



HAL
open science

Quantification des contraintes métaboliques et physiologiques liées à la surproduction de protéines recombinantes par Escherichia coli : amélioration des performances et de la robustesse du système d'expression et du procédé de production

Clement Patacq

► **To cite this version:**

Clement Patacq. Quantification des contraintes métaboliques et physiologiques liées à la surproduction de protéines recombinantes par Escherichia coli : amélioration des performances et de la robustesse du système d'expression et du procédé de production. Biochimie [q-bio.BM]. INSA de Toulouse, 2018. Français. NNT : 2018ISAT0040 . tel-04416430

HAL Id: tel-04416430

<https://theses.hal.science/tel-04416430>

Submitted on 25 Jan 2024

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

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

Université Fédérale



Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut National des Sciences Appliquées de
Toulouse

Présentée et soutenue par

Clément PATACQ

Le 23 octobre 2018

**Quantification des contraintes métaboliques et
physiologiques liées à la surproduction de protéines
recombinantes par *Escherichia coli* : amélioration des
performances et de la robustesse du système d'expression
et du procédé de production.**

École doctorale : **SEVAB – Sciences Ecologiques, Vétérinaires,
Agronomiques et Bioingénieries**

Spécialité : **Ingénieries microbienne et enzymatique**

Unité de recherche :

LISBP - Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés

Thèse dirigée par

Fabien LETISSE

Jury

M. Stéphane DELAUNAY, Rapporteur

M. Johannes GEISELMANN, Rapporteur

Mme Emmanuelle BOUVERET, Examinatrice

M. Fabien LETISSE, Directeur de thèse

M. Nicolas CHAUDET, Co-directeur de thèse

*« Apercevoir une difficulté et s'étonner,
c'est reconnaître sa propre ignorance » - Aristote*

Je voudrais tout d'abord remercier les membres de mon jury de thèse, Stéphane Delaunay, Hans Geiselmann, Emmanuelle Bouveret, Nicolas Chaudet et Fabien Létisse pour avoir accepté d'évaluer ce travail. Merci Emmanuelle d'avoir éclairé par tes conseils et soutenu ce travail au cours des comités de thèse.

Cette thèse a été financée par Sanofi Pasteur dans le cadre du dispositif CIFRE de l'ANRT en collaboration avec le laboratoire LISBP de l'INSA de Toulouse. Ce projet a été initié par Nicolas Chaudet et Fabien Létisse lors de trois stages réalisés successivement par Vincent Templier, Luis Ricardo Maldonado Ramirez et Elodie Guémard. A partir des résultats générés vous avez décidé de prolonger ce projet sous la forme de cette thèse. Une telle collaboration entre partenaires public et privé, longue et difficile à organiser, a été possible grâce à votre entier investissement et votre ténacité. C'est également grâce au soutien et l'aide d'Etienne Boutry, Jean-Marc Guillaume, Isabelle Fugier, Marjorie Monnet, Jean-Charles Portais que vous avez pu me confier votre projet. Pour tout cela je vous adresse mes plus sincères remerciements ainsi qu'aux personnes ayant participé à la mise en place de cette thèse.

Au cours des huit premiers mois de thèse, j'ai eu la chance d'être parfaitement accueilli dans l'équipe BRD de Sanofi Pasteur à Marcy l'Etoile. Je tiens à remercier Nicolas Chaudet qui m'a fait confiance du début à la fin de ce projet ; je te remercie pour tes conseils, ton pragmatisme et ton engagement personnel, indispensables à la réussite de ce projet. Merci à toute l'équipe du département BRD pour votre aide et l'ambiance chaleureuse que vous savez créer au quotidien. Merci à Marjorie Monnet d'être garante de cette cohésion par un management humain et bienveillant. Je voudrais également remercier Raphaël Esson et plus généralement le département ARD pour leur soutien analytique, source de résultats centraux dans l'avancé du projet. Enfin, je remercie Fabien Barbirato pour son implication lors de l'écriture de ce manuscrit.

Je remercie tout particulièrement mon directeur de thèse Fabien Létisse. Ton accompagnement exemplaire et la richesse de ton enseignement m'ont donné la possibilité d'atteindre un niveau qui me semblait initialement hors de portée. Merci pour ta confiance et ton écoute qui m'ont permis de m'épanouir dans la gestion de ce projet. Je remercie toute l'équipe MetaSys et la plateforme Metatoul d'avoir rendu le quotidien si agréable, merci pour le partage de votre expertise, pour les discussions passionnantes, les soirées amicales. Un grand merci à Mickaël Dinclaux et Mickaël Labrit, d'indispensables partenaires expérimentateurs. Merci à Pierre Millard pour ton aide, nos échanges m'ont toujours fait progresser. Merci à Sandrine Uttenweiler pour ta formation précieuse à l'écriture de cette thèse.

Ces trois ans m'ont énormément enrichi d'un point de vue scientifique et humain. J'ai été fier de partager le quotidien de personnes aussi compétentes et profondément bienveillantes. Vous m'avez tous beaucoup appris et fait progresser.

Je remercie tout spécialement mes parents qui m'ont permis de m'épanouir et qui ont fait bien plus que le nécessaire pour me rendre libre de suivre mes passions. Mon frère, pour son indéfectible présence et soutien. Tracy qui m'encourage constamment à donner le meilleur de moi-même.

Résumé

La production de protéines hétérologues permet de développer une nouvelle génération de vaccins. La bactérie *Escherichia coli* est l'un des organismes hôtes les plus utilisés pour la production de protéines hétérologues, appelées également protéines recombinantes. Le déclenchement de la production de protéine altère la croissance bactérienne en réponse à la réallocation des ressources métaboliques vers la synthèse de la protéine ; ce qui peut conduire à l'arrêt complet de la croissance. Le maintien de la croissance bactérienne durant la production de la protéine recombinante est pourtant essentiel pour améliorer significativement la quantité et la fonctionnalité des protéines produites. L'équilibre entre croissance et production est parfois obtenu après une longue et complexe démarche empirique du système d'expression génique, de la souche de production, du milieu de culture et du procédé de production.

Dans une démarche rationnelle visant à développer un système biologique robuste et performant pour la production d'une grande diversité de protéines recombinantes chez *E. coli*, les contraintes métaboliques liées à leur production ont été quantifiées. Pour cela, un procédé de production basé sur l'utilisation du système d'expression T7 et de la souche *E. coli* BL21 (DE3) a été mis au point en milieu minimum et en bioréacteur pour étudier les cinétiques de croissance et de production de deux protéines modèles. Des analyses fonctionnelles du métabolisme ont ensuite été effectuées afin de déterminer les perturbations métaboliques majeures et les limitations nutritionnelles associées à la production de protéines recombinantes. L'étude de ces contraintes a permis de mettre en évidence l'activation d'un mécanisme d'adaptation de la bactérie, appelé réponse stringente, lors de l'apparition d'une limitation en acides aminés résultant de la production des protéines recombinantes étudiées. Une méthode de quantification absolue des molécules centrales de la réponse stringente, le ppGpp et le pppGpp, a été développée par chromatographie ionique couplée à la spectrométrie de masse. Cette méthode a été utilisée pour étudier l'influence de la réponse stringente sur la croissance et la production de la souche *E. coli* BL21 (DE3). Les résultats obtenus mettent en évidence que l'évolution de la concentration intracellulaire de ppGpp intervient dans la compétition entre la croissance bactérienne et la production pour la consommation des acides aminés intracellulaires. De ce constat, le système d'expression T7 a été intégré à la régulation métabolique et traductionnelle de la bactérie *E. coli* BL21 (DE3) afin d'adapter la vitesse de production avec les capacités métaboliques de la souche. Ces modifications ont permis d'obtenir une preuve de concept de l'efficacité de cette stratégie visant à conserver l'équilibre entre croissance et production. Ce nouveau système biologique de production a ainsi permis d'augmenter considérablement les quantités de protéines produites et offre la possibilité de développer de nouveaux procédés performants de production semi-continu et continu en milieu chimiquement défini.

Abstract

The production of heterologous proteins offers the ability to develop a new generation of vaccines. The most used organism for the production of heterologous proteins, also called recombinant proteins, is the bacterium *Escherichia coli*. However, the induction of the production often alleviates the bacterial growth by the new allocation of metabolic resources toward the production of the recombinant protein. Even, this may also lead to growth arrest. The production of high quantities of functional recombinant proteins requires a good balance between of bacterial growth and production of the recombinant protein. Such a balance is generally obtained empirically through a long and laborious process aiming to optimize the gene expression system, the strain of production, the culture medium and the production process.

In order to rationally develop a robust and an efficient biological system for the production of a large variety of recombinant proteins in *E. coli*, the metabolic constraints associated to their production have been quantified. To achieve this, a production method based on the use of the T7 expression system and the *E. coli* BL21 (DE3) strain was developed on a minimal medium in bioreactor to study growth and production kinetics of two model proteins. Functional analyzes of metabolism were then performed to determine the major metabolic constraints and nutritional limitations associated with the production of recombinant proteins. This study revealed that the stringent response, a well know mechanism of adaptation, is activated as a response to amino acid limitation, provoked by the production of the recombinants proteins. An absolute quantification method of the key molecules of the stringent response, ppGpp and pppGpp, was developed by ion chromatography coupled to mass spectrometry. This method was used to study the influence of stringent response on the growth and the production of *E. coli* BL21 (DE3) strain. The results show that the variations of ppGpp concentrations play a crucial role in the competition between bacterial growth and protein production for the consumption of intracellular amino acids. From this observation, the T7 expression system has been integrated into the metabolic and translational regulation of the *E. coli* BL21 (DE3) in order to adjust as perfect as possible the protein production rate to the metabolic capacities of the strain. A proof-of-concept of the innovative design to be capable to balance the resources between growth and production was obtained. This new biological production system has made it possible to significantly increase the quantities of proteins produced and opens up the possibility of developing performant semi-continuous and continuous production processes in a chemically defined medium.

Sommaire

Sommaire	1
Table des illustrations	6
Index des tableaux	9
Abréviations	10
Chapitre 1 – Introduction bibliographique	12
1.1. Système d’expression génique de la protéine recombinante chez E. coli	15
1.1.1. Vecteur d’expression	16
1.1.2. Présentation du système d’expression T7	18
1.1.3. Origine de réplication de type ColE1	19
1.1.4. Plasmide de Sanofi Pasteur	21
1.1.5. Conception de la séquence du gène d’intérêt	22
1.1.6. « Etiquettes » d’expression	24
1.1.7. Souches d’Escherichia coli	27
1.2. Intégration du système d’expression T7 dans une souche d’E. coli	29
1.2.1. Importance de l’équilibre entre croissance et production	29
1.2.2. Charges métaboliques de la production	32
1.2.3. Influence de la production de protéines recombinantes sur le taux de croissance	37
1.3. Réponse stringente chez E. coli	43
1.3.1. ppGpp et pppGpp : synthèse et dégradation	43
1.3.2. Régulation cycle cellulaire	46
1.3.3. Régulation de la transcription	47
1.3.4. Autres régulations par interaction protéiques	50
1.4. Objectif du travail de thèse	51
Chapitre 2 – Matériels et Méthodes	55
2.1. Milieux et conditions de cultures	55
2.1.1. Milieux de culture solides :	55
2.1.2. Milieux de cultures liquides :	55

2.2.	Souches	56
2.3.	Plasmides	56
2.4.	Séquences des promoteurs du système MetaProd	58
2.4.1.	Souches MetaProd	58
2.4.2.	Plasmides MetaProd	59
2.5.	Conditions de culture	60
2.6.	Préparation de cellules BL21 compétentes	61
2.7.	Transformation	62
2.8.	Techniques analytiques	62
2.8.1.	Détermination de la concentration en biomasse	62
2.8.2.	Stabilité plasmidique et quantification du nombre de copies de plasmide par bactérie	63
2.8.3.	Analyse de la production de protéines recombinantes	64
2.9.	Purification de la protéine UL130	65
2.10.	Préparation des échantillons pour la RMN	65
2.11.	Préparation des échantillons pour analyses en spectrométrie de masse	65
2.12.	Préparation des métabolites marqués au ¹³C et à ¹⁵N	66
2.13.	Analyses RMN des métabolites extracellulaires	66
2.14.	Analyses HPLC-OrbiTrap des acides aminés libres	67
2.15.	Analyses IC-OrbiTrap	68

Chapitre 3 – Quantification absolue du ppGpp et du pppGpp par LC-MS : Développement d'une méthode de dilution isotopique à deux dimensions **69**

3.1.	Conclusion :	96
-------------	---------------------	-----------

Chapitre 4 : Etude des cinétiques de croissance et de production des protéines UL130 et Hla avec le procédé de référence. **97**

4.1.	Méthodes d'analyse des cinétiques de croissance et de production.	99
4.1.1.	Détermination des biais d'estimation de la concentration en biomasse par turbidimétrie en condition de production	99

4.1.2.	Quantification de la protéine UL130 _____	102
4.2.	Modification du procédé standard vers le procédé de référence avec la protéine UL130	
	104	
4.2.1.	Etude du procédé standard du département BRD Sanofi Pasteur _____	104
4.2.2.	Induction précoce de la production de la protéine UL130 en milieu riche _____	106
4.2.3.	Formulation du milieu minimum chimiquement défini _____	108
4.2.4.	Procédé de production de la protéine UL130 en milieu minimum _____	109
4.2.5.	Marquage isotopique de la protéine d'intérêt _____	112
4.3.	Procédé modèle de production de la protéine Hla _____	114
4.4.	Discussions _____	119
Chapitre 5 – Analyse des contraintes métaboliques liées à la production des protéines		
UL130 et Hla au cours du procédé de référence. _____		
		125
5.1.	Analyse non ciblée de l'exométabolome d'<i>E. coli</i> BL21. _____	126
5.1.1.	Procédé de culture sans induction _____	126
5.1.2.	Procédé de production des protéines UL130 et Hla _____	130
5.2.	Quantification des acides aminés libres en condition de production _____	136
5.3.	Comparaison de la proportion en acides aminés des protéines UL130 et Hla avec la proportion en acides aminés des protéines de la biomasse _____	140
5.4.	Quantification du ppGpp au cours des procédés de production _____	144
5.5.	Discussion _____	147
Chapitre 6 – Etude de l'influence de la réponse stringente sur la souche <i>E. coli</i> BL21 sans		
ou avec production de protéines recombinantes _____		
		151
6.1.	Etude de l'influence de la réponse stringente sur la souche <i>E. coli</i> BL21 sans production de protéine recombinante. _____	154
6.1.1.	Crucial role of ppGpp – and not pppGpp – on resilience capability of Escherichia coli to the disruption in amino acids homeostasis _____	154
6.1.2.	Quantification des acides aminés d' <i>E. coli</i> BL21 WT lors de l'ajout de SHX. _____	178
6.2.	Etude de l'influence de la réponse stringente sur la souche <i>E. coli</i> BL21 en condition de production de protéines recombinantes _____	182

Sommaire

6.2.1.	Influence de la concentration intracellulaire en ppGpp sur la croissance et la production de la protéine UL130.	182
6.2.2.	Influence de la concentration intracellulaire de ppGpp sur la croissance et la production de la protéine Hla.	185
6.2.3.	Détermination du nombre de copies de plasmides au cours du procédé de production de la protéine UL130	188
6.2.4.	Quantification de l'exométabolome de souches <i>E. coli</i> BL21 mutées ($\Delta gppA$, $\Delta relA$, $\Delta relA : \Delta gppA$)	193

6.3. Discussion : _____ 196

6.3.1.	Vers un système biologique robuste et performant pour la production de protéines recombinantes	198
--------	--	-----

Chapitre 7 - Optimisation de la performance et de la robustesse du système biologique de production 201

7.1. Principe des souches MetaProd _____ 202

7.1.1.	Choix des promoteurs d'intérêts	203
7.1.2.	Caractéristiques des promoteurs des opérons ribosomiaux d' <i>E. coli</i> .	204
7.1.3.	Description des promoteurs <i>rrn</i> P1	206
7.1.4.	Etude de la performance des souches MetaProd	208

7.2. Stabilisation du nombre de copies de plasmide par supplémentation en acides aminés 213

7.2.1.	Effet du nombre de copies de plasmide sur la performance des souches MetaProd.	213
7.2.2.	Effet de la supplémentation en acides aminés sur le procédé de production de la protéine UL130 par la souche WT.	215
7.2.3.	Comparaison de la souche WT avec les souches MetaProd <i>rrnE</i> et <i>rrnB</i> en condition de production de la protéine UL130 lors de la stabilisation du plasmide par supplémentation en acides aminés.	220
7.2.4.	Comparaison des souches WT et MetaProd <i>rrnE</i> en condition de production de la protéine Hla lors de la stabilisation du plasmide par supplémentation en acides aminés.	223

7.3. Principe des plasmides MetaProd _____ 225

7.3.1.	Choix des promoteurs d'intérêt	227
7.3.2.	Evaluation des performances de la souche WT transformée avec les plasmides MetaProd au cours de la production de la protéine UL130	229
7.3.3.	Evaluation des performances des souches MetaProd <i>rrnE</i> et <i>rrnB</i> transformées par les plasmides MetaProd	231

7.4. Discussion	236
Conclusions et perspectives	241
8.1. Conclusions	241
8.1.1. Importance de l'équilibre entre croissance et production de protéines recombinantes	241
8.1.2. Influence de la production de protéines recombinantes sur la croissance bactérienne.	241
8.1.3. Analyse des contraintes métaboliques liées à la production de protéines recombinantes chez <i>E. coli</i> .	243
8.1.4. Comparaison des contraintes métaboliques de la production entre une culture en milieu riche et une culture en milieu minimum	244
8.1.5. Etude fonctionnelle du métabolisme en condition de production de protéines recombinantes en milieu minimum	244
8.1.6. Système MetaProd	248
8.1.7. Influence de la réponse stringente sur le métabolisme et la croissance	249
8.2. Perspectives	250
8.2.1. Optimisation du système MetaProd	250
8.2.2. Mise en œuvre d'un procédé semi-continu avec le système MetaProd	252
8.2.3. Modèle prédictif de production de protéines recombinantes avec le système MetaProd	253
Bibliographie	255

Table des illustrations

Figure 1.1 : Séquences d'acides désoxyribonucléiques (ADN) d'un vecteur d'expression assurant le minimum de fonctions nécessaires à l'expression de protéines recombinantes chez <i>E. coli</i> .	17
Figure 1.2 : Régulation du système d'expression T7 chez <i>E. coli</i> BL21 (DE3) avant ajout d'IPTG (gauche) et après ajout d'IPTG (droite).	19
Figure 1.3 : Régulation de l'origine de répllication de type ColE1 par les ARN structurés ARN I et ARN II.	20
Figure 1.4 : Plasmide de Sanofi Pasteur sur la base du plasmide pET.	22
Figure 1.5 : Influence de l'usage des codons sur la vitesse de traduction des ARNm.	23
Figure 1.6 : Eléments de contrôle de la vitesse de transcription du gène d'intérêt (vert), de la vitesse de traduction de la protéine d'intérêt (rouge) ou des deux vitesses (noir).	30
Figure 1.7 : Compétition entre production de protéines recombinantes et croissance bactérienne pour la consommation des ressources intracellulaires. <i>La traduction des protéines cellulaires et des protéines recombinantes nécessite la consommation d'acides aminés (violet), apportés par les ARN de transfert (ARNt) (bleu), d'énergie (vert) et l'utilisation de ribosomes (orange).</i>	33
Figure 1.8 : Modification de la taille des bactéries et des proportions des protéines impliquées dans le métabolisme (PM), des ribosomes (Rib) et des autres protéines (AP) parmi les protéines totales en fonction du taux de croissance.	38
Figure 1.9 : Augmentation du nombre de fourches de répllication lors de l'augmentation du taux de croissance d' <i>E. coli</i> .	40
Figure 1.10 : Régulation des proportions en protéines impliquées dans le métabolisme d' <i>E. coli</i> ou des ribosomes lors d'une culture en milieu riche (A) ou une culture en milieu minimum (B).	42
Figure 1.11 : Voies de synthèse et de dégradation du ppGpp et du pppGpp chez <i>E. coli</i> .	44
Figure 1.12 : Contrôle de la répllication cellulaire par le (p)ppGpp chez <i>E. coli</i> .	46
Figure 1.13 : Association des facteurs de transcription avec l'ARN polymérase (RNAP) d' <i>E. coli</i> en fonction de la concentration intracellulaire de ppGpp et du taux de croissance.	49
Figure 4.1 : Corrélation entre le poids sec par unité de DO_{600nm} (CDW (g) / DO_{600nm}) et le taux de croissance (μ , h^{-1}) (A) ou le nombre de cellules par unité de DO_{600nm} (nombre de cellules. 10^8 / DO_{600nm}) et le taux de croissance (B).	100
Figure 4.2 : Mesure par comptage du pourcentage de cellules viables par cytométrie après coloration au cours du procédé de production de la protéine UL130 en milieu minimum (induction précoce).	101
Figure 4.3 : Réalisation d'une droite de calibration de la quantité d'UL130 par quantité de biomasse (gUL130/gCDW) en fonction de la densitométrie (DO/mm^2) sur un gel SDS-PAGE.	103
Figure 4.4 : Comparaison des cinétiques de croissance bactérienne et de production de la protéine UL130 des cultures de la souche <i>E. coli</i> BL21 en milieu riche avec induction tardive (IT) ou sans induction (NI).	105
Figure 4.5 : Evolution de la croissance bactérienne et de la production de la protéine UL130 en milieu riche lors d'une induction précoce de la production.	107

Figure 4.6 : Etude de l'influence de l'utilisation d'un milieu minimum pour la production de la protéine UL130 en induction précoce.	110
Figure 4.7 : Gel SDS-PAGE des échantillons déposés au cours du procédé de purification de la protéine UL130 (18,4 kDa).	113
Figure 4.8 : Evolution de la croissance d' <i>E. coli</i> BL21 et de la production de la protéine Hla au cours du procédé de référence (milieu minimum, induction précoce).	115
Figure 4.9 : Corrélation entre le pourcentage de protéine insoluble parmi la production spécifique soluble et insoluble en fonction de la vitesse de production spécifique de la protéine Hla en milieu minimum ($R^2 = 0,69$).	117
Figure 4.10 : Analyse de corrélation entre la proportion de la fraction insoluble parmi la production spécifique soluble et insoluble et la production spécifique ($R^2 = 0,64$).	118
Figure 4.11 : Corrélations entre la production spécifique et le taux de croissance au cours des procédés de production des protéines UL130 (A) et Hla (B).	122
Figure 5.1 : Activité métabolique d' <i>E. coli</i> BL21 (pSP) en milieu minimum sans induction.	129
Figure 5.2 : Analyses de l'exométabolome des procédés de production des protéines UL130 et Hla.	132
Figure 5.3 : Evolution des concentrations en acides aminés au cours des procédés de production des protéines UL130 (A) et Hla (B).	138
Figure 5.4 : Influence de la composition en acides aminés de la protéine recombinante sur la consommation de chaque acide aminé.	141
Figure 5.5 : Calcul de la modification des flux de synthèses des acides aminés à partir de la différence de composition en acides aminés entre la protéine UL130 (rouge) ou Hla (bleu) et les protéines de la biomasse pour une vitesse de production des protéines recombinantes égale à la vitesse de production de la biomasse.	143
Figure 5.6 : Quantification du ppGpp au cours des procédés des productions des protéines UL130 (A) et Hla (B).	145
Figure 6.1 : Evolution des concentrations en acides aminés libres lors de l'activation de la réponse stringente causant un arrêt de croissance après ajout de SHX lors d'une culture en milieu minimum de la souche <i>E. coli</i> BL21.	179
Figure 6.2 : Evolution des concentrations en acides aminés libre lors de l'arrêt de croissance au cours de la production de la protéine UL130 lors d'une culture en milieu minimum de la souche <i>E. coli</i> BL21.	181
Figure 6.3 : Procédés de production de la protéine UL130 avec la souche <i>E. coli</i> BL21 WT (ronds noirs) et avec les souches $\Delta reIA$ (triangles noirs) et $\Delta gppA : \Delta reIA$ (triangles blancs).	183
Figure 6.4 : Etude de l'influence de la concentration en ppGpp intracellulaire au cours des procédés de production de la protéine Hla avec la souche <i>E. coli</i> BL21 WT et les souches $\Delta gppA$, $\Delta reIA$ et $\Delta gppA : \Delta reIA$.	186
Figure 6.5 : Quantification du nombre de copies de plasmide (PCN) par bactérie au cours des procédés de production de la protéine UL130 en milieu minimum par la souche <i>E. coli</i> WT ou mutées.	190
Figure 6.6 : Corrélation entre la production spécifique de la protéine UL130 en milieu minimum par la souche <i>E. coli</i> WT et le nombre de copies de plasmides (PCN/bactérie).	191
Figure 6.7 : Analyse de l'exométabolome en condition de production de la protéine Hla avec la souche <i>E. coli</i> WT et les souches mutées.	195

Figure 7.1 : Principe du système d'expression T7 (haut) avec la souche MetaProd (bas).	202
Figure 7.2 : Force d'expression des promoteurs des sept opérons ribosomaux d' <i>E. coli</i> (<i>rrn</i>).	205
Figure 7.3 : Séquences des promoteurs <i>rrnB</i> , <i>rrnD</i> et <i>rrnE</i> choisis pour remplacer le promoteur <i>lacUV5</i> avant le gène de la T7 RNAP de la souche <i>E. coli</i> BL21 (DE3).	208
Figure 7.4 : Analyse des cinétiques de croissance et de production spécifique au cours des procédés de production des protéines UL130 (gauche) et Hla (droite) par les souches <i>E. coli</i> BL21 WT, WT-400 pb et MetaProd (FIS- <i>rrnB</i> , <i>rrnB</i> , <i>rrnD</i> , <i>rrnE</i>) avec ajout de 1 mM d'IPTG (haut) ou des souches <i>E. coli</i> BL21 WT-400 pb et MetaProd (FIS- <i>rrnB</i> , <i>rrnB</i> , FIS- <i>rrnD</i> , <i>rrnD</i> , <i>rrnE</i>) avec ajout de 0,1 mM d'IPTG (bas) à T=0 h.	210
Figure 7.5 : Quantification du nombre de copies de plasmide (PCN) par bactérie et analyse du CER lors des procédés de production des protéines UL130 ou Hla par les souches <i>E. coli</i> BL21 WT (ronds noirs) ou <i>rrnE</i> (ronds blancs).	214
Figure 7.6 : Quantification des acides aminés dans le surnageant par RMN au cours des procédés de production de la protéine UL130 avec la souche WT en condition AA PCN (A), ou AA TOT (B) ou avec la souche <i>rrnE</i> en condition AA PCN (C).	216
Figure 7.7 : Effet de la supplémentation en acides aminés sur le procédé de production de la protéine UL130 par la souche WT après induction à l'IPTG (0,1mM).	217
Figure 7.8 : Comparaison des souches WT, <i>rrnB</i> et <i>rrnE</i> avec ajout d'acides aminés stabilisant le nombre de copies de plasmide (AA PCN) au cours du procédé de production de la protéine UL130 (IPTG 0.1mM).	221
Figure 7.9 : Comparaison des souches WT et <i>rrnE</i> avec ajout d'acides aminés (AA PCN) permettant de stabiliser le nombre de copies de plasmide (PCN) par bactérie au cours du procédé de production de la protéine Hla (IPTG 0.1mM).	224
Figure 7.10 : Principe du système d'expression MetaProd.	226
Figure 7.11 : Force d'expression des promoteurs d'enzymes impliquées dans la synthèse des acides aminés chez <i>E. coli</i> après ajout de SHX à t=0min (blanc), t=5min (gris clair), t=10min (gris foncé), t=30min (noir).	227
Figure 7.12 : Comparaison de la souche WT transformée par les plasmides modifiés (WT -PMP1 à 6) ou non modifié (WT -pSP) au cours du procédé de production de la protéine UL130 (IPTG 0.1mM).	230
Figure 7.13 : Comparaison des souches MetaProd <i>rrnE</i> ou <i>rrnB</i> transformées par les plasmides modifiés (PMP1, PMP2, PMP3, PMP4, PMP6) ou non modifié (pSP) au cours du procédé de production de la protéine UL130 (IPTG 0.1mM).	233
Figure 7.14 : Détermination du nombre de copie de plasmides (PCN) par bactérie au cours des procédés de production de la protéine UL130 avec la souche WT ou MetaProd <i>rrnE</i> et <i>rrnB</i> transformées par les plasmides MetaProd (PMP1, PMP2 et PMP6) ou le plasmide non modifié (pSP).	235
Figure 8.1 : Régulation du flux de synthèse et de consommation des acides aminés (violet) par le ppGpp à partir de la détection de la disponibilité en acides aminés intracellulaires.	242
Figure 8.2 : Intervention du ppGpp dans la réparation des flux de consommation des acides aminés entre la croissance (protéines de la biomasse) et la production de protéines recombinantes.	245
Figure 8.3 : Contraintes limitant la production de la protéine d'intérêt.	247

Index des tableaux

Tableau 1.1 : Classification des tags les plus utilisés suivant leur(s) fonction(s) et leur(s) propriété(s)	26
Tableau 1.2 : Composition macro-chimique d' <i>E. coli</i> en fonction du taux de croissance	37
Tableau 4.1 : Cinétiques de croissance et de production de la protéine UL130 obtenus lors de l'étude des procédés en milieu riche induction tardive, milieu riche induction précoce et milieu minimum induction précoce.	119
Tableau 4.2 : Phases caractéristiques des cinétiques de croissance et de production des procédés de production des protéines UL130 et Hla en milieu minimum.	121
Tableau 5.1 : Vitesses de consommation et d'excrétions des métabolites (mmol/g/h) au cours de la culture de la souche <i>E. coli</i> BL21 en milieu minimum.	127
Tableau 5.2 : Phases caractéristiques des cinétiques de croissance et de production des procédés de production des protéines UL130 et Hla en milieu minimum.	131
Tableau 6.1 : Vitesses spécifiques (qGlucose, $\mu\text{...}$) déterminées lors d'une croissance de la souche <i>E. coli</i> WT et des mutants ΔrelA , ΔgppA et $\Delta\text{relA} : \Delta\text{gppA}$.	193

Abréviations

μ	Taux de croissance
ACP	Acide gras synthase
ACP	Analyse des composantes principales
ADN	Acide désoxyribonucléique
AMP	Adénosine monophosphate
AMPc	Adénosine monophosphate cyclique
ARD	Recherche et développement analytique
ARN	Acides ribonucléiques
ARNm	Acides ribonucléiques messagers
ARNt	Acides ribonucléiques de transfert (ARNt)
ATP	Adénosine triphosphate
AU	Unité arbitraire
AUG	Codon d'initiation de la traduction
BM	Biomasse
BRD	Recherche et développement des bioprocédés
CDW	Biomasse sèche
CER	Taux d'enrichissement de l'air en dioxyde de carbone
CRP	C-réactive protéine
Da	Masse molécule en dalton
DO	Densité optique
ELISA	Immuno-absorption par enzyme liée
EMP	Emden-Meyerhof-Parnas
FAD	Flavine adénine dinucléotide
GC-MS	Chromatographie en phase gazeuse couplée à la spectrométrie de masse
GDP	Guanosine diphosphate
GMP	Guanosine monophosphate
GOGAT	Glutamine 2-oxoglutarate
GS	Glutamine synthase
GTP	Guanosine triphosphate
HPLC	Chromatographie en phase liquide à haute pression
HSQC	Corrélation des déplacements chimiques des ¹ H et ¹³ C par couplages directs
IC	Chromatographie ionique
IPTG	β-D-1-thiogalactopyranoside
Kan	Kanamycine
LB	Luria Bertani
Lrp	Protéine régulatrice de la réponse-leucine
MCS	Site multiple de clonage
MP	MetaProd
MR	Milieu riche

Abréviations

MS	Spectrométrie de masse
MT	Marqueur de taille
NADPH	Nicotinamide adénine dinucléotide phosphate
Ntr	Régulation nitrogène
ORF	Cadre de lecture ouvert
OUR	Taux de consommation de l'oxygène de l'air
pb	Paires de bases
PC	Composante principale
PCN	Nombre de copie de plasmide
PMP	Plasmide MetaProd
PS	Production spécifique
ppGpp	Guanosine 5'diphosphate,3'-diphosphate
pppGpp	Guanosine guanosine-3'-diphosphate,5'-triphosphate
PR	Protéine recombinante
pSP	Plasmide de Sanofi Pasteur
qPS	Vitesse de production spécifique
QR	Quotient respiratoire
RBS	Site de fixation des ribosomes
RMN	Résonance magnétique nucléaire
RNAP	Acides ribonucléiques polymérase
SD	Séquence Shine-Dalgarno
SD	Ecart type de la moyenne
SHX	Sérine hydroxamate
T7 RNAP	Acides ribonucléiques polymérase du bactériophage T7
Tag	Etiquette
TGS	Cible guide de synthèse
TMSP-d4	Triméthylsilylpropionate de sodium tétradeutééré
uDO	Unité de densité optique
UGA	Codon d'arrêt de la traduction
UTR	Région non traduit
WT	Souche sauvage

Acides aminés :

F	Phénylalanine
G	Glycine
f-Met	N-formylméthionine
H	Histidine
K	Lysine
L	Leucine
R	Arginine
T	Thréonine
Y	Tyrosine

Chapitre 1 – Introduction bibliographique

Depuis la découverte de son principe en 1879, la vaccination est l'intervention sanitaire la plus efficace et la moins coûteuse pour protéger l'Homme contre les infections bactériennes et virales (poliomyélite, tétanos, diphtérie, hépatite A ou B...) (Nokes & Anderson 1991; Tengs et al. 1995; Chen & Orenstein 1996). Le principe de la vaccination est basé sur l'injection d'agents infectieux inactivés, atténués ou d'antigènes purifiés dans l'organisme humain permettant au système immunitaire de générer des anticorps spécifiques dirigés contre le pathogène, protégeant préventivement contre son infection (Nascimento & Leite 2012). Cette stratégie rencontre cependant des limites pour conférer une immunité durable contre des pathogènes à grande variabilité génétique, contre des organismes à multiplication intracellulaire responsables le plus souvent de maladies chroniques ou pour déclencher une réponse immunitaire suffisante pour permettre la génération d'anticorps spécifiques (Lemaire et al. 2012). De plus, la purification des antigènes d'intérêts vaccinaux à partir d'organismes pathogènes conduit généralement à l'obtention de faibles quantités de protéines pouvant s'agréger avec des protéines non désirables, contaminant la purification (Nascimento & Leite 2012). Lorsque le vaccin repose sur l'injection d'un agent pathogène inactivé entièrement ou partiellement, un risque de forte virulence est présent chez certains patients. Une méthode alternative à la production traditionnelle de vaccins est basée sur l'introduction de la séquence génique de la protéine immunogène d'intérêt dans un organisme dit « hôte », non pathogène, chargé de produire en quantité élevée la protéine d'intérêt. La possibilité de produire ces antigènes dits « recombinants », sélectionnés rationnellement et parfois modifiés à façon, de combiner plusieurs variants d'une ou de plusieurs protéines, est une alternative ouvrant des solutions aux limites de la méthode traditionnelle de production des vaccins. La production de vaccins recombinants permet d'obtenir de fortes productions de l'antigène et ainsi d'enrichir la biomasse de l'hôte en protéine d'intérêt facilitant leur purification. L'utilisation d'organismes non pathogènes dans les procédés de production d'antigènes recombinants supprime les risques liés à la contamination des manipulateurs ou du produit final par des pathogènes virulents. Les organismes hôtes de production étant beaucoup plus étudiés, de nombreux outils sont disponibles pour optimiser les procédés de production de protéines recombinantes par l'intermédiaire de l'ingénierie protéique, ingénierie métabolique, la biologie de synthèse, etc... Dans cette nouvelle ère de

séquençage à haut débit de génomes d'organismes pathogènes, de développement de nouveaux outils statistiques et bio-informatiques, ces outils de production rendent applicables les recherches d'antigènes conservés au sein d'une espèce bactérienne ou virale, capables d'activer une réponse immunitaire dans le but de développer de nouveaux vaccins (Sette & Rino 2010).

Les organismes hôtes utilisés pour la production de protéines recombinantes sont regroupés en deux catégories : les cellules eucaryotes (cellules d'ovaires de hamster, humaines, d'insectes ou levures) et les cellules procaryotes (essentiellement *Escherichia coli*) (Assenberg et al. 2013). Ces deux types cellulaires ont des champs d'utilisations différents. La culture cellulaire eucaryote permet d'optimiser la production de protéines hétérologues avec des structures tertiaires complexes (multimère, haute masse moléculaire...) ou des modifications post-traductionnelles telles que les glycosylations (Demain & Vaishnav 2009). Cependant, la culture d'organismes eucaryotes à des échelles industrielles est coûteuse et les quantités de protéines recombinantes générées sont en moyenne relativement faibles (Legastelois et al. 2017). *E. coli* n'est pas capable de réaliser des glycosylations, ni de créer des ponts disulfures en dehors de son périplasme (à l'exception de quelques souches commerciales modifiées). Ces contraintes restreignent son utilisation à la production de protéines peu complexes structurellement et sans ou peu de modifications post-traductionnelles. La culture de cet organisme de production est cependant très économique et permet de produire des protéines recombinantes en grandes quantités.

Entre 2010 et 2014, 29 % de l'ensemble des produits biopharmaceutiques mis sur le marché ont été produit chez *E. coli* (Walsh 2014). Ces productions sont principalement réalisées avec le système d'expression génique T7 et la souche *E. coli* BL21 dans un procédé type discontinu (« batch »). Ce système, détaillé plus tard dans l'introduction, est basé sur l'utilisation d'un système d'expression fort garantissant la production rapide de grandes quantités de protéines d'intérêts (Rosano & Ceccarelli 2014). Cependant, la vitesse de production engendrée par ce système exerce d'importantes perturbations physiologiques et métaboliques qui peuvent se traduire par un arrêt de croissance (Cserjan-Puschmann et al. 1999). Les vitesses élevées de synthèse protéique peuvent également entraîner des problèmes de qualité de production telles que la formation de corps d'inclusions non souhaités, la production de protéines inactives ou tronquées (Baneyx & Mujacic 2004). Pour plusieurs

raisons explicitées dans la suite de l'introduction, une production de qualité nécessite le maintien d'une croissance bactérienne et d'une vitesse de production stable aux cours du procédé. Pour augmenter les quantités de protéines produites le procédé « batch » peut alors être prolongé en procédé semi-continu ou de type continu. Le principe des procédés semi-continus ou continus est d'alimenter la culture en substrat carboné pour prolonger la croissance bactérienne. Cette alimentation peut être réalisée de telle sorte que la vitesse d'alimentation soit inférieure à la vitesse maximale de consommation du substrat par les bactéries, permettant de contrôler le taux de croissance par limitation de la disponibilité du substrat et ainsi contrôler les mécanismes cellulaires associés à la croissance. L'équilibre entre croissance et production nécessaire à la mise en place de ce type de procédé est difficile à obtenir car la stabilisation de cet état nécessite une importante optimisation du système de production de façon spécifique aux particularités des protéines recombinantes à produire.

Dans un premier temps, le système d'expression du gène d'intérêt le plus utilisé pour la production de protéine recombinante (système T7) est présenté. Puis les caractéristiques de la souche hôte utilisée seront exposées. Enfin, les connaissances sur les contraintes métaboliques liées à la production de protéines recombinantes chez *E. coli* permettront de comprendre les stratégies de production permettant d'optimiser la qualité et la quantité des productions et les moyens à déployer pour y parvenir. Cette compréhension des contraintes métaboliques de la production permettra d'ouvrir l'introduction sur la présentation du mécanisme de gestion du taux de croissance de la bactérie régulé en fonction de la disponibilité des ressources issues du métabolisme bactérien.

1.1. Système d'expression génique de la protéine recombinante chez *E. coli*

La production de protéines recombinantes chez *E. coli* est basée sur l'intégration d'un système d'expression du gène d'intérêt dans une souche hôte. La souche est chargée de transcrire le gène d'intérêt puis de traduire les acides ribonucléiques messagers (ARNm) pour synthétiser la protéine d'intérêt. En général, la mise en place d'une production de protéine recombinante est initiée par la conception du système de transcription du gène d'intérêt (vecteur d'expression génique), de la

séquence du gène d'intérêt puis par le choix de la souche. La production de la protéine est alors évaluée en termes de quantité et de qualité, puis les éléments du système de production sont modifiés empiriquement dans le but de résoudre les problèmes identifiés. Des plateformes de production par cultures parallélisées sont parfois mises en place afin d'évaluer à haut débit de nombreux systèmes de production et de souches d'*E.coli* (Jia & Jeon 2016). Cette optimisation peut parfois être très longue en fonction des spécificités de la protéine recombinante à produire.

1.1.1. Vecteur d'expression

Les plasmides, supports de l'information génétique de la protéine à produire, permettent l'expression du gène, c'est-à-dire la transcription du gène d'intérêt et sont appelés vecteurs d'expression. La quantité d'ARNm du gène d'intérêt transcrit est dépendante de la force d'expression du vecteur. Au minimum, ces vecteurs sont composés d'une origine de réplication, d'un gène rapporteur permettant la sélection des clones transformés par le plasmide, d'un promoteur permettant l'expression du gène d'intérêt, d'un site de fixation du ribosome (RBS, « ribosome binding site ») suivi d'un site de clonage (MCS, « multiple cloning site ») composé en générale de plusieurs sites de restriction (figure 1.1). Dans ce site de clonage est introduit le cadre de lecture ouvert (ORF, « open reading frame ») du gène d'intérêt. L'ORF est suivi d'un site de terminaison de la transcription (figure 1.1).

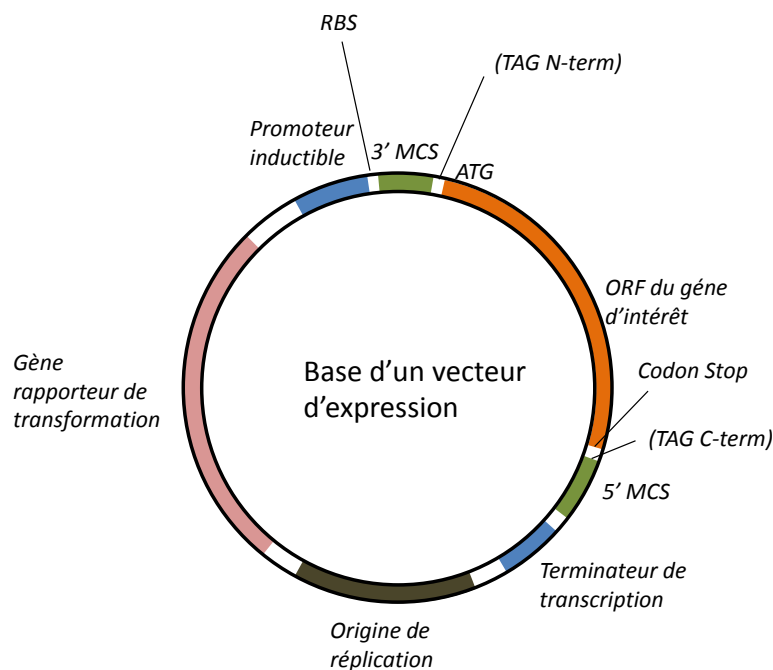


Figure 1.1 : Séquences d'acides désoxyribonucléiques (ADN) d'un vecteur d'expression assurant le minimum de fonctions nécessaires à l'expression de protéines recombinantes chez *E. coli*. Légende : site de fixation du ribosome (RBS, « ribosome binding site »), cadre de lecture ouvert (ORF, « open reading frame ») site multiple de clonage (MCS, « multiple cloning site »).

L'origine de répllication permet la multiplication du plasmide et son maintien dans la bactérie au cours de la division cellulaire. Un nombre élevé de copies de plasmide (PCN, « plasmid copy number ») engendre une forte expression mais ce nombre est soumis à une grande variabilité au cours de la culture (Rosano & Ceccarelli 2014; Gopal & Kumar 2013). A l'inverse, un faible PCN engendre une plus grande stabilité du nombre de copies mais avec une plus faible expression du gène d'intérêt. La force du système d'expression est déterminée par la force d'expression du promoteur et le nombre de copies de plasmide qui dépend des propriétés de l'origine de répllication. Selon la force d'expression désirée, le choix peut se porter sur une origine de répllication à haut PCN (50-100 copies de plasmide par bactérie), à PCN moyen (10-50 copies de plasmide par bactérie), ou faible PCN (1-10 copies de plasmide par bactérie).

Le deuxième élément du plasmide modulant la force d'expression du vecteur est le promoteur en amont du gène d'intérêt. Pour la production de protéine recombinante, ce promoteur est souvent inducible, c'est-à-dire non exprimé constitutivement et activé par un élément spécifique appelé

inducteur (β -D-1-thiogalactopyranoside (IPTG), arabinose, tétracycline, tryptophane...). Le choix du promoteur est déterminant pour la force d'expression du gène d'intérêt et la méthode de déclenchement de l'expression (Rosano & Ceccarelli 2014). Historiquement, le promoteur utilisé pour l'expression de protéines recombinantes est le promoteur *lac* de l'opéron lactose d'*E. coli* (Sørensen & Mortensen 2005). Ce promoteur, activé exclusivement en présence de lactose ou de son homologue non métabolisable l'IPTG, permet une expression modérée du gène d'intérêt. Son utilisation a été abandonnée à cause de sa sensibilité à la répression catabolique (par l'adénosine monophosphate cyclique (AMPC)-C-réactive protéine (CRP) en présence de glucose) et sa trop faible répression de l'expression du gène en absence de lactose engendrant des fuites d'expression du gène d'intérêt. La mutation de ce promoteur a permis de supprimer la sensibilité du promoteur à la répression catabolique (*lacUV5*) mais pas les fuites d'expression. Afin de résoudre ce problème tout en amplifiant la force d'expression du gène d'intérêt, le système T7 a été développé (Studier & Moffatt 1986). Ce système est actuellement le plus utilisé pour la production de protéines recombinantes et en particulier, est celui utilisé dans le cadre de ce travail de thèse.

1.1.2. Présentation du système d'expression T7

Le système d'expression T7 permet une forte production de protéines recombinantes avec une très bonne répression du système d'expression en absence de lactose, ou de son homologue l'IPTG (Studier & Moffatt 1986). Ce système, libre de droit depuis 2014, est basé sur une intégration du système transcriptionnel du bactériophage T7 au système transcriptionnel d'*E. coli* (Studier et al. 1990; Studier et al. 1997; Studier & Dubendorff 1998). Le gène codant pour l'ARN polymérase viral T7 (T7 RNAP) est inséré dans la zone du prophage DE3 du génome d'*E. coli* BL21 sous la régulation du promoteur *lacUV5* (figure 1.2).

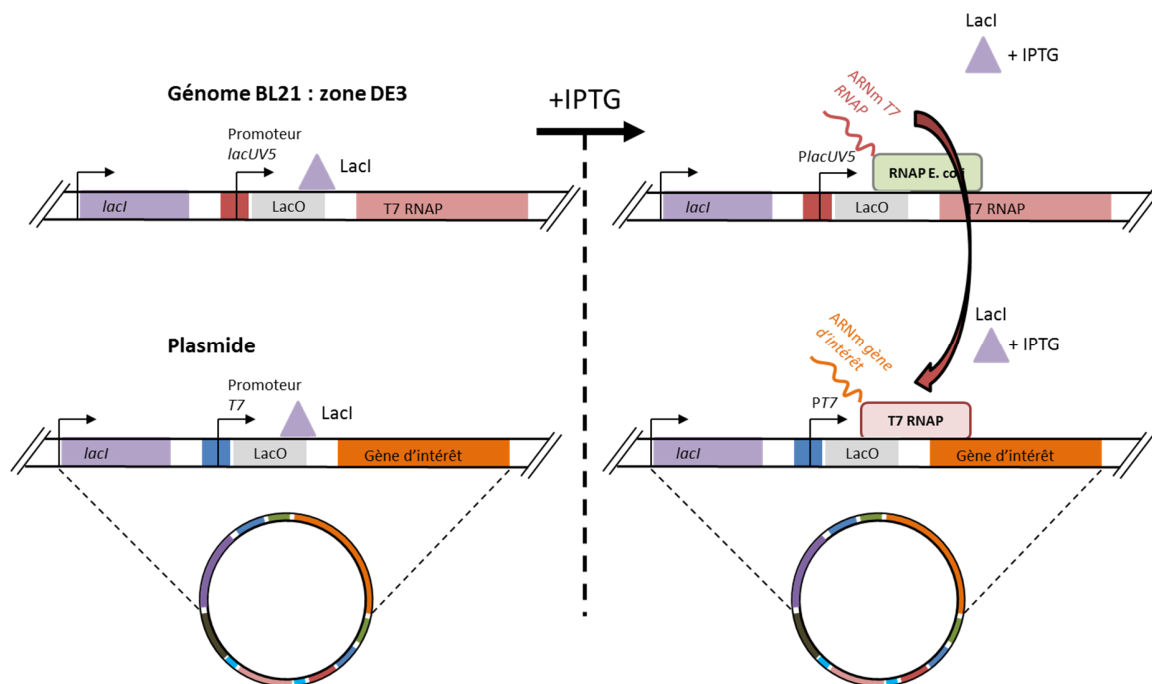


Figure 1.2 : Régulation du système d'expression T7 chez *E. coli* BL21 (DE3) avant ajout d'IPTG (gauche) et après ajout d'IPTG (droite).

Le promoteur *lacUV5* permet la transcription de la T7 RNAP par l'ARN polymérase (RNAP) d'*E. coli*. Il est suivi par le site d'inhibition *lacO*, site de fixation de la protéine LacI. Cette protéine permet l'inhibition de la transcription en absence d'IPTG. Le gène *lacI* est présent sur le plasmide et en aval du gène *T7 RNAP* dans le génome. L'IPTG, ajouté dans le milieu, se lie à la protéine LacI induisant un changement de conformation de cette protéine qui libère alors les sites *lacO*. Ainsi la T7 RNAP est produite permettant la transcription du gène d'intérêt placé sur le plasmide sous le contrôle du promoteur T7 (Rosano & Ceccarelli 2014).

1.1.3. Origine de répllication de type ColE1

Les vecteurs d'expression couramment utilisés avec le système d'expression T7 possèdent une origine de répllication pBR322 classée dans la famille des origines de répllication ColE1. Cette famille est caractérisée par une séquence *ori* très conservée. Ce type d'origine de répllication permet le maintien de 10 à 50 copies de plasmide par cellule mais il est soumis à une forte variabilité de PCN en condition de carence en acides aminés (Camps 2010a).

Le mécanisme d'initiation de la réplication fait intervenir un grand nombre d'étapes et de protéines. Succinctement, la réplication débute par la transcription du promoteur P2 générant un brin d'ARN anti-sens appelé ARN II (Polisky 1988) (figure 1.3).

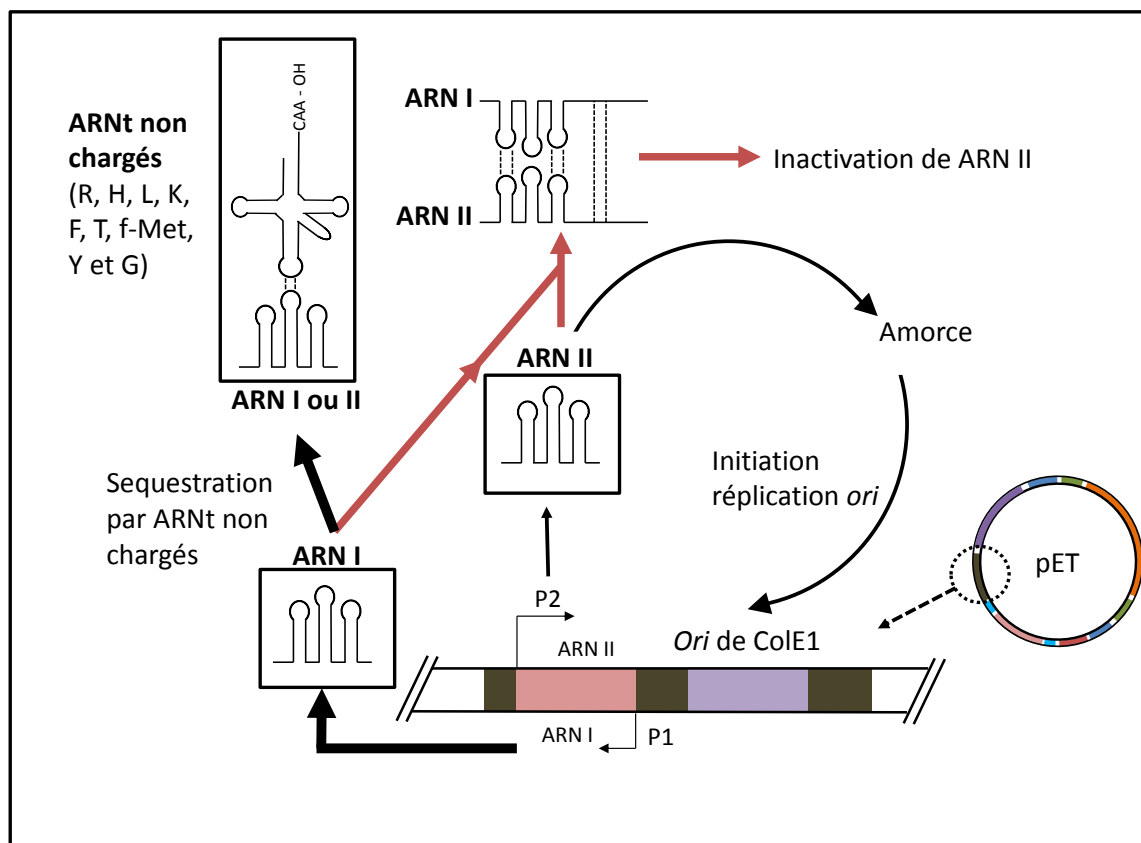


Figure 1.3 : Régulation de l'origine de réplication de type ColE1 par les ARN structurés ARN I et ARN II. Influence des acides ribonucléiques de transfert (ARNt) non chargés sur la régulation de l'origine de réplication par les ARN I et II conduisant à l'augmentation de la réplication du plasmide pET (R : arginine, H : histidine, L : leucine, K : lysine, F : phénylalanine, T : thréonine, f-Met : N-formylméthionine Y : tyrosine, G : glycine).

Le brin ARN II, après plusieurs étapes supplémentaires, est à l'origine de la synthèse d'une amorce initiant la réplication du plasmide. La séquence ARN II est également transcrite par un deuxième promoteur sur le brin sens, promoteur P1, 100 fois plus fort que le promoteur P2. Le deuxième brin sens, appelé ARN I, s'hybride avec l'ARN II pour former un complexe entre les ARN I et II. La formation de ce complexe empêche à l'ARN II d'initier la réplication du plasmide et cela conduit à l'inhibition de la réplication maintenant un nombre de PCN à environ 10 à 50 par bactérie. Ces brins d'ARN, structurés en trois boucles, sont très proches des acides ribonucléiques de transfert (ARNt).

Une homologie de plus de 40% a été montrée entre la boucle d'anticodon des ARNt et la deuxième boucle de l'ARN I (pour les acides aminés arginine, histidine, leucine, lysine, phénylalanine et thréonine), de l'ARN II (N-formylméthionine, tyrosine et glycine) ou des deux (méthionine, valine). Par cette homologie, certains ARNt seraient capables d'inhiber la formation du duplex ARN I et ARN II conduisant à une augmentation des PCN. Des études ont démontré une augmentation des PCN en condition de carence en acides aminés et donc en présence d'ARNt non chargés plus nombreux (Stifling et al. 1988; Wrobel & We 1998; Wang et al. 2002; Camps 2010a). Des modèles tentent d'expliquer pourquoi les ARNt non chargés auraient un pouvoir plus important de déstabilisation du duplex ARN I/ ARN II que les ARNt chargés, mais à ce jour aucune donnée n'a permis d'expliquer ce mécanisme.

Un grand nombre de copies de plasmide entraîne une importante force d'expression du gène d'intérêt. Si une augmentation des PCN est causée par la présence d'ARNt non chargés, et donc par une limitation en acides aminés, l'augmentation de l'expression de la protéine recombinante aura l'effet d'un cercle vicieux sur la limitation en acides aminés.

1.1.4. Plasmide de Sanofi Pasteur

Dans un contexte de production de protéines recombinantes, les antibiotiques sont utilisés dans un double intérêt. Le premier est de sélectionner les bactéries ayant intégrées le plasmide d'expression. Le deuxième permet d'appliquer une pression de sélection au cours du procédé pour garantir le maintien du plasmide dans les bactéries au cours de leur croissance. La perte de plasmide est principalement causée par un problème de ségrégation des plasmides répliqués qui, une fois la bactérie divisée, sont retenus par une seule des deux cellules filles. Cette mauvaise ségrégation du plasmide est induite par l'entrelacement des plasmides répliqués, phénomène appelé « dimère catastrophe » (Summers et al. 1993; Field & Summers 2011). Malgré cette utilité des antibiotiques, les procédés de production de protéines à intérêt biopharmaceutique doivent exclure leur emploi pour des raisons réglementaires.

Le plasmide de Sanofi Pasteur (pSP) est un vecteur utilisant le système d'expression pET avec l'ajout d'un système évitant la perte du plasmide lors de la division cellulaire (figure 1.4).

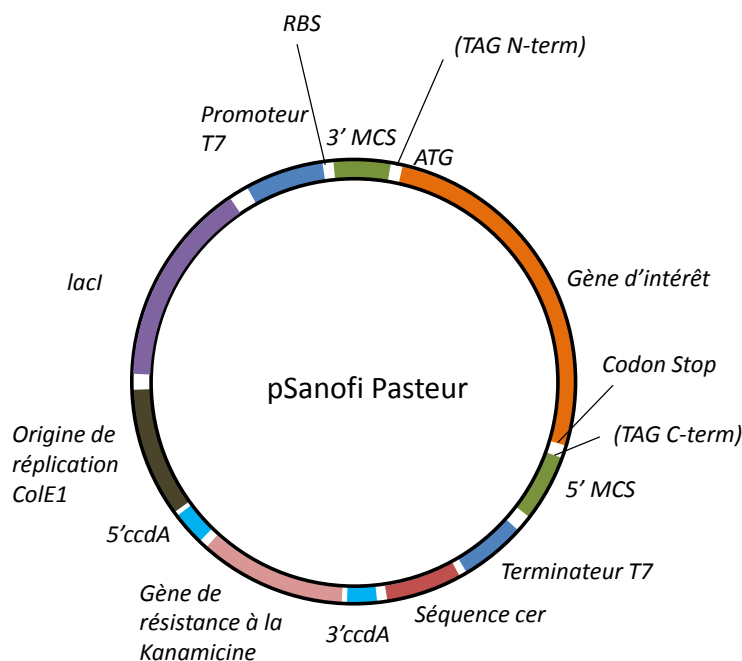


Figure 1.4 : Plasmide de Sanofi Pasteur sur la base du plasmide pET. Le locus *cer* a été ajouté pour garantir la conservation du plasmide par les bactéries lors de la division cellulaire ainsi que les séquences 5' et 3' *ccdA* (antitoxine *CcdA*) permettant l'excision du gène de résistance à la kanamycine *in vivo*.

Ce système est basé sur l'ajout d'une séquence *cer* permettant d'améliorer la ségrégation des plasmides répliqués et favorisant ainsi la présence du plasmide dans chaque cellule fille, même en absence d'antibiotique (Peubez et al. 2010). Le locus *cer* est un site reconnu par la recombinase XerC-XerD capable de couper la séquence *cer* spécifiquement puis d'effectuer la ligation afin de libérer les plasmides dimérisés suite à la répllication (Stifling et al. 1988). La recombinase, codée dans le génome d'*E. coli*, a pour fonction première de résoudre les problèmes de dimérisation du chromosome bactérien lors de sa répllication (Peubez et al. 2010).

1.1.5. Conception de la séquence du gène d'intérêt

Au cours de la traduction des ARNm, chaque acide aminé peut-être codé, le plus souvent, par plusieurs triplets d'acides nucléiques (codon). Chaque organisme possède une utilisation préférentielle de codons pour chaque acide aminé appelé biais d'usage du code. Ce biais de codon est un critère très important dans la production de protéines recombinantes puisqu'il s'agit de faire produire une

protéine souvent synthétisée par un organisme différent de l'organisme hôte recombinant. L'usage des codons du gène d'intérêt a un impact très important sur la vitesse de transcription du gène et de traduction des ARNm (Demain & Vaishnav 2009; Sørensen & Mortensen 2005; Huang et al. 2012; Graumann & Premstaller 2006; Gopal & Kumar 2013; Rosano & Ceccarelli 2014; Baneyx & Mujacic 2004; Assenberg et al. 2013).

Chez *E. coli*, la fréquence d'usage de chaque codon est proportionnelle au nombre d'ARNt possédant l'anticodon associés (Stoletzki & Eyre-walker 2007). La vitesse de traduction est influencée par la disponibilité des ARNt chargés en acide aminé. Un usage de codon peu fréquent, avec des ARNt associés en faible nombre, conduit à un temps d'attente des ribosomes pour obtenir l'acide aminé associé au codon (Yu et al. 2015). Ce temps d'attente est susceptible d'entraîner de mauvaises incorporations d'acides aminés par utilisation du mauvais ARNt. Ces temps de pause permettent cependant de garantir des temps de repliement à des domaines caractéristiques de la protéine en ralentissant la vitesse de traduction. L'usage des codons les plus fréquents conduit à de grandes vitesses de traduction de l'ARNm car les ARNt associés sont disponibles. Si les ARNt restent disponibles, les erreurs associées à de mauvaises incorporations sont réduites (figure 1.5).

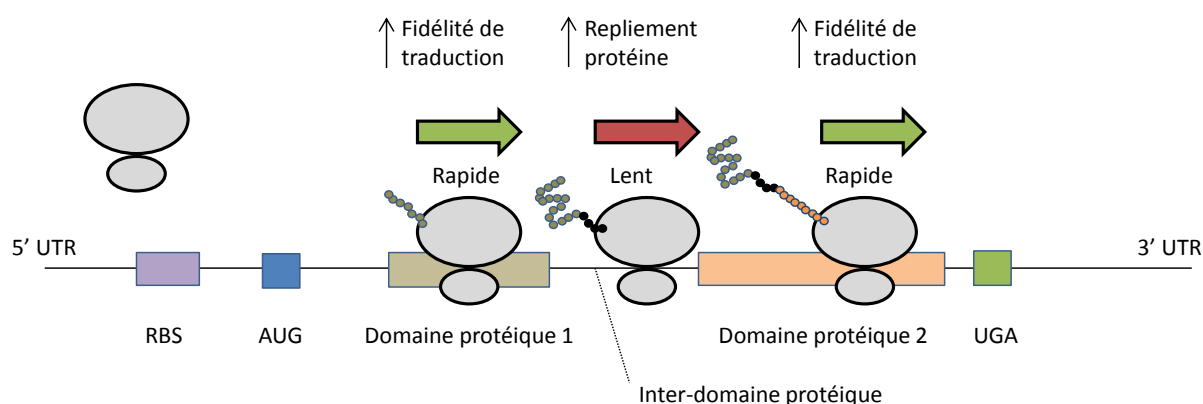


Figure 1.5 : Influence de l'usage des codons sur la vitesse de traduction des ARNm. L'usage des codons les plus utilisés dans les séquences codant les domaines protéiques augmente la vitesse de traduction et évite les erreurs d'incorporation des acides aminés grâce à un grand nombre d'ARNt-chargés disponibles. Au contraire, l'usage de codons rares engendre un ralentissement de la traduction améliorant le repliement de la protéine. L'utilisation de codons rares peut engendrer des erreurs d'incorporation des acides aminés sans influence sur l'activité de la protéine d'intérêt lorsqu'elles sont présentes sur les inter-domaines de protéine. Légende : 5' UTR (région 5' non traduite), RBS (ribosome

binding site), AUG (codon d'initiation de la traduction), UGA (codon d'arrêt de traduction), 3' UTR (région 3' non traduite). Schéma inspiré de Hanson & Collier 2017.

L'usage des codons majoritaires augmente la vitesse de traduction et donc la vitesse de production de la protéine d'intérêt. La pratique actuellement conseillée est d'utiliser des codons majoritaires pour la traduction des domaines protéiques afin d'éviter les erreurs d'incorporation sur les domaines importants pour l'activité et, d'utiliser des codons plus rares dans les domaines interprotéiques afin d'optimiser le repliement des domaines protéiques (Hanson & Collier 2017) (figure 1.5). L'usage des codons n'est pas le seul critère pouvant influencer la vitesse de traduction. La formation de structures secondaires au niveau de l'ARNm, la présence de RBS internes à la séquence, la richesse en GC, la stabilité de l'ARNm sont autant de critères importants à considérer pour le choix des codons (Hanson & Collier 2017).

La conception de la séquence a également une influence au niveau transcriptionnel. Le choix des codons ne doit pas entraîner l'apparition de séquence Shine-Dalgarno (SD) interne, de TATA box, de séquences terminales ou des répétitions inversées caractéristiques des séquences d'insertion. La proportion de chaque acide nucléique (G et C particulièrement) ou de motifs répétés sont également des éléments de régulation de la vitesse transcriptionnelle.

En dehors de l'ORF, la spécificité du RBS influence l'affinité des ribosomes pour la fixation sur l'ARNm, et les séquences non codantes en 3' et 5' de l'ARNm modifient la stabilité de l'ARNm et sa disponibilité pour la fixation des ribosomes (Sørensen & Mortensen 2005).

L'optimisation de l'usage des codons est en général réalisée par des algorithmes essayant d'obtenir le meilleur score en donnant un poids à chacun de ces critères. Le poids des critères peut être modifié en fonction de la protéine à produire.

1.1.6. « Etiquettes » d'expression

Les étiquettes (tags) sont des peptides ou des protéines pouvant être placées en C-terminal ou en N-terminal des protéines recombinantes par ajout de leur séquence en 5' ou en 3' de la séquence du gène d'intérêt. Elles sont utilisées pour faciliter la purification de la protéine d'intérêt, pour leur

détection et pour résoudre des problèmes de qualité des protéines produites (solubilité, repliement...) (Kimple et al. 2015; Young et al. 2012).

De façon générale, les tags de purification et/ou de détection sont basés sur l'ajout d'un épitope ou peptide spécifique reconnu par un anticorps et ajoutent une forte affinité à la protéine pour un support chromatographique (tableau 1)(Zakeri et al. 2012; Kimple et al. 2015). La détection des tags est permise par une activité catalytique de luminescence ou de fluorescence ou d'un lien covalent avec une autre protéine spécifiquement. D'autres tags permettent d'ajouter à la protéine une affinité spécifique à une matrice permettant l'utilisation d'une chromatographie d'affinité. Malgré l'absence d'explication claire quant à leur mécanisme, certains tags sont utilisés pour solubiliser la protéine d'intérêt. La prise en charge par des protéines chaperonnes ou la prise en charge par des protéines chargées d'exporter la protéine dans le périplasme ou dans le milieu extracellulaire se fait grâce à l'ajout de tags et sont également beaucoup utilisées chez *E. coli* pour la formation des ponts disulfures et le repliement de protéines complexes (telles que les anticorps) (Malik 2016). Des sites de clivages sont également ajoutés après ou avant les tags en C- ou N-terminal respectivement (Enterokinase, thrombine, facteur Xa and tobacco etch virus) car la présence des tags peut parfois engendrer des problèmes d'activité, de structure ou d'immunogénicité de la protéine d'intérêt.

Tableau 1.1 : Classification des tags les plus utilisés suivant leur(s) fonction(s) et leur(s) propriété(s), (synthèse des tableaux présentés dans Zakeri et al. 2012 et Kimple et al. 2015).

Fonctions	Propriétés	tags
Purification/détection	Epitope pour reconnaissance par anticorps	FLAG, Green Fluorescent Protein (GFP), Glutathione S-transferase (GST), 6xHIS, Human influenza Hematogglutinine (HA), Myc, bacteriophage V5 epitope, epitope HSV, Horseradish peroxidase (HRP)...
	Forte affinité	Biotine-Streptavidine
Détection	Activité catalytique	LacZ, Alkaline Phosphatase (AP), Chloramphenicol Acetyl Transferase (CAT), GST, HRP
	Luminescence	luciférase
	Fluorescence	GFP, Red Fluorescent Protein
	Lien covalent	HaloTag, CnaB2
Purification	Affinité support chromatographie	albumin binding protein (ABP), AP, CAT, cellulase binding domain, choline binding domain, galactose binding domain, maltose binding domain (MBP), polyhistidine, polyarginine, polycysteine, polyphenylalanine...
Solubilisation de la protéine d'intérêt	Inconnu	Trx, SUMO, GST, MBP, NusA

Le tag histidine (enchaînement de 6 histidines, 6xHis) est le plus utilisé pour la détection et la purification de la protéine d'intérêt. Il permet de fixer la protéine par interaction des noyaux imidazoles des histidines avec des ions métalliques divalent (cuivre, nickel, cobalt, zinc) de la phase stationnaire. L'élution de la protéine de la colonne chromatographique se fait par compétition avec de l'imidazole. Des anticorps spécifiques anti-poly histidines permettent la détection des protéines et leur quantification par dosage d'immuno-absorption par enzyme liée (ELISA) ou western-blot. Enfin, le tag 6xHis entraîne une utilisation rapide des anticodons ARNt-histidines anti-CAC ou anti-CAU. Le faible

nombre d'ARNt anti-CAC et le plus faible encore nombre d'ARNt anti-CAU permet de ralentir la traduction du tag 6xHis (Navon & Pilpel 2011). Le tag 6xHIS n'est pas connu pour améliorer la solubilité de la protéine.

1.1.7. Souches d'*Escherichia coli*

1.1.7.1. Comparaison des souches E. coli BL21 et K-12

Les souches *E. coli* BL21 et K-12 sont les souches d'*E. coli* les plus utilisées pour la production industrielle de protéines recombinantes. Ces souches ont une homologie génomique supérieure à 99.1% avec des génomes de tailles similaires (4,6 millions de paires de bases (pb) pour la souche K-12 MG1655 et 4,5 Mpb pour la souche *E. coli* BL21 (www.genome.jp)). Les gènes impliqués dans la synthèse du flagelle sont absents dans la souche *E. coli* BL21 (Yoon et al. 2012). Les gènes de la protéase ribosomique Lon et de la protéase extracellulaire OmpT sont inactivés dans cette souche par des séquences d'insertion (éléments génétiques mobiles) ce qui a pour conséquence d'améliorer la stabilité des protéines recombinantes produites. Les souches *E. coli* K-12 présentent davantage de gènes impliqués dans le catabolisme des molécules aromatiques et dans la biosynthèse des lipopolysaccharides (Li et al. 2014a)(Yoon et al. 2012). Au niveau transcriptionnel, une différence d'expression de 347 gènes sur les 3882 gènes partagés par les deux souches a été observée (Marisch, Bayer, Cserjan-Puschmann, et al. 2013). Ces gènes sont impliqués dans le métabolisme du glucose, l'acquisition du fer et la mobilité. La faible sécrétion d'acétate chez *E. coli* BL21 a été montrée comme le résultat de quelques mutations sur les promoteurs de facteurs de régulation du métabolisme central (ArcA, CRP et Cra) permettant un métabolisme catabolique du glucose plus efficace et plus adapté à la production protéique (Yoon et al. 2012; Waegeman et al. 2011; Wittmann et al. 2007). Lors de cultures en bioréacteur, les analyses macro cinétiques associées à la culture de la souche *E. coli* BL21 ou de souches *E. coli* K-12 montrent un taux de croissance supérieur pour la souche *E. coli* BL21, ainsi qu'un meilleur rendement carboné en biomasse (avec 60% pour *E. coli* BL21 contre 50% pour les souches K-12) qui serait en partie dû à une plus faible sécrétion d'acétate (1.3% du carbone consommé pour *E. coli* BL21 contre 10 à 15% pour les souches *E. coli* K-12) (Marisch, Bayer, Scharl, et al. 2013). En condition de production de protéines recombinantes dans un environnement régulé, le système

d'expression T7 permet à la souche *E. coli* BL21 la synthèse d'une plus grande quantité de protéines avec une meilleure qualité de protéine produite de façon soluble par rapport à la souche *E. coli* K-12 (Marisch, Bayer, Cserjan-Puschmann, et al. 2013).

1.1.7.2. Souche *E. coli* BL21

La souche *E. coli* BL21 (B F⁻ *ompT gal dcm lon hsdS_B(r_B⁻m_B⁻) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB⁺]_{K-12}(λ^S)), étudiée au cours de cette thèse est la souche la plus utilisée pour la production de protéines recombinantes (Rosano & Ceccarelli 2014; Studier & Moffatt 1986). Cette souche dérive de la souche *E. coli* B, sélectionnée par Studier et Moffatt en 1986 pour servir de modèle d'étude à la sensibilité de la souche aux phages, aux systèmes de restriction, à l'évolution bactérienne et à la production de protéines recombinantes. La souche *E. coli* BL21 montre de très bonnes performances pour l'expression de protéines recombinantes. En effet, en aérobie, cette souche se caractérise par une faible accumulation d'acétate dans le milieu de culture malgré une forte concentration en glucose. De plus des mutations ponctuelles sur les promoteurs de régulation de cofacteurs transcriptionnels (ArcA, Crp, Cra) engendrent une plus grande activité des voies anaérobies du cycle de Krebs améliorant la disponibilité des précurseurs pour la synthèse des acides aminés par rapport à d'autres souches d'*E. coli* (Waegeman et al. 2011; Castaño-Cerezo et al. 2015; Yoon et al. 2012; Marisch, Bayer, Cserjan-Puschmann, et al. 2013; Li et al. 2014a; Noronha et al. 2000).*

De nombreuses autres souches d'*E. coli*, dérivées de la souche *E. coli* BL21, sont commercialisées et revendiquées comme étant capables de résoudre des problèmes spécifiques de production grâce à une meilleure stabilité des ARNm (Star™), à des croissances à faible température (ArticExpress), à une plus grande richesse d'ARNt (Rosetta™, Condon plus®), à une meilleure formation de ponts disulfures (Shuffle®), à une plus grande résistance à la toxicité des protéines (Origami™) ou à une capacité à stocker les protéines membranaires (C41 et C43) (Dumon-seignovert et al. 2004; Gopal & Kumar 2013; Jia & Jeon 2016).

1.2. Intégration du système d'expression T7 dans une souche d'*E. coli*

Une fois le système d'expression T7 intégré dans la souche, la régulation de la force d'expression du gène est exclusivement soumise à la concentration intracellulaire d'IPTG. Le système d'expression est donc entièrement découplé de la régulation bactérienne. Le développement de procédé de production performant et de qualité repose sur la mise en accord du système d'expression et les capacités métaboliques de l'hôte pour chaque protéine.

1.2.1. Importance de l'équilibre entre croissance et production

Généralement, deux approches sont utilisées pour le développement de la production de protéines recombinantes. La première nécessite un temps de développement moins important du système d'expression et de la souche en fonction de la protéine. Cette approche consiste à induire fortement la production à une biomasse élevée en fin de procédé « batch ». Dans ce cas, la croissance est perturbée par la forte production entraînant parfois l'arrêt total de la croissance. Cette approche permet de générer de grandes quantités de protéines dans un temps court mais elle ne permet pas de se placer dans les meilleures conditions pour assurer la qualité de synthèse de la protéine, *a contrario* de la deuxième approche (Sandén et al. 2003). Cette deuxième approche consiste à mettre en adéquation la force d'expression génique avec les capacités de synthèse de l'hôte afin que l'induction des protéines recombinantes n'entraîne pas une consommation trop importante des ressources bactériennes. Cette approche nécessite une importante optimisation du système de production destinée à maintenir la croissance de la bactérie avec une vitesse de production amoindrie. Dans ce cas, l'induction est réalisée de façon à laisser un temps important de croissance et de production qualitative de protéines en grande quantité (Sharma et al. 2011; Norsyahida et al. 2009; Boström et al. 2005). De plus, le maintien d'une croissance bactérienne lors de la production de la protéine recombinante ouvre la possibilité de maîtriser le taux de croissance par une alimentation limitante en substrat au cours d'un procédé semi-continu ou continu et ainsi de contrôler la vitesse de production et donc la qualité de la protéine (Marschall et al. 2016).

Une protéine recombinante de qualité est une protéine recombinante en tout point identique à la protéine native produite par l'hôte d'origine. Plus particulièrement, il s'agit de respecter la structure de la protéine, son intégrité et la fidélité de sa séquence. La vitesse de production de la protéine recombinante peut influencer la qualité de production. La vitesse de production dépend de la vitesse de transcription du gène d'intérêt puis de la vitesse de traduction de la protéine d'intérêt.

La vitesse de transcription est définie par le nombre d'ARNm transcrits par unité de temps. Celui-ci dépend de la fréquence de l'initiation de transcription du gène d'intérêt (force du promoteur du gène d'intérêt), de la vitesse de transcription du gène d'intérêt par l'ARN polymérase, du nombre de copies de plasmide et de la quantité d'inducteur du système d'expression ajoutée à la culture (figure 1.6).

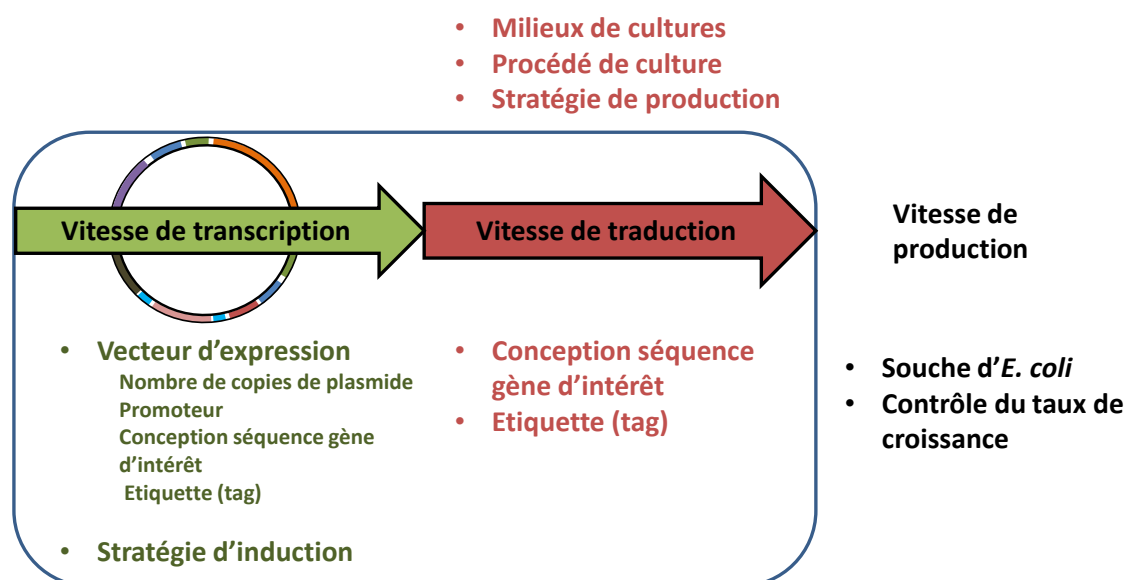


Figure 1.6 : Éléments de contrôle de la vitesse de transcription du gène d'intérêt (vert), de la vitesse de traduction de la protéine d'intérêt (rouge) ou des deux vitesses (noir). (Graumann & Premstaller 2006; Choi et al. 2006; Rosano & Ceccarelli 2014; Young et al. 2012).

La vitesse de traduction dépend de la vitesse de traduction des ARNm, du nombre d'ARNm traduits simultanément. La vitesse de traduction de chaque ARNm est influencée par l'usage des codons de la séquence d'intérêt. Comme précédemment explicité, cet usage influence la fidélité de

traduction et le repliement de la protéine en fonction de la vitesse de traduction. Le nombre d'ARNm va dépendre de la vitesse de transcription du gène d'intérêt.

La vitesse de traduction dépend du nombre de ribosomes disponibles, de la vitesse de traduction de chaque ARNm par les ribosomes et de la capacité de l'hôte à apporter toutes les ressources nécessaires à la traduction (ARNt, acides aminés, énergie). La vitesse de traduction de chaque ARNm est influencée par l'usage des codons de la séquence d'intérêt comme explicité précédemment. Une vitesse de traduction trop élevée conduit à une baisse de la disponibilité des ressources cellulaires ce qui peut entraîner des erreurs d'incorporation des acides aminés (problème de structure, d'activité, d'immunogénicité) ou des arrêts de traduction précoces (protéines tronquées) (Marschall et al. 2016). Cet aspect de disponibilité des ressources est peu étudié dans la littérature et sera développé dans le paragraphe suivant ainsi que le principe de capacité de synthèse de la protéine par l'hôte.

Une vitesse de transcription plus grande que la vitesse de traduction nécessaire pour traduire tous les ARNm d'intérêt va conduire à une accumulation des ARNm d'intérêt puis à une compétition en faveur des ARNm de la protéine d'intérêt plutôt que des ARNm des protéines de la biomasse bactérienne. Une traduction trop faible des ARNm des protéines cellulaires va entraîner une diminution de la quantité de protéines nécessaires à la croissance et donc un arrêt de croissance.

La maîtrise d'une production de grande quantité de protéines de qualité ne peut être atteinte que par l'obtention d'un équilibre durable entre vitesse de production en accord avec la capacité de synthèse de la protéine par l'hôte. De par la connaissance fine du génome d'*E. coli* et de ses mécanismes cellulaires, de nombreux leviers d'actions ont été utilisés pour contrôler les vitesses de transcription et de traduction de la protéine d'intérêt dans cette souche (figure 1.6). Ces moyens de contrôle, développés dans la première partie de l'introduction, reposent sur le choix du vecteur d'expression (incluant le promoteur du gène d'intérêt, l'origine de répllication, l'ajout d'une étiquette, la modification de l'usage des codons de la séquence du gène), mais aussi sur le choix du milieu de culture, du procédé de culture employé, du contrôle du taux de croissance ou du type de souches utilisé (Graumann & Premstaller 2006; Choi et al. 2006; Rosano & Ceccarelli 2014; Young et al. 2012) (figure 1.6). Chaque protéine recombinante possède sa spécificité de séquence et de structure. Ces

moyens de contrôle de la vitesse de production sont donc à adapter empiriquement pour chaque protéine. Certaines productions de protéines sont toxiques pour la bactérie et conduisent à un arrêt de croissance rapide lors de la production de faibles quantités de protéines. La cause de la toxicité des protéines étant dépendante de chaque protéine, cette problématique ne sera pas prise en compte dans cette introduction.

De façon générale, les revues scientifiques et la recherche dans le domaine de la production de protéines recombinantes s'arrêtent à la présentation ou à l'étude des mécanismes explicités précédemment. L'étude de l'influence de la production de protéines recombinantes sur le métabolisme bactérien est en général peu étudiée et encore moins avec la souche *E. coli* BL21 comme modèle d'étude par rapport à la souche *E. coli* K-12. L'analyse des procédés de production s'arrête au constat que si la production entraîne un arrêt de croissance alors la vitesse de production est trop élevée pour la bactérie. Une baisse de la vitesse de production est alors appliquée au procédé par plusieurs options. La suite de cette introduction a pour objectif de présenter l'ensemble des travaux portant sur l'influence de la production de protéines recombinantes sur le métabolisme d'*E. coli*. Ensuite, la régulation de la croissance d'*E. coli* est présentée afin de comprendre les conséquences des contraintes métaboliques engendrées par la production sur la croissance bactérienne.

1.2.2. Charges métaboliques de la production

Les protéines sont des polymères d'acides aminés liés par des liaisons peptidiques. Leur synthèse est assurée par les ribosomes qui permettent de coupler la traduction des ARNm avec la liaison ordonnée des acides aminés apportés par les ARNt. Les protéines sont les composants principaux de la biomasse bactérienne (Bremer & Dennis 2008; Scott et al. 2010). La croissance bactérienne sera déterminée par la vitesse de transformation des nutriments de l'environnement en acides aminés puis en protéines de la biomasse (Waschina et al. 2016; Hui et al. 2015; Bosdriesz et al. 2015; Scott et al. 2014; Bremer & Dennis 2008; Mahalik et al. 2014). Le déclenchement de la synthèse d'une protéine recombinante en plus de la synthèse des protéines de la biomasse repose sur l'utilisation de la machinerie traductionnelle (ARNt, ribosomes) et des ressources cellulaires (acides aminés, énergie) pour la synthèse protéique. La synthèse de la protéine recombinante n'apportant

aucune fonction pour la cellule, elle représente alors un fardeau appelée charge métabolique (figure 1.7).

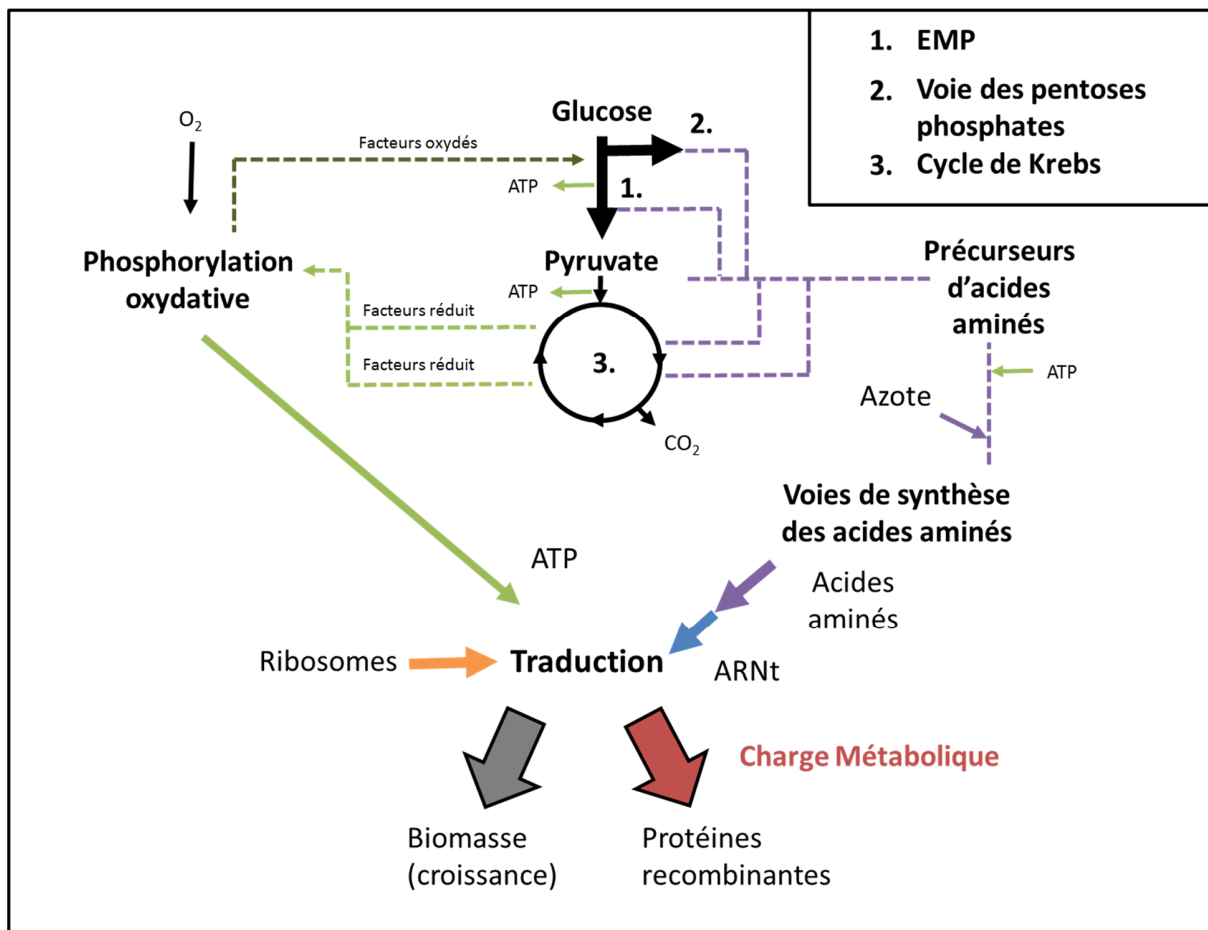


Figure 1.7 : Compétition entre production de protéines recombinantes et croissance bactérienne pour la consommation des ressources intracellulaires. La traduction des protéines cellulaires et des protéines recombinantes nécessite la consommation d'acides aminés (violet), apportés par les ARN de transfert (ARNt) (bleu), d'énergie (vert) et l'utilisation de ribosomes (orange). Les acides aminés sont synthétisés à partir des intermédiaires du métabolisme central (glycolyse d'Embden-Meyerhof-Parnas (EMP), cycle de Krebs, voie des pentoses phosphates) et d'azote. L'énergie chimique (ATP) ou d'oxydo-réduction est générée par la glycolyse, le cycle de Krebs et la chaîne respiratoire lors du catabolisme du glucose. La charge métabolique (rouge) représente la consommation supplémentaire de ces ressources pour la production de la protéine recombinante (Scott et al. 2014; Bosdriesz et al. 2015).

1.2.2.1. *Consommation des acides aminés pour la production de protéines recombinantes*

La synthèse de protéines recombinantes entraîne une consommation importante des acides aminés intracellulaires aboutissant parfois à des limitations nutritionnelles (Cserjan-Puschmann et al. 1999; Bachinger et al. 2001; Boström et al. 2005; Funabashi et al. 2002; Pramanik & Keasling 1997). Une limitation en acides aminés apparaît lorsque leur vitesse de consommation est supérieure à leur vitesse de synthèse ou d'assimilation à partir du milieu de culture (Cserjan-Puschmann et al. 1999; Bachinger et al. 2001).

La vitesse de consommation de chaque acide aminé pour la synthèse de la protéine recombinante dépend de la vitesse de traduction de la protéine et de la proportion que représente chaque acide aminé dans sa séquence primaire. La consommation d'acides aminés nécessaire pour la synthèse de la protéine d'intérêt s'ajoute à celle d'acides aminés pour la synthèses des protéines cellulaires (Cserjan-Puschmann et al. 1999; Bachinger et al. 2001). En milieu minimum, la vitesse de synthèse de chaque acide aminé protéinogénique (constitutif de la biomasse) dépend du taux de croissance de la bactérie et de la proportion que représente chaque acide aminé dans la biomasse (Pramanik & Keasling 1997).

La synthèse des acides aminés est effectuée à partir d'intermédiaire de la glycolyse Embden-Meyerhof-Parnas (EMP) (3-phosphoglycérate, pyruvate), de la voie des pentoses phosphates (erythrose-4-phosphate), et du cycle de Krebs (oxaloacétate, α -cétoglutarate).

Les vitesses de synthèse des acides aminés peuvent évoluer en fonction de l'évolution de la concentration intracellulaire de chacun d'entre eux. L'adaptation des flux de biosynthèse des acides aminés permet parfois d'atteindre un équilibre entre consommation et synthèse des acides aminés en condition de production de protéine recombinante. Cependant, le réarrangement des flux présente des limites. Les limites sont contraintes par la production d'énergie, les flux du métabolisme central, la quantité de protéines impliquées dans les voie de synthèse des acides aminés et par la vitesse de consommation de l'azote (Kaleta et al. 2013).

Les acides aminés sont composés de carbone et d'azote. L'azote minéral est assimilé par la bactérie par l'intermédiaire de la voie de la glutamine synthase (GS) et glutamate synthase (glutamine 2-oxoglutarate aminotransferase (GOGAT)) ou la voie glutamate déshydrogénase (GDH) (Helling 1998). Le glutamate et la glutamine transfèrent l'azote assimilé aux autres acides aminés en se convertissant respectivement en glutamine et 2-oxoglutarate. Une limitation en azote se traduit par une limitation en glutamate et glutamine et entraîne une diminution de la synthèse des autres acides aminés (Reitzer 2003). Des limitations en azote peuvent donc entraîner un défaut de synthèse des acides aminés et donc engendrer une limitation en acides aminés.

1.2.2.2. Consommation d'énergie pour la production de protéines recombinantes

Les voies de biosynthèse de chaque acide aminé sont bien décrites chez *E. coli*. Le coût énergétique engendré par la synthèse des acides aminés est associé à la consommation d'adénosine triphosphate (ATP) des voies de biosynthèse des acides aminés, d'agents réducteurs sous forme de nicotinamide adénine dinucléotide phosphate (NADPH) et du coût énergétique (redox et chimique) pour la synthèse des précurseurs (Kaleta et al. 2013). La synthèse des acides aminés, leur chargement sur les ARNt et la formation de liaisons peptidiques nécessaires à la synthèse de protéines recombinantes entraînent une surconsommation énergétique pouvant perturber significativement le métabolisme (Heyland et al. 2011). Des cartes de flux ont été générées en condition de production de protéines recombinantes induite par une augmentation de la température de la culture en milieu minimum (Wittmann et al. 2007). Dans ces conditions de culture une limitation en énergie redox a été observée à travers la sécrétion d'acétate, de pyruvate et de lactate. La sécrétion d'acétate permet de récupérer de l'ATP et la sécrétion de lactate d'oxyder des facteurs réduits. Au cours de la production, une forte activation des voies anapérotyques a été observée en particulier du shunt glyoxylique. L'activation du shunt glyoxylique permet de générer de l'oxaloacétate, précurseur d'acides aminés, et pourrait être responsable de l'accumulation d'acétate (Wittmann et al. 2007).

1.2.2.3. *Mesure de la charge métabolique et de la capacité de synthèse de la protéine recombinante par l'hôte*

La charge métabolique imposée par la production de protéines recombinantes provient de la consommation des ressources cellulaires par la synthèse supplémentaire de protéines imposée à la bactérie. M. Cserjan et M. Puschmann ont cherché un moyen de mesurer en temps réel la charge métabolique de la souche hôte. Dans leurs travaux de 1999 et 2001, cette équipe s'est attachée à développer une méthode d'optimisation des procédés de production basée sur la détection de molécules rapporteuses de la charge métabolique (Cserjan-Puschmann et al. 1999; Bachinger et al. 2001). La molécule rapporteuse utilisée dans leur papier est la guanosine 5'diphosphate,3'-diphosphate (ppGpp). Cette molécule appartient au mécanisme appelé réponse stringente et permet l'adaptation du taux de croissance à la variabilité nutritionnelle de l'environnement. Le mécanisme d'adaptation associé au ppGpp sera développé dans la deuxième partie de l'introduction. Cette méthode a permis de développer des modèles de compréhension basés sur l'équilibre entre production et capacités métaboliques de l'hôte. Cela a également permis de déterminer des phases de production trop fortes par rapport aux capacités métaboliques de l'hôte et de déterminer rapidement les systèmes d'expression génique les plus adaptés à la spécificité de la protéine recombinante.

La capacité de synthèse de la souche hôte est sa capacité à fournir l'ensemble des ressources nécessaires à la synthèse de la protéine recombinante et donc à supporter la charge métabolique de la production. Une méthode pour mesurer la capacité de synthèse d'*E. coli* a été développée par Ceroni et al. 2015. Cette méthode consiste à intégrer dans le génome le gène de la protéine GFP, exprimé constitutivement, en parallèle d'une protéine RFP produite grâce à un plasmide et utilisée comme une protéine recombinante d'intérêt. Les chercheurs ont constaté que la production de GFP est inversement corrélée à la production de la protéine recombinante. Ainsi, l'intégration du gène de la GFP dans le génome permet de mesurer en temps réel la capacité traductionnelle de la souche pour ses protéines utiles de la biomasse. Les chercheurs ont constaté que plus la production de la protéine recombinante est grande, plus la capacité de traduction des protéines intracellulaires est faible et plus le taux de croissance est faible. Les chercheurs ont ainsi lié la mesure de la GFP à la capacité de synthèse de la souche (Ceroni et al. 2015).

Le lien entre la capacité de traduction des protéines intracellulaires et le taux de croissance est fondamental. Il s'agit d'une stratégie centrale dans la sélection évolutive d'*E. coli* qui ne peut survivre durablement que par sa capacité à croître rapidement dans des environnements riches et à survivre dans des environnements pauvres par modulation de sa composition macro-chimique et en particulier sa composition en protéines (Bosdriesz et al. 2015).

1.2.3. Influence de la production de protéines recombinantes sur le taux de croissance

Le taux de croissance d'*E. coli* augmente lorsque la richesse nutritive de son environnement augmente. La richesse nutritive de l'environnement représente la diversité en sources de carbone de l'environnement. Par exemple un milieu contenant des acides aminés, des vitamines et du glucose sera dit riche et un environnement ne contenant qu'une seule source de carbone sera dit pauvre ou minimum. Depuis 1996, la modification de la composition macro-chimique (proportion de protéine, d'ADN, d'ARN dans la biomasse) d'*E. coli* en fonction de la richesse nutritive de l'environnement et donc du taux de croissance a été observée (Bremer & Dennis 1996) (tableau 2).

Tableau 1.2 : Composition macro-chimique d'*E. coli* en fonction du taux de croissance (d'après Bremer & Dennis 2008).

	Unité	$\mu = 0.6 \text{ h}^{-1}$	$\mu = 1 \text{ h}^{-1}$	$\mu = 1,5 \text{ h}^{-1}$	$\mu = 2 \text{ h}^{-1}$	$\mu = 2,5 \text{ h}^{-1}$	$\mu = 3 \text{ h}^{-1}$
% Proteines/Biomasse	%	60	56	53	50	47	42
% ARN/Biomasse	%	10	12	14	17	20	21
% ADN/Biomasse	%	5	3	2	2	2	2
% Autres/Biomasse	%	27	29	31	31	31	35
Proteines ribosomales/Proteines totales	%	7,7	9,2	11,6	15	18,8	22,7
Nombres de bactéries/Biomasse	10^3 bactéries/OD ₄₆₀	7,7	4,6	3,1	2,2	1,9	1,7
Protéines/bactérie	$\mu\text{g}/10^9$ bactéries	136	214	295	387	431	426
ARN/bactérie	$\mu\text{g}/10^9$ bactéries	23	44	76	128	180	214
ADN/bactérie	$\mu\text{g}/10^9$ bactéries	7,6	9,5	12	14,7	17,2	19,4
gènes ribosomales/bactérie	No./bactérie	12,4	16,5	22	27,6	32,9	37,5
Ribosomes/bactérie	10^3 ribosomes/ bactérie	8	15	26	44	61	73

De façon remarquable, la proportion de ribosomes parmi la quantité totale de protéines est corrélée de façon linéaire au taux de croissance dans chacun des environnements (Scott et al. 2014). De plus, si des antibiotiques capables d’inhiber la vitesse de traduction sont ajoutés, le taux de croissance est corrélé de façon linéaire avec la concentration en antibiotique et donc à la vitesse de traduction (Scott et al. 2010). Ces expériences ont révélé que le taux de croissance d’*E. coli* est exclusivement contraint par la vitesse de transformation des nutriments de l’environnement en protéines intracellulaires (Scott et al. 2010; Scott et al. 2014; Bosdriesz et al. 2015; Shachrai et al. 2010). La vitesse de transformation des nutriments est composée de deux mécanismes biologiques majeurs. Le premier est le métabolisme qui convertit les nutriments de l’environnement en acides aminés. La quantité de protéines impliquées dans le métabolisme est modifiée en fonction de la richesse nutritive de l’environnement (Traxler, Summers, H. Nguyen, et al. 2008) (figure 1.8). Le deuxième mécanisme est la traduction qui synthétise des protéines à partir des acides aminés. Les ribosomes et les protéines impliquées dans le métabolisme représentent environ 50% des protéines totales d’*E. coli* (Scott et al. 2010). La deuxième moitié des protéines est composée de protéines structurales dont la proportion ne varie pas en fonction du taux de croissance (figure 1.8).

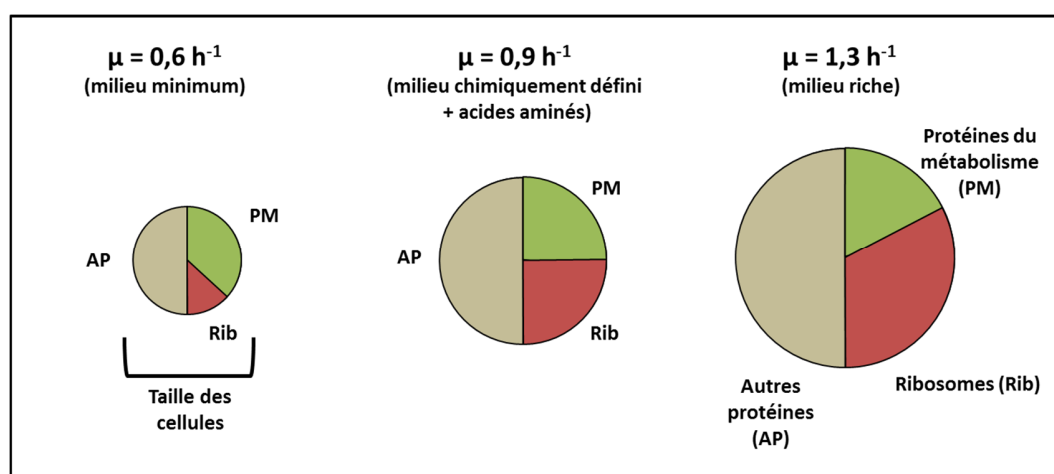


Figure 1.8 : Modification de la taille des bactéries et des proportions des protéines impliquées dans le métabolisme (PM), des ribosomes (Rib) et des autres protéines (AP) parmi les protéines totales en fonction du taux de croissance. Inspiré de Scott et al. 2010.

Dit autrement, le taux de croissance est contraint par le temps nécessaire pour synthétiser toutes les protéines à chaque génération. En milieu minimum, la transformation de la source carbonée du milieu en acides aminés nécessite un grand nombre de protéines impliquées dans le métabolisme. De fait, dans cette condition, la proportion en ribosomes parmi les protéines totales diminue et le temps pour synthétiser l'ensemble des protéines par les ribosomes augmente ainsi que le temps de doublement des bactéries (figure 1.8). En milieu riche, les acides aminés sont présents dans l'environnement, les protéines du métabolisme sont en moindre proportion et la proportion de ribosomes augmente. Le temps nécessaire à la synthèse de l'ensemble des protéines diminue et le temps de doublement est plus faible qu'en milieu minimum (figure 1.8). Ce phénomène, appelé « fitness du μ », est mis en évidence à partir des proportions de chaque catégorie de protéines. Cependant, la modification du taux de croissance en fonction de son environnement implique également une modification de la taille des cellules. Plus le taux de croissance est élevé plus la taille des cellules augmente (tableau 1.2 et figure 1.8). De fait, la quantité de protéines par bactérie (g/bactérie) augmente lorsque le taux de croissance augmente alors que la proportion (g/g) de protéines par unité biomasse diminue dans le même temps (tableau 2). Malgré cette variable, le principe expliqué avec les proportions en protéines est donc toujours observé.

Une autre conséquence de la modification de la taille des cellules en fonction du taux de croissance est que le pourcentage d'ADN par gramme de biomasse diminue lorsque le taux de croissance augmente alors que la quantité d'ADN par bactérie (μg ADN/bactérie) augmente (voir tableau 2).

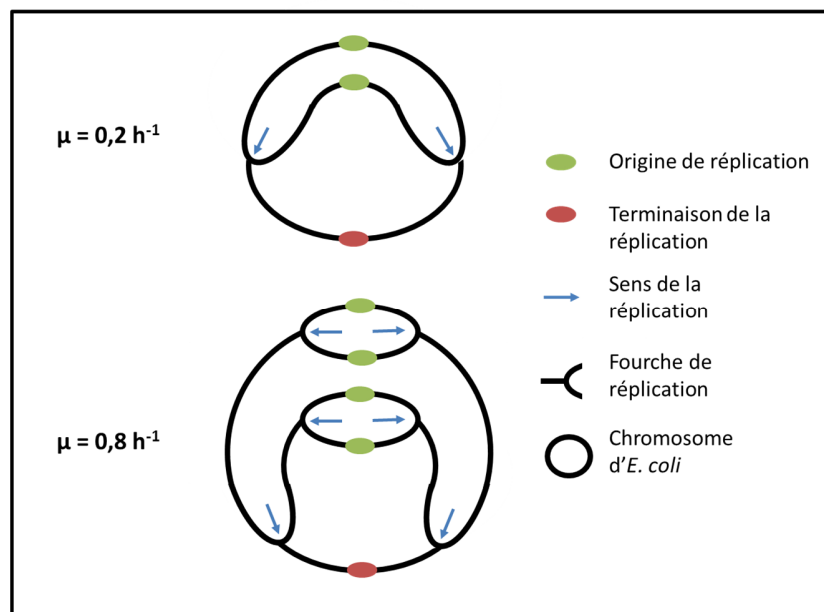


Figure 1.9 : Augmentation du nombre de fourches de réplication lors de l'augmentation du taux de croissance d'*E. coli*. Inspiré de Bremer & Dennis 2008

Lors d'un fort taux de croissance le temps de réplication du chromosome est plus long que le temps de doublement des bactéries. Pour compenser ce phénomène, l'initiation de la réplication se produit 2 à 3 doublements avant la fin de la réplication de l'ADN (Bremer & Dennis 2008) (figure 1.9). Ce phénomène entraîne l'apparition de plusieurs fourches de réplication et donc une plus grande quantité d'ADN par cellule lorsque le taux de croissance augmente. Par ailleurs, étant donné que le chromosome d'*E.coli* ne possède qu'une seule origine de réplication, ce phénomène de réplication engendre une augmentation du nombre de copies de l'ADN plus importante à proximité de l'origine de réplication qu'au niveau du site de terminaison de la réplication. Il est ainsi logique de constater la présence des gènes ribosomaux à proximité de l'origine de réplication afin de coupler l'augmentation du taux de croissance avec l'augmentation de la quantité de ribosomes par cellule.

Puisque le taux de croissance dépend de la vitesse de transformation des nutriments en protéines de la biomasse, chaque protéine synthétisée représente une charge métabolique pour la croissance. Cette charge est rentabilisée par la fonction qu'apporte ou pourrait apporter la protéine à la cellule. Les protéines recombinantes sont des protéines sans aucune fonction pour la cellule. Leur synthèse conduira à l'utilisation d'une part non rentabilisée des ressources bactériennes et de l'activité

des ribosomes entraînant une diminution du taux de croissance. Il est décrit que la part de l'activité des ribosomes allouée à la synthèse de la protéine recombinante ne peut pas dépasser 30 à 50% des protéines totales à synthétiser sans engendrer un arrêt total de la croissance (Bosdriesz et al. 2015; Scott et al. 2010).

L'optimisation du fitness du μ a été fortement sélectionnée au cours de l'évolution car il s'agit d'une stratégie centrale dans la survie d'*E. coli* (Bosdriesz et al. 2015). Cette évolution a abouti à une forte régulation de l'ensemble des protéines impliquées dans ce fitness (protéines du métabolisme et ribosomes) par une unique molécule qu'est le ppGpp. Le rôle de cette molécule chez *E. coli* sera développé dans la deuxième partie de l'introduction. Pour l'instant, il faut savoir que le ppGpp est un nucléotide synthétisé lors de la diminution de la concentration intracellulaire en acides aminés. Son accumulation conduit à l'inhibition de la synthèse des ribosomes et à l'augmentation de la synthèse des protéines impliquées dans le métabolisme. De fait, le ppGpp est impliqué dans la régulation du fitness du μ (figure 1.10). De plus, la production de protéines recombinantes entraîne une consommation des acides aminés cellulaires. Cette consommation engendre parfois l'accumulation de ppGpp dans la cellule (Bachinger et al. 2001). De fait, la production de protéines recombinantes va engendrer une perturbation du fitness du μ d'*E. coli*. Ce lien entre fitness du μ , ppGpp et production de protéine recombinante est utilisé dans les modèles de lois régissant le taux de croissance (Scott et al. 2010) mais est très peu étudié expérimentalement.

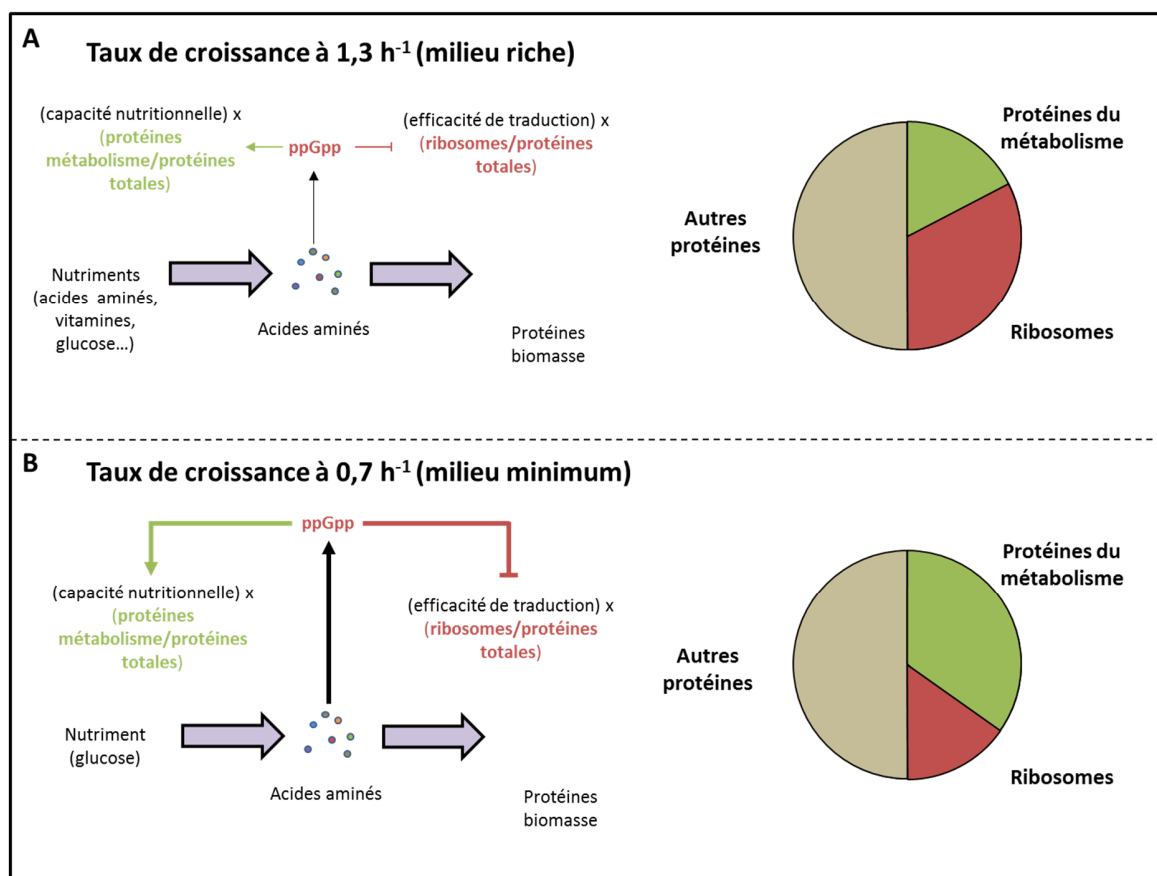


Figure 1.10 : Régulation des proportions en protéines impliquées dans le métabolisme d’*E. coli* ou des ribosomes lors d’une culture en milieu riche (A) ou une culture en milieu minimum (B). Les flux de synthèse et d’assimilation des acides aminés à partir du milieu extérieur dépendent de la capacité nutritionnelle de la souche (topologie du réseau métabolique, efficacité des enzymes...) et de la proportion de protéines impliquées dans le métabolisme parmi la quantité totale de protéines. Le flux de synthèse des protéines dépend de l’efficacité de traduction et de la proportion de ribosomes parmi la quantité totale de protéines. Lors d’une culture de bactéries en milieu minimum, une concentration plus élevée de ppGpp intracellulaire est observée permettant la synthèse de protéines du métabolisme et l’inhibition de la synthèse de ribosomes, modifiant les proportions de ces protéines dans la biomasse. Inspiré de Scott et al. 2010.

La modification macro-chimique d’*E. coli* en fonction de son taux de croissance est prise en charge par une régulation systémique à partir de la réplication jusqu’à la traduction. Comme il a été expliqué précédemment, le changement de taux de croissance engendre un changement profond de la composition bactérienne en protéines, en ARN et ADN. Ainsi, le type de protéines à traduire devient différent, la force d’expression des gènes change radicalement ainsi que la capacité de traduction de la bactérie. Ces régulations entraînent une modification des voies métaboliques exprimées et des flux

métaboliques. Les modifications de la réplication, de la transcription, de la traduction et du métabolisme doivent être mises en place de façon cohérentes et corrélées pour garantir à chaque moment un équilibre du système. Cette adaptation est contrôlée par la concentration en ppGpp intracellulaire (mécanisme appelé réponse stringente) (Magnusson et al. 2005; Potrykus & Cashel 2008; Srivatsan & Wang 2008; Iyer et al. 2018). Par ailleurs, la production de protéines recombinantes entraîne la consommation des acides aminés intracellulaires et utilise une part des ribosomes causant directement une inhibition du taux de croissance et interférent, de fait, avec le fitness du μ et la réponse stringente associée à sa régulation.

1.3. Réponse stringente chez *E. coli*

La réponse stringente a été définie dans un premier temps comme l'inhibition rapide de la synthèse des ARN stables (ARNt et ARN ribosomaux), suite à une carence en acides aminés, entraînant un arrêt de synthèse des protéines ribosomales et un arrêt de la croissance (Cashel et al. 1996). Cette définition a ensuite été étendue à toutes les limitations nutritionnelles (carbone, azote, phosphate, fer, oxygène) et causes environnementales (choc osmotique, dommages aux ultra-violet, acidification...) provoquant un arrêt de croissance (Potrykus & Cashel 2008; Dalebroux & Swanson 2012; Srivatsan & Wang 2008; Azir & Arinarayanan 2016; Liu et al. 2015). Depuis 2008, le rôle de la réponse stringente a été élargi dans la coordination du fitness du μ lors de la modification de la richesse nutritive environnementale (Traxler, Summers, H. Nguyen, et al. 2008; Scott et al. 2010; Scott et al. 2014).

1.3.1. ppGpp et pppGpp : synthèse et dégradation

Le ppGpp et le pppGpp (guanosine-3'-diphosphate,5'-triphosphate) dérivent respectivement d'une estérification du ribose 3' de la guanosine diphosphate (GDP) et de la guanosine triphosphate (GTP) par consommation d'adénosine triphosphate (ATP) transformé en adénosine monophosphate (AMP)(Cashel et al. 1996)(figure 1.11). Chez *E. coli*, seules deux enzymes sont capables de catalyser cette réaction, les protéines RelA et SpoT. Ces enzymes de la famille RSH (RelA/SpoT Homologue) possèdent les mêmes domaines protéiques avec une partie hydrolase, une partie synthase, un site TGS (cible guide de synthèse) de fonction inconnue et un site ACT (appelé ainsi car retrouvé sur Aspartate

kinase, Chorismate mutase et TyrA). La partie hydrolase est inactive chez RelA, laissant SpoT la seule enzyme capable de dégrader respectivement le pppGpp et le ppGpp en GTP et GDP, (Shyp et al. 2012; Loveland et al. 2016) (figure 1.11). La protéine GppA intervient dans la voie de dégradation du pppGpp par son activité pppGpp pyrophosphatase, capable de transformer le pppGpp en ppGpp (DeNapoli et al. 2013; Mechold et al. 2013) (figure 1.11). Les différences dans la capacité d'action du ppGpp comparé au pppGpp ont commencé à être évaluées seulement depuis 2013 (Mechold et al. 2013; DeNapoli et al. 2013). Dans la littérature scientifique, ces deux alarmones sont souvent regroupés en un seul terme, le (p)ppGpp, ou résumés au nucléotide seul ppGpp. Il ressort de ces travaux que le ppGpp a une action plus forte que le pppGpp au niveau de la régulation transcriptionnelle ainsi que dans sa capacité à inhiber la croissance. Cependant, par manque de consensus scientifique clair sur la question, le ppGpp et le pppGpp seront regroupés sous le terme de (p)ppGpp dans la suite de cette introduction.

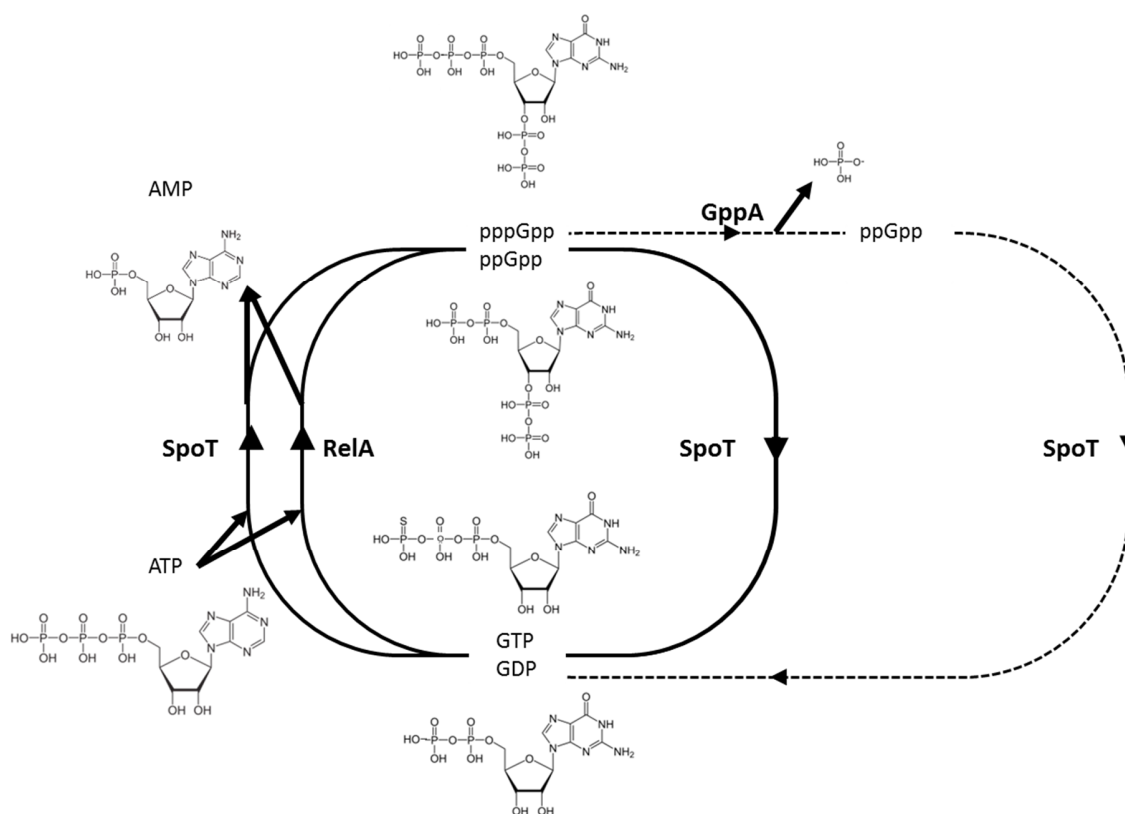


Figure 1.11 : Voies de synthèse et de dégradation du ppGpp et du pppGpp chez *E. coli*, d'après Hauryliuk 2015.

La réponse stringente est un mécanisme impliqué dans le contrôle du taux de croissance. Le taux de croissance étant limité par la capacité bactérienne à synthétiser les protéines il est cohérent de retrouver la protéine RelA en interaction avec les ribosomes. Plus particulièrement, la protéine RelA est localisée dans la zone d'interaction entre les ARNt et les ribosomes et cette interaction rend la protéine inactive. La protéine RelA se sépare des ribosomes lorsque des ARNt non chargés en acide aminé se présentent aux ribosomes activant la synthèse (p)ppGpp par cette protéine. Ce phénomène permet une réponse graduelle en fonction du nombre d'ARNt non chargés, marqueurs de l'équilibre de vitesses d'apport des acides aminés et de consommation. La localisation de la protéine RelA démontre la place centrale des acides aminés dans la détection des carences nutritionnelles chez *E. coli*. Lors d'une carence en azote, la voie des Ntr (régulation nitrogène) est activée. Par ce mécanisme, les protéines Ntr-B/Ntr-C permettent la transcription d'une centaine de gènes impliqués dans l'absorption de l'azote mais aussi la surexpression du gène *relA*. La surexpression de ce gène par la voie Ntr couple ainsi la réponse stringente à la carence en azote. Ce couplage permet donc d'augmenter la sensibilité de la réponse stringente lors d'une carence en azote (Brown et al. 2014; Carneiro et al. 2012).

La réponse stringente permet à la bactérie de détecter des carences de composés nutritifs essentiels à la croissance. La protéine RelA intervient dans la détection des limitations en acides aminés et en azote, essentiels à la synthèse protéique, mais une détection d'un ensemble plus grand de composés a été observée pour la protéine SpoT (Sarubbi et al. 1988; Hogg et al. 2004; Vinella et al. 2005; Battesti & Bouveret 2006; Boehm et al. 2009). En effet, cette protéine serait capable de détecter des limitations en acides gras, phosphate et fer et elle est également sensible à l'oxydation et aux chocs osmotiques. La détection de la limitation en acides gras repose sur l'interaction entre SpoT et le complexe d'acide gras synthase (ACP) impliqué dans la synthèse des acides gras (Battesti & Bouveret 2006). Les mécanismes de détection des autres éléments n'ont pas été révélés jusqu'à maintenant.

La régulation du fitness du μ est une régulation globale du fonctionnement bactérien (transcriptomique, protéomique, métabolomique) qui doit être modifiée de façon cohérente avec toutes les autres. Les cibles du (p)ppGpp sont présentes à toutes ces échelles et déclenchent des cascades de régulation permettant une conservation de l'équilibre du système.

1.3.2. Régulation cycle cellulaire

Les moyens de contrôle du taux de croissance d'*E. coli* par le (p)ppGpp sont en partie observés au niveau des trois phases du cycle cellulaire : i) la phase B correspond à la phase d'initiation de la réplication de l'ADN, directement après la naissance (« Birth ») de la bactérie et l'initiation de la formation de la cellule, ii) la phase C correspond à la synthèse de l'ADN jusqu'à la fin de réplication, iii) la phase D correspond à la ségrégation des deux cellules filles et de l'ADN (Azir & Arinarayanan 2016; Nazir & Harinarayanan 2016).

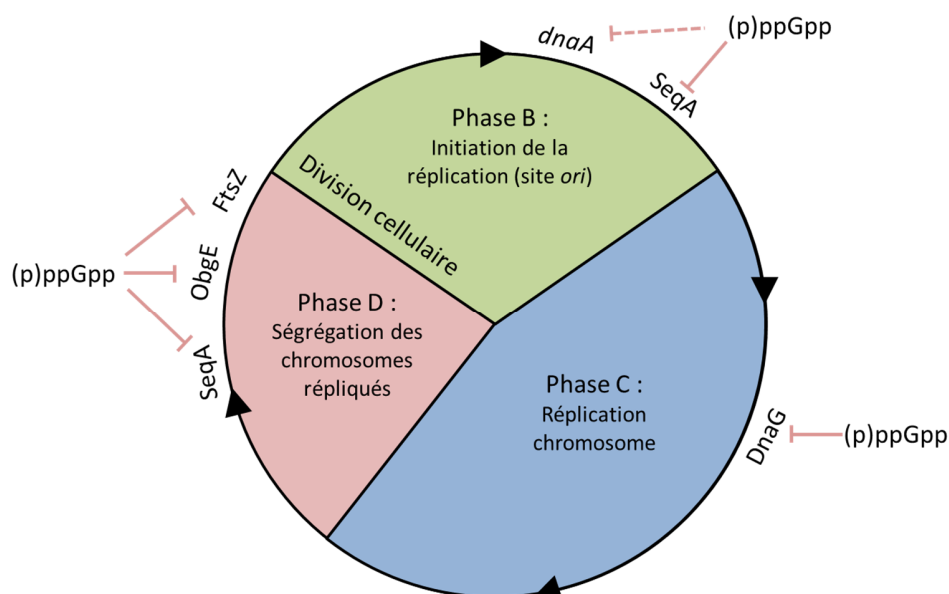


Figure 1.12 : Contrôle de la réplication cellulaire par le (p)ppGpp chez *E. coli*. Les flèches en pointillées représentent l'inhibition de la transcription du gène et les flèches pleines, l'inhibition par interaction directe avec une protéine, d'après Nazir & Harinarayanan 2016

Au niveau de la phase B, le (p)ppGpp inhibe la transcription du gène *dnaA* par interaction avec l'ARN polymérase (figure 1.12). Chez *E. coli*, la protéine *DnaA* régule l'initiation de la réplication lorsqu'elle est complexée à de l'ATP. Une diminution de la quantité de *DnaA* entraîne une diminution de l'initiation de la phase B. De nombreuses publications suggèrent que la modification de la taille des bactéries en fonction du taux de croissance est une conséquence de l'implication du (p)ppGpp dans l'initiation de la formation de la cellule à la phase B (Nazir & Harinarayanan 2016). En effet, l'activation expérimentale de la réponse stringente entraîne une diminution de la taille des cellules de 30%. Plus

spécifiquement, le (p)ppGpp inhiberait la protéine SeqA impliquée dans l'initiation de la réplication (Nazir & Harinarayanan 2016). Au cours de la phase C, le (p)ppGpp possède une forte activité d'inhibition sur les primases, protéines synthétisant les amorces nécessaires à la réplication du brin anti-sens. Il existe une interaction directe avec DnaG, inhibant son activité de primase chez *E. coli* (Maciag et al. 2010; Wang et al. 2007). Enfin, la ségrégation des chromosomes au cours de la phase D est inhibée par le (p)ppGpp par interaction avec la GTPase ObgE. Cette inhibition est également réalisée par interaction avec la protéine FtsZ impliquée dans la division cellulaire. Plusieurs publications mettent en évidence l'implication du (p)ppGpp dans l'inhibition de la ségrégation des cellules filles suite à l'observation de très long bacilles non ségrégués après induction de la réponse stringente.

Ces mécanismes mettent en évidence la capacité de contrôle du (p)ppGpp sur le cycle cellulaire et donc sur le taux de croissance. Cependant, une modification du taux de croissance doit s'accompagner d'une adaptation du métabolisme bactérien de façon proportionnelle et équilibrée. Comme expliqué ci-dessus, le (p)ppGpp agit en ralentissant la division cellulaire mais cette alarmone est également impliquée dans l'adaptation transcriptionnelle de la bactérie lors de l'inhibition du taux de croissance.

1.3.3. Régulation de la transcription

Le (p)ppGpp permet d'adapter la transcription à la modification du taux de croissance en interagissant directement avec l'holoenzyme RNA polymérase (RNAP) d'*E. coli* (Zuo et al. 2013; Dalebroux & Swanson 2012; Sharma & Chatterji 2010; Carneiro et al. 2011; Srivatsan & Wang 2008). La RNAP est l'enzyme permettant la transcription des gènes d'*E. coli*. L'interaction du (p)ppGpp avec RNAP change sa conformation structurale et son affinité avec les motifs spécifiques des promoteurs. La transcription des gènes par RNAP ne peut se faire qu'en présence d'un facteur sigma, cofacteur transcriptionnel protéique. Les principaux facteurs sigma sont au nombre de sept (s70 (RpoD), s54 (RpoN), s38 (RpoS), s32 (RpoH), s28 (RpoF), s24 (RpoE) et s19 (FecI)), et peuvent également modifier structurellement RNAP par interaction directe avec l'holoenzyme (Battesti et al. 2015; Srivatsan & Wang 2008; Battesti et al. 2011; Magnusson et al. 2005). Ces interactions RNAP-facteur sigma permettent la reconnaissance spécifique de promoteurs qui dépend de la nature du facteur sigma.

Succinctement, le principal facteur sigma RpoD est retrouvé en phase exponentielle de croissance et permet la synthèse des gènes impliqués dans cette phase. Le facteur sigma RpoS est impliqué dans la phase stationnaire, dans la résistance à l'acidité de l'environnement, et plus généralement dans la résistance aux environnements hostiles. Il est le deuxième facteur sigma le plus abondant chez *E. coli* après RpoD (Sharma & Chatterji 2010; Battesti et al. 2011). Le facteur sigma RpoN est impliqué dans la réponse à une carence en azote, RpoH dans la réponse à un choc thermique, RpoF dans la mobilité, Fecl dans la carence en fer et RpoE dans la sécrétion de facteurs extracellulaires et dans l'intégrité membranaire. Le (p)ppGpp a un rôle majeur dans la compétition de ces facteurs sigma pour la fixation à RNAP. A un niveau faible en (p)ppGpp, l'interaction de RpoD-RNAP est favorisée. Lors de l'augmentation de la concentration intracellulaire en (p)ppGpp, les autres facteurs sigma interagissent préférentiellement avec RNAP. C'est en partie par ce mécanisme de compétition entre les facteurs sigma pour RNAP que la réponse stringente permet une adaptation transcriptionnelle de la bactérie à un large spectre de carences (Sharma & Chatterji 2010).

L'interaction entre le (p)ppGpp et RNAP est stabilisée grâce à la protéine DksA (Kolmsee et al. 2011; Brown et al. 2002; Paul et al. 2005; Kolmsee et al. 2011; Roghanian et al. 2015; Magnusson et al. 2007; Zhang et al. 2014). Ce cofacteur transcriptionnel, exprimé de façon constitutive, a la propriété de pouvoir se fixer à RNAP en maintenant l'interaction entre le (p)ppGpp et RNAP ; cette action sur RNAP est dite synergique à l'action du (p)ppGpp. En absence de DksA, l'action transcriptionnelle du ppGpp existe mais est très faible.

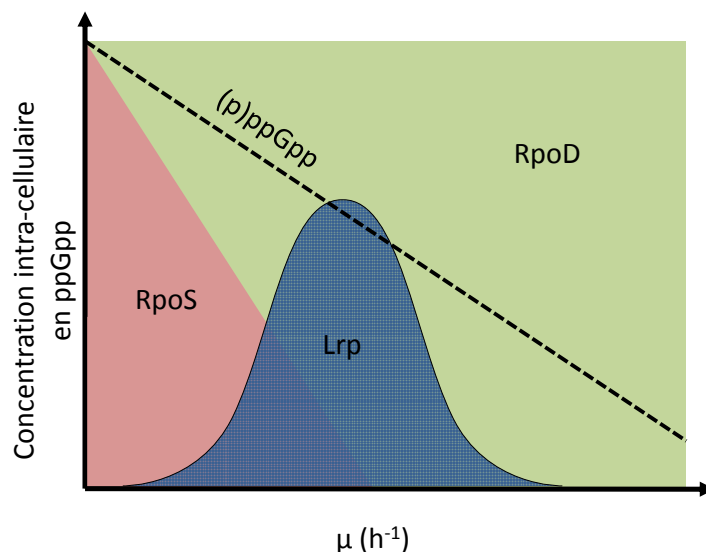


Figure 1.13 : Association des facteurs de transcription avec l'ARN polymérase (RNAP) d'*E. coli* en fonction de la concentration intracellulaire de ppGpp et du taux de croissance, d'après Traxler et al., 2008.

Chez *E. coli*, les travaux de Traxler ont mis en évidence une régulation graduelle du (p)ppGpp sur la régulation transcriptionnelle en fonction du taux de croissance dans trois environnements de croissance : i) un environnement à forte richesse nutritive (diversité de nutriments disponibles en forte abondance), ii) un environnement à faible richesse nutritive (milieu minimum de croissance), iii) un environnement sans source de carbone (Traxler, Summers, H. Nguyen, et al. 2008; Traxler et al. 2011) (figure 1.13). En condition de croissance exponentielle en milieu riche (i), la disponibilité des acides aminés est élevée, diminuant la quantité d'ARNt non chargé et par conséquent, le niveau de (p)ppGpp dans la bactérie par rapport à une croissance en milieu pauvre en nutriment (figure 1.13). Ce faible niveau de (p)ppGpp permet une interaction de RpoD avec RNAP. Dans ce cas (i), les gènes impliqués dans le doublement rapide des cellules sont exprimés, le nombre de ribosomes est augmenté suite à la faible inhibition du (p)ppGpp et les protéines impliquées dans la synthèse des acides aminés sont faiblement exprimées. De fait, le taux de croissance augmente jusqu'au point d'équilibre entre capacité d'apport des acides aminés et consommation des acides aminés, ce point d'équilibre étant détecté par la concentration intracellulaire de (p)ppGpp. Lorsque les bactéries se multiplient en phase exponentielle en milieu minimum (ii), le niveau de (p)ppGpp est légèrement plus élevé dans la cellule par rapport à une croissance en milieu riche (figure 1.13). Cette concentration intracellulaire de

(p)ppGpp entraîne l'activation de la transcription de cofacteurs transcriptionnels tels que la « protéine régulatrice de la réponse-leucine » (Lrp), impliquée dans l'activation des voies de biosynthèse des acides aminés, ou CsrA, responsable de la régulation des protéines impliquées dans la glycolyse. Une inhibition du taux de croissance est observée ainsi qu'une modification du profil de reconnaissance des promoteurs par RNAP, avec le maintien de la fixation de RpoD. Ainsi, les gènes impliqués dans la phase exponentielle sont transcrits et l'interaction RNAP-ppGpp entraîne une diminution de la transcription des gènes ribosomiaux ainsi qu'une augmentation de la transcription de gènes codant des protéines des voies de synthèse des acides aminés. Graduellement, le passage de la phase exponentielle de croissance à une phase stationnaire (iii) se fait par augmentation de la concentration en (p)ppGpp dans la cellule (figure 1.13). A un fort niveau intracellulaire de (p)ppGpp, la transcription du facteur sigma RpoS est activée et la fixation entre RpoD et RNAP est déstabilisée. Le profil de reconnaissance des promoteurs par RNAP change radicalement, inhibant la transcription des gènes ribosomiaux, et activant la transcription des gènes responsables de la résistance à un environnement acide, ceux impliqués dans le recyclage des protéines intracellulaires, la motilité, la formation de biofilm, etc... L'implication de la réponse stringente dans ces mécanismes est actuellement très étudiée car il a été démontré que ce mécanisme intervient dans la persistance des bactéries aux antibiotiques (Harms et al. 2016). En effet, l'activation de la réponse stringente permet la survie de faibles proportions de bactéries placées en état végétatif lors du traitement antibiotique. Ces bactéries sont capables de se multiplier lors de l'arrêt du traitement (Harms et al. 2016).

Le (p)ppGpp permet d'inhiber directement le taux de croissance par action sur le cycle cellulaire mais il permet également aux bactéries de s'adapter à cette diminution de croissance en intervenant à l'échelle transcriptionnelle. Ces modifications transcriptionnelles ont un effet sur la traduction par régulation du nombre de ribosomes, et sur le métabolisme. Cependant une régulation directe du (p)ppGpp à ces différents niveaux de régulation cellulaire est également présente.

1.3.4. Autres régulations par interaction protéiques

Les GTPases, protéines capables de synthétiser ou dégrader le GMP/GDP/GTP chez *E. coli*, possèdent des sites catalytiques de structures très proches (Kanjee et al. 2012; Verstraeten et al. 2015;

Carneiro et al. 2011; Hogg et al. 2004). Ainsi, il a été remarqué par criblage *in silico*, que l'affinité de ces GTPases pour le GDP ou le GTP est également retrouvée pour le (p)ppGpp (Kanjee et al. 2012). Le (p)ppGpp, lors de son accumulation, entrerait alors en compétition avec le GDP ou GTP, ce qui pourrait inhiber l'activité de GTPases. Ce mécanisme a été très bien décrit avec les GTPases IF2 et EF-G, respectivement impliquées dans l'initiation et l'élongation de la traduction, dont l'activité est inhibée *in vitro* et *in vivo* par le (p)ppGpp (Caserta et al. 2006).

1.4. Objectif du travail de thèse

La majorité des productions de protéines recombinantes chez *E. coli* est réalisée avec le système d'expression T7 et la souche *E. coli* BL21 en « batch » et en milieu riche. Ce procédé est reconnu pour optimiser les chances de produire une grande quantité de protéines recombinantes (Rosano & Ceccarelli 2014). Cependant, une forte production conduit le plus souvent à l'arrêt précoce de croissance (Rosano & Ceccarelli 2014). Dans ces conditions, la production de grandes quantités de protéines peut être obtenue au détriment de la qualité. De plus, l'arrêt de croissance ne permet pas de mettre en place des procédés maîtrisés à plus haut rendement comme des procédés semi-continus ou continus. Ces types de procédés offrent pourtant la possibilité d'améliorer les qualités de production par la limitation du taux de croissance *via* l'alimentation contrôlée en source de carbone. Cela permet également de mieux maîtriser la vitesse de traduction de la protéine recombinante. En général, un procédé performant en termes de qualité et de quantité est obtenu en ajustant la vitesse de production des protéines aux capacités de synthèse de l'organisme hôte. Cet équilibre se mesure lorsque la croissance et la production sont stabilisées durant plusieurs heures - voire plusieurs jours. Le développement de ce type de procédé nécessite d'optimiser le *design* du système d'expression pour chaque problème rencontré et pour chaque protéine à produire (Mairhofer et al. 2013; Wang et al. 2014; Dong et al. 1995; Baig et al. 2015). L'analyse bibliographique met en évidence que le nombre de paramètres pouvant être modifiés est très conséquent. Tester chacune des possibilités est long, complexe et coûteux, en plus d'être fastidieux. Cependant cette stratégie empirique est jusqu'à maintenant, celle la plus utilisée, ne permettant pas systématiquement d'obtenir la stabilité du procédé conditionnant le passage d'un procédé « batch » à un procédé semi-continu ou continu. La

très grande majorité des articles présentant les méthodes et les outils de développement de production de protéines recombinantes se limite à cette approche. Les travaux de Puschmann et collaborateurs (Cserjan-Puschmann et al. 1999; Bachinger et al. 2001) ou Ceroni et collaborateurs (Ceroni et al. 2015) font exception dans la mesure où une approche différente, basée sur l'analyse des contraintes métaboliques provoquées par la production, vise à améliorer de manière rationnelle le système d'expression intégré à l'hôte. En effet, identifier et quantifier les contraintes métaboliques provoquées par la production de protéine sont des conditions primordiales pour optimiser de façon rationnelle l'équilibre entre croissance et production. Comme mentionné ci-avant, cette approche est essentielle pour identifier les raisons de l'arrêt de croissance consécutif à l'induction de la production. C'est cette stratégie qui a été mise en place dans le cadre de ce travail de thèse. Elle s'appuie sur une analyse métabolique et fonctionnelle détaillée de la réponse de la bactérie à la production de protéine recombinante afin d'identifier des cibles et de proposer des solutions à l'optimisation du procédé.

Aucune méthode de quantification absolue du ppGpp et du pppGpp n'est pas rapportée dans la littérature. Or, compte tenu du rôle central que ces métabolites peuvent jouer dans le cadre de la production de protéine (Bachinger et al. 2001) et dans le contrôle du taux de croissance (Scott et al. 2014), le développement d'une telle méthode est un prérequis. Cette méthode, présentée Chapitre 3, est basée sur l'utilisation d'une approche originale de dilution isotopique par spectrométrie de masse permettant de corriger la totalité des biais analytiques. Ce volet fait l'objet d'une publication publiée dans la revue *Analytical Chemistry*.

Le procédé de production de protéines recombinantes développé par Sanofi Pasteur repose sur l'utilisation d'un système d'expression T7 et de la souche *E. coli* BL21 en milieu riche et en procédé de type « batch ». Dans un premier temps, les cinétiques de croissance bactérienne et de production de ce procédé « standard » ont été étudiées afin de définir les performances et caractéristiques initiales du procédé (Chapitre 4). Dans le but d'améliorer la réalisation des analyses fonctionnelles du métabolisme (métabolomique et fluxomique), ce procédé a été modifié en un procédé dit « de référence » consistant en une modification des conditions d'induction et en la substitution du milieu riche par un milieu minimum. Le transfert de technologie de l'échelle pilote à l'échelle du laboratoire a été dans un premier temps validé par une analyse des modifications réalisées. Le procédé de référence

est alors utilisé pour étudier les cinétiques de croissance de la souche *E. coli* BL21 et les cinétiques de production de deux protéines modèles. L'analyse macro-cinétique de ces procédés a conduit à définir plusieurs phases caractéristiques du procédé, mises ultérieurement en lien avec les analyses fonctionnelles du métabolisme (voir ci-dessous).

Dans la deuxième partie de cette étude, une analyse non ciblée de l'exométabolome par RMN a été réalisée pour identifier et quantifier les métabolites extracellulaires (Chapitre 5). Cette approche permet de mieux comprendre le fonctionnement du métabolisme en condition de production. Les résultats obtenus, au regard des phases des cinétiques de croissance et de production, permettent de guider la mise en place d'analyses plus ciblées des métabolites intracellulaires. Ainsi, l'étude s'est ensuite focalisée sur le métabolisme des acides aminés. La quantification des acides aminés au cours des procédés de production des protéines modèles, combinée à celle du ppGpp et du pppGpp, a permis de mettre en évidence l'activation de la réponse stringente en réponse à une limitation intracellulaire en acides aminés.

Dans un premier temps, l'effet de la réponse stringente sur la croissance et le métabolisme de la souche *E. coli* a été caractérisé en condition de non production de protéine (Chapitre 6). La réponse stringente est alors déclenchée par l'ajout d'un analogue de la sérine, la serine hydroxamate (SHX) à des cultures d'*E. coli* BL21 ou K-12 sauvages (WT) ou à des mutants dont des gènes clefs dans la synthèse du ppGpp et du pppGpp ont été supprimés du génome. Cette étude est une étape préliminaire à l'étude de l'influence de la réponse stringente au cours du procédé de production des protéines recombinantes. Ainsi, dans un second temps, cette étude a été élargie à l'effet de la réponse stringente au cours de la phase de production, en comparant les profils de croissance et de production de la souche BL21 WT à ceux de ces mêmes mutants, mais transformés du plasmide contenant le gène des protéines d'intérêt.

L'étude de l'effet de la réponse stringente au cours de la phase de production, a permis d'identifier une piste d'optimisation originale de la protéine par modification de la souche *E. coli* BL21 et/ou du système d'expression T7 (Chapitre 7). L'objectif est ici de développer un système biologique de production permettant de garantir le maintien de la croissance en condition de production, et cela de façon générique à toutes les protéines recombinantes entraînant une forte charge métabolique pour *E.*

coli. La preuve de concept du fonctionnement de ce nouveau système d'expression a été démontrée lors de la production des protéines modèles. Conjointement, l'étude s'est portée sur la formulation du milieu de culture afin de répondre à l'objectif mentionné ci-dessus.

Chapitre 2 – Matériels et Méthodes

2.1. Milieux et conditions de cultures

2.1.1. Milieux de culture solides :

Les cultures en milieu solides sont réalisées en milieu riche de type Luria Bertani agar (LBA) et sont utilisées pour l'isolement de colonies.

- LBA : Tryptone : Tryptone 10 g/L ; Extrait de levure 5 g/L ; NaCl 10 g/L ; Agar 15 g/L.

Si nécessaire de la kanamycine 50 µg/mL est ajoutée au milieu.

2.1.2. Milieux de cultures liquides :

L'ensemble des précultures est réalisé en milieu riche (LB2Xm) ou minimum (PréPAT) pour l'inoculation des bactéries en bioréacteur en milieu riche (GlusKYE4) ou en milieu minimum (PAT) respectivement. Le lot de semence de bactéries est réalisé en milieu riche Luria Bertani (LB) contenant 25 % (v/v) de glycérol final et stocké à -80 °C.

-Milieux riches complexes :

- LB2Xm : extrait de levure 30 g/L ; NaCl 10 g/L ; Tris Base 2,43 g/L ; MgCl₂, 6 H₂O 0,406 g/L.
- LB : Tryptone 10 g/L ; Extrait de levure 5 g/L ; NaCl 10 g/L
- GlusKYE4 : extrait de levure 40 g/L ; NaCl 1 g/L ; K₂HPO₄ 0,84 g/L ; MgCl₂, 6 H₂O 0,6 g/L ; K₂SO₄ 0,5 g/L, glucose 40 g/L.

- Milieux minimums :

- PAT : KCl 0,769 g/L ; (NH₄)₂HPO₄ 4 g/L ; MgSO₄, 7H₂O 1,2 g/L ; EDTA 84 mg/L ; CoCl₂, 6H₂O 2,5 mg/L ; MnCl₂, 4H₂O 15 mg/L ; CuCl₂, 2H₂O 1,5 mg/L ; H₃BO₃ 3 mg/L ; Na₂MoO₄, 2 H₂O 2,5 mg/L ; ZnSO₄, 7 H₂O 16,4 mg/L ; FeSO₄ 21 mg/L, glucose 30 à 50 g/L.

- PréPAT : KH_2PO_4 13 g/L ; $(\text{NH}_4)_2\text{HPO}_4$ 4 g/L ; $\text{MgSO}_4, 7 \text{ H}_2\text{O}$ 1,2 g/L ; EDTA 84 mg/L ; $\text{CoCl}_2, 6 \text{ H}_2\text{O}$ 2,5 mg/L ; $\text{MnCl}_2, 4 \text{ H}_2\text{O}$ 15 mg/L ; $\text{CuCl}_2, 2 \text{ H}_2\text{O}$ 1,5 mg/L ; H_3BO_3 3 mg/L ; $\text{Na}_2\text{MoO}_4, 2 \text{ H}_2\text{O}$ 2,5 mg/L ; $\text{ZnSO}_4, 7 \text{ H}_2\text{O}$ 16,4 mg/L ; FeSO_4 21 mg/L, glucose 10 g/L.

De la thiamine est ajouté à une concentration finale de 0,1 g/L pour la culture des souches *E. coli* BW25113. Le glucose, le MgSO_4 et/ou la thiamine sont stérilisée par filtration (Minisart polyamide 0,2 μm , Sartorius, Göttingen, Germany) et ajouté dans le milieu, stérilisé également par filtration, avant inoculation.

2.2. Souches

Les souches utilisées pour cette étude sont la souche *E. coli* BL21 (F-, ompT dcm hsdS (rB- mB-) gal (DE3)) (Novagen). Les mutants de délétion dans ce fond génétique ($\Delta relA$, $\Delta gppA$, $\Delta gppA : \Delta relA$) ont été construits par un prestataire extérieur. La souche *E. coli* BW25141 (*rrnB3 DElacZ4787 DEphoBR580 hsdR514 DE(ara- BAD)567 DE(rhaBAD)568 galU95 DEendA9::FRT DEuidA3::pir(wt) recA1 rph-1*) provient de la banque Keio (Baba et al. 2006), ainsi que les mutants de délétion dans ce fond génétique ($\Delta relA$ et $\Delta gppA$). La souche Top10 F- (*mcrA Δ (mrr-hsdRMS-mcrBC) Φ 80lacZ Δ M15 Δ lacX74 recA1 araD139 Δ (araleu)7697 galU galK rpsL (StrR) endA1 nupG*) est utilisée pour le clonage des plasmides.

2.3. Plasmides

Le vecteur d'expression est un plasmide dérivé d'un plasmide de type pET propriété de Sanofi Pasteur (plasmide pSP) (Peubez et al. 2010). Dans le site de clonage du plasmide est inséré le gène codant pour la protéine UL130 ou pour la protéine Hla.

ORF de la protéine UL130 : 6xHisTag, protéine : 18,4 kDa.

```
ATGGCGGAGCTGAATTACATTCCCAACCGCGTGGCACAACAACCTGGCGGGCAAACAGTCGTTGCTGATTGGCGT
TGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGG
TGCCAGCGTGGTGGTGTGATGGTAGAACGAAGCGGCGTGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTC
```

GCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTG
CACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAGACG
GTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGT
TCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAA
CGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCAC
TGCATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTG
GTGCGGATATCTCGGTAGTGGGATACGACGATACCGAAGACAGCTCATGTTATATCCCGCCGTTAACCACCATCA
AACAGGATTTTCGCTGCTGGGGCAAACAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAG
GGCAATCAGCTGTTGCCGTCTCACTGGTAAAAGAAAAACCACCTGGCGCCAATACGCAAACCGCCTCTCCC
CGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGA

ORF de la protéine Hla, protéine : 33,3 kDa.

ATGGGTCATCACCACCATCATCACGGCAGCGGCGCAGATAGCGACATTAACATCAAGACCGGTACCACCGACAT
TGGCAGCAATACGACGGTCAAGACGGGCGATCTGGTTACTTACGACAAAGAGAACGGTATGTTAAAAAAGTTT
TCTACAGCTTTATCGATGACAAAACTTAAACAAGAACTGTTGGTGATTTCGCACGAAGGGTACCATTGCAGGCC
AATACCGTGTGATTCCGAAGAGGGCGCGAATAAGTCTGGTCTGGCGTGGCCGAGCGCCTTCAAGGTCCAACCTG
CAGCTGCCGGATAATGAAGTGGCGCAAATCTCTGATTACTATCCGCGTAATAGCATCGATACTAAAGAGTACATG
AGCACGCTGACGTACGGTTTTAACGGCAACGTGACCGGCGACGATACCGGTAAGATTGGCGGTCTGATTGGTGC
GAATGTCAGCATCGGTCACACCCTGAAATATGTGCAGCCGGACTTCAAGACCATTCTGGAAAGCCCTACCGACAA
AAAAGTTGGCTGGAAGGTTATCTTCAACAACATGGTTAATCAAACTGGGGTCCGTACGATCGTGATTCTGGAA
TCCAGTCTATGGTAACCAGCTGTTTCATGAAAACCCGTAATGGTAGCATGAAGGCAGCCGATAACTTTCTGGACCC
GAACAAGGCGAGCTCCCTGTTGAGCAGCGGTTTTAGCCCGACTTTGCTACGGTTACTACTATGGATCGTAAAGC
TTCTAAGCAGCAAACCAATATCGATGTCATCTATGAGCGCGTGCGGACGACTACCAGCTGCACTGGACCAGCAC
CAATTGGAAGGTACCAACACCAAAGACAAGTGGACGGACCGTAGCTCGGAACGCTATAAGATCGACTGGGAG
AAAGAAGAGATGACGAATTAA

2.4. Séquences des promoteurs du système MetaProd

2.4.1. Souches MetaProd

Les séquences des promoteurs des opérons ribosomiaux *rrnB*, *rrnD* et *rrnE* ou *FIS-rrnB*, *FIS-rrnD* et *FIS-rrnE* ont été obtenus à partir de la publication de Hirvonen et al. 2001 (Hirvonen et al. 2001). Le promoteur *lacUV5* a été remplacé par les promoteurs des opérons ribosomiaux par une société prestataire.

Séquences de recombinaison homologue :

5' : CATTAAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGA

3' : TCCGATTTACTAACTGGAAGAGGCACTAAATGAACACGATTAACATCGC

Promoteur *lacUV5* :

GCGCAACGCAATTAATGTAAGTTAGCTCACTCATTAGGCACCCCAGGCTTACACTTTATGCTTCCGGCTCGTATA
ATGTGTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGCTATGACC

Promoteur *rrnB_FIS* :

GATTTGGTTGAATGTTGCGCGGTCAGAAAATTATTTTAAATTTCTCTTGTGTCAGGCCGGAATAACTCCCTATAATG
CGCCACCACTGACAGGAATTGTGAGCGGATAACAATT

Promoteur *rrnB* :

GGTCAGAAAATTATTTTAAATTTCTCTTGTGTCAGGCCGGAATAACTCCCTATAATGCGCCACCACTGACAGGAATT
GTGAGCGGATAACAATT

Promoteur *rrnD_FIS* :

AAAAAGGGTGAAAAACAACAACAGAAAAAAGATCAAAAAATACTTGTGCAAAAAATTGGGATCCCTATAA
TGCGCCTCCGTTGAGAGGAATTGTGAGCGGATAACAATT

Promoteur *rrnD* :

AAACAGAAAAAAGATCAAAAAATACTTGTGCAAAAAATTGGGATCCCTATAATGCGCCTCCGTTGAGAGGAA
TTGTGAGCGGATAACAATT

Promoteur *rrnE_FIS* :

TTTTGCTGGGGAATGCAGCAGTCAGTATTTTTCTGCAATTTTTCTATTGCGGCCTGCGGAGAACTCCCTATAATG
CGCCTCCGTTGAGAGGAATTGTGAGCGGATAACAATT

Promoteur *rrnE* :

AGTCAGTATTTTTCTGCAATTTTTCTATTGCGGCCTGCGGAGAACTCCCTATAATGCGCCTCCATCGACAGGAAT
TAGCGGATAACAATT

2.4.2. Plasmides MetaProd

Le promoteur du gène *lacI* du plasmide pSP a été éliminé par les enzymes de restriction Apal (1µl) (New England Biolabs, Ipswich, MA, Etat-Unis) avec une incubation à 25 °C pendant 15 min puis ajout de l'enzyme XbaI (1µl) (New England Biolabs, Ipswich, MA, Etat-Unis) avec incubation 37 °C pendant 15 min dans le tampon fournit avec l'enzyme. Les promoteurs présents dans des plasmides pUC sont digérés avec les mêmes enzymes. Les produits de digestion enzymatique du vecteur et des inserts sont placés sur gel d'agarose 1 %, migrés à 135 V pendant 25min, puis les bandes d'intérêt sont découpées puis purifiées par un kit de purification d'ADN sur gel d'agarose (Sigma Aldrich, St. Quentin-Fallavier, France). Le plasmide linéarisé et les inserts digérés sont ligués avec l'enzyme T4 DNA ligase dans le tampon fournit avec l'enzyme pendant 10 min à température ambiante puis inactivé à 65 °C pendant 10 min. Les produits de ligation sont transformés dans des souches Top10. Les plasmides sont extraits suivant le protocole fourni par le kit d'extraction de plasmides (Sigma Aldrich, St. Quentin-Fallavier, France).

Promoteur *lacI* (plasmide pSP) :

TCGAGATCCCGACACCATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCC

Promoteur du gène *iraP* (PMP1) :

GCTGGTAATCAAACAAAAAATATTTGCGCAAAGTATTCCTTTGTCATAAAAAATAACTTCC

Promoteur du gène *ilvL* (PMP2) :

ATAAAAATCGGCCAAAAAATATCTTGTACTATTTACAAAACCTATGGTAACTCTTTAGGCATT

Promoteur du gène *leuL* (PMP3) :

GGGACGTTTTTATTGCGTCAGGGTTGACATCCGTTTTTGTATCCAGTAACTCTAAAAGCATA

Promoteur du gène *talA-ktb* (PMP4) :

CTATAACTGATGTTACCTGCTTAATCCAGCAATACCATGCCTGTCTGCTATGCTTTTTTGA

Promoteur du gène *thrA* (PMP5) :

GCCGTGAGTAAATTAATAATTTTATTGACTTAGGTCACTAAATACTTTAACCAATATAGGCATA

Promoteur P3 du gène *csrA* (PMP6) :

AACGCTATCGACAACGATAAAGTCAGGTTGAAGTTGTGTATATCGGCTAACTTAGGTTTAAC

2.5. Conditions de culture

Préculture : Les précultures sontensemencées à 0,1 % v/v en milieu riche ou 1 % v/v en milieu minimum à partir du congelât dans une fiole bafflé de 250 mL contenant 1/5^{ème} de milieu, agité à 210 rpm pendant 15 h à 37 °C en milieu riche et 30 °C en milieu minimum.

Culture en bioréacteur : Avant l'inoculation à 0,5 unité de densité optique (uDO) les cellules sont centrifugées 10 minutes à 10 000g à température ambiante (centrifugeuse Sigma 3-18 K, Sigma Aldrich, Seelze, Germany). Les cultures batch en bioréacteurs ont été réalisées en système Ambr250® 12x250 mL (200 mL de milieu) (Sartorius, Göttingen, Allemagne) ou en bioréacteurs Multifors® 2x500 mL (300 mL de milieu) (Infors HT, Bottmingen, Suisse) régulés (pH 7, pO₂ 30 % de la saturation, température 37 °C). Le pH est régulé à l'ammoniac 14 % g/g (VWR, Fontenay-sous-Bois,

France) et à l'acide phosphorique 11 % g/g (Sigma Aldrich, St. Quentin-Fallavier, France). L'aération, l'enrichissement en oxygène de l'air d'entrée du bioréacteur et l'agitation sont contrôlés de façon à maintenir une pO₂ égale à 30 % de la saturation. L'analyse des gaz d'entrées et de sorties des bioréacteurs est réalisée avec un spectromètre de masse Dycor Proline (Ametek, Berwyn, PA, US).

Culture induite : La production de protéine recombinante est déclenchée par ajout d'IPTG (1 ou 0,1 mM en concentration finale) quand la biomasse atteint une valeur comprise entre 3,5 et 5 uDO (induction précoce) ou entre 25 et 35 uDO (induction tardive).

Perturbation à la serine hydroxamate (SHX) : La réponse stringente est activée par ajout de SHX à une concentration finale de 0,8 mM, 0,4 mM ou 0,2 mM quand la culture atteint 3,5 uDO.

Supplémentation en acides aminés : Au moment de l'induction, 1 heure après l'induction et toutes les 30 minutes pendant 1 h, une concentration finale en acides aminés de 2mM est ajoutée dans le bioréacteur.

- Acides aminés de la condition appelée « AA PCN » : histidine, tyrosine, glycine, méthionine, de lysine, arginine, thréonine, valine, isoleucine et leucine
- Les acides aminés de la condition « AA TOT » correspondent aux 20 acides aminés naturels d'*E. coli*.

Ajout du glucose ¹³C pour marquage isotopique au moment de la reprise de croissance : Une culture d'*E. coli* BL21 en milieu minimum avec induction précoce est réalisée comme décrit ci-dessus avec une concentration en glucose non marqué ¹²C de 20 g/L. Au moment de la reprise de croissance visible avec la mesure de la densité optique (DO), 10 g/L de glucose uniformément marqué U-¹³C est ajouté dans le milieu de culture.

2.6. Préparation de cellules BL21 compétentes

Une culture de la souche *E. coli* BL21 est réalisée dans 50 mL de milieu riche (LB2Xm ou LB) en erlenmeyer jusqu'à une densité optique comprise entre 0,3 uDO et 0,5 uDO. Le milieu est placé dans la glace et centrifugé (5 min, 4000 rpm, 4 °C). Le surnageant est éliminé et le culot est repris dans 40 mL d'une solution froide de CaCl₂ 0,1 M. Après une heure d'incubation dans la glace, un second lavage

au CaCl₂ est effectué. Le culot est resuspendu dans 2 mL de solution de CaCl₂ 0,1 M puis placé dans de la glace entre 16 h à 24 h.

2.7. Transformation

Pour la transformation, 50 ng d'ADN plasmidique est ajouté à 50 µL de culture compétente. La solution est incubée 30 min dans la glace puis est soumise à un choc thermique: 30 secondes à 42 °C, suivi de 2 min dans la glace. Un ajout de milieu LB (950 µL) incubé préalablement à 37 °C est ajouté à la solution. Les cellules sont incubées une heure (37 °C, 250 rpm) puis étalées sur gélose LBA/kanamycine (50 µg/mL). Parmi les clones obtenus, 3 sont sélectionnés aléatoirement et évalués en termes de croissance et de production de protéines recombinantes dans le but de vérifier la variabilité de production entre les clones. Le lot de semence de bactéries est réalisé en milieu LB contenant 25 % (v/v) de glycérol final et stocké à -80 °C.

2.8. Techniques analytiques

2.8.1. Détermination de la concentration en biomasse

2.8.1.1. *Mesure de la turbidimétrie*

La biomasse cellulaire est estimée par mesure de l'absorbance à 600 nm à l'aide d'un spectrophotomètre Genesys 6 (Thermo, Carlsbad, CA, USA). Des dilutions sont éventuellement effectuées pour respecter la gamme de linéarité de l'appareil (0,1 à 0,6 uDO) dans une solution de NaCl 0,9 % (g/g).

2.8.1.2. *Mesure Poids secs*

Les poids secs (« cell dry weight », CDW) sont réalisés par prélèvement de 2mL de culture. Les échantillons sont centrifugés à 13 000g pendant 5 min, repris dans du NaCl 0,9 % et avant d'être séchés à 75 °C à 100mBa pendant 24 h. La masse est déterminée avec une balance de précision (Sartorius, Göttingen, Allemagne) par différentiel entre le tube vide et le tube avec le culot séché.

2.8.1.3. Dénombrement des bactéries

Le dénombrement des bactéries et l'analyse de leur viabilité sont réalisés avec un cytomètre en flux Guava® easyCyte 5 HT (Merck, Molsheim, France) suivant les instructions et avec les réactifs fournis par le constructeur. Le fluorochrome Syto9 (Merck, Molsheim, France) et le fluorochrome iodure de propidium (Merck, Molsheim, France) sont utilisés pour colorer respectivement les bactéries viables et non viables. Les prélèvements sont dilués quatre fois au dixième dans du PBS 1X (Life Technology, Carlsbad, Etats-Unis) et l'ensemble des dilutions est analysée. Le traitement des données est réalisé à l'aide du logiciel Guava Soft 3.1.1 (Merck, Molsheim, France).

2.8.2. Stabilité plasmidique et quantification du nombre de copies de plasmide par bactérie

Des dilutions successives d'échantillons de culture sont effectuées pour ensemercer des boîtes de Pétri contenant du LBA et les boîtesensemencées sont incubées à 37 °C O/N. Cent colonies sont repiquées sur gélose LBA/kanamycine. Seules les bactéries porteuses du plasmide sont capables de se diviser en présence de kanamycine, le nombre de colonies présentes après 24 heures d'incubation à 37 °C permet donc d'estimer le pourcentage de bactéries ayant gardées le plasmide au cours des cultures.

La quantification du nombre de copies de plasmide par bactérie est effectuée par « digital droplet PCR » (ddPCR). Le nombre de bactéries est déterminé à l'aide de la quantification du nombre de gènes *dxs* (une copie par chromosome) et le nombre de copies de plasmide par quantification du nombre de gènes *kan* (une copie par plasmide). Pour cela, l'ADN génomique et plasmidique d'un culot bactérien sont extraits à l'aide de la solution InstaGene™ (Biorad, Marnes-La-Coquette, France) puis dilués successivement jusqu'à 10^{-6} dans la solution InstaGene™. Les trois dernières dilutions sont indépendamment analysées pour chaque échantillon. Les dilutions sont ajoutées à un mélange réactionnel composé des amorces de chaque gène à quantifier, des sondes fluorescentes spécifiques aux gènes à quantifier, du tampon ddPCR supermix (Biorad, Marnes-La-Coquette, France) et de l'enzyme de restriction PstI (New England Biolabs, Ipswich, MA, Etat-Unis) puis le mélange est placé dans les conditions adéquates pour que l'enzyme PstI soit active (15 min, 56 °C). Après inactivation de

l'enzyme PstI (8 min, 100 °C), le mélange réactionnel est transformé en gouttelettes grâce à un générateur de gouttelettes (Biorad, Marnes-La-Coquette, France) et aux réactifs associés à l'appareil. Une amplification PCR est réalisée sur les gouttelettes puis celles-ci sont dénombrées par cytométrie (QX100, Biorad, Marnes-La-Coquette, France) et la fluorescence de chaque gouttelette est mesurée. Le logiciel QuantaSoft™ (Biorad, Marnes-La-Coquette, France) est utilisé pour déterminer le rapport entre le nombre de gouttelettes fluorescentes aux longueurs d'ondes correspondantes aux sondes fluorescentes des gènes *dxs* et *kan* afin de déterminer le nombre de copies de plasmide par chromosome de la bactérie.

2.8.3. Analyse de la production de protéines recombinantes

Une quantité de biomasse équivalente à 5 uDO/mL est prélevée et centrifugée (10 min, 14 500 rpm, température ambiante) au cours des cultures. Les culots sont conservés à -20 °C. Les culots décongelés sont lysés par reprise dans un mélange TRIS-EDTA/ Lysosyme (40 U/μL) (Sigma Aldrich, St. Quentin-Fallavier, France)/ Benzonase (250 U/μL) (Sigma Aldrich, St. Quentin-Fallavier, France) puis incubés pendant 20 min à 37 °C, sous agitation rotative. Si nécessaire, les fractions solubles et insolubles sont séparées par centrifugation (10 min, 14 500 rpm, 4 °C) et la fraction insoluble reprise dans du tampon phosphate salin (PBS). Les échantillons sont mélangés au tampon de migration (Biorad, Marnes-La-Coquette, France) et à du dithiothréitol (DTT) (0,083 g/L) (Sigma Aldrich, St. Quentin-Fallavier, France, Marnes-La-Coquette, France) puis incubées 5 min à 95 °C. Les échantillons et un marqueur de taille (All Blue, BioRad, Marnes-La-Coquette, France) sont déposés sur gel de polyacrylamide 12 % (Gel Criterion, BioRad, Marnes-La-Coquette, France). Après migration dans du tampon MOPS (Biorad, Marnes-La-Coquette, France) à 200 V, les gels sont colorés 1 h au bleu de Coomassie (Instant Blue™, Expedeon, Sigma Aldrich, St. Quentin-Fallavier, France), lavés à l'eau par bains successifs. La densitométrie des bandes est obtenus par scan des gels à l'aide d'un densitomètre (GS-800, BioRad, St. Quentin-Fallavier, France) ou par l'analyse des photos de gels prises par Gel Doc™ EZ System (Biorad, St. Quentin-Fallavier, France) et traitées par le logiciel ImageLab™ (Biorad, St. Quentin-Fallavier, France).

La surface des bandes de la protéine recombinante, des gels scannés par le densitomètre GS-800, est délimitée manuellement sur l'interface du logiciel ImageLab, permettant de calculer l'aire du rectangle et l'intensité optique de la bande (mm^2/DO). La densitométrie de la bande à 20 kDa du marqueur de taille All Blue (BioRad, Marnes-La-Coquette, France) est utilisée comme référence pour normaliser les densitométries des bandes de protéines recombinantes. Les bandes formées par les protéines de la biomasse sont utilisées pour normaliser la quantité de biomasse déposée par puits.

2.9. Purification de la protéine UL130

Une lyse chimique est réalisée sur les culots de biomasse enrichie en protéine recombinante de façon identique à la méthode employée pour l'analyse de la production. Le lavage du culot et la purification de la protéine UL130 sont ensuite réalisés suivant le protocole indiqué dans Palmer & Wingfield 2004 (p3, « Prepare washed pellets » et p4, « Extract recombinant protein from washed pellets with guanidine-HCl »).

2.10. Préparation des échantillons pour la RMN

Dans des tubes RMN de 3 mm de diamètre, 50 μL de TMSP-d4 (triméthylsilylpropionate de sodium tétradeutééré) à 4 mM dans de l'eau deutérées sont ajoutés à 150 μL de surnageant de culture pour servir de référence au déplacement chimique (en partie par million (ppm)) et d'étalon interne aux quantifications. Les analyses RMN sont réalisées sur spectromètre un Bruker Avance III 800 Mhz NMR équipé d'une cryo sonde 5mm QCI-P.

2.11. Préparation des échantillons pour analyses en spectrométrie de masse

Des échantillons de culture (400 μL) sont prélevés du bioréacteur et mélangés vigoureusement avec 4,5mL d'une solution d'acétonitrile/méthanol/ H_2O (4:4:2) à -40°C pour figer l'activité métabolique (Millard et al. 2014). Pour une quantification absolue, 100 μL de métabolites marqué au

^{13}C et si nécessaire, 100 μL de métabolites ^{15}N , sont ajoutés comme standards internes. Les tubes sont placés dans un bain d'éthanol à $-40\text{ }^\circ\text{C}$ puis une évaporation à sec est réalisée dans un SpeedVac SpeedVac (SC110A SpeedVac Plus, ThermoSavant, Waltham, MA, USA) sous vide pendant 4 h et conservés à sec à $-20\text{ }^\circ\text{C}$ jusqu'à l'analyse.

2.12. Préparation des métabolites marqués au ^{13}C et à ^{15}N

Les métabolites entièrement marqués au ^{13}C sont obtenus à partir d'une culture d'*E. coli* BW25113 $\Delta gppA$. Le glucose non marqué (^{12}C) est remplacé par du glucose uniformément marqué (U- ^{13}C ; Euriso-Top, Saint Aubin, France). Comme explicité au-dessus, du SHX à 0,8 mM est ajouté au milieu de culture à 3,5 μDO et l'échantillonnage est effectué 15 min après dans une solution d'acétonitrile/méthanol/ H_2O (4:4:2) à $-40\text{ }^\circ\text{C}$. Les tubes sont placés dans un bain d'éthanol à $-40\text{ }^\circ\text{C}$ puis évaporés à sec dans un SpeedVac (SC110A SpeedVac Plus, ThermoSavant, Waltham, MA, USA) sous vide pendant 4 h. Les métabolites marqués au ^{13}C sont repris dans de l'eau déminéralisée, rassemblés dans un même flacon, aliquotés puis conservés à $-80\text{ }^\circ\text{C}$ jusqu'à utilisation. Le même protocole est réalisé pour les métabolites marqués en ^{15}N avec utilisation de la souche BL21 (DE3) au lieu de la souche BW25113 et le $^{14}\text{NH}_4\text{Cl}$ est remplacé par du $^{15}\text{NH}_4\text{Cl}$ (Euriso-Top, Saint Aubin, France).

2.13. Analyses RMN des métabolites extracellulaires

Les analyses en résonance magnétique nucléaire (RMN) sont réalisées sur un spectromètre Bruker Avance III 800 Mhz NMR équipé d'une cryo sonde 5 mm QCI-P. Les spectres 1D proton sont obtenus à 280 K, avec un pulse de 30° , un temps de relaxation de 10 s et une accumulation de 128 scans. Pour les HSQC 4096 x 128 points sont réalisés avec 16 scans sur une fenêtre de 13,35 x 60 ppm. Les spectres sont analysés avec le logiciel topspin 3.5 de Bruker (Billerica, Massachusetts, États-Unis).

2.14. Analyses HPLC-OrbiTrap des acides aminés libres

Pour l'analyse des acides aminés libres, les échantillons séchés et stockés à -20 °C sont resuspendus dans de l'eau déminéralisée à un volume final de 500 µL, les débris cellulaires sont éliminés par centrifugation 10 000 g pendant 10 min à 4 °C.

Pour l'analyse des acides aminés protéinogéniques, les échantillons séchés et stockés à -20 °C sont resuspendus dans 500 µL de solution d'acide chlorhydrique 5N (Sigma Aldrich, St. Quentin-Fallavier, France) et placés dans un incubateur à 70°C pendant une nuit. Les échantillons sont alors séchés à vide au RotaVap (ThermoSavant, Waltham, MA, USA). Les échantillons secs sont repris dans 500 µL d'eau déminéralisée pour l'analyse par chromatographie en phase liquide couplée à un spectromètre de masse LTQ Orbitrap Velos (Thermo Fisher Scientific, Waltham, MA, USA).

Les acides aminés sont séparés sur une colonne PFP (150 x 2.1 mm, taille des particules de la colonne 5 µM ; Supelco Bellefonte, PEN, USA) dans un système Ultimate 3000 HPLC (Dionex, CA, USA). Le solvant A est de l'acide formique 0,1 % (v/v) dans de l'eau et le solvant B est de l'acide formique à 0,1 % (v/v) dans de l'acétonitrile (Sigma) à un débit de 250 µL/min. Débit du solvant B dans le solvant A: 0 min, 2 % ; 2 min, 2 % ; 10 min, 5 % ; 16 min, 35 % ; 20 min, 100 %. La colonne est équilibrée 6min aux conditions initiales avant chaque échantillon. Le volume d'injection est de 20 µL.

Les analyses en scan entiers (fullscans) sont réalisées par un spectromètre de masse LTQ Orbitrap Velos (Thermo Fisher Scientific, Waltham, MA, USA) équipé d'une source de type électro spray (ESI). Les analyses sont réalisées en mode FTMS positif avec une résolution de 60 000 (à m/z 400) avec les paramètres de sources suivantes : température du capillaire 275 °C, température de la source à 250 °C, un débit de gaz principal à 45, des gaz auxiliaires à 20, un niveau de S Lens RF à 40 % et un voltage à 5 kV. Les données sont acquises avec le logiciel Xcalibur (Thermo Fisher Scientific, Waltham, MA, Etat-Unis) et traitées avec le logiciel TraceFinder (Thermo Fisher Scientific, Waltham, MA, Etat-Unis).

2.15. Analyses IC-OrbiTrap

Après reprise des cellules extraites dans un tampon acétate 20 mM à pH 9 (Sigma Aldrich, St. Quentin-Fallavier, France) à un volume final de 500 μ L, les débris cellulaires sont éliminés par centrifugation 10 000g pendant 10 min à 4°C. Les échantillons sont analysés par couplage d'une chromatographie ionique (IC) (Thermo Scientific Dionex ICS-5000+ system, Dionex, Sunnyvale, CA, Etat-Unis) couplé à un spectromètre de masse LTQ Orbitrap Velos (Thermo Fisher Scientific, Waltham, MA, USA) équipé d'une source d'ionisation de type électrospray.

La méthode en chromatographie ionique est une variante de la méthode de (Kiefer et al. 2007), visant à optimiser la séparation chromatographique du ppGpp et de pppGpp. Le ppGpp et le pppGpp sont séparés des autres métabolites par un gradient de 35 min avec un gradient de KOH sur une colonne IonPax AS11 (250 x 2 mm, Dionex) équipée d'une pré-colonne AG11 (50 x 2 mm, Dionex) à un débit de 0,35 ml/min. Le gradient est le suivant : 0 min, 0.5 mM; 5 min, 0.5 mM; 10 min, 10 mM; 15 min, 20 mM; 26 min, 20 mM; 31 min, 100 mM; 38 min, 100 mM; 38 min, 0.5 mM; 43 min, 0.5 mM. La colonne et le passeur automatique d'échantillon sont réglés à une température de 29 °C et de 4 °C respectivement. Le volume d'injection est de 15 μ L.

Les analyses en scan entiers (fullscans) sont réalisées par un spectromètre de masse LTQ Orbitrap Velos (Thermo Fisher Scientific, Waltham, MA, USA) équipé d'une source de type électrospray (ESI). Les analyses sont réalisées en mode FTMS négatif à une résolution de 30 000 (à $m/Z = 400$) en scan entier (fullscan), avec les paramètres de sources suivants: température du capillaire , 350 °C ; température de la source, 300 °C ; débit du gaz de la gaine, 50 AU (unité arbitraire); débit du gaz auxiliaire, 5 AU; niveau du S-Lens RF, 60 %; voltage du spray à ion, 3,5 kV. Les données sont acquises avec le logiciel Xcalibure (Thermo Fisher Scientific, Waltham, MA, Etat-Unis) et traitées avec le logiciel TraceFinder (Thermo Fisher Scientific, Waltham, MA, Etat-Unis).

Chapitre 3 – Quantification absolue du ppGpp et du pppGpp par LC-MS : Développement d'une méthode de dilution isotopique à deux dimensions

La croissance bactérienne est contrainte par la capacité des bactéries à transformer les nutriments de leur environnement en protéines (Bremer & Dennis 1996; Scott et al. 2010). Ce principe fait du métabolisme et de la traduction, les mécanismes limitants la croissance bactérienne. En effet, le métabolisme transforme les nutriments de l'environnement cellulaire en bases élémentaires (acides aminés, nucléotides, acyl-CoA...) et en énergie (oxydatif ou chimique), exploités ensuite pour la synthèse de macromolécules (protéines, ADN, lipides...). La traduction utilise les acides aminés et l'énergie produits par le métabolisme pour synthétiser les protéines intracellulaires. La synthèse de la protéine recombinante entraîne à la fois la consommation d'acides aminés et d'énergie ainsi que l'utilisation de la machinerie traductionnelle en compétition avec la synthèse des protéines constitutives de la biomasse (Scott et al. 2014). La compétition imposée par la production de protéines recombinantes conduit donc *de facto* à l'inhibition partielle de la croissance. La détection de la disponibilité des nutriments et la régulation de leur transformation en protéines intracellulaire est coordonnée par la réponse stringente (Cashel et al. 1996). Ce mécanisme s'avère central dans l'étude des contraintes métaboliques liées à la production de protéines recombinantes. Le contrôle de la réponse stringente sur ces processus bactériens est dépendant de la concentration intracellulaire en ppGpp et en pppGpp (Traxler, Summers, H.-T. Nguyen, et al. 2008; Mechold et al. 2013). La quantification de leurs concentrations intracellulaires au cours des procédés de production est donc une étape majeure à l'étude des contraintes métaboliques liées à la production de protéine recombinante chez *E. coli* BL21.

Dans ce chapitre est présentée une méthode de quantification absolue du ppGpp et du pppGpp publiée dans la revue *Analytical Chemistry*. Il s'agit de disposer d'une méthode basée sur la spectrométrie de masse, qui permette à la fois de mesurer la concentration de ppGpp mais également celle de pppGpp. Une telle méthode n'est pas décrite dans la littérature. Un problème auquel le développement de cette méthode s'est heurté, est la dégradation partielle des deux nucléotides et

plus précisément, la dégradation de la forme pentaphosphorylée en la forme tetraphosphorylée par hydrolyse d'un groupement phosphate. D'une manière générale, la dégradation des métabolites au cours du processus d'analyse est corrigée par l'utilisation d'étalons internes. En métabolomique, l'ensemble des métabolites servant d'étalon interne, est obtenu par culture de micro-organisme sur des substrats totalement marqués, en général au ^{13}C . Cela est parfaitement décrit dans les articles du groupe de S. Heijnen à l'université de Delft (Wu et al. 2005). Or, cette approche, bien que très efficace et utilisée en routine dans de nombreux laboratoires dans le monde, ne permet pas de corriger au cours du processus de traitement de l'échantillon, la formation d'un métabolite en un autre, en plus de la dégradation de ces même métabolites. Aussi, pour remédier à cette difficulté, une nouvelle méthode (décrite ci-dessous) a été développée. Cette méthode repose non pas sur l'addition d'un extrait de métabolites marqués (au ^{13}C), mais par l'addition de deux extraits de métabolites marqués (au ^{13}C et au ^{15}N) en tant qu'étalons internes.

Absolute quantification of ppGpp and pppGpp by double-spike isotope dilution ion-chromatography-high-resolution mass spectrometry

Clément PATACQ^{1,3}, Nicolas CHAUDET³, Fabien LETISSE^{1,2*}

1 : LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

2 : Université Paul Sabatier, Université de Toulouse, Toulouse, France.

3 : SANOFI PASTEUR, Département de Bioprocédés R&D, Marcy-L'Etoile, France

*fabien.letisse@insa-toulouse.fr

Abstract:

Guanosine 5'-diphosphate 3'-diphosphate (ppGpp) and guanosine 5'-triphosphate 3'-diphosphate (pppGpp) play a central role in the adaptation of bacterial and plant cells to nutritional and environmental stresses, and in bacterial resistance to antibiotics. These compounds have historically been detected and quantified by two-dimensional thin-layer chromatography of ³²P radiolabeled nucleotides. We report a new method to quantify ppGpp and pppGpp in complex biochemical matrix using ion-chromatography coupled to high-resolution mass spectrometry. The method is based on isotopic dilution mass spectrometry (IDMS) using ¹³C to accurately quantify the nucleotides. However, the loss of a phosphate group from pppGpp during the sample preparation process results in the erroneous quantification of ppGpp. This bias was corrected by adding an extra ¹⁵N isotope dilution dimension. This double-spike IDMS method was applied to quantify the ppGpp and pppGpp in *Escherichia coli* and in a mutant strain deleted for *gppA* (encoding the ppGpp phosphohydrolase) before and after exposure of both strains to serine hydroxamate, known to trigger the accumulation of these nucleotides.

Introduction:

Stringent response is a widespread bacterial adaptive mechanism that allows bacteria to overcome limited nutrition, stressful environments and antibiotic treatments. It consists in the homeostatic control of replication, transcription and translation through modulations of the intracellular concentrations of guanosine 5'-diphosphate 3'-diphosphate (ppGpp) and guanosine 5'-triphosphate 3'-diphosphate (pppGpp). The biosynthesis and degradation of both nucleotides are mediated by RSH (Rel Spo Homolog) enzymes, whose names derive from the RelA and SpoT proteins of *Escherichia coli* in which (p)ppGpp was originally detected. Synthesis by RSHs involves the transfer of a pyrophosphoryl group from ATP to the 3'-hydroxyl group of GDP or GTP, yielding ppGpp or pppGpp, respectively, whereas their degradation proceeds via the hydrolysis of the pyrophosphoryl group in 3' yielding GDP and GTP, respectively, and a pyrophosphate (Cashel et al. 1996). RSHs have also been identified in plants and algae (Takahashi et al. 2004; Field 2017) and (p)ppGpp has been recently shown to accumulate in chloroplasts with so far no clearly identified targets (Boniecka et al. 2017). In *E. coli*,

the contribution of (p)ppGpp to regulation and its effect on bacterial physiology have been better characterized (Potrykus & Cashel 2008; Dalebroux & Swanson 2012; Liu et al. 2015). However, although both ppGpp and pppGpp accumulate in *E. coli* cells (DeNapoli et al. 2013; Mechold et al. 2013), they are considered as a single metabolic pool acting together as secondary messengers. ppGpp and pppGpp accumulate in different ratios in different bacterial species and under different mutation and starvation conditions (DeNapoli et al. 2013; Mechold et al. 2013). More importantly, it has recently been reported that pppGpp and ppGpp are not equally potent regulators of growth rate (Mechold et al. 2013). This highlights the need to individually quantify ppGpp and pppGpp to provide further insights into their regulation of bacteria, plants and other organisms.

The generally accepted method for (p)ppGpp level analysis involves radiolabeling (p)ppGpp with ^{32}P and separating the different forms by thin-layer chromatography (Battesti & Bouveret 2006; Mechold et al. 2013; Reidl et al. 2017). This technique is sensitive but is essentially qualitative. HPLC-based methods have also been developed to quantify just ppGpp but these require large amounts of biological material for ppGpp to be detected because of poor sensitivity (Cserjan-Puschmann et al. 1999; Simen et al. 2017; Varik et al. 2017). This has recently been drastically improved by coupling a UPLC system with a tandem quadrupole mass spectrometer (Ihara et al. 2015). None of the methods mentioned above use internal standards to correct for inherent biases from sample preparation and/or the analytical techniques. In this paper, we report the development of an ion-chromatography coupled to high resolution mass spectrometry (IC-ESI-HRMS) method to accurately quantify ppGpp and pppGpp using isotope dilution mass spectrometry (IDMS). To achieve this, we biologically produced fully ^{13}C labeled species of both nucleotides using *Escherichia coli* cultivated on $\text{U-}^{13}\text{C}$ glucose and exposed to serine hydroxamate (SHX), which is known to trigger stringent response and thus lead to the accumulation of ppGpp and pppGpp. Spiking the ^{13}C -labeled metabolites produced in this way into the samples before metabolite extraction allows experimental (partial loss of metabolites during the extraction) and analytical (mainly ion suppression and matrix effects) biases to be corrected (Mashego et al. 2004; Wu et al. 2005). However, we also observed the loss of a phosphate group from the pentaphosphate form of the nucleotide during the sample preparation process, converting a significant proportion of the pppGpp to ppGpp. This cannot be corrected by only spiking fully ^{13}C labeled nucleotides into the samples and requires adding a second dimension of IDMS (García Alonso &

Rodríguez-Gonzalez 2013). So in addition to ^{13}C -labeled metabolites, we spiked into the samples ^{15}N -labeled metabolites from an *E. coli* strain that accumulates ppGpp. The combination of both isotope dilution dimensions fully corrects for experimental (degradation and formation) and analytical biases and offers accurate individual quantification of ppGpp and pppGpp. This novel method integrates thus the principles of double-spike isotope dilution mass spectrometry (García Alonso & Rodríguez-Gonzalez 2013) with the principles for the absolute quantification of intracellular metabolites by the use of labeled metabolite extracts as internal standards (Wu et al. 2005). Finally, the benefits of this variant of double-spike IDMS method were demonstrated by quantifying these nucleotides in *E. coli* BW25113 and in a mutant deleted for the gene *gppA*—which encodes the *pppGpp pyrophosphatase*—before and after exposure to SHX.

Experimental Section

Chemicals and reagents:

ppGpp and pppGpp standards were purchased from Jena Bioscience (Jena, Germany) and DL-Serine Hydroxamate (SHX) was purchased from Sigma Aldrich (St. Quentin-Fallavier, France). LC-MS grade solvents (methanol, acetonitrile) were obtained from Instrumentation Consommables et Service (ICS, Lapeyrouse-Fossat, France).

Bacterial strains and growth conditions

Escherichia coli strain BW25113 and the *gppA*-deleted mutant BW25113 (both from the Keio collection), and *E. coli* BL21 (DE3) were cultured on minimal synthetic medium containing 5 mM KH_2PO_4 , 10 mM Na_2HPO_4 , 9 mM NaCl, 40 mM NH_4Cl , 0.8 mM MgSO_4 , 0.1 mM CaCl_2 , and 30 g/L glucose. Thiamine was added to a final concentration of 0.1 g/L to cultivate the BW25113 strains. Glucose, thiamine and MgSO_4 were sterilized by filtration (Minisart polyamide 0.2 μm , Sartorius, Göttingen, Germany), and other solutions were autoclaved separately. All stock cultures were stored at -80°C in Luria-Bertani (LB) medium containing glycerol (40%, v/v). For the cultures, 5 mL of overnight cultures in LB were used as inoculum and then sub-cultured in shake-flasks containing 50 mL of

minimum medium with 3 g/L glucose starting at OD_{600nm} = 0.05 and incubated at 30°C and 210 rpm in an orbital shaker (Inova 4230, New Brunswick Scientific, New Brunswick, NJ, USA). Cells were harvested during the exponential growth phase by centrifugation for 10 min at 10,000 *g* at room temperature with a Sigma 3-18K centrifuge (Sigma Aldrich, Seelze, Germany), washed with the same volume of fresh medium (without glucose or thiamine), and used to inoculate 500 mL bioreactors (Multifors, Infors HT, Bottmingen, Switzerland) containing 300 mL of 30 g/L minimal medium with an OD_{600nm} of 0.15. The temperature was set to 37°C and the pH was maintained at 7 by automatically adding 14% (g/g) ammonia (VWR, Fontenay-sous-Bois, France) or 11% (g/g) phosphoric acid (Sigma Aldrich, St. Quentin-Fallavier, France). Aeration and the stirrer speed were controlled to maintain adequate aeration (DOT > 30% saturation). Cell growth was monitored through the optical density at 600 nm with a Genesys 6 spectrophotometer (Thermo, Carlsbad, CA, USA). Stringent response was triggered by adding SHX to the culture (final SHX concentration of 0.8 mM when the OD reached 3.5).

Sampling and metabolite extraction

Culture medium (400 µL) was withdrawn from the bioreactor and vigorously mixed with 4.5 mL of a pre-cooled acetonitrile/methanol/H₂O (4:4:2) solution at -40°C to rapidly quench metabolic activity (Millard et al. 2014). Immediately thereafter, 100 µL of ¹³C labeled metabolites and 100 µL of ¹⁵N labeled metabolites (see below for the preparation of the ¹³C- and ¹⁵N- labeled samples) were added to the latter mixture as internal standards. The tubes were then placed in a cooling bath of ethanol pre-cooled at -40°C and evaporated to dryness in a SpeedVac (SC110A SpeedVac Plus, ThermoSavant, Waltham, MA, USA) under vacuum for 4 h and then stored at -80°C until needed.

Biological production of ¹³C- and ¹⁵N- labeled metabolites

¹³C labeled metabolites were obtained from a culture of *E. coli* BW25113 Δ *gppA*. The unlabeled glucose (¹²C-glucose) in the medium was replaced by U-¹³C labeled glucose (Euriso-Top, Saint Aubin, France). As described above, SHX was added to the culture medium when the OD reached 3.5, and 15 min after exposure to SHX, the samples of the culture were quenched in a pre-cooled acetonitrile/methanol/H₂O (4:4:2) solution at -40°C. The tubes were then placed in a cooling bath of ethanol precooled at -40°C and evaporated to dryness in a SpeedVac (SC110A SpeedVac Plus, ThermoSavant, Waltham, MA, USA) under vacuum for 4 h. The ¹³C labeled metabolites were

resuspended in deionized water and pooled together before being aliquoted and then stored at -80°C until further usage. The same protocol was used to produce the ^{15}N labeled metabolites by using *E. coli* strain BL21 instead of *E. coli* strain BW25113, and by replacing unlabeled NH_4Cl ($^{14}\text{NH}_4\text{Cl}$) with ^{15}N labeled NH_4Cl (Euriso-Top, Saint Aubin, France).

IC-ESI-HRMS analysis of ppGpp and pppGpp

After resuspension of the cell extract samples in 20 mM ammonium acetate buffer at pH 9 (Sigma Aldrich, St. Quentin-Fallavier, France) to a final volume of 500 μL , cell debris were removed by centrifugation at 10,000 g for 10 min at 4°C . The samples were then analyzed by ion chromatography (Thermo Scientific Dionex ICS-5000+ system, Dionex, Sunnyvale, CA, USA) coupled to a LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) equipped with an electrospray ionization probe. The ion chromatography method is a variant of the method described by Kiefer et al. (Kiefer et al. 2007) to improve the chromatographic separation of ppGpp and pppGpp. The KOH gradient was modified as follows: 0 min, 0.5 mM; 5 min, 0.5 mM; 10 min, 10 mM; 15 min, 20 mM; 26 min, 20 mM; 31 min, 100 mM; 38 min, 100 mM; 38 min, 0.5 mM; 43 min, 0.5 mM. The gradients were all linear. For background suppression, an Anionic Electrolytically Regenerated Suppressor (AERS 300 – 2 mm) from Dionex was used. The device was operated at constant electrolysis current of 87 mA, in external-water mode with ultrapure water and regenerant was delivered by an external AXP pump at a flow rate of 1 mL/min. In all cases, the volume of the injected sample was 15 μL . Mass spectrometry analysis was performed in the negative FTMS mode at a resolution of 30,000 (at $m/z = 400$) in fullscan mode, with the following source parameters: capillary temperature, 350°C ; source heater temperature, 300°C ; sheath gas flow rate, 50 a.u. (arbitrary unit); auxiliary gas flow rate, 5 a.u.; S-Lens RF level, 60%; and ion spray voltage, 3.5 kV. The data were acquired using the Xcalibur software (Thermo Fisher Scientific, Waltham, MA, USA).

Results

1. Mass spectrometry and ion chromatography

We first analyzed standard solutions of ppGpp and pppGpp by direct infusion electrospray ionization (ESI) on a LTQ-Orbitrap operated in negative mode to find the major ions and to optimize

the electrospray ionization parameters and MS settings. Figure 1 shows the MS spectra obtained for each nucleotide at a resolution of 30,000 (at 400 m/z) after the parameters were optimized.

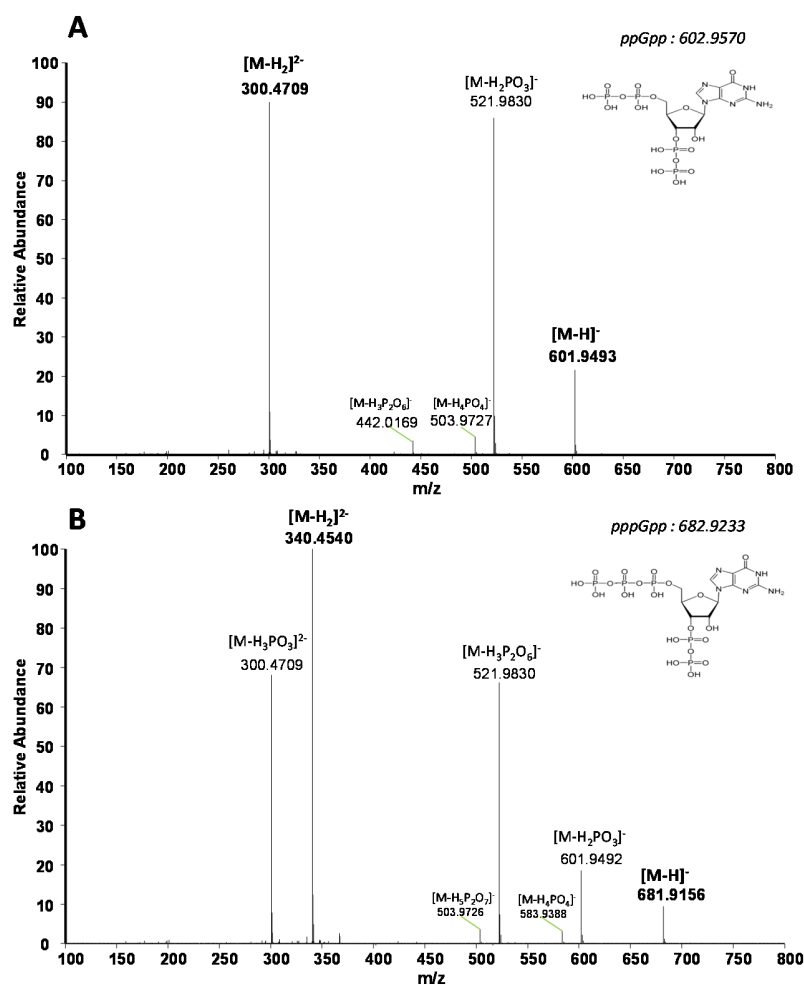


Figure 1. ESI LTQ-Orbitrap XL mass spectra of ppGpp (A) and pppGpp (B) (5 μ M in 0.1% (v/v) aqueous formic acid solution) recorded in negative mode. The *m/z* values in bold correspond to molecular ions at *m/z* = 601.9493 and 681.9156, and doubly charged forms of ppGpp and pppGpp at m/z_2 = 300.4709 and 340.4540, respectively.

Both ppGpp and pppGpp were detected as singly and doubly charged anionic forms, $[M-H]^{-1}$ at *m/z* = 601.9493 and *m/z* = 681.9156; $[M-2H]^{2-}$ at *m/z* = 300.4709 and *m/z* = 340.4540, respectively, the latter ions giving the most intense signals. Additionally, in-source fragmentation with neutral losses of

phosphate moieties (79.9663 Da and 97.9769 Da) resulted in the detection of ions at $m/z = 521.9830$ and 442.0169 for ppGpp, and at $m/z = 601.9492$, $m/z = 300.4709$ and $m/z = 521.9830$ for pppGpp.

We then used ion-chromatography (IC), whose efficiency for the separation of nucleotides and more generally, phosphorylated compounds, is proven (Enjalbert et al. 2013; Van Dam et al. 2002; Burgess et al. 2011). The KOH gradient applied to elute the compounds was optimized to separate ppGpp and pppGpp within a reasonable amount of time (see Material and Methods section for details). The optimized gradient fully separates the two nucleotides, with retention times of 25.8 min for ppGpp and 26.5 min for pppGpp (Figure 2; panels A and D).

2. (p)ppGpp degradation during the extraction procedure

It is known that extraction procedures can chemically degrade nucleotides because of the labile phosphoryl group (Rabinowitz & Kimball 2007; Gil et al. 2017). Various phosphorylated guanosine derivatives can thus be formed during the extraction of (p)ppGpp. We therefore used an extraction procedure for phosphorylated metabolites with a cold (-40°C) acetonitrile/methanol/water mixture (Millard et al. 2014; Enjalbert et al. 2013) for single standard solutions of ppGpp and pppGpp at $2.5\ \mu\text{M}$. The samples were then analyzed using the IC-ESI-MS method described above. Figure 2 shows extracted ion chromatograms (XICs) of phosphorylated guanosine derivatives, namely guanosine bisphosphate (G+2p), triphosphate (G+3p), tetraphosphate (G+4p) and pentaphosphate (G+5p).

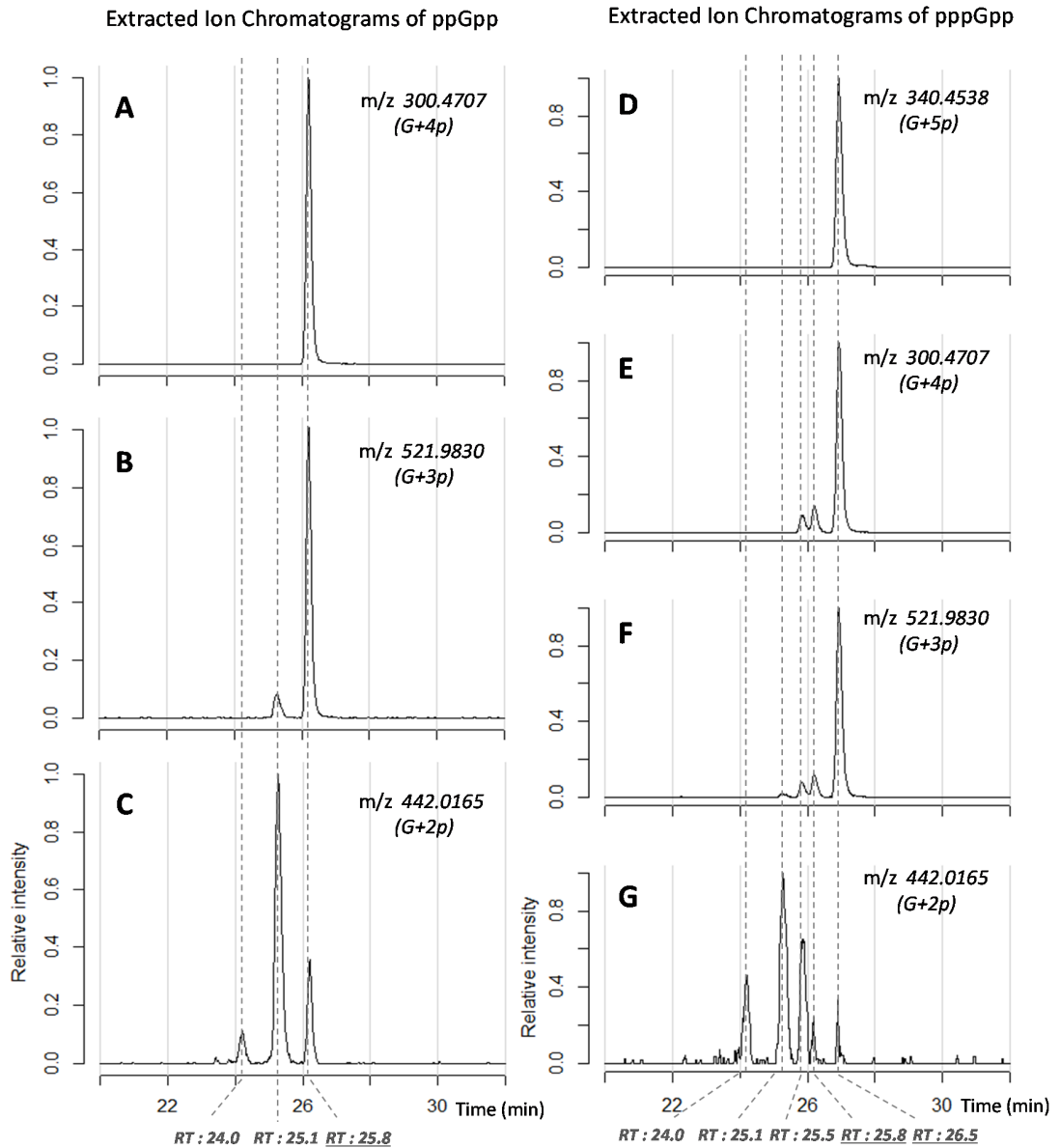


Figure 2. Representative IC-MS chromatograms of 2.5 μ M standard solutions of ppGpp (A,B,C) or pppGpp (D,E,F,G) submitted to the metabolite extraction procedure. The IC-MS chromatograms are extracted ion chromatograms obtained at $m/z = 300.4707$ (A,E), $m/z = 521.9830$ (B,F), $m/z = 442.0165$ (C,G), and $m/z = 340.4538$ (D). The retention times of ppGpp and pppGpp are underlined. G refers to Guanosine and p to a phosphoryl group.

The XICs of the ppGpp standard solution showed several separated peaks in addition to that of ppGpp resulting from the loss of phosphoryl group(s) in either the 3' or 5' position. Similarly, guanosine

derivatives were also detected after pppGpp extraction, including a peak at $m/z = 300.4707$ in the XIC whose retention time (25.8 min) matches that of ppGpp. This suggests that pppGpp may be converted into ppGpp during extraction. Note that we also detected a peak at $m/z = 601.9492$ with the same retention time, corresponding to the $[M-H]^-$ ions of ppGpp (data not shown). This decomposition of ppGpp and pppGpp during cold-solvent based extraction by cleavage of one or several phosphodiester bonds is not surprising since it has been reported for other nucleotides by Kimball & Rabinowitz (Kimball & Rabinowitz 2006) and more recently by Gil et al. (Gil et al. 2017). However, the decomposition of pppGpp into ppGpp prevents the reliable and accurate quantification of the latter as demonstrated below. To overcome this quantification bias—in addition to those inherent to the LC-ESI-MS technique, we developed an isotopic dilution-based method with two isotopic dimensions.

3. A double-spike IDMS method for accurate quantification of ppGpp and pppGpp

General principle

Isotope dilution mass spectrometry (IDMS) is a powerful method to overcome the drawbacks of LC-ESI-MS analysis, such as ion suppression and matrix effects, as well as the partial loss of metabolites during extraction procedures (Wu et al. 2005). Briefly, concentrations of metabolites are quantified relative to the concentration of their $U-^{13}C$ labeled forms (Mashego et al. 2004). Analytical and experimental biases are eliminated because the two isotopic forms (^{12}C and ^{13}C) of a given compound are chemically identical (figure 3; panel A, left hand – panel C, IDMS equation). This $U-^{13}C$ internal standard method is thus routinely and successfully applied to quantify diverse metabolomes in various organisms (Wu et al. 2005; Bennett et al. 2009; Enjalbert et al. 2013; Kim et al. 2015; Wang et al. 2017). *Although this is not explicitly mentioned in the literature*, this approach cannot however correct for the partial formation of a given metabolite (X) from the degradation of another metabolite (Y). Indeed, this contributes to the peak area of the $U-^{12}C$ ($(Ax)^{12}C$) metabolite and to that of its $U-^{13}C$ counterpart ($(Ax)^{13}C$) but in an independent manner. As illustrated in figure 3 (panel A, right hand), the $(Ax)^{12}C / (Ax)^{13}C$ ratio ceases to be representative of the metabolite concentration. Nevertheless, one can correct for the partial formation of a given metabolite provided i) the metabolite(s) contributing to its formation is (are) known and, ii) it (they) can be quantified. The contribution resulting from the degradation of metabolite Y could be determined by performing the extraction procedure with a

standard solution of Y and then quantifying the amount of metabolite X formed during its degradation. Due to sample matrix effects however, the degradation of a metabolite during the extraction procedure and its contribution to the formed metabolite may vary from one sample to another, meaning that internal standards are required to properly correct for the formation of X.

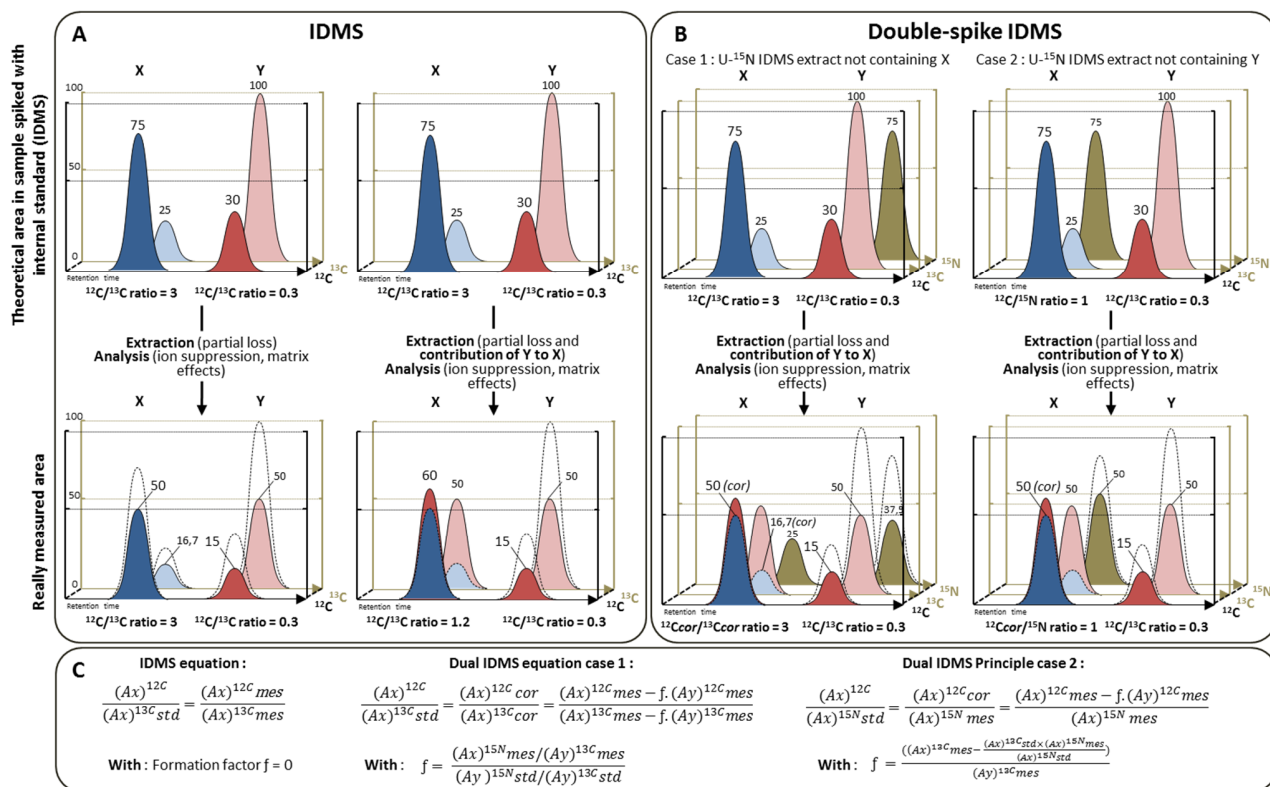


Figure 3. Schematic descriptions of the principles of IDMS and double-spike IDMS. (A, left) IDMS correction for the drawbacks of LC-ESI-MS and the partial degradation of metabolites X and Y. (A, right) IDMS correction also considering the partial formation of a given metabolite (here X) from the degradation of another metabolite (here Y). (B) Double-spike IDMS correction using a second isotopic dimension (¹⁵N), shown for metabolite X according to whether the ¹⁵N labeled metabolite extract contains only Y or only X. cor refers to the corrected peak area. (C) Equations used for bias correction where (A_i)^j refers to the theoretical peak area in sample spiked with internal standard (std), measured peak area (mes) or corrected peak area (cor) of metabolite i, for the isotopic dimension j. The principles of IDMS and double-spike IDMS are highlighted by applying degradation coefficients on metabolites X and Y of 2/3 and 1/2 respectively, and by considering that the degradation of 1 unit of Y creates 1 unit

of X.

Introducing metabolites labeled with a second isotope as internal standards makes it possible to correct for partial formation of metabolites during extraction procedures. This approach has been known for years to be efficient to account specifically for chemical interconversion (Ouerdane & Meija 2009; García Alonso & Rodríguez-Gonzalez 2013) and its efficiency has recently been demonstrated for organic compounds (Fernández-Fernández, Mario Rodríguez-González et al. 2015). Known as “multiple-spike isotope dilution”, it relies on the addition to the sample of species labeled with a different isotope of the same element at the beginning of the sample preparation procedure for all the interconverting molecules. However, there is so far no application of this method - or variations thereof - to simultaneously account for partial loss of metabolites during the sample preparation process. Moreover, it requires the additional labeled forms to be commercially available, or otherwise (bio)chemically or biologically synthesizable. But for organic compounds (e.g. nucleotides), the required isotopic forms can rarely be purchased and their chemical synthesis and purification can be very complex and in some cases impossible. Unless a form fully labeled with a particular element is used, the biological alternative leads to isotopologue distributions that are not predictable since they depend in most cases on the metabolic state of the organism i.e. on the distribution of metabolic fluxes (Millard et al. 2014).

To overcome these difficulties, we developed a novel variation of multiple-spike IDMS. This method is well suited for solving the absolute quantification problem that arises from the partial formation of a given metabolite (i.e. X) from the degradation of another (i.e. Y), while correcting for its partial loss during sample processing. To achieve this, samples were spiked with two labeled metabolite extracts. Importantly, one spiked sample should neither contain the nascent metabolite nor the one that gives rise to it via degradation. Furthermore, instead of using metabolites labeled with two different isotopes of the same element, we exploit the fact that organic compounds contain several chemical elements by adding isotopes of two different elements (namely carbon and nitrogen). The biological synthesis of the isotopologues of interest (see below) is thus achieved by using fully labeled forms of glucose ($U\text{-}^{13}\text{C}$) and ammonium chloride (^{15}N) which are commercially available and affordable. The novelty of the method lies thus on the combined uses of multiple-spike isotope dilution

and the principle of metabolites absolute quantification based on the addition of fully labeled metabolite extracts as internal standards. This method relies on the addition of two isotopic dimensions (here ^{13}C and ^{15}N) and therefore referred to hereafter as “double-spike IDMS”.

To illustrate the principle of this method, let us consider a first isotopic dimension obtained with U- ^{13}C labeled metabolites and a second with U- ^{15}N labeled metabolites. We will then distinguish two cases based on whether the second isotopic dimension contains solely Y or X. In the former case (Figure 3; panel B, case 1), because the U- ^{15}N dimension does not contain X, the U- ^{15}N -X peak ($(\text{Ax})^{15\text{N}}$) comes exclusively from the decomposition of Y. One can thus correct the $(\text{Ax})^{12\text{C}}$ and $(\text{Ax})^{13\text{C}}$ peak areas for its contribution using the following equation (equation 1 – Figure 3; panel C, double-spike IDMS equation case 1 – see supplementary Note 1 for details).

$$(\text{Ax})^{12\text{C} \text{ (or } 13\text{C)}} \text{ cor} = (\text{Ax})^{12\text{C} \text{ (or } 13\text{C)}} \text{ mes} - \frac{(\text{Ax})^{15\text{N}} \text{ mes} / (\text{Ay})^{13\text{C}} \text{ mes}}{(\text{Ay})^{15\text{N}} \text{ std} / (\text{Ay})^{13\text{C}} \text{ std}} \cdot (\text{Ay})^{12\text{C} \text{ (or } 13\text{C)}} \text{ mes} \quad (1)$$

where Ax is the peak area of metabolite X and Ay the peak area of metabolite Y, measured in the sample (*mes*) or in the standard solution spiked with ^{13}C - and ^{15}N -*E. coli* extracts (*sdt*) or corrected for experimental biases (*cor*). The isotopic dimensions of the peaks are shown as superscripts.

The absolute concentration of X can therefore be deduced from the $(\text{Ax})^{12\text{C}} \text{ cor} / (\text{Ax})^{13\text{C}} \text{ cor}$ ratio of peak areas using the appropriate calibration curve (e.g plotting the ratio of peak areas $(\text{Ax})^{12\text{C}} \text{ sdt} / (\text{Ax})^{13\text{C}} \text{ sdt}$ against the known concentration of U- ^{12}C -X).

In the second case (Figure 3; panel B, case 2), in which the second isotopic dimension contains X only, the corrected peak area for X ($(\text{Ax})^{12\text{C}} \text{ cor}$) can be obtained by subtracting the contribution from the decomposition of Y from the peak area measured for X as per equation (2) (see also figure 3; panel C, double-spike IDMS equation case 2 and supplementary Note 1).

$$(\text{Ax})^{12\text{C}} \text{ cor} = (\text{Ax})^{12\text{C}} \text{ mes} - \frac{((\text{Ax})^{13\text{C}} \text{ mes} - \frac{(\text{Ax})^{13\text{C}} \text{ sdt} \times (\text{Ax})^{15\text{N}} \text{ mes}}{(\text{Ax})^{15\text{N}} \text{ sdt}})}{(\text{Ay})^{13\text{C}} \text{ mes}} \cdot (\text{Ay})^{12\text{C}} \text{ mes} \quad (2)$$

Since the U- ^{15}N dimension does not contain Y, the U- ^{15}N -X metabolite is only subjected to degradation and can thus be used as an internal standard for X, to correct for ion suppression, matrix

effects, partial loss, etc. Its absolute concentration can thus be determined from a calibration curve plotting the ratio of peak areas $(A_x)^{12C_{sdt}}/(A_x)^{15N_{sdt}}$ against the known concentration of U-¹²C-X.

In summary, this double-spike IDMS method allows X and Y to be quantified, while correcting for all analytical and experimental biases. This approach is therefore ideal to individually quantify pppGpp and ppGpp.

Biological production of IDMS standards

To do this, we first produced U-¹³C labeled metabolites biologically using *E. coli* BW25113 deleted for the gene *gppA* (which encodes the pppGpp 5'-gamma phosphate hydrolase that converts pppGpp to ppGpp), since this strain is known to accumulate high concentrations of pppGpp and ppGpp during stringent response (DeNapoli et al. 2013). *E. coli* BW25113 $\Delta gppA$ was thus grown in minimal medium containing 100% U-¹³C glucose. ppGpp and pppGpp accumulation was triggered by adding SHX. U-¹³C labeled metabolites were extracted 30 min after adding SHX using the cold-solvent based extraction procedure. The isotopologue distributions of ppGpp and pppGpp present in the U-¹³C labeled extract are shown in figure 4.

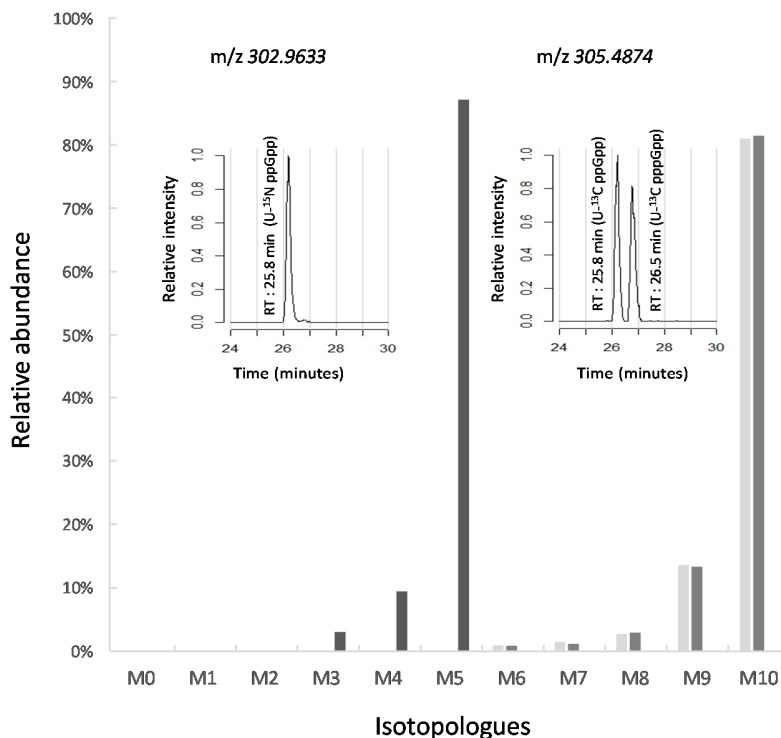


Figure 4. Isotopologue distribution of ^{13}C labeled ppGpp (light grey) and pppGpp (dark grey), and ^{15}N labeled ppGpp (black). The insets are the extracted ion chromatograms obtained at $m/z = 305.4874$ from a ^{13}C labeled metabolite extract, showing two peaks corresponding to $\text{U-}^{13}\text{C}$ -ppGpp (RT = 25.8 min) and $\text{U-}^{13}\text{C}$ -pppGpp (RT = 26.5 min), and that obtained at $m/z = 302.9633$ from a ^{15}N labeled metabolite extract, showing a single peak from $\text{U-}^{15}\text{N}$ -ppGpp (RT = 25.8 min).

To obtain the second isotopic dimension, we had to find a condition where only one of the two nucleotides accumulates in the cells. We therefore screened for several *E. coli* strains and derivatives and found that the strain *E. coli* BL21 only accumulated ppGpp in the presence of SHX. Note that we did not find any strains that only accumulate pppGpp. Using the same procedure as described above, $\text{U-}^{15}\text{N}$ labeled metabolites were extracted from *E. coli* BL21 cells grown in minimal medium containing 100% $^{15}\text{NH}_4\text{Cl}$ as the only source of nitrogen and exposed to SHX. The isotopologue distribution of ppGpp present in the $\text{U-}^{15}\text{N}$ labeled extract is shown in figure 4. The ^{13}C - and ^{15}N -labeled extracts were then used as internal standards and correspond to the two isotopic dimensions. Because the latter extract did not contain pppGpp, it corresponds to *case 2* described above for the use of double-spike IDMS and we applied this method accordingly.

4. Quantification by double-spike IDMS of purified ppGpp and pppGpp

First, the limits of quantification (LOQ), defined here as the lowest concentration of standard solution that can be determined with a standard deviation of less than 20%, were estimated using standard solutions of ppGpp and pppGpp spiked with ^{13}C - and ^{15}N -*E. coli* extracts mixed in equal volumes (20 μL each). We determined LOQ of 11.7 nM for both ppGpp and pppGpp. High linearity ($R^2 > 0.99$) was observed for ppGpp and pppGpp up to at least 6 μM (Figure 5).

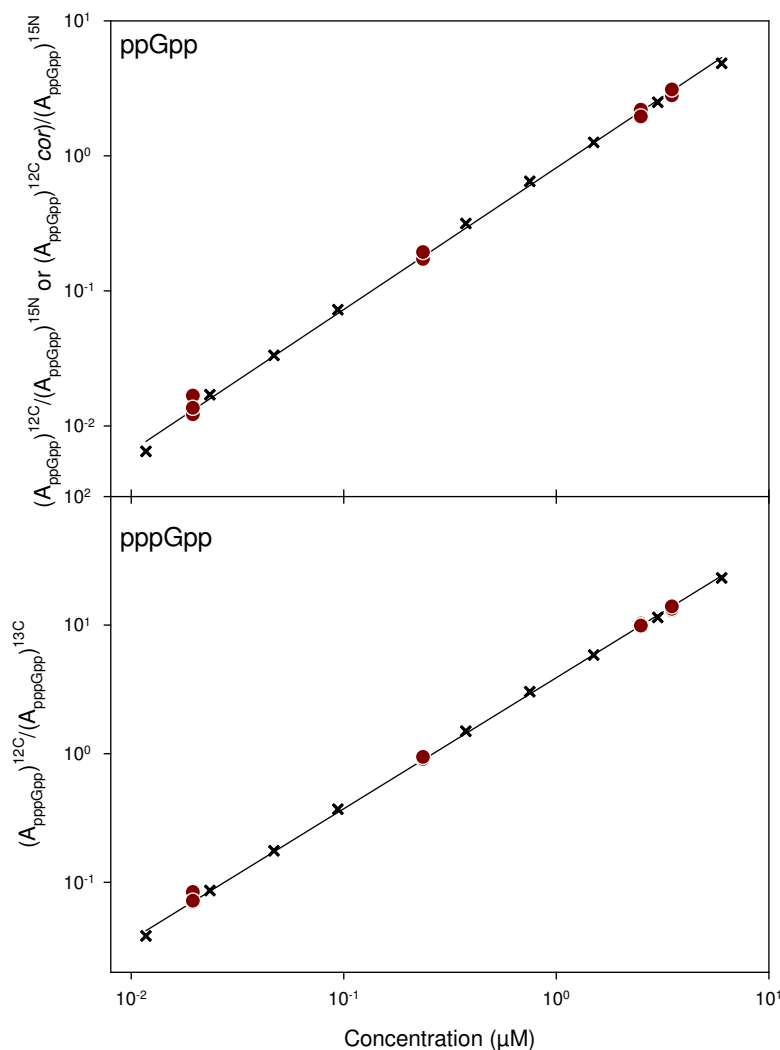


Figure 5. Calibration curves (x) and recovery rates (●) for ppGpp and pppGpp. For the calibration curves, standards were freshly diluted and spiked with U-¹³C- and U-¹⁵N-labeled metabolite extracts. The ratios of peak areas for ppGpp and pppGpp with respect to internal standards ($(A_{ppGpp})^{12C}/(A_{ppGpp})^{15N}$ or $(A_{pppGpp})^{12C}/(A_{pppGpp})^{13C}$ respectively) are plotted against the concentrations of the standards. For the recovery rates, known amounts of ppGpp and pppGpp in mixtures spiked with U-¹³C- and U-¹⁵N-labeled metabolite extracts were extracted using the cold solvent method. Ratios of peak areas for ppGpp (corrected for the contribution from pppGpp degradation) and pppGpp with respect to internal standards ($(A_{ppGpp})^{12C}cor/(A_{ppGpp})^{15N}$ or $(A_{pppGpp})^{12C}/(A_{pppGpp})^{13C}$ respectively) are plotted against the theoretical concentrations.

To test whether double-spike IDMS allows to fully correct biases occurring during the extraction procedure (degradation and formation), we then spiked known amounts of ppGpp and pppGpp mixed with the same volumes of ^{13}C - and ^{15}N -*E. coli* extracts as described above. These solutions (4 concentrations/3 replicates each) were submitted to the extraction procedure and then analyzed by IC-ESI-MS. The concentrations of ppGpp and pppGpp were chosen to cover the linearity ranges determined for each nucleotide, i.e. roughly twice the LOQ, 5%, 40% and 60% of the upper LOQ (6 μM). Figure 5 shows that the $(\text{Ax})^{12}\text{C}_{cor}/(\text{Ax})^{15}\text{N}_{mes}$ and the $(\text{Ay})^{12}\text{C}_{mes}/(\text{Ay})^{13}\text{C}_{mes}$ ratios measured respectively for ppGpp and pppGpp are on the regression lines for all the chosen concentrations meaning that double-spike IDMS corrects for experimental and analytical biases very effectively.

5. Application of double-spike IDMS to the quantification of intracellular levels of ppGpp and pppGpp in *E. coli*

We then used double-spike IDMS to quantify intracellular levels of ppGpp and pppGpp in *E. coli*. This bacteria is known to accumulate (p)ppGpp during amino acid starvation, which can be mimicked by exposing *E. coli* to SHX (Li et al. 2016; Samarrai et al. 2011; Traxler, Summers, H.-T. Nguyen, et al. 2008). SHX (0.8 mM) was thus added to exponentially growing cultures of *E. coli* BW25113 and *E. coli* BW25113 $\Delta gppA$. The intracellular levels of pppGpp should be higher in the mutant (with the pppGpp pyrophosphatase gene knocked out) than in the WT strain. Figure 6 shows the absolute concentrations (3 replicates) of both nucleotides before and 30 min after SHX exposure, determined using double-spike IDMS. The absolute concentrations of ppGpp and pppGpp are respectively deduced from the $(\text{Ax})^{12}\text{C}_{cor}/(\text{Ax})^{15}\text{N}_{mes}$ ratios and from the $(\text{Ay})^{12}\text{C}_{mes}/(\text{Ay})^{13}\text{C}_{mes}$ ratios. To assess whether the level of pppGpp has an impact on the quantification of ppGpp, we also determined the contribution of the ppGpp formed from the degradation of pppGpp, which can be deduced from the $(\text{Ax})^{12}\text{C}_{mes}/(\text{Ax})^{15}\text{N}_{mes}$ ratio using equation 2.

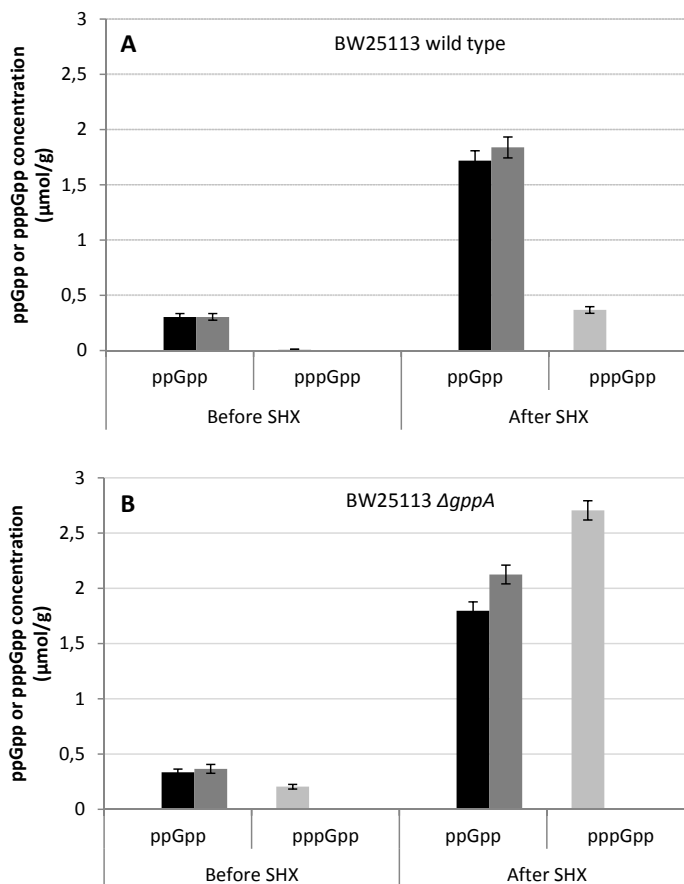


Figure 6. Absolute concentrations of ppGpp (black bars) and pppGpp (light grey bars) quantified using double-spike IDMS in wild type *E. coli* BW25113 (A) and in *E. coli* BW25113 $\Delta gppA$ (B) before and 30 min after adding serine hydroxamate. The dark grey bars show the absolute concentration of ppGpp not corrected for its formation from the degradation of pppGpp.

Firstly, the exposure of *E. coli* BW25113 to SHX results clearly in the accumulation of ppGpp, as expected, and also of pppGpp to a significant level (DeNapoli et al. 2013; Mechold et al. 2013). In the wild type strain (figure 6; panel A), the basal ppGpp level and that determined after adding SHX (0.30 μmol/g and 1.71 μmol/g, respectively, or approximately 0.14 mM and 0.80 mM assuming a cell volume of 2.15 mL/g) (Pramanik & Keasling 1997) are similar to those reported previously (Cashel et al. 1996; Simen et al. 2017). The absolute concentrations of pppGpp before and after exposure to SHX are respectively 0.010 μmol/g and 0.366 μmol/g. As expected, the intracellular concentration of pppGpp in the $\Delta gppA$ mutant (figure 6; panel B), is *higher, with an absolute concentration seven times higher than*

in the WT strain, without any significant variation in the ppGpp level. In the mutant, the concentration of the pentaphosphate form is even higher than that of the tetraphosphate form. Interestingly, while pppGpp decomposition makes an insignificant contribution to the concentration of ppGpp in the WT strain, it accounts for 18% of the ppGpp in the $\Delta gppA$ mutant. This demonstrates i) that the quantification method developed in this study can be used to individually quantify ppGpp and pppGpp in cell extracts and, ii) that this double-spike IDMS method has to be used to accurately quantify ppGpp levels.

Conclusion

A sensitive and robust IC-ESI-HRMS method has been developed for the accurate individual quantification of ppGpp and pppGpp in cell extracts. This method includes correction for ppGpp formed from the degradation of pppGpp, thereby avoiding any erroneous quantification of the tetraphosphate form. This was achieved by introducing a second isotopic dimension to the IDMS method. As a proof of principle of this double-spike IDMS approach, the two nucleotides were quantified in *E. coli* extracts. This method could also be applied to quantify ppGpp and pppGpp in other organisms and should thus contribute to clarifying the distinct physiological roles of pppGpp and ppGpp in stringent response, which still remains unclear. This study shows that this double spike IDMS method is a powerful and robust innovation for the accurate quantification of metabolites in cell extracts. Its principle could be extended to other metabolites likely to be formed in part by the degradation of other metabolites (Kimball & Rabinowitz 2006; Gil et al. 2015). We therefore believe that double-spike IDMS for quantitative purpose is a promising new tool in metabolomics for the comprehensive quantitative analysis of cellular metabolism.

Associated content

Supporting Information

Additional information as indicated in the text. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

Author information

Corresponding Author

* Telephone: +33(0)5-61-55-94-07. E-mail: fabien.letisse@insa-toulouse.fr.

Notes

The authors declare no competing financial interest.

Acknowledgment

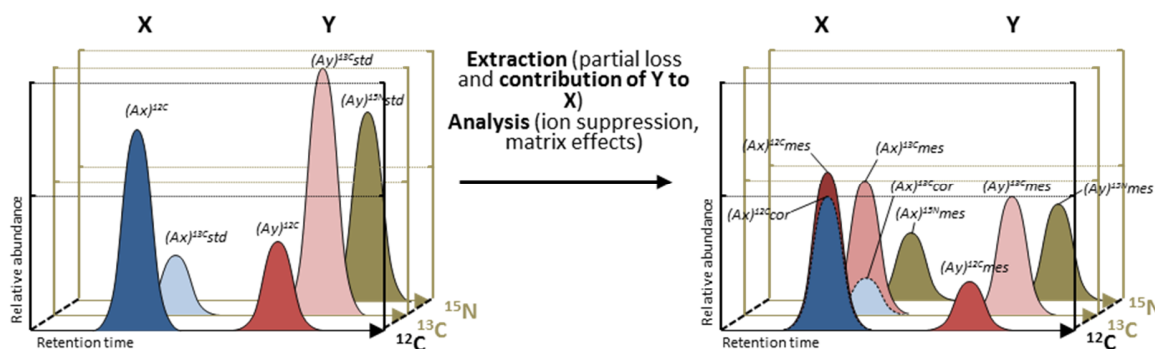
The authors gratefully acknowledge financial support from Bioprocess R&D department of Sanofi Pasteur and from l'Association Nationale de la Technologique (ANRT). The authors also thank MetaToul (Metabolomics & Fluxomics Facilities, Toulouse, France, www.metatoul.fr) and its staff for technical support and access to NMR and mass spectrometry facilities. MetaToul is part of the MetaboHUB-ANR-11-INBS-0010 national infrastructure (www.metabohub.fr).

Supplementary Note 1

Establishment of the equations used for bias correction

Let consider two chromatographically resolved compounds, X and Y, both subjected to partial degradation during extraction and sampling procedure. Let also consider that the decomposition of Y can give X whereas X cannot be converted into Y. All the biases (partial degradation of X and Y, partial formation of X from Y, ion suppression, matrix effects, etc.) can be eliminated based on IDMS principle applied to intracellular metabolites (Mashego et al., 2003; Wu et al., 2005) using metabolite extracts labeled with of 2 different isotopes (¹³C and ¹⁵N). An important prerequisite is that one of the two labeled metabolite extracts should not contain either the metabolite which is formed or the one that contributes to the formation of the latter by degradation. We thus distinguish two cases based on whether the second isotopic dimension contains solely Y or X.

Case 1: The ¹³C-labeled extract contains both ¹³C-X and ¹³C-Y, the ¹⁵N-labeled extract contains only ¹⁵N-Y (see figure below)



For Y:

$$\frac{(Ay)^{12C}}{(Ay)^{13Cstd}} = \frac{(Ay)^{12Cmes}}{(Ay)^{13Cmes}} = \frac{(Ay)^{12Cmes}}{(Ay)^{15Nmes}} \tag{eq. 1.1}$$

where Ax is the peak area of metabolite X and Ay the peak area of metabolite Y, measured in the sample (mes) or in the standard solution spiked with ¹³C- and ¹⁵N-metabolite extracts (std). The isotopic dimensions of the peaks are shown as superscripts.

For X:

$$\frac{(Ax)^{12C}}{(Ax)^{13Cstd}} = \frac{(Ax)^{12Ccor}}{(Ax)^{13Ccor}} \tag{eq. 1.2}$$

where (Ax)^{12Ccor} and (Ax)^{13Ccor} refer to the area corrected for the contribution of Y decomposition into X applied either on unlabeled X or on ¹³C-X respectively.

In the unlabeled dimension, $(Ax)^{12C}cor$ is obtained by subtracting to $(Ax)^{12C}mes$ the contribution of Y decomposition into X by the following equation:

$$(Ax)^{12C}cor = (Ax)^{12C}mes - \alpha \times \pi \times (Ay)^{12C} \quad (\text{eq. 1.3})$$

where α is the proportion of mole of Y (ny) being decomposed into mole of X (nx) defined as $\alpha = \frac{nx}{ny}$ and, π the ratio between the response factor (RF) of X and that of Y as given above:

$$\pi = \frac{RFx}{RFy} \quad (\text{eq. 1.4})$$

RF is defined as follow:

$$RFx \text{ (or } y) = \frac{Ax \text{ (or } y)}{nx \text{ (or } y)} \quad (\text{eq. 1.5})$$

where nx (or y) is the number of mole of X or Y

Here, because the ^{15}N dimension contains only Y, $\alpha \cdot \pi$ can be determined in each sample, using the following equation:

$$\alpha \times \pi = \frac{(Ax)^{15N}mes}{(Ay)^{15N}std} \quad (\text{eq. 1.6})$$

where $(Ax)^{15N}mes$ is the peak area of X measured in the ^{15}N dimension - resulting exclusively from Y decomposition and, $(Ay)^{15N}std$ the peak area of ^{15}N -Y, measured in the standard solution spiked with the ^{15}N -metabolite extract.

Combining equation 1.3 with equations 1.1 and 1.6,

$$(Ax)^{12C}cor = (Ax)^{12C}mes - \frac{(Ax)^{15N}mes}{(Ay)^{15N}std} \times \frac{(Ay)^{13C}std}{(Ay)^{13C}mes} \times (Ay)^{12C}mes \quad (\text{eq. 1.7})$$

which can be rearranged as:

$$(Ax)^{12C}cor = (Ax)^{12C}mes - \frac{(Ax)^{15N}mes/(Ay)^{13C}mes}{(Ay)^{15N}std/(Ay)^{13C}std} \cdot (Ay)^{12C}mes \quad (\text{eq. 1.8})$$

This equation corresponds to the equation 1 in the manuscript.

From this equation, a formation factor f_i can be defined as:

$$f = \frac{(Ax)^{15N}mes/(Ay)^{13C}mes}{(Ay)^{15N}std/(Ay)^{13C}std} \quad (\text{eq. 1.9})$$

which is given in figure 3, panel C of the manuscript.

The formation factor can be applied for ^{13}C -X in order to calculate $(Ax)^{13C}cor$:

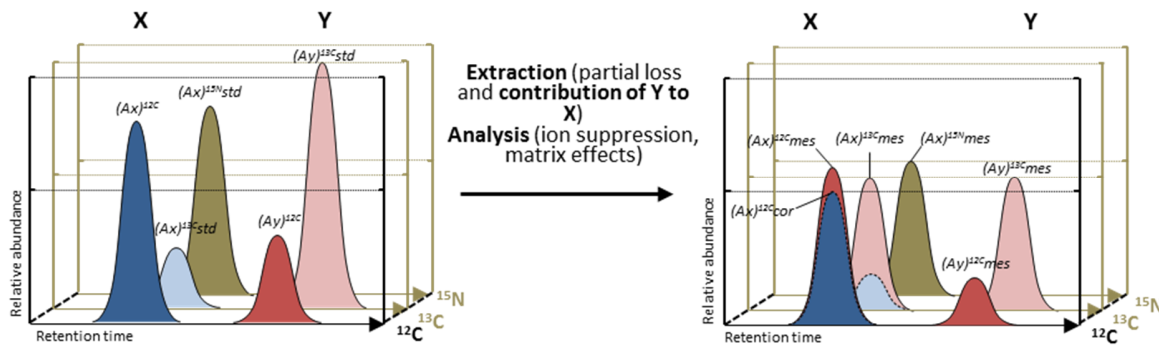
$$(Ax)^{13C} cor = (Ax)^{13C} mes - f \cdot (Ay)^{13C} mes \quad (\text{eq. 1.10})$$

Therefore, equation 1.2 can be rewritten as follow using equations 1.8 and 1.10:

$$\frac{(Ax)^{12C} cor}{(Ax)^{13C} cor} = \frac{(Ax)^{12C} mes - f \times (Ay)^{12C} mes}{(Ax)^{13C} mes - f \times (Ay)^{13C} mes} \quad (\text{eq. 1.11})$$

as given in figure 3, panel C of the manuscript.

Case n°2 : The ¹³C-labeled extract contains both ¹³C-X and ¹³C-Y, the ¹⁵N-labeled extract contains only ¹⁵N-X (see figure below)



For Y:

$$\frac{(Ay)^{12C}}{(Ay)^{13C} std} = \frac{(Ay)^{12C} mes}{(Ay)^{13C} mes} \quad (\text{eq. 2.1})$$

where Ax is the peak area of metabolite X and Ay the peak area of metabolite Y, measured in the sample (mes) or in the standard solution spiked with ¹³C- and ¹⁵N-metabolite extracts (std). The isotopic dimensions of the peaks are shown as superscripts.

and for X:

$$\frac{(Ax)^{12C}}{(Ax)^{15N} std} = \frac{(Ax)^{12C} cor}{(Ax)^{15N} mes} \quad (\text{eq. 2.2})$$

where $(Ax)^{12C} cor$ refers to the area corrected for the contribution of Y decomposition into X applied either on unlabeled X, $(Ax)^{15N} mes$ is the peak area of X measured in the ¹⁵N dimension - resulting exclusively from partial degradation of X, ion suppression, matrix effects, etc. and, $(Ay)^{15N} std$ the peak area of ¹⁵N-Y, measured in the standard solution spiked with the ¹⁵N-metabolite extract.

In the unlabeled dimension, $(Ax)^{12C} cor$ is obtained by subtracting to $(Ax)^{12C} mes$, the contribution of Y decomposition into X by the following equation:

$$(Ax)^{12C}cor = (Ax)^{12C}mes - \alpha \times \pi \times (Ay)^{12C} \quad (\text{eq. 2.3})$$

where α is the proportion of Y being decomposed into X and, π the ratio between the response factor (RF) of X and that of Y (see equations 1.4 and 1.5).

Similarly, in the ^{13}C -labeled dimension, $(Ax)^{13C}cor$ is obtained by subtracting to $(Ax)^{13C}mes$, the contribution of Y decomposition into X, by the following equation:

$$(Ax)^{13C}cor = (Ax)^{13C}mes - \alpha \times \pi \times (Ay)^{13C}sdt \quad (\text{eq. 2.4})$$

Because in this case, ^{15}N -labeled metabolite extract does not contain Y, the ^{15}N -X is only subjected to degradation and can thus also be used as an internal standard for ^{13}C -X, to correct for ion suppression, matrix effects, partial loss, etc. by the equation below:

$$\frac{(Ax)^{13C}sdt}{(Ax)^{15N}std} = \frac{(Ax)^{13C}cor}{(Ax)^{15N}mes} = \frac{(Ax)^{13C}mes - \alpha \times \pi \times (Ay)^{13C}sdt}{(Ax)^{15N}mes} \quad (\text{eq. 2.5})$$

This equation can be rearranged as follow:

$$\alpha \times \pi = \frac{(Ax)^{13C}mes}{(Ay)^{13C}sdt} - \frac{(Ax)^{15N}mes \times (Ax)^{13C}sdt}{(Ax)^{15N}std \times (Ay)^{13C}sdt} \quad (\text{eq. 2.6})$$

Combining equation 2.3 with equation 2.6, we obtained:

$$(Ax)^{12C}cor = (Ax)^{12C}mes - \left(\frac{(Ax)^{13C}mes}{(Ay)^{13C}sdt} - \frac{(Ax)^{15N}mes \times (Ax)^{13C}sdt}{(Ax)^{15N}std \times (Ay)^{13C}sdt} \right) \times (Ay)^{12C} \quad (\text{eq. 2.7})$$

$(Ay)^{12C}$ can be replaced in this equation using equation 2.1,

$$(Ax)^{12C}cor = (Ax)^{12C}mes - \left(\frac{(Ax)^{13C}mes}{(Ay)^{13C}sdt} - \frac{(Ax)^{15N}mes \times (Ax)^{13C}sdt}{(Ax)^{15N}std} \right) \times \frac{(Ay)^{12C}mes}{(Ay)^{13C}mes} \times (Ay)^{13C}sdt \quad (\text{eq. 2.8})$$

which can be rearranged as follow:

$$(Ax)^{12C}cor = (Ax)^{12C}mes - \left((Ax)^{13C}mes - \frac{(Ax)^{15N}mes \times (Ax)^{13C}sdt}{(Ax)^{15N}std} \right) \times \frac{(Ay)^{12C}mes}{(Ay)^{13C}mes}$$

$$(Ax)^{12C}cor = (Ax)^{12C}mes - \left(\frac{(Ax)^{13C}mes - \frac{(Ax)^{15N}mes \times (Ax)^{13C}sdt}{(Ax)^{15N}std}}{(Ay)^{13C}mes} \right) \times (Ay)^{12C}mes \quad (\text{eq. 2.9})$$

This equation corresponds to the equation 2 in the manuscript.

From this equation, a formation factor f can be defined as:

$$f = \frac{(Ax)^{13C}mes - \frac{(Ax)^{15N}mes \times (Ax)^{13C}sdt}{(Ax)^{15N}std}}{(Ay)^{13C}mes} \quad (\text{eq. 2.10})$$

which is given in figure 3, panel C of the manuscript.

Therefore, equation 2.9 can be rewritten as follow using equation 2.2 and 2.10:

$$\frac{(Ax)^{12C}_{cor}}{(Ax)^{15N}_{mes}} = \frac{(Ax)^{12C}_{mes} - f \times (Ay)^{12C}_{mes}}{(Ax)^{15N}_{mes}} \quad (\text{eq. 2.11})$$

as given in figure 3, panel C of the manuscript.

3.1. Conclusion :

Le travail présenté ci-dessus présente une nouvelle méthode permettant de quantifier de façon absolue le ppGpp et le pppGpp. Le développement de cette méthode a été finalisé en fin de thèse. Les quantifications effectuées avant la finalisation de la méthode sont réalisées de manière relative par rapport à la concentration en ppGpp avant induction ou de façon absolue par utilisation d'une approche IDMS à une dimension. Le pppGpp n'est pas détectable dans la souche *E. coli* BL21. De fait, la quantification relative du ppGpp ne présente pas de biais liée à la dégradation du pppGpp marqué ou non marqué, en ppGpp. En revanche, la méthode IDMS utilisée pour la quantification absolue du ppGpp contient du ¹³C-pppGpp. Ainsi, la quantification absolue du ¹²C-ppGpp est potentiellement sous-évaluée. La même quantité de ¹³C-pppGpp étant ajouté à chaque échantillon, le biais est néanmoins équivalent pour tous les échantillons.

Cet article présente une méthode basée sur la spectrométrie de masse, permettant de quantifier la concentration en pppGpp et en ppGpp dans des extraits de métabolites par une approche IDMS à deux dimensions. Cette approche repose sur la capacité des micro-organismes à se développer sur une seule source de carbone et une seule source d'azote, pour produire les différents extraits de métabolites marquées. Le développement de cette méthode peut être étendue aux cas où la cascade de dégradation des métabolites ne se limitent pas à un seul métabolite mais à plusieurs. Cela peut être réalisé soit par l'utilisation d'extraits de métabolites marqués à l'aide d'un seul isotope (*e.g.* le ¹³C) mais avec l'incorporation d'un nombre différent d'isotopes soit par l'utilisation d'autres isotopes en plus de ceux du C et de l'N (*e.g.* le deutérium).

Chapitre 4 : Etude des cinétiques de croissance et de production des protéines UL130 et Hla avec le procédé de référence.

Le premier objectif de ce travail est de mettre en place un procédé de production de référence à partir du procédé standard du département BRD de Sanofi Pasteur. Chaque modification apportée au procédé standard est analysée par l'intermédiaire des cinétiques de croissance bactérienne et de production de la protéine recombinante modèle UL130. Une fois le procédé de référence obtenu, celui-ci sera caractérisé en phases de croissance et de production de deux protéines recombinantes modèles (UL130 et Hla). Ces phases serviront dans le prochain chapitre à mettre en lien les analyses fonctionnelles du métabolisme avec les cinétiques de croissance et de production des protéines recombinantes.

La protéine UL130 est une protéine native du cytomégalovirus humain (HCMV) (Patrone et al. 2005). Chez Sanofi Pasteur, sa production atteint plusieurs grammes par litre et entraîne une forte charge métabolique conduisant à un arrêt rapide de la croissance d'*E.coli* BL21 après induction à l'IPTG. Sa synthèse conduit à la formation de corps d'inclusions, c'est-à-dire d'agrégats de protéines insolubles dans le cytosol. La protéine UL130 est utilisée comme première protéine modèle d'étude du fait de l'importante charge métabolique observée durant sa production en milieu riche puis est comparée au procédé de production de la protéine Hla, une protéine native de *Staphylococcus aureus*. Chez Sanofi Pasteur, cette protéine est également produite en forte quantité mais provoque une plus faible charge métabolique pour *E. coli* BL21 en milieu riche permettant la prolongation de la croissance en condition de production. Hla est produite majoritairement de façon soluble avec une plus faible proportion de protéines insolubles. Si les différences en termes de charges métaboliques provoquées par la production des deux protéines recombinantes sont retrouvées dans le cadre du procédé de référence de la thèse, une comparaison des analyses fonctionnelles du métabolisme de la souche *E. coli* BL21 produisant chacune des protéines permettra de mettre en évidence les déterminants métaboliques à l'origine de l'arrêt de croissance.

Tout d'abord, les méthodes d'analyses mises en place pour la quantification des cinétiques de croissance bactérienne et de production sont étudiées. L'étude des biais associés à ces méthodes

permet la mise en place des moyens de correction et lorsque cela est impossible, de fixer des limites dans l'interprétation des données obtenues.

Le procédé standard du département BRD de Sanofi Pasteur est modifié pour tendre vers un double objectif. Le premier est de garantir que les résultats générés dans un contexte de laboratoire soient transférables pour une application industrielle. Le deuxième objectif est de réaliser des cultures de la souche *E. coli* BL21 en condition de production qui permettent la mise en place d'analyses fonctionnelles de son métabolisme (métabolomique, fluxomique...). Le procédé standard du département BRD est un procédé batch en milieu riche dont la production de protéines recombinantes est déclenchée par ajout de l'inducteur (IPTG) en milieu de phase exponentielle de croissance (induction tardive). La première modification consiste à avancer le moment de l'induction en début de phase exponentielle de croissance (induction précoce). Une induction précoce de la synthèse de la protéine recombinante permet d'augmenter le temps d'observation du procédé dans le cas où la croissance est maintenue en condition de production. Dans un deuxième temps, le milieu riche est remplacé par un milieu minimum chimiquement défini. Cette substitution de milieu est nécessaire à la mise en place des analyses fonctionnelles du métabolisme en condition de production des protéines recombinantes. En effet, ce type d'analyse serait difficile à effectuer en présence d'une complexité biologique amené par l'utilisation de milieu riche. Au niveau industriel ce changement de milieu permet d'améliorer la maîtrise de la qualité des composants du milieu. La complexité de la composition des milieux riches explique souvent une variabilité de la quantité et de la qualité des composants du milieu entre les lots, ce qui impact la reproductibilité des procédés de production. De plus, le milieu minimum servira de base d'optimisation à la formulation du milieu en fonction des contraintes métaboliques détectées en condition de production. Ce type de milieu permet également le développement de procédés semi-continus ou continus avec une croissance limitée par l'alimentation en substrat carboné. Cependant, les bactéries cultivées en milieu minimum sont contraintes de synthétiser l'ensemble de leur biomasse à partir du substrat carboné plutôt que d'utiliser directement des métabolites contenus dans un milieu complexe (ex : acides aminés). De fait, un taux de croissance inférieur est attendu et une charge métabolique plus importante devrait être engendrée par la production de protéines recombinantes. L'influence de la substitution du milieu sur

les cinétiques de croissance de la souche *E. coli* BL21 et de production de la protéine UL130 est ensuite étudiée.

Une fois ces optimisations faites, les procédés de production des protéines modèles UL130 et Hla sont étudiés en utilisant le procédé de référence finalisé. Les cinétiques de croissance bactérienne et de production obtenues sont comparées afin de déterminer les phases caractéristiques de chaque procédé.

Dans cette thèse, le gène de résistance à la kanamycine est conservé mais aucun antibiotique n'est utilisé durant les cultures. L'antibiotique est éliminé en prévision d'une utilisation du procédé à une échelle industrielle. En effet, l'absence d'antibiotique est un impératif réglementaire dans un contexte de production de protéines d'intérêt vaccinale. De plus, la résistance à un antibiotique est une charge métabolique supplémentaire pour la bactérie (Andersson & Levin 1999). L'absence d'antibiotique permet de diminuer la charge métabolique provoquée par la mise en place du mécanisme de résistance à l'antibiotique.

4.1. Méthodes d'analyse des cinétiques de croissance et de production.

4.1.1. Détermination des biais d'estimation de la concentration en biomasse par turbidimétrie en condition de production

La quantification de la biomasse par la densité optique est basée sur la mesure de la turbidimétrie du milieu de culture appelée densité optique (DO). Dans des conditions dynamiques de croissance et de production, la taille des cellules peut évoluer lors du changement du taux de croissance (Bremer & Dennis 2008). De plus une déformation morphologique pourrait-être associée à la production. Ces modifications de la taille des bactéries au cours du procédé conduiraient à une modification du nombre de cellules par unité de DO mesurée. La quantification précise de la concentration en biomasse est une donnée importante pour caractériser l'adaptation bactérienne à la production. Afin de déterminer l'influence de la production sur la mesure de la DO, une mesure des

poids secs et une quantification du nombre de cellules par cytométrie ont été réalisées durant la production de la protéine UL130 en milieu minimum.

La quantification du nombre de cellules totales est rapportée aux mesures de DO effectuées durant le procédé de production. Ce rapport permet d'obtenir l'évolution du nombre de cellules par unité de DO au cours de la production. La mesure des poids secs est également utilisée afin de mesurer la quantité de biomasse sèche (cell dry weight, CDW) par unité de DO (uDO) au cours de la production. Avec cette méthode, la biomasse pesée sèche est composée de la biomasse bactérienne et de protéine recombinante en condition de production (figure 4.1).

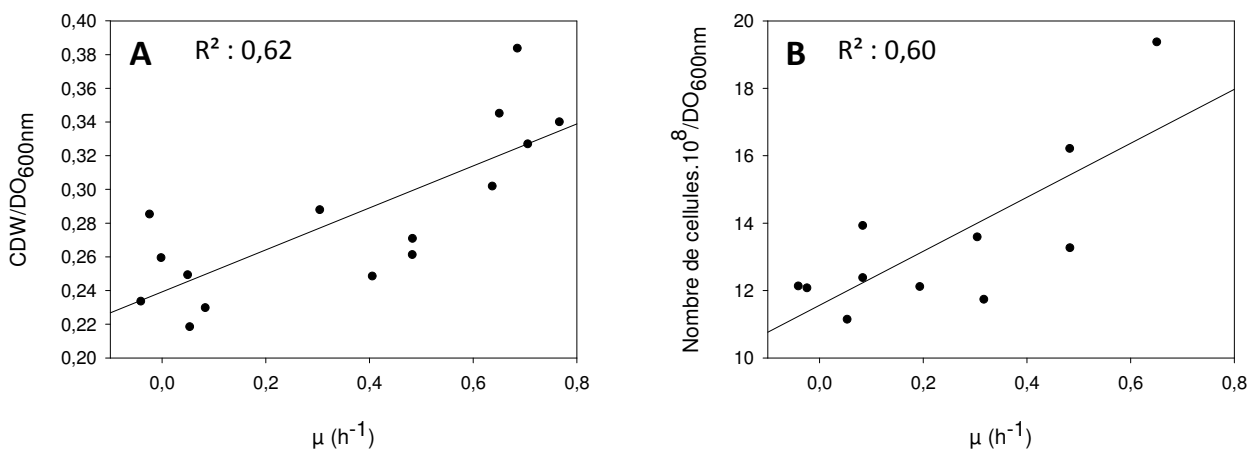


Figure 4.1 : Corrélation entre le poids sec par unité de DO_{600nm} (CDW (g) / DO_{600nm}) et le taux de croissance (μ , h⁻¹) (A) ou le nombre de cellules par unité de DO_{600nm} (nombre de cellules · 10⁸ / DO_{600nm}) et le taux de croissance (B). Les pesées sur culots (poids sec mesuré par pesée de culots bactériens secs) et la cytométrie (nombre de cellules par mL mesuré par cytométrie) sont chacune réalisées sur une culture différente de la souche *E. coli* BL21 en condition de production de la protéine UL130 en milieu minimum.

Des corrélations entre la biomasse sèche par unité de DO ou le nombre de cellules par unité de DO en fonction du taux de croissance sont obtenues au cours du procédé de production de la protéine UL130 par la souche *E. coli* BL21 en milieu minimum et en condition d'induction précoce (figure 4.1). La première corrélation (figure 4.1, A) indique qu'en condition de production, plus le taux de croissance est faible et plus la quantité de biomasse sèche par unité de DO est faible. De façon similaire la deuxième corrélation (figure 4.1, B) indique que plus le taux de croissance est faible, plus le nombre de

bactérie par unité de DO est faible. Le nombre de cellules ou la quantité de biomasse sèche par unité de DO diminue de 37 % et 33 % respectivement lors du passage d'un taux de croissance de $0,74 \text{ h}^{-1}$ à 0 h^{-1} (figure 4.1).

Ces deux mesures supplémentaires à la turbidimétrie semblent indiquer que la taille des cellules augmente d'environ 35 % lorsque le taux de croissance diminue de $0,74 \text{ h}^{-1}$ à 0 h^{-1} . Ces résultats sont inverses aux données présentées par Bremer & Dennis, 2008 qui indique, par extrapolation, que la taille des bactéries peut être divisée par un facteur 2 lors d'un passage d'un taux de croissance de $0,6 \text{ h}^{-1}$ à 0 h^{-1} . De fait, la production de la protéine UL130 ne semble pas augmenter le biais sur la mesure de la turbidimétrie engendrée par la modification de la taille des cellules lors de l'évolution du taux de croissance.

En plus du comptage cellulaire, l'analyse cytométrique a permis de séparer la quantité de cellules suivant leur viabilité après induction par coloration des cellules viables ou non viables spécifiquement (figure 4.2).

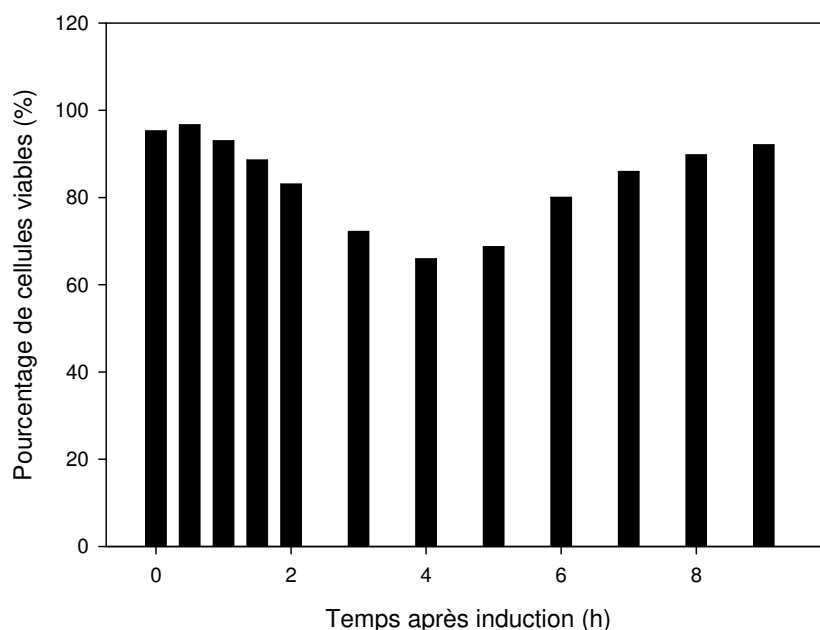


Figure 4.2 : Mesure par comptage du pourcentage de cellules viables par cytométrie après coloration au cours du procédé de production de la protéine UL130 en milieu minimum (induction précoce).

Cette analyse permet de mettre en évidence le fort impact de la production sur la viabilité cellulaire (figure 4.2). Une baisse de 34 % de la viabilité est observée entre 1 h et 4 h suivant l'induction.

La perte de viabilité observée met en évidence l'impact important de la production sur la physiologie cellulaire. La production de la protéine UL130 entraîne une perturbation majeure du fonctionnement cellulaire.

4.1.2. Quantification de la protéine UL130

Dans le but de pouvoir estimer le plus précisément possible la quantité de protéines produites, une gamme de calibration de la protéine UL130 purifiée a été analysée après dépôt et coloration sur gel SDS-PAGE. La quantification par densitométrie des bandes colorées sur gel SDS-PAGE est une méthode d'analyse semi-quantitative et peut-être adaptée à l'analyse des cinétiques de production lorsque les biais inhérents à la méthode sont corrigés (Marisch, Bayer, Scharl, et al. 2013; Bruschi et al. 2014; Ali et al. 2015).

Tout d'abord les protéines d'une quantité de biomasse fixe et définie par lecture de la DO sont séparées par migration sur gel SDS-PAGE. Au moment de l'arrêt de croissance des bactéries, la quantité de biomasse déposée peut être sous-estimée. Ce biais peut être corrigé par une normalisation de la bande d'intérêt par mesure de l'intensité des bandes des protéines de la biomasse dans chaque piste du gel SDS-PAGE. De plus, les quantifications réalisées par cette méthode sont soumises à une variabilité entre les gels. Cette variabilité est due à une différence de coloration ou de décoloration des gels, des différences de temps de migration ou de qualité de gels utilisés. Ce biais sera éliminé dans l'ensemble des quantifications réalisées par la normalisation de la mesure de la densitométrie de la bande d'intérêt sur la mesure de la densitométrie de la bande à 25 kDa du marqueur de taille déposée en quantité fixe sur tous les gels.

La gamme a été réalisée par un ajout de standard de la protéine UL130 purifiée et quantifié par ELISA, dans des échantillons de lysat cellulaire sans protéine recombinante afin de prendre en compte l'effet associé à une superposition de bandes de protéines de la biomasse avec la bande de la protéine

UL130 après migration (Bruschi et al. 2014). De plus, la quantité de protéine UL130 déposée a été déterminée en fonction de la quantité de biomasse sèche (CDW) déposée afin d'avoir une mesure d'enrichissement (gUL130/gCDW) et ainsi de faciliter l'évaluation de la production spécifique de la protéine UL130 réalisées au cours des procédés (figure 4.3).

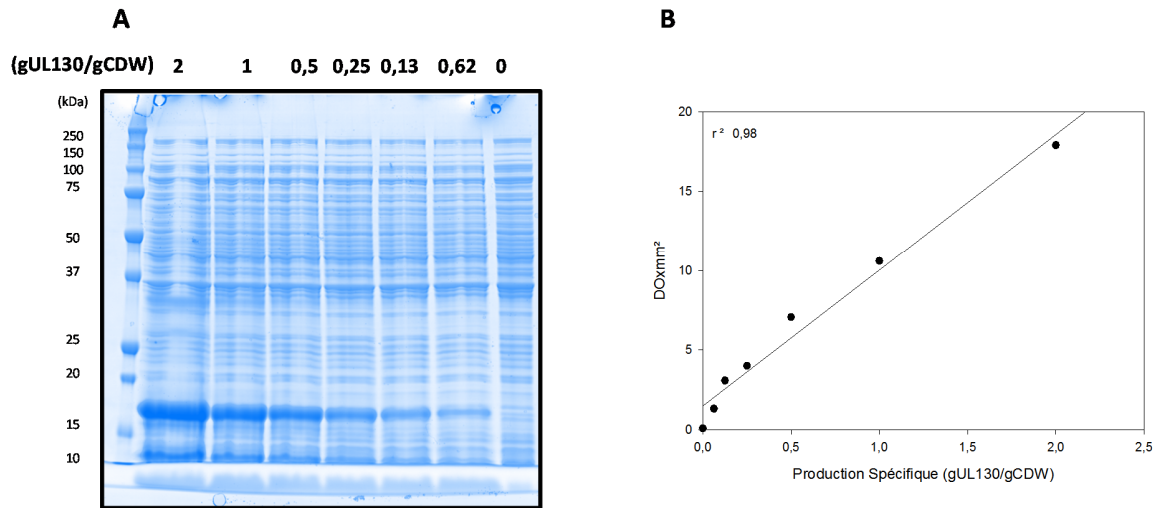


Figure 4.3 : Réalisation d'une droite de calibration de la quantité d'UL130 par quantité de biomasse (gUL130/gCDW) en fonction de la densitométrie (DO/mm²) sur un gel SDS-PAGE. (A) Gel SDS-PAGE après migration et coloration au bleu de Coomassie des échantillons de biomasse avec ajout d'une gamme de dilution de la protéine UL130 purifiée et quantifiée par ELISA et ajoutée à du lysat bactérien sans protéine recombinante. (B) Droite de calibration obtenus après quantification densitométrique (DO/mm²) des bandes formées par le standard de la protéine UL130 sur le gel SDS-PAGE en fonction de la quantité de protéine déposée (gUL130/gCDW).

Une droite de calibration a été obtenue entre la quantité de protéines déposées (gUL130/gCDW) et la mesure densitométrique des bandes (DO/mm²) de la protéine UL130 à 18,4 kDa après migration (densité optique de la bande/mm²) (figure 4.3, A et B).

Cette méthode de quantification des protéines recombinantes permet de suivre de façon relative les cinétiques de productions au cours des procédés en éliminant les biais associées à cette technique (quantité de biomasse déposée et coloration des protéines) et d'estimer les quantités absolues de protéines recombinantes produites grâce à la droite de calibration.

Au cours des procédés de production, la quantification de la production spécifique (quantité de protéine recombinante par unité de biomasse) est effectuée par dépôt d'une quantité fixe de biomasse

sur un gel d'acrylamide. La production spécifique dépend de la vitesse de production spécifique de la protéine recombinante par rapport au taux de croissance (μ) des bactéries. La vitesse de production spécifique de la protéine représente le flux de synthèse de protéines recombinantes par unité de biomasse (g/gCDW/h) et peut-être calculée par l'équation suivante :

$$q_{PS} = \frac{dPS}{dt} + \mu \times PS$$

Dans cette équation q_{PS} est la vitesse de production spécifique, dPS/dt est la dérivé de la production spécifique en fonction du temps (g/gCDW/h), μ est le taux de croissance (h^{-1}) et PS la production spécifique (g/gCDW).

Une vitesse de production spécifique plus élevée que le μ entraîne une augmentation de la production spécifique. Inversement une vitesse de production spécifique plus faible que le μ entraîne une diminution de la production spécifique. Les productions totales en protéine recombinante des cultures sont calculées à partir de la production spécifique rapportée à la quantité de biomasse mesurée au moment du prélèvement de l'échantillon. Les quantités absolues (g/L) de protéines recombinantes sont calculées à partir de la droite de calibration présentée ci-dessus (figure 4.3, B).

4.2. Modification du procédé standard vers le procédé de référence avec la protéine UL130

4.2.1. Etude du procédé standard du département BRD Sanofi Pasteur

Le procédé standard de production du département BRD de Sanofi Pasteur est étudié dans le cadre de la production de la protéine UL130. Cette étude a pour objectif de définir les cinétiques de croissance bactérienne et de production du procédé tel que défini au démarrage de la thèse. Ces cinétiques permettront, par la suite, d'analyser l'influence de toutes les modifications liées à la mise en place du procédé de référence de la thèse.

Pour cette étude une culture non induite est comparée à une culture induite de la souche *E. coli* BL21 en milieu riche. Le suivi de la croissance est effectué par mesure de la turbidimétrie (DO_{600nm})

et les productions spécifiques et totales sont effectuées par lectures densitométriques des bandes formées par la protéine recombinantes après migration et coloration au bleu de coomassie sur un gel SDS-PAGE (figure 4.4).

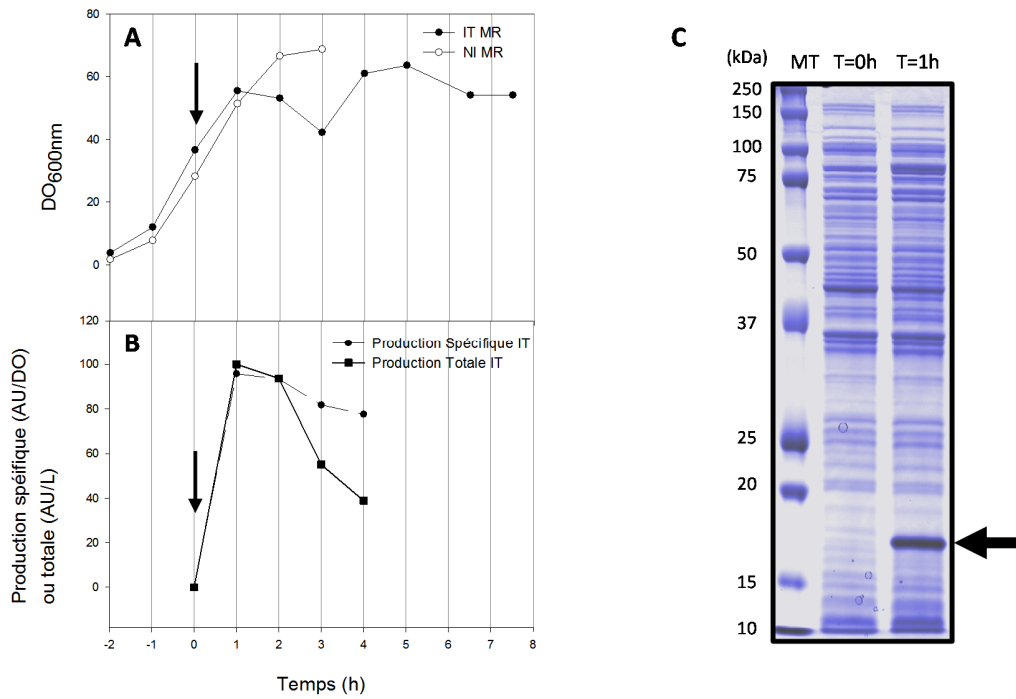


Figure 4.4 : Comparaison des cinétiques de croissance bactérienne et de production de la protéine UL130 des cultures de la souche *E. coli* BL21 en milieu riche avec induction tardive (IT) ou sans induction (NI). (A) Biomasse (DO) au cours d’une culture induite tardivement (rond noir) et d’une culture non induite (rond blanc) en milieu riche (MR). (B) La production spécifique (unité arbitraire (AU)/DO) de la protéine UL130 dans la culture IT (rond noir, pointillés) représente la quantité de protéines par unité de biomasse au cours de la culture et la production totale est la quantité de protéine par litre de milieu de culture au cours du temps (carré noir, trait). (C) Biomasse déposée à quantité de DO équivalente et colorée au bleu de Coomassie sur un gel SDS-PAGE à T=0h et T=1h avec un marqueur de taille (MT) (la flèche bleue signale la bande formée par la protéine UL130). L’échelle de temps en normalisé par rapport au moment de l’ajout de l’IPTG à 1mM (T=0h) pour déclencher le début de la production (induction, flèche noire).

L’utilisation d’un milieu riche permet à la souche *E. coli* BL21 de passer de 0,8 à 70 unité de DO (uDO) (soit 0,27 à 23,8 gCDW/L) en 4 heures de culture ($\mu = 1,3 \text{ h}^{-1}$ en phase exponentielle de croissance) en absence d’induction (figure 4.4, A). Quand l’IPTG est ajouté à 35 uDO pour la condition induite, le système d’expression T7 est activé, déclenchant l’expression du vecteur et la synthèse de la protéine UL130. Une heure après l’ajout d’IPTG, la diminution de la DO indique un arrêt de croissance

pendant 2 heures (figure, A). La croissance reprend ensuite pendant 1 heure jusqu'à 64 uDO (soit environ 22 gCDW/L). La mesure de la production spécifique révèle que la protéine UL130 est produite en très grande quantité pendant la 1^{ère} heure après induction (figure 4.4, panel B). Une estimation de la production totale à partir de la mesure densitométrique de la bande formée par la protéine UL130 colorée présente à 18,4 kDa est comprise entre 5 à 10 g/L de protéines UL130 (figure 4.4, C). Après cette 1^{ère} heure, c'est-à-dire pendant l'arrêt de croissance bactérienne, la production spécifique diminue légèrement. Pendant cette période, la production totale diminue également jusqu'à la fin du procédé révélant un arrêt ou une très faible production de la protéine UL130. Durant ce procédé, la protéine UL130 est présente sous forme de corps d'inclusion dans la bactérie.

L'arrêt de croissance observé après 1 heure d'induction est concomitant à la production élevée de protéines UL130. Il semble donc que la vitesse élevée de production d'UL130 entraîne l'arrêt de croissance après une heure d'induction. Durant la phase de reprise de croissance, il semble que l'augmentation de la biomasse ne soit pas accompagnée d'une production significative. Les bactéries, en fin de culture, ont été isolées sur boîtes de Pétri sans antibiotique puis réensemencées sur une gélose enrichie en kanamycine. Toutes les colonies réensemencées étaient résistantes à la kanamycine. L'arrêt de production n'est donc pas provoqué par une perte complète du nombre de copies de plasmide par bactérie.

Ce procédé est caractérisé par une production d'une heure conduisant à un arrêt de croissance de la souche. Dans ce contexte, l'induction tardive est en effet la meilleure stratégie pour optimiser la quantité de protéines produite. Cette thèse a pour objectif final d'optimiser le procédé et le système d'expression dans le but d'allonger la phase de croissance en condition de production. Afin d'évaluer les optimisations sur la base de ce critère dans le procédé de référence en batch, il est nécessaire d'avancer le moment de l'induction de la production à une phase de croissance correspondant au début de la phase exponentielle.

4.2.2. Induction précoce de la production de la protéine UL130 en milieu riche

Dans le but d'augmenter le temps d'observation entre l'induction et la consommation totale du glucose, la DO cible pour l'ajout d'IPTG est fixée entre 3,5 et 5 uDO. L'influence de cette modification

est étudiée par la mesure des cinétiques de croissance bactérienne et de production de la protéine UL130 en milieu riche, l'objectif étant de retrouver des cinétiques de croissances et de production relativement proches du procédé standard de production du département BRD de Sanofi Pasteur (figure 4.5).

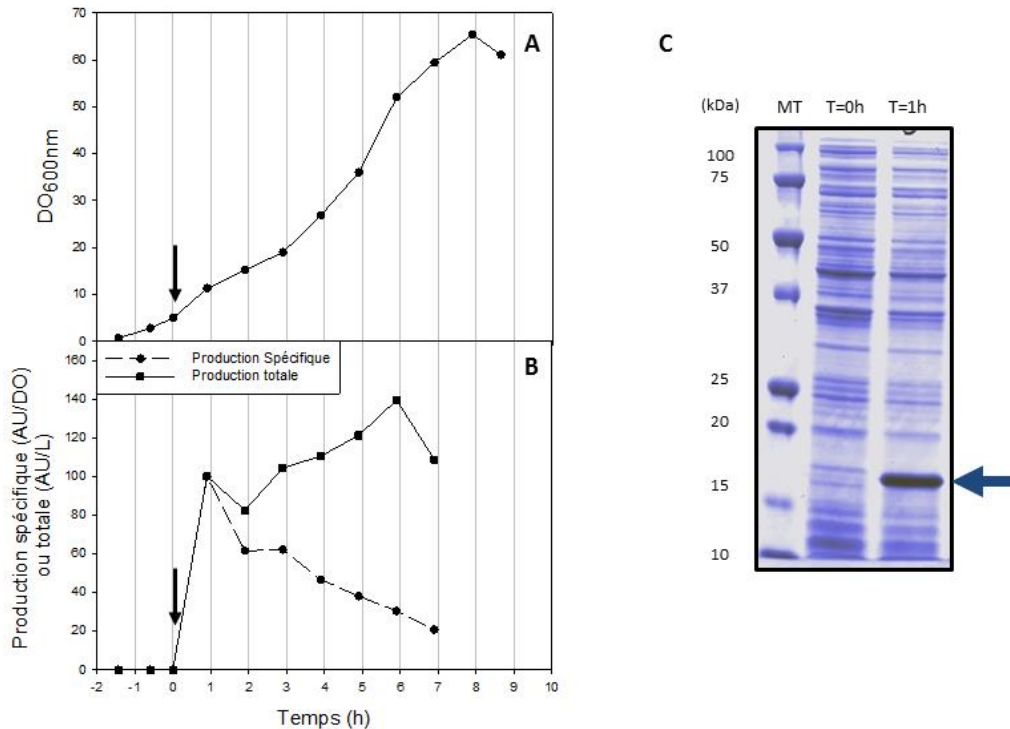


Figure 4.5 : Evolution de la croissance bactérienne et de la production de la protéine UL130 en milieu riche lors d'une induction précoce de la production. (A) Biomasse (DO) (B) production spécifique de la protéine UL130 (AU/DO) (point noir, pointillés) et production totale de la protéine UL130 (AU/L) (carré noir, trait plein) au cours du temps. (C) Biomasse déposée à quantité de DO équivalente et colorée au bleu de Coomassie sur un gel SDS-PAGE à T=0h et T=1h avec un marqueur de taille (MT) (la flèche bleue signale la bande formée par la protéine UL130). L'induction est réalisée à t=0h (flèche noire) par ajout d'IPTG 1mM.

Après inoculation des bactéries à 0,7 uDO, l'induction de la production de la protéine UL130 est réalisée après 1h20 de culture, à l'atteinte de 5 uDO (figure 4.5, A). Suite à l'induction, la croissance semble moins inhibée par la production que lors de l'induction tardive. La phase exponentielle de croissance est infléchiée après induction mais la croissance se poursuit jusqu'à 65 uDO. Une production spécifique maximale est atteinte dès la première heure d'induction (figure 4.5, B et C). La production

spécifique diminue ensuite rapidement. Après la première heure d'induction, la production totale n'évolue que faiblement pouvant révéler un arrêt ou une faible production après la première heure d'induction (figure 4.5, B). La résistance des bactéries à la kanamycine lors de leur étalement sur gélose a permis de démontrer que cette perte de production n'est pas due à une perte de plasmide.

Le passage d'une induction tardive à une induction précoce n'a pas drastiquement changé les cinétiques de croissance bactérienne et de production de la protéine UL130. Le suivi de la DO révèle que la phase d'arrêt de croissance engendrée par la production n'a pas été retrouvée au cours de ce procédé. Comme lors de l'induction tardive, la production s'est déroulée durant la première heure après induction puis est devenue très faible voire nulle. Le profil global de croissance et de production est conservé permettant de répondre à l'objectif de cette étude. La dernière étape consiste maintenant à utiliser un milieu minimum chimiquement défini pour la production.

4.2.3. Formulation du milieu minimum chimiquement défini

La dernière modification apportée au procédé standard de production consiste à remplacer le milieu riche par un milieu minimum chimiquement défini. Cette modification du procédé permet de garantir la faisabilité des analyses fonctionnelles du métabolisme et pourra servir de base au développement d'un procédé semi-continu. Les contraintes inhérentes à l'atteinte de ces objectifs sont, i) la présence d'une source unique de carbone dans le milieu (ici le glucose), ii) la formulation d'un milieu n'entraînant aucune limitation en absence de production et permettant ainsi d'atteindre un μ équivalent au μ sur glucose décrit dans la littérature avec la souche *E. coli* BL21 ($\mu = 0,7 \text{ h}^{-1}$) (Marisch, Bayer, Scharl, et al. 2013), iii) la formulation d'un milieu chimiquement défini compatible avec des analyses effectuées par RMN ou par spectrométrie de masse (faibles concentrations en sels, bonne stabilité des métabolites produits dans le surnageant...). Succinctement, la formulation du milieu est basée sur le milieu chimiquement défini Korz (Korz et al. 1995). Tous les éléments carbonés autres que le glucose (acétate de zinc, citrate de fer, acide citrique, thiamine-HCl) ont été éliminés ou remplacés par leur analogue non carboné (sulfate de zinc, sulfate de fer). Le dihydrogénophosphate de potassium (13 g/L) est remplacé par du chlorure de potassium (0,7 g/L), seule source de potassium du milieu. Ce milieu a permis la culture de la souche *E. coli* BL21 à $\mu = 0,74$

h^{-1} jusqu'à DO 70 (soit 26 g/L de biomasse sèche) avec 50 g/L de glucose. Ce milieu a donc pu être utilisé pour la production de protéines recombinantes en batch.

4.2.4. Procédé de production de la protéine UL130 en milieu minimum

L'utilisation d'un milieu minimum plutôt qu'un milieu riche pour la culture d'*E. coli* change profondément l'état du métabolisme bactérien et de la composition macro-chimique des cellules (Traxler, Summers, H. Nguyen, et al. 2008; Bremer & Dennis 2008). Ce changement de milieu a pour conséquence majeure de baisser le taux de croissance, de diminuer le nombre de ribosomes par bactérie et de conduire à l'expression des voies de biosynthèse nécessaires à la synthèse de tous les composants de la biomasse à partir de la consommation du glucose et de l'azote minérale. En particulier, les acides aminés sont synthétisés à partir du glucose et de l'azote du milieu minimum plutôt que d'être assimilés à partir d'un milieu riche. Il est donc attendu que cette modification du milieu entraîne une diminution de la disponibilité des acides aminés et donc une augmentation de la charge métabolique associée à la synthèse de la protéine recombinante. Cette partie de l'étude a pour objectif d'étudier l'influence de la modification du milieu de culture sur le procédé de production de la protéine UL130. Pour cela les cinétiques de croissance bactérienne et de production de la protéine UL130 sont analysées (figure 4.6).

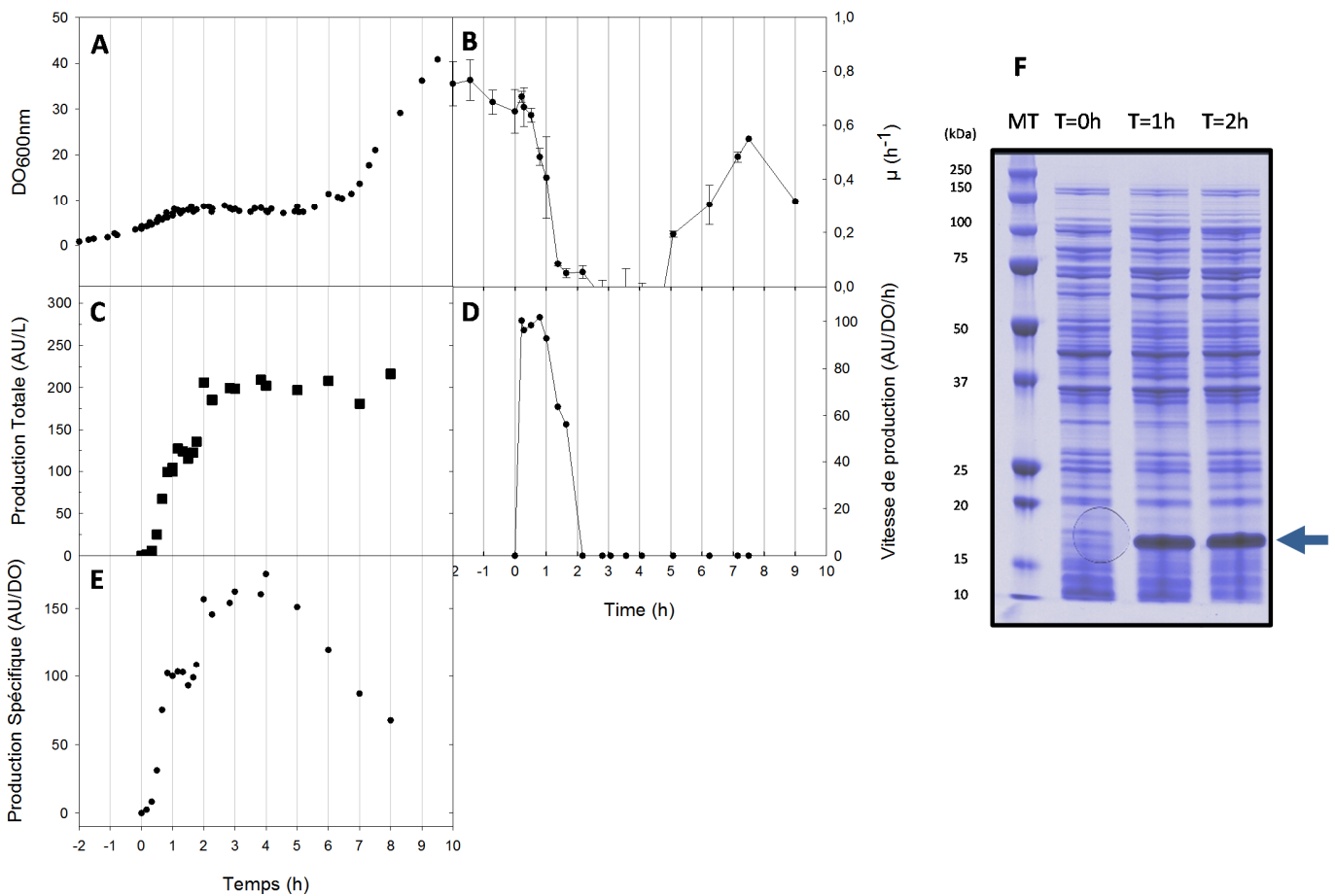


Figure 4.6 : Etude de l'influence de l'utilisation d'un milieu minimum pour la production de la protéine UL130 en induction précoce. (A) Biomasse (DO), (B) taux de croissance (h^{-1}), (C) production totale (AU/L), (D) vitesse de production spécifique (AU/DO/h), et (E) production spécifique (AU/DO) au cours du temps après induction à T=0h par ajout d'IPTG 1mM. (F) Biomasse déposée à quantité de DO équivalente et colorée au bleu de Coomassie sur un gel SDS-PAGE à T=0h, T=1h et T=2h avec un marqueur de taille (MT) (la flèche bleue signale la bande formée par la protéine UL130).

Une diminution rapide du taux de croissance est observée dès l'ajout d'IPTG à la culture de la souche *E. coli* BL21 en milieu minimum (T=0h, figure 4.6, A et B). Un arrêt de croissance apparaît 1 à 2 heures après induction. Une reprise de la croissance est observée 5 heures après induction.

L'analyse par densitométrie des protéines colorées au bleu de Coomassie après séparation sur SDS-PAGE révèle une forte production de la protéine UL130 dès la 1^{ère} heure d'induction avec l'apparition d'une bande intense sur le gel (à 18,4 kDa) (figure 4.6, F). Les productions spécifiques et totales augmentent durant 2 heures après l'induction (figure 4.6, C et E). La production totale peut être

estimée entre 0,25 et 0,5 gUL130/gCDW avec la droite de calibration effectuée avec la protéine purifiée (figure 4.3). De façon surprenante, la production spécifique devient 60% plus élevée en milieu minimum qu'en milieu riche par comparaison avec l'échantillon de référence commun aux gels (figure 4.6, E). Cette augmentation de la production spécifique engendre une production totale supérieure en moyenne à la culture en milieu riche (figure 4.6, C). Cependant, comme en milieu riche, une diminution de la production spécifique apparaît lors de la reprise de croissance, conduisant à une production totale constante 2 heures après d'induction. Les bactéries, en fin de culture, ont été isolées sur boîtes de Pétri puis réensemencées sur une gélose enrichie en kanamycine. L'ensemble des clones réensemencés était résistants démontrant la présence du plasmide dans chaque bactérie en fin de procédé.

Le calcul de la vitesse de production spécifique permet de définir quatre phases dans le procédé. La première phase se caractérise par une vitesse de production spécifique très forte de T=0h à T=1h (figure 4.6, D). Cette période de forte production est concomitante à la baisse rapide du taux de croissance. La seconde phase se caractérise par une diminution de la vitesse de production spécifique de T=1h à T=2h. Pendant cette période, la croissance est nulle. La troisième phase est caractérisée par l'absence de production et de croissance de T=2h à T=5h. Une quatrième phase se caractérise par le retour de croissance avec une vitesse de production faible ou nulle (T=5h à T=9h).

L'arrêt de croissance lorsque la vitesse de production de la protéine UL130 est maximale indique que cette vitesse est trop élevée pour être maintenue par la souche *E. coli* BL21. L'objectif de cette thèse est d'étudier les conséquences de cette forte vitesse de production sur le métabolisme. Cette étude sera portée sur les phases du procédé durant lesquelles la production des protéines recombinantes est présente. Le manque de précision de la quantification de la protéine ne permet pas de déterminer avec certitude si le retour de croissance est accompagné d'une production de la protéine UL130. Il est essentiel de définir les phases de production de la protéine afin de focaliser les analyses fonctionnelles du métabolisme sur les périodes de production de la protéine recombinante.

4.2.5. Marquage isotopique de la protéine d'intérêt

Dans le but de délimiter clairement la période de production, une expérience de marquage isotopique de la protéine UL130 a été réalisée au cours de la reprise de croissance.

Du glucose entièrement marqué au carbone 13 (glucose ^{13}C) est ajouté au moment de la reprise de croissance et un prélèvement est réalisé trois heures après cet ajout. Ainsi, durant trois heures, toute protéine nouvellement synthétisée sera marquée en ^{13}C dans la même proportion que le ratio glucose ^{13}C /glucose ^{12}C dans le milieu (70 % de glucose ^{13}C). La proportion de marquage de la protéine d'intérêt et des protéines de la biomasse représente donc la quantité de protéines synthétisées après ajout du glucose marqué (à 70%) par rapport à la quantité de protéines synthétisées avant l'ajout.

La première étape a été de mettre au point la purification de la protéine UL130. La protéine UL130 est produite en corps d'inclusion. La purification commence par la séparation des protéines solubles et insolubles après lyse cellulaire. Les protéines solubles sont utilisées pour représenter les protéines de la biomasse car ne contiennent pas de trace de protéines recombinantes. La fraction insoluble, contenant les corps d'inclusions formés majoritairement par la protéine UL130, est lavée puis solubilisée avec des solutions d'urée et de DTT concentrées (Palmer & Wingfield 2004). Une colonne de cobalt est utilisée pour la purification de la protéine fusionnée avec le marqueur 6xhistidine.

Avant analyse, les protéines de la biomasse ou la protéine recombinantes sont hydrolysées par une solution HCl concentrée afin de rompre les liaisons peptidique et ainsi de libérer les acides aminés protéinogéniques. L'analyse de l'enrichissement en carbone 13 des acides aminés protéinogéniques est réalisée par LC-MS. Ce type couplage permet de détecter les acides aminés avec une grande sensibilité. La mesure de l'enrichissement en ^{13}C étant une proportion d'acides aminés marqués au ^{13}C parmi l'ensemble des acides aminés détectés, elle ne sera pas impactée par le rendement de purification.

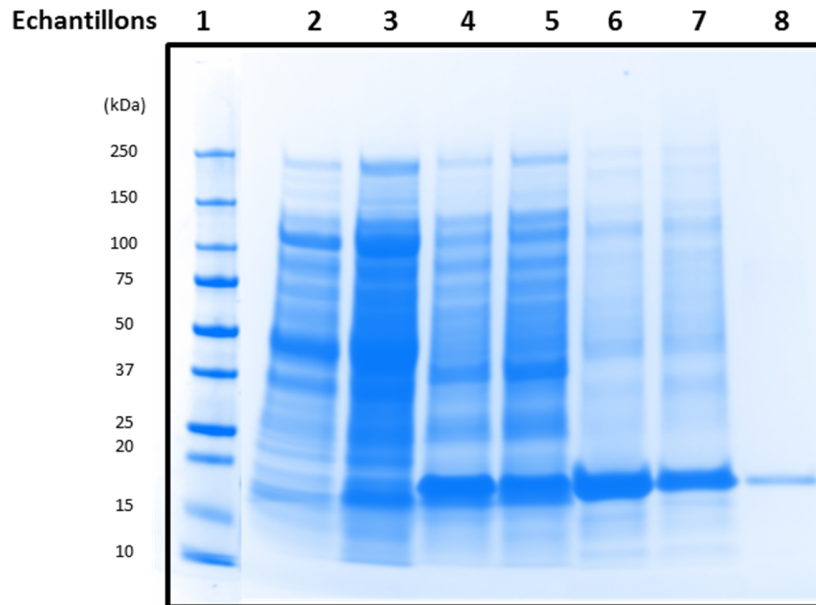


Figure 4.7 : Gel SDS-PAGE des échantillons déposés au cours du procédé de purification de la protéine UL130 (18,4 kDa). Colonne 1 : marqueur de taille, colonne 2 : protéines solubles de la biomasse, colonne 3 : protéines solubles de la biomasse (quantité déposée x2), colonne 4 : protéines insolubles (biomasse et protéine UL130), colonne 5 : protéines insolubles (quantité déposée x2), colonne 6 : corps d'inclusions de la fraction insoluble après sonication et lavage, colonne 7 : corps d'inclusions de la fraction insoluble après sonication et lavage une deuxième fois, colonne 8 : Protéine UL130 après purification.

Le gel SDS-PAGE révèle que la protéine recombinante est absente de la fraction soluble (figure 4.7, puits 1). Ainsi les protéines de la biomasse ne sont pas contaminées par la protéine recombinante (figure 4.7, puits 2 et 3 respectivement). Après lavage des corps d'inclusion (figure 4.7, puits 4, 5, 6 et 7) la protéine est purifiée (figure 4.7, puits 8). L'échantillon d'éluât de la purification de la protéine UL130 montre que la protéine est purifiée sans contamination apparente.

Les enrichissements obtenus pour l'ensemble des acides aminés ont été moyennés pour les protéines solubles de la biomasse ou la protéine UL130. Après 3 heures de reprise de croissance, les acides aminés protéinogéniques de la biomasse sont marqué en C^{13} à 55 % ($\pm 1,8$ %) et les acides aminés de la protéine UL130 à 2,8 % ($\pm 1,6$ %). Ainsi, les protéines UL130 synthétisées en trois heures à partir de l'ajout du glucose marqué représente au maximum 3,64 % de la quantité totale synthétisée depuis l'induction si la protéine UL130 est purifiée à 100 %.

Ce résultat indique que durant le procédé de production en milieu minimum le retour de croissance est accompagné d'un flux de production négligeable. La phase de production est définie pendant les 2 premières heures suivant l'ajout d'IPTG. En milieu riche, la production totale n'évolue plus de manière significative après une heure d'induction. De fait, par extrapolation des résultats de marquage isotopique, la production dans les procédés en milieu riche ne dure probablement qu'une heure après induction.

Le procédé de référence généré a permis d'étudier et de définir les cinétiques de croissance et de production de la protéine UL130. La même approche a été employée pour étudier l'influence de la production d'une deuxième protéine recombinante sur les cinétiques de croissance.

4.3. Procédé modèle de production de la protéine Hla

La protéine Hla est utilisée comme modèle car son étude chez Sanofi Pasteur a révélé que sa production en quantité élevée engendre une faible diminution de la croissance d'*E. coli* BL21 en milieu riche contrairement à la protéine UL130. Si ce phénotype est retrouvé dans le cadre du procédé de référence de la thèse, cette protéine permettrait de mettre en évidence les contraintes métaboliques conduisant à l'arrêt de croissance par comparaison avec le procédé de production de la protéine UL130.

Les cinétiques de croissance bactérienne et de production sont analysées avec la même approche que pour l'étude de la production d'UL130 (figure 4.8).

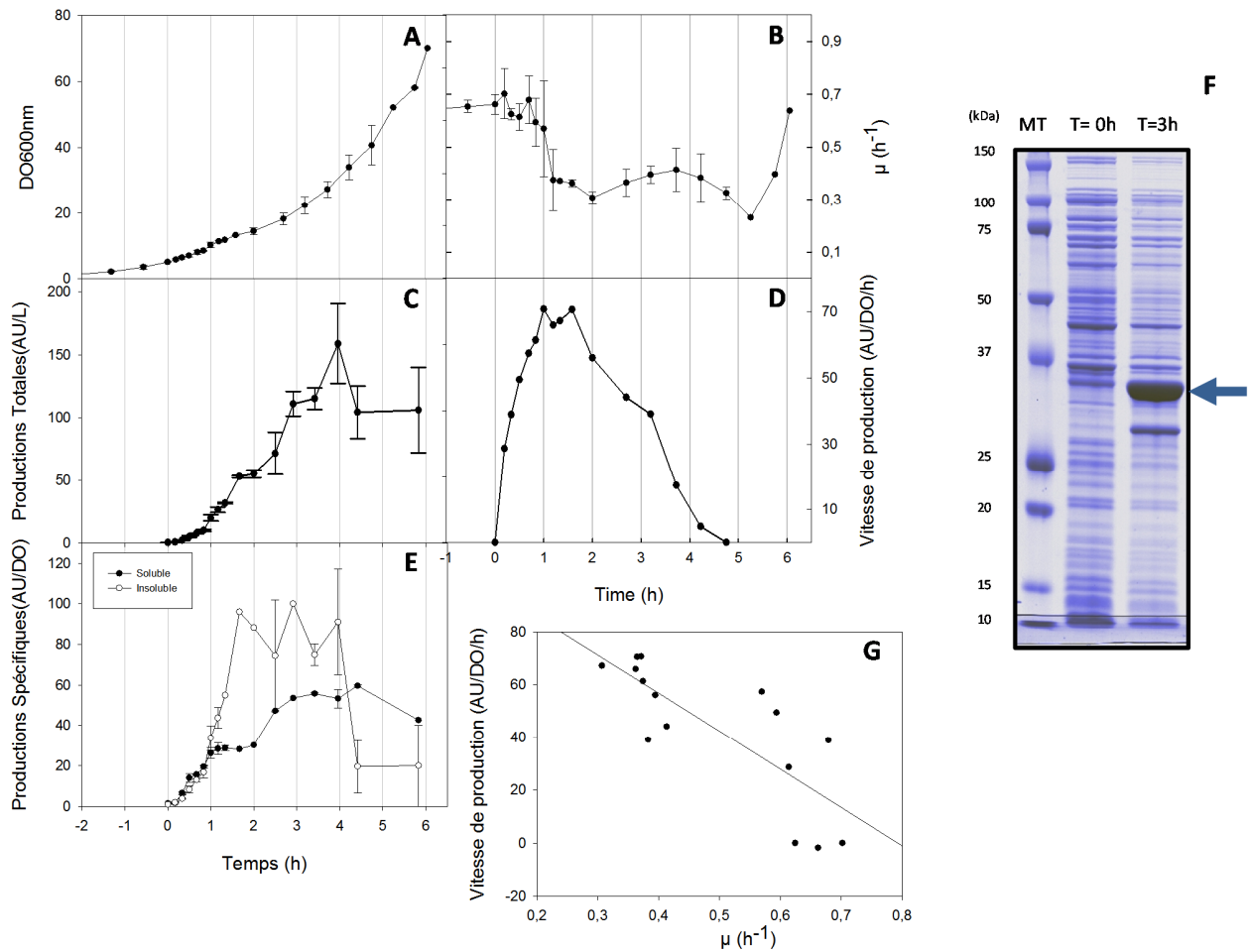


Figure 4.8 : Evolution de la croissance d'*E. coli* BL21 et de la production de la protéine Hla au cours du procédé de référence (milieu minimum, induction précoce). (A) Biomasse (DO), (B) taux de croissance (h^{-1}), (C) production totale (AU/L), (D) vitesse de production spécifique (AU/DO/L), (E) production spécifique (AU/DO), au cours du temps avec induction à $t=0h$. (F) Biomasse déposée à quantité de DO équivalente et colorée au bleu de Coomassie sur un gel SDS-PAGE à $T=0h$ et $T=3h$ avec un marqueur de taille (MT) (la flèche bleu signale la bande formée par la protéine Hla). (G) Corrélation obtenue entre le taux de croissance et la vitesse de production spécifique au cours de la période de production du procédé (entre 0 et 3h30 après induction) ($R^2 = 0,68$).

L'ajout d'IPTG à 5 uDO déclenche la synthèse de la protéine Hla (figure 4.8). Le taux de croissance de la souche passe de $0,7$ à $0,3 h^{-1}$, 1 heure après l'induction (figure 4.8, B). Le taux de croissance est ensuite constant à $0,3 h^{-1}$ jusqu'à 5 heures après l'induction puis atteint $0,7 h^{-1}$. Le procédé se termine 6 heures après induction à une DO finale de 65 uDO (figure 4.8, A).

La détermination de la vitesse de production spécifique révèle trois phases de production durant le procédé (figure 4.8, D). La première phase se caractérise par une augmentation progressive de la vitesse de production spécifique (entre l'induction et la première heure de production). Durant cette phase le taux de croissance diminue jusqu'à $0,3 \text{ h}^{-1}$. La deuxième phase de $t = 1 \text{ h}$ à $t = 4 \text{ h}$, se caractérise par une vitesse de production spécifique élevée et un taux de croissance constant entre $0,3 \text{ h}^{-1}$ et $0,4 \text{ h}^{-1}$. Enfin, la quantité de protéines produites reste stable à partir de 4 heures d'induction et donc la vitesse de production spécifique est nulle durant la troisième phase de $t = 4 \text{ h}$ à $t = 6 \text{ h}$ (figure 4.8, C). Ainsi, durant la période de production, une corrélation négative entre le taux de croissance et la vitesse de production spécifique est observée (figure, panel G).

Ces trois phases de production (augmentation de la vitesse de production spécifique, vitesse de production spécifique élevée puis nulle) se caractérisent également par les évolutions des productions spécifiques de protéines solubles et insolubles (figure 4.8, panel E). En effet, durant la première phase, les productions spécifiques soluble et insoluble augmentent de manière similaire. Puis, lors de la seconde phase, la fraction insoluble augmente plus rapidement que la fraction soluble. La fraction insoluble reste alors élevée et constante durant toute la seconde phase et la fraction soluble augmente lentement. Lors de la troisième phase, la fraction insoluble diminue rapidement et la fraction soluble reste stable jusqu'à la fin du procédé. Durant cette phase, la vitesse de production spécifique est nulle conduisant à une diminution de la production spécifique (soluble et insoluble) (figure 4.8, panel D et E).

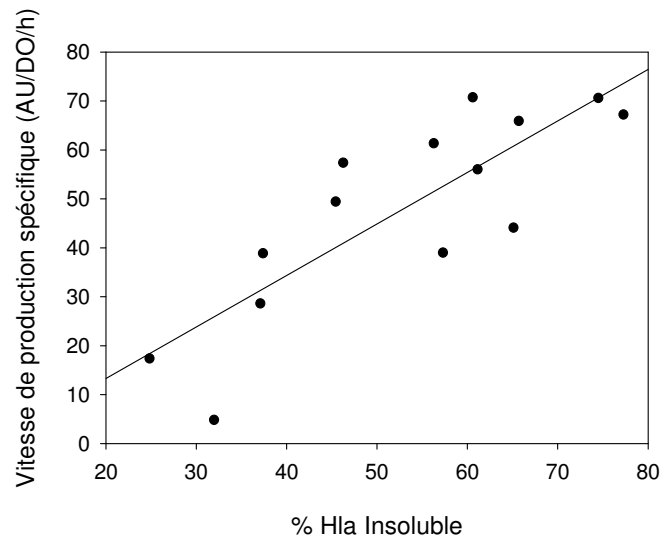


Figure 4.9 : Corrélation entre le pourcentage de protéine insoluble parmi la production spécifique soluble et insoluble en fonction de la vitesse de production spécifique de la protéine H1a en milieu minimum ($R^2 = 0,69$).

Au cours des trois phases du procédé, la vitesse de production spécifique est corrélée au pourcentage de protéines insolubles parmi la production spécifique de H1a (figure 4.9). Le pourcentage de protéines insolubles est calculé par rapport à la production spécifique totale, c'est-à-dire la somme des productions spécifiques solubles et insolubles. Une forte vitesse de production semble donc conduire à l'apparition de protéines insolubles. Cette interprétation est largement décrite dans la littérature et pourrait expliquer la corrélation entre les vitesses de production spécifiques et pourcentages de protéines insolubles observées (Baneyx & Mujacic 2004).

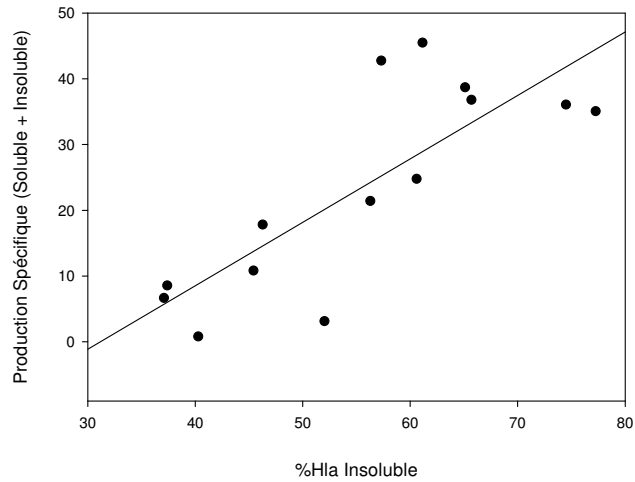


Figure 4.10 : Analyse de corrélation entre la proportion de la fraction insoluble parmi la production spécifique soluble et insoluble et la production spécifique ($R^2 = 0,64$).

Une corrélation apparaît également entre le pourcentage de protéines insolubles et la production spécifique (soluble et insoluble) (figure 4.10). Cette corrélation indique que la part de protéines insolubles augmente lorsque la production spécifique augmente. Cette corrélation pourrait être expliquée par une trop forte concentration intracellulaire de la protéine Hla conduisant à son agrégation. Ainsi, l'augmentation de la proportion de protéine insoluble peut-être expliquée par l'augmentation de vitesse de production et/ou la saturation intracellulaire en protéines.

D'après l'analyse de la densitométrie des bandes obtenues par migration des lysats cellulaires sur gel SDS-PAGE, la production spécifique de Hla est au moins égale à la production spécifique d'UL130 (figure 4.8, panel F comparée à figure 4.6, panel F). Si ces productions spécifiques sont considérées comme égales, alors la production totale est estimée entre 2,5 et 4,5g/L de culture après 3 heures d'induction. Le standard de la protéine Hla purifiée et quantifié n'est pas disponible, sa quantification est donc issue uniquement d'une estimation.

L'analyses des cinétiques de croissance bactérienne et de production des protéines recombinantes démontrent que malgré une production de la protéine Hla à des productions spécifiques similaires voir supérieures à UL130, la croissance bactérienne est moins impactée par l'induction. La protéine Hla permet de fournir une condition de comparaison entre un procédé à forte

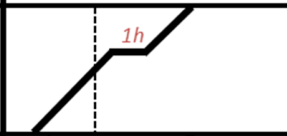

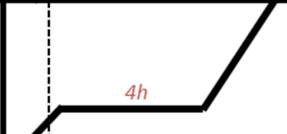
charge métabolique (UL130) et un procédé un plus faible charge métabolique (Hla) afin de déterminer les contraintes métaboliques liées à la surproduction de protéines recombinantes.

4.4. Discussions

L'objectif du travail présenté dans ce chapitre est de mettre en place un procédé de référence pour répondre aux besoins de la thèse à partir du procédé standard du département BRD de Sanofi Pasteur. Les modifications apportées ont été étudiées et le procédé de référence a permis de caractériser les productions des protéines UL130 et Hla par l'analyse des cinétiques de croissance et de production des deux protéines. Cette caractérisation a été réalisée par la détermination de phases des procédés qui serviront à comparer les analyses fonctionnelles du métabolisme avec les cinétiques de croissance et de production des protéines recombinantes.

L'analyse du procédé standard a révélé qu'une forte production de la protéine UL130 engendre un arrêt rapide de la croissance (tableau 1).

Tableau 4.1 : Cinétiques de croissance et de production de la protéine UL130 obtenues lors de l'étude des procédés en milieu riche induction tardive, milieu riche induction précoce et milieu minimum induction précoce.

Conditions	Croissance <i>(temps d'arrêt de croissance)</i>	Production totale <i>(temps de production)</i>
Milieu Riche <i>Induction Tardive</i>	 1h	+ + 1h
Milieu Riche <i>Induction précoce</i>		+ + 1h
Milieu Minimum <i>Induction précoce</i>	 4h	+ + + 2h

Le procédé standard n'est pas compatible avec une analyse des cinétiques de croissance et de production sur des temps longs car l'induction tardive est effectuée en milieu de phase exponentielle de croissance. La première modification apportée est la mise en place d'une induction plus précoce, c'est-à-dire à en début de phase exponentielle de croissance. Cette induction n'engendre pas de changement de la cinétique de production. L'arrêt de croissance n'est en revanche pas observé dans cette condition mais un ralentissement significatif du taux de croissance est visible. La dernière étape a consisté à remplacer le milieu riche par un milieu minimum pour pouvoir mettre en place les analyses fonctionnelles du métabolisme. Ce changement de milieu a accentué l'impact négatif de la production sur la croissance mais a également permis d'améliorer la production spécifique. Ce dernier point est surprenant car la disponibilité des ressources et en particulier des acides aminés semble intuitivement supérieure en milieu riche. Une observation intéressante est que l'augmentation de la production spécifique pendant deux heures est concomitante avec un arrêt total et prolongé de la croissance en milieu minimum. Cet arrêt de croissance prolongé orienterait l'utilisation des ressources au bénéfice de la production de la protéine et au détriment de la croissance. La modification du milieu de culture est une modification profonde du cadre d'étude de la production de protéine recombinante. En effet, il faut s'attendre à ce que les contraintes métaboliques apparaissant en condition de production en milieu riche ou minimum soient différentes. Au vu de la compréhension apportée par l'utilisation de ce milieu, de l'absence de variabilité des lots de composés utilisés, des performances similaires en terme de production et des moyens d'optimisation du milieu qu'offre cette base, l'utilisation de milieu chimiquement minimum, incrémentée par la suite d'une diversité de sources de carbone, est une base intéressante à l'optimisation de procédé de production de protéines recombinantes dans un contexte industriel.

La reprise de croissance en milieu minimum n'est pas accompagnée de production de la protéine UL130. L'absence de production est étonnante au vu de la présence de plasmide dans chaque bactérie. Une expérience, dont les résultats ne sont pas présentés ici, a consisté à récupérer des bactéries à la fin du procédé de production de la protéine UL130 en milieu minimum puis à les inoculer à 0,8 uDO dans un nouveau procédé en milieu minimum. A 5 uDO, l'IPTG à 1mM a été ajouté sans qu'un impact significatif sur la croissance ne soit observé indiquant l'absence probable de production. Ce résultat suggère que les bactéries, après la phase d'arrêt de croissance perde irréversiblement leur

capacité de production malgré la présence avérée du plasmide. L'hypothèse d'un arrêt de l'expression du système T7 par mutation est étonnant au vu de la répétabilité observée concernant le moment de reprise de croissance. Un séquençage de la zone DE3 du génome d'*E. coli* BL21 et du plasmide en fin de culture permettrait peut-être de comprendre cette absence de production après la reprise de croissance.

Les procédés de production des protéines UL130 et Hla en milieu minimum ont été définies par des phases (tableau 2).

Tableau 4.2 : Phases caractéristiques des cinétiques de croissance et de production des procédés de production des protéines UL130 et Hla en milieu minimum.

Conditions	Phase 1	Phase 2	Phase 3	Phase 4
Milieu Minimum <i>Induction précoce</i> UL130	0,7 - 0 h ⁻¹ 70 % 0 h - 1 h	0 h ⁻¹ 30 % 1 h - 2 h	0 h ⁻¹ 0 % 2 h - 5 h	0,7 h ⁻¹ 0 % 5 h - 9 h
Milieu Minimum <i>Induction précoce</i> Hla	0,7 - 0,5 h ⁻¹ 20 % 0 h - 1 h	0,3 h ⁻¹ 80 % 1 h - 4 h	0,7 h ⁻¹ 0% 4 h - 6 h	
Légende : Taux de croissance (vert), Pourcentage de la production total maximale (bleu), Temps à partir de l'induction				

Ces phases sont caractérisées par les différences de cinétiques de croissance et de production observées dans chacun des procédés. Dans chacune des phases des deux procédés, la production est inversement corrélée au taux de croissance de la bactérie. La phase 1 du procédé de production des protéines UL130 et la phase 2 du procédé Hla sont caractérisées par des vitesses de productions spécifiques élevées. Ces phases sont suivit d'un arrêt de la production pour le procédé UL130 et Hla. La comparaison de ces phénomènes entre les deux procédés met en évidence l'apparition de mécanismes cellulaires communs aux deux procédés entraînant l'inhibition du taux de croissance ou l'arrêt de production. La définition de ces phases permettra donc d'étudier les déterminants métaboliques liés à la production des protéines sur la cinétique de croissance au cours des procédés de production des protéines UL130 et Hla.

La courbe de calibration effectuée avec la protéine UL130 purifiée et quantifiée par ELISA révèle que la production spécifique maximale en milieu minimum peut être estimée à 0,25 - 0,5 g de protéines UL130 par gramme de biomasse sèche. La comparaison de la densitométrie de la bande formée par la protéine UL130 avec celle formée par la protéine Hla permet d'estimer que la production spécifique maximale est similaire entre les deux procédés. Cependant, la production de la protéine Hla n'entraîne pas d'arrêt prolongé de croissance. Cette différence de cinétique de croissance pour une production similaire de protéines met en évidence que la charge métabolique engendrée par la production de la protéine UL130 est plus importante que la charge métabolique engendrée par la production de Hla. Les analyses fonctionnelles du métabolisme de la souche *E. coli* BL21 au cours de la production de ces deux protéines pourront être comparées afin de mettre en évidence les déterminants métaboliques à l'origine de cette différence de charge métabolique.

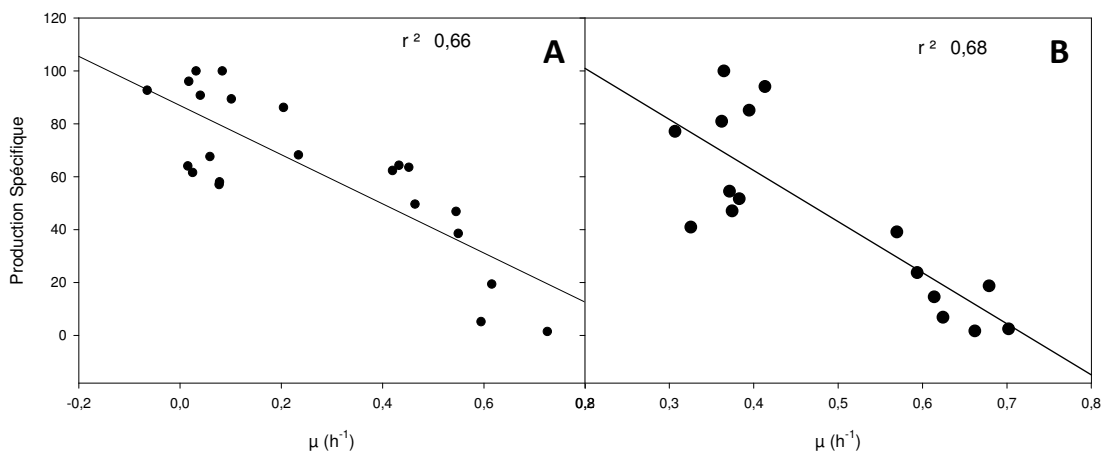


Figure 4.11 : Corrélations entre la production spécifique et le taux de croissance au cours des procédés de production des protéines UL130 (A) et Hla (B).

De manière très intéressante, une corrélation négative est observée entre la production spécifique et le taux de croissance lors des procédés de production des protéines UL130 et Hla (figure 4.11). Il est décrit dans la littérature que la production spécifique ne peut pas dépasser durablement 30% des protéines totales sans conduire à un arrêt de la croissance (Dong et al. 1995). Ce résultat de 30% a également été estimé par modélisation de l'interférence de la vitesse de production sur les lois régissant le taux de croissance (Scott et al. 2010). Au-delà de 30% de protéine d'intérêt parmi les protéines totales, les ARNm de la protéine recombinante exercent une compétition trop importante

Chapitre 4 : Etude des cinétiques de croissance et de production des protéines UL130 et Hla avec le procédé de référence.

avec les autres ARNm de la cellule et en particulier les ARNm ribosomiaux. Une réduction trop importante de la traduction des ARNm ribosomiaux entraîne obligatoirement un arrêt de la croissance. La corrélation observée serait donc le résultat de la compétition entre la croissance et la production pour les ressources bactériennes (acides aminés, énergie...) et la traduction. Il est même surprenant de constater que la production de telles quantités de protéines Hla ne conduise pas à l'arrêt total de la croissance. Le prochain chapitre est porté sur l'étude des contraintes métaboliques à l'origine des différences de charges métaboliques engendrées par la production des protéines UL130 et Hla.

Chapitre 4 : Etude des cinétiques de croissance et de production des protéines UL130 et H1a avec le procédé de référence.

Chapitre 5 – Analyse des contraintes métaboliques liées à la production des protéines UL130 et Hla au cours du procédé de référence.

Le procédé standard de production du département BRD de Sanofi Pasteur a été modifié dans le but d'obtenir un procédé de référence aux études à venir. Ce nouveau procédé permet d'effectuer des analyses fonctionnelles du métabolisme dans un contexte de surproduction de protéines recombinantes en milieu minimum. Le procédé de référence a été utilisé pour comparer les productions des protéines UL130 et Hla et ainsi étudier l'influence de la production sur la croissance de la souche *E. coli* BL21. Les résultats de cette étude mettent en évidence que la production de la protéine UL130 engendre une plus forte charge métabolique pour la bactérie que la production de la protéine Hla. L'objectif du travail présenté dans ce chapitre est de comparer les analyses fonctionnelles du métabolisme du procédé de production de chaque protéine modèle afin de déterminer et quantifier les contraintes métaboliques majeures liées à la production de la protéine UL130 par *E. coli* BL21 dans un milieu minimum. Il s'agit ici de caractériser la perturbation du métabolisme causée par la production et d'identifier quelles ressources cellulaires deviennent limitantes pour la croissance lors de la production de protéines recombinantes.

Dans un premier temps, une analyse non ciblée de l'exométabolome est effectuée par résonance magnétique nucléaire (RMN) au cours des cultures. Les vitesses de consommation du glucose et d'excrétion des produits du métabolisme de la souche *E. coli* BL21 sans production sont d'abord définies afin de pouvoir comparer dans un second temps l'influence de la production sur le métabolisme de la bactérie. Cette analyse non ciblée permet d'étudier sans *a priori* l'impact de la production sur le métabolisme. En effet, l'excrétion de métabolites dans le surnageant peut révéler la présence de goulots d'étranglement de flux métaboliques et des modifications plus globales du fonctionnement métabolique.

Par ailleurs une quantification ciblée des acides aminés cellulaires est réalisée en condition de production par LC-MS dans le but d'étudier l'évolution des concentrations intracellulaires en acides aminés aux cours des productions des protéines UL130 et Hla.

Un modèle descriptif des mécanismes observés est ensuite présenté. Ce modèle permet de mettre en évidence l'influence de la différence entre la proportion en acides aminés de la protéine recombinante et celle des protéines de la biomasse sur les concentrations intracellulaires en acides aminés au cours des procédés de production.

La limitation en acide aminé au cours d'une production avec un système T7 a été plusieurs fois rapportée dans la littérature (Sandén et al. 2003; Bachinger et al. 2001; Dong et al. 1995). Ce type de contrainte engendre l'augmentation de la proportion d'ARNt non chargés de certains acides aminés limitants. La présence d'ARNt non chargés est détectée par la protéine RelA (Potrykus & Cashel 2008). Cette protéine, fixée sur les ribosomes, synthétise du ppGpp et du pppGpp lors de la détection d'ARNt non chargés, déclenchant la réponse stringente. Dans le but de vérifier si la réponse stringente est activée au cours des procédés de production suite à l'apparition de la limitation en acides aminés suspectée, le ppGpp est quantifié au cours des procédés de production des protéines Hla et UL130.

5.1. Analyse non ciblée de l'exométabolome d'*E. coli* BL21.

5.1.1. Procédé de culture sans induction

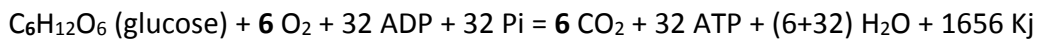
Dans cette étude, les surnageants obtenus durant la culture de la souche *E. coli* BL21 sans induction sont analysés par RMN (1D-proton). L'analyse non ciblée des molécules du surnageant permet d'identifier l'ensemble des métabolites majoritaires, de quantifier la vitesse spécifique de consommation du glucose et les vitesses spécifiques d'excrétion des produits majeurs du métabolisme en absence de production. Une analyse RMN est ensuite effectuée sur les surnageants de culture obtenus en condition de production des protéines Hla et UL130 afin de comparer les différences d'accumulation causées par la charge métabolique associée à la production.

La mesure de la vitesse de production de CO₂ par les bactéries à l'échelle du bioréacteur (CER) (mmol/L/h) est utilisée pour analyser l'activité métabolique des bactéries au cours de la culture. Le CO₂ est principalement le produit du catabolisme du glucose, en particulier au niveau du cycle de Krebs. Le cycle de Krebs est amphibolique, c'est-à-dire qu'il possède à la fois une fonction catabolique et

anabolique. La fonction catabolisme du cycle de Krebs permet l'oxydation de l'acétyl-CoA en CO₂ pour récupérer de l'énergie d'oxydo/réduction (ox/red) sous forme de NADH et FADH₂. Ces cofacteurs réduits alimentent la chaîne oxydative permettant la phosphorylation de l'ADP en ATP par l'ATPase. Le cycle de Krebs est également dit anabolique car il fournit des précurseurs à la synthèse des acides aminés et d'acides gras. La perte des intermédiaires du cycle de Krebs, utilisés pour l'anabolisme, est compensée par les voies anaplérotiques.

La vitesse de consommation d'O₂ par les bactéries à l'échelle du bioréacteur (OUR) (mmol/L/h) a été mesurée de la même façon que la production de CO₂ lors des cultures. L'oxydation des cofacteurs réduits pour la phosphorylation de l'ADP en ATP conduit à la consommation d'oxygène. Au final, le catabolisme du glucose conduit à une production de CO₂ équivalente à la consommation d'O₂ (mol/mol).

Bilan du catabolisme du glucose :



Avec un nombre d'ATP synthétisé par atome d'oxygène réduit (P/O) = 2,5 pour NADH et 1,5 pour FADH₂

La biomasse issue de l'anabolisme possède un degré d'oxydation équivalent au glucose et représente pour la souche *E. coli* BL21 entre 54 et 59 % du carbone issu du glucose (30 à 38 % pour le CO₂) (Marisch, Bayer, Scharl, et al. 2013; Marisch, Bayer, Cserjan-Puschmann, et al. 2013). L'anabolisme ne modifie pas le bilan oxydatif du catabolisme du glucose en terme de différence entre le CO₂ produit et l'O₂ consommé permettant de conserver un quotient respiratoire (QR = CER/OUR) égale à 1 en condition aérobie sur glucose. Ce ratio peut différer lorsque le potentiel d'oxydation du substrat diffère du glucose ou lorsque la composition de la biomasse est modifiée.

Tableau 5.1 : Vitesses de consommation et d'excrétions des métabolites (mmol/g/h) au cours de la culture de la souche *E. coli* BL21 en milieu minimum. Les vitesses spécifiques sont calculées à l'aide du logiciel Physiofit développé dans le laboratoire (<https://imetatoul.insa-toulouse.fr/software/physiofit/>) sur un duplicat de culture. La vitesse de production du CO₂ a été réalisée par intégration de la courbe exponentielle formée par le CER au cours du temps et rapporté à la biomasse sèche. Tous les métabolites majoritaires identifiés dans les surnageants de culture sont présentés dans le tableau.

μ (h ⁻¹)	0,76 ± 0,04
CO ₂ (mmol/g/h)	17,01 ± 0,51
qGlc (mmol/g/h)	-8,24 ± 0,16
qFormate (mmol/g/h)	0,15 ± 0,01
qEthanol (mmol/g/h)	0,08 ± 0,01
qAcetate (mmol/g/h)	0,04 ± 0,03
qSuccinate (mmol/g/h)	0,03 ± 0,01
qValine (mmol/g/h)	0,03 ± 0,00
qPyruvate (mmol/g/h)	0,02 ± 0,00
qGlutamate (mmol/g/h)	0,03 ± 0,01
qAlanine (mmol/g/h)	0,02 ± 0,00
qUracil (mmol/g/h)	0,01 ± 0,00
qPhénylalanine (mmol/g/h)	0,01 ± 0,00
qLeucine (mmol/g/h)	>0,01 ± 0,00
qPrephénate (mmol/g/h)	>0,01 ± 0,01

En phase exponentielle de croissance sans induction de la souche *E. coli* BL21 (avec le plasmide pSP), le taux de croissance est de 0,76 h⁻¹ (± 0,04 h⁻¹) et la vitesse de consommation du glucose de 8,24 mmol/g/h (± 0,16 mmol/g/h) (tableau 5.1). La production de CO₂ représente 38 % (± 1 %) de carbone du glucose consommé (mol/mol) avec une vitesse d'excrétion de 17,81 mmol/g/h (± 0,51 mmol/g/h). L'excrétion des autres métabolites représente moins de 2% (mol/mol) du glucose consommé (biomasse 59 %, CO₂ : 38 %, autres 2 %). Le formate, l'éthanol et l'acétate sont les métabolites les plus excrétés dans le surnageant avec des vitesses d'excrétion de 0,17, 0,08 et 0,04 mmol/g/h respectivement. De plus faibles vitesses d'excrétions sont mesurées pour les acides aminés tels que la valine, le glutamate, l'alanine, la phénylalanine et la leucine au cours de la croissance (0,03 à 0,01 mmol/g/h). De faibles concentrations en uracile sont également détectées au cours de la culture de la souche (0,01mmol/g/h).

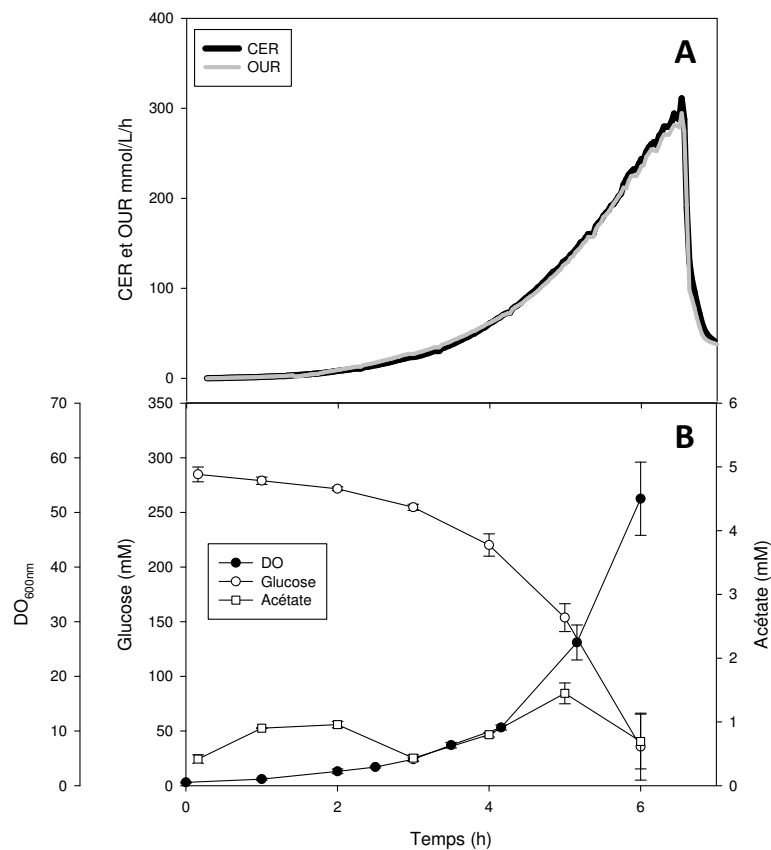


Figure 5.1 : Activité métabolique d'*E.coli* BL21 (pSP) en milieu minimum sans induction. (A) CER et de l'OUR (mmol/L/h) d'une culture représentative des cultures non induite d'*E. coli* BL21 sur milieu minimum. (B) Quantification du glucose et de l'acétate par RMN (1D-proton) et mesure de la DO au cours de la culture.

L'OUR est superposé à la courbe exponentielle du CER indiquant un QR de 1 durant toute la phase exponentielle de croissance (figure 5.1, A). La diminution rapide de l'OUR et du CER 6,5 h après inoculation révèle la consommation totale du glucose du milieu et donc la fin du procédé. Ainsi la consommation totale du glucose peut-être déterminée à l'aide de cette diminution caractéristique du CER et de l'OUR (figure 5.1, A). La mesure de la concentration d'acétate au cours du procédé révèle des périodes d'excrétion et de consommation d'acétate en présence de glucose, rendant difficile une détermination précise de la vitesse d'excrétion (figure 5.1, B).

Cette étude des caractéristiques du métabolisme de la souche *E. coli* BL21 (pSP) (activité métabolique à un quotient respiratoire de 1, excrétion d'acides organiques, d'acides aminés tels que

valine ou alanine, faible excrétion d'acétate) permet ensuite d'étudier l'influence de la production de protéines recombinantes sur le métabolisme. La présence de formate, d'éthanol et d'acétate est courante lors de la culture d'*E. coli* (Marisch, Bayer, Scharl, et al. 2013; Simen et al. 2017). En revanche, l'accumulation de certains métabolites n'est pas explicitée dans la littérature (la valine, le glutamate, l'alanine, la phénylalanine, leucine, uracile, le préphénate). Du fait que les cultures soient réalisées en bioréacteurs, les hautes densités cellulaires générées rendent détectables ces métabolites excrétés très faiblement dans le surnageant cellulaire. En effet, ces excrétions représentent au total moins de 0,5 % du carbone consommé.

5.1.2. Procédé de production des protéines UL130 et Hla

En condition de production, la croissance n'est pas exponentielle et les vitesses de consommation ou d'excrétion des métabolites évoluent dans le temps rendant leur comparaison moins pertinente. Afin de pouvoir comparer les cinétiques d'accumulation des métabolites des conditions induites à celles de la condition non induite, les quantités de métabolites sécrétés sont comparées à la quantité de glucose consommée dans la même période. Ainsi, les rendements d'excrétion des métabolites par rapport au glucose consommé (mmol/mol) sont présentés plutôt que les vitesses de consommation. Les périodes sont définies pour chaque métabolite entre l'induction et le moment où le métabolite est retrouvé en concentration maximale dans le surnageant.

Au cours du chapitre précédent, des phases caractéristiques des procédés ont été définies en fonctions des cinétiques de croissance de la souche *E. coli* BL21 et de production des protéines UL130 ou Hla (tableau 2).

Tableau 5.2 : Phases caractéristiques des cinétiques de croissance et de production des procédés de production des protéines UL130 et Hla en milieu minimum. Les phases en rouge ne sont pas analysées durant cette étude.

Conditions	Phase 1	Phase 2	Phase 3	Phase 4
Milieu Minimum <i>Induction précoce</i> UL130	0,7 - 0 h ⁻¹ 70 % 0 h – 1 h	0 h ⁻¹ 30 % 1 h – 2 h	0 h ⁻¹ 0 % 2 h – 5 h	0,7 h ⁻¹ 0 % 5 h – 9 h
Milieu Minimum <i>Induction précoce</i> Hla	0,7 – 0,5 h ⁻¹ 20 % 0 h – 1 h	0,3 h ⁻¹ 80 % 1 h – 4 h	0,7 h ⁻¹ 0% 4 h- 6 h	
Légende : Taux de croissance (vert), Pourcentage de la production total maximale (bleu), Temps à partir de l'induction				

Ces phases permettent de mettre en lien l'analyse de l'exométabolome avec les cinétiques de croissance et de production observées dans chaque procédé. La phase 4 du procédé UL130 et la phase 3 du procédé Hla ne sont pas analysées car ces phases sont caractérisées par une croissance sans production (figure 5.2).

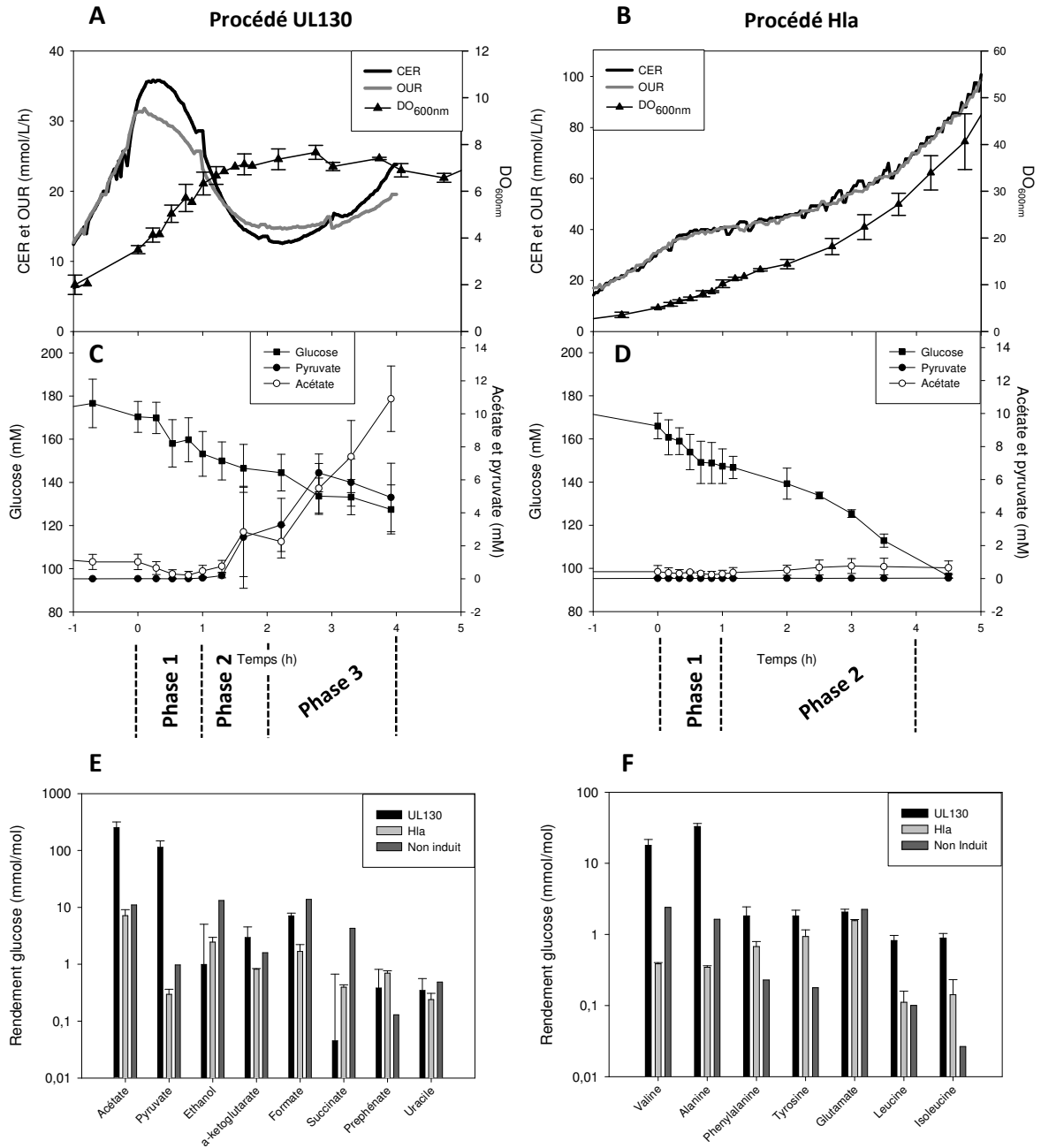


Figure 5.2 : Analyses de l'exométabolome des procédés de production des protéines UL130 et Hla. Evolution du CER (mmol/L/h), de l'OUR (mmol/L/h) et de la DO des procédés de production des protéines UL130 (A) et Hla (B) au cours du temps. Dans un but de visibilité des données, les cinétiques du CER et de l'OUR sont issues d'une culture représentative du duplicat biologique utilisé dans les autres panels. Quantification de la concentration en glucose, en pyruvate et en acétate dans le surnageant (mmol/L) des procédés de production des protéines UL130 (C) et Hla (D) au cours du temps. Rendement d'excrétion des acides organiques (E) ou des acides aminés (F) (mmol/mol) à partir du

début de la phase 1 jusqu'à la fin de la phase 3 du procédé UL130 et du début de la phase 1 jusqu'à la fin de la phase 2 du procédé Hla (voir tableau 2).

Au cours des deux procédés de production, une baisse significative de la concentration en glucose durant la phase 1 met en évidence que la vitesse de consommation en glucose augmente en début de production des protéines UL130 et Hla (figure 5.2, C et D).

Cette augmentation de la consommation du glucose révèle une probable augmentation de la demande en carbone durant la phase 1 des procédés où la production est déclenchée par l'ajout d'IPTG en condition de croissance.

Pour le procédé de production de la protéine UL130, les analyses du CER et de l'OUR révèlent un découplage de la production de CO₂ et de la consommation d'O₂ dès l'ajout de l'IPTG durant la phase 1 du procédé (figure 5.2, A). La production de CO₂ augmente 30 minutes après ajout d'IPTG puis diminue ensuite durant 1h30 (phase 1 et 2). Par contre, la consommation d'oxygène décroît dès l'ajout d'IPTG et cela pendant 2 heures (phase 1 et 2). Une décorrélation apparaît également entre le CER, l'OUR et la quantité de biomasse (DO_{600nm}) durant la phase 1 du procédé. En effet une croissance a été démontrée durant toute la phase 1 du procédé alors que le CER n'augmente que durant 30 min puis diminue et l'OUR diminue durant cette période. Pour le procédé de production de Hla, le déclenchement de la production engendre un ralentissement similaire entre le CER et l'OUR au milieu de la phase 1 du procédé (figure 5.2, B). Ce ralentissement du CER et de l'OUR est observé avant le ralentissement de la DO qui n'intervient qu'à la fin de la phase 1. Après la première phase, l'activité métabolique augmente de façon exponentielle jusqu'à la fin du procédé.

Au vue de la quantité de protéine UL130 synthétisée durant la phase 1 du procédé, il est probable que le découplage entre l'OUR et le CER soit la conséquence de la production élevée de la protéine recombinante durant cette phase et d'une perturbation majeure du métabolisme (figure 5.2, A et tableau 2). En effet, une plus forte production de CO₂ par rapport à la consommation d'oxygène peut mettre en évidence soit une perturbation du catabolisme avec une oxydation incomplète du glucose soit une modification profonde de la composition chimique de la biomasse. Etant donné que la formule brut de la protéine UL130 (C₁H_{1,84}O_{0,47}N_{0,25}) est proche de la formule brut de la biomasse (C₁H_{1,77}O_{0,49}N_{0,24}) (Stockar & Liu 1999), un fort enrichissement de la biomasse en protéine UL130 ne

semble pas expliquer une telle différence entre l'OUR et le CER. La modification du catabolisme du glucose est donc plus probable pour expliquer ce phénomène. Plus particulièrement, une limitation en cofacteurs réduits (ex : NADH) au niveau de la phosphorylation oxydative pourrait entraîner une baisse de la consommation d'oxygène par rapport à la production de CO₂. Au cours du procédé de production de la protéine Hla le découplage entre le CER et l'OUR n'est pas observée. Ainsi, il est possible que ce découplage soit un indicateur d'une importante charge métabolique de la production. Un découplage apparaît entre le CER ou l'OUR qui diminuent (procédé UL130) ou n'augmentent plus (procédé Hla) à la fin de la phase 1 et la biomasse qui augmente jusqu'à la fin de la phase 1 pour les deux procédés. Ce résultat signifie que de la biomasse est générée avec une production plus faible de CO₂ et une consommation plus faible d'O₂. Cette observation appuie un phénomène de découplage de l'anabolisme et du catabolisme apparaissant dans la phase 1 du procédé de production de la protéine UL130 et apparaissant en moindre proportion dans le procédé de production de la protéine Hla.

Au cours du procédé de production de la protéine UL130, une excrétion importante d'acétate et de pyruvate est observée durant les phases 2 et 3, atteignant 10 mM d'acétate et 6 mM de pyruvate dans les surnageants au bout de 4 heures d'induction (figure 5.2, C). Cette accumulation représente un rendement molaire par rapport au glucose consommé de 252 mmol/mol pour l'acétate et 113 mmol/mol pour le pyruvate soit une excrétion totale de 8% du carbone issu du glucose consommée dans ces phases (figure 5.2, E). Le lactate n'est pas détecté et aucune accumulation significative d'éthanol, de succinate, d' α -cétoglutarate, ou d'uracile n'a été observée (figure 5.2, E). En revanche le préphénate est produit avec un rendement 10 fois plus élevé que la souche non induite. Une excrétion importante d'alanine, de valine, de tyrosine, de phénylalanine, de leucine et d'isoleucine est également observée avec une évolution de la concentration dans le surnageant comparables à celles observées pour l'acétate et le pyruvate (figure 5.2,C et F). Ces rendements d'excrétions sont 10 à 20 fois plus élevés que ceux de la souche non induite. Au cours du procédé de production de la protéine Hla, seules des accumulations de préphénate, de tyrosine, de phénylalanine et d'isoleucine sont observée avec un rendement 5 à 10 fois plus important que ceux de la souche non induite. Au contraire, pour certains métabolites la production de la protéine Hla conduit à une diminution du rendement d'excrétion telle que la valine et l'alanine et leur précurseur commun, le pyruvate par rapport à ceux de la condition non induite.

L'absence de détection du lactate ou d'accumulation d'éthanol et de succinate révèle l'absence d'oxydation des cofacteurs red/ox par excrétion de ces acides organiques dans le surnageant aux cours des deux procédés. Ainsi, en condition de production de la protéine UL130 et Hla la limitation en cofacteurs oxydés ne semblent pas être une contrainte métabolique de la production.

L'excrétion importante de pyruvate et d'acétate en condition de production a été observé par Wittman et al., 2007 dans un procédé de production d'une protéine recombinante (souche TG1, production de hFGF-2 en milieu minimum sur glucose). Dans ces travaux, l'analyse fluxomique des souches en condition de production a révélé que l'activité du cycle de Krebs semblait insuffisante pour utiliser l'ensemble de l'acétyl-CoA. De fait, le pyruvate et l'acétate s'accumule dans la cellule et sont excrétés. Les flux présentés dans ces travaux démontrent une activation du shunt glyoxylique, voie anaplérotiques du cycle de Krebs. La comparaison de ces résultats et ceux obtenues durant la production de la protéine UL130 permettent d'envisager que le cycle de Krebs n'est pas assez efficace pour utiliser l'ensemble de l'acétyl-CoA. Ce phénomène pourrait être la conséquence d'une utilisation massive des précurseurs métaboliques, dont des intermédiaires du cycle de Krebs, transformés en acides aminés et consommés pour la forte synthèse de protéine recombinante durant la phase 1 et 2 du procédé de production de la protéine UL130. Cette perte des précurseurs pourrait être insuffisamment compensée par les voies anaplérotiques conduisant à un dysfonctionnement du cycle de Krebs, une baisse de son activité catabolique et une accumulation de pyruvate et d'acétate.

Dans les deux procédés de production, une accumulation 10 fois plus importante de préphénate est observée dans le surnageant comparée aux rendements de la culture non induite. Ce métabolite est un intermédiaire des voies de biosynthèse de la phénylalanine et de la tyrosine. De façon intéressante, les excrétions significative de ces acides aminés dans le surnageant en condition de production sont également observées révélant une perturbation de cette voie de biosynthèse. L'importante excrétion d'alanine, de valine, de leucine et d'isoleucine au cours du procédé de production de la protéine UL130 est cohérente avec l'accumulation de pyruvate, précurseur de ces acides aminés branchés. Ces accumulations sont surprenantes étant donné que la production de protéines recombinantes entraîne un surcroît de consommation des acides aminés. L'accumulation de

ces acides aminés et des précurseurs associés met en évidence la forte perturbation des voies de biosynthèse des acides aminés au cours de la production.

L'ensemble des résultats obtenus durant le procédé de production de la protéine UL130 semblent indiquer une déstabilisation du cycle de Krebs entraînant une accumulation d'acétate et de pyruvate, une possible limitation en cofacteurs réduits et une perturbation de l'homéostasie des voies de synthèse des acides aminés. Certaines de ces observations sont retrouvées avec le procédé de production de Hla (accumulation de certains acides aminés, découplage entre la DO et le CER) mais en plus faibles proportions. L'absence d'accumulation d' α -cétoglutarate ou de glutamate, métabolites clés de l'assimilation de l'azote minéral et de son transfert vers les acides aminés, ne permet pas de mettre en évidence une limitation en azote. L'absence d'accumulation de lactate, d'éthanol et de succinate rend peu probable une limitation en facteurs oxydés.

5.2. Quantification des acides aminés libres en condition de production

L'objectif est maintenant de vérifier si la forte production de la protéine UL130 ou Hla conduit à une perturbation significative des concentrations intracellulaires en acides aminés. Les concentrations intracellulaires en acides aminés libres sont stables en condition de croissance exponentielle des bactéries car la synthèse de chaque acide aminés s'équilibre avec leur consommation. De fait, une modification de la synthèse ou de la consommation des acides aminés peut impacter leur concentration intracellulaire.

La quantification des acides aminés est réalisée par couplage d'une chromatographie en phase liquide à haute pression (HPLC) à un spectromètre de masse à haute résolution (Orbitrap®). Ce type d'instrument possède une sensibilité supérieure à un spectromètre RMN de 800Mhz permettant de détecter des diminutions significatives de concentrations intracellulaires des acides aminés. Certains acides aminés tels que la cystéine, l'arginine ou la valine sont cependant très mal ionisés par le spectromètre de masse conduisant à l'absence de leur détection au cours des analyses.

En condition de production, la concentration intracellulaire d'acides aminés libres peut évoluer très rapidement. La méthode d'échantillonnage est basée sur un arrêt immédiat du métabolisme en ajoutant le prélèvement de culture (bactéries et surnageant de culture) à un mélange de solvants organiques à -40°C (acétonitrile, méthanol, eau). Ce mélange permet à la fois de bloquer toutes les réactions métaboliques et de casser les membranes cellulaires libérant les acides aminés intracellulaires dans le mélange. Une méthode courante pour quantifier les acides aminés intracellulaires consiste à quantifier d'une part les acides aminés du surnageant (extracellulaires) et d'autre part les acides aminés totaux du milieu de culture (intracellulaires et extracellulaires) (Millard et al. 2014). La différence entre la quantité d'acides aminés totaux et la quantité d'acides aminés extracellulaires permet de quantifier les acides aminés intracellulaires. Cependant, les quantifications par RMN ont montré de fortes accumulations des acides aminés dans le surnageant. Si la quantité d'acides aminés intracellulaires est trop faible comparée à la quantité d'acides aminés extracellulaires alors la précision des mesures devient insuffisante pour les distinguer. Aussi, cette analyse a pour objectif principal de détecter des diminutions de concentrations en acides aminés libres intracellulaires. Une baisse significative de la concentration d'un acide aminé dans un extrait total de milieu de culture (extracellulaire et intracellulaire) révèle sa diminution au niveau intracellulaire et extracellulaire, et donc révèle une concentration intracellulaire en acide aminé très faible. En revanche, lors d'une accumulation des acides aminés, aucun discernement ne peut être effectué entre une accumulation des acides aminés de façon intracellulaire ou extracellulaire. La quantification des acides aminés au cours des productions est réalisée de façon relative à leur concentration avant induction. Ces quantifications ont été effectuées sur un triplicat biologique du procédé de production de la protéine UL130 et en duplicat pour la protéine Hla. Les moyennes sont présentées sous la forme de heatmaps (figure 5.3).

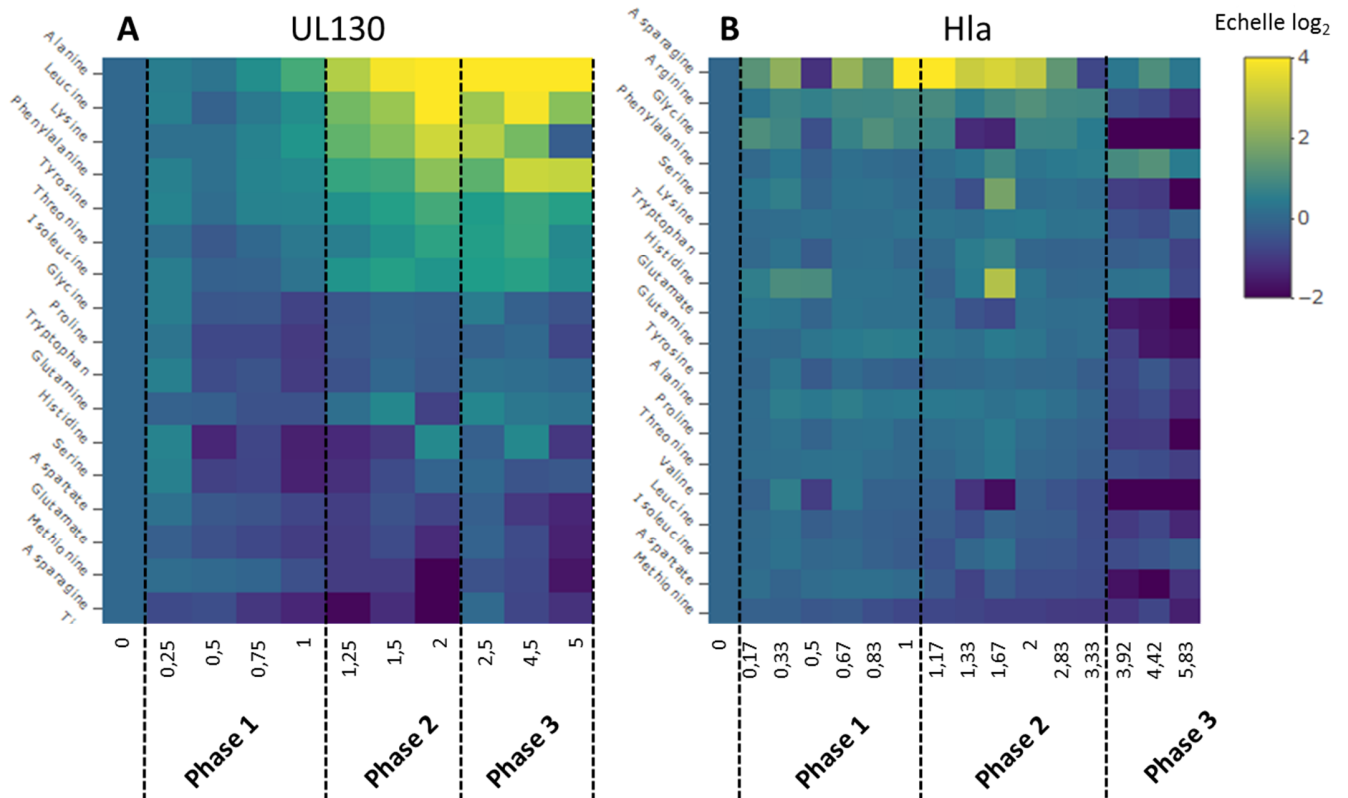


Figure 5.3 : Evolution des concentrations en acides aminés au cours des procédés de production des protéines UL130 (A) et Hla (B). Les heatmaps représentent l'évolution des concentrations en acides aminés totaux (intracellulaire et extracellulaire) de façon relative à leur concentration au moment de l'induction (échelle Log₂). Les moyennes d'un triplicat biologique du procédé de production de la protéine UL130 et d'un duplicat biologique du procédé de production de la protéine Hla sont présentées en fonction du temps normalisé par rapport à l'induction (T=0h). Les phases caractéristiques des cinétiques de croissance et de production sont ajoutées (voir tableau 2).

Lors de la production de la protéine UL130, une diminution des concentrations de l'ensemble des acides aminés est observée durant la phase 1 du procédé. Au cours de la phase 2, la concentration de certains acides aminés augmente jusqu'à des concentrations bien supérieures à la concentration d'avant induction (alanine, leucine, lysine, phénylalanine, tyrosine, thréonine). Les concentrations d'un autre groupe d'acides aminés augmente jusqu'à retrouver une concentration similaire par rapport au moment de l'induction (glycine, proline, tryptophane, histidine, sérine) et la concentration d'une dernière partie des acides aminés continue à diminuer par rapport à leur concentration initiale (aspartate, glutamate, méthionine et asparagine).

La quantification des acides aminés lors de la production de la protéine Hla révèle une plus faible variation de leurs concentrations. La concentration de certains acides aminés diminue 30 minutes après induction tel que l'asparagine, la glycine et la valine de façon concomitante avec le ralentissement du CER (figure 5.2). Cependant un retour au niveau basal est rapidement observé. Seule l'asparagine semble s'accumuler significativement durant le procédé. Au cours de la phase 3 du procédé, une diminution de la concentration de l'ensemble des acides aminés est observée. Ce moment est concomitant un arrêt de la production de Hla sans impact sur la croissance.

Durant le procédé de production de la protéine UL130, de faibles concentrations en certains acides aminés sont mesurées lors des changements de phases du procédé. En effet, certains acides aminés (glycine, proline, tryptophane, histidine et sérine) sont quantifiés à de très faibles concentrations à la fin de la phase 1. Ces acides aminés pourraient donc être limitants et conduire à l'arrêt de croissance. D'autres acides aminés sont retrouvés en très faibles concentrations à la fin de la phase 2 (aspartate, glutamate, méthionine et asparagine). Ces acides aminés pourraient être limitants à la fin de cette phase et engendrer l'arrêt de la production après l'arrêt de croissance. L'aspartate, l'asparagine et la méthionine ont comme précurseur commun l'oxaloacétate. La limitation en oxaloacétate pourrait donc apparaître au cours du procédé de production de la protéine UL130.

La comparaison des évolutions des concentrations des acides aminés libres entre le procédé de production de la protéine UL130 et Hla révèle une perturbation bien plus importante des concentrations dans le procédé UL130 qu'au cours du procédé Hla. Ainsi la forte consommation des acides aminés par la production de la protéine UL130 pourrait être à l'origine de la différence de charge métabolique entre ces deux procédés. Au cours des deux premières phases du procédé Hla, une réorganisation des flux de synthèse des acides aminés a dû s'opérer afin de permettre la production de grandes quantités de protéine sans perturbation majeure des concentrations en acides aminés. L'étude de la proportion en acides aminés des séquences primaires des protéines recombinantes permet de connaître les quantités relatives en acides aminés nécessaires à la synthèse des protéines recombinantes et de mettre en évidence l'impact de la production de protéines recombinantes sur les concentrations intracellulaires en acides aminés.

5.3. Comparaison de la proportion en acides aminés des protéines UL130 et Hla avec la proportion en acides aminés des protéines de la biomasse

L'étude de la proportion en acides aminés dans les protéines UL130 et Hla permet d'évaluer l'impact de la production de ces protéines sur le métabolisme des acides aminés. En particulier, cette étude permet de calculer la modification des flux de biosynthèse des acides aminés qui doit s'opérer chez *E. coli* pour être en adéquation avec les cinétiques de croissance et de production de protéines recombinantes observées.

Durant les procédés, la biomasse des bactéries est enrichie de 30 à 50% par les protéines UL130 et Hla. Du fait que la biomasse avant induction n'est pas enrichie en protéine recombinante, l'atteinte de telles productions spécifiques nécessite d'atteindre une vitesse de synthèse de protéine recombinante au moins équivalente à la vitesse de synthèse de la biomasse. Pour faciliter la compréhension et se placer dans des conditions proches de celles observées expérimentalement, il est défini dans le modèle ci-dessous que la vitesse de synthèse des protéines recombinantes est égale à la vitesse de synthèse des protéines intracellulaires de la bactérie.

Lors d'une croissance exponentielle des bactéries, la vitesse de synthèse de chaque acide aminé est corrélée au taux de croissance et à la proportion que représente chaque acide aminé dans les protéines de la biomasse. De plus, la concentration intracellulaire en acides aminés est stable au cours de la croissance avec une vitesse de synthèse égale à leur vitesse de consommation. Dans le cas d'une production de protéines recombinantes, supposons que la vitesse de production de la protéine recombinante est égale à la vitesse de production des protéines de la biomasse. Le flux de consommation de l'ensemble des acides aminés est alors réparti à part égale pour la synthèse de la biomasse et pour celle de la protéine recombinante. Cette répartition du flux de consommation conduit à une diminution par deux du taux de croissance sans modification nécessaire du flux de synthèse des acides aminés. La répartition égale du flux de consommation sera retrouvée pour la consommation de chaque acide aminé dans le cas où la protéine recombinante est composée de la même proportion en acide aminés que les protéines de la biomasse de la bactérie (figure 5.4, cas n°1).

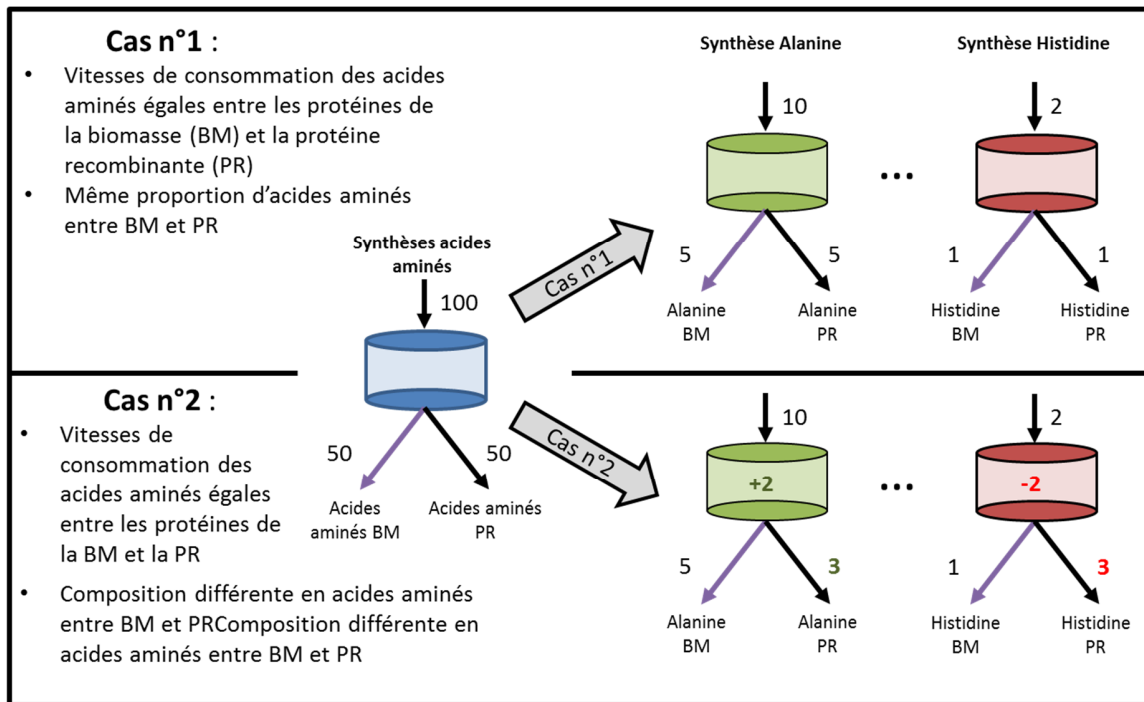


Figure 5.4 : Influence de la composition en acides aminés de la protéine recombinante sur la consommation de chaque acide aminé. Dans cet exemple, la vitesse de production spécifique de la protéine recombinante est égale à la vitesse de synthèse des protéines de la biomasse. Cas n°1 : Une composition en acides aminés identique entre la protéine recombinante (PR) et les protéines de la biomasse (BM) conduit à un partage équitable du flux de synthèse de chaque acide aminé entre les flux de consommation destinés à la synthèse de la biomasse ou à la protéine recombinante. Cas n°2 : si la protéine recombinante possède une composition différente en acides aminés que les protéines de la biomasse alors le flux de synthèse de l'ensemble des acides aminés est équitablement réparti entre les deux flux de consommations. En revanche, la répartition de la vitesse de consommation de chaque acide aminé destinée à la synthèse de la protéine recombinante est modifiée. Ces différences conduisent à l'accumulation des concentrations intracellulaires en certains acides aminés (ex : alanine) ou à la diminution d'autres (ex : histidine).

Si la protéine recombinante possède une composition en acide aminés différente de la biomasse, alors une répartition différente des flux de synthèse de chaque acide aminé est nécessaire pour garantir leur disponibilité (figure 5.4, cas n°2). En effet, si la proportion d'un acide aminé de la protéine recombinante est inférieure à sa proportion dans la biomasse, alors le flux de consommation est inférieur au flux de synthèse de l'acide aminé. Dans ce cas, la concentration intracellulaire de cet acide aminé augmente. Par contre, si la proportion d'un acide aminé de la protéine recombinante est supérieure à sa proportion dans la biomasse alors le flux de consommation total pour cet acide aminé est augmenté et donc le flux de synthèse de cet acide aminé doit être augmenté au risque de devenir

limitant. Plus la différence de proportion des acides aminés entre les protéines de la biomasse et la protéine est importante, plus un réarrangement des flux de synthèse de chaque acide aminé doit s'opérer. Si ce réarrangement n'est pas effectué assez efficacement la concentration intracellulaire de certains acides aminés augmente et la concentration intracellulaire d'autres acides aminés diminue jusqu'à devenir limitant et entraîner un arrêt de croissance.

En comparant la proportion en acides aminés des séquences primaires des protéines UL130 et Hla à la proportion en acides aminés des protéines de la biomasse d'*E. coli* il est possible de déterminer les réarrangements des flux de synthèses des acides aminés à opérer par la bactérie pour produire les protéines recombinantes en conservant des concentrations intracellulaires en acides aminés stables. En effet, la proportion de chaque acide aminé des protéines de la biomasse permet de connaître le flux de synthèse de chaque acide aminé de façon relatif dans la bactérie avant le début de la production, lors d'une croissance exponentielle (Pramanik & Keasling 1997). La moyenne entre la proportion en acides aminés des protéines de la biomasse et la proportion en acides aminés de la protéine recombinante permet de connaître le flux de consommation relatifs à chaque acide aminé lorsque la vitesse de production de la protéine recombinante est égale à la vitesse de production de la biomasse. Ainsi, la comparaison du flux de synthèse de chaque acide aminé de la bactérie avant le début de la production avec le flux de consommation des acides aminés inhérent à une croissance en condition de production de protéine recombinante, indique l'adaptation à opérer pour chaque flux afin d'obtenir un système stable.

$$\text{Modification du flux de synthèse} = - \frac{\text{Flux de synthèse} - \text{Flux de consommation}}{\text{Flux de synthèse}} \times 100$$

Un flux de consommation plus élevé qu'un flux de synthèse de synthèse des acides aminés nécessite une augmentation du flux de synthèse (+X% de la valeur du flux de synthèse) pour conserver une concentration intracellulaire en acide aminé stable. A l'inverse, une vitesse de consommation plus faible que le flux de synthèse nécessite une inhibition du flux de synthèse (-X% de la valeur du flux de synthèse) (figure 5.5).

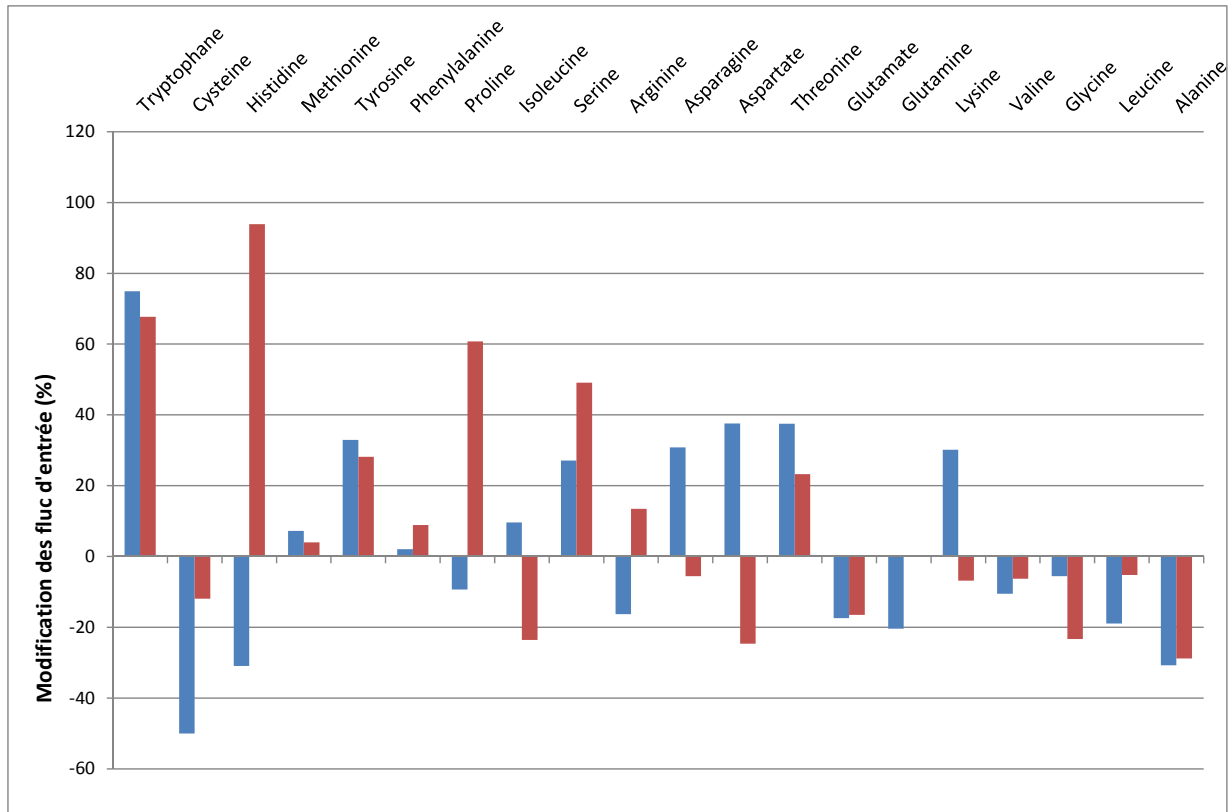


Figure 5.5 : Calcul de la modification des flux de synthèses des acides aminés à partir de la différence de composition en acides aminés entre la protéine UL130 (rouge) ou Hla (bleu) et les protéines de la biomasse pour une vitesse de production des protéines recombinantes égale à la vitesse de production de la biomasse. Une valeur positive indique une augmentation du flux de synthèse à réaliser et une valeur négative une inhibition du flux de synthèse.

La production de la protéine UL130 nécessite une augmentation importante (supérieure à 45%) des flux de synthèse du tryptophane, de l’histidine, de la proline et de la sérine et une augmentation plus modérée des flux de synthèse de la tyrosine et de la thréonine (inférieur à 45%) (figure 5.5). Sauf pour la tyrosine, la quantification par spectrométrie de masse de ces acides aminés à révéler une forte diminution de leur concentration lors de la phase 1 du procédé (figure 5.3). L’analyse de la séquence de Hla met en évidence une demande importante en tryptophane pour la synthèse de la protéine et une demande plus faible de tyrosine, de sérine, d’asparagine, d’aspartate, de thréonine et de lysine (figure 5.5).

Le réarrangement des flux de biosynthèse des acides aminés nécessaire pour une production identique de la protéine UL130 et de la protéine Hla est en moyenne plus important pour la production

de la protéine UL130 à cause de grands écarts entre la proportion d'acides aminés de cette protéine et de la biomasse (moyennes : 42 ± 26 % d'augmentation de flux, 15 ± 8 % de diminution de flux). Au contraire, le réarrangement des flux de synthèse pour la production de Hla est plus faible (moyennes : 29 ± 15.7 % d'augmentation de flux, 21 ± 9 % de diminution de flux).

La quantification des acides aminés au cours des procédés de production des protéines UL130 et Hla a mis en évidence une diminution des concentrations en acides aminés beaucoup plus importante lors de la synthèse de la protéine UL130 que lors de la synthèse de la protéine Hla. L'analyse des séquences primaires de ces deux protéines confirme que la synthèse de la protéine UL130 est susceptible d'engendrer une charge métabolique plus importante au niveau de la consommation des acides aminés que la protéine Hla du fait des différences de proportions en acides aminés des séquences primaires.

5.4. Quantification du ppGpp au cours des procédés de production

Une diminution significative des concentrations intracellulaires des acides aminés a été observée au cours du procédé de production d'UL130 (figure 5.3). La réponse stringente est un mécanisme d'adaptation activé lors de la détection de carences nutritionnelles dans la bactérie (Hauryliuk 2015). En particulier, une carence en acide aminé déclenche la synthèse de ppGpp par la protéine RelA (Svitils et al. 1993). Dans le but de vérifier si la réponse stringente est activée, le ppGpp est quantifié durant les procédés de production de la protéine UL130 et Hla. La réponse stringente peut être activée par la détection de nombreux autres types de carences (carbone, azote, phosphate, fer, oxygène) (Potrykus & Cashel 2008). L'accumulation de ppGpp au cours des procédés de production ne permet pas de démontrer qu'une limitation en acide aminé est présente durant les procédés. Cependant, l'accumulation de ppGpp intracellulaire peut influencer fortement les cinétiques de croissance et de production des protéines recombinantes car ce métabolite est crucial dans les mécanismes d'inhibition de la croissance, dans la régulation de la synthèse des ribosomes et celle des voies de biosynthèse des acides aminés (Liu et al. 2015). Sa quantification est donc essentielle à la détermination des contraintes métaboliques liées à la production de protéines recombinantes (figure 5.6).

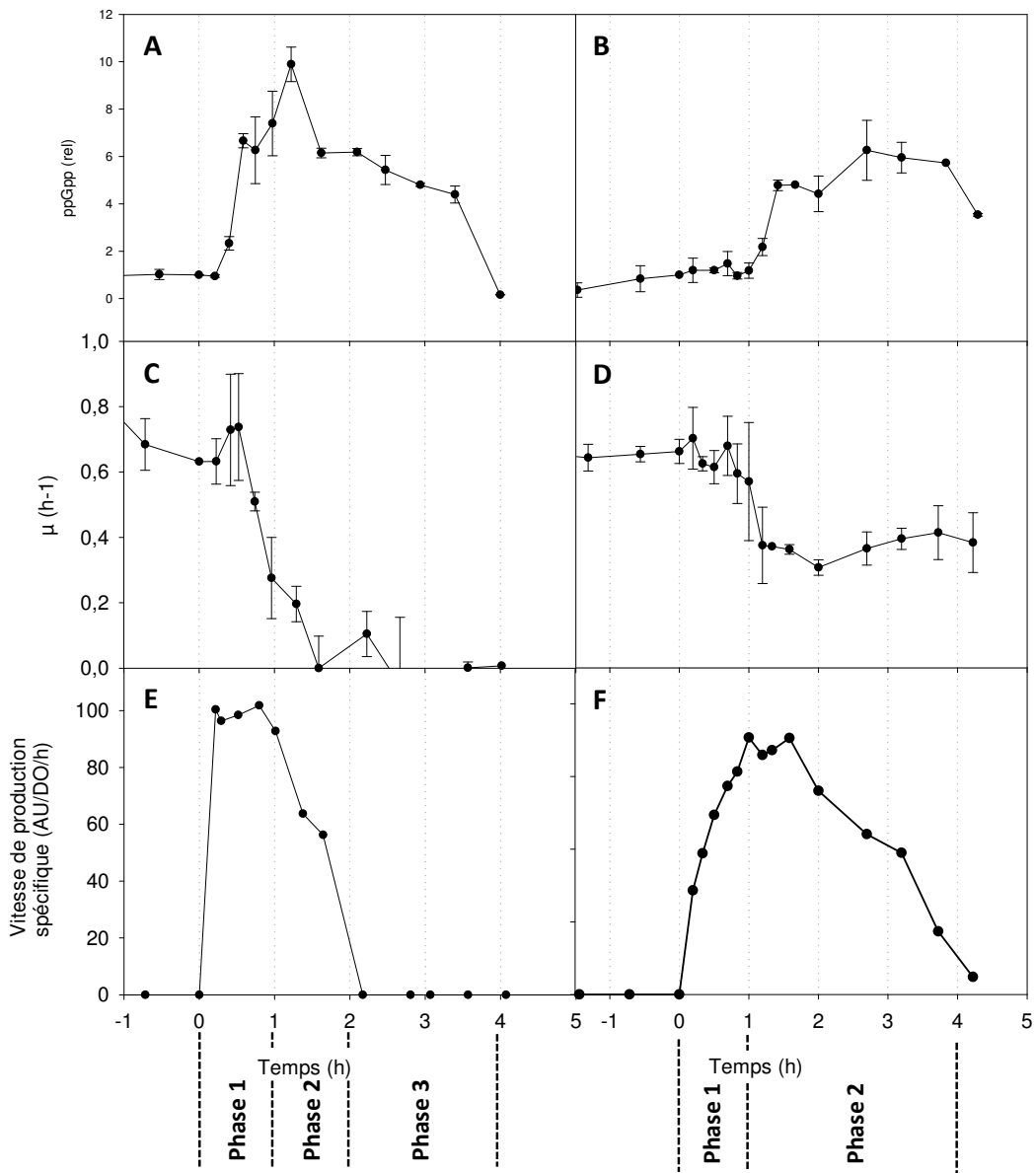


Figure 5.6 : Quantification du ppGpp au cours des procédés des productions des protéines UL130 (A) et Hla (B). Le ppGpp est quantifié par couplage entre une chromatographie ionique et un spectromètre de masse relativement à sa concentration avant induction. Taux de croissance (h^{-1}) au cours du procédé de production des protéines UL130 (C) et Hla (D). Vitesse de production spécifique (AU/DO/h) des protéines recombinantes au cours du procédé de production des protéines UL130 (E) et Hla (F). Les phases caractéristiques des cinétiques de croissances et de production sont ajoutées (voir tableau 2).

Les quantifications du ppGpp réalisées au cours des procédés de production de la protéine UL130 et Hla révèlent une augmentation de la concentration intracellulaire du nucléotide dans les deux conditions.

Pour le procédé de production de la protéine UL130, la concentration maximale en ppGpp est atteinte à la fin de la phase 1 du procédé et représente 11 fois la concentration détectée avant l'ajout d'IPTG (figure 5.6, A). Cette augmentation du ppGpp durant la phase 1 est concomitante à la diminution du taux de croissance et à l'augmentation de la vitesse de production spécifique (figure 5.6, A - C et E). La concentration en ppGpp diminue ensuite et se stabilise durant la phase 2 à un niveau correspondant à 6 fois sa concentration avant induction. Une diminution similaire est retrouvée avec les mesures de vitesse de production spécifique (figure 5.6, A et E). La concentration en ppGpp diminue faiblement pendant toute la phase 3 puis retourne à un niveau basal à la fin de la phase 3.

Pour le procédé de production de la protéine Hla, la concentration en ppGpp n'évolue pas de manière significative durant la phase 1 puis augmente au début de la phase 2 (figure 5.6, B). Durant toute cette phase de forte production, la concentration en ppGpp se maintient à un niveau 5 à 6 fois plus élevée par rapport à son niveau basal. Il est intéressant de constater que le taux de croissance diminue de façon simultanée avec l'augmentation de la concentration en ppGpp (figure 5.6, B et D). De plus, une tendance similaire est retrouvée entre la concentration en ppGpp et la vitesse de production spécifique de la protéine Hla (figure 5.6, B et F). Lors de la production de la protéine UL130, la quantité de ppGpp atteint un maximum 2 fois plus important que lors de la production de la protéine Hla (figure 5.6, A et B). Ce plus fort niveau de ppGpp est cohérent avec la plus forte charge métabolique observée au cours du procédé de production de la protéine UL130 par rapport à celle du procédé de production de la protéine Hla.

L'inhibition de la croissance par le ppGpp est un fait établi(e) dans la littérature (Srivatsan & Wang 2008; Mechold et al. 2013; Liu et al. 2015; Nazir & Harinarayanan 2016). L'augmentation de la concentration en ppGpp durant le procédé de production des protéines UL130 et Hla engendre probablement une diminution du taux de croissance. Du fait que la concentration maximale en ppGpp soit deux fois plus importante au cours du procédé de production de la protéine UL130 que lors de la production de la protéine Hla l'accumulation de ppGpp semble être un marqueur de la charge

métabolique de la production. Il est très probable que la synthèse de ppGpp soit causée par une baisse de la disponibilité en acides aminés dans le procédé de production de la protéine UL130.

5.5. Discussion

Dans ce chapitre les résultats de quantifications des contraintes métaboliques obtenues au cours des procédés de production des protéines UL130 et Hla ont été présentés. Les différentes phases des procédés définies par les cinétiques de croissance et de production sont également caractérisées par les cinétiques observées à partir des analyses fonctionnelles du métabolisme. L'analyse fonctionnelle du métabolisme à chaque phase des procédés permet donc de mettre en évidence les perturbations majeures du métabolisme liées aux cinétiques de croissances et de production. Les perturbations sont caractérisées par une excrétion importante de métabolites et de certains acides aminés dans le surnageant, une diminution des concentrations intracellulaires d'autres acides aminés et l'activation de la réponse stringente.

La production en quantité importante de la protéine UL130 perturbe fortement le métabolisme de la souche conduisant à son arrêt de croissance. Cette période de production et d'arrêt de la croissance entraîne une excrétion de certains acides aminés (acides aminés branchés, alanine, tyrosine, phénylalanine) ainsi que de leur précurseur (pyruvate et préphénate). Ce type d'excrétion en condition de production n'est pas présenté dans la littérature et n'est pas attendu car la synthèse en quantités élevées de protéines recombinantes conduit nécessairement à une consommation importante d'acides aminés. Cependant, l'excrétion de certains acides aminés est observée de façon simultanée à la diminution de la concentration intracellulaire d'autres. La comparaison des proportions entre les acides aminés composants les protéines recombinantes UL130 et Hla à la proportion en acides aminés des protéines de la biomasse a permis d'expliquer ce phénomène. La production de protéines recombinantes en plus de la synthèse des protéines de la biomasse conduit à une demande en acides aminés différente à celle présente avant le déclenchement de la production. Le temps nécessaire à la réorganisation des flux de synthèse de chaque acide aminé par rapport à la nouvelle demande créée par la synthèse de la protéine recombinante peut entraîner des phénomènes d'augmentation et de diminution des concentrations intracellulaires en acides aminés. La consommation importante d'acides

aminés pour la production de protéine recombinante entraîne également une consommation des précurseurs métaboliques des voies de synthèse des acides aminés. Les précurseurs des acides aminés étant des intermédiaires du métabolisme central, une diminution de leur disponibilité peut entraîner des perturbations majeures du métabolisme. En particulier, les intermédiaires du cycle de Krebs tels que l' α -cétoglutarate et l'oxaloacétate sont des précurseurs importants dans la synthèse d'acides aminés. Les quantifications de l' α -cétoglutarate effectuées dans les surnageants de culture indiquent que ce métabolite ne serait pas limitant pour la synthèse de glutamate, de glutamine ou de proline. L'oxaloacétate est le précurseur de l'aspartate, l'asparagine, la méthionine et la thréonine. La concentration intracellulaire de ces acides aminés diminue lors de la production de la protéine UL130 indiquant que l'oxaloacétate pourrait être limitant pour leur synthèse. Ce manque de disponibilité de l'oxaloacétate signifierait que l'activité des voies anaplérotiques est insuffisante pour compenser sa consommation conduisant à une perturbation du fonctionnement du cycle de Krebs. Cette perturbation pourrait expliquer l'excrétion importante de pyruvate et d'acétate observée dans le surnageant mais également le découplage entre le catabolisme et l'anabolisme révélé par la modification du coefficient respiratoire ($QR = CER/OUR$).

La synthèse de protéines entraîne la consommation d'acides aminés et d'énergie. Les acides aminés sont des métabolites clés dans la détection des limitations pour la bactérie (Scott et al. 2014; Traxler, Summers, H.-T. Nguyen, et al. 2008; Potrykus & Cashel 2008). La synthèse de ces métabolites nécessite l'utilisation d'une grande diversité de composés (carbone, azote, soufre, fer pour certaines enzymes impliquées dans leur synthèse). Un manque de disponibilité de ces composés peut donc se traduire par une carence en acides aminés. Il est alors cohérent de constater que l'adaptation des bactéries par rapport aux carences nutritionnelles est effectuée en partie par synthèse de (p)ppGpp lors de la détection des ARNt non chargés. La quantification du ppGpp en condition de production de protéines recombinantes a permis d'observer son accumulation en lien avec les phases caractéristiques des procédés. Dans le cadre de cette étude, la concentration en ppGpp intracellulaire varie de manière similaire à la vitesse de production spécifique et au taux de croissance durant les procédés de production des protéines UL130 et Hla. Ces observations ouvrent de nombreuses questions sur l'implication de la réponse stringente durant la période de production. L'étude de son

influence sur les cinétiques de croissance et de production semble donc nécessaire à la compréhension des dynamiques observées.

De façon surprenante la production spécifique de la protéine UL130 en milieu minimum est similaire, voir supérieure, à la production spécifique de la protéine en milieu riche (figure 4.6). Ce résultat signifie que la disponibilité intracellulaire des acides aminés est similaire lorsqu'ils sont assimilés à partir du milieu extérieur ou lorsqu'ils sont synthétisés par la bactérie. Pourtant, l'utilisation d'un milieu riche rend accessible une grande quantité d'acides aminés pour les bactéries. Dans ces conditions de culture, les voies de biosynthèse des acides aminés sont inhibées. A l'inverse, un milieu minimum engendre l'activation des voies de biosynthèses des acides aminés car les acides aminés ne sont pas disponibles dans le milieu de culture. L'observation de la diminution des concentrations intracellulaires en acides aminés lors de la production en milieu minimum témoigne d'une diminution de leur disponibilité intracellulaire pour la production et la croissance. Sur la base des résultats observés, une stratégie d'amélioration du procédé a été établie. Cette stratégie consiste à débiter le procédé en milieu minimum afin d'obliger la bactérie à activer les voies de biosynthèse des acides aminés. Au moment de l'induction, une supplémentation en acides aminés dans le milieu de culture permettrait d'améliorer la disponibilité intracellulaire des acides aminés par leur assimilation à partir du milieu de culture. Il est attendu que cette supplémentation ne cause pas d'inhibition des voies de biosynthèses des acides aminés car les niveaux intracellulaires d'acides aminés diminuent au cours de la production. Cette méthode complémenterait la biosynthèse des acides aminés par leur assimilation du milieu et augmenterait la disponibilité intracellulaire des acides aminés par rapport à un procédé en milieu riche ou un procédé en milieu minimum. Cette hypothèse sera testée au cours du chapitre 7.

Chapitre 6 – Etude de l'influence de la réponse stringente sur la souche *E. coli* BL21 en condition de production de protéines recombinantes

La réponse stringente est activée au cours des procédés de production des protéines UL130 et Hla (chapitre 4). La quantification du ppGpp a permis d'observer une évolution de sa concentration intracellulaire proche de l'évolution de la vitesse de production spécifique et inverse à l'évolution du taux de croissance au cours des procédés. La réponse stringente est un mécanisme connu depuis les années 1970 et a été depuis fortement étudiée. Son rôle dans l'adaptation globale du fonctionnement bactérien lors d'une modification nutritionnelle ou chimique de l'environnement a été démontré (Svitils et al. 1993; Potrykus & Cashel 2008; Nazir & Harinarayanan 2016). En particulier l'implication de la réponse stringente dans le contrôle du taux de croissance, de la transcription des ARNt, des ARN ribosomiaux (ARNr) et des voies de synthèse des acides aminés ainsi que les résultats obtenus durant l'analyse des procédés de production ont orienté les investigations vers l'étude de l'influence de la réponse stringente sur la souche en condition de production. Ce type d'étude a été très peu entrepris depuis ces 50 dernières années (Bachinger et al. 2001; Grabherr et al. 2002). Ce n'est que plus récemment que des modélisations ont été effectuées mettant en évidence l'implication de la réponse stringente dans l'adaptation d'*E. coli* à la production de protéines recombinantes (Shachrai et al. 2010; Scott et al. 2010). L'objectif ici est d'étudier l'influence de la réponse stringente sur la croissance, la production et plus généralement sur le métabolisme de la souche *E. coli* BL21 au cours des procédés de production des protéines UL130 et Hla.

La production d'une protéine recombinante est une perturbation évoluant de façon dynamique au cours du temps. Ainsi, le lien entre la cinétique de production et l'évolution de la concentration intracellulaire en ppGpp rend complexe l'étude de l'influence de chaque mécanisme de façon indépendante sur la souche *E. coli* BL21. C'est pourquoi, la première partie de ce chapitre a pour objectif d'étudier l'influence de la réponse stringente sur la croissance et le métabolisme bactérien lors d'une perturbation ponctuelle de façon indépendante de la production afin de déterminer les caractéristiques de la réponse stringente de la souche *E. coli* BL21 dans un modèle d'étude simplifié. Ce

modèle d'étude repose sur la simulation d'une carence en acides aminés par ajout de sérine hydroxamate (SHX) dans le milieu de culture. Cet analogue de la sérine est la molécule la plus utilisée dans la bibliographie pour activer la synthèse de ppGpp par la protéine RelA (Riesenberg et al. 1984; Bougdour & Gottesman 2007; Traxler et al. 2008; Durfee, A. M. Hansen, et al. 2008; Samarrai et al. 2011; Kolmsee et al. 2011; Hauryliuk 2015). La synthèse de ppGpp serait causée par une inhibition de la seryl-transférase empêchant le chargement de la sérine sur les ARNt^{ser} et activant la protéine RelA. Dans la littérature, l'ajout de SHX est effectué le plus souvent sur des cultures de la souche *E. coli* K-12. C'est pourquoi, dans un premier temps, le modèle d'étude est mis en place avec cette souche d'*E. coli*. L'utilisation de cette souche permet de comparer les résultats obtenus avec ceux présentés dans la littérature. De plus, la souche K-12 est couramment utilisée comme hôte pour la production de protéines recombinantes (Özkan et al. 2005; Graumann & Premstaller 2006; Huang et al. 2012; Mahalik et al. 2014). L'influence de la réponse stringente dans la résilience à la perturbation déclenchée par ajout de SHX permet de caractériser les spécificités des souches d'*E. coli* BL21 et K-12 et, dans ce contexte, de comparer leur capacité d'adaptation. Une modulation du niveau de ppGpp intracellulaire des souches est effectuée par suppression des gènes codant pour les protéines impliquées dans la synthèse de ppGpp (protéine RelA) ou dans la transformation du pppGpp en ppGpp (protéine GppA) pour générer les mutants $\Delta relA$ et $\Delta gppA$. L'absence de transformation du pppGpp en ppGpp par suppression de la protéine GppA conduit à une diminution de la quantité de ppGpp synthétisé au moment de l'activation de la réponse stringente (Mechold et al. 2013). L'étude est présentée sous forme de publication en vue de sa soumission à un éditeur.

Dans le but d'approfondir l'étude de l'influence de la réponse stringente sur le métabolisme bactérien, la première partie de chapitre est complétée par une quantification des acides aminés libres, réalisée après ajout de SHX. Les concentrations en acides aminés obtenues dans ce contexte d'arrêt de croissance sans production sont étudiées puis comparées aux concentrations en acides aminés quantifiées lors d'un arrêt de croissance en condition de production de la protéine UL130. Cette comparaison permet de distinguer l'effet de l'arrêt de croissance d'une part et l'effet de l'arrêt de croissance en condition de production d'autre part sur l'évolution des concentrations en acides aminés.

La caractérisation de l'influence de la réponse stringente sur la croissance et le métabolisme de la souche BL21 permet dans une seconde partie du chapitre d'étudier son influence en condition de production des protéines UL130 et Hla. Cette étude est réalisée avec la souche *E. coli* BL21 sauvage (WT) et les souches *E. coli* BL21 mutées $\Delta gppA$, $\Delta relA$ et le double mutant $\Delta gppA : \Delta relA$.

Dans la seconde partie du chapitre, le nombre de copies de plasmides par bactérie (PCN) est mesuré au cours du procédé de production de la protéine UL130. Cette quantification est effectuée sur des procédés de production en milieu riche et en milieu minimum afin d'étudier l'influence de la charge métabolique (caractérisée comme étant supérieure en milieu minimum) sur le nombre de copies de plasmide par bactérie. De façon complémentaire, des quantifications du nombre de copies de plasmide par bactérie sont effectuées avec la souche *E. coli* BL21 $\Delta relA$ dans les mêmes conditions afin d'étudier l'influence de la réponse stringente sur l'évolution du PCN.

Ce chapitre se termine sur la quantification de l'exométabolome par RMN de la souche *E. coli* BL21 et des mutants au cours de la production de protéine. Préalablement, l'analyse de l'exométabolome des souches en condition non induite par RMN permet de fixer un cadre de référence à l'évaluation de l'impact de la production sur le métabolisme.

6.1. Etude de l'influence de la réponse stringente sur la souche *E. coli* BL21 sans production de protéines recombinantes.

6.1.1. Crucial role of ppGpp – and not pppGpp – on resilience capability of *Escherichia coli* to the disruption in amino acids homeostasis

INTRODUCTION

Cellular metabolism is composed of thousands of reactions acting in concert to allow cells to grow and survive in changing environments. For autotroph or heterotroph organisms able to utilize nitrogen mineral sources such as nitrate and ammonium, a significant part of the cellular metabolism is dedicated to the biosynthesis of the diverse amino acids used as building blocks of proteins. In the bacterium *Escherichia coli*, protein content is roughly 60 % (w/w) when the bacterium is grown on mineral medium and the amino acids metabolism represents 20 % of the total proteome (Bremer & Dennis 2008; Scott et al. 2014; Bosdriesz et al. 2015). In balanced growth condition, metabolic fluxes through amino acids biosynthetic pathways are finely regulated to perfectly supply the translational machinery with each required amino acid. This homeostatic behavior is highly robust as long as no environmental or nutritional changes have occurred.

In response to nitrogen starvation, bacteria adapt to the nutritional stress by rapidly modulating the expression of hundreds of gene, *via* the action of (p)ppGpp. The latter nucleotides act as second messengers in the so-called stringent response to massively reprogram cellular physiology from rapid growth to survival and adaptation (Potrykus & Cashel 2008; Srivatsan & Wang 2008), along with other important roles in the regulation of bacterial virulence (Hüttener et al. 2018), survival during host invasion, antibiotic resistance and persistence (Wu et al. 2015; Verstraeten et al. 2015; Piskunova et al. 2017). Nitrogen starved bacterial cells rapidly accumulate both nucleotides, ppGpp and pppGpp (Villadsen & Michelsen 1977; Brown et al. 2014; Iyer et al. 2018). Intracellular levels of (p)ppGpp are modulated by RSH (Rel Spo Homolog) enzymes, whose names derive from the RelA and SpoT proteins of *Escherichia coli* in which (p)ppGpp was originally detected (Cashel et al. 1996). In addition to the (p)ppGpp synthetase relA and the (p)ppGpp synthetase/hydrolase SpoT, the pppGpp

pyrophosphatase GppA also contribute in *E. coli* to modulate the (p)ppGpp intracellular levels by converting pppGpp into ppGpp. To date, the physiological role of pppGpp remains elusive especially as it emerges from literature that pppGpp is a less potent regulator than ppGpp of the stringent response in *E. coli* (DeNapoli et al. 2013; Mechold et al. 2013). It is well established that basal levels of (p)ppGpp control growth by modulating the number of ribosomes (Potrykus and Cashel, 2008). How quasi-immediate inhibition of growth and protein synthesis (Cashel et al. 1996; Durfee, A. Hansen, et al. 2008) by high levels of ppGpp remains in contrast highly elusive although dozens of (p)ppGpp binding proteins have been recently identified (Zhang et al. 2018) that might contribute to slow down the growth and reduce translational activity.

The response to amino-acid insult is essentially studied at the molecular level and are regularly reviewed and discussed (Carneiro et al. 2011; Dalebroux & Swanson 2012; Kanjee et al. 2012). In this study, we addressed i) the ability of *E. coli* to cope with a failure in the amino acid homeostasis at the cellular level and, ii) the contribution of (p)ppGpp to the adaptive capacity of *E. coli* to face such a disruptive event. To achieve this, we take benefit of the serine hydroxamate (SHX), an analogue of serine, to induce amino acid starvation (Cashel et al. 1996). As a consequence of the presumed inhibition of serine binding to the seryl-tRNA synthetase (Tosa & Pizer 1971; Weygand-durasevic et al. 1993; Zeng et al. 2009), SHX addition causes RelA-dependant (p)ppGpp accumulation and growth arrest (Tosa 1971; Weygand-durasevic et al. 1993; Svitils et al. 1993; Zeng et al. 2009; Li et al. 2016; Ross et al. 2017). The response to homeostasis disruption by SHX addition were dissected by analyzing the dynamics of the growth rate and measuring the intracellular levels of ppGpp and pppGpp in perturbation experiments using the model organism *E. coli* K-12 WT strain and $\Delta relA$ and $\Delta gppA$ mutants. The study was then extended to the *E. coli* BL21 strain (WT strain and $\Delta relA$ or $\Delta gppA$ mutants) which is a widely used host for recombinant protein overproduction among other biotechnological products. In this study, we obtained clear evidence for the resilience of *E. coli* to disruption of amino acid homeostasis and demonstrated the key role played by ppGpp – and not pppGpp – on the ability of *E. coli* to recover its full growth capacity.

EXPERIMENTAL SECTION

Chemicals and reagents:

DL-Serine Hydroxamate (SHX) was purchased from Sigma Aldrich (St. Quentin-Fallavier, France). LC-MS grade solvents (methanol, acetonitrile) were obtained from Instrumentation Consommables et Service (ICS, Lapeyrouse-Fossat, France).

Bacterial strains and growth conditions

Strains were derived from *Escherichia coli* K-12 strain BW25113 from the Keio collection, and from *E. coli* BL21 (DE3) strain and were listed in Supplementary Table 1. Cells were cultured on minimal synthetic medium containing 5 mM KH_2PO_4 , 10 mM Na_2HPO_4 , 9 mM NaCl, 40 mM NH_4Cl , 0.8 mM MgSO_4 , 0.1 mM CaCl_2 , and 30 g/L glucose. Thiamine was added to a final concentration of 0.1 g/L to cultivate the *E. coli* BW25113 WT and derivative strains. Glucose, thiamine and MgSO_4 were sterilized by filtration (Minisart polyamide 0.2 μm , Sartorius, Göttingen, Germany), and other solutions were autoclaved separately. All stock cultures were stored at -80°C in Luria-Bertani (LB) medium containing glycerol (40 %, v/v). For the cultures, 5 mL of overnight cultures in LB were used as inoculum and then sub-cultured in shake-flasks containing 50 mL of minimum medium with 3 g/L glucose starting at $\text{OD}_{600\text{nm}} = 0.05$ and incubated at 30°C and 210 rpm in an orbital shaker (Inova 4230, New Brunswick Scientific, New Brunswick, NJ, USA). Cells were harvested during the exponential growth phase by centrifugation for 10 min at 10,000 g at room temperature with a Sigma 3-18K centrifuge (Sigma Aldrich, Seelze, Germany), washed with the same volume of fresh medium (without glucose or thiamine), and used to inoculate 500 mL bioreactors (Multifors, Infors HT, Bottmingen, Switzerland) containing 300 ml of minimal medium with 30 g/L glucose at $\text{OD}_{600\text{nm}} = 0.15$. The temperature was set to 37°C and the pH was maintained at 7 by automatically adding 14% (g/g) ammonia (VWR, Fontenay-sous-Bois, France) or 11 % (g/g) phosphoric acid (Sigma Aldrich, St. Quentin-Fallavier, France). Aeration and the stirrer speed were controlled to maintain adequate aeration ($\text{DOT} > 30\%$ saturation). Cell growth was monitored through the optical density at 600 nm with a Genesys 6 spectrophotometer (Thermo, Carlsbad, CA, USA). Simultaneously, the metabolic activity of the cultures was estimated indirectly by observing the Oxygen Uptake Rate (OUR) and Carbon Emission Rate (CER)

which was measured by a Dycor ProLine process mass spectrometers (Ametek, Berwyn, PA, USA). Stringent response was triggered by adding SHX to the culture when the OD reached 3.5.

Calculation of an approximate instantaneous growth rate

To determine the instantaneous growth rate, the specific absorbance rate (μ) was calculated with the following formula according to numerical differentiation method:

$$\mu(t_n) = \left(\frac{1}{y_n}\right)(y_{n+1} - y_{n-1}) / (t_{n+1} - t_n) \quad (\text{eq. 1})$$

where $\mu(t_n)$ is the specific growth rate (per hour) at time t_n and y_n is the measured absorbance at t_n .

Sampling and (p)ppGpp extraction

Culture medium (400 μ L) was withdrawn from the bioreactor and vigorously mixed with 4.5 mL of a pre-cooled acetonitrile/methanol/H₂O (4:4:2) solution at -40°C to rapidly quench metabolic activity (Millard et al. 2014). Immediately thereafter, 100 μ L of ¹³C labeled metabolites were added to the latter mixture as internal standards. The tubes were then placed in a cooling bath of ethanol pre-cooled at -40°C and evaporated to dryness in a SpeedVac (SC110A SpeedVac Plus, ThermoSavant, Waltham, MA, USA) under vacuum for 4 h and then stored at -80°C until needed.

IC-ESI-HRMS quantification of (p)ppGpp

After resuspension of the cell extract samples in 20 mM ammonium acetate buffer at pH 9 (Sigma Aldrich, St. Quentin-Fallavier, France) to a final volume of 500 μ L, cell debris were removed by centrifugation at 10,000 g for 10 min at 4°C. The samples were then analyzed by ion chromatography (Thermo Scientific Dionex ICS-5000+ system, Dionex, Sunnyvale, CA, USA) coupled to a LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) equipped with an electrospray ionization probe. The ion chromatography method is a variant of the method described by Kiefer et al. (Kiefer et al. 2007) to improve the chromatographic separation of ppGpp and pppGpp. The KOH gradient was modified as follows: 0 min, 0.5 mM; 5 min, 0.5 mM; 10 min, 10 mM; 15 min, 20 mM; 26 min, 20 mM; 31 min, 100 mM; 38 min, 100 mM; 38 min, 0.5 mM; 43 min, 0.5 mM. The gradients were all linear. In all cases, the volume of the injected sample was 15 μ L. Mass spectrometry analysis was performed in the negative FTMS mode at a resolution of 30,000 (at $m/z = 400$) in fullscan mode, with

the following source parameters: capillary temperature, 350°C; source heater temperature, 300°C; sheath gas flow rate, 50 a.u. (arbitrary unit); auxiliary gas flow rate, 5 a.u.; S-Lens RF level, 60%; and ion spray voltage, 3.5 kV. The data were acquired using the Xcalibur software (Thermo Fisher Scientific, Waltham, MA, USA). Three samples from three independent biological replicates were analyzed.

NMR analysis of culture supernatant

Exocellular metabolites were identified and quantified by nuclear magnetic resonance (NMR). Broth samples were collected at different times and filtered (Minisart 0.2 µM filter from Sartorius, Göttingen, Germany). The supernatants, corresponding to the culture medium, were mixed with 100 µL of D₂O containing 2.35 g/L of TSPd4 as the standard (3-(trimethylsilyl)-1-propanesulfonic acid-tetra deuterated). Proton-NMR spectra were recorded on an Avance III 800 MHz NMR spectrometer equipped with a 5 mM QCI-P cryo probe (Bruker, Rheinstatten, Germany). Quantitative ¹H-NMR was performed at 280 K, using a 30° pulse and a relaxation delay of 10 s. Acquisition of Heteronuclear Single Quantum Coherence (HSQC) 2D ¹H–¹³C HSQC spectra were performed to quantify SHX in supernatant. Scans (16) were acquired at 280K with 4096 x 128 points on 13.35 x 60 ppm windows. Spectral processing and metabolite quantification were performed using Topspin 3.1 (Bruker, Rheinstatten, Germany). Extracellular metabolites from three independent biological replicates were analyzed.

RESULT SECTION

Dynamic response characteristics to SHX addition.

Detailed macroscopic characterization of serine hydroxamate (SHX) addition was obtained from *Escherichia coli* K-12 BW25113 cells exponentially growing in M9 mineral medium (Enjalbert et al., 2015) containing 166mM glucose (30g.l⁻¹) as sole carbon source. Growth were performed aerobically in bioreactors at 37°C, pH 7, under sufficient oxygenation (DOT>30%). SHX (0.8 mM) was added after 2 hours of exponential growth.

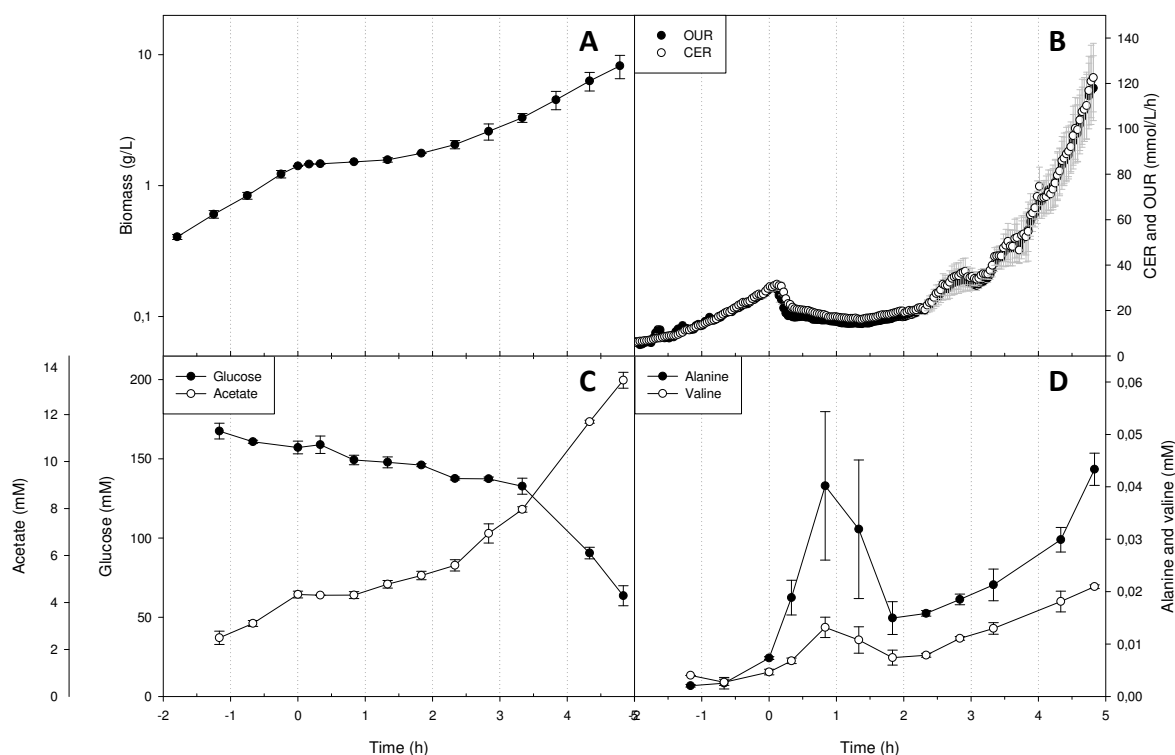


Figure 1. Macroscopic characterization of *E. coli* K-12 BW25113 cells during SHX perturbation. (A) Biomass (g/L), (B) oxygen uptake rate (OUR) and carbon enrichment rate (CER) (mmol/L/h), (C) glucose and acetate in supernatant (mM), (D) alanine and valine in supernatant. Time is normalized in function of SHX addition (0,8mM) at t=0 h.

As expected, SHX addition provokes a severe growth inhibition (figure 1A). Concomitantly, we clearly observe a strong drop of respiratory activity (figure 1B) and a stop of acetate production (figure 1C), reflecting the reduction in metabolic activity. However, growth inhibition and metabolic activity reduction appear to be transient events since about 1.5 hours after SHX addition, we observe an acceleration of growth rate correlated with the increase of respiratory activity and acetate production (figure 1A). Beside acetate, other fermentation products such as formate, ethanol and orotate are detected from the NMR analysis of cultivation medium during the exponential growth phase (data not shown) in agreement with previous published data (Enjalbert et al., 2013). Surprisingly, alanine and

valine accumulate transiently in the cultivation medium, their maximal concentrations being reached 1 hour after SHX.

***Escherichia coli* is resilient to SHX addition.**

Although SHX treatment causes a profound metabolic perturbation, these results reveal that *E. coli* recovers - at least partially - its growth capacity. To better characterize the *E. coli*'s resilience to SHX addition, we then calculated the instantaneous growth rate (figure 2A).

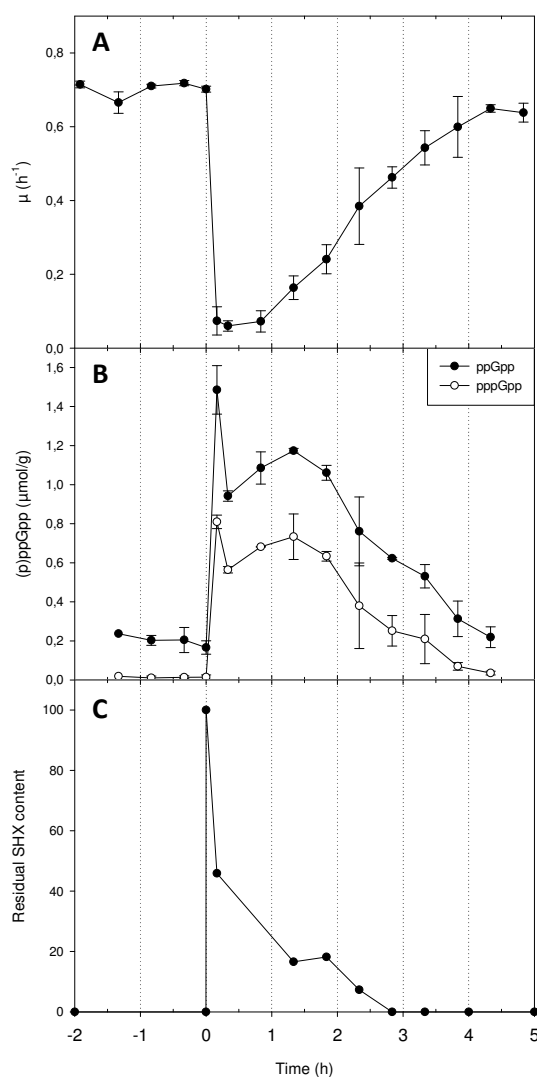


Figure 2. Dynamics of (p)ppGpp intracellular levels and SHX in the medium. (A) Instantaneous growth rate (h^{-1}), (B) ppGpp and pppGpp concentrations ($\mu mol/gCDW$) and (C) residual SHX content in supernatant (%) measured before and after SHX treatment. Time 0 was defined as the time at which SHX (0.8mM) was added in the bioreactor.

The growth rate before SHX addition is $0.68 \pm 0.04 \text{ h}^{-1}$ which is in agreement with published data of *E. coli* K-12 growing in minimal medium (Marisch, Bayer, Scharl, et al. 2013). The addition of SHX leads to a sudden drop of the growth rate, reaching a value close to zero. Afterwards, the growth rate increases constantly until reaching a new exponential regime 4 hours after SHX addition, the growth rate being close to the one measured before SHX addition ($\mu = 0.66 \pm 0.03 \text{ h}^{-1}$). From this curve, we defined 3 metrics to describe the resilience dynamics to SHX addition (Yodo 2018)(Yodo 2018) i) the “robustness”, which refers here to the residual growth rate following the SHX addition, ii) the “recovery rate” which is the slope of the increase in growth rate, and iii) the “recovered state” which corresponds to the new exponential regime reached after completing the recovery process. The values of these metrics estimated from this experiment, are given in table 1.

Table 1. Resilience characteristics of *E. coli* K-12 WT strain to SHX addition.

	SHX (mM)	Robustness (%)	Adaptation rate	Recovery (%)
BW25113 WT	0.8	9.00 ± 3.3	0.17 ± 0.01	95.23 ± 0.37
BW25113 WT	0.4	36.55 ± 7.63	0.64 ± 0.05	112.58 ± 2.09
BW25113 WT	0.2	59.95	0.60	108

Dynamics of (p)ppGpp intracellular levels and SHX in the medium.

We then wonder if the cells recover their growth capacity as a result of stringent response allowing the cells to divide despite of the presence of SHX in the medium. To investigate that question, intracellular content of ppGpp and pppGpp were determined over the cultivation time by LC-MS/MS (Patacq et al. Submitted) and in parallel, the SHX into the cultivation medium was quantified by NMR (figure 2).

Surprisingly, we observed that SHX concentration immediately decreases in the cultivation medium after the serine analog was added, becoming undetectable in the medium after 2.8 h. The disappearance of SHX is clearly biotic since the SHX concentration does not decrease in similar conditions without cells (data not shown). To our knowledge, this has not been reported in literature so far.

As expected, the addition of SHX leads to the intracellular accumulation of ppGpp of which the concentration is measured to be $1.48 \mu\text{mol/gCDW}$ 10 minutes after SHX addition. Then, ppGpp content

decreases and reaches a plateau for approximately 2 hours. In a last phase, it decreases continuously and reaches a value close to the basal value measured during the exponential phase (0.2 $\mu\text{mol/gCDW}$). Interestingly, pppGpp content follows a similar profile, albeit at lower level.

Because both the extracellular SHX content and the intracellular (p)ppGpp diminish during the recovery phase, it is difficult to distinguish between their respective contribution to the restoration of the growth. To do so, we attempted to decouple SHX effect from the stringent response by varying the intensity of SHX disturbance. This was achieved by decreasing the final concentration of SHX added in the culture (figure 3).

The resilience to SHX depends on its added amount.

To investigate the effect of the SHX concentration on the *E. coli*'s resilience, we added two- and four-fold less of SHX (0.4 mM and 0.2 mM respectively). We first observed that the growth is less seriously altered with the exposure to lower concentration of SHX (figure 3A). This is even more obvious looking at the respiratory activity profiles for which the disruptive effect of SHX addition is less pronounced; the respiratory activity being recovered after 2 hours, 30 minutes and 15 minutes after the addition of 0.8, 0.4 and 0.2 mM of SHX (figure 3B).

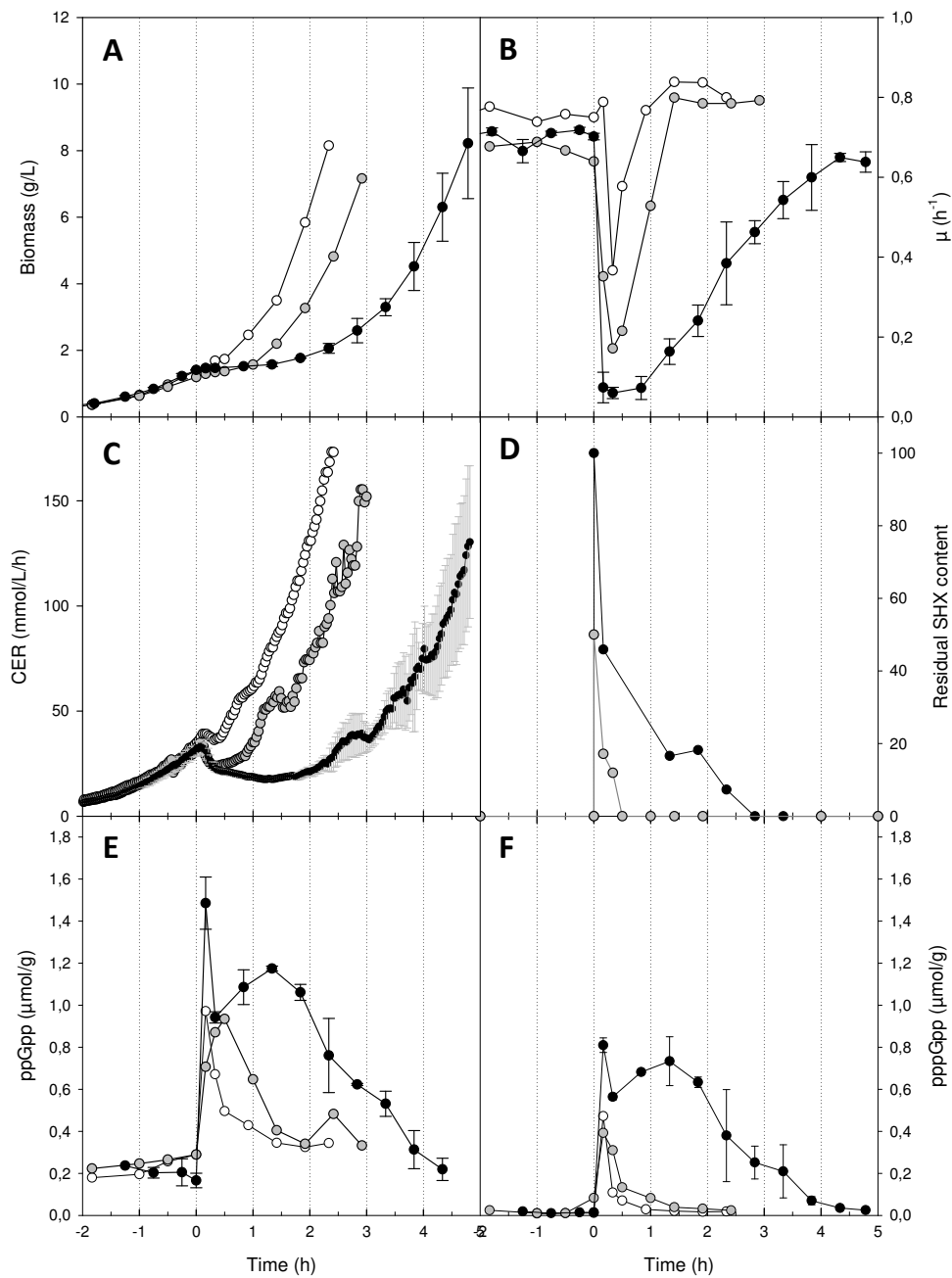


Figure 3. Resilience of *E. coli* to SHX treatment. (A) Biomass (gCDW/L), (B) instantaneous growth rate (h^{-1}), (C) carbon enrichment rate (CER) (mmol/L/h), (D) residual SHX content (%), (E) ppGpp and (F) pppGpp concentrations (μ mol/gCDW) measured before and after SHX treatment with 0.8 mM (black), 0.4 mM (grey) or 0.2 mM (white) of SHX. Time 0 was defined as the time at which SHX was added in the bioreactor.

The SHX in the medium is only detectable few minutes (less than 30 minutes) after its addition when the culture is exposed to 0.4 mM and under the limit of detection when it is exposed to 0.2 mM, showing a rapid disappearance of the serine analog from the cultivation medium. However, these additions are sufficient to induce the stringent response as testified by the (p)ppGpp accumulation (figure 3E and F). The intracellular ppGpp and pppGpp pool dynamics are very similar for both conditions. The intracellular levels of ppGpp and pppGpp first increase (4 -fold and 30 fold respectively) after SHX addition and returns back close to the basal value in 2-3 hours for ppGpp and less than 1 hour for pppGpp. It could however be noticed that the intracellular level of ppGpp returns back to the basal value faster when 0.2mM of SHX are added instead of 0.4mM. More importantly, the intracellular levels of ppGpp and pppGpp show different dynamic responses depending on whether the concentration of SHX added is 0.8 mM or less. The maximum accumulation of both nucleotides is nearly the double after the addition of 0.8 mM SHX than those measured after the addition of 0.2 or 0.4 mM SHX. In the latter experiments, (p)ppGpp intracellular concentrations decrease continuously with no plateau as it is observed when 0.8 mM is added.

Moreover, we detect a dose-response relationship between the robustness and the amount of SHX added (Table 1) meaning that the latter controls the disturbance intensity. However, such a relationship is not found for the recovery rate. This suggests that the type of adaptive response depends on the disturbance intensity. This is fully consistent with the dynamics of (p)ppGpp accumulation showed on figure 3 where two types of (p)ppGpp accumulation profiles are clearly observed whether the amount of SHX added is high (0.8 mM) or low (0.2 or 0.4 mM). Furthermore, the amount of SHX added appears to have no impact on the recovery state since they are similar in any conditions. We also notice that the recovery takes place mainly during the phase where (p)ppGpp is decreasing over the time from 0.33 h to 1 h, from 0.33 h to 1.5 h and from 0.33 to 4 h whether 0.2, 0.4 or 0.8 mM of SHX has been added (Figure 3B). This occurs whether or not SHX is still present in the cultivation medium. These results indicate that restoration of growth is mainly correlated with (p)ppGpp intracellular level and not with the disappearance of SHX from the cultivation medium.

The stringent response is crucial to ensure growth recovery.

To make sure that the growth recovery is a stringent-response associated phenomenon, similar experiments (adding 0.8 mM SHX) were carried out with mutants deleted for stringent-response associated genes, namely a mutant deleted for the gene *relA*, encoding a GTP pyrophosphokinase which catalyzes *via* a ribosome associated mechanism, the formation of ppGpp or pppGpp whether GDP or GTP are respectively the substrate (Haseltine & Block 1973; Mechold et al. 2002) and, a mutant deleted for the gene *gppA*, encoding the pppGpp 5'-gamma phosphate hydrolase which converts pppGpp to ppGpp. The $\Delta gppA$ mutant is known to accumulate high concentrations of pppGpp and ppGpp after SHX addition (DeNapoli et al. 2013) unlike the $\Delta relA$ mutant which does not (Traxler, Summers, H.-T. Nguyen, et al. 2008).

In $\Delta relA$ mutant, we do not detect pppGpp and ppGpp level is lower than the basal level measured in the WT strain (close to the limit of detection). As expected, we do not detect transient accumulation of the alarmone after SHX addition. This confirms that the synthase activity of SpoT is mainly silent in this condition and there is no other RSH is involved in this response in agreement with previous studies (Mechold et al. 2002; Traxler, Summers, H.-T. Nguyen, et al. 2008). Additionally, $\Delta gppA$ mutant accumulates higher intracellular level of pppGpp than ppGpp (Figure 4) in contrast to the WT strain for which the accumulation of ppGpp is twice that of pppGpp. Indeed, the ppGpp intracellular level are rather similar between both strains, but the pppGpp level is 3-fold higher in the $\Delta gppA$ mutant compared to the WT strain.

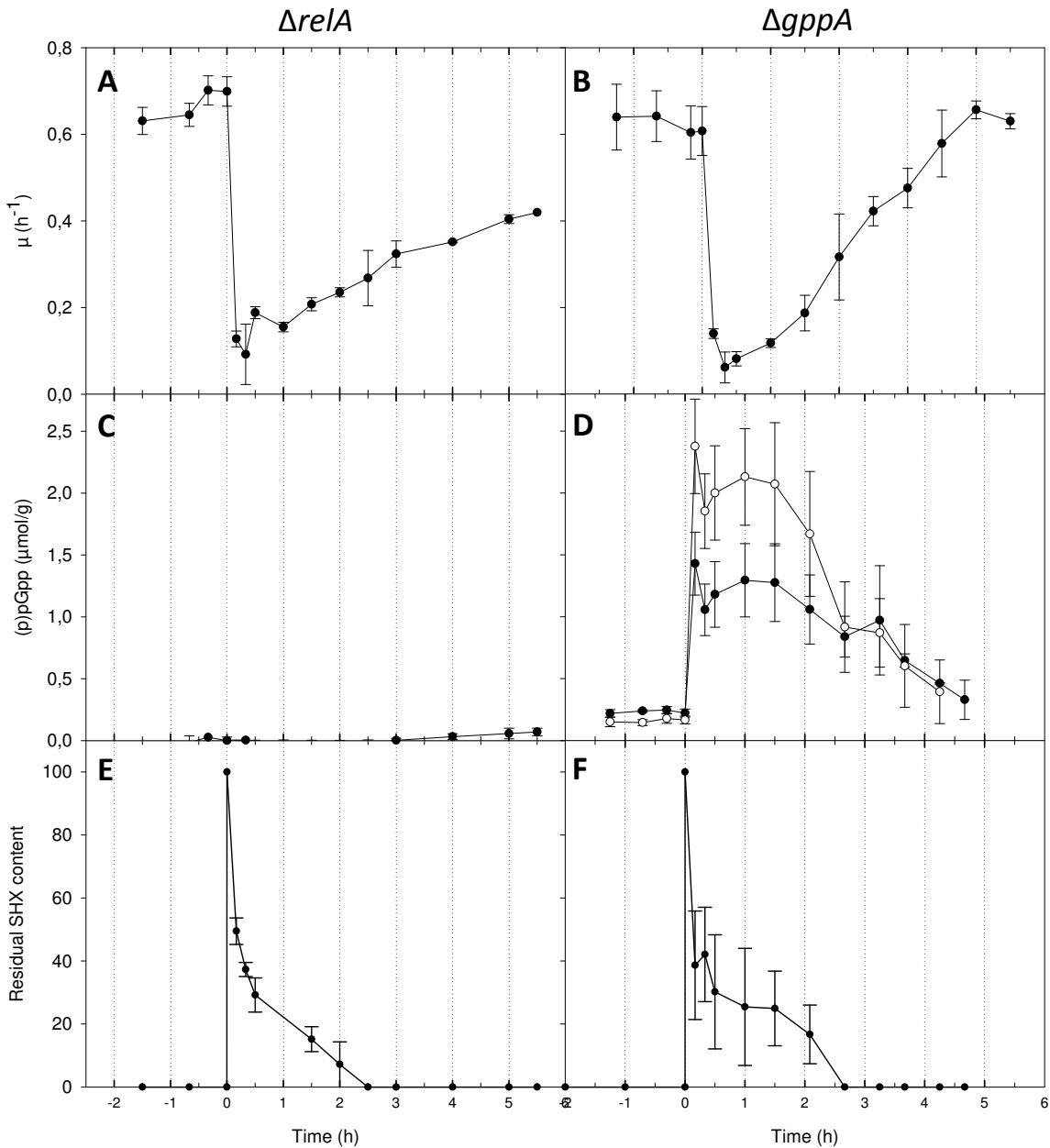


Figure 4. Involvement of ppGpp or pppGpp in the resilience of *E. coli* to SHX perturbation. Instantaneous growth rate (h^{-1}) (A - B), ppGpp and pppGpp concentrations ($\mu\text{mol/gCDW}$) (C – D) and residual SHX content (%) (E – F) measured before and after 0.8 mM SHX treatment during cultivations of *E. coli* K-12 BW25113 ΔgppA (left) or ΔrelA (right) strains. Time 0 was defined as the time at which SHX was added in the bioreactor.

With the $\Delta relA$ mutant, whereas the SHX is fully depleted from the medium after 2 hours, cells fail to fully recover a stable growth rate 6 hours after SHX addition. This clearly highlights the importance of the stringent response in the ability of *E. coli* to adapt to the disruptive effect caused by SHX addition on growth. Of note, we observe a slight recovery rate for this mutant suggesting that a rather inefficient adaptive mechanism occurs in a stringent-independent manner. With the $\Delta gppA$ mutant, the growth rate profile is similar to that with the WT strain (figure 4 and table 2) meaning that the level of the pentaphosphate form of the alarmone has no effect on growth recovery rate.

Table 2. Resilience characteristics of *E. coli* K-12 WT and mutant strains to SHX addition

	SHX (mM)	Robustness (%)	Adaptation rate	Recovery (%)
BW25113 WT	0.8	9.00 ± 3.3	0.17 ± 0.01	95.23 ± 0.37
BW25113 $\Delta relA$	0.8	20.14 ± 4.12	0.08 ± 0.03	75.19 ± 13.26
BW25113 $\Delta gppA$	0.8	13.12 ± 1.96	0.16 ± 0.01	105.08 ± 3.48

Involvement of stringent response on resilience in *E. coli* BL21 strain.

This study was then extended to the *E. coli* BL21 strain which is a widely used host for recombinant protein overproduction and other biotechnological products. Interestingly, this strain exhibits numerous of well characterized differences with *E. coli* K-12 strain at the metabolic level (Shiloach et al. 2010; Marisch, Bayer, Scharl, et al. 2013) (REFs Shiloach ou autres), notably in the greater capacity of *E. coli* BL21 strain for amino acids biosynthesis (Yoon et al. 2012; Li et al. 2014b) . Therefore, in the context of this study, we were then interested in comparing the response of *E. coli* BL21 strain to SHX perturbation with that of *E. coli* K-12 strain, both sharing a high protein sequence identity (>99% - The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45: D158-D169 (2017)) of RelA, SpoT, DksA and GppA, which are encoded by central stringent-response associated genes. To do so, the dynamic response to SHX addition (using 0.8 mM) was characterized for the *E. coli* BL 21 WT strain as well as for *E. coli* BL 21 $\Delta relA$ and $\Delta gppA$ mutants.

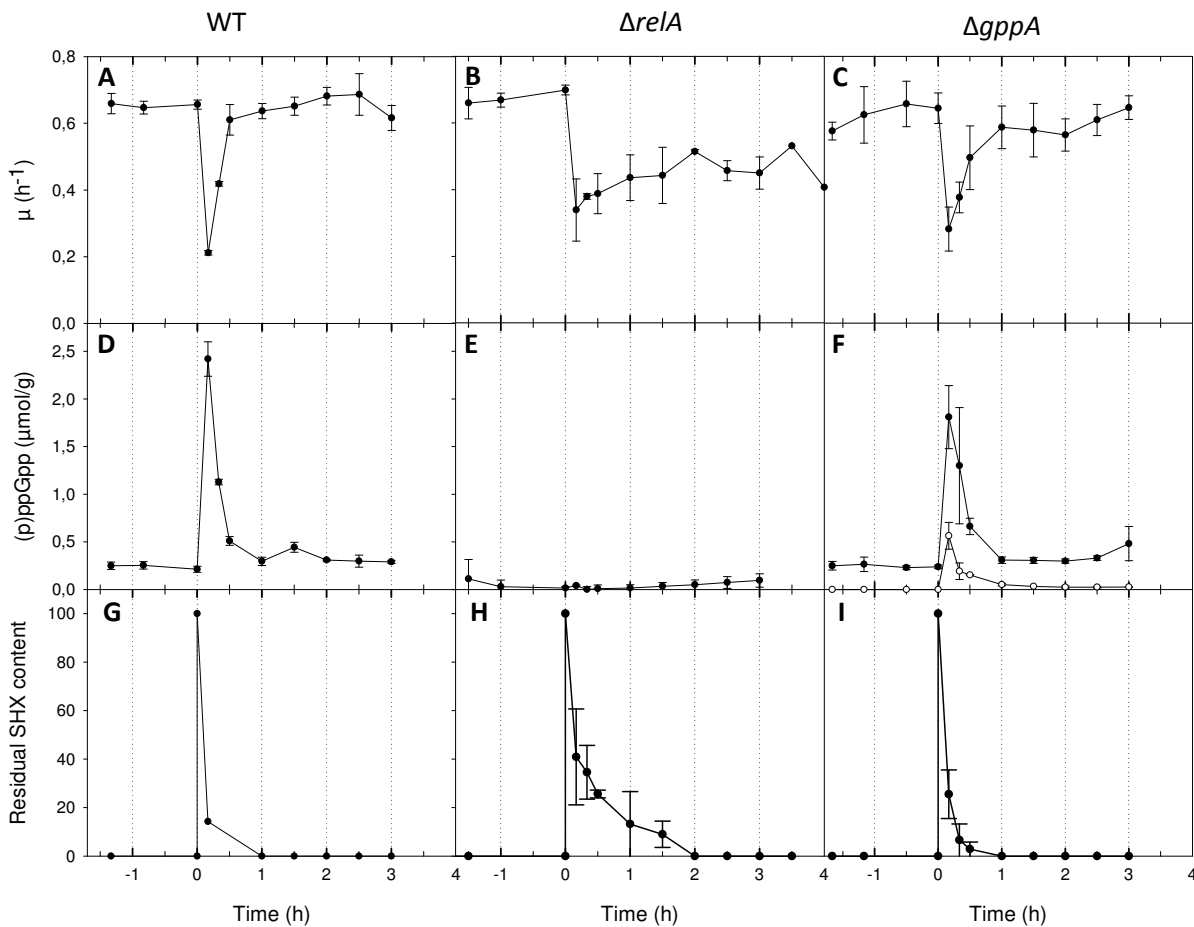


Figure 5. Resilience of *E. coli* BL21 to SHX perturbation. Instantaneous growth rate (h^{-1}) (A – B – C), ppGpp and pppGpp concentrations ($\mu mol/gCDW$) (D – E – F) and residual SHX content (%) (G – F – I) measured before and after 0.8 mM SHX treatment during cultivations of *E. coli* BL21 WT (left), $\Delta gppA$ (middle) or $\Delta relA$ (right) strains. Time 0 was defined as the time at which SHX was added in the bioreactor.

Surprisingly, we observe that the dynamic response of *E. coli* BL21 WT strain to SHX addition is very different compared to that of *E. coli* K-12 WT strain (Figure 5). First, the disappearance of SHX is very fast and takes place in less than 1 hour (instead of 3 hours with *E. coli* K-12). The cells respond to the addition of SHX by the sole accumulation of ppGpp (pppGpp level is under the limit of detection) that reaches approximatively twice the maximal level measured in the *E. coli* K-12 strain (Figure 5). Afterwards, the ppGpp content diminishes and returns back in one hour to the basal value (0.25

μmol/g CDW) which is similar to that measured in *E. coli* K-12. Finally, the cells recover an exponential growth rate in roughly 30 minutes (Figure 5) although it takes 4 hours for *E. coli* K-12 to recover its growth rate. These results highlight the high resilience capacity of the *E. coli* BL21 strain to SHX disturbance compared to *E. coli* K-12 strain. This is especially well illustrated by the robustness and the recovery rate which are much higher for *E. coli* BL21 strain compared to *E. coli* K-12 strain (table 3). Despite these significant discrepancies between both *E. coli* strains, it is interesting to note that we also detect significant amounts of valine and alanine that accumulate transiently in the cultivation medium during the pause in growth.

Table 3. Resilience characteristics of *E. coli* BL21 WT and mutant strains to SHX addition

	SHX (mM)	Robustness (%)	Adaptation rate	Recovery (%)
BL21 WT	0.8	30.78 ± 3.18	1.22 ± 0.11	98.82 ± 5.12
BL21 $\Delta relA$	0.8	47.82 ± 3.55	0.09 ± 0.11	71.63 ± 5.96
BL21 $\Delta gppA$	0.8	45.78 ± 13.46	0.55 ± 0.19	96.64 ± 5.12

In contrast, the *E. coli* BL21 $\Delta relA$ mutant is poorly resilient to SHX addition since no recovered state may be observed 4 hours after SHX addition (Figure 5). This is perfectly in line with the results obtained with the *E. coli* K-12 $\Delta relA$ mutant demonstrating that ppGpp is a key actor in the recovery process to SHX perturbation. It is worth mentioning that no ppGpp (nor pppGpp) was detected throughout this experiment, meaning that RelA is also in *E. coli* BL21 strain, the unique RelA/SpoT Homologue (RSH) protein synthesizing ppGpp during amino acids starvation.

Finally, *E. coli* BL21 $\Delta gppA$ mutant faces the SHX addition in a similar way as the *E. coli* BL21 WT strain although the recovery rate appears to be slightly higher for the WT strain (table 3). More interestingly, we detected a slight accumulation of pppGpp in this mutant indicating that GppA is also involved in the conversion of pppGpp into ppGpp in *E. coli* BL21 strain. Nevertheless, the intracellular level of pppGpp remains quite low compared to that measured in *E. coli* K-12 strain. This tends to suggest that the low pppGpp level in *E. coli* BL21 WT strain does not result from a high GppA hydrolase activity. Because of the high gene sequence identity of *relA* between *E. coli* K-12 and BL21 strains and the presence of *gppA* in the genome of both strains as a single copy gene, it is more likely that the pppGpp accumulation is dependent of GTP pool within the cells. However, further investigations are required to elucidate

definitively the difference in pppGpp accumulation between both strains. Nevertheless, despite the pppGpp in *E. coli* BL 21 $\Delta gppA$ mutant, we do not notice significant difference in growth profile of this mutant compared to the WT strain suggesting that pppGpp does not play a crucial role in growth recovery. We draw a similar conclusion in respect with the results obtained in *E. coli* K-12 although the latter accumulates higher level of intracellular pppGpp.

DISCUSSION

In this study, we investigated the dynamic response of *E. coli* to a severe disruption in amino-acids homeostasis and the role of the stringent response in the ability of *E. coli* to recover its growth. To achieve this, we monitor the growth, consumed or excreted metabolites and intracellular level of ppGpp and pppGpp throughout *E. coli* cultures subjected to a sudden SHX addition.

First, we demonstrated that *E. coli* K-12 WT strain is resilient to the disruption of amino acids homeostasis since it recovers its growth capacities few hours after SHX addition. Such a recovery has been shown in the pioneer work of Tosa & Pizer (1971) where the relief of growth inhibition was achieved by adding serine. In our work, the ability of *E. coli* to restore its growth appears to be self-contained. Intriguingly, we also observed that SHX disappears rapidly from the medium as a result of biological activity without one knowing whether the SHX is degraded (of note, we did not detect products of degradation) or simply internalized into the cells. Intuitively, one might assume that resuming of growth depends on whether SHX is present or absent in the medium. However, the results of the experiment with *E. coli* K-12 $\Delta relA$ mutant indicate that this is not true. In this experiment, although SHX disappears similarly as the WT, the $\Delta relA$ mutant does not recover growth despite the absence of SHX in the medium. This allows us to conclude that the stringent response is absolutely crucial for *E. coli* to recover its growth after the disruption of amino acids homeostasis.

Additionally, an important finding in this study is that the pentaphosphate form, pppGpp, has no clear impact on the resilience of *E. coli* to SHX addition. By eliminating the GppA, we deeply modified the relative accumulation of pppGpp vs. ppGpp in *E. coli*. Accordingly in a $\Delta gppA$ mutant, the ratio of pppGpp vs. ppGpp is approximatively 3-fold higher than that measured in the WT strain, ppGpp level being similar in both strain. However, both strains appear to be resilient to SHX the same way

indicating pppGpp has no significant effect on the growth recovery. It is in well in agreement with the conclusion of Mechold *et al.* (2013) stating that pppGpp is a less potent regulator than ppGpp on growth rate. Interestingly, we also observe that the decrease in pppGpp level two hours after the SHX addition, is highly correlated to that of ppGpp in the WT strain. One may thus assume that active degradation process by hydrolysis of (p)ppGpp into GD(T)P and pyrophosphate by SpoT (Mechold et al. 2002), acts in identical manner on ppGpp or on pppGpp. Another hypothesis is that the decrease is simply the result of the dilution of (p)ppGpp intracellular pool by the growth. Actually, we calculated the (p)ppGpp intracellular levels assuming only a dilution by growth (all fluxes of (p)ppGpp formation *via* RelA or GppA or degradation are considered to be null during that particular phase). On figure 6A, solid line is the ppGpp or pppGpp intracellular levels so calculated. If measured intracellular levels of ppGpp or pppGpp fit with this line, it means that decrease in ppGpp level is due to dilution of the pool by growth. If measured ppGpp level are up or down of this line, this means that an active formation or degradation process occurs respectively. We found that the measured ppGpp and pppGpp levels fit very well with solid line tending to show that the decrease of (p)ppGpp levels in *E. coli* K-12 may be mainly explained by the growth and not to by an active degradation process. This rules out the involvement of SpoT in the degradation of ppGpp.

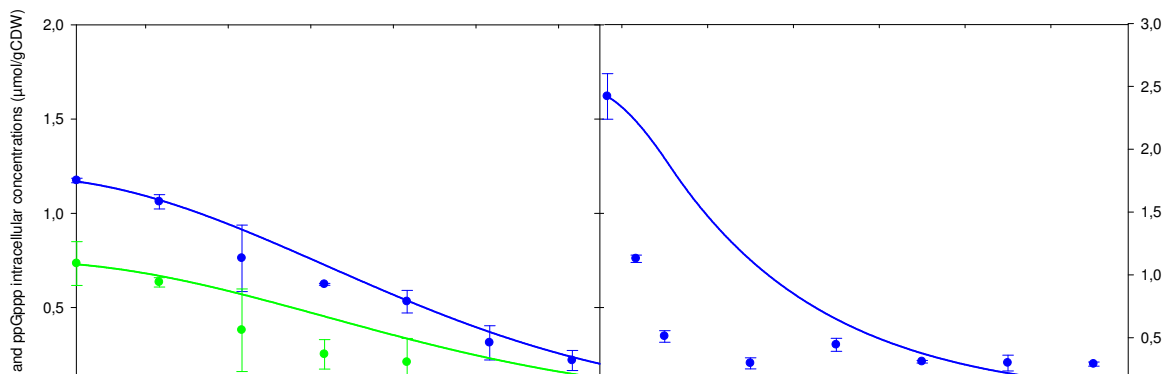


Figure 6. Intracellular concentration of ppGpp (solid blue line) and pppGpp calculated (solid green line) in *E. coli* K-12 (A) and in *E. coli* BL21 (B) assuming only a dilution of the pool by growth. Measured intracellular concentration of ppGpp and pppGpp are given in blue and green circles respectively.

Accumulation of both pppGpp and ppGpp occurred quickly after addition of SHX (less than 10 minutes) in line with earlier study (Riesenberg et al. 1984), in another bacterium albeit. As mentioned above, the ratio pppGpp vs. ppGpp is in favor of ppGpp in the WT strain in agreement with previous report (DeNapoli et al. 2013) as a result of RelA activity. In contrast, in the $\Delta gppA$ mutant, the ratio is clearly in favor of pppGpp in the 2.5 first hours after SHX addition. This indicates that *in vivo* GTP vs. GDP ratio favor the synthesis of pppGpp as regard to ppGpp (Varik et al. 2017). This results suggests that GppA activity might act to adjust the level of ppGpp which is the most potent regulator. This is in line with previous conclusion drawn by Mechold *et al.* (2013).

The present study points also out important differences between *E. coli* K-12 and BL21 strains to face the disruption in amino acids homeostasis. Although *coli* K and B strains are often considered closely related, sharing the same genome organization with >99 % sequence identity over ≈ 92 % of the genome, it is however well established that both strains exhibit significant physiological differences at the metabolic (Swartz 1996; Shiloach et al. 2010; Meier et al. 2012; Marisch, Bayer, Scharl, et al. 2013), proteomic (Yoon et al. 2012) and transcriptomic (Negrete et al. 2010) levels. Nonetheless, the magnitude of the differences is quite surprising as regards to RelA and SpoT identity between both strains. First, solely the tetraphosphate form of the alarmone is detected in the BL21 WT strain, in which the ppGpp accumulation reached approximatively twice that measured in K-12 strain. This indicates the high activity of RelA in BL21 strain. In contrast, the accumulation of pppGpp remains low in the *E. coli* BL21 $\Delta gppA$ mutant while exceeding the ppGpp level in K-12 strain. Because of the high identity between *relA* genes of *E. coli* K-12 and BL21 strains, both enzymes should share the same enzymatic properties in respect with substrate specificity and affinity. It is thus hard to think that the absence of pppGpp accumulation in BL21 strain compared to K-12 would be the result of different enzymatic properties. One can instead assume that GDT/GTP ratio might differ between the strains or that SpoT of BL21 strain might be highly active to hydrolyze (p)ppGpp (Somerville & Ahmed 1979). Second, the BL21 strain appears to be highly resilient to disruption of amino acids homeostasis since this strain recovers its growth rate in approximatively half hour (instead of 4 for *E. coli* K-12 strain). Importantly, the resilience ability relies undoubtedly on the stringent response since the $\Delta relA$ mutant failed to recover its maximal growth rate. This is common and important trait shared by the both K-12 and BL21 strains. This result also confirms that the ppGpp is the major regulator since the BL21 strain

faces efficiently the perturbation caused by SHX addition whereas pppGpp is absent. An interesting observation is the difference in (p)ppGpp accumulation profile between the K-12 and BL21 strains. In both strains, the addition of SHX triggers the rapid accumulation of ppGpp, the accumulation in BL21 strain being roughly twice that measured in K-12 strain. But in the former strain, the content in ppGpp returns straight back to the basal value after the build-up in contrast to what it is observed in K-12 strain, where ppGpp level shows a plateau for approximately 2 hours, prior to decrease to the basal value. It well admitted that quantitative differences in the ppGpp intracellular level could shape different pattern of gene expression, where low levels of ppGpp activate the Lrp regulon while high levels activate RpoS and the general stress response (Balsalobre 2011; Traxler, Summers, H.-T. Nguyen, et al. 2008). In addition, the hydrolase activity of SpoT is supposed to be inhibited in condition of physiological stress and in particular, in high level of uncharged tRNA (Murray & Bremer 1996). Based on this, it is tempting to think that the discrepancies in resilience dynamics observed for the two strains, could be attributed to hydrolase activity of SpoT. When using the same approach as above (figure 6B), it appears clearly that the decrease of the ppGpp pool is not due to the dilution by growth but by an active degradation system. Therefore in BL21 strain, the SpoT hydrolase would be highly active despite of the high accumulation of ppGpp after SHX addition. This may also explain why the pppGpp pool measured in this strain is low. In contrast, in K-12 strain, the accumulation of ppGpp following the addition of SHX, would induce stress survival genes in a RpoS-dependent manner, inhibiting the hydrolase activity of SpoT. This inhibition would explain the plateau of ppGpp pool as well as the gently decrease of the ppGpp level that is mostly correlated with the recovery of growth (see above). This suggests that K-12 strain fail to overcome the disruption of amino acids homeostasis as fast as BL21 strain. It is noteworthy that *E. coli* K-12 responds differently to SHX in respect with the amount of SHX added. Specifically, the profile of ppGpp accumulation does not show a plateau when the amount of SHX added is at least halved, similarly to that observed with the BL21 strain. In such case, one can assume that the accumulation of ppGpp does not reach a threshold level triggering the induction of stress survival genes. The SpoT hydrolase would remain active to degrade (p)ppGpp. Numerous studies with BL21 strain imply that this *E. coli* strain has metabolic characteristics adapted for amino acids biosynthesis (Phue et al. 2005; Li et al. 2014a) - making indeed this bacteria a

workhorse for heterologous protein overproduction. One can think that these characteristics contribute the BL21 strain to better succeed in facing the effect of SHX on amino acids metabolism.

Furthermore, together with the data obtained with the K-12 ΔreA mutant, it also suggests that SHX disappearance is mostly independent of the stringent response although the latter accelerates this phenomenon, and confirms that the absence of SHX in the medium is not a sufficient condition for *E. coli* to resume its growth. The pioneer work of Toza and Pizer (1971a & b) on the effect of SHX on *E. coli*'s growth demonstrated that growth inhibition exerted by SHX is rapidly reversed by the addition of L-serine and that SHX is a competitive inhibitor of the seryl-tRNA synthetase activity. Therefore, the sensitivity of *E. coli* would be due to the intracellular ratio of SHX vs. L-serine. It is well known that the stringent response activates amino acids biosynthesis *via* Lrp dependent gene expression, even at low ppGpp level (Balsalobre 2011). In our experiments, the overflow of alanine and valine observed few minutes after ppGpp accumulation in both K-12 and BL21 WT strains pleads in that senses. Therefore, one can assume that activation of serine biosynthesis by the stringent response via Lrp may contribute to dilute the SHX pool within the cell and to relieve the growth inhibition exerted by the serine analog, the growth recovery contributing to shift the balance in favor of L-serine by growth rate dilution effect of SHX pool.

PERSPECTIVES

L'étude de l'influence de la réponse stringente sur la résilience d'*E. coli* à une perturbation au SHX permet d'ouvrir de nombreuses perspectives à ce travail. Dans cette étude, le SHX disparaît du surnageant après son ajout dans le milieu de culture. Il est probable que ce composé soit dégradé ou s'accumule dans les cellules. Deux questions se posent alors : que devient le SHX et quelle mécanisme conduit à sa disparition du surnageant ? Pour répondre à la première question une expérience par marquage isotopique peut être réalisée. Pour cela, du SHX non marqué peut être ajouté à une culture d'*E. coli* en milieu minimum avec du glucose ^{13}C comme source de carbone. Ainsi, la recherche de carbones non marqués dans le surnageant, les acides aminés protéogéniques ou acides aminés libres serait facilitée. Certaines études montrent que le SHX est capable d'inhiber l'activité de la seryl-transférase créant ainsi une augmentation de la quantité d'ARNt non chargé dans la bactérie (Tosa & Pizer 1971; Tosa 1971; Weygand-durasevic et al. 1993; Zeng et al. 2009). Cependant, un tel mécanisme

ne permet pas d'expliquer pourquoi le SHX disparaît du milieu de culture. La découverte de la raison pour laquelle le SHX disparaît du surnageant pourrait permettre d'améliorer la compréhension de son mécanisme d'action.

La deuxième perspective est d'étudier le mécanisme d'adaptation des bactéries à l'ajout de SHX. La quantification des acides aminés libres a permis d'observer une accumulation de sérine de façon transitoire au moment de l'arrêt de croissance de la souche BL21. De plus, une reprise de croissance est observée lorsque de la L-sérine est ajoutée dans le milieu de culture après un ajout de SHX (Tosa 1971). L'hypothèse est donc que l'ajout de L-sérine dans le milieu de culture permet d'augmenter la concentration intracellulaire en L-sérine entraînant une compétition avec le SHX pour l'interaction avec la seryl-transférase permettant le chargement des ARNt. Il est donc probable que la reprise de croissance de la souche BL21 soit causée par une augmentation de la concentration intracellulaire de L-sérine lors de l'arrêt de croissance. Cette hypothèse pourrait être démontrée par une quantification absolue de la L-sérine et du SHX intracellulaire après ajout de SHX à une culture de la souche *E. coli* BL21 et la souche *E. coli* BW25113. Cependant, ce mécanisme d'adaptation ne permet pas d'expliquer pourquoi le SHX disparaît du surnageant des cultures.

L'adaptation transcriptionnelle pourrait également être à l'origine ou participer à l'adaptation des bactéries à la perturbation causée par le SHX. Pour vérifier cela, un mutant du gène *dksA* de la souche *E. coli* BL21 et BW25113 pourrait être utilisé lors d'une culture avec ajout de SHX (mutants disponibles au laboratoire). La protéine DksA est indispensable à l'efficacité de l'action du ppGpp au niveau transcriptionnel (Kolmsee et al. 2011; Paul et al. 2005; Roghanian et al. 2015; Brown et al. 2002). La comparaison de la résilience des souches $\Delta dksA$ avec celles des souches WT permettrait d'étudier l'implication de la régulation transcriptionnelle par le (p)ppGpp sur l'adaptation à une perturbation au SHX.

La quantité de ppGpp et de pppGpp est très différente entre la souche *E. coli* BL21 et la souche *E. coli* BW25113 même lorsque la réponse stringente n'est pas activée. L'étude de l'origine des différences entre les concentrations basales de ppGpp ou de pppGpp entre les souches pourrait être entreprise par deux approches différentes. Les séquences primaires des protéines RelA, SpoT et GppA de la souche *E. coli* BL21 et *E. coli* BW25113 ont été comparées et de faibles différences apparaissent

entre les mêmes protéines des deux souches (1 à 2% des acides aminés, résultats non montrés). La première approche consiste à échanger les gènes codant pour la protéine RelA ou SpoT ou GppA entre la souche *E. coli* BL21 et la souche *E. coli* BW25113. En quantifiant la quantité de ppGpp et de pppGpp synthétisée par chacune des souches modifiées il serait alors possible de déterminer si la concentration intracellulaire plus élevée en pppGpp de la souche *E. coli* BW25113 est dû à une différence d'activité de la protéine RelA, SpoT (différence de dégradation), de GppA ou encore à une différence de l'environnement cellulaire des protéines. Une deuxième approche consiste à étudier l'activité des protéines *in-vivo*. Cette approche nécessite d'effectuer un ajout de substrat marqué (¹³C, ¹⁵N, ou ¹³C et ¹⁵N) au cours d'une culture de la souche *E. coli* BL21 ou BW25113 et des mutants $\Delta relA$, $\Delta gppA$, $\Delta SpoTS$ (mutation dans la partie synthase de SpoT) en condition de croissance exponentielle. Les flux de synthèse et de dégradation du ppGpp et du pppGpp dans chaque souche pourraient alors être déterminés en déduisant l'implication de chaque protéine à l'aide des mutants. Les deux approches (échange de gènes et analyse de l'activité des protéines *in-vivo*) peuvent également être combinées.

SUPPLEMENTARY DATA : TABLE 1

Strain	Relevant genotype	Reference/source
BW25113	rrnB3 DElacZ4787 DEphoBR580 hsdR514 DE(ara- BAD)567 DE(rhaBAD)568 galU95 DEendA9::FRT DEuidA3::pir(wt) recA1 rph-1	Baba et al. 2016
BW25113 <i>relA</i>	rrnB3 DElacZ4787 DEphoBR580 hsdR514 DE(ara- BAD)567 DE(rhaBAD)568 galU95 DEendA9::FRT DEuidA3::pir(wt) recA1 rph-1	Baba et al. 2016
BW25113 <i>gppA</i>	rrnB3 DElacZ4787 DEphoBR580 hsdR514 DE(ara- BAD)567 DE(rhaBAD)568 galU95 DEendA9::FRT DEuidA3::pir(wt) recA1 rph-1	Baba et al. 2016
BL21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] DEhsdS λ DE3 = λ sBamHlo DEEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 DEin5	This study
BL21 (DE3) <i>relA</i>	fhuA2 [lon] ompT gal (λ DE3) [dcm] DEhsdS λ DE3 = λ sBamHlo DEEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 DEin5	This study
BL21 (DE3) <i>gppA</i>	fhuA2 [lon] ompT gal (λ DE3) [dcm] DEhsdS λ DE3 = λ sBamHlo DEEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 DEin5	This study

6.1.2. Quantification des acides aminés d'*E. coli* BL21 WT lors de l'ajout de SHX.

L'étude de l'influence de la réponse stringente dans la résilience de la souche *E. coli* BL21 et de la souche *E. coli* K-12 à une perturbation liée à l'ajout de SHX met en évidence que la réponse stringente permet une adaptation plus rapide de la souche *E. coli* BL21 que la souche *E. coli* K-12. Cette étude met également en évidence que l'accumulation de ppGpp entraîne une inhibition du taux de croissance conformément à ce qui est rapporté dans la littérature (Mechold et al. 2013). De plus, l'arrêt de croissance est concomitant à l'accumulation de valine et d'alanine dans le milieu extracellulaire. Cette excrétion a également été observée lors des analyses de l'exométabolome de la souche *E. coli* BL21 en condition de production de la protéine UL130. Une analyse avec une plus grande sensibilité de détection a permis d'observer au cours de la production des évolutions de concentrations en acides aminés corrélées pour certains à la cinétique de croissance et pour d'autres à la cinétiques de production. Dans le but d'étudier l'effet de l'arrêt de croissance en absence de production sur l'évolution des concentrations en acide aminés, un arrêt de croissance a été causé par l'activation de la réponse stringente suite à l'ajout de SHX à une culture de la souche *E. coli* BL21 sans production, et les acides aminés libres ont été quantifiés par LC-MS (figure 6.1).

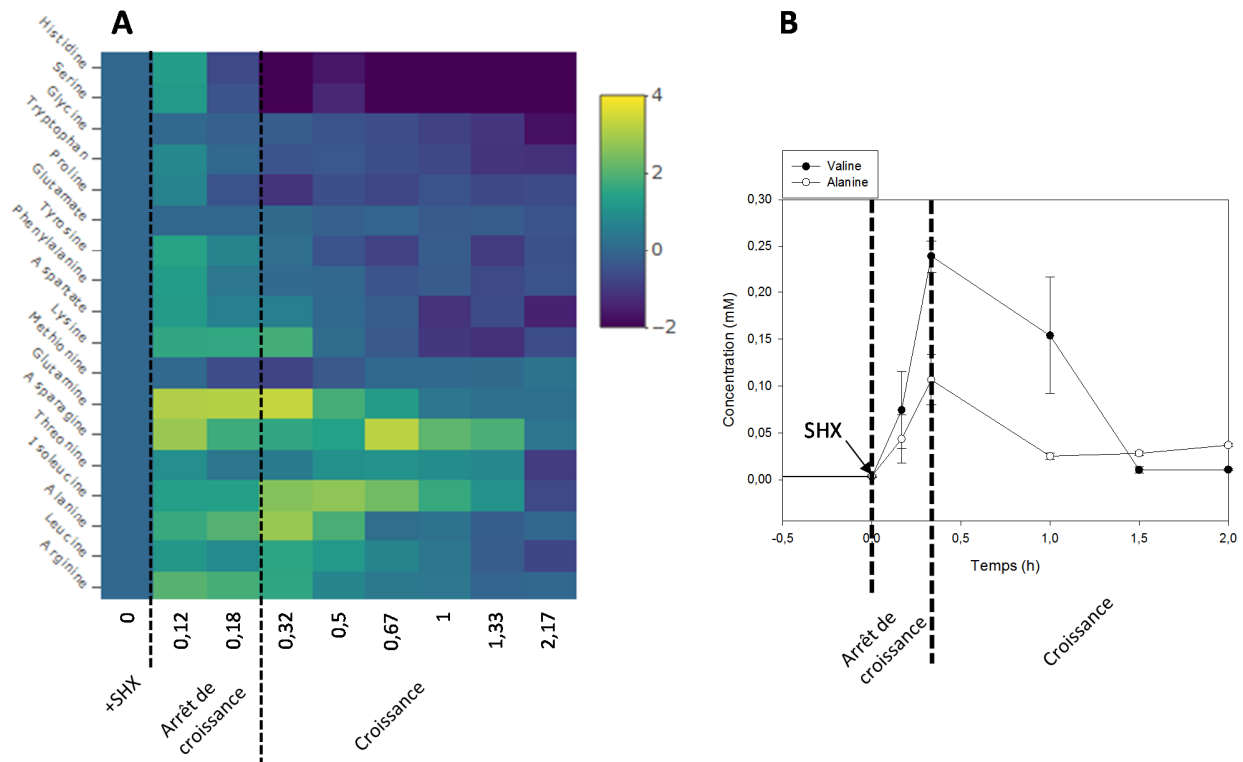


Figure 6.1 : Evolution des concentrations en acides aminés libres lors de l'activation de la réponse stringente causant un arrêt de croissance après ajout de SHX lors d'une culture en milieu minimum de la souche *E. coli* BL21. (A) Les concentrations en acides aminés libres sont quantifiées par LC-MS de façon relative à leur concentration avant l'ajout du SHX (simplicat biologique). (B) Concentrations en valine et alanine dans le surnageant des cultures, quantifiées par RMN (duplicat biologique).

Les résultats obtenus par quantification des acides aminés lors de l'ajout de SHX à une culture d'*E. coli* BL21 WT mettent en évidence une accumulation de tous les acides aminés (sauf la thréonine, la glycine, et le glutamate) au moment de l'arrêt de croissance (figure, A et B). Les concentrations maximales détectées et les durées d'accumulation sont plus importantes pour l'arginine, la leucine, l'alanine, l'isoleucine, l'asparagine, glutamine et la lysine que pour les autres acides aminés. La valine est difficile à ioniser en spectrométrie de masse mais son accumulation est détectée en RMN dans le surnageant indiquant une forte accumulation de cet acide aminé durant l'arrêt de croissance causé par l'ajout de SHX (figure 6.1, B).

L'augmentation de la concentration en sérine durant l'arrêt de croissance pourrait faire partie du mécanisme permettant l'adaptation de la souche *E. coli* BL21 à l'ajout de SHX. En effet, il a été démontré que l'ajout de L-sérine dans le milieu de culture durant un arrêt de croissance causé par le SHX entraîne une reprise immédiate de la croissance (Tosa 1971). La reprise de croissance résulterait de la compétition de la L-sérine avec son analogue pour l'interaction avec la seryl transférase. Ainsi, une augmentation de la concentration intracellulaire de L-serine lors de l'arrêt de croissance pourrait conduire au même mécanisme de compétition avec le SHX et permettre le retour de croissance. De plus, ce mécanisme expliquerait pourquoi l'activation de la réponse stringente de la souche *E. coli* BL21 permet une meilleure adaptation de la souche au SHX. En effet, il a été montré que cette souche est plus performante que la souche *E. coli* K-12 pour synthétiser des acides aminés grâce à des voies anaplérotiques plus actives (Noronha et al. 2000; Waegeman et al. 2011; Marisch, Bayer, Cserjan-Puschmann, et al. 2013). Pour démontrer cette hypothèse, une quantification absolue de la concentration intracellulaire en L-sérine devrait être effectuée sur la souche *E. coli* BL21 et la souche *E. coli* K-12. Si la concentration intracellulaire en L-sérine est plus élevée dans la souche *E. coli* BL21 que dans la souche *E. coli* K-12 au moment de l'arrêt de croissance, un résultat serait avancé pour expliquer l'adaptation des souches au SHX et l'origine de l'efficacité de l'adaptation de la souche *E. coli* BL21 par rapport à celle de la souche *E. coli* K-12.

La quantification des acides aminés met en évidence que l'arrêt de croissance engendre une augmentation de la concentration de presque tous les acides aminés. L'inhibition de la croissance par le ppGpp augmente donc la disponibilité intracellulaire des acides aminés de façon transitoire. Cette accumulation est transitoire probablement parce que des régulations de rétrocontrôle de la synthèse des acides aminés sont mises en place suite à l'augmentation de leur concentration intracellulaire.

Pour rappel, les quantifications des acides aminés réalisées au cours du procédé de production de la protéine UL130 sont présentées (figure 6.2).

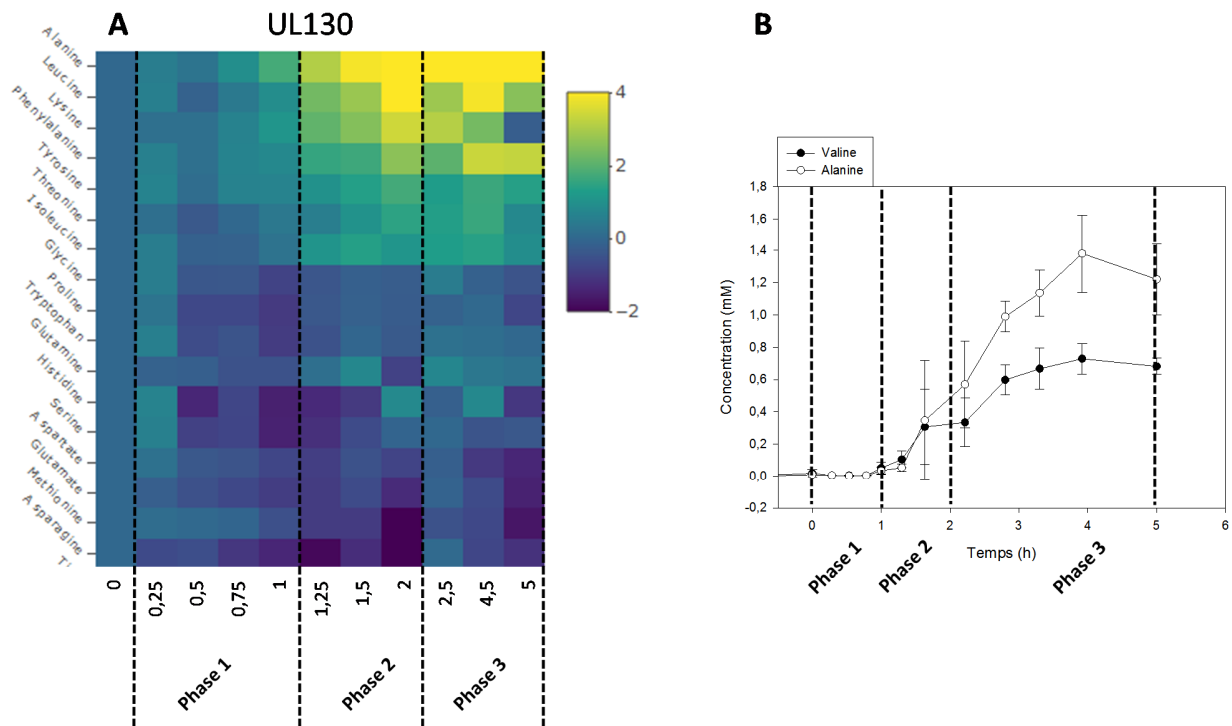


Figure 6.2 : Evolution des concentrations en acides aminés libres lors de l'arrêt de croissance au cours de la production de la protéine UL130 lors d'une culture en milieu minimum de la souche *E. coli* BL21. (A) Les concentrations en acides aminés libres sont quantifiées par LC-MS de façon relative à leur concentration avant l'ajout d'IPTG (triplicat biologique). (B) Concentrations en valine et alanine dans le surnageant des cultures, quantifiées par RMN (triplicat biologique). La phase 1 correspond à une période de croissance et de production, la phase 2 à une période d'arrêt de croissance en condition de production et la phase 3 à un arrêt de la croissance sans production.

Au cours du procédé de production de la protéine UL130, l'accumulation d'une partie des acides aminés est également concomitante à l'arrêt de croissance (figure 6.2, A et B, phase 2). Cependant, la concentration de certains acides aminés diminue durant cette phase d'arrêt de croissance et de production (proline, tryptophane, aspartate, méthionine et asparagine) (figure 6.2, A). Très probablement ces acides aminés ne s'accumulent pas car la synthèse de la protéine recombinante consomme les acides aminés qui ne sont plus utilisés pour la synthèse de la biomasse. Du fait de l'inhibition du taux de croissance par le ppGpp, la part du flux de consommation des acides aminés utilisé normalement pour la synthèse des protéines de la biomasse semble orienté vers pour la synthèse de la protéine recombinante

Pour conclure, ces analyses mettent en évidence que l'accumulation des acides aminés est caractéristique d'un arrêt de croissance lors de l'activation de la réponse stringente. Ce phénomène semble être la conséquence d'une inhibition de la croissance par l'accumulation de ppGpp conduisant à une période où les flux de synthèses des acides aminés sont supérieurs aux flux de consommations. Dans un contexte de production, le flux important de consommation de certains acides aminés destiné à la synthèse de la protéine recombinante compenserait la diminution du flux de consommation pour la synthèse de la biomasse.

6.2. Etude de l'influence de la réponse stringente sur la souche *E. coli* BL21 en condition de production de protéines recombinantes

Les travaux présentés dans ce chapitre ont été réalisés pour étudier l'influence de la concentration en ppGpp sur la production de la protéine UL130 et Hla. L'étude présentée ci-dessous s'attache à distinguer l'influence de l'accumulation de ppGpp et de la production sur la croissance. Pour cela, des souches mutées de gènes impliqués dans la réponse stringente sont utilisées pour moduler le niveau de ppGpp dans la cellule au cours de la production. L'effet de cette modulation est analysé avec la même approche que celle utilisée au cours des chapitres 3 et 4 de la thèse.

6.2.1. Influence de la concentration intracellulaire en ppGpp sur la croissance et la production de la protéine UL130.

Pour le procédé de production de la protéine UL130 les mutants $\Delta reIA$ et le double mutant $\Delta reIA : \Delta gppA$ sont utilisés pour moduler la concentration intracellulaire de ppGpp. Le mutant $\Delta gppA$ n'est pas présenté à cause de problèmes techniques de quantification du ppGpp ne permettant pas une exploitation pertinente des résultats.

Pour rappel, la production spécifique représente la quantité de protéine recombinante par unité de biomasse et la production totale représente la quantité de protéine par litre de culture (figure 6.3).

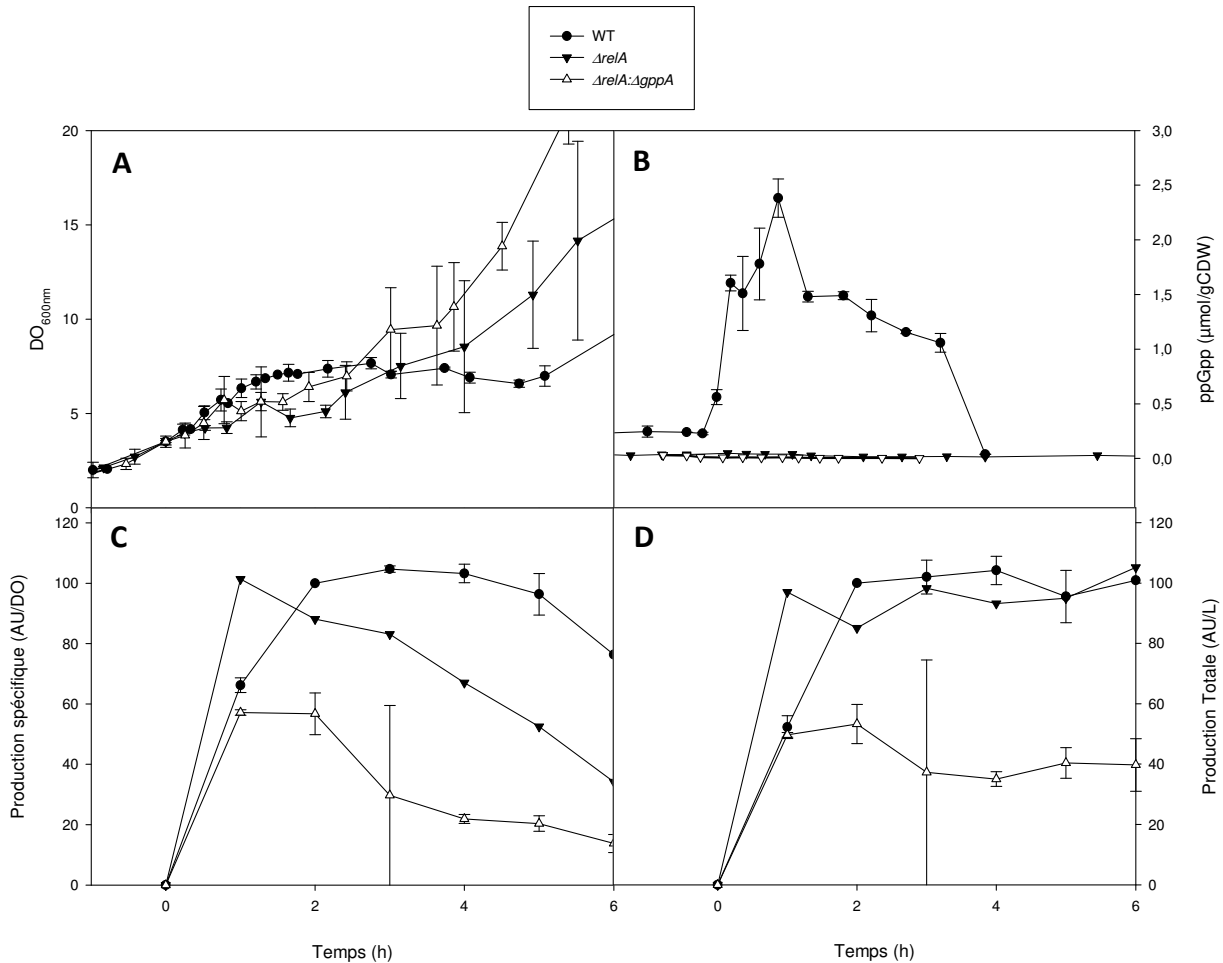


Figure 6.3 : Procédés de production de la protéine UL130 avec la souche *E. coli* BL21 WT (ronds noirs) et avec les souches $\Delta relA$ (triangles noirs) et $\Delta gppA : \Delta relA$ (triangles blancs). (A) Biomasse (DO), (B) concentration en ppGpp ($\mu mol/gCDW$), (C) production spécifique (AU/DO), (D) production totale (AU/L) au cours du temps. Le temps est normalisé à partir de l'ajout de l'IPTG 1 mM ($T = 0$ h).

Un arrêt de croissance apparaît pour l'ensemble des souches 1 heure après induction (figure 6.3, A). Cependant une reprise de croissance est observée respectivement 0,5 h et 1 h après l'arrêt de croissance pour les souches $\Delta gppA : \Delta relA$ et $\Delta relA$ contre 4 h après l'arrêt de croissance pour la souche WT. La reprise de croissance des souches WT, $\Delta gppA : \Delta relA$ et $\Delta relA$ n'est pas accompagnée de production (figure 6.3, C et D). La production des souches $\Delta relA$ et $\Delta gppA : \Delta relA$ est observée durant la première heure après induction et durant les deux premières heures pour la souche WT (figure 6.3, D). Les productions spécifiques et totales maximales de la souche $\Delta gppA : \Delta relA$ sont 40% à 50% plus faibles que celles des souches WT et $\Delta relA$ (figure 6.3, C et D). Concernant la quantification du ppGpp

au cours du procédé, une faible concentration est retrouvée avant induction qui n'évolue pas après induction pour les souches $\Delta gppA$: $\Delta relA$ et $\Delta relA$ contrairement à celle de la souche WT (figure 6.3, B).

Chez *E. coli*, la protéine RelA est activé lorsqu'un ARNt non chargé se fixe sur un ribosome déclenchant la synthèse de ppGpp (Loveland et al. 2016; Li et al. 2016). En condition de production, la concentration intracellulaire de ppGpp n'évolue pas avec la souche $\Delta relA$ contrairement à une accumulation significative observée avec la souche WT. La protéine RelA est donc responsable de la synthèse de ppGpp dans la souche WT. De fait, dans la souche WT, la présence d'ARNt non chargés conduit à l'activation de RelA démontrant une limitation en acides aminés causée par la production de la protéine UL130.

Le retour de croissance de la souche $\Delta relA$ et de la souche $\Delta gppA$: $\Delta relA$, rapidement après leur arrêt de croissance, met en évidence que la limitation en acides aminés n'entraîne pas d'inhibition significative de la croissance de ces souches. De fait, l'arrêt de croissance prolongé de la souche WT est une conséquence de l'accumulation du ppGpp. Lorsque ces résultats sont mis en perspectives avec le fait que l'accumulation de ppGpp inhibe la croissance, orientant ainsi la consommation des acides aminés destinées à la synthèse de la biomasse vers leur utilisation pour la synthèse de protéine recombinante, une hypothèse apparait. L'hypothèse est que la durée de l'arrêt de croissance serait en lien avec la quantité de ppGpp synthétisé au début de la production. Cet arrêt de croissance permettrait à la production de se poursuivre une heure en consommant les acides aminés intracellulaires qui ne sont plus consommés par la croissance. Cette prolongation de la production consommant les acides aminés libérés par l'arrêt de croissance, maintient une limitation en certains acides aminés et donc une forte concentration en ppGpp, d'où un arrêt de croissance long.

La suppression du gène *gppA* à la souche $\Delta relA$ semble diminuer la production spécifique et totale de la protéine UL130 au profit d'une croissance plus importante après induction. La protéine GppA, impliquée dans la transformation du pppGpp en ppGpp, semble donc avoir une influence sur le taux de croissance et la production après induction. Cependant les concentrations similaires en ppGpp et l'absence de pppGpp détecté au cours des cultures des souches $\Delta relA$ et $\Delta gppA$: $\Delta relA$ ne permettent pas d'apporter d'explication à ces différences de phénotypes.

Il est cependant clair que la suppression de la capacité à synthétiser du ppGpp avec la protéine RelA entraîne une diminution du temps d'arrêt de croissance après induction et modifie la dynamique de production de la protéine UL130. Cependant, la reprise de croissance n'est jamais accompagnée de production démontrant que la perte de capacité de production des souches n'est pas liée ni à la durée de l'arrêt de croissance ni à l'accumulation de ppGpp.

6.2.2. Influence de la concentration intracellulaire de ppGpp sur la croissance et la production de la protéine Hla.

Durant le procédé de production de la protéine Hla avec la souche *E. coli* BL21 WT, la phase de production et de croissance est maintenue sur un temps plus long que lors de la production de la protéine UL130. Au cours de cette période, la modulation du ppGpp permettrait de distinguer son influence sur la croissance et la production de façon plus précise qu'avec le procédé de production de la protéine UL130.

Pour rappel, La production spécifique représente la quantité de protéine par unité de biomasse (AU/DO) et la production totale la quantité de protéine par litre de culture (AU/L). La vitesse de production spécifique représente le flux de synthèse de protéines recombinantes par unité de biomasse et par heure (AU/DO/h) (figure 6.4).

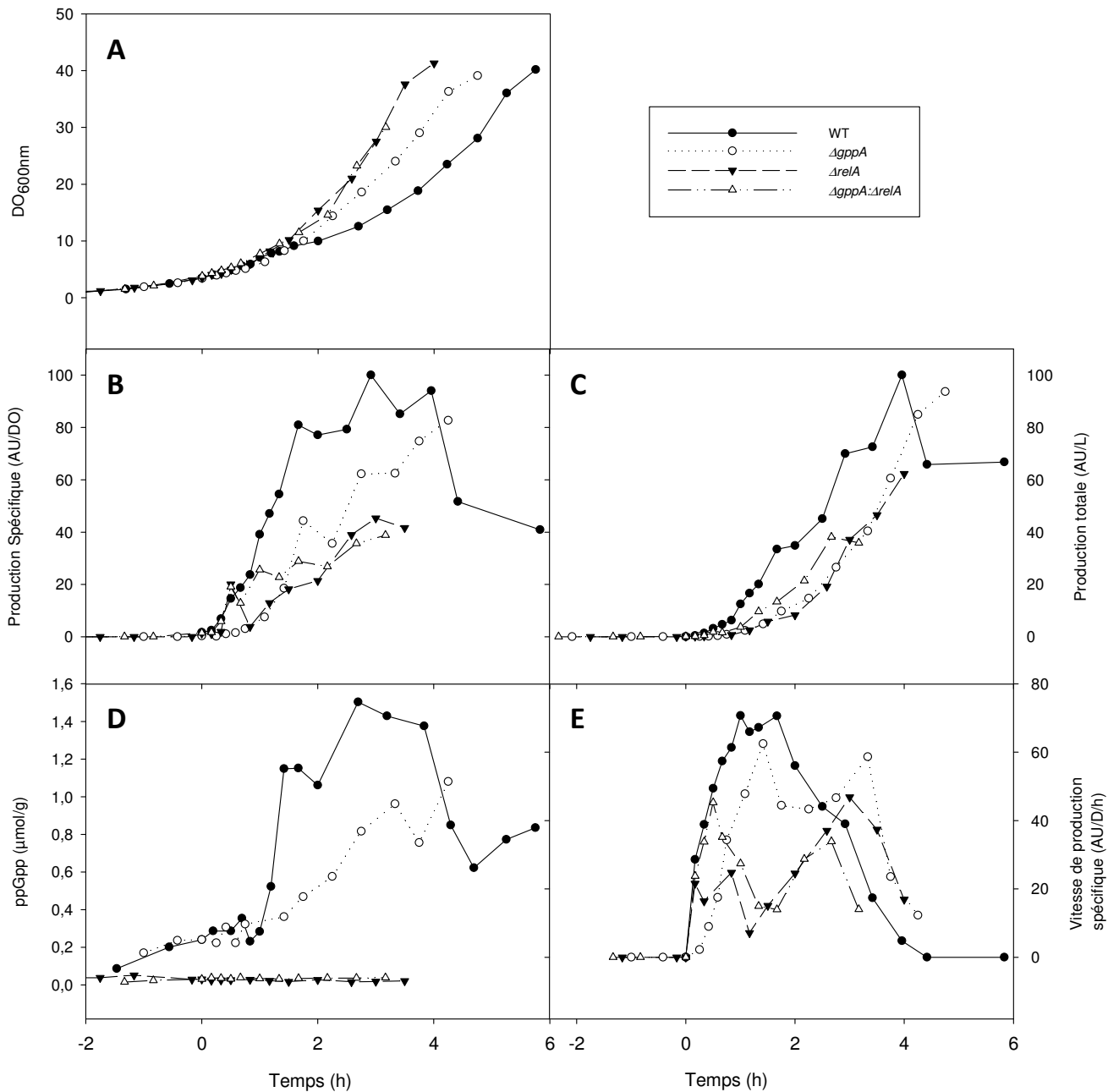


Figure 6.4 : Etude de l'influence de la concentration en ppGpp intracellulaire au cours des procédés de production de la protéine Hla avec la souche *E. coli* BL21 WT et les souches $\Delta gppA$, $\Delta relA$ et $\Delta gppA : \Delta relA$. (A) Biomasse (DO), (B) productions spécifiques (AU/DO), (C) production totales (AU/L), (D) concentration en ppGpp ($\mu mol/g$) et (E) vitesse de production au cours du temps. Le temps est normalisé à partir de l'ajout de l'IPTG 1 mM ($T = 0$ h).

Au cours du procédé de production de la protéine Hla, la délétion des gènes *gppA*, *relA* ou *gppA* et *relA* engendre une augmentation plus rapide de la biomasse après 1,5h d'induction par rapport à celle de la souche WT (figure 6.4, A). En effet, la souche WT atteint 30 uDO environ 4h45min après induction, la souche $\Delta gppA$ 3h45min après induction et les souches $\Delta relA$ et $\Delta gppA : \Delta relA$ atteignent cette DO au bout de 3 h. La quantification du ppGpp est cohérente avec les croissances observées (figure 6.4, D). En effet, aucune accumulation de ppGpp n'est détectée pour les souches $\Delta relA$ ou $\Delta gppA : \Delta relA$ et une accumulation intermédiaire de ppGpp est observée avec la souche $\Delta gppA$ par rapport à celle observée avec la souche WT. Une production spécifique plus faible a été déterminée pour l'ensemble des souches mutées au cours du procédé (figure 6.4, B). Les évolutions des productions spécifiques semblent donc coïncider avec l'évolution de la concentration intracellulaire en ppGpp dans chacune des souches (figure 6.4, B et D). Cependant, durant la période de production la plus grande quantité de biomasse générée pour les souches mutées permet d'obtenir des productions totales proches de celles observées avec la souche WT (figure 6.4, C). Les vitesses de production spécifique semblent être influencées par la concentration intracellulaire en ppGpp (figure 6.4, D et E). En effet, la production atteint des vitesses spécifiques plus élevées avec les souches WT et $\Delta gppA$ entre 1 heure et 3 heures après l'induction. Cette phase du procédé correspond au moment où le ppGpp est retrouvé en plus forte concentration intracellulaire dans ces souches (figure 6.4, D). Les souches mutées semblent produire la protéine Hla jusqu'à la fin de la phase exponentielle de croissance (4 h après induction pour la souche $\Delta relA$ ou $\Delta gppA : \Delta relA$ et 4,75 h après induction pour la souche $\Delta gppA$). La souche WT ne semble plus produire de protéine Hla 2 heures avant la fin de la phase exponentielle de croissance (figure 6.4, C). Cet arrêt de production pourrait être lié au fait que cette souche atteigne de fortes productions spécifiques.

L'absence d'accumulation de ppGpp pour les souches mutées du gène *relA* démontre que le ppGpp observé avec la souche WT est exclusivement synthétisé via la détection d'ARNt non chargé par RelA. De fait, au cours de la production de la protéine Hla par la souche WT, la synthèse de ppGpp est provoquée par une limitation en acides aminés.

Les productions spécifiques des souches WT et $\Delta gppA$ sont deux fois plus élevées que les souches $\Delta relA$ et $\Delta gppA : \Delta relA$ en fin de procédé. Cette observation peut être expliquée par deux phénomènes.

Tout d'abord, une inhibition du taux de croissance par le ppGpp est observée avec la souche WT et $\Delta gppA$ par rapport aux souches mutées du gène *relA*. De plus, une vitesse de production spécifique supérieure est observée avec les souches WT et $\Delta gppA$ par rapport à celle des souches $\Delta relA$ et $\Delta gppA : \Delta relA$. Ainsi, la vitesse de production spécifique plus importante pour les souches WT et $\Delta gppA$ et l'inhibition du taux de croissance par le ppGpp expliquent les productions spécifiques plus importantes observée avec ces souches. Ce phénomène met donc en évidence l'intervention du ppGpp dans la compétition pour la consommation des acides aminés entre la croissance et la production. L'inhibition de la croissance par le ppGpp rendrait disponible le flux de consommation des acides aminés pour la synthèse de la protéine recombinante.

La vitesse de production spécifique de la protéine Hla par la souche WT semble devenir nulle 2 heures avant la fin de la phase exponentielle de croissance. Cette perte de la capacité à produire la protéine recombinante n'est pas retrouvée au cours des cultures avec les souches mutées. Un lien entre la perte de la capacité de production et la forte production spécifique est observé avec la souche WT. Ce phénomène est également retrouvé au cours du procédé de production de la protéine UL130. En effet, au cours du procédé de production de la protéine UL130, les souches $\Delta relA$ et $\Delta gppA : \Delta relA$ produisent la protéine jusqu'à de fortes productions spécifiques puis continuent leur croissance sans produire de protéines recombinantes (figure 6.3). De même en milieu riche, la souche WT produit la protéine UL130 durant 1 heure atteignant une forte production spécifique puis continue de croître en absence de production mais sans qu'un arrêt de croissance ne soit observé (figure 3.5). La perte de la capacité de production des souches n'est donc pas en lien avec l'apparition d'un arrêt de croissance au cours du procédé. Il semblerait plutôt que l'arrêt de production soit la conséquence d'une production spécifique trop élevée.

6.2.3. Détermination du nombre de copies de plasmides au cours du procédé de production de la protéine UL130

Il est établi dans la littérature que la présence d'ARNt non chargés peut conduire à une accumulation du nombre de copies de plasmide (plasmid copy number, PCN) par perturbation de la régulation de l'origine de répllication ColE1 (Camps 2010b). Les résultats du chapitre 4 et les résultats

présentés ci-dessus suggèrent une augmentation du nombre d'ARNt non chargés au cours des procédés de production des protéines UL130 et Hla. Un tel événement pourrait conduire à la mise en place d'un cercle vicieux. En effet, l'augmentation du nombre d'ARNt non chargés causée par une limitation en acides aminés engendrée par la production de protéines recombinantes conduit à une augmentation de PCN par bactérie. Cette augmentation conduirait à une plus forte expression du gène d'intérêt et donc à une augmentation de la production accentuant la limitation en acides aminés et augmentant le nombre d'ARNt non chargés. Pour vérifier si un tel mécanisme apparaît durant le procédé de production de la protéine UL130, les PCN sont quantifiés. Ces quantifications sont réalisées dans plusieurs conditions afin de vérifier si une limitation en acides aminés peut-être à l'origine d'une augmentation de PCN.

Afin de savoir si l'évolution du PCN a un lien avec la présence d'ARNt non chargés, des quantifications des PCN sont réalisées lors de la production de la protéine UL130 en milieu minimum et en milieu riche avec la souche *E. coli* BL21 WT, la souche *E. coli* BL21 $\Delta gppA$ et la souche $\Delta relA$. La production de la protéine UL130 par la souche WT en milieu riche engendre une faible inhibition de la croissance mettant en évidence une plus faible charge métabolique qu'en milieu minimum (figures 3.5 et 3.6). Etant donné que la quantité d'acides aminés pouvant être assimilée dans le milieu de culture est élevée en milieu riche, la comparaison des quantifications des PCN lors d'une production en milieu riche et en milieu minimum permet de mettre en évidence l'influence de la disponibilité des acides aminés sur les PCN. De plus, la suppression du gène *relA* d'une souche d'*E. coli* entraîne une incapacité de cette souche à détecter la présence d'ARNt non chargés (Cashel et al. 1996). De fait, un mutant $\Delta relA$ est dans l'incapacité de mettre en place les mécanismes d'adaptation adéquats pour réduire le nombre d'ARNt non chargés dans la cellule (Traxler, Summers, H.-T. Nguyen, et al. 2008). Ainsi, la délétion du gène *relA* engendrerait une quantité supérieure d'ARNt non chargés dans la bactérie (Camps 2010b). De fait, si les PCN varient en fonction du nombre d'ARNt non chargés, alors la suppression de *relA* en condition de limitation en acides aminés devrait augmenter les PCN par bactérie.

La quantification de PCN par bactérie est effectuée par le département « analytical research and development » (ARD) de Sanofi Pasteur Europe. La méthode consiste à effectuer une ddPCR (« digital

droplet PCR ») sur le gène *dxs* (gène de la Deoxy-xylulose-P synthase) du génome d'*E. coli* afin de déterminer le nombre de chromosomes et en déduire le nombre de bactéries. Sur le même échantillon, une ddPCR est réalisée sur le gène de résistance à la kanamycine pour quantifier le PCN dans l'échantillon. Le rapport des quantités des deux gènes permet de connaître le PCN moyen par bactérie.

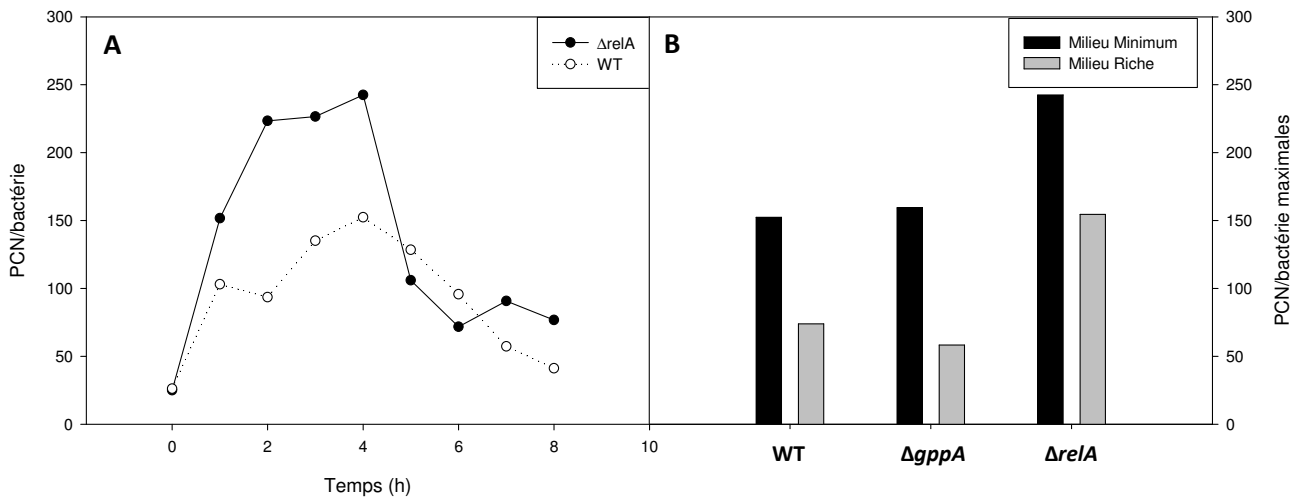


Figure 6.5 : Quantification du nombre de copies de plasmide (PCN) par bactérie au cours des procédés de production de la protéine UL130 en milieu minimum par la souche *E. coli* WT ou mutées. (A) Le PCN par bactérie est quantifié par ddPCR des gènes *kan* et *dxs* au cours du procédé de production de la protéine UL130 avec les souches WT et $\Delta relA$. (B) Valeurs maximales des quantifications du PCN par bactéries obtenues au cours des procédés de production de la protéine UL130 en milieu minimum (noir) ou milieu riche (gris) obtenues avec les souches WT, $\Delta gppA$ et $\Delta relA$.

La mesure du PCN lors de la production de la protéine UL130 en milieu minimum avec les souches *E. coli* BL21 WT et $\Delta relA$ révèle une augmentation du PCN suite à l'induction (figure 6.5, A). La souche *E. coli* BL21 $\Delta relA$ a un profil similaire de variation du PCN mais avec un maximum de PCN 60% plus élevé que la souche WT. La production de la protéine UL130 dans un milieu riche conduit à une diminution du nombre maximal de copies de plasmide retrouvé au cours du procédé par rapport aux quantifications de PCN par bactérie réalisées au cours des procédés en milieu minimum avec les souches WT, $\Delta gppA$ et $\Delta relA$ (figure 6.5, B). Quel que soit le milieu, la souche $\Delta relA$ a un nombre de copies de plasmide plus élevée que celui des souches WT et $\Delta gppA$ dans un même milieu (figure 6.5, A et B).

Le fait que le PCN soit toujours plus élevé en milieu minimum par rapport au milieu riche pour chaque souche révèle que la charge métabolique causée par la production de la protéine recombinante est impliquée dans l'augmentation du PCN. Au vue des quantifications des contraintes métaboliques au cours du procédé de production de la protéine UL130 en milieu minimum, il est très probable que la limitation en acides aminés soit indirectement la cause de cette augmentation du PCN en entraînant l'augmentation du nombre d'ARNt non chargés.

La souche $\Delta relA$ ne détecte pas les ARNt non chargés et n'active pas les mécanismes permettant leur diminution (activation de la synthèse des acides aminés, baisse de la transcription des gènes ribosomiaux, baisse du taux de croissance) (Traxler, Summers, H.-T. Nguyen, et al. 2008). Le nombre plus élevée de copies de plasmide retrouvée avec cette souche dans chaque milieu en comparaison des souches WT et $\Delta gppA$ met en évidence que l'absence d'activation de la réponse stringente conduit à augmenter le PCN par bactérie au cours de la production de la protéine UL130 comme rapporté dans la littérature (Camps 2010b).

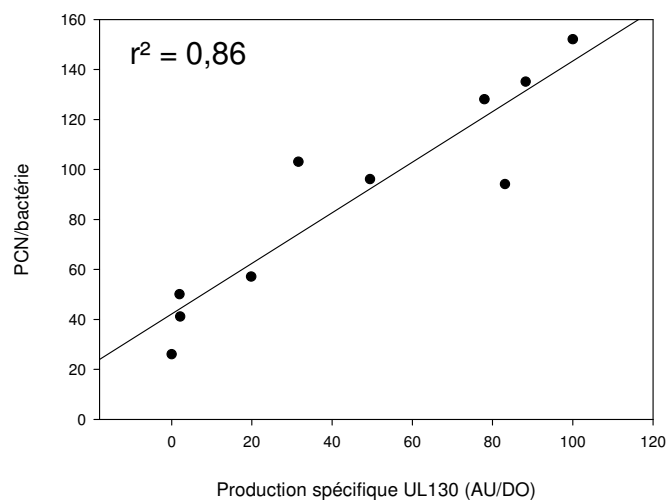


Figure 6.6 : Corrélation entre la production spécifique de la protéine UL130 en milieu minimum par la souche *E. coli* WT et le nombre de copies de plasmides (PCN/bactérie).

Pour la souche WT en milieu minimum, le nombre de PCN est corrélé avec la production spécifique ($r^2 = 0,86$) (figure 6.6). Ce résultat est tout à fait cohérent étant donné que l'augmentation du PCN et la quantité de protéines par unité de biomasse sont liées tous les deux à la vitesse de production et au taux de croissance.

Tous ces résultats mettent donc en évidence que l'augmentation du nombre d'ARNt non chargés conduit à l'augmentation du PCN contenant une origine de réplication de type ColE1. L'augmentation de PCN peut conduire potentiellement à une augmentation de l'expression du gène d'intérêt et donc de la vitesse de production. Cette amplification de la force d'expression du gène par une limitation en acides aminés peut aboutir à un emballement du système d'expression et à l'arrêt de croissance de la souche.

Si le PCN par bactérie augmente consécutivement à l'augmentation du nombre d'ARNt non chargés dans la bactérie et que le nombre d'ARNt non chargés est lié à une limitation en acides aminés alors le PCN peut être utilisé comme indicateur de la limitation en acides aminés dans la bactérie. Ainsi, le PCN plus important observé avec la souche $\Delta relA$ comparé aux souches WT et $\Delta gppA$ durant la production démontre que la limitation en acides aminés est plus importante avec la souche $\Delta relA$. Pourtant, la croissance de la souche $\Delta relA$ est peu inhibée par la production démontrant que l'arrêt de croissance prolongé observée avec les souches WT et $\Delta gppA$ est causé directement par l'accumulation de ppGpp et, dans une bien plus faible mesure, par des dysfonctionnements provoqués par la limitation en acides aminés. Enfin, le nombre élevé de copies de plasmide par bactérie observé avec la souche $\Delta relA$ au cours du procédé de production en milieu riche révèle qu'une limitation en acide aminé intracellulaire apparaît dans cette condition malgré la forte disponibilité en acides aminés dans le milieu de culture.

6.2.4. Quantification de l'exométabolome de souches *E. coli* BL21 mutées ($\Delta gppA$, $\Delta relA$, $\Delta relA : \Delta gppA$)

6.2.4.1. Quantification de l'exométabolome en condition non induite

Dans le but d'étudier le fonctionnement métabolique des souches mutées de gènes impliqués dans la réponse stringente, le taux de croissance, la vitesse spécifique de consommation du glucose (qGlucose) et la vitesse spécifique d'excrétion de l'ensemble des produits sont déterminés lors d'une culture en milieu minimum sans induction. L'objectif est de caractériser le fonctionnement métabolique des souches mutées dans le but de déterminer ensuite si des différences métaboliques entre les souches peuvent influencer leur comportement lors de la production de protéines recombinantes.

Les analyses ont été réalisées par RMN (1D-proton) sur les surnageants de culture. Les vitesses spécifiques de consommation et d'excrétion des métabolites sont déterminées grâce au logiciel Physiofit développé dans le laboratoire (<https://imetatoul.insa-toulouse.fr/software/physiofit/>) (tableau 6.1).

Tableau 6.1 : Vitesses spécifiques (qGlucose, μ ...) déterminées lors d'une croissance de la souche *E. coli* WT et des mutants $\Delta relA$, $\Delta gppA$ et $\Delta relA : \Delta gppA$.

	WT	$\Delta relA$	$\Delta gppA$	$\Delta gppA:\Delta relA$ NI
μ (h^{-1})	0,71 \pm 0,02	0,70 \pm 0,05	0,66 \pm 0,06	0,67 \pm 0,11
qGlc (mmol/L/h)	-7,28 \pm 0,16	-7,39 \pm 0,21	-7,41 \pm 2,8	-7,31 \pm 0,78
qFormate (mmol/L/h)	0,15 \pm 0,01	0,21 \pm 0,03	0,22 \pm 0,11	0,10 \pm 0,01
qPyruvate (mmol/L/h)	0,02 \pm 0,00	0,01 \pm 0,00	0,02 \pm 0,01	0,01 \pm 0,00
qAcetate (mmol/L/h)	0,04 \pm 0,03	0,16 \pm 0,06	0,22 \pm 0,09	0,09 \pm 0,01
qEthanol (mmol/L/h)	0,08 \pm 0,01	0,16 \pm 0,05	0,14 \pm 0,06	0,09 \pm 0,01
qValine (mmol/L/h)	0,03 \pm 0,00	0,05 \pm 0,01	0,04 \pm 0,02	0,01 \pm 0,00

Aucune différence significative n'est observée entre les vitesses spécifiques de la souche WT et celles des souches mutées ($\Delta relA$, $\Delta gppA$, $\Delta relA : \Delta gppA$) (test de student avec $p > 0,05$) (tableau 6.1). Ce résultat permet de mettre en évidence que l'exométabolisme des souches en condition de croissance

en milieu minimum est similaire, ce qui est plutôt cohérent puisque la réponse stringente n'est pas activée dans cette condition.

6.2.4.2. Quantification de l'exométabolome en condition de production de la protéine Hla.

L'étude de l'influence de la concentration en ppGpp sur les cinétiques de croissance et de production de la protéine UL130 met en évidence une phase de production des mutants similaire à la souche WT c'est-à-dire 1 à 2 heures de production après induction. Après cette phase, la reprise de croissance des mutants n'est pas accompagnée de production. La courte phase de production de la protéine UL130 (1 à 2 h) ne permet pas d'observer l'influence de la réponse stringente sur l'exométabolisme en condition de production de la protéine UL130 avec les souches mutées. De plus, les différences pouvant apparaître après la phase de production (accumulation de valine, alanine...) sont engendrées par des différences de croissance entre les souches qui ne présentent pas d'intérêt dans le cadre de cette étude. C'est pourquoi, l'étude de l'influence du ppGpp sur le métabolisme en condition de production sera présentée exclusivement avec le procédé de production de la protéine Hla par la souche WT et les souches mutées ($\Delta relA$, $\Delta gppA$ et $\Delta relA : \Delta gppA$). Ce procédé permet une croissance en condition de production sur une longue période et laisse un temps d'analyse suffisant pour étudier l'influence de la réponse stringente sur le métabolisme en condition de production avec les souches mutées.

Les rendements des métabolites produits dans le surnageant par rapport au glucose consommé obtenus avec les souches mutées en condition de production sont comparés à ceux obtenus avec la souche WT en condition non induite afin d'étudier l'influence de la production et de la réponse stringente sur l'exométabolome. Ces rendements sont également comparés à ceux obtenus avec la souche WT en condition d'induction afin d'étudier plus spécifiquement l'influence de la réponse stringente sur l'exométabolome. Tous les métabolites majoritaires des surnageants de cultures ont été quantifiés et sont présentés (figure 6.7).

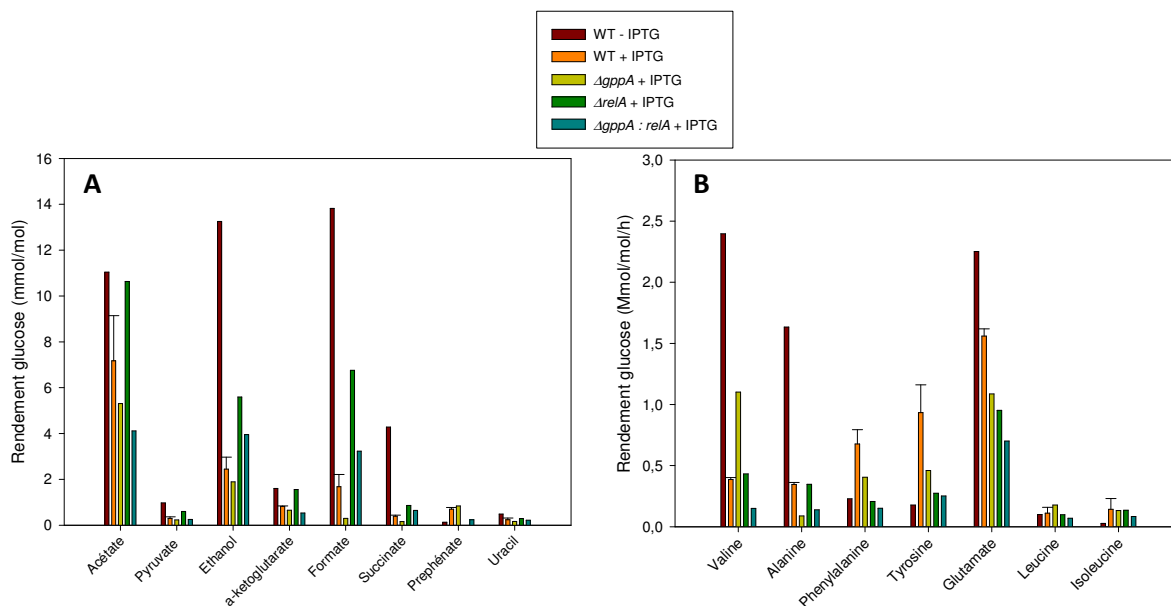


Figure 6.7 : Analyse de l'exométabolome en condition de production de la protéine Hla avec la souche *E. coli* WT et les souches mutées. (A) Rendements des acides organiques et (B) des acides aminés excrétés dans le surnageant par rapport au glucose consommé (mmol/mol).

Les souches mutées et la souche WT en condition d'induction (avec IPTG), ont des rendements de production des métabolites dans le surnageant (mmol/mol) inférieurs ou similaires à ceux observés avec la souche WT en condition non induite (sans IPTG) (figure 6.7, A et B). Des excrétions d'acétate, d'éthanol et de formate dans le surnageant de culture plus importante sont observées avec la souche $\Delta relA$ comparée à la souche WT induite mais ces excrétion restent inférieures à la souche non induite (figure 6.7, A). Les plus faibles rendements d'excrétions d'éthanol, de formate, de valine, d'alanine ou de glutamate observés avec l'ensemble des souches induites comparés aux rendements observés avec la souche non induite (figure 6.7, B) tendent à montrer que la synthèse de la protéine Hla conduit à diminuer l'excrétion de ces métabolites de façon indépendante à l'accumulation de ppGpp.

Ces résultats révèlent que la modification de la réponse stringente ne change pas fortement les rendements d'excrétion des métabolites dans le surnageant et ne permet d'observer l'apparition d'autres métabolites que ceux retrouvés avec la souche WT au cours de la production de la protéine Hla. Il semblerait donc que la modulation du ppGpp n'implique pas de perturbation majeure de

l'exométabolome en condition de production. Il est pourtant connu que la réponse stringente est impliqué dans la régulation du métabolisme et en particulier des voies de biosynthèse des acides aminés lors de son activation (Traxler, Summers, H.-T. Nguyen, et al. 2008). Il n'est donc pas à exclure qu'une modification des flux métaboliques apparaissent suite à l'activation de la réponse stringente en condition de production de la protéine Hla sans que des modifications majeures de l'exométabolome ne soient observables.

6.3. Discussion :

L'objectif du travail présenté dans ce chapitre est d'étudier l'influence de l'augmentation de la concentration en ppGpp dans la souche *E. coli* BL21 lors des procédés de production des protéines UL130 et Hla. Cette étude a mis en évidence les conséquences de l'inhibition de croissance par le ppGpp sur le métabolisme de la souche et sur la production des protéines.

Le modèle d'étude basé sur l'activation de la réponse stringente d'une culture d'*E. coli* BL21 par ajout de SHX a permis d'étudier l'influence de l'accumulation de ppGpp sur la croissance et le métabolisme de la souche sans production de protéines recombinantes. L'inhibition de la croissance suite à l'ajout de SHX met évidence une augmentation des concentrations de presque tous les acides aminés et une excrétion importante dans le surnageant de valine et d'alanine. L'accumulation de ces acides aminés est probablement due à une période durant laquelle la synthèse des acides aminés est supérieure à leur consommation ; leur consommation étant destinée à la synthèse des protéines de la biomasse, inhibée par l'accumulation de ppGpp.

En condition de production de la protéine UL130, l'arrêt de croissance observé avec la souche WT est la conséquence de l'inhibition de la croissance par accumulation de ppGpp. Cette condition peut être comparée à une inhibition de croissance déclenchée par l'accumulation de ppGpp suite à l'ajout de SHX. En effet, dans les deux cas, l'arrêt de croissance est causé principalement par l'accumulation de ppGpp dû à l'activité de la protéine RelA. La comparaison de ces deux conditions permet d'étudier l'effet de la production sur le métabolisme de la souche dans un contexte d'arrêt de croissance causé par une accumulation de ppGpp. Lors de l'ajout de SHX à une culture de la souche *E. coli* BL21 WT (sans production), une accumulation de l'ensemble des acides aminés est observée. Cette accumulation est

probablement le résultat de l'inhibition de la croissance se traduisant par l'arrêt de consommation des acides aminés destinés à la synthèse des protéines de la biomasse. Lors de la production de la protéine UL130 par la souche *E. coli* BL21 WT, une accumulation de certains acides aminés est observée au moment de l'arrêt de croissance alors que la concentration d'autres acides aminés diminue. Si tous les acides aminés ne s'accumulent pas en condition d'arrêt de croissance et de production de la protéine UL130 alors la diminution de la concentration d'une partie des acides aminés est expliquée par leur consommation destinée à la synthèse de la protéine recombinante. De fait, l'accumulation de ppGpp inhibe la croissance et rend disponible le flux de consommation des acides aminés destinée à la croissance au profit de la synthèse de la protéine recombinante. La disponibilité du flux des acides aminés destinée à la croissance permet à la production de continuer durant 1 heure après l'arrêt de croissance de la souche WT. En effet, lorsque le ppGpp n'est pas synthétisé (souche $\Delta relA$ et $\Delta gppA$: $\Delta relA$) la production s'arrête dès la fin de la première heure d'induction. De plus, cette disponibilité du flux de synthèse des acides aminés par inhibition de la croissance par l'accumulation de ppGpp permet également d'augmenter la vitesse de production spécifique au cours du procédé de production de la protéine Hla. De fait, le ppGpp interfère dans la compétition pour la consommation des acides aminés en faveur de la production et en défaveur de la croissance. Ce phénomène se poursuit probablement jusqu'à ce que la vitesse de synthèse de la protéine recombinante soit limitée par la vitesse de transcription du gène d'intérêt. Dans le cas du procédé de production de la protéine Hla par la souche WT, l'inhibition de la croissance par le ppGpp se stabilise à un μ de $0,3 \text{ h}^{-1}$ signifiant qu'à ce moment la vitesse de production est limitée par un autre facteur, probablement par la vitesse de transcription du gène *hla*. Dans le cas de la production de la protéine UL130, la vitesse de transcription du gène d'intérêt semble toujours plus élevée que la vitesse de traduction conduisant à un arrêt de croissance de la souche.

Dans le procédé de production de la protéine Hla, une quantité plus importante de ppGpp dans la souche WT comparée aux souches mutées entraîne un taux de croissance plus faible durant la production. La diminution du taux de croissance engendre une augmentation de la production spécifique. De fortes productions spécifiques sont intéressantes pour un procédé car cela permet d'augmenter la productivité et de faciliter la purification de la protéine recombinante d'intérêt. En revanche, de trop fortes productions spécifiques semblent conduire à la perte de la capacité de

production des souches. Comme décrit dans la littérature, une production spécifique trop importante conduit à un dysfonctionnement traductionnelle de la bactérie (Scott et al. 2010; Bosdriesz et al. 2015). Ainsi, l'accumulation de ppGpp est propice pour maximiser les rendements de production d'un procédé en permettant d'atteindre de fortes production spécifique mais est également négatif pour l'objectif de maintenir une croissance en condition de production sur des temps longs. A partir de ces observations, le contrôle de la concentration intracellulaire en ppGpp au cours de la production paraît crucial afin d'optimiser la production spécifique de la protéine recombinante tout en assurant la conservation de la croissance jusqu'à la fin programmée du procédé.

Les quantifications de plasmides réalisées au cours du procédé de production d'UL130 tendent à montrer l'apparition d'un autre mécanisme participant à l'arrêt de croissance par une forte production. La production de la protéine UL130 conduit à l'augmentation du nombre d'ARNt non chargés par une consommation forte des acides aminés. Ces ARNt non chargés entraînent l'accumulation de plasmides et ainsi une potentielle augmentation de l'expression du gène *UL130*. Cette augmentation de l'expression du gène d'intérêt peut conduire à l'augmentation de la production et à l'augmentation de la quantité d'ARNt non chargés. Ce mécanisme conduit donc à une amplification de la force d'expression de gène d'intérêt lorsque la production entraîne une limitation en acides aminés amplifiant de fait la limitation et conduisant à l'arrêt de croissance.

6.3.1. Vers un système biologique robuste et performant pour la production de protéines recombinantes

Les résultats de l'étude de l'influence de l'accumulation de ppGpp sur la croissance et la production de protéine recombinante permettent de formuler 3 conclusions importantes pour la suite du travail : i) une production spécifique très élevée au cours du procédé conduit à un arrêt définitif de la production, ii) l'augmentation de la concentration intracellulaire en ppGpp intervient dans la compétition entre la croissance et la production pour la consommation des acides aminés libres en faveur de la production et au détriment de la croissance et, iii) une limitation en acides aminés libres conduit à une augmentation du nombre de copies de plasmide par bactérie pouvant entraîner une plus grande force d'expression du vecteur et aboutir à la première conclusion.

Ces 3 phénomènes peuvent être expliqués et corrigés dans le but d'améliorer les procédés. En ce qui concerne la première observation, une production spécifique excessive est causée par une vitesse de production trop élevée comparée au taux de croissance. La force d'expression du gène d'intérêt est donc trop élevée dès le début du procédé.

L'intervention du ppGpp dans la compétition entre la croissance et production pour la consommation des ressources intracellulaire peut avoir un aspect positif et négatif. Cette action est positive lors d'une accumulation intermédiaire du ppGpp (correspondant à 1,4 $\mu\text{mol/g}$ suivant le maximum atteint durant la production de la protéine Hla par la souche WT) car elle permet de libérer des ressources pour la production et d'augmenter la production spécifique. Elle est également négative car elle peut conduire à un arrêt total de la croissance lorsque la concentration intracellulaire en ppGpp dépasse un certain seuil (entre 1,4 $\mu\text{mol/g}$ et 2,5 $\mu\text{mol/g}$ suivant les concentrations en ppGpp maximales respectivement atteintes au cours des productions des protéines Hla et UL130). Ce problème est dû au fait que les systèmes d'expression ne sont soumis à aucun mécanisme de régulation cellulaire et s'exprime de façon indépendante du fonctionnement cellulaire.

L'augmentation du nombre de copies de plasmide est probablement liée à l'apparition d'ARNt non chargés lorsque la compétition entre croissance et production pour la consommation des acides aminés est importante. Cette accumulation des ARNt non chargés est causée par une vitesse de production trop importante par rapport aux capacités de la souche à synthétiser des acides aminés. L'accumulation de PCN pourrait être résolue par une modification de l'origine de réplication ou en évitant l'apparition d'une limitation en acides aminés au cours du procédé. Dans la suite du travail, le maintien d'une vitesse de production spécifique en accord avec les capacités métaboliques de l'hôte a été l'option privilégiée.

L'amélioration de performance de robustesse du système d'expression et du procédé pourrait donc être obtenue par une diminution de la force d'expression du système d'expression lorsque la vitesse de production dépasse les capacités métaboliques de l'hôte grâce une intégration de sa régulation au fonctionnement cellulaire. Le rôle de la réponse stringente est de réguler le taux de croissance, le métabolisme et la traduction en fonction de la disponibilité des ressources intracellulaires (Potrykus & Cashel 2008). De plus, le ppGpp est l'une des rares molécules à posséder une action de régulation

transcriptionnelle (Zuo et al. 2013). Son accumulation pourrait être utilisée pour contrôler négativement l'expression génique du système T7. Le système T7 repose sur l'intégration du système transcriptionnel du prophage T7 dans le système transcriptionnel de la bactérie *E.coli*. L'objectif est d'intégrer la régulation transcriptionnelle du système d'expression aux mécanismes de régulation traductionnelle et métabolique déjà en place dans la bactérie. Ainsi l'expression du système T7 serait régulée en fonction de la disponibilité des ressources métaboliques intracellulaires de façon à coupler la croissance et la vitesse de production en fonction des capacités métaboliques instantanées de la bactérie. L'accumulation de ppGpp étant déclenchée par des carences en acides aminés, en azote, en phosphate, en fer, par déformation de la membrane de la bactérie et par choc osmotique, le couplage de la croissance avec la production serait alors générique à l'ensemble de ces conséquences potentiellement induites sur l'hôte par la production de protéines recombinantes. Cette intégration de la régulation transcriptionnelle du système T7 à la réponse stringente est présentée dans le chapitre 7.

Chapitre 7 - Optimisation de la performance et de la robustesse du système biologique de production

Les études réalisées dans les chapitres précédents ont mis en évidence que la vitesse de production de la protéine UL130 est trop élevée comparée à la capacité de la souche *E. coli* BL21 à fournir les ressources nécessaires sa synthèse. En particulier, une limitation en certains acides aminés au cours du procédé déclenche la synthèse de ppGpp par la protéine RelA. L'accumulation de ppGpp conduit à une inhibition de la croissance bactérienne. Le flux de consommation des acides aminés utilisé normalement pour la synthèse des protéines de la biomasse est alors réorienté vers la synthèse de la protéine recombinante. L'intervention du ppGpp dans la compétition pour la consommation des acides aminés entre croissance et production conduit à une augmentation de la vitesse de production spécifique. L'augmentation de la vitesse de production spécifique en protéine recombinante combinée à une croissance inhibée conduit à augmenter l'enrichissement de la biomasse en protéine recombinante. Une production spécifique trop élevée conduit à l'arrêt de croissance et l'arrêt définitif de la production. De façon parallèle à ce mécanisme, la limitation en acides aminés conduit indirectement à une augmentation du nombre de copies de plasmide et donc de l'expression du gène d'intérêt, accentuant le déplacement de l'équilibre entre croissance et production vers la production.

L'objectif du travail présenté dans ce chapitre est de développer un système biologique de production de protéine recombinante « générique », c'est à dire capable de conserver un équilibre entre croissance et production quelle que soit la protéine recombinante. Cet aspect est essentiel pour i) éviter l'arrêt de production de la protéine, ii) augmenter la production totale, iii) maîtriser la qualité de la protéine produite et, iv) développer des procédés robustes de production de types « semi-continus » ou « continus ». La stratégie innovante qui a été mise en œuvre pour atteindre cet objectif consiste à intégrer au système de régulation transcriptionnelle et traductionnelle de la bactérie, le système de régulation de l'expression de la protéine recombinante afin que la vitesse de production de la protéine s'adapte en temps réel, aux capacités instantanées de synthèse de la bactérie. L'objectif est donc de disposer d'un outil générique et robuste pour la production de protéines hétérologues chez *E. coli*.

7.1. Principe des souches MetaProd

Une série de souches (dénommées souches MetaProd dans le reste du chapitre pour illustrer leur capacité à coupler la production avec les capacités métaboliques) ont été générées en y intégrant de nouveaux circuits de régulation de la production de la protéine d'intérêt, circuits régulés par la concentration intracellulaire en ppGpp. La croissance et la machinerie traductionnelle étant inhibées par le ppGpp, il est essentiel que le ppGpp puisse contrôler la production en cas de dépassement des capacités de synthèse de protéines (figure 7.1).

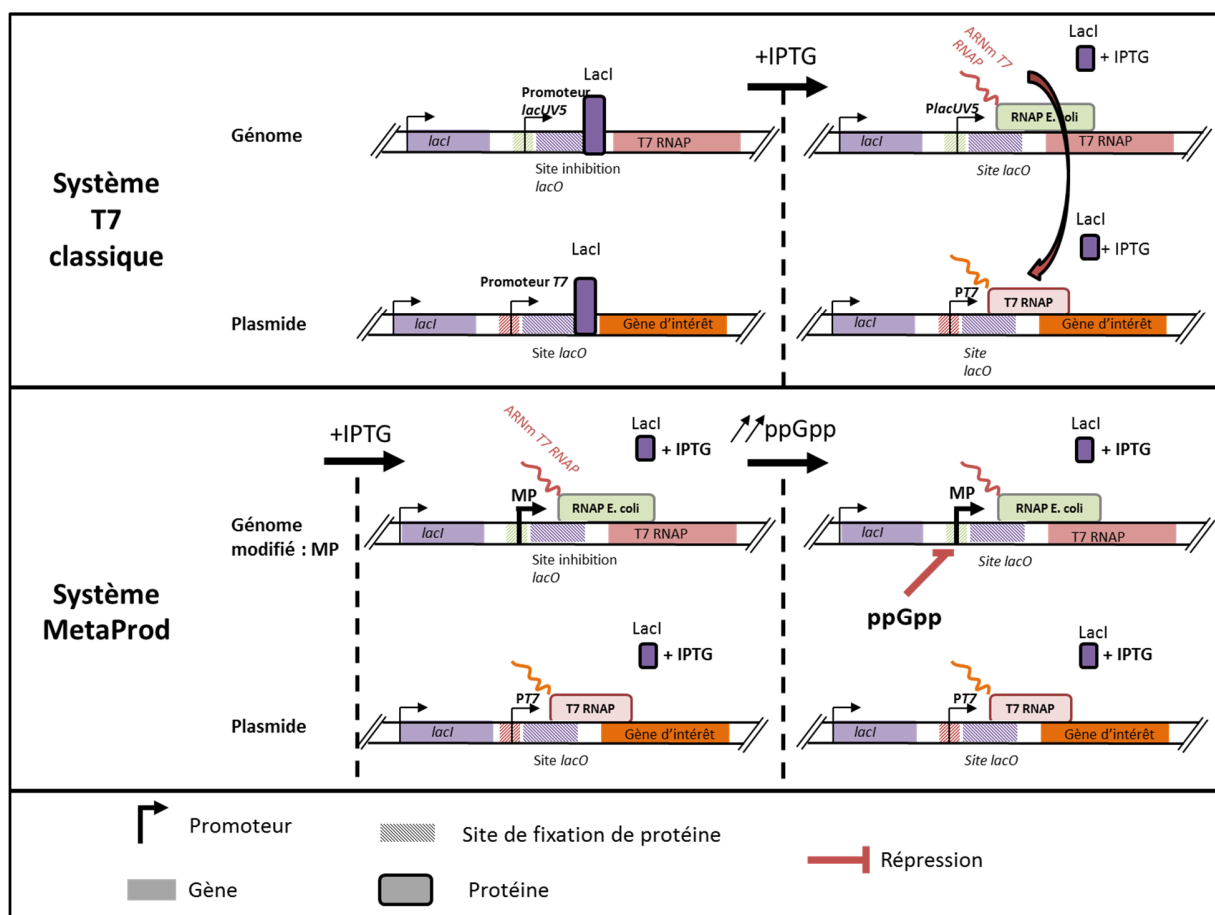


Figure 7.1 : Principe du système d'expression T7 (haut) avec la souche MetaProd (bas). Le promoteur de la T7 RNAP (promoteur *lacUV5*) est remplacé par un promoteur d'opéron ribosomal (*rrn P1*) inhibé par le *ppGpp*.

Le *ppGpp* est l'une des rares molécules ayant une activité transcriptionnelle (Potrykus & Cashel 2008). La variation de sa concentration entraîne une modification transcriptionnelle profonde *via* son

interaction avec la protéine DksA et l'ARN polymérase d'*E.coli*. Ainsi, de nombreux promoteurs sont soit activés soit inhibés lorsque la concentration intracellulaire en ppGpp augmente. La force d'expression des promoteurs des gènes impliqués dans la phase exponentielle de croissance est modifiée par la concentration intracellulaire en ppGpp lorsque le facteur sigma RpoD interagit avec RNAP (Potrykus & Cashel 2008). La stratégie est donc d'utiliser ces promoteurs et de les intégrer dans le système d'expression T7. Dans un premier temps, le promoteur *lacUV5* de la T7 ARN polymérase (RNAP) est remplacé par un promoteur inhibé par le ppGpp (figure 7.1).

7.1.1. Choix des promoteurs d'intérêts

A une concentration basale de ppGpp intracellulaire, la force d'expression du promoteur utilisé pour remplacer le promoteur *lacUV5* de la T7 RNAP doit être similaire au promoteur *lacUV5*. Ce promoteur doit également être inhibé lors de l'accumulation de ppGpp. De préférence, le promoteur ne doit pas être soumis à des régulations transcriptionnelles autres que le niveau intracellulaire de ppGpp et la présence du facteur sigma de la phase exponentielle de croissance RpoD. La sélection du promoteur est effectuée sur la base des études existantes dans la littérature scientifique.

Sur la base d'une analyse bibliographique et d'une veille technologique, des promoteurs possédant les propriétés explicitées ci-dessus, les promoteurs des ARN ribosomiaux (ARNr) ont été sélectionnés pour remplacer le promoteur *lacUV5*. Ce choix se justifie par plusieurs raisons. La première est que les promoteurs ribosomiaux ont été fortement étudiés par la communauté scientifique en raison de leur importance pour la physiologie bactérienne. Leur force d'expression à une concentration basale de ppGpp a été comparée ainsi que l'inhibition de leur expression par accumulation de ppGpp (Barker et al. 2001; Bradley et al. 2017; Hirvonen et al. 2001; Kolmsee et al. 2011; Lemke et al. 2011; Maeda et al. 2015; Meng et al. 2001; Murray et al. 2003; Zhang et al. 2002; Zhao et al. 2011; Zhang et al. 2006). Deuxièmement, la croissance est limitée par la capacité des bactéries à transformer les nutriments de l'environnement en protéines intracellulaires (mécanisme appelé fitness du μ). Le fitness du μ est donc caractérisé par une adaptation de la quantité de ribosomes par cellule en fonction du taux de croissance (Bosdriesz et al. 2015; Bremer & Dennis 2008; Hui et al. 2015; Scott et al. 2014; Scott et al. 2010; Waschina et al. 2016). Les promoteurs ribosomiaux

sont très sensibles à la concentration intracellulaire en ppGpp avec une forte répression lors de l'accumulation de ppGpp. Dans le but de coupler la croissance avec la production, il est intéressant de coupler la régulation de la machine traductionnelle avec la force d'expression du gène d'intérêt. Enfin, l'étude de la force d'expression de 18 promoteurs impliqués dans la synthèse de protéines et d'ARN constituant les ribosomes a révélé que les promoteurs des ARNr constitutifs des ribosomes étaient plus inhibés par la réponse stringente que les promoteurs des gènes codant pour les protéines ribosomales (Lemke et al. 2011). Cette analyse bibliographique permet donc de postuler que l'utilisation des promoteurs des ARNr pour contrôler l'expression de la T7 RNAP permettraient de diminuer la production de façon efficace et simultanée à la synthèse de ribosomes (et donc au taux de croissance) lors du dépassement de la capacité traductionnelle de la bactérie caractérisé par l'accumulation de ppGpp.

7.1.2. Caractéristiques des promoteurs des opérons ribosomaux d'*E. coli*.

Chez *E. coli*, les gènes des ARNr (16S, 23S et 5S) sont regroupés en sept opérons (*rrnA*, *rrnB*, *rrnC*, *rrnD*, *rrnE*, *rrnG* et *rrnH*). Ces opérons ont des séquences très proches. Des micro-hétérogénéités de séquences, en particulier au niveau des promoteurs de ces opérons, provoquent des forces d'expression différentes (Maeda et al. 2015). Ces opérons sont contrôlés au niveau transcriptionnel par deux promoteurs positionnés en tandem, le promoteur P1 et le promoteur P2. L'expression du promoteur P2 est peu sensible à la concentration en ppGpp (Zhang et al. 2002). Le promoteur P1 est celui contrôlant l'expression des opérons en condition exponentielle de croissance et c'est lui qui est le plus soumis à la régulation du ppGpp (Zhang et al. 2006).

Dans la publication de Kolmsee et al. 2011, la force de transcription des promoteurs P1 de 7 opérons présents dans *E. coli* MG1655 est déterminée (figure 7.2, A).

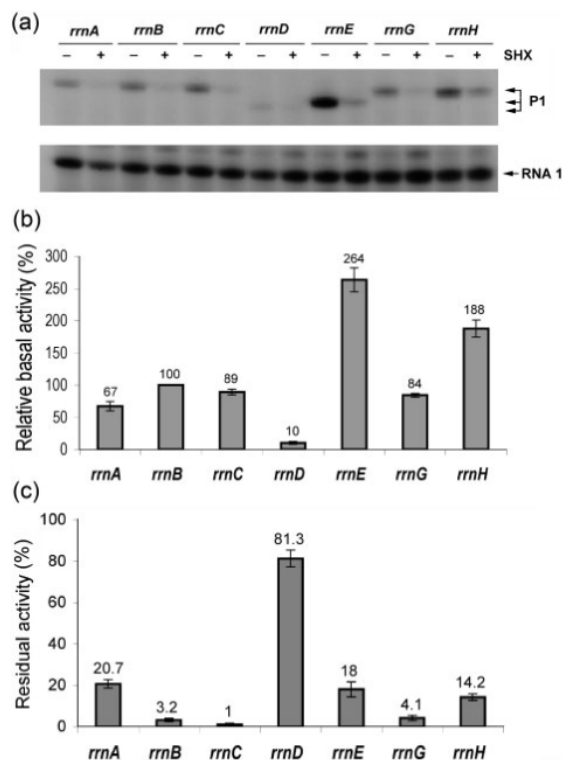


Figure 7.2 : Force d'expression des promoteurs des sept opérons ribosomiaux d'*E. coli* (*rrn*), issus de Kolmsee et al. 2011. (a) La force d'expression de chaque promoteur a été quantifiée après migration sur gel des produits d'amplification des ARNm (b) avant ou (c) après ajout de SHX à 1mg/mL de façon relative à leur niveau basal.

La figure 7.2, (a et b) met en évidence la grande différence d'expression entre les promoteurs de chaque opéron, avec une expression maximale observée pour *rrnE* et minimale pour *rrnD*. Les promoteurs des opérons *rrnA*, *B*, *C*, *G* et *H* ont des expressions intermédiaires. En activant la réponse stringente grâce à l'ajout de SHX, les différences de répression des promoteurs par le ppGpp sont évaluées (figure 7.2, (a) et (c)). Il apparaît que les promoteurs des opérons *rrnB*, *C* et *G* sont les plus réprimés à l'inverse du promoteur de l'opéron *rrnD*.

A partir de ces résultats, les promoteurs des opérons *rrnB*, *rrnD* et *rrnE* (appelés par la suite promoteurs *rrnB*, *rrnD* et *rrnE*) possèdent 3 niveaux de transcription basals très différents. En effet, le promoteur *rrnB* est très fortement inhibé par le ppGpp, le promoteur *rrnE* moyennement et le promoteur *rrnD* faiblement. Ces trois promoteurs possèdent des caractéristiques différentes qui

permettront d'optimiser les chances d'obtenir des caractéristiques compatibles avec le rétrocontrôle du système T7.

7.1.3. Description des promoteurs *rrn* P1

Les promoteurs *rrnB*, *rrnD* et *rrnE* natifs d'*E. coli* possèdent des caractéristiques différentes. Une modification de ces promoteurs pourrait permettre d'augmenter le spectre de force d'expression basale des promoteurs pour augmenter encore les chances d'obtenir un promoteur avec une force d'expression initiale similaire au promoteur *lacUV5*.

Les promoteurs P1 ribosomaux sont composés de 3 domaines : l'élément UP (car présent en amont (« upstream ») de nombreux promoteurs), le « core » et des séquences de fixation du cofacteur transcriptionnel FIS (factor for inversion stimulation). A un taux de croissance élevé, cette protéine permet d'augmenter l'expression des gènes impliqués dans la croissance exponentielle de la bactérie (dont les ribosomes) (Bradley et al. 2017). Plusieurs sites de fixation de FIS sont présents au niveau des promoteurs ribosomaux. Le site I de fixation de FIS est le site avec le plus d'influence sur la régulation positive de la transcription comparé aux autres sites de fixation de FIS (Hirvonen et al. 2001). L'élément UP est présent en amont de nombreux promoteurs d'*E. coli* (Meng et al. 2001). Sa séquence permet une amélioration de l'affinité avec RNAP et en particulier avec la sous unité α de sa partie C-terminale (Hirvonen et al. 2001). Enfin, le « core » est la partie du promoteur permettant la fixation de RNAP avant l'initiation de la transcription. Le travail de Hirvonen et al., a permis de caractériser la force de transcription des 7 promoteurs des opérons ribosomaux d'*E.coli* en enlevant séquentiellement les sites de fixation FIS et l'UP élément. Ces résultats ont été générés en plaçant les promoteurs en amont du gène codant pour la β -galactosidase. Ainsi, l'activité mesurée de cette protéine est directement liée à la force de son promoteur. Les résultats mettent en évidence que la suppression d'un élément des promoteurs ribosomaux diminue d'un facteur 10 la force d'expression du gène contrôlé par les promoteurs à un niveau basal de ppGpp (Hirvonen et al. 2001). De plus, la suppression de la séquence de fixation FIS engendre une baisse de l'expression des promoteurs *rrnB*, *rrnD* et *rrnE* tout en conservant leur différences relatives de force d'expression.

Dans la publication de Barker et al. 2001 la force d'expression des promoteurs *LacUV5* et *rrnB* de -66 à +9 (élément UP et le « core » de *rrnB* P1) est quantifiée en mesurant l'activité de la β -galactosidase placée en aval des promoteurs. Une inhibition de l'expression du promoteur *rrnB* P1 est mise en évidence lorsque le taux de croissance diminue alors que l'expression du promoteur *lacUV5* reste constante. Les résultats indiquent qu'à un taux de croissance de $0,7 \text{ h}^{-1}$ la force d'expression du promoteur *rrnB* est trois fois plus faible que celle du promoteur *lacUV5*. Un résultat équivalent est retrouvé dans la publication de Lemke et al. 2011. Ces études indiquent que l'utilisation du promoteur P1 *rrnB* sans séquence FIS ou avec séquence FIS permettrait d'obtenir respectivement une force d'expression plus faible ou supérieure au promoteur *lacUV5* en absence d'accumulation de ppGpp en condition de croissance exponentielle.

Sur la base de la recherche bibliographique, deux types de construction de promoteurs ont été réalisés : les promoteurs contenant le site I FIS, l'élément UP et le site « core » et des promoteurs sans séquence FIS, avec l'élément UP et « core ». Ces deux constructions de promoteurs sont appliquées sur trois variants de promoteurs P1 d'opérons ribosomiaux. Ces promoteurs ont été sélectionnés en fonction de leur différence d'expression à un niveau basal et leur force d'inhibition : *rrnB*, *rrnD*, *rrnE*. Un site de 400pb sans utilité apparente entre le promoteur *lacUV5* et le gène T7 RNAP est supprimé lors des constructions. La suppression de cet élément pourrait engendrer une augmentation de la force d'expression du promoteur c'est pourquoi une construction sans modification du promoteur *lacUV5* sera effectué avec une suppression des 400pb (souche appelée WT-400pb) (figure 7.3).

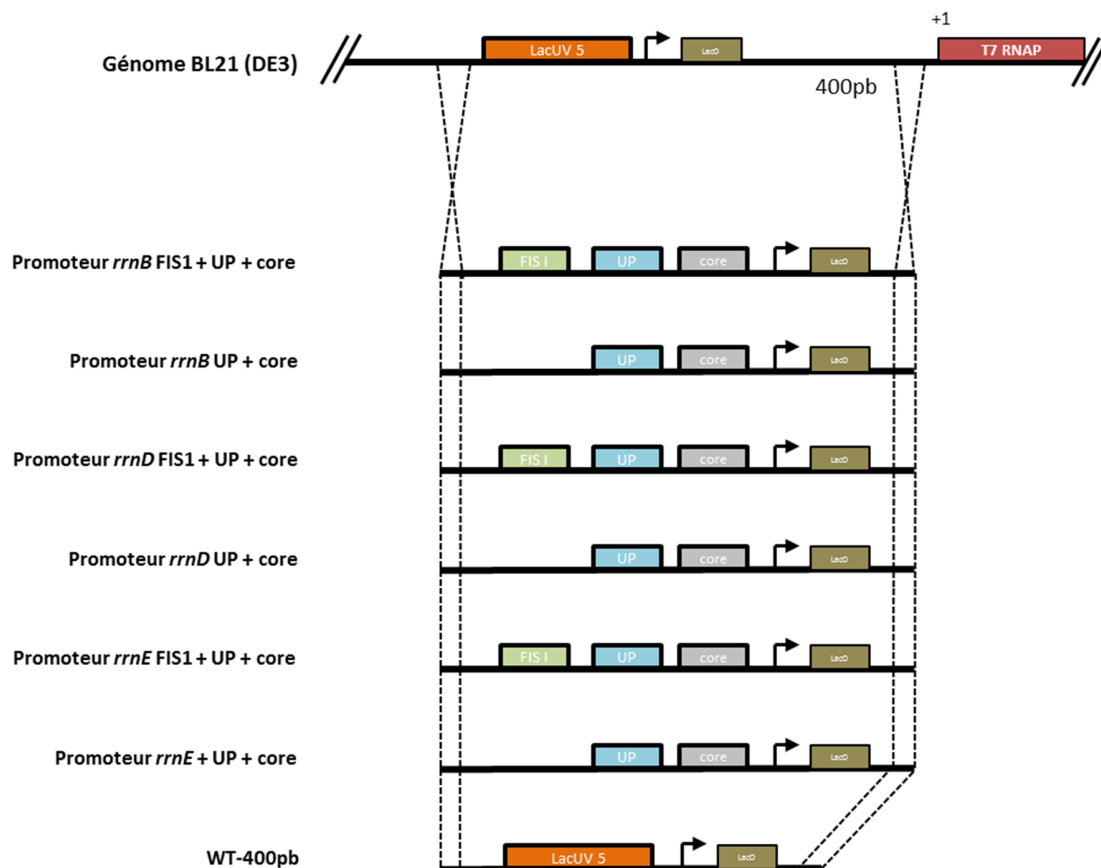


Figure 7.3 : Séquences des promoteurs *rrnB*, *rrnD* et *rrnE* choisis pour remplacer le promoteur *lacUV5* avant le gène de la T7 RNAP de la souche *E. coli* BL21 (DE3). Les rectangles représentent les éléments fonctionnels des séquences (site de fixation de FIS, élément UP, « core » et le site d'inhibition lacO) et la flèche le site d'initiation de la transcription.

7.1.4. Etude de la performance des souches MetaProd

Dans le but d'éviter des problèmes de variabilités pouvant apparaître entre différents lots de transformation et ainsi de pouvoir comparer les souches entre elles, les souches MetaProd et les souches WT (WT et WT-400pb) ont été transformées en parallèles par les plasmides codant les protéines UL130 et Hla. Suite à cette nouvelle transformation, un impact plus important de la production de la protéine UL130 et Hla sur la croissance est observé sur les souches WT.

L'objectif est de caractériser la croissance et la capacité de production des souches MetaProd en condition de production des protéines UL130 et Hla selon, i) leur capacité à maintenir la phase de

croissance et de production après induction et, ii) la quantité de protéines recombinantes produites. Du fait que ces souches possèdent un promoteur de la T7 RNAP activé en présence d'IPTG et inhibé lors de l'augmentation de la concentration en ppGpp, ces souches devraient maintenir une plus faible vitesse production spécifique que la souche WT ou la réduire lorsque la croissance devient trop faible.

Le changement du promoteur *lacUV5* par les promoteurs des opérons ribosomiaux engendre potentiellement une modification du niveau basal d'expression de la T7 RNAP pour une même quantité d'IPTG. Ainsi, la réalisation des productions à 2 concentrations d'IPTG permet d'évaluer la sensibilité des promoteurs au changement de concentration en inducteur. Aucune fuite d'expression n'a été observée avant ajout d'IPTG dans les souches MetaProd ou WT

La souche WT et la souche WT-400pb sont comparées en condition de production avec 1mM d'IPTG. Dans cette condition les souches *rrnD*, FIS-*rrnB*, *rrnB* et *rrnE* ont été comparées aux souches WT. La souche FIS - *rrnE* n'a pas été obtenue. Du fait que la suppression des 400pb de la souche WT n'est pas révélée d'influence sur la croissance et la production, seule la souche WT-400pb est utilisée dans le procédé de production de protéines induites avec 0,1 mM d'IPTG pour servir de témoins à l'ensemble des souches MetaProd obtenues (figure 7.4).

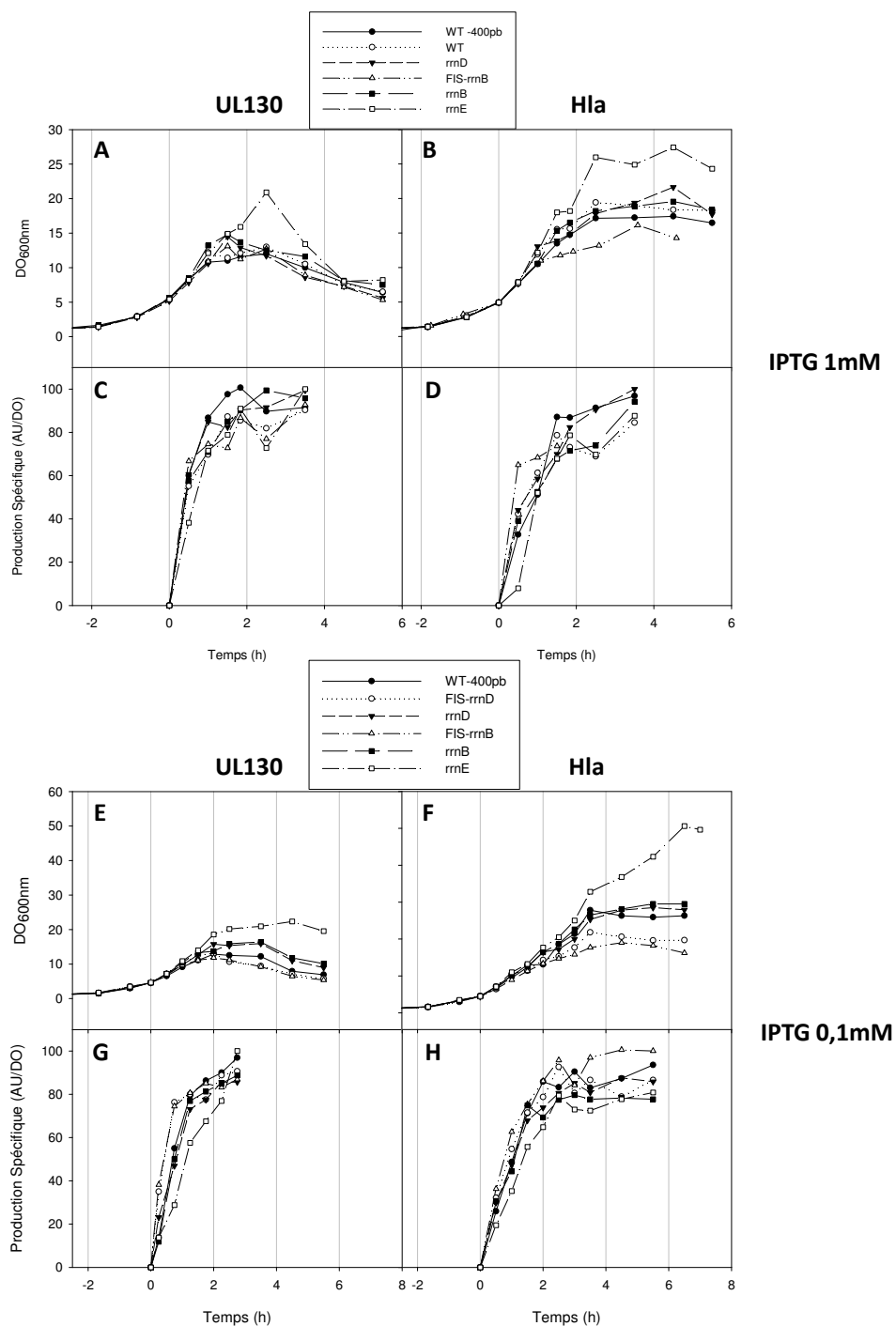


Figure 7.4 : Analyse des cinétiques de croissance et de production spécifique au cours des procédés de production des protéines UL130 (gauche) et Hla (droite) par les souches *E. coli* BL21 WT, WT-400 pb et MetaProd (FIS-*rrnB*, *rrnB*, *rrnD*, *rrnE*) avec ajout de 1 mM d’IPTG (haut) ou des souches *E. coli* BL21 WT-400 pb et MetaProd (FIS-*rrnB*, *rrnB*, FIS-*rrnD*, *rrnD*, *rrnE*) avec ajout de 0,1 mM d’IPTG (bas) à T=0 h. Biomasse (DO) (A, B, E, F) et productions spécifiques (AU/DO) (C, D, G, H) en fonction du temps.

Les cinétiques de croissance et de production des souches WT et WT-400 pb sont identiques lors de la production de la protéine UL130 et Hla après induction à 1mM d'IPTG (figure 7.4, A-B-C-D). La suppression de la séquence de 400 pb entre le promoteur et le gène codant pour la T7RNAP n'influence donc pas significativement la force d'expression du promoteur.

Au cours des procédés de production des protéines UL130 ou Hla, un arrêt de croissance est observé pour l'ensemble des souches 1 à 3 h après l'induction à 1 mM ou 0,1 mM d'IPTG, sauf pour la souche *rrnE* pour laquelle la croissance dure 1 h de plus que pour les autres souches lors de la production des protéines UL103 ou Hla induite avec 1 mM d'IPTG et 3 h de plus lors de l'induction de la production des protéines à 0,1mM d'IPTG. A l'inverse, un ralentissement de croissance apparaît plus précocement pour les souches FIS-*rrnB* et FIS-*rrnD* lors de la production de Hla à 0,1mM. Sur la base du premier critère (maintien de la croissance), la souche *rrnE* semble être la plus performante.

La quantification de la production spécifique au cours du temps permet d'observer que l'ensemble des souches ont des valeurs de production spécifique maximales proches lors de leur arrêt de croissance. La souche *rrnE* semble se démarquer par une production spécifique plus basse que les autres souches durant la première heure d'induction lors de la production de la protéine UL130 et Hla après induction à 1 mM d'IPTG et lors des 2 premières heures lors de la production de ces protéines après induction à 0,1 mM d'IPTG. Les productions spécifiques maximales étant proches entre toutes les souches, la production totale dépend surtout de la DO maximale atteinte au moment de l'arrêt de croissance. Puisque la souche *rrnE* atteint une biomasse 30 à 70% plus élevée que les autres souches, cette souche produit 30 à 70 % de plus de protéines UL130 ou Hla que la souche WT dans chaque condition. Ainsi, la souche *rrnE* maintient une phase de croissance et de production plus longue que les autres souches permettant d'atteindre une production totale plus élevée.

Cette évaluation met en évidence un allongement de la phase de production et de croissance de la souche *rrnE* après induction par rapport à celle de l'ensemble des souches, la rendant plus performante sur la base du premier critère d'évaluation. L'allongement de la phase de croissance et de production permet à la souche *rrnE* d'atteindre une production totale plus élevée que les autres souches, la rendant de fait plus performante sur la base du deuxième critère d'évaluation.

La prolongation de la phase de croissance pourrait être en lien avec une production spécifique plus faible dans les premiers instants après induction. La mise en place rapide de la production spécifique dans l'ensemble des souches indique que les forces des promoteurs *rrn* P1 du gène de la T7 RNAP sont assez proches de la force du promoteur *lacUV5* de la souche WT. D'après les résultats obtenus lors de l'induction à 0,1 mM de la protéine Hla, la présence de la séquence FIS dans les promoteurs des opérons ribosomiaux conduit à une augmentation plus rapide de la production spécifique après ajout d'IPTG que celle de la souche WT et engendre un arrêt plus précoce de la croissance. L'ajout de la séquence FIS aux promoteurs ribosomiaux est connu pour augmenter leur force d'expression (Hirvonen et al. 2001). La présence de la séquence FIS semble donc engendrer une force d'expression du promoteur plus forte que le promoteur *lacUV5* de la souche WT. Un arrêt de croissance est visible pour toutes les souches après l'induction. Ce résultat signifie que l'inhibition de ppGpp sur les promoteurs des souches MetaProd est soit, pas assez efficace, ou s'applique tardivement par rapport à la dynamique d'arrêt de croissance.

Pour la suite de l'étude, la souche *rrnE* est sélectionnée sur la base des critères cités précédemment. La souche *rrnB* (sans séquence FIS) est également sélectionnée car les données bibliographiques mettent en évidence la capacité du promoteur *rrnB* à être efficacement inhibé lors de l'accumulation de ppGpp et donc son fort potentiel pour un système MetaProd (Hirvonen et al. 2001; Meng et al. 2001; Zhang et al. 2002; Zhang et al. 2006; Zhao et al. 2011; Kolmsee et al. 2011; Lemke et al. 2011). La baisse de concentration en IPTG permet de faire apparaître de plus grandes différences entre les souches. Afin de pouvoir évaluer si un rétrocontrôle du ppGpp sur la production apparaît durant les procédés, la concentration en IPTG sera diminuée à 0,1mM pour le reste du chapitre.

7.2. Stabilisation du nombre de copies de plasmide par supplémentation en acides aminés

7.2.1. Effet du nombre de copies de plasmide sur la performance des souches MetaProd.

Le chapitre 6 a révélé que le nombre de copies de plasmide (PCN) par bactérie augmente au moment de la production de la protéine UL130 par la souche WT. L'augmentation du PCN pourrait engendrer une perte d'efficacité de l'inhibition du ppGpp sur le système d'expression. En effet, l'augmentation du nombre de copies de plasmide est connu pour être un facteur permettant d'augmenter la force d'expression du gène d'intérêt (Rosano & Ceccarelli 2014). De fait, l'augmentation de la force d'expression génique pourrait s'opposer à l'inhibition de la production par l'accumulation de ppGpp intracellulaire. Les quantifications du PCN par bactérie au cours des procédés de production des protéines UL130 et Hla par les souches *rrnE* et WT sont comparées. L'objectif est d'analyser l'évolution du PCN au cours des procédés de production avec les deux souches afin de vérifier si la cinétique du PCN est en lien avec les cinétiques de croissances et de production analysées précédemment.

Seules les conditions de production des protéines UL130 et Hla par les souches *rrnE* et WT induit à 0,1mM d'IPTG sont comparées car les croissances observées dans ces conditions sont suffisamment discriminantes pour mettre en évidence une différence au niveau du PCN par bactérie au cours du temps. Les échantillons utilisés pour la quantification des PCN ont été prélevés au cours des cultures présentées précédemment et les mesures ont été effectuées au moins jusqu'à l'arrêt de croissance des souches (figure 7.5).

La mesure du CER est exploitée pour mesurer l'évolution de l'activité métabolique inhérente à la croissance durant cette étude (figure 7.5). Cette analyse a l'avantage d'être beaucoup plus résolutive que la mesure de la DO grâce à une quantification effectuée toutes les deux minutes.

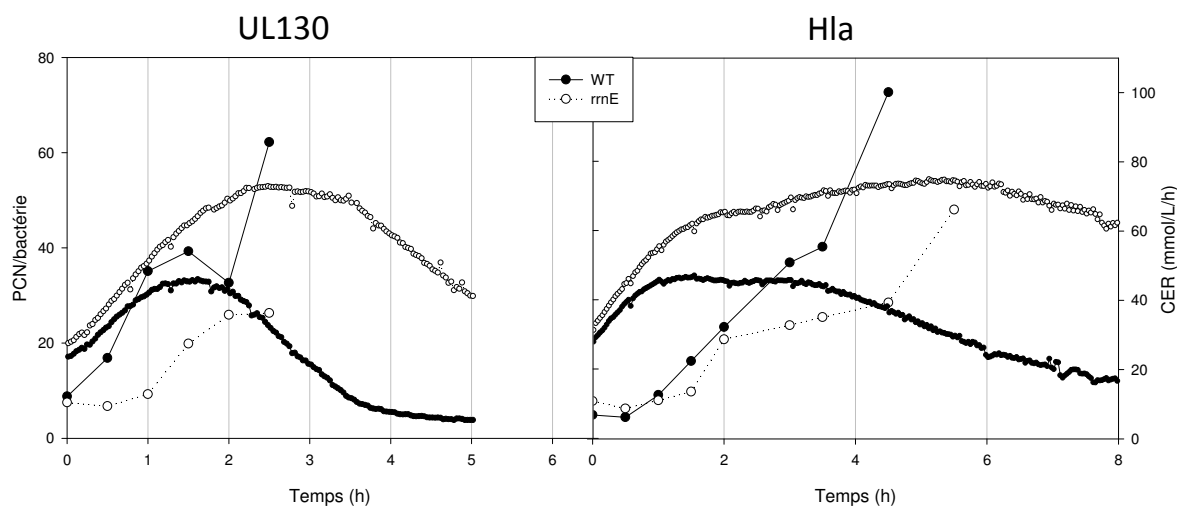


Figure 7.5 : Quantification du nombre de copies de plasmide (PCN) par bactérie et analyse du CER lors des procédés de production des protéines UL130 ou Hla par les souches *E. coli* BL21 WT (ronds noirs) ou *rrnE* (ronds blancs).

Ces résultats montrent une augmentation du PCN par bactérie pour les deux souches après induction de la production des protéines UL130 et Hla quand le CER augmente (figure 7.5). De manière générale, il est observé que le PCN augmente plus rapidement en fin de culture lorsque le CER diminue ou se stabilise. Toutes les quantifications du PCN effectuées sur la souche *rrnE* après 1 heure d'induction sont plus faibles que pour la souche WT.

La forte augmentation du PCN lors de l'arrêt de croissance observé avec le CER appuie fortement l'hypothèse que l'augmentation des PCN conduit à une augmentation de la production contribuant à l'inhibition de croissance. La répression de la production par une accumulation de ppGpp pourrait être insuffisante pour contrebalancer l'augmentation de la force d'expression du gène d'intérêt liée au PCN.

L'augmentation en deux temps des PCN par bactérie durant le procédé de production de Hla par la souche *rrnE* tend à montrer qu'un mécanisme permet de ralentir l'augmentation de PCN par bactérie en comparaison à la souche WT (figure 7.5). Un ralentissement de l'augmentation des PCN signifie que la quantité d'ARNt non chargés est maintenue plus basse dans la souche *rrnE* probablement grâce à un ralentissement de la production de Hla. Ce mécanisme pourrait être la conséquence recherchée du rétrocontrôle de la production par le ppGpp.

7.2.2. Effet de la supplémentation en acides aminés sur le procédé de production de la protéine UL130 par la souche WT.

De nombreux articles mettent en évidence que l'augmentation des PCN par bactérie est causée par la présence d'ARNt non chargés interférant avec l'origine de réplication du plasmide de type ColE1 (Camps 2010b; Cserjan-Puschmann et al. 1999; Rodriguez et al. 2017; Wang et al. 2002; Wrobel & We 1998). Les travaux présentés dans le chapitre 6 vont clairement dans cette direction (figure 6.5). En particulier, les ARNt de 10 acides aminés sont décrits comme étant responsables de l'augmentation du PCN lorsqu'ils ne sont pas chargés (Camps 2010). L'ajout, au moment de l'induction, d'un mélange d'acides aminés dont les ARNt non chargés causent l'augmentation des PCN pourrait permettre d'éviter l'apparition de ce phénomène et donc de stabiliser le PCN par bactérie.

L'objectif de ce travail est de maintenir l'efficacité de contrôle du ppGpp sur la production et d'améliorer la stabilité de la croissance et de la production des souches MetaProd par supplémentation en acides aminés au moment de l'induction. L'ajout des 10 acides aminés décrits comme étant responsables en cas de carence de l'augmentation du nombre de copies de plasmide par bactérie est comparé dans un premier temps à un ajout des 20 acides aminés au cours d'une production d'UL130 avec la souche WT. Cet ajout d'acides aminés a un double objectif. Le premier est de retarder l'apparition d'une carence des acides aminés dont les ARNt non chargés entraînent l'augmentation des PCN. Si cette méthode permet de freiner l'augmentation des PCN lors d'une production par la souche WT, alors elle pourra être employée avec les souches MetaProd afin d'améliorer l'efficacité de l'inhibition du ppGpp sur l'expression du système T7. Le deuxième objectif est d'étudier l'effet d'une supplémentation en acides aminés sur la cinétique de croissance et la production de la souche WT. Cette stratégie pourrait être une base d'optimisation des performances des procédés de production basée sur l'addition d'acides aminés.

L'ajout des acides aminés est réalisé au moment de l'induction puis 1 heure plus tard, puis toutes les 30 minutes durant 1 heure. La supplémentation de 10 acides aminés (histidine, tyrosine, glycine, méthionine, lysine, arginine, thréonine, valine, isoleucine et leucine) dont les ARNt non chargés sont décrits comme responsables de l'augmentation du PCN est appelée condition « AA PCN » et la

supplémentation des 20 acides aminés est appelée condition « AA TOT ». Le surnageant a été prélevé au cours de la culture et les acides aminés ont été quantifiés par RMN (figure 7.6).

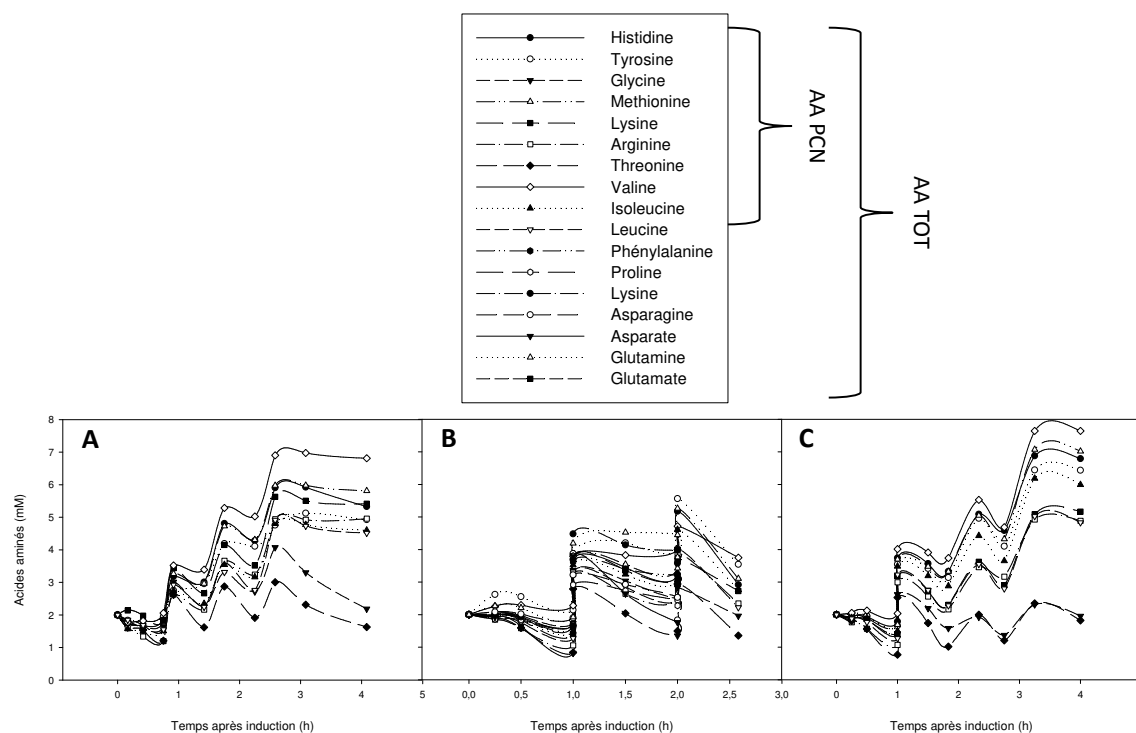


Figure 7.6 : Quantification des acides aminés dans le surnageant par RMN au cours des procédés de production de la protéine UL130 avec la souche WT en condition AA PCN (A), ou AA TOT (B) ou avec la souche *rrnE* en condition AA PCN (C). La condition AA PCN correspond à l'ajout d'histidine, de tyrosine, de glycine, de méthionine, de lysine, d'arginine, de thréonine, de valine, d'isoleucine et de leucine au moment de l'induction (T=0h), une heure après induction et toutes les demis heures pendant 1h30. La condition AA TOT correspond à l'ajout de tous les acides aminés aux mêmes temps que la condition AA PCN.

Parmi les acides aminés détectés (10/10 acides aminés pour AA PCN et 17/20 acides aminés pour AA TOT : sérine, tryptophane et alanine non mesurés), les concentrations mesurées ne sont jamais inférieures à 0,7mM durant toutes les cultures à partir de l'induction (figure 7.6).

Le ppGpp a également été quantifié de façon absolue par la méthode dual-IDMS par LC-MS au cours des cultures sans ajout d'acides aminés et avec l'ajout des acides aminés dit AA PCN (figure 7.7).

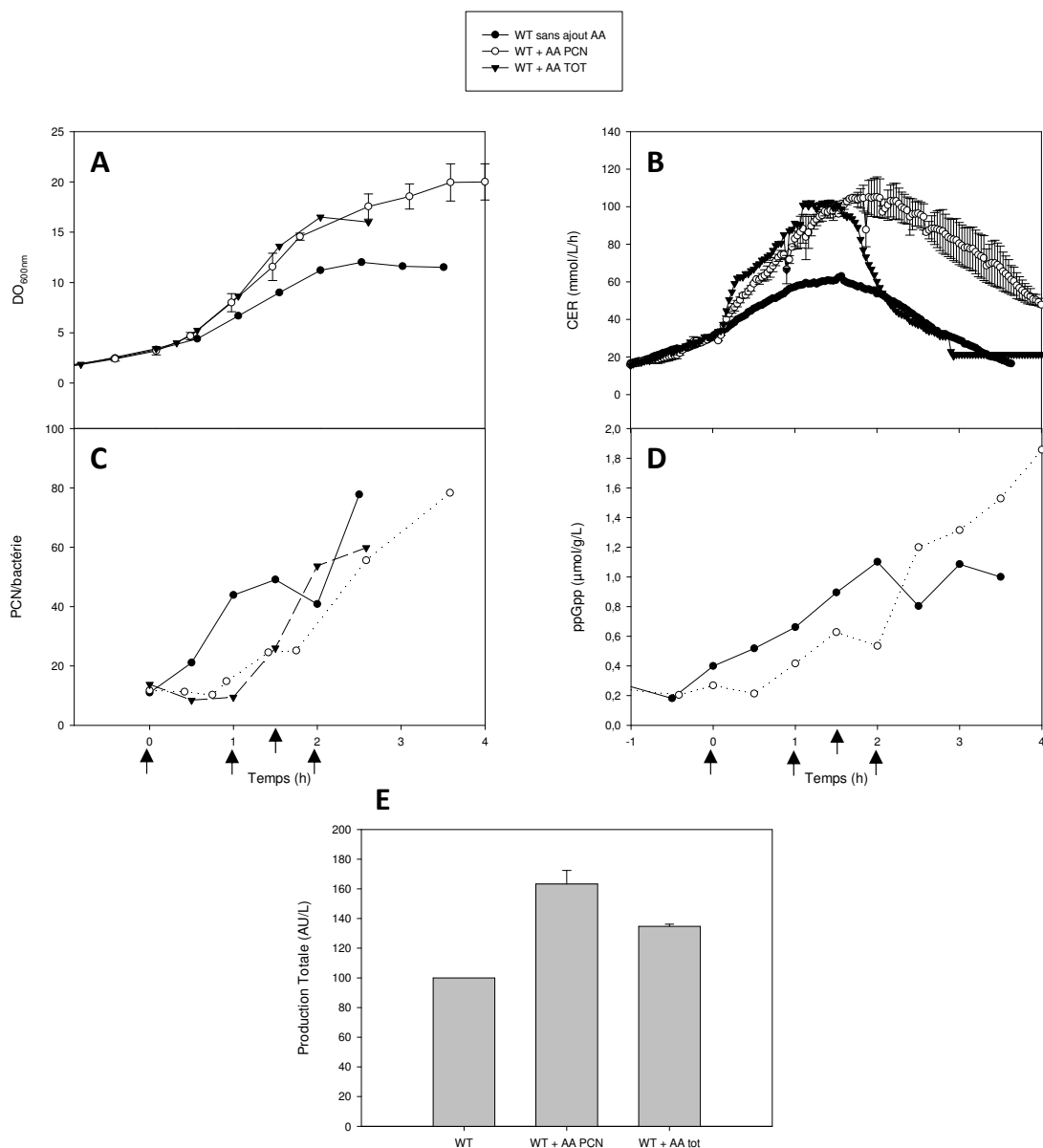


Figure 7.7 : Effet de la supplémentation en acides aminés sur le procédé de production de la protéine UL130 par la souche WT après induction à l’IPTG (0,1mM). (A) Biomasse (DO), (B) CER (mmol/L/h), (C) nombre de copies de plasmide (PCN), (D) concentration en ppGpp intracellulaire ($\mu\text{mol/g}$) en fonction du temps et production totale finales (AU/L). Ces mesures ont été réalisées au cours du procédé de production de la protéine UL130 sans ajout d’acides aminés (« sans ajout AA »), avec l’ajout des 10 acides aminés dont les ARNt non chargés sont impliqués dans l’augmentation du nombre de copies de plasmide par interaction avec l’origine de répllication *ColE1* (« AA PCN » : L-valine, L-méthionine, glycine, L-tyrosine, L-leucine, L-isoleucine, L-lysine, L-arginine, L-thréonine, L-histidine) ou les 20 acides aminés (« AA TOT »). Les flèches noires définissent les moments d’ajout des acides aminés. La culture de la souche WT avec ajout des acides aminés AA PCN a été réalisée en duplicat..

L'ajout de l'ensemble des acides aminés au moment de l'induction (AA TOT) conduit à une augmentation plus rapide de la DO après induction par rapport à celle de la condition sans ajout d'acides aminés (figure 7.7, A). Cependant, l'arrêt de croissance dans cette condition est concomitant à celui de la condition sans ajout d'acides aminés (figure 7.7, B). Comme dans l'ensemble des procédés en milieu minimum présentés dans cette thèse, l'arrêt de croissance est concomitant avec une production spécifique maximale et similaire entre les souches (données non présentées). Ainsi, la quantité de biomasse plus importante en fin de procédé permet de générer une quantité de protéines plus importante (figure 7.7, E). L'ajout d'un nombre d'acides aminés plus restreint décrits comme étant capables de stabiliser les PCN (AA PCN) permet de prolonger de 1h30 la phase de croissance et de production par rapport à la condition sans ajout d'acides aminés. Cette condition permet alors d'augmenter de 60 % la production totale en protéine UL130 par rapport à celle de la condition sans acides aminés et de 20 % par rapport celle de la condition AA TOT. La mesure du CER permet d'observer une forte augmentation de la quantité de CO₂ produit après ajout des acides aminés. De façon surprenante, le CER diminue beaucoup plus rapidement au moment de l'arrêt de croissance lorsque les 20 acides aminés ont été ajoutés plutôt que 10. La quantification des PCN permet d'observer que l'ajout des 10 acides aminés PCN ou des 20 acides aminés permet une stabilisation significative du PCN pendant 1 heure de plus que sans ajout d'acides aminés (figure 7.7). L'accumulation de ppGpp intracellulaire est observée 1 heure plus tard dans la condition AA PCN que dans la condition témoin sans ajout d'acides aminés. Ce retard d'une heure représente une stabilisation significative du PCN et du ppGpp au regard du temps total de croissance en condition de production des procédés (2h30 en condition sans ajout d'acides aminés et 4h dans la condition AA PCN).

Ces résultats démontrent l'efficacité de la méthode d'ajout des acides aminés pour stabiliser le PCN durant les premières heures de production de la protéine UL130 par la souche WT. L'ajout d'acides aminés au moment de l'induction a également un effet important sur la croissance après induction. Cet effet est clairement visible sur le CER avec une augmentation de l'activité métabolique dès l'ajout des acides aminés.

Une production totale 60% plus élevée dans la condition AA PCN par rapport à celle en absence d'ajout d'acides aminés est atteinte par une augmentation de 60% de la biomasse générés avec une production spécifique identique. La plus grande quantité de biomasse générée avant l'arrêt de croissance permet donc d'augmenter significativement la quantité totale de protéine recombinante produite. Le fait que la supplémentation en acides aminés conduise à une augmentation de l'activité métabolique et de la production démontre que la disponibilité des acides aminés est une contrainte majeure de la croissance et de la production. Des investigations supplémentaires seraient nécessaires pour expliquer la meilleure croissance et la production supplémentaire obtenue lorsque 10 acides aminés sont ajoutés plutôt que les 20.

Au cours du chapitre 3, la comparaison du procédé de production de référence en milieu minimum et en milieu riche a révélé que les productions totales dans ces conditions étaient similaires, voire supérieures en milieu minimum (figures 4.5 et 4.6). Une augmentation de 60% de la production avec une supplémentation en acides aminés à un milieu minimum signifie que cette méthode permet d'obtenir des quantités supérieures de production par rapport à celle d'un procédé en milieu riche. Les 40 g/L d'extrait de levures composant le milieu riche représente une quantité d'acides aminés disponible dans le milieu largement supérieure à la quantité d'acides aminés ajoutés en milieu minimum. La disponibilité des acides aminés dans le milieu de culture ne peut donc pas expliquer la plus forte production totale obtenue en milieu minimum supplémenté en acides aminés. Cette forte production pourrait provenir du fait que la culture en milieu minimum a permis aux bactéries de développer les voies de biosynthèse des acides aminés avant l'induction. Au moment de l'induction, la production engendre une baisse des concentrations intracellulaires de certains acides aminés. Cependant, puisque des acides aminés sont ajoutés au moment de l'induction dans le milieu minimum, la baisse des concentrations intracellulaires pourrait être momentanément compensée par une assimilation des acides aminés du milieu s'ajoutant à leur biosynthèse. L'ajout de la biosynthèse des acides aminés, en plus de leur assimilation, pourrait donc expliquer l'augmentation de la croissance et de la production du procédé en milieu minimum par rapport à la croissance et la production retrouvée avec un procédé en milieu riche.

Il est surprenant de constater que la DO continue d'augmenter malgré la diminution du CER 2 heures après induction dans la condition AA PCN. Un tel découplage entre la cinétique de croissance et le CER pourrait être expliqué par un biais de DO à partir de 2 heures d'induction ou/et par une modification profonde du métabolisme entraînant un découplage important entre l'anabolisme et le catabolisme.

7.2.3. Comparaison de la souche WT avec les souches MetaProd *rrnE* et *rrnB* en condition de production de la protéine UL130 lors de la stabilisation du plasmide par supplémentation en acides aminés.

L'étude de l'effet de la supplémentation en acides aminés au cours du procédé de production de la protéine UL130 par la souche WT a permis de démontrer que l'ajout de 10 acides aminés permet de limiter l'augmentation du PCN par bactérie durant le procédé. Cette méthode est maintenant appliquée aux souches MetaProd *rrnE* et *rrnB* lors de la production de la protéine UL130. L'objectif est d'étudier l'influence de l'accumulation de ppGpp sur la production et la croissance afin de déterminer si la stabilisation des PCN engendrée par l'ajout des acides aminés de la condition AA PCN permet d'améliorer le mécanisme de rétrocontrôle du ppGpp sur le système d'expression génique.

La concentration intracellulaire de ppGpp sera un indicateur de la charge métabolique dans la bactérie et a été quantifiée par la méthode dual-IDMS présentée chapitre 3 (figure 7.8).

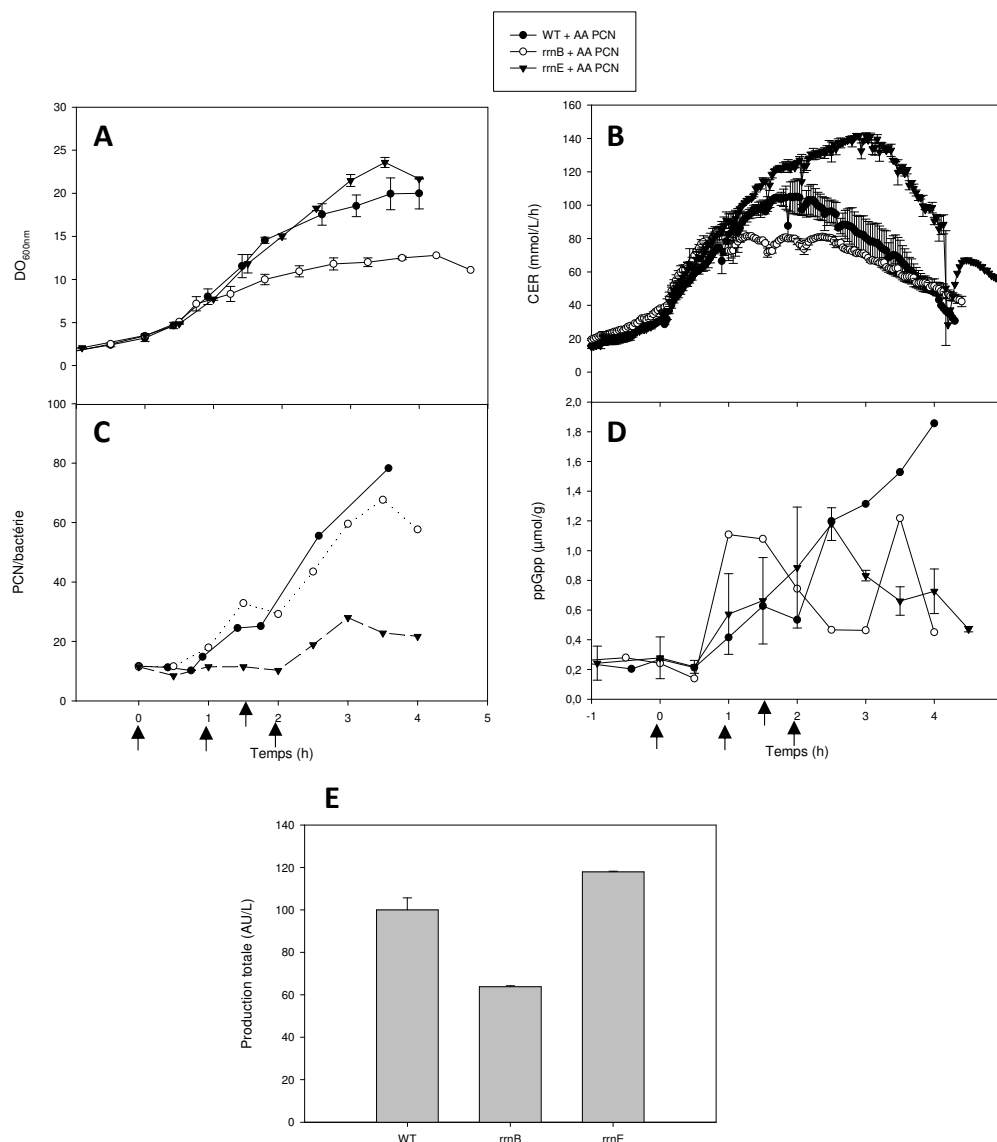


Figure 7.8 : Comparaison des souches WT, *rrnB* et *rrnE* avec ajout d'acides aminés stabilisant le nombre de copies de plasmide (AA PCN) au cours du procédé de production de la protéine UL130 (IPTG 0.1mM). (A) Biomasse (DO), (B) CER (mmol/L/h), (C) PCN par bactérie (AU/L), (D) concentration en ppGpp ($\mu\text{mol/gCDW}$) en fonction du temps et, (E) production totale finale (AU/L). La condition « AA PCN » correspond à l'ajout des 10 acides aminés dont les ARNt non chargés sont impliqués dans l'augmentation du nombre de copies de plasmide par interaction avec l'origine de réplique *ColE1* (« AA PCN » : L-valine, L-méthionine, glycine, L-tyrosine, L-leucine, L-isoleucine, L-lysine, L-arginine, L-thréonine, L-histidine). Les flèches noires définissent les moments d'ajout des acides aminés. La culture de la souche WT avec ajout des acides aminés AA PCN a été réalisée en duplicat. Les procédés de production ont été réalisés en duplicat avec les souches WT et *rrnE* et la quantification du ppGpp est réalisé en duplicat pour le procédé de production avec la souche *rrnE*.

Les PCN au cours des procédés de production de la protéine UL130 supplémentées en acides aminés sont, par rapport à ceux de la souche WT, similaires pour la souche *rrnB* ou plus faibles avec les souches *rrnE* (figure 7.8, C). Ainsi, la méthode de supplémentation en acides aminés PCN au cours des procédés permet de limiter l'augmentation du PCN au moins aussi efficacement avec les souches MetaProd qu'avec la souche WT (figure 7.8, C et 7.5).

Le suivi de la croissance par mesure de la DO ne permet pas de distinguer de forte différence entre la souche *rrnE* et la souche WT (figure 7.8, A). Via cette mesure, la croissance de la souche *rrnB* est en revanche fortement réduite suite à l'induction. L'évolution du CER permet de constater le maintien de l'augmentation de l'activité métabolique 2 heures de plus pour la souche *rrnE* que pour la souche WT (figure 7.8, B). Le CER de la souche *rrnB* est constant après 1 heure d'induction pendant 2 heures. Ce plateau atteint par le CER n'a encore jamais été observée avec la souche WT pour laquelle le CER augmente ou diminue mais n'est jamais stabilisé.

De manière très intéressante, l'évolution des concentrations intracellulaires en ppGpp sont très différentes entre la souche WT et les souches MetaProd (figure 7.8, D). Pour la souche WT, la concentration intracellulaire de ppGpp ne cesse d'augmenter au cours de la production de la protéine UL130 jusqu'à 1,95 $\mu\text{mol/g}$ de ppGpp. En revanche, la concentration en ppGpp des souches MetaProd ne dépasse jamais 1,2 $\mu\text{mol/g}$ et a tendance à diminuer une fois cette valeur atteinte.

L'efficacité du rétrocontrôle est démontrée par une diminution du niveau de ppGpp après son accumulation dans les souches *rrnE* et *rrnB*. La quantification du ppGpp permet d'observer l'adaptation des souches *rrnE* et *rrnB* à la production permettant, pour *rrnE*, la stabilisation de la croissance et de la production durant 1h30 et pour *rrnB* le maintien d'une activité métabolique constante. Les mécanismes décrits dans le chapitre précédent, entraînant l'arrêt de croissance prolongé de la souche WT lors de la production de protéine UL130 causés par l'accumulation de ppGpp sont retardés grâce à l'utilisation des souches MetaProd dans ces conditions. La meilleure adaptation de la souche *rrnE* à la production est retrouvée de façon encore plus évidente avec la production de la protéine HLA.

7.2.4. Comparaison des souches WT et MetaProd *rrnE* en condition de production de la protéine Hla lors de la stabilisation du plasmide par supplémentation en acides aminés.

Un arrêt de croissance a été constaté lors de la production de la protéine Hla par *rrnE*, 2,5 h plus tard que la souche WT (sans ajout d'acides aminés) (figure 7.4, F). L'ajout d'acides aminés PCN a démontré son efficacité pour retarder l'augmentation du PCN après induction dans les deux souches et de façon plus efficace dans la souche *rrnE* lors de la production de la protéine UL130. Cette méthode est appliquée au procédé de production de la protéine Hla avec exclusivement les souches WT et *rrnE* afin d'évaluer l'efficacité d'inhibition du ppGpp durant la production de Hla par les souches MetaProd lorsque le plasmide est stabilisé.

Les déterminations du PCN n'ont pas été réalisées durant cette culture. Seuls les résultats de croissance et les quantifications de ppGpp sont présentés (figure 7.9).

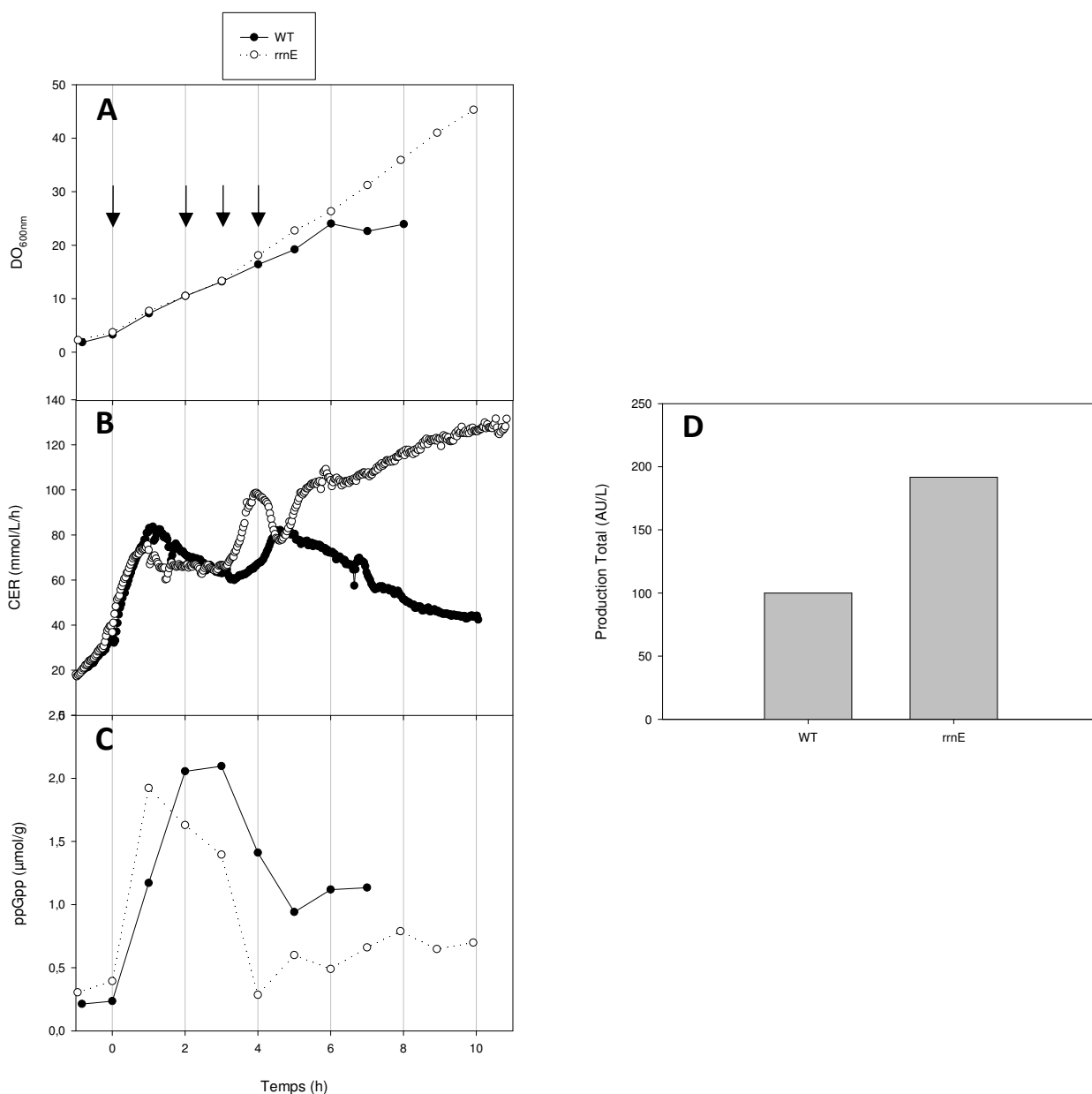


Figure 7.9 : Comparaison des souches WT et *rrnE* avec ajout d'acides aminés (AA PCN) permettant de stabiliser le nombre de copies de plasmide (PCN) par bactérie au cours du procédé de production de la protéine Hla (IPTG 0.1mM). (A) Biomasse (DO), (B) CER (mmol/L/h), (C) concentration en ppGpp (μmol/gCDW) en fonction du temps, (D) et production totale finale des procédés (AU/L).

L'ajout des acides aminés dans la condition AA PCN conduit à une différence de comportement très significative entre la souche WT et la souche MetaProd *rrnE* (figure 7.9). En effet, la stabilisation de la croissance est observée pendant 4 heures supplémentaires par rapport à la croissance de la souche WT (figure 7.9, A-B). Après 5 heures d'induction, la stabilisation du procédé est nettement visible avec l'analyse du CER, diminuant pour la souche WT et augmentant pour la souche *rrnE* (figure 7.9, B). La quantification du ppGpp permet de mettre en évidence la capacité de la souche *rrnE* à s'adapter durant le procédé (figure 7.9, C). En effet, l'accumulation de ppGpp est visible pour les deux souches en début de procédé. Cependant, la concentration en ppGpp diminue rapidement à un niveau plus bas pour la souche *rrnE* que la souche WT, se maintenant par la suite à ce niveau. L'adaptation de la souche MetaProd à la production de la protéine Hla permet alors de multiplier la production par un facteur 2 (figure 7.9, D).

Ces résultats prouvent l'efficacité de la répression du promoteur de l'opéron ribosomal *rrnE* par le ppGpp et l'intérêt de ce concept pour obtenir la robustesse et la stabilité de la production. La cause principale de l'arrêt de croissance de la souche *rrnE* a été clairement identifiée. Il s'agit de l'augmentation du nombre de copies de plasmide par bactérie diminuant l'efficacité de répression du ppGpp sur l'expression du gène d'intérêt. Ainsi, une stabilisation du nombre de copie de plasmides et l'intégration de la régulation du système T7 du plasmide au système MetaProd semble nécessaire pour garantir la robustesse de ce châssis pour la production de protéines recombinantes.

7.3. Principe des plasmides MetaProd

La génération des souches MetaProd et le ralentissement de l'augmentation du PCN par supplémentation en acides aminés ont apporté les preuves de l'intérêt d'un rétrocontrôle du ppGpp sur la production. Cependant le maintien d'une production stable n'est pas assuré en raison de l'augmentation du PCN. Dans le but d'assurer la stabilité de la production et de la croissance, une deuxième stratégie MetaProd a été appliquée par modification du plasmide (figure 7.10).

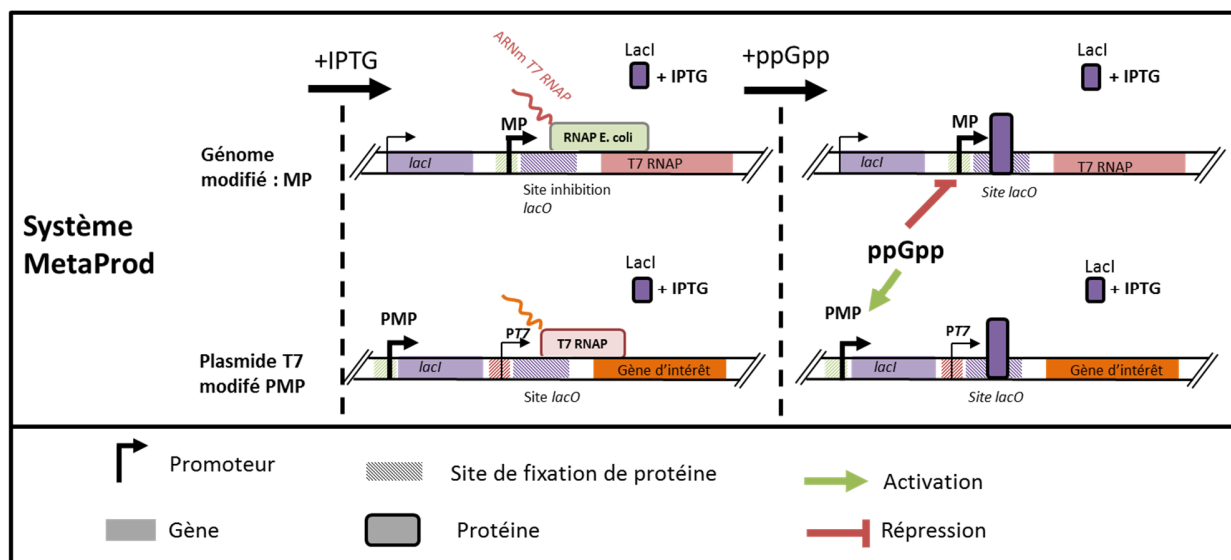


Figure 7.10 : Principe du système d'expression MetaProd. En plus du remplacement de promoteur lacUV5 par un promoteur des opérons ribosomiaux (souches MetaProd, MP), le promoteur du gène *lacI* est remplacé par une série de promoteurs activés par le ppGpp. L'activation de *lacI* engendre une inhibition de la transcription de la T7 RNAP et du gène d'intérêt par l'intermédiaire du site d'inhibition *lacO*.

En absence d'IPTG, l'expression du système T7 est totalement inhibée par la protéine LacI. Le gène *lacI* est exprimé à la fois sur le génome et sur le plasmide de façon constitutive. La protéine LacI se fixe sur le site d'inhibition *lacO* situé après les promoteurs du gène T7 RNAP ou du gène d'intérêt. L'IPTG, ajouté dans le milieu, va se lier à la protéine LacI induisant un changement de conformation de cette protéine qui va alors libérer les sites *lacO* et permettre la transcription du gène T7RNAP et finalement du gène d'intérêt (figure 7.10).

Le promoteur du gène *lacI*, présent sur le plasmide, a été remplacé par un promoteur activé par le ppGpp (figure 7.10). Ainsi, une accumulation de ppGpp entrainera une augmentation de la transcription du gène *lacI* et de fait, une inhibition de l'expression du système T7. De plus, en cas d'augmentation des PCN, le nombre de copies de *lacI* augmentera, préservant l'efficacité d'inhibition.

7.3.1. Choix des promoteurs d'intérêt

La sélection des promoteurs activés par l'accumulation de ppGpp a été réalisée suivant les critères ci-dessous :

- Promoteurs référencés comme étant activés par le ppGpp dans la base de données BioCyc.
- Activation des promoteurs par le ppGpp en présence du sigma facteur RpoD (cofacteur transcriptionnel impliqué dans la régulation de la transcription des gènes en phase exponentielle de croissance).
- Régulation exclusive par le ppGpp (en présence de DksA et RpoD) connue dans la littérature.
- Données expérimentales solides démontrant l'activation du promoteur par le ppGpp.
- Grande amplitude d'activation du promoteur entre niveau basal de ppGpp et haut niveau de ppGpp intracellulaire.

Les promoteurs d'enzymes impliquées dans la synthèse des acides aminés chez *E. coli* sont activés lors de l'augmentation de la concentration en ppGpp dans la cellule (figure 7.11).

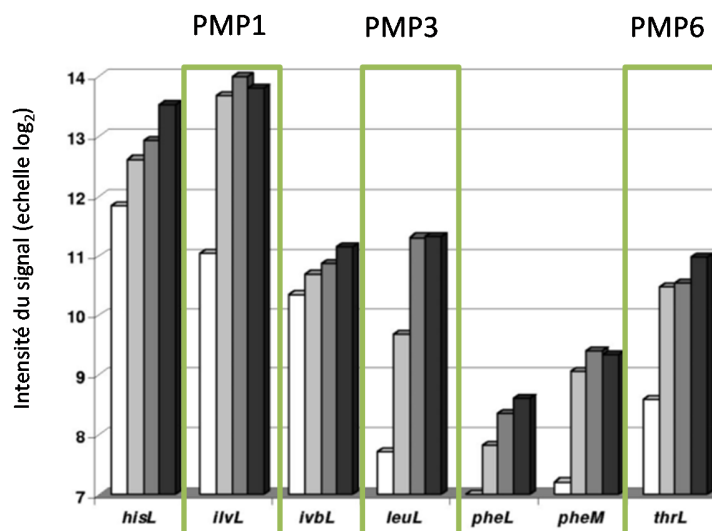


Figure 7.11 : Force d'expression des promoteurs d'enzymes impliquées dans la synthèse des acides aminés chez *E. coli* après ajout de SHX à t=0min (blanc), t=5min (gris clair), t=10min (gris foncé),

t=30min (noir). La force des promoteurs est quantifiée par la technique d'extension d'amorces marquées par la biotin-N6-ddATP. Résultats issus de la publication de Durfee et al. 2008.

Les résultats de la publication de Durfee et al. 2008 ont permis d'observer une augmentation significative de l'expression des promoteurs des gènes *ilvL*, *leuL* et *thrL* après ajout de SHX lors de la culture d'*E.coli* MG1655 (figure 7.11). De plus ces promoteurs ont différents niveaux d'expression basale et permettront d'optimiser les chances de sélectionner un promoteur avec un niveau d'expression basale similaire à celui du promoteur du gène *lacI*.

Sur les 60 promoteurs référencés dans EcoCyc comme étant régulés par le ppGpp, 6 promoteurs (appelés PMP pour plasmides MetaProd) remplissent les critères précédemment cités :

- Promoteur du gène *ilvL* (Plasmide MetaProd 1 (PMP1)) : impliqué dans la synthèse de l'isoleucine et de la valine (figure 7.11)(Durfee, A. M. Hansen, et al. 2008).
- Promoteur du gène *iraP* (PMP2) : impliqué dans la régulation de RpoS (Bougdour & Gottesman 2007).
- Promoteur du gène *leuL* (PMP3) : impliqué dans la synthèse de l'isoleucine et de la valine (figure 7.11)(Durfee, A. M. Hansen, et al. 2008).
- Promoteur P3 du gène *csrA* (PMP4) : impliqué dans la régulation globale du métabolisme (glycolyse, gluconéogenèse, glycogène...) (Yakhnin et al. 2012; Edwards et al. 2011).
- Promoteur des gènes *talA-ktkB* (PMP5) : impliqué dans la régulation de la voie des pentoses phosphates (Rajendran et al. 2008).
- Promoteur du gène *thrL* (PMP6) : impliqué dans la synthèse de la thréonine (figure 7.11)(Durfee, A. M. Hansen, et al. 2008).

Ces promoteurs sont caractérisés par leur force d'expression à un niveau basal de ppGpp et leur capacité à être activé par le ppGpp. La force de répression sans IPTG du système T7 modifié dépend de la force d'expression basale du gène *lacI* du plasmide et de la force d'expression du gène *lacI* du génome exprimé constitutivement. Le gène *lacI* du génome doit assurer la répression du système T7 si la force des promoteurs des « plasmides MetaProd » est trop faible à un niveau basal de ppGpp. En revanche, si la force des promoteurs est supérieure au promoteur original du gène *lacI* alors la force de répression du système T7 sera supérieure et une concentration d'IPTG plus élevée devra être ajoutée

pour lever cette inhibition. La capacité d'activation du promoteur PMP par le ppGpp permettra d'inhiber l'expression du gène de la T7RNAP et le gène d'intérêt en augmentant la concentration intracellulaire de protéine LacI par rapport à la concentration intracellulaire d'IPTG.

La complexité d'évaluation de ces promoteurs apparaît lorsque ces deux aspects de régulation s'additionnent aux caractéristiques du promoteur ribosomal de *rrnE* ou *rrnB* possédant chacun un niveau basal sans doute différent du promoteur *lacUV5* et une capacité à être inhibée par le ppGpp différent. C'est pourquoi, dans un premier temps, la suite de cette étude a pour objectif d'évaluer l'impact de la modification du promoteur de *lacI* par un promoteur PMP sur la souche WT. Une fois les caractéristiques des promoteurs décrites, les plasmides PMP sont évalués dans les souches *rrnE* et *rrnB*.

7.3.2. Evaluation des performances de la souche WT transformée avec les plasmides MetaProd au cours de la production de la protéine UL130

La stabilité du taux de croissance de la souche après induction et la quantité totale de la protéine recombinante produite sont utilisées comme critères d'évaluation de la souche WT transformée par les plasmides PMP 1 à 6. Les performances attendues seront obtenues si : i) le niveau basal d'expression du vecteur avec IPTG est similaire au système T7 et, ii) la capacité d'inhibition de l'expression du système T7 par le ppGpp est suffisant pour stabiliser l'équilibre entre production et croissance.

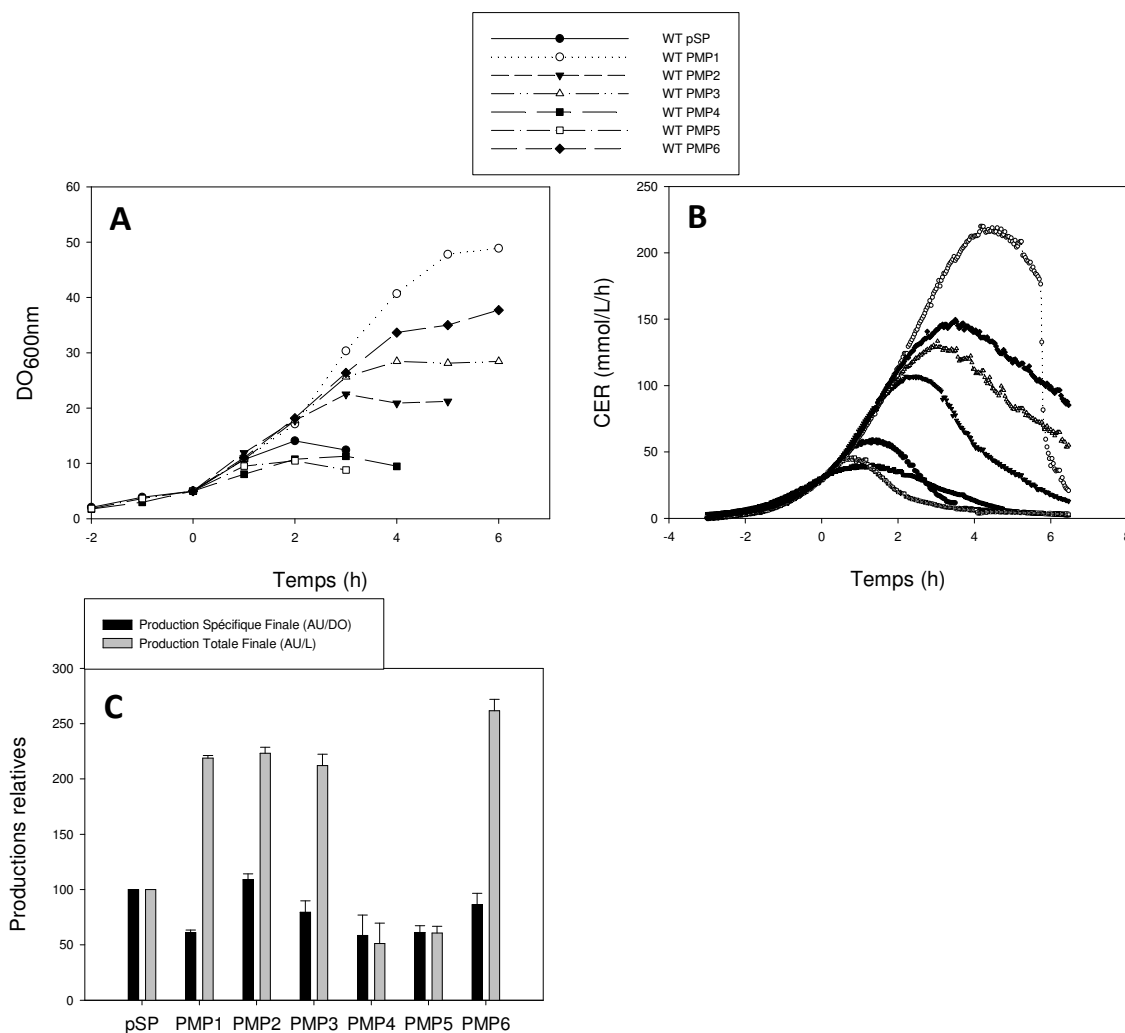


Figure 7.12 : Comparaison de la souche WT transformée par les plasmides modifiés (WT -PMP1 à 6) ou non modifié (WT - pSP) au cours du procédé de production de la protéine UL130 (IPTG 0.1mM). (A) Biomasse (DO) et (B) CER en fonction du temps, (C) productions spécifiques et totales (AU/L) finales en protéine UL130 (duplicat technique).

Les résultats obtenus indiquent que la croissance des souches WT transformées par les plasmides PMP1, PMP2, PMP3 et PMP6 sont significativement maintenues pendant 2 à 4 heures de plus que la souche WT transformée avec le plasmide pSP (plasmide non modifié codant pour la protéine UL130) (figure 7.12, A). Pour ces souches, la quantité plus élevée de biomasse permet de générer deux fois plus de protéines UL130 au moment de l'arrêt de croissance (figure 7.12, C). Seuls les souches WT transformées avec les plasmides PMP4 et PMP5 ne permettent pas d'augmenter la biomasse générée ni la quantité de protéines produite. Une production spécifique plus faible est

retrouvée avec tous les plasmides PMP par rapport à celle de la souche avec le plasmide pSP (à l'exception de la condition avec PMP2) (figure 7.12, C). La diminution rapide du CER, caractéristique de la consommation totale du glucose dans le milieu n'étant pas observé, la croissance des bactéries s'arrête avant la consommation totale du glucose (figure 7.12, B).

L'évaluation du procédé de la souche WT avec le plasmide pSP non modifiée produisant la protéine UL130 a permis de déterminer que 60% de la production spécifique finale est atteinte avant l'arrêt de croissance de la souche et 40% après arrêt de la croissance (figure 4.6, C). Le fait que les souches PMP1, PMP3, PMP4, PMP5 et PMP6 possèdent une production spécifique finale de 60 % par rapport à celle de la souche WT tend à montrer que la production a été réprimée à partir de l'arrêt de croissance (figure 7.12, C).

Il est intéressant de constater que les plasmides ayant obtenus une meilleure stabilité de croissance et de production (PMP1, PMP3 et PMP6) sont des constructions dont le promoteur du gène *lacI* a été remplacé par des promoteurs des voies de biosynthèse des acides aminés (figure 7.11 et 7.12).

Les souches ont arrêté de croître avant la consommation totale du glucose du milieu, déterminé à partir de l'analyse du CER. En revanche, la modification du plasmide a permis un gain de stabilité du procédé se traduisant par un gain significatif de production. Ces modifications ont tout de même permis de générer des gains plus nets de production et de croissance que la stratégie précédente, sans ajout d'acides aminés dans le milieu.

7.3.3. Evaluation des performances des souches MetaProd *rrnE* et *rrnB* transformées par les plasmides MetaProd

De façon indépendante, la modification du génome ou du plasmide ont permis une amélioration de la stabilité de croissance et donc de la production finale de protéines. Afin d'évaluer la mise en synergie des deux stratégies de modifications, les souches MetaProd *rrnE* et *rrnB* ont été transformés par les plasmides pSP, PMP1, PMP2, PMP3, PMP4 ou PMP6. La performance de ces souches est évaluée suivant les mêmes critères que précédemment, c'est-à-dire la stabilité du taux de

croissance de la souche après induction et la quantité total en protéine recombinante produite à la fin du procédé (figure 7.13). Le système d'expression n'était pas réprimé en absence d'IPTG après transformation des souches MetaProd par le plasmide PMP5, c'est pourquoi ces conditions n'ont pas été évaluées.

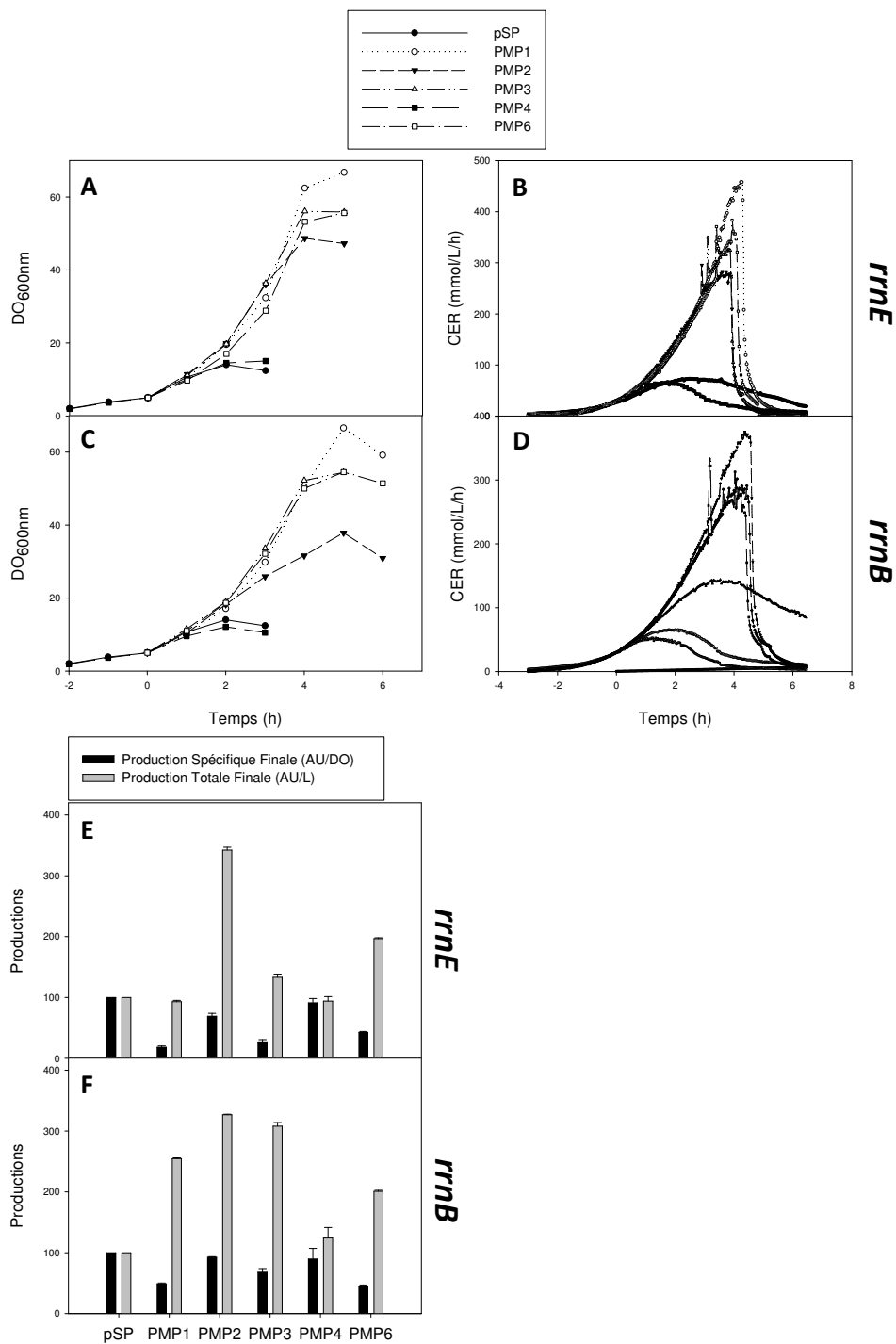


Figure 7.13 : Comparaison des souches MetaProd *rrnE* ou *rrnB* transformées par les plasmides modifiés (PMP1, PMP2, PMP3, PMP4, PMP6) ou non modifié (pSP) au cours du procédé de production de la protéine UL130 (IPTG 0.1mM). (A,C) Biomasse (DO) et (B,D) CER en fonction du temps, (E,F) production spécifique (AU/DO) et totale (AU/L) finales en protéines UL130 (duplicat technique).

La mise en synergie des deux stratégies de rétrocontrôle par le ppGpp a permis d'améliorer considérablement le maintien de la croissance durant la phase de production des souches *rrnE* et *rrnB* PMP1, PMP2, PMP3 et PMP6 (figure 7.13, A-B-E-F). Ces souches maintiennent leur capacité de croissance jusqu'à la consommation totale du glucose caractérisé par la diminution rapide du CER (à l'exception de la condition *rrnB* PMP2) (figure 7.13, B-D). Cette stabilité de la croissance conduit à un gain significatif de la production des souches *rrnE* PMP2, *rrnB* PMP1, 2 et 3 (3 fois plus de protéines UL130 produites), la souche *rrnE* PMP6 et *rrnB* PMP6 (2 fois plus de protéines UL130 produites) par rapport à la production des souches transformées avec pSP (figure 7.13, B-D). La faible production spécifique des souches *rrnE* PMP1 et PMP3 conduit à plus faible production totale de la protéine par rapport à la condition pSP malgré le fait qu'une biomasse 3 fois plus élevée ai été obtenue en fin de procédé.

Durant ces expériences, les souches *rrnE* PMP1, 2, 3 et 6 et *rrnB* PMP 1, 3 et 6 maintiennent une capacité de croissance et une production stable jusqu'à la fin du procédé. La production de ces souches s'est arrêtée lorsque le glucose a été entièrement consommé (déterminé à partir de l'analyse du CER) (décrit figure 5.1) (données non montrées). De fait, des productions totales supérieures auraient été obtenues avec ces souches si une quantité plus importante de substrat carboné avait été utilisée ou si une alimentation en substrat avait été mise en place à la fin du procédé.

Les constructions réalisées avec les promoteurs des gènes *ilvL*, *leuL* et *thrL* (PMP1, PMP3 et PMP6) transformées dans les souches WT, *rrnE* et *rrnB* ont permis d'obtenir les meilleures performances relatives aux critères de croissance et de production définis. Ce résultat signifie que ces promoteurs sont soumis à une activation efficace et contrôlée de leur expression lors de l'accumulation de ppGpp. Les résultats obtenus avec ces promoteurs démontrent donc que l'accumulation de ppGpp conduit à une activation des voies de synthèse des acides aminés de la valine, de la leucine et de la thréonine et probablement d'autres. Cette activation pourrait également permettre une adaptation du métabolisme à la production.

Le PCN augmente lors de l'augmentation du nombre d'ARNt non chargés dans la bactérie (Camps 2010)(figure 6.6). L'évolution du PCN est donc un indicateur de l'apparition d'une limitation en

acides aminés au cours de la production. Une détermination des PCN a été réalisée sur les cultures de les souches WT, *rrnE* et *rrnB* transformées avec les plasmides MetaProd 1, 2 et 6 (PMP) (figure 7.14).

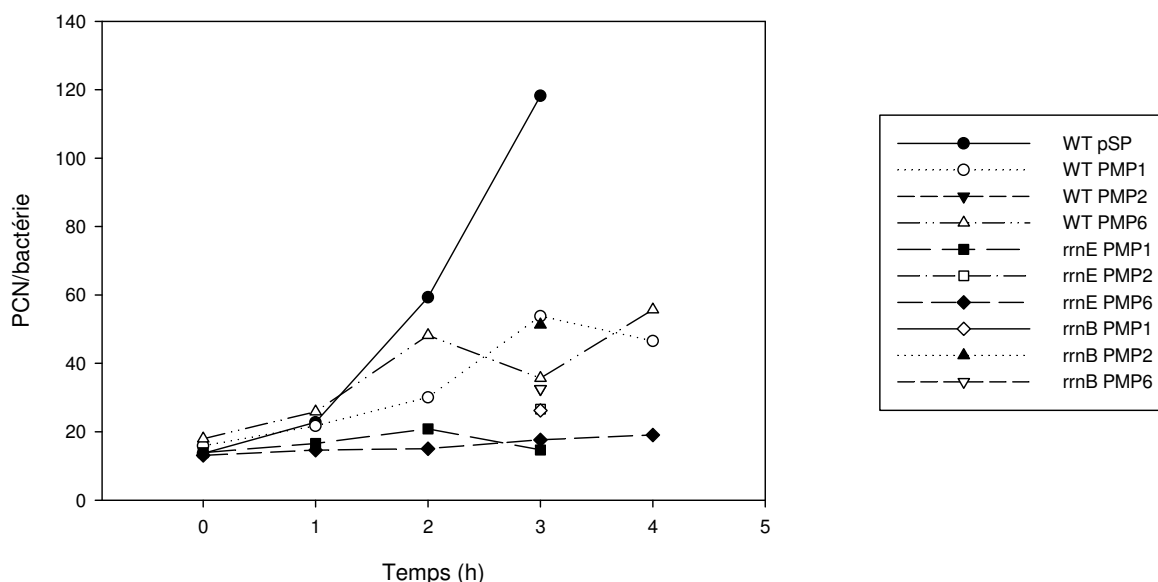


Figure 7.14 : Détermination du nombre de copie de plasmides (PCN) par bactérie au cours des procédés de production de la protéine UL130 avec la souche WT ou MetaProd *rrnE* et *rrnB* transformées par les plasmides MetaProd (PMP1, PMP2 et PMP6) ou le plasmide non modifié (pSP).

Ces résultats indiquent que l'utilisation d'un plasmide MetaProd stabilise complètement ou limite fortement l'augmentation des PCN (maximum à 50 PCN par bactérie pour toutes les combinaisons MetaProd testées au lieu de 120 pour la souche WT avec le plasmide non modifié pSP) (figure 7.14). Le couplage des plasmides MetaProd avec la souche MetaProd *rrnE* permet une stabilisation totale des PCN à environ 20 PCN/bactérie (figure 7.14).

Ce résultat est une preuve supplémentaire de l'efficacité du rétrocontrôle par une adaptation de la production en fonction de la charge métabolique car cela indique que la limitation en acides aminés n'est probablement pas apparue au cours de ces cultures. La stabilisation du procédé permet de synthétiser deux fois plus de quantités de protéines sans augmentation de la force d'expression du vecteur par augmentation de PCN et laisse envisager la prolongation du procédé par alimentation

permettant une augmentation exponentielle de la biomasse et donc une augmentation exponentielle de la production.

7.4. Discussion

L'objectif des travaux présentés dans ce chapitre est d'évaluer des solutions permettant d'obtenir un système de production robuste et performant pour la production de protéines recombinantes.

La stratégie d'ajout d'acides aminés au moment de l'induction permet d'augmenter significativement le taux de croissance et la vitesse de production après induction. Cet ajout est une stratégie de production développée à partir du fait que la production totale est similaire en milieu riche et en milieu minimum avec la même protéine recombinante au cours du procédé de référence. Pourtant, les sources d'approvisionnement en acides aminés pour la synthèse protéique sont différentes dans ces deux milieux. En milieu riche, les acides aminés sont assimilés à partir du milieu par la cellule et leurs voies de biosynthèse sont réprimées (Traxler, Summers, H.-T. Nguyen, et al. 2008). En milieu minimum, l'ensemble des acides aminés sont synthétisés à partir du glucose. Le principe de la stratégie mise en place pour augmenter la performance du procédé repose sur la supplémentation en acides aminés au moment de l'induction d'une culture d'*E. coli* réalisée sur milieu minimum. Parce qu'une diminution des concentrations intracellulaires en acides aminés a été observé en milieu minimum au cours de la production, il est attendu que l'ajout d'acides aminés dans le milieu conduise à une assimilation des acides aminés en plus de leur biosynthèse. Via cette approche les performances de production sont meilleures que celles obtenues en milieu riche et en milieu minimum. En effet l'ajout des acides aminés au moment de l'induction a permis d'augmenter la quantité de protéines générées de 40 à 60 % par rapport à celles des conditions sans supplémentation en acides aminés.

Un système biologique performant pour la production de protéines recombinantes signifie une vitesse de production maîtrisée avec une force du système d'expression constamment en adéquation avec la capacité de synthèse de l'hôte. Un tel système permet d'obtenir une croissance et une production stable au cours du temps. Cette stabilité aboutit également à la possibilité de développer

des procédés de type semi-continu ou continu en condition de production. Comme présenté dans l'introduction, la mise au point d'un système d'expression équilibré passe normalement par le choix du système d'expression (séquence, origine de réplication, promoteurs, étiquettes), par le choix de la souche, du milieu de culture et du procédé. Les conséquences de la production de chaque protéine recombinante étant très variables ce travail d'optimisation peut être long et sans garanti de réussite. Le principe du système MetaProd est d'intégrer la régulation du système d'expression T7 à la régulation transcriptionnelle, traductionnelle et métabolique de la bactérie afin générer un système d'expression générique au développement des productions des protéines recombinantes. Le système T7 est un système d'expression permettant une transcription très forte du gène d'intérêt. C'est également un système d'expression possédant un verrou d'inhibition puissant grâce à la répression transcriptionnelle exercée par la protéine LacI sur le site *lacO*. Ce système est donc tout à fait adapté pour servir de base au développement d'une plateforme d'expression robuste et performante à la production d'une grande diversité de protéines recombinantes. En effet, son système de régulation de l'expression du gène d'intérêt doit permettre de contrôler la vitesse de production de protéine recombinante sur un large spectre de forces d'expressions. Les études réalisées dans les chapitres précédents ont mis en évidence que le ppGpp via la réponse stringente est un excellent candidat pour prendre en charge cette régulation. Le ppGpp est une molécule synthétisée lors de la détection de carences en carbone, en acides aminés, en azotes, en phosphate, en fer par *E. coli* ou lorsque sa membrane est fragilisée. Il a été montré que la production des protéines recombinantes UL130 et Hla conduit à une forte consommation en acides aminés entraînant des carences intracellulaires déclenchant la synthèse de ppGpp. De façon plus large, la production de protéine recombinante peut engendrer des limitations en azotes, en carbone ou des déformations de membrane. Cette détection d'un large spectre de limitation est propice au développement de production performante de protéines recombinantes diverses et donc une application générique au système d'expression. Outre la modification de la concentration en ppGpp résultant de ces carences, le ppGpp possède la capacité d'interagir directement avec l'ARN polymérase d'*E. coli*. Cette modification de la reconnaissance des promoteurs cibles par l'ARNP est rarement déclenchée par une molécule mais le plus souvent par un facteur de transcription. L'ensemble des biosenseurs rapportés dans la littérature jusqu'à maintenant repose sur l'activation de la transcription d'un cofacteur chargé ensuite d'interagir avec la RNAP pour

activer les cibles d'intérêt. L'interaction directe du ppGpp avec la RNAP évite de passer par une étape de transcription et de traduction du cofacteur transcriptionnel entre le moment où la limitation est détectée et le moment où la réponse est apportée. La synthèse rapide de ppGpp lors de la détection d'une limitation offre une rapidité de régulation qui sera extrêmement utile pour la régulation de la vitesse de transcription du gène d'intérêt en temps réel. De plus, le ppGpp peut être rapidement dégradé suite à son action et donc sa concentration intracellulaire peut évoluer significativement sur des temps courts. La synthèse et la dégradation des cofacteurs de transcription est par contre probablement plus longue.

Le remplacement du promoteur *lacUV5* de la T7 RNAP par le promoteur de l'opéron ribosomal *rrnE* (identifié comme étant le plus prometteur des promoteurs des opérons ribosomaux) n'a pas suffi pour inhiber le système efficacement lors de l'accumulation de ppGpp. L'augmentation du nombre de copies de plasmides (PCN) par bactérie engendre une force d'expression du système trop importante. Un essai a été effectué pour apporter des mutations à la boucle 2 de l'ARN II de l'origine de réplication ColE1. Cette boucle serait responsable de l'interaction avec les ARNt non chargés et de l'augmentation du nombre de PCN par bactérie. Aucune colonie n'a été obtenue avec ces plasmides dont l'origine de réplication a été mutée et cette piste a été abandonnée (résultats non présentés). Le changement d'origine de réplication a également été une piste envisagée. Cependant ce changement devait être effectué par une origine de réplication régulée par une protéine et non par des ARN pour éviter la perturbation de sa régulation par des ARNt non chargés. Les origines de réplication régulées par des protéines sont plus stables. Cependant ces origines de réplication maintiennent moins de 10 copies de plasmides par bactérie. Une baisse trop importante du nombre de copies de plasmides par bactérie pourrait engendrer une force d'expression plus faible du système T7. De plus la séquence de la protéine chargée de réguler l'origine de réplication ajoute plus de 1kb au vecteur d'expression. Pour ces raisons, l'ajout d'un deuxième levier d'inhibition du système T7 par le ppGpp a été privilégié. Ce deuxième levier consiste à utiliser le système d'inhibition déjà présent dans le système T7. Ce système d'inhibition est basé sur l'activation de l'expression du gène *lacI* lors de l'accumulation de ppGpp. Ce gène est présent sur le génome et sur le plasmide. La modification du promoteur du gène présent sur le plasmide permet ainsi de compenser l'augmentation du nombre de plasmide par l'augmentation de l'expression du gène *lacI* présent à un plus grand nombre de copies. Dans la souche WT, le

remplacement du promoteur *lacI* par un promoteur activé par le ppGpp ne suffit pas à contrôler efficacement l'expression du système T7 lors de l'augmentation de la concentration en ppGpp pour prolonger la croissance. Cependant, les bénéfices de croissance et de production observés sont significatifs avec la plupart des promoteurs sélectionnés. La mise en synergie des deux systèmes d'inhibition a permis d'atteindre l'objectif fixé, c'est-à-dire d'obtenir une croissance et une production stable et importante tout au long du procédé batch avec une induction précoce.

La modification du promoteur du gène *lacI* du plasmide a permis de maintenir la phase de croissance et de production plus longtemps avec la souche WT que la modification du promoteur *lacUV5* de la T7 RNAP. Il est possible qu'un facteur autre que l'augmentation du PCN conduise à une faible efficacité de la modification du promoteur *lacUV5*. Le promoteur *lacUV5* a été remplacé par un promoteur ribosomal pour inhiber la synthèse de T7 RNAP lors de l'accumulation de ppGpp. Cependant, lors de cette accumulation, la T7 RNAP est déjà produite et présente dans la cellule. L'arrêt de sa synthèse n'aura d'effet que lorsque la T7 RNAP sera dégradée ou diluée par division cellulaire. Or, l'accumulation de ppGpp conduit à l'inhibition de croissance ce qui empêche la dilution de la T7 RNAP par division cellulaire. En revanche, l'augmentation de la synthèse de protéines LacI engendrée par l'accumulation de ppGpp, produit le phénomène inverse. La quantité de protéine LacI par bactérie est susceptible d'augmenter rapidement lors du ralentissement de la division cellulaire. L'augmentation du rapport entre la concentration intracellulaire de protéines LacI et la concentration intracellulaire d'IPTG rend immédiat l'action d'inhibition de la transcription de la T7 RNAP et du gène d'intérêt par fixation de la protéine LacI sur les sites de répression *lacO*. Ces deux approches sont pourtant complémentaires puisque seule l'association des deux modifications a permis de stabiliser la croissance et la production durant tout le procédé. Il est probable que la modification du promoteur du gène *lacI* permette une meilleure réactivité du système et que le couplage des deux modifications, une meilleure efficacité.

Conclusions et perspectives

8.1. Conclusions

8.1.1. Importance de l'équilibre entre croissance et production de protéines recombinantes

La surproduction de protéines recombinantes chez *E. coli* demande une allocation de la machinerie cellulaire et des ressources bactériennes pour la synthèse de la protéine d'intérêt (Cserjan-Puschmann et al. 1999). Cette synthèse conduit à l'apparition d'une charge métabolique qui provoque nécessairement une diminution du taux de croissance de la bactérie pouvant aller jusqu'à l'arrêt de la croissance et de la production (Scott et al. 2010). Pour améliorer la productivité et la qualité de la protéine produite (intégrité, activité, solubilité...), le système d'expression, la souche, le milieu de culture et le procédé sont habituellement optimisés empiriquement pour tendre vers un équilibre entre la vitesse de production de la protéine et la capacité de synthèse de l'hôte (Marschall et al. 2016). Cette démarche visant à obtenir une croissance et une production équilibrée est essentielle au développement d'un procédé de production performant (productivité, production spécifique élevée...) telles que des cultures semi-continues et continues. L'étude du lien entre le taux de croissance et la production est au centre de ce travail avec comme objectif final de générer un système biologique de production robuste et performant pour la production de protéines recombinantes chez *E. coli*.

8.1.2. Influence de la production de protéines recombinantes sur la croissance bactérienne.

Le taux de croissance est contraint par la capacité de la bactérie à transformer les nutriments du milieu de culture en protéines de la biomasse (Bremer & Dennis 1996; Scott et al. 2010; Ding Jun et al. 2012; Scott et al. 2014; Bosdriesz et al. 2015). Les composés présents dans un milieu riche ne nécessitent qu'une faible transformation pour être utilisés au cours de la synthèse des protéines de la biomasse. De fait, la capacité des bactéries à transformer les nutriments en protéines de la biomasse

est donc élevée dans un milieu riche et la quantité de protéines à synthétiser à chaque génération est plus faible qu'en milieu minimum, contraignant moins la croissance bactérienne. A l'inverse, en milieu minimum, la capacité des bactéries à transformer les nutriments en protéines de la biomasse est plus faible et la quantité de protéines à synthétiser à chaque génération est plus élevée, augmentant le temps de doublement des bactéries. Ce mécanisme, appelé fitness du μ , est régulé par la réponse stringente. Ce mécanisme permet de mettre en adéquation le fonctionnement métabolique et la vitesse de traduction des protéines à la disponibilité en acides aminés intracellulaires (figure 8.1) (Scott et al. 2014). La régulation est mise en place à un niveau transcriptionnel et traductionnel entraînant une modification de la proportion en protéines impliquées dans le métabolisme et en ribosomes parmi la quantité totale de protéines de la biomasse (figure 8.1) (Magnusson et al. 2005; Sherlock et al. 2018).

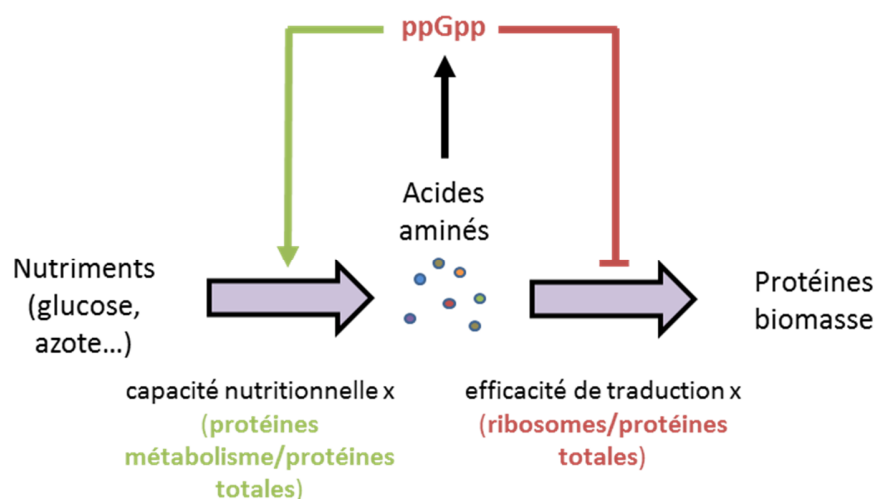


Figure 8.1 : Régulation du flux de synthèse et de consommation des acides aminés (violet) par le ppGpp à partir de la détection de la disponibilité en acides aminés intracellulaires. Inspiré de Scott et al. 2010.

Dans ce contexte, l'ajout de la synthèse d'une protéine recombinante, sans fonction pour la bactérie, entraîne une consommation supplémentaire de ressources cellulaires et une utilisation additionnelle de la machinerie traductionnelle. De fait, la production de protéines recombinantes est une contrainte pour la croissance bactérienne (Scott et al. 2014).

8.1.3. Analyse des contraintes métaboliques liées à la production de protéines recombinantes chez *E. coli*.

L'approche originale de cette thèse est d'analyser les contraintes métaboliques liées à la production de deux protéines recombinantes modèles chez *E. coli*. Il s'agit en réalité d'étudier au niveau métabolique l'influence de la production de protéines recombinantes sur le fitness du μ . Pour cela, les cinétiques de croissance et de production de protéines recombinantes ont été caractérisées dans le procédé de référence en milieu minimum, développé au cours de ce travail de thèse. Les analyses ont permis de confirmer les corrélations décrites dans la littérature avec une diminution du taux de croissance lors de l'augmentation de la vitesse de production spécifique en condition de production des deux protéines modèles (Scott et al. 2014). Des analyses du métabolisme ont ensuite permis d'étudier la perturbation et les contraintes métaboliques en lien avec les cinétiques de croissance et de production. Les perturbations du métabolisme observées au cours de ces analyses coïncident avec les phases caractéristiques du procédé définies avec le taux de croissance et la vitesse de production spécifique. En particulier, l'excrétion de métabolites dans le surnageant et l'évolution des concentrations intracellulaires en acides aminés sont corrélées à des phases de fortes productions ou d'arrêt de croissance de la souche. Ce changement de la disponibilité des acides aminés a appuyé l'approche mis en place dans cette thèse consistant à étudier le lien entre la production de protéines recombinantes et le contrôle de la croissance bactérienne. Le suivi de la cinétique de concentration intracellulaire du ppGpp au cours du procédé de production a permis de mettre en évidence son accumulation et donc de supporter la fait que la disponibilité des acides aminés est le facteur majeur limitant la production des protéines recombinantes et impactant fortement la croissance. La concentration intracellulaire de ppGpp varie en lien avec le taux de croissance et la vitesse de production des deux protéines recombinantes étudiées. L'étude de l'influence de la réponse stringente sur les cinétiques de croissance et de production a permis de révéler que l'accumulation intracellulaire de ppGpp interfère dans la compétition entre la croissance et la production pour la consommation des acides aminés en faveur de la production et en défaveur de la croissance. De cette observation un nouveau système d'expression a été développé permettant à la bactérie de contrôler le système

d'expression de la protéine d'intérêt en fonction de la disponibilité des ressources intracellulaires évitant l'arrêt de la croissance au cours de la production.

8.1.4. Comparaison des contraintes métaboliques de la production entre une culture en milieu riche et une culture en milieu minimum

La mise en place du procédé de référence de production a débuté par la formulation du milieu minimum chimiquement défini. Ce milieu a permis de mettre en place les analyses fonctionnelles du métabolisme des souches en condition de surproduction de protéines recombinantes. Ce milieu a conditionné les types de contraintes détectées dans la suite de la thèse. En effet, en milieu riche, les ressources sont présentes en grande quantité dans l'environnement et la bactérie possède une plus grande proportion de ribosomes que de protéines fonctionnelles du métabolisme parmi la quantité totale de protéines (Bremer & Dennis 2008). La production d'une protéine recombinante dans la bactérie risque donc de ralentir la croissance de par l'utilisation d'une partie des ribosomes et dans une plus faible mesure par une limitation nutritionnelle. En milieu minimum, la part des protéines impliquées dans le métabolisme de la bactérie est plus importante que la part de ribosomes parmi la quantité totale de protéines (Bremer & Dennis 2008). De fait, la synthèse de protéines recombinantes engendre une contrainte à la fois par consommation des ressources intracellulaires issues du métabolisme du glucose et par utilisation des ribosomes pour la traduction. L'utilisation d'un système de rétrocontrôle de la production par rapport aux limitations nutritionnelles engendrées par la production est donc plus pertinente dans un milieu minimum que dans un milieu riche.

8.1.5. Etude fonctionnelle du métabolisme en condition de production de protéines recombinantes en milieu minimum

L'analyse fonctionnelle du métabolisme des souches en condition de production en milieu minimum a révélé l'apparition d'une limitation en acides aminés au cours de la production des protéines UL130 et Hla en milieu minimum. Cette limitation a été démontrée grâce à la quantification du ppGpp de la souche *E. coli* BL21 WT et $\Delta relA$ en condition de production. Plus généralement, cette étude a permis de mettre en évidence un fort bouleversement du métabolisme de la souche *E. coli*

BL21 WT lors de la production de la protéine UL130. Ce bouleversement est caractérisé par la forte accumulation d'acétate et de pyruvate concomitante à un arrêt de croissance en condition de production. De plus, une importante excrétion d'acides aminés dans le surnageant a été observée durant l'arrêt de croissance en condition de production. De façon intéressante, la concentration intracellulaire de seulement une partie des acides aminés augmente lors d'un arrêt de croissance en condition de production de protéines recombinantes, tandis que la concentration intracellulaire des autres acides aminés tend à diminuer. En revanche, la concentration intracellulaire de l'ensemble des acides aminés augmente lors d'un arrêt de croissance sans production de protéine, provoqué par une activation de la réponse stringente (ajout de sérine hydroxamate). Il est donc probable que l'absence d'accumulation et la diminution de la concentration intracellulaire de certains acides aminés durant l'arrêt de croissance soient dues à leur consommation par la synthèse de la protéine recombinante. De fait, il semble que l'inhibition du taux de croissance par l'accumulation de ppGpp oriente le flux de consommation des acides aminés destinés à la synthèse des protéines de la biomasse vers leur consommation pour la synthèse des protéines recombinantes (figure 8.2).

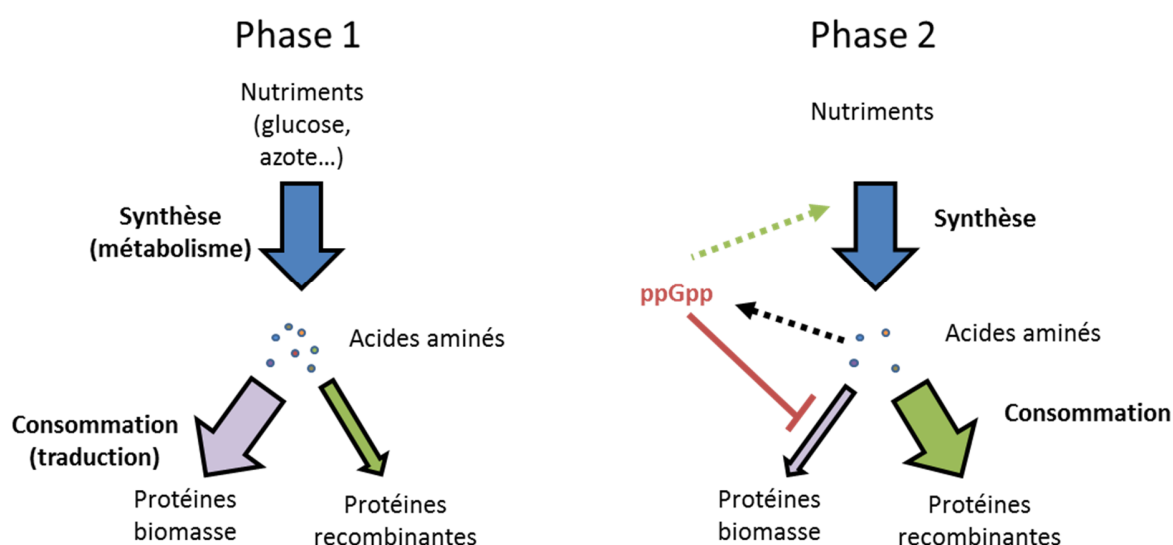


Figure 8.2 : Intervention du ppGpp dans la répartition des flux de consommation des acides aminés entre la croissance (protéines de la biomasse) et la production de protéines recombinantes. La phase 1 correspond à la mise en place de la production après ajout de l'IPTG et la phase 2 correspond à une période d'accumulation de ppGpp lors d'une forte vitesse de production.

La répartition des flux de consommation de chaque acide aminé entre croissance et production a été analysée à partir de la proportion qu'ils représentent dans les protéines de la biomasse (valeur moyenne) et dans la séquence primaire des protéines recombinantes. La différence de composition en acides aminés entre la biomasse et la protéine recombinante conditionne la réorganisation des flux de synthèse des acides aminés devant se mettre en place pour éviter des limitations en acides aminés lors de la production de protéines recombinantes. La réorganisation des flux signifie qu'une inhibition ou une activation de la synthèse de certains acides aminés doit s'opérer au cours de la production. Il est observé que cette réorganisation des flux de synthèse des acides aminés doit être plus importante lors de la production de la protéine UL130 que lors de la production de la protéine H1a pour conserver la croissance de la souche pour une production spécifique similaire. Cela pourrait être dû au fait que la composition en acides aminés de la séquence primaire de H1a est plus proche de la composition en acides aminés des protéines de la biomasse que ne l'est celle de la protéine UL130. Ce mécanisme de réorientation des flux doit être mis en place par la cellule en parallèle de la dynamique de compétition entre la production et la croissance qui s'opère durant le procédé. Sur la base de ce postulat, le procédé de production peut-être décrit en deux phases. Il faut d'abord rappeler que la vitesse de production spécifique dépend de la vitesse de transcription du gène d'intérêt (force d'expression du vecteur) puis de la vitesse de traduction des ARNm (figure 8.3, B).

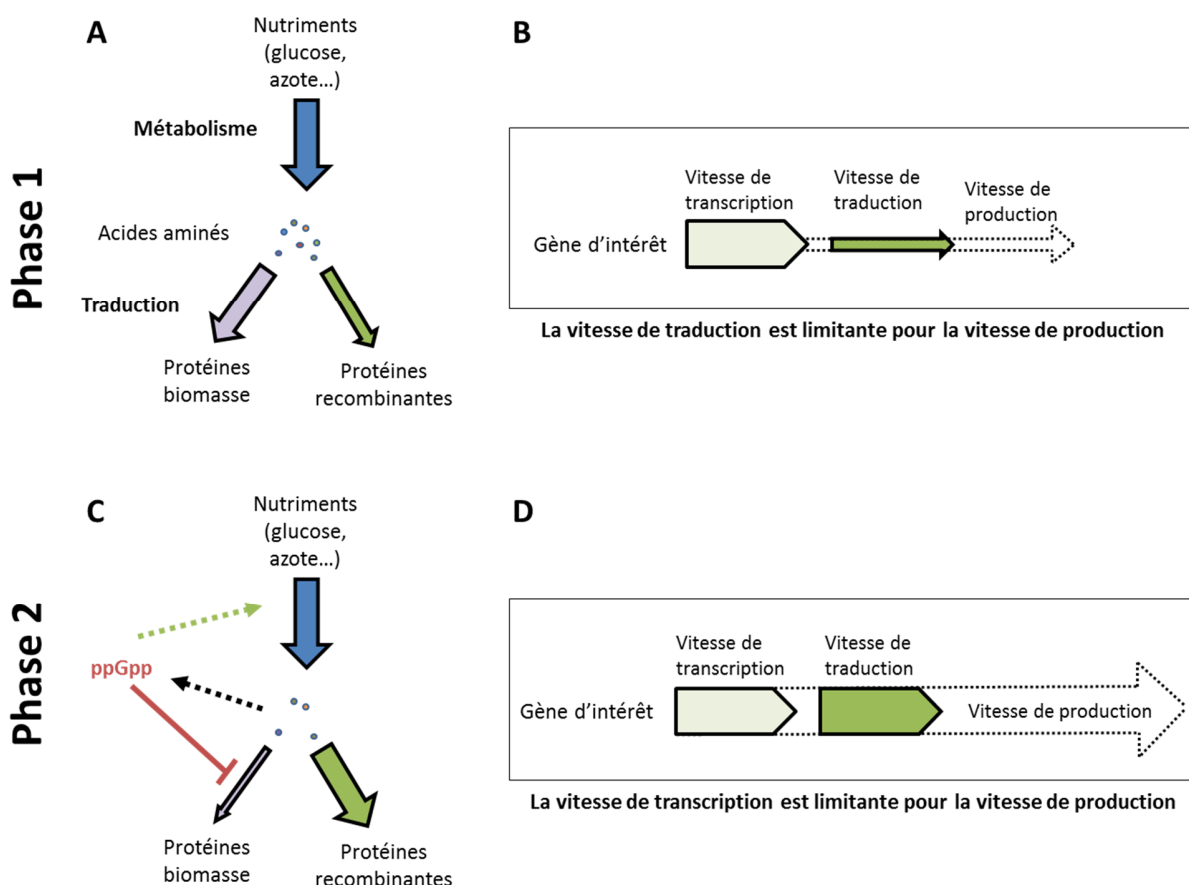


Figure 8.3 : Contraintes limitant la production de la protéine d'intérêt. La phase 1 des procédés de production est limitée par la vitesse de traduction (A, B), et la phase 2 par la vitesse de transcription (C, D).

La première phase est caractérisée par un vecteur d'expression induit très fortement dès le début du procédé, une importante quantité d'ARNm est transcrite et la vitesse de production est limitée par la capacité traductionnelle de la bactérie (figure 8.2 A, B). A ce moment, une compétition au niveau de la machinerie traductionnelle apparaît pour l'utilisation des ribosomes et la consommation des acides aminés (figure 8.3, A). Les flux de consommation des acides aminés évoluent et les flux de synthèse des acides aminés doivent être réorganisés. Si des acides aminés deviennent limitants, le ppGpp s'accumule, inhibant la croissance et orientant le flux de consommation des acides aminés associé à la croissance vers la production de protéines recombinantes (figure 8.3, C). Un équilibre entre croissance et production peut être atteint seulement si la vitesse de transcription du

gène d'intérêt devient limitante à la vitesse de production (figure 8.3, D). Si la vitesse de transcription ne devient pas limitante alors l'inhibition de la croissance continue jusqu'à arrêt de la croissance.

Le fait que la production de la protéine Hla ne conduise pas nécessairement à un arrêt de croissance peut être le résultat de plusieurs facteurs. D'abord la proportion en acides aminés de la protéine est plus proche que celle de la biomasse rendant la réorientation des flux de consommation en acides aminés par le ppGpp plus efficace pour alimenter la production car plus adaptés à la demande en acides aminés pour la synthèse de la protéine Hla. Il est également possible que la vitesse de transcription du gène *hla* soit plus faible que celle du gène *UL130* (taille de la séquence génique, structures d'ADN internes à la séquence, enrichissement en GC, usage des codons...). Une limitation de la vitesse de production par la vitesse de transcription du gène d'intérêt au moment où le taux de croissance devient trop faible semble indispensable pour obtenir un équilibre entre croissance et production de protéine recombinante.

La limitation en acides aminés conduit également à augmenter le nombre de copies de plasmide par bactérie pendant la production. L'augmentation du nombre de copies de plasmide augmente la force d'expression du gène d'intérêt, empêchant qu'un équilibre entre croissance et production soit atteint. L'augmentation du nombre de copies de plasmide est une conséquence de la vitesse de production trop élevée par rapport aux capacités de synthèse de la souche *E. coli* BL21. Dans le but d'éviter qu'une augmentation du nombre de copies de plasmide apparaisse au cours de la production, l'option consistant à résoudre la limitation en acides aminés a été privilégiée à celle consistant à modifier l'origine de répllication. En effet, d'autres conséquences apparaissent du fait que la vitesse de production soit trop élevée comme l'arrêt de croissance suite à l'induction ou la production de protéines insolubles. Cette vitesse élevée de production est causée par le fait que le système d'expression T7, comme tous les autres systèmes d'expression, soit découplé des régulations métaboliques et traductionnelles de la bactérie.

8.1.6. Système MetaProd

L'approche consiste à intégrer le système d'expression génique de la protéine hétérologue à la régulation métabolique et traductionnelle de la bactérie afin que la bactérie puisse adapter, par

autorégulation, la force d'expression du gène d'intérêt à sa capacité maximale de synthèses de la protéine recombinante en s'appuyant sur un mécanisme de détection de la disponibilité en ressources intracellulaires. Pour cela, un nouveau circuit de régulation transcriptionnelle a été implémenté dans la bactérie. Ce circuit de régulation met en jeu le ppGpp comme élément central. Ce dernier joue le rôle de senseur de la disponibilité des ressources pour un contrôle adéquat de la production de la protéine d'intérêt. Ce senseur est en particulier capable de ralentir la production de la protéine lorsque les ressources deviennent limitantes (concentration en ppGpp élevée). *A contrario*, si les ressources sont disponibles (concentration en ppGpp faible), alors la production de la protéine n'est plus ralentie. Ce nouveau système d'expression, appelé système « MetaProd », a démontré son efficacité et ses bénéfices avec les deux protéines modèles choisies pour la thèse (UL130 et H1a), avec une augmentation allant jusqu'au triplement de la production totale. La version la plus performante de ce système a été obtenues grâce à l'utilisation de promoteurs de gènes impliqués dans le fitness du μ (promoteurs d'opérons ribosomiaux et des voies de synthèse d'acides aminés). La quantification du ppGpp intracellulaire a permis de prouver que la bactérie s'adapte à une production trop élevée et est capable de maintenir une croissance et une production stables durant plusieurs heures jusqu'à épuisement des sources carbonées contrairement à la souche non modifiée dont la croissance et la production sont rapidement stoppées. L'obtention d'une stabilité de la croissance, de la production et du nombre de copies de plasmide avec la version la plus récente du système MetaProd permet de remplir les objectifs fixés par le travail de thèse. La création de ce système d'expression innovant MetaProd est donc extrêmement prometteuse pour la mise en place de procédés optimisés de types semi-continus voire de types continus.

8.1.7. Influence de la réponse stringente sur le métabolisme et la croissance

En parallèle, l'adaptation des souches *E. coli* BL21 et K-12, sauvages ou mutées ($\Delta relA$, $\Delta gppA$) a été étudié lors de l'activation de la réponse stringente déclenchée par l'ajout d'un analogue de la sérine, le SHX. Ces expériences ont permis de montrer que le SHX est soit dégradé soit stocké dans les cellules après son ajout dans le milieu de culture. Ce phénomène est indépendant de l'activation de la réponse stringente et n'a jamais été décrit auparavant. Le (p)ppGpp a été quantifié de façon absolue grâce à une méthode mise au point par le travail de thèse. Grâce à cette méthode, l'influence de la

réponse stringente dans la résilience des souches *E. coli* BL21 et K-12 à la perturbation causée par l'ajout de SHX a été analysée. Cette étude a permis de discerner l'influence du SHX et de l'accumulation du (p)ppGpp sur l'inhibition du taux de croissance (sensibilité des souches) et sur leur capacité à retrouver une croissance maximale (adaptation des souches) après l'ajout du SHX. Cette étude a montré que le (p)ppGpp permet aux bactéries de s'adapter plus rapidement à l'ajout du SHX tout en inhibant fortement leur croissance. Dans ce contexte, la souche *E. coli* BL21 est plus robuste et s'adapte plus efficacement à l'ajout de SHX par rapport à la souche *E. coli* K-12. Des différences majeures d'accumulation de ppGpp et de pppGpp entre les deux souches ont également été observées. Il s'agit maintenant de savoir si les différences d'accumulation de ppGpp et de pppGpp observées entre les deux souches sont à l'origine des différences de sensibilité et d'adaptation des bactéries à l'ajout de SHX.

8.2. Perspectives

8.2.1. Optimisation du système MetaProd

Le système MetaProd a permis de générer une preuve de concept de l'efficacité du système de rétrocontrôle de ppGpp sur la vitesse de production pour maintenir la croissance en condition de forte production. L'étape suivante est de générer un système capable produire un plus large panel de protéines recombinantes. Pour cela, les cinétiques de croissance et de production doivent être finement étudiées avec un plus grand nombre de protéines recombinantes. Le système pourra être évalué en terme de performance et de robustesse afin d'étudier la capacité d'adaptation de la vitesse de production à la diversité de protéines à produire.

Les optimisations auront ensuite comme objectif de concevoir un système MetaProd avec une vitesse de production optimale dans une gamme de variation physiologique de la concentration en ppGpp. En effet, une concentration élevée en ppGpp conduisant à un arrêt total de la croissance devrait engendrer un arrêt total de la production. A l'inverse, une faible concentration en ppGpp devrait permettre une forte expression du gène d'intérêt. Entre ces deux concentrations, l'inhibition de la vitesse de production ne doit pas forcément être linéaire, laissant envisager plusieurs évolutions de

la régulation. Cette optimisation de la souche MetaProd passe par la construction étape par étape du système d'expression avec une évaluation des propriétés des promoteurs activés ou inhibés par le ppGpp avant leur intégration dans le système d'expression. Les promoteurs devront être caractérisés en terme de force d'expression en fonction de la concentration en ppGpp. De nombreux promoteurs possèdent une force d'expression dépendant de la concentration en ppGpp. Dans l'étude présentée dans ce manuscrit, les promoteurs sélectionnés régulent des gènes impliqués dans le fitness du μ . Au vu des résultats obtenus, cette stratégie semble pertinente. Cependant un grand nombre de promoteurs impliqués dans le fitness du μ reste à étudier dans le cadre de cette application. En effet, 55 promoteurs codant des protéines formant les ribosomes sont connus chez *E. coli* (Lemke et al. 2011). De plus, certains promoteurs impliqués dans la transcription de gènes codant pour des protéines des voies de synthèse des acides aminés restent à étudier. En fonction des moyens déployés pour l'évaluation des promoteurs, le nombre de promoteurs à étudier peut être augmenté par des mutations ou des combinaisons de séquences entre promoteurs. L'étude de la force d'expression de ces promoteurs *in vivo* nécessite d'obtenir une souche avec des concentrations intracellulaires de ppGpp différentes et constantes sur des temps suffisamment longs pour analyser le rétrocontrôle du ppGpp sur la force d'expression des promoteurs. Dans ce but, la souche *E. coli* BW25113 WT est plus pertinente à utiliser que la souche *E. coli* BL21 car l'accumulation de ppGpp suite à l'ajout de SHX se maintient à un fort niveau assez longtemps pour pouvoir évaluer la force du promoteur par gène rapporteur ou par qPCR. Les différentes concentrations en ppGpp pourront donc être obtenues avec la souche *E. coli* BW25113 $\Delta relA$ (faible niveau basal de ppGpp), la souche *E. coli* BW25113 WT (niveau basal normal de ppGpp), la souche *E. coli* BW25113 WT avec ajout de SHX (haut niveau de ppGpp pendant 2 heures). Le promoteur étudié devra être suivi de la séquence d'inhibition *lacO* afin de coupler l'influence de la concentration en ppGpp avec l'influence de la concentration en IPTG. Un couplage des promoteurs les plus pertinents pourrait permettre d'obtenir une version 2.0 du système MetaProd. Ces souches seront évaluées en condition de production « batch » et semi-continu avec une grande diversité de protéines recombinantes. Ces études aboutiront sur la sélection d'une souche MetaProd 2.0 optimisée.

8.2.2. Mise en œuvre d'un procédé semi-continu avec le système MetaProd

Les perspectives d'optimisation du procédé autour du système MetaProd sont nombreuses. Une étude des capacités de production et de croissance en procédé semi-continu pourra être réalisée en condition de limitation en substrat carboné. Lors d'un procédé semi-continu de production utilisant une souche MetaProd, la limitation du taux de croissance de la souche par une limitation en substrat devrait conduire à une limitation de la vitesse de production. Ainsi, la gestion du taux de croissance permettrait une gestion directe de la vitesse de production par l'intermédiaire de la réponse stringente. Cette relation ne sera peut-être pas linéaire et le lien entre limitation du taux de croissance, quantité et qualité de production sera à étudier. Des procédés semi-continus avec du glycérol comme source de carbone pourront également être des pistes d'optimisation du procédé de production. En effet, la charge métabolique liée à la production de protéines recombinante diffère en fonction de la source de carbone. Ce phénomène est en partie lié au fait que le coût énergétique associé à la synthèse de chaque acide aminé est modifié en fonction de la source de carbone consommé par la bactérie (Kaleta et al. 2013).

La mise en place de procédés semi-continus reposant sur l'utilisation d'un système MetaProd ou T7 nécessite également de connaître l'effet de l'évolution du rapport de concentration entre l'IPTG (fixe) et la biomasse sur la force d'expression du gène d'intérêt. En effet, la force d'expression génique peut diminuer lorsque la quantité d'IPTG par bactérie diminue (Striedner et al. 2017). Une étude pourra être réalisée sur la vitesse d'ajout de l'IPTG au moment de l'induction. En effet, la mise en place lente d'une vitesse de production par augmentation progressive de la concentration en IPTG dans le milieu de culture pourrait permettre de laisser un temps d'adaptation plus long à la bactérie et ainsi diminuer l'apparition de limitations métaboliques. Une alimentation en IPTG permettrait également de maintenir constant le rapport entre la quantité d'IPTG et le nombre de bactéries dans le bioréacteur.

8.2.3. Modèle prédictif de production de protéines recombinantes avec le système MetaProd

Les études menées sur la production des protéines UL130 et Hla par la souche *E. coli* BL21 ont permis de d'identifier les contraintes métaboliques liées à la production. En particulier, ces études ont permis de mieux caractériser la charge métabolique associée à la production de protéines recombinantes, par la mesure de la vitesse de production spécifique, du taux de croissance et la quantification de la concentration intracellulaire en ppGpp. Il s'agit là d'une première base pour définir un modèle prédictif de la production de protéine recombinante chez *E. coli*, visant à l'optimisation rationnelle du système MetaProd. Pour cela, il convient de compléter les analyses réalisées dans le cadre de ce projet par la mesure de la force d'expression du gène d'intérêt, activée en fonction de la concentration en inducteur et inhibée en fonction de la concentration en ppGpp. L'objectif est de déterminer le point d'équilibre que va atteindre le système entre la vitesse de production spécifique et le taux de croissance contrôlées par la concentration en ppGpp lors de la modification de la concentration en inducteur. A terme, un tel modèle permettrait de distinguer les données caractéristiques de chaque protéine (usage des codons, proportion en acides aminés, structure...) permettant de prédire au mieux la concentration en ppGpp associée à leur vitesse de production. Ces données permettraient de déterminer la meilleure stratégie à adopter pour la production de ces protéines.

Bibliographie

- Ali, S.A. et al., 2015. Use of FabV-Triclosan Plasmid Selection System for Efficient Expression and Production of Recombinant Proteins in Escherichia coli. , pp.1–15.
- Andersson, D.I. & Levin, B.R., 1999. The biological cost of antibiotic resistance. *Antimicrobials*, pp.489–493.
- Assenberg, R. et al., 2013. Advances in recombinant protein expression for use in pharmaceutical research. *Current Opinion in Structural Biology*, 23(3), pp.393–402. Available at: <http://dx.doi.org/10.1016/j.sbi.2013.03.008>.
- Azir, A.A.N. & Arinarayanan, R.A.H., 2016. Mini-review (p) ppGpp and the bacterial cell cycle. *Journal of Biosciences*, 41(June), pp.277–282.
- Baba, T. et al., 2006. Construction of Escherichia coli K-12 in-frame , single-gene knockout mutants : the Keio collection. , 4474.
- Bachinger, T. et al., 2001. Non-invasive detection of the metabolic burden on recombinant microorganisms during fermentation processes. *Journal of Chemical Technology & Biotechnology*, 76(8), pp.885–889. Available at: <http://doi.wiley.com/10.1002/jctb.463> [Accessed April 20, 2015].
- Baig, F. et al., 2015. Dynamic Transcriptional Response of Escherichia coli to Inclusion Body Formation. *Biotechnology and Bioengineering*, 111(5), pp.980–999.
- Balsalobre, C., 2011. Concentration matters!! ppGpp, from a whispering to a strident alarmone. *Molecular microbiology*, 79(4), pp.827–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21299641> [Accessed January 6, 2016].
- Baneyx, F. & Mujacic, M., 2004. Recombinant protein folding and misfolding in Escherichia coli. *Nature biotechnology*, 22(11), pp.1399–1408.
- Barker, M.M. et al., 2001. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. *Journal of molecular biology*, 305(4), pp.673–

688. Available at:
<http://www.ncbi.nlm.nih.gov/pubmed/11162084>\n<http://www.ncbi.nlm.nih.gov/pubmed/11162085>.
- Battesti, A. & Bouveret, E., 2006. Acyl carrier protein/SpoT interaction, the switch linking SpoT-dependent stress response to fatty acid metabolism. *Molecular microbiology*, 62(4), pp.1048–63. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17078815> [Accessed March 20, 2015].
- Battesti, A., Majdalani, N. & Gottesman, S., 2015. Stress sigma factor RpoS degradation and translation are sensitive to the state of central metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, pp.1–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25847996> [Accessed April 13, 2015].
- Battesti, A., Majdalani, N. & Gottesman, S., 2011. The RpoS-mediated general stress response in Escherichia coli. *Annual review of microbiology*, 65, pp.189–213. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21639793> [Accessed January 8, 2015].
- Bennett, B.D. et al., 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature Chemical Biology*, 5(8), pp.593–599.
- Boehm, A. et al., 2009. Second messenger signalling governs Escherichia coli biofilm induction upon ribosomal stress. *Molecular Microbiology*, 72(6), pp.1500–1516.
- Boniecka, J. et al., 2017. Within and beyond the stringent response - RSH and (p) ppGpp in plants. *Planta*, 246, pp.817–842.
- Bosdriesz, E. et al., 2015. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *The FEBS journal*, 282(10), pp.2029–44. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25754869> [Accessed November 4, 2015].
- Boström, M. et al., 2005. Effect of substrate feed rate on recombinant protein secretion , degradation and inclusion body formation in Escherichia coli. *Applied microbiology and biotechnology*, 68, pp.82–90.

- Bougdour, A. & Gottesman, S., 2007. ppGpp regulation of RpoS degradation via anti-adaptor protein IraP. *Proceedings of the National Academy of Sciences of the United States of America*, 104(31), pp.12896–12901.
- Bradley, M.D. et al., 2017. Effects of Fis on Escherichia coli gene expression during different growth stages. , (2007), pp.2922–2940.
- Bremer, H. & Dennis, P.P., 2008. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus*, 3(1), pp.1–56. Available at: <http://www.asmscience.org/content/view.action> \n <http://ctbp.ucsd.edu/qbio/beemer96.pdf> \n <http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosal.5.2.3> \n http://ecosal.org/images/A_images/chapter_pdf/5.2.3_module.pdf \n (null).
- Bremer, H. & Dennis, P.P., 1996. *Modulation of chemical composition and other parameters of the cell by growth rate* Escherichi. F. C. Neidhardt, ed., Washington, D.C.
- Brown, D.R. et al., 2014. Nitrogen stress response and stringent response are coupled in Escherichia coli. *Nature communications*, 5(May), p.4115. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4066584&tool=pmcentrez&rendertype=abstract> [Accessed March 5, 2015].
- Brown, L. et al., 2002. DksA Affects ppGpp Induction of RpoS at a Translational Level. *Journal of Bacteriology*, 184(16), pp.4455–4465. Available at: <http://jb.asm.org/cgi/doi/10.1128/JB.184.16.4455-4465.2002>.
- Bruschi, M. et al., 2014. Production of the short peptide surfactant DAMP4 from glucose or sucrose in high cell density cultures of Escherichia coli BL21(DE3). *Microbial cell factories*, 13, p.99. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4229601&tool=pmcentrez&rendertype=abstract> [Accessed June 16, 2015].
- Burgess, K. et al., 2011. Semi-targeted analysis of metabolites using capillary- flow ion chromatography coupled to high-resolution mass spectrometry. *Rapid Commun. Mass Spectrom.*, 25, pp.3447–

3452.

Camps, M., 2010. Modulation of ColE1-Like Plasmid Replication for Recombinant Gene Expression. *Recent patents on DNA & gene sequences*, 4, pp.58–73.

Carneiro, S. et al., 2012. Influence of the RelA Activity on E. coli Metabolism by Metabolite Profiling of Glucose-Limited Chemostat Cultures. *Metabolites*, 2(4), pp.717–32. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3901239&tool=pmcentrez&rendertype=abstract> [Accessed April 20, 2015].

Carneiro, S. et al., 2011. Stringent response of Escherichia coli: revisiting the bibliome using literature mining. *Microbial Informatics and Experimentation*, 1(1), p.14. Available at: <http://www.microbialinformaticsj.com/content/1/1/14>.

Caserta, E. et al., 2006. The nucleotide-binding site of bacterial translation initiation factor 2 (IF2) as a metabolic sensor. *PNAS*, 103(38).

Cashel, M. et al., 1996. *Escherichia coli and Salmonella Molecular and Cell Biology*, ASM Press, Washington DC.

Castaño-Cerezo, S. et al., 2015. Regulation of acetate metabolism in Escherichia coli BL21 by protein N(ε)-lysine acetylation. *Applied microbiology and biotechnology*, 99(8), pp.3533–45. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25524697> [Accessed November 27, 2015].

Ceroni, F. et al., 2015. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nature methods*, 12(5), pp.415–418.

Chen, R.T. & Orenstein, W.A., 1996. Epidemiologic Methods in Immunization Programs. *Epidemiologic Reviews*, 18(2).

Choi, J.H., Keum, K.C. & Lee, S.Y., 2006. Production of recombinant proteins by high cell density culture of Escherichia coli. *Chemical Engineering Science*, 61(3), pp.876–885. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0009250905002691> [Accessed April 9, 2015].

Cserjan-Puschmann, M. et al., 1999. Metabolic approaches for the optimisation of recombinant

- fermentation processes. *Applied Microbiology and Biotechnology*, 53(1), pp.43–50. Available at: <http://link.springer.com/10.1007/s002530051612>.
- Dalebroux, Z.D. & Swanson, M.S., 2012. ppGpp: magic beyond RNA polymerase. *Nature Reviews Microbiology*, 10(3), pp.203–212. Available at: <http://dx.doi.org/10.1038/nrmicro2720>.
- Van Dam, J.C. et al., 2002. Analysis of glycolytic intermediates in *Saccharomyces cerevisiae* using anion exchange chromatography and electrospray ionization with tandem mass spectrometric detection. *Analytica Chimica Acta*, 460, pp.209–218.
- Demain, A.L. & Vaishnav, P., 2009. Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances*, 27(3), pp.297–306. Available at: <http://dx.doi.org/10.1016/j.biotechadv.2009.01.008>.
- DeNapoli, J., Tehranchi, A. & Wang, J., 2013. Dose-dependent reduction of replication elongation rate by (p) ppGpp in *Escherichia coli* and *Bacillus subtilis*. *Molecular microbiology*, 88(1), pp.93–104. Available at: <http://onlinelibrary.wiley.com/doi/10.1111/mmi.12172/full> [Accessed January 6, 2016].
- Ding Jun, J., Cedric, C. & Yan Ning, Z., 2012. Growth rate regulation in *Escherichia coli*. *FEMS Microbiology Reviews*, 36(2), pp.269–287.
- Dong, H., Nilsson, L. & Kurland, C.G., 1995. Gratuitous Overexpression of Genes in *Escherichia coli* Leads to Growth Inhibition and Ribosome Destruction. *Journal of bacteriology*, 177(6), pp.1497–1504.
- Dumon-seignovert, L., Cariot, G. & Vuillard, L., 2004. The toxicity of recombinant proteins in *Escherichia coli* : a comparison of overexpression in BL21 (DE3), C41 (DE3), and C43 (DE3). *Protein Expression and Purification*, 37, pp.203–206.
- Durfee, T., Hansen, A.M., et al., 2008. Transcription profiling of the stringent response in *Escherichia coli*. *Journal of Bacteriology*, 190(3), pp.1084–1096.
- Edwards, A.N. et al., 2011. Circuitry Linking the Csr and Stringent Response Global Regulatory Systems.

Mol Microbiol., 80(6), pp.1561–1580.

Enjalbert, B., Letisse, F. & Portais, J., 2013. Physiological and Molecular Timing of the Glucose to Acetate Transition in *Escherichia coli*. *Metabolites*, 3, pp.820–837.

Fernández-Fernández, Mario Rodríguez-González, P. et al., 2015. Simultaneous Determination of Creatinine and Creatine in Human Serum by Double-Spike Isotope Dilution Liquid Chromatography– Tandem Mass Spectrometry (LC-MS/MS) and Gas Chromatography– Mass Spectrometry (GC-MS). *Analytical Chemistry*, 87, pp.3755–3763.

Field, C.M. & Summers, D.K., 2011. Multicopy plasmid stability : Revisiting the dimer catastrophe. *Journal of Theoretical Biology*, 291, pp.119–127. Available at: <http://dx.doi.org/10.1016/j.jtbi.2011.09.006>.

Funabashi, H., Haruyama, T. & Mie, M., 2002. Non-destructive monitoring of rpoS promoter activity as stress marker for evaluating cellular physiological status. *Journal of bacteriology*, 95, pp.85–93.

García Alonso, J.I. & Rodríguez-Gonzalez, P., 2013. *Isotope Dilution Mass Spectrometry*, Royal Society of Chemistry: Cambridge, U.K.

Gil, A. et al., 2015. Stability of energy metabolites — An often overlooked issue in metabolomics studies : *Electrophoresis*, 36, pp.2156–2169.

Gil, A., Siegel, D. & Hjalmar, S.B., 2017. The degradation of nucleotide triphosphates extracted under boiling ethanol conditions is prevented by the yeast cellular matrix. *Metabolomics*, 13(1), pp.1–10. Available at: "<http://dx.doi.org/10.1007/s11306-016-1140-4>.

Gopal, G.J. & Kumar, A., 2013. Strategies for the production of recombinant protein in *Escherichia coli*. *The protein journal*, 32(6), pp.419–25. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23897421>.

Grabherr, R. et al., 2002. Stabilizing Plasmid Copy Number to Improve Recombinant Protein Production.

Graumann, K. & Premstaller, A., 2006. Manufacturing of recombinant therapeutic proteins in microbial

systems. *Biotechnology journal*, 1, pp.164–186.

Hanson, G. & Collier, J., 2017. Codon optimality, bias and usage in translation and mRNA decay. *Nature Publishing Group*, 19(1), pp.20–30. Available at: <http://dx.doi.org/10.1038/nrm.2017.91>.

Harms, A., Maisonneuve, E. & Gerdes, K., 2016. Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science (New York, N.Y.)*, 4268.

Haseltine, W.A. & Block, R., 1973. Synthesis of Guanosine Tetra- and Pentaphosphate Requires the Presence of a Codon-Specific, Uncharged Transfer Ribonucleic Acid in the Acceptor Site of Ribosomes. *Proceedings of the National Academy of Sciences*, 70(5), pp.1564–1568. Available at: <http://www.pnas.org/cgi/doi/10.1073/pnas.70.5.1564>.

Haurlyuk, V., 2015. Recent functional insights into the role of (p) ppGpp in bacterial physiology. *Nature Publishing Group*, 13(5), pp.298–309. Available at: <http://dx.doi.org/10.1038/nrmicro3448>.

Helling, R.B., 1998. Pathway Choice in Glutamate Synthesis in Escherichia coli. *Journal of bacteriology*, 180(17), pp.4571–4575.

Heyland, J., Blank, L.M. & Schmid, A., 2011. Quantification of metabolic limitations during recombinant protein production in Escherichia coli. *Journal of biotechnology*, 155(2), pp.178–84. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21723332> [Accessed April 3, 2015].

Hirvonen, C.A. et al., 2001. Contributions of UP elements and the transcription factor FIS to expression from the seven rrn P1 promoters in Escherichia coli. *Journal of Bacteriology*, 183(21), pp.6305–6314.

Hogg, T. et al., 2004. Conformational Antagonism between Opposing Active Sites in a Bifunctional RelA / SpoT Homolog Modulates (p) ppGpp Metabolism during the Stringent Response. *Cell*, 117, pp.57–68.

Huang, C.-J., Lin, H. & Yang, X., 2012. Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. *Journal of industrial microbiology & biotechnology*, 39(3),

- pp.383–99. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22252444> [Accessed November 21, 2014].
- Hui, S. et al., 2015. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular systems biology*, 11(1), p.784. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4358657&tool=pmcentrez&rendertype=abstract>.
- Hüttener, M. et al., 2018. Stringent Response and AggR-Dependent Virulence Regulation in the Enterocoagulative Escherichia coli Strain 042. *Frontiers in Microbiology*, 9(April), pp.1–9.
- Ihara, Y., Ohta, H. & Masuda, S., 2015. A highly sensitive quantification method for the accumulation of alarmone ppGpp in Arabidopsis thaliana using UPLC-ESI-qMS/MS. *Journal of Plant Research*, 128, pp.511–518.
- Iyer, S. et al., 2018. Distinct mechanisms coordinate transcription and translation under carbon and nitrogen starvation in Escherichia coli. *Nature Microbiology*. Available at: <http://dx.doi.org/10.1038/s41564-018-0161-3>.
- Jia, B. & Jeon, C.O., 2016. High-throughput recombinant protein expression in Escherichia coli : current status and future perspectives. *Open Biology*, 6, pp.1–17.
- Kaleta, C. et al., 2013. Metabolic costs of amino acid and protein production in Escherichia coli. *Biotechnology journal*, 8(9), pp.1105–14. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23744758> [Accessed November 27, 2015].
- Kanjee, U., Ogata, K. & Houry, W. a, 2012. Direct binding targets of the stringent response alarmone (p)ppGpp. *Molecular microbiology*, 85(6), pp.1029–43. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22812515> [Accessed March 26, 2015].
- Kiefer, P. et al., 2007. Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry. *Analytical Biochemistry*, 360, pp.182–188.

- Kim, D. et al., 2015. LC – MS-based absolute metabolite quantification : application to metabolic flux measurement in trypanosomes. *Metabolomics*, 11(6), pp.1721–1732. Available at: "http://dx.doi.org/10.1007/s11306-015-0827-2.
- Kimball, E. & Rabinowitz, J.D., 2006. Identifying Decomposition Products in Extracts of Cellular Metabolites. *Analytical biochemistry*, 358(2), pp.273–280.
- Kimple, M.E., Brill, A.L. & Pasker, R.L., 2015. Overview of Affinity Tags for Protein Purification. *Curr Protoc Protein Sci*, pp.608–616.
- Kolmsee, T. et al., 2011. Differential stringent control of Escherichia coli rRNA promoters: Effects of ppGpp, DksA and the initiating nucleotides. *Microbiology*, 157(10), pp.2871–2879.
- Korz, D.J. et al., 1995. Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. *Journal of Biotechnology*, 39(1), pp.59–65. Available at: <http://www.sciencedirect.com/science/article/B6T3C-3YRSMBG-V/2/b58596472a11d7582e15b10d45b790b2>.
- Legastelois, I. et al., 2017. Non-conventional expression systems for the production of vaccine proteins and immunotherapeutic molecules. *HUMAN VACCINES & IMMUNOTHERAPEUTICS*, 13(4), pp.947–961.
- Lemaire, D. et al., 2012. Coping with genetic diversity : the contribution of pathogen and human genomics to modern vaccinology. *Brazilian Journal of Medical and Biological Research*, 45(5), pp.376–385.
- Lemke, J.J. et al., 2011. Direct regulation of Escherichia coli ribosomal protein promoters by the transcription factors ppGpp and DksA. *PNAS*, 108(18), pp.5712–5717.
- Li, W. et al., 2016. Effects of amino acid starvation on RelA diffusive behavior in live Escherichia coli. *Mol Microbiol.*, 99(3), pp.571–585.
- Li, Z., Nimtz, M. & Rinas, U., 2014a. The metabolic potential of Escherichia coli BL21 in defined and rich medium. *Microbial cell factories*, 13(45), pp.1–17.

- Li, Z., Nimtz, M. & Rinas, U., 2014b. The metabolic potential of Escherichia coli BL21 in defined and rich medium. *Microbial cell factories*, 13(1), p.45. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4021462&tool=pmcentrez&rendertype=abstract>.
- Liu, K., Bittner, A.N. & Wang, J.D., 2015. Diversity in (p)ppGpp metabolism and effectors. *Current opinion in microbiology*, 24C, pp.72–79. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25636134> [Accessed February 2, 2015].
- Loveland, A.B. et al., 2016. Ribosome • RelA structures reveal the mechanism of stringent response activation. *eLIFE*, 5, pp.1–61.
- Maciag, M. et al., 2010. ppGpp inhibits the activity of Escherichia coli DnaG primase. *Plasmid*, 63(1), pp.61–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19945481> [Accessed November 27, 2015].
- Maeda, M., Shimada, T. & Ishihama, A., 2015. Strength and Regulation of Seven rRNA Promoters in Escherichia coli. *PLoS ONE*, 10(12), pp.1–19.
- Magnusson, L.U. et al., 2007. Identical, independent, and opposing roles of ppGpp and DksA in Escherichia coli. *Journal of bacteriology*, 189(14), pp.5193–202. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1951846&tool=pmcentrez&rendertype=abstract> [Accessed April 20, 2015].
- Magnusson, L.U., Farewell, A. & Nyström, T., 2005. ppGpp: a global regulator in Escherichia coli. *Trends in microbiology*, 13(5), pp.236–42. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15866041> [Accessed July 10, 2014].
- Mahalik, S., Sharma, A.K. & Mukherjee, K.J., 2014. Genome engineering for improved recombinant protein expression in Escherichia coli. *Microbial cell factories*, 13(1), p.177. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4300154&tool=pmcentrez&rendertype=abstract> [Accessed April 3, 2015].
- Mairhofer, J. et al., 2013. Comparative Transcription Profiling and In-Depth Characterization of

- Plasmid-Based and Plasmid-Free Escherichia coli Expression Systems under Production Conditions. *Applied and environmental microbiology*, 79(12), pp.3802–3812.
- Malik, A., 2016. Protein fusion tags for efficient expression and purification of recombinant proteins in the periplasmic space of E . coli. *3 Biotech*, 6(1), pp.1–7. Available at: "<http://dx.doi.org/10.1007/s13205-016-0397-7>.
- Marisch, K., Bayer, K., Scharl, T., et al., 2013. A Comparative Analysis of Industrial Escherichia coli K – 12 and B Strains in High-Glucose Batch Cultivations on. *PLoS one*, 8(8).
- Marisch, K., Bayer, K., Cserjan-Puschmann, M., et al., 2013. Evaluation of three industrial Escherichia coli strains in fed-batch cultivations during high-level SOD protein production. *Microbial cell factories*, 12(1), p.58. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3698069&tool=pmcentrez&rendertype=abstract> [Accessed February 3, 2015].
- Marschall, L., Sagmeister, P. & Herwig, C., 2016. Tunable recombinant protein expression in E . coli : enabler for continuous processing ? *Applied microbiology and biotechnology*, 100, pp.5719–5728.
- Mashego, M.R. et al., 2004. MIRACLE : Mass Isotopomer Ratio Analysis of U- 13 C-Labeled Extracts . A New Method for Accurate Quantification of Changes in Concentrations of Intracellular Metabolites. *Biotechnology and Bioengineering*, 85(6), pp.620–628.
- Mechold, U. et al., 2013. Differential regulation by ppGpp versus pppGpp in Escherichia coli. *Nucleic Acids Research*, 41(12), pp.6175–6189.
- Mechold, U. et al., 2002. Intramolecular Regulation of the Opposing (p) ppGpp Catalytic Activities of Rel Seq , the Rel / Spo Enzyme from Streptococcus equisimilis. *Journal of bacteriology*, 184(11), pp.2878–2888.
- Meier, S., Jensen, P.R. & Duus, J.Ø., 2012. Direct Observation of Metabolic Differences in Living Escherichia Coli Strains K-12 and BL21. *Chembiochem*, (Figure 1), pp.308–310.
- Meng, W. et al., 2001. UP element-dependent transcription at the Escherichia coli rrnB P1 promoter:

- positional requirements and role of the RNA polymerase alpha subunit linker. *Nucleic acids research*, 29(20), pp.4166–78. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=60210&tool=pmcentrez&rendertype=abstract>.
- Millard, P. et al., 2014. Sampling of intracellular metabolites for stationary and non-stationary (13)C metabolic flux analysis in *Escherichia coli*. *Analytical biochemistry*, 465, pp.38–49. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25102204> [Accessed June 26, 2015].
- Murray, H.D., Appleman, J.A. & Gourse, R.L., 2003. Regulation of the *Escherichia coli* rrnB P2 promoter. *Journal of Bacteriology*, 185(1), pp.28–34.
- Murray, K.D. & Bremer, H., 1996. Control of spoT -dependent ppGpp Synthesis and Degradation in *Escherichia coli*. *Journal of molecular biology*, 259, pp.41–57.
- Nascimento, I.P. & Leite, L.C.C., 2012. Recombinant vaccines and the development of new vaccine strategies Recombinant vaccines and the development of new vaccine strategies. *Brazilian Journal of Medical and Biological Research*, 45(12), pp.1102–1111.
- Navon, S. & Pilpel, Y., 2011. The role of codon selection in regulation of translation efficiency deduced from synthetic libraries. *Genome Biology*, 12(2), p.R12. Available at: <http://genomebiology.com/2011/12/2/R12>.
- Nazir, A. & Harinarayanan, R., 2016. (p)ppGpp and the bacterial cell cycle. *Journal of Biosciences*, 41(2), pp.277–282.
- Negrete, A., Ng, W. & Shiloach, J., 2010. Glucose uptake regulation in *E. coli* by the small RNA SgrS : comparative analysis of *E. coli* K-12 (JM109 and MG1655) and *E. coli* B (BL21). *Microbial cell factories*, 3(75), pp.1–9.
- Nokes, D.J. & Anderson, R.M., 1991. Vaccine safety versus vaccine efficacy in mass immunisation programmes In. *The Lancet*, 338, pp.1309–1312.
- Noronha, S.B. et al., 2000. Investigation of the TCA cycle and the glyoxylate shunt in *Escherichia coli*

- BL21 and JM109 using ¹³C-NMR/MS. *Biotechnology and Bioengineering*, 68(3), pp.316–327.
- Norsyahida, a, Rahmah, N. & Ahmad, R.M.Y., 2009. Effects of feeding and induction strategy on the production of BmR1 antigen in recombinant *E. coli*. *Letters in applied microbiology*, 49(5), pp.544–50. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19832937> [Accessed April 20, 2015].
- Ouerdane, L. & Meija, J., 2009. General Equation for Multiple Spiking Isotope Dilution Mass Spectrometry. *Analytical Chemistry*, 81(12), pp.5075–5079.
- Özkan, P. et al., 2005. Metabolic flux analysis of recombinant protein overproduction in *Escherichia coli*. *Biochemical Engineering Journal*, 22(2), pp.167–195. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1369703X04002360> [Accessed April 3, 2015].
- Palmer, I. & Wingfield, P.T., 2004. Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from *Escherichia coli*. *Curr Protoc Protein Sci*, (6.3), pp.1–25.
- Patrone, M. et al., 2005. Human Cytomegalovirus UL130 Protein Promotes Endothelial Cell Infection through a Producer Cell Modification of the Virion †. *Journal of virology*, 79(13), pp.8361–8373.
- Paul, B.J., Berkmen, M.B. & Gourse, R.L., 2005. DksA potentiates direct activation of amino acid promoters by ppGpp. *PNAS*, 102(22), pp.7823–7828.
- Peubez, I. et al., 2010. Antibiotic-free selection in *E. coli*: new considerations for optimal design and improved production. *Microbial Cell Factories*, 9(65), pp.1–10.
- Phue, J.N. et al., 2005. Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: Differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and Northern blot analyses. *Biotechnology and Bioengineering*, 90(7), pp.805–820.
- Piskunova, J. et al., 2017. Peptide-nucleotide Antibiotic Microcin C Is a Potent Inducer of Stringent Response and Persistence in Both Sensitive and Producing Cells. *Molecular Microbiology*, 64(8), pp.2482–2488. Available at: <http://doi.wiley.com/10.1111/mmi.13640>.
- Polisky, B., 1988. ColEI Replication Control Circuitry: Sense from Antisense Minireview. *Cell*, 55, pp.929–932.

- Potrykus, K. & Cashel, M., 2008. (p)ppGpp: still magical? *Annual review of microbiology*, 62, pp.35–51. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18454629> [Accessed December 28, 2014].
- Pramanik, J. & Keasling, J.D., 1997. Stoichiometric Model of Escherichia coli Metabolism : Incorporation of Growth-Rate Dependent Biomass Composition and Mechanistic Energy Requirements. *Biotechnology and Bioengineering*, 56(4), pp.398–421.
- Rabinowitz, J.D. & Kimball, E., 2007. Acidic Acetonitrile for Cellular Metabolome Extraction from Escherichia coli. *Analytical Chemistry*, 79(16), pp.6167–6173.
- Rajendran, H., Helen, M. & Michael, C., 2008. Synthetic growth phenotypes of E.coli lacking ppGpp and transketolase A (tktA) are due to ppGpp-mediated transcriptional regulation of tktB. *Molecular microbiology*, 69(4), pp.882–894.
- Reidl, J. et al., 2017. Stringent factor and proteolysis control of sigma factor RpoS expression in Vibrio cholerae. *Int. J. Med. Microbiology*, 16, pp.1438–4221. Available at: <http://dx.doi.org/10.1016/j.ijmm.2017.01.006>.
- Reitzer, L., 2003. Nitrogen assimilation and global regulation in Escherichia coli. *Annual review of microbiology*, 57, pp.155–76. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12730324> [Accessed December 28, 2014].
- Riesenber, D., Bergter, F. & Csaba, K., 1984. Effect of Serine Hydroxamate and Methyl a-D-glucopyranoside Treatment on Nucleoside Polyphosphate P o o l s , RNA and Protein Accumulation in Streptomyces hygroscopicus. *Journa of General Microbiology*, 130, pp.2549–2558.
- Rodriguez, A. et al., 2017. Plasmid-Encoded Biosynthetic Genes Alleviate Metabolic Disadvantages While Increasing Glucose Conversion to Shikimate in an Engineered Escherichia coli Strain. , 114(6), pp.1319–1330.
- Roghanian, M., Zenkin, N. & Yuzenkova, Y., 2015. Bacterial global regulators DksA/ppGpp increase fidelity of transcription. *Nucleic acids research*, 43(3), pp.1529–36. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4330370&tool=pmcentrez&renderty>

pe=abstract [Accessed November 27, 2015].

- Rosano, G.L. & Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli* : advances and challenges. *Frontiers in Microbiology*, 5(172), pp.1–17.
- Ross, W. et al., 2017. ppGpp binding to a site at the RNAP-DksA interface accounts for its dramatic effects on transcription initiation during the stringent response. , 62(6), pp.811–823.
- Samarrai, W. et al., 2011. Differential responses of *Bacillus subtilis* rRNA promoters to nutritional stress. *Journal of Bacteriology*, 193(3), pp.723–733.
- Sandén, A.M. et al., 2003. Limiting factors in *Escherichia coli* fed-batch production of recombinant proteins. *Biotechnology and Bioengineering*, 81(2), pp.158–166.
- Sarubbi, E., Rudd, K.E. & Cashel, M., 1988. Basal ppGpp level adjustment shown by new *spoT* mutants affect steady state growth rates and *rrnA* ribosomal promoter regulation in *Escherichia coli*. *Molecular genetics and genomics : MGG*, 213, pp.214–222.
- Scott, M. et al., 2014. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Molecular systems biology*, 10(8), p.747. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25149558> \n <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4299513> \n <http://msb.embopress.org/cgi/doi/10.15252/msb.20145379>.
- Scott, M. et al., 2010. Interdependence of Cell Growth Origins and Consequences. *Science*, 330(November), pp.1099–1102.
- Sette, A. & Rino, R., 2010. Reverse Vaccinology: Developing Vaccines in the Era of Genomics. *Immunity*, 33(4), pp.530–541.
- Shachrai, I. et al., 2010. Article Cost of Unneeded Proteins in *E. coli* Is Reduced after Several Generations in Exponential Growth. *Molecular Cell*, 38(5), pp.758–767. Available at: <http://dx.doi.org/10.1016/j.molcel.2010.04.015>.
- Sharma, A.K. et al., 2011. Comparative transcriptomic profile analysis of fed-batch cultures expressing different recombinant proteins in *Escherichia coli*. *AMB Express*, 1(1), p.33. Available at:

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3214799&tool=pmcentrez&rendertype=abstract> [Accessed April 20, 2015].

Sharma, U.K. & Chatterji, D., 2010. Transcriptional switching in *Escherichia coli* during stress and starvation by modulation of sigma activity. *FEMS microbiology reviews*, 34(5), pp.646–57. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20491934> [Accessed November 25, 2015].

Sherlock, M.E., Sudarsan, N. & Breaker, R.R., 2018. Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. *PNAS*, pp.1–6.

Shiloach, J. et al., 2010. Analyzing metabolic variations in different bacterial strains , historical perspectives and current trends – example *E. coli*. *Current Opinion in Biotechnology*, 21(1), pp.21–26. Available at: <http://dx.doi.org/10.1016/j.copbio.2010.01.001>.

Shyp, V. et al., 2012. Positive allosteric feedback regulation of the stringent response enzyme RelA by its product. *EMBO reports*, 13(9), pp.835–839.

Simen, J.D. et al., 2017. Transcriptional response of *Escherichia coli* to ammonia and glucose fluctuations. *Microbial Biotechnology*, 10, pp.858–872.

Somerville, C.R. & Ahmed, A., 1979. Mutants of *Escherichia coli* Defective in the Degradation of Guanosine 5' -Triphosphate , 3' -Diphosphate (pppGpp). , 323, pp.315–323.

Sørensen, H.P. & Mortensen, K.K., 2005. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of biotechnology*, 115(2), pp.113–28. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15607230> [Accessed July 10, 2014].

Srivatsan, A. & Wang, J.D., 2008. Control of bacterial transcription, translation and replication by (p)ppGpp. *Current opinion in microbiology*, 11(2), pp.100–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18359660> [Accessed December 30, 2014].

Stifling, C.J., Stewart, G. & Sherratt, D.J., 1988. Multicopy plasmid stability in *Escherichia coli* requires host-encoded functions that lead to plasmid site-specific recombination. *Molecular genetics and genomics : MGG*, 214, pp.80–84.

- Stockar, U. Von & Liu, J., 1999. Does microbial life always feed on negative entropy ? Thermodynamic analysis of microbial growth. *Biochimica et biophysica acta*, 1412, pp.191–211.
- Stoletzki, N. & Eyre-walker, A., 2007. Synonymous Codon Usage in Escherichia coli : Selection for Translational Accuracy. *Molecular biology Evolution*, 24(2), pp.374–381.
- Striedner, G., Huber, J. & Reinisch, D., 2017. Method for producing a recombinant protein on a manufacturing scale.
- Studier, F.W. et al., 1990. Cloning and expression of the gene for bacteriophage T7 RNA polymerase.
- Studier, F.W. et al., 1997. Cloning and expression of the gene for bacteriophage T7 RNA polymerase.
- Studier, F.W. & Dubendorff, J.W., 1998. Cloning and expression of autogenes encoding RNA polymerase of T7-like bacteriophage.
- Studier, F.W. & Moffatt, B.A., 1986. Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes. *Journal of molecular biology*, 189, pp.113–130.
- Summers, D.K., Beton, C.W.H. & Withers, H.L., 1993. Multicopy plasmid instability : the dimer catastrophe hypothesis. *Molecular microbiology*, 8(6), pp.1031–1038.
- Svitils, A.L., Cashel, M. & Zyskind, J.W., 1993. Guanosine Tetraphosphate Inhibits Protein Synthesis in Vivo. *The Journal of biological chemistry*, 268(4), pp.2307–2311.
- Swartz, J., 1996. Escherichia coli recombinant DNA technology. In *Escherichia coli and Salmonella*. pp. 1693–1712.
- Takahashi, K., Kasai, K. & Ochi, K., 2004. Identification of the bacterial alarmone guanosine 5'-diphosphate 3'-diphosphate (ppGpp) in plants. *PNAS*, 101(12), pp.4320–4324.
- Tengs, T. et al., 1995. Five-Hundred Life-Saving Interventions and Their Cost- Effectiveness. *Risk Analysis*, 15(3), pp.369–390.
- Tosa, T., 1971. Effect of Serine Hydroxamate on the Growth of Escherichia coli. *Journal of bacteriology*, 106(3), pp.966–971.

- Tosa, T. & Pizer, L., 1971. Biochemical Bases for the Antimetabolite Action of L-Serine Hydroxamate. *Journal of bacteriology*, 106(3), pp.972–982.
- Traxler, M.F. et al., 2011. Discretely calibrated regulatory loops controlled by ppGpp partition gene induction across the “feast to famine” gradient in *Escherichia coli*. *Mol Microbiol.*, 79(4), pp.830–845.
- Traxler, M.F., Summers, S.M., Nguyen, H.-T., et al., 2008. The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Molecular microbiology*, 68(5), pp.1128–48. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3719176&tool=pmcentrez&rendertype=abstract> [Accessed October 27, 2014].
- Varik, V. et al., 2017. HPLC-based quantification of bacterial housekeeping nucleotides and alarmone messengers ppGpp and pppGpp. *Scientific Reports*, 7(11022), pp.1–12. Available at: <http://dx.doi.org/10.1038/s41598-017-10988-6>.
- Verstraeten, N. et al., 2015. Obg and Membrane Depolarization Are Part of a Microbial Bet-Hedging Strategy that Leads to Antibiotic Tolerance. *Molecular Cell*, 59(1), pp.9–21.
- Villadsen, I.S. & Michelsen, O.L.E., 1977. Regulation of PRPP and Nucleoside Tri- and Tetraphosphate Pools in *Escherichia coli* Under Conditions of Nitrogen Starvation. *Journal of bacteriology*, 130(1), pp.136–143.
- Vinella, D. et al., 2005. Iron limitation induces SpoT-dependent accumulation of ppGpp in *Escherichia coli*. *Molecular microbiology*, 56(4), pp.958–70. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15853883> [Accessed April 14, 2015].
- Waegeman, H. et al., 2011. Effect of *iclR* and *arcA* knockouts on biomass formation and metabolic fluxes in *Escherichia coli* K12 and its implications on understanding the metabolism of *Escherichia coli* BL21 (DE3). *BMC microbiology*, 11(1), p.70. Available at: <http://www.biomedcentral.com/1471-2180/11/70>.
- Walsh, G., 2014. Biopharmaceutical benchmarks 2014. *Nature biotechnology*, 32(10), pp.992–1002.

- Wang, H. et al., 2014. Improving the expression of recombinant proteins in *E. coli* BL21 (DE3) under acetate stress: an alkaline pH shift approach. *PLoS one*, 9(11), p.e112777. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4234529&tool=pmcentrez&rendertype=abstract> [Accessed April 3, 2015].
- Wang, J.D., Sanders, G.M. & Grossman, A.D., 2007. Nutritional Control of Elongation of DNA Replication by (p)ppGpp. *Cell*, 128(5), pp.865–875.
- Wang, L., Cardenas, R.B. & Watson, C., 2017. An isotope dilution ultra high performance liquid chromatography-tandem mass spectrometry method for the simultaneous determination of sugars and humectants in tobacco products. *Journal of Chromatography A*, 1514, pp.95–102. Available at: <http://dx.doi.org/10.1016/j.chroma.2017.07.079>.
- Wang, Z. et al., 2002. A Model for Regulation of ColE1-like Plasmid Replication by Uncharged tRNAs in Amino Acid-Starved *Escherichia coli* Cells. *Plasmid*, 47, pp.69–78.
- Waschina, S. et al., 2016. Metabolic network architecture and carbon source determine metabolite production costs. *FEBS Journal*, 283(11), pp.2149–2163.
- Weygand-durasevic, I. et al., 1993. Yeast seryl-tRNA synthetase expressed in *Escherichia coli* recognizes bacterial serine-specific tRNAs in vivo. *European journal of Biochemistry*, 877(214), pp.869–877.
- Wittmann, C. et al., 2007. Response of fluxome and metabolome to temperature-induced recombinant protein synthesis in *Escherichia coli*. *Journal of biotechnology*, 132(4), pp.375–84. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17689798> [Accessed April 3, 2015].
- Wrobel, B. & We, G., 1998. Replication Regulation of ColE1-like Plasmids in Amino Acid-Starved *Escherichia coli*. *Plasmid*, 39, pp.48–62.
- Wu, L. et al., 2005. Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly ¹³C-labeled cell extracts as internal standards. *Analytical Biochemistry*, 336(2), pp.164–171.
- Wu, N. et al., 2015. Ranking of persister genes in the same *Escherichia coli* genetic background

- demonstrates varying importance of individual persister genes in tolerance to different antibiotics. *Frontiers in Microbiology*, 6(SEP), pp.1–11.
- Yakhnin, H. et al., 2012. Complex regulation of the global regulatory gene *csrA*: CsrA-mediated translational repression, transcription from five promoters by σ^{70} and σ^S , and indirect transcriptional activation by CsrA. *Molecular microbiology*, 81(3), pp.689–704.
- Yodo, N., 2018. Engineering Resilience Quantification and System Design Implications : A Literature Survey. , 138(November 2016).
- Yoon, S.H. et al., 2012. Comparative multi-omics systems analysis of Escherichia coli strains B and K-12. *Genome biology*, 13(5), p.R37. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3446290&tool=pmcentrez&rendertype=abstract> [Accessed November 27, 2015].
- Young, C.L., Britton, Z.T. & Robinson, A.S., 2012. Recombinant protein expression and purification : A comprehensive review of affinity tags and microbial applications. *Biotechnology journal*, 7, pp.620–634.
- Yu, C. et al., 2015. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. *Molecular Cell*, 59(5), pp.744–754.
- Zakeri, B. et al., 2012. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *PNAS*, 109(12).
- Zeng, Y. et al., 2009. Characterization of Two Seryl-tRNA Synthetases in Albomycin-Producing. *Antimicrobial agents and chemotherapy*, 53(11), pp.4619–4627.
- Zhang, X. et al., 2002. Kinetic properties of *rrn* promoters in Escherichia coli. *Biochimie*, 84(10), pp.981–996.
- Zhang, X., Liang, S.-T. & Bremer, H., 2006. Feedback control of ribosome synthesis in Escherichia coli is dependent on eight critical amino acids. *Biochimie*, 88, pp.1145–1155.
- Zhang, Y. et al., 2014. DksA Guards Elongating RNA Polymerase Against Ribosome- Stalling-Induced

Bibliographie

Arrest. *Molecular Cell*, 53(5), pp.766–778.

Zhang, Y., Rejman, D. & Gerdes, K., 2018. Novel (p)ppGpp Binding and Metabolizing Proteins of Escherichia coli. *mBio*, 9(2), pp.1–20.

Zhao, X. et al., 2011. A loop-controlled rrnB P1 promoter for high-level expression of heterologous proteins in Escherichia coli. *Biotechnology Letters*, 33(2), pp.327–332.

Zuo, Y., Wang, Y. & Steitz, T.A., 2013. The Mechanism of E. coli RNA Polymerase Regulation by ppGpp is Suggested by the Structure of Their Complex. *Molecular cell*, 50(3), pp.430–436.