectinomycine variant entre 200 μ g/mL et 900 μ g/mL (**Figures 53A**).

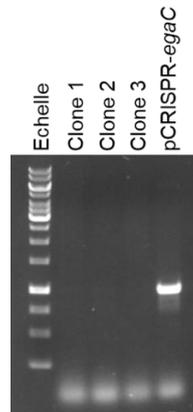


Figure 52 : *Présence du plasmide pCRISPR-egaC chez les souches mutées après transformation. Electrophorèse en gel d'agarose à 1 % pour vérifier, après amplification spécifique, la présence du plasmide pCRISPR-egaC dans les souches présentant la délétion chromosomique d'intérêt.*

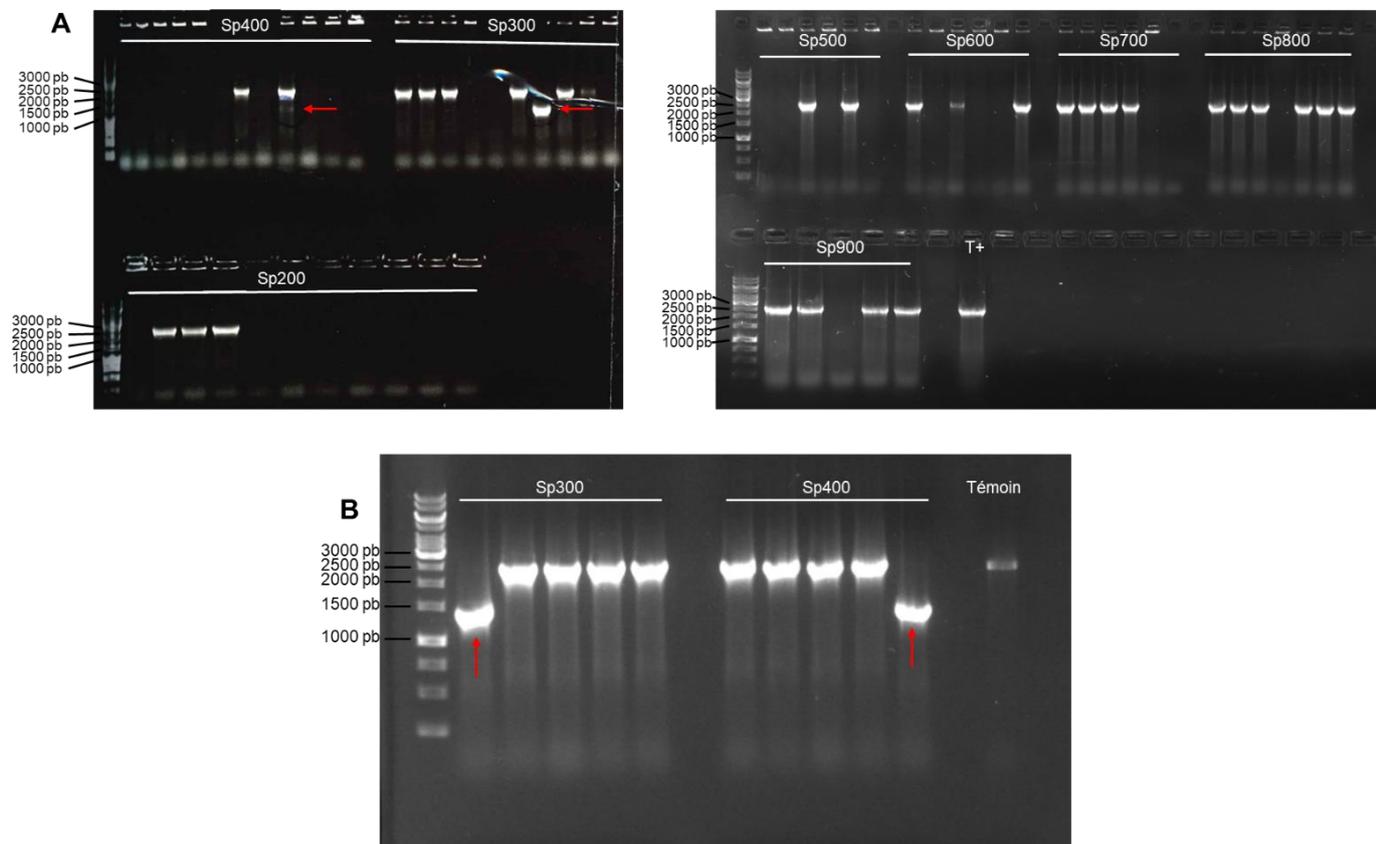




Figure 53 : Vérification de l'outil CRISPR-Cas9. **A-B** : Electrophorèse en gel d'agarose 1 % après PCR pour vérifier la présence de la délétion chromosomique d'intérêt sur des souches ensemencées à différentes concentrations de spectinomycine. L'abréviation *Sp* signifie spectinomycine, le nombre indique la concentration en $\mu\text{g}/\text{mL}$. Le témoin positif correspond au produit d'amplification obtenu à partir d'ADN génomique de la souche OG1RF WT. Les flèches rouges indiquent la présence d'une colonie délétée pour le gène *egaC*. **A** : Dans chaque puits a été déposée une PCR effectuée sur un mélange de 5 clones. **B** : Dans chaque puit a été déposée une PCR effectuée sur 1 clone isolé. **C** : Localisation du plasmide pCRISPR-*egaC* chez différents clones testés.

Des colonies présentant la délétion d'intérêt sont obtenues uniquement pour des concentrations de spectinomycine comprises entre 300 et 400 $\mu\text{g}/\text{mL}$: une très faible efficacité de mutagenèse est cependant obtenue, où seulement 2% (1/50) et 1.6% (1/60) des colonies testées présentent la mutation d'intérêt respectivement (**Figure 53B**). En dehors de ces concentrations en antibiotique, aucun clone ayant la mutation n'a été obtenu. Il est toutefois notable que pour chacune de ces concentrations en antibiotique, des colonies de type sauvage (≈ 2500 pb) ont pu se développer, suggérant que le plasmide pCRISPR-*egaC* s'était inséré dans le chromosome (**Figure 53C**). Des colonies sauvages ont été testées, des bandes parasites sont obtenues sur le gel d'électrophorèse. Il semble donc que le plasmide pCRISPR-*egaC* ne soit pas intégré dans le chromosome. Le plasmide n'a pas non plus été détecté dans le cytoplasme. L'absence du plasmide pCRISPR-*egaC* dans la cellule indique donc que la cassette de résistance n'est pas efficace pour sélectionner et maintenir le plasmide. Nous avons donc changé de stratégie. La cassette de résistance spectinomycine du plasmide pVDM1001 a été changée par une cassette de résistance au chloramphénicol, ce qui a permis d'obtenir le plasmide pCRISPR(Ch) (**Figure 54A**). Le changement de cassette de résistance semble permettre une meilleure stabilité du plasmide dans la cellule comportant déjà le plasmide pVPL3004 (**Figure 54B**). Par la suite, le plasmide pCRISPR(Ch)-*egaC* a été construit (**Figure 54A**) et transformé dans la souche OG1RF WT comportant le plasmide pVPL3004. Malgré les différentes tentatives de transformation et de vérification par colonie PCR, nous n'avons pas pu obtenir de mutant présentant la délétion d'intérêt (**Figure 54C**).

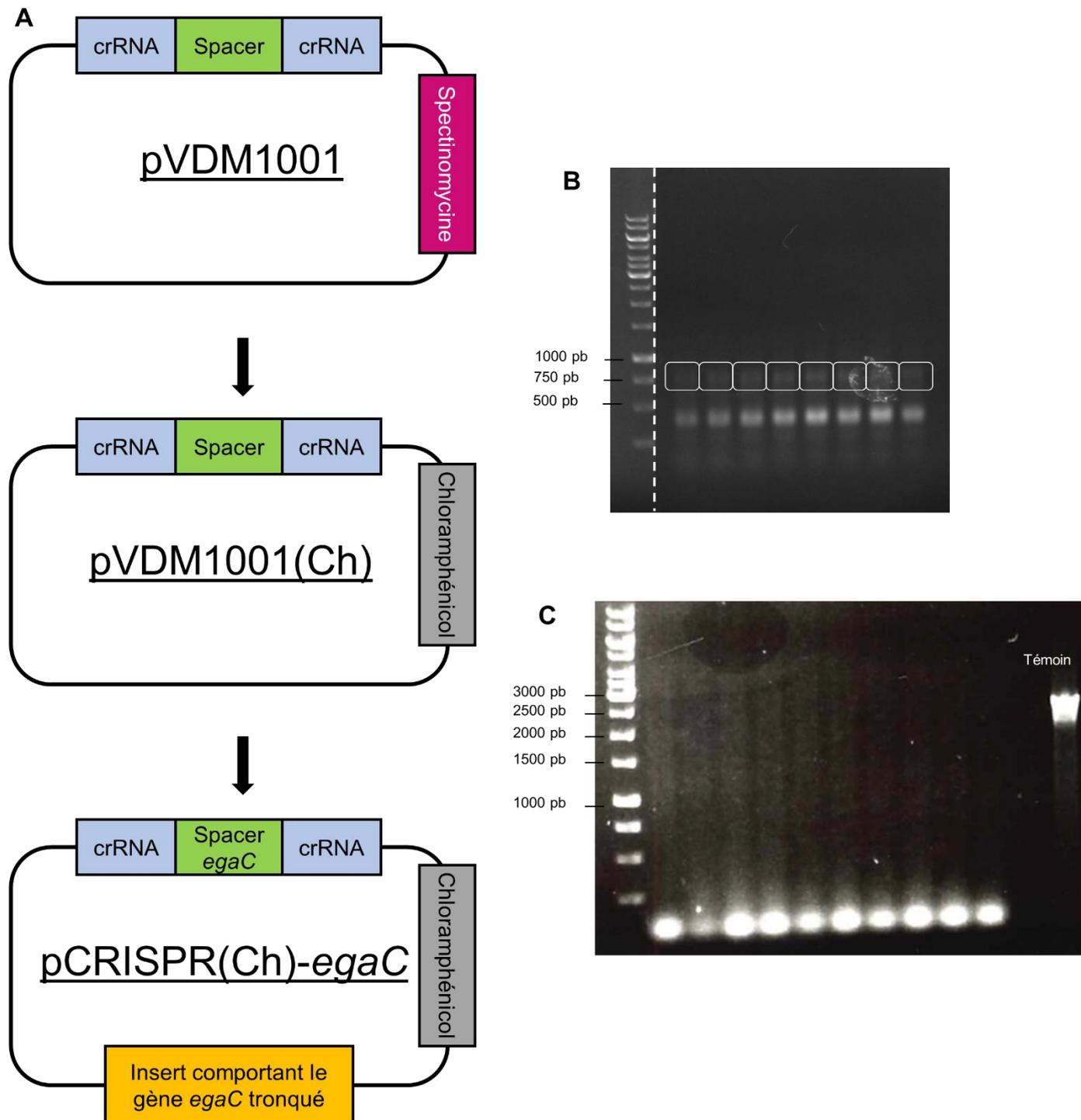


Figure 54 : Efficacité de l'outil CRISPR-Cas9 après modification de la cassette de résistance. **A** : Représentation schématique du plasmide pCRISPR(Ch)-*egaC* obtenu après le changement de cassette de résistance à la spectinomycine contre la résistance au chloramphénicol. **B** : Electrophorèse en gel d'agarose 1 % après PCR individuelle ayant pour objectif de vérifier la présence du plasmide pCRISPR(Ch)-*egaC* après transformation dans la souche OG1RF WT + pVPL3004. **C** : Efficacité de l'outil CRISPR-Cas9 après modification de la cassette de résistance.

Discussion et perspectives

1 Caractérisation du locus glycosylasparaginase chez *E. faecalis* OG1RF

Dans ce travail de thèse, nous avons étudié différents gènes de *E. faecalis* supposés être impliqués dans l'utilisation de produits de déglycosylation suite à l'action d'une glycosylasparaginase putative. Les résultats obtenus nous ont permis d'affiner en partie le modèle proposé initialement en **Figure 28** (page 64). Ces travaux ont mis en avant la capacité d'*E. faecalis* à utiliser le GlcNAc-L-Asn comme source de carbone. Ce composé est retrouvé dans la structure des protéines N-glycosylées de l'hôte. Or, il est estimé que la moitié des protéines produites chez les eucaryotes sont glycosylées [130]. Au niveau du TGI où l'on retrouve les entérocoques, diverses protéines N-glycosylées sont présentes telles que les e-cadhérines, les immunoglobulines, la protéine BCF-1, les récepteurs humains de la bradykinin B2, le collagène ou encore la lactoferrine [222–226]. L'utilisation, par *E. faecalis*, du GlcNAc-L-Asn des protéines N-glycosylées de l'hôte nécessite par conséquent une première étape de dégradation, séparant ainsi ce composé du reste de la glycoprotéine. Cette première étape était supposée être réalisée par EgaP mais les analyses et les résultats obtenus nous indiquent qu'EgaP n'est pas sécrétée et ne peut donc pas venir dégrader les protéines glycosylées. Néanmoins, cela peut être réalisé par une ou plusieurs autres protéases. Plusieurs protéases sont retrouvées chez *E. faecalis*, dont les plus étudiées sont GelE et SprE. Une étude publiée par Mäkinen et collaborateurs a montré que GelE est capable de dégrader le collagène [227]. En plus de ces coupures par les protéases, des coupures par les ENGases comme la protéine EndoE d'*E. faecalis* doivent avoir lieu. Cette enzyme est capable de cliver la liaison entre deux résidus GlcNAc des protéines N-glycosylées [157,158]. Cependant, l'ordre d'action des différentes enzymes peut avoir son importance. En effet, il se pourrait que l'action des peptidases nécessite un découplage préalable de la glycosylation par des ENGases pour dégrader la protéine. Ces travaux n'ont cependant pas permis d'identifier la/les protéase(s) impliquée(s). Un pré-traitement de la RNaseB avec le surnageant d'une culture d'*E. faecalis* est envisagé avant d'utiliser la RNaseB comme substrat pour la peptidase et la glycosylasparaginase. Si ce pré-traitement permet l'utilisation de la RNaseB comme substrat pour EgaG, une analyse du sécrétome d'*E. faecalis* sera réalisée pour déterminer les potentielles peptidases impliquées.

Les résultats obtenus concernant la peptidase n'ont pas permis la libération dans le milieu de substrat utilisable par la glycosylasparaginase. Le fait que la chaîne N-glycane de la

RNase B ne soit pas disponible à cause de la conformation tridimensionnelle est peu probable dû au traitement thermique réalisée. Nous pourrions éventuellement utiliser la RNase B sans prétraitement comme substrat potentiel pour la glycosylasparaginase. Contrairement à ce qui était envisagé, la peptidase ne serait pas sécrétée dans le milieu extracellulaire mais aurait plutôt une action intracellulaire. Dans le contexte de l'étude, il est supposé que cette peptidase permette la libération de substrat utilisable par la glycosylasparaginase, donc une action sur des protéines N-glycosylées. À ce jour, il n'a pas été mentionné de protéines glycosylées chez *E. faecalis* dans la littérature. Néanmoins, l'analyse Blast du génome d'*E. faecalis* OG1RF révèle que les protéines OG1RF_11721 et OG1RF_11725 présentent des homologies de séquences avec les protéines PglC (45,41 %) et PglB (34,21 %) retrouvées chez *C. jejuni* impliquées dans la glycosylation des protéines. La présence de ces gènes laisse supposer une possible glycosylation des protéines chez *E. faecalis*. L'opéron *ega* servirait dans ce contexte à la régulation du pool de protéines glycosylées chez *E. faecalis*. Ces hypothétiques protéines glycosylées ne contiendraient pas de core GlcNAc, donc ne permettraient pas la libération de substrat utilisable pour la glycosylasparaginase. L'identification de ces hypothétiques protéines glycosylées chez *E. faecalis* reste toutefois une piste intéressante d'étude car, d'après nos connaissances, il n'a jamais été montré un tel type de protéine chez ce microorganisme. Pour montrer un tel type de protéine chez *E. faecalis*, il est envisageable de déléter les gènes *pglC* et *pglB* de cette dernière. Le protéome de cette souche sera alors comparé au protéome de la souche sauvage par spectrométrie de masse. Si ces enzymes sont impliquées dans la glycosylation et qu'elles sont fonctionnelles, alors nous verrons des différences de poids moléculaires dans les deux protéomes pour une même protéine.

Une autre possibilité de la fonction de cette peptidase serait une action non pas sur des protéines N-glycosylées entières, mais sur des éléments plus petits tels que des dipeptides-glycosylés. Il a été montré chez *L. lactis* que des protéinases de surface permettaient la libération de peptides, dont des di-peptides [228]. Ces fragments sont ensuite internalisés par le transporteur DtpT dans la cellule et ensuite dégradés, entre autre, par la dipeptidase PepV conduisant à la libération d'AA [228–230]. Cette libération est nécessaire car les bactéries lactiques sont poly-auxotrophes pour certains AA comme le glutamate, la sérine, la méthionine et les AA à chaîne ramifiée [231]. Un fonctionnement similaire chez *E. faecalis* peut être supposé : c'est également une bactérie auxotrophe pour les AA : arginine, histidine,

méthionine, tryptophane, isoleucine, leucine, valine et les acides aminés à chaîne ramifiée [232]. Les analyses bio-informatiques ont montré qu'EgaP présente une homologie de séquence avec la protéine PepV de *Lactobacillus delbrueckii subsp. sunkii* (32,4 % d'identité sur 99 % de couverture). La caractérisation de cette protéine chez diverses bactéries lactiques (*L. delbrueckii subsp. lactis DSM7290*, *L. lactis subsp. lactis*, *L. lactis subsp. cremoris* et *L. lactis subsp. hordniae*) a montré que cette enzyme avait la capacité de cliver les liaisons d'une large variété de di-peptides et certains tripeptides libérant ainsi des AA [229,233,234]. De plus, *E. faecalis* possède le gène *ntpT*, dont la protéine NtpT a été montrée comme étant impliquée dans l'internalisation de dipeptide chez *L. lactis* [230]. Ces éléments nous conduisent à penser que, suite à l'action des protéases sécrétées par *E. faecalis*, des di-peptides contenant un résidu GlcNAc-L-Asn peuvent être libérés, internalisés dans la cellule *via* le transporteur NtpT, puis clivés par EgaP libérant ainsi un substrat utilisable par la glycosylasparaginase. Cela constituerait une voie parallèle d'utilisation de molécules possédant un résidu GlcNAc, dont le substrat d'intérêt de cette étude GlcNAc-L-Asn. Il est également possible que ces potentiels di-peptides puissent être internalisés *via* le PTS EgaC, EgaB, et OG1RF_10751. Les données de structure suggèrent que les PTS de type lactose/cellobiose sont capables de transporter des hydrates de carbone avec des aglycones relativement importants [235].

Suite à la libération de GlcNAc-L-Asn, ce composé est internalisé et phosphorylé par les domaines PTS EgaB et EgaC et OG1RF_10751. Cependant nos résultats ont montré que le composé GlcNAc-L-Asn non phosphorylé n'est pas seulement hydrolysable par la glycosylasparaginase EgaG, mais qu'il joue également le rôle d'inducteur en se liant au répresseur EgaR. Les résultats de retard sur gel indiquent une concentration en substrat GlcNAc-L-Asn supérieure à 7,5 mM pour interagir avec le répresseur. Dans la littérature, des concentrations de substrat nécessaire pour qu'une interaction avec un régulateur HutC sont comprises entre 0,1 μ M et 500 μ M. L'exemple de NagR dans la littérature indique qu'il faut cependant une concentration en GlcNAc phosphorylé de l'ordre du mM pour interagir avec le répresseur [71,236–238]. Dans le cadre de l'utilisation du maltose chez *E. faecalis*, il a également été démontré que la forme non phosphorylée avait une meilleure affinité avec le régulateur MalR que la forme phosphorylée [94]. De plus, une fois le maltose internalisé et phosphorylé dans la cellule, l'enzyme MapP catalyse le clivage du groupement phosphate du

maltose-6P [239]. Ces données nous interpellent quant au rôle et à la nécessité de l'état de phosphorylation du GlcNAc-L-Asn. Pour déterminer si l'état de phosphorylation joue un rôle vis-à-vis de l'affinité de ce substrat à EgaR et EgaG, des expériences avec ce substrat devront être réalisées. Nous avons précédemment abordé le fait que le GlcNAc-L-Asn sous sa forme phosphorylée n'était pas disponible dans le commerce : durant ces travaux, la synthèse d'un tel composé a été entreprise. En effet, Thompson et collaborateurs ont identifié une β -glucoside kinase chez *Klebsiella pneumoniae*, BglK, pouvant phosphoryler une large gamme de composés de type β -glucoside [240]. Une collaboration avec cette équipe a permis l'obtention du composé GlcNAc-L-Asn phosphorylé. Cependant, après phosphorylation, la forme non phosphorylée du composé GlcNAc-L-Asn est encore détectée. La proportion ainsi que la concentration de ces deux formes n'ont pas pu être déterminées. De plus, certaines impuretés (matériels ayant permis la phosphorylation) persistaient après purification. Une fois que la mise au point de ces étapes permettra d'obtenir seulement la forme phosphorylée, il pourra être envisagé de confirmer si l'état de phosphorylation du GlcNAc influe sur l'affinité avec le régulateur EgaR et la glycosylasparaginase EgaG.

L'activité glycosylasparaginase chez *E. faecalis* a pu être mise en évidence : elle est capable d'utiliser le GlcNAc-Asn comme substrat avec une activité spécifique de $19,35 \mu\text{mol.mg protéine}^{-1}.\text{min}^{-1}$. Des valeurs très similaires sont décrites pour les glycosylasparaginases de *Lc. casei* et *E. meningoseptica* avec une activité spécifique pour le GlcNAc-Asn, respectivement de $20,91 \mu\text{mol.mg protéine}^{-1}.\text{min}^{-1}$ et $25 \mu\text{mol.mg protéine}^{-1}.\text{min}^{-1}$ [176,212]. Suite au changement de fournisseur de l'enzyme Glutamate-Oxaloacétate Transaminase pour le test de la glycosylasparagine, dû à un arrêt de production par le fournisseur initial, nous n'avons pas pu caractériser tous les paramètres cinétiques et biochimiques de la glycosylasparaginase (V_{max} , K_m , température optimale, pH, Inhibiteur DONV). Néanmoins, nous avons récemment réussi à surexprimer et purifier l'enzyme Glutamate-Oxaloacétate Transaminase : cela devrait nous permettre d'analyser les paramètres cités précédemment. Enfin, la glycosylasparaginase d'*E. faecalis* subit un processus d'autoprotéolyse pour acquérir sa forme active. Les AA His150, Arg151 et Thr152 sont conservés, et semblent être impliqués dans le processus d'autoprotéolyse d'après les résultats de spectrométrie de masse. Des expériences de mutagenèse dirigée sont envisagées pour affirmer cette hypothèse.

Finalement, nous supposons que le 1-amino-6-P-GlcNAc était hydrolysé par les réductases EgaD1D2, permettant ainsi la libération de 6-P-GlcNAc pouvant être utilisé dans le métabolisme bactérien. Les différentes analyses *in silico* ont mis en doute cette hypothèse. Les réductases EgaD1D2 présentent des homologues de séquences avec GrdB, protéine d'un complexe permettant de métaboliser la glycine en acétyl-phosphate et en ammoniacque [213]. Cependant, les protéines GrdA et GrdC impliquées dans le complexe glycine/sarcosine/bétaïne réductase ne sont pas retrouvées chez *E. faecalis*, ce qui sous-entend que les enzymes EgaD1D2 ne sont pas impliquées dans cette voie métabolique. Les gènes *egaD1* et *egaD2* sont issus d'une mutation du gène *egaD*, forme tronquée de ce gène que l'on ne retrouve chez aucun autre microorganisme. Cette mutation semble impacter la traduction de la protéine car les analyses de spectrométrie de masse n'ont pas permis d'identifier EgaD1 et EgaD2 dans le sécrétome ou le protéome cytoplasmique d'*E. faecalis* OG1RF. L'analyse des génomes de *Lc. casei* BL23, *E. meningoseptica* ou encore *V. lutrae* a montré que ces micro-organismes ne possèdent pas d'équivalents des gènes *egaD* ou *egaD1D2* alors qu'elles possèdent une glycosylasparaginase. Nous pouvons donc supposer que, chez ces organismes, l'hydrolyse du 1-amino-6-P-GlcNAc en ammoniacque et 6-P-GlcNAc ne soit pas due à une action enzymatique. Cette hypothèse est supportée par l'étude de Makino et collaborateurs qui ont montré qu'en milieu aqueux, le 1-amino-GlcNAc est instable et s'hydrolyse spontanément en ammoniacque et GlcNAc [241]. L'ensemble des résultats obtenus et des points abordés dans la discussion nous ont permis d'affiner le modèle initial (Figure 55).

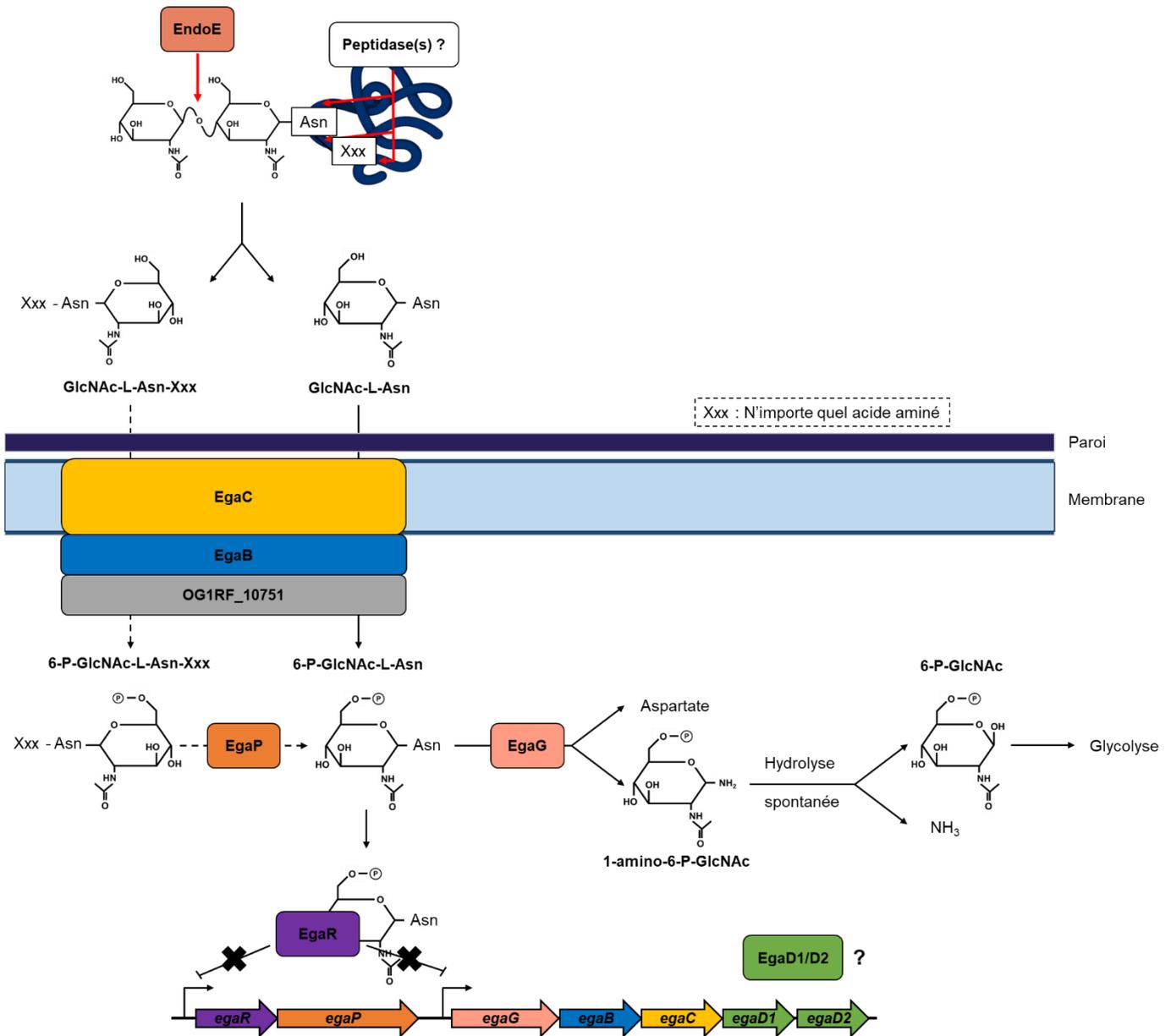


Figure 55 : *Modèle proposé de la voie catabolique du GlcNAc-L-Asn chez E. faecalis.* Les étapes indiquées par les flèches pleines sont étayées par des données expérimentales. Les étapes indiquées par des flèches en pointillées indiquent des hypothèses. Les flèches rouges indiquent le site de coupure de l'enzyme. Le nom des molécules est indiqué en gras.

Une enzyme jusque-là non identifiée permettrait le clivage d'une protéine N-glycosylée de part et d'autre de l'asparagine. En parallèle, une ENGase comme EndoE pourrait cliver la chaîne glycane comme décrit précédemment. Les résidus GlcNAc-L-Asn ou un dipeptide N-glycosylé GlcNAc-L-Asn-Xxx est transporté et phosphorylé par les protéines PTS EgaB, EgaC et OG1RF_10751. Dans le cas où l'Asn serait liée à un AA supplémentaire, le substrat intracellulaire devrait être transformé en 6-P-GlcNAc-L-Asn et Xxx, probablement par la dipeptidase EgaP. Le Xxx libéré entre dans le métabolisme des AA. Le 6-P-GlcNAc-L-Asn agit

probablement comme inducteur des deux unités de transcription Ega en se liant au répresseur EgaR et en l'empêchant d'interagir avec ses sites opérateurs de type HutC. Cependant, des expériences de retard sur gel ont montré que le GlcNAc-L-Asn non phosphorylé interagit également avec EgaR. Le 6-P-GlcNAc-L-Asn est ensuite hydrolysé en 1-amino-6-P-GlcNAc et en aspartate par la glycosylasparaginase EgaG. L'aspartate entre également dans le métabolisme des AA. Enfin, le 1-amino-6-P-GlcNAc est spontanément hydrolysé en 6-P-GlcNAc et en ammoniacque.

La capacité à utiliser les produits de dégradation des protéines N-glycosylées constitue un avantage évident au cours de l'infection par les entérocoques, comme le démontrent les résultats obtenus avec le modèle animal *G. mellonella*. En absence du locus *ega*, donc sans certains produits que l'on pense issus de la déglycosylation, la bactérie présente une virulence réduite. A notre connaissance, c'est la première fois qu'il est démontré que l'action d'une glycosylasparaginase en condition infectieuse augmente la virulence bactérienne. Néanmoins, le fait que la suppression du gène *egaP* réduise également la virulence de l'entérocoque est surprenante : la suppression du gène *egaP* n'a eu aucun effet sur la croissance en présence de GlcNAc-L-Asn, mais a atténué la virulence des entérocoques. Une explication possible de ce résultat contradictoire pourrait être que la dégradation des protéines N-glycosylées de *G. mellonella* produit principalement du GlcNAc-L-Asn-Xxx, dont le catabolisme nécessite une EgaP fonctionnelle, alors que pendant la croissance sur GlcNAc-L-Asn, EgaP n'est pas nécessaire. Cependant, une limite à ce modèle est l'absence de système immunitaire adaptatif chez les larves [242]. Il serait intéressant de réaliser des expériences de survie en modèle animal souris pour voir si des résultats similaires sont obtenus en présence d'un système immunitaire adaptatif, ce qui se rapproche de ce qui est retrouvé chez l'Homme.

Les différents points soulevés et hypothèses avancées dans cette partie ouvrent des perspectives de recherches qui permettront une meilleure compréhension du fonctionnement et du rôle du locus glycosylasparaginase chez *E. faecalis* en particulier lors de son processus infectieux.

2 L'outil de mutagenèse CRISPR-Cas9 chez *E. faecalis*.

Depuis sa découverte, CRISPR-Cas a été détourné pour en faire un outil de biologie moléculaire utilisé dans de nombreuses applications : édition du génome, contrôle de la

transcription, lutte contre l'antibiorésistance, agriculture ou encore thérapie génique [187,198,243–247]. Un aspect annexe des travaux de thèse présenté dans ce manuscrit a été la mise en place de l'outil CRISPR-Cas9 en tant que nouvel outil de mutagenèse chez *E. faecalis*. Pour cela, nous nous sommes basés sur l'outil précédemment développé chez *E. faecium* par de Maat et collaborateurs en 2019 [201]. Bien que cet outil ait permis l'obtention de la souche délétée Δ *egaC*, son efficacité n'est pas optimale. En effet, un très faible pourcentage de souches testées présentait la délétion d'intérêt. Nous avons pu mettre en évidence que la cassette de résistance à la spectinomycine portée par le plasmide pCRISPR-*egaC* ne suffisait probablement pas à maintenir ce plasmide dans la cellule. Suite au changement de la cassette de résistance spectinomycine par la cassette de résistance chloramphénicol, nous avons pu attester de la présence simultanée des plasmides pCRISPR(Ch)-*egaC* et pVPL3004 dans la cellule. Cependant, malgré la présence simultanée de ces plasmides, aucun mutant de délétion n'a été obtenu.

Bien que nous n'ayons pas réussi à identifier la raison pour laquelle cet outil ne fonctionne pas pleinement chez *E. faecalis*, plusieurs hypothèses ont été avancées. Il a été décrit dans la littérature des protéines anti-CRISPR (Acr) [248,249]. Ces protéines peuvent inhiber l'interaction du complexe CRISPR-Cas à l'ADN, inhiber la formation du complexe CRISPR-Cas, ou encore le clivage de l'ADN cible [248,249]. Il a été montré que le transfert d'un plasmide ciblé par l'outil CRISPR-Cas9 dans certaines souches d'*E. faecalis* est renforcé en présence des protéines anti-CRISPR [249]. Cependant, les gènes codant les protéines anti-CRISPR ne sont pas portés par le génome d'*E. faecalis*, mais par des plasmides. La souche OG1RF utilisée lors de cette étude est curée de tout plasmide, invalidant donc cette hypothèse. Une autre hypothèse expliquant l'échec de la mise en place de l'outil CRISPR-Cas9 chez *E. faecalis* concerne l'origine de répllication des plasmides. Il est évident que les plasmides pVPL3004 et pVDM1001 possèdent des origines de répllications différentes, sans quoi l'outil n'aurait pu fonctionner chez *E. faecium* [201]. Cependant, pour se répliquer, les plasmides utilisent généralement les protéines de l'hôte. Il est envisageable que les réplicons de ces plasmides ne soient pas aussi bien reconnus chez *E. faecalis* que chez *E. faecium* ou *L. lactis*, entraînant ainsi le non fonctionnement du plasmide et sa perte. Il est également connu que l'outil CRISPR-Cas9 a un effet « off-target » qui correspond au clivage du double brin d'ADN en dehors de la zone définie par le spacer [250]. Cette coupure non désirée du chromosome

entraînerait la mort de la cellule bactérienne. Enfin, le séquençage du génome d'*E. faecalis* a permis de mettre en évidence la présence des gènes CRISPR-Cas, sans pour autant que le complexe CRISPR-Cas en découlant ne soit fonctionnel. Ce système peut être réactivé lorsque l'on fournit à la souche les gènes *cas*, dont *cas9*. Bien que non décrit dans la littérature, nous pouvons avancer l'hypothèse que la présence de deux systèmes CRISPR dans la cellule peut altérer son fonctionnement.

Une nouvelle approche a été envisagée pour mettre en place un outil de mutagenèse CRISPR-Cas9 chez *E. faecalis*. En 2015, Rodrigues et collaborateurs ont développé un outil CRISPR-Cas9 porté par le plasmide conjugatif (pKH88) qui permet de cibler efficacement des souches d'*E. faecalis* portant le gène de résistance à l'érythromycine [251]. Contrairement à l'outil développé par de Maat et collaborateurs, l'outil développé par Rodrigues et collaborateurs ne permet pas la réparation du chromosome après cassure par l'endonucléase Cas9, et donc n'est pas destiné à de la mutagenèse. Nous envisageons de modifier ce plasmide de sorte à insérer un spacer contre un gène d'intérêt ainsi qu'un fragment ADN comportant la version du gène d'intérêt tronqué pour permettre la double recombinaison homologue nécessaire pour en faire un outil de mutagenèse. Ce nouvel outil ne serait porté que par un seul plasmide, évitant donc une double pression de sélection dans le milieu, ce qui réduirait l'environnement stressant des bactéries et favoriserait ainsi leur développement.

Références bibliographiques

- [1] Cattoir V. The multifaceted lifestyle of enterococci: genetic diversity, ecology and risks for public health. *Curr Opin Microbiol* 2021;65:73–80. <https://doi.org/10.1016/j.mib.2021.10.013>.
- [2] Arias CA, Murray BE. The rise of the *Enterococcus*: beyond vancomycin resistance. *Nat Rev Microbiol* 2012;10:266–78. <https://doi.org/10.1038/nrmicro2761>.
- [3] García-Solache M, Rice LB. The *Enterococcus*: a Model of Adaptability to Its Environment. *Clin Microbiol Rev* 2019;32. <https://doi.org/10.1128/CMR.00058-18>.
- [4] Muller C, Cacaci M, Sauvageot N, Sanguinetti M, Rattei T, Eder T, *et al.* The Intraperitoneal Transcriptome of the Opportunistic Pathogen *Enterococcus faecalis* in Mice. *PLoS ONE* 2015;10:e0126143. <https://doi.org/10.1371/journal.pone.0126143>.
- [5] Grand M, Aubourg M, Pikis A, Thompson J, Deutscher J, Hartke A, *et al.* Characterization of the gen locus involved in β -1,6-oligosaccharide utilization by *Enterococcus faecalis*. *Mol Microbiol* 2019;112:1744–56. <https://doi.org/10.1111/mmi.14390>.
- [6] Aebi M. N-linked protein glycosylation in the ER. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 2013;1833:2430–7. <https://doi.org/10.1016/j.bbamcr.2013.04.001>.
- [7] Hounsell EF, Davies MJ, Renouf DV. O-linked protein glycosylation structure and function. *Glycoconjugate J* 1996;13:19–26. <https://doi.org/10.1007/BF01049675>.
- [8] Maeda M, Kimura Y. Structural features of free N-glycans occurring in plants and functional features of de-N-glycosylation enzymes, ENGase, and PNGase: the presence of unusual plant complex type N-glycans. *Front Plant Sci* 2014;5:429. <https://doi.org/10.3389/fpls.2014.00429>.
- [9] Bozon D, Tarentino AL, Trimble RB, Maley F. Characterization of cellular oligosaccharides from normal and cystic fibrotic fibroblasts using sequential endoglycosidase digestions. *Archives of Biochemistry and Biophysics* 1986;249:546–56. [https://doi.org/10.1016/0003-9861\(86\)90032-9](https://doi.org/10.1016/0003-9861(86)90032-9).
- [10] Tarentino AL, Plummer TH. The first demonstration of a procaryotic glycosylasparaginase. *Biochem Biophys Res Commun* 1993;197:179–86. <https://doi.org/10.1006/bbrc.1993.2457>.
- [11] Devriese LA, Pot B. The genus *Enterococcus*. In: Wood BJB, Holzapfel WH, editors. *The Genera of Lactic Acid Bacteria*, Boston, MA: Springer US; 1995, p. 327–67. https://doi.org/10.1007/978-1-4615-5817-0_10.
- [12] Andrewes FW, Horder TJ. A STUDY OF THE STREPTOCOCCI PATHOGENIC FOR MAN. *The Lancet* 1906;168:852–5. [https://doi.org/10.1016/S0140-6736\(01\)43302-2](https://doi.org/10.1016/S0140-6736(01)43302-2).
- [13] Foulquié Moreno MR, Sarantinopoulos P, Tsakalidou E, De Vuyst L. The role and application of enterococci in food and health. *International Journal of Food Microbiology* 2006;106:1–24. <https://doi.org/10.1016/j.ijfoodmicro.2005.06.026>.
- [14] Devriese L, Baele M, Butaye P. The Genus *Enterococcus*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. *The Prokaryotes: Volume 4: Bacteria: Firmicutes, Cyanobacteria*, New York, NY: Springer US; 2006, p. 163–74. https://doi.org/10.1007/0-387-30744-3_5.
- [15] Schleifer KH, Kilpper-Bälz R 1984. Transfer of *Streptococcus faecalis* and *Streptococcus faecium* to the Genus *Enterococcus* nom. rev. as *Enterococcus faecalis* comb. nov. and *Enterococcus faecium* comb. nov. *International Journal of Systematic and Evolutionary Microbiology* n.d.;34:31–4. <https://doi.org/10.1099/00207713-34-1-31>.
- [16] Scheidegger EMD, Fracalanza S a. P, Teixeira LM, Cardarelli-Leite P. RFLP analysis of a PCR-amplified fragment of the 16S rRNA gene as a tool to identify *Enterococcus* strains. *Mem Inst Oswaldo Cruz* 2009;104:1003–8. <https://doi.org/10.1590/S0074-02762009000700011>.
- [17] Flahaut S, Auffray Y, Boutibonnes P. Les entérocoques dans l'environnement proche de l'homme. *Can J Microbiol* 1997;43:699–708. <https://doi.org/10.1139/m97-101>.
- [18] Qadri SMH, deSilva MJ, Qadri SGM, Villarreal A. Presumptive identification of Enterococci from other D Streptococci by a rapid sodium chloride tolerance test. *Med Microbiol Immunol* 1979;167:197–203. <https://doi.org/10.1007/BF02121186>.

- [19] Knudtson LM, Hartman PA. Routine procedures for isolation and identification of enterococci and fecal streptococci. *Appl Environ Microbiol* 1992;58:3027–31. <https://doi.org/10.1128/aem.58.9.3027-3031.1992>.
- [20] Parte AC. LPSN - List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *Int J Syst Evol Microbiol* 2018;68:1825–9. <https://doi.org/10.1099/ijsem.0.002786>.
- [21] Hardie JM, Whiley RA. Classification and overview of the genera *Streptococcus* and *Enterococcus*. *J Appl Microbiol* 1997;83:1S-11S. <https://doi.org/10.1046/j.1365-2672.83.s1.1.x>.
- [22] Murray BE. The life and times of the *Enterococcus*. *Clin Microbiol Rev* 1990;3:46–65. <https://doi.org/10.1128/CMR.3.1.46>.
- [23] Lebreton F, Willems RJL, Gilmore MS. *Enterococcus* Diversity, Origins in Nature, and Gut Colonization. In: Gilmore MS, Clewell DB, Ike Y, Shankar N, editors. *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection*, Boston: Massachusetts Eye and Ear Infirmary; 2014.
- [24] Sherman JM. The streptococci. *Bacteriol Rev* 1937;1:3–97.
- [25] Selleck EM, Van Tyne D, Gilmore MS. Pathogenicity of Enterococci. *Microbiol Spectr* 2019;7:10.1128/microbiolspec.GPP3-0053–2018. <https://doi.org/10.1128/microbiolspec.GPP3-0053-2018>.
- [26] Bulacio M de LÁ, Galván LR, Gaudio C, Cangemi R, Erimbaue MI. *Enterococcus faecalis* Biofilm. Formation and Development in Vitro Observed by Scanning Electron Microscopy. *Acta Odontol Latinoam* 2015;28:210–4.
- [27] Goh HMS, Yong MHA, Chong KKL, Kline KA. Model systems for the study of Enterococcal colonization and infection. *Virulence* 2017;8:1525–62. <https://doi.org/10.1080/21505594.2017.1279766>.
- [28] Wheeler AL, Hartel PG, Godfrey DG, Hill JL, Segars WI. Potential of *Enterococcus faecalis* as a Human Fecal Indicator for Microbial Source Tracking. *Journal of Environmental Quality* 2002;31:1286–93. <https://doi.org/10.2134/jeq2002.1286>.
- [29] Ostrolenk M, Kramer N, Cleverdon RC. Comparative Studies of Enterococci and *Escherichia coli* as Indices of Pollution. *J Bacteriol* 1947;53:197–203. <https://doi.org/10.1128/jb.53.2.197-203.1947>.
- [30] Franz CMAP, Huch M, Abriouel H, Holzapfel W, Gálvez A. Enterococci as probiotics and their implications in food safety. *International Journal of Food Microbiology* 2011;151:125–40. <https://doi.org/10.1016/j.ijfoodmicro.2011.08.014>.
- [31] Cox MA, Jackson J, Stanton M, Rojas-Triana A, Bober L, Lavery M, *et al.* Short-chain fatty acids act as antiinflammatory mediators by regulating prostaglandin E2 and cytokines. *World J Gastroenterol* 2009;15:5549–57. <https://doi.org/10.3748/wjg.15.5549>.
- [32] Bourgogne A, Garsin DA, Qin X, Singh KV, Sillanpaa J, Yerrapragada S, *et al.* Large scale variation in *Enterococcus faecalis* illustrated by the genome analysis of strain OG1RF. *Genome Biol* 2008;9:R110. <https://doi.org/10.1186/gb-2008-9-7-r110>.
- [33] Wang X, Yang Y, Huycke MM. Risks associated with enterococci as probiotics. *Food Research International* 2020;129:108788. <https://doi.org/10.1016/j.foodres.2019.108788>.
- [34] Wells CL, Jechorek RP, Erlandsen SL. Evidence for the Translocation of *Enterococcus faecalis* across the Mouse Intestinal Tract. *The Journal of Infectious Diseases* 1990;162:82–90. <https://doi.org/10.1093/infdis/162.1.82>.
- [35] Agudelo Higueta NI, Huycke MM. Enterococcal Disease, Epidemiology, and Implications for Treatment. In: Gilmore MS, Clewell DB, Ike Y, Shankar N, editors. *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection*, Boston: Massachusetts Eye and Ear Infirmary; 2014.
- [36] Gordon S, Swenson JM, Hill BC, Pigott NE, Facklam RR, Cooksey RC, *et al.* Antimicrobial susceptibility patterns of common and unusual species of enterococci causing infections in the United States. Enterococcal Study Group. *J Clin Microbiol* 1992;30:2373–8.

- [37] Mohanty S, Behera B. Antibiogram Pattern and Virulence Trait Characterization of *Enterococcus* Species Clinical Isolates in Eastern India: A Recent Analysis. *J Lab Physicians* 2022;14:237–46. <https://doi.org/10.1055/s-0042-1750085>.
- [38] SPF. Bilan 2001-2017 des signalements externes d'infections nosocomiales. Part des signalements impliquant une bactérie multirésistante, hautement résistante-émergente ou un *Clostridium difficile* n.d. <https://www.santepubliquefrance.fr/import/bilan-2001-2017-des-signalements-externes-d-infections-nosocomiales.-part-des-signalements-impliquant-une-bacterie-multiresistante-hautement-resis> (accessed September 6, 2023).
- [39] Leblanc DJ. *Enterococcus*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. *The Prokaryotes: Volume 4: Bacteria: Firmicutes, Cyanobacteria*, New York, NY: Springer US; 2006, p. 175–204. https://doi.org/10.1007/0-387-30744-3_6.
- [40] Miller WR, Murray BE, Rice LB, Arias CA. Resistance in Vancomycin-Resistant *Enterococci*. *Infect Dis Clin North Am* 2020;34:751–71. <https://doi.org/10.1016/j.idc.2020.08.004>.
- [41] Ahmed MO, Baptiste KE. Vancomycin-Resistant *Enterococci*: A Review of Antimicrobial Resistance Mechanisms and Perspectives of Human and Animal Health. *Microbial Drug Resistance* 2018;24:590–606. <https://doi.org/10.1089/mdr.2017.0147>.
- [42] Stogios PJ, Savchenko A. Molecular mechanisms of vancomycin resistance. *Protein Sci* 2020;29:654–69. <https://doi.org/10.1002/pro.3819>.
- [43] Geraldes C, Tavares L, Gil S, Oliveira M. *Enterococcus* Virulence and Resistant Traits Associated with Its Permanence in the Hospital Environment. *Antibiotics (Basel)* 2022;11:857. <https://doi.org/10.3390/antibiotics11070857>.
- [44] Garsin DA, Frank KL, Silanpää J, Ausubel FM, Hartke A, Shankar N, *et al.* Pathogenesis and Models of Enterococcal Infection. In: Gilmore MS, Clewell DB, Ike Y, Shankar N, editors. *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection*, Boston: Massachusetts Eye and Ear Infirmary; 2014.
- [45] Soussan D, Salze M, Ledormand P, Sauvageot N, Boukerb A, Lesouhaitier O, *et al.* The NagY regulator: A member of the BglG/SacY antiterminator family conserved in *Enterococcus faecalis* and involved in virulence. *Front Microbiol* 2022;13:1070116. <https://doi.org/10.3389/fmicb.2022.1070116>.
- [46] Mohamed JA, Huang DB. Biofilm formation by enterococci. *Journal of Medical Microbiology* 2007;56:1581–8. <https://doi.org/10.1099/jmm.0.47331-0>.
- [47] Michaelis C, Grohmann E. Horizontal Gene Transfer of Antibiotic Resistance Genes in Biofilms. *Antibiotics (Basel)* 2023;12:328. <https://doi.org/10.3390/antibiotics12020328>.
- [48] Pinkston KL, Gao P, Diaz-Garcia D, Sillanpää J, Nallapareddy SR, Murray BE, *et al.* The Fsr Quorum-Sensing System of *Enterococcus faecalis* Modulates Surface Display of the Collagen-Binding MSCRAMM Ace through Regulation of *gelE*. *J Bacteriol* 2011;193:4317–25. <https://doi.org/10.1128/JB.05026-11>.
- [49] McBrayer DN, Gantman BK, Tal-Gan Y. N-methylation of amino acids in Gelatinase Biosynthesis-Activating Pheromone (GBAP) Identifies Key Site for Stability Enhancement with Retention of the *Enterococcus faecalis* Fsr Quorum Sensing Circuit Response. *ACS Infect Dis* 2019;5:1035–41. <https://doi.org/10.1021/acsinfecdis.9b00097>.
- [50] Yue J, Hua M, Chen N, Li J, Liu X, Duan A, *et al.* Role of the *fsr* Quorum-Sensing System in *Enterococcus faecalis* Bloodstream Infection. *Appl Environ Microbiol* n.d.;88:e01551-22. <https://doi.org/10.1128/aem.01551-22>.
- [51] Madsen KT, Skov MN, Gill S, Kemp M. Virulence Factors Associated with *Enterococcus Faecalis* Infective Endocarditis: A Mini Review. *Open Microbiol J* 2017;11:1–11. <https://doi.org/10.2174/1874285801711010001>.
- [52] Ali L, Goraya M, Arafat Y, Ajmal M, Chen J-L, Yu D. Molecular Mechanism of Quorum-Sensing in *Enterococcus faecalis*: Its Role in Virulence and Therapeutic Approaches. *IJMS* 2017;18:960. <https://doi.org/10.3390/ijms18050960>.
- [53] Van Tyne D, Martin M, Gilmore M. Structure, Function, and Biology of the *Enterococcus faecalis* Cytolysin. *Toxins* 2013;5:895–911. <https://doi.org/10.3390/toxins5050895>.

- [54] Shankar N, Baghdayan AS, Gilmore MS. Modulation of virulence within a pathogenicity island in vancomycin-resistant *Enterococcus faecalis*. *Nature* 2002;417:746–50. <https://doi.org/10.1038/nature00802>.
- [55] Salze M. Identification et caractérisation d'ARN régulateurs impliqués dans la réponse au stress et la virulence chez *Enterococcus faecalis*. phdthesis. Normandie Université, 2019.
- [56] Vergis EN, Shankar N, Chow JW, Hayden MK, Snyderman DR, Zervos MJ, *et al.* Association between the Presence of Enterococcal Virulence Factors Gelatinase, Hemolysin, and Enterococcal Surface Protein and Mortality among Patients with Bacteremia Due to *Enterococcus faecalis*. *Clinical Infectious Diseases* 2002;35:570–5. <https://doi.org/10.1086/341977>.
- [57] Kao PHN, Kline KA. Dr. Jekyll and Mr. Hide: How *Enterococcus faecalis* Subverts the Host Immune Response to Cause Infection. *Journal of Molecular Biology* 2019;431:2932–45. <https://doi.org/10.1016/j.jmb.2019.05.030>.
- [58] Giridhara Upadhyaya PM, Ravikumar KL, Umopathy BL. Review of virulence factors of enterococcus: an emerging nosocomial pathogen. *Indian J Med Microbiol* 2009;27:301–5. <https://doi.org/10.4103/0255-0857.55437>.
- [59] Palmer KL, van Schaik W, Willems RJL, Gilmore MS. Enterococcal Genomics. In: Gilmore MS, Clewell DB, Ike Y, Shankar N, editors. *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection*, Boston: Massachusetts Eye and Ear Infirmary; 2014.
- [60] Mylonakis E, Engelbert M, Qin X, Sifri CD, Murray BE, Ausubel FM, *et al.* The *Enterococcus faecalis* *fsrB* Gene, a Key Component of the *fsr* Quorum-Sensing System, Is Associated with Virulence in the Rabbit Endophthalmitis Model. *Infect Immun* 2002;70:4678–81. <https://doi.org/10.1128/IAI.70.8.4678-4681.2002>.
- [61] Singh KV, Coque TM, Weinstock GM, Murray BE. In vivo testing of an *Enterococcus faecalis* *efaA* mutant and use of *efaA* homologs for species identification. *FEMS Immunol Med Microbiol* 1998;21:323–31. <https://doi.org/10.1111/j.1574-695X.1998.tb01180.x>.
- [62] Singh KV, Nallapareddy SR, Murray BE. Importance of the *ebp* (endocarditis- and biofilm-associated pilus) locus in the pathogenesis of *Enterococcus faecalis* ascending urinary tract infection. *J Infect Dis* 2007;195:1671–7. <https://doi.org/10.1086/517524>.
- [63] Tsai CJ-Y, Loh JMS, Proft T. *Galleria mellonella* infection models for the study of bacterial diseases and for antimicrobial drug testing. *Virulence* 2016;7:214–29. <https://doi.org/10.1080/21505594.2015.1135289>.
- [64] Suvorova IA, Korostelev YD, Gelfand MS. GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution. *PLOS ONE* 2015;10:e0132618. <https://doi.org/10.1371/journal.pone.0132618>.
- [65] Finn RD. Pfam: the protein families database. In: Jorde LB, Little PFR, Dunn MJ, Subramaniam S, editors. *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*, Chichester, UK: John Wiley & Sons, Ltd; 2005, p. g306303. <https://doi.org/10.1002/047001153X.g306303>.
- [66] Hoskisson PA, Rigali S. Chapter 1: Variation in form and function the helix-turn-helix regulators of the GntR superfamily. *Adv Appl Microbiol* 2009;69:1–22. [https://doi.org/10.1016/S0065-2164\(09\)69001-8](https://doi.org/10.1016/S0065-2164(09)69001-8).
- [67] Jain D. Allosteric control of transcription in GntR family of transcription regulators: A structural overview. *IUBMB Life* 2015;67:556–63. <https://doi.org/10.1002/iub.1401>.
- [68] Haydon DJ, Guest JR. A new family of bacterial regulatory proteins. *FEMS Microbiology Letters* 1991;79:291–6. <https://doi.org/10.1111/j.1574-6968.1991.tb04544.x>.
- [69] Vindal V, Suma K, Ranjan A. GntR family of regulators in *Mycobacterium smegmatis*: a sequence and structure based characterization. *BMC Genomics* 2007;8:289. <https://doi.org/10.1186/1471-2164-8-289>.
- [70] Blancato VS, Repizo GD, Suárez CA, Magni C. Transcriptional regulation of the citrate gene cluster of *Enterococcus faecalis* Involves the GntR family transcriptional activator CitO. *J Bacteriol* 2008;190:7419–30. <https://doi.org/10.1128/JB.01704-07>.

- [71] Fillenberg SB, Grau FC, Seidel G, Muller YA. Structural insight into operator *dre*-sites recognition and effector binding in the GntR/HutC transcription regulator NagR. *Nucleic Acids Res* 2015;43:1283–96. <https://doi.org/10.1093/nar/gku1374>.
- [72] MMDB and VAST+: tracking structural similarities between macromolecular complexes - PubMed n.d. <https://pubmed.ncbi.nlm.nih.gov/24319143/> (accessed October 18, 2023).
- [73] Aravind L, Anantharaman V. HutC/FarR-like bacterial transcription factors of the GntR family contain a small molecule-binding domain of the chorismate lyase fold. *FEMS Microbiology Letters* 2003;222:17–23. [https://doi.org/10.1016/S0378-1097\(03\)00242-8](https://doi.org/10.1016/S0378-1097(03)00242-8).
- [74] Resch M, Schiltz E, Titgemeyer F, Muller YA. Insight into the induction mechanism of the GntR/HutC bacterial transcription regulator YvoA. *Nucleic Acids Res* 2010;38:2485–97. <https://doi.org/10.1093/nar/gkp1191>.
- [75] Milano T, Angelaccio S, Tramonti A, Di Salvo ML, Contestabile R, Pascarella S. Structural properties of the linkers connecting the N- and C- terminal domains in the MocR bacterial transcriptional regulators. *Biochim Open* 2016;3:8–18. <https://doi.org/10.1016/j.biopen.2016.07.002>.
- [76] Kundig W, Ghosh S, Roseman S. Phosphate bounds to histidine in a protein as an intermediate in a novel phospho-transferase system. *Proc Natl Acad Sci U S A* 1964;52:1067–74. <https://doi.org/10.1073/pnas.52.4.1067>.
- [77] Barabote RD, Saier MH. Comparative Genomic Analyses of the Bacterial Phosphotransferase System. *Microbiology and Molecular Biology Reviews* 2005;69:608–34. <https://doi.org/10.1128/mubr.69.4.608-634.2005>.
- [78] Postma PW, Lengeler JW, Jacobson GR. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev* 1993;57:543–94. <https://doi.org/10.1128/mr.57.3.543-594.1993>.
- [79] Galinier A, Deutscher J. Sophisticated Regulation of Transcriptional Factors by the Bacterial Phosphoenolpyruvate: Sugar Phosphotransferase System. *Journal of Molecular Biology* 2017;429:773–89. <https://doi.org/10.1016/j.jmb.2017.02.006>.
- [80] Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;45:D353–61. <https://doi.org/10.1093/nar/gkw1092>.
- [81] Gonzy-Tréboul G, Steinmetz M. Phosphoenolpyruvate:sugar phosphotransferase system of *Bacillus subtilis*: cloning of the region containing the *ptsH* and *ptsI* genes and evidence for a *crr*-like gene. *J Bacteriol* 1987;169:2287–90. <https://doi.org/10.1128/jb.169.5.2287-2290.1987>.
- [82] De Reuse H, Kolb A, Danchin A. Positive regulation of the expression of the *Escherichia coli pts* operon. Identification of the regulatory regions. *J Mol Biol* 1992;226:623–35. [https://doi.org/10.1016/0022-2836\(92\)90620-y](https://doi.org/10.1016/0022-2836(92)90620-y).
- [83] Gagnon G, Vadeboncoeur C, Gauthier L, Frenette M. Regulation of *ptsH* and *ptsI* gene expression in *Streptococcus salivarius* ATCC 25975. *Molecular Microbiology* 1995;16:1111–21. <https://doi.org/10.1111/j.1365-2958.1995.tb02336.x>.
- [84] Christensen DP, Benson AK, Hutkins RW. Cloning and expression of the *Listeria monocytogenes* scott A *ptsH* and *ptsI* genes, coding for HPr and enzyme I, respectively, of the phosphotransferase system. *Appl Environ Microbiol* 1998;64:3147–52. <https://doi.org/10.1128/AEM.64.9.3147-3152.1998>.
- [85] Vadeboncoeur C, Frenette M, Lortie LA. Regulation of the *pts* operon in low G+C Gram-positive bacteria. *J Mol Microbiol Biotechnol* 2000;2:483–90.
- [86] Lai X, Ingram LO. Discovery of a *ptsHI* operon, which includes a third gene (*ptsT*), in the thermophile *Bacillus stearothermophilus*. *Microbiology (Reading)* 1995;141 (Pt 6):1443–9. <https://doi.org/10.1099/13500872-141-6-1443>.
- [87] Nguyen CC, Saier MH. Phylogenetic analysis of the putative phosphorylation domain in the pyruvate kinase of *Bacillus stearothermophilus*. *Research in Microbiology* 1995;146:713–9. [https://doi.org/10.1016/0923-2508\(96\)81067-9](https://doi.org/10.1016/0923-2508(96)81067-9).

- [88] Patel HV, Vyas KA, Mattoo RL, Southworth M, Perler FB, Comb D, *et al.* Properties of the C-terminal Domain of Enzyme I of the *Escherichia coli* Phosphotransferase System *. Journal of Biological Chemistry 2006;281:17579–87. <https://doi.org/10.1074/jbc.M508966200>.
- [89] Chauvin F, Brand L, Roseman S. Sugar transport by the bacterial phosphotransferase system. Characterization of the *Escherichia coli* enzyme I monomer/dimer equilibrium by fluorescence anisotropy. J Biol Chem 1994;269:20263–9.
- [90] Hu K-Y, Saier MH. Phylogeny of phosphoryl transfer proteins of the phosphoenolpyruvate-dependent sugar-transporting phosphotransferase system. Res Microbiol 2002;153:405–15. [https://doi.org/10.1016/s0923-2508\(02\)01339-6](https://doi.org/10.1016/s0923-2508(02)01339-6).
- [91] Jia Z, Vandonselaar M, Hengstenberg W, Wilson Quail J, Delbaere LTJ. The 1.6 Å structure of histidine-containing phosphotransfer protein HPr from *Streptococcus faecalis*. Journal of Molecular Biology 1994;236:1341–55. [https://doi.org/10.1016/0022-2836\(94\)90062-0](https://doi.org/10.1016/0022-2836(94)90062-0).
- [92] Kravanja M, Engelmann R, Dossonnet V, Blüggel M, Meyer HE, Frank R, *et al.* The *hprK* gene of *Enterococcus faecalis* encodes a novel bifunctional enzyme: the HPr kinase/phosphatase. Molecular Microbiology 1999;31:59–66. <https://doi.org/10.1046/j.1365-2958.1999.01146.x>.
- [93] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30. <https://doi.org/10.1093/nar/28.1.27>.
- [94] Grand M, Blancato VS, Espariz M, Deutscher J, Pikis A, Hartke A, *et al.* *Enterococcus faecalis* MalR acts as a repressor of the maltose operons and additionally mediates their catabolite repression via direct interaction with seryl-phosphorylated-HPr. Mol Microbiol 2020;113:464–77. <https://doi.org/10.1111/mmi.14431>.
- [95] Deutscher J, Galinier A, Martin-Verstraete I. Carbohydrate Uptake and Metabolism. *Bacillus subtilis* and Its Closest Relatives, John Wiley & Sons, Ltd; 2001, p. 129–50. <https://doi.org/10.1128/9781555817992.ch11>.
- [96] Deutscher J, Francke C, Postma PW. How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria. Microbiol Mol Biol Rev 2006;70:939–1031. <https://doi.org/10.1128/MMBR.00024-06>.
- [97] Galinier A. La répression catabolique ou comment les bactéries choisissent leurs sucres préférés. Med Sci (Paris) 2018;34:531–9. <https://doi.org/10.1051/medsci/20183406012>.
- [98] Kumar S, Bhadane R, Shandilya S, Salo-Ahen OMH, Kapila S. Identification of HPr kinase/phosphorylase inhibitors: novel antimicrobials against resistant *Enterococcus faecalis*. J Comput Aided Mol Des 2022;36:507–20. <https://doi.org/10.1007/s10822-022-00461-6>.
- [99] Nessler S. The bacterial HPr kinase/phosphorylase: A new type of Ser/Thr kinase as antimicrobial target. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 2005;1754:126–31. <https://doi.org/10.1016/j.bbapap.2005.07.042>.
- [100] Mijakovic I, Poncet S, Galinier A, Monedero V, Fieulaine S, Janin J, *et al.* Pyrophosphate-producing protein dephosphorylation by HPr kinase/phosphorylase: A relic of early life? Proceedings of the National Academy of Sciences 2002;99:13442–7. <https://doi.org/10.1073/pnas.212410399>.
- [101] Deutscher J, Engelmann R. Purification and characterization of an ATP-dependent protein kinase from *Streptococcus faecalis*. FEMS Microbiology Letters 1984;23:157–62. <https://doi.org/10.1111/j.1574-6968.1984.tb01053.x>.
- [102] Stülke J, Schmalisch MH. The bacterial phosphotransferase system: a perfect link of sugar transport and signal transduction. Molecular Mechanisms Controlling Transmembrane Transport, Berlin, Heidelberg: Springer; 2004, p. 179–205. <https://doi.org/10.1007/b95776>.
- [103] Deutscher J, Kessler U, Alpert CA, Hengstenberg W. Bacterial phosphoenolpyruvate-dependent phosphotransferase system: P-Ser-HPr and its possible regulatory function? Biochemistry 1984;23:4455–60. <https://doi.org/10.1021/bi00314a033>.
- [104] Leboeuf C, Leblanc L, Auffray Y, Hartke A. Characterization of the *ccpA* Gene of *Enterococcus faecalis*: Identification of Starvation-Inducible Proteins Regulated by CcpA. Journal of Bacteriology 2000. <https://doi.org/10.1128/JB.182.20.5799-5806.2000>.

- [105] Opsata M, Nes IF, Holo H. Class IIa bacteriocin resistance in *Enterococcus faecalis* V583: The mannose PTS operon mediates global transcriptional responses. *BMC Microbiology* 2010;10:224. <https://doi.org/10.1186/1471-2180-10-224>.
- [106] Jones BE, Dossonnet V, Küster E, Hillen W, Deutscher J, Klevit RE. Binding of the catabolite repressor protein CcpA to its DNA target is regulated by phosphorylation of its corepressor HPr. *J Biol Chem* 1997;272:26530–5. <https://doi.org/10.1074/jbc.272.42.26530>.
- [107] Schumacher MA, Allen GS, Diel M, Seidel G, Hillen W, Brennan RG. Structural basis for allosteric control of the transcription regulator CcpA by the phosphoprotein HPr-Ser46-P. *Cell* 2004;118:731–41. <https://doi.org/10.1016/j.cell.2004.08.027>.
- [108] Hueck CJ, Hillen W, Saier MH. Analysis of a cis-active sequence mediating catabolite repression in Gram-positive bacteria. *Research in Microbiology* 1994;145:503–18. [https://doi.org/10.1016/0923-2508\(94\)90028-0](https://doi.org/10.1016/0923-2508(94)90028-0).
- [109] Kim J-H, Voskuil MI, Chambliss GH. NADP, corepressor for the *Bacillus* catabolite control protein CcpA. *Proc Natl Acad Sci U S A* 1998;95:9590–5.
- [110] Müller W, Horstmann N, Hillen W, Sticht H. The transcription regulator RbsR represents a novel interaction partner of the phosphoprotein HPr-Ser46-P in *Bacillus subtilis*. *FEBS J* 2006;273:1251–61. <https://doi.org/10.1111/j.1742-4658.2006.05148.x>.
- [111] Monedero V, Yebra MJ, Poncet S, Deutscher J. Maltose transport in *Lactobacillus casei* and its regulation by inducer exclusion. *Res Microbiol* 2008;159:94–102. <https://doi.org/10.1016/j.resmic.2007.10.002>.
- [112] Cook GM, Kearns DB, Russell JB, Reizer J, Saier MH. Regulation of the lactose phosphotransferase system of *Streptococcus bovis* by glucose: independence of inducer exclusion and expulsion mechanisms. *Microbiology (Reading)* 1995;141 (Pt 9):2261–9. <https://doi.org/10.1099/13500872-141-9-2261>.
- [113] Luesink EJ, Kuipers OP, Vos WM de. Regulation of the carbohydrate metabolism in *Lactococcus lactis* and other lactic acid bacteria. *Lait* 1998;78:69–76. <https://doi.org/10.1051/lait:199819>.
- [114] Deutscher J, Aké FMD, Derkaoui M, Zébré AC, Cao TN, Bouraoui H, *et al*. The Bacterial Phosphoenolpyruvate:Carbohydrate Phosphotransferase System: Regulation by Protein Phosphorylation and Phosphorylation-Dependent Protein-Protein Interactions. *Microbiol Mol Biol Rev* 2014;78:231–56. <https://doi.org/10.1128/MMBR.00001-14>.
- [115] Ye JJ, Saier MH. Purification and characterization of a small membrane-associated sugar phosphate phosphatase that is allosterically activated by HPr(Ser(P)) of the phosphotransferase system in *Lactococcus lactis*. *J Biol Chem* 1995;270:16740–4. <https://doi.org/10.1074/jbc.270.28.16740>.
- [116] Reizer J, Novotny MJ, Panos C, Saier MH. Mechanism of inducer expulsion in *Streptococcus pyogenes*: a two-step process activated by ATP. *J Bacteriol* 1983;156:354–61. <https://doi.org/10.1128/jb.156.1.354-361.1983>.
- [117] Heravi KM, Altenbuchner J. Regulation of the *Bacillus subtilis* mannitol utilization genes: promoter structure and transcriptional activation by the wild-type regulator (MtlR) and its mutants. *Microbiology (Reading)* 2014;160:91–101. <https://doi.org/10.1099/mic.0.071233-0>.
- [118] Rothe FM, Wrede C, Lehnik-Habrink M, Görke B, Stülke J. Dynamic localization of a transcription factor in *Bacillus subtilis*: the LicT antiterminator relocates in response to inducer availability. *J Bacteriol* 2013;195:2146–54. <https://doi.org/10.1128/JB.00117-13>.
- [119] van Tilbeurgh H, Le Coq D, Declerck N. Crystal structure of an activated form of the PTS regulation domain from the LicT transcriptional antiterminator. *EMBO J* 2001;20:3789–99. <https://doi.org/10.1093/emboj/20.14.3789>.
- [120] Deutscher J. The mechanisms of carbon catabolite repression in bacteria. *Current Opinion in Microbiology* 2008;11:87–93. <https://doi.org/10.1016/j.mib.2008.02.007>.
- [121] Grand M. Régulation des opérons Maltose/Maltodextrines et Gentiobiose induits en contexte d'infection chez *Enterococcus faecalis*. Theses. Normandie Université, 2019.

- [122] Roy DJ, Casabon I, Vaillancourt K, Huot JL, Vadeboncoeur C. Streptococci and lactococci synthesize large amounts of HPr(Ser-P)(His~P). *Can J Microbiol* 2008;54:941–9. <https://doi.org/10.1139/W08-085>.
- [123] Chen J-K, Shen C-R, Liu C-L. N-Acetylglucosamine: Production and Applications. *Mar Drugs* 2010;8:2493–516. <https://doi.org/10.3390/md8092493>.
- [124] Keffeler EC, Parthasarathy S, Abdullahi ZH, Hancock LE. Metabolism of Poly- β 1,4-N-Acetylglucosamine Substrates and Importation of N-Acetylglucosamine and Glucosamine by *Enterococcus faecalis*. *J Bacteriol* 2021;203:e0037121. <https://doi.org/10.1128/JB.00371-21>.
- [125] Vaaje-Kolstad G, Bøhle LA, Gåseidnes S, Dalhus B, Bjørås M, Mathiesen G, *et al.* Characterization of the Chitinolytic Machinery of *Enterococcus faecalis* V583 and High-Resolution Structure of Its Oxidative CBM33 Enzyme. *Journal of Molecular Biology* 2012;416:239–54. <https://doi.org/10.1016/j.jmb.2011.12.033>.
- [126] Chen J, Tsai Y-H. Applications of Genetic Code Expansion in Studying Protein Post-translational Modification. *Journal of Molecular Biology* 2022;434:167424. <https://doi.org/10.1016/j.jmb.2021.167424>.
- [127] Abeywickrama TD, Perera IC. In Silico Characterization and Virtual Screening of GntR/HutC Family Transcriptional Regulator MoyR: A Potential Monooxygenase Regulator in *Mycobacterium tuberculosis*. *Biology (Basel)* 2021;10:1241. <https://doi.org/10.3390/biology10121241>.
- [128] Knorre DG, Kudryashova NV, Godovikova TS. Chemical and functional aspects of posttranslational modification of proteins. *Acta Naturae* 2009;1:29–51.
- [129] Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 2002;12:43R-56R. <https://doi.org/10.1093/glycob/12.4.43R>.
- [130] Apweiler R. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1999;1473:4–8. [https://doi.org/10.1016/S0304-4165\(99\)00165-8](https://doi.org/10.1016/S0304-4165(99)00165-8).
- [131] Rademacher TW, Parekh RB, Dwek RA. Glycobiology. *Annual Review of Biochemistry* 1988;57:785–838. <https://doi.org/10.1146/annurev.bi.57.070188.004033>.
- [132] Hirata T, Kizuka Y. N-Glycosylation. In: Lauc G, Trbojević-Akmačić I, editors. *The Role of Glycosylation in Health and Disease*, Cham: Springer International Publishing; 2021, p. 3–24. https://doi.org/10.1007/978-3-030-70115-4_1.
- [133] Stanley P, Taniguchi N, Aebi M. N-Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, *et al.*, editors. *Essentials of Glycobiology*. 3rd ed., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015.
- [134] Steen PV den, Rudd PM, Dwek RA, Opdenakker G. Concepts and Principles of O-Linked Glycosylation. *Critical Reviews in Biochemistry and Molecular Biology* 2008. <https://doi.org/10.1080/10409239891204198>.
- [135] Perdivara I, Yamauchi M, Tomer KB. Molecular Characterization of Collagen Hydroxylysine O-Glycosylation by Mass Spectrometry: Current Status. *Aust J Chem* 2013;66:760–9. <https://doi.org/10.1071/CH13174>.
- [136] Shpak E, Leykam JF, Kieliszewski MJ. Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes. *Proc Natl Acad Sci U S A* 1999;96:14736–41.
- [137] Thompson N, Wakarchuk W. O-glycosylation and its role in therapeutic proteins. *Biosci Rep* 2022;42:BSR20220094. <https://doi.org/10.1042/BSR20220094>.
- [138] Brockhausen I, Stanley P. O-GalNAc Glycans. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, *et al.*, editors. *Essentials of Glycobiology*. 3rd ed., Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015.
- [139] Nothaft H, Szymanski CM. New discoveries in bacterial N-glycosylation to expand the synthetic biology toolbox. *Current Opinion in Chemical Biology* 2019;53:16–24. <https://doi.org/10.1016/j.cbpa.2019.05.032>.

- [140] Vershinin Z, Zaretsky M, Guan Z, Eichler J. Revisiting N-glycosylation in *Halobacterium salinarum*: Characterizing a dolichol phosphate- and glycoprotein-bound tetrasaccharide. *Glycobiology* 2021;31:1645–54. <https://doi.org/10.1093/glycob/cwab080>.
- [141] Szymanski CM. Bacterial glycosylation, it's complicated. *Front Mol Biosci* 2022;9:1015771. <https://doi.org/10.3389/fmolb.2022.1015771>.
- [142] Szymanski CM, Yao R, Ewing CP, Trust TJ, Guerry P. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol Microbiol* 1999;32:1022–30. <https://doi.org/10.1046/j.1365-2958.1999.01415.x>.
- [143] Chen MM, Weerapana E, Ciepichal E, Stupak J, Reid CW, Swiezewska E, et al. Polyisoprenol Specificity in the *Campylobacter jejuni* N-linked Glycosylation Pathway. *Biochemistry* 2007;46:14342–8. <https://doi.org/10.1021/bi701956x>.
- [144] Srichaisupakit A, Ohashi T, Fujiyama K. Identification of a protein glycosylation operon from *Campylobacter jejuni* JCM 2013 and its heterologous expression in *Escherichia coli*. *Journal of Bioscience and Bioengineering* 2014;118:256–62. <https://doi.org/10.1016/j.jbiosc.2014.02.011>.
- [145] Cain JA, Dale AL, Sumer-Bayraktar Z, Solis N, Cordwell SJ. Identifying the targets and functions of N-linked protein glycosylation in *Campylobacter jejuni*. *Mol Omics* 2020;16:287–304. <https://doi.org/10.1039/d0mo00032a>.
- [146] Salah Ud-Din AIM, Roujeinikova A. Flagellin glycosylation with pseudaminic acid in *Campylobacter* and *Helicobacter*: prospects for development of novel therapeutics. *Cell Mol Life Sci* 2018;75:1163–78. <https://doi.org/10.1007/s00018-017-2696-5>.
- [147] McNally DJ, Hui JPM, Aubry AJ, Mui KKK, Guerry P, Brisson J-R, et al. Functional characterization of the flagellar glycosylation locus in *Campylobacter jejuni* 81-176 using a focused metabolomics approach. *J Biol Chem* 2006;281:18489–98. <https://doi.org/10.1074/jbc.M603777200>.
- [148] Szymanski CM, Wren BW. Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol* 2005;3:225–37. <https://doi.org/10.1038/nrmicro1100>.
- [149] Zebian N, Merckx-Jacques A, Pittock PP, Houle S, Dozois CM, Lajoie GA, et al. Comprehensive analysis of flagellin glycosylation in *Campylobacter jejuni* NCTC 11168 reveals incorporation of legionaminic acid and its importance for host colonization. *Glycobiology* 2016;26:386–97. <https://doi.org/10.1093/glycob/cwv104>.
- [150] Moormann C, Benz I, Schmidt MA. Functional Substitution of the TibC Protein of Enterotoxigenic *Escherichia coli* Strains for the Autotransporter Adhesin Heptosyltransferase of the AIDA System. *Infect Immun* 2002;70:2264–70. <https://doi.org/10.1128/IAI.70.5.2264-2270.2002>.
- [151] Berger S, Menudier A, Julien R, Karamanos Y. Do de-N-glycosylation enzymes have an important role in plant cells? *Biochimie* 1995;77:751–60. [https://doi.org/10.1016/0300-9084\(96\)88193-4](https://doi.org/10.1016/0300-9084(96)88193-4).
- [152] Yin J, Li L, Shaw N, Li Y, Song JK, Zhang W, et al. Structural Basis and Catalytic Mechanism for the Dual Functional Endo- β -N-Acetylglucosaminidase A. *PLoS One* 2009;4:e4658. <https://doi.org/10.1371/journal.pone.0004658>.
- [153] Sjögren J, Struwe WB, Cosgrave EFJ, Rudd PM, Stervander M, Allhorn M, et al. EndoS2 is a unique and conserved enzyme of serotype M49 group A *Streptococcus* that hydrolyses N-linked glycans on IgG and α 1-acid glycoprotein. *Biochem J* 2013;455:107–18. <https://doi.org/10.1042/BJ20130126>.
- [154] Shadnezhad A, Naegeli A, Sjögren J, Adamczyk B, Leo F, Allhorn M, et al. EndoSd: an IgG glycan hydrolyzing enzyme in *Streptococcus dysgalactiae* subspecies *dysgalactiae*. *Future Microbiol* 2016;11:721–36. <https://doi.org/10.2217/fmb.16.14>.
- [155] Roberts G, Tarelli E, Homer KA, Philpott-Howard J, Beighton D. Production of an Endo- β -N-Acetylglucosaminidase Activity Mediates Growth of *Enterococcus faecalis* on a High-Mannose-Type Glycoprotein. *J Bacteriol* 2000;182:882–90.
- [156] KOBATA A. Exo- and endoglycosidases revisited. *Proc Jpn Acad Ser B Phys Biol Sci* 2013;89:97–117. <https://doi.org/10.2183/pjab.89.97>.
- [157] Garbe J, Sjögren J, Cosgrave EFJ, Struwe WB, Bober M, Olin AI, et al. EndoE from *Enterococcus faecalis* Hydrolyzes the Glycans of the Biofilm Inhibiting Protein Lactoferrin and Mediates Growth. *PLoS ONE* 2014;9:e91035. <https://doi.org/10.1371/journal.pone.0091035>.

- [158] Collin M, Fischetti VA. A Novel Secreted Endoglycosidase from *Enterococcus faecalis* with Activity on Human Immunoglobulin G and Ribonuclease B. *Journal of Biological Chemistry* 2004;279:22558–70. <https://doi.org/10.1074/jbc.M402156200>.
- [159] Fairbanks AJ. The ENGases: versatile biocatalysts for the production of homogeneous N-linked glycopeptides and glycoproteins. *Chem Soc Rev* 2017;46:5128–46. <https://doi.org/10.1039/c6cs00897f>.
- [160] Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* 1991;280 (Pt 2):309–16. <https://doi.org/10.1042/bj2800309>.
- [161] Keffeler EC, Iyer VS, Henderson AJ, Huck IL, Schwarting N, Cortez A, *et al.* Activity of CcpA-Regulated GH18 Family Glycosyl Hydrolases That Contributes to Nutrient Acquisition and Fitness in *Enterococcus faecalis*. *Infect Immun* 2021;89:e00343-21. <https://doi.org/10.1128/IAI.00343-21>.
- [162] Collin M, Olsén A. EndoS, a novel secreted protein from *Streptococcus pyogenes* with endoglycosidase activity on human IgG. *EMBO J* 2001;20:3046–55. <https://doi.org/10.1093/emboj/20.12.3046>.
- [163] Naegeli A, Bratanis E, Karlsson C, Shannon O, Kalluru R, Linder A, *et al.* *Streptococcus pyogenes* evades adaptive immunity through specific IgG glycan hydrolysis. *Journal of Experimental Medicine* 2019;216:1615–29. <https://doi.org/10.1084/jem.20190293>.
- [164] Makino M, Kojima T, Yamashina I. Enzymatic cleavage of glycopeptides. *Biochemical and Biophysical Research Communications* 1966;24:961–6. [https://doi.org/10.1016/0006-291X\(66\)90344-5](https://doi.org/10.1016/0006-291X(66)90344-5).
- [165] Aronson NN. Aspartylglycosaminuria: biochemistry and molecular biology. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1999;1455:139–54. [https://doi.org/10.1016/S0925-4439\(99\)00076-9](https://doi.org/10.1016/S0925-4439(99)00076-9).
- [166] Takahashi N. Demonstration of a new amidase acting on glycopeptides. *Biochem Biophys Res Commun* 1977;76:1194–201. [https://doi.org/10.1016/0006-291x\(77\)90982-2](https://doi.org/10.1016/0006-291x(77)90982-2).
- [167] Hirayama H, Hosomi A, Suzuki T. Physiological and molecular functions of the cytosolic peptide:N-glycanase. *Seminars in Cell & Developmental Biology* 2015;41:110–20. <https://doi.org/10.1016/j.semcdb.2014.11.009>.
- [168] Gosain A, Lohia R, Shrivastava A, Saran S. Identification and characterization of peptide: N-glycanase from *Dictyostelium discoideum*. *BMC Biochem* 2012;13:9. <https://doi.org/10.1186/1471-2091-13-9>.
- [169] Chantret I, Couvineau A, Moore S. Nouvelles fonctions de la peptide N-glycanase indépendantes de sa capacité de déglycosylation. *Med Sci (Paris)* 2014;30:47–54. <https://doi.org/10.1051/medsci/20143001013>.
- [170] Arvio M, Mononen I. Aspartylglycosaminuria: a review. *Orphanet J Rare Dis* 2016;11:162. <https://doi.org/10.1186/s13023-016-0544-6>.
- [171] Sun G, Yu X, Bao C, Wang L, Li M, Gan J, *et al.* Identification and characterization of a novel prokaryotic peptide: N-glycosidase from *Elizabethkingia meningoseptica*. *J Biol Chem* 2015;290:7452–62. <https://doi.org/10.1074/jbc.M114.605493>.
- [172] Tarentino AL, Quinones G, Trumble A, Changchien LM, Duceman B, Maley F, *et al.* Molecular cloning and amino acid sequence of peptide-N4-(N-acetyl-beta-D-glucosaminyl)asparagine amidase from *Flavobacterium meningosepticum*. *Journal of Biological Chemistry* 1990;265:6961–6. [https://doi.org/10.1016/S0021-9258\(19\)39244-0](https://doi.org/10.1016/S0021-9258(19)39244-0).
- [173] Suzuki T, Park H, Kitajima K, Lennarz WJ. Peptides Glycosylated in the Endoplasmic Reticulum of Yeast Are Subsequently Deglycosylated by a Soluble Peptide: N-Glycanase Activity. *Journal of Biological Chemistry* 1998;273:21526–30. <https://doi.org/10.1074/jbc.273.34.21526>.
- [174] Risley JM, Huang DH, Kaylor JJ, Malik JJ, Xia Y-Q, York WM. Glycosylasparaginase Activity Requires the α -Carboxyl Group, but Not the α -Amino Group, on N4-(2-Acetamido- 2-deoxy- β -d-glucopyranosyl)-l-asparagine. *Archives of Biochemistry and Biophysics* 2001;391:165–70. <https://doi.org/10.1006/abbi.2001.2416>.

- [175] Liu Y, Dunn GS, Aronson NN Jr. Purification, biochemistry and molecular cloning of an insect glycosylasparaginase from *Spodoptera frugiperda*. *Glycobiology* 1996;6:527–36. <https://doi.org/10.1093/glycob/6.5.527>.
- [176] Becerra JE, Rodríguez-Díaz J, Gozalbo-Rovira R, Palomino-Schätzlein M, Zúñiga M, Monedero V, *et al.* Unique Microbial Catabolic Pathway for the Human Core N-Glycan Constituent Fucosyl- α -1,6-N-Acetylglucosamine-Asparagine. *mBio* 2020;11:e02804-19. <https://doi.org/10.1128/mBio.02804-19>.
- [177] Xuan J, Tarentino AL, Grimwood BG, Plummer TH, Cui T, Guan C, *et al.* Crystal structure of glycosylasparaginase from *Flavobacterium meningosepticum*. *Protein Sci* 1998;7:774–81. <https://doi.org/10.1002/pro.5560070327>.
- [178] Guan C, Cui T, Rao V, Liao W, Benner J, Lin C-L, *et al.* Activation of Glycosylasparaginase FORMATION OF ACTIVE N-TERMINAL THREONINE BY INTRAMOLECULAR AUTOPROTEOLYSIS. *J Biol Chem* 1996;271:1732–7. <https://doi.org/10.1074/jbc.271.3.1732>.
- [179] Qian X, Guan C, Guo H-C. A dual role for an aspartic acid in glycosylasparaginase autoproteolysis. *Structure* 2003;11:997–1003. [https://doi.org/10.1016/s0969-2126\(03\)00150-3](https://doi.org/10.1016/s0969-2126(03)00150-3).
- [180] Wang Y, Guo H-C. Two-step Dimerization for Autoproteolysis to Activate Glycosylasparaginase *. *Journal of Biological Chemistry* 2003;278:3210–9. <https://doi.org/10.1074/jbc.M210431200>.
- [181] Wang Y, Guo H-C. Crystallographic Snapshot of a Productive Glycosylasparaginase–Substrate Complex. *Journal of Molecular Biology* 2007;366:82–92. <https://doi.org/10.1016/j.jmb.2006.09.051>.
- [182] Guan C, Liu Y, Shao Y, Cui T, Liao W, Ewel A, *et al.* Characterization and Functional Analysis of the Cis-autoproteolysis Active Center of Glycosylasparaginase *. *Journal of Biological Chemistry* 1998;273:9695–702. <https://doi.org/10.1074/jbc.273.16.9695>.
- [183] Noronkoski T, Mononen I. Influence of L-fucose attached α 1 \rightarrow 6 to the asparagine-linked N-acetylglucosamine on the hydrolysis of the N-glycosidic linkage by human glycosylasparaginase. *Glycobiology* 1997;7:217–20. <https://doi.org/10.1093/glycob/7.2.217>.
- [184] Salze M, Giard J-C, Riboulet-Bisson E, Hain T, Rincé A, Muller C. Identification of the general stress stimulon related to colonization in *Enterococcus faecalis*. *Arch Microbiol* 2020;202:233–46. <https://doi.org/10.1007/s00203-019-01735-8>.
- [185] Gao P, Pinkston KL, Bourgogne A, Cruz MR, Garsin DA, Murray BE, *et al.* Library Screen Identifies *Enterococcus faecalis* CcpA, the Catabolite Control Protein A, as an Effector of Ace, a Collagen Adhesion Protein Linked to Virulence. *J Bacteriol* 2013;195:4761–8. <https://doi.org/10.1128/JB.00706-13>.
- [186] Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* 1987;169:5429–33.
- [187] Mojica FJM, Juez G, Rodríguez-Valera F. Transcription at different salinities of *Haloferox mediterranei* sequences adjacent to partially modified PstI sites. *Molecular Microbiology* 1993;9:613–21. <https://doi.org/10.1111/j.1365-2958.1993.tb01721.x>.
- [188] Mojica FJ, Ferrer C, Juez G, Rodríguez-Valera F. Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferox mediterranei* and *Haloferox volcanii* and could be involved in replicon partitioning. *Mol Microbiol* 1995;17:85–93. https://doi.org/10.1111/j.1365-2958.1995.mmi_17010085.x.
- [189] Mojica FJM, Díez-Villaseñor C, Soria E, Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular Microbiology* 2000;36:244–6. <https://doi.org/10.1046/j.1365-2958.2000.01838.x>.
- [190] Adli M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* 2018;9:1911. <https://doi.org/10.1038/s41467-018-04252-2>.
- [191] Jansen R, Embden JDA van, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 2002;43:1565–75. <https://doi.org/10.1046/j.1365-2958.2002.02839.x>.

- [192] Haft DH, Selengut J, Mongodin EF, Nelson KE. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* 2005;1:e60. <https://doi.org/10.1371/journal.pcbi.0010060>.
- [193] Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 2012;337:816–21. <https://doi.org/10.1126/science.1225829>.
- [194] Amitai G, Sorek R. CRISPR–Cas adaptation: insights into the mechanism of action. *Nature Reviews Microbiology* 2016;14:67–76. <https://doi.org/10.1038/nrmicro.2015.14>.
- [195] Wang Y, Wang S, Chen W, Song L, Zhang Y, Shen Z, *et al.* CRISPR-Cas9 and CRISPR-Assisted Cytidine Deaminase Enable Precise and Efficient Genome Editing in *Klebsiella pneumoniae*. *Applied and Environmental Microbiology* 2018;84. <https://doi.org/10.1128/AEM.01834-18>.
- [196] Hille F, Richter H, Wong SP, Bratovič M, Ressel S, Charpentier E. The Biology of CRISPR-Cas: Backward and Forward. *Cell* 2018;172:1239–59. <https://doi.org/10.1016/j.cell.2017.11.032>.
- [197] Barrangou R. Cas9 Targeting and the CRISPR Revolution. *Science* 2014;344:707–8. <https://doi.org/10.1126/science.1252964>.
- [198] Bikard D, Euler CW, Jiang W, Nussenzweig PM, Goldberg GW, Duportet X, *et al.* Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials. *Nature Biotechnology* 2014;32:1146–50. <https://doi.org/10.1038/nbt.3043>.
- [199] Yosef I, Manor M, Kiro R, Qimron U. Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 2015;112:7267–72. <https://doi.org/10.1073/pnas.1500107112>.
- [200] Wu Z-Y, Huang Y-T, Chao W-C, Ho S-P, Cheng J-F, Liu P-Y. Reversal of carbapenem-resistance in *Shewanella algae* by CRISPR/Cas9 genome editing. *Journal of Advanced Research* 2019;18:61–9. <https://doi.org/10.1016/j.jare.2019.01.011>.
- [201] de Maat V, Stege PB, Dedden M, Hamer M, van Pijkeren J-P, Willems RJL, *et al.* CRISPR-Cas9-mediated genome editing in vancomycin-resistant *Enterococcus faecium*. *FEMS Microbiol Lett* 2019;366:fnz256. <https://doi.org/10.1093/femsle/fnz256>.
- [202] Bertani G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* 1951;62:293–300. <https://doi.org/10.1128/jb.62.3.293-300.1951>.
- [203] Leenhouts K, Buist G, Bolhuis A, ten Berge A, Kiel J, Mierau I, *et al.* A general system for generating unlabelled gene replacements in bacterial chromosomes. *Mol Gen Genet* 1996;253:217–24. <https://doi.org/10.1007/s004380050315>.
- [204] Terzaghi BE, Sandine WE. Improved Medium for Lactic Streptococci and Their Bacteriophages. *Applied Microbiology* 1975;29:807–13. <https://doi.org/10.1128/am.29.6.807-813.1975>.
- [205] La Carbona S, Sauvageot N, Giard J-C, Benachour A, Posteraro B, Auffray Y, *et al.* Comparative study of the physiological roles of three peroxidases (NADH peroxidase, Alkyl hydroperoxide reductase and Thiol peroxidase) in oxidative stress response, survival inside macrophages and virulence of *Enterococcus faecalis*. *Mol Microbiol* 2007;66:1148–63. <https://doi.org/10.1111/j.1365-2958.2007.05987.x>.
- [206] Dunny GM, Brown BL, Clewell DB. Induced cell aggregation and mating in *Streptococcus faecalis*: evidence for a bacterial sex pheromone. *Proc Natl Acad Sci U S A* 1978;75:3479–83.
- [207] Arnaud M, Chastanet A, Débarbouillé M. New Vector for Efficient Allelic Replacement in Naturally Nontransformable, Low-GC-Content, Gram-Positive Bacteria. *Appl Environ Microbiol* 2004;70:6887–91. <https://doi.org/10.1128/AEM.70.11.6887-6891.2004>.
- [208] Bryan EM, Bae T, Kleerebezem M, Dunny GM. Improved Vectors for Nisin-Controlled Expression in Gram-Positive Bacteria. *Plasmid* 2000;44:183–90. <https://doi.org/10.1006/plas.2000.1484>.
- [209] Holo H, Nes IF. High-Frequency Transformation, by Electroporation, of *Lactococcus lactis subsp. cremoris* Grown with Glycine in Osmotically Stabilized Media. *Applied and Environmental Microbiology* 1989. <https://doi.org/10.1128/aem.55.12.3119-3123.1989>.
- [210] Walhout AJM, Temple GF, Brasch MA, Hartley JL, Lorson MA, van den Heuvel S, *et al.* [34] GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading

- frames or ORFeomes. In: Thorner J, Emr SD, Abelson JN, editors. *Methods in Enzymology*, vol. 328, Academic Press; 2000, p. 575-IN7. [https://doi.org/10.1016/S0076-6879\(00\)28419-X](https://doi.org/10.1016/S0076-6879(00)28419-X).
- [211] Aranda PS, LaJoie DM, Jorcyk CL. Bleach Gel: A Simple Agarose Gel for Analyzing RNA Quality. *Electrophoresis* 2012;33:366–9. <https://doi.org/10.1002/elps.201100335>.
- [212] Liu Y, Guan C, Aronson NN. Site-directed Mutagenesis of Essential Residues Involved in the Mechanism of Bacterial Glycosylasparaginase. *Journal of Biological Chemistry* 1998;273:9688–94. <https://doi.org/10.1074/jbc.273.16.9688>.
- [213] Andreesen JR. Glycine reductase mechanism. *Current Opinion in Chemical Biology* 2004;8:454–61. <https://doi.org/10.1016/j.cbpa.2004.08.002>.
- [214] Michaux C, Hansen EE, Jenniches L, Gerovac M, Barquist L, Vogel J. Single-Nucleotide RNA Maps for the Two Major Nosocomial Pathogens *Enterococcus faecalis* and *Enterococcus faecium*. *Front Cell Infect Microbiol* 2020;10:600325. <https://doi.org/10.3389/fcimb.2020.600325>.
- [215] You D, Zhang B-Q, Ye B-C. GntR Family Regulator DasR Controls Acetate Assimilation by Directly Repressing the *acsA* Gene in *Saccharopolyspora erythraea*. *J Bacteriol* 2018;200:e00685-17. <https://doi.org/10.1128/JB.00685-17>.
- [216] Zhang M, Kang Z, Guo X, Guo S, Xiao D, Liu Y, et al. Regulation of Glutarate Catabolism by GntR Family Regulator CsiR and LysR Family Regulator GcdR in *Pseudomonas putida* KT2440. *mBio* 2019;10:e01570-19. <https://doi.org/10.1128/mBio.01570-19>.
- [217] Gaugué I, Oberto J, Putzer H, Plumbridge J. The use of amino sugars by *Bacillus subtilis*: presence of a unique operon for the catabolism of glucosamine. *PLoS One* 2013;8:e63025. <https://doi.org/10.1371/journal.pone.0063025>.
- [218] Matilla MA, Velando F, Martín-Mora D, Monteagudo-Cascales E, Krell T. A catalogue of signal molecules that interact with sensor kinases, chemoreceptors and transcriptional regulators. *FEMS Microbiology Reviews* 2022;46:fuab043. <https://doi.org/10.1093/femsre/fuab043>.
- [219] Alice AF, Pérez-Martínez G, Sánchez-Rivas C. Phosphoenolpyruvate phosphotransferase system and N-acetylglucosamine metabolism in *Bacillus sphaericus*. *Microbiology (Reading)* 2003;149:1687–98. <https://doi.org/10.1099/mic.0.26231-0>.
- [220] Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;30:1236–40. <https://doi.org/10.1093/bioinformatics/btu031>.
- [221] Ramarao N, Nielsen-Leroux C, Lereclus D. The insect *Galleria mellonella* as a powerful infection model to investigate bacterial pathogenesis. *J Vis Exp* 2012:e4392. <https://doi.org/10.3791/4392>.
- [222] de Freitas Junior JCM, Morgado-Díaz JA. The role of N-glycans in colorectal cancer progression: potential biomarkers and therapeutic applications. *Oncotarget* 2015;7:19395–413. <https://doi.org/10.18632/oncotarget.6283>.
- [223] He Y, Hao F, Fu H, Tian G, Zhang Y, Fu K, et al. N-glycosylated intestinal protein BCF-1 shapes microbial colonization by binding bacteria via its fimbrial protein. *Cell Rep* 2023;42:111993. <https://doi.org/10.1016/j.celrep.2023.111993>.
- [224] Pierce A, Legrand D, Mazurier J. La lactoferrine : une protéine multifonctionnelle. *Med Sci (Paris)* 2009;25:361–9. <https://doi.org/10.1051/medsci/2009254361>.
- [225] Michineau S, Muller L, Pizard A, Alhenc-Gélas F, Rajerison RM. N-linked glycosylation of the human bradykinin B2 receptor is required for optimal cell-surface expression and coupling. *Biol Chem* 2004;385:49–57. <https://doi.org/10.1515/BC.2004.007>.
- [226] Hennet T. Collagen glycosylation. *Current Opinion in Structural Biology* 2019;56:131–8. <https://doi.org/10.1016/j.sbi.2019.01.015>.
- [227] Mäkinen PL, Clewell DB, An F, Mäkinen KK. Purification and Substrate Specificity of a Strongly Hydrophobic Extracellular Metalloendopeptidase (“Gelatinase”) from *Streptococcus faecalis* (Strain OG1-10). *Journal of Biological Chemistry* 1989;264:3325–34. [https://doi.org/10.1016/S0021-9258\(18\)94069-X](https://doi.org/10.1016/S0021-9258(18)94069-X).
- [228] Smid EJ, Plapp R, Konings WN. Peptide uptake is essential for growth of *Lactococcus lactis* on the milk protein casein. *J Bacteriol* 1989;171:6135–40.

- [229] Mori S, Miyamoto M, Kaneko S, Nirasawa S, Komba S, Kasumi T. Characterization and kinetic analysis of enzyme-substrate recognition by three recombinant lactococcal PepVs. *Archives of Biochemistry and Biophysics* 2006;454:137–45. <https://doi.org/10.1016/j.abb.2006.08.001>.
- [230] 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. *J Biol Chem* 1994;269:11391–9.
- [231] Coccagn-Bousquet M, Garrigues C, Novak L, Lindley N d., Loublere P. Rational development of a simple synthetic medium for the sustained growth of *Lactococcus lactis*. *Journal of Applied Bacteriology* 1995;79:108–16. <https://doi.org/10.1111/j.1365-2672.1995.tb03131.x>.
- [232] Veith N, Solheim M, van Grinsven KWA, Olivier BG, Levering J, Grosseholz R, *et al.* Using a Genome-Scale Metabolic Model of *Enterococcus faecalis* V583 To Assess Amino Acid Uptake and Its Impact on Central Metabolism. *Appl Environ Microbiol* 2015;81:1622–33. <https://doi.org/10.1128/AEM.03279-14>.
- [233] Hellendoorn MA, Franke-Fayard BM, Mierau I, Venema G, Kok J. Cloning and analysis of the pepV dipeptidase gene of *Lactococcus lactis* MG1363. *J Bacteriol* 1997;179:3410–5. <https://doi.org/10.1128/jb.179.11.3410-3415.1997>.
- [234] Vongerichten KF, Klein JR, Matern H, Plapp R. Cloning and nucleotide sequence analysis of pepV, a carnosinase gene from *Lactobacillus delbrueckii subsp. lactis* DSM 7290, and partial characterization of the enzyme. *Microbiology (Reading)* 1994;140 (Pt 10):2591–600. <https://doi.org/10.1099/00221287-140-10-2591>.
- [235] Cao Y, Jin X, Levin EJ, Huang H, Zong Y, Quick M, *et al.* Crystal structure of a phosphorylation-coupled saccharide transporter. *Nature* 2011;473:50–4. <https://doi.org/10.1038/nature09939>.
- [236] Naren N, Zhang X-X. Global Regulatory Roles of the Histidine-Responsive Transcriptional Repressor HutC in *Pseudomonas fluorescens* SBW25. *J Bacteriol* 2020;202:e00792-19. <https://doi.org/10.1128/JB.00792-19>.
- [237] Bandera AM, Bartho J, Lammens K, Drexler DJ, Kleinschwärzer J, Hopfner K-P, *et al.* BusR senses bipartite DNA binding motifs by a unique molecular ruler architecture. *Nucleic Acids Res* 2021;49:10166–77. <https://doi.org/10.1093/nar/gkab736>.
- [238] Zeng L, Burne RA. NagR Differentially Regulates the Expression of the *glmS* and *nagAB* Genes Required for Amino Sugar Metabolism by *Streptococcus mutans*. *J Bacteriol* 2015;197:3533–44. <https://doi.org/10.1128/JB.00606-15>.
- [239] Mokhtari A, Blancato VS, Repizo GD, Henry C, Píkis A, Bourand A, *et al.* *Enterococcus faecalis* utilizes maltose by connecting two incompatible metabolic routes via a novel maltose 6'-phosphate phosphatase (MapP). *Mol Microbiol* 2013;88:234–53. <https://doi.org/10.1111/mmi.12183>.
- [240] Thompson J, Lichtenthaler FW, Peters S, Píkis A. Beta-glucoside kinase (BglK) from *Klebsiella pneumoniae*. Purification, properties, and preparative synthesis of 6-phospho-beta-D-glucosides. *J Biol Chem* 2002;277:34310–21. <https://doi.org/10.1074/jbc.M206397200>.
- [241] MAKINO M, KOJIMA T, OHGUSHI T, YAMASHINA I. Studies on Enzymes Acting on Glycopeptides*. *The Journal of Biochemistry* 1968;63:186–92. <https://doi.org/10.1093/oxfordjournals.jbchem.a128760>.
- [242] Singkum P, Suwanmanee S, Pumeesat P, Luplertlop N. A powerful in vivo alternative model in scientific research: *Galleria mellonella*. *Acta Microbiologica et Immunologica Hungarica* 2019;66:31–55. <https://doi.org/10.1556/030.66.2019.001>.
- [243] Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. *Nat Methods* 2013;10:957–63. <https://doi.org/10.1038/nmeth.2649>.
- [244] Barrangou R. The roles of CRISPR-Cas systems in adaptive immunity and beyond. *Curr Opin Immunol* 2015;32:36–41. <https://doi.org/10.1016/j.coi.2014.12.008>.
- [245] Goren M, Yosef I, Qimron U. Sensitizing pathogens to antibiotics using the CRISPR-Cas system. *Drug Resistance Updates* 2017;30:1–6. <https://doi.org/10.1016/j.drug.2016.11.001>.

- [246] Tian P, Wang J, Shen X, Rey JF, Yuan Q, Yan Y. Fundamental CRISPR-Cas9 tools and current applications in microbial systems. *Synthetic and Systems Biotechnology* 2017;2:219–25. <https://doi.org/10.1016/j.synbio.2017.08.006>.
- [247] Asmamaw M, Zawdie B. Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing. *Biologics* 2021;15:353–61. <https://doi.org/10.2147/BTT.S326422>.
- [248] Bondy-Denomy J, Garcia B, Strum S, Du M, Rollins MF, Hidalgo-Reyes Y, *et al.* Multiple mechanisms for CRISPR–Cas inhibition by anti–CRISPR proteins. *Nature* 2015;526:136–9. <https://doi.org/10.1038/nature15254>.
- [249] Mahendra C, Christie KA, Osuna BA, Pinilla-Redondo R, Kleinstiver BP, Bondy-Denomy J. Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer. *Nat Microbiol* 2020;5:620–9. <https://doi.org/10.1038/s41564-020-0692-2>.
- [250] Guo C, Ma X, Gao F, Guo Y. Off-target effects in CRISPR/Cas9 gene editing. *Front Bioeng Biotechnol* 2023;11:1143157. <https://doi.org/10.3389/fbioe.2023.1143157>.
- [251] Rodrigues M, McBride SW, Hullahalli K, Palmer KL, Duerkop BA. Conjugative Delivery of CRISPR-Cas9 for the Selective Depletion of Antibiotic-Resistant Enterococci. *Antimicrobial Agents and Chemotherapy* 2019;63. <https://doi.org/10.1128/AAC.01454-19>.

Annexes

1 Tableaux supplémentaires

Tableau S1 : Composition du milieu M17.

Composés	Concentration
Tryptone USP	5 g/L
Peptone de Soja	5 g/L
Extrait de viande	5 g/L
Extrait de levure	2,5 g/L
B-glycérophosphate de sodium	19 g/L
Acide L-ascorbique	0,5 g/L
Sulfate de magnésium (heptahydrate)	0,25 g/L

Tableau S2 : Composition du milieu cdM17.

Composés	Concentration
Tryptone USP	5 g/L
Peptone de Soja	5 g/L
Extrait de viande	5 g/L
Extrait de levure	2,5 g/L
MOPS	42 g/L
Acide L-ascorbique	0,5 g/L
Sulfate de magnésium (heptahydrate)	0,25 g/L

Tableau S3 : Composition du milieu LB.

Composés	Concentration
Tryptone USP	5 g/L
Extrait de levure	5 g/L
Chlorure de sodium	5 g/L
pH 7,1	

Tableau S4 : Oligonucléotides utilisés pour la construction des mutants et des complémentations.

Nom	Séquence 5' → 3'	Fonction
CV_egaR(1)_for	CTGCAACTTTTTCTGTTCCGAA	Construction du mutant de délétion <i>egaR</i>
CV_egaR(1)_rev	TTGGTTGTTTTTCAGACGGGCTTACATACAGACCGTCTCGAAT	
CV_egaR(2)_for	ATTCGAGACGGTCTGTATGTAAGCCCGTCTGAAAACAACCAA	
CV_egaR(2)_rev	ACCCCAAGTCATTTTCCATCC	
CV_39	TCTAGCTAATGTTACGTTACACATTAAGTAGAGCTTTCGGCGAACCGGCTTT	
CV_40	AGAATCATAATGGGGAAAGGCCATCCAGCCTACAATCCCTCGCTCCCAT	
CV_93	TTAAGTGCGCAAGCAACCGAT	Construction du mutant de délétion <i>egaP</i>
CV_94	TGCAATGCCTTTTGACCAGGAAACCTTTTCACTTTCGAATCGC	
CV_95	GCGATTGCAAGTGTGAAAGGGTTTCTGGTCAAAAAGGCATTGCA	
CV_96	AGGAATCCCTTGAGGAGCAAATTG	
CV_117	TCTAGCTAATGTTACGTTACACATTAAGTAAAGTTCGAAAGAACTCTGG	

CV_118	AGAATCATAATGGGGAAGGCCATCCAGCCAGAAAGACAGCCTTTCATTAA	
ef1158_for	AGGTCCTATCTTGGCTTGCC	Construction du mutant de délétion <i>egaG</i>
ef1158_mut_rev	TCTCTGCTAGCTTTTAGTAACCCATCATG	
ef1158_mut_for_bis	GCGTATCTAGAAATGTGGAATTTACTTTTAC	
ef1158_rev	GACTTAACCACTGTTGGTGGC	
pMAD_1158_for	AAAACGGATCCGTTTTTGGATTACTCCAG	
pMAD_1158_rev	AACGAGTCGACATTAATGCGACAATCATT	
CV_67	GTGTTATCATTGGCCGTGCG	Construction du mutant de délétion <i>egaB</i>
CV_68	TGCTTAATTAACCGCAGCGCACTCATGTTCCCTCACTCTCCTC	
CV_69	GAGGAGAGTGAAGGAACATGAGTGCCTGCGGTTAATTAAGCA	
CV_70	AAAGAACCGCATCGGCATGA	
CV_119	TCTAGCTAATGTTACGTTACACATTAACCTCACGTATGAGAAGAAGCCAAT	
CV_120	AGAATCATAATGGGGAAGGCCATCCAGCCAGACAATGCCCGCAAACAGG	
Spacer1160F	AAACTGAATTTCTACAAACGCCCTTATTGAGTCTG	Construction du mutant de délétion <i>egaC</i>
Spacer1160R	AAAACAGACTCAATAAGGGCGTTTGTAGAAATCA	
CV_43F	TTAACGAAAGGGCGGATGGA	
CV_44R	ACTCTCTTCTTGTGCTTTCCAAGACATCATTGTCGCCAGT	
CV_45F	ACTGGGCGACAATGATGCTTGGAAAGCACAAGAAGAAGAGAGT	
CV_46R	TTGGTGCTTTTGTGCATACCG	
CV_47F	GTGAGTGGAATTCCTGCAGCCCATACACAGAGATTCAGCCG	
CV_48R	TCTAGAACTAGTGGATCCCCCTGCTTTTGCACGACCAATTG	
CV_49F	CCTCAAGGGATTCTGTGCG	
CV_50R	TCCGCTCCTTGTCTGTAA	
CV_73F	GGCGAAAAGCATTGAAAGAAGC	
CV_74R	ATTTGCTTTTGCACGACCAATTCATCACTGCCAAATCCTGC	
CV_75F	GCAGGATTTGGCAGTGATGAATGGTCTGGCAAAGCAAAT	
CV_76R	ATCCGTGTCTTTTCTTACAG	
CV_121F	TCTAGCTAATGTTACGTTACACATTAACCTGGGTGGACAATAAAAAATGCCA	
CV_122R	AGAATCATAATGGGGAAGGCCATCCAGCCGTCAGCAAACGGGCAATGTT	
CV117	CAAAATAAATTATAAGGAGGCACTCAAATGACAGAAAATGCACCCAAAT	Complémentation du mutant Δ <i>egaR</i>
CV118	GAAGCGCTGTCGACGAGCTCGCATGCGAATCATTGTTCCAAGAACCCTCC	
CV119	CAAAATAAATTATAAGGAGGCACTCAAATGAAGATCAAAGAAGAAATAG	Complémentation du mutant Δ <i>egaP</i>
CV120	GAAGCGCTGTCGACGAGCTCGCATGCGAATTATGCTTCTGTTAATGCAAT	
CV123	CAAAATAAATTATAAGGAGGCACTCAAATGAGTAAAAACAAATTTATC	Complémentation du mutant Δ <i>egaB</i>
CV124	GAAGCGCTGTCGACGAGCTCGCATGCGAATTATTTTTATTTTGGCTTAAT	
CV125	CAAAATAAATTATAAGGAGGCACTCAAAGTGCAAAAATTTTGGATTGGT	Complémentation du mutant Δ <i>egaC</i>
CV126	GAAGCGCTGTCGACGAGCTCGCATGCGAATTATAAATCATCGAAGTCAAA	
CV_127	CAAAATAAATTATAAGGAGGCACTCAAATGAAAAAATCCTTTTAGTCT	Complémentation du mutant Δ <i>egaD1D2</i>
CV_128	GAAGCGCTGTCGACGAGCTCGCATGCGAATTAAAAACAACCTGTTTCTTC	
P3535 3'	CTCGAGTCTAGAAGCGCTGTC	Vérification de l'insert dans pMSP3535
P3535 5'	TTAGATACAATGATTTGTTCCG	
pU	TGTAACGACGGCCAGT	Vérification de l'insert dans pMAD
pR	CAGGAAACAGCTATGACC	

Tableau S5 : Oligonucléotides utilisés pour la construction des plasmides de surexpression.

Nom	Séquence 5' → 3'	Fonction
CV_62	TTGTGTTAACTTTAAGAAGGAGATATACATATGACAGAAAATGCACCCAA	Insertion du gène <i>egaR</i> dans pET29b+
CV_63	AGCCGGATCTCAGTGGTGGTGGTGGTGGTTCGTTCCAAGAACCTCCTA	
OCG_232	CACCACCACCACCACCAC	
OCG_233	CATATGTATATCTCCTTCTTAAAGTTAAACAAA	
OCG_240	CGCGAAATTAATACGACTCACTATA	Vérification de l'insertion d'insert dans pET29b+
OCG_241	GGGGTTATGCTAGTTATTGCTCA	
pQE_1157for	TCTTGGGATCCATGAAGATCAAAGAAGAAATAGC	Insertion d' <i>egaP</i> dans pQE30
pQE_1157rev	TTCTAAAGCTTTTATGCTTCTGTTAATGCAATC	
pQE_1158for	GCGGAGGATCCATGACTTGGGGTGCAATTGC	Insertion d' <i>egaG</i> dans pQE30
pQE_1158rev	ACTCTAAGCTTTTATTCAATTGGTGCTTTGATTC	
pQE_TypeIII/IV	CGGATAACAATTTACACAG	Vérification du clonage dans pQE30
pQE_rev	GTTCTGAGGTCATTA CTGG	

Tableau S6 : Oligonucléotides utilisés pour les expériences de RT-qPCR.

Nom	Séquence 5' → 3'	Gène
RT_CV_1	AAGTGGTTGAAGCAGAGAAG	<i>egaB</i>
RT_CV_2	GGAATCCCTTGAGGAGCAAA	
RT_CV_3	AAGGAGTGCCACCAACAGTGG	<i>egaC</i>
RT_CV_4	AAGGAGTTAAAGCAGCCAGCCA	
RT_CV_9	GCAAGGAGCGGAAATTGTTGC	<i>egaD1</i>
RT_CV_10	CCGCATCGGCATGAAACTCCAA	
RT_CV_7	GGCTATGCGCAATGGGGAACA	<i>egaP</i>
RT_CV_8	TGAAAGGGCGGCACTGACCAA	
RT_CV_11	CCCCTGAGAAAATTCGGGA	<i>egaR</i>
RT_CV_12	CAACTGTCATTGTTTCGCCG	
EF1158_RTF	AGAACGCGCTAATCAATGGT	<i>egaG</i>
EF1158_RTR	AACGTAATAGCGCCGACAGT	

Tableau S7 : Oligonucléotides utilisés pour l'expérience de 5'RACE-PCR.

Nom	Séquence 5' → 3'	Fonction
1158_race1	GCACCTCCTGTTGTAATTC	Utilisé pour le séquençage
1158_race2	TTAGGTAAACCGCCGTAGCC	Utilisé pour amplifier l'ADNc-queue polyA et la Nested-PCR
1158_race3	TTGCAAATGTGTCTCCATCC	Utilisé lors de la synthèse d'ADNc

Tableau S8 : *Autres oligonucléotides.*

Nom	Séquence 5' → 3'	Fonction
CV_56	TGAAGCCTGTCCGCTACTCT	Clonage de P1 dans pGEM-T
CV_57	GTGCATTTTCTGTCATTTTCATCGC	
CV_58	TGCAAGCAATGATTGCATTAAC	
CV_59	TTTAGTAACCCATCATGTGCC	Clonage de P2 dans pGEM-T
pU	TGTAAAACGACGGCCAGT	
pR	CAGGAAACAGCTATGACC	Vérification de l'insert dans pGEM-T

2 Publication scientifique : « Utilization of glycoprotein-derived N-acetylglucosamine-L-asparagine during *Enterococcus faecalis* infection depends on catabolic and transport enzymes of the glycosylasparaginase locus »

Utilization of glycoprotein-derived N-acetylglucosamine-L-asparagine during *Enterococcus faecalis* infection depends on catabolic and transport enzymes of the glycosylasparaginase locus

Victor Combret¹, Isabelle Rincé¹, Aurélie Budin-Verneuil¹, Cécile Muller¹, Josef Deutscher^{2,3}, Axel Hartke¹, Nicolas Verneuil¹ and Nicolas Sauvageot^{1*}

¹Normandie Université, UNICAEN, CBSA, F-14000 Caen, France.

²Université Paris Saclay, INRAE, Micalis Institute, 78350 Jouy en Josas, France.

³CNRS, Institut de Biologie Physico-Chimique UMR8261, Expression Génétique Microbienne, 75005 Paris, France.

* Corresponding Author:

Email: nicolas.sauvageot@unicaen.fr

Tel: (+33)2 31 56 56 61

Fax: (+33)2 31 56 53 11

Abstract

Enterococcus faecalis is a Gram-positive clinical pathogen causing severe infections. Its survival during infection depends on its ability to utilize host-derived metabolites, such as protein-deglycosylation products. We have identified in *E. faecalis* OG1RF a locus (*ega*) involved in the catabolism of the glycoamino acid N-acetylglucosamine-L-asparagine. This locus is separated into two transcription units, genes *egaRP* and *egaGBCD1D2*, respectively. RT-qPCR experiments revealed that the expression of the *ega* locus is regulated by the transcriptional repressor EgaR. Electromobility shift assays evidenced that N-acetylglucosamine-L-asparagine interacts directly with the EgaR protein, which leads to the transcription of the *ega* genes. Growth studies with *egaG*, *egaB* and *egaC* mutants confirmed that the encoded proteins are necessary for N-acetylglucosamine-L-asparagine catabolism. This glycoamino acid is transported and phosphorylated by a specific phosphotransferase system EIIABC components (OG1RF_10751, EgaB, EgaC) and subsequently hydrolyzed by the glycosylasparaginase EgaG, which generates aspartate and 6-P-N-acetyl- β -D-glucosaminylamine. The latter can be used as a fermentable carbon source by *E. faecalis*. Moreover, *Galleria mellonella* larvae had a significantly higher survival rate when infected with *ega* mutants compared to the wild-type strain, suggesting that the loss of N-acetylglucosamine-L-asparagine utilization affects enterococcal virulence.

Key Words: *Enterococcus faecalis*; N-acetylglucosamine-L-asparagine; glycosylasparaginase; phosphotransferase system (PTS); metabolism; *ega*

Introduction

Enterococci are Gram positive, catalase positive, facultative anaerobic, non-spore-forming opportunistic pathogens belonging to the lactic acid bacteria. These bacteria are commensal microorganisms inhabiting the gastro-intestinal tract of humans and animals, but can be also found in natural environments or fermented food products [1–3]. In the past 50 years, especially two species, *i.e.* *Enterococcus faecalis* and *Enterococcus faecium*, emerged as leading causes of hospital-acquired infections [4,5] such as endocarditis, meningitis, septicemia, pneumonia or urinary tract infections [6,7]. Moreover, enterococci can persist in hospital environments due to their high antibiotic resistance, tolerance to desiccation, pH and temperature [2,6,8–10]. Numerous virulence factors were identified and their roles in pathogenicity defined [2,3,8]. However, little is known about the acquisition of carbon sources and their metabolism, which is important for the first step of infection that is the colonization of the host. Various studies highlighted that utilization of glycans represents an important factor for host colonization for several bacteria such as *Listeria monocytogenes*, *Escherichia coli*, *Streptococcus thermophilus*, *E. faecalis*, *Streptococcus pyogenes* or *Edwardsiella piscicida* [11–17]. An RNA-Seq study with *E. faecalis* V583 demonstrated that expression of a significant number of genes was induced during intraperitoneal infection in a mouse model [18]. Among them, 19% are linked to the category “metabolism”, especially genes is involved in the catabolism of β -glucosides, suggesting that the bacterium adapts to the conditions of the host environment. It was also shown that when *E. faecalis* was grown on the 1,6- β -glucoside gentiobiose, numerous genes involved in the transport and degradation of this disaccharide were induced [19]. This indicates that β -glucoside utilization is important in *E. faecalis*. Some of these genes encode for proteins with homologies to β -glucosidases degrading a wide range of polysaccharides liberated by the breakdown of glycoproteins. Proteins in eukaryotic cells are generally glycosylated including those involved in adaptive and innate immunity. Two types of glycosylation are observed, the O-glycosylation which corresponds to the addition of sugars to serine/threonine residues of proteins, and N-glycosylation where the glycan is linked to the NH_2 of the carboxylamide of an asparagine (Asn) residue via N-acetylglucosamine (GlcNAc). Although it was originally thought that glycoproteins were exclusively present in eukaryotic cells, we now know that they are also present in archaea and bacteria [20]. Until today, no N-glycosylation has been found in Gram-positive bacteria. Although N-glycosylations are quite different in yeast, insects and mammals, the core group is similar in all eucaryotes and corresponds usually to a trimannosylchitobiosyl (Man3-GlcNAc-GlcNAc) residue attached to the asparagine of a polypeptide. Bacterial enzymes able to cleave the N-glycan chains of glycosylated proteins

seem to be important for virulence. For example, the secreted *E. faecalis* endo- β -N acetylglucosaminidase (EndoE) removes the glycans from lactoferrin, RNaseB and IgG which, for the latter impedes its ability to recruit the complement system [21–23]. This has a double advantage for the bacterium: weakening of the host's immune system and mobilizing sugars for its carbon and energy supply allowing them to survive and colonize the host. Besides this activity, two other enzymes possess enzymatic activities able to disconnect the N-glycan by cleaving the N-glycosidic linkage between L-asparagine and the GlcNAc. The first enzyme, the peptide N-glycanase (PNGase), acts on glycan-GlcNAc- β -1-N-Asn-peptide and releases 1-amino-glycan and aspartate-peptide [24] (Fig. 1). The second one corresponds to the N(4)-(β -N-acetylglucosaminy)-L-asparaginase or glycosylasparaginase (GA), an amidohydrolase which cleaves the same linkage as PNGase but is only active on the degradation product GlcNAc- β -1-N-Asn resulting from proteolysis of the polypeptide [25–27]. To date, the only bacterial GAs characterized are those of *Lactocaseibacillus casei* BL23 and *Elizabethkingia meningoseptica* [26,28–32]. GA is a zymogen belonging to the N-terminal nucleophile (Ntn) hydrolases. This proenzyme undergoes a process of autoproteolysis generating the α and β subunits of 18 kDa and 16 kDa, respectively. The two subunits are held together by strong non-covalent forces [26], and it has been shown that residue Thr152 is important for autoproteolysis [33]. Furthermore, crystallographic studies suggested that His150 and Asp151 are also needed in this process [32]. In addition to its natural substrate N4-(β N-acetylglucosaminy)-L-asparagine, GAs cleave several other substrates as long as the asparagine has free α -amino and α -carboxyl groups [26,27,30]. Whereas the GA of *Lc. casei* (AsnA2) is active on substrates with fucosylated GlcNAc residues, no information is available regarding the *E. meningoseptica*'s GA [28]. Analysis of the genome sequence of *E. faecalis* revealed the presence of a gene showing high sequence homologies with GAs. The composition and regulation of this locus has been studied as well as its regulation. Also, glycosylasparaginase assay was performed. Finally, since endo- β -N-acetylglucosaminidase and PNGase have been linked to virulence, we decided to test whether GA of *E. faecalis* might be also important for pathogenicity [23,24].

Experimental procedures

Bacterial strains and growth conditions

E. faecalis OG1RF [34] wild-type and mutant strains were grown routinely at 37°C without shaking in M17 medium containing 0.5% glucose (wt/vol). Growth studies were carried out in

carbon-depleted M17MOPS (M17cc) medium, which has previously been described [35]. For growth experiments, M17cc was supplemented with 0.3% glucose (wt/vol) or 0.3% 2-acetamido-2-deoxy- β -D-glucopyranosyl L-asparagine (wt/vol). The O.D._{600nm} was measured on an Infinite M Nano multiplate reader (Tecan). When needed, antibiotics were added to the medium: erythromycin (50 μ g per mL), chloramphenicol (25 μ g per mL). *E. faecalis* strains used in this study are described in Table S1.

E. coli BL21(DE3) (New England Biolabs) and TOP10 strains (ThermoFisher Scientific) used for protein purifications and cloning experiments were grown at 37°C with shaking at 160 rpm. Growth was followed by measuring the O.D._{600nm} in a BioPhotometer (Eppendorf). Chloramphenicol, ampicillin, spectinomycin and kanamycin were added when needed at concentrations of 10, 100, 100 and 25 μ g per mL, respectively. *E. coli* strains used in this study are described in Table S1.

General molecular methods

Restriction endonucleases, shrimp alkaline phosphatase and T4 DNA ligase were purchased from Promega and used according to the manufacturer's instructions. PCR experiments were realized with GoTaq (Promega) or Q5 polymerase (New England Biolabs) in a thermocycler Biometra TAdvanced (Analytik Jena). If needed, PCR products were purified using the QIAquick® PCR purification kit (Qiagen). Plasmid extractions were realized by using the QIAprep Spin Miniprep Kit (Qiagen). Plasmids used in this study are described in Table S2. Transformations of *E. coli* cells were performed by heat shock and transformations of *E. faecalis* by electroporation with the Gene Pulser Apparatus (Bio-Rad Laboratories) [36,37]. 5'-RACE-PCR was performed with the "5'/3'-RACE Kit, 2nd Generation" (Roche®) according to the manufacturer's recommendations. Primers used for PCR, constructions of mutants or proteins overexpression are listed in Table S3, and those used for the 5'-RACE-PCR are listed in Table S4.

Construction of deleted strains, complemented strains and whole genome sequencing

E. faecalis OG1RF mutants were obtained by the deletion of the central part of genes. For each gene, two DNA fragments of approximately 1000 bp flanking the region to be deleted were amplified by using chromosomal DNA as template. A third PCR was carried out by using the two PCR products as templates and the two distal primers to generate an amplicon lacking the region aimed to be deleted. These amplicons were cloned into the pMAD vector [38]. Plasmids

were transformed into *E. faecalis* OG1RF in order to replace the wild-type gene with the partially deleted gene by double crossing over.

In order to complement mutants, the entire corresponding gene was amplified by PCR and cloned behind the nisine inducible promoter of pMSP3535 [39]. The resulting plasmids were introduced into the corresponding mutant strains. As control, mutated strains were transformed with empty pMSP3535.

Chromosomal DNA of OG1RF and OG1RF Δ *egaG* were extracted by using the Nucleospin Microbial DNA (Macherey-Nagel) according the manufacturer recommendations. DNA quantities was quantified using NanoDrop One Spectrophotometer (Thermofisher Scientific) and DNA integrity was checked by electrophoresis. Illumina sequencing was performed by Eurofins. Sequence analysis was performed using “CLC Genomics Workbench” software (version 12.0.2) with the variant detection tool by comparing the genome of Δ *egaG* to the wild type strain. All programs were used with default parameters. The sequencing resulted in an output containing 14538084 and 13928060 paired-end reads sets for the WT and Δ *egaG* strains respectively. Differences between WT and Δ *egaG* are listed in Table S5.

RNA extraction and RT-qPCR

Total RNA was extracted from a 2 mL bacterial culture by using the Direct-zolTM RNA Miniprep (Zymo Research). Isolated RNA purity and concentration were measured with a NanoDrop 2000 Spectrophotometer (Thermofisher Scientific). 2 μ g were used for reverse transcription with the QuantiTect[®] Reverse Transcription kit (Qiagen). qPCR was realized in a 96 wells microplate by using the GoTaq[®] qPCR Master Mix (Promega). Quantitative real-time PCR (RT-qPCR) was realized with SYBR Green fluorescence using a CFX96 real-time PCR machine (Bio Rad). Primers used for the RT-qPCR are listed in Table S4. The 5S gene was used as reference.

Purification of His-tagged E. faecalis proteins

The pET29b+ vector was amplified by PCR with primers OCG233 and OCG232 and the amplicon was purified and digested with restriction endonuclease *DpnI* for 2h. The gene *egaR* was amplified from *E. faecalis* OG1RF with the primers CV_62 and CV_63 (Table S3). These primers possess a complementary region with primers OCG233 and OCG232. The digested plasmid and the *egaR* PCR product were introduced and ligated in the *E. coli* TOP10 strain. The pET29b+::*egaR* vector was used to transform the *E. coli* BL21(DE3) strain for EgaR

overexpression. The gene *egaG* was amplified from *E. faecalis* OG1RF with the primers pQE_1158for and pQE_1158rev (Table S3). The PCR product and pQE30 vector were digested with restriction endonucleases *Bam*HI and *Hind*III and then ligated. The pQE30::*egaG* was introduced in the *E. coli* M15(pRep4) strain.

EgaR and EgaG production was carried out on Terrific Broth with 0.5mM of isopropyl- β -D-1-thiogalactopyranoside (IPTG) for 2 h at 37°C under vigorous agitation. Cells were washed with 50 mM sodium phosphate buffer pH 7.6 and treated for 30 min with lysozyme 1mg/mL. Lysis was performed with FastPrep-24TM5G (mpbio). After centrifugation, purification was carried out with AKTA under non denaturing conditions. The protein concentration was determined using the PierceTM BCA Protein Assay Kit (Thermofisher Scientific) with a BSA standard curve.

Electromobility shift assays

DNA fragments of 173 pb and 169 pb corresponding to the promoter region upstream from *egaR* and *egaG*, respectively, were amplified with primers pU (labelled with Cy5 fluorophore) and pR and then purified with the QIAquick® PCR purification kit (Qiagen). Purified His-EgaR protein (500 nM - 50 nM) was added to 10 μ L of 2x dilution buffer (40 mM Tris-HCl, pH 7.4; 40 mM KCl; 20 mM MgCl₂; 2 mM DTT; 0.2 mM EDTA; 10% glycerol; 0.4 μ g/mL herring sperm), 5 nM DNA. The solution was incubated for 30 min at 37°C. Samples were loaded on 8% non denaturing polyacrylamide gels 0.5X Tris-borate/EDTA buffer, pH 7.4. Electrophoresis was run for 1h. DNA fragments were visualized by fluorescence using an Amersham Typhoon Biomolecular Imager. A 204 pb internal fragment of the *gusA* gene was used as negative control.

Glycosylasparaginase assay

Glycosylasparaginase activity was determined using a three-step coupled enzymatic method [33]. First, hydrolysis of GlcNAc-L-Asn by the glycosylasparaginase releases 1-amino-GlcNAc and aspartate. Subsequently, transamination of aspartate by glutamate-oxaloacetate transaminase (GOT) in the presence of α -ketoglutarate generates glutamate and oxaloacetate. Finally, malate dehydrogenase (MDH) converts oxaloacetate and NADH into malate and NAD⁺. The GOT and MDH enzymes are in excess with respect of the GA, thus making this latter the reaction's limiting element. Consumption of NADH was followed at O.D._{340nm} with a TECAN microplate reader. Reactions were performed in 150 μ L medium containing 20 mM

sodium phosphate, pH 7.5, 0.1 mM α -ketoglutarate, 0.2 mg of NADH, 7.5 U of GOT and MDH and various amounts of glycosylasparaginase and GlcNAc-L-Asn.

The determination of the specific activity was obtained as follow. The kinetic NADH consumption was calculated during the maximal NADH oxidation (curve's slope). Then, this value was divided by the protein amount (mg protein) in order to obtain the specific activity in $\mu\text{mol mg protein}^{-1} \text{ min}^{-1}$.

Infection of Galleria mellonella

Wild-type, mutants, and complemented strains were cultured as previously described in GM17 broth at 37°C, harvested at mid-log phase and washed twice with saline solution (0.9% NaCl). Cultures were adjusted to inject around 9.10^5 CFU/larvae into the last proleg of 7 weeks old *G. mellonella* larvae using a syringe pump (Thermofisher Scientific) and a 26G needle (Terumo). Five larvae were hosted per Petri dish and incubated at 37°C. Alive and dead larvae were enumerated every 2 h for a 24 h post-infection period. Three replicates were performed for a total of at least 60 animals per condition. Data are represented on Kaplan-Meier curves.

Bioinformatic analysis

Statistical analyzes were carried out with GraphPad Prism 8.0.1.

Results and Discussion

Identification of the E. faecalis OG1RF glycosylasparaginase gene locus

The analysis of the genome sequence of *E. faecalis* OG1RF revealed the presence of a gene (OG1RF_10935) annotated to encode an N4-(β -N-acetylglucosaminy1)-L-asparaginase. The gene seems to be functionally linked to several other genes in this locus (Fig. 2). Therefore, we propose to name these genes *ega* for *enterococcal glycosylasparaginase locus* and its coding gene OG1RF_10935 *egaG*. Two genes are present upstream from *egaG*: *egaR* (OG1RF_10933) and *egaP* (OG1RF_10934) coding for proteins annotated to encode a GntR family regulator (transcriptional factor) and a Mn^{2+} -dependent dipeptidase, respectively.

The genes downstream of *egaG*, named *egaB* (OG1RF_10936) and *egaC* (OG1RF_10937) encode the EIIB and EIIC permease components of a lactose/cellobiose class phosphoenolpyruvate:glycose phosphotransferase system (PTS), respectively. The PTS components transport and simultaneously phosphorylate a large number of carbohydrates and also carry out numerous regulatory functions [40]. Two other genes, OG1RF_10938 and

OG1RF_10939, seem to be co-transcribed with *egaG*, *egaB* and *egaC* and were named *egaD1* and *egaD2*. They are annotated as hypothetical proteins. Analyses of the genome sequences of other *E. faecalis* strains (V583, JH2-2, EnGen0311), showed that in all tested strains *egaD1* and *egaD2* are fused to one gene, named *egaD*. A closer look at the OG1RF sequence revealed that the splitting into two ORFs is the result of an authentic frameshift and that the *egaD2* ORF starts with an alternative start codon TTG at the end of *egaD1* (Fig. 3A). This strongly suggests that in OG1RF the *egaD* gene is naturally mutated. Blast analysis with the amino acid sequence of *egaD* of strain V583 revealed significant homology (54% identity; 100% coverage) to GrdB, which is part of the glycine/sarcosine/betaine reductase complex of *Clostridium sp.* This complex is composed of the three proteins GrdA, GrdB and GrdC. GrdB is a selenoprotein that metabolizes glycine, sarcosine or betaine to acetyl-phosphate and ammonia [41]. A Blast search with GrdA and GrdC of *Clostridium sp.* as queries did not reveal significant similarity with *E. faecalis* proteins. Identically organized homologues of the *ega*-locus of *E. faecalis* were found in strain V583, *Vagococcus lutrae* and *Lactocaseibacillus casei* BL23 (Fig. 3A and 3B) indicating that products of the *ega* genes might be functionally linked.

The ega genes are organized in two transcriptional units

The overall organization of the *ega* genes including the presence of a putative transcriptional terminator between *egaP* and *egaG* genes suggested that they are likely separated into two transcriptional units although only the transcription start site (TSS) upstream from *egaR* has been identified in a recent global analysis of TSS of *E. faecalis* [42]. We confirmed the dual operon organization experimentally by RT-PCR and 5'-RACE-PCR (Supplementary Fig. S1 and S2). The promoters P1 and P2 were mapped 49 and 34 bp upstream from the *egaR* and *egaG* start codons, respectively (Fig. 2). The previously identified TSS of the *ega* locus [42] is identical to the TSS of P1.

Construction of ega mutants and growth kinetics

To carry out a functional analysis of the *ega* genes, mutants were constructed by in-frame deletion of the central part of each gene in OG1RF strain, with the exception of *egaD1* and *egaD2*, which are the results of a natural mutation. Due to the expected role of the glycosylasparaginase, the OG1RF strain and the different mutants were grown on M17cc medium complemented with molecules mimicking the products released from protein deglycosylation. Two molecules were tested as substrates. Both contain the N-glycan core

structure N-acetylglucosamine bound to an asparagine (2-acetamido-2-deoxy- β -D-glucopyranosyl L-asparagine, GlcNAc-L-Asn). The difference between the two substrates is the presence or absence of a fucose residue linked to the GlcNAc. As shown in Fig. 4A, strain OG1RF was able to grow only on the GlcNAc-L-Asn. No growth was observed with the fucosylated glycan. This indicates that this compound cannot be internalized or metabolized by the bacteria. Three mutants, Δ *egaG*, Δ *egaB* and Δ *egaC* exhibited growth defects suggesting that the glycosylasparaginase and the PTS components EIIB and EIIC are involved in the utilization of this compound (Fig. 4B, 4C and 4D). Deletion of the *egaR* gene did not impact growth of the mutant (Supplementary Fig. S3A). This result is consistent with its presumed role as transcriptional repressor. In this mutant the relief of repression allows expression of *egaG* and the PTS genes and consequently GlcNAc-L-Asn utilization. Growth of the dipeptidase mutant was similar to that of the parental strain, indicating that EgaP is not required for growth on GlcNAc-L-Asn (Supplementary Fig. S3B'). Similarly, the *egaD1D2* genes are not essential for the utilization of GlcNAc-L-Asn, because strain OG1RF grew on this substrate as the sole carbon source although its *egaD* gene is disrupted by a frameshift mutation giving rise to *egaD1* and *egaD2* (Fig. 2).

The EIIA of a cellobiose-type PTS is required for the transport of the glycosylasparaginase substrate.

The EgaB and EgaC proteins are two elements of a PTS transporter belonging to the lactose/cellobiose class. To be functional, the permeases of this class are composed of 3 EII components, EIIA, EIIB and EIIC. However, a gene encoding an EIIA protein is absent in the *ega* locus, and hence should be located elsewhere in the genome. Analysis of the genome sequence revealed a single gene coding for an EIIA component belonging to the same class as the EgaBC proteins. This EIIA protein, OG1RF_10751, has previously been shown to complement the EIIB/EIIC components catalyzing the uptake of gentiobiose, a 1,6- β -diglucoside [19]. The EIIA/OG1RF_10751 mutant and the complemented strain were therefore tested for growth on GlcNAc-L-Asn (Fig. 4E). This mutant did not grow on GlcNAc-L-Asn, whereas complementation partially restored growth. This result strongly suggests that together with EgaB and EgaC the EIIA^{Cel} protein OG1RF_10751 catalyzes the PTS-mediated transport and phosphorylation of GlcNAc-L-Asn.

The glycosylasparaginase locus is induced during growth on GlcNAc-L-Asn

In order to confirm that the *ega* locus is required for the utilization of GlcNAc-L-Asn by *E. faecalis*, the expression of the six genes (*egaR*, *P*, *G*, *B*, *C* and *DI*) was assayed by RT-qPCR with cells cultivated on M17cc complemented with 0.3% glucose or GlcNAc-L-Asn. The expression of all genes was very low when cells were cultivated on glucose alone. Growth on GlcNAc-L-Asn strongly induced expression of the *ega* genes (Fig. 5A). Compared to glucose grown cells, expression of *egaR* and *P* was increased between 20 to 30-fold and expression of *egaG*, *B*, *C* and *DI* between 200 to 300-fold.

EgaR is the transcriptional repressor of the glycosylasparaginase locus

The constant localization of the *egaR* gene in the vicinity of *egaG* suggested that EgaR might be the regulator of the expression of the *egaGBCD* operon. EgaR resembles transcription regulators belonging to the GntR family. They usually function as repressors and control the expression of numerous genes involved in metabolic pathways [43–46]. Inside this large family, EgaR belongs to the HutC subfamily according to interproscan analysis [47]. Members of the HutC subfamily are usually regulated by binding small molecules to their C-terminal domain, such as histidine (HutC) or gluosamine-6-P (*B. subtilis* NagR/YvoA and GamR/YbgA) [48], whereas GntR from other subfamilies respond to fatty acids (FarR), sugars (TreR) and alkylphosphonate (PhnF). In order to experimentally confirm the role of EgaR in the regulation of the *egaGBCD* operon, we measured the expression levels of the glycosylasparaginase locus genes by RT-qPCR in the *egaR* deletion mutant of strain OG1RF (Fig. 5B). Expression of all genes was significantly higher than in the wild-type strain in agreement with a relief of repression. This was also the case for the *egaP* gene, suggesting that EgaR autoregulates its expression. Complementation with the *egaR* wild-type allele restored repression of all genes (Fig. 5B).

EgaR directly regulates expression of the *ega*-operons.

To show that EgaR directly interacts with the two promoter regions of the *ega* locus, purified EgaR was used for electromobility shift assays (EMSA). Three fluorescently labelled DNA fragments containing either the *egaRP* or the *egaGBCD* promoter region or a small part of the *gus* gene (control DNA) were incubated in the presence of EgaR. Shifted bands appeared for both promoter regions at EgaR concentrations higher than 200 nM, suggesting a direct specific interaction between EgaR and the DNA fragments (Fig. 6A). These interactions are considered

specific because the control DNA was not shifted under the same conditions, indicating that EgaR directly regulates expression of the *ega*-operons.

The consensus binding sequence of the HutC/GntR subfamily has been determined. It is composed of an 8-bp conserved repeated motif GTMTAKAC separated by 5-6 bp [49]. Such conserved motifs are present in the promoter regions of the *ega* locus (Fig. 1). Binding assays with EgaR in the presence of a mutated HutC repeated motif are presented in Fig. 6B. No band shift was visible, indicating that EgaR specifically binds to the GTMTAKAC motifs, which are located close to the -10 box of the transcription start site consistent with the repressor function of EgaR.

Interproscan analysis [47] of the EgaR regulator revealed a UbiC transcription regulator-associated (UTRA) domain, a conserved ligand-binding domain which is believed to modulate the activity of bacterial transcription factors in response to binding of small molecules such as sugar phosphates [50]. In the case of EgaR, GlcNAc-L-Asn phosphorylated during PTS-catalyzed transport was expected to bind to the repressor and to prevent it from interacting with its operator site. Because the phosphorylated compound is not commercially available, EMSA were performed in the presence of unphosphorylated GlcNAc-L-Asn (Fig. 6C). Starting from a condition (250 nM EgaR), where all DNA was shifted (Fig. 6C, lanes 2), increasing amounts of GlcNAc-L-Asn were added. When the concentration reached 7.5 mM, the DNA was no longer shifted, confirming that GlcNAc-L-Asn prevents EgaR from binding to the promoter region of both *ega* operons. This result is consistent with the induction of the *ega* genes observed when carrying out RT-qPCR with cells grown on GlcNAc-L-Asn (Fig. 5A). Interestingly, while carbohydrates transported by the PTS usually function as inducers only when they are phosphorylated, GlcNAc-L-Asn prevented EgaR/DNA interaction when it was unphosphorylated. Moreover, because EgaR is a member of the HutC subfamily, which is known to primarily bind amino acids, we suspected that binding of GlcNAc-L-Asn to EgaR is mainly mediated via the asparagine part. To test this possibility, EMSA were carried out with either asparagine or GlcNAc (Supplementary Fig. S4). Protein-DNA interaction (band shift) was prevented by both, asparagine and GlcNAc. This indicates that both parts of GlcNAc-L-Asn contribute to its binding to EgaR.

EgaG is a glycosylasparaginase.

To test whether EgaG exhibits the predicted glycosylasparaginase activity, the proenzyme was overexpressed in *E. coli* and purified (Supplementary Fig. S5). After electrophoresis, two bands

were observed on SDS gels with estimated molecular weights of 18 kDa and 16 kDa, likely corresponding to the N-terminal and C-terminal domains of the active enzyme after autoproteolysis. A BLAST-P research with EgaG evidenced 72.47% and 38.40% sequence identity to the glycosylasparaginase of *Lc. casei* (AsnA2) and *E. meningoseptica*, respectively [28,29]. Sequence alignments revealed that the amino acids involved in autoproteolysis in AsnA2 are conserved in EgaG (His149, Asp150 and Thr151).

We then determined the activity of EgaG by measuring the consumption of NADH in a coupled enzymatic assay (see Experimental procedures) (Fig. 7). When GlcNAc-L-Asn and EgaG are present in the mixture, a sharp decrease of absorbance is observed indicating that EgaG liberates aspartate from the substrate, which is further metabolized to oxaloacetate and malate under NADH consumption. This result confirms that hydrolysis of GlcNAc-L-Asn occurs between the NH₂ and the γ -carbon of asparagine and thus releases aspartate and 1-amino-GlcNAc instead of GlcNAc and asparagine. This is in agreement with the activity of the previously described GAs of *Lc. casei* and *E. meningoseptica*. The GA of *E. faecalis* has a specific activity of 19.35 $\mu\text{mol mg protein}^{-1} \text{ min}^{-1}$ for GlcNAc-L-Asn as the substrate. This value is very similar to those previously described for GAs of *Lc. casei* and *E. meningoseptica* with 20.91 $\mu\text{mol mg protein}^{-1} \text{ min}^{-1}$ and 25 $\mu\text{mol mg protein}^{-1} \text{ min}^{-1}$ respectively [28,33]. The near identity of the specific activities of the three GAs was surprising. The substrates of the PTS are usually phosphorylated during their uptake at the CH₂-OH group of the glucose moiety. We had therefore expected that EgaG would exhibit low specific activity with unphosphorylated GlcNAc-L-Asn and full activity only with the phosphorylated substrate. Because phosphorylated GlcNAc-L-Asn is not commercially available, this hypothesis could not be tested.

EgaP is an intracellular protein

Growth studies (Supplementary Fig. S3B) revealed that deletion of the dipeptidase-like EgaP does not affect utilization of GlcNAc-L-Asn. A similar result was obtained with *Lc. casei*, where a related peptidase, PepV, was not required for the utilization of fucosylated GlcNAc-L-Asn [28]. These peptidases belong to the M20 family, the members of which can non-specifically cleave peptide bonds of dipeptides [51,52]. It was therefore likely that EgaP is required for the catabolism of GlcNAc-L-Asn carrying at least one additional amino acid linked to Asn via a peptide bond. This raised the question, whether EgaP participates in the extracellular or intracellular degradation of GlcNAc-L-Asn-Xxx (Xxx corresponding to an amino acid linked to Asn). In fact, no signal peptide could be found in the amino acid sequence

of EgaP, suggesting that this protein is located in the cytoplasm. This assumption was supported by results obtained by mass spectrometry of the proteins produced by the deregulated Δ *egaR* mutant (Supplementary Table S6). EgaP was only detected in the cytoplasmic and not the extracellular protein fraction. This result suggests that EgaC is able to transport and to phosphorylate not only GlcNAc-L-Asn, but also substrates containing one or possibly more additional amino acids bound to Asn. The transport protein EgaC is probably able to bind substrates carrying additional small to medium-sized amino acids. The crystal structure of an EIIC of the cellobiose/lactose class [53] and photoinactivation experiments with azidophenyl-substrates [54] had shown that these EIICs contain a large pocket next to the sugar binding site, where big aglycons bound to C-1 of the sugar moiety can be placed and freely rotated. Once arrived in the cell as P-GlcNAc-L-Asn-Xxx, the Xxx needs to be removed, because, as already mentioned, EgaG belongs to the glycosyl asparaginase family that hydrolyzes substrates containing only Asn as amino acid. This peptidase reaction is most likely catalyzed by EgaP. The peptidase producing GlcNAc-L-Asn or GlcNAc-L-Asn-Xxx by degrading glycoproteins outside of bacterial cells (Fig. 1) has not yet been identified.

The ega locus is important for E. faecalis virulence.

To test whether EgaG might have a potential role in virulence, the *egaG* as well as *egaP*, *egaB*, and *egaC* deletion mutants were tested in the *G. mellonella* animal model (Fig. 8). Compared to the wild-type strain, for which the survival rate 24 h post injection reaches lowered to 25% of the infected larvae, more than 50% of the insects infected with the Δ *egaG* mutant were still alive at this time point. Similar results were obtained with the Δ *egaP*, Δ *egaB* and Δ *egaC* mutants. This strongly suggests that utilization of glycoprotein-derived GlcNAc-L-Asn-like substrates is important for *E. faecalis* virulence. Nevertheless, the finding that deletion of *egaP* also lowered enterococcal virulence was surprising. Deletion of the *egaP* gene had no effect on the uptake of GlcNAc-L-Asn, but attenuated enterococcal virulence. A possible explanation for this contradictory result might be that the degradation of *G. mellonella* N-glycosylated proteins mainly produces GlcNAc-L-Asn-Xxx, the catabolism of which requires functional EgaP, while during growth on GlcNAc-L-Asn EgaP is not needed.

Conclusion

We observed that *E. faecalis* is able to use GlcNAc-L-Asn, a degradation product of N-glycosylated proteins, as carbon and nitrogen source. The results obtained and the proposed annotation of the proteins encoded by the glycosylasparaginase locus allow us to propose a catabolic pathway for GlcNAc-L-Asn utilization by *E. faecalis* (Fig. 9). GlcNAc-L-Asn or an N-glycosylated di-peptide GlcNAc-L-Asn-Xxx are probably transported and phosphorylated by the PTS proteins EgaB, EgaC and OG1RF_10751. In the case where the Asn is linked to an additional amino acid, the intracellular substrate needs to be transformed to 6-P-GlcNAc-L-Asn and Xxx, presumably by the dipeptidase EgaP. The liberated Xxx enters the amino acid metabolism. 6-P-GlcNAc-L-Asn probably acts as inducer of the two *ega* transcription units by binding to the repressor EgaR and preventing it from interacting with its HutC-like operator sites. However, EMSA experiments had shown that unphosphorylated GlcNAc-L-Asn also interacts with EgaR. 6-P-GlcNAc-L-Asn is then hydrolyzed to 1-amino-6-P-GlcNAc and aspartate by the glycosylasparaginase EgaG. Aspartate also joins the metabolism of amino acids. Finally, 1-amino-6-P-GlcNAc is hydrolyzed to 6-P-GlcNAc and ammonia. Normally, this reaction is probably catalyzed by the deaminase encoded by the gene downstream from *egaC* (Fig. 2). However, in the OG1RF strain, this gene is disrupted by a frame shift mutation giving rise to the two small protein fragments EgaD1 and EgaD2. Deamination of 1-amino-6-P-GlcNAc is therefore assumed to occur non-enzymatically [55,56]. Interestingly, the glycosylasparaginase operon of *Lc. paracaesi* BL23 lacks a deaminase and hydrolysis of 1-amino-6-P-GlcNAc in this organism was also proposed to occur spontaneously [28].

However, several aspects in this model remain unclear. First, secreted enterococcal peptidases degrading proteoglycans have not been identified. The reaction products of such enzymes could be complex glycans linked to L-Asn or L-Asn-(Xxx)_n. To obtain GlcNAc-L-Asn, the core structure of the glycans needs to be cleaved between the two GlcNAc residues connected to Asn. The *E. faecalis* endo-β-N-acetylglucosaminidase EndoE has been reported to carry out this reaction [21–23]. Peptidase and glycosidase reactions might occur sequentially or simultaneously and should produce GlcNAc-L-Asn and/or GlcNAc-L-Asn-(Xxx)_n. Our results established that GlcNAc-L-Asn can be used by *E. faecalis* as the sole carbon source. However, experiments with *G. mellonella* rather demonstrated that during infection *E. faecalis* preferentially takes up GlcNAc-L-Asn-Xxx or GlcNAc-L-Asn-Xxx-(Xxx)_n. Structure data suggest that lactose/cellobiose type PTS are able to transport carbohydrates with relatively large aglycons [53]. These substrates are expected to arrive in the bacterial cell in phosphorylated

form and to be transformed by the intracellular dipeptidase EgaP and possibly other peptidases to P-GlcNAc-L-Asn, which is subsequently hydrolyzed by the glycosylasparaginase EgaG (Fig. 9). However, our results showed that unphosphorylated GlcNAc-L-Asn is not only hydrolyzed by EgaG, but also functions as inducer by binding to the repressor EgaR. Phosphorylation of GlcNAc-L-Asn might therefore not be important for its hydrolysis by EgaG and its inducer function or P-GlcNAc-L-Asn might be dephosphorylated before it is hydrolyzed by EgaG to Asp and N-acetylglucosaminyamine, which spontaneously decomposes to ammonia and GlcNAc. However, the latter scenario is unlikely, because in most bacteria GlcNAc is catabolized in phosphorylated form to fructose-6-P by the enzymes GlcNAc-6-phosphate deacetylase (NagA) and glucosamine-6-phosphate deaminase (NagB). Homologues of these two enzymes are present in *E. faecalis* and are required for the catabolism of GlcNAc and poly- β 1,4- GlcNAc [57]. The ability to utilize N-glycosylation products provides an obvious advantage during enterococcal infection, as demonstrated by the results obtained with the animal model *G. mellonella*. Without the capacity to use N-glycosylation products, bacterial fitness was diminished, resulting in reduced virulence. To our knowledge, this is the first time that the PTS-catalyzed uptake of probably GlcNAc-L-Asn-Xxx and its catabolism by a dipeptidase and glycosylasparaginase has been shown to enhance bacterial virulence.

Acknowledgements

We are grateful to IMOGERE and PROTEOGEN platforms who let us use their equipments. We thank the Normandy Region for the doctoral fellowship of V.C. This work was funded by the French Ministry of Higher Education.

Conflicts of Interest

The authors declare no conflict of interest.

Bibliography

- [1] Cattoir V. The multifaceted lifestyle of enterococci: genetic diversity, ecology and risks for public health. *Curr Opin Microbiol* 2021;65:73–80. <https://doi.org/10.1016/j.mib.2021.10.013>.
- [2] Fisher K, Phillips C. The ecology, epidemiology and virulence of *Enterococcus*. *Microbiology* 2009;155:1749–57. <https://doi.org/10.1099/mic.0.026385-0>.

- [3] Goh HMS, Yong MHA, Chong KKL, Kline KA. Model systems for the study of Enterococcal colonization and infection. *Virulence* 2017;8:1525–62. <https://doi.org/10.1080/21505594.2017.1279766>.
- [4] Gillespie SH, Hawkey PM, Nannini EC, Murray BE, editors. *Enterococcus spp. Principles and practice of clinical bacteriology*. 2nd ed, Chichester, West Sussex, England; Hoboken: John Wiley & Sons; 2006, p. 59–71.
- [5] Weiner-Lastinger LM, Haass K, Gross C, Leaptrot D, Wong E, Wu H, *et al.* Pathogens attributed to central line-associated bloodstream infections in US acute care hospitals during the first year of the COVID-19 pandemic. *Infect Control Hosp Epidemiol* 2022:1–13. <https://doi.org/10.1017/ice.2022.16>.
- [6] Agudelo Higuaita NI, Huycke MM. Enterococcal Disease, Epidemiology, and Implications for Treatment. In: Gilmore MS, Clewell DB, Ike Y, Shankar N, editors. *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection*, Boston: Massachusetts Eye and Ear Infirmary; 2014.
- [7] García-Solache M, Rice LB. The *Enterococcus*: a Model of Adaptability to Its Environment. *Clin Microbiol Rev* 2019;32. <https://doi.org/10.1128/CMR.00058-18>.
- [8] Ceci M, Delpech G, Sparo M, Mezzina V, Bruni SS, Baldaccini B. Clinical and microbiological features of bacteremia caused by *Enterococcus faecalis*. *The Journal of Infection in Developing Countries* 2015;9:1195–203. <https://doi.org/10.3855/jidc.6587>.
- [9] Grand M, Riboulet-Bisson E, Deutscher J, Hartke A, Sauvageot N. *Enterococcus faecalis* Maltodextrin Gene Regulation by Combined Action of Maltose Gene Regulator MalR and Pleiotropic Regulator CcpA. *Appl Environ Microbiol* 2020;86:e01147-20. <https://doi.org/10.1128/AEM.01147-20>.
- [10] Kristich CJ, Rice LB, Arias CA. Enterococcal Infection—Treatment and Antibiotic Resistance. In: Gilmore MS, Clewell DB, Ike Y, Shankar N, editors. *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection*, Boston: Massachusetts Eye and Ear Infirmary; 2014.
- [11] Cao TN, Joyet P, Aké FMD, Milohanic E, Deutscher J. Studies of the *Listeria monocytogenes* Cellobiose Transport Components and Their Impact on Virulence Gene Repression. *J Mol Microbiol Biotechnol* 2019;29:10–26. <https://doi.org/10.1159/000500090>.
- [12] Chang D-E, Smalley DJ, Tucker DL, Leatham MP, Norris WE, Stevenson SJ, *et al.* Carbon nutrition of *Escherichia coli* in the mouse intestine. *PNAS* 2004;101:7427–32. <https://doi.org/10.1073/pnas.0307888101>.

- [13] Thomas M, Wrzosek L, Ben-Yahia L, Noordine M-L, Gitton C, Chevret D, *et al.* Carbohydrate Metabolism Is Essential for the Colonization of *Streptococcus thermophilus* in the Digestive Tract of Gnotobiotic Rats. *PLOS ONE* 2011;6:e28789. <https://doi.org/10.1371/journal.pone.0028789>.
- [14] Keffeler EC, Iyer VS, Henderson AJ, Huck IL, Schwarting N, Cortez A, *et al.* Activity of CcpA-Regulated GH18 Family Glycosyl Hydrolases That Contributes to Nutrient Acquisition and Fitness in *Enterococcus faecalis*. *Infect Immun* 2021;89:e00343-21. <https://doi.org/10.1128/IAI.00343-21>.
- [15] Sjögren J, Struwe WB, Cosgrave EFJ, Rudd PM, Stervander M, Allhorn M, *et al.* EndoS2 is a unique and conserved enzyme of serotype M49 group A *Streptococcus* that hydrolyses N-linked glycans on IgG and α 1-acid glycoprotein. *Biochem J* 2013;455:107–18. <https://doi.org/10.1042/BJ20130126>.
- [16] Braza RE, Silver AB, Sundar GS, Davis SE, Razi A, Islam E, *et al.* Phosphotransferase System Uptake and Metabolism of the β -Glucoside Salicin Impact Group A Streptococcal Bloodstream Survival and Soft Tissue Infection. *Infect Immun* 2020;88:e00346-20. <https://doi.org/10.1128/IAI.00346-20>.
- [17] Mao Q, Jiang J, Wu X, Xu R, Ma Y, Zhang Y, *et al.* Coordinate regulation of carbohydrate metabolism and virulence by PtsH in pathogen *Edwardsiella piscicida*. *Appl Microbiol Biotechnol* 2022;106:2063–77. <https://doi.org/10.1007/s00253-022-11848-8>.
- [18] Muller C, Cacaci M, Sauvageot N, Sanguinetti M, Rattei T, Eder T, *et al.* The Intraperitoneal Transcriptome of the Opportunistic Pathogen *Enterococcus faecalis* in Mice. *PLoS ONE* 2015;10:e0126143. <https://doi.org/10.1371/journal.pone.0126143>.
- [19] Grand M, Aubourg M, Pikis A, Thompson J, Deutscher J, Hartke A, *et al.* Characterization of the gen locus involved in β -1,6-oligosaccharide utilization by *Enterococcus faecalis*. *Mol Microbiol* 2019;112:1744–56. <https://doi.org/10.1111/mmi.14390>.
- [20] Nothaft H, Szymanski CM. Bacterial protein N-glycosylation: new perspectives and applications. *J Biol Chem* 2013;288:6912–20. <https://doi.org/10.1074/jbc.R112.417857>.
- [21] Garbe J, Sjögren J, Cosgrave EFJ, Struwe WB, Bober M, Olin AI, *et al.* EndoE from *Enterococcus faecalis* Hydrolyzes the Glycans of the Biofilm Inhibiting Protein Lactoferrin and Mediates Growth. *PLoS ONE* 2014;9:e91035. <https://doi.org/10.1371/journal.pone.0091035>.
- [22] Roberts G, Tarelli E, Homer KA, Philpott-Howard J, Beighton D. Production of an Endo- β -N-Acetylglucosaminidase Activity Mediates Growth of *Enterococcus faecalis* on a High-Mannose-Type Glycoprotein. *J Bacteriol* 2000;182:882–90.

- [23] Collin M, Fischetti VA. A Novel Secreted Endoglycosidase from *Enterococcus faecalis* with Activity on Human Immunoglobulin G and Ribonuclease B. *Journal of Biological Chemistry* 2004;279:22558–70. <https://doi.org/10.1074/jbc.M402156200>.
- [24] Sun G, Yu X, Bao C, Wang L, Li M, Gan J, *et al.* Identification and characterization of a novel prokaryotic peptide: N-glycosidase from *Elizabethkingia meningoseptica*. *J Biol Chem* 2015;290:7452–62. <https://doi.org/10.1074/jbc.M114.605493>.
- [25] Mononen IT, Kaartinen VM, Williams JC. A Fluorometric Assay for Glycosylasparaginase Activity and Detection of Aspartylglycosaminuria. *Analytical Biochemistry* 1993;208:372–4. <https://doi.org/10.1006/abio.1993.1063>.
- [26] Wang Y, Guo H-C. Crystallographic Snapshot of a Productive Glycosylasparaginase–Substrate Complex. *Journal of Molecular Biology* 2007;366:82–92. <https://doi.org/10.1016/j.jmb.2006.09.051>.
- [27] Risley JM, Huang DH, Kaylor JJ, Malik JJ, Xia Y-Q, York WM. Glycosylasparaginase Activity Requires the α -Carboxyl Group, but Not the α -Amino Group, on N4-(2-Acetamido-2-deoxy- β -d-glucopyranosyl)-l-asparagine. *Archives of Biochemistry and Biophysics* 2001;391:165–70. <https://doi.org/10.1006/abbi.2001.2416>.
- [28] Becerra JE, Rodríguez-Díaz J, Gozalbo-Rovira R, Palomino-Schätzlein M, Zúñiga M, Monedero V, *et al.* Unique Microbial Catabolic Pathway for the Human Core N-Glycan Constituent Fucosyl- α -1,6-N-Acetylglucosamine-Asparagine. *MBio* 2020;11:e02804-19. <https://doi.org/10.1128/mBio.02804-19>.
- [29] Xuan J, Tarentino AL, Grimwood BG, Plummer TH, Cui T, Guan C, *et al.* Crystal structure of glycosylasparaginase from *Flavobacterium meningosepticum*. *Protein Sci* 1998;7:774–81. <https://doi.org/10.1002/pro.5560070327>.
- [30] Tarentino AL, Plummer TH. The first demonstration of a prokaryotic glycosylasparaginase. *Biochem Biophys Res Commun* 1993;197:179–86. <https://doi.org/10.1006/bbrc.1993.2457>.
- [31] Guan C, Cui T, Rao V, Liao W, Benner J, Lin C-L, *et al.* Activation of Glycosylasparaginase: formation of active N-terminal threonine by intramolecular autoproteolysis. *J Biol Chem* 1996;271:1732–7. <https://doi.org/10.1074/jbc.271.3.1732>.
- [32] Wang Y, Guo H-C. Crystallographic Snapshot of Glycosylasparaginase Precursor Poised for Autoprocessing. *Journal of Molecular Biology* 2010;403:120–30. <https://doi.org/10.1016/j.jmb.2010.08.038>.

- [33] Liu Y, Guan C, Aronson NN. Site-directed Mutagenesis of Essential Residues Involved in the Mechanism of Bacterial Glycosylasparaginase. *Journal of Biological Chemistry* 1998;273:9688–94. <https://doi.org/10.1074/jbc.273.16.9688>.
- [34] Dunny GM, Brown BL, Clewell DB. Induced cell aggregation and mating in *Streptococcus faecalis*: evidence for a bacterial sex pheromone. *Proc Natl Acad Sci U S A* 1978;75:3479–83.
- [35] La Carbona S, Sauvageot N, Giard J-C, Benachour A, Posteraro B, Auffray Y, *et al.* Comparative study of the physiological roles of three peroxidases (NADH peroxidase, Alkyl hydroperoxide reductase and Thiol peroxidase) in oxidative stress response, survival inside macrophages and virulence of *Enterococcus faecalis*. *Mol Microbiol* 2007;66:1148–63. <https://doi.org/10.1111/j.1365-2958.2007.05987.x>.
- [36] Green R, Rogers EJ. Transformation of chemically competent *E. coli*. *Methods Enzymol* 2013;529:329–36. <https://doi.org/10.1016/B978-0-12-418687-3.00028-8>.
- [37] Holo H, Nes IF. High-Frequency Transformation, by Electroporation, of *Lactococcus lactis subsp. cremoris* Grown with Glycine in Osmotically Stabilized Media. *Applied and Environmental Microbiology* 1989. <https://doi.org/10.1128/aem.55.12.3119-3123.1989>.
- [38] Arnaud M, Chastanet A, Débarbouillé M. New Vector for Efficient Allelic Replacement in Naturally Nontransformable, Low-GC-Content, Gram-Positive Bacteria. *Appl Environ Microbiol* 2004;70:6887–91. <https://doi.org/10.1128/AEM.70.11.6887-6891.2004>.
- [39] Bryan EM, Bae T, Kleerebezem M, Dunny GM. Improved Vectors for Nisin-Controlled Expression in Gram-Positive Bacteria. *Plasmid* 2000;44:183–90. <https://doi.org/10.1006/plas.2000.1484>.
- [40] Deutscher J, Aké FMD, Derkaoui M, Zébré AC, Cao TN, Bouraoui H, *et al.* The Bacterial Phosphoenolpyruvate:Carbohydrate Phosphotransferase System: Regulation by Protein Phosphorylation and Phosphorylation-Dependent Protein-Protein Interactions. *Microbiol Mol Biol Rev* 2014;78:231–56. <https://doi.org/10.1128/MMBR.00001-14>.
- [41] Andreesen JR. Glycine reductase mechanism. *Current Opinion in Chemical Biology* 2004;8:454–61. <https://doi.org/10.1016/j.cbpa.2004.08.002>.
- [42] Michaux C, Hansen EE, Jenniches L, Gerovac M, Barquist L, Vogel J. Single-Nucleotide RNA Maps for the Two Major Nosocomial Pathogens *Enterococcus faecalis* and *Enterococcus faecium*. *Front Cell Infect Microbiol* 2020;10:600325. <https://doi.org/10.3389/fcimb.2020.600325>.

- [43] You D, Zhang B-Q, Ye B-C. GntR Family Regulator DasR Controls Acetate Assimilation by Directly Repressing the *acsA* Gene in *Saccharopolyspora erythraea*. *J Bacteriol* 2018;200:e00685-17. <https://doi.org/10.1128/JB.00685-17>.
- [44] Blancato VS, Repizo GD, Suárez CA, Magni C. Transcriptional regulation of the citrate gene cluster of *Enterococcus faecalis* Involves the GntR family transcriptional activator CitO. *J Bacteriol* 2008;190:7419–30. <https://doi.org/10.1128/JB.01704-07>.
- [45] Zhang M, Kang Z, Guo X, Guo S, Xiao D, Liu Y, et al. Regulation of Glutarate Catabolism by GntR Family Regulator CsiR and LysR Family Regulator GcdR in *Pseudomonas putida* KT2440. *MBio* 2019;10:e01570-19. <https://doi.org/10.1128/mBio.01570-19>.
- [46] Gaugué I, Oberto J, Putzer H, Plumbridge J. The use of amino sugars by *Bacillus subtilis*: presence of a unique operon for the catabolism of glucosamine. *PLoS One* 2013;8:e63025. <https://doi.org/10.1371/journal.pone.0063025>.
- [47] Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;30:1236–40. <https://doi.org/10.1093/bioinformatics/btu031>.
- [48] Bertram R, Rigali S, Wood N, Lulko AT, Kuipers OP, Titgemeyer F. Regulon of the N-acetylglucosamine utilization regulator NagR in *Bacillus subtilis*. *Journal of Bacteriology* 2011;193:3525–36. <https://doi.org/10.1128/JB.00264-11>.
- [49] Suvorova IA, Korostev YD, Gelfand MS. GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution. *PLOS ONE* 2015;10:e0132618. <https://doi.org/10.1371/journal.pone.0132618>.
- [50] Matilla MA, Velando F, Martín-Mora D, Monteagudo-Cascales E, Krell T. A catalogue of signal molecules that interact with sensor kinases, chemoreceptors and transcriptional regulators. *FEMS Microbiology Reviews* 2022;46:fuab043. <https://doi.org/10.1093/femsre/fuab043>.
- [51] Hellendoorn MA, Franke-Fayard BM, Mierau I, Venema G, Kok J. Cloning and analysis of the *pepV* dipeptidase gene of *Lactococcus lactis* MG1363. *J Bacteriol* 1997;179:3410–5. <https://doi.org/10.1128/jb.179.11.3410-3415.1997>.
- [52] Jozic D, Bourenkow G, Bartunik H, Scholze H, Dive V, Henrich B, et al. Crystal structure of the dinuclear zinc aminopeptidase PepV from *Lactobacillus delbrueckii* unravels its preference for dipeptides. *Structure* 2002;10:1097–106. [https://doi.org/10.1016/s0969-2126\(02\)00805-5](https://doi.org/10.1016/s0969-2126(02)00805-5).

- [53] Cao Y, Jin X, Levin EJ, Huang H, Zong Y, Quick M, *et al.* Crystal structure of a phosphorylation-coupled saccharide transporter. *Nature* 2011;473:50–4. <https://doi.org/10.1038/nature09939>.
- [54] Sossna-Wunder G, Hengstenberg W, Briozzo P, Deutscher J. Photoinactivation of the *Staphylococcus aureus* Lactose-Specific EIICB Phosphotransferase Component with p-azidophenyl- β -D-Galactoside and Phosphorylation of the Covalently Bound Substrate. *J Mol Microbiol Biotechnol* 2018;28:147–58. <https://doi.org/10.1159/000494433>.
- [55] Makino M, Kojima T, Yamashina I. Enzymatic cleavage of glycopeptides. *Biochemical and Biophysical Research Communications* 1966;24:961–6. [https://doi.org/10.1016/0006-291X\(66\)90344-5](https://doi.org/10.1016/0006-291X(66)90344-5).
- [56] Aronson NN. Aspartylglycosaminuria: biochemistry and molecular biology. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1999;1455:139–54. [https://doi.org/10.1016/S0925-4439\(99\)00076-9](https://doi.org/10.1016/S0925-4439(99)00076-9).
- [57] Keffeler EC, Parthasarathy S, Abdullahi ZH, Hancock LE. Metabolism of Poly- β 1,4-N-Acetylglucosamine Substrates and Importation of N-Acetylglucosamine and Glucosamine by *Enterococcus faecalis*. *J Bacteriol* 2021;203:e0037121. <https://doi.org/10.1128/JB.00371-21>.

Fig. 3: **A:** Structural organization of the *ega* locus in *E. faecalis* strain V583 and *V. lutrae*. **B:** genetic organization of the glycosylasparaginase locus of *Lc. casei* and *E. meningoseptica*. Percentage identity of AA sequences of related proteins is indicated above the homolog.

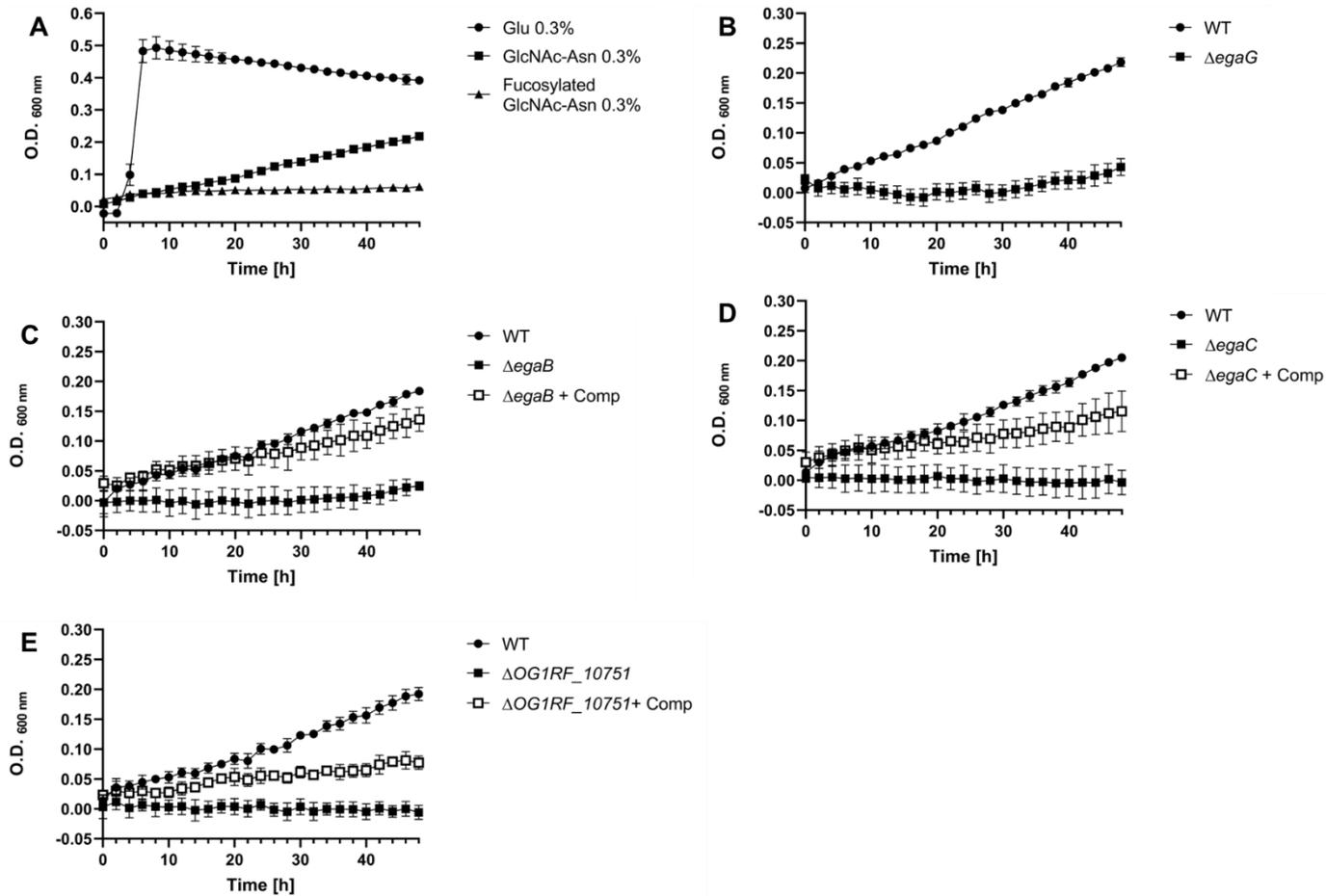


Fig. 4: **A:** Growth studies with *E. faecalis* wild-type OG1RF strain on 0.3% glucose (circle), 0.3% GlcNAc-L-Asn (square) or 0.3% Fucosylated-GlcNAc-L-Asn (triangle). **B:** Growth studies of ΔegaG on GlcNAc-L-Asn 0.3%. **C:** Growth studies of ΔegaB on GlcNAc-L-Asn 0.3%. **D:** Growth studies of ΔegaC on GlcNAc-L-Asn 0.3%. **E:** Growth studies of $\Delta\text{OG1RF}_{10751}$ on GlcNAc-L-Asn 0.3%. **B-E:** Wild type strains are indicated with circle, mutated strains with black square and complemented strains with white square.

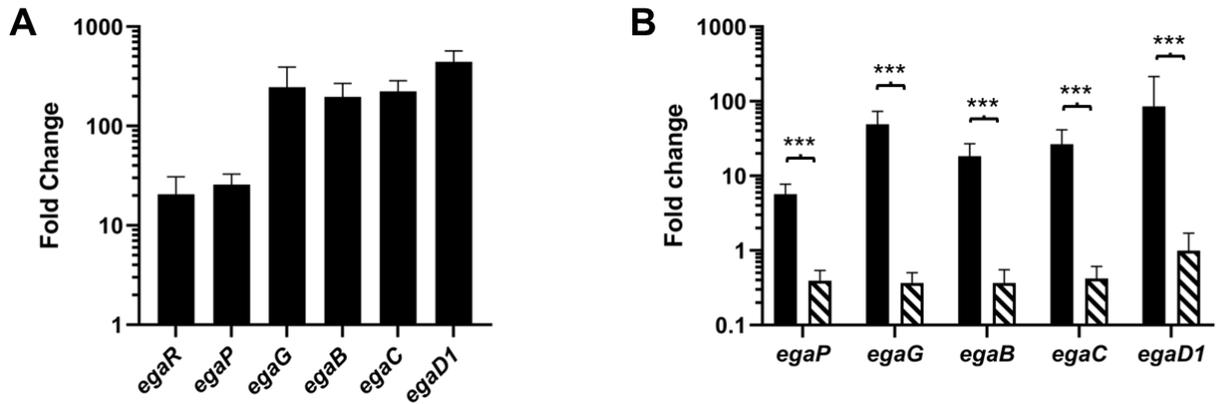


Fig. 5: **A**: Relative expression of the *egaR*, *egaP*, *egaG*, *egaB*, *egaC* and *egaD1* genes in the wild-type OG1RF strain determined by RT-qPCR assays. The wild-type strain was grown in M17cc medium containing 0.3% glucose or 0.3% GlcNAc-L-Asn. The fold change was calculated by dividing the expression level determined for a mutant during growth on GlcNAc-L-Asn by the expression level observed during its growth on glucose. **B**: Expression of the *egaP*, *egaG*, *egaB*, *egaC* and *egaD1* genes in glucose-grown Δ *egaR* (black boxes) or complemented Δ *egaR* (hatched boxes) relative to their expression in glucose-grown wild-type strain. Fold changes represent the ratio of the expression of a gene in Δ *egaR* or complemented Δ *egaR* divided by its expression in the wild-type strain, which was always set to 1. The mean values calculated from three independent experiments are presented. Standard deviations are indicated.

Figure 6 :

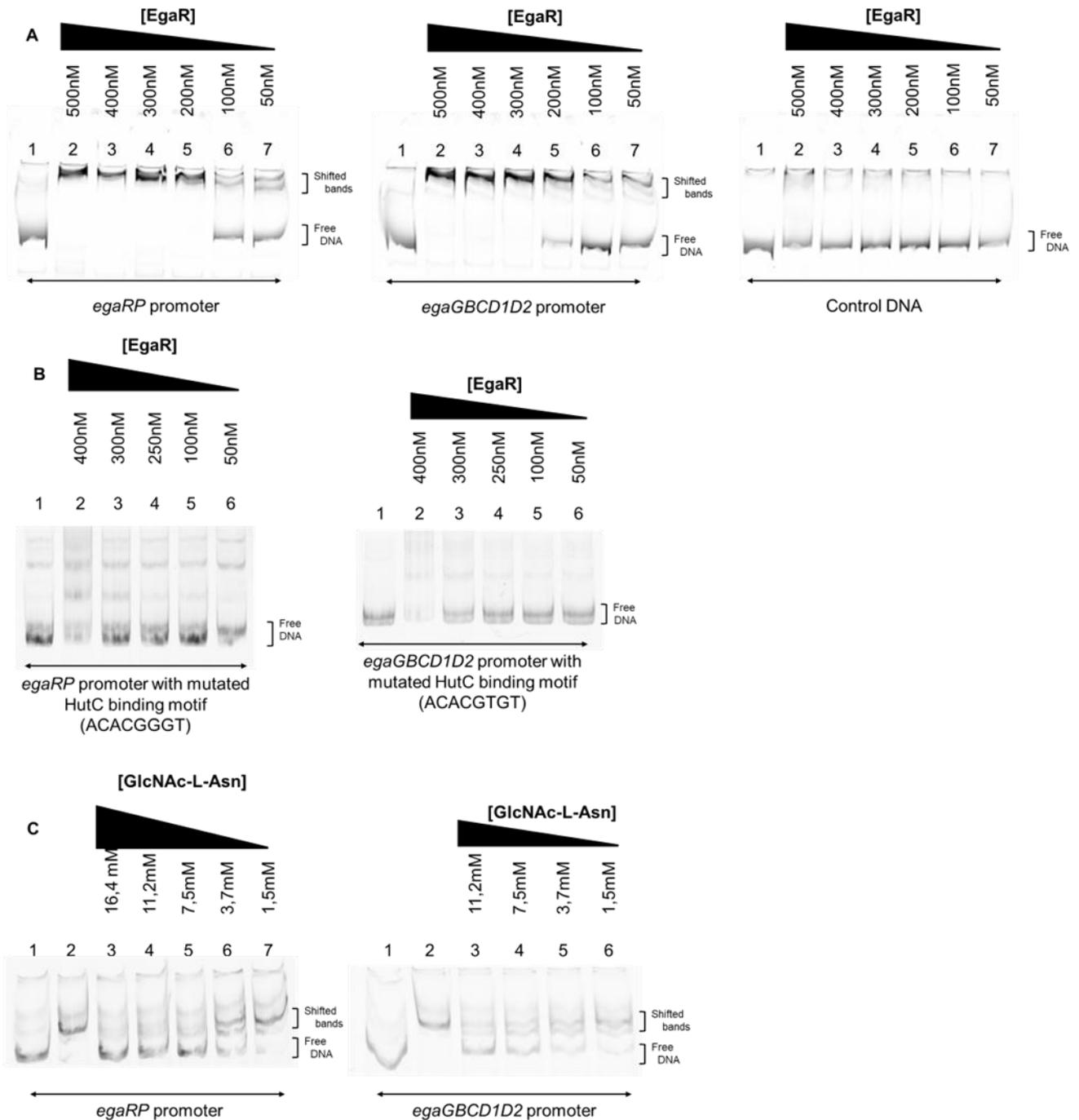


Fig. 6: **A**: EMSA of the promoter regions of the *egaRP* and *egaGBCD1D2* genes. EMSAs were performed with 5 nM of labelled DNA and various concentrations of EgaR ranging from 500 nM to 50 nM (lanes 2 to 7) as indicated above the gel. A negative control was carried out with an endogenous DNA fragment from the *gus* gene. DNA without protein was loaded on lane 1. **B**: EMSA of the *egaRP* and *egaGBCD1D2* promoters with mutated HutC binding motif. EMSA was performed with 5 nM of labelled DNA and concentration of EgaR ranging from 400 nM to 50 nM (lanes 2 to 6) as indicated above the gel. DNA without protein was loaded on lane 1.

C: EMSA of the promoter region of the *egaRP* genes and *egaGBCD1D2* genes in the presence of GlcNAc-L-Asn. EMSAs were performed with 5 nM of labelled DNA, 250 nM of EgaR and various concentrations of GlcNAc-L-Asn ranging from 16.4 mM to 1.5 mM (lanes 3 to 7) as indicated above the gel. DNA without EgaR and GlcNAc-L-Asn was loaded on lane 1 and DNA with 250 nM EgaR but without GlcNAc-L-Asn on lane 2.

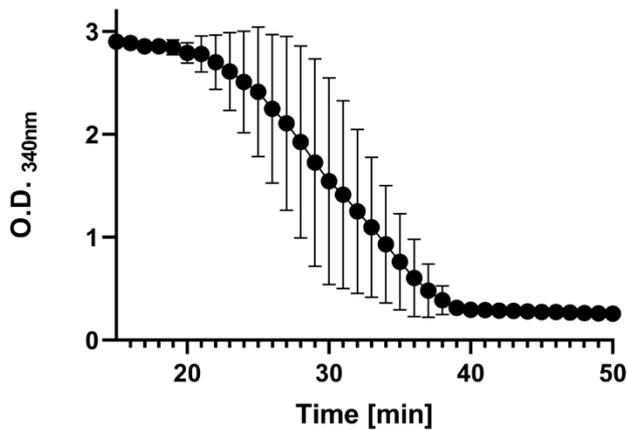


Fig. 7: Kinetic analysis of EgaG's glycosylasparaginase activity. Aspartate produced by the degradation of GlcNAc-L-Asn by the glycosylasparaginase is used by the GOT and MDH enzymes. The latter uses NADH as a cofactor. Progress of the reaction is monitored by measuring the absorbance of NADH at O.D._{340nm}. The mean values calculated from three independent experiments are presented. Standard deviations are indicated.

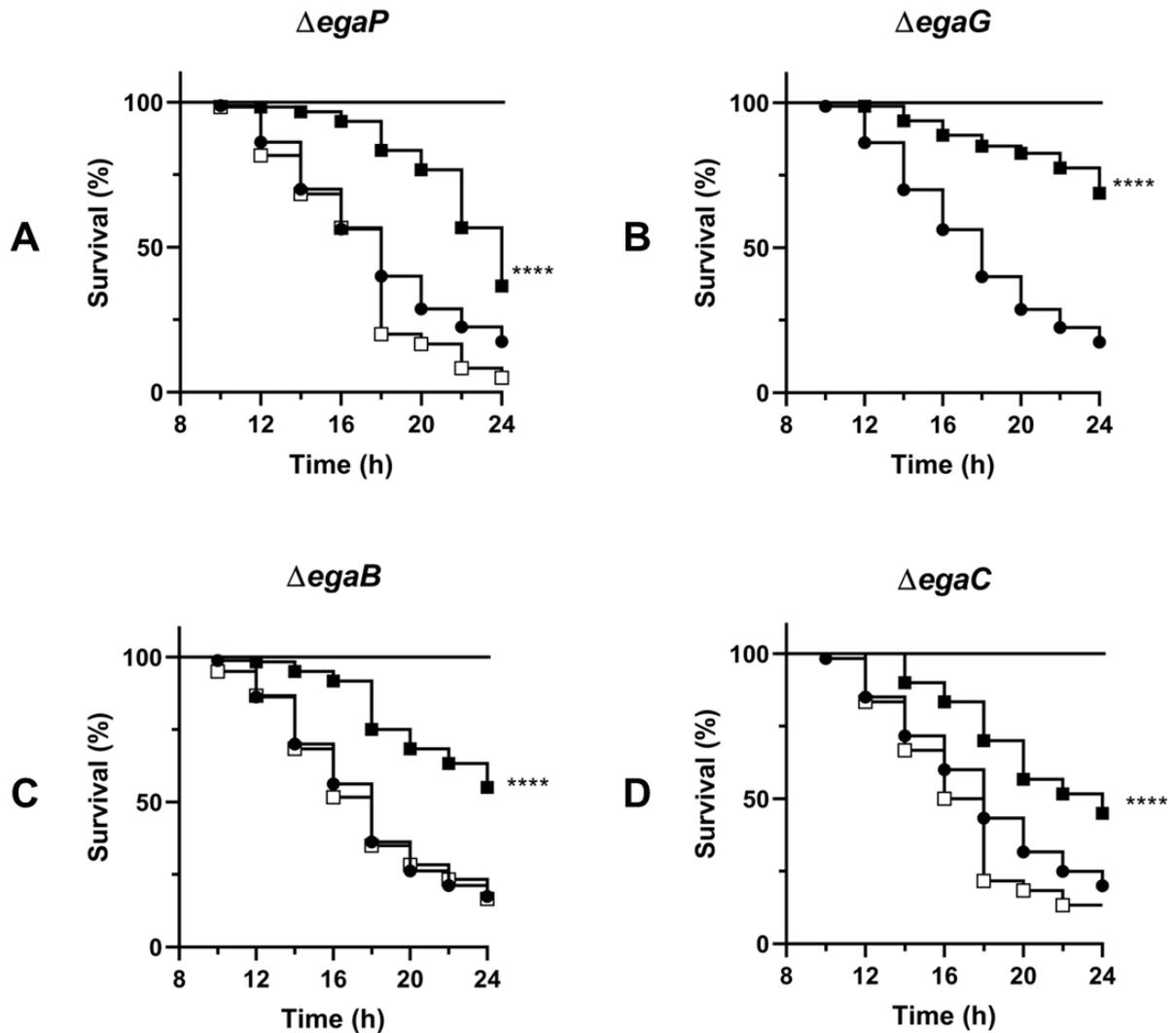


Fig 8: Survival of *G. mellonella* larvae after infection with various *E. faecalis* strain. *G. mellonella* larvae were infected with wild-type *E. faecalis* OG1RF (black circle), mutant strains (black square), mutants complemented with plasmid pMSP3535 carrying the corresponding gene (white square) or saline solution without bacteria (black line). The effect of the gene deletion on larvae survival was monitored with **A** : ΔegaP , **B** : ΔegaG , **C** : ΔegaB and **D** : ΔegaC . Living and dead larvae were enumerated every 2 h post-infection. A fourfold asterisk indicates a significant difference (P < 0.001, log-rank test) of virulence between wild-type and mutant.

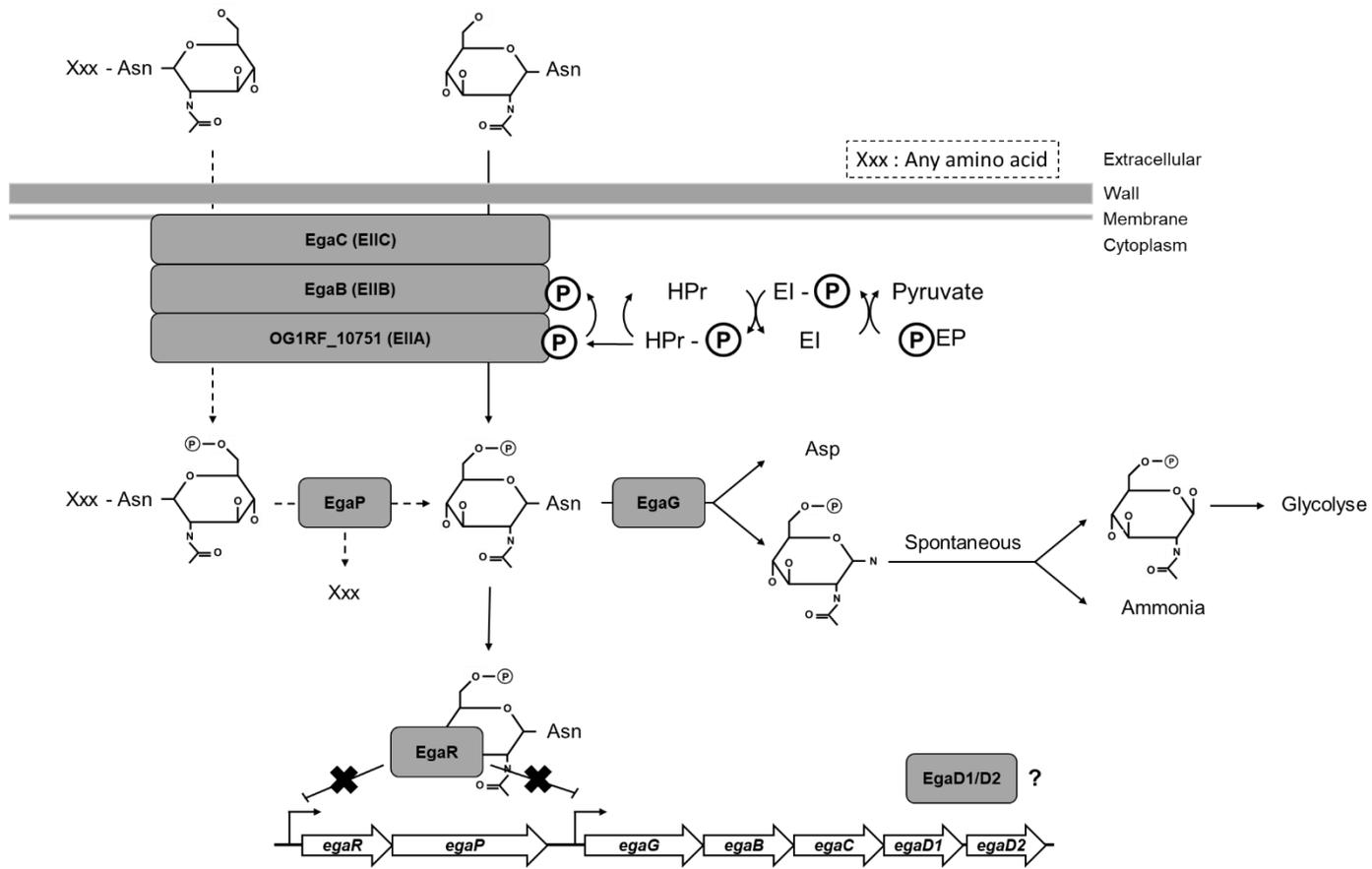


Fig. 9: Proposed model of the catabolic pathway of GlcNAc-L-Asn utilization in *E. faecalis*. Steps indicated by solid lines and arrows are supported by experimental data. Dotted lines and arrows indicate hypothetical pathways.

Supplementary data

Table S1: Strains used in this study.

Strains	Relevant genotype	Reference or source
<i>Enterococcus faecalis</i>		
OG1RF	Fus ^r , Rif ^r , plasmid-free wild-type strain	[34]
OG1RF pMSP3535	WT strain carrying the pMSP3535 plasmid	This study
Δ <i>egaR</i> mutant	<i>egaR</i> deletion mutant of OG1RF parental strain	This study
Δ <i>egaR</i> pMSP3535:: <i>egaR</i> mutant	<i>egaR</i> mutant carrying the pMSP3535:: <i>egaR</i> plasmid	This study
Δ <i>egaR</i> pMSP3535 mutant	<i>egaR</i> mutant carrying the pMSP3535 empty plasmid	This study
Δ <i>egaP</i> mutant	<i>egaP</i> deletion mutant of OG1RF parental strain	This study
Δ <i>egaP</i> pMSP3535:: <i>egaP</i> mutant	<i>egaP</i> mutant carrying the pMSP3535:: <i>egaP</i> plasmid	This study
Δ <i>egaP</i> pMSP3535 mutant	<i>egaP</i> mutant carrying the pMSP3535 empty plasmid	This study
Δ <i>egaG</i> mutant	<i>egaG</i> deletion mutant of OG1RF parental strain	This study
Δ <i>egaB</i> mutant	<i>egaB</i> deletion mutant of OG1RF parental strain	This study
Δ <i>egaB</i> pMSP3535:: <i>egaB</i> mutant	<i>egaB</i> mutant carrying the pMSP3535:: <i>egaB</i> plasmid	This study
Δ <i>egaB</i> pMSP3535 mutant	<i>egaB</i> mutant carrying the pMSP3535 empty plasmid	This study
Δ <i>egaC</i> mutant	<i>egaC</i> deletion mutant of OG1RF parental strain	This study
Δ <i>egaC</i> pMSP3535:: <i>egaC</i> mutant	<i>egaC</i> mutant carrying the pMSP3535:: <i>egaC</i> plasmid	This study
Δ <i>egaC</i> pMSP3535 mutant	<i>egaC</i> mutant carrying the pMSP3535 empty plasmid	This study
Δ OG1RF_10751 mutant	OG1RF_10751 deletion mutant of OG1RF parental strain	[19]
Δ OG1RF_10751:: <i>OG1RF_10751</i> mutant	Complemented strain with OG1RF_10751 gene of <i>E. faecalis</i> strain V583	[19]
<i>Escherichia coli</i>		
TOP10	F- <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ (<i>ara-leu</i>) 7697 <i>galU</i> <i>galK</i> <i>rpsL</i> (Str ^R) <i>endA1</i> <i>nupG</i> λ	Thermo Fischer Scientific

NEB5 α	<i>fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17</i>	New England Biolabs
BL21(DE3)	F- <i>ompT hsdS_B (r_BM_B-) gal dcm</i> (DE3)	New England Biolabs
BL21(DE3) pET29b+:: <i>egaR</i>	<i>E. coli</i> BL21 carrying the pET29b+:: <i>egaR</i> overexpression plasmid	This study
M15pRep4	F-, Φ 80 Δ <i>lacM15</i> , <i>thi</i> , <i>lac</i> -, <i>mtl</i> -, <i>recA</i> +, Km ^R	Qiagen
M15pRep4 pQE30:: <i>egaG</i>	<i>E. coli</i> M15pRep4 carrying the pQE30:: <i>egaR</i> overexpression plasmid	This study

Table S2: Plasmids used in this study.

Plasmids name	Relevant properties	Reference or source
pMAD	Origins of replication pE194ts and pBR322, Em ^r , Amp ^r , <i>bgaB</i>	[38]
pMAD:: Δ <i>egaR</i>	pMAD containing the <i>egaR</i> gene region with a deletion of the <i>egaR</i> gene	This study
pMAD:: Δ <i>egaP</i>	pMAD containing the <i>egaP</i> gene region with a deletion of the <i>egaP</i> gene	This study
pMAD:: Δ <i>egaG</i>	pMAD containing the <i>egaG</i> gene region with a deletion of the <i>egaG</i> gene	This study
pMAD:: Δ <i>egaB</i>	pMAD containing the <i>egaB</i> gene region with a deletion of the <i>egaB</i> gene	This study
pMAD:: Δ <i>egaC</i>	pMAD containing the <i>egaC</i> gene region with a deletion of the <i>egaC</i> gene	This study
pMSP3535	pNisA promoter, Em ^r , <i>nisR</i> , <i>nisK</i>	[39]
pMSP3535:: <i>egaR</i>	pMSP3535 containing the <i>egaR</i> gene downstream the nisine inducible promoter	This study
pMSP3535:: <i>egaP</i>	pMSP3535 containing the <i>egaP</i> gene downstream the nisine inducible promoter	This study
pMSP3535:: <i>egaB</i>	pMSP3535 containing the <i>egaB</i> gene downstream the nisine inducible promoter	This study
pMSP3535:: <i>egaC</i>	pMSP3535 containing the <i>egaC</i> gene downstream the nisine inducible promoter	This study
pGEM-T Easy	Cloning of P1, P2 and an internal fragment of the <i>gusA</i> gene to produce a labelled PCR product for EMSA, Amp ^R	Promega
pGEMT P1	pGEM-T Easy vector carrying the <i>egaRP</i> promoter	This study
pGEMT P2	pGEM-T Easy vector carrying the <i>egaGBCDE</i> promoter	This study
pGEMT <i>gus</i>	pGEM-T Easy vector containing an internal fragment of the <i>gusA</i> gene	[9]
pET29b+	Expression of C-ter-6x-His fusion proteins, Km ^R , Ori ColE1	Novagen

pET29b+:: <i>egaR</i>	Expression of C-ter-6xHis tagged EgaR protein of <i>E. faecalis</i> OG1RF	This study
pQE30	Expression of N-ter-6xHis fusion proteins, Amp ^R , Cm ^R , Ori ColE1	Qiagen
pQE30:: <i>egaG</i>	Expression of N-ter-6xHis tagged EgaG protein of <i>E. faecalis</i> OG1RF	This study

Table S3: Primers used in this study.

Name	Sequence 5' → 3'	Used for
CV_egaR(1)_for	CTGCAACTTTTTTCGTTTCGCAA	Construction of <i>egaR</i> mutant
CV_egaR(1)_rev	TTGGTTGTTTTTCAGACGGGCTTACATACAGACCGTCTCGAAT	
CV_egaR(2)_for	ATTCGAGACGGTCTGTATGTAAGCCCGTCTGAAAACAACCAA	
CV_egaR(2)_rev	ACCCAAGTCATTTTCCATCC	
CV_39	TCTAGCTAATGTTACGTTACACATTAAGTAGAGCTTTCGGCGAACCGGCTTT	
CV_40	AGAATCATAATGGGGAAGGCCATCCAGCCTAACCAATCCCTCGCTCCCAT	
CV_93	TTAAGTGCAGCAAGCAACCGAT	Construction of <i>egaP</i> mutant
CV_94	TGCAATGCCTTTTTGACCAGGAAACCCTTTCACACTTCGAATCGC	
CV_95	GCGATTTCGAAGTGTGAAAGGGTTTCTGGTCAAAAAAGCATTGCA	
CV_96	AGGAATCCCTTGAGGAGCAAATTG	
CV_117	TCTAGCTAATGTTACGTTACACATTAAGTAAAGTTCCGAAAGAACTCTGG	
CV_118	AGAATCATAATGGGGAAGGCCATCCAGCCAGAAAGACAGCCTTTCATTAA	
efl158_for	AGGTCCTATCTTGGCTTGCC	Construction of <i>egaG</i> mutant
efl158_mut_rev	TCTCTGCTAGCTTTTAGTAACCCATCATG	
efl158_mut_for_bis	GCGTATCTAGAAATGTGGAATTTACTTTTAC	
efl158_rev	GACTTAACCACTGTTGGTGGC	
pMAD_1158_for	AAAACGGATCCGTTTTTGGATTTACTCCAG	
pMAD_1158_rev	AACGAGTCGACATTAATGCGACAATCATTCC	
CV_67	GTGTTATCATTGGCCGTGCG	Construction of <i>egaB</i>
CV_68	TGCTTAATTAACCGCAGCGCACTCATGTTTCCTTCACTCTCCTC	
CV_69	GAGGAGAGTGAAGGAACATGAGTGCGCTGCGGTAAATAAGCA	
CV_70	AAAGAACCGCATCGGCATGA	

CV_119	TCTAGCTAATGTTACGTTACACATTA ACTCACGTATGA GAAGAAGCCAAT	
CV_120	AGAATCATAATGGGGAAGGCCATCCAGCCAGACAATG CCCGCAAAACAGG	
CV_43	TTAACGAAAGGGCGGATGGA	Construc tion of <i>egaC</i>
CV_44	ACTCTCTTCTTCTTGTGCTTTCCAAGACATCATTGTCG CCCAGT	
CV_45	ACTGGGCGACAATGATGTCTTGGAAGCACAAGAAG AAGAGAGT	
CV_46	TTGGTGCTTTTGTGCATACCG	
CV_47	GTGAGTGGAATTCCTGCAGCCCCATACCACAGAGATT CAGCCG	
CV_48	TCTAGA ACTAGTGGATCCCCCTGCTTTTGCCACGACC AATTG	
CV117	CAAATAAATTATAAGGAGGCACTCAAATGACAGAA AATGCACCCAAAT	
CV118	GAAGCGCTGTGACGAGCTCGCATGCGAATCATTTCGT TCCAAGAACCTCC	
CV119	CAAATAAATTATAAGGAGGCACTCAAATGAAGATC AAAGAAGAAATAG	Cloning of <i>egaP</i> in pMSP35 35
CV120	GAAGCGCTGTGACGAGCTCGCATGCGAATTATGCTT CTGTTAATGCAAT	
CV123	CAAATAAATTATAAGGAGGCACTCAAATGAGTAAA AAACAAATTTATC	Cloning of <i>egaB</i> in pMSP35 35
CV124	GAAGCGCTGTGACGAGCTCGCATGCGAATTATTTTTT ATTTTGCTTAAT	
CV125	CAAATAAATTATAAGGAGGCACTCAAAGTGCAAAAA TTTTTGATTGGT	Cloning of <i>egaC</i> in pMSP35 35
CV126	GAAGCGCTGTGACGAGCTCGCATGCGAATTATAAAT CATCGAAGTCAA	
P3535 3'	CTCGAGTCTAGAAGCGCTGTC	Verificati on of insertion into pMSP35 35
P3535 5'	TTAGATACAATGATTTCGTTTCG	
CV_62	TTTGTTTAACTTTAAGAAGGAGATATACATATGACAGA AAATGCACCCAA	Cloning of <i>egaR</i> in pET29b+
CV_63	AGCCGGATCTCAGTGGTGGTGGTGGTGGTGGTTCGTTTC CAAGAACCTCCTA	
OCG_232	CACCACCACCACCACCAC	
OCG_233	CATATGTATATCTCCTTCTTAAAGTTAAACAAA	
OCG_240	CGCGAAATTAATACGACTCACTATA	Verificati on of insertion
OCG_241	GGGGTTATGCTAGTTATTGCTCA	

		into pET29b+
pQE_1158for	GCGGAGGATCCATGACTTGGGGTGCAATTGC	Cloning of <i>egaG</i> in pQE30
pQE_1158rev	ACTCTAAGCTTTTATTCAATTGGTGCTTTGATTC	
pQE_TypeIII/IV	CGGATAACAATTTACACACAG	Verification of insertion into pQE30
pQE_rev	GTTCTGAGGTCATTACTGG	
CV_56	TGAAGCCTGTCCGCTACTCT	Cloning of P1 in pGEM-T
CV_57	GTGCATTTTCTGTCATTTTCATCGC	
CV_58	TGCAAGCAATGATTGCATTAAC	Cloning of P2 in pGEM-T
CV_59	TTTAGTAACCCATCATGTGCC	
pU	TGTA AACGACGGCCAGT	Verification of insertion into pMAD and pGEM-T
pR	CAGGAAACAGCTATGACC	

Table S4: Primers used for RT-qPCR and 5'-RACE-PCR.

Name	Sequence 5' → 3'	Gene / Used for
<u>RT-qPCR</u>		
RT CV 1	AAGTGGTTGAAGCAGAGAAG	<i>egaB</i>
RT CV 2	GGAATCCCTTGAGGAGCAA	
RT CV 3	AAGGAGTGCCACCAACAGTGG	<i>egaC</i>
RT CV 4	AAGGAGTTAAAGCAGCCAGCCA	
RT CV OG1RF 9	GCAAGGAGCGGAAATTGTTGC	<i>egaD1</i>
RT CV OG1RF 10	CCGCATCGGCATGAAACTCCAA	
RT CV 7	GGCTATGCGCAATGGGGAACA	<i>egaP</i>
RT CV 8	TGAAAGGGCGGCACTGACCAA	
RT CV 11	CCCGCTGAGAAAATTTCGGGA	<i>egaR</i>
RT CV 12	CAACTGTCATTGGTTCGCCG	
EF1158 RTF	AGAACGCGCTAATCAATGGT	<i>egaG</i>
EF1158 RTR	AACGTAATAGCGCCGACAGT	
<u>RACE PCR</u>		
1158_race1	GCACCTCCTTGTTGTAATTC	Used for sequencing
1158_race2	TTAGGTAAACCGCCGTAGCC	Used during cDNA-tail polyA

		amplification and <i>Nested-PCR</i>
1158_race3	TTGCAAATGTGTCTCCATCC	Used during cDNA synthesis

Table S5: Results of the variant detection analysis of the Δ *egaG* genome. MNV: Multiple Nucleotide Variant; SNV: Single Nucleotide Variant.

Genome Position	Location and annotation	Variant Type	Nucleotide and amino-acid changes	Comments
978560 / 978561	<i>egaG</i> /OG1RF N4-(beta-N-acetylglucosaminy)-L-asparaginase	MNV	ACG to GAA Thr to Glu	This change occurred into the untruncated 3' part of the mutated gene. Didn't change the reading frame.
1460011	<i>apt</i> /OG1RF_11400 adenine phosphoribosyltransferase	SNV	TAC to TAT Tyr to Tyr	No amino-acid change

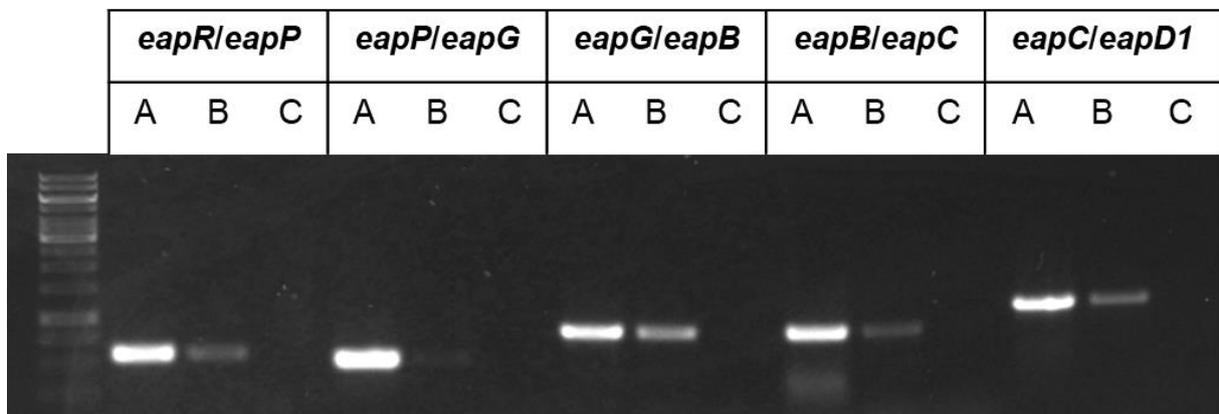


Fig. S1: RT-PCR analysis of the operon's structure with A : Positif control (gDNA) ; B : cDNA samples ; C : Negatif control (DNase-treated RNA).

(white square), **B** : mutated strain strain Δ *egaP* (black square) and the corresponding complemented strain (white square).

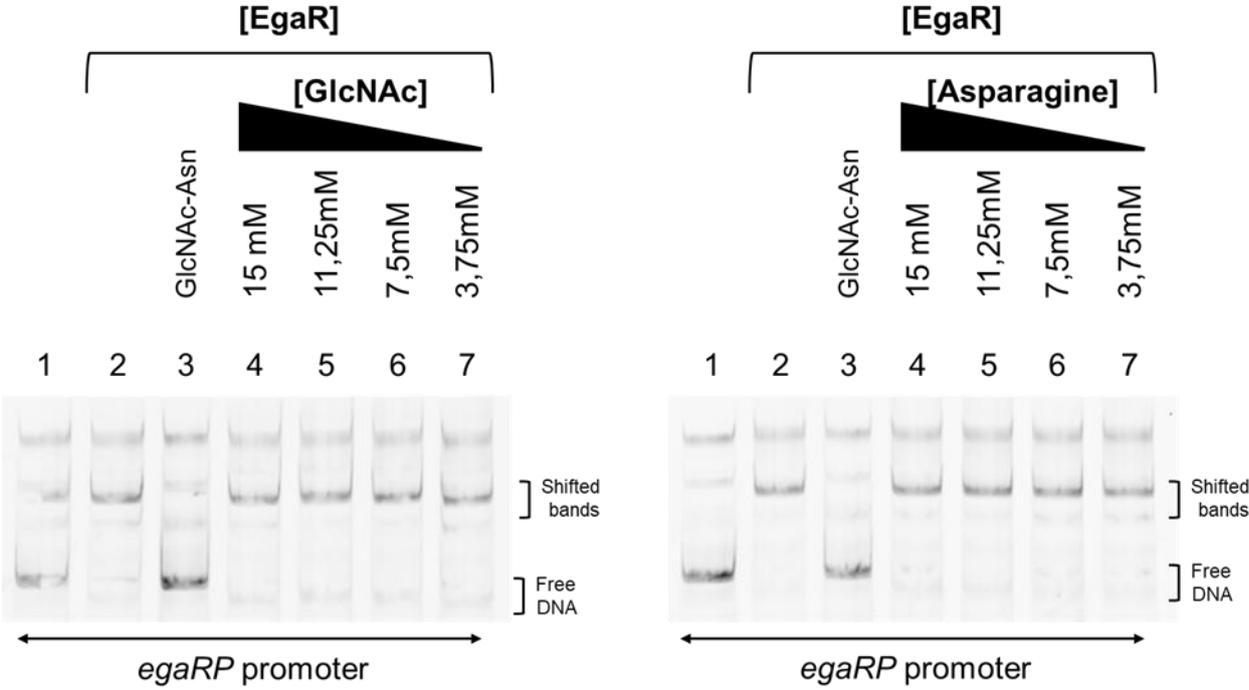


Fig. S4: EMSA in presence of asparagine or GlcNAc. EMSAs were performed with 5 nM of labelled DNA, 250 nM of EgaR and various concentration of asparagine or GlcNAc ranging from 15 mM to 3.75 mM (lanes 4 to 7) and are indicated above the gel. Lane 1 corresponds to DNA alone. Lane 2 corresponds to DNA with EgaR. Lane 3 corresponds to DNA with EgaR and GlcNAc-L-Asn 10 mM.

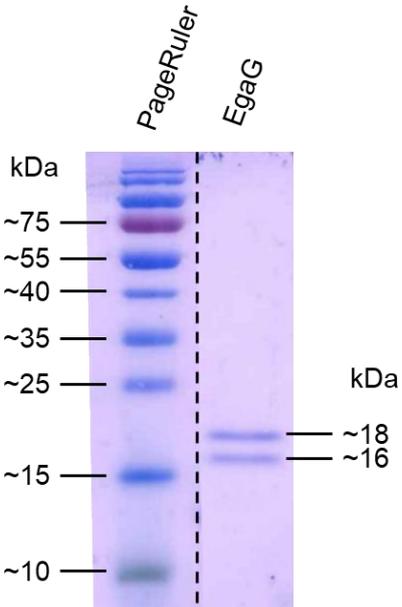


Fig. S5: Protein purification of glycosylasparaginase (EgaG). Two subunits are obtained after purification, indicating the protein undergoes a process of autocleavage. Image was obtained from results on the same gel but not on continuous wells; the vertical broken lines indicate the cropped regions.

Résumé

Caractérisation du locus glycosylasparaginase chez *E. faecalis* et son implication potentielle dans la virulence.

Enterococcus faecalis est une bactérie opportuniste retrouvée dans le tractus gastro-intestinal de l'Homme responsable de multiples infections. Des études menées dans l'unité de recherche Communication Bactérienne et Stratégies Anti-infectieuses (CBSA) de l'université de Caen Normandie ont permis de montrer qu'un certain nombre de gènes impliqués dans l'utilisation des polysaccharides sont induits en conditions d'infection. Dans cette étude, un gène présentant des homologies de séquence avec une glycosylasparaginase a été identifié chez *E. faecalis* OG1RF. Ce gène est situé dans le locus *ega* comportant 7 autres gènes, et est impliqué dans le catabolisme du N-acétylglucosamine L-asparagine (GlcNAc-L-Asn). Ce travail de thèse s'est également intéressé au fonctionnement de ce locus. Il a ainsi été montré qu'il est séparé en deux unités transcriptomiques avec d'une part les gènes *egaRP* et d'autre part les gènes *egaGBCD1D2*. L'expression de ce locus est régulée par le répresseur EgaR qui se fixe directement aux motifs de liaisons 5' - GTMTAKAC - 3' retrouvés au niveau des régions promotrices. Des croissances réalisées avec GlcNAc-L-Asn comme source de carbone ont montré que les gènes *egaGBC* sont nécessaires pour le catabolisme de cet acide aminoglycane. Ce composé est internalisé par le PTS EgaB, EgaC et OG1RF_10750 et dégradé par la glycosylasparaginase EgaG. L'hydrolyse du N-acétylglucosamine L-asparagine génère de l'aspartate et du 1-amino-GlcNAc qui sera utilisé comme source de carbone par *E. faecalis*. Enfin, les tests de virulence en modèle *Galleria mellonella* ont montré une meilleure survie des larves infectées avec les mutants de délétion Δ *egaP* (Peptidase), Δ *egaG*, Δ *egaB* et Δ *egaC* qu'avec la souche sauvage, indiquant un rôle dans la virulence des produits de ce locus.

Mots clés : *Enterococcus*, métabolisme, glycosylasparaginase, N-acétylglucosamine L-asparagine, virulence

Abstract

Characterization of the glycosylasparaginase locus in *E. faecalis* and its potential involvement in virulence.

Enterococcus faecalis is an opportunistic bacterium found in the human gastrointestinal tract and responsible for a wide range of infections. A study conducted by the research unit "Communication Bactérienne et Stratégies Anti-infectieuses (CBSA) of the Caen Normandy university, showed that a number of genes involved in polysaccharide utilization are overexpressed under conditions of infection. In this work, we characterized a gene in *E. faecalis* OG1RF showing sequence homologies with a glycosylasparaginase. This gene is located within a locus (*ega*) and is involved in the catabolism of N-acetylglucosamine L-asparagine. Function of this locus was also investigated. This locus is separated into two transcriptomic units, composed of genes *egaRP* and *egaGBCD1D2*, respectively. This locus is regulated by the EgaR repressor which binds directly to the 5' - GTMTAKAC - 3' linker motifs found in the promoter regions. Growth studies with GlcNAc-L-Asn as carbon source showed that the genes *egaGBC* are required for the catabolism of this glycoamino acid. This compound is first internalized by the PTS formed by EgaB, EgaC, OG1RF_10751 and then catabolized by the glycosylasparaginase EgaG. N-acetylglucosamine-L-asparagine hydrolysis generates aspartate and 1-amino-GlcNAc which can be used as fermentable carbon sources by *E. faecalis*. The same compound inhibits DNA/EgaR interaction. Finally, virulence tests in the *Galleria mellonella* model showed better survival of larvae infected with the Δ *egaP* (Peptidase), Δ *egaG*, Δ *egaB* and Δ *egaC* deletion mutants than with the wild-type strain, indicating a role for products of this locus in virulence.

Keywords : *Enterococcus*, metabolism, glycosylasparaginase, N-acetylglucosamine L-asparagine, virulence