



# Analyse de l'impact de mutations dans différentes souches bactériennes apparentées à *Bacillus subtilis* (strain 168)

Martin SCHROBILTGEN

Bloc 3 du cursus Bachelier en Biotechnique  
HEH – Département des Sciences et technologies  
Année académique 2022-2023

Confidentiel

Analyse de l'impact de mutations dans  
différentes souches bactériennes  
apparentées à *Bacillus subtilis* (strain 168)

Martin SCHROBILTGEN

Bloc 3 du cursus Bachelier en Biotechnique  
HEH – Département des Sciences et technologies  
Année académique 2022-2023

**Avertissement**

Les informations contenues dans ce document doivent être considérées comme confidentielles et ne doivent en aucun cas être transmises à des tiers non concernés par son évaluation sans autorisation écrite préalable du représentant de Puratos. En aucun cas, les informations contenues dans ce document ne peuvent être considérées comme un rapport public qui pourrait détruire la nouveauté en cas de demande de brevet.

### **Remerciements**

Je voudrais remercier Monsieur Guillemin, manager de la plateforme bioinformatique de l'entreprise Puratos à Andenne, de la confiance qu'il m'a accordée en me confiant cette tâche. Je remercie également les membres de l'équipe de la R&D Enzymes pour leur gentillesse et leurs encouragements.

Je souhaite également exprimer ma gratitude au corps enseignant de la HEH et de Condorcet pour tout ce qu'ils m'ont appris durant ces trois années de Bachelier. Je remercie tout particulièrement M. Coornaert, M. Haquin et Mme Léonet pour leurs conseils et relecture de ce travail. Un tout grand merci également à M. Branders pour ses conseils et encouragements durant l'année.

## Résumé

La souche 168 de *Bacillus subtilis* est hautement utilisée dans l'industrie notamment agroalimentaire. Ce travail porte sur l'étude bioinformatique sans a priori de 26 gènes atteints de frameshift, perte/gain de codon stop ou perte de codon start apparus sur une des souches de *Bacillus subtilis* précédant celle appelée souche finale. Pour réaliser cette étude, les diverses mutations détectées sur les génomes ont d'abord dû être classifiées selon l'impact potentiel des mutations sur l'activité des gènes. Ensuite, l'activité de ces gènes a été analysée. Enfin, l'impact sur la physiologie de l'organisme a été étudiée. Pour ce faire, les réactions et pathways impactés ont été analysés. Pour finir ce travail, qui n'est que la première brique de la compréhension des causes des problèmes rencontrés par l'utilisation de la souche finale, des arguments en faveur de la création d'une nouvelle souche ont été avancés.

## Abstract

*Bacillus subtilis* (strain 168) is highly used in industry, in particular in the food industry. This work focuses on the bioinformatics study without a priori of 26 genes affected by frameshift, loss/gain of stop codon or loss of start codon which appeared on one of the strains of *Bacillus subtilis* preceding the strain called final strain. To carry out this study, the various mutations detected on the genomes were first classified according to the potential impact of these mutations on gene activity. Gene activity was then analyzed. Finally, the impact on the physiology of the organism was studied. To do so, the different impacted reactions and pathways were analyzed. To finish this work, which is only the first step towards the understanding of the causes of the problems encountered in the use of the final strain, some arguments in favour of creating a new strain were put forward.

## Table des matières

Lexique .....	1
Préambule .....	2
Contexte .....	3
L'entreprise Puratos.....	3
<i>Bacillus subtilis</i> .....	3
Ses intérêts en industrie .....	3
Intérêt de l'utilisation des enzymes en production du pain .....	4
Historique des souches.....	4
Pourquoi effectuer ce travail ? .....	6
Travail effectué .....	6
Matériel et méthode .....	8
Résultats et interprétations.....	11
Frameshifts .....	11
Frameshift et perte codon start .....	28
Gain codon stop.....	29
Perte de codon stop .....	35
Récapitulatif.....	36
Conclusion .....	40
Bibliographie.....	41
Annexes .....	44

## Table des figures

Figure 1 Logo de l'entreprise .....	3
Figure 2 Historique des souches internes à Puratos .....	5
Figure 3 Les 10 premières lignes du fichier mutation à haut risque .....	7
Figure 4 Les 10 premières lignes du fichier des mutations à risque variable .....	7
Figure 5 Les 10 premières lignes du fichier des mutations à risque paisible .....	7
Figure 6 Workflow du travail .....	10
Figure 7 Réaction catalysée par UDP-glucose:polyglycerol phosphate glucosyltransferase. Source BsubCyc .....	12
Figure 8 Pathway de synthèse de l'acide téichoïque. Source BsubCyc .....	13
Figure 9 Zoom sur la zone de réaction au sein du pathway de biosynthèse de l'acide téichoïque .....	14
Figure 10 Réaction catalysée par N-acetyl-S-(2-succino)-L-cysteine monooxygenase (EC number 1.14.13). Source BsubCyc .....	14
Figure 11 Pathway de dégradation du « S-(2-succinyl)-L-cysteine » en « L-cystéine ». Source BsubCyc .....	15
Figure 12 Réaction catalysée par thréonine synthétase. Source BsubCyc .....	15
Figure 13 Pathway de synthèse de la L-thérionine. Source BsubCyc .....	16
Figure 14 Pathway de synthèse de la L-isoleucine. Source BsubCyc .....	17

Figure 15 Réaction d'hydrolyse effectuée par l'ATPase synthétisé par l'ATP. Source Rhea	18
Figure 16 ATPase spoIVA prédite et récupérée sur le site AlphaFold	18
Figure 17 Protéine prédite et récupérée via AlphaFold	19
Figure 18 Analyse interPro de la séquence du gène ypgR. Réalisé sur EMBL-EBI	21
Figure 19 réaction catalysée par « unsaturated rhamnogalacturonyl hydrolase ». Source BsubCyc	21
Figure 20 Pathway de dégradation du rhamnogalacturonan. Source BsubCyc	21
Figure 21 Réaction catalysée par la Putative arginine/ornithine antiporter. Source Rhea	22
Figure 22 Prédiction du repliement de la protéine hypothétique yqph effectuée via AlphaFold. Source Uniprot	23
Figure 23 Réaction de synthèse du L-glutamate, catalysée par le glutamate synthase. Source BsubCyc	23
Figure 24 Cycle d'assimilation de l'ammoniac. Source BsubCyc	24
Figure 25 Protéine prédite via AlphaFold. Source UniProt	25
Figure 26 Réaction catalysée par l'histidine kinase (EC number 2.7.13.3). Source BRENDA	25
Figure 27 Protéine prédite via AlphaFold. Source Uniprot	26
Figure 28 Réaction catalysée par la « acyl-CoA dehydrogenase ». Source BsubCyc	27
Figure 29 Pathway de $\beta$ -oxydation des acides gras. Source BsubCyc	27
Figure 30 Réaction catalysée par la « secreted alkaliphilic lipase ». Source BsubCyc	27
Figure 31 Protéine prédite via AlphaFold. Source UniProt	29
Figure 32 Protéine prédite via AlphaFold. Source UniProt	30
Figure 33 Réaction catalysée par la mannan endo-1,4-beta-mannosidase EC number 3.2.1.78. Source BsubCyc	30
Figure 34 Pathway du fructose et mannose. Source BRENDA	31
Figure 35 Zoom et mise en évidence de la voie secondaire dans le pathway du fructose et mannose. Source BRENDA	32
Figure 36 Protéine prédite via AlphaFold. Source UniProt	33
Figure 37 Réaction catalysée par la extracellular neutral metalloprotease. Source BsubCyc	33
Figure 38 Réaction catalysée par la (S)-benzoin forming benzil reductase EC 1.1.320. Source BsubCyc	34
Figure 39 Réaction catalysée par la (S)-benzoin forming benzil reductase EC 1.1.320. Source BsubCyc	34
Figure 40 Réaction catalysée par la « 6-phospho-alpha-glucosidase" EC 3.2.1.122. Source BsubCyc	36
Figure 41 Comparaison du gène malA muté prédit via AlphaFold en bleu et sauvage prédit et récupéré sur AlphaFold en rouge. Comparaison effectuée avec Pymol	36

## Table des tableaux

Tableau 1 Récapitulatif de la méthodologie utilisée pour réaliser ce travail	9
Tableau 2 Récapitulatif des gènes	37

## Table des Annexes

Annexe I Tableau de données des gènes	44
Annexe II Tableau de similitude à 90% pour le gène ypgR. Source NCBI	45



## Lexique

**Pathway** : une série d'interactions et réactions cellulaires définissant une partie du métabolisme / catabolisme.

**Codon** : séquence de trois nucléotides sur un ARN messager qui, au cours de la traduction, seront traduits en l'acide aminé correspondant.

**Frameshift** : mutation causée par une insertion ou une délétion non multiple de trois induisant un décalage au sein des nucléotides altérant ainsi tous les acides aminés traduits après.

**Missense** : mutation au cours de laquelle un nucléotide est modifié suite au changement d'une seule base, ce qui entraîne la transformation de l'acide aminé après l'étape de traduction.

## Préambule

L'unité de production de l'entreprise Puratos a rencontré des problèmes dans l'utilisation de la souche interne créée à partir de *Bacillus subtilis strain 168*. Il était donc nécessaire de déterminer l'origine de ces difficultés. Les recherches effectuées ont permis de retracer l'historique génétique des mutations intervenues dans les différentes souches de *Bacillus subtilis 168* ayant servi à créer la souche actuellement utilisée. La méthodologie mise en place lors du stage a permis de mettre en évidence une série de gènes mutés. Les gènes mutés les plus significatifs, leur expression et les pathways impactés par ces gènes seront analysés dans ce travail qui fait suite au stage au sein de l'entreprise Puratos et au rapport de stage qui en découle.

## Contexte

### L'entreprise Puratos

PURATOS est une société familiale qui a été fondée à Bruxelles en 1919 par M. Groot dans le but de créer, produire et vendre des matières premières destinées aux boulangers et pâtisseries.



Figure 1 Logo de l'entreprise

Actuellement, PURATOS est un grand groupe agroalimentaire international avec plus de 9.500 employés et plus de 1.100 chercheurs en recherche et développement et conseillers techniques. PURATOS dispose actuellement de filiales dans 83 pays, 72 sites de production dans 50 pays et 94 centres d'innovation dans le monde.

PURATOS fabrique notamment des produits de :

- Boulangerie : mixes, levures, matières grasses, émulsifiants, fourrages, enzymes, etc.
- Pâtisserie : mixes, glaçages, sauces, ganaches, caramel, etc.
- Chocolaterie : chocolat et produits dérivés.

L'usine de Seilles a été créée en 1994 dans le but de produire des enzymes. Elle est toujours actuellement l'unique site producteur d'enzymes du groupe PURATOS. L'usine produit notamment des extraits de levure et des levains séchés.

### *Bacillus subtilis*

#### Ses intérêts en industrie

*Bacillus subtilis* est un organisme de référence qui présente de nombreux avantages pour l'industrie notamment alimentaire. D'abord, la culture de cette bactérie est très simple et peut être aisément effectuée à grande échelle. Ensuite, grâce à sa physiologie, il est possible de la cultiver dans de nombreux milieux et dans des conditions variées. (Martinez, 2013). Enfin, son utilisation ayant commencé dans les années 40, les recherches dans lesquelles elle apparaît sont nombreuses et son génome a été totalement cartographié. En outre, son utilisation est idéale car *Bacillus subtilis* est jugée sans risque pour la santé. En effet, *Bacillus subtilis* est reprise dans la liste des micro-organismes qui

disposent du statut QPS (Qualified Presumption of Safety), c'est-à-dire un statut de présomption d'innocuité reconnue. Cette liste est établie par l'autorité européenne de sécurité des aliments (EFSA) qui évalue la sécurité des micro-organismes mis sur le marché sous forme d'additifs destinés à l'alimentation humaine et animale, d'enzymes alimentaires, ... (Zenodo, s.d.) (European Food Safety Authority, s.d.)

Ces caractéristiques en font un sujet idéal pour la bioinformatique et permettent la création de nouvelles souches que les industries utilisent pour augmenter la rentabilité du processus de production industrielle. Elle est surtout un outil efficace d'expression et de production d'enzymes ou de protéines. Ces enzymes sont notamment utilisées dans la fabrication de produits alimentaires (Martinez, 2013) et sont d'une importance capitale pour l'amélioration de farines ou mixes.

### Intérêt de l'utilisation des enzymes en production du pain

Les entreprises agroalimentaires spécialisées dans la production du pain comme Puratos ont recours à de nombreuses enzymes dans leur farine dans le but de produire le meilleur pain possible : un meilleur goût, une mie aérée, une belle couleur, ... mais aussi le plus rapidement possible. Tout cela influence positivement le rendement de la production et ainsi la rentabilité des produits pour l'entreprise. En plus de tous les avantages cités ci-dessus, l'utilisation des enzymes permet de stabiliser la production et de garder ainsi une qualité stable et ce malgré les différences entre les produits utilisés pour la production. Cette stabilité est entièrement due au fait que les processus industriels sont maîtrisés par la domestication des enzymes. Ces enzymes sont utiles à la dégradation des sucres complexes en glucose pour faciliter l'action des levures et les rendre ainsi plus productives en augmentant l'émission de CO<sub>2</sub> pour donner une pâte plus volumineuse avec une quantité plus faible de produit. Elles permettent également de conserver un pain aux caractéristiques constantes en tirant le meilleur des céréales utilisées (BoulangerieNet, s.d.) (Pain suisse, 2018).

### Historique des souches

Puratos, dans le cadre de sa production industrielle, a créé dans les années 80/90 des souches à partir de *Bacillus subtilis strain 168*. Pour arriver à la souche utilisée actuellement appelée dans ce travail souche finale, de nombreuses souches ont créées comme expliqué dans le schéma ci-dessous.

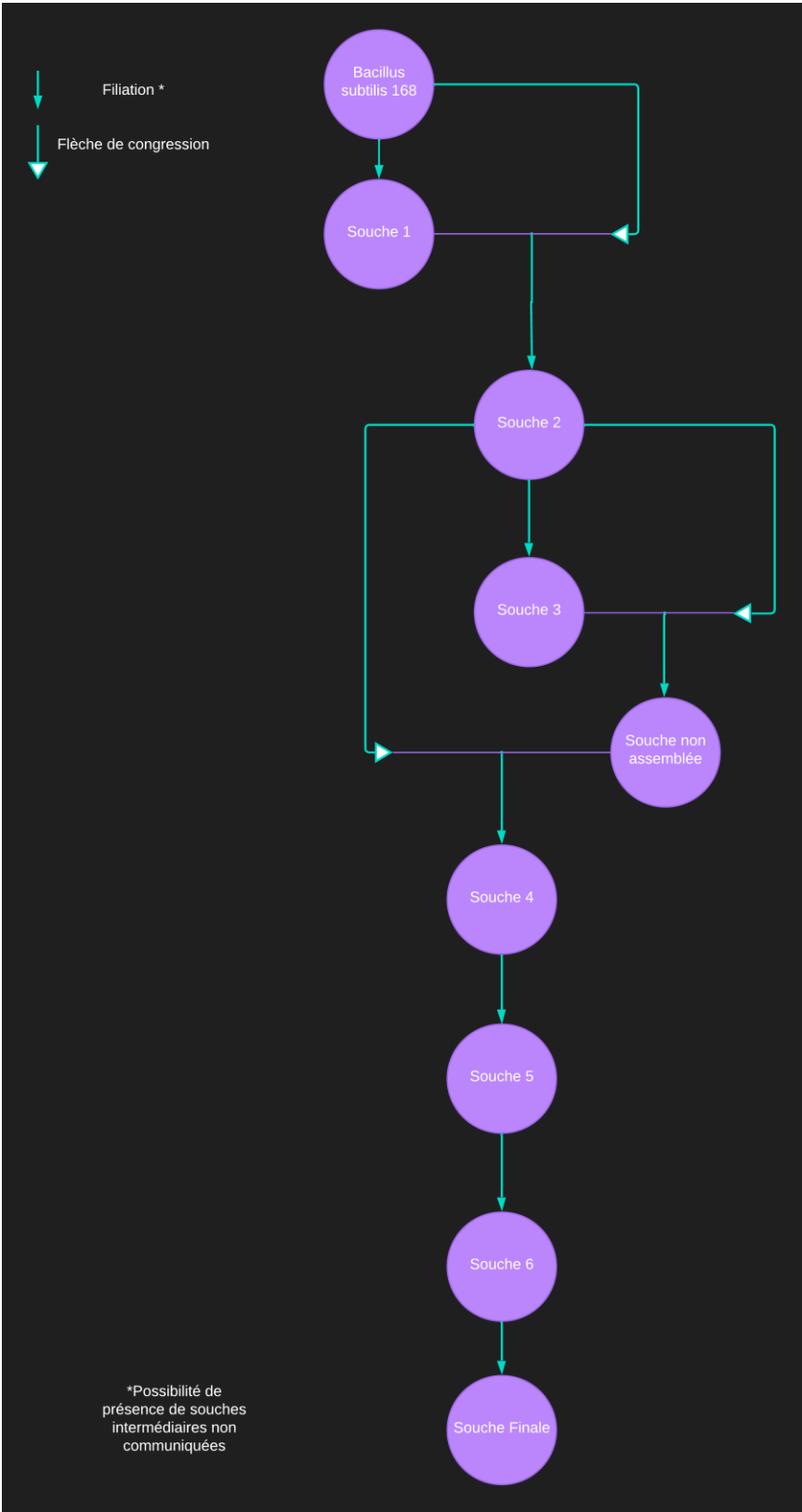


Figure 2 Historique des souches internes à Puratos

Sur le schéma ci-dessus, nous pouvons voir le développement de la souche finale apparentée à *Bacillus subtilis* 168. Au cours de sa création, plusieurs fusions entre souches ont été effectuées. C'est le cas de la souche 1 fusionnée avec *Bacillus subtilis* 168, de la souche 2 avec la souche 3 et de la souche 2 avec une souche intermédiaire non étudiée appelée souche non assemblée. Cette souche n'a pas pu être étudiée à cause d'un problème apparu lors de l'assemblage, ce qui a rendu impossible l'analyse de son génome.

### Pourquoi effectuer ce travail ?

L'unité production de l'entreprise Puratos a rencontré des problèmes lors de l'utilisation de la souche finale *Bacillus subtilis* 168. Il était donc nécessaire d'effectuer un travail de recherche pour tenter de résoudre ces difficultés.

Pour comprendre l'origine de ces problèmes, il était important de cartographier les mutations apparues lors de la création des souches dans le but d'identifier les moyens de palier leur impact.

Ce travail a été réalisé dans une volonté de l'entreprise d'optimiser la production et de comprendre les sources des problèmes nuisant au rendement.

### Travail effectué

Pour identifier l'origine des problèmes et obtenir des résultats fiables, ce travail a été réalisé sans tenir compte des informations découlant de l'utilisation de la souche finale de *Bacillus subtilis* interne à l'entreprise Puratos ni des notes de l'équipe en charge de la création des souches. Les souches ayant été construites dans les années 1980/90, certaines informations sont partielles ou tout simplement manquantes. Les conclusions de cette étude sont exclusivement basées sur des examens bioinformatiques des génomes et ne sont donc qu'une première étape. L'étude se poursuivra par la suite en interne par des études en laboratoire. De plus, les problèmes n'ayant pas été précisés, le travail est donc une analyse de toutes les sources potentielles de problèmes.

Il a d'abord fallu retracer l'historique génétique des souches employées pour la création de la souche finale utilisée dans la production. Pour ce faire, une cartographie des génomes a été effectuée à la recherche des mutations / variations apparues à chaque étape.

Une fois cet historique terminé, les mutations ont été classées en trois groupes en fonction des chances d'impacts sur les gènes. Les mutations à haut risque sont constituées des frameshifts, des pertes / gains de codon stop et des pertes de codon start (Figure 3). Les mutations à risque indéterminé sont quant à elles constituées des missense variants (Figure 4). Enfin, les mutations dites paisibles correspondent aux mutations silencieuses (Figure 5).

```
614655 snp C T stop_gained ydgK
633067 snp G A stop_gained gmuG
835824 snp C T stop_gained yf1N
891369 snp T C stop_lost&splice_region_variant mala
1342073 snp C T stop_gained xkdT
1541526 snp G A stop_gained nprE
3265963 snp G A stop_gained bznD
3513133 snp G A stop_gained sigL
3809891 snp G A stop_gained spo0F
239554 ins A AT frameshift_variant&start_lost ybFF
```

Figure 3 Les 10 premières lignes du fichier mutation à haut risque

```
2771 snp C T missense_variant dnaN
239501 snp T C missense_variant ybFF
241921 snp G A missense_variant ybFH
241927 complex ATT GTA missense_variant ybFH
241938 complex GTTCTCCAA ATTCCCCAT missense_variant ybFH
280370 snp A T missense_variant ycbN
342290 snp G A missense_variant ycgK
467284 snp G T missense_variant mtlR
467950 snp T A missense_variant mtlR
505679 snp C G missense_variant ydbM
```

Figure 4 Les 10 premières lignes du fichier des mutations à risque variable

```
22001 snp C T synonymous_variant serS
234552 snp A C synonymous_variant glpT
239406 snp A G synonymous_variant ybFF
239415 snp C A synonymous_variant ybFF
239445 snp C A synonymous_variant ybFF
239451 snp A G synonymous_variant ybFF
239484 snp C T synonymous_variant ybFF
239650 snp A T synonymous_variant ybFG
239668 snp G A synonymous_variant ybFG
241988 snp A G synonymous_variant ybFH
```

Figure 5 Les 10 premières lignes du fichier des mutations à risque paisible

Une fois les mutations classées, la recherche d'informations sur les gènes (outils utilisés illustrés au Tableau 1) a pu commencer. Cette recherche a pour but de comprendre quelles activités cellulaires ont pu être impactées d'une manière ou d'une autre par l'apparition des modifications de leur génome.

C'est donc la partie la plus importante et qui sera la base des recherches effectuées par d'autres équipes permettant ainsi de confirmer les résultats et hypothèses. Les résultats de cette recherche sont détaillés dans la partie résultats et interprétations.

## Matériel et méthode

L'ensemble des données analysées et résumées en Annexe I, résulte des analyses pratiquées au cours du stage durant lequel la méthodologie a été mise en place et appliquée à chacun des génomes des souches internes à l'entreprise Puratos et dont l'historique est représenté à la Figure 2. Le schéma du déroulement de ce travail de recherche est représenté à la Figure 6.



Tableau 1 Récapitulatif de la méthodologie utilisée pour réaliser ce travail

Stage	Assemblage	Pipeline	FastQC LoRDEC-correct Unicycler Quast
	Annotation	Prokka	
		Abricate	
	Cartographie des mutations	ProgressiveMauve	
		Snippy	
	Analyse des mutations	Artemis	
		BRIG	
AlphaFold et PyMOL			
Gene ontology	Geneontology		
	Revigo		
TFE	Obtention d'informations sur les gènes	Subtiwiki	
		NCBI	
		BsubCyc	
		BRENDA	
		UniProt	
		Google scholar	
		PubMed	

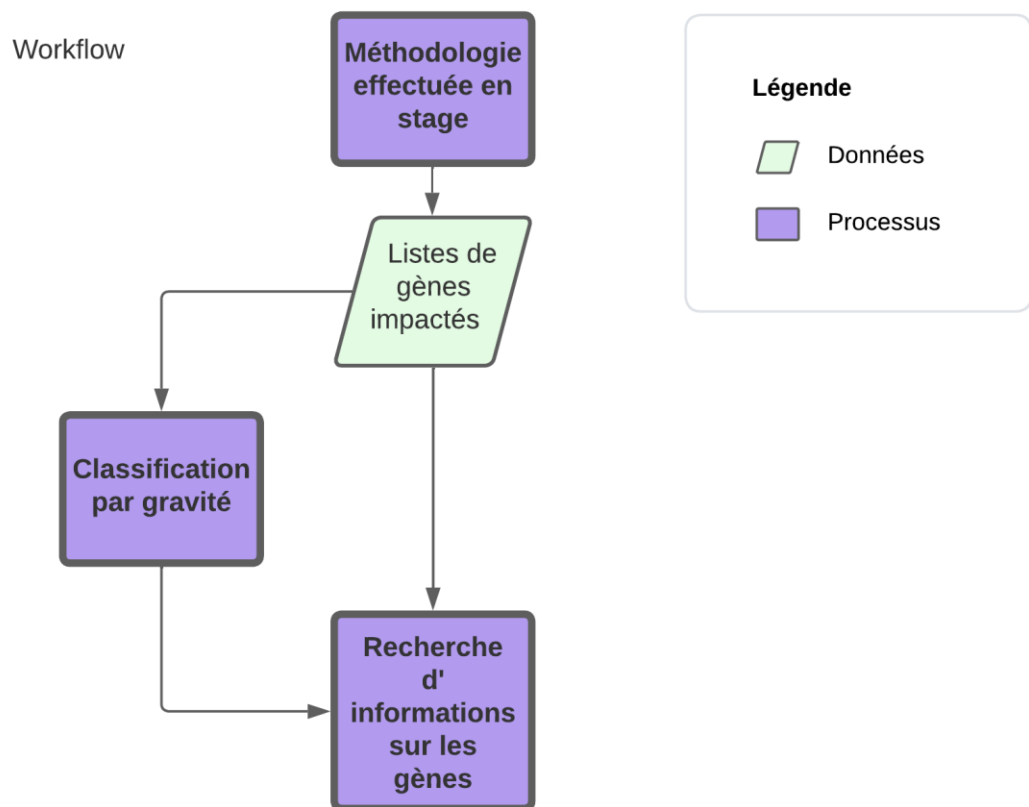


Figure 6 Workflow du travail

## Résultats et interprétations

Au cours des analyses, des mutations sont apparues sur différentes souches apparentées à *Bacillus subtilis* 168 et utilisées en industrie agroalimentaire. Ces mutations ont été conservées de souche en souche jusqu'à la souche finale. Plusieurs gènes se sont avérés impactés d'une manière ou d'une autre.

Le nombre de mutations sur la souche finale s'élève à plus de 800 dont 26 correspondent aux mutations aux risques les plus élevés (frameshift / perte ou gain de codon stop / perte de codon start) et 225 à des missense variants, le reste étant des mutations silencieuses. Le tout sans compter les gènes se trouvant sur les parties délétées dans lesquelles beaucoup de gènes codaient des protéines hypothétiques ou avaient une activité présumée mais non confirmée.

La liste des 26 gènes impactés par les mutations à haut risque et leur rôle sur la souche finale utilisée par cette industrie sont développées ci-dessous.

### Frameshifts

#### Le gène *tagE*

Mutation causée par l'apparition en souche 5 de l'insertion d'une base (T), une thymine.

Le gène *tagE*, aussi appelé *rodD*, *gtaA*, *gtaD*, intervient dans la biosynthèse de l'acide téichoïque via la synthèse d'une transférase : la « UDP-glucose:polyglycerol phosphate glucosyltransferase » (EC Number 2.4.1.52). Cette enzyme va intervenir dans la réaction illustrée ci-dessous qui elle-même intervient dans la voie de biosynthèse de l'acide téichoïque (il joue un rôle dans les parois cellulaires notamment en déterminant la forme que prendra la cellule). En plus de son rôle de glycosyltransférase, le gène joue un rôle dans la résistance aux bactériophages et est une étape essentielle dans la glycosylation de l'acide téichoïque de la paroi de *Bacillus subtilis* 168. L'étude démontre également que sans cette modification de l'acide effectuée par la transférase, la croissance de l'organisme est altérée. Cependant, les cultures de cette souche étant viables, une conclusion s'impose, à savoir que l'altération subie par le gène n'est pas suffisante pour l'inactiver (Allison, D'Elia, Arar, Monteiro, & Brown, 2011).

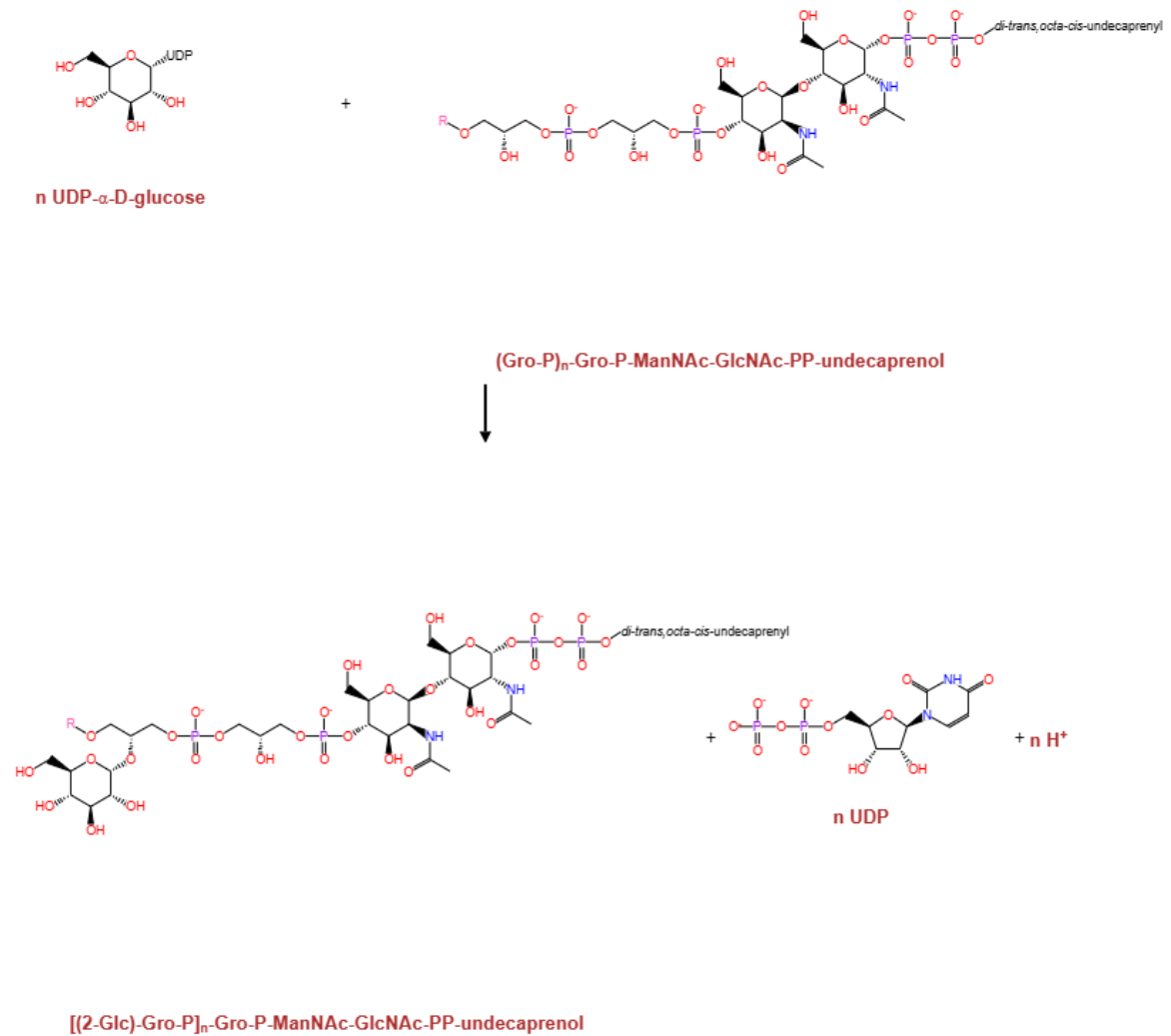


Figure 7 Réaction catalysée par UDP-glucose:polyglycerol phosphate glucosyltransferase. Source BsubCyc

Il s'agit de la réaction du transfert de l' $\alpha$  UDP-glucose à la position deux du polymère catalysé par la glucosyltransférase synthétisée par le gène. Cette réaction est la dernière à se dérouler dans le cytosol et résulte au produit (4-O-{poly[(2R)-2- $\alpha$ -D-glucosyl-1-glycerophospho]-(2R)-glycerophospho}-N-acetyl- $\beta$ -D-mannosaminyl-(1 $\rightarrow$ 4)-N-acetyl- $\alpha$ -D-glucosaminyl-diphospho-ditrans,octacis-undecaprenol) qui, lui, sera transporté vers le milieu extra-cellulaire.

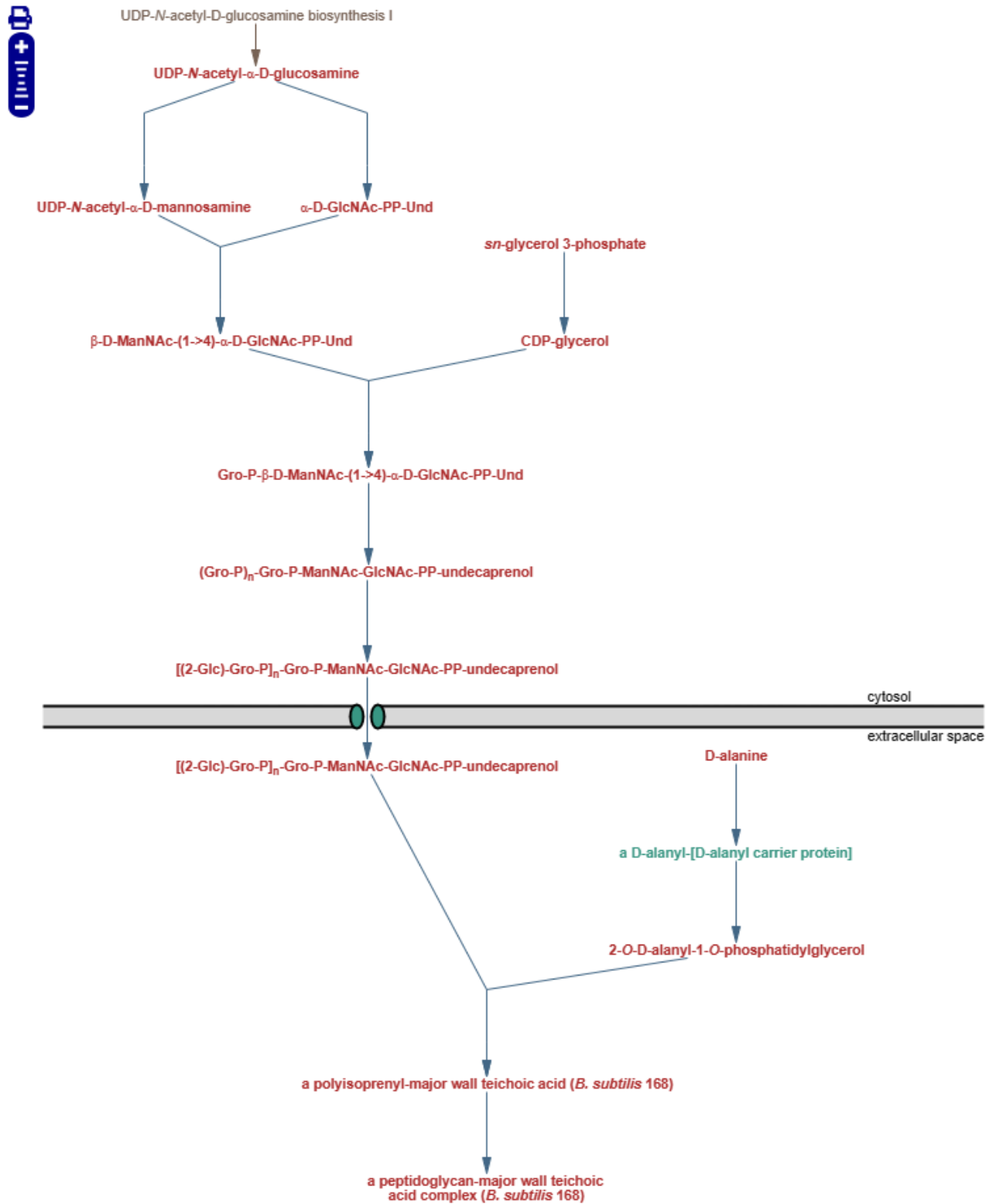


Figure 8 Pathway de synthèse de l'acide téichoïque. Source BsubCyc

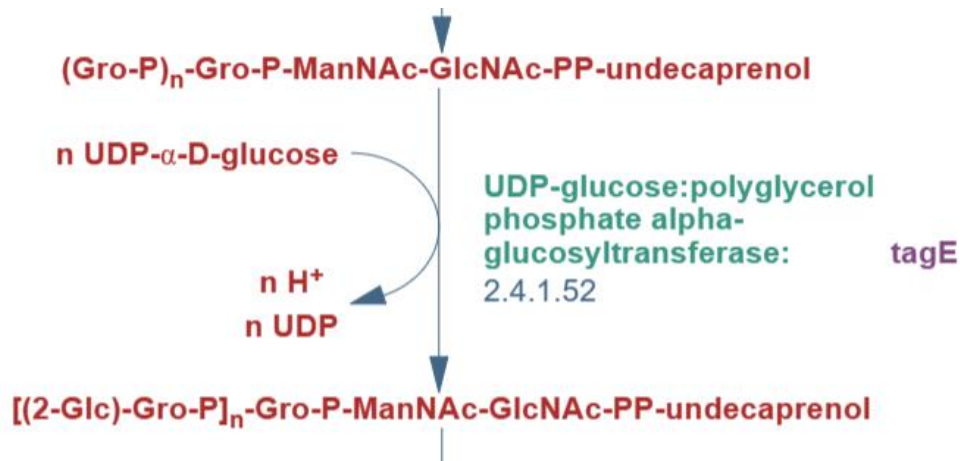


Figure 9 Zoom sur la zone de réaction au sein du pathway de biosynthèse de l'acide téichoïque

### Le gène *yxeK*

Également identifié par scmK, il intervient dans le pathway de dégradation du « S-(2-succinyl)-L-cystéine » en « L-cystéine » en synthétisant l'enzyme « N-acetyl-S-(2-succino)-L-cystéine monooxygénases » (EC number 1.14.13).

Même si la délétion des 13 bases (GAAGCTTCGGCGTC → G) apparue à la souche 3 a conduit à un frameshift pouvant causer l'inactivation du gène, une étude de 2022 a démontré l'existence d'une autre voie basée sur une lyase permettant quand même de synthétiser la cystéine (Hillmann, Goethel, Erickson, & Niehaus, 2022) .

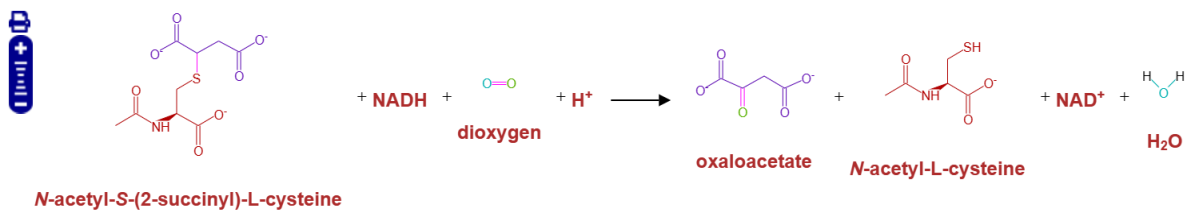


Figure 10 Réaction catalysée par N-acetyl-S-(2-succino)-L-cystéine monooxygénase (EC number 1.14.13). Source BsubCyc

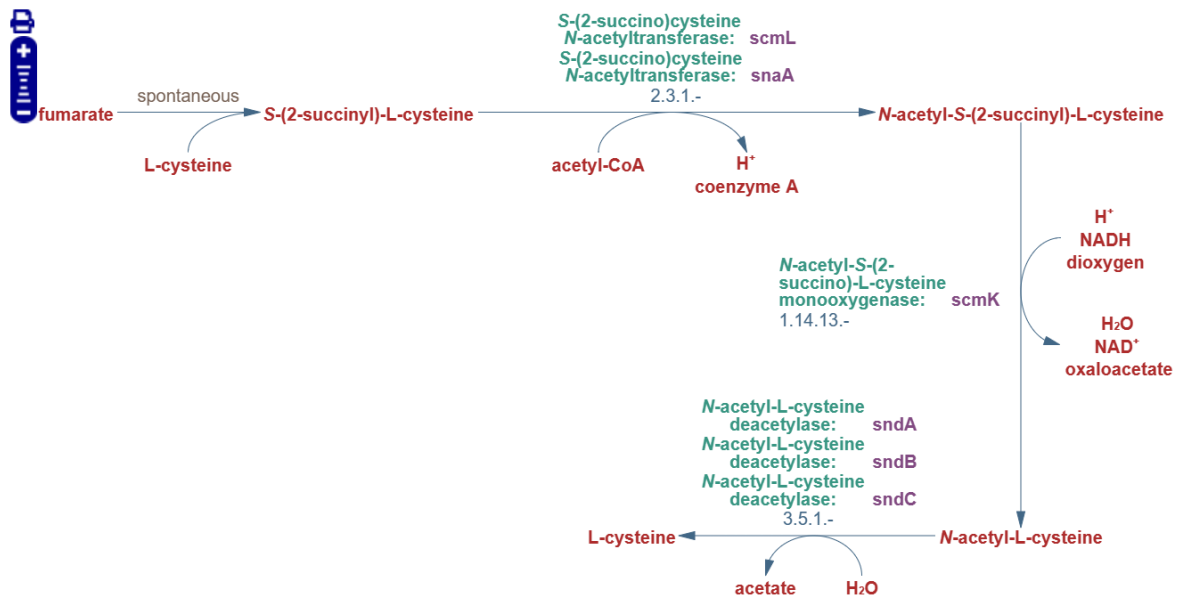


Figure 11 Pathway de dégradation du « S-(2-succinyl)-L-cysteine » en « L-cystéine ». Source BsubCys

Le gène thrC

Mutation causée par une insertion d'une base (C), une cytosine, en souche 1.

C'est le gène responsable de la synthèse de la thréonine synthétase » une enzyme catalysant la synthèse de la « L-théronine » (EC number 4.2.3.1) en suivant la réaction Figure 12. Cette réaction va intervenir dans la voie de biosynthèse de la thréonine Figure 13. Cependant, comme illustré à la Figure 14, ce pathway va lui-même faire partie de la voie de synthèse de la « L-isoleucine »

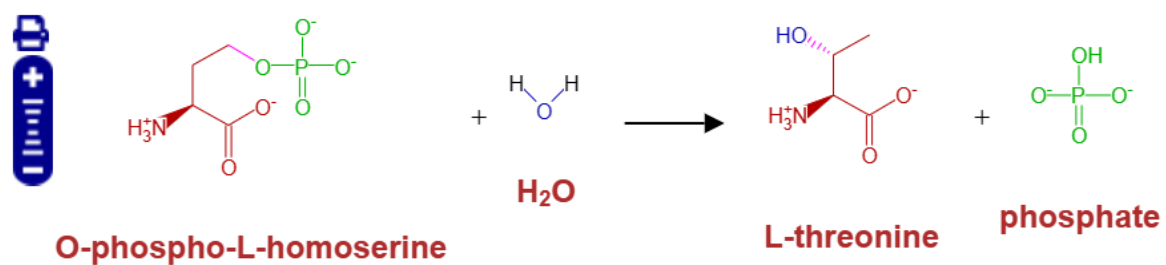


Figure 12 Réaction catalysée par thréonine synthétase. Source BsubCys

C'est par cette réaction d'hydrolyse que la L-thréonine est produite via l'intermédiaire de la thréonine synthétase en dégradant la « O-phospho-L-homoserine » libérant une molécule de phosphate.

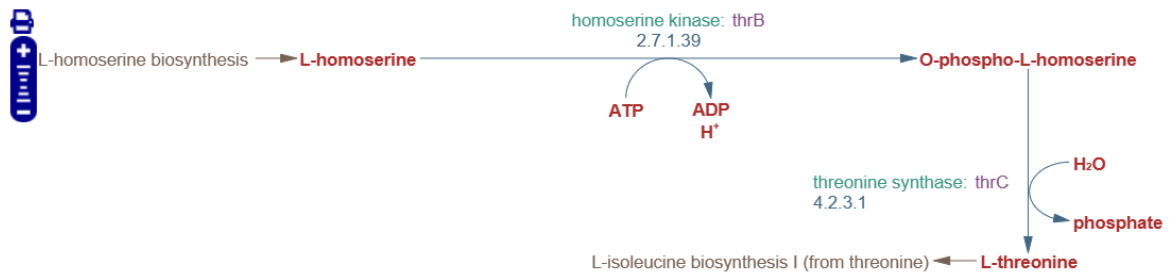


Figure 13 Pathway de synthèse de la L-threonine. Source BsubCyc



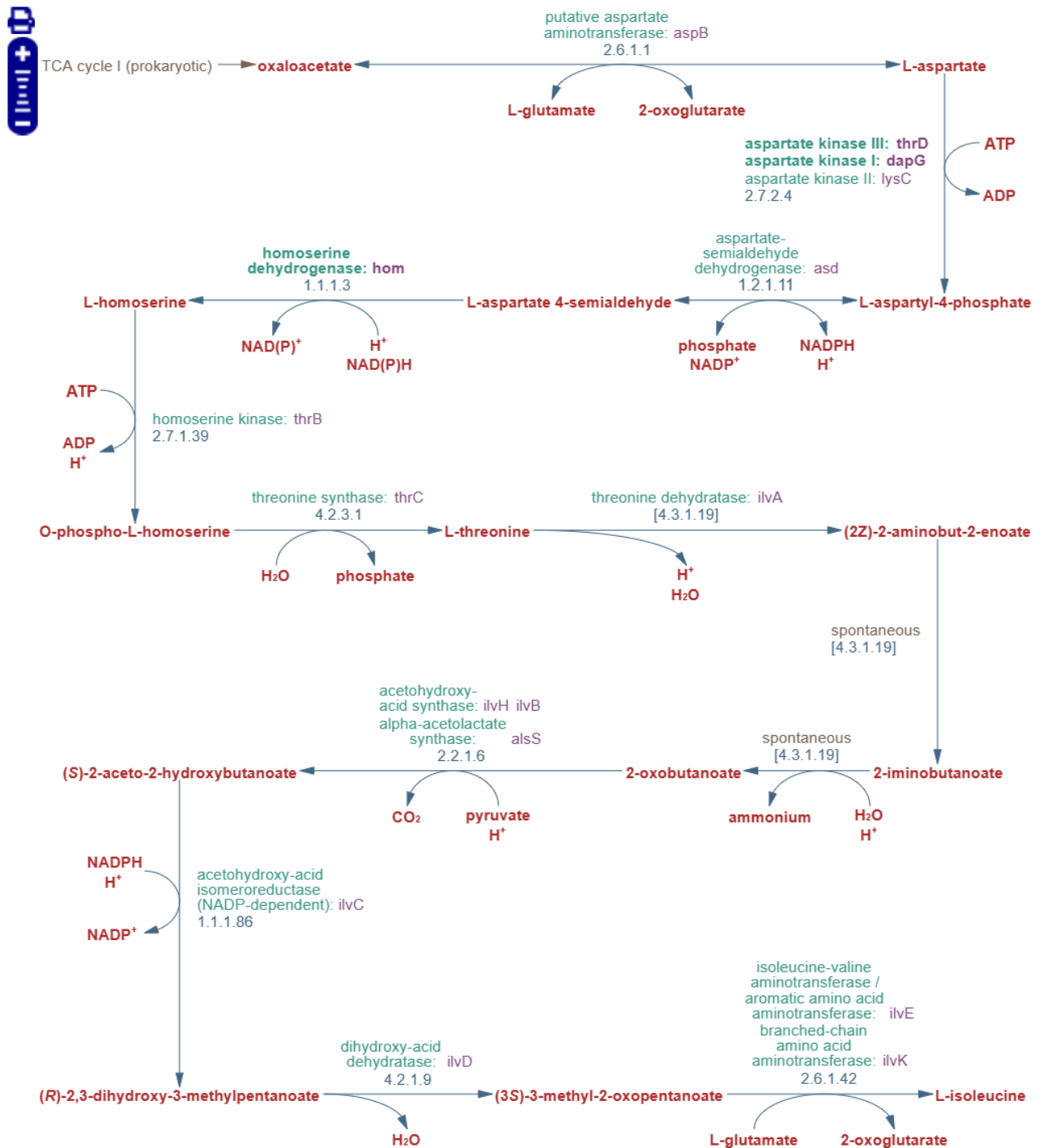


Figure 14 Pathway de synthèse de la L-isoleucine. Source BsubCyc

### Le gène *spoIVA*

Mutation apparue à cause de l'insertion d'une base (T), une thymine, en souche 6.

La protéine *spoIVA* joue un rôle très important dans la sporulation de *Bacillus subtilis* 168. Cette protéine est une ATPase qui va créer une couche sur laquelle va se construire une autre couche de

sporulation appelée manteau. Cette enzyme va également s'autoassembler et non se désassembler en un polymère statique en utilisant l'hydrolyse de l'ATP. Cet assemblage va s'effectuer à cet endroit précis grâce à un marqueur (Castaing, Nagy, Anantharaman, Aravind, & Ramamurthi, 2012).

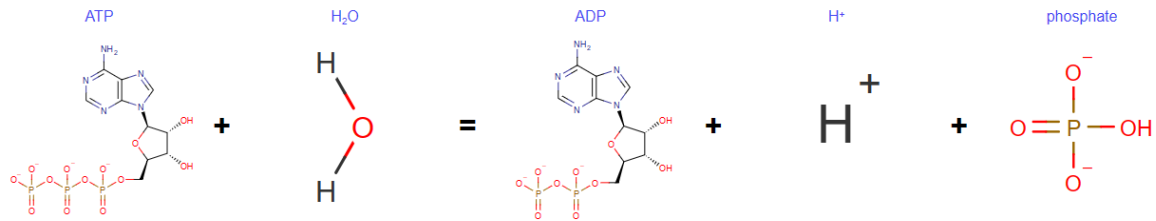


Figure 15 Réaction d'hydrolyse effectuée par l'ATPase synthétisé par l'ATP. Source Rhea

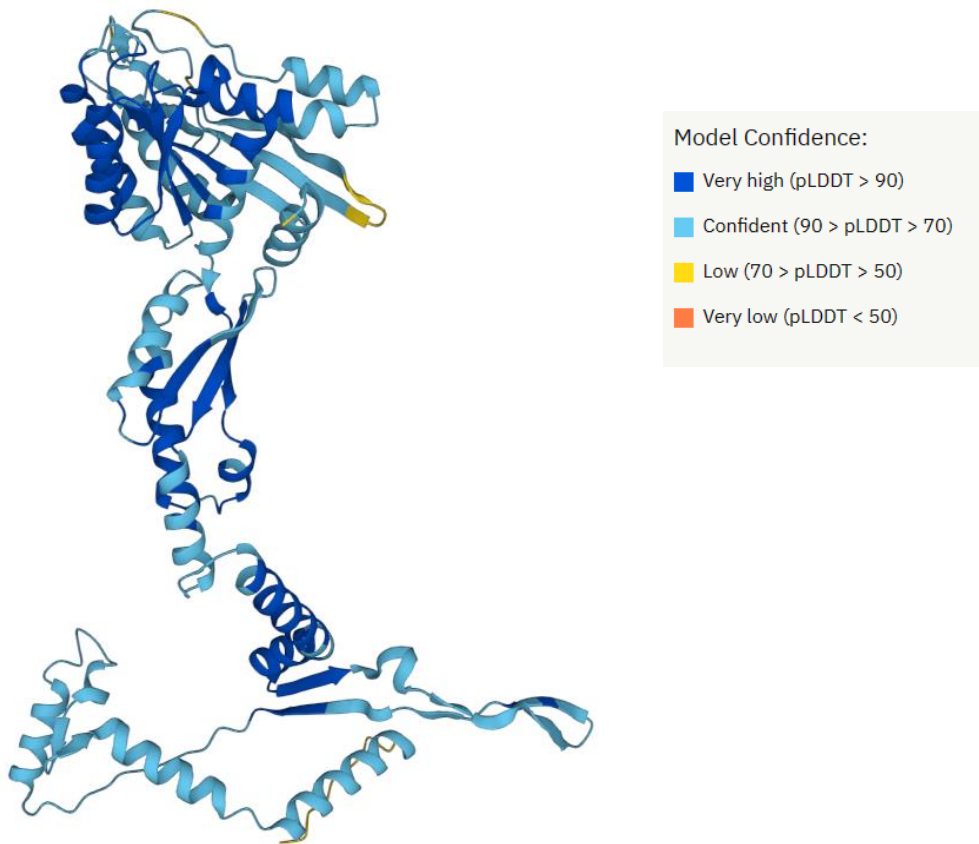


Figure 16 ATPase spoIVA prédite et récupérée sur le site AlphaFold

Le gène ypgR

Mutation apparue en souche 1 à cause de la délétion d'une base (CT → C)

ypgR est une protéine qui n'a pas encore été caractérisée, son activité nous est par conséquent totalement inconnue. Cependant, en comparant ses similarités avec d'autres protéines (Annexe II), on

peut observer que la majorité (16/28) dont la similitude est de 90% sont des « Virulence factor » soit des molécules issues d'une infection virale. Cependant, après une analyse InterPro Figure 18, la « pbs heat like protein » est également une fonction probable au vu de la proportion du gène concerné par cette ressemblance.

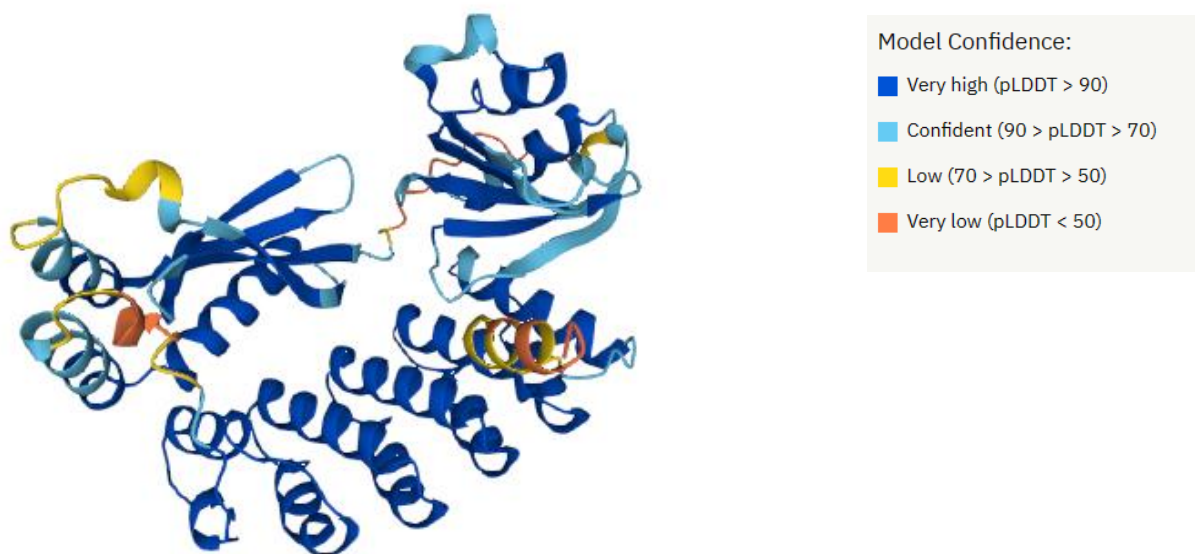


Figure 17 Protéine prédite et récupérée via AlphaFold

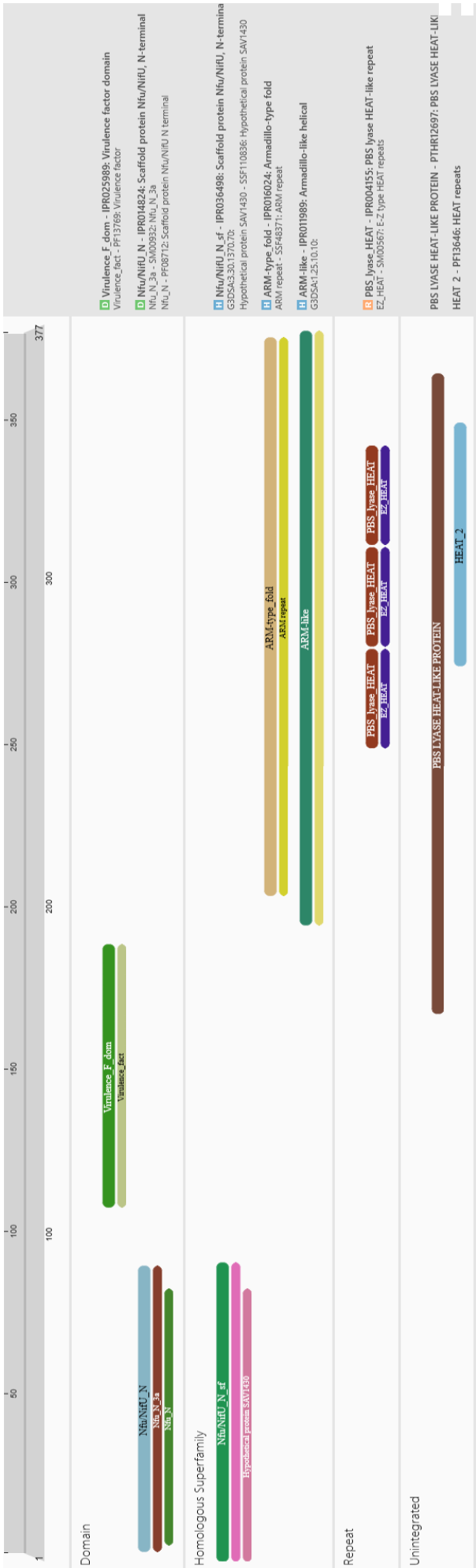


Figure 18 Analyse interPro de la séquence du gène *ypgR*. Réalisé sur EMBL-EBILe gène *rmgQ*

Mutation apparue des suites de la délétion d'une base (GC → G) à la souche 3.

Le gène *rmgQ* code pour une enzyme, une hydrolase : la « unsaturated rhamnogalacturonyl hydrolase ». (EC Number 3.2.1.172 ). Cette enzyme va venir découper la molécule de « 2-O-(4-deoxy-β-L-threo-hex-4-enopyranuronosyl)-α-L-rhamnopyranose » en présence d'eau en deux molécules : la « 4-deoxy-L-threo-hex-4-enopyranuronate » et la « L-rhamnopyranose ».

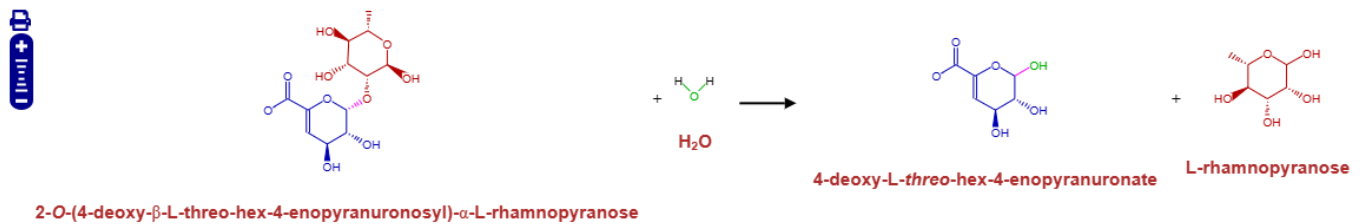


Figure 19 réaction catalysée par « unsaturated rhamnogalacturonyl hydrolase ». Source BsubCyc

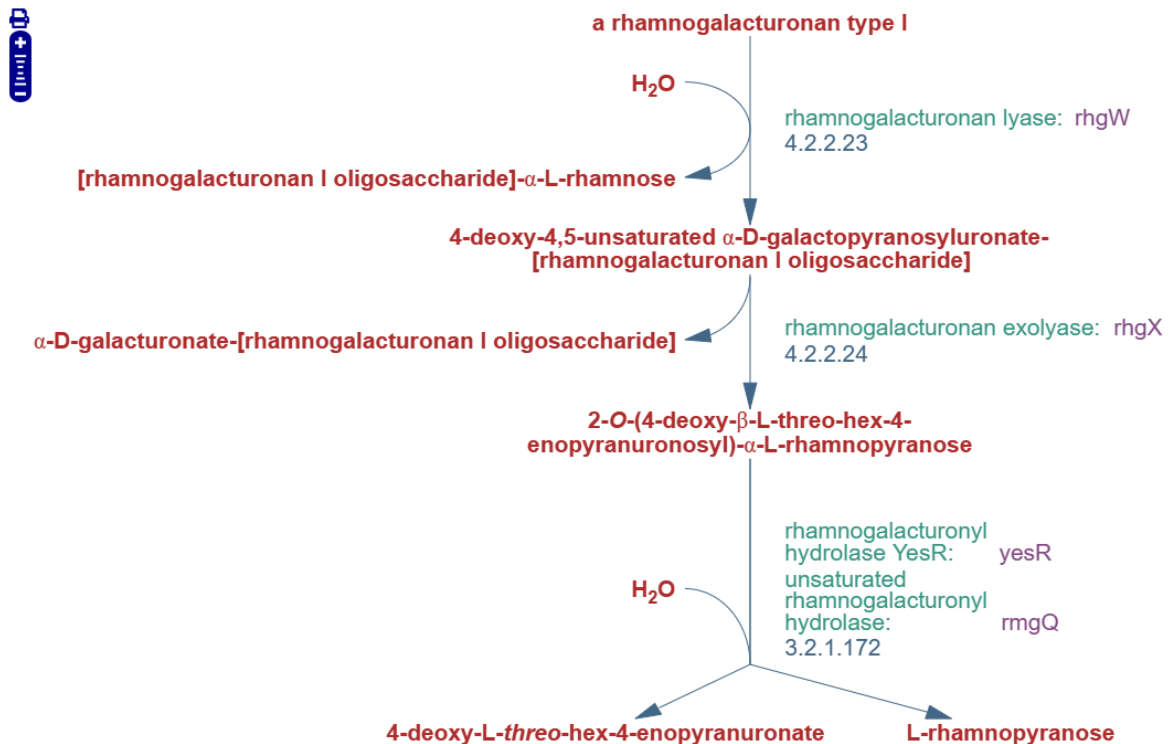


Figure 20 Pathway de dégradation du rhamnogalacturonan. Source BsubCyc

Le gène *Cged*

Mutation causée par l'insertion de deux bases (G → GTA), une thymine et une adénine, en souche 6.

Le gène *cgeD* coderait selon BsubCyc pour une glycosyltransférase putative. Celle-ci est soupçonnée de jouer un rôle dans la maturation de la couche extérieure de la spore de *Bacillus subtilis* 168.

#### Le gène *lysP*

Mutation causée par la délétion d'une base (AT → A) à la souche 6.

LysP, selon plusieurs sources telles que BsubCyc ou NCBI, coderait pour « lysine permease ». Une étude sur le gène chez *Echerichia Coli* tendrait à confirmer ce résultat. (Steffes, Ellis, Wu, & Rosen, 1992) Cependant, ce gène est également appelé *yvsh* pour *Bacillus subtilis* 168 sur UniProt et aurait pour activité la synthèse d'une « Putative arginine/ornithine antiporter » qui catalyserait la réaction ci-dessous par sa similitude de séquence. Toutefois, UniProt met en garde sur le fait que ce gène a également été décrit comme une lysine perméase *lysP* due en partie à la présence d'éléments LYS dans la région régulatrice.

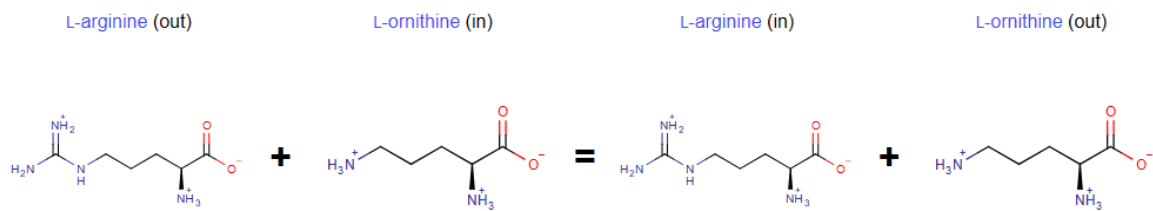


Figure 21 Réaction catalysée par la Putative arginine/ornithine antiporter. Source Rhea

#### Le gène *yqhP*

YqhP a subi une délétion à la souche 4 en perdant une base (GA → G) ce qui a causé le frameshift. Cependant, aucune information sur son activité n'est répertoriée sur Uniprot. Elle y est notamment identifiée comme une protéine non catégorisée.

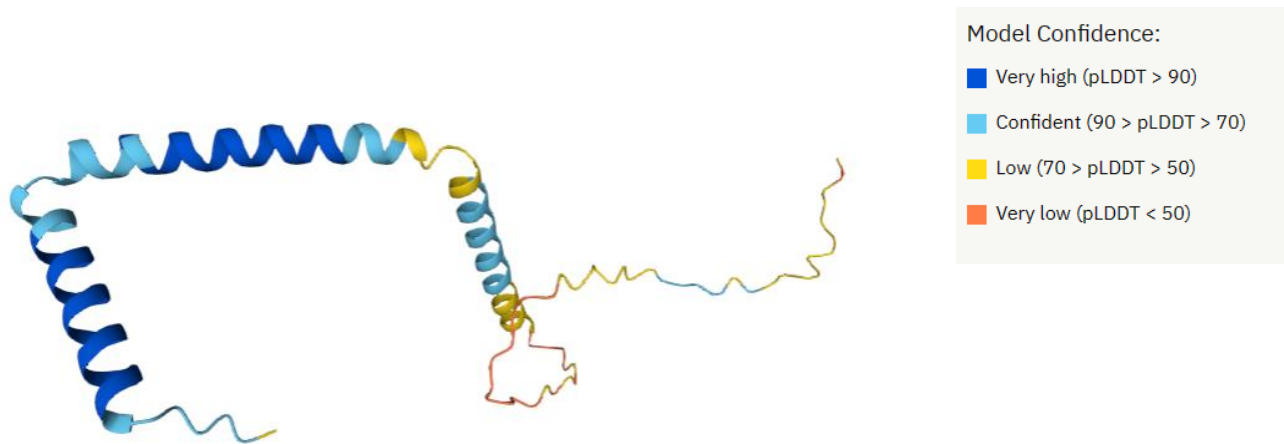


Figure 22 Prédiction du repliement de la protéine hypothétique yqph effectuée via AlphaFold. Source Uniprot

### Le gène *gltA*

Mutation causée par une insertion d'une base (T), une thymine à la souche 5.

Le gène *gltA* permet la synthèse d'une enzyme qui a pour but de catalyser la réaction de synthèse du glutamate. Cette enzyme est la « glutamate synthase (large subunit, NADP-dependent) » (EC Number 1.4.1.13). La transcription de la protéine est effectuée grâce à GltC (Lin, Chen, Chang, Tschen, & Liu, 1999)

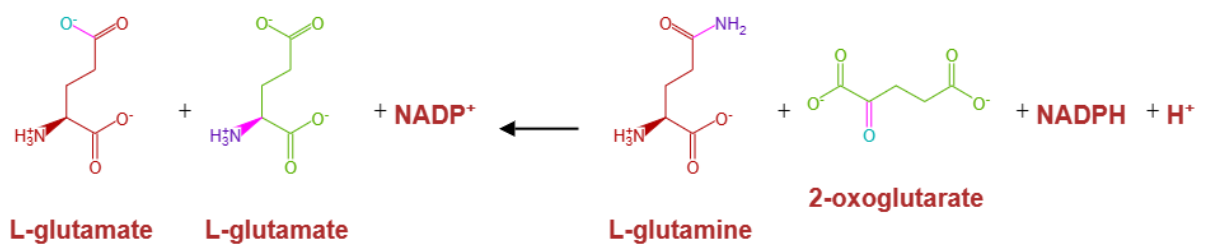


Figure 23 Réaction de synthèse du L-glutamate, catalysée par le glutamate synthase. Source BsubCyc

Comme présenté dans l'équation ci-dessus, la réaction va permettre la synthèse de deux molécules de glutamate, une première à partir de glutamine et une seconde issue de la 2-oxoglutarate. Cette réaction va être effectuée en présence de NADPH qui va être réduit pour obtenir un NADP<sup>+</sup>. Cette réaction et ce gène vont, selon le site BsubCyc, être effectués dans plusieurs pathways comme le cycle d'assimilation de l'ammoniac présenté ci-dessous ou encore de la dégradation de la L-glutamine.

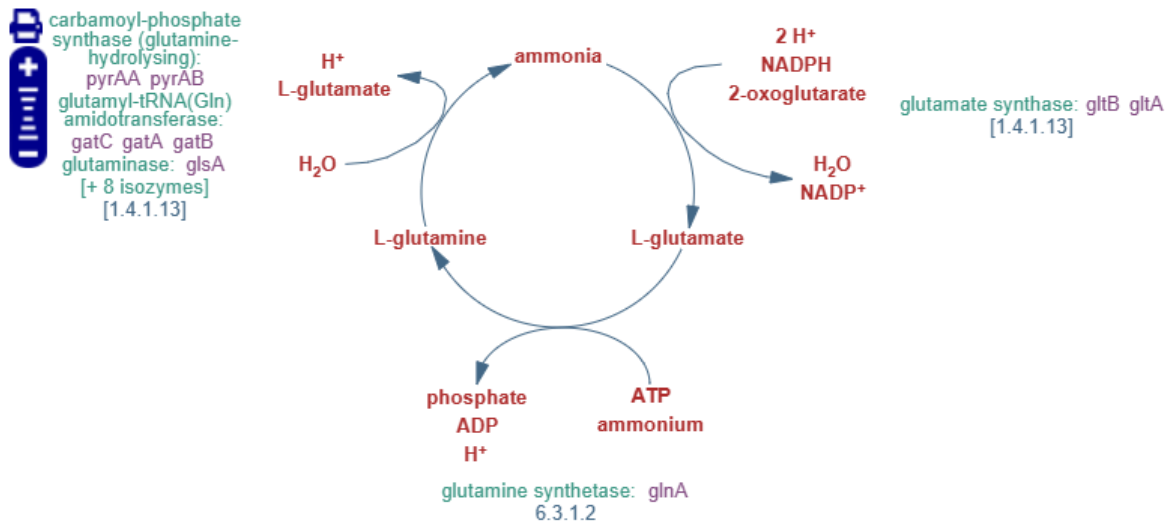


Figure 24 Cycle d'assimilation de l'ammoniac. Source BsubCyc

#### Le gène *ylxY*

Mutation causée par la délétion d'une base (CT → C) en souche 1.

Également appelé *swsB* et *ymxI*, il serait selon le site BsubCys à l'origine de la synthèse d'une enzyme bien que cette fonction ne semble pas encore sûre. Cette enzyme se nomme donc « putative sugar deacetylase ». Cependant, selon une étude réalisée en 2020, ce gène jouerait un rôle clé dans la germination, c'est-à-dire la sortie de la phase de dormance et ce en jouant un rôle dans la dégradation de la couche cortex de la spore de *Bacillus subtilis*. (Amon, et al., 2020). Ces données seront donc à vérifier au cours des prochaines études dans les années à venir.

#### Le gène *sigP*

Mutation causée par la délétion d'une base (CA → C) en souche 1.

Il est également appelé *ylaD*, et va coder pour « Anti-sigma-YlaC factor YlaD ». Ce facteur est ancré dans la membrane de *Bacillus subtilis* 168 et contiendrait un motif HX3CXXC qui lui permettrait d'agir comme un domaine de détection d'oxydoréduction. Il contrôlerait également la transcription du *ylaC*. (Kwak, Ryu, Song, Lee, & Kang, 2018)



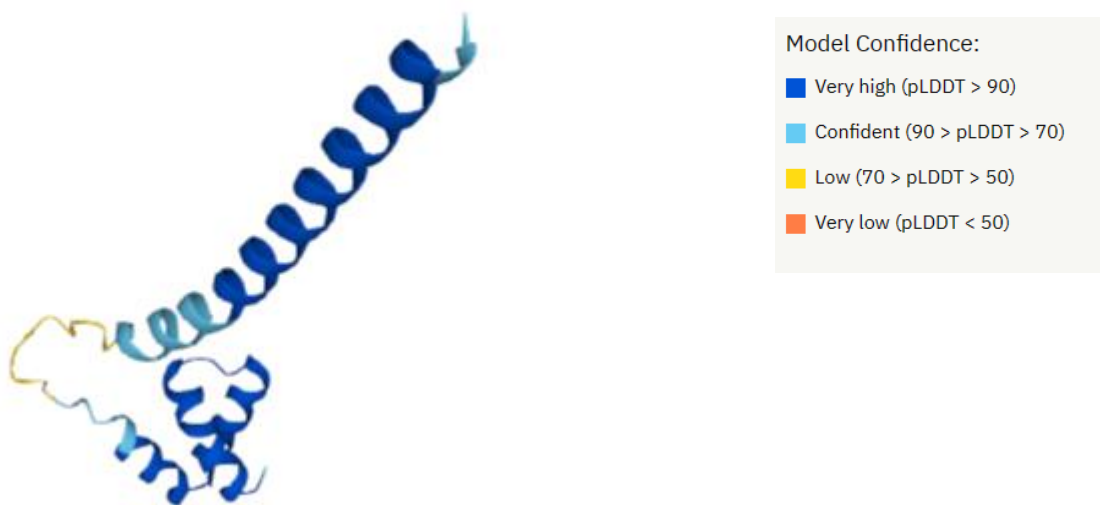


Figure 25 Protéine prédite via AlphaFold. Source UniProt

### Le gène *yhcY*

Mutation causée par la délétion d'une base (CA → C) en souche 2.

Il code pour une protéine jouant le rôle d'histidine kinase (EC number 2.7.13.3). Cette enzyme catalyserait d'ailleurs, selon BRENDA, la réaction ci-dessous au cours de laquelle une molécule de L-histidine est transformée en une « N-phospho-L-histidine » en présence d'ATP. De plus, BsubCyc la décrit comme une « two-component sensor histidine kinase [YhcZ] ». Cela signifierait que cette kinase intervient dans le système à deux composantes et jouerait le rôle de capteur de signal (Ryzhykau, Orekhov, Rulev, & et al, 2021)

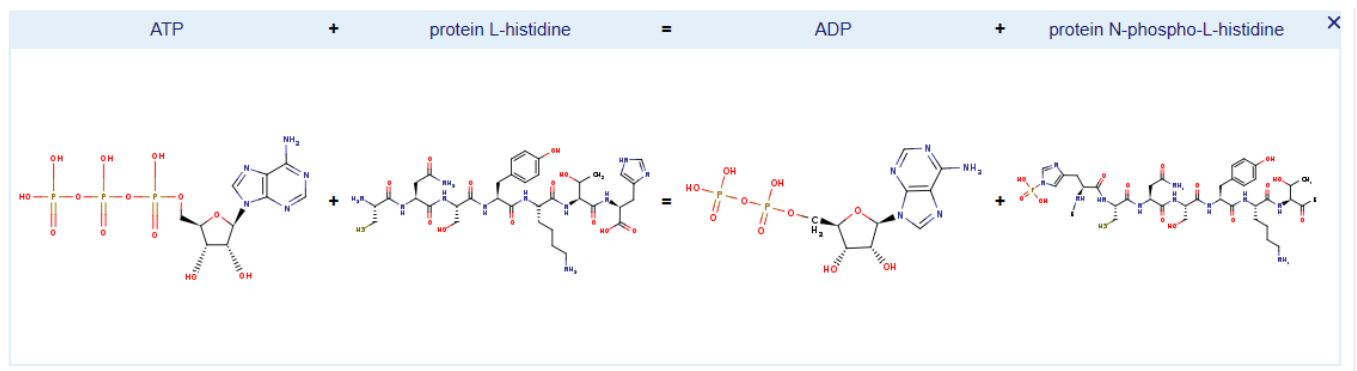


Figure 26 Réaction catalysée par l'histidine kinase (EC number 2.7.13.3). Source BRENDA

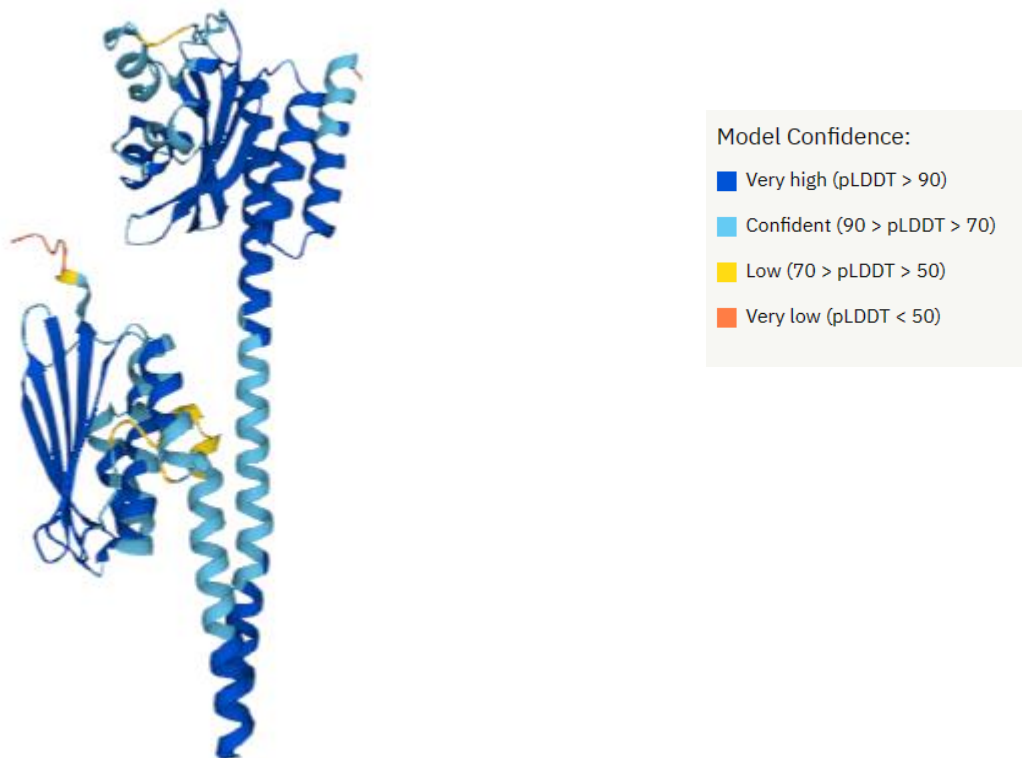


Figure 27 Protéine prédite via AlphaFold. Source Uniprot

### Le gène *ydbM*

Selon Snippy, la mutation considérée comme complexe serait causée par la délétion de deux bases suivies de l'insertion d'une base en souche 1 (AA →G). Cependant, il est hautement probable qu'il n'y ait eu qu'une seule délétion et que la seconde base se soit modifiée, hypothèse plus que plausible étant donné que ce changement aurait eu lieu entre deux bases purines (A →G).

Il coderait selon UniProt et BsubCyc pour une enzyme la « acyl-CoA dehydrogenase » (EC number 1.3.99). BsubCyc nous dit également que cette enzyme catalyserait la réaction suivante dans laquelle une flavoprotéine va être réduite et un acide gras sera transformé en enoyl-CoA. De plus, il dit également que cette réaction interviendrait dans le pathway de  $\beta$ -oxydation des acides gras. Cependant, plusieurs études nous disent que ce gène interviendrait dans le métabolisme du soufre (Auger, Danchin, & Martin-Verstraete, 2002). Une autre étude réalisée en 2006 la décrit quant à elle comme une putative butyryl CoA déshydrogénase à cause de sa similitude avec elle (Even, et al., 2006).

## Confidentiel

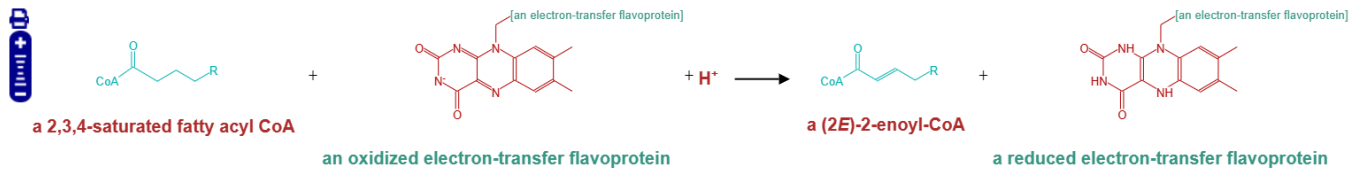


Figure 28 Réaction catalysée par la « acyl-CoA dehydrogenase ». Source BsubCyc

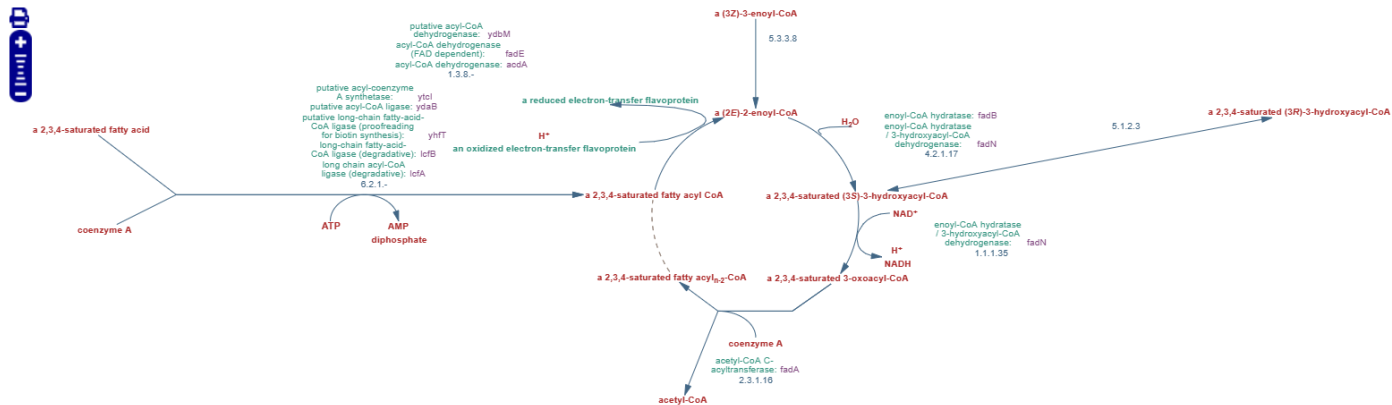


Figure 29 Pathway de  $\beta$ -oxydation des acides gras. Source BsubCyc

## Le gène *estA*

Mutation causée par l'insertion d'une base (T  $\rightarrow$  TA), une adénine, en souche 4.

Ce gène est également appelé *lipA* et coderait pour une lipase, la « secreted alkaliphilic lipase » (EC Number 3.1.1.3). Cette enzyme a pour but d'effectuer l'hydrolyse de la molécule comme montré sur la réaction ci-dessous dans laquelle une molécule de triacyl-sn-glycerol est décomposée pour libérer la molécule d'acide gras en libérant un  $\text{H}^+$ . Il est à noter que, selon BsubCyc et confirmé par BRENDA, une autre enzyme (secreted esterase / acylglycerol lipase) permettrait de catalyser la même réaction.

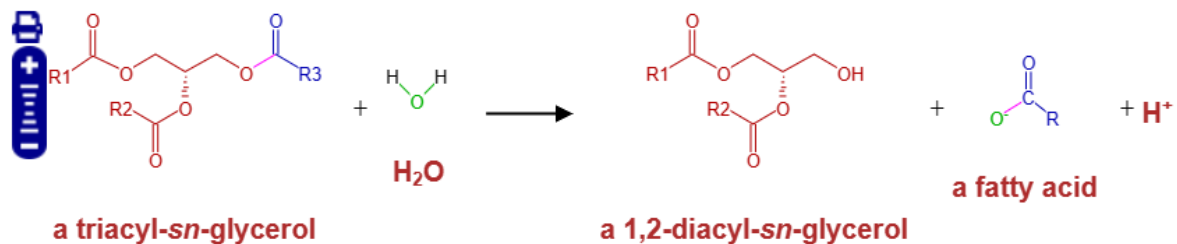


Figure 30 Réaction catalysée par la « secreted alkaliphilic lipase ». Source BsubCyc

### Le gène *flvL*

Délétion d'une base apparue à la souche 4 (CT → C) ce qui a provoqué le frameshift. Ce gène coderait pour « transcriptional regulator (FlvL-flavonoids) » selon NCBI, informations de février 2021 (Gene ID : 938778). Aucune autre information pour un gène nommé *flvL* n'a été trouvée ailleurs. Cependant, l'accèsion ID (O31541) de la protéine renvoie à une autre protéine appelée *yetL*, qui coderait pour la « HTH-type transcriptional repressor » sur UniProt et pour la « transcriptional regulator (FlvL-flavonoids) » sur BsubCyc qui est exactement la même enzyme que celle indiquée par NCBI pour notre gène d'intérêt. Une étude de 2009 suppose également l'activité de régulateur transcriptionnel putatif pour le gène *yetL* (Hirooka, et al., 2009). Tout cela pousse à croire que le gène *flvL* correspond au gène *yetL* et coderait pour un régulateur transcriptionnel.

### Frameshift et perte codon start

#### Le gène *ybfF*

Il coderait pour une protéine non catégorisée selon UniProt (O31446). Cependant, ce gène souffre d'un frameshift et d'une perte de codon start apparus en souche 1 causés par l'insertion d'une base (A → AT).

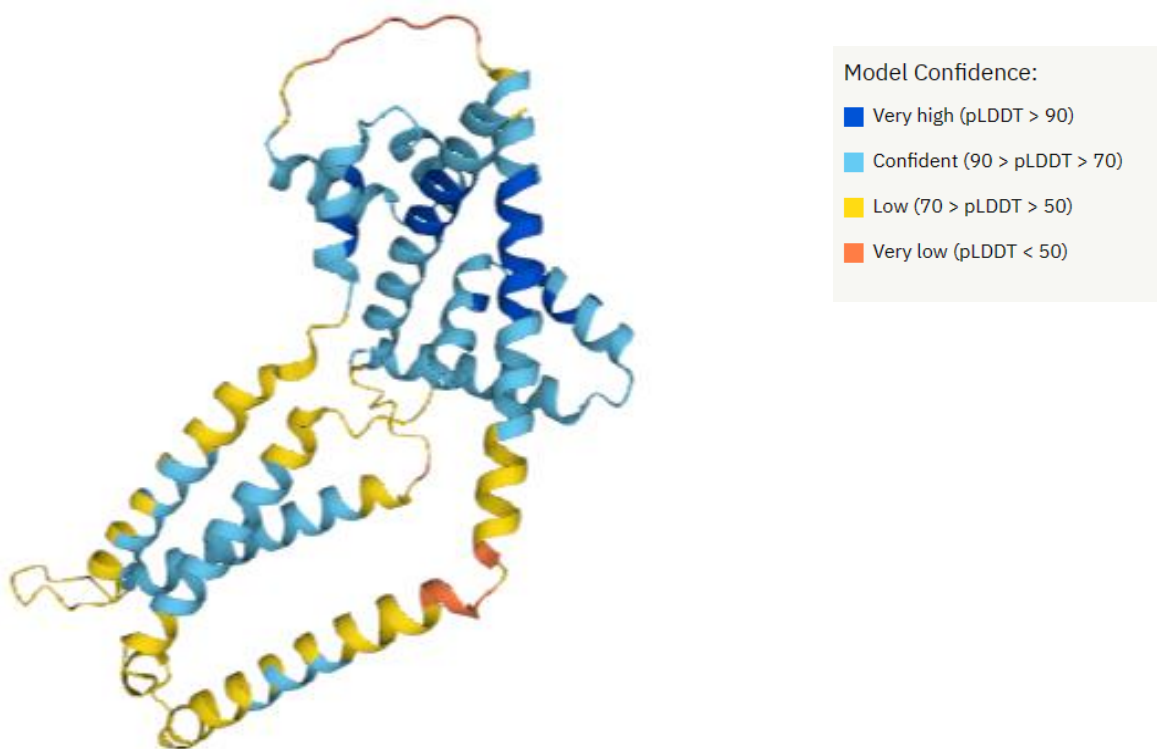


Figure 31 Protéine prédite via AlphaFold. Source UniProt

## Gain codon stop

### Le gène *ydgK*

Mutation apparue en souche 1 des suites du changement d'une base (C), une cytosine, en une autre (T), une thymine.

C'est, selon BsubCyc et Uniprot, une potentielle molécule *transporteur*. Par ailleurs, BsubCyc la catégorise « *efflux transporter* ». Cela signifie que ce gène interviendrait dans le processus d'extrusion de molécules toxiques de la cellule (Efflux, 2021). Uniprot quant à lui le définit comme une molécule *transporteur* MFS non caractérisée. Ce type de transporteur permet de faire sortir les petites molécules de l'intérieur de la protéine pour stabiliser les gradients. (Drew, North, Nagarathinam, & Tanabe, 2021). L'apparition du codon stop a lieu au milieu du gène ce qui a pour effet de l'inactiver. Cette inactivation a un impact incertain sur la bactérie à cause du flou d'activité du gène. La seule certitude est que la bactérie perd une activité de *transporteur*.

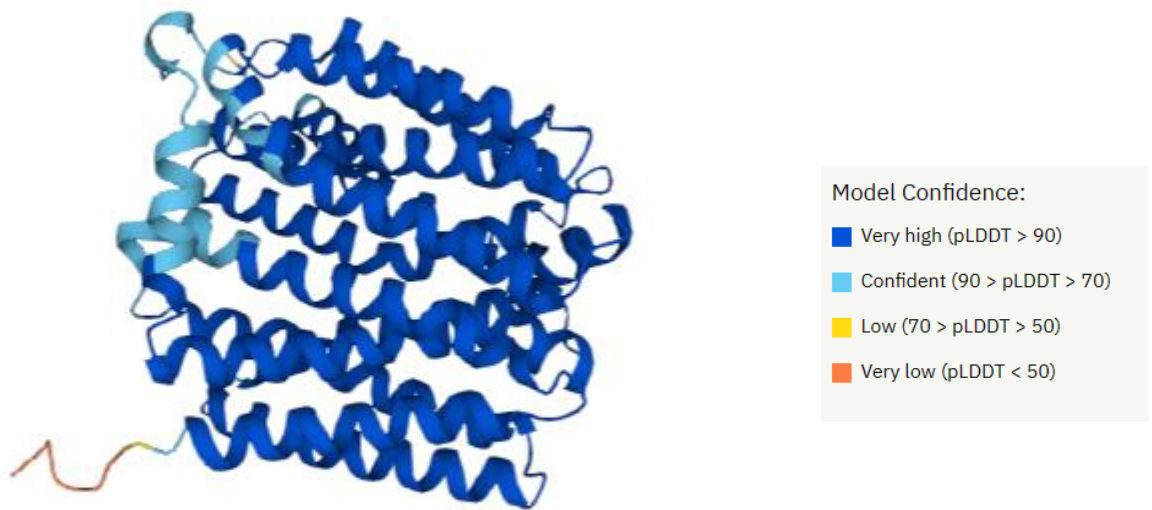


Figure 32 Protéine prédite via AlphaFold. Source UniProt

### Le gène *gmuG*

Le gène *gmuG* (EC number 3.2.1.78) est inactivé par l'apparition à la souche 1 d'un codon stop au milieu du gène. Cette mutation est apparue à la suite du changement d'une seule base (G), une guanine, en une autre (A), une adénine. Ce codon stop a causé la séparation du gène initial en 2 parties, ce qui fait supposer que la protéine initialement exprimée n'est plus fonctionnelle. Ce gène synthétise une hydrolase, la « mannan endo-1,4-beta-mannosidase », qui va permettre de catalyser la réaction suivante :

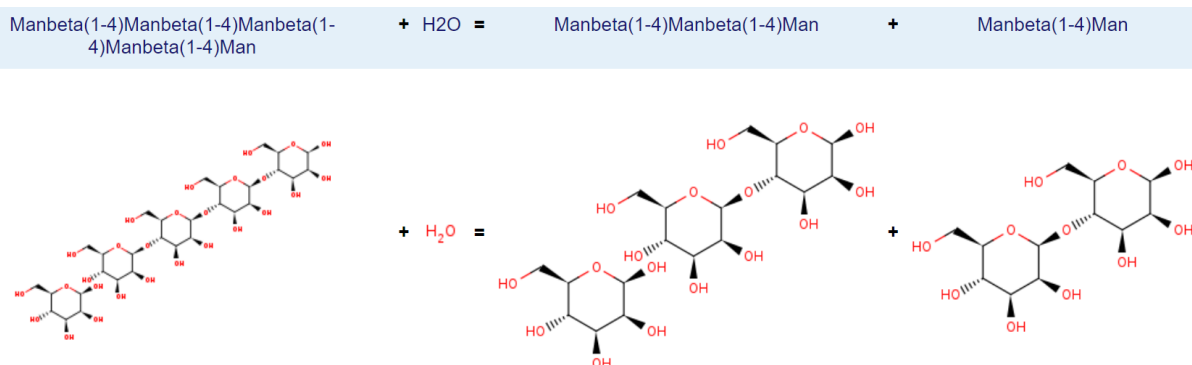
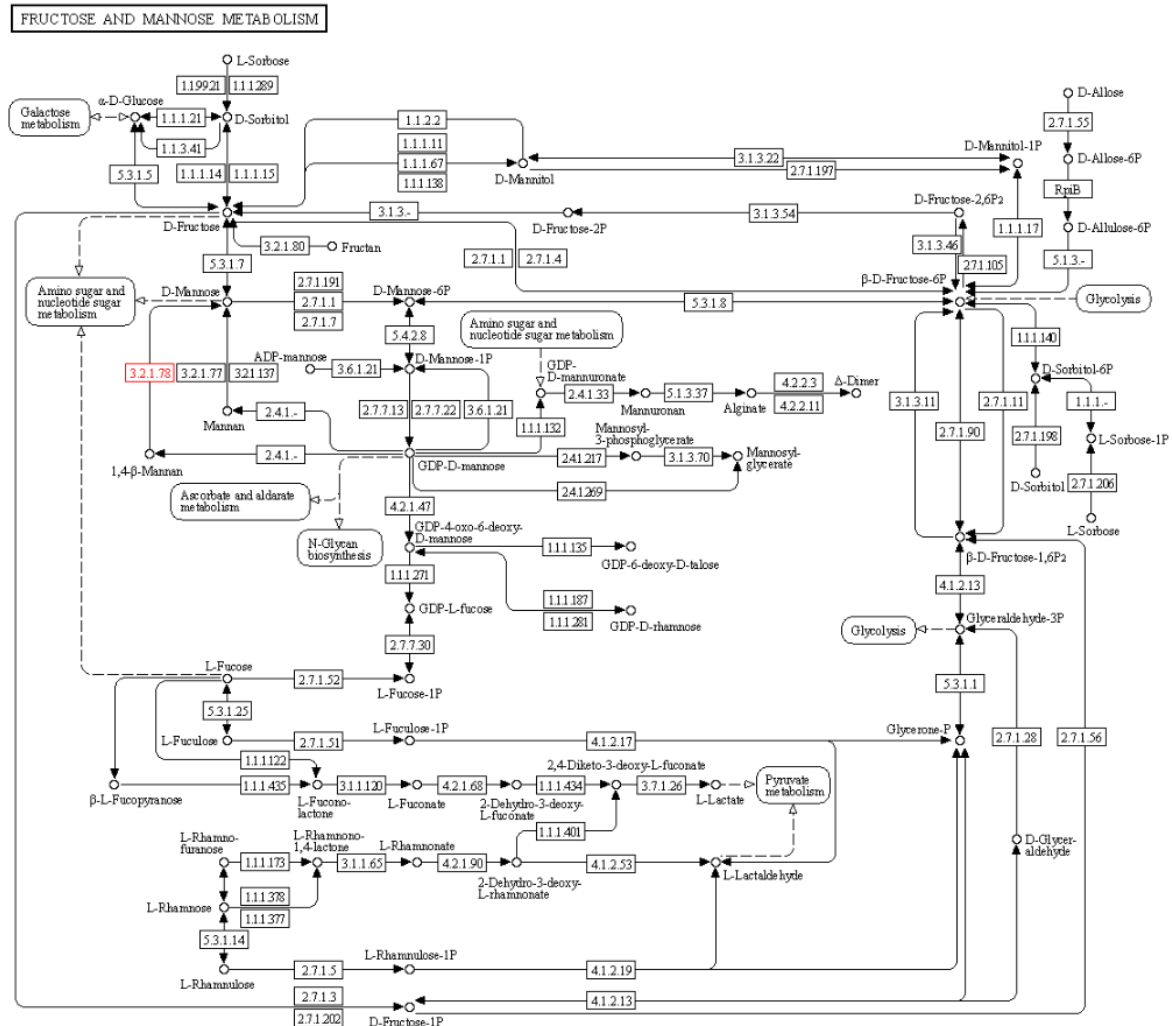


Figure 33 Réaction catalysée par la mannan endo-1,4-beta-mannosidase EC number 3.2.1.78. Source BsubCyc

Figure 34 Pathway du fructose et mannose. Source BRENDA



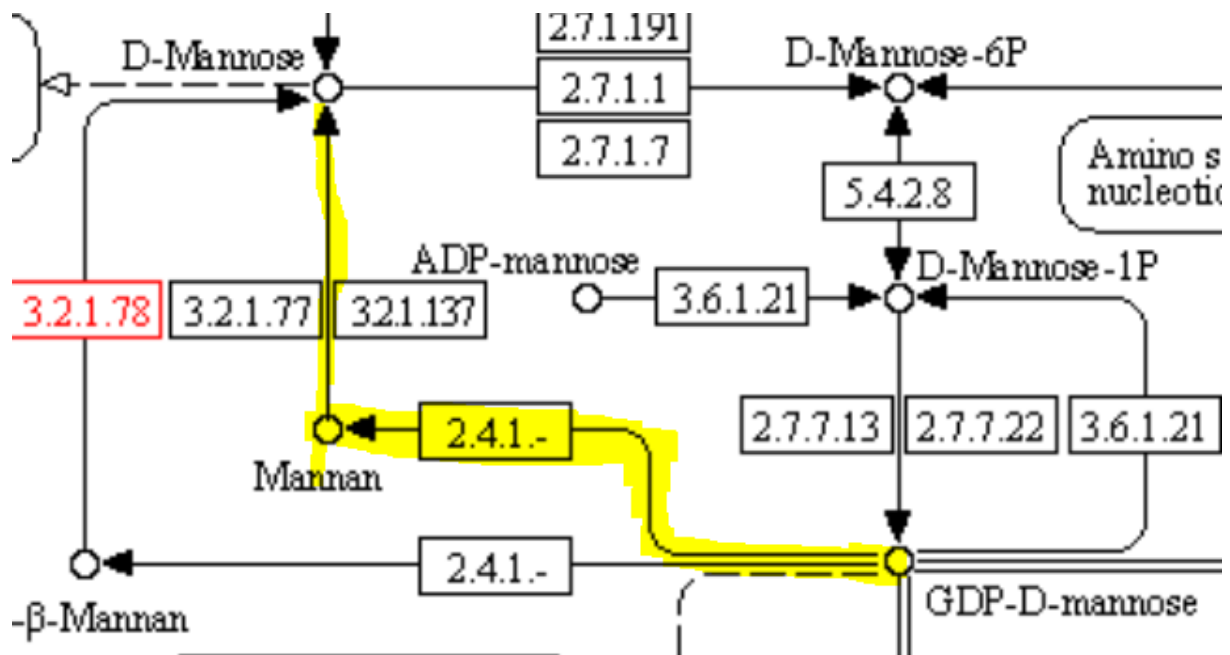


Figure 35 Zoom et mise en évidence de la voie secondaire dans le pathway du fructose et mannose. Source BRENDA

#### Le gène *yflN*

Mutation apparue en souche 3 des suites du changement d'une seule base pyrimidine (C), une cytosine vers une autre pyrimidine (T), une thymine.

Le gène *yflN* d'une longueur de 795 bp permettrait la synthèse d'une protéine de 264 acide aminé, la « putative metal-dependent hydrolase » selon Uniprot(O34409). Cette protéine est une hydrolase métalo dépendante, cela signifie que pour catalyser une réaction celle-ci a besoin d'un ou plusieurs cofacteurs métalliques (Hernick & Fierke, 2010). Cependant, ce gène a muté et obtenu un gain de codon stop en son sein, ce qui a sans doute pour effet de l'inactiver.

#### Le gène *xkdT*

Mutation apparue en souche 6 à cause du changement d'une base pyrimidine (T), une thymine en une autre (C), une cytosine.

Il coderait selon Uniprot et BsubCyc pour une protéine décrite comme « phage PBSX, putative base plate assembly protein ». Elle est donc une molécule virale issue du phage PBSX. Cependant, l'apparition du codon stop au milieu du gène a pour effet de l'inactiver.



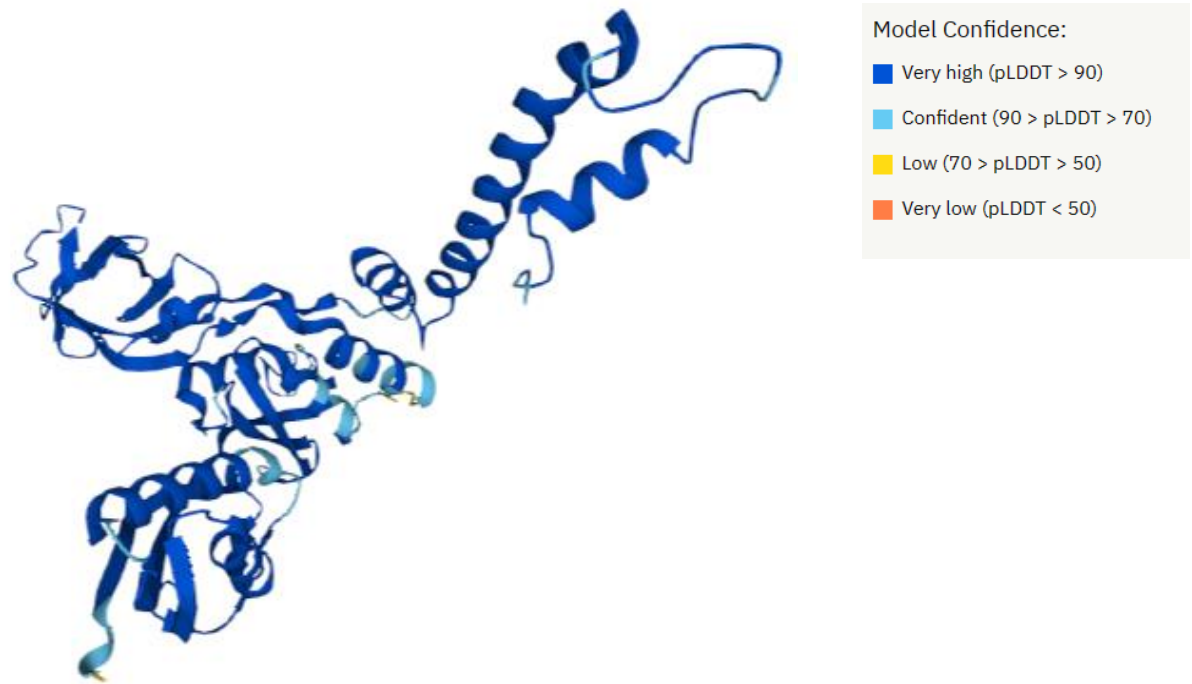


Figure 36 Protéine prédite via ALphaFold. Source UniProt

### Le gène *nprE*

Mutation causée en souche 1 par le changement d'une base purine (G), une guanine, en une autre purine (A), une adénine.

C'est un gène codant une enzyme, la « extracellular neutral metalloprotease » (EC Number 3.4.24.28). Cette protéase est une enzyme dite non essentielle pour la croissance ou pour la sporulation de *Bacillus subtilis* 168. (Barbieri, Albertini, Ferrari, Sonenshein, & Belitsky, 2016) . Ses réactions sont réprimées par le gène *ScoC* et activées via le gène *degU*.

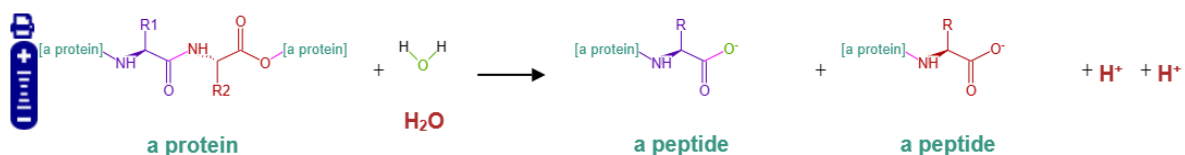


Figure 37 Réaction catalysée par la extracellular neutral metalloprotease. Source BsubCyc

Cette réaction montre que cette enzyme catalyse une activité visant à décomposer les protéines en peptides en présence d'eau. Cependant, la présence du gain de codon stop désactive le gène, cela a pour effet de priver l'organisme d'une protéase rendant par conséquent le recyclage des acides aminés plus compliqué.

Le gène *bznD*

Mutation causée en souche 4 par le changement d'une seule base purine (G), une guanine, en une autre (A), une adénine.

BsubCyc ne fournissant aucune information concernant ce gène pour *Bacillus subtilis* 168, elles ont été recherchées pour *Bacillus* 168G, cette souche étant très proche. D'après ces informations, le gène coderait pour la « (S)-benzoin forming benzil reductase » (EC number 1.1.1.320.) Cette enzyme catalyserait les deux réactions ci-dessous. Cependant, ces réactions sont celles du même enzyme codé par le gène *yueD* de *Bacillus cereus* et ont été récupérées sur BioCyc qui créditait le SRI pour ces informations. De plus, l'équation illustrée à la Figure 38 est confirmée par BRENDA. Le gène étant inactivé par l'apparition du codon stop en son milieu, l'activité de réductase est compromise.

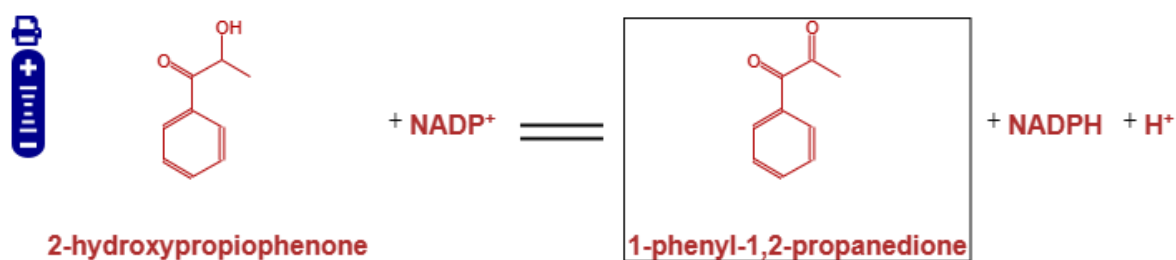


Figure 38 Réaction catalysée par la (S)-benzoin forming benzil reductase EC 1.1.1.320. Source BsubCyc

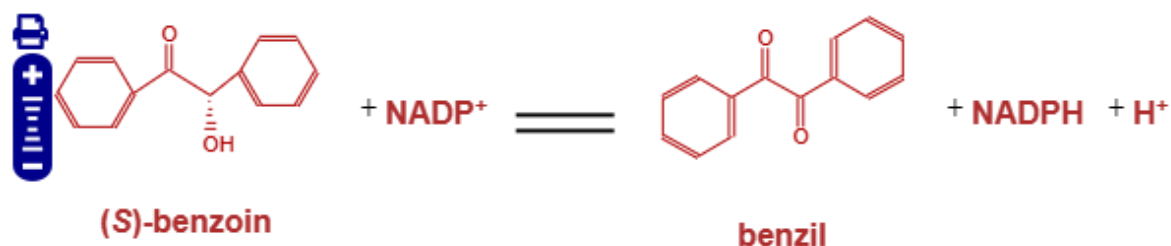


Figure 39 Réaction catalysée par la (S)-benzoin forming benzil reductase EC 1.1.1.320. Source BsubCyc

Le gène *sigL*

Mutation causée par le changement d'une base (G), une guanine, en une adénine (A) à la souche 5

Le gène *sigL* est un gène permettant la synthèse d'une ARN polymérase, « RNA polymerase sigma-54 factor (sigma-L) ». Cette enzyme va permettre la transcription d'opérons dans lesquels se trouvent des

gènes permettant l'absorption de l'arginine. (Commichau, et al., 2006) (Gardan, Rapoport, & Débarbouillé, 2003)

#### Le gène *spoOF*

Mutation causée par le changement d'une base, une guanine (G) en adénine (A) à la souche 4.

Il code la « two-component response regulator of sporulation initiation ». Il va intervenir dans l'initiation de la sporulation due au transfert de groupement phosphoryle provenant de plusieurs kinases (KinA et KinB). (Tzeng, Feher, Cavanagh, Perego, & Hoch, 1998) Il va transférer ce groupement en se phosphorylant lui-même et transférant à un autre réducteur Spo0B qui va lui-même le transférer à Spo0A qui va activer la sporulation (Gallegos-Monterrosa & Kovács, 2023) Par l'ajout du codon stop au centre du gène, celui-ci se trouve désactivé. La sporulation ne peut donc se réaliser complètement laissant la bactérie vulnérable dans les milieux incompatibles à sa survie.

#### Perte de codon stop

##### Le gène *malA* (EC number 3.2.1.122)

Le gène *malA* est également appelé *glvG*, *glvA*, *mdxK* selon subtiwiki . Il synthétise une protéine : la « 6-phospho-alpha-glucosidase » (Yamamoto , Serizawa , Thompson, & Sekiguchi, 2001) Cette enzyme joue un rôle dans l'utilisation du maltose, à savoir dégrader, en présence d'eau, le maltose en deux composantes, le D-glucopyranose et le D-glucopyranose 6-phosphate. Selon BsubCyc, cette enzyme a pour cofacteur le NAD<sup>+</sup> et le MN<sup>2+</sup>. Le gène a subi une perte de codon stop à cause de la transformation d'une base (G), une guanine, en une autre (A), une arginine. Cette perte a provoqué l'ajout de 45 acides aminés. Cependant, malgré l'élongation de la chaîne C terminale, la protéine synthétisée reste globalement inchangée (Figure 41) : aucun site fonctionnel n'est affecté. De ce fait, le gène reste actif. Cela a pour effet de permettre à l'entreprise de pouvoir utiliser du maltose comme source de sucre lors de la culture de *Bacillus subtilis* 168. Cette molécule une fois décomposée va permettre à l'organisme de posséder deux molécules de glucose comme montré dans l'équation ci-dessous.

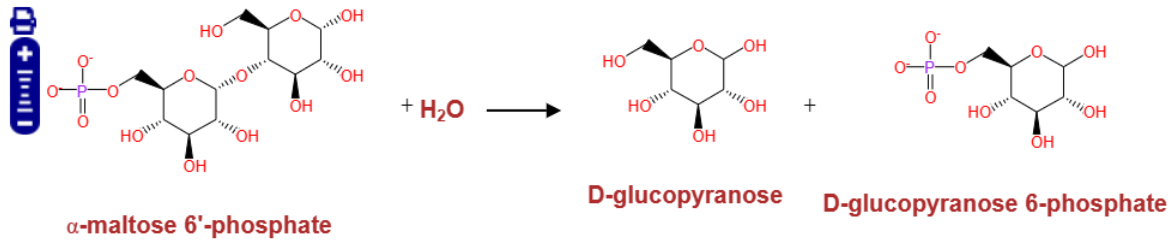


Figure 40 Réaction catalysée par la « 6-phospho-alpha-glucosidase » EC 3.2.1.122. Source BsubCyc

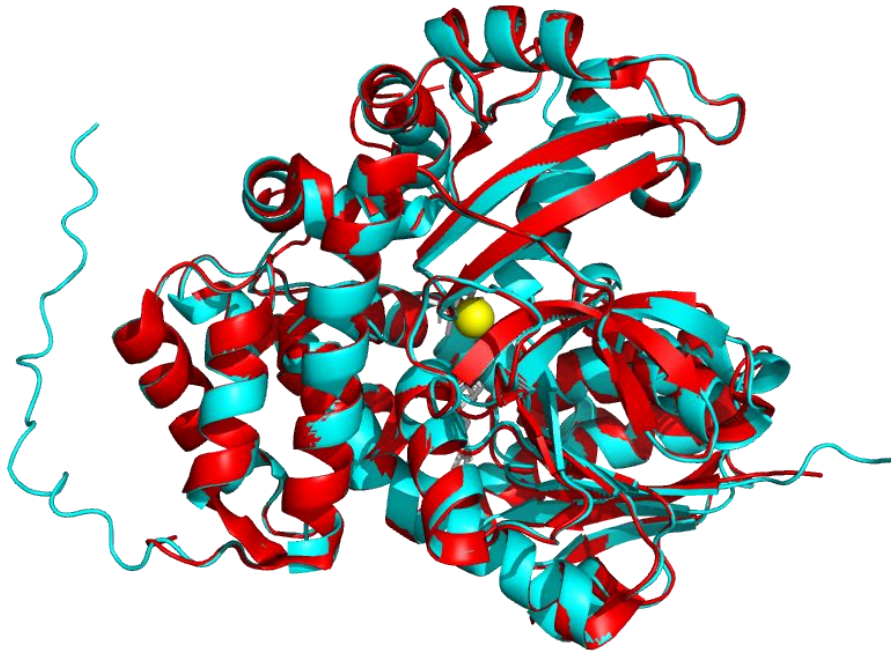


Figure 41 Comparaison du gène *malA* muté prédit via AlphaFold en bleu et sauvage prédit et récupéré sur AlphaFold en rouge. Comparaison effectuée avec Pymol

## Récapitulatif

Au sein du tableau récapitulatif ci-dessous, les données dont l'état est indiqué comme "inconnu / manque d'information" découlent du fait que les génomes n'étaient plus accessibles<sup>1</sup> au cours de l'analyse de l'impact pour déterminer le niveau d'altération des gènes impactés par un frameshift.

<sup>1</sup> Cette inaccessibilité est causée par la confidentialité découlant de ce travail et à l'interdiction de sortir des génomes des locaux de l'entreprise.

Tableau 2 Récapitulatif des gènes

Gène	Produit	Mutation	Apparition (Souche)	État
<b>tagE</b>	UDP-glucose:polyglycerol phosphate glucosyltransferase	Frameshift	5	Actif
<b>yxkK</b>	N-acetyl-S-(2-succino)-L-cysteine monooxygenase	Frameshift	3	Inconnu / manque d'informations
<b>thrC</b>	thréonine synthétase	Frameshift	1	Inconnu / manque d'informations
<b>spoIV A</b>	ATPase couche basale du manteau de sporulation	Frameshift	6	Inconnu / manque d'informations
<b>ypgR</b>	Inconnu mais « Virulence factor » proposé ici	Frameshift	1	Inconnu / manque d'informations
<b>rmgQ</b>	unsaturated rhamnogalacturonyl hydrolase	Frameshift	3	Inconnu / manque d'informations
<b>cgeD</b>	« glycosyltransférase putative » en sporulation	Frameshift	6	Inconnu / manque d'informations
<b>lysP</b>	Lysine perméase ou Putative arginine/ornithine antiporter	Frameshift	6	Inconnu / manque d'informations
<b>yqhP</b>	une protéine non catégorisée	Frameshift	4	Inconnu / manque d'informations
<b>gltA</b>	« glutamate synthase (large subunit, NADP-dependent) »	Frameshift	5	Inconnu / manque d'informations

<b>ylxY</b>	« putative sugar deacetylase »	Frameshift	1	Inconnu / manque d'informations
<b>sigP</b>	« Anti-sigma-YlaC factor YlaD ».	Frameshift	1	Inconnu / manque d'informations
<b>yhcY</b>	histidine kinase ou « two-component sensor histidine kinase [YhcZ] ».	Frameshift	2	Inconnu / manque d'informations
<b>ydbM</b>	« acyl-CoA dehydrogenase » Ou putative butyryl CoA déshydrogénase	Frameshift	1	Inconnu / manque d'informations
<b>estA</b>	« secreted alkaliphilic lipase » Ou « secreted esterase / acylglycerol lipase »	Frameshift	4	Inconnu / manque d'informations
<b>flvL</b>	transcriptional regulator (FlvL- flavonoids) ou « HTH-type transcriptional repressor »	Frameshift	4	Inconnu / manque d'informations
<b>ybfF</b>	protéine non catégorisée	Frameshift et perte codon start	1	Inconnu / manque d'informations
<b>ydgK</b>	efflux transporter ou transporter MFS non caractérisée	Gain codon stop	1	Inactivé
<b>gmuG</b>	« mannan endo-1,4-beta- mannosidase »	Gain codon stop	1	Inactivé
<b>yfIN</b>	« putative metal-dependent »	Gain codon stop	3	Inactivé
<b>xkdT</b>	« phage PBSX, putative base plate assembly protein ».	Gain codon stop	1	Inactivé

<b>nprE</b>	« extracellular neutral metalloprotease »	Gain codon stop	1	Inactiv�
<b>bznD</b>	« (S)-benzoin forming benzil reductase »	Gain codon stop	4	Inactiv�
<b>sigL</b>	RNA polymerase sigma-54 factor (sigma-L) »	Gain codon stop	5	Inactiv�
<b>spo0F</b>	« two-component response regulator of sporulation initiation »	Gain codon stop	4	Inactiv�
<b>malA</b>	« 6-phospho-alpha-glucosidase »	Perte codon stop	1	Actif

## Conclusion

Pour conclure cette analyse qui a porté sur 26 gènes répartis sur tout le génome, il a été établi que de nombreuses activités métaboliques avaient été impactées. Parmi les fonctions altérées, on peut notamment citer l'inactivation de la sporulation ou la perte de la protection contre des bactériophages via l'inactivation en souche 1 du gènes *nprE*, ... Cependant, malgré les mutations, certains gènes restent actifs. C'est le cas du gène *tagE* qui a subi un frameshift et qui est malgré tout considéré comme toujours actif, son activité étant dite vitale pour l'organisme.

Certains de ces gènes vont nécessiter d'autres études plus approfondies, des informations manquant pour conclure à l'impact qu'ont eu les mutations ou l'activité des gènes n'étant tout simplement pas encore connue. En plus de ces analyses complémentaires, d'autres analyses devront être effectuées pour identifier l'activité des gènes touchés non étudiés dans ce travail : les 225 missense et les gènes se trouvant dans les parties délétées pour pouvoir ensuite déterminer l'impact qu'ils auraient sur la physiologie de la souche de *Bacillus subtilis* utilisée par l'industrie.

Le grand nombre de mutations qui touchent presque l'ensemble du génome et la grande quantité de parties délétées affectent les voies métaboliques complexifiant ainsi les interactions chimiques, fermant ou altérant certaines voies ce qui prive l'organisme de pathways très intéressants pour la production industrielle. Dans le souci d'une meilleure productivité et d'un rendement supérieur, il serait avantageux pour l'entreprise de recréer une souche. Ce travail à court terme aura besoin d'un investissement humain et financier certain. En effet, cette étape nécessiterait d'abord d'affecter des équipes de recherche pour effectuer la création et le contrôle de la qualité des nouvelles souches. Ensuite, ces nouvelles souches devront être soumises à l'EFSA (European Food Safety Authority) pour contrôle. Pour ce faire, des dossiers de régulations devront être élaborés. Malgré tout, cela permettrait de repartir sur des bases plus stables et de probablement diminuer les coûts de production, ce qui serait, à plus ou moins long terme, un gain de rentabilité non négligeable pour l'entreprise.

En effet, en augmentant davantage la production des souches ou en permettant à celles-ci d'utiliser l'ensemble des voies métaboliques qui ont été inactivées de manière fortuite, l'organisme pourra se développer de manière plus simple, les contraintes ayant été en partie effacées. Produire une nouvelle souche permettrait d'effectuer des milieux de culture plus simples, économisant ainsi coût, produit et personnel.



## Bibliographie

- Allison, S. E., D'Elia, M. A., Arar, S., Monteiro, M. A., & Brown, E. D. (2011, juillet). Studies of the Genetics, Function, and Kinetic Mechanism of TagE, the Wall Teichoic Acid Glycosyltransferase in *Bacillus subtilis* 168\*. *Journal of Biological Chemistry*, 286, 23708-23716. doi:<https://doi.org/10.1074/jbc.M111.241265>
- Amon, J. D., Yadav, A. K., Ramirez-Guadiana, F. H., Meeske, A. J., Cava, F., & Rudner, D. Z. (2020, Février 25). SwsB and SafA Are Required for CwlJ-Dependent Spore Germination in *Bacillus subtilis*. doi:<https://doi.org/10.1128/jb.00668-19>
- Auger, S., Danchin, A., & Martin-Verstraete, I. (2002, septembre). Global expression profile of *Bacillus subtilis* grown in the presence of sulfate or methionine. *Journal of bacteriology*, 184, 5179–5186. doi:<https://doi.org/10.1128%2FJB.184.18.5179-5186.2002>
- Barbieri, G., Albertini, A. M., Ferrari, E., Sonenshein, A. L., & Belitsky, B. R. (2016, février 29). Interplay of CodY and ScoC in the Regulation of Major Extracellular Protease Genes of *Bacillus subtilis*. *Journal of Bacteriology*, 198(6). doi:<https://doi.org/10.1128/jb.00894-15>
- BioCyc. (s.d.). *bznD*. Récupéré sur BioCyc: [https://biocyc.org/gene?orgid=GCF\\_001703495&id=BFI33\\_RS17320&sid=biocyc15-3898251089#tab=RXNS](https://biocyc.org/gene?orgid=GCF_001703495&id=BFI33_RS17320&sid=biocyc15-3898251089#tab=RXNS)
- BoulangerieNet. (s.d.). *Dossiers techniques de BoulangerieNet* < Les enzymes. Récupéré sur BoulangerieNet: <https://www.boulangerienet.fr/bn/bnweb/dt/enzymes.php>
- BRENDA. (s.d.). *Information on EC 3.2.1.78 - mannan endo-1,4-beta-mannosidase*. Récupéré sur BRENDA: <https://www.brenda-enzymes.org/enzyme.php?ecno=3.2.1.78>
- BRENDA. (s.d.). *Information on EC 1.1.1.320 - benzil reductase [(S)-benzoin forming]*. Récupéré sur BRENDA: <https://www.brenda-enzymes.org/enzyme.php?ecno=1.1.1.320>
- BsubCyc. (s.d.). *Bacillus subtilis subtilis 168 Pathway: L-threonine biosynthesis*. Récupéré sur <https://bsubcyc.org/BSUB/NEW-IMAGE?type=PATHWAY&object=HOMOSER-THRESYN-PWY>
- BsubCyc. (s.d.). *Bacillus subtilis subtilis 168 Pathway: superpathway of L-isoleucine biosynthesis I*. Récupéré sur <https://bsubcyc.org/BSUB/NEW-IMAGE?type=PATHWAY&object=PWY-3001>
- BsubCyc. (s.d.). *gltA*. Récupéré sur BsubCyc: <https://bsubcyc.org/gene?id=BSU18450#tab=RXNS>
- BsubCyc. (s.d.). *lysP*. Récupéré sur BsubCyc: <https://bsubcyc.org/gene?orgid=BSUB&id=BSU33330>
- BsubCyc. (s.d.). *malA*. Récupéré sur BsubCyc: <https://bsubcyc.org/gene?id=BSU08180#tab=RXNS>
- BsubCyc. (s.d.). *nprE*. Récupéré sur bsubCyc: <https://bsubcyc.org/gene?id=BSU14700>
- Castaing, J.-P., Nagy, A., Anantharaman, V., Aravind, L., & Ramamurthi, K. S. (2012, Décembre 24). ATP hydrolysis by a domain related to translation factor GTPases drives polymerization of a static bacterial morphogenetic protein. doi:<https://doi.org/10.1073%2Fpnas.1210554110>
- Commichau, F. M., Wacker, I., Schleider, J., Blencke, H.-M., Reif, I., Tripal, P., & Stülke, S. (2006, décembre 21). Characterization of *Bacillus subtilis* Mutants with Carbon Source-Independent Glutamate Biosynthesis. *Journal of Molecular Microbiology and Biotechnology*, 12, 106–113. doi:<https://doi.org/10.1159/000096465>

- Drew, D., North, R. A., Nagarathinam, K., & Tanabe, M. (2021, mai). Structures and General Transport Mechanisms by the Major Facilitator Superfamily (MFS). doi:<https://doi.org/10.1021/acs.chemrev.0c00983>
- Efflux*. (2021, mai 1). Récupéré sur Wikipedia: <https://fr.wikipedia.org/w/index.php?title=Efflux&oldid=182487892>
- European Food Safety Authority. (s.d.). *Qualified presumption of safety (QPS)*. Récupéré sur efsa: <https://www.efsa.europa.eu/en/topics/topic/qualified-presumption-safety-qps#list-of-microorganisms-with-qps-status-qps-list-and-list-of-notifications>
- Even, S., Burguière, P., Auger, S., Soutourina, O., Danchin, A., & Martin-Verstraete, I. (2006, mars). Global Control of Cysteine Metabolism by CymR in *Bacillus subtilis*. *Journal of bacteriology*, 188, 2184–2197. doi:<https://doi.org/10.1128/JB.188.6.2184-2197.2006>
- Gallegos-Monterrosa, R., & Kovács, Á. T. (2023, avril 11). Phenotypic plasticity: The role of a phosphatase family Rap in the genetic regulation of Bacilli. *Molecular Microbiology*, 120, 20–31. doi:<https://doi.org/10.1111/mmi.15060>
- Gardan, R., Rapoport, G., & Débarbouillé, M. (2003, octobre 31). Role of the transcriptional activator RocR in the arginine-degradation pathway of *Bacillus subtilis*. *Journal of bacteriology*, 191, 3685–3697. doi:<https://doi.org/10.1046/j.1365-2958.1997.3881754.x>
- Hernick, M., & Fierke, C. (2010). Mechanisms of Metal-Dependent Hydrolases in Metabolism. doi:<https://www.sciencedirect.com/science/article/pii/B9780124095472028821>
- Hillmann, K. B., Goethel, M. E., Erickson, N. A., & Niehaus, T. D. (2022, Décembre). Identification of a S-(2-succino)cysteine breakdown pathway that uses a novel S-(2-succino) lyase. *Journal of biological chemistry*, 298, 102639. doi:<https://doi.org/10.1016/j.jbc.2022.102639>
- Hirooka, H., Danjo, Y., Hanano, Y., Kunikane, S., Matsuoka, H., Tojo, S., & Fujita, Y. (2009, juin). Regulation of the *Bacillus subtilis* Divergent yetL and yetM Genes by a Transcriptional Repressor, YetL, in Response to Flavonoid. doi:10.1128/JB.00202-09
- Kwak, M.-K., Ryu, H.-B., Song, S.-H., Lee, J.-W., & Kang, S.-O. (2018, juillet 05). Anti- $\sigma$  factor YlaD regulates transcriptional activity of  $\sigma$  factor YlaC and sporulation via manganese-dependent redox-sensing molecular switch in *Bacillus subtilis*. *Biochemical Journal*, 475, 2127–2151. doi:<https://doi.org/10.1042/BCJ20170911>
- Lin, T.-P., Chen, C.-L., Chang, L.-K., Tschén, J. S.-M., & Liu, S.-T. (1999, Août 15). Functional and Transcriptional Analyses of a Fengycin Synthetase Gene, fenC, from *Bacillus subtilis*. Récupéré sur <https://journals.asm.org/doi/full/10.1128/jb.181.16.5060-5067.1999>
- Martinez, R. M. (2013). *Bacillus subtilis*. *Brenner's Encyclopedia of Genetics*, 246–248. doi:10.1016/b978-0-12-374984-0.00125-x
- Pain suisse. (2018, septembre 23). Le processus de fermentation de la pâte – le travail des enzymes et des bactéries. *Pain suisse*.
- RHEA. (s.d.). RHEA:13065. Récupéré sur SIB Rhea: <https://www.rhea-db.org/rhea/13065>
- RHEA. (s.d.). RHEA:34991. Récupéré sur SIB Rhea: <https://www.rhea-db.org/rhea/34991>

- Ryzhykau, Y. L., Orekhov, P. S., Rulev, M. I., & et al. (2021, mai 24). Molecular model of a sensor of two-component signaling system. *scientific reports*, 11. doi:<https://doi.org/10.1038/s41598-021-89613-6>
- Steffes, C., Ellis, J., Wu, J., & Rosen, B. P. (1992, mai). THE LYSP GENE ENCODES THE LYSINE-SPECIFIC PERMEASE. doi:10.1128/jb.174.10.3242-3249.1992
- SubtiWiki. (s.d.). *malA*. Récupéré sur SubtiWiki: <http://subtiwiki.uni-goettingen.de/v4/gene?id=C650160C9E5B5F8587CA9E80A57BCCE97D701037>
- Tzeng, Y.-L., Feher, V. A., Cavanagh, J., Perego, M., & Hoch, J. A. (1998, novembre 6). Characterization of Interactions between a Two-Component Response Regulator, Spo0F, and Its Phosphatase, RapB. *Biochemistry*, 37(47), 16538-16545. doi:<https://doi.org/10.1021/bi981340o>
- UniProt. (s.d.). 007527 · YHCY\_BACSU. Récupéré sur UniProt: <https://www.uniprot.org/uniprotkb/O07527/entry>
- UniProt. (s.d.). 007628 · YLAD\_BACSU. Récupéré sur UniProt: <https://www.uniprot.org/uniprotkb/O07628/entry>
- UniProt. (s.d.). 031446 · YBFF\_BACSU. Récupéré sur UniProt: <https://www.uniprot.org/uniprotkb/O31446/entry>
- UniProt. (s.d.). 034409 · YFLN\_BACSU. Récupéré sur UniProt: <https://www.uniprot.org/uniprotkb/O34409/entry>
- UniProt. (s.d.). P54169 · YPGR\_BACSU. Récupéré sur UniProt: <https://www.uniprot.org/uniprotkb/P54169/entry>
- UniProt. (s.d.). P54339 · XKDT\_BACSU. Récupéré sur UniProt: <https://www.uniprot.org/uniprotkb/P54339/entry>
- UniProt. (s.d.). P54514 · YQHP\_BACSU. Récupéré sur UniProt: <https://www.uniprot.org/uniprotkb/P54514/entry>
- UniProt. (s.d.). P96709 · YDGK\_BACSU. Récupéré sur UniProt: <https://www.uniprot.org/uniprotkb/P96709/entry>
- Yamamoto , h., Serizawa , M., Thompson, J., & Sekiguchi, J. (2001, septembre). Regulation of the glv Operon in Bacillus subtilis: YfiA (GlvR) Is a Positive Regulator of the Operon That Is Repressed through CcpA and cre. *Journal of bacteriology*, 183, 5110–5121. doi:<https://doi.org/10.1128%2FJB.183.17.5110-5121.2001>
- Zenodo. (s.d.). *Microbial species as notified to EFSA*. Récupéré sur Zenodo: <https://zenodo.org/record/3607184>

## Annexes

Annexe I Tableau de données des gènes

Mutation	Référence	Résultat	Impact	ID
<b>snp</b>	C	T	stop_gained	ydgK
<b>snp</b>	G	A	stop_gained	gmuG
<b>snp</b>	C	T	stop_gained	yflN
<b>snp</b>	T	C	stop_lost&splice_region_variant	malA
<b>snp</b>	C	T	stop_gained	xkdT
<b>snp</b>	G	A	stop_gained	nprE
<b>snp</b>	G	A	stop_gained	bznD
<b>snp</b>	G	A	stop_gained	sigL
<b>snp</b>	G	A	stop_gained	spo0F
<b>ins</b>	A	AT	frameshift_variant&start_lost	ybfF
<b>ins</b>	T	TA	frameshift_variant	estA
<b>complex</b>	AA	G	frameshift_variant&synonymous_variant	ydbM
<b>del</b>	CT	C	frameshift_variant	flvL
<b>del</b>	CA	C	frameshift_variant	yhcY
<b>del</b>	CA	C	frameshift_variant	sigP
<b>del</b>	CT	C	frameshift_variant	ylxY
<b>ins</b>	C	CT	frameshift_variant	gltA
<b>ins</b>	G	GTA	frameshift_variant	cgeD
<b>del</b>	CT	C	frameshift_variant	ypgR
<b>ins</b>	A	AT	frameshift_variant	spoIVA
<b>del</b>	GA	G	frameshift_variant	yqhP
<b>del</b>	GC	G	frameshift_variant	rmgQ
<b>ins</b>	G	GC	frameshift_variant	thrC
<b>del</b>	AT	A	frameshift_variant	lysP
<b>ins</b>	C	CT	frameshift_variant	tagE
<b>del</b>	GAAGCTTCGGCGTC	G	frameshift_variant	yxek
<b>ins</b>	A	AT	frameshift_variant&start_lost	ybfF

Annexe II Tableau de similitude à 90% pour le gène *ypgR*. Source NCBI

Entry	Reviewed	Entry Name	Protein names	Gene Names	Organism	Length
AOA0D1JF7	unreviewed	AOA0D1JF7_BACIU	Scaffold protein Nfu/NifU N-terminal domain-containing protein	SC09_Contig24orf00248	Bacillus subtilis	377
AOA0H3E0P4	unreviewed	AOA0H3E0P4_BACA1	Putative lyase	BATR1942_08790	Bacillus atrophaeus (strain 1942)	376
AOA16AXFM8	unreviewed	AOA16AXFM8_BACIU	PBS lyase HEAT-like repeat domain protein	B4417_1259	Bacillus subtilis	377
AOA165AEY5	unreviewed	AOA165AEY5_BACIU	Conserved virulence factor C family protein (PBS lyase HEAT-like repeat domain protein)	B4122_2057 J5227_09280	Bacillus subtilis	377
AOA2M8T2K8	unreviewed	AOA2M8T2K8_9BACI	Virulence factor	CVV77_02025	Bacillus sp. SM1	377
AOA3A5I8J4	unreviewed	AOA3A5I8J4_BACIU	Virulence factor	CI481_15300	Bacillus subtilis	377
AOA410WFT6	unreviewed	AOA410WFT6_BACVA	Virulence factor	BV11031_07060	Bacillus vallismortis	377
AOA4R6HYI1	unreviewed	AOA4R6HYI1_BACIU	HEAT repeat protein	DF069_2119	Bacillus subtilis	377
AOA5D4NCV7	unreviewed	AOA5D4NCV7_BACIU	Virulence factor	FZC70_00780	Bacillus subtilis	377
AOA5F2KIW1	unreviewed	AOA5F2KIW1_BACIU	Virulence factor	FAL52_16840	Bacillus subtilis	377
AOA6H0H3S0	unreviewed	AOA6H0H3S0_BACIU	Virulence factor	HCN55_11845	Bacillus subtilis subsp. subtilis str. SMY	377
AOA6H0WIK2	unreviewed	AOA6H0WIK2_9BACI	Virulence factor	G4P54_11475	Bacillus tequilensis	377
AOA6H2RIL5	unreviewed	AOA6H2RIL5_BACMO	PBS lyase HEAT-like repeat domain protein	YPgR HC660_20670	Bacillus mojavensis	376
AOA6M4JII1	unreviewed	AOA6M4JII1_BACSU	Virulence factor	HIR78_13175	Bacillus subtilis (strain 168)	377
AOA7G7JUL8	unreviewed	AOA7G7JUL8_9BACI	Conserved virulence factor C family protein	H7F27_10145	Bacillus sp. PAMC26543	376
AOA7U5BV05	unreviewed	AOA7U5BV05_BACIU	Virulence factor	A1D11_08790	Bacillus subtilis subsp. globigii	376
AOA7V8GHZ5	unreviewed	AOA7V8GHZ5_9BACI	Virulence factor	BTW01_09225	Bacillus sp. SKDUJ12	377
AOA809FIK4	unreviewed	AOA809FIK4_BACIU	Virulence factor	D9C10_09285	Bacillus subtilis subsp. subtilis	377
AOA837XQP0	unreviewed	AOA837XQP0_BACAT	Scaffold protein Nfu/NifU N-terminal domain-containing protein	B4144_2926	Bacillus atrophaeus	376
AOA857HI51	unreviewed	AOA857HI51_BACIU	Virulence factor C	Bateq7P116_2403	Bacillus subtilis	377
AOA8B5NMV9	unreviewed	AOA8B5NMV9_BACIU	Virulence factor	C6V43_06650	Bacillus subtilis	377
AOA8E0VQA7	unreviewed	AOA8E0VQA7_9BACI	Virulence factor	FQ086_07020	Bacillus sp. ANT_WA51	377
D4FXV1	unreviewed	D4FXV1_BACNB	Scaffold protein Nfu/NifU N-terminal domain-containing protein	YPgR BSNT_08577	Bacillus subtilis subsp. natto (strain BEST195)	377
EOU0I5	unreviewed	EOU0I5_BACSH	Putative lyase	YPgR BSUW23_10710	Bacillus spizizenii (strain ATCC 23059 / NRRL B-14472 / W23) (Bacillus subtilis subsp. spizizenii)	377
G4EUO2	unreviewed	G4EUO2_BACIU	Scaffold protein Nfu/NifU N-terminal domain-containing protein	BSSC8_20820	Bacillus subtilis subsp. subtilis str. SC-8	363
G4NSH0	unreviewed	G4NSH0_BACCS4	YpgR	GYO_2411	Bacillus spizizenii (strain DSM 15029 / JCM 12233 / NBRC 101239 / NRRL B-23049 / TU-B-10) (Bacillus subtilis subsp. spizizenii)	377
M4KW61	unreviewed	M4KW61_BACIU	Putative lyase	YPgR C663_2060	Bacillus subtilis XF-1	377
P54169	reviewed	YPGR_BACSU	Uncharacterized protein YpgR	YPgR BSU21880	Bacillus subtilis (strain 168)	377