
THESE DE DOCTORAT

Spécialité : Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingénieries

Filière : Microbiologie et Biocatalyse industrielle

Présentée le 07 juillet 2009

à

l'Institut National des Sciences Appliquées de Toulouse

par

Clémentine DRESSAIRE

Comprendre l'adaptation de *Lactococcus lactis* par une
approche de biologie intégrative à l'échelle du génome

Jury

| | | |
|-------------------------|--|---------------------|
| Philippe BESSE | Professeur, INSA, Toulouse | Examineur |
| Muriel COCAIGN-BOUSQUET | Chargé de recherche INRA, LISBP, Toulouse | Directrice de thèse |
| Pascale DARAN-LAPUJADE | Assistant Professor, TU Delft, Hollande | Rapportrice |
| Philippe GAUDU | Directeur de recherche INRA, Jouy-en-Josas | Rapporteur |
| Nic LINDLEY | Directeur de recherche CNRS, LISBP, Toulouse | Président |
| Emmanuelle MAGUIN | Directeur de recherche INRA, Jouy-en Josas | Examinatrice |

Ce projet de thèse a été préparé au Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés. UMR-CNRS 5504, UR-INRA 792 du Département de Génie Biochimique et Alimentaire de l'INSA de Toulouse.

Résumé :

L'adaptation de *Lactococcus lactis* à différentes conditions de culture a été appréhendée grâce à une démarche de biologie intégrative. Cette approche intègre les données issues de différents niveaux de régulation et combine diverses techniques de mesure à l'échelle globale (transcriptome, protéome, stabilité des ARN messagers) et locale (suivi des paramètres de culture). Plusieurs outils mathématiques de modélisation (tels que la modélisation numérique et la modélisation statistique) ont été développés pour intégrer l'ensemble de ces données hétérogènes.

Une culture continue de *L. lactis* à différents taux de dilution a permis d'étudier l'influence du taux de croissance sur la physiologie de la bactérie, un paramètre qui n'est jamais distingué de la réponse au stress lors des études dynamiques de l'adaptation. La réponse à la variation du taux de croissance implique majoritairement les fonctions associées à la biogenèse mais demeure extrêmement étendue puisqu'elle affecte l'expression de 30 % des gènes de *L. lactis*. Cette réponse concerne les niveaux d'ARN messagers et de protéines mais aussi les processus cellulaires majeurs que sont la traduction, la dilution et la dégradation. Il a été montré, par une approche de modélisation, que les efficacités de traduction et les vitesses de dégradation des protéines étaient en effet inversement proportionnelles au taux de croissance. Au final, l'influence des différents processus cellulaires a pu être quantifiée par des calculs de coefficients de contrôle.

L'imposition progressive d'une carence en isoleucine lors d'une culture discontinuée en batch a permis de caractériser la réponse, encore peu étudiée, de *L. lactis* à une carence en acide aminé. L'adaptation à ce stress nutritionnel entraîne une vaste réorganisation de la physiologie cellulaire qui se divise en trois types de réponses : une répression globale des principales fonctions biologiques associées à la croissance, une réponse propre au stress imposé visant à lutter spécifiquement contre la carence en isoleucine, ainsi qu'une activation inexplicite de mécanismes en lien avec le stress oxydatif. L'implication de différents mécanismes (réponse stringente, mécanisme lié au taux de croissance, régulations par CodY, GlnR et CcpA) dans la régulation de cette réponse a été évaluée par transcriptomique comparative.

Les déterminants majeurs des concentrations en protéines au sein de la cellule ont été recherchés mathématiquement grâce à un algorithme de sélection de modèles de covariances. Le biais de codons (CAI) s'est avéré être un paramètre majeur, plus important que les concentrations en ARN messagers, suggérant l'existence d'un contrôle génétique prépondérant sur l'adaptation transcriptionnelle. Enfin, il a pu être démontré que le degré d'implication des différents déterminants varie en fonction du mode d'adaptation.

L'approche de biologie intégrative suivie au cours de cette thèse a permis une meilleure compréhension des mécanismes d'adaptation de *L. lactis* et est aujourd'hui entièrement généralisable à d'autres processus comme à d'autres microorganismes.

Mots clés :

Lactococcus lactis, adaptation, biologie intégrative, taux de croissance, niveaux de protéines

Abstract:

A systems biology approach was implemented to study *Lactococcus lactis* adaptation to various growing conditions. This method combines growth parameter monitoring and genome-wide measurement technologies (transcriptome, proteome, messenger RNA stability). Data from these diverse regulation levels were integrated thanks to mathematical tools developed on purpose.

Growth rate influence on *L. lactis* physiology, which is never dissociated from stress responses when studying dynamic adaptation processes, was analysed through continuous culture at various growth rates. This widespread response mainly involves biogenesis-related functions and affects the expression of 30 % of *L. lactis* genes. Both messenger RNA and protein levels are modified but cellular processes such as translation, dilution and degradation are also concerned. As a matter of fact, translation efficiency and protein degradation rates were proved to be inversely proportional to growth rate by a modelling approach. Control coefficient calculations enabled the quantification of cellular processes influences.

The dynamic response of *L. lactis* to isoleucine starvation was studied by the progressive consumption of this amino-acid in a discontinuous batch fermentation. This poorly characterized adaptation process triggers a wide reorganization of cellular physiology that could be divided in three parts: a global repression of the main biological functions related to growth, a response more specific to the encountered stress to struggle against isoleucine starvation and an unexplained activation of oxidative stress-related cellular functions. Comparative transcriptomics allowed the implication of various mechanisms to be quantified in the regulation of this adaptation response (stringent response, growth rate adaptation mechanism, CodY, GlnR and CcpA regulation).

The major biological determinants of protein intracellular concentration were mathematically investigated thanks to a covariance model selection algorithm. Codons bias (CAI) was found to be the most influent parameter, even more than mRNA concentrations, which suggests that genetic control is stronger than transcriptional adaptation. The weight of the different determinants was also found to depend on adaptation modes.

The systems biology approach applied in this work enabled a better understanding of *L. lactis* adaptation mechanisms and will be entirely transposable to other cellular processes as well as other microorganisms.

Key words:

Lactococcus lactis, adaptation, systems biology, growth rate, protein levels

Remerciements

La thèse est autant une aventure humaine que scientifique, c'est la raison pour laquelle je souhaiterai commencer par remercier l'ensemble des personnes qui, de près ou de loin, ont contribué au bon déroulement de ce projet.

En premier lieu je remercie bien sûr Nic Lindley, le directeur du Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés pour m'avoir permis de réaliser mes travaux de thèse dans son unité et surtout d'avoir accepté de présider le jury lors de ma soutenance. Je remercie également Pascale Daran-Lapujade ainsi que Philippe Gaudu qui ont accepté d'être rapporteurs et ont agréablement animé la discussion lors de la soutenance. J'adresse des remerciements particuliers à Emmanuelle Maguin pour m'avoir fait l'honneur de sa participation au jury de thèse. Merci également à Philippe Besse pour avoir apporté une touche de maths dans un monde de bio lors de ce jury !

Je tiens également à adresser mes sincères remerciements à l'ensemble des personnes avec qui j'ai eu l'occasion de collaborer au cours de ces travaux que ce soit avec Christophe Gitton et Véronique Monnet de l'INRA de Jouy-en-Josas, avec Béatrice Laurent, Hélène Milhem et Philippe Besse du Laboratoire de Statistiques et de Probabilités ou encore avec Isabelle Queinnec du LAAS de Toulouse. Cette expérience de collaboration pluridisciplinaire fut pour moi des plus enrichissantes et m'a permis d'avoir une vision plus critique de mes travaux en particulier et de la biologie en général.

J'adresse aussi un grand merci à mes précieux relecteurs et correcteurs du manuscrit pour le temps qu'ils ont bien voulu me consacrer et grâce à qui de nombreuses « coquilles » ont pu être évitées : Laurence Girbal et Sébastien Nouaille.

Au cours de ma thèse j'ai eu l'occasion de participer à un raid sportif et, étrangement, j'ai trouvé qu'il y avait pas mal de liens entre ces deux épreuves : au départ c'est juste un rêve que l'on voudrait réaliser par passion mais aussi par défi ; puis, quand le véritable challenge commence, la mise en route est un peu désorganisée, on est très motivé mais on ne sait pas vraiment dans quelle direction aller ni comment gérer son effort ; au fur et à mesure que le temps défile on apprend à s'organiser et à se gérer ; bien sûr il y a des hauts et des bas, des moments où l'on se demande pourquoi on a voulu faire ça, qu'est ce qui nous est passé par la tête quand on a décidé de s'engager, où on se dit que c'est dur, très dur et où on redoute de ne pas arriver au bout ; mais les coéquipiers sont là pour vous donner de l'aide, du réconfort et vous encourager alors on n'abandonne pas ; et enfin, voilà on arrive au bout, un dernier coup d'adrénaline pour ne pas faiblir avant d'avoir franchit la ligne d'arrivée, l'équipe est soudée, victorieuse quelle que soit sa place, la joie est présente et le soulagement est intense ; quelques temps plus tard la pression retombe mais la fierté demeure, les souvenirs perdurent et le goût du challenge donne envie de repartir pour une nouvelle aventure,

différente mais aussi motivante... Merci donc à Pascal Loubière pour son accueil au sein de l'équipe du Génie du Métabolisme des Procaryotes ainsi qu'à tous les membres (passés ou présents, dans le désordre), Emma, qui m'a transmis le relai auprès de Muriel CB, Myriam, Valérie qui a ressuscité l'Amino-Quant pour que je puisse terminer mes dosages plus souvent que je ne l'aurai voulu, Marie-Pierre, Flora, Chonticha, Puntip et Panadda, mes thais préférées dont les petits présents m'ont parfois redonné le sourire, Marianne, Lam, Delphine, Laetitia, Sophie, Sébastien, qui est arrivé au labo le même jour que moi et, non, ce n'est pas un thésard, mais un jeune chercheur qui m'a accompagnée et épaulée tout au long de cette épreuve, Julien, Mathieu, Laurence, Muriel MB dont le soutien et la bonne humeur communicative ou tout simplement la compagnie m'ont apporté bien plus qu'ils ne le pensent.

Ces remerciements seraient incomplets si j'oubliais les encouragements des « spectateurs », ma famille que je remercie toute entière. Je ne peux pas citer tout le monde mais j'ai une pensée plus particulière pour Tonton Guy qui a été mon pilier familial à Toulouse depuis le début, Mamie Simone dont la fierté pour ses petits enfants n'a d'égal que le stress qu'elle a ressenti lors de la soutenance ; Mamie Jackie qui en plus de quelques fautes d'orthographe dans le manuscrit a découvert que les bactéries pouvaient être « stressées » ; Papa qui, bien que n'ayant jamais vraiment compris où tout ça allait me mener (mais quelqu'un le sait-il ?), m'a toujours soutenue, tout en m'encourageant à toujours garder un œil tourné vers l'avenir ; Maman qui, de l'avis de tous, a passé avec succès son examen de préparation de pot de thèse !

Je remercie également Christel, passée par là quelques quatre ans auparavant, quasiment jour pour jour, ses conseils ont toujours été des plus précieux et nos repas l'occasion de faire un break tout comme de renforcer notre amitié.

Un grand merci aussi à Rémi qui a eu la malchance de rentrer dans ma vie en fin de thèse ! Comme chacun le sait ce n'est pas la période la plus agréable, surtout pour les conjoints, mais il a non seulement su s'en accommoder avec patience et se montrer solidaire, attentif et encourageant à la fois.

Enfin, mon dernier et plus grand merci s'adresse à Muriel Cocaign-Bousquet, ma directrice de thèse, l'instigatrice de cet ambitieux projet qui perdure depuis déjà trois générations de thésards et devrait (c'est tout le mal que je souhaite) se poursuivre encore pour de nombreuses années. Sans elle je n'en serai pas là aujourd'hui et même si parfois elle est encore plus enthousiaste ou pessimiste que moi, ses encouragements pendant la thèse comme à l'heure actuelle sont toujours d'un précieux réconfort ! Depuis le début, sans le savoir, elle a été mon mentor, mon modèle, un exemple de réussite professionnelle, sociale et familiale. Travailler avec elle fut un réel plaisir et j'ose espérer que, quoi qu'il advienne, ça se reproduira dans l'avenir.

Merci à tous

Sommaire

| | |
|---|-----------|
| Introduction | 1 |
| Etude bibliographique | 7 |
| I. Comment étudier un stress ? | 9 |
| I.1. Les principaux stress étudiés | 9 |
| I.2. Modes de culture | 10 |
| I.3. Outils d'analyse des stress | 13 |
| II. Biologie intégrative et mécanisme de traduction | 20 |
| II.1. Qu'est-ce que la biologie intégrative ? | 20 |
| II.2. Comparaison des données de transcriptome et de protéome | 24 |
| II.3. Régulations des niveaux de transcrits et de protéines | 29 |
| II.4. Modélisation de la traduction | 35 |
| III. <i>Lactococcus lactis</i> et ses stress | 39 |
| III.1. Intérêt de <i>Lactococcus lactis</i> et de la souche IL1403 | 39 |
| III.2. Les réponses au stress de <i>Lactococcus lactis</i> caractérisées à l'heure actuelle | 43 |
| IV. Positionnement, objectifs et démarche de la thèse | 50 |
| Matériel et Méthodes | 53 |
| I. Souche et conditions de culture | 55 |
| I.1. Souche | 55 |
| I.2. Suivi des cultures | 55 |
| I.3. Culture discontinue (batch) | 55 |
| I.4. Culture continue (chémostat) | 57 |
| II. Transcriptome | 57 |
| II.1. Description du support | 57 |
| II.2. Extraction et quantification de l'ARN total | 58 |
| II.3. Rétro-transcription, marquage, hybridation et détection | 58 |
| II.4. Normalisation des données | 59 |
| III. Protéome | 59 |
| III.1. Prélèvements | 59 |
| III.2. Extraction des protéines | 60 |
| III.3. Séparation sur gel d'électrophorèse | 60 |
| III.4. Détection et quantification des spots | 61 |
| III.5. Identification des protéines | 61 |
| III.6. Traitement des données | 62 |
| IV. Mesure de la stabilité des ARN messagers | 63 |

| | |
|--|------------|
| V. Recherche de motifs | 64 |
| VI. Traitement statistique des données | 64 |
| VI.1. Exploration des données | 66 |
| VI.2. Analyse statistique des données | 67 |
| VI.3. Modélisation | 67 |
| <i>Chapitre I : Réponse de <i>Lactococcus lactis</i> à la variation du taux de croissance et implication dans ses réponses au stress</i> | 69 |
| <i>Chapitre II : Adaptation dynamique de <i>Lactococcus lactis</i> à la carence en isoleucine</i> | 105 |
| <i>Chapitre III : Modélisation des efficacités de traduction et de la stabilité des protéines de <i>Lactococcus lactis</i></i> | 155 |
| <i>Chapitre IV : Identification des déterminants des concentrations protéiques de <i>Lactococcus lactis</i> par modèles de covariance</i> | 187 |
| <i>Conclusion et perspectives</i> | 219 |
| DEMARCHE ET PRINCIPAUX RESULTATS | 221 |
| Impact de la diminution du taux de croissance sur la physiologie cellulaire | 222 |
| Caractérisation de la réponse de <i>L. lactis</i> à la carence en isoleucine | 223 |
| Contrôle des concentrations intracellulaires en protéines chez <i>L. lactis</i> | 225 |
| QUESTIONNEMENTS ET PERSPECTIVES | 228 |
| Existe-t-il un mécanisme général de réponse au stress chez <i>L. lactis</i> ? | 228 |
| Quels sont les facteurs prépondérants dans l'adaptation de <i>L. lactis</i> ? | 230 |
| Le contrôle de l'adaptation peut-il s'effectuer à d'autres niveaux ? | 232 |
| Comment caractériser au mieux l'adaptation ? | 234 |
| <i>Références bibliographiques</i> | 235 |
| <i>Publications et communications</i> | 255 |
| <i>Annexes</i> | 259 |

Introduction

Couramment employé dans le langage quotidien, le terme de stress n'est pourtant pas clairement défini et relève plus du concept que du phénomène tangible. Du point de vue de l'Homme, le mot stress est utilisé à tout va pour définir une sensation, un ressenti par rapport à une agression extérieure qu'elle soit de nature physique ou, plus souvent, émotionnelle. Bien que défini de manière intuitive, le stress est de plus en plus considéré comme le mal du siècle pour l'Homme car il peut entraîner des manifestations physiologiques et psychologiques. En ce sens, le terme de stress est généralement associé à la notion d'adaptation car la perturbation de l'équilibre nécessite le rétablissement de l'homéostasie. Pour les microorganismes, auxquels nous nous intéressons dans cette étude, la notion de stress est beaucoup moins connue du grand public mais n'est finalement pas très éloignée. Elle reprend en effet l'idée d'adaptation en réponse à une agression environnementale. Cette dernière est liée à la présence de conditions de croissance non optimales, rencontrées aussi bien au sein d'un écosystème naturel que lors d'une exploitation industrielle, et peut être de nature nutritionnelle ou physico-chimique. Comprendre les effets engendrés par un stress et élucider les différents mécanismes de régulations mis en jeu est un défi scientifique de taille aussi bien pour son aspect cognitif que dans l'optique de réduire et/ou de maîtriser l'impact d'un stress sur un microorganisme d'intérêt.

Pour comprendre le fonctionnement biologique des microorganismes, les outils de caractérisation de la physiologie cellulaire sont nombreux. L'entrée de la microbiologie dans l'ère de la post-génomique a d'ailleurs contribué au développement de techniques de mesure à l'échelle globale telles que le transcriptome et le protéome. Chaque technologie disponible permet de mesurer un niveau de régulation différent dont la réconciliation avec les autres est nécessaire pour obtenir une vision globale, cohérente et intégrée de la cellule. La biologie intégrative est une démarche scientifique qui consiste à mettre en œuvre des outils mathématiques afin d'intégrer ces différents niveaux d'analyse. Elle se présente donc comme une méthode de choix pour la compréhension globale de la physiologie des microorganismes, en particulier lors de leur réponse au stress.

L'objectif de cette thèse est de mieux comprendre l'adaptation de *Lactococcus lactis* en condition de stress par une démarche de biologie intégrative. *L. lactis*, le modèle des bactéries lactiques, a été choisi comme microorganisme d'étude en raison de la simplicité relative de son métabolisme et de la petite taille de son génome. Plusieurs techniques de mesure à

l'échelle du génome (transcriptome, protéome, stabilité des ARN messagers) ont été employées afin d'appréhender sa physiologie d'adaptation de manière intégrée et de caractériser ses réponses aux variations des conditions environnementales. Ces données globales ont été associées à l'analyse cinétique des paramètres de fermentation. Des outils mathématiques ont été développés pour analyser de façon intégrée l'ensemble des données mesurées. L'adaptation a été abordée au travers de deux types de fermentations maîtrisées et rigoureusement contrôlées. Une culture discontinue en réacteur de type batch a permis d'étudier la réponse dynamique de *L. lactis* à l'épuisement progressif de l'isoleucine initialement présente dans le milieu de culture. Cette carence nutritionnelle n'a encore jamais été caractérisée pour ce microorganisme dont les auxotrophies sont pourtant multiples. L'imposition d'un stress s'accompagne systématiquement d'une réduction du taux de croissance. Néanmoins, les études d'adaptation en mode dynamique ne dissocient jamais l'effet du stress de celui de la chute du taux de croissance. La culture continue en chémostat permet un découplage et un contrôle des différents paramètres de culture. Cet outil permet notamment d'étudier l'influence du taux de croissance sans modifier l'intensité du stress imposé par la limitation. Afin de visualiser spécifiquement la réponse à la carence en isoleucine de *L. lactis*, nous avons choisi de caractériser la réponse engendrée par la diminution du taux de croissance dans un chémostat aussi limité par l'isoleucine. Ainsi, la part de la réponse de *L. lactis* en carence en acide aminé liée à la baisse du taux de croissance a pu être déterminée grâce à la comparaison des deux réponses transcriptomiques. Celle spécifiquement dépendante de la carence isoleucine a donc pu en être déduite par soustraction de l'effet du taux de croissance à la réponse dynamique observée en batch.

Au cours de cette thèse, nous nous sommes intéressés plus particulièrement à la régulation des niveaux de protéines au cours de l'adaptation. En tant qu'effecteurs de la cellule, ces biomolécules jouent un rôle biologique central. Trois processus cellulaires interviennent dans le contrôle des niveaux de protéines : la traduction des ARN messagers, la dilution par la croissance et la dégradation des protéines. L'efficacité de traduction et la vitesse de dégradation des protéines sont des paramètres difficilement mesurables à l'échelle du génome et encore jamais évalués pour une bactérie. Leur estimation a néanmoins été possible pour des états d'équilibre grâce à une approche mathématique couplant les équations décrivant les niveaux de protéines à une modélisation des résultats de mesure. Nous avons d'autre part recherché les déterminants biologiques majeurs des niveaux de protéines par une approche de modélisation statistique, basée sur la régression linéaire multiple, dans les deux modes de culture, continu et discontinu.

Dans ce manuscrit, les résultats sont présentés en anglais sous forme de publications avec un résumé étendu en français proposé en préambule de chaque article. L'introduction bibliographique exposée dans la première partie de ce manuscrit vise à resituer le sujet dans son contexte. Elle récapitule d'abord les différentes méthodes d'analyse d'un stress puis présente la notion de biologie intégrative avant de faire un point sur *Lactococcus lactis* et les différents stress auxquels il peut être confronté. Enfin, les objectifs de cette thèse seront présentés en établissant le lien entre les différents points bibliographiques abordés. Les deux premiers chapitres de résultats s'attachent à la caractérisation des réponses de *L. lactis* dans les deux environnements particuliers de stress correspondant aux deux modes de culture, chemostat et batch. Ces deux fermentations correspondent à une limitation et à une carence en isoleucine et donc à des modes d'adaptation très différents. Les deux chapitres suivants sont consacrés à des approches intégratives avec l'utilisation d'outils de modélisation pour tenter de comprendre le fonctionnement physiologique de *L. lactis* et des processus de régulation des concentrations en protéines. Au final, l'ensemble des résultats obtenus sera repris dans une conclusion générale mettant en exergue de nouvelles perspectives scientifiques.

Etude bibliographique

I. Comment étudier un stress ?

La notion de stress intervient dès que les conditions optimales de croissance d'un micro-organisme ne sont pas réunies. Par conséquent, quelle que soit sa nature, toute modification environnementale conduisant à des conditions de croissance non favorables va perturber la physiologie cellulaire et sera donc perçue comme un stress. La capacité d'adaptation des microorganismes aux différentes conditions de vie qu'ils rencontrent va être déterminante pour leur survie et leur développement. Elucider les mécanismes cellulaires et les régulations permettant de résister aux stress reste depuis de nombreuses années le cheval de bataille de nombreux microbiologistes. L'objet de cette première partie est de faire le point sur les principaux stress étudiés mais surtout sur les méthodes employées pour y aboutir que ce soit en termes de mode de culture ou d'outils d'analyse. Les mécanismes mis en jeu lors des réponses aux différents stress ne seront pas abordés ici mais font l'objet de trois revues parues en 2007 pour les bactéries (Giuliodori *et al.*, 2007; Marles-Wright & Lewis, 2007; Schimel *et al.*, 2007).

I.1. Les principaux stress étudiés

Au sein de leurs écosystèmes naturels et éventuellement lors de leur exploitation industrielle, les microorganismes sont confrontés à de multiples stress liés à la disponibilité des nutriments, à la nature physico-chimique du milieu ou encore à l'interaction avec d'autres organismes. Pour les levures comme pour les bactéries, la littérature à ce sujet est particulièrement abondante et les différentes informations recueillies sont donc souvent synthétisées sous forme de revues. Les stress nutritionnels sont certainement les plus étudiés (McDougald *et al.*, 2002; Spector, 1998) notamment la carence en carbone induite la plupart du temps par une limitation en sucre (Gelade *et al.*, 2003) ou encore la carence en azote principalement simulée par une restriction de l'apport en acides aminés et plus connue chez les bactéries comme réponse stringente (Jain *et al.*, 2006). Si les chocs de température restent les stress physico-chimiques les plus étudiés (Harwood, 2007; Inouye & Phadtare, 2004; Schumann, 2003; Weber & Marahiel, 2003), les effets de divers autres paramètres tels que le pH (Hall *et al.*, 1995), les facteurs provoquant des lésions de l'ADN et conduisant à la mise en place de la réponse SOS (Shinagawa, 1996), l'osmolarité (Record *et al.*, 1998), le stress

oxydatif (Lushchak, 2001), les composés nitrés (Poole, 2005) sont également décrits pour différentes espèces bactériennes. Le recoupement de certaines données associées à différents stress (Gasch & Werner-Washburne, 2002) a également permis de proposer une réponse générale au stress, en particulier chez *Escherichia coli* (Hengge-Aronis, 2002) et *Bacillus subtilis* (Hecker *et al.*, 2007), modèles des bactéries Gram négatives et Gram positives respectivement. Bien que n'étant pas un stress à part entière, le taux de croissance et ses variations y sont indéniablement associés, ce qui rend intéressantes les études de ses effets sur la physiologie de la levure *Saccaromyces cerevisiae*, récemment regroupées afin de déterminer les mécanismes sous-jacents (Ramanathan & Schreiber, 2007).

I.2. Modes de culture

I.2.1. Culture en fiole erlenmeyer

De tout temps, la culture en milieu liquide en fiole erlenmeyer a été utilisée pour produire de la biomasse. Il reste que cette méthode de culture rudimentaire ne permet pas de contrôler les conditions de croissance qui vont être particulièrement dépendantes de l'homogénéisation du système (Irvine, 1990). Malgré le manque de maîtrise inhérent à ce type de culture, leur utilisation reste toujours très répandue puisque même de récents travaux d'analyse de réponse au stress par des techniques de pointe se basent sur des cellules cultivées en fioles erlenmeyers (Ferullo & Lovett, 2008; Melamed *et al.*, 2008; Tam le *et al.*, 2007).

I.2.2. Culture en réacteurs contrôlés

Au-delà de l'aspect industriel, l'intérêt de réaliser des fermentations en réacteur réside dans la maîtrise des conditions de culture (Irvine, 1990). En effet, les bioréacteurs sont généralement équipés de capteurs et associés à des modules de régulation permettant ainsi le contrôle des paramètres de température, agitation, pH et aération. Certains fermenteurs permettent même un suivi en ligne de différents paramètres de croissance dont la quantité de biomasse en particulier. D'autre part, un dispositif d'échantillonnage évite l'ouverture du réacteur à chaque prélèvement ce qui minimise les risques de contamination. Maîtriser les procédés fermentaires constitue un atout majeur pour la reproductibilité des phénomènes et la compréhension des réponses mises en place. Ne sont décrits ici que les deux principaux types de réacteurs biologiques à savoir le batch et le chémostat correspondant respectivement à des cultures en mode discontinu et continu.

I.2.2.1. Culture discontinue (batch)

Lors des cultures discontinues, le milieu de culture et le microorganisme à cultiver sont initialement introduits dans le réacteur mais aucun élément supplémentaire n'est rajouté en cours de fermentation. Ainsi, l'évolution de la croissance en fonction de la composition du milieu va conduire à la division de la fermentation en différentes phases : selon ses conditions de développement avant son inoculation dans le fermenteur, le microorganisme peut présenter une phase de latence caractérisée par une croissance très lente et durant laquelle il s'adapte à son nouvel environnement ; lors de la phase exponentielle de croissance tous les nutriments sont en excès et la croissance est alors optimale ; vient ensuite une phase de ralentissement de croissance liée à l'épuisement de l'un des nutriments ou à l'inhibition par l'un des produits de fermentation ; le microorganisme entre ensuite en phase stationnaire : la biomasse est maintenue mais il n'y a plus de croissance ; enfin peut apparaître une phase de lyse au cours de laquelle la concentration en biomasse décroît (Enfors, 1991). La spécificité de ce type de fermentation réside dans l'état d'adaptation dynamique dans lequel se trouve le microorganisme. Il doit en effet constamment réajuster ses processus cellulaires afin de s'adapter à la perpétuelle évolution de l'environnement de culture. De part la possibilité de suivi de la dynamique de croissance ainsi que des paramètres physiologiques qui vont être associés aux différentes phases, la culture en batch est donc couramment utilisée comme un outil pour déterminer la réponse à divers stress de type nutritionnel (Albers *et al.*, 2007; Colombie *et al.*, 2005; Gyaneshwar *et al.*, 2005b; Moritz *et al.*, 2002; Ryu *et al.*, 2007). La comparaison de deux phases exponentielles dans des conditions différentes est également envisageable afin d'étudier l'effet d'un paramètre particulier comme cela a été fait pour l'oxygène (Zhang *et al.*, 2008).

I.2.2.2. Culture continue (chémostat)

Dans les fermenteurs de type chémostat, l'apport de milieu se fait de façon continue. Les débits d'alimentation et de soutirage étant égaux Q (L/h), le volume V (L) du fermenteur est maintenu constant. L'équation d'accumulation de la biomasse au sein du réacteur est donc la suivante : $dX / dT = X (\mu - D)$ où μ (h^{-1}) est le taux de croissance et $D = Q/V$ le taux de dilution (h^{-1}). Lorsque le régime permanent est atteint, les concentrations en substrats (S , g/L), produits (P , g/L) et biomasse (X , g/L) sont constantes dans le chémostat. Dans ce type de fermenteur, le taux de croissance est donc fixé par le taux de dilution. Quant à la concentration en biomasse, X , elle est déterminée par l'un des constituants du milieu (facteur

limitant), les autres constituants étant en excès (Enfors, 1991). Contrairement à la fermentation en batch, il n'est plus question d'adaptation dynamique mais plutôt d'état d'équilibre avec des cellules adaptées à des conditions environnementales constantes. Le maintien de la biomasse dans un état stable garantit non seulement la reproductibilité des expériences, mais permet également d'étudier l'effet d'un paramètre spécifique. En effet, lors des études de réponse au stress en culture discontinue, plusieurs effets peuvent se superposer (Hoskisson & Hobbs, 2005). Lors d'un stress nutritionnel par exemple, les effets carence et croissance sont totalement indissociables dans un réacteur de type batch. *A contrario*, la culture en chémostat permet le découplage des différents paramètres et leurs études séparées ce qui la rend particulièrement intéressante par rapport au batch (Hoskisson & Hobbs, 2005) : les stress nutritionnels vont pouvoir être étudiés au travers du choix de l'élément limitant (Bacon *et al.*, 2007; Dauner *et al.*, 2001; Haydock *et al.*, 2004; Johansson & Liden, 2006) ; la modification des paramètres physico-chimiques va permettre d'identifier les réponses qui y sont associées (Tai *et al.*, 2007; Zhang *et al.*, 2008) ; enfin, les effets du taux de croissance vont être déterminés grâce à des modifications du taux de dilution (Castrillo *et al.*, 2007; Regenber *et al.*, 2006). Il est cependant important de noter que l'étude d'un stress nutritionnel se fait en condition de limitation et non de carence comme c'est le cas lors d'une fermentation en batch. En effet, si dans ces deux types de culture l'élément nutritif d'intérêt est absent du milieu réactionnel, il reste consommé par les cellules adaptées dans le réacteur continu alors qu'il est épuisé dans le réacteur discontinu où les cellules s'adaptent progressivement.

L'outil chémostat, et surtout l'état quasi-stationnaire des cultures qui y sont réalisées, ont récemment été remis en cause par Ferenci. Ce microbiologiste considère en effet que les divers stress qui vont être appliqués au microorganisme en culture continue vont induire des fonctions favorisant la variabilité génétique (Ferenci, 2008). En ce sens, la pression de sélection exercée par l'imposition du taux de croissance va favoriser l'évolution génétique de la biomasse pour une meilleure adaptation. Toutefois, l'hétérogénéité de la population au sein du réacteur ne devient significative qu'à partir d'une centaine de générations et une fois l'équilibre atteint (5 générations sont généralement attendues mais les paramètres macro-cinétiques se stabilisent généralement avant) la marge de manœuvre pour l'étude d'un stress particulier reste suffisamment importante en terme d'échelle de temps. A partir du moment où la culture n'est pas indéfiniment prolongée, il n'est donc pas déraisonnable de considérer que les cellules sont dans un état constant.

I.3. Outils d'analyse des stress

Chez les microorganismes, l'effet d'un stress peut affecter différents niveaux de la physiologie cellulaire et se mesure donc par différentes méthodes.

I.3.1. Outils physiologiques

La première étape vers la compréhension des effets d'un stress consiste à observer le comportement du microorganisme qui le subit, ce qui requiert le suivi de divers paramètres macromoléculaires.

I.3.1.1. Viabilité

La survie cellulaire est généralement le premier des paramètres mesurés suite à l'imposition d'un stress. La mesure de l'aptitude d'une cellule à se multiplier par le comptage des unités formant des colonies sur boîte de Pétri est la technique la plus simple et la plus largement répandue pour évaluer la viabilité aujourd'hui encore (Sikic Pogacar *et al.*, 2009). D'autres méthodes, rendant compte de différents états physiologiques tels que la capacité métabolique (test d'acidification, synthèse d'acide nucléique, synthèse d'ATP, respiration...) ou encore l'intégrité membranaire (méthodes colorimétriques) peuvent également servir à mesurer la viabilité cellulaire (Oliver, 1993). Avec la culture en bioréacteurs, certaines de ces techniques, et en particulier les méthodes photométriques, ont été adaptées pour la mesure en ligne de la quantité de biomasse et éventuellement de sa viabilité (Madrid & Felice, 2005).

I.3.1.2. Constituants cellulaires

Les différents stress que peut rencontrer un microorganisme peuvent altérer sa composition en macromolécules. Là encore, plusieurs techniques sont à disposition pour mesurer les différents constituants cellulaires. Après extraction appropriée, souvent en utilisant le phénol, les ADN et ARN totaux, éventuellement purifiés à l'aide de kits commerciaux, sont quantifiés par mesure de l'absorbance à 260 nm tandis que leur contamination protéique peut être évaluée par le rapport entre l'absorbance à 260 nm et celle à 280 nm. Les stress peuvent modifier la topologie de l'ADN. L'état de sur-enroulement connu pour être sensible aux stress thermiques (Lopez-Garcia & Forterre, 2000), pourra par exemple être déterminé par des techniques faisant appel à l'électrophorèse (Keller, 1975). Plusieurs protocoles peuvent être appliqués pour le dosage des protéines totales et si la méthode de Bradford est couramment employée en raison de sa simplicité, il ne faut pas négliger que sa sensibilité varie en fonction du pH auquel s'effectue le dosage (Lovrien & Matulis, 2005).

Les stress n'épargnent pas non plus l'enveloppe cellulaire, entraînant des modifications de la membrane cellulaire (Mykytczuk *et al.*, 2007). Ces variations de la composition lipidique (lipides, acides gras, taux d'insaturation) vont affecter la fluidité membranaire. Plusieurs méthodes sont disponibles pour mesurer cette fluidité mais depuis de nombreuses années maintenant, c'est la polarisation de fluorescence qui prévaut. Cette technique repose sur l'utilisation de lipides fluorescents comme sonde (Denich *et al.*, 2003). L'analyse de la perméabilité membranaire à différentes sondes fluorescentes a par exemple permis de mettre en évidence chez *Geobacter sulfurreducens* à la fois une modification de la structure de la membrane et une perte de viabilité en réponse au stress hyper-osmotique (Ragoonanan *et al.*, 2008).

I.3.1.3. Métabolisme et énergétique cellulaire

En altérant leur croissance, la plupart des stress modifient également le comportement métabolique des cellules. L'utilisation de techniques de dosage chromatographique (gaz ou liquide) est un excellent moyen de mettre en évidence l'accumulation de produits toxiques mais surtout de révéler quels substrats sont consommés et à la formation de quels produits ils aboutissent (Fox *et al.*, 1990). A partir de ces données, la connaissance des réseaux métaboliques permet de repérer quelles voies sont spécifiquement utilisées. Des dosages en chromatographie liquide et gazeuse ont par exemple permis de repérer une modification du catabolisme des sucres pour les souches de *Bifidobacterium animalis* résistantes aux sels biliaires (Ruas-Madiedo *et al.*, 2005). Le dosage des activités de certaines enzymes clés peut également permettre de confirmer la nature du métabolisme mis en place. Les mesures d'activité enzymatique se font la plupart du temps grâce au suivi spectrophotométrique en continu de la disparition d'un substrat, d'un cofacteur ou de la formation d'un produit. Lorsqu'un suivi direct de la réaction n'est pas possible, des prélèvements réguliers sont effectués et couplés à d'autres méthodes de détection impliquant parfois une réaction enzymatique supplémentaire (Dobrogosz, 1981).

L'évaluation de l'état énergétique de la cellule repose également sur la mesure de son activité ATPase ainsi que de la force protomotrice. L'activité ATPase est évaluée par la mesure de la vitesse d'hydrolyse de l'ATP en ADP + Pi qui peut se faire à l'aide de kits commerciaux, tandis que la détermination de la force protomotrice requiert à la fois la mesure du gradient de protons (ΔpH) et du gradient électrique ($\Delta\Psi$) (Mitchell & Moyle, 1967). Les méthodes de détermination de ces paramètres impliquent l'établissement d'un équilibre entre l'intérieur et l'extérieur de la membrane qui pourra être suivi grâce à des composés marqués,

souvent de manière radioactive, parfois fluorescente. Pour le ΔpH , des acides ou des bases faibles sont souvent utilisés tandis que pour le $\Delta\Psi$, l'équilibration des ions de part et d'autre de la membrane est directement suivie (Kaback, 1990).

I.3.2. Les outils de mesure à l'échelle globale

Le séquençage des génomes a beaucoup apporté à la biologie et a notamment permis le développement d'outils de mesure globale, ou à haut débit, regroupés sous l'appellation de technologies post-génomiques (ou parfois génomique fonctionnelle). Depuis leur avènement il y a une vingtaine d'années, et bien qu'apportant des informations différentes, ces techniques de mesure ont supplanté la plupart des outils de mesure de la physiologie générale des cellules dans l'étude des réponses au stress (Schweder & Hecker, 2004) et ont permis l'accumulation de données à l'échelle du génome ou données en « omes » (figure 1).

I.3.2.1. Transcriptome

Introduit par Velculescu *et al.* en 1997, le terme de transcriptome se définit comme la mesure de l'expression des gènes et se présente comme un maillon essentiel entre le génome d'un organisme et sa physiologie (Velculescu *et al.*, 1997). A l'époque de cette définition, le transcriptome avait déjà été mesuré par deux méthodes différentes : la technique SAGE, pour Serial Analysis of Gene Expression, qui repose sur la quantification de courtes séquences d'ADN complémentaire (ADNc), spécifiques d'un gène et obtenues lors de la rétro-transcription des messagers (Velculescu *et al.*, 1995), et une technique d'hybridation sur membrane baptisée puce à ADN et reposant sur l'hybridation d'ADNc marqués par des fluorochromes avec des fragments d'ADN fixés sur une lame de verre (Schena *et al.*, 1995). Depuis 1995, le transcriptome n'a eu de cesse d'être mesuré et diverses méthodologies, souvent dérivées des premières approches, ont été et sont encore développées voire commercialisées (Heller, 2002; Peeters & Van der Spek, 2005). Dans l'ensemble, les procédés les plus couramment utilisés sont de type puce à ADN. Ces puces sont composées de sondes gène-spécifiques fixées sur un support solide et avec lesquelles des ADNc (cibles) marqués vont venir s'hybrider. Les deux techniques les plus couramment utilisées se différencient par la nature des sondes : produits de PCR directement déposés (ou spottés) sur le support solide ou oligonucléotides souvent synthétisés *in situ* (Katagiri & Glazebrook, 2009; Mandruzzato, 2007). Le marquage des ADNc peut également se faire de différentes manières, les plus courantes étant l'incorporation d'une base marquée par un élément radioactif ou par un fluorochrome lors de la rétro-transcription (Do & Choi, 2006).

Bien qu'utilisées en routine pour de multiples applications (Wheelan *et al.*, 2008), les puces à ADN sont fréquemment employées pour l'étude de réponses au stress (Gracey & Cossins, 2003; Jaluria *et al.*, 2007; Peeters & Van der Spek, 2005). La mesure du transcriptome est d'ailleurs le principal outil d'analyse des réponses au stress, en témoignent les 3666 références proposées à ce jour (1 avril 2009) par PubMed lors de la recherche par mots clés avec « transcriptome and stress ».

Si la technologie de type puce à ADN reste la plus utilisée, de nouvelles techniques d'analyse du transcriptome ont vu le jour au cours des dernières années. Basées sur le séquençage à haut débit (Wold & Myers, 2008), ces méthodes présentent plusieurs avantages : tout d'abord elles ne requièrent pas la connaissance préalable du génome auquel elles sont appliquées et permettent de détecter l'ensemble des séquences codantes, y compris lorsque les messagers sont peu abondants ; d'autre part, elles sont également quantitatives et très reproductibles (R_p de 0.99 pour la répétition d'une expérience donnée) (Cloonan & Grimmond, 2008). Ces études transcriptomiques de nouvelle génération nécessitent toutefois encore des quantités d'ARN total comprises entre 20 μ g et 1 mg (Cloonan & Grimmond, 2008) soit environ de 2 à 10 fois plus que la quantité utilisée sur les puces à ADN. Enfin, l'étape de rétro-transcription aléatoire pose un problème chez les bactéries car, contrairement aux eucaryotes, les ARN messagers ne sont pas polyadénylés et ne peuvent donc pas être spécifiquement ciblés. L'élimination préalable des ARN ribosomiaux est donc nécessaire afin d'enrichir la fraction étudiée en ARN messagers et de minimiser la conversion des ARNr en ADNc (Mao *et al.*, 2008).

I.3.2.2. Protéome

Défini comme le complément protéique du génome (Wasinger *et al.*, 1995), le protéome a donné naissance à la protéomique qui, par analogie, correspond à de la génomique fonctionnelle mais au niveau des protéines (Blackstock & Weir, 1999). Si l'analyse du protéome s'est longtemps faite et se fait encore grâce à la séparation des protéines par électrophorèse en deux dimensions (2D) (Kalia & Gupta, 2005; Yarmush & Jayaraman, 2002), de nombreuses améliorations ont été apportées en terme de marquage (Bachi & Bonaldi, 2008), d'identification des protéines (Kalia & Gupta, 2005) et d'analyse d'image (Berth *et al.*, 2007) afin de rendre cette technique quantitative.

Ces dernières années, plusieurs approches dites « gel-free » permettant d'éviter la séparation sur gel d'électrophorèse ont également été mises au point et se posent comme des alternatives à l'électrophorèse bidimensionnelle classique (Dhingra *et al.*, 2005; Lambert *et*

al., 2005; Roe & Griffin, 2006). L'ensemble de ces nouvelles techniques, répertoriées dans une revue de Gevaert *et al.* (2007), repose sur la mesure, non plus des protéines directement, mais des peptides qui les composent. La méthodologie la plus couramment utilisée propose une digestion protéolytique des échantillons protéiques à doser suivie d'un couplage entre la séparation par chromatographie liquide avec la spectrométrie de masse en tandem ou LC-MS/MS encore appelée protéomique shotgun (Canas *et al.*, 2006; Pisitkun *et al.*, 2007). Lu *et al.* (2007) ont même proposé une amélioration des résultats issus de la protéomique shotgun *via* la prise en compte des propriétés physicochimiques des peptides. Cette technique de normalisation des résultats, baptisée APEX pour Absolute Protein Expression, vise à améliorer la quantification des données. Un logiciel de traitement des données par cette technique a d'ailleurs été récemment développé (Braisted *et al.*, 2008).

Les nouvelles techniques sont généralement décrites comme plus performantes que la protéomique à deux dimensions car elles permettent de mesurer un plus grand nombre de protéines et sont donc plus exhaustives ; par exemple, l'étude du protéome de *S. cerevisiae* en condition de limitation en carbone ou en azote par la technique shotgun a permis de quantifier 759 protéines mais les auteurs estiment qu'une analyse 2D classique en aurait manqué 29 % car elles sont en dehors des gammes de poids moléculaires et de points isoélectriques classiques (Kolkman *et al.*, 2006). Similairement, chez la bactérie *Bacillus subtilis*, la protéomique « gel free » a également permis d'identifier 473 protéines supplémentaires (ajoutées aux 745 identifiées sur gel 2D) (Wolff *et al.*, 2006). Toutefois, il est important de relever que deux types de préparation des protéines pour les analyses « gel free » ont été pratiquées, afin de collecter d'une part les protéines cytoplasmiques et de l'autre les protéines membranaires. Employer différents processus d'extraction des protéines selon la fraction à étudier n'est pas exclusivement réservé aux nouvelles approches et peut être étendu à la protéomique 2D même si les tentatives de caractérisation du protéome membranaire par une analyse 2D restent encore incomplètes à l'heure actuelle (Poetsch & Wolters, 2008). Dans le même ordre d'idée, il n'est pas exclu de « dilater » les gels d'électrophorèse en multipliant le nombre de gammes de pH utilisées tout en diminuant la taille de chacune afin d'obtenir une meilleure résolution. De façon intéressante, l'étude menée chez *B. subtilis* a aussi montré que 193 protéines ont été spécifiquement identifiées par l'électrophorèse bidimensionnelle (Wolff *et al.*, 2006), montrant ainsi que les différentes approches peuvent se compléter. D'autre part, les analyses issues des données de surexpression se recoupent quelle que soit la technique utilisée révélant ainsi la cohérence de chacune. De plus, malgré le fait que les techniques sur gel soient plus restrictives en terme de nombre de protéines identifiées et

surtout quantifiées, elles s'avèrent être de bons outils pour la détermination des réponses au stress, du moins pour les bactéries gram positives (Hecker *et al.*, 2008). Il reste que les nouvelles techniques, comme les anciennes, dépendent encore fortement de la préparation des échantillons ainsi que des performances des instruments techniques auxquels elles font appel (Garbis *et al.*, 2005).

Enfin, l'idée d'utiliser des puces à protéines pour une mesure exhaustive du protéome comme alternative aux techniques actuelles n'est pas nouvelle mais sa réalisation constitue un énorme challenge technique (Freire & Wheeler, 2006). Les progrès en microfluidique permettent déjà de réaliser des études protéomiques sur puce (Franke & Wixforth, 2008) mais le volume de l'échantillon à analyser est faible, toutes les protéines ne sont pas encore détectées et les applications restent pour l'instant confinées au domaine médical (Fan *et al.*, 2008).

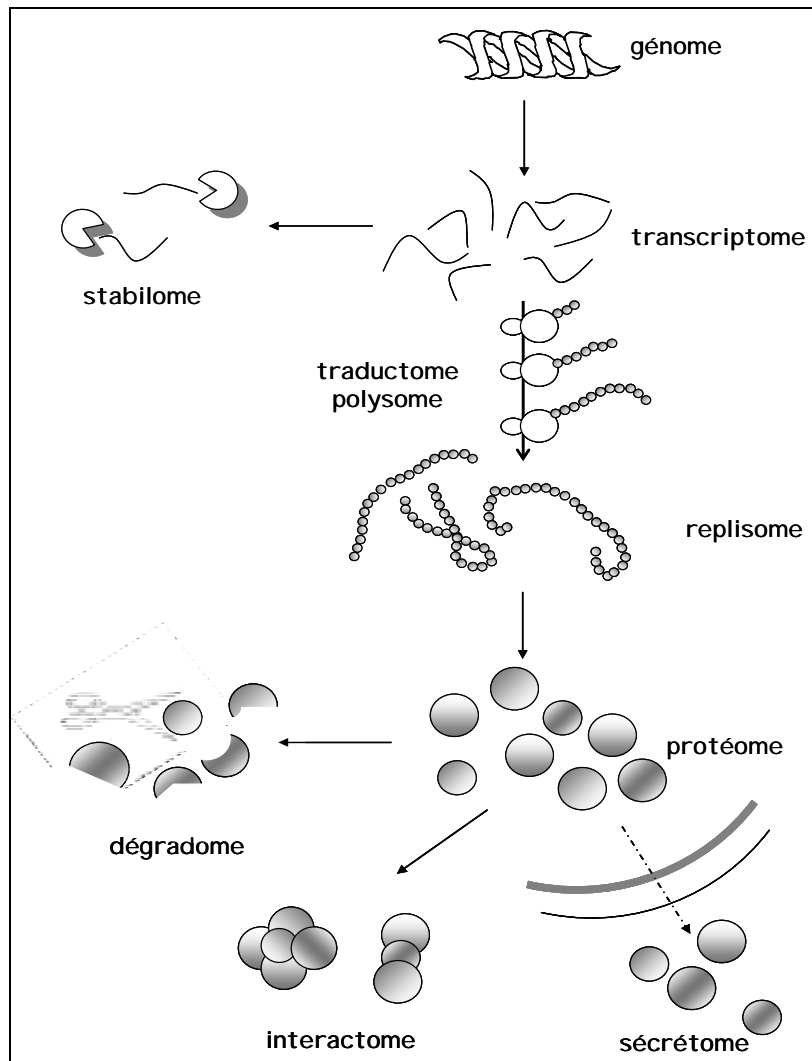


Figure 1 : Une vision en « ome » de la cellule.

I.3.2.3. Métabolome

Par extension des termes en « ome », celui de métabolome a été avancé pour définir l'ensemble des métabolites présents dans une cellule et leur quantification (Oliver *et al.*, 1998). Autrefois dosés de manière individuelle par des réactions enzymatiques, les métabolites peuvent être de nature très variée. De fait, la diversité de leurs propriétés physiques et chimiques fait qu'il n'existe pas une technique de choix pour leur mesure contrairement aux transcrits ou aux protéines (Dunn *et al.*, 2005). La plupart des méthodes reposent cependant sur le couplage entre une technique chromatographique pour la séparation (chromatographie gazeuse, liquide, électrophorèse capillaire) et de la spectrométrie de masse pour l'identification (Dettmer *et al.*, 2007; Garcia *et al.*, 2008). Bien que moins sensible, la spectroscopie par résonance magnétique nucléaire (RMN) est également largement employée pour les analyses métabolomiques car elle présente l'avantage d'être très reproductible et spécifique (Dunn *et al.*, 2005; Pan & Raftery, 2007). Comme la plupart des techniques issues de la génomique, la mesure du métabolome s'applique à de nombreux domaines de recherche allant de la détermination de la réponse à un stimulus au diagnostic médical (Gomase *et al.*, 2008). En ce qui concerne l'apport lors de l'étude de stress, la métabolomique a par exemple permis de comprendre l'aspect métabolique de l'adaptation au tellurite chez *Pseudomonas pseudoalcaligenes* (Tremaroli *et al.*, 2009), adaptation qui, jusqu'alors, n'était que partiellement expliquée par l'induction de la réponse au stress oxydatif (Tremaroli *et al.*, 2007).

I.3.2.4. Autres outils

Dans la continuité de la subdivision du contenu cellulaire en « ome », Greenbaum *et al.* ont proposé la définition de nombreux autres termes correspondant à différents niveaux d'information biologique (figure 1, (Greenbaum *et al.*, 2001)). Par exemple, l'interactome représenterait la mesure de l'ensemble des interactions cellulaires tandis que le sécrétome quantifierait les molécules sécrétées. Ces sous-populations sont encore loin d'être toutes quantifiables en routine mais certaines mesures sont d'ores et déjà réalisables grâce à l'utilisation de méthodes de type microarray (Mata *et al.*, 2005). Ainsi, il est possible de mesurer la stabilité des messagers (Bernstein *et al.*, 2002; Grigull *et al.*, 2004), d'estimer leur état de traduction grâce aux profils polysomiques (Beilharz & Preiss, 2004; Melamed & Arava, 2007), d'identifier les cibles des protéines de liaison à l'ARN (RBP pour Ribosome Binding Protein) (Sanchez-Diaz & Penalva, 2006; Tenenbaum *et al.*, 2000). La technologie

Phenotype MicroArray développée par les laboratoires Biolog est également remarquable car elle permet de réaliser une caractérisation phénotypique d'une bactérie sur une microplaque 96 puits. Encore à ses débuts, cette technique reste principalement utilisée dans des conditions de référence mais présente un véritable potentiel pour l'étude et la caractérisation de réponses au stress (Bochner, 2009).

II. Biologie intégrative et mécanisme de traduction

Devant l'accumulation de données en « omes » générées par le développement de ces technologies de mesure à l'échelle du génome, il devient intéressant sinon primordial de réconcilier les informations issues des différents niveaux d'analyse biologique afin d'avancer dans la compréhension globale des fonctionnements biologiques, lors de la réponse à un stress par exemple. Chaque donnée en « omes » permet en effet de décrire de manière quasi-exhaustive un phénomène moléculaire mais c'est l'ensemble de ces niveaux de régulation qui participe à l'expression d'un phénotype particulier.

II.1. Qu'est-ce que la biologie intégrative ?

II.1.1. Origine et nature

L'intégration de données post-génomiques pour déchiffrer les systèmes biologiques dans leur ensemble (Joyce & Palsson, 2006) a donné naissance au début du siècle à la notion de biologie des systèmes (Ideker *et al.*, 2001; Kitano, 2002b) plus souvent nommée biologie intégrative par les chercheurs francophones. Ce terme ne désigne pas une nouvelle discipline scientifique à part entière mais plutôt une démarche pluridisciplinaire dans le sens où il englobe l'utilisation d'outils informatiques et de modèles mathématiques pour l'exploitation de données expérimentales issues de différents niveaux biologiques (Kitano, 2002a; Williamson, 2005). D'autre part, l'obtention de ces données repose sur des développements techniques, aussi bien au niveau des nanotechnologies, impliquant les domaines de la biologie mais également de la chimie et de la physique (Cheung *et al.*, 1999; Ivanov *et al.*, 2006), qu'au niveau des outils statistiques pour le traitement des données de grande dimension (Chich *et al.*, 2007; Jaluria *et al.*, 2007; Nie *et al.*, 2008).

II.1.2. Définition et objectifs

En 2004 et 2005, plusieurs chercheurs regrettaient qu'il n'existe pas de définition claire et consensuelle du terme biologie des systèmes dont l'emploi se banalise pourtant (Aderem, 2005; Bork & Serrano, 2005; Cowley, 2004; Kirschner, 2005; Levesque & Benfey, 2004; Liu, 2005; Poetsch & Wolters, 2008). Tous s'accordent néanmoins sur le fait que la complexité biologique s'étend au-delà de la simple addition des propriétés déduites des différents niveaux hiérarchiques. Les fonctions cellulaires vont naître des interactions entre les divers éléments qui les composent justifiant une approche pluridisciplinaire du problème. La quantité et l'hétérogénéité des données collectées grâce aux technologies post-génomiques introduisent par exemple une forte contrainte mathématique pour leur intégration. En ce sens, la biologie des systèmes se situe à l'interface entre plusieurs domaines scientifiques ; elle vise à analyser et réconcilier les données issues des différents niveaux de régulation afin d'obtenir une vision globale et compréhensive de la physiologie cellulaire par la modélisation des fonctionnements biologiques.

II.1.3. Les différentes approches en biologie intégrative

Bien que poursuivant le même but de compréhension globale de la cellule dans son ensemble, deux approches de modélisation sont utilisées en biologie des systèmes (figure 2, (Bruggeman & Westerhoff, 2007; Coveney & Fowler, 2005; de Lorenzo & Galperin, 2009)).

II.1.3.1. Top-down

L'approche dite « top-down » ou inductive consiste à utiliser des outils mathématiques pour construire un modèle à partir de l'ensemble des données globales disponibles (transcriptome, protéome, métabolome...). Ce type de modélisation est le plus classiquement utilisé car il ne requiert aucune connaissance *a priori* sur le fonctionnement du système mais permet d'en identifier les déterminants. Si les modèles obtenus ne décrivent pas forcément la réalité biologique du système, ils apportent néanmoins une vision globale et fournissent des informations quant aux divers composants impliqués (Bruggeman & Westerhoff, 2007; Coveney & Fowler, 2005). Ainsi, l'approche « top-down » est avant tout une démarche à visée explicative.

II.1.3.2. Bottom-up

A l'inverse, la démarche déductive ou « bottom-up » débute par la description mathématique d'un système réalisée sur la base des connaissances biologiques ou de la compilation des données issues de la littérature. Les données expérimentales sont alors utilisées pour tester la validité du modèle établi et éventuellement en estimer les paramètres. Cette approche est moins utilisée que celle du « top-down » car elle nécessite un niveau de connaissance plus élevé afin de se baser sur les mécanismes plutôt que sur les phénomènes (Bruggeman & Westerhoff, 2007; Coveney & Fowler, 2005).

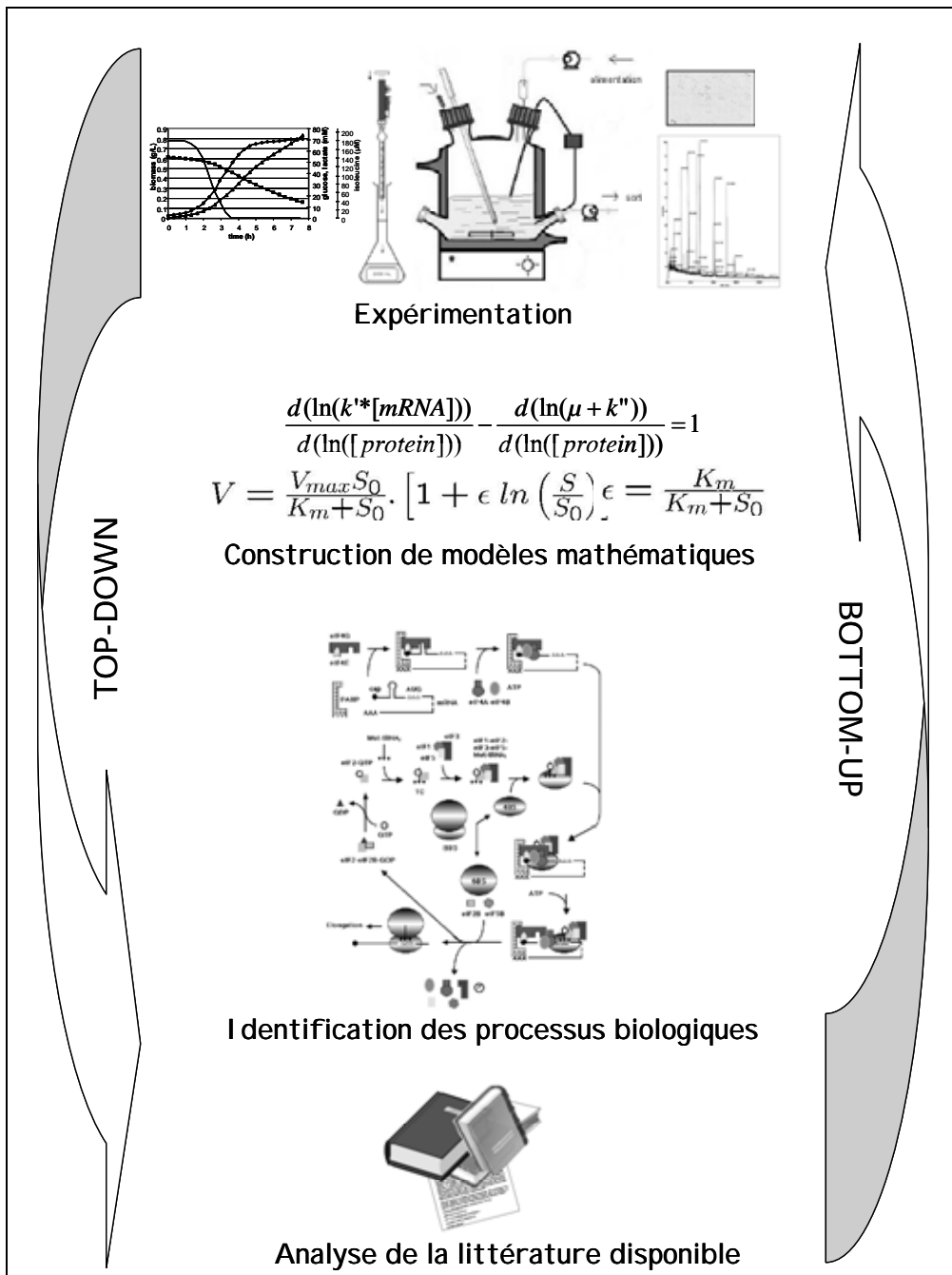


Figure 2 : Approches top-down et bottom-up de la biologie des systèmes.

II.1.4. Challenges

Depuis de nombreuses années, la biologie a souvent été considérée comme la matière de prédilection des scientifiques réfractaires aux mathématiques (Cassman, 2005). Pourtant, de par son caractère intégratif, la biologie des systèmes tend à réunir de nombreuses disciplines scientifiques historiquement dissociées et indépendantes. Cette séparation académique des sciences constitue une première épreuve pour la biologie intégrative. Le franchissement de cet obstacle passe en premier lieu par l'établissement d'une communication entre scientifiques de diverses formations et requiert donc l'établissement d'un langage commun : chacun se doit de fournir des efforts pour tenter de comprendre les grandes lignes des autres disciplines tout en essayant de vulgariser la sienne afin de la rendre accessible. En ce sens, la réorganisation des structures de recherche permettant la mise en place d'unités pluridisciplinaires devrait favoriser la collaboration entre biologistes, chimistes, physiciens, bioinformaticiens, mathématiciens et tout autre spécialiste dont les connaissances pourraient être utiles (Aderem, 2005; Cowley, 2004). Plusieurs entités de recherche en biologie des systèmes ont déjà été créées depuis 2002 au Japon, aux Etats-Unis et dans différents pays d'Europe (<http://systems-biology.org/links/corporate-and-institute-sites/>). En France, des unités spécifiques de biologie intégrative commencent à se structurer avec, par exemple, le Laboratoire de Biologie Intégrative (LBI) de Saclay ou encore l'Institut de Biologie Intégrative (IFR83) de Jussieu. L'approche biologie intégrative est aussi fortement encouragée *via* le soutien à des projets collaboratifs et le lancement de programmes fédérateurs tel que le projet européen BaSysBio qui vise à étudier le fonctionnement global de la bactérie modèle *Bacillus subtilis* (<http://www.basysbio.eu/>). Enfin, parce que la biologie des systèmes est perçue comme une démarche d'avenir, l'accent est aussi mis sur la formation avec l'ouverture de cursus portant la mention biologie intégrative.

A ce challenge intellectuel de taille, s'ajoutent des défis d'ordre technique. La biologie intégrative nécessite en effet la mesure globale d'un maximum de niveaux ce qui requiert à la fois l'amélioration des technologies existantes et le développement de nouveaux outils et méthodologies de mesure (Aderem, 2005; Bork & Serrano, 2005; Levesque & Benfey, 2004), comme de traitement des données (Burrage *et al.*, 2006). D'autre part, la mise en place de standards en termes de normalisation et de collecte des résultats issus de niveaux d'analyses différents ou identiques, mais surtout obtenus par diverses méthodes, devrait faciliter le partage des données et stimuler les avancées en biologie des systèmes (Brazma *et al.*, 2006). Pour cela, le développement de bases de données et de ressources mathématiques est

également un point critique qui nécessite d'être perfectionné et sans cesse remis à jour (van Gend & Snoep, 2008). Il existe par exemple pour *Escherichia coli* un portail web rassemblant l'essentiel des bases de données disponibles pour cette bactérie modèle et permettant donc d'avoir accès à l'ensemble des connaissances actuellement publiques concernant ce microorganisme (<http://www.eid.ac.cn/MirrorResources/3637/ieca.html>).

Au-delà de la compréhension globale des mécanismes cellulaires *via* la réconciliation de données hétérogènes, la biologie des systèmes s'inscrit également dans une optique prédictive. Cet objectif à plus long terme reste en ligne de mire car un tel niveau de connaissance pourrait permettre de déterminer de manière rationnelle les techniques d'ingénierie à appliquer au système pour l'adapter aux nécessités productives aussi bien que thérapeutiques (Friboulet & Thomas, 2005; Liu, 2005).

II.2. Comparaison des données de transcriptome et de protéome

Bien que l'intégration simultanée des données de transcriptome et de protéome ne soit pas une fin en soi pour la biologie intégrative, relever ce premier challenge permettrait déjà d'apporter des éléments de réponse concernant les mécanismes qui sous-tendent la physiologie des cellules (Cox *et al.*, 2005). De plus, ces deux types de données en « omes » servent souvent de point de départ en raison de leur « accessibilité ». En effet, leur mesure s'effectue maintenant presque en routine et de nombreuses données sont déjà disponibles dans la littérature.

II.2.1. Analyses indépendantes et comparaison des résultats

Les analyses transcriptomiques comme protéomiques constituent d'excellents outils pour l'étude globale de réponses au stress (Hecker *et al.*, 2008; Schweder & Hecker, 2004; Silberbach & Burkovski, 2006; Yoshida *et al.*, 2001). De fait, l'utilisation conjointe de ces deux méthodes pour la caractérisation d'une réponse n'est pas rare. Toutefois, la prise en compte de ces deux jeux de données se fait généralement sous forme de comparaisons non quantitatives et ne se réfère qu'aux messagers et/ou aux protéines différenciellement exprimés afin d'obtenir une vision panoramique des stratégies d'adaptation à un stress particulier. Ce sont donc directement les résultats, voire même les interprétations, comme l'activation d'une voie métabolique par exemple, issus de chacune des deux approches qui sont comparés. Dans ce cas, c'est donc surtout la complémentarité de ces deux techniques qui est mise en avant : elles permettent d'identifier des biomolécules appartenant à des mêmes groupes fonctionnels ce

qui confirme et complète à la fois les résultats qui auraient été obtenus en utilisant une seule des deux méthodes (Griffin *et al.*, 2002; Mader *et al.*, 2002a; Sampathkumar *et al.*, 2006).

Devant l'influence que peuvent avoir les différentes phases de culture sur la physiologie cellulaire (Chaussee *et al.*, 2008; Coleman *et al.*, 2004; Ramanathan & Schreiber, 2007), il est regrettable que les comparaisons de transcriptome et de protéome ne soient pas toujours effectuées sur des cellules provenant rigoureusement du même état physiologique. En effet, si certaines équipes ont recours à des cultures maîtrisées et régulées en fermenteurs afin d'assurer le contrôle et la reproductibilité des états cellulaires (Jurgen *et al.*, 2005; Silberbach *et al.*, 2005; Yoon *et al.*, 2003), d'autres se contentent simplement de définir la composition initiale du milieu, la température de croissance et l'absorbance à laquelle les prélèvements doivent être effectués (Eymann *et al.*, 2002; Mader *et al.*, 2002b; Mostertz *et al.*, 2004; Tam le *et al.*, 2007). En outre, il a déjà été démontré que la reproductibilité, et, par conséquent, la qualité des données, étaient bien meilleures lorsque les conditions de culture sont strictement contrôlées (Piper *et al.*, 2002).

II.2.2. Comparaison directe des niveaux de transcrits et de protéines

Si les données de transcriptome et de protéome sont souvent utilisées à des fins de validation ou de complémentation, il existe des études portant sur leur recoupement quantitatif. Principalement menées sur la levure modèle *Saccharomyces cerevisiae*, ces analyses quantifient le lien mathématique, couramment appelé corrélation, entre les deux jeux de données au travers du calcul de différents coefficients de corrélation qu'il convient de définir afin de mieux en percevoir la signification.

II.2.2.1. Les outils statistiques : coefficients de corrélation

Le calcul de la covariance entre deux variables quantitatives, c'est-à-dire leurs variations simultanées par rapport à leurs moyennes respectives, est un moyen de mesurer le lien entre ces deux variables aléatoires. Plus la valeur de la covariance sera forte plus la relation linéaire entre les deux variables d'intérêt sera élevée. Toutefois, cet indice n'étant pas normalisé il est dépendant des unités de chacune des variables ce qui le rend difficilement comparable avec d'autres mesures. Diviser la covariance par l'écart type (mesure de la dispersion des observations) de chacune des variables permet d'obtenir un nouvel indice standardisé, insensible aux unités de mesures, appelé coefficient de corrélation de Pearson ou R_p . Compris entre -1 et 1, ce coefficient permet donc de mesurer le degré de dépendance linéaire entre deux variables quantitatives (Valleron, 2007). Le coefficient de corrélation de Spearman ou

R_S , se base quand à lui non plus sur les valeurs exactes des observations mais sur leurs rangs. Le calcul de R_S requiert donc une classification des observations et permet de repérer les corrélations monotones entre variables sans pour autant que ce ne soit des droites. De la même manière, le coefficient de corrélation de Kendall, ou τ , est un coefficient non-paramétrique qui mesure le degré de correspondance entre deux classements.

Les coefficients de corrélation de Pearson et de Spearman sont les plus couramment utilisés lors de la comparaison de données biologiques aussi bien que pour la plupart des analyses de corrélations. Dans le cas de corrélations linéaires, le coefficient de détermination (R^2) qui correspond au carré de R_P est aussi fréquemment employé car, bien que faisant perdre le sens de la corrélation (en temps que carré d'un nombre compris entre -1 et 1, il est lui-même entre 0 et 1), il indique directement le pourcentage de variation expliqué par la relation entre deux variables (Valleron, 2007).

II.2.2.2. Recherche de corrélations

La plupart des recherches de corrélations entre les données de transcriptome et de protéome sont directement effectuées sur les valeurs d'abondances de chacune des biomolécules mais peuvent également se faire sur les ratios de variation des niveaux entre deux conditions (Feder & Walser, 2005). La majorité de ces comparaisons a été menée sur la levure modèle *Saccharomyces cerevisiae* et les valeurs de corrélation obtenues restent relativement modestes avec des R_P qui oscillent généralement entre 0.45 et 0.75 (Feder & Walser, 2005; Hack, 2004). Pour les bactéries, la littérature s'avère être encore plus pauvre ; une étude menée chez *Desulfovibrio vulgaris* cultivé dans trois conditions différentes rapporte néanmoins des coefficients de détermination (R^2) de 0.28, 0.20 et 0.25 (correspondant donc à des R_P de 0.53, 0.45 et 0.50 respectivement) (Nie *et al.*, 2006a). Plus récemment, une étude transcriptomique et protéomique de la cinétique de croissance de la bactérie productrice d'antibiotiques *Streptomyces coelicolor* a établi un coefficient de corrélation de Pearson de 0.63 entre les ratios de variation d'expression des protéines et des messagers en phase stationnaire par rapport à la phase exponentielle de croissance (Jayapal *et al.*, 2008).

Selon les études, différentes méthodes d'analyse du transcriptome et du protéome ont été employées. De même, le nombre de protéines mesurées, les souches étudiées et les conditions de culture peuvent varier, ce qui rend délicate la comparaison des corrélations obtenues par divers groupes, qui plus est, sur des microorganismes différents. Les résultats rapportés suggèrent toutefois que les niveaux de messagers sont mieux corrélés aux niveaux de protéines chez la levure (eucaryote) que chez les bactéries (procaryotes). Cette hypothèse

est contradictoire avec l'idée selon laquelle chez les bactéries transcription et traduction sont couplées (El-Sharoud & Graumann, 2007; Gowrishankar & Harinarayanan, 2004), couplage qui devrait être d'autant plus facilité par l'absence de compartimentation chez les procaryotes par rapport aux eucaryotes. Pourtant, cette hypothèse est corroborée par une étude de 2007 dans laquelle les auteurs utilisent la même technique protéomique (Lu *et al.*, 2007). En effet, ils ont obtenu des coefficients de détermination de $R^2 = 0.43$ et $R^2 = 0.76$ pour la bactérie et la levure modèles *E. coli* et *S. cerevisiae* respectivement. Au passage, il est intéressant de noter que, dans cette étude qui rapporte les meilleurs coefficients jamais observés jusqu'alors, la réalisation du protéome utilise la technologie APEX précédemment décrite.

II.2.3. Amélioration des corrélations

La nature globalement modeste des corrélations décrites dans la littérature est en partie liée au bruit de mesure des données inhérent aux méthodes utilisées. Une étude de 2006 a d'ailleurs révélé que les variations de mesure pouvaient modifier significativement la corrélation ARNm-protéine, jusqu'à 44 % (Nie *et al.*, 2006a). Il reste que les données sont généralement traitées de manière à minimiser les bruits de mesure qui ne sont donc probablement pas entièrement responsables du manque de corrélation. D'autres paramètres influencent certainement aussi le degré de corrélation entre protéome et transcriptome.

II.2.3.1. Nature des techniques employées

Alors que globalement les techniques de transcriptomique sont couramment reconnues comme performantes, celles de protéomiques sont plutôt pointées pour expliquer le manque de corrélation entre messagers et protéines. Leur premier défaut est, contrairement aux méthodologies employées en transcriptome, de ne pas permettre une analyse exhaustive de l'ensemble des protéines de la cellule. En particulier, les protéines membranaires sont souvent exclues par les méthodes classiques basées sur la séparation par électrophorèse 2D (Braun *et al.*, 2007). Toutefois, Hack a remarqué que les corrélations obtenues par des techniques utilisant la séparation 2D étaient un peu meilleures que celles obtenues par d'autres méthodes mais il explique cette amélioration par le biais de la 2D qui a tendance à détecter les protéines les plus abondantes (Hack, 2004).

Outre cet aspect technique, il est important de comprendre que les limitations des techniques protéomiques sont directement liées à la complexité même des échantillons protéiques. En effet, ces derniers renferment des protéines de nature diverse et variée, aussi bien en termes de taille que de charge, d'hydrophobicité, de conformation, de complexation

avec d'autres molécules ou encore d'abondance (Garbis *et al.*, 2005). En ce sens, la protéomique se doit de résoudre au mieux les mélanges protéiques, en permettant la détection d'un maximum de protéines, sans pour autant compromettre la quantification des plus abondantes (saturation) et tout en restant quantitative. De plus, nous avons précédemment noté que les meilleures corrélations entre messagers et protéines ont été reportées par Lu *et al.* Ces derniers ont quantifié les protéines par la technique APEX décrite précédemment et qui introduit un facteur correctif pour l'estimation des niveaux de protéines (Lu *et al.*, 2007), ce qui souligne l'influence que peut avoir la normalisation dans le traitement des données. Dans le même ordre d'idée, le manque de corrélation pourrait aussi résulter des différents traitements statistiques appliqués aussi bien dans les étapes de traitement des données de protéome (Nie *et al.*, 2008) et de transcriptome (Baldi & Long, 2001), que lors de la recherche de corrélations (Nie *et al.*, 2007).

II.2.3.2. Catégorisation des données

Dès 1999, une étude sur la levure a montré que des gènes présentant des niveaux de messagers similaires pouvaient être associés à des niveaux de protéines très variables (Gygi *et al.*, 1999) suggérant des différences de corrélation selon les gènes étudiés. Sans qu'il n'y ait de changement de méthode de mesure ou d'analyse, des variations importantes du degré de corrélation entre transcriptome et protéome lors de l'étude de certains groupes de gènes particuliers ont en effet pu être relevées depuis. En 2003, Washburn *et al.* ont trouvé une corrélation globale modeste entre les niveaux de protéines et de messagers mais ils se sont également aperçus que les corrélations étaient bien meilleures lorsque les données étaient regroupées par voie de biosynthèse des acides aminés (Washburn *et al.*, 2003). Depuis, cette dépendance de la corrélation avec la fonction cellulaire a de nouveau été reportée pour les gènes de la catégorie régulation de *S. cerevisiae* (Beyer *et al.*, 2004) et ceux liés au métabolisme chez *Schizosaccharomyces pombe* (Schmidt *et al.*, 2007). Similairement, chez *S. cerevisiae*, il semblerait qu'il y ait même un lien entre la corrélation et la localisation cellulaire de la protéine (Beyer *et al.*, 2004). Kolkman *et al.* (2006) ont également révélé que la corrélation ARNm-protéine pouvait varier selon les ratios d'expression des messagers comparant deux conditions de croissance : elle est meilleure pour les gènes surexprimés en réponse à la limitation en glucose alors, qu'à l'inverse, elle diminue pour les gènes surexprimés en réponse à la limitation en azote.

II.3. Régulations des niveaux de transcrits et de protéines

La faiblesse du lien entre les niveaux de transcrits et de protéines au sein d'une cellule remet en cause le dogme central de la biologie moléculaire d'après lequel une forte corrélation est attendue. Il stipule en effet que l'ADN code pour des ARN messagers codant eux-même pour des protéines, lesquelles vont remplir les diverses fonctions cellulaires. Le fait que les niveaux de messagers ne corrént pas parfaitement avec les niveaux de protéines qui sont pourtant les véritables effecteurs cellulaires remet donc en question la fiabilité des données de transcriptome pour rendre compte de l'adaptation à un stress (Feder & Walser, 2005; Nie *et al.*, 2007). Il ne faut cependant pas oublier que l'expression des gènes, bien que n'impliquant *a priori* que deux étapes, la transcription et la traduction, n'est pas un mécanisme linéaire et met également en jeu d'autres niveaux de régulations (Mata *et al.*, 2005) (figure 3). Face au manque de corrélation entre transcrits et protéines, nous nous intéressons en particulier aux mécanismes de régulation post-transcriptionnels susceptibles d'affecter les niveaux de ces deux types de molécules.

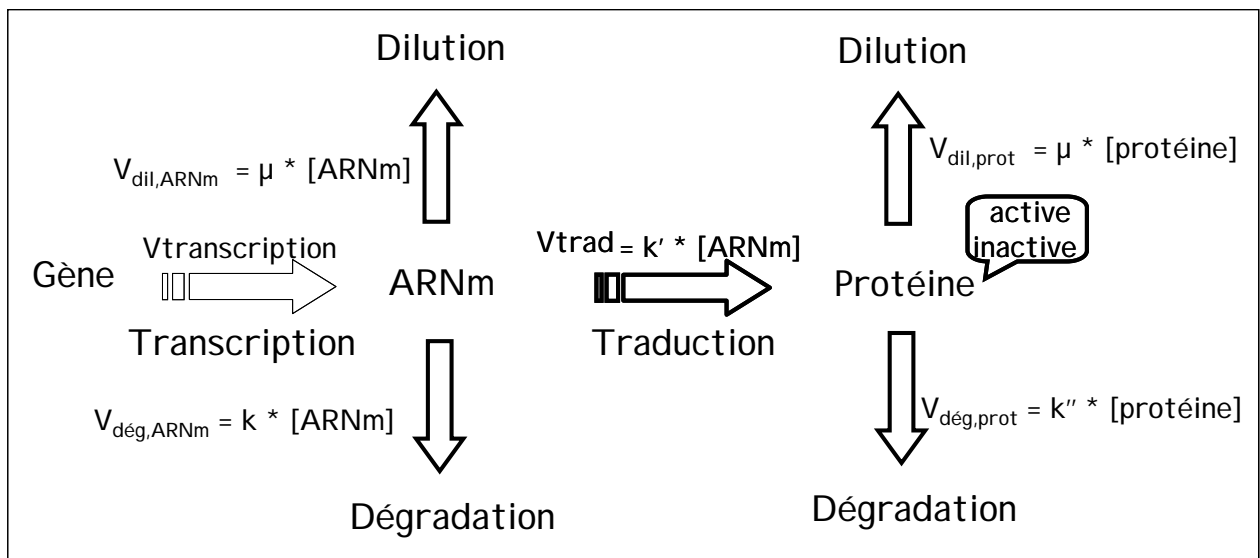


Figure 3 : Les différentes étapes impliquées dans l'expression des gènes.

μ : taux de croissance; k : constante de dégradation des ARNm, k' : efficacité de traduction; k'' : constante de dégradation des protéines, les termes entre crochet ([]) désignent des concentrations

II.3.1. Stabilité des messagers

La mesure du transcriptome d'une cellule est généralement considérée comme le reflet du contrôle transcriptionnel de l'expression des gènes. La quantité mesurée pour un ARN messager résulte pourtant d'un équilibre entre sa synthèse, sa dilution par la croissance et sa stabilité (figure 3).

II.3.1.1. Données de stabilité des messagers

Le phénomène de dilution des messagers est souvent considéré comme négligeable devant leur vitesse de dégradation. De fait, les études menées à l'échelle globale ont rapporté des temps de demi-vie moyens compris entre 1.5 et 10 min pour les bactéries (Bernstein *et al.*, 2002; Hambræus *et al.*, 2003; Redon *et al.*, 2005a; Selinger *et al.*, 2003) et de 3 à 100 min pour la levure modèle *S. cerevisiae* (Wang *et al.*, 2002) à comparer respectivement à des taux de croissance (donc de dilution) maximaux de l'ordre de 45 min et 2 h respectivement. D'autre part, une grande variabilité de la stabilité au sein des transcrits a pu être observée (Bernstein *et al.*, 2002; Redon *et al.*, 2005a), ce qui indique un fin contrôle individuel. Enfin, la dégradation des transcrits n'est pas constante dans le temps ; c'est donc un mécanisme d'adaptation permettant d'ajuster les niveaux de messagers au besoin de la cellule afin de rapidement répondre aux changements environnementaux (Redon *et al.*, 2005a; Takayama & Kjelleberg, 2000). Chez les eucaryotes, c'est également un mécanisme de régulation en conditions de stress qui s'oppose à la transcription et permet donc d'en contrebalancer l'effet (Shalem *et al.*, 2008) mais qui assure aussi l'élimination des ARNm défectueux (Amrani *et al.*, 2006; Shyu *et al.*, 2008).

II.3.1.2. Mécanismes de dégradation des messagers

Plusieurs mécanismes de dégradation existent conjointement au sein des cellules : ils peuvent différer d'un microorganisme à l'autre, surtout entre eucaryotes et procaryotes (Kushner, 2004) mais sont peu coûteux en énergie et impliquent tous l'intervention de ribonucléases (RNAses) capables de réaliser des clivages à l'intérieur des messagers (endoribonucléases) ou au niveau d'une de leurs extrémités (exoribonucléases). Chez les eucaryotes, il existe deux mécanismes majeurs (Garneau *et al.*, 2007; Mata *et al.*, 2005; Meyer *et al.*, 2004). Les molécules d'ARN messenger eucaryotes sont polyadénylées au niveau de leur extrémité 3' mais leur extrémité 5' est libre pour une attaque par des exoribonucléases. D'autre part, la dégradation de la queue polyA par une enzyme ayant une activité déadénylase favorise l'attaque en 3' par un complexe multi-protéique appelé exosome. Ce complexe de dégradation est composé de nombreuses protéines de liaison à l'ARN mais ses constituants majeurs sont une exoribonucléase, une hélicase, une émolase et une polynucléotide phosphorylase (ou PNPase = endoribonucléase) (Schmid & Jensen, 2008). L'attaque directe par une endoribonucléase est un mécanisme de dégradation plus rarement rencontré chez les eucaryotes (Garneau *et al.*, 2007; Mata *et al.*, 2005; Meyer *et al.*, 2004). Chez les procaryotes, l'initiation de la dégradation des messagers passe généralement par un clivage

endoribonucléasique dans les régions riches en adénine (A) et uridine (U) (Rauhut & Klug, 1999). La dégradation est ensuite assurée par un complexe protéique appelé dégradosome et équivalent à un exosome bien que présentant une structure beaucoup moins complexe (Carpousis, 2007).

II.3.1.3. Facteurs influençant la stabilité des messagers

Les différences de valeurs de stabilité entre les messagers suggèrent un lien entre leurs caractéristiques intrinsèques et leur vitesse de dégradation. Ainsi, la présence de structures secondaires vers l'extrémité 5', tout comme la position du site de liaison des ribosomes (RBS) influencent la stabilité des messagers (Sharp & Bechhofer, 2005). Le RBS, aussi appelé séquence Shine-Dalgarno, est une région riche en purine située environ 10 paires de bases en amont du codon start. Chez les procaryotes, le recrutement et l'assemblage des ribosomes sur les ARN messagers dépendent de la complémentarité entre le RBS et une séquence située à proximité de l'extrémité 3' de l'ARN ribosomique 16S, l'un des constituants de la sous-unité 30S des ribosomes. D'après des travaux menés sur *E. coli*, Sharp et Bechhofer (2005) ont cependant conclu que la stabilité des messagers est principalement influencée par la présence de structures secondaires (tiges-boucles essentiellement) et que la force et la position du RBS avaient un effet moindre. D'autre part, la queue poly(A) située en 3' des ARN messagers eucaryotes a un effet protecteur (Meyer *et al.*, 2004). *A contrario*, elle a plutôt un effet déstabilisateur en 3' des ARNm procaryotes (Mohanty & Kushner, 1999), probablement parce que la polyadénylation limite l'appariement des nucléotides de l'extrémité 3' favorisant ainsi l'attaque par les exoribonucléases (Deutscher, 2006).

L'interaction entre les messagers et les ribosomes lors de la traduction a également un impact sur leur stabilité (Deana & Belasco, 2005), en particulier parce que les ribosomes fixés sur les messagers restreignent leur accessibilité pour un clivage par les ribonucléases (Condon, 2003). Dès 1997, un modèle mécanistique de la dégradation des messagers prédisait qu'un espacement entre les ribosomes supérieur à 24 nucléotides était nécessaire pour permettre le clivage des transcrits d'*E. coli* par la RNAse E (Carrier & Keasling, 1997). Enfin, si cet effet protecteur des ribosomes sur les messagers est couramment admis, des pauses en cours de traduction peuvent, à l'inverse, avoir un effet déstabilisateur. Une étude menée chez *E. coli* a en effet démontré qu'un ralentissement de la traduction en raison de la présence d'un codon rare induisait le clivage endonucléolytique du messenger à proximité du site de pause du ribosome (Sunohara *et al.*, 2004).

II.3.2. Régulation de la traduction

Bien que rarement évoquée comme facteur responsable de la discordance entre les niveaux de messagers et de protéines, la traduction, en tant qu'intermédiaire entre les deux (figure 3), joue un rôle central et toute régulation y étant associée aura forcément un impact sur le lien entre transcrits et protéines. Grâce à la séparation des polysomes (messagers + ribosomes) sur gradient de saccharose suivie d'une analyse de type puce à ADN mise au point par Arava (2003) pour la levure modèle *S. cerevisiae*, les différences de densité en ribosomes sur les messagers ont clairement démontré que l'expression des gènes était régulée au niveau de la traduction (Melamed & Arava, 2007). Chez les procaryotes (Kaczanowska & Ryden-Aulin, 2007) comme chez les eucaryotes (Kozak, 2007), la traduction est un mécanisme complexe qui se divise en trois étapes : l'initiation au cours de laquelle le ribosome est recruté par le RBS et s'assemble au niveau du codon start ; l'élongation, qui est la véritable étape de lecture des ARN messagers par les ribosomes qui avancent de l'extrémité 5' vers l'extrémité 3' à une vitesse dépendante des codons qu'ils rencontrent ; l'étape de terminaison met fin à la lecture et entraîne le détachement du ribosome suite à la rencontre d'un codon stop. La régulation de la traduction est donc susceptible de se jouer au niveau de ces différentes étapes et implique des caractéristiques liées à la séquence des ARN messagers mais peut également faire intervenir des molécules régulatrices.

II.3.2.1. Déterminants de séquence

L'efficacité de l'initiation de la traduction est évaluée par la force d'interaction entre le RBS et le ribosome. Chez *Desulfovibrio vulgaris*, ATG est à la fois le codon start le plus utilisé et le plus efficace pour initier la traduction. En ce qui concerne l'étape d'élongation, la comparaison des niveaux de protéines avec leur composition en acides aminés a révélé un usage préférentiel des acides aminés présentant un faible coût énergétique dans les protéines les plus abondantes de *B. subtilis* et *E. coli* (Akashi & Gojobori, 2002). Ce biais d'acides aminés est couplé à un biais de codons. De la même façon, des travaux conduits chez la levure (Brockmann *et al.*, 2007; Tuller *et al.*, 2007) ont montré que l'efficacité de traduction, bien sûr dépendante de la quantité de messenger et du taux d'occupation par les ribosomes, était également influencée par des paramètres liés à la séquence intrinsèque du gène exprimé comme son CAI (Codon Adaptation Index, (Sharp & Li, 1987)) ou son tAI (tRNA Adaptation Index, (dos Reis *et al.*, 2003)), tous deux reflétant un biais d'usage de codons. La contribution des biais de codons et d'acides aminés a d'ailleurs été évaluée à 8.9 et 7.7 % respectivement chez *S. cerevisiae* (Wu *et al.*, 2008). L'environnement du codon stop est,

quant à lui, déterminant dans l'étape de terminaison (Nie *et al.*, 2006b; Rocha *et al.*, 1999). Chez *B. subtilis* la terminaison est facilitée par la présence de codons UAA (Rocha *et al.*, 1999) tandis que chez la levure un codon TAG suivi d'un autre G est plus efficace (Nie *et al.*, 2006b).

Bien que l'initiation soit encore considérée comme l'étape limitante de la traduction (Simonetti *et al.*, 2008), une étude comparative de l'influence de divers paramètres affectant les trois étapes de la traduction chez la bactérie *Desulfovibrio vulgaris*, tend à prouver que la limitation proviendrait plutôt de l'élongation. En effet, les auteurs ont trouvé un contrôle prépondérant des niveaux de protéines par les biais d'usage de codons et d'acides aminés par rapport aux environnements des codons stop et start respectivement (Nie *et al.*, 2006b). D'autre part, la forte corrélation entre le CAI et les niveaux de protéines, meilleure qu'avec les caractéristiques de séquences liées à d'autres étapes de la traduction, n'est pas nouvelle (Lithwick & Margalit, 2003) ce qui laisse entrevoir la possibilité d'utiliser ce paramètre comme une mesure indirecte de la vitesse de traduction (Brockmann *et al.*, 2007). Toutefois, le CAI ne permet pas de prédire plus de 65 % de l'expression des protéines de *S. cerevisiae* et n'est donc que partiellement fiable (Friberg *et al.*, 2004). Finalement, il est difficile de trancher quant à l'étape limitante de la traduction, d'autant plus qu'une récente approche de modélisation (Zouridis & Hatzimanikatis, 2008) indique que tous les cas de figure sont possibles : si la densité des ribosomes est faible, l'initiation est limitante, si elle est forte, c'est la terminaison qui devient le facteur limitant et une densité intermédiaire révèle un contrôle par l'élongation. Les données expérimentales de *S. cerevisiae* indiquent cependant que la densité ribosomale en croissance exponentielle est inférieure à la densité maximale suggérant un contrôle par l'initiation (Arava *et al.*, 2003).

II.3.2.2. Biomolécules régulatrices

Si le nombre de ribosomes fixés sur un messenger donné peut varier en fonction des conditions de croissance, c'est également le cas des protéines qui vont interagir aussi bien avec l'ARNm qu'avec les ribosomes. En particulier, certaines protéines, qualifiées de ribosome modulation factor (RMF), limitent l'activité des ribosomes *via* leur dimérisation (El-Sharoud, 2004; Yoshida *et al.*, 2004). En outre, les protéines de liaison à l'ARN, ou RBP (pour RNA Binding Proteins), vont affecter la traduction mais également la stabilité des messagers (Glisovic *et al.*, 2008). Par exemple, lorsque les protéines ribosomales sont en excès par rapport aux ARN ribosomaux, elles se fixent à leurs propres ARNm et répriment leur traduction (Dennis *et al.*, 2004). De la même façon, la fixation des petits ARN non codant

aussi appelés miRNA (surtout pour les eucaryotes), influence à la fois la traduction et la stabilité des transcrits (Brantl, 2007; Wassarman, 2002). Certaines structures secondaires présentes au niveau de l'extrémité 5' des ARN messagers sont appelées « riboswitches » en raison de leur propriétés : suite à leur interaction avec un ligand (acide aminé, sucre ou base nucléique par exemple), elles peuvent changer de conformation et ainsi moduler la traduction (Henkin, 2008). La formation de structures secondaires et notamment d'hélices peut endosser le rôle de terminateur de transcription ou encore séquestrer le site d'initiation de la traduction (Grundy & Henkin, 2006).

II.3.3. Dégradation des protéines

Les mécanismes de dégradation des protéines varient légèrement selon que l'on s'intéresse à des microorganismes eucaryotes ou procaryotes. Dans tous les cas, c'est néanmoins un processus cellulaire coûteux en énergie car il implique des protéases ayant une activité ATPase (Ogura & Wilkinson, 2001).

II.3.3.1. Machineries de dégradation des protéines

Chez les eucaryotes, la dégradation des protéines est assurée par un complexe protéique appelé le protéasome (Tanaka, 2009). Composé d'un assemblage cylindrique de protéases (20S) et d'une partie régulatrice renfermant l'activité ATPase et les domaines de reconnaissance de l'ubiquitine (19S), cette machinerie (26S au total) dégrade spécifiquement les protéines marquées par de l'ubiquitine (Murata *et al.*, 2009). Chez les bactéries, un complexe protéasique de type 20S est capable de s'assembler de manière autonome (Murata *et al.*, 2009). Le temps de demi-vie des protéines procaryotes est indépendant de l'ubiquitination et serait déterminé par la nature de leur résidu N-terminal spécifiquement reconnu par le protéasome (Mogk *et al.*, 2007). En effet, certains résidus N-terminaux sont connus pour leur effet stabilisateur ou déstabilisateur mais leur classification dépend des espèces étudiées. Chez *E. coli*, ce sont par exemple les acides aminés hydrophobes (leucine, phénylalanine, tyrosine et tryptophane) qui ont le plus grand effet déstabilisateur suivis par les acides aminés basiques (arginine et lysine) (Varshavsky, 1996).

II.3.3.2. Données de dégradation des protéines

A l'heure actuelle, seules deux études à grande échelle sont publiées sur la stabilité des protéines (Belle *et al.*, 2006; Pratt *et al.*, 2002). Menés chez la levure, ces travaux montrent une grande hétérogénéité des taux de dégradation au sein de l'ensemble des protéines pour lesquelles ils ont pu être mesurés. La plus ancienne publication, basée sur la disparition

progressive d'un marquage par de la leucine deutériée lors d'une culture à faible taux de croissance en chémostat, rapporte un temps de demi-vie moyen de 31 heures (Pratt *et al.*, 2002), tandis que celui calculé pour des protéines marquées par une étiquette fluorescente et prélevées en phase exponentielle de croissance n'est que de 45 minutes (Belle *et al.*, 2006). L'hétérogénéité de ces estimations souligne l'impact que peuvent avoir les techniques de mesure ainsi que les conditions de croissance sur les résultats. Malgré cette très grande différence en termes de temps de demi-vie, les deux équipes s'accordent néanmoins à dire que la stabilité des protéines est un phénomène dont l'influence sur les niveaux de protéines ne doit pas être négligée.

Nous avons précédemment vu que la dilution des messagers était négligeable devant leur vitesse de dégradation. D'après les ordres de grandeur que nous venons de donner pour leur temps de demi-vie, nous pouvons supposer qu'il n'en va pas de même pour les protéines.

II.3.4. Modulation des activités protéiques

Bien que quelque peu en marge de l'intégration des données de transcriptome et de protéome pour la compréhension du mécanisme de traduction, il est important de garder à l'esprit que plusieurs paramètres peuvent moduler l'activité protéique et donc influencer la physiologie cellulaire. En effet, la présence d'une protéine au sein de la cellule ne garantit pas qu'elle soit active. L'activité d'une protéine va, de fait, être contrôlée et modulée par divers paramètres tels que les modifications post-traductionnelles (Jensen, 2006), qui sont courantes chez les eucaryotes mais existent aussi chez les bactéries (Deutscher & Saier, 2005), la formation de complexes ou encore les régulations allostériques (Swain & Gierasch, 2006).

II.4. Modélisation de la traduction

La traduction permet de convertir l'information contenue dans les ARN messagers en protéines, les véritables effecteurs cellulaires. En ce sens, c'est un processus majeur qui va notamment jouer un rôle important dans l'adaptation. Devant la complexité de ce phénomène biologique et des différents niveaux de régulation qui peuvent y être associés, la modélisation mathématique se pose comme un outil intégratif susceptible d'offrir une vision globale de la traduction. Son objectif premier consiste à décrire et comprendre les mécanismes impliqués dans la régulation de ce processus. A plus long terme, ce ne sont plus des modèles explicatifs mais des modèles prédictifs qui sont attendus. Le but optimal serait même d'être capable de prévoir les comportements cellulaires en réponses à divers stimuli afin de pouvoir contrôler

les microorganismes et d'exploiter au mieux leurs potentialités. A l'heure actuelle les modèles proposés dans la littérature sont essentiellement explicatifs et sont construits sur la base de données expérimentales voire seulement théoriques.

II.4.1. Modèles intégratifs basés sur les données génomiques

Nous avons vu précédemment qu'en raison de sa bonne corrélation avec les niveaux de protéines, le CAI pouvait éventuellement être utilisé comme un indicateur de traduction (Brockmann *et al.*, 2007). Il a cependant été démontré que la capacité de ce paramètre à représenter l'expression des protéines était limitée, en tout cas chez la levure (Friberg *et al.*, 2004). Depuis deux ou trois ans, ce constat, additionné à celui de plus en plus évident du faible recoupement entre les données de transcriptome et de protéome, a conduit quelques groupes à travailler sur l'intégration simultanée de plusieurs variables, dont le CAI et l'abondance en messenger, grâce à des régressions multiples. Peu de travaux sont publiés à l'heure actuelle mais les recherches menées ont abouti à la construction de modèles intégratifs ayant pour but de prédire au mieux les niveaux de protéines, la pertinence des modèles étant évaluée par le calcul d'un coefficient de corrélation comme pour la recherche de corrélation entre les niveaux de protéines et de transcrits (paragraphe II.2.2.1).

II.4.1.1. Modèles disponibles et variables prises en compte

Les deux modèles bactériens proposés portent sur des données de transcriptome et de protéome recueillies pour la bactérie *D. vulgaris* placée dans trois conditions de croissance différentes (Nie *et al.*, 2006a; Nie *et al.*, 2006b). Le premier des deux modèles utilise l'abondance en messenger mais pas le CAI et englobe également les catégories fonctionnelles, la stabilité des messagers, la stabilité des protéines, la longueur des gènes et la variabilité des mesures, estimée à partir de la reproductibilité des répliques (Nie *et al.*, 2006a). Globalement, ce modèle explique 52 à 61 % (R^2) des variations d'abondances de protéines observées alors que les valeurs de transcrits seules ne permettraient d'en expliquer que 20 à 28 %. L'autre modèle disponible pour *D. vulgaris* ne cherche pas vraiment à prédire les niveaux de protéines mais plutôt à mesurer la contribution de l'efficacité de traduction dans les niveaux atteints (Nie *et al.*, 2006b). Pour cela, plusieurs variables liées à l'initiation (séquence Shine-Dalgarno, nature et environnement du codon start), l'élongation (biais de codons et d'acides aminés) ou la terminaison (nature et contexte du codon stop) de la traduction sont prises en compte. Bien que ce ne soit pas directement rapporté dans la publication, il est possible de calculer que l'ensemble de ces variables plus les niveaux de messagers expliquent au total 36

à 41 % des variations des niveaux protéiques. Cette approche de modélisation par régression multiple a été étendue à la levure par le même groupe à partir de données disponibles dans la littérature (Wu *et al.*, 2008). Dans ce cas, l'intégration de données d'usages de codons, d'acides aminés, de densité ribosomale, de taux d'occupation par les ribosomes, de stabilité des protéines et des messagers permettent aux auteurs d'expliquer 33 % des variations de niveaux de protéines. En 2007, deux études sur *S. cerevisiae* (Brockmann *et al.*, 2007; Tuller *et al.*, 2007) avaient abouti à la construction de deux modèles corrélant environ à 50 % avec les données de protéome (valeur estimée à partir des R_S fournis dans ces deux publications). Ces deux modèles prennent pourtant en compte des paramètres différents : l'un n'intègre que le tAI et le taux d'évolution des protéines (calculé à partir de la variabilité génétique entre protéines homologues de deux sous-espèces) en plus des niveaux de messagers (Tuller *et al.*, 2007) tandis que l'autre propose d'utiliser une combinaison non linéaire (mais linéarisable par transformation logarithmique) du CAI avec l'abondance en messenger, la densité ribosomale et le taux d'occupation par les ribosomes. Aucun de ces modèles n'arrive à expliquer plus de 60 % des variations de niveaux protéiques suggérant donc l'implication d'autres phénomènes de régulation des quantités de protéines.

II.4.1.2. Limitations des modèles existants

Au final, l'ensemble de ces modèles met en lumière divers facteurs, différents des messagers, mais influençant aussi les niveaux de protéines. Selon la nature de ces paramètres, il est donc possible de se faire une idée des étapes soumises à régulation et des caractéristiques génétiques qui peuvent avoir de l'influence. Il est néanmoins regrettable, que l'importance relative de chacun des paramètres intégrés dans un modèle ne soit pas toujours estimée. Seules les publications du groupe de Zhang proposent une classification des paramètres par ordre d'importance (Nie *et al.*, 2006a; Nie *et al.*, 2006b; Wu *et al.*, 2008). Cette hiérarchisation est basée sur le calcul individuel pour chacun des paramètres d'un

coefficient de détermination ajusté : $R_{paramètre}^2 = \frac{R_{ARNm,paramètre}^2 - R_{ARNm}^2}{1 - R_{ARNm}^2}$. Ce R^2 ajusté permet

donc de rendre compte de la contribution du paramètre étudié sur la corrélation ARN-protéine. Toutefois, les effets des différents paramètres ne sont pas forcément additifs et la somme des corrélations individuelles est d'ailleurs toujours supérieure à la corrélation globale associée au modèle (Nie *et al.*, 2006a; Nie *et al.*, 2006b; Wu *et al.*, 2008). Classer les variables de cette manière conduit donc à la perte de la « dimension intégrative » introduite par les modèles de régressions multiples et ne permet pas la comparaison des effets des

paramètres. En d'autres termes, l'étude individuelle des différents paramètres ne tient pas compte du fait que deux paramètres peuvent être équivalents, comme avoir des effets synergiques ou antagonistes.

II.4.2. Modèles basés sur la description des processus biologiques

La mise en équations mathématiques des phénomènes biologiques tels que la traduction est un moyen de les appréhender sans donnée biologique. En effet, la connaissance des fonctionnements biologiques peut permettre d'écrire des modèles pour rendre compte des mécanismes mis en jeu. La résolution théorique des équations grâce à des algorithmes de simulation ouvre ensuite la voie à des interprétations biologiques dont la validité peut ensuite être testée par l'expérimentation.

Plusieurs équations sont décrites dans la littérature pour modéliser le processus de traduction. Selon les phénomènes pris en compte, les modèles diffèrent, tout comme les informations biologiques qui en sont retirées. En conceptualisant la traduction comme une vague de ribosomes traversant les ARNm, Gilchrist & Wagner (2006) ont par exemple mis à jour l'impact que pourraient avoir les erreurs de lecture et le recyclage des ribosomes sur son efficacité. De façon semblable, modéliser la traduction comme un problème de circulation a apporté un nouvel éclairage sur le lien entre le CAI et l'efficacité de traduction en montrant qu'un fort biais d'usage de codons permettait d'éviter les queues et les collisions de ribosomes et ainsi d'optimiser l'efficacité de traduction (Mitarai *et al.*, 2008). En 2006, un modèle basé sur la théorie de l'efficacité maximale a permis de réévaluer le taux d'élongation maximal et le taux d'erreur de lecture chez *E. coli* (Lovmar & Ehrenberg, 2006). Les résultats obtenus diffèrent des estimations habituelles mais ont pu être validés car ils révèlent une variation de la traduction liée au taux de croissance totalement en accord avec des données expérimentales.

Parmi les différents modèles de la traduction proposés, le plus abouti est certainement celui proposé en 2008 par une équipe américaine. Ce modèle est en effet le dernier volet de l'évolution d'un modèle de base proposé en 2003 pour représenter la traduction (Mehra & Hatzimanikatis, 2006; Mehra *et al.*, 2003; Zouridis & Hatzimanikatis, 2007; Zouridis & Hatzimanikatis, 2008). Depuis sa formulation initiale, ce modèle a subi de nombreuses évolutions et s'est fortement complexifié. Il prend maintenant en compte de multiples phénomènes tels que l'assemblage des ribosomes, la taille des polysomes, la formation des complexes ternaires (ribosomes + facteur d'élongation lié au GTP + tRNA^{aa}) et l'inhibition compétitive par les ARN de transfert chargés (tRNA^{aa}) au cours des différentes étapes de la

traduction tout au long du messager. L'analyse biologique des équations posées par ce modèle confirme que les efficacités de traduction dépendent des séquences des messagers mais révèle surtout quels codons influent sur cette efficacité : à faible densité de ribosomes la séquence en 5' est primordiale, à moyenne densité ce sont tous les codons le long du messager qui influent sur la vitesse de traduction et à forte densité les codons en 3' sont très importants.

Il reste que la construction, voire même l'application, de ce genre de modèle souvent complexe requièrent des connaissances mathématiques approfondies et la maîtrise d'outils de résolution d'équations. De fait, ils sont généralement élaborés et résolus par des mathématiciens et ne sont que très peu utilisés par les biologistes.

III. *Lactococcus lactis* et ses stress

Nous venons de voir que la biologie intégrative présente un véritable potentiel pour l'étude des réponses au stress. La mise en place d'une démarche de biologie intégrative requiert néanmoins la disponibilité de nombreux outils d'analyse de la physiologie cellulaire. Parce qu'elle répond à ce critère, la bactérie lactique *Lactococcus lactis* se présente comme modèle pertinent, d'autant plus que son petit génome (2.35 Mb) est entièrement séquencé (Bolotin *et al.*, 2001). Certaines de ses réponses au stress sont d'ailleurs déjà partiellement caractérisées mais les mécanismes d'adaptation sous-jacents ne sont encore que peu décrits. Dans ce chapitre, nous verrons pourquoi *L. lactis* présente un intérêt certain pour l'étude de réponses au stress puis nous ferons le point sur les différentes réponses décrites à l'heure actuelle et les mécanismes de régulation potentiellement impliqués.

III.1. Intérêt de *Lactococcus lactis* et de la souche IL1403

L. lactis est un microorganisme procaryote, Gram-positif, anaérobie facultatif, non sporulant et non mobile qui se présente sous forme de cellules sphériques ou ovoïdes, isolées ou formant de courtes chainettes. Comme toutes les bactéries lactiques, il est doté de capacités cataboliques lui permettant de convertir les sucres fermentescibles en acide lactique.

III.1.1. Conditions de vie et domaines d'application

Naturellement présent dans le sol et sur les végétaux, *L. lactis* est très largement utilisé dans l'industrie agro-alimentaire pour la fabrication de produits laitiers fermentés et en particulier de fromage. Sa capacité métabolique de conversion des sucres en acide lactique est en effet exploitée pour assurer le caillage du lait et la formation de la matrice fromagère. Cette propriété acidifiante est également un atout majeur pour la conservation des aliments car l'acidification inhibe la croissance de microorganismes non désirables voire pathogènes (Pitt *et al.*, 2000). En outre, certaines souches de *L. lactis* sont capables de synthétiser des bactériocines, de petites protéines toxiques pour les autres bactéries qui vont jouer un rôle important dans la sécurité alimentaire (Galvez *et al.*, 2007). Le rôle de *L. lactis* ne se limite pas à éviter la prolifération des microorganismes indésirables, il contribue aussi à la texture ainsi qu'à la qualité organoleptique des produits *via*, respectivement, la production d'exopolysaccharides (Dabour *et al.*, 2006) et d'arômes (Ayad *et al.*, 2001). Depuis quelques années, les bactéries lactiques ont connu un regain d'intérêt non plus pour leurs applications dans l'industrie agroalimentaire mais pour leurs applications dans le domaine de la santé. Certaines bactéries lactiques présentent en effet des propriétés probiotiques (Kaur *et al.*, 2002) bien que cela n'ait jamais pu être démontré pour *L. lactis* dont le caractère inoffensif en fait néanmoins un candidat idéal pour le rôle de vecteur de protéines d'intérêt thérapeutique ou de vaccin vivant (Nouaille *et al.*, 2003).

III.1.2. Les différents stress rencontrés par *L. lactis*

Dans son environnement naturel aussi bien que lors de son utilisation industrielle ou thérapeutique, *L. lactis* est régulièrement confronté à divers stress nutritionnels ou physico-chimiques pouvant même intervenir de manière concomitante (Duwat *et al.*, 2000; van de Guchte *et al.*, 2002). Dans son écosystème naturel (sol ou surface des plantes), la bactérie est exposée à l'oxygène, à de nombreuses variations de température et à de longues phases de carences nutritionnelles. En industrie agroalimentaire, le principal stress rencontré par *L. lactis* est le stress acide qu'il génère lui-même de part la production d'acide lactique lors sa croissance. De plus, un procédé fromager tel que la fabrication du cheddar le soumet successivement à des stress acide, hyperthermique, osmotique, nutritionnels, hypothermique et oxydatif lors des étapes de caillage, pressage, affinage et stockage. De la même façon, la traversée du tractus digestif fortement acide du fait de la présence de sucs gastriques et l'interaction avec la flore intestinale vont multiplier les stress subis par la bactérie.

III.1.3. La souche IL1403

Largement étudié depuis de nombreuses années, *L. lactis*, et plus particulièrement la souche IL1403 de la sous-espèce *lactis*, se pose comme un modèle des lactocoques et même plus largement des bactéries lactiques. Cette souche originellement isolée d'un milieu laitier mais curée de ses plasmides a d'ailleurs été la première bactérie lactique séquencée (Bolotin *et al.*, 2001). Son génome, relativement petit, compte 2.35 Mb et présente un pourcentage moyen en GC de 35.4 %. L'annotation automatique a montré que 86 % des régions du génome codent pour des protéines et 1.4 % pour des ARN stables, les 12.6 % restantes étant des régions non codantes. Parmi les 2310 gènes identifiés, une fonction biologique a pu être attribuée à 64.2 % d'entre eux sur la base de connaissances biologiques sur cette bactérie ou d'homologies avec des gènes déjà séquencés pour d'autres espèces. D'après ces études d'homologies, il semblerait d'ailleurs que 15.7 % des gènes soient spécifiques des lactocoques (Bolotin *et al.*, 2001).

III.1.3.1. Caractéristiques métaboliques

Grâce au séquençage de la souche IL 1403, les potentialités métaboliques de *L. lactis* ont pu être évaluées plus précisément (Bolotin *et al.*, 2001). L'analyse fonctionnelle des gènes a, en effet, permis de confirmer la présence de la majorité des gènes codant pour les enzymes de la glycolyse et de la voie des pentoses phosphates. De nombreux gènes liés à l'utilisation de sources de carbone alternatives tels que le fructose, le galactose, le maltose, le mannose, le cellobiose ou encore le xylose ont été identifiés et la fonctionnalité de certaines voies a depuis été confirmée (Barriere *et al.*, 2005; Cocaign-Bousquet *et al.*, 2002; Ohara *et al.*, 2007). Déjà décrites dans la littérature (Cocaign-Bousquet *et al.*, 1996), les différentes voies d'utilisation du pyruvate ont été confirmées par l'identification des gènes correspondants (Bolotin *et al.*, 2001). Toujours d'après l'annotation, le cycle du citrate est incomplet et une partie des gènes codant pour les enzymes de la néoglucogénèse et des voies anaplérotiques est absente.

La présence de gènes dont les protéines associées sont impliquées dans les voies de biosynthèse des acides aminés est d'autant plus surprenante que certaines (la voie de l'histidine (Delorme *et al.*, 1993) et celle de l'isoleucine (Godon *et al.*, 1993)) ont préalablement été caractérisées comme non fonctionnelles et que les exigences nutritionnelles de *L. lactis* en termes d'acides aminés sont multiples (Cocaign-Bousquet *et al.*, 1995). De façon étonnante pour une bactérie anaérobie facultative, des gènes nécessaires à la respiration ont été identifiés. Depuis, les capacités respiratoires de *L. lactis* en présence d'hème ont d'ailleurs été prouvées (Duwat *et al.*, 2001) et leur régulation a été étudiée (Gaudu *et al.*,

2003). Enfin, bien qu'aucun état de compétence naturelle n'ait jamais été décrit pour cette bactérie, l'annotation du génome a révélé la présence de gènes de compétence (Bolotin *et al.*, 2001) dont une partie serait manquante et qui seraient sous le contrôle de circuits de régulation inconnus (Claverys & Martin, 2003).

III.1.3.2. Caractéristiques régulatrices

Le séquençage du génome de *L. lactis* IL1403 a permis d'annoter 138 régulateurs potentiels (Bolotin *et al.*, 2001) dont certains paraissent particulièrement pertinents en tant que modulateurs des réponses à des changements environnementaux. Si le répresseur catabolique CcpA est le principal régulateur du métabolisme carboné (Luesink *et al.*, 1998), de nombreux régulateurs secondaires contrôlent aussi l'expression des gènes codant pour l'assimilation de divers substrats carbonés tels que le lactose (LacR), le saccharose (SacR), le fructose (FruR), le xylose (XylR), ou le maltose (MalR) (Kowalczyk & Bardowski, 2007). Le métabolisme azoté serait, quant à lui, principalement sous le contrôle de CodY (Guedon *et al.*, 2005) et GlnR (Larsen *et al.*, 2006). Par ailleurs, le système protéolytique, qui participe également au métabolisme azoté, est régulé non seulement par CodY mais aussi par CtsR, le régulateur des protéases Clp (Savijoki *et al.*, 2006). Le métabolisme des bases nucléotidiques est, quant à lui, soumis au contrôle de PurR (Kilstrup & Martinussen, 1998) et PyrR (Switzer *et al.*, 1999), les répresseurs de la biosynthèse des purines et des pyrimidines respectivement. Enfin, HrcA est un répresseur de transcription contrôlant l'expression des chaperonnes (GroEL, GroES, GrpE, DnaJ, DnaK) (Guedon *et al.*, 2002; van de Guchte *et al.*, 2002). Initialement identifiées suite à leur induction en réponse au stress hyperthermique, les chaperonnes assurent le repliement des protéines naissantes ou dénaturées et jouent un rôle majeur dans l'adaptation cellulaire (Sugimoto *et al.*, 2008).

Les facteurs sigma (σ) sont des molécules qui se fixent au cœur de l'ARN polymérase, déterminant ainsi sa spécificité de reconnaissance des séquences promotrices, ce qui leur confère un rôle essentiel dans la régulation de l'expression des gènes bactériens. D'une manière générale, les promoteurs consensus, principalement retrouvés en amont des gènes dits de ménage, sont reconnus par un facteur σ végétatif alors que des facteurs σ alternatifs interviennent pour répondre à différents stress (Gruber & Gross, 2003). De façon très étonnante, le génome de *L. lactis* code seulement pour 3 facteurs σ alors que celui de *Bacillus subtilis* en code 18 (Guedon *et al.*, 2002). *rpoD* est le gène codant pour le facteur σ végétatif tandis que ComX interviendrait dans la régulation de l'expression des gènes de compétence (Wydau *et al.*, 2006) alors que le rôle de SigX reste hypothétique. Toutefois, en tant que

facteur extracytoplasmique, ce dernier est probablement impliqué dans la régulation de fonctions spécifiques de la membrane, du périplasma ou encore de fonctions extracellulaires (Helmann, 2002). L'absence de facteur σ alternatif de réponse au stress est donc une particularité de *L. lactis*. De part cette originalité, l'étude de l'organisation des réseaux de régulation de réponse à différents stress revêt un intérêt particulier.

L'intervention d'autres niveaux de régulation de l'expression des gènes de *L. lactis* est également envisageable. Les pistes d'un contrôle par le (p)ppGpp (Srivatsan & Wang, 2008) ou encore des modifications post-traductionnelles (Soufi *et al.*, 2008) ont d'ailleurs récemment vu le jour.

III.2. Les réponses au stress de *Lactococcus lactis* caractérisées à l'heure actuelle

En tant qu'organisme modèle, *L. lactis* a depuis de nombreuses années fait l'objet de plusieurs études visant à caractériser son comportement physiologique en réponse à divers stress. Ces réponses ont été abordées et décrites à différents niveaux allant de la survie de la bactérie aux régulations contrôlant les mécanismes mis en place. Les réponses à différents stress ont même été comparées ; une étude transcriptomique utilisant des puces oligonucléotidiques partielles contenant seulement 375 gènes cibles répartis dans plusieurs catégories fonctionnelles (dégradation des protéines, métabolisme des sucres, des acides gras, des acides nucléiques et des acides aminés) compare les stress thermique, osmotique et acide chez *L. lactis* IL1403 (Xie *et al.*, 2004) et une revue plus générale répertorie les stress rencontrés par les bactéries lactiques et compare leurs réponses associées (van de Guchte *et al.*, 2002).

Dans cette partie, nous nous proposons de regrouper les connaissances acquises quant aux modifications physiologiques mises en place par *L. lactis* pour faire face à des conditions environnementales adverses. Les données présentées n'ont pas toutes été recueillies pour la souche IL1403 mais concernent néanmoins des souches de l'espèce *Lactococcus lactis*.

III.2.1. Stress thermique

Parmi les stress thermiques, on distingue le stress hypothermique (ou choc froid) du stress hyperthermique (ou choc chaud). L'une des premières publications a rapporté que *L. lactis*, qui est une bactérie mésophile et dont la température optimale de croissance se situe autour des 30 °C, était capable de survivre et même de croître jusqu'à des températures de 4 °C, même si c'était à un taux de croissance lent (inférieur à 0.4 h⁻¹) (Panoff *et al.*, 1994).

Essentiellement décrite par Wouters *et al.*, la réponse à un stress hypothermique entraîne à la fois la production de protéines spécifiques appelées Csp pour Cold Shock Protein (Wouters *et al.*, 1998) et une réduction de l'activité glycolytique (Wouters *et al.*, 2000). Le stress hyperthermique se caractérise quant à lui par l'induction de protéines participant à la fois au repliement et à la dégradation des protéines cellulaires, les chaperonnes, appelées Hsp pour Heat Shock Protein (Whitaker & Batt, 1991). L'étude menée avec les puces à ADN partielles a aussi révélé une répression des gènes codant pour le transport d'acides aminés et de peptides (système *opt*) et pour le système protéolytique (Xie *et al.*, 2004). Les mécanismes de régulation des stress thermiques ne sont pas complètement élucidés à ce jour mais impliquent probablement le système à deux composants constitué par l'histidine kinase KinC et le régulateur de réponse LlrC (Wouters *et al.*, 2001).

III.2.2. Stress osmotique

Le rôle du système à deux composants LlrD/KinD dans la perception du stress osmotique a précédemment été mis en évidence (O'Connell-Motherway *et al.*, 2000). Comme le stress hyperthermique, le stress osmotique (NaCl 2.5 %) induit la synthèse de chaperonnes (Kilstrup *et al.*, 1997) ce qui suggère qu'il affecte également l'intégrité des protéines. De plus, un changement d'osmolarité du milieu provoque une perte en eau de la cellule afin de maintenir une pression de turgescence constante (Csonka, 1989). L'accumulation de solutés compatibles ou osmolytes, physiologiquement inactifs, permet de maintenir la pression osmotique intracellulaire tout en évitant la perte d'eau. Ce maintien des concentrations évite donc de créer un déséquilibre intracellulaire. Chez *L. lactis*, c'est la bétaïne qui joue le rôle d'osmolyte puisque la tolérance au stress osmotique dépend fortement de l'activité de son système de transport (Obis *et al.*, 2001). Le régulateur BusR associé à ces transporteurs, joue donc un rôle important dans la réponse au stress osmotique (Romeo *et al.*, 2003). L'induction du système de transport de la bétaïne (*busAA*, *busAB*) a d'ailleurs été observée au niveau transcriptomique en réponse à la présence de 4 % de NaCl (Xie *et al.*, 2004). Cette dernière étude a aussi mis en évidence une activation des gènes codant pour la biosynthèse du peptidoglycane suggérant une modification de l'enveloppe cellulaire.

III.2.3. Stress oxydatif

Un bilan des connaissances concernant le stress oxydatif chez *L. lactis* a été dressé dans une revue de Myoshi *et al.* (2003). Parce qu'il cause des dommages aux niveaux métabolique et moléculaire, le stress oxydatif est probablement l'un des plus délétères pour la cellule. La formation de radicaux libres est en effet toxique car ces molécules sont susceptibles de réagir avec la plupart des composés organiques et en particulier d'altérer l'ADN (Miyoshi *et al.*, 2003). En ce sens, la protéine de réparation de l'ADN, RecA, va avoir un rôle de protection important. La présence de certains composés permet néanmoins de réduire la toxicité du stress oxygène. La glutathione protège en effet *L. lactis* contre le stress oxydatif, certainement grâce à l'activation du système glutathione réductase capable de catalyser la réduction du peroxyde d'hydrogène (Li *et al.*, 2003). De la même manière, une accumulation de thiorédoxine va avoir un effet protecteur (Vido *et al.*, 2005). D'autre part, la présence d'hème couplée à l'oxygène favorise le passage de la fermentation à la respiration (Rezaiki *et al.*, 2004). La consommation d'oxygène limite alors les effets du stress oxydatif.

Une étude transcriptomique complète comparant la fermentation anaérobie avec la fermentation aérobie et la respiration a d'ailleurs récemment permis de mieux comprendre l'effet de l'oxygène (Pedersen *et al.*, 2008). L'induction de l'expression des NADH oxydases confirme la transition métabolique décrite au niveau du pyruvate (Cocaign-Bousquet *et al.*, 1996) puisque le NADH consommé par les NADH oxydases n'est plus disponible pour la lactate déshydrogénase (Ldh) (figure 4). En outre, la pyruvate formate lyase (Pfl) n'est active qu'en condition anaérobie alors la pyruvate déshydrogénase (Pdh) prend le relai en présence d'oxygène (Henriksen & Nilsson, 2001) ; la conversion du pyruvate en éthanol est donc inhibée et les productions de diacétyle et d'acétate favorisées (figure 4). D'autre part, l'atténuation du stress par la respiration a été confirmée par la surexpression spécifique en culture aérée, mais pas lors de la respiration en présence d'hème, de gènes dont les enzymes sont connues pour leur rôle protecteur dans les réponses aux stress tels *ahpCF*, *gshR*, *hslB*, *noxE*, *osmC*, *pmsX*, *qor*, *sodA*, ou *ygmK* (Pedersen *et al.*, 2008).

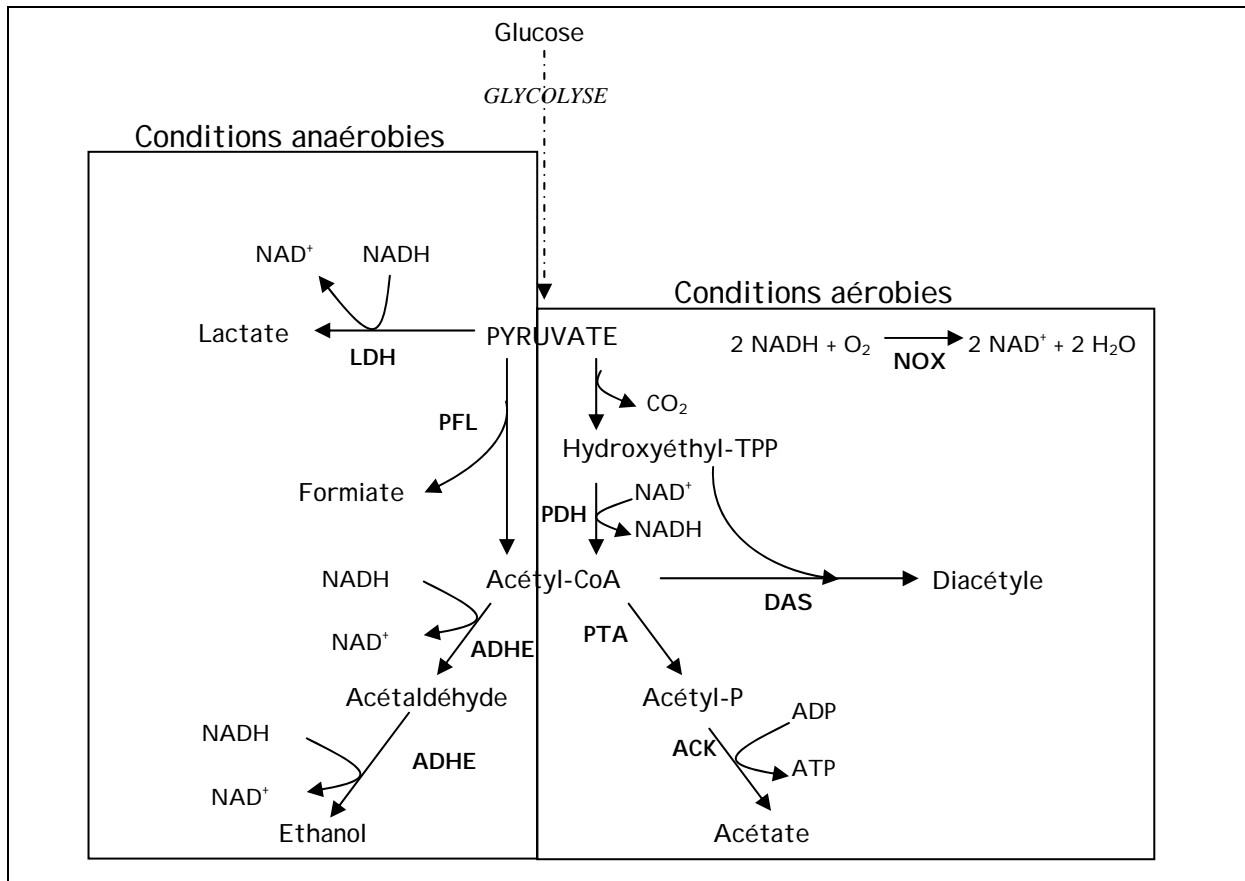


Figure 4 : Métabolisme du pyruvate en fonction des conditions d'aération.

LDH : lactate déshydrogénase, PFL : pyruvate formate lyase, ADHE : alcool déshydrogénase, NOX : NADH oxydase, PDH : pyruvate déshydrogénase, PTA : phosphotransacétylase, ACK : acétate kinase, DAS : diacétyle synthase.

Le système à deux composants LlrF/KinF a été proposé comme senseur de stress oxydatif (O'Connell-Motherway *et al.*, 2000). Les protéines Flp semblent quant à elles impliquées dans la résistance à ce stress vraisemblablement *via* le contrôle de l'assimilation de Zn^{2+} , un ion qui pourrait protéger les groupes thiols de l'oxydation (Miyoshi *et al.*, 2003). Le transport des ions métalliques en général pourrait d'ailleurs jouer sur la sensibilité au stress oxydatif de *L. lactis* (Turner *et al.*, 2007). En outre, CcpA étant impliqué dans la régulation de la respiration (Gaudu *et al.*, 2003), il va forcément avoir aussi un rôle dans la réponse au stress oxydatif.

III.2.4. Stress acide

Les capacités acidifiantes des bactéries lactiques les confrontent fréquemment à des stress acides. En ce sens, déchiffrer les réponses engendrées a fait l'objet de nombreuses études, ce qui en fait certainement le stress le mieux caractérisé chez *L. lactis*.

III.2.4.1. Réponse physiologique au stress acide

La production d'acide lactique conduit à une acidification progressive du cytoplasme mais cette bactérie neutrophile, dont le pH optimal se situe aux alentours de 6.6, est capable de se développer à des pH intracellulaires avoisinant 5.5 (Mercade *et al.*, 2000). Par contre, sa viabilité est très fortement affectée, sinon perdue, par l'imposition de pH inférieurs à 4 (Rallu *et al.*, 1996), car elle devient incapable de maintenir un pH intracellulaire compatible avec les processus biologiques. Une étude globale de la réponse de *L. lactis* à sa propre autoacidification parue en 2002, rapporte une forte diminution de l'activité glycolytique liée à la perte d'activité des enzymes en raison de l'acidification du cytoplasme (Even *et al.*, 2002). En parallèle, le ralentissement du métabolisme carboné ainsi que la lutte contre l'acidification du cytoplasme provoquent une déficience énergétique et un ralentissement de la croissance. Bien que réduite, l'activité glycolytique se poursuit même après l'arrêt de la croissance grâce au maintien de la transcription des gènes et à la stabilisation des transcrits concernés. L'activité de traduction semble également maintenue puisque la concentration intracellulaire de certaines protéines est accrue, probablement pour compenser leur perte d'activité. Dans leur étude partielle du transcriptome, Xie *et al.* (2004) ont observé une répression du catabolisme des acides aminés avec la diminution du niveau des transcrits impliqués dans le transport des acides aminés et des peptides tout comme de ceux liés au système protéolytique. Cette étude a aussi révélé l'induction du système de transport de la bêtaïne comme lors du stress osmotique. Au niveau protéomique, Frees *et al.* (2003) ont mis en évidence la surexpression de plusieurs chaperonnes (DnaK, GroEL, GroES, GrpE) et protéases (ClpE, ClpP) ainsi que de protéines de stress (SodA, AhpC et LuxS) dont les expressions sont respectivement sous le contrôle de HrcA (Guedon *et al.*, 2002) et CtsR (Varmanen *et al.*, 2000). Ces résultats ont d'ailleurs été confirmés et complétés lors d'une étude protéomique plus récente (Budin-Verneuil *et al.*, 2005).

III.2.4.2. Mécanismes mis en place pour répondre au stress acide

Une étude phénotypique de mutants des systèmes à deux composants indique que LlrA/KinA et LlrB/KinB pourraient être impliqués dans la lutte contre le stress acide *via* le métabolisme de l'arginine (O'Connell-Motherway *et al.*, 2000). Une activation de la voie arginine déiminase (ADI) a d'ailleurs été observée en réponse à l'autoacidification de *L. lactis* cultivé en milieu synthétique (Even *et al.*, 2002) comme lors de sa croissance sur lait (Raynaud *et al.*, 2005). Cette voie couple la désamination de l'arginine en ornithine à un

système d'antiport ainsi qu'à la production d'ammoniaque et d'ATP (figure 5). L'ATP est alors disponible pour l'ATPase assurant l'expulsion de protons et minimisant ainsi l'acidification du cytoplasme (Koebmann *et al.*, 2000). Le fonctionnement de ce mécanisme de lutte contre le stress est d'ailleurs corroboré par la diminution du rendement en biomasse par rapport à la production d'ATP lors de l'autoacidification de *L. lactis* confirmant bien un détournement de l'utilisation d'ATP (Even *et al.*, 2002). A partir d'un certain degré d'acidification, l'activité de l'ATPase est néanmoins réduite ce qui peut s'expliquer par des pertes d'activité, soit de l'ATPase directement, soit des enzymes des voies métaboliques assurant la production d'ATP. Un contrôle impliquant la phosphofructokinase (Pfk) et/ou des enzymes liées au métabolisme des phosphoglycérates (Pgm, Pfk, EnoA) pourrait effectivement expliquer ce phénomène (Even *et al.*, 2002; Raynaud *et al.*, 2005). Une étude récente, basée sur de la modélisation, laisse supposer que la réduction du flux glycolytique n'est pas uniquement liée à une perte d'activité enzymatique mais est aussi limitée par le système PTS de transport des sucres. L'activité PTS est en effet réduite en raison de la faible disponibilité du phosphoenolpyruvate qui est très rapidement converti en pyruvate, ce qui génère de l'ATP et permet ainsi de réduire la limitation énergétique (Andersen *et al.*, 2009). La glutathione aurait également un effet protecteur de stress acide sur les constituants cellulaires de *L. lactis* mais son mécanisme d'action reste encore hypothétique (Zhang *et al.*, 2007).

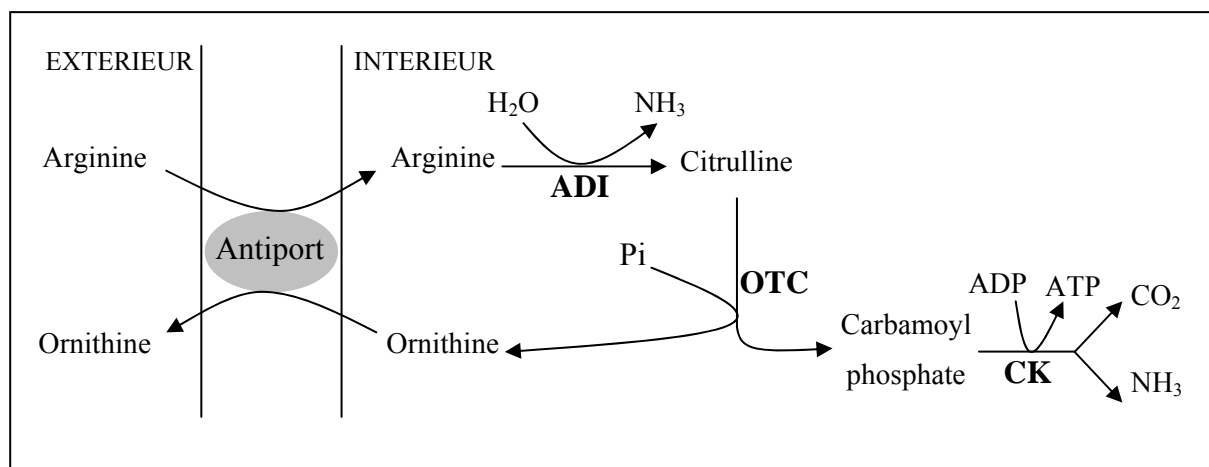


Figure 5 : Voie de l'arginine déiminase chez *Lactococcus lactis*. ADI : arginine déiminase, OTC : ornithine carbamoyltransférase, CK : carbamate kinase.

III.2.5. Carence en carbone

Bien que tout aussi importants pour *L. lactis*, les stress nutritionnels sont bien moins étudiés que les stress physico-chimiques. Hormis une analyse protéomique de la carence en nucléotides (Beyer *et al.*, 2003), la réponse à la carence en carbone est d'ailleurs quasiment le seul stress nutritionnel caractérisé chez cette bactérie.

III.2.5.1. Réponse physiologique à la carence en carbone

Alors que les stress physico-chimiques semblent permettre la croissance tant que des niveaux létaux de stress ne sont pas atteints, les stress nutritionnels de type carence en carbone provoquent un ralentissement brutal non seulement du flux glycolytique mais plus généralement du métabolisme et de la croissance mais ne causent pas de dommages majeurs dans la cellule. D'autre part, une carence prolongée en sucre rend *L. lactis* non cultivable mais ne détruit pas pour autant son activité métabolique : les gènes impliqués dans le métabolisme des sucres, la division cellulaire et l'autolyse sont réprimés mais la transcription, le métabolisme des acides aminés et la production d'énergie sont maintenus (Ganesan *et al.*, 2007). Dans une étude à l'échelle génomique de l'adaptation progressive de *L. lactis* à la carence en glucose, Redon *et al.* (2005b) ont mis en évidence trois types de réponse. En premier lieu, la réponse globale se caractérise par la diminution des transcrits codant pour des protéines associées à la croissance cellulaire ce qui se traduit par une diminution des processus de transcription, de traduction et de division cellulaire. Une réponse plus spécifique de la carence en carbone se met également en place avec la répression des gènes codant pour les enzymes de la glycolyse (qui ne semble toutefois pas se répercuter au niveau protéique) et une activation de ceux codant pour l'utilisation des sucres alternatifs. La voie arginine déiminase est aussi activée au niveau transcriptionnel et pourrait permettre de pallier au déficit énergétique inhérent à la réduction de l'activité glycolytique. Enfin, des réponses plus inattendues ont été observées, comme l'induction de gènes de transports divers, de compétence, phagiques et prophagiques.

III.2.5.2. Régulation de la réponse à la carence en carbone

Les mécanismes de régulations impliqués dans le contrôle du métabolisme des sucres sont recensés dans la revue de Kowalczyk et Bardowski (2007). Ils se divisent globalement en deux groupes selon qu'ils sont de nature biochimique (modulation des activités enzymatiques) ou génétique (induction ou répression des gènes). La régulation la plus connue et la plus étudiées de toutes reste néanmoins la répression catabolique par CcpA, car c'est visiblement

un mécanisme de contrôle commun à l'ensemble des bactéries Gram positives (Titgemeyer & Hillen, 2002). La détermination du régulon CcpA par une étude transcriptomique comparative entre la souche *L. lactis cremoris* MG1363 et le mutant $\Delta ccpA$ correspondant a d'ailleurs révélé l'étendue du contrôle catabolique par CcpA (Zomer *et al.*, 2007). Loin de se limiter au métabolisme carboné, celui-ci semble en effet s'étendre au métabolisme des bases et de l'azote et pourrait même jouer sur le pH intracellulaire au travers du contrôle de la voie ADI (figure 5). Une étude à l'échelle globale de la stabilité des transcrits en réponse à la carence en carbone a mis à jour un mode de régulation de l'expression des gènes jusqu'à présent peu exploré chez les bactéries lactiques : la modulation de la dégradation des messagers (Redon *et al.*, 2005a).

IV. Positionnement, objectifs et démarche de la thèse

En raison de sa relative simplicité (petit génome, absence de compartimentation, métabolisme simple) par rapport aux autres microorganismes modèles que sont *E. coli*, *B. subtilis* et *S. cerevisiae*, *Lactococcus lactis* se pose comme un modèle intéressant en biologie intégrative. Comprendre son fonctionnement global revêt d'autant plus d'intérêt qu'il est utilisé en industrie dans les procédés de fabrication de produits laitiers fermentés et se présente aussi comme un vecteur de choix pour la délivrance de molécules thérapeutiques en raison de son innocuité. Dans son environnement naturel comme lors de son exploitation industrielle, *L. lactis* est confronté à diverses contraintes environnementales (stress pH, osmotique, thermique, carences nutritionnelles). Son adaptation à la carence en azote n'a encore jamais été caractérisée de façon complète. Le catabolisme des acides aminés chez les bactéries lactiques est souvent étudié car il permet la production de composés aromatiques qui modifient les propriétés organoleptiques des produits fermentés (Fernandez & Zuniga, 2006). Le rôle des acides aminés est pourtant bien plus étendu puisqu'ils sont également impliqués dans la production d'énergie, le contrôle du pH intracellulaire et la régénération des cofacteurs (Ardo, 2006). De plus, les acides aminés sont des constituants cellulaires majeurs que les bactéries lactiques, souvent auxotrophes (Cocaign-Bousquet *et al.*, 1995), puisent dans le milieu de culture.

Il est important de noter que les stress, quelle que soit leur nature, s'accompagnent d'une réduction plus ou moins importante de la vitesse de croissance. L'impact de cette diminution du taux de croissance sur la physiologie cellulaire n'est jamais dissocié de la réponse au stress lors de cultures discontinues. Pourtant, deux études en chémostat menées sur la levure modèle *S. cerevisiae* ont récemment montré son importance : la seule variation du taux de croissance peut en effet entraîner la modification de l'expression de 27 à 50 % des gènes (Castrillo *et al.*, 2007; Regenber *et al.*, 2006). A ce jour, aucune étude semblable n'a encore été publiée pour un procaryote et l'effet que peut avoir la variation du taux de croissance sur la physiologie cellulaire des bactéries reste à déterminer.

L'objet de cette thèse vise à mettre en place une démarche de biologie intégrative pour étudier l'adaptation de *L. lactis* à son environnement, en particulier au travers des processus de régulation des concentrations en protéines, les effecteurs majeurs de la physiologie cellulaire. Différents phénomènes impliqués dans la régulation des concentrations protéiques intracellulaires, à savoir, la traduction, mais aussi la dilution par la croissance et la dégradation des protéines seront intégrés pour obtenir une vision cohérente de l'adaptation. Nous nous intéressons à la réponse de *L. lactis* à la carence en acide aminé et plus particulièrement à la carence en isoleucine car c'est un acide aminé central, particulièrement abondant dans les protéines, notamment dans les régions hydrophobes. Deux conditions environnementales, maîtrisées et contrôlées, correspondant à un état de stress et à un état adapté, seront étudiées. La carence en isoleucine sera imposée lors d'une culture discontinue par la consommation progressive de cet acide aminé par la bactérie. Une culture continue à différents taux de dilution en chémostat limité en isoleucine permettra la détermination spécifique de l'effet du taux de croissance sur la physiologie de cette bactérie. Différentes techniques d'analyses seront employées afin de caractériser différents niveaux biologiques des réponses observés. L'ensemble des données recueillies sera ensuite analysé de façon intégrée grâce à l'utilisation d'outils mathématiques adaptés.

Ce projet s'intègre dans le programme fédérateur agroBI, un programme INRA de biologie intégrative visant à améliorer la compréhension globale des microorganismes par une approche pluridisciplinaire. Quatre partenaires, deux biologistes et deux mathématiciens, ont participé à la réalisation de ces travaux. Les cultures de *L. lactis* et les analyses transcriptomiques ont été réalisées au sein du laboratoire de biotechnologie et bioprocédés (LISBP, UMR CNRS/INSA 5504 - UMR INRA/INSA 792) sur le site de l'INSA de Toulouse

tandis que la partie protéomique a été menée à l'INRA de Jouy en Josas dans l'unité de Biochimie Bactérienne (BioBac, UR477). Le traitement des données s'est ensuite poursuivi au LISBP avec l'appui de l'Institut de Mathématiques de Toulouse. Enfin, l'aspect modélisation a été développé au LISBP toujours en collaboration avec l'IMT mais aussi avec le groupe Méthodes et Algorithmes en Commandes (MAC) du Laboratoire d'Analyse et d'Architecture des Systèmes de Toulouse (LAAS-CNRS).

Matériel et Méthodes

I. Souche et conditions de culture

I.1. Souche

L'ensemble des expérimentations présentées dans ce manuscrit ont été conduites sur la bactérie lactique *Lactococcus lactis* sous-espèce *lactis* IL1403. Initialement isolée en milieu laitier cette souche a été curée de ses plasmides et est couramment utilisée comme référence car elle est entièrement séquencée depuis 2001 ((Bolotin *et al.*, 2001), numéro d'accèsion GenBank AE005176).

I.2. Suivi des cultures

L'évolution de la biomasse au sein des bioréacteurs a été suivie par mesure de spectrophotométrique de l'absorbance à 580 nm. Une corrélation DO/poids sec de 0,3 g/L par unité de DO a ensuite été appliquée pour remonter à la concentration.

Le suivi cinétique des concentrations en substrats et produits a été effectué par chromatographie liquide haute pression (HPLC) tandis que les acides aminés ont été mesurés par AminoQuant HP1090 comme décrit dans le premier article de cette thèse (Dressaire *et al.*, 2008).

I.3. Culture discontinue (batch)

Afin d'étudier la réponse de *L. lactis* IL1403 à la carence en isoleucine, cette bactérie a été cultivée lors de fermentations de type batch permettant l'imposition progressive de la carence. Les détails du milieu utilisés et du mode de culture sont disponibles dans la littérature (Redon *et al.*, 2005a). Brièvement, des réacteurs de 2L équipés de modules de régulation permettant de mesurer et éventuellement de contrôler les paramètres de cultures (température, agitation, pH) ont été utilisés. Le milieu de culture initial est dérivé du milieu chimiquement défini contenant l'ensemble des nutriments nécessaires à la croissance de *L. lactis* (CDM, (Poolman & Konings, 1988)) ; sa principale différence résidant dans le fait qu'il contient dix fois moins d'acides aminés branchés (valine, leucine et isoleucine, tableau 1), ce qui permet de provoquer une carence en isoleucine avant la carence en glucose.

Tableau 1 : Composition des différents milieux de culture utilisés.

| produit | concentrations (g/L) dans le milieu MCD original | concentrations (g/L) dans le milieu utilisé en batch | concentrations (g/L) dans le milieu utilisé en chémostat |
|---------------------------------------|---|---|---|
| sucré (glucose) | 10 | 10 | 20 |
| acétate de sodium | 1 | 1 | 1 |
| citrate d'ammonium | 0.6 | 0.6 | 0.6 |
| KH ₂ PO ₄ | 3 (9) | 3 (9) | 3 (9) |
| K ₂ HPO ₄ | 2,5 (7,5) | 2,5 (7,5) | 2,5 (7,5) |
| MgCl ₂ , 6H ₂ O | 0.2 | 0.2 | 0.2 |
| FeSO ₄ , 7H ₂ O | 0.011 | 0.011 | 0.011 |
| CaCl ₂ , 2H ₂ O | 0.05 | 0.05 | 0.05 |
| ZnSO ₄ , 7H ₂ O | 0.005 | 0.005 | 0.005 |
| COCl ₂ , 6H ₂ O | 0.0025 | 0.0025 | 0.0025 |
| alanine | 0.24 | 0.24 | 0.24 |
| arginine | 0.12 | 0.12 | 0.12 |
| asparagine | 0.34 | 0.34 | 0.34 |
| glutamine | 0.51 | 0.51 | 0.51 |
| glycine | 0.17 | 0.17 | 0.17 |
| histidine | 0.11 | 0.11 | 0.11 |
| lysine | 0.35 | 0.35 | 0.35 |
| methionine | 0.12 | 0.12 | 0.12 |
| proline | 0.68 | 0.68 | 0.68 |
| sérine | 0.34 | 0.34 | 0.34 |
| thréonine | 0.23 | 0.23 | 0.23 |
| tryptophane | 0.05 | 0.05 | 0.05 |
| isoleucine | 0.2 | 0.02 | 0.025 |
| leucine | 0.47 | 0.047 | 0.05875 |
| valine | 0.33 | 0.033 | 0.04125 |
| cystéine | 0.17 | 0.17 | 0.17 |
| phénylalanine | 0.28 | 0.28 | 0.28 |
| tyrosine | 0.29 | 0.29 | 0.29 |
| adénine | 0.01 | 0.01 | 0.05 |
| guanine | 0.01 | 0.01 | 0.05 |
| uracile | 0.01 | 0.01 | 0.05 |
| xanthine | 0.01 | 0.01 | 0.05 |
| acide P-aminobenzoïque | 0.01 | 0.01 | 0.01 |
| biotine | 0.01 | 0.01 | 0.01 |
| cyano-cobalamine (B12) | 0.001 | 0.001 | 0.001 |
| acide folique | 0.001 | 0.001 | 0.01 |
| inosine | 0.005 | 0.005 | 0.025 |
| acide nicotinique | 0.001 | 0.001 | 0.001 |
| acide orotique | 0.005 | 0.005 | 0.005 |
| panthoténate de Ca | 0.001 | 0.001 | 0.001 |
| pyridoxamine | 0.005 | 0.005 | 0.005 |
| pyridoxine (B6) | 0.002 | 0.002 | 0.002 |
| riboflavine (B2) | 0.001 | 0.001 | 0.001 |
| thiamine | 0.001 | 0.001 | 0.001 |
| acide D,L-6,8-thioctique | 0.0025 | 0.0025 | 0.0025 |
| thymidine | 0.005 | 0.005 | 0.025 |

I.4. Culture continue (chémostat)

Les cultures continues sont réalisées dans des fermenteurs de 300 mL équipés avec des électrodes de niveaux permettant d'y conserver un volume utile de 200 mL. Un bain Marie à circulation d'eau branché sur la double enveloppe du chémostat permet de maintenir une température de 30°C au sein du fermenteur. Le pH est régulé à 6,6 par ajout automatique de KOH 10 N. Une surpression d'azote est appliquée au réacteur ainsi qu'au bidon d'alimentation tout au long de la culture afin de maintenir une atmosphère dépourvue d'oxygène.

Le fermenteur et le bidon d'alimentation sont d'abord autoclavés (121 °C, 20 min) deux fois avec de l'eau. Puis le milieu est introduit stérilement dans le bidon d'alimentation par filtration à travers une cartouche Sartorius 0,2 µm, lequel est ensuite dégazé à l'azote pendant 30 minutes. Le chémostat est également dégazé à l'azote et vidé stérilement de son eau avant d'être branché sur le bidon d'alimentation pour être rempli.

Le milieu d'alimentation utilisé est également dérivé du milieu chimiquement défini (Poolman & Konings, 1988) et a été dimensionné de manière à ce que l'isoleucine soit le seul élément limitant : les concentrations en acides aminés branchés ont été réduites mais celle en glucose a été doublée (tableau 1) .

II. Transcriptome

Le protocole utilisé pour les expérimentations de transcriptome est celui décrit par Redon *et al.* (2005a) et les principales étapes sont rappelées ici. Un minimum de trois répliques biologiques est respecté pour chacune des conditions.

II.1. Description du support

Afin de permettre la comparaison simultanée de plusieurs échantillons correspondant à différentes conditions, un support de type membrane de nylon plutôt que lame de verre a été adopté. 1949 produits de PCR fournis par la société Eurogentech et couvrant environ 85 % du génome de *L. lactis* IL1403 sont spottés en duplicates sur chaque membrane de nylon chargée positivement.

II.2. Extraction et quantification de l'ARN total

Lorsque les conditions de croissance à examiner sont atteintes, les cellules fraîches sont directement immergées dans de l'azote liquide avant de subir un cassage mécanique par des billes de verre dans du phénol acide. Intervient ensuite une purification sur colonne (RNeasy Midi Kit, Quiagen) incluant un traitement à la DNase. La quantification de l'ARN total extrait se fait par mesure de l'absorbance à 260 nm. Les concentrations moyennes en ARN total obtenues pour les différents points de prélèvements présentés dans ce manuscrit sont récapitulées dans le tableau 2. La qualité de l'ARN est vérifiée soit par électrophorèse en conditions dénaturantes soit avec le kit commercial Bioanalyzer d'Agilent Technologies.

Tableau 2 : Concentrations en ARN total et rendement d'extraction moyens obtenus aux différents taux de croissance pour les différentes cultures.

| type de culture | taux de croissance (h ⁻¹) | [ARN] µg/µl | écart sur [ARN] (%) | rendement d'extraction (%) | écart sur extraction (%) |
|-----------------|---------------------------------------|-------------|---------------------|----------------------------|--------------------------|
| batch | 0.88 | 1.56 | 4.45 | 11.06 | 2.12 |
| | 0.34 | 1.18 | 17.88 | 8.74 | 17.45 |
| | 0.06 | 1.03 | 12.40 | 6.95 | 11.01 |
| | 0.04 | 0.95 | 3.34 | 6.95 | 5.35 |
| chémostat | 0.09 | 0.51 | 9.78 | 3.58 | 11.01 |
| | 0.24 | 0.70 | 12.78 | 4.92 | 10.62 |
| | 0.35 | 0.96 | 7.64 | 6.90 | 6.22 |
| | 0.47 | 0.98 | 3.73 | 7.34 | 3.83 |

II.3. Rétro-transcription, marquage, hybridation et détection

L'étape de rétro-transcription est menée sur une quantité constante d'ARN total en présence d'un marqueur radioactif : le dCTP-[³³P] (Amersham Biosciences). Les ADNc radioactifs sont ensuite dénaturés puis leur hybridation sur les membranes de nylon s'effectue sur une nuit dans un four à hybridation à 68 °C. Après lavage, les membranes sont exposées à un écran autoradiographique 3 jours durant. Enfin, les écrans sont scannés sur un Phosphoimager (Storm 860, Molecular Dynamics). La détection spots et la quantification des images sont réalisées à la Plateforme Biopuce de Toulouse (<http://biopuces.genotoul.fr/>).

II.4. Normalisation des données

Selon l'utilisation à laquelle elles sont destinées, deux types de normalisation des données sont pratiqués.

II.4.1. Abondance

Le mode de normalisation le plus couramment utilisé dans la littérature consiste à corriger l'intensité des spots correspondant à un gène par l'intensité totale de la membranes, comme précédemment décrit (Redon *et al.*, 2005a). Cette méthode permet d'observer l'abondance d'un messager au sein de l'ensemble des ARNm de la cellule. Par la suite, l'analyse des variations de niveau entre deux conditions expérimentales sera basée sur ces données d'abondance.

II.4.2. Concentration

L'autre procédé de normalisation consiste à tenir compte de l'intensité totale de la membrane de référence, qui sert de témoin de marquage, mais également de la variation de l'ARN total au sein de la cellule. En effet, cette quantité peut varier de façon significative selon les conditions de culture (tableau 2) et refléter des différences de concentration intracellulaire. Ce traitement des données, précédemment décrit par Redon *et al.* (2005a), permet de se rapprocher des concentrations cellulaires de chacun des ARNm mesurés. Les données ainsi normalisées sont utilisées lorsqu'elles sont comparées à d'autres données biologiques à savoir dans toutes les expériences de modélisation décrites dans ce manuscrit.

III. Protéome

Le protocole utilisé pour les expérimentations de protéome dont les principales étapes sont rappelées ici et décrit plus en détails dans plusieurs publications soumises (Dressaire *et al.* 2009a et 2009b, Beganovic *et al.* 2009). Trois répliques biologiques sont au minimum réalisés pour chacune des conditions.

III.1. Prélèvements

Lorsque les conditions à analyser sont atteintes, les échantillons prélevés sont immédiatement centrifugés 10 minutes à 4 °C et 8 000 rpm. Le culot de bactéries est lavé dans 100 mL de tampon phosphate de sodium (0.2 mol/L, pH 6.4) EDTA (1mM) préalablement refroidi à 4°C puis centrifugé 10 min à 4 °C et 8 000 rpm. Le culot cellulaire

est re-suspendu dans 5mL de tampon de lavage (phosphate de sodium EDTA) puis congelé et conservé à -20 °C avant d'être expédié à l'équipe BioBac de l'INRA de Jouy en Josas où sont effectuées les analyses protéomiques.

III.2. Extraction des protéines

Les culots cellulaires sont repris dans du tampon NaPO₄ (20 mM) à pH 6.4 contenant 10 mM de tributylphosphine, un cocktail d'inhibiteur de protéase (dilué au 1/20) et de la catalase à 40 U/mL, afin de limiter la formation d'isoformes. Après cassage mécanique au BASIC Z cell disrupter (Celld, Warwickshire, United Kingdom), l'extrait protéique est centrifugé 20 min à 4 °C et 5000 rpm afin d'éliminer les gros débris cellulaires. La fraction cytosolique (surnageant) est ultra-centrifugée (220000g, 30 min, 4 °C) avant dosage des protéines totales par la méthode de Bradford avec le réactif de BioRad puis congélation à -20 °C.

III.3. Séparation sur gel d'électrophorèse

III.3.1. Isoélectrofocalisation

La fraction cytosolique est incubée 30 min à 37 °C avec de la RNase (25 U pour 100 µL) puis précipitée à l'éthanol (75 % V/V) et resuspendue dans du tampon d'électrophorèse (urée 7 M, thio-urée 2 M, CHAPS urée 4 %, dithiothreitol 100 mM (ou tributylphosphine 4 mM et DeStreak (1.2 % v/v, Amersham Biosciences, GE Healthcare), pour les gels basiques), et du tampon IPG (pour Immobilized pH Gradient) au pH approprié (Amersham Biosciences, GE Healthcare). La séparation se fait sur des gels de polyacrylamide commerciaux contenant des gradients de pH immobilisés. La migration se fait dans un plateau IEF Cell ou un sarcophage IPG PHOR. Les gels sont ensuite lavés avec une solution de Tris HCl 2 M pH 8.8, SDS 20 %.

III.3.2. Electrophorèse SDS-page

Les protéines sont chargées négativement par le SDS et sont séparées par migration à 15 °C et 1 W sur la nuit, dans un maillage de gel de polyacrylamide (Acrylamide 30%, Tris 2M pH 8.8, SDS 10 %, Temed, APS 10%) déposé dans une cuve de migration Ettan DALT (Amersham) et recouvert de tampon TGS (Tris-Glycine-SDS, solution commerciale concentrée). Les gels sont rincés dans un bain d'eau miliQ puis colorés à l'aide du bleu de Coomassie colloïdal et rincés à l'eau déionisée jusqu'à disparition de la coloration de fond.

III.4. Détection et quantification des spots

Les gels sont numérisés en 256 niveaux de gris (8 bits par pixel) avec un scanner Epson expression 1640XL contrôlé par le logiciel Siver Fast. Manipulations et analyses d'images sont réalisées avec le logiciel Phoretix 2D. Les volumes de chaque spot sont calculés en unité arbitraire : aire du spot * intensité en pixel - bruit de fond.

III.5. Identification des protéines

L'identification des protéines se fait sur le Plateau d'Analyse Protéomique par Séquençage et Spectrométrie de masse (PAPSS), à l'INRA de Jouy en Josas.

III.5.1. Spectrométrie de masse MALDI-TOF

Les spots à identifier sont découpés et lavés dans une solution d'acétonitrile-hydrogénocarbonate (50/50) avant digestion à la trypsine. Le surnageant est associé à de la matrice composée d'acide α -cyano-4-hydroxycinnamique à 4mg/mL dans du TFA 0.1%, et de l'acétonitrile 50 %. L'analyse est ensuite directement réalisée par spectrométrie de masse MALDI-TOF sur l'appareil Voyager DE STR Instrument (Applied Biosystems, Framingham, CA, USA). La calibration interne se fait avec les peptides à 842.5 et 2211.1 Da. L'identification se fait alors par recherche directe dans les bases de données spécifiques de *L. lactis* ou à l'aide du logiciel MS-Fit (<http://prospector.ucsf.edu>).

III.5.2. LC-MS/MS

Les protéines qui n'ont pas pu être identifiées par MALDI-TOF sont analysées par un système Ultimate 3000 LC (Dionex, Voisins le Bretonneux, France) connecté à spectromètre de masse (LTQ, Thermo Fisher, USA). Le gel polyacrylamide est préalablement extrait des digestions triptyques par une solution d'acide formique (0.1 %) acétonitrile (50 %). Le surnageant est alors mélangé à du tampon de charge (TFA 0.08 %, acétonitrile 2 %) puis déposé dans la pré-colonne en amont de la nano-colonne Pepmap C18 (0.075 x 150 mm, 100 Å, 3 μ m, Dionex). L'ionisation est assurée au sein d'un capillaire (PicoTip EMITER 10 μ m ID; New Objective, USA). L'identification des protéines se fait grâce au logiciel Bioworks 3.3 TM.

III.6. Traitement des données

III.6.1. Normalisation et traitement des spots

Le volume de chaque spot est normalisé par l'intensité totale de la membrane correspondante afin de s'affranchir des variations de coloration entre répétitions. Sur un gel donné, les différents spots correspondants à une même protéine sont sommés. Les protéines identifiées sur deux gammes de pH différentes (4.5-5.5 et 5-6 par exemple) sont analysées comme deux protéines différentes. Les spots mixtes, au sein desquels plusieurs protéines ont été identifiées, ne sont pas retenus pour la suite des analyses.

III.6.2. Mesure des protéines totales

La concentration en protéine totale est déterminée par la méthode de Lowry *et al.* (1951) après solubilisation de protéines dans du NaOH 1 M (10 min à 90 °C). Cette méthode est utilisable pour des concentrations en protéine comprises entre 0,01 à 0,1 g/L. La gamme étalon est, pour chaque dosage, réalisée à partir d'une solution de BSA (Bovine Serum Albumine) à 0,2 g /L. Le protocole de dosage est le suivant: 1 mL d'extrait à doser, convenablement dilué, est incubé pendant 10 minutes à l'obscurité avec 3 ml de réactif (50 mL de solution de NaOH (0,1 N) avec 2 % (P/V) de Na₂CO₃ et 0,02 % (P/V) de tartrate de Na, K +1 mL de solution de CuSO₄, 5H₂O (0,5 % (P/V)), contenant une goutte de H₂SO₄ concentré pour 100 mL de solution). 0,3 ml de réactif de Folin-Ciocalteu dilué deux fois, sont ensuite ajoutés. Le mélange est agité et incubé 30 minutes à l'obscurité avant de lire l'absorbance à 750 nm.

Tableau 3 : Concentrations moyennes en protéines totales aux différents taux de croissance pour les différentes cultures.

| type de culture | taux de croissance (h ⁻¹) | [protéine] g de protéine/g de biomasse | écart sur [protéine] (%) |
|-----------------|---------------------------------------|--|--------------------------|
| batch | 0.88 | 0.40 | 14 |
| | 0.34 | 0.40 | 21 |
| | 0.06 | 0.38 | 50 |
| | 0.04 | 0.38 | 12 |
| chémostat | 0.09 | 0.42 | 11 |
| | 0.24 | 0.44 | 7 |
| | 0.35 | 0.52 | 33 |
| | 0.47 | 0.40 | 18 |

Le dosage des protéines totales nous a permis de constater que, contrairement à la concentration en ARN total, celle en protéines totales était constante quel que soit le taux de croissance ou les conditions de culture (tableau 3). En conséquence, abondances et concentrations de protéines sont équivalentes.

IV. Mesure de la stabilité des ARN messagers

Déjà décrite dans la littérature (Redon *et al.*, 2005a), la mesure de la stabilité des transcrits repose sur le blocage de la transcription par un antibiotique (la rifampicine). La mesure de la cinétique de disparition de chacun ARNm par une technique transcriptomique permet ensuite d'accéder directement à leur constante de dégradation. Au minimum trois répétitions indépendantes sont toujours réalisées. La détection des spots se fait comme pour le transcriptome mais les analyses se font à partir des données brutes non normalisées. Ces données ne sont cependant exploitées que lorsqu'elles présentent une intensité supérieure au seuil de détection fixé à partir des intensités des spots vides. 4 points de cinétiques de dégradation correspondant à divers instants après le blocage de la transcription sont ensuite choisis de manière à observer une décroissance significative des intensités associées à chaque messenger à ces différents temps. Puis des régressions linéaires permettent de déterminer les temps de demi-vie des transcrits. Seuls les résultats associés à des coefficients de détermination (R^2) supérieurs à 0.7 sont considérés.

La détermination des temps de demi-vie requiert d'effectuer des prélèvements après le blocage antibiotique de la transcription en respectant une échelle de temps adaptée au suivi cinétique de la disparition des transcrits. Les mesures doivent en effet être effectuées sur un intervalle de temps proche des temps de demi-vie, *a priori* inconnus, d'un maximum d'ARN messagers : des plages de mesures trop courtes ne permettraient pas de visualiser la dégradation de manière significative tandis que si elles sont trop étendues le taux de dégradation risque d'être trop important et les intensités des spots représentant chacun des transcrits pourraient descendre en dessous du seuil de détection.

Afin de vérifier que des prélèvements sur une durée de 20 minutes seraient suffisants, les stabilités globales de l'ensemble des transcrits ont d'abord été. Cette estimation a été faite à partir des intensités totales de chacune des membranes de nylon correspondant à un instant de prélèvement. Le tracé de la diminution de cette intensité en fonction du temps (figure 6), permet de valider l'étendue de la zone de prélèvement car la cinétique de disparition y reste linéaire ($R^2 \geq 0.7$). D'autre part, les pentes obtenues restent relativement constantes pour les différentes répétitions ce qui souligne la reproductibilité des expériences (seule la première diffère des autres pour le cas présenté ci-dessous).

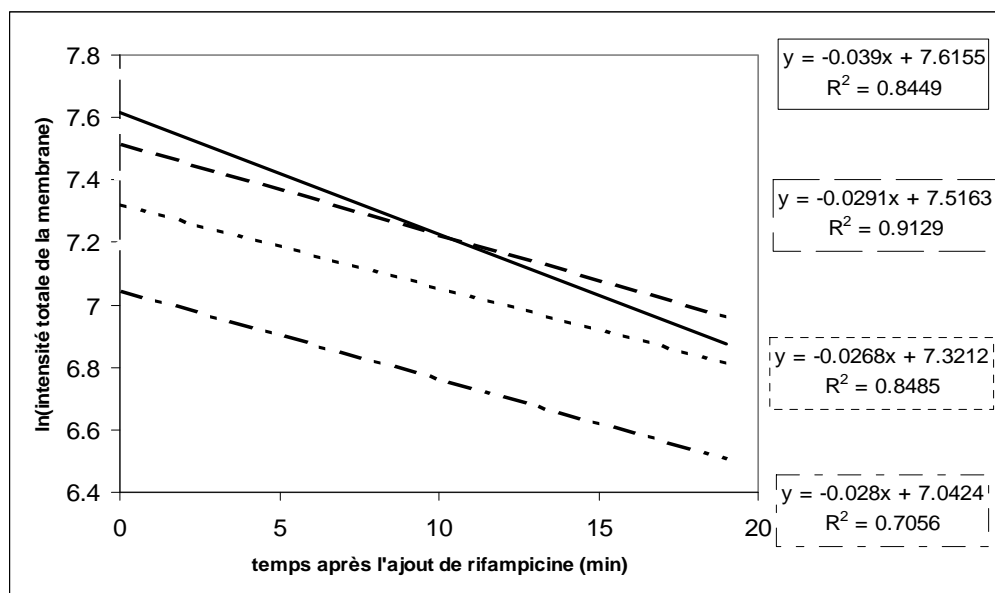


Figure 6 : Estimation globale des stabilités moyennes à un taux de croissance de 0.11 h^{-1} . Chaque style de trait correspond à une répétition incluant un point de référence avant le blocage de la transcription (à 0 minutes) et 3 points espacés dans le temps entre 0 et 20 minutes.

V. Recherche de motifs

Deux logiciels disponibles sur le web ont été utilisés pour la recherche de motifs dans les séquences non codantes en amont des gènes de *L. lactis* IL1403 (de -400 à +1) : RSA tools pour les motifs de 4 à 8 nucléotides (<http://rsat.scmdbb.ulb.ac.be/rsat/>) et MEME pour ceux de 4 à 25 nucléotides (<http://meme.sdsc.edu/meme/meme.html>). D'une manière générale, les motifs découverts souvent avec MEME, ont ensuite été recherchés sur certains gènes ou sur l'ensemble du génome avec RSAtools.

VI. Traitement statistique des données

Plusieurs manipulations statistiques des données ont pu être réalisées grâce à l'utilisation du logiciel de statistiques R qui est en accès libre (<http://www.r-project.org/index.html>). La description de l'ensemble des fonctions utilisées est disponible dans le tableau 4.

Tableau 4 : Les principales fonctions de R utilisées lors de cette étude.
Ces fonctions sont disponibles dans les packages MASS, stat, multtest et randomForest.

| FONCTION | DESCRIPTION |
|----------------|---|
| help() | aide en ligne |
| help.search() | recherche par mots clé dans l'aide en ligne |
| read.table() | lecture des fichiers de données |
| length() | renvoie la longueur d'un vecteur |
| dim() | renvoie les dimensions d'une matrice |
| rep() | répète un vecteur |
| log() | application du logarithme népérien (ln) |
| cbind() | colle des colonnes |
| rbind() | colle des lignes |
| sum() | somme |
| mean() | calcul de la moyenne |
| median() | calcul de la médiane |
| var() | calcul de la variance |
| cov() | calcul de la covariance |
| sd() | calcul de l'écart type |
| hist() | représentation des données sous forme d'histogramme |
| plot() | représentation des données sous forme de scatter plot |
| sploM() | représentation de données matricielle sous forme de scatter plot |
| boxplot() | représentation des données sous forme de boîtes à moustaches |
| heatmap() | double classification hiérarchique des données |
| hclust() | classification ascendante hiérarchique |
| scale() | centrage des colonnes d'une matrice |
| prcomp() | analyse en composantes principales (ACP) |
| par() | partage la fenêtre graphique |
| summary() | fournit l'ensemble des détails disponibles pour l'objet appelé |
| class() | détermine la nature d'un objet |
| apply() | applique une fonction choisie à un élément donné |
| sink() | dirige les résultats vers un fichier externe |
| t.test() | test d'égalité des moyennes (Student) ou des variances (Fischer) si var.equal = T |
| mt.rawp2adjp() | calcul des p-values ajustées |
| cor() | recherche de corrélation entre deux variables |
| cor.test() | recherche de corrélation entre deux variables et p-values associées |
| ks.test | test de répartition uniforme (Kolmogorov-Smirnov) |
| lm() | construction de modèle linéaire |
| res() | résidus du modèle |
| stepAIC() | algorithme de sélection de variables |
| for(), while() | permet de faire des boucles de calcul |
| dist() | calcul des distances euclidiennes |
| randomForest() | recherche des données les plus discriminantes |

VI.1. Exploration des données

L'exploration des données est effectuée pour permettre leur validation. Ce préambule consiste essentiellement à s'appropriier les données de manière visuelle grâce à différentes représentations graphiques (figure 7).

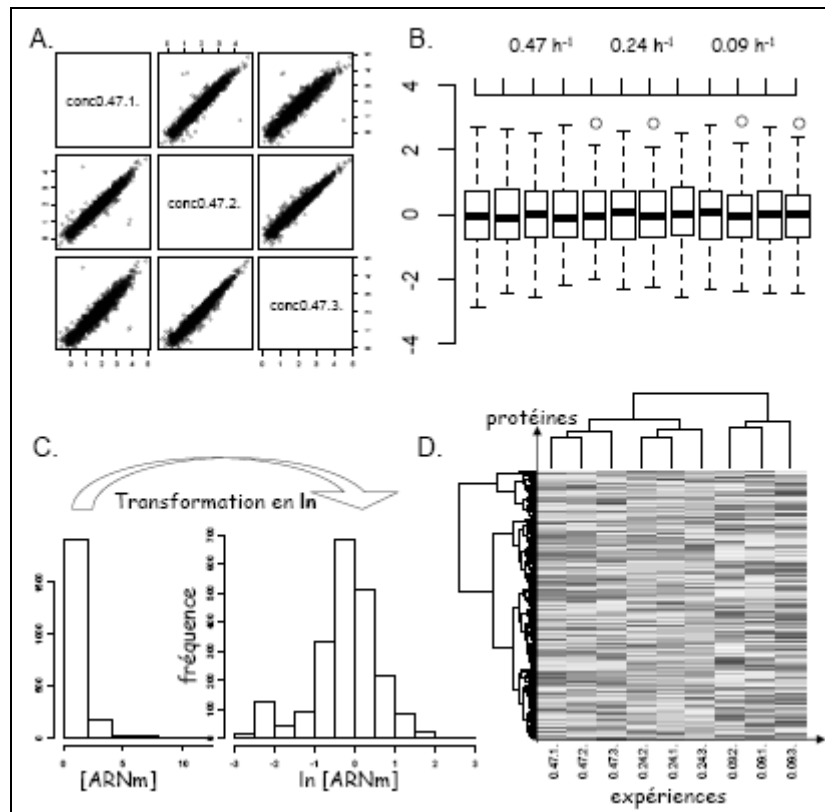


Figure 7 : Représentations graphiques des données.

Les représentations sous forme de scatter-plot (figure 7.A) et de boîtes à moustache (figure 7.B) permettent de comparer les répétitions d'une même expérience et donc de vérifier la reproductibilité des données. Ces techniques servent aussi à repérer les valeurs atypiques qui devront éventuellement être éliminées par la suite afin de ne pas fausser les analyses. Le tracé de l'histogramme des données détermine leur allure indiquant ainsi si leur transformation est nécessaire (figure 7.C). En effet, de nombreux tests statistiques (comme le test de Student par exemple) requièrent une répartition gaussienne des données auxquels ils sont appliqués. Les données brutes de transcriptome comme de protéome présentent une allure gaussienne uniquement si elles sont transformées par le logarithme. Enfin, la double classification hiérarchique (figure 7.D) reclasse les différentes expériences en fonction de leur proximité de réponse et permet de repérer les répétitions aberrantes.

VI.2. Analyse statistique des données

VI.2.1. Recherche des données qui varient entre deux conditions

Accéder aux variations d'expression des ARNm comme des protéines passe par le calcul des ratios de valeurs entre deux conditions. La significativité des variations observées est ensuite estimée par le test de Student. Ce test estime pour chaque ratio la probabilité (p-value) qu'il soit égal à 1. Toutefois, devant le nombre de données recueillies en protéome et en transcriptome il était nécessaire d'effectuer des tests multiples et de définir un seuil d'acceptation des erreurs pour l'ensemble des données et non pour chacune individuellement. Nous avons donc choisi pour cela de calculer un taux de faux positif (FDR) par la méthode de Storey (2002) pour les ARN messagers en plaçant la limite de significativité à 10 %. La méthode de calcul du FDR proposée par Benjamini et Hochberg (1995) a été appliquées aux protéines. Pour ces données de protéines qui, pour des raisons techniques, sont un peu moins reproductibles, la limite de significativité à été rehaussée à 20 %.

VI.2.2. Recherche de corrélations

La recherche de corrélations entre données quantitatives s'est faite par le calcul des coefficients de corrélation de Pearson et de Spearman. Le premier mesure la force d'un lien linéaire entre les variables étudiées tandis que le second reflète la corrélation entre leurs rangs (Valleron, 2007). Le seuil de significativité de chacun de ces tests a été fixé à 5 % pour la p-value associée.

VI.3. Modélisation

La modélisation statistique est une démarche à visée explicative qui consiste à rechercher les déterminants majeurs permettant d'expliquer le comportement d'une variable d'entrée. Nous avons cherché à modéliser les niveaux de protéines en utilisant les données de transcriptome mais aussi d'autres paramètres quantitatifs ou qualitatifs par construction de modèles de covariance. Nous avons supposé que les influences de ces différents paramètres sur les niveaux de protéines étaient multiplicatives et avons choisi de modéliser le logarithme népérien des niveaux de protéines afin de pouvoir utiliser des modèles linéaires (la transformation logarithmique d'une multiplication se traduit par une addition). Ce genre de modèle s'écrit de la façon suivante :

$$\ln[prot_i] = \alpha + \sum_j \beta_j \ln(var_{i,j}) + \sum_k \lambda_{k(i)} + \varepsilon_i$$

où $[prot_i]$ représente la concentration de la, α la pente, $var_{i,j}$ la valeur du $j^{ième}$ paramètre quantitatif pour la $i^{ième}$ protéine, β_j et $\lambda_{k(i)}$ les coefficients respectivement associés au $j^{ième}$ paramètre quantitatif et au $k^{ième}$ paramètre qualitatif de la $i^{ième}$ protéine et ε_i le terme d'erreur pour la $i^{ième}$ protéine. Les conditions d'identifiabilité requièrent que la somme des coefficients $\lambda_{k(i)}$ soit égale à zéro.

Les coefficients associés aux différents paramètres du modèles sont déterminés par la méthode des moindres carrés ; c'est-à-dire qu'ils sont choisis de manière à minimiser la somme suivante :

$$\sum_i (\ln[prot_i]_{\text{mod}} - \ln[prot_i])^2$$

où $\ln[prot_i]_{\text{mod}} = \hat{\alpha} + \sum_j \hat{\beta}_j \ln(var_{i,j}) + \sum_k \hat{\lambda}_{k(i)}$ est la valeur estimée à partir du modèle.

La qualité d'ajustement du modèle est quand à elle estimée par le calcul du coefficient de détermination :

$$R^2 = 1 - \frac{\sum (\ln[prot_i]_{\text{mod}} - \ln[prot_i])^2}{\sum (\ln[prot_i] - \overline{\ln[prot_i]})^2}.$$

où $\overline{\ln[prot_i]}$ est la valeur moyenne de $\ln[prot_i]$. R^2 représente le pourcentage de variance de la variable $\ln[prot]$ expliqué par le modèle.

Une sélection statistiques des paramètres du modèles selon le critère d'Akaike (AIC pour Akaike Selection Criterion, (Burnham & Anderson, 2002)) permet de faire un compromis entre la qualité d'ajustement des modèles et le nombre de variables qui y sont introduites. Cet algorithme n'est autre qu'un critère des moindres carrés pénalisé et va permettre la sélection des variables les plus pertinentes :

$$n \log\left(\sum_i (\ln[prot_i]_{\text{mod}} - \ln[prot_i])^2 / n\right) + 2N$$

où N représente le nombre de paramètres à estimer par le modèle.

Enfin, les paramètres quantitatifs introduits dans le modèle sont préalablement normalisés (division par la variance après soustraction de la moyenne) de façon à pouvoir ensuite les hiérarchiser selon les valeurs absolues des coefficients du modèle qui leur sont associés.

**Chapitre I : Réponse de *Lactococcus lactis* à
la variation du taux de croissance et
implication dans ses réponses au stress**

Les études de stress consistent généralement à comparer des états « stressés » avec des états de croissance optimale. Ces comparaisons physiologiques peuvent s'effectuer à divers niveaux allant de la survie à la mesure des activités enzymatiques jusqu'à des niveaux de plus en plus globaux avec les mesures de transcriptome et de protéome. Bien que l'imposition d'un stress s'accompagne systématiquement d'une réduction de la croissance bactérienne, l'unique impact du taux de croissance sur la physiologie cellulaire n'a jamais été décrit pour une bactérie. Seules des données pour la levure modèle *Saccharomyces cerevisiae* sont disponibles (Castrillo *et al.*, 2007; Regenber *et al.*, 2006).

Dans cette publication nous nous sommes intéressés à l'impact que pouvait avoir la variation du taux de croissance sur la bactérie lactique *Lactococcus lactis* au travers d'une étude métabolique et transcriptomique. Les principaux résultats obtenus au cours de cette étude et les méthodes pour y parvenir sont récapitulés en figure 8. L'analyse spécifique de l'effet du taux de croissance a été menée en culture continue en chémostat. Ce type de fermentation est un précieux outil car il permet le découplage des différents paramètres de culture et en particulier le contrôle du taux de croissance par l'imposition du taux de dilution. L'isoleucine, pour laquelle la bactérie étudiée est auxotrophe (Cocaign-Bousquet *et al.*, 1995), a été choisie comme élément nutritif limitant. La limitation en carbone, plus habituelle, n'a pas été retenue car elle oriente le métabolisme vers des voies particulières, plus complexes (métabolisme mixte).

Le métabolisme carboné de *L. lactis* est resté homolactique aux quatre taux de croissance étudiés (0.09, 0.24, 0.35 et 0.47 h⁻¹). Les résultats transcriptomiques ont été comparés en utilisant comme référence l'image transcriptomique obtenue au taux de croissance le plus lent (0.09 h⁻¹). Globalement, la réponse est très étendue puisqu'elle affecte l'expression de 30 % des gènes de *L. lactis*. Les gènes codant pour des fonctions physiologiques associées à la biogénèse telles que la transcription, la traduction, la division ou encore le métabolisme des acides gras et des phospholipides, sont surexprimés en réponse à l'augmentation du taux de croissance. La répression massive des gènes associés aux phages, prophages et transposons est plus surprenante et pourrait être reliée à la capacité d'évolution de la bactérie. Ces éléments génétiques mobiles sont en effet impliqués dans la modification des génomes bactériens (Canchaya *et al.*, 2004).

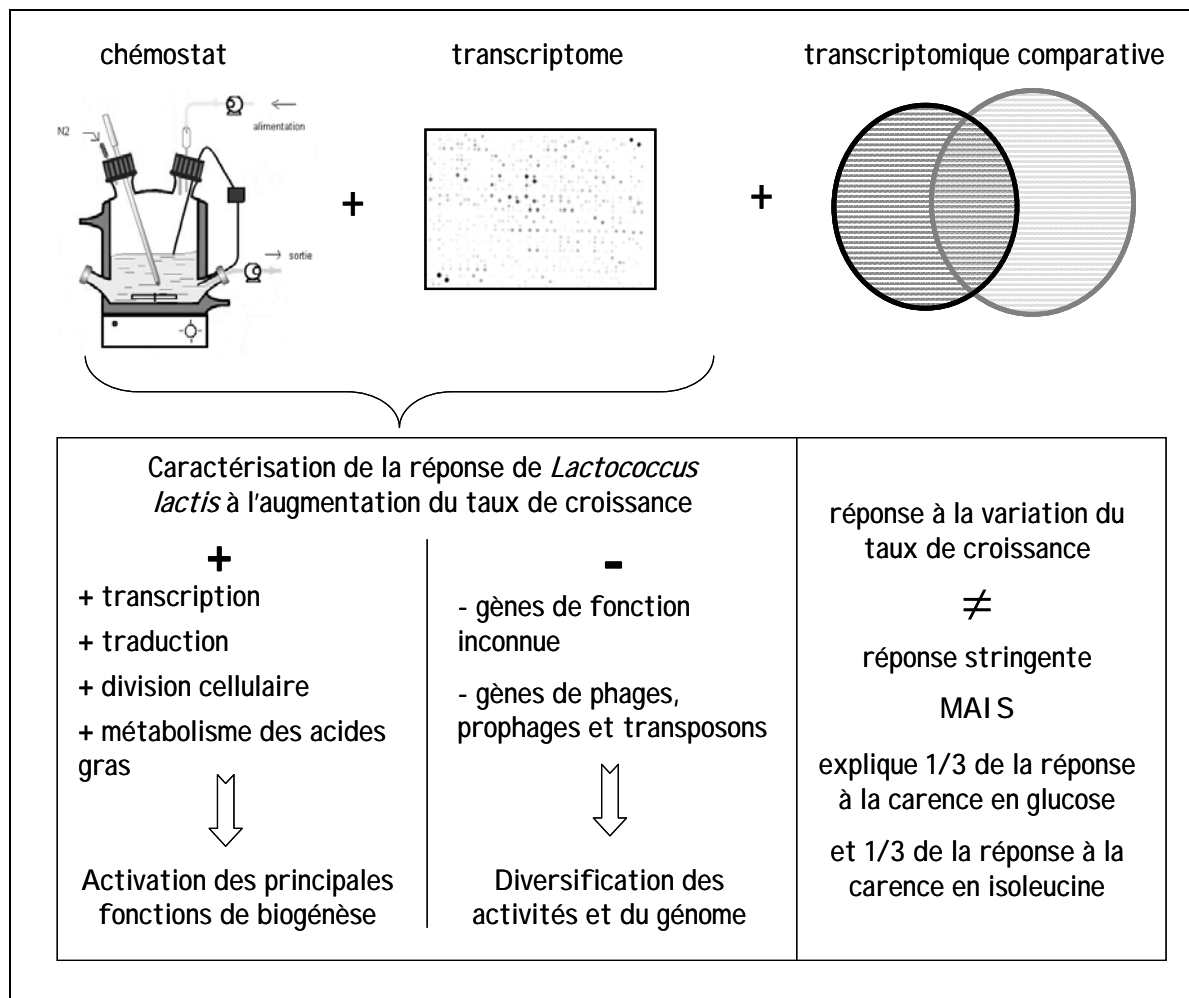


Figure 8 : Effet taux de croissance et réponses au stress chez *Lactococcus lactis*.

La comparaison transcriptomique de la réponse à la variation du taux de croissance avec celles de la carence en glucose et en acide aminé a pointé des recouvrements significatifs révélant ainsi qu'environ 1/3 de la réponse observée lors des carences en glucose et en acide aminé est directement associée à la variation du taux de croissance. La part de la réponse liée au taux de croissance n'est donc pas négligeable, soulignant l'importance de ce phénomène.

La réponse stringente est un mécanisme général de plus en plus considéré comme gouvernant la réduction du taux de croissance chez les bactéries (Jain *et al.*, 2006). La réponse stringente de *L. lactis* a été induite par l'ajout de norvaline, un inhibiteur du chargement des ARN de transfert (Eymann *et al.*, 2002), et comparée à l'effet du taux de croissance. Le faible recouvrement des profils transcriptomiques de la réponse stringente et de l'impact du taux de croissance implique que ces deux mécanismes sont clairement distincts.

Deux régulateurs potentiels de cet effet taux de croissance, llrE et yabB, ont pu être identifiés mais leur rôle biologique n'a pas encore été démontré. De façon remarquable, il a été observé que les gènes dont l'expression est proportionnelle au taux de croissance sont préférentiellement localisés à proximité de l'origine de réplication. *A contrario*, ceux dont l'expression est inversement proportionnelle au taux de croissance se répartissent plutôt aux alentours de la zone de terminaison de la réplication, à l'autre extrémité du chromosome. Cette observation suggère que l'expression des gènes liés au taux de croissance est déjà influencée par leur localisation chromosomique et n'est pas uniquement une réponse adaptative.

Version publiée par BMC Genomics en juillet 2008

Growth rate regulated genes and their wide involvement in the *Lactococcus lactis* stress responses

Clémentine Dressaire^{†1,2}, Emma Redon^{†1,3}, Helene Milhem^{4,5}, Philippe Besse^{4,5}, Pascal Loubière^{1,2}, Muriel Coccain-Bousquet^{1,2*}

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

(2) UMR5504, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, CNRS, INRA, 31400 Toulouse, France

(3) Current address : Laboratoire de Biologie et Biotechnologies marines, IBFA, UMR100 IFREMER, Physiologie et Ecophysiologie des Mollusques Marins, Université de Caen Basse-Normandie, Esplanade de la Paix, F14032 Caen Cedex, France

(4) Université de Toulouse ; INSA ; F-31077 Toulouse France

(5) CNRS ; IMT (Institut de Mathématiques de Toulouse) ; F-31077 Toulouse France

[†] These authors contributed equally to this work

*correspondent foot-note: e-mail: cocaign@insa-toulouse.fr

ISBP, INSA 135 Avenue de Rangueil,
31077 Toulouse cedex 4, France.

Tel: (33) 561 559 438

Fax: (33) 561 559 400

Abstract

Background: The development of transcriptomic tools has allowed exhaustive description of stress responses. These responses always superimpose a general response associated to growth rate decrease and a specific one corresponding to the stress. The exclusive growth rate response can be achieved through chemostat cultivation, enabling all parameters to remain constant except the growth rate.

Results: We analysed metabolic and transcriptomic responses of *Lactococcus lactis* in continuous cultures at different growth rates ranging from 0.09 to 0.47 h⁻¹. Growth rate was conditioned by isoleucine supply. Although carbon metabolism was constant and homolactic, a widespread transcriptomic response involving 30 % of the genome was observed. The expression of genes encoding physiological functions associated with biogenesis increased with growth rate (transcription, translation, fatty acid and phospholipids metabolism). Many phages, prophages and transposon related genes were down regulated as growth rate increased. The growth rate response was compared to carbon and amino-acid starvation transcriptomic responses, revealing constant and significant involvement of growth rate regulations in these two stressful conditions (overlap 27 %).

Two regulators potentially involved in the growth rate regulations, *llrE* and *yabB*, have been identified. Moreover it was established that genes positively regulated by growth rate are preferentially located in the vicinity of replication origin while those negatively regulated are mainly encountered at the opposite, thus indicating the relationship between genes expression and their location on chromosome. Although stringent response mechanism is considered as the one governing growth deceleration in bacteria, the rigorous comparison of the two transcriptomic responses clearly indicated the mechanisms are distinct.

Conclusions: This work of integrative biology was performed at the global level using transcriptomic analysis obtained in various growth conditions. It raised the importance of

growth rate regulations in bacteria but also participated to the elucidation of the involved mechanism. Though the mechanism controlling growth rate is not yet fully understood in *L. lactis*, one expected regulatory mechanism has been ruled out, two potential regulators have been pointed out and the involvement of gene location on the chromosome has also been found to be involved in the expression regulation of these growth related genes.

Background

Natural free living bacteria, as well as industrially exploited strains, have to face many challenging situations such as chemical (pH, salinity) and physical (temperature) changes of the environment or nutrient deprivation. Pooled under the generic term of stress, these environmental challenges share a common feature: they all affect the cellular growth rate. Many studies, mostly using transcriptomic approach, have exhaustively described the responses to various stresses in different bacteria [1-4]. However these studies performed in batch cultures do not allow the growth rate response to be specifically described because influences of the specific stress and the growth rate are superimposed. The exclusive cellular response associated to growth rate changes can be achieved through chemostat cultivation. This fermentation mode enables the growth rate to be changed by the dilution rate monitoring without modifying environmental or nutritional parameters [5]. Transcriptomic analysis in chemostat have been performed mainly in *Saccharomyces cerevisiae* [6-10] and the specific response associated to growth rate modifications has been characterized. The authors demonstrated that a large part of the genome expression, ranging between 27 and 50%, was affected by the growth rate [7-9]. Growth rate influence is also major in bacteria and Egli and co-workers [11] demonstrated in chemostat culture that the sensitivity of *Escherichia coli* to stress was dependent on the specific growth rate. However a complete analysis combining

chemostat cultivation and transcriptomic approach to clearly study the growth rate influence at the global level has never been performed in bacteria.

Currently, the existence of a general mechanism activated by the bacteria to cope with any stressful condition slowly develops and a rising hypothesis considers this biological process to be the stringent response [12, 13]. Chang and colleagues even go ahead by clearly proposing stringent response to be triggered by the cells when growth rate decreases independently of this decrease origin [14]. Stringent response is a well known mechanism involved in bacteria mediated by ppGpp (see [13] for review), however its strict comparison to growth rate response has never been reported, the unique growth rate study being available for *S. cerevisiae* [7-9], a micro-organism lacking the stringent mechanism. Thus, the assumption that the regulation involved in the growth rate control may be identical to the stringent response is emerging though no clear evidence is available in the literature.

In order to characterize growth rate effect in the model of lactic acid bacteria, *Lactococcus lactis*, we have performed chemostat cultures at various growth rates and analysed the cellular response to this growth rate changes by transcriptomic approach. The transcriptomic response was then compared to the one obtained when the growth rate was modified by stress (carbon and amino acid starvations) in batch cultures. Lastly the norvaline response, a chemical inductor previously used in *Bacillus subtilis* [15] to induce the stringent response, was characterised and compared to the growth rate response.

Methods

Strain and growing conditions

The strain *Lactococcus lactis* IL1403, whose genome is available [16] was grown in anaerobic chemostat cultures (under nitrogen atmosphere) on a chemically defined medium. The medium composition (table 1) is derived from the CDM medium [17, 18] with a ten-fold reduction of Branched-Chain Amino Acid (BCAA) concentrations *i.e.*, valine, leucine, isoleucine. Continuous cultures were performed in a 0.5 L fermentor (Verrerie Wagner, Toulouse, France) maintained at a constant temperature of 30 °C and under nitrogen atmosphere. The pH was regulated at 6.6 by automatic addition of KOH (10 N). Four different growth rates have been studied, namely 0.09, 0.24, 0.35 and 0.47 h⁻¹. For each steady-state studied, samples have been harvested in at least quadruplicate with a minimum delay of five doubling time between each sampling.

To study isoleucine starvation response, *L. lactis* was grown in batch cultures in similar medium (table 1). Cultures were grown under nitrogen atmosphere in a 2 L fermentor (Setric Génie Industriel, Toulouse, France) at 30 °C with an agitation speed of 250 rpm. The pH was maintained at 6.6 by automatic addition of KOH 10 N. To study the stringent response, cells were grown similarly in batch cultures in a medium containing two-fold increase of BCAA compared to the medium described in table 1 and norvaline was added at a final concentration of 10 g.l⁻¹ during the exponential growth phase.

Analytical methods

Bacterial growth was estimated by spectrophotometric measurements at 580 nm (Hitachi U1100, 1 Unity of absorbance is equivalent to 0.3 g·L⁻¹). Glucose, lactate, acetate and ethanol concentrations were measured by high pressure liquid chromatography (Hewlett Packard

1050; Waters 717 autosampler; Biorad microguard; H+ column Biorad HPX87H; Perkin Elmer LC90Bio UV detector; Hewlett Packard 1047A refractometer). Amino-acids concentrations in culture supernatants were measured via the AminoQuant HP1090 system (column: Hypersil-AA-ODS, 200*2.1 mm, 5µm, Agilent).

Transcriptomic analyses

Membrane spotting and analytical support were provided by the Biochips Platform (Genopole Toulouse, France). Cells were harvested from steady state continuous cultures (growth rate 0.09, 0.24, 0.35 and 0.47h⁻¹), from batch culture in exponential growth phase and after 30 min, 1.7 h and 3.5 h of BCAA starvation (isoleucine), and from batch culture in exponential growth phase and 1.6 h after norvaline addition. Cell lysis and total RNA extraction were performed as previously described [19]. RNA was quantified at 260 nm and RNA quality was controlled on electrophoresis agarose gel in denaturing conditions. Gene expression was measured using nylon arrays containing the PCR fragments (EUROGENTEC) for all genes of *L. lactis* IL1403 [16]. A constant amount of 10 µg of total RNA was used to perform the retrotranscription. Synthesis of radiolabelled cDNA, nylon arrays hybridizations and washings were performed as previously described [19]. Membranes were exposed to a phosphoimager screen for three days and scanned with a phosphofluoroimager (Storm 860, Molecular Dynamics). For each condition, three independent repetitions were performed.

Data analyses and statistical treatment

Hybridization signals were quantified, assigned to gene names, and statistically treated with the Bioplot software (developed by S. Sokol in Plateforme Génomique, Toulouse, <http://biopuce.insa-toulouse.fr/>). Local background was removed and signals were normalized

by the mean intensity of the membrane. For chemostat samples, expression ratios were calculated relative to the slowest growth rate of 0.09 h^{-1} . Student tests were performed and the statistical significance of expression ratios was evaluated using Student test ($p\text{-value} < 0.05$) and a false discovery rate ($\text{FDR} < 10 \%$). This selection method was slightly more stringent than the one based on a FDR limit of 10%.

For isoleucine starvation and stringent response studies, expression ratios were calculated using the exponential phase as a reference. Similar statistical treatment was performed.

For chemostat and isoleucine starvation data, genes whose expression is significantly modified compared to the reference in at least one of the three growth conditions have been selected in order to provide an exhaustive description of the response. Clustering analyses were performed with Genespring software (AGILENT) and R free statistical software (<http://www.r-project.org/>) was used for random forest analyses.

Motif research

The presence of DNA pattern in un-translated region of genes has been explored using either RSAtools (<http://rsat.ulb.ac.be/rsat/>) or MEME (<http://meme.sdsc.edu/meme/meme.html>) softwares. In any cases these sequences were obtained from RSAtools (retrieve sequence section, default parameters).

Data availability

Raw data of each experiment are available on GEO database (series accession number: GSE10256, GSE10254, GSE4872, GSE5761 respectively for growth rate, stringent, isoleucine starvation and glucose starvation responses).

Results

Growth rate response

Culture parameters

The growth rate of *Lactococcus lactis* was 5 fold increased in a chemostat culture from 0.09 to 0.47 h⁻¹ which corresponds to a decrease of the generation time from 7.7 to 1.5 h. The residual concentration of isoleucine, unlike all other amino acids, remains lower than the detection threshold (10 μM) whereas the glucose one is always high (table 2). Biomass concentration is independent of the growth rate (table 2) and the biomass production yield values from the different consumed amino acids are identical at the four different growth rates (figure 1). A very weak ornithine production is observed at $\mu=0.09$ h⁻¹ (result not shown). The lactate residual concentration decreases when growth rate is increased (table 2) but lactate remains the main fermentation product. The lactate production yield is indeed constant in the different growth conditions and the average value of 1.7 ± 0.1 mol of lactate per mol of glucose consumed is close to the maximum theoretical yield of 2 lactates per glucose. The production of formate, acetate and ethanol is independent of the growth rate and never exceeds 10 % of this lactate production (result not shown).

The relationship between glucose consumption rate and dilution rate is linear pointing out that the maintenance energy coefficient is constant and independent of the growth rate ($5\text{mmol}_{\text{Gluc}}/\text{gX}/\text{h}$).

Transcriptional response

As described in material and method, the slowest steady state ($\mu=0.09 \text{ h}^{-1}$) was chosen for the reference for transcriptomic analysis. Expression ratios between the various growth rates and this reference have been calculated (0,24/0,09 ; 0,35/0,09 and 0,47/0,09). Raw data are available on GEO database (GSE10256). Functional analysis was based on Bolotin *et al.* classification [16]. The over- or under-expressed genes enrichment in the various (sub) categories was determined at each steady state by Wilcoxon test on the whole category (p-value < 0.05 for significance). Main results corresponding to functional category tendencies are described in the following sections and individual genes are discussed when the statistical criteria on the gene are filled in (Student test p-value < 0.05 and FDR < 0.1).

When growth rate increases, the categories fatty acid and phospholipid metabolism (*accB*, *C*, *D*, *acpA*, *fabD*, *F*, *G1*, *Z1*, *Z2*, *plsX* and *yeaG*), cell division (*ezrA*, *ftsA*, *H*, *gid* and *rodA*) and transcription (*rnhA*, *B* (RNA degradation), *nusB*, *G*, *rpoA*, *B*, *C*, *D*, *E* and *smpB*) are significantly up-regulated. A wide up-regulation is also observed for the category translation (Wilcoxon p-value < 10^{-6}). The expression of various genes involved in translation process is indeed increased: 6 genes encoding amino-acyl tRNA synthetase (*argS*, *asnS*, *ileS*, *metS*, *pheT* and *tyrS*), 3 genes whose products are involved in protein degradation (*pepC*, *T* and *yuhB*), 4 genes encoding translation factors (*infA*, *B*, *rbfA* and *tsf*) and 44 genes encoding ribosomal proteins (*gatB*, *prmA*, *rnpA*, *rplA*, *B*, *M*, *N*, *O*, *Q*, *S*, *T*, *U*, *V*, *X*, *rpmA*, *B*, *D*, *E*, *F*, *GB*, *I*, *J*, *rpsA*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, *J*, *K*, *L*, *N*, *O*, *Q*, *R*, *S*, *T*, *U*).

The category energy metabolism is mostly down-regulated when the growth rate is increased according to Wilcoxon test. Nevertheless control of genes related to this category differs between the different sub categories. On one hand, genes encoding protein involved in sugar

metabolism (*bglA*, *galM*, *T*, *gntK*, *Z*, *rbsK*, *scrK*, *uxuA*, *B*, *xylA*, *B*, *xynB* and *yidC*) are chiefly under-expressed when growth rate increases. Other genes involved in aerobic (*noxC*, *E*, *poxL*, *yddB*, *ymgK*, *ypgB* *yphA*, *yrjB*, *C*, *yxdE*), anaerobic (*dhaL*, *M*, *glpD*) and amino acids and amines (*arca*, *B*, *ipd*, *pdC*, *yciA*, *yjiB*) metabolisms are also under-expressed. On the other hand, RNA messengers encoding proteins involved in tricarboxylic acids metabolism (*citC*, *D*, *E* and *F*) and ATP conversion (*atpB*, *E*, *F*, *G* and *H*) are mostly over-expressed when growth rate increases and some genes encoding glycolytic enzymes (*enoA*, *gapA*, *B*, *pgk*, *pgmB*, *pmg*, *pyk*, *tpiA*, *yjhF*) are also over-expressed.

Regarding purine and pyrimidine metabolisms, the expression of genes encoding proteins involved in interconversion of nucleotide and nucleoside (*rmlA*, *B*, *C*) globally increases under growth acceleration whereas, unexpectedly, those encoding enzymes from purine biosynthesis pathway (*purD*, *E*, *L*, *M*, *N*) are mostly under-expressed. Genes involved in pyrimidine metabolism do not seem to be affected in a specific manner by growth rate variations (no significant Wilcoxon p-value).

The category named other is strongly down-regulated (Wilcoxon p-value $<10^{-15}$). Expression of phages and prophages related genes (63 *pi* and *ps*), and genes encoding transposition proteins (*tra981C*, *983L*, *1077B*, *yajE*, *ybdK* and *L*) are mainly under-expressed. Growth rate increase also reduces the expression of genes classified in the category transport and binding proteins and notably for the subclasses multi-drug resistance, carbohydrates, organic alcohols and acids, cations and anions.

No general tendency can be observed for the category amino acids biosynthesis: genes of histidine family (*hisC*, *G*, *Z*, and *D*) are mainly under-expressed whilst those of branched amino acids family (*ilvB*, *D*, *N*, *leuC* and *D*) are mostly over-expressed when growth rate is increased. More particularly some genes of the glutamate family (*argB*, *D*, *J*, *proC*) are under

expressed. Lastly, the two categories, regulatory and unknown functions, are significantly down-regulated.

Genes whose expression is significantly modified compared to the reference ($\mu = 0.09 \text{ h}^{-1}$) in at least one of the three growth conditions have been selected with usual statistical criteria (student test p -value < 0.05 and FDR $< 10 \%$). The expression of 722 genes (377 up- and 345 down-regulations) varied significantly with the growth rate increase. These genes have been clustered with GeneSpring software in five different groups (see figure 2 for average expression ratios). Cluster A, B and C enclose respectively 115, 149 and 169 mRNA over-expressed at every dilution rates. Genes of cluster A display a maximum over-expression at 0.24 h^{-1} and expression ratios tend to decrease when growth rate further increases. In cluster B, over-expression increases with the growth rate though an inflexion is observed in the curve at $\mu=0.35 \text{ h}^{-1}$. Cluster C includes transcripts whose expression variations follow growth rate evolution. For genes ranked in this cluster, the biggest the difference between growth rate and the reference, the more important is the over-expression. This cluster encloses the main part of mRNA related to the translation (*argS*, *metS*, *pepC*, *T*, *pheT*, *prfB*, *rbfA*, *tyrS* and 35 genes encoding ribosomal proteins) and transcription (*nusG*, *rnhA*, *B*, *rpoA*, *C*, *E* and *smpB*) functions. Cluster D (199 genes) and E (138 genes) enclose the genes under-expressed when the growth rate increases. A stronger under-expression is observed in cluster D compared to cluster E at high growth rate. This cluster D includes transcripts belonging mainly to the category named other with more specifically 34 phages or prophages related genes (*pi* and *ps*) and 5 related to transposition (*tra981C*, *983L*, *1077B*, *ybdK*, *ybdL*). Moreover 31 mRNA encoding transporters and 6 transcripts belonging to the subclass sugar of the energy metabolism functional category (*bglA*, *galM*, *rbsK*, *scrK*, *xynB*, *yidC*) are also included in this cluster D.

Determinants of the growth rate response

Pattern research has been carried out in un-translated regions (upstream and downstream) of the genes belonging to the five different clusters using both RSAtools and MEME softwares. MEME identified in downstream regions a conserved pattern ($5.6 \times 10^{-47} < \text{E-value} < 2.2 \times 10^{-5}$) in each of the five groups analysed separately: gtcagtaa (or its reverse complement ttactgac). Then the frequency of the pattern has been searched in the different clusters but also in the entire genome of *L. lactis* IL1403 with RSAtools software. The average representation of the pattern in the different clusters ($19.6 \pm 6 \%$) is similar to the one in the entire genome (19 %) indicating that this pattern is not specifically involved in the growth rate response. One can however notice a significant under-representation of the pattern in cluster D (only 11 % of genes flanked with the pattern). This result can be related to the over-representation in this cluster of genes encoding phages, prophages and transposons functions. Indeed, though the pattern is uniformly spread out in the different functional categories ($21 \pm 7 \%$), a strong under-representation can be seen in the category other due to the nearly systematic lack of the pattern in the genes related to phages, prophages and transposons functions (4% of the genes). This pattern, previously identified as a highly repetitive motif [20] involved in mRNA destabilisation [19] in *L. lactis* can thus correspond to a specific feature of *L. lactis* genes; its absence could, at the opposite, be a sign of recent genetic acquisition.

As transcriptional regulations could be influenced by chromosomal organisation, we evaluated the potential link between genes involved in the growth rate response and their distance from the replication origin. In order to determine whether the location of genes belonging to each cluster is similar or not to the global distribution of the genes along the chromosome, the test of Kolmogorov-Smirnov was used. For data independency requirements

of the test, the genes belonging to the tested cluster were removed from the global distribution. The null hypothesis is rejected only for the clusters C and D (Kolmogorov p-value of 0.02 and 1.1 %) indicating an atypical distribution of the genes belonging to these clusters. To complete these investigations, *L. lactis* IL1403 chromosome was divided into 8 equal parts starting from its origin (8 intervals of 295699nt covering in average 266 genes). *L. lactis* unique chromosome being circular, these parts were then paired so as to congregate parts at the same distance from the replication origin. The distributions of the transcripts from clusters C and D as well as the global distribution in those intervals are reported in the figure 3. One can notice that genes from cluster D are rather located far from the replication origin while genes from cluster C are mainly located close to it. A thinner division confirms this distribution profile and enables a better evaluation of gene locations (figure3).

Growth rate impact in the different stress responses

Stringent response

As literature suggests stringent response to be a general mechanism used by cells to face growth rate reduction [12-14], stringent response and growth rate transcriptional responses have been compared. Stringent mechanism, firstly evidenced in *L. lactis* by Rallu *et al.* [21] was induced *via* norvaline addition. This leucine analogue that inhibits the amino-acylation of tRNA^{ile} and tRNA^{leu} has previously been used to induce the stringent response in *Bacillus subtilis* [15]. Norvaline was added during exponential growth of *L. lactis* in batch fermentation. Culture medium was similar to the one used for the characterization of growth rate response except that isoleucine and other branched chain amino-acid concentrations were two-fold increased in order to avoid any nutritional limitation. After norvaline addition, growth stopped for 45 min and then resumed at constant but slower rate (0.17 h⁻¹ compared to 0.78 h⁻¹ before norvaline addition). The main feature of the stringent response, consisting in

the negative control of rRNA transcription [15, 22], was confirmed here by the decrease of the rate of [³H]-uridine incorporation into RNA after norvaline addition (results not shown). Gene expression was measured in the exponential growth phase (reference sample) and 1.6 h after norvaline addition (biomass concentration of 0.82 g.l⁻¹) and 461 differentially expressed genes representing the stringent response were identified. Raw data are available on Geo database (GSE10254).

The growth rate expression data and the stringent response have been compared with Venn diagrams. Only 75 genes were common to the stringent and growth rate responses (figure 4a), thus revealing a weak overlap between the two responses (10 % of the growth rate response). Stringent response and growth rate effect were also compared at gene functional level. The main common feature between these two responses is a generalised control of transcripts encoding translation apparatus (ribosomal proteins in essence but also transfer RNA and translation factors). However the regulation of translation machinery encoding genes seems more pronounced in continuous culture than in stringent response (57 compared to 25 genes positively regulated with growth rate) although the growth rate variations are similar in the two experiments (approximately five-fold). Similar differences are obtained with the regulation of phage related genes which is massive in continuous culture but very limited in the stringent response (63 compared to 7 genes negatively regulated with growth rate). Inversely, the stringent response pointed out a large over-expression of genes involved in stress protection but the phenomenon is more restricted in the growth rate response in continuous culture (27 and 18 genes respectively). Lastly, 88 genes are regulated in a completely opposite way in the two compared conditions.

Nutrient starvations

In order to evaluate growth rate impact in stress responses, the growth rate regulated genes have been compared to transcriptomic analyses obtained during glucose or isoleucine starvation. These two experiments were performed in batch cultures with the same strain, similar culture medium and environmental conditions. For isoleucine starvation experiment, growth was exponential and decelerated strongly after isoleucine was exhausted. Transcriptomic analysis was performed during the exponential growth phase and 30 min, 1.7 h and 3.5 h after isoleucine starvation. Expression ratios were calculated relative to the exponential phase. Raw data are available on GEO database (GSE4872). The carbon starvation response was previously studied in our team [23] with similar dynamic approach (GEO database, accession number : GSE5761). Transcriptomic analysis and statistical treatment were similar in these experiments compared to the chemostat study, thus allowing a rigorous comparison of the various data sets (figure 4a and b). A number of 175 genes were common to isoleucine starvation and growth rate responses, which represents 24 % of the whole growth rate stimulon. Similarly 212 genes accounting for 29 % of the growth rate response also belong to growth and carbon starvation responses.

The global comparison between glucose, isoleucine starvations and growth rate effect shows a “core” common response including 70 genes. Among these common genes, 24 genes positively regulated by growth rate belong to the translation functional category and 3 genes negatively regulated by growth rate are phage or prophage genes. It is important to note, in this gene core, the presence of two genes encoding (putative) regulatory proteins, *llrE* and *yabB*, which are negatively regulated by growth rate.

The two starvation conditions were also compared to the stringent response (figure 5). The involvement of the stringent response highly differs between the two nutrient limitations

though constant implication (mean 27 %) has been observed above for growth rate regulation. A strong overlap between stringent response and isoleucine starvation is indeed observed (69 %) while the overlap is weak with carbon starvation (14 %). Furthermore 55 genes are common to the two nutrient starvations and the stringent response. However in this set of genes a strong enrichment in growth rate related genes is measured (40 % of these genes belong to the growth rate response in chemostat compared to 17 % obtained for the global stringent response).

Robustness analysis

When two or more conditions are compared, the list of (un)common genes may vary with the selection criteria initially applied on transcriptomic data (such as p-value or ratio filter), even if the same criterion is applied to all conditions. In order to test whether the overlapping percentages calculated here can depend or not on these data selection criteria, the selectivity on expression data was increased either by increasing ratio threshold or by decreasing p-value. The number of genes whose expression significantly varies in each condition decreases when the selectivity is increased, provoking a decrease in the number of transcripts in common between the different conditions (results not shown). However overlapping percentages between the responses are not strongly affected when the data selection criteria are changed (table 3). The robustness of the comparative analysis is therefore ascertained and the conclusions drawn previously can thus be reinforced.

Lastly, in order to provide a deeper analysis of the robustness of the overlapping method, other types of statistical tools (clustering or random forest analysis) have been used. The growth rate response has been discriminated into the two clusters C and D, chosen for their direct relationship with growth rate modifications, and overlaps with other responses have

been estimated in each case. The overlapping of clusters C and D with carbon starvation, isoleucine starvation and stringent response are respectively 27, 31 and 15 %, which confirm the previous figures. Lastly, a random forest analysis was performed on the growth rate data to select the most discriminatory genes ([24], <http://cran.r-project.org/>). Unlike other statistical treatment as multiple tests, this random forest approach considers data without making any assumption on genes independency and leads to complementary results [24]. The list of discriminatory transcripts changes with every new round of random forest testing. Nevertheless, comparing the various lists with other responses gave overlapping percentages in the same order of magnitude than the previous ones.

Discussion

Growth rate response: metabolism

Nitrogen metabolism is crucial in lactic acid bacteria and high amino acid requirements have been described for *L. lactis* [25]. The growth of *L. lactis* IL1403 when limited by isoleucine supply in chemostat culture is characterized by homolactic metabolism. Lactate concentration remains always lower than the growth inhibiting threshold value of 150-200 mM [26, 27]. This homolactic metabolism is different from the mixed metabolism observed for *L. lactis* when grown in similar conditions but with carbon limitation [26]. This result is however consistent with the work of Garrigues *et al.* [28] showing that mixed metabolism is dependant on the unbalance between catabolism and anabolism. In the chemostat limited by isoleucine supply, anabolic limitation rather than catabolic limitation is occurring. In these conditions the mixed metabolism, governed by energetic demand, is not activated due to a relative carbon flux excess. As a consequence, the metabolism remains homolactic. A deviation of the metabolism towards the mixed metabolism would have provoked a radical change in the

energetic status of the cells since homolactic and mixed metabolism are not equivalent in their ATP yield. In the chemostat controlled by isoleucine supply, the energetic status of the cells is constant since maintenance and biomass yields do not vary with growth rate modifications. Isoleucine limitation and its corresponding homolactic metabolism are thus well adapted for studying the growth rate impact on cells because they avoid metabolic and energetic interferences.

Growth rate response: transcriptomic analysis

Widespread transcriptional response is observed when growth rate is modified since more than 30 % of *L. lactis* genome shows significant expression changes. Growth rate increase is accompanied by over-expression of genes involved in transcription, translation, cell division, fatty acids and phospholipids metabolism. This, and especially the massive up-regulation of 44 genes encoding ribosomal proteins, is consistent with the results observed for *S. cerevisiae* [7, 9] and should be related to the acceleration of biogenesis at high growth rate. Unlike in the yeast, amino-acid and nucleotide biosynthesis pathways are not up-regulated in *L. lactis*. Such differences could be related to *L. lactis* multiple nutritional requirements and the presence of the various bases and amino-acids in the culture medium. Some genes involved in detoxification (aerobic metabolism) or in multi-drug resistance (transport) are under-expressed when the growth rate is increasing. However the expression of genes encoding important functions involved in stress protection such as chaperone, ATPase, proteases or general stress proteins, is not modified. Thus, the wide down-regulation of the genes involved in stress protection, which is observed in the yeast [7, 9] when the growth rate is increased, does not occur in *L. lactis*. Finally, genes with unknown function are massively under expressed when growth rate is increased in the yeast [9] as well as in *L. lactis*. Though these genes do not necessarily share the same functions in the two micro-organisms, it can be

ascertained that the cellular role of these genes is yet ignored in the two micro-organisms because they have been poorly studied due to their weak expression at high growth rates, conditions generally used in laboratories.

Genes encoding enzymes involved in ornithine biosynthesis through the arginine deiminase pathway (*arcA*, *arcB*) and from glutamate metabolism (*argB*, *argD*, *argJ*) are under-expressed (figure 6) when the growth rate increases. This profile has been confirmed at macromolecular level since ornithine production, though weak, was detected at the slowest growth rate. This result raises the importance of ornithine metabolism in the growth rate adaptation of *L. lactis*. However ornithine metabolism should be disconnected from cellular energy requirements since the mixed metabolism governed by energetic demand is not activated and maintenance is also constant.

Lastly, massive under-expression of genes encoding phage and prophage related functions and transposition proteins is observed in *L. lactis* when the growth rate is increased. Phages, prophages and transposons functions can play a role in bacterial genome diversification [29]. We thus hypothesise the cells to enter, at low growth rate, a state favouring DNA rearrangements. This assumption agrees with the theory developed for *E. coli* considering that mechanisms increasing genetic adaptation through variability may provide advantages in stressful conditions [1]. The *mutX* under-expression at low growth rate could also be associated to this state favouring genetic changes since disruption of this gene homologue in *Streptococcus pneumoniae* has been proved to increase A/T to C/G transversions [30].

Growth rate response: regulation

In order to better understand the growth rate regulations, the transcriptomic response to growth rate modifications has been compared to other transcriptomic responses through overlap calculations (Venn diagrams). The overlaps results were demonstrated to be

consistent through robustness analysis since they hardly differ with the statistical data selection criteria.

Cross-comparisons of growth rate response with isoleucine and carbon starvation responses give similar overlapping percentages (mean 27 %), indicating constant involvement of this mechanism in the different stress responses. The expression control of about 200 genes involved in these stress responses was thus related to growth rate modifications rather than to the stress itself. This result hence demonstrates the importance of the growth rate influence in the interpretation of all biological experiments involving growth rate decrease, *i.e.* when two environments are compared but also when two strains are analysed since mutants are most often affected in their growth capacity. Studying *E. coli* responses to nitrogen and sulphur limitations, Gyaneshwar *et al.* have found a number of genes with similar down-regulated expression profiles [31, 32]. Among these common genes, some, like those encoding ribosomal proteins or those involved in fatty acids metabolism, could thus have been linked to growth rate variations. Such a comparison with the growth rate influence would have allowed a more precise characterisation of the general stress response in the yeast.

Stringent response is considered to be a general mechanism allowing cells to adjust their major physiological process to growth deceleration, independently of the growth decrease origin [12-14]. However, the comparison of growth rate and stringent responses did not reveal the deep overlap expected if these mechanisms were identical or closely related. Indeed a weak overlap (10 % of the growth rate response) was obtained though the growth rate changes were similar in the two conditions. Moreover, more discrepancies (88 genes) than consistencies (75 genes) have been observed between the two responses. Functional analysis also revealed strong differences between the two responses and the under-regulation of

translation function, which is considered as the main trait of the stringent response [13], is even deeper extended in the growth rate response than in the stringent response. Thus stringent and growth rate responses correspond to two distinct mechanisms and the stringent response is not the general mechanism controlling growth rate modifications in *L. lactis*. This conclusion is also supported by the findings that stringent response overlaps with carbon starvation and isoleucine starvation responses are significantly different. With a 69 % overlap between stringent and isoleucine starvation responses, we can state that stringent response is included in the response to isoleucine starvation, which is not surprising since stringent response has been firstly described in *E. coli* response to amino acids starvation [22]. However, stringent response is only slightly involved in the response of *L. lactis* to glucose starvation (14 % of overlapping) though this mechanism is believed to be involved in carbon starvation in *B. subtilis* [33].

Two genes encoding (putative) regulatory proteins, *llrE* and *yabB*, belong to the growth rate core. Both are poorly studied: *llrE* encodes a regulatory protein of the OmpR family found to be involved phosphate activity regulation [34] and *yabB* encodes a putative protein that could belong to the Cro/Ci regulator family. Their expression is negatively controlled by growth rate in the various conditions. Bearing in mind that *L. lactis* lacks the alternative sigma factors which ensures, in other Gram positive bacteria, extended transcriptomic responses during stress, these two genes and their regulatory function deserve to be further explored. So far, no putative regulatory motif was identified since no specific DNA pattern could have been identified for growth rate regulated genes. Gene location on the chromosome appeared to be involved in the transcription regulation of the genes belonging to the growth rate response. This is notably the case for the genes whose expression follows the growth rate evolution. The distribution on the chromosome of this genes positively regulated by growth rate (cluster C)

indicates that genes whose expression increases with growth rate are mainly located close to the replication origin. The increased number of replication forks and thus the copy number of genes located in the vicinity of replication origin may contribute to this positive response. From the data obtained for *E. coli* [35], an increase of 1.5 of the replication origin copy number can be expected between 0.09 and 0.47 h⁻¹, which is closely similar to the expression changes measured in the cluster C. At the opposite, genes negatively controlled by growth rate are mostly located far away from the replication origin. A different conclusion had been drawn in *S. cerevisiae* since the genes positively and negatively regulated by growth rate had been found to be located adjacent to replication origins [9]. *L. lactis* specific organisation of the growth related genes suggests a link between replication and transcription control and opens new perspectives for gene expression studies considering that gene location on the chromosome can be an actor of expression regulation though it is generally neglected.

Conclusions

This work of integrative biology was performed at the global level using transcriptomic analysis obtained in various growth conditions. It raised the importance of growth rate regulations and provided, for the first time in a bacterium, the full description of the growth rate response. This work also participated to the elucidation of the mechanism involved in growth rate control. Though the mechanism is not yet fully understood in *L. lactis*, one expected regulatory mechanism (the stringent response) has been ruled out, two potential regulators (*llrE* and *yabB*) have been identified. Finally the involvement of gene location on the chromosome has also been found to be involved in the expression regulation of these growth related genes, opening new perspectives for gene expression studies.

Authors' contributions

CD, ER, PL and MC: conception, data acquisition, analysis of the data, drafting of the manuscript; CD, MC, HM and PB: statistical treatment of the data; all the authors read and approved the final manuscript.

Acknowledgments

We thank Nic Lindley for useful discussion.

This work was supported by the INRA Integrative Biology Program (agroBI).

References

1. Foster PL: **Stress responses and genetic variation in bacteria.** *Mutat Res* 2005, **569**(1-2):3-11.
2. Gupta S, Chatterji D: **Stress responses in mycobacteria.** *IUBMB Life* 2005, **57**(3):149-159.
3. Saito H, Kobayashi H: **Bacterial responses to alkaline stress.** *Sci Prog* 2003, **86**(Pt 4):271-282.
4. Wick LM, Egli T: **Molecular components of physiological stress responses in Escherichia coli.** *Adv Biochem Eng Biotechnol* 2004, **89**:1-45.
5. Hoskisson PA, Hobbs G: **Continuous culture--making a comeback?** *Microbiology* 2005, **151**(Pt 10):3153-3159.
6. Boer VM, Tai SL, Vuralhan Z, Arifin Y, Walsh MC, Piper MD, de Winde JH, Pronk JT, Daran JM: **Transcriptional responses of Saccharomyces cerevisiae to preferred and nonpreferred nitrogen sources in glucose-limited chemostat cultures.** *FEMS Yeast Res* 2007, **7**(4):604-620.
7. Brauer MJ, Huttenhower C, Airoidi EM, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D: **Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast.** *Mol Biol Cell* 2007.
8. Castrillo JI, Zeef LA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell MJ, Petty J, Hakes L, Wardleworth L *et al*: **Growth control of the eukaryote cell: a systems biology study in yeast.** *J Biol* 2007, **6**(2):4.
9. Regenberg B, Grotkjaer T, Winther O, Fausboll A, Akesson M, Bro C, Hansen LK, Brunak S, Nielsen J: **Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae.** *Genome Biol* 2006, **7**(11):R107.
10. Tai SL, Daran-Lapujade P, Walsh MC, Pronk JT, Daran JM: **Acclimation of Saccharomyces cerevisiae to low temperature: a chemostat-based transcriptome analysis.** *Mol Biol Cell* 2007, **18**(12):5100-5112.
11. Berney M, Weilenmann HU, Ihssen J, Bassin C, Egli T: **Specific growth rate determines the sensitivity of Escherichia coli to thermal, UVA, and solar disinfection.** *Appl Environ Microbiol* 2006, **72**(4):2586-2593.
12. Braeken K, Moris M, Daniels R, Vanderleyden J, Michiels J: **New horizons for (p)ppGpp in bacterial and plant physiology.** *Trends Microbiol* 2006, **14**(1):45-54.

13. Jain V, Kumar M, Chatterji D: **ppGpp: stringent response and survival.** *J Microbiol* 2006, **44**(1):1-10.
14. Chang DE, Smalley DJ, Conway T: **Gene expression profiling of Escherichia coli growth transitions: an expanded stringent response model.** *Mol Microbiol* 2002, **45**(2):289-306.
15. Eymann C, Homuth G, Scharf C, Hecker M: **Bacillus subtilis functional genomics: global characterization of the stringent response by proteome and transcriptome analysis.** *J Bacteriol* 2002, **184**(9):2500-2520.
16. Bolotin A, Wincker P, Mauger S, Jaillon O, Malarme K, Weissenbach J, Ehrlich SD, Sorokin A: **The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403.** *Genome Res* 2001, **11**(5):731-753.
17. Otto R, ten Brink V, Veldkamp H, Koenings WN: **The relationship between growth rate and electrochemical proton gradient of Streptococcus cremoris.** *FEMS Microbiol Lett* 1983(16):69-74.
18. Poolman B, Konings WN: **Relation of growth of Streptococcus lactis and Streptococcus cremoris to amino acid transport.** *J Bacteriol* 1988, **170**(2):700-707.
19. Redon E, Loubiere P, Coccagn-Bousquet M: **Role of mRNA stability during genome-wide adaptation of Lactococcus lactis to carbon starvation.** *J Biol Chem* 2005, **280**(43):36380-36385.
20. Mrazek J, Gaynon LH, Karlin S: **Frequent oligonucleotide motifs in genomes of three streptococci.** *Nucleic Acids Res* 2002, **30**(19):4216-4221.
21. Rallu F, Gruss A, Maguin E: **Lactococcus lactis and stress.** *Antonie Van Leeuwenhoek* 1996, **70**(2-4):243-251.
22. Cashel M, Rudd K: **The stringent response.** In: *In Escherichia coli and Salmonella typhimurium*. Edited by F.C., vol. 2: Neidhart; 1987: 1410-1438.
23. Redon E, Loubiere P, Coccagn-Bousquet M: **Transcriptome analysis of the progressive adaptation of Lactococcus lactis to carbon starvation.** *J Bacteriol* 2005, **187**(10):3589-3592.
24. Le Cao KA, Goncalves O, Besse P, Gadat S: **Selection of biologically relevant genes with a wrapper stochastic algorithm.** *Stat Appl Genet Mol Biol* 2007, **6**:Article29.
25. Coccagn-Bousquet M, Garrigues C, Novak L, Lindley ND, Loubière P: **Rational development of a simple synthetic medium for the sustained growth of Lactococcus lactis.** *Journal of applied bacteriology* 1995, **79**:108:116.

26. Even S, Lindley ND, Coccagn-Bousquet M: **Transcriptional, translational and metabolic regulation of glycolysis in *Lactococcus lactis* subsp. cremoris MG 1363 grown in continuous acidic cultures.** *Microbiology* 2003, **149**(Pt 7):1935-1944.
27. Loubière P, Coccagn-Bousquet M, Matos J, Goma G, Lindley ND: **Influence of end-products inhibition and nutrient limitations on the growth of *Lactococcus lactis* subsp. *lactis*.** *Journal of applied microbiology* 1997, **82**:95-100.
28. Garrigues C, Mercade M, Coccagn-Bousquet M, Lindley ND, Loubiere P: **Regulation of pyruvate metabolism in *Lactococcus lactis* depends on the imbalance between catabolism and anabolism.** *Biotechnol Bioeng* 2001, **74**(2):108-115.
29. Canchaya C, Fournous G, Brussow H: **The impact of prophages on bacterial chromosomes.** *Mol Microbiol* 2004, **53**(1):9-18.
30. Bullions LC, Mejean V, Claverys JP, Bessman MJ: **Purification of the MutX protein of *Streptococcus pneumoniae*, a homologue of *Escherichia coli* MutT. Identification of a novel catalytic domain for nucleoside triphosphate pyrophosphohydrolase activity.** *J Biol Chem* 1994, **269**(16):12339-12344.
31. Gyaneshwar P, Paliy O, McAuliffe J, Jones A, Jordan MI, Kustu S: **Lessons from *Escherichia coli* genes similarly regulated in response to nitrogen and sulfur limitation.** *Proc Natl Acad Sci U S A* 2005, **102**(9):3453-3458.
32. Gyaneshwar P, Paliy O, McAuliffe J, Popham DL, Jordan MI, Kustu S: **Sulfur and nitrogen limitation in *Escherichia coli* K-12: specific homeostatic responses.** *J Bacteriol* 2005, **187**(3):1074-1090.
33. Bernhardt J, Weibezahn J, Scharf C, Hecker M: ***Bacillus subtilis* during feast and famine: visualization of the overall regulation of protein synthesis during glucose starvation by proteome analysis.** *Genome Res* 2003, **13**(2):224-237.
34. O'Connell-Motherway M, van Sinderen D, Morel-Deville F, Fitzgerald GF, Ehrlich SD, Morel P: **Six putative two-component regulatory systems isolated from *Lactococcus lactis* subsp. cremoris MG1363.** *Microbiology* 2000, **146** (Pt 4):935-947.
35. Bipatnath M, Dennis PP, Bremer H: **Initiation and velocity of chromosome replication in *Escherichia coli* B/r and K-12.** *J Bacteriol* 1998, **180**(2):265-273.

Figures legends

Figure 1: Biomass production yields from amino acids in *L. lactis* IL1403 grown in continuous culture at different growth rates.

Amino acids with positive yield are consumed by the bacteria whereas those with negative yields are produced.

Figure 2: Average expression profiles of genes classified in the different clusters during the *L. lactis* IL1403 continuous culture at different growth rate.

X axis: compared growth rate; Y axis: average expression ratios; $\mu=0.09 \text{ h}^{-1}$ is considered as the reference. Error bars represent the 5 % confidence interval.

Figure 3: Distribution of genes from clusters C and D along *Lactococcus lactis* IL1403 chromosome.

Dark bars indicate the global distribution of genes along the chromosome and the straight lines represent the normal one. Light grey and white bars are respectively representatives of the chromosomal localization of genes belonging to clusters C and D. The chromosome has been divided in 4 (upper graph) and 15 (lower graph) parts.

Figure 4: Venn diagrams comparing the growth rate transcriptomic response with other stress responses in *L. lactis* IL1403.

Growth rate (μ) is compared with stringent response (SR) (a), isoleucine starvation (IS) (b), and carbon starvation (CS) (c). + and – respectively represents genes whose expression increases or decreases when growth rate decreases.

The missing probes of the various transcriptomic datasets were removed from the diagrams.

Figure 5: Venn diagrams comparing the stringent transcriptomic response with other stress responses in *L. lactis* IL1403.

Stringent response (SR) is compared with isoleucine starvation (IS) response (a) and carbon starvation (CS) response (b). + and – respectively represents over- and under- expressed genes in the different responses.

The missing probes of the various transcriptomic datasets were removed from the diagrams.

Figure 6: Ornithine biosynthesis pathways in *L. lactis*.

Ornithine can be synthesised from glutamate or from arginine through the arginine deiminase pathway (ADI pathway). Thick arrows indicate up regulation of the corresponding genes upon growth rate decrease.

Table 1: Composition of the culture medium

The lines separating the different products indicate those that are dissolved together. Values into brackets in front of KH_2PO_4 and K_2HPO_4 indicate the required concentrations to obtain a buffered media for pre-cultures

| Product | Concentration (g/L) | Product | Concentration (g/L) |
|--------------------------------------|---------------------|------------------------|---------------------|
| Glucose | 20 | Isoleucine | 0.025 |
| Sodium acetate | 1 | Leucine | 0.05875 |
| Ammonium citrate | 0.6 | Valine | 0.04125 |
| KH_2PO_4 | 3 (9) | Cysteine | 0.17 |
| K_2HPO_4 | 2,5 (7,5) | Phenylalanine | 0.28 |
| $\text{MgCl}_2, 6\text{H}_2\text{O}$ | 0.2 | Tyrosine | 0.29 |
| $\text{FeSO}_4, 7\text{H}_2\text{O}$ | 0.011 | Adenine | 0.05 |
| $\text{CaCl}_2, 2\text{H}_2\text{O}$ | 0.05 | Guanine | 0.05 |
| $\text{ZnSO}_4, 7\text{H}_2\text{O}$ | 0.005 | Uracil | 0.05 |
| $\text{COCl}_2, 6\text{H}_2\text{O}$ | 0.0025 | Xanthin | 0.05 |
| Alanine | 0.24 | P-aminobenzoic acid | 0.01 |
| Arginine | 0.12 | Biotin | 0.01 |
| Asparagine | 0.34 | Cyano-cobalamine (B12) | 0.001 |
| Glutamine | 0.51 | Folic acid | 0.01 |
| Glycine | 0.17 | Inosine | 0.025 |
| Histidine | 0.11 | Nicotinic acid | 0.001 |
| Lysine | 0.35 | Orotic acid | 0.005 |
| Methionine | 0.12 | Ca panthotenate | 0.001 |
| Proline | 0.68 | Pyridoxamine | 0.005 |
| Serine | 0.34 | Pyridoxin (B6) | 0.002 |
| Threonine | 0.23 | Riboflavin (B2) | 0.001 |
| Tryptophan | 0.05 | Thiamine | 0.001 |
| | | D,L-6,8-thioctic acid | 0.0025 |
| | | Thymidine | 0.025 |

Table 2: Macrokinetic parameters of *L. lactis* continuous cultures calculated from at least 12 different samples collected at the different growth rates.

| D (h ⁻¹) | 0.09 | 0.24 | 0.35 | 0.47 |
|--|---------------|---------------|---------------|---------------|
| [X] (g/L) | 0.76 +/- 0.02 | 0.79 +/- 0.01 | 0.75 +/- 0.01 | 0.75 +/- 0.01 |
| [isoleucine] (μM) | <10 | <10 | <10 | <10 |
| [glucose] (mM) | 43.23 +/- 4 | 70.9 +/- 2 | 81.05 +/- 5 | 84.41 +/- 2 |
| [lactate] (mM) | 116.52 +/- 6 | 77.57 +/- 4 | 57.12 +/- 6 | 51.85 +/- 2 |
| Y _{Gluc, Lact} (mol Lact/mol Gluc) | 1.7 | 1.8 | 1.7 | 1.6 |
| qGluc (mmol/gX/h) | 8.28 +/- 1 | 12.99 +/- 1 | 15.81 +/- 2 | 20.25 +/- 1 |
| vLact (mmol/gX/h) | 13.33 +/- 1 | 22.47 +/- 1 | 25.42 +/- 3 | 31.54 +/- 1 |

D = dilution rate, [] = residual concentration, X = biomass, Y_{Gluc, Lact} = lactate yield from glucose, qGluc = specific glucose consumption rate, vLact = specific lactate production rate.

Table 3: Overlapping of the different transcriptional responses with different selection criteria on datasets

| Compared conditions | Overlapping percentages | | | | | |
|---------------------|-------------------------|---------------------|----------------------|-----------------|---------------|---------------|
| | p-values <0.05 | | | ratios <1 or >1 | | |
| | ratios <1 or >1 | ratios <0.8 or >1.2 | ratios <0.67 or >1.5 | p-values<0.05 | p-values<0.03 | p-values<0.02 |
| μ and CS | 29% | 29% | 31% | 29% | 25% | 21% |
| μ and IS | 24% | 23% | 20% | 24% | 22% | 22% |
| μ and SR | 10% | 9% | 5% | 10% | 11% | 9% |
| SR and μ | 16% | 15% | 8% | 16% | 18% | 14% |
| SR and CS | 15% | 15% | 17% | 15% | 14% | 11% |
| SR and IS | 68% | 70% | 72% | 68% | 65% | 59% |

Values in percentage represent the fraction of common genes compared to the whole growth rate response (for the three first lines) or the whole stringent response (for the three last lines). Abbreviations: SR = stringent response regulon, CS = carbon starvation response, IS = isoleucine starvation response, μ = growth rate response

**Chapitre II : Adaptation dynamique de
Lactococcus lactis à la carence en isoleucine**

L'assimilation des acides aminés est cruciale pour la survie et le développement des bactéries. Principales sources d'azote, ils sont aussi impliqués dans la production d'énergie, le contrôle du pH intracellulaire et la régénération des cofacteurs (Ardo, 2006). Ce besoin est d'autant plus important pour les bactéries lactiques qui présentent de nombreuses auxotrophies vis-à-vis des acides aminés (Cocaign-Bousquet *et al.*, 1995). Toutefois, pour ces bactéries d'intérêt industriel, l'étude du catabolisme des acides aminés est essentiellement restée centrée sur leur assimilation lors des culture en lait, souvent limitées en acides aminés libres, et leur rôle dans la production de composés aromatiques (Fernandez & Zuniga, 2006).

Dans cette publication, nous nous sommes attachés à la caractérisation de la réponse de *Lactococcus lactis*, modèle des bactéries lactiques, à la carence en isoleucine. Cette réponse a été approchée de manière globale et dynamique, au travers du suivi cinétique, macroscopique, transcriptomique et protéomique. La carence a été imposée progressivement suite à la consommation naturelle de l'isoleucine par la bactérie lors d'une culture discontinue en batch. L'implication de différents mécanismes susceptibles de réguler cette réponse a été évaluée au travers de comparaisons transcriptomiques avec des régulons connus : la réponse stringente et la réponse à la diminution du taux de croissance ont été définies dans nos travaux antérieurs (cf. chapitre précédent) tandis que les régulons CodY (Guedon *et al.*, 2001), GlnR (Larsen *et al.*, 2006) et CcpA (Zomer *et al.*, 2007) ont déjà été décrits dans la littérature. Bien que ces deux derniers régulons aient été obtenus avec la souche MG1363 que les autres études sont réalisées avec la souche *L. lactis* IL1403, la proximité phylogénétique de ces deux souches (Kok *et al.*, 2005) suggère que les différences de composition des régulons restent mineures.

A l'échelle macro-cinétique, l'adaptation de *L. lactis* à l'épuisement progressif de l'isoleucine du milieu de culture se traduit par l'apparition d'un ralentissement de croissance avant même l'épuisement complet de l'isoleucine. De façon étonnante en regard de l'auxotrophie de *L. lactis* pour l'isoleucine (Cocaign-Bousquet *et al.*, 1995), une croissance, certes très faible ($\mu = 0.05 \text{ h}^{-1}$), se poursuit même lorsque l'état de carence est atteint. La confirmation de ce résultat par transfert successif dans un milieu dépourvu d'isoleucine suggère que la voie de biosynthèse de cet acide aminé initialement décrite comme inactive (Godon *et al.*, 1993) pourrait néanmoins présenter une activité résiduelle suffisante pour permettre une croissance lente mais continue de la bactérie.

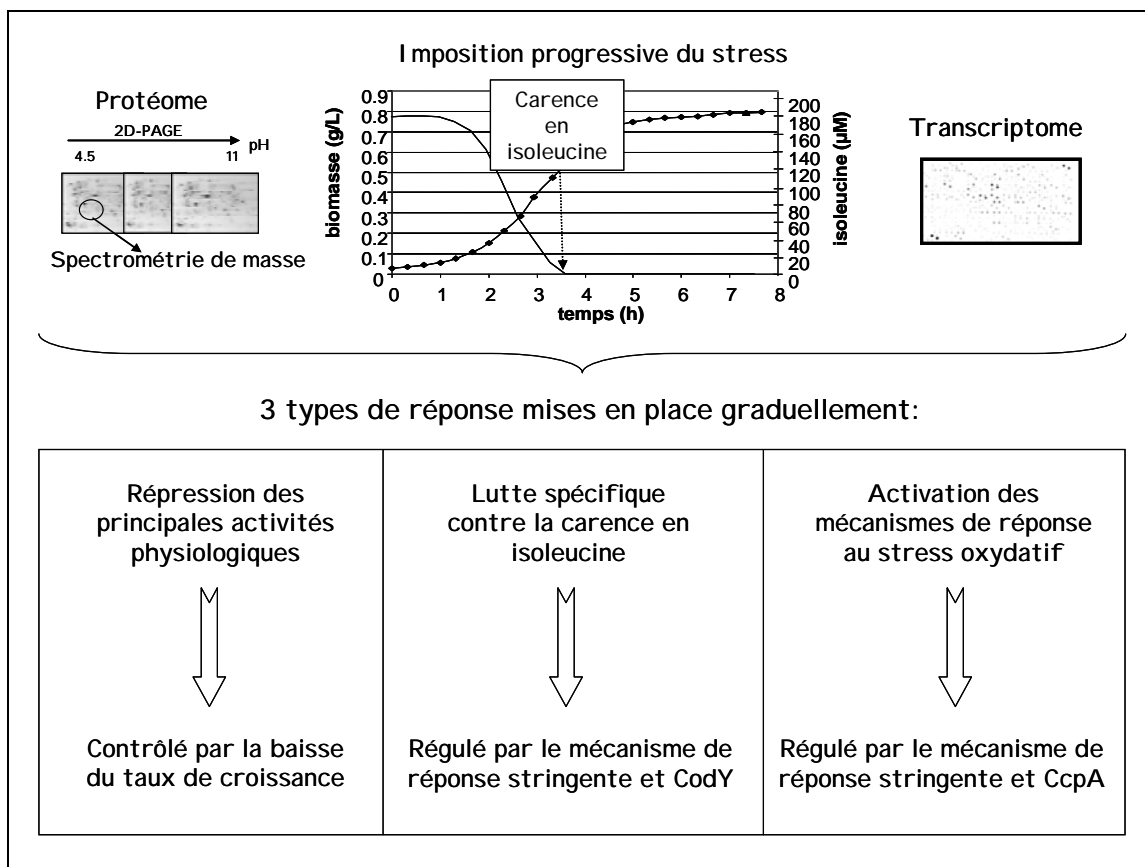


Figure 9 : Analyse globale de l'adaptation de *L. lactis* à la carence en isoleucine et de son contrôle.

Les résultats de transcriptome, corroborés par ceux de protéome, ont permis de visualiser la mise en place graduelle d'une réponse qui peut être scindée en trois mécanismes principaux (figure 9). Tout d'abord une répression globale des activités cellulaires liées à la croissance bactérienne et au catabolisme a pu être observée avec la diminution des niveaux de transcrits et de protéines de gènes associés à la traduction, le métabolisme des pyrimidines et des acides gras ainsi qu'au transport et au métabolisme des sucres. La mise en place de cette réponse apparaît contrôlée par la diminution du taux de croissance mais également par le mécanisme de réponse stringente. L'activation de l'activité peptidase, du transport des acides aminés, et de plusieurs voies associées aussi bien au métabolisme azoté que carboné tend à orienter le métabolisme global vers la production d'isoleucine et peut donc être considérée comme une réponse spécifique à la carence imposée. Cette mise en place vise à contrecarrer l'effet du stress subi et semble régulée à la fois par le mécanisme de réponse stringente, connu pour son implication dans la réponse à la carence en acide aminé (Traxler *et al.*, 2008) et CodY, le répresseur de transcription associé au métabolisme azoté. L'activité de ce dernier chez *L. lactis* est d'ailleurs dépendante du pool d'acides aminés branchés (Guedon *et al.*, 2001).

Enfin, de façon plus inattendue dans des conditions de culture strictement anaérobies, une surexpression de transcrits et de protéines habituellement impliqués dans la réponse au stress oxydatif a pu être notée. En effet le métabolisme aérobie, le transport d'électrons, le métabolisme de la thiorédoxine et les différentes sous-unités de la pyruvate déshydrogénase sont apparus comme induits en réponse à la carence en isoleucine. La régulation de cette dernière réponse implique encore une fois le mécanisme stringent. Elle semble également en partie contrôlée par le régulateur de la répression catabolique et du métabolisme carboné CcpA dont l'implication dans la réponse à la carence en acide aminé a aussi été suggérée chez *Bacillus subtilis* (Tam le *et al.*, 2007).

La complémentarité des analyses protéomiques et transcriptomiques associées à un suivi macro-cinétique de la culture a donc permis d'obtenir un vision globale de l'adaptation de *L. lactis* à la carence en isoleucine. En outre, l'approche de transcriptomique comparative proposée dans cette publication a mis à jour des réseaux de régulation complexes. Cette approche, très informative, présente l'avantage d'être entièrement applicable à d'autres conditions environnementales aussi bien qu'à d'autres microorganismes dont on chercherait à déterminer les mécanismes de régulation.

Dynamic transcriptome and proteome integration to investigate the adaptation of *Lactococcus lactis* to isoleucine starvation

**Clémentine Dressaire^{†1,2,3}, Emma Redon^{†1,2,3}, Christophe Gitton⁴, Pascal Loubière^{1,2,3},
Véronique Monnet⁴, Muriel Cocaign-Bousquet^{1,2,3*}**

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

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

(3) CNRS, UMR5504, F-31400 Toulouse, France

(4) INRA, Unité de Biochimie Bactérienne, UR477, F-78350 Jouy en Josas, France

[†] These authors contributed equally to this work.

*corresponding author: e-mail: cocaign@insa-toulouse.fr

ISBP, INSA 135 Avenue de Rangueil,

31077 Toulouse cedex 4, France.

Tel: (33) 561 559 438

Fax: (33) 561 559 400

ABSTRACT

Amino acids assimilation is crucial for bacterial growth and survival and this is particularly obvious for lactic acid bacteria (LAB) that are generally auxotroph for amino acids. The global response of the LAB model *Lactococcus lactis* was characterized during progressive isoleucine starvation in batch culture. The stress was imposed by isoleucine natural consumption in an initially rich chemically defined medium. Dynamic analyses were performed using both transcriptomic and proteomic approaches. The response was found to occur gradually and could be divided into three major mechanisms that were firstly deduced from transcriptomic analysis and generally corroborated by proteomic results: (i) a global down-regulation of processes linked to bacterial growth and catabolism (transcription, translation, carbon metabolism and transport, pyrimidines and fatty acid metabolisms), (ii) a specific positive response related to the limiting nutrient (activation of pathways belonging to carbon or nitrogen metabolism and leading to isoleucine supply) and (iii) an unexpected oxidative stress response (positive regulation of aerobic metabolism, electron transport, thioredoxin metabolism and pyruvate dehydrogenase). Then, the involvement of various regulatory mechanisms (growth rate and stringent related regulations, CodY, GlnR and CcpA regulations) during this adaptation was analysed on the basis of transcriptomic data comparisons.

KEYWORDS

Transcriptome, proteome, dynamic adaptation, amino-acid starvation, *Lactococcus lactis*, regulation and metabolism

INTRODUCTION

Amino acid catabolism has been extensively studied in lactic acid bacteria (LAB) because of their ability to produce flavor compounds through amino acid degradation^{1, 2}. These studies mainly focused on relationships between amino acids and organoleptic properties of fermented food. Amino acids are also of physiological importance because of their involvement in intracellular pH control, energy production and cofactor regeneration¹⁻³. Moreover, amino acids are required for protein synthesis and their assimilation is especially crucial for LAB that have numerous nutritional auxotrophies⁴⁻⁶. More particularly in the LAB model, *Lactococcus lactis* IL1403, no less than 7 amino acid auxotrophies have been detected⁴. These auxotrophies are generally related to punctual mutations or gene inactivations^{5, 6}.

In natural environments as well as in industrial complex media, LAB have to face many amino acid deprivations, even though they manage to metabolize complex nitrogen sources^{7, 8}. However, a complete study of the response of a LAB during amino acid starvation is still lacking and more generally amino acid assimilation is poorly characterized in bacteria. Genome wide analysis of amino acid starvation has been recently reported for *Bacillus subtilis*⁹, *Bordetella pertusis*¹⁰ and *Escherichia coli*¹¹ revealing an unexpectedly complex transcript pattern. The well known general mechanism of stringent response, which is mediated by RelA protein and (p)ppGpp alarmone¹²⁻¹⁵, was involved but other mechanisms such as sigma like factor regulations or other unidentified regulatory mechanisms were also assumed to interfere. The regulatory networks coordinating the expression of genes involved in amino acid starvation adaptation process are only partially understood in bacteria. This lack of knowledge is particularly obvious in *L. lactis* since it does not encode any stress-response like sigma factor¹⁶. The mechanism of stringent response, recently investigated in *L. lactis* by transcriptomic approach after norvaline induction¹⁷, could participate to the amino acid starvation response as well as a general mechanism associated to growth rate regulation since

depletion should be associated to growth deceleration ¹⁷. Moreover the CodY regulator, whose activity was shown to be mediated by the branched chain amino acids (BCAA) pool in *L. lactis* ¹⁸, could also be involved, as well as the GlnR protein which was recently identified as a nitrogen metabolism regulator in *L. lactis* ¹⁹. Furthermore, CcpA, the carbon catabolite repressor, could also participate in the response, since it was found to be involved in *B. subtilis* response to nitrogen starvation ⁹. The belief of a multi-regulated response is thus emerging today though more evidences should be given in the future to be completely admitted.

The aim of the paper is to provide a full description of the response of *L. lactis* to isoleucine starvation and to analyse the involvement in this response of various potential regulations. Isoleucine with leucine and valine, the two other branched chain amino acids (BCAA), constitute the most abundant amino acids in proteins and are major components of the hydrophobic protein domains ^{20, 21}. The response of *L. lactis* to isoleucine starvation was studied dynamically with an integrative approach combining macro kinetic parameters monitoring and both transcriptome and proteome measurements. The stress was gradually imposed in discontinuous cultures by bacterial BCAA consumption. Then, using previously described regulons, the involvement of various potential regulations in this response has been estimated through comparative transcriptomic.

EXPERIMENTAL SECTION

Organism and growth conditions

Lactococcus lactis ssp. *lactis* IL1403, whose genome has been entirely sequenced ¹⁶, was used throughout this study. Bacteria were grown in modified chemically defined medium ²² with ten-fold reduced BCAA concentrations ([isoleucine] = 150 μ M, [leucine] = 360 μ M, [valine] = 280 μ M). At least 3 independent repetitions of the culture were analyzed. Cultures were performed at 30 °C under anaerobic conditions in a 2 L fermentor (Setric Génie Industriel,

Toulouse, France) with an agitation speed of 250 rpm. The pH was maintained at 6.6 by automatic addition of KOH 10 N. Bacterial growth was estimated by spectrophotometric measurement at 580 nm (1 OD unit is equivalent to 0.3 g.l⁻¹). Inoculation was realized with exponential phase cells from pre-cultures grown in the same medium so as to obtain an initial OD₅₈₀ of 0.08.

Analytical methods

Glucose and fermentation products (lactate, formate, acetate and ethanol) concentrations were measured by high-pressure liquid chromatography as previously described ²³. Amino acids concentrations were determined with an AminoQuant 1090 high-pressure liquid chromatography (Hewlett Packard). After precipitation of proteins at 4 °C by addition of four volumes of methanol, samples were chemically modified (derivatization in presence of 3-mercaptopropionic acid by ortho-phthalaldehyde and 9-fluorenylmethyl chloroformate for primary and secondary amino acids respectively), separated with a C18 column and detected by spectrophotometry at 338 and 262 nm.

Transcription activity assay

To evaluate transcription activity, [³H]-uridine incorporation rate into RNA was measured with a method adapted from Lin *et al.* ²⁴. In exponential growth phase or during isoleucine starvation, [³H]-uridine (5 μCi.ml⁻¹) and nalidixic acid (20 μg.ml⁻¹, avoid uridine incorporation into DNA) were added in the culture. [³H]-uridine incorporation was monitored over a time course of 10 min at 30 seconds intervals. 200 μl-samples were collected, precipitated in 800 μl of 20 % (w/v) TCA, filtered on GF/C Whatman papers after 30-min incubation on ice, washed with 5 % (w/v) ice-cold TCA (3 × 5 ml) and absolute ethanol (1 × 5 ml) and then dried. Radioactivity incorporated into precipitated RNA was measured with a Beckman scintillation counter. The [³H]-uridine incorporation into RNA rate was defined as

the slope of the plot of the radioactivity level versus time of incorporation, divided by the initial biomass concentration.

Transcriptomic analyses

Membrane spotting and analytical support were provided by the Biochips Platform (Genopole Toulouse, France). Cell samples were harvested from culture in exponential growth phase and after 20 min, 1.34 h and 3 h of amino acid starvation. Cell lysis and total RNA extraction were performed as previously described 25. RNA was quantified at 260 nm and RNA quality was controlled both on electrophoresis agarose gel in denaturing conditions and on Agilent Bioanalyzer. Gene expression was measured by using nylon arrays displaying 1948 ORFs of *L. lactis* IL1403 out of the 2310 ORFs identified in the genome 16. The PCR set specific for *L. lactis* IL1403 was provided by Eurogentec. A constant amount of 10 µg of total RNA was retro-transcribed. Synthesis of radio-labeled cDNA, nylon arrays hybridizations and washings were performed as previously described 25. Membranes were exposed to a phosphoimager screen for three days and scanned with a phosphofluoroimager (Storm 860, Molecular Dynamics). Hybridization signals were quantified, assigned to gene names, and statistically treated with the Biplot software (developed by S. Sokol in the Biochips Platform, Toulouse, see <http://biopuce.insa-toulouse.fr>). For each condition, three repetitions were performed with independent cultures, extractions, labeling and hybridizations. After background removal, raw data, available online in GEO database (series accession number: GSE12962, the already available platform GSE4872 also corresponding to isoleucine starvation was not used because biological replicates were not independent), were normalized by the mean intensity of the corresponding membrane. Expression ratios were calculated using the exponential phase as a reference. The statistical significance of expression ratios were evaluated using False Discovery Rate (FDR) calculations. Genes with an expression ratio associated to a FDR lower than 10 % were considered differentially expressed. Each of these genes displayed a p-value

inferior to 0.05. In order to determine expression changes at the level of the functional (sub)categories (global tendencies), over- or under-expressed genes enrichment in the different groups were calculated with the Wilcoxon test and a p-value lower than 0.05 was considered to be significant.

Quantitative RT-PCR

In order to confirm transcriptomic results RT-PCR measurements were performed in exponential phase and after 1.34 h of isoleucine starvation. Nine over- or under-expressed genes and involved in various functional categories (*adhE*, *busAB*, *clpE*, *glpF1*, *ilvD*, *optC*, *ptnAB*, *rplM* and *rpoB*) were tested. 10 µg of RNA samples extracted from cells were retro-transcribed, RNase H-treated and purified as previously described²⁵ except that unlabelled dCTP and only random primers were used. Primers pairs were designed with Beacon Designer Biorad software to fulfill the following criteria: length of 22 ± 2 bp, T_m of 60 ± 1 °C, GC content superior to 50 % and amplicons length ranging from 75 to 150 bp (table 1). Primers specificity was controlled with Blast analyses. Real-time PCR was carried out on a Biorad-MyIQ with the IQTM SYBR[®] supermix in 96-well plates. 5 µl of diluted cDNA were added to 20 µl of PCR mixture (12.5 µl of IQTM SYBR[®] supermix, 2 µl of each primer at 5 µM and 3.5 µl of Rnase-free water). Amplification program consisted in a denaturation at 95 °C for 3 min followed by 40 cycles at 95 °C for 15 s and 60 °C for 45 s. Fluorescence was measured during each annealing step and melting curves were performed from 60 to 95 °C (0.05 °C.s^{-1}) to validate the specificity of PCR reaction. In each run, three dilutions of cDNA were analyzed to determine the PCR efficiency and negative controls were included. Three independent measurements were performed for each gene and culture condition. The threshold cycle values (C_T) were determined with a baseline set manually at 135 relative fluorescence units (baseline above background and corresponding to PCR efficiencies ranging from 90 to 110 %). Results were analyzed using the comparative critical threshold method

($\Delta\Delta C_T$) in which the amount of target RNA is adjusted to an internal reference²⁶. *ldh* gene, encoding lactate dehydrogenase, did not show significant expression variation in these experiments and was used as an internal reference to normalize the results. Expression ratios were expressed as $2^{\Delta\Delta C_T}$.

Proteomic analyses

For each condition, three repetitions were performed with independent cultures, extractions and electrophoresis. Bacteria were harvested from and cell pellets were washed twice with ice-cold 200 mM Na-phosphate, pH 6.4 and re-suspended in 4 ml of 20 mM Na-phosphate buffer, pH 6.4, 1 mM EDTA, 10 mM tributylphosphine, a cocktail of protease inhibitors (P8465; Sigma Aldrich, St Louis, MO) 20-fold diluted and catalase 40 U/mL (C3155; Sigma Aldrich, St Louis, MO) to limit isoform formation. The cell suspension (approximately 35 units of optical density at 600 nm [OD600]/ml) was transferred to the pre-cooled chamber of a BASIC Z cell disrupter (CellD, Warwickshire, United Kingdom) and was subjected to a pressure of 2,500 bars. The suspension was centrifuged at 5,000 x g for 20 min at 4 °C to remove unbroken cells and large cellular debris. The supernatant was collected and centrifuged at 220,000 x g for 30 min at 4 °C. The total protein concentration in the resulting supernatant (cytosolic fraction) was determined with the Coomassie protein assay reagent (Pierce, Rockford, IL) using bovine serum albumin as standard and was included between 1 and 2 mg/mL. The cytosolic fraction was aliquoted and stored frozen at -20 °C.

2-Dimension electrophoresis: A volume of cytosolic fraction corresponding to 350 µg or 500 µg (for basic gels) of proteins was incubated with nuclease (benzonase, Novagen 70664-3; 25 U for 100 µL of cytosolic fraction) 30 min at 37 °C and then thawed on ice and precipitated with 75 % (vol/vol) methanol. The protein pellet was resuspended in 500 µl (for pH 4.5-5.5 and 5-6 gels) or 100 µl (for pH 6-11 gels) of isoelectric focusing (IEF) buffer 1, consisting of 7 M urea, 2 M thiourea, 4% CHAPS was not, 100 mM dithiothreitol or 4 mM

tributylphosphine and DeStreak (1.2 % v/v, Amersham Biosciences, GE Healthcare) (for basic gels), and 0.5 % pH 4.5 to 5.5 or 5 to 6 or 6 to 11 immobilized pH gradient (IPG) buffer (Amersham Biosciences, GE Healthcare). The sample was loaded on 24 cm pH 4.5 to 5.5 or 5 to 6 immobilized pH gradient (IPG) strip (Amersham Biosciences, GE Healthcare) which was previously rehydrated at 50 V for 11 h. IEF was carried out for 65,000 Vh at a maximum of 8,000 V, using the Protean II IEF cell (Bio-Rad, Hercules, CA). Analysis of basic proteins was performed with 18 cm pH 6-11 (IPG) strip (Amersham Biosciences, GE Healthcare). After passive rehydration of the strip in buffer 1; the protein sample was loaded on sample cups and IEF was carried out for 20000 V.h at a maximum of 3500 V using the IPGphor device (Amersham Biosciences, GE Healthcare). Before the second dimension, IPG strips were incubated for 15 min with shaking in 150 mM Tris-HCl pH 8.8, 0.1 % w/v SDS. The IPG strip was then positioned on sodium dodecyl sulfate-polyacrylamide gels, using 1% low-melting-point agarose in 150 mM Tris-HCl, pH 8.8. Second-dimension electrophoresis was performed on 12% polyacrylamide gels (24 by 20 by 0.1 cm) in 25 mM Tris, 192 mM glycine, 0.1% sodium dodecyl sulfate, pH 8.3, using the Ettan-Dalt II apparatus. Electrophoresis was run at 1 W/gel for 16 h at 15 °C. The gels were stained with BioSafe colloidal Coomassie blue (Bio-Rad) for 1 h and destained with three successive washes in deionised water.

Gel images were generated using an Epson Expression 1640XL scanner controlled by Silver Fast software. Image files were recorded at 256 gray levels (8 BitsPerPixel). Image manipulation and analysis were performed with Phoretix 2D v software (Nonlinear Dynamics). Protein abundances were given using arbitrary units which correspond to spot volumes and which were calculated as follows: spot area x spot pixel intensity - background intensity.

Protein identification: Protein identification was carried out at the PAPPSO platform (INRA, Jouy-en-Josas). Most proteins were identified by MALDI-TOF mass spectrometry. Protein spots were excised from Coomassie blue-stained gels and in-gel digested with trypsin. Gel pieces were placed in Eppendorf tubes and washed with 30 μ L 25 mM ammonium carbonate, 50 % acetonitrile. The supernatants were discarded and gel pieces were dried at 37 °C for 15 min. The gels were rehydrated with 20 μ L 50 mM ammonium carbonate containing 100 ng of porcine trypsin (Promega, Madison, WI, USA). The solutions were incubated overnight at 37 °C. The supernatants containing peptides were directly analyzed by MALDI-TOF Mass spectrometry on a Voyager DE STR Instrument (Applied Biosystems, Framingham, CA, USA). The α -cyano-4-hydroxycinnamic acid matrix was prepared at 4mg/mL in 0.1% TFA, 50 % acetonitrile. An equal volume (1 μ L) of matrix and sample were spotted onto the MALDI-TOF target plate. Spectra were acquired in the reflector mode with the following parameters: 2000 laser intensity, 20 kV accelerating voltage, 62 % grid voltage, 120 ns delay. The mass gates used were 840–3500 Da. Internal calibration was performed by using the trypsin peptides at 842.5 and 2211.1 Da. Database searches were conducted with the MS-Fit software (<http://prospector.ucsf.edu>) either on an *L. lactis*-specific protein database.

The few spots which remained unidentified by MALDI-TOF were analysed by LC-MS/MS. LC-MS/MS analysis was performed on an Ultimate 3000 LC system (Dionex, Voisins le Bretonneux, France) connected to a linear ion trap mass spectrometer (LTQ, Thermo Fisher, USA) by a nanoelectrospray interface to realize the separation, ionisation and fragmentation of peptides, respectively.

The supernatant of trypsin hydrolysis was transferred to a new tube and the gel pieces were extracted with a) 25 μ l of buffer B (50 mM ammonium carbonate) and b) two times buffer C (Formic acid 0.1 %, acetonitrile 50 %). For each extraction, the gel pieces were incubated for 15 min at room temperature while shaking. The supernatants of each extraction were pooled

with original trypsin digest supernatants and dried for 2 h in a Speed-Vacuum concentrator. The peptides were then re-suspended in 25 μ l of precolumn loading buffer (0.08 % TFA and 2 % ACN in water). LC-MS/MS analysis was performed on an Ultimate 3000 LC system (Dionex, Voisins le Bretonneux, France) connected to linear ion trap mass spectrometer (LTQ, Thermo Fisher, USA) by nanoelectrospray interface for separation, ionisation and fragmentation of all peptides. Four μ l of tryptic peptide mixtures were loaded at flow rate 20 μ l/min onto precolumn Pepmap C18 (0.3 x 5 mm, 100 Å, 3 μ m; Dionex). After 4 min, the precolumn was connected with the separating nanocolumn Pepmap C18 (0.075 x 150 mm, 100 Å, 3 μ m, Dionex) and the gradient was started at 300 nL/min. All peptides were separated onto the nanocolumn using a linear gradient from 2 to 36 % of buffer B, over 18 min (eluting buffer A: 0.1 % Formic acid, 2 % acetonitrile and eluting buffer B: 0.1 % Formic acid, 80 % acetonitrile). Including the regeneration step, one run was 50 min length. Ionization was performed on liquid junction with a spray voltage of 1.3 KV applied to non-coated capillary probe (PicoTip EMITER 10 μ m ID; New Objective, USA). Peptides ions were analysed by the Nth-dependent method as follows: (i) full MS scan (m/z 300-2000), (ii) ZoomScan (scan of the 3 major ions), (iii) MS/MS on these 3 ions with classical peptides fragmentation parameter: $Q_z = 0.25$, activation time = 30 ms, collision energy = 40 %. Proteins identifications were performed with Bioworks 3.3 TM software. The raw data were converted and filtered in peak lists with default data generation parameters for LTQ mass spectrometer. All peak lists of precursor and fragment ions were matched automatically against a *Lactococcus lactis* IL 1403 protein database. The Bioworks search parameter included: trypsin specificity with one missed cleavage, variable oxydation of methionine and the mass tolerance was fixed to 1.4 Da for precursor ion and 0.5 Da for fragment ions. The search results were filtered using Bioworks 3.3. A multiple threshold filter applied at the peptide level consisted of the following criteria: Xcorr magnitude up to 1.7, 2.5 and 3.0 for

respectively mono-, di- and tri-charged peptides; peptide probabilities lower than 0.01; ΔC_n greater than 0.1 and only the first match result for each identified peptide.

Statistical treatment: Raw spot volumes were normalized by the mean intensity of the corresponding gel. Only the spots containing a single protein were considered. When a protein was present in several spots of a same gel, the protein abundance corresponded to the sum of the intensity of the different spots. Ratios of protein abundance were calculated using the slowest growth phase as a reference. Proteins identified both on 4.5-5.5 and 5-6 pH ranges were considered and analysed independently as two unique entities. The statistical significance of expression ratios was evaluated using Student test. Proteins with an abundance ratio associated to a p-value lower than 0.05 were considered differentially abundant. The mean False Discovery Rate (FDR) of all the selected proteins was lower than 20%.

RESULTS AND DISCUSSION

Isoleucine starvation and kinetic adaptation of *L. lactis*

L. lactis dynamic adaptation to amino acid starvation was studied in batch culture. This stress was imposed progressively and naturally by the consumption of BCAA during cell growth. Biomass, glucose, end products and amino acids concentrations were monitored during 8 h (Figure 1). Isoleucine was exhausted at 3.66 h of culture, when glucose and biomass concentrations were respectively 41 mM and 0.56 g.L⁻¹. Leucine and valine reached a concentration threshold of 100 μ M at 4.2 h and 4.3 h respectively. Below this level, these two amino acids could not be accurately quantified however they were not exhausted at the end of the culture. All other amino acids were still present at the end of the culture with concentrations ranging from 160 to 2500 μ M (data not shown). Glutamate was even produced during the fermentation (final concentration of ~130 μ M). From these results, it was established that only isoleucine starvation occurred during the culture.

Growth of *L. lactis* was exponential until 3.3 h with a rate of 0.88 h^{-1} , characteristic of a non-limited growth state. From 3.3 h, as isoleucine concentration became limiting, growth progressively decelerated to reach a constant rate of 0.05 h^{-1} at 6 h. This extremely low growth rate (generation time of 13.9 h) was maintained when the cells were transferred in fresh CDM medium lacking isoleucine (two successive transfers corresponding to at least 10 generations more). Constant glucose consumption (q_S) and lactate production (v_L) rates were observed during both the exponential ($22 \pm 3 \text{ mmol.g}^{-1}.\text{h}^{-1}$ and $40 \pm 2 \text{ mmol.g}^{-1}.\text{h}^{-1}$ respectively) and the slow growth ($8 \pm 1 \text{ mmol.g}^{-1}.\text{h}^{-1}$ and $14 \pm 1 \text{ mmol.g}^{-1}.\text{h}^{-1}$ respectively) phases. In both phases, lactate production accounted for almost 90 % of the glucose consumption, indicating a constant homolactic metabolism. Although *L. lactis* IL1403 was previously reported as an isoleucine auxotroph^{4,6}, it was shown in this study that this strain could sustain a constant growth in the absence of isoleucine, while at an extremely low rate. Isoleucine biosynthesis pathway therefore seems to be barely functional, but enough to maintain a highly limited anabolic activity. Godon *et al.*⁶ considered *L. lactis* as auxotroph for an amino acid when no colony was formed after 24 h incubation on agar plate lacking this specific amino acid. They have nevertheless observed small colony formation after 3 days without isoleucine which could correlate with our findings.

Overview of transcriptomic and proteomic responses

To characterize the dynamic adaptation of *L. lactis* to isoleucine starvation, both transcriptome and proteome analyses were performed during exponential phase (2.7 h of culture) and after 20 min, 1.34 h and 3 h of isoleucine starvation (respectively 4 h, 5 h and 6.66 h of culture). Raw transcriptomic data are available online in GEO database (series accession number: GSE12962) and the list of all identified proteins with their mass spectrometry characteristics are provided as supplementary data. For technical reasons,

transcriptome has delivered much more information than proteome (detection of 1948 genes compared to 341 proteins), according to the well documented literature ²⁷⁻²⁹. All the transcripts and proteins considered as differentially abundant during growth kinetics, with the statistical criteria described in materials and methods, are listed in table 2.

The response of *L. lactis* to isoleucine starvation is progressive. Indeed, when compared to the exponential phase (reference), 309, 420 and 587 genes were differentially expressed after 20 min, 1.34 h and 3 h of isoleucine starvation respectively. Similarly the levels of 30, 38 and 41 proteins were significantly modified while starvation occurred. Clearly, the longer the time after isoleucine starvation, the more genes and proteins were involved in the response. This shows that the dynamic approach is useful to provide the most complete description of the response.

When protein and transcript ratios were compared individually, only weak statistical correlations were obtained (Spearman coefficient of 0.26, p-value = 10^{-11}). The general reliability of macroarray data was however confirmed experimentally by the strong correlation with qRT-PCR results (Figure 2; Pearson coefficient of 0.98, p-value = 10^{-6}). Weak correlations between transcript and protein levels were similarly obtained for other bacteria, *Desulfovibrio vulgaris* ³⁰ and *Escherichia coli* ³¹ and this can be the consequence of post-transcriptional regulations ³². The functional analysis of the responses was however consistent with both methodologies. Global tendency for each functional category was estimated using the Wilcoxon test (Table 2) for transcriptomic results. For proteomic results, this was not possible due to the small size of the dataset but individual protein changes generally confirmed transcriptomic analysis tendencies (Table 2). For many metabolic pathways up-regulated at the transcriptomic level, *i.e.* in the categories “amino acid biosynthesis”, “thioredoxin glutathione and glutaredoxin”, “detoxification” or “aerobic metabolism”, punctual protein increases in the same metabolism could have been found. Similarly global down-regulations at transcriptomic level *i.e.* for “glycolysis”, “ribosomal proteins” and “translation factors” or “PTS systems”, were generally associated to some

protein level decreases. Thus combining these two independent approaches allowed biological results to be strongly consolidated. Less frequently, transcriptomic and proteomic techniques did not give the same type of results. Some pathways down-regulated at transcriptomic level were indeed associated to no significant variation at proteomic level or even weak up-regulations (*i.e.* “menaquinone ubiquinone” and “transport and binding proteins” or “fatty acid metabolism” and “pyrimidines” respectively). Part of these observed inconsistencies could be related to inability, with the 2D gel based approach, to detect all proteins and notably carrier proteins localized in the membrane. However the lag between transcripts and proteins to reach a new functioning state in a dynamic growth deceleration may also participate to this phenomenon. Indeed transcript abundances probably change more rapidly than protein abundances since protein degradation is expected to be slower than transcript degradation principally due to its energetic dependency (involvement of ATP-dependant enzymes^{33, 34}). These opposite results thus highlighted the complementarity of transcriptomic and proteomic approaches and the complexity of protein regulation.

Three major mechanisms are involved in isoleucine starvation adaptation

The transcriptomic and proteomic responses of *L. lactis* to isoleucine starvation (Table 2) can be summarized as three major mechanisms.

i) Negative control of major physiological activities. A general negative response was observed on genes and proteins involved into major physiological processes. The functional category "translation" was mostly, and very significantly (Wilcoxon p-value $<10^{-10}$), down-regulated at transcriptional level in response to isoleucine starvation. Globally the expression of 27 genes encoding ribosomal proteins as well as 6 genes encoding translation factors and 8 encoding amino-acyl tRNA synthetases decreased (Table 2). Among the 52 translation related proteins that were detected in proteomic analysis, 11 showed significant level variations and most of them were decreased (the 3 ribosomal proteins RplJ, RplL and RpsA as well as the

transcription factors Frr and Tsf, the amino-acyl tRNA synthetase ArgS, and the protein KsgA also implicated in protein synthesis). In the category "energy metabolism", the expression of genes related to carbon metabolism, more specifically those belonging to "glycolysis" (*enoA*, *ldhX* and *tpiA*), "fermentation" (*ackA1*, *adhE* and *frdC*) and "sugars" (*bglH*, *gntK*, *lacC*, *scrK*, *uxaC*, *uxuB*, *ypcA* and *ypdB*) subcategories, decreased. Among the measured proteins related to these subclasses, EnoA, Pyk and YpdD were less abundant too. At the same time, one can notice that CcpA, the carbon catabolism repressor³⁵, was down-regulated both at transcript and protein levels. Carbon transport was also affected by isoleucine depletion, since genes linked to "carbohydrate, organic alcohol and acids transport" (*rbsC*, *yngE*, *yngF*, *ypcG*, *ypcH* and *ypdA*) and "PTS transport" (*celB*, *ptcA*, *ptcB*, *ptnAB*, *ptnC*, *ptnD*, *ptsH* and *ptsI*) were under-expressed. Cytoplasmic fractions of the PTS system were detected and the level of PtnAB and PtsI proteins was decreased.

For the functional category "transcription", no expression variation tendency was observed (only few genes were found to be under-expressed, see table 2). This result is corroborated by proteomic analyses, since no variation of protein level was detected in this particular category. However, a strong decrease of transcriptional activity during isoleucine starvation was observed. A three-fold reduction of [³H]-uridine incorporation rate from $1.02 \cdot 10^9$ cpm.min⁻¹.g⁻¹ in the exponential phase to $3.34 \cdot 10^8$ cpm.min⁻¹.g⁻¹ in the slow growth phase was indeed measured.

Lastly, most of the messengers categorized in "fatty acid metabolism" (*accB*, *accC*, *accD*, *acpA*, *fabF*, *fabG1*, *fabI*, *fabZ1*, *fabZ2*, *fadD*, *lplL*, *plsX* and *thiL*) and "pyrimidines biosynthesis" (*pydB*, *pyrC*, *pyrF*, *pyrG* and *pyrR* encoding the pyrimidines biosynthesis regulator³⁶) were significantly under-expressed in response to isoleucine starvation, though these expression decreases were not reflected at proteomic level (see before).

This general decrease of translation, transcription, pyrimidines and fatty acid metabolisms are consistent with the growth rate reduction observed during isoleucine starvation. Similarly the wide down-regulation of carbon transport and metabolism can be associated to a reduced catabolic rate since a strong decrease of glucose consumption rate was also measured during isoleucine starvation. The response involved in *L. lactis* adaptation to isoleucine starvation is thus significantly different from the one of *E. coli* for which over-expression of genes of central metabolism had been observed during isoleucine starvation¹¹. The negative control of the catabolism observed in *L. lactis* will allow metabolic activities to be adjusted to the anabolic demand and avoid energy resources wastage. Such catabolic and anabolic balance will probably allow essential metabolic activities such as cell maintenance and survival to be favored.

ii) Specific stress response related to isoleucine starvation. *An extended positive control aiming at supplying cell with isoleucine was observed. Isoleucine, or BCAA, biosynthesis was above all stimulated by the strong induction of the isoleucine biosynthesis pathway (Table 2). The expression of the complete leu-ilv operon encoding BCAA biosynthesis pathway was strongly increased during isoleucine starvation. With the exception of ilvA and leuA whose targets were not available on the array, mean expression ratios in comparison with exponential phase ranged from 1.9 to 5.2 for ilvB, ilvC, ilvD, ilvN, leuB, leuC and leuD. The expression of bcaT, encoding the amino-transferase catalyzing the last step of BCAA biosynthesis, was also increased. In addition, the level of two proteins from this pathway (IlvD, LeuC) was increased.*

The response of *L. lactis* to isoleucine starvation was not limited to the stimulation of the isoleucine biosynthetic pathway but was more largely characterized by a wide and complex metabolic reorganization of the global amino acid metabolism (Figure 3). Most of these metabolic adaptations are connected by the carbon metabolism and seem to be dedicated to

increase isoleucine production in order to struggle against isoleucine depletion. First, threonine synthesis was favored, probably in order to increase the ketobutyrate pool which is one of the isoleucine precursors. Direct threonine biosynthesis from aspartate was enhanced since genes involved in threonine biosynthesis (*thrA* and *thrC*) were over-expressed and the Hom protein level was also increased (Table 2). An alternative pathway of threonine production was also envisaged from serine/glycine synthesis and catabolism. Serine/glycine metabolism was postulated to increase since genes involved in serine biosynthesis from 3P-glycerate (*serB* and *serC*) and more largely from glycerol catabolism (*yihF* an hypothetical protein sharing protein domain homology with glycerate kinase, *ypjA*, and the transporter *glpF1*) were over-expressed and the level of GlyA protein increased (see table 2). Secondly, pyruvate synthesis was also increased, since two genes potentially involved in L-alanine catabolism, *dal* and *ytjE*, were over-expressed. These transcriptomic results were confirmed experimentally by metabolite quantification since alanine consumption rate increased from 50 $\mu\text{mol.g}^{-1}.\text{h}^{-1}$ to 170 $\mu\text{mol.g}^{-1}.\text{h}^{-1}$ after isoleucine exhaustion and traces of pyruvate were detected. Increase of isoleucine biosynthesis is thus expected since pyruvate enters the isoleucine metabolic pathway. These results are consistent with what has recently been found in *E. coli* during isoleucine starvation¹¹ in which alanine, serine and threonine metabolisms dedicated to isoleucine formation were observed to be up-regulated.

Beside this direct effect of improving the metabolism of isoleucine precursors, common to *E. coli* and *L. lactis*, other effects were shown to be specific to *L. lactis*. The gene *gltD*, involved in glutamate biosynthesis from glutamine, which is available in the culture medium, was over-expressed during isoleucine starvation (Table 2). As a direct consequence, glutamate biosynthesis was found to be enhanced: initially lacking from the culture medium, a concentration of about 130 μM was measured at the end of the culture (6.66 h after inoculation). Since glutamate is also required for BCAA synthesis, this metabolic adaptation

can also be considered as a way to increase isoleucine production. Consistent with this metabolic organization favoring glutamate supply, genes *argG*, *argH* and *glnA*, encoding glutamate consuming enzymes for the production of metabolites other than isoleucine, were under-expressed (Table 2). Since this response was evidenced in *L. lactis* but not in *E. coli*, it could be related to the lack of glutamate dehydrogenase in LAB and their inability to synthesize glutamate directly from α -ketoglutarate.

Other responses possibly aiming at increasing glutamate supply were observed: the over-expression of genes involved in histidine biosynthesis pathways (*hisA*, *hisC*, *hisD* and *hisI*), the over-expression of genes involved in tyrosine and phenylalanine biosynthesis (*aroA*, *aroB*, *aroC*, *aroD* and *aroE*) and of the AroB protein, and the over-expression of the *gltA* gene encoding citrate synthase and of other genes upstream in the carbon metabolism (*mae*, *mleS*, *pdhA*, *pdhB*, *pdhC* and *pdhD*) (Table 2). From a theoretical point of view, a metabolic link could be drawn between these metabolic pathways and glutamate biosynthesis (Figure 3). However, since the TCA cycle is considered to be non operative in LAB, it is thus probable that aromatic amino acids and citrate were not able to provide supplemental glutamate. These observed expression changes may rather correspond to ancestral expression regulatory mechanisms.

Consistent with this amino acid metabolism reorientation, a response at the level of peptides was also observed (Table 2). The peptidases encoding genes *htrA*, *pepDA*, *pepO* and *pepXP* were over-expressed. The protein abundance of the peptidase PepO also increased. In addition, genes encoding oligo-peptides transporters (*oppA*, *oppC*, *oppD*, *oppF*, *optA*, *optB*, *optC*, *optD*, *optF* and *optS*) were over-expressed. Lamarque *et al.* did not control the pH of their cultures and failed to detect the *Opp* transcripts in the same strain³⁷. This was probably due to acidification since *Opp* is expected to be repressed at low pH [40]. Up-regulation of peptides transport and metabolism in response to isoleucine starvation will favor the

assimilation of extra- or intra- cellular peptides and is thus expected to lead to an increase of isoleucine supply.

iii) Additional response connected to oxidative stress. Genes belonging to the functional category “adaptation to atypical conditions” (*clpE*, *clpX*, *cpo*, *dinF* and *tpx*) were globally over-expressed (Table 2) while the chaperones encoding genes *dnaK* and *groES* as well as their transcriptional regulator *hrcA*³⁸ were under-expressed. Similarly, at protein level, stress related proteins amounts were increased for ClpE and decreased for GrpE and CspE. These different data seem to indicate that no general stress response was induced during isoleucine starvation. In contrast, in *E. coli*, the general stress response was massively up-regulated¹¹. This major difference highlights *L. lactis* regulation network originality, which is probably linked to its lack of stress sigma like factor.

Isoleucine starvation however provoked the induction of genes and proteins usually related to oxidative stress response. Genes belonging to the sub-categories “thioredoxin, glutaredoxin and glutathione” (*gpo*, *trxB1* and *trxH*) and “detoxification” (*ahpF*) were globally over-expressed (Table 2). Correspondingly, the proteins TrxB1 and SodA from these sub-categories were increased. Unlike other genes belonging to the category “energy metabolism”, genes involved in “aerobic” subcategory (*cbr*, *yddB*, *ymgK*, *ypaI*, *ypgB*, *yphA*, *yphC*, *ypjA*, *ypjH*, *yrfB*, *yrjC* and *yugC*, but not *noxA* and *noxB*) were also massively over-expressed (Table 2) and the protein level of the oxidoreductase YpjH increased strongly. Similarly, the pyruvate dehydrogenase metabolism expected when oxygen is present in *L. lactis* culture³⁹ was over-expressed at transcriptome level (*pdhA*, *pdhB*, *pdhC* and *pdhD*). Finally, “electron transport” subcategory genes (*cydA*, *ndrH*, *ndrI* and *yviC*) were also mainly over-expressed. These oxygen response-like inductions are not related to oxygen presence since anaerobic conditions were maintained during the culture (nitrogen atmosphere), and could

rather belong to isoleucine starvation cross-response. Induction of the oxidative stress response has already been observed in *E. coli* at the onset of stationary phase independently of oxygen supply⁴⁰. This adaptative response is probably dedicated to cells protection from subsequent stress, like a preventive mechanism.

Regulations involved in the isoleucine starvation response

To further investigate regulations that could occur and coordinate this adaptation process, growth rate and stringent responses¹⁷, GlnR regulon¹⁹, CcpA regulon³⁵ and aeration stimulon⁴¹, available in the literature, were compared to our data. Genes differentially up- or down- regulated at least one time during the isoleucine starvation were searched in the different regulons (see table 3). Genes expected to be controlled during isoleucine starvation by a particular regulator are expected to be common between isoleucine starvation response and its regulon. It should however be noticed that GlnR and CcpA regulons were defined with strain MG1363 of *L. lactis* whereas IL1403 was used in the other studies. Nevertheless, these two strains are phylogenetically closely related and share about 85 % sequence similarity⁴².

Regulations of the three major responses involved in the isoleucine starvation have been elucidated with this approach. First, major physiological activities were negatively regulated both by growth rate related and stringent regulations. Under-expression of genes involved in “transcription”, “translation” and “fatty acid metabolism” were, for a majority, related to growth rate regulation, while “pyrimidines” and “carbon metabolism and transport” (“glycolysis”, “fermentation”, “sugars”, “carbohydrate, organic alcohol and acids transport” and “PTS transport” subcategories) are rather linked to the stringent response (Table 3). Secondly, the response specifically dedicated to cope with the isoleucine depletion seemed to be under stringent and CodY controls. Metabolic pathways dedicated to glutamate production are not significantly controlled by GlnR (Table 3) although GlnR was found to control glutamate and glutamine metabolism in *L. lactis*¹⁹. They are however under CodY control

like all other amino acid biosynthesis pathways (except for aromatic amino acid) and amino acid and peptide carriers. On the other hand, genes involved in aromatic amino acid biosynthesis as well as those encoding peptidases are rather controlled by stringent mechanism (Table 3). Lastly, genes involved in the oxygen cross protection and belonging the sub-categories “thioredoxin, glutaredoxin and glutathione”, “detoxification”, “aerobic energy metabolism” and “electron transport pathways” are generally controlled both by stringent response and growth rate related regulators. Genes classified in “pyruvate dehydrogenase” subcategory are however not regulated by any one of these mechanisms but are rather controlled by CcpA. This result agrees with the CcpA involvement in *B. subtilis* response during amino acid starvation ⁹. However, though CcpA is down-regulated, the catabolic repression is only partially alleviated since the overlap between the isoleucine starvation response and CcpA regulon is weak (4 %). Finally from this analysis one can notice that the overlap between isoleucine starvation response and the one observed in aerated cultures of *L. lactis* ⁴¹ is weak (6 %) which confirmed again the absence of oxygen.

CONCLUSION

The combination of proteomic and transcriptomic dynamic analyses associated with fermentation kinetic study provided a sound description of *L. lactis* response to isoleucine starvation and has extended the physiological understanding of the metabolism of this bacterium. A regulatory network more complex than previously thought linking nitrogen and carbon metabolisms and involving various regulations was revealed. The dynamic response was found to occur gradually and could be divided into three main mechanisms. Firstly, a reduction of major physiological activities presumably regulated both by growth rate and stringent associated regulators was observed. Secondly, a response specifically dedicated to cope with the imposed starvation was defined. This process seemed to be under stringent and

CodY controls. Lastly, a response related to oxidative stress controlled by the growth rate related regulator, stringent response and CcpA, was described. Finally, growth rate related and stringent responses were by far the most involved global regulatory mechanisms in isoleucine starvation response. However, all the regulons described to date (Table 3) accounted only for 39 % of isoleucine starvation response, indicating that the molecular mechanism of this response still needs to be explored. Moreover, such integrated and comparative approach will allow, by its implementation to other regulations and environmental conditions, the whole regulatory network of *L. lactis* or any other microorganism to be deciphered.

ACKNOWLEDGMENTS

This work was supported by the INRA Integrative Biology Program (agroBI).

REFERENCES

1. Ardo, Y., Flavour formation by amino acid catabolism. *Biotechnol Adv* **2006**, 24, (2), 238-42.
2. Fernandez, M.; Zuniga, M., Amino acid catabolic pathways of lactic acid bacteria. *Crit Rev Microbiol* **2006**, 32, (3), 155-83.
3. Ganesan, B.; Stuart, M. R.; Weimer, B. C., Carbohydrate starvation causes a metabolically active but nonculturable state in *Lactococcus lactis*. *Appl Environ Microbiol* **2007**, 73, (8), 2498-512.
4. Cocaign-Bousquet, M.; Guarrigues, C.; Novak, L.; Lindley, N.; Loubiere, P., Rational development of a simple synthetic medium for the sustained growth of *Lactococcus lactis*. *J Appl Bacteriol* **1995**, 79, 108-116.
5. Delorme, C.; Godon, J. J.; Ehrlich, S. D.; Renault, P., Gene inactivation in *Lactococcus lactis*: histidine biosynthesis. *J Bacteriol* **1993**, 175, (14), 4391-9.
6. Godon, J. J.; Delorme, C.; Bardowski, J.; Chopin, M. C.; Ehrlich, S. D.; Renault, P., Gene inactivation in *Lactococcus lactis*: branched-chain amino acid biosynthesis. *J Bacteriol* **1993**, 175, (14), 4383-90.
7. Savijoki, K.; Ingmer, H.; Varmanen, P., Proteolytic systems of lactic acid bacteria. *Appl Microbiol Biotechnol* **2006**, 71, (4), 394-406.
8. Doeven, M. K.; Kok, J.; Poolman, B., Specificity and selectivity determinants of peptide transport in *Lactococcus lactis* and other microorganisms. *Mol Microbiol* **2005**, 57, (3), 640-9.
9. Tam le, T.; Eymann, C.; Antelmann, H.; Albrecht, D.; Hecker, M., Global gene expression profiling of *Bacillus subtilis* in response to ammonium and tryptophan starvation as revealed by transcriptome and proteome analysis. *J Mol Microbiol Biotechnol* **2007**, 12, (1-2), 121-30.
10. Nakamura, M. M.; Liew, S. Y.; Cummings, C. A.; Brinig, M. M.; Dieterich, C.; Relman, D. A., Growth phase- and nutrient limitation-associated transcript abundance regulation in *Bordetella pertussis*. *Infect Immun* **2006**, 74, (10), 5537-48.
11. Traxler, M. F.; Summers, S. M.; Nguyen, H. T.; Zacharia, V. M.; Hightower, G. A.; Smith, J. T.; Conway, T., The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol Microbiol* **2008**, 68, (5), 1128-48.
12. Braeken, K.; Moris, M.; Daniels, R.; Vanderleyden, J.; Michiels, J., New horizons for (p)ppGpp in bacterial and plant physiology. *Trends Microbiol* **2006**, 14, (1), 45-54.
13. Jain, V.; Kumar, M.; Chatterji, D., ppGpp: stringent response and survival. *J Microbiol* **2006**, 44, (1), 1-10.

14. Magnusson, L. U.; Farewell, A.; Nystrom, T., ppGpp: a global regulator in Escherichia coli. *Trends Microbiol* **2005**, 13, (5), 236-42.
15. Srivatsan, A.; Wang, J. D., Control of bacterial transcription, translation and replication by (p)ppGpp. *Curr Opin Microbiol* **2008**, 11, (2), 100-5.
16. Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malmme, K.; Weissenbach, J.; Ehrlich, S. D.; Sorokin, A., The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403. *Genome Res* **2001**, 11, (5), 731-53.
17. Dressaire, C.; Redon, E.; Milhem, H.; Besse, P.; Loubiere, P.; Coccagn-Bousquet, M., Growth rate regulated genes and their wide involvement in the Lactococcus lactis stress responses. *BMC Genomics* **2008**, 9, (1), 343.
18. Guedon, E.; Serror, P.; Ehrlich, S. D.; Renault, P.; Delorme, C., Pleiotropic transcriptional repressor CodY senses the intracellular pool of branched-chain amino acids in Lactococcus lactis. *Mol Microbiol* **2001**, 40, (5), 1227-39.
19. Larsen, R.; Kloosterman, T. G.; Kok, J.; Kuipers, O. P., GlnR-mediated regulation of nitrogen metabolism in Lactococcus lactis. *J Bacteriol* **2006**, 188, (13), 4978-82.
20. Garault, P.; Letort, C.; Juillard, V.; Monnet, V., Branched-chain amino acid biosynthesis is essential for optimal growth of Streptococcus thermophilus in milk. *Appl Environ Microbiol* **2000**, 66, (12), 5128-33.
21. Shivers, R. P.; Sonenshein, A. L., Bacillus subtilis ilvB operon: an intersection of global regulons. *Mol Microbiol* **2005**, 56, (6), 1549-59.
22. Poolman, B.; Konings, W. N., Relation of growth of Streptococcus lactis and Streptococcus cremoris to amino acid transport. *J Bacteriol* **1988**, 170, (2), 700-7.
23. Garrigues, C.; Goupil-Feuillerat, N.; Coccagn-Bousquet, M.; Renault, P.; Lindley, N. D.; Loubiere, P., Glucose metabolism and regulation of glycolysis in Lactococcus lactis strains with decreased lactate dehydrogenase activity. *Metab Eng* **2001**, 3, (3), 211-7.
24. Lin, C.; Kovalsky, O.; Grossman, L., Transcription coupled nucleotide excision repair by isolated Escherichia coli membrane-associated nucleoids. *Nucleic Acids Res* **1998**, 26, (6), 1466-72.
25. Redon, E.; Loubiere, P.; Coccagn-Bousquet, M., Role of mRNA stability during genome-wide adaptation of Lactococcus lactis to carbon starvation. *J Biol Chem* **2005**, 280, (43), 36380-5.
26. Livak, K. J.; Schmittgen, T. D., Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **2001**, 25, (4), 402-8.

27. Fenselau, C., A review of quantitative methods for proteomic studies. *J Chromatogr B Analyt Technol Biomed Life Sci* **2007**, 855, (1), 14-20.
28. Garbis, S.; Lubec, G.; Fountoulakis, M., Limitations of current proteomics technologies. *J Chromatogr A* **2005**, 1077, (1), 1-18.
29. Hegde, P. S.; White, I. R.; Debouck, C., Interplay of transcriptomics and proteomics. *Curr Opin Biotechnol* **2003**, 14, (6), 647-51.
30. Nie, L.; Wu, G.; Zhang, W., Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: a multiple regression to identify sources of variations. *Biochem Biophys Res Commun* **2006**, 339, (2), 603-10.
31. Lu, P.; Vogel, C.; Wang, R.; Yao, X.; Marcotte, E. M., Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **2007**, 25, (1), 117-24.
32. Mata, J.; Marguerat, S.; Bahler, J., Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci* **2005**, 30, (9), 506-14.
33. Ingmer, H.; Vogensen, F. K.; Hammer, K.; Kilstrup, M., Disruption and analysis of the *clpB*, *clpC*, and *clpE* genes in *Lactococcus lactis*: ClpE, a new Clp family in gram-positive bacteria. *J Bacteriol* **1999**, 181, (7), 2075-83.
34. Ogura, T.; Wilkinson, A. J., AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells* **2001**, 6, (7), 575-97.
35. Zomer, A. L.; Buist, G.; Larsen, R.; Kok, J.; Kuipers, O. P., Time-resolved determination of the CcpA regulon of *Lactococcus lactis* subsp. *cremoris* MG1363. *J Bacteriol* **2007**, 189, (4), 1366-81.
36. Martinussen, J.; Schallert, J.; Andersen, B.; Hammer, K., The pyrimidine operon *pyrRPB-carA* from *Lactococcus lactis*. *J Bacteriol* **2001**, 183, (9), 2785-94.
37. Lamarque, M.; Charbonnel, P.; Aubel, D.; Piard, J. C.; Atlan, D.; Juillard, V., A multifunction ABC transporter (Opt) contributes to diversity of peptide uptake specificity within the genus *Lactococcus*. *J Bacteriol* **2004**, 186, (19), 6492-500.
38. Guedon, E.; Jamet, E.; Renault, P., Gene regulation in *Lactococcus lactis*: the gap between predicted and characterized regulators. *Antonie Van Leeuwenhoek* **2002**, 82, (1-4), 93-112.
39. Cocaign-Bousquet, M.; Garrigues, C.; Loubiere, P.; Lindley, N. D., Physiology of pyruvate metabolism in *Lactococcus lactis*. *Antonie Van Leeuwenhoek* **1996**, 70, (2-4), 253-67.
40. Dukan, S.; Nystrom, T., Oxidative stress defense and deterioration of growth-arrested *Escherichia coli* cells. *J Biol Chem* **1999**, 274, (37), 26027-32.

41. Pedersen, M. B.; Garrigues, C.; Tophile, K.; Brun, C.; Vido, K.; Bennedsen, M.; Mollgaard, H.; Gaudu, P.; Gruss, A., Impact of aeration and heme-activated respiration on *Lactococcus lactis* gene expression: identification of a heme-responsive operon. *J Bacteriol* **2008**, *190*, (14), 4903-11.
42. Kok, J.; Buist, G.; Zomer, A. L.; van Hijum, S. A.; Kuipers, O. P., Comparative and functional genomics of lactococci. *FEMS Microbiol Rev* **2005**, *29*, (3), 411-33.

FIGURES LEGENDS

Figure 1: Kinetic profile of *L. lactis* IL1403 during isoleucine starvation.

Concentrations of biomass (◆), glucose (■), lactate (▲) and isoleucine (black line).

Figure 2: Macroarray and qRT-PCR comparison in *L. lactis* IL1403.

Log-log scatter plot of expression ratios between exponential phase and 1.34 h of isoleucine starvation obtained by macroarray (y-axis) and qRT-PCR (x-axis) analyses for 9 genes (*adhE*, *busAB*, *clpE*, *glpF1*, *ilvD*, *optC*, *ptnAB*, *rplM* and *rpoB*).

Figure 3: Coordinated expression of genes and proteins involved in amino acids metabolism during isoleucine starvation in *L. lactis* IL1403.

One arrow represents one metabolic reaction and dashed line arrows correspond to more than one reaction. Protein names, if they are known, are indicated beside the arrows. Brackets stand for genes that are absent from the macro-array. Thick black arrows indicate up-regulation or, if slashed, down-regulation at the transcriptional level. White arrows stand for over-expression of the corresponding protein. Framed amino acids are those that can be synthesised by *L. lactis* IL1403 while the other ones correspond to natural auxotrophies of the strain.

TABLES

Table 1: Primer sequences used for qRT-PCR.

| Target gene | Forward primer | Reverse primer | Amplicon length (bp) |
|--------------|-------------------------|-------------------------|----------------------|
| <i>adhE</i> | CTGACCCAAC TTTGAGCGAAGC | AGAGCAGAACCACCACCAAGAC | 94 |
| <i>busAB</i> | TTCCAGCGGTTGCATTCTTTGG | TTCTGACTGTTGGTGGGAGTGC | 84 |
| <i>clpE</i> | GCAGCAAGCACTCAA ACTCCAC | ACGACGAGCTGACTCTGTGATG | 132 |
| <i>glpF1</i> | AATGGGGTTGTCGCAGCAAATG | AGAAGCCA ACTACCCAGACAGC | 118 |
| <i>ilvD</i> | CAACACAACCAGCGACTCAAGC | ATGCTGACGATTCCGACCTGAG | 88 |
| <i>ldh</i> | ATGGTGTCGCTGTAGCTCTTG | GCAGTCGTTACGCCATATTG | 110 |
| <i>optC</i> | TTTGCATGGACCGGGATGGC | TGAGGCCAAGACGAAGTCACG | 78 |
| <i>ptnAB</i> | CTGCTGTTCAAGGTGCGATTCC | ACGATGATGCGGTTTGCTCTTG | 145 |
| <i>rplM</i> | TGGTACGTCGTCGATGCAACTG | TTTTCCACGAAGTACGCTTGCG | 75 |
| <i>rpoB</i> | TCCTGATGCAGAATGCCGTGTC | TCAGCAAGACCAAGCCAGTAGC | 75 |

Table 2: Transcriptomic and proteomic responses of *L. lactis* to isoleucine starvation

List of transcripts and proteins significantly up or down regulated 20 min, 1.34 h and 3 h after isoleucine exhaustion compared to the exponential phase and classified as defined by Bolotin¹⁶. Ratio values are indicated as exponent and data are ranged by decreasing ratios. Signs + or – followed by a number in brackets represent respectively significant up or down-regulation and the associated p-value of the functional (sub) category according to Wilcoxon test.

| FUNCTIONAL CATEGORIES or sub-categories | TRANSCRIPTOME | | | PROTEOME | | |
|---|---|---|--|---|----------------------|----------------------|
| | 20 min | 1.34 h | 3.00 h | 20 min | 1.34 h | 3.00 h |
| AMINO ACID BIOSYNTHESIS | + (1.2x10⁻⁴) | + (3.3x10⁻⁵) | + (2.1x10⁻⁴) | | | |
| Aromatic amino acid family | + (2.8x10⁻²) aroC ^{1.88} , aroA ^{1.57} , aroE ^{1.49*} , aroH ^{0.7} | + (1.9x10⁻²) aroC ^{1.7} , aroA ^{1.39} , aroD ^{1.37} , aroH ^{0.43} | + (7.8x10⁻³) trpA ^{1.77} , trpB ^{1.76} , aroC ^{1.62} , aroA ^{1.38} , aroB ^{1.37} , aroD ^{1.26} , aroH ^{0.38} | AroB ^{1.7} | AroB ^{2.81} | |
| Aspartate family | + (1.7x10⁻²) dapA ^{1.51} | + (1.8x10⁻²) thrA ^{2.08} , dapA ^{1.7} | + (2.9x10⁻²) thrA ^{2.08} , dapA ^{1.85} , thrC ^{1.29} | Hom ^{1.4} | LysA ^{2.07} | Hom ^{1.98} |
| Branched chain family | + (9.7x10⁻⁵) leuD ^{6.8} , leuC ^{6.35} , ilvD ^{4.71} , ilvB ^{3.57} , leuB ^{3.24} , ilvN ^{3.23} , ilvC ^{2.2} | + (3.0x10⁻⁵) leuC ^{4.95} , leuD ^{4.93} , ilvD ^{3.85} , leuB ^{2.58} , ilvN ^{2.5} , ilvB ^{2.28} | + (5.4x10⁻⁴) leuD ^{3.8} , leuC ^{3.74} , ilvD ^{2.96} , leuB ^{2.04} | IlvD ^{2.56} , LeuC ^{2.5} | | IlvD ^{4.74} |
| Glutamate family | gltD ^{1.77*} , gltA ^{1.35} , glnA ^{0.5} | argG ^{0.74} , argH ^{0.6} , glnA ^{0.28} | argG ^{0.76} , argH ^{0.55} , glnA ^{0.29} | | | |
| Histidine family | + (4.2x10⁻²) hisI ^{1.54} , hisA ^{1.5} , hisC ^{1.31} | + (7.2x10⁻³) hisI ^{1.59} , hisC ^{1.5} , hisA ^{1.44} , hisD ^{1.4} | + (3.5x10⁻²) hisC ^{1.64} , hisD ^{1.56} , hisA ^{1.52} , hisI ^{1.48} | | | |
| Serine family | serB ^{1.9} , serC ^{1.63} | serB ^{1.71} , serC ^{1.47} | serB ^{1.73} , serC ^{1.43} | | GlyA ^{2.59} | GlyA ^{2.1} |
| BIOSYNTHESIS OF COFACTORS | | | | | | |

| FUNCTIONAL CATEGORIES or sub-categories | TRANSCRIPTOME | | | PROTEOME | | |
|---|--|--|---|----------|-----------------------|-----------------------|
| | 20 min | 1.34 h | 3.00 h | 20 min | 1.34 h | 3.00 h |
| Folic acid | pabB ^{1.87} | pabB ^{2.11} | pabB ^{1.92} , folP ^{0.77} , fhs ^{0.69} | | | |
| Heme and porphyrin | | | | | | |
| Menaquinone and ubiquinone | - (2.1x10 ⁻²) menX ^{0.57} , menD ^{0.5} , menF ^{0.36} | - (9.8x10 ⁻³) preA ^{0.58} , menX ^{0.55} , menD ^{0.52} , menF ^{0.29} | - (1.0x10 ⁻²) ispB ^{0.82} , preA ^{0.63} , menX ^{0.62} , menD ^{0.5} , menF ^{0.32} | | | |
| Pantothenate | | - (2.9x10 ⁻²) coaA ^{0.61} , panE ^{0.47} | dfpB ^{1.21} , coaA ^{0.59} , panE ^{0.55} ribA ^{2.27} , ribC ^{0.71} | | | |
| Riboflavin and cobalamin | ribA ^{1.32} , ribC ^{0.63} | ribC ^{0.66} | | | | |
| Thioredoxin, glutaredoxin and glutathione | + (1.0x10 ⁻²) trxB1 ^{1.84} | + (2.9x10 ⁻³) trxB1 ^{2.26} , trxH ^{1.58} | + (1.9x10 ⁻³) trxB1 ^{2.6} , trxH ^{1.65} , gpo ^{1.64} | | TrxB1 ^{2.02} | TrxB1 ^{2.44} |
| Thiamin | apbE ^{0.6} | apbE ^{0.51} | apbE ^{0.53} | | | |
| Pyridine nucleotides | | | | | | |
| CELL ENVELOPE | | | | | | |
| Membranes, lipoproteins and porins | - (1.9x10 ⁻²) dgkA ^{1.25} , plpA ^{0.67} , plpC ^{0.63} , bmpA ^{0.55} | - (1.9x10 ⁻²) plpC ^{0.68} , bmpA ^{0.52} | - (2.3x10 ⁻²) bmpA ^{0.49} | | GlmU ^{3.05} | |
| Murein sacculus and peptidoglycan | dal ^{1.41} , murC ^{0.68} , murB ^{0.53} | dal ^{1.58} , pbp2A ^{1.56} , murC ^{0.68} , murB ^{0.53} | dal ^{1.69} , pbp2A ^{1.68} , acmD ^{1.62} , murD ^{1.46} , acmB ^{1.34} , murI ^{1.2} , murC ^{0.59} , murB ^{0.52} | | | MurB ^{3.66} |
| Surface polysaccharides and antigens | dltC ^{2.34} , icaC ^{1.8} | mycA ^{2.28} , dltC ^{2.11} , icaC ^{1.48} , ycbF ^{0.79} , ycbB ^{0.67} | mycA ^{2.72} , dltC ^{1.98} , yjg ^{1.66} , icaC ^{1.6} , ysfC ^{1.42} , ycbF ^{0.75} , ycbD ^{0.74} , pspB ^{0.72} , ycbB ^{0.7} | | | |
| CELLULAR PROCESS | | | | | | |
| Cell division | rodA ^{0.57} | gidB ^{0.68} , rodA ^{0.48} | gidB ^{0.69} , rodA ^{0.48} | | | |

| FUNCTIONAL CATEGORIES or sub-categories | TRANSCRIPTOME | | | PROTEOME | | |
|---|--|---|---|----------------------|-----------------------|-----------------------|
| | 20 min | 1.34 h | 3.00 h | 20 min | 1.34 h | 3.00 h |
| Cell killing | | | | | | |
| Chaperones | - (4.0x10⁻²) groES ^{0.64} , dnaK ^{0.48} | - (2.0x10⁻²) groES ^{0.69} , dnaK ^{0.47} | dnaK ^{0.53} | | | |
| Detoxification | + (1.9x10⁻²) ahpF ^{1.78} | + (1.9x10⁻²) ahpF ² | + (2.9x10⁻²) ahpF ^{2.23} secE ^{1.95} | | | SodA ^{1.79} |
| Protein and peptide secretion | | | | | | |
| Transformation | comFC ^{1.48} , comFA ^{1.25} | comFC ^{1.63} , comFA ^{1.38} | comFC ^{2.36} , comFA ^{1.72} | | | |
| CENTRAL INTERMEDIARY METABOLISM | | | | | | |
| General | | | pcaC ^{1.28} | | | MetK ^{1.66} |
| Amino sugars | ypcD ^{0.25} | nagA ^{1.59} , ylfH ^{0.82} , ypdC ^{0.27} | nagA ^{1.619} , ylfH ^{0.84} , ypdC ^{0.26} | FemD ^{0.43} | | FemD ^{0.34} |
| Degradation of polysaccharides | amyL ^{2.15} , dexB ^{0.47} , lnbA ^{0.24} | amyL ^{2.31} , dexB ^{0.45} , lnbA ^{0.23} | amyL ^{2.63} , xynD ^{2.19} , dexB ^{0.49} , lnbA ^{0.33} | | | |
| Phosphorus compounds | | | | | | |
| Polyamine biosynthesis | | | | | | |
| Other | + (3.9x10⁻²) | | glgA ^{1.37} | | | |
| ENERGY METABOLISM | - (1.5x10⁻²) | | | | | |
| Aerobic | | + (2.2x10⁻³) | + (5.6x10⁻³) | | | |
| | ypjH ^{2.93} , yphA ^{2.39} , yphC ^{2.19} , yugC ^{1.87} , ymgK ^{1.68} , yrfB ^{1.54} , ypal ^{1.34} , ypjA ^{1.33} , noxB ^{0.49} | ypjH ^{2.74} , yphA ^{2.44} , yphC ^{1.96} , yrfB ^{1.85} , ymgK ^{1.81} , yugC ^{1.74} , ypgB ^{1.48} , ypal ^{1.44} , yddB ^{1.43} , ypjA ^{1.23} , noxC ^{1.22} , ycdG ^{0.69} , noxA ^{0.52} , noxB ^{0.41} | yphA ^{2.43} , ypjH ^{2.37} , yphC ^{2.32} , yrfB ^{2.15} , yugC ^{1.88} , ybgA ^{1.84} , ymgK ^{1.75} , cbr ^{1.54} , ypal ^{1.46} , yrc ^{1.39} , ypjA ^{1.31} , ycdG ^{0.66} , noxA ^{0.55} , noxB ^{0.39} | YpjH ^{8.81} | YpjH ^{15.55} | YpjH ^{17.81} |
| Amino acids and amines | yciA ^{2.55} , araT ^{2.27} , ytiE ^{2.11} , bcaT ^{1.42} , arcC2 ^{0.87} | yciA ^{2.65} , araT ^{1.97} , ytiE ^{1.85} , pfs ^{1.54} | yciA ^{2.89} , araT ^{2.15} , ytiE ^{2.14} , yijB ^{1.79} | | | |

| FUNCTIONAL CATEGORIES or sub-categories | TRANSCRIPTOME | | | PROTEOME | | |
|---|--|--|--|----------|--|--|
| | 20 min | 1.34 h | 3.00 h | 20 min | 1.34 h | 3.00 h |
| Anaerobic ATP-PMF conversion | | dhaM ^{1.43} | | | GpdA ^{2.92} | |
| Electron transport | + (5.4x10⁻³) ndrH ^{2.35} , ndrI ² , yviC ^{1.38} | + (3.7x10⁻²) ndrH ^{2.65} , ndrI ^{2.65} , fer ^{0.76} | + (1.9x10⁻²) ndrH ^{2.66} , ndrI ^{2.44} , yviC ^{1.49} , cydA ^{1.32} , yfjE ^{0.73} , fer ^{0.72} | | | |
| Entner-Doudoroff Fermentation | - (6.0x10⁻³) aldB ^{3.76} , ackA1 ^{0.74} , frdC ^{0.68} , adhE ^{0.15} | - (3.5x10⁻²) aldB ^{3.03} , mleS ^{1.39} , ackA1 ^{0.67} , frdC ^{0.59} , adhE ^{0.14} | aldB ^{2.78} , mleS ^{1.57} , mae ^{1.19} , ackA1 ^{0.69} , frdC ^{0.62} , adhE ^{0.15} | | | |
| Gluconeogenesis | fbp ^{0.58} | fbp ^{0.61} | fbp ^{0.65} | | | |
| Glycolysis | ldhX ^{0.59} , enoA ^{0.56} | - (1.6x10⁻²) ldhX ^{0.58} , enoA ^{0.5} | - (2.7x10⁻²) tpiA ^{0.76} , ldhX ^{0.55} , enoA ^{0.49} | | Pyk ^{0.86} , EnoA ^{0.74} | EnoA ^{0.66} |
| Pentose phosphate pathway | gnd ^{0.74} | ptk ^{1.45} , gnd ^{0.57} , tkt ^{0.51} | | | | |
| Pyruvate dehydrogenase | + (8.2x10⁻³) pdhD ^{1.74} | + (1.6x10⁻²) pdhD ^{1.66} , pdhC ^{1.58} | + (1.8x10⁻²) pdhD ^{1.67} , pdhA ^{1.48} , pdhC ^{1.48} , pdhB ^{1.28} | | | |
| Sugars | - (4.8x10⁻³) uxuB ^{0.7} , uxaC ^{0.49} , bglH ^{0.46} , scrK ^{0.38} , ypdB ^{0.35} , ypcA ^{0.24} | - (3.2x10⁻²) lacC ^{0.73} , bglH ^{0.48} , scrK ^{0.39} , ypdB ^{0.36} , ypcA ^{0.27} | - (4.7x10⁻²) xylX ^{2.01} , galK ^{1.69} , galM ^{1.28} , gntK ^{0.73} , pmi ^{0.69} , bglH ^{0.62} , scrK ^{0.44} , ypdB ^{0.34} , ypcA ^{0.21} , citC ^{0.78} | | YpdD ^{0.73} | |
| TCA cycle | citC ^{0.81} | | | | | |
| FATTY ACID AND PHOSPHOLIPID METABOLISM | - (4.6x10⁻⁶) fbal ^{0.76} , plsX ^{0.67} , accC ^{0.55} , accD ^{0.55} , fabF ^{0.52} , fabG1 ^{0.46} , accB ^{0.43} , fadD ^{0.26} | - (3.0x10⁻⁶) plsX ^{0.79} , lpil ^{0.63} , fabZ1 ^{0.49} , accD ^{0.4} , acpA ^{0.4} , fadD ^{0.4} , accC ^{0.38} , fabF ^{0.36} , fabZ2 ^{0.33} , accB ^{0.31} , fabG1 ^{0.29} | - (1.6x10⁻⁶) acpD ^{1.43} , fabI ^{0.78} , plsX ^{0.76} , lpil ^{0.67} , thiL ^{0.62} , fabZ1 ^{0.53} , fadD ^{0.41} , acpA ^{0.38} , fabF ^{0.38} , accC ^{0.37} , accD ^{0.37} , fabZ2 ^{0.32} , accB ^{0.31} , | | AccC ^{1.54} , FabF ^{1.35} | ThiL ^{2.99} , YdiD ^{2.26} |

| FUNCTIONAL CATEGORIES or sub-categories | TRANSCRIPTOME | | | PROTEOME | | | |
|--|--|---|---|--|---|---|---|
| | 20 min | 1.34 h | 3.00 h | 20 min | 1.34 h | 3.00 h | |
| PURINES, PYRIMIDINES, NUCLEOSIDES AND NUCLEOTIDES 2'deoxyribonucleotide metabolism Nucleotide and nucleoside interconversion Purine ribonucleotide biosynthesis Pyrimidine ribonucleotide biosynthesis Salvage of nucleosides and nucleotides Sugar-nucleotide biosynthesis and interconversions REGULATORY FUNCTIONS | | | fabG1 ^{0.27} | | | | |
| | | | - (4.6x10 ⁻²) | | | | |
| | | nrdE ^{2.82} | nrdF ^{3.51} , nrdE ^{3.48} | nrdF ^{3.29} , nrdE ^{3.23} | | | |
| | | nucA ^{2.27} , dukA ^{1.75} | nucA ^{2.34} , dukA ^{1.59} , cmk ^{0.87} | nucA ^{1.92} , dukA ^{1.81} , dukB ^{0.74} | | | PyrH ^{4.91} |
| | | guaB ^{1.87} | guaB ^{2.15} , purM ^{1.1} , purN ^{1.1} , guaC ^{0.45} | guaB ^{2.42} , guaA ^{1.4} , guaC ^{0.5} | PurM ^{21.38} , PurL ^{11.71} , PurH ^{6.05} , PurB ^{1.39} | PurM ^{27.98} , Pur ^{9.93} , PurH ^{5.67} , PurB ^{1.51} | PurM ^{27.01} , PurL ^{9.85} , PurH ^{5.32} |
| | | - (3.0x10 ⁻²) pyrG ^{0.54} | - (1.7x10 ⁻²) thyA ^{1.37} , pydB ^{0.68} , pyrC ^{0.66} , pyrG ^{0.56} , pyrF ^{0.42} | PyrB ^{2.54} , ThyA ^{2.47} , PyrC ^{1.86} | ThyA ³ , PyrB ^{2.63} , PyrC ^{1.82} | ThyA ^{2.5} , PyrB ^{2.23} , PyrC ^{1.71} | |
| | xpt ^{4.65} , add ^{2.35} , deoC ^{0.46} | xpt ^{3.93} , add ^{2.3} , udk ^{1.38} , adk ^{0.77} , apt ^{0.73} , deoC ^{0.56} | xpt ^{3.08} , add ^{2.24} , udk ^{1.46} , apt ^{0.71} , deoC ^{0.52} | | | DeoD ^{0.59} | |
| | - (7.9x10 ⁻³) | - (6.8x10 ⁻³) | - (6.1x10 ⁻³) rmlC ^{0.69} | | | RmlB ^{1.38} | |

| FUNCTIONAL CATEGORIES or sub-categories | TRANSCRIPTOME | | | PROTEOME | | |
|---|--|---|--|----------------------|---------------------|--|
| | 20 min | 1.34 h | 3.00 h | 20 min | 1.34 h | 3.00 h |
| General | aldR ^{3.65} , yfbM ^{2.15} , ywdE ^{1.77} , yohC ^{1.65} , copR ^{1.5} , yrfA ^{1.34} , ywjD ^{1.33} , argR ^{0.64} , pyrR ^{0.64} , glnR ^{0.49} | aldR ^{3.1} , ywdE ^{1.96} , yfbM ^{1.92} , yohC ^{1.78} , yhgC ^{1.32} , ywjD ^{1.32} , ysgA ^{1.3} , ydcG ^{1.26} , yjaD ^{1.26} , argR ^{0.6} , glnR ^{0.29} | ynaB ^{3.69} , aldR ^{2.71} , ywdE ^{2.17} , yohC ^{2.08} , yfbM ^{2.02} , nadR ^{1.79} , yrbI ^{1.79} , copR ^{1.59} , yrfA ^{1.54} , yhgC ^{1.5} , ysfD ^{1.46} , yjaD ^{1.44} , fur ^{1.42} , rmeC ^{1.38} , ywjD ^{1.36} , birA2 ^{1.3} , ydcG ^{1.28} , ysgA ^{1.26} , codZ ^{0.82} , ahrC ^{0.77} , yxcB ^{0.67} , pyrR ^{0.64} , argR ^{0.59} , glnR ^{0.27} | | PyrR ^{1.2} | PyrR ^{1.43} |
| Two-component systems | lIrc ^{0.75} | lIrc ^{0.68} | lIrf ^{1.45} , lIrc ^{0.69} , lIrh ^{0.47} | | | LlrA ^{2.58} , LlrC ^{1.44} |
| Lac-lfamily regulators | rliA ^{1.87} | rliA ^{1.65} , ccpA ^{0.73*} | rliA ^{1.9} | Ccpa ^{0.79} | | |
| LysR-family regulators | rIrc ^{1.69} | rIrc ^{1.88} , mIeR ^{1.22} | rIrc ^{2.03} | FhuR ^{2.09} | | |
| AraC-family regulators | adaA ^{2.44} | adaA ^{2.18} | adaA ^{2.2} | | | |
| GntR-family regulators | rgrA ^{0.69} | rgrA ^{0.67} | rgrA ^{0.72} | | | |
| DeoR-family regulators | | | lacR ^{0.27} | | | |
| MarR-family regulators | rmaG ^{0.45} | rmaG ^{0.35} | rmaE ^{1.74} , rmaF1 ^{1.42} , rmaJ ^{1.39} , zitR ^{0.52} , rmaG ^{0.35} | | | |
| BglG-family regulators | | | | | | |
| GTP-binding proteins | | | ylqL ^{1.15} | | | TypA ^{0.71} |
| REPLICATION | | | | | | |
| Degradation of DNA | | rexA ^{0.69} | nth ^{1.38} , exoA ^{1.27} , rexA ^{0.69} | | | |

| FUNCTIONAL CATEGORIES or sub-categories | TRANSCRIPTOME | | | PROTEOME | | |
|---|---|---|--|---|---|--|
| | 20 min | 1.34 h | 3.00 h | 20 min | 1.34 h | 3.00 h |
| DNA replication, modification, recombination and repair | mutM ^{1.79} , cshA ^{1.39} , recA ^{1.24} | ogt ^{2.11} , mutM ^{1.75} , pcrA ^{1.14} , parC ^{0.79} | umuC ^{3.8} , ogt ^{1.98} , mutM ^{1.72} , cshA ^{1.46} , radC ^{1.39} , parC ^{0.67} | DnaA ^{2.82} | | HslA ^{1.35} |
| TRANSCRIPTION | | | | | | |
| Degradation of RNA | vacB1 ^{1.93} | vacB1 ^{1.68} | vacB1 ^{1.81} , vacB2 ^{1.6} | | | |
| RNA synthesis, modification and DNA transcription | rpoD ^{1.72} , rpoB ^{0.54} | rpoD ^{1.7} , yfjD ^{1.61} , rsuA ^{0.8} , RpoA ^{0.75} , rpoB ^{0.47} | rpoD ^{1.81} , yfjD ^{1.77} , nusG ^{1.64} , rsuA ^{0.73} , rpoA ^{0.71} , rpoC ^{0.6} , rpoB ^{0.44} | | | |
| RNA processing | | rimM ^{0.78} | rheB ^{0.128} | | | |
| TRANSLATION | | | | | | |
| Amino acyl tRNA synthetases | - (6.3x10 ⁻¹³) - (1.7x10 ⁻³) ileS ^{2.06} , proS ^{1.2} , glyS ^{0.75} , argS ^{0.65} , pheS ^{0.55} , thrS ^{0.48} | - (5.5x10 ⁻¹⁵) - (5.5x10 ⁻⁴) ileS ^{1.9} , argS ^{0.76} , alaS ^{0.74} , valS ^{0.68} , glyS ^{0.63} , pheS ^{0.51} , thrS ^{0.5} | - (5.3x10 ⁻¹⁶) - (3.0x10 ⁻⁴) ileS ^{1.94} , alaS ^{0.8} , argS ^{0.79} , lysS ^{0.76} , glyS ^{0.71} , pheS ^{0.52} , thrS ^{0.49} , tyrS ^{0.45} | Gltx ^{2.32} , ArgS ^{0.84} | TyrS ^{1.29} , ArgS ^{0.87} | TyrS ^{1.27} |
| Degradation of proteins, peptides and glycopeptides | pepXP ^{1.93} , pepO ^{1.56} , htrA ^{1.43} | pepXP ^{1.86} , pepDA ^{1.47} , htrA ^{1.41} , yueF ^{0.63} | pepXP ^{2.09} , pepDA ^{1.52} , pepO ^{1.49} , htrA ^{1.35} , yueF ^{0.59} | | PepO ^{1.58} | PepO ^{1.78} |
| Protein modification | pmsR ^{1.89} , ytaD ^{1.4} - (1.2.x10 ⁻¹¹) | ytaD ^{1.66} , pmsR ^{1.54} - (1.3.x10 ⁻¹¹) | pmsR ^{1.84} , ytaD ^{1.82} - (1.8.x10 ⁻¹³) | | | |
| Ribosomal proteins: synthesis and modification | rpmGA ^{1.45} , rpsP ^{1.39} , gatB ^{1.34} , rplQ ^{0.74} , rpsK ^{0.74} , rpsG ^{0.63} , rpsS ^{0.61} , rpsU ^{0.61} , rplN ^{0.6} , rpsJ ^{0.6} , rpmB ^{0.59} , rpsQ ^{0.59} , rpsI ^{0.58} , rpsC ^{0.55} , rpsL ^{0.55} , rpmA ^{0.54} , rpml ^{0.52} | prmA ^{0.76} , rplQ ^{0.7} , rpsK ^{0.7} , rpsG ^{0.67} , rpsC ^{0.62} , rpsB ^{0.6} , rplS ^{0.57} , yhdC ^{0.56} , rpsI ^{0.55} , rpsJ ^{0.55} , rplN ^{0.54} , rpmB ^{0.54} , rpsQ ^{0.54} , rplB ^{0.53} , rpsS ^{0.53} , rpsF ^{0.51} , rpsU ^{0.5} , rplM ^{0.46} , rpsL ^{0.46} , rpml ^{0.39} | trmU ^{1.58} , yhjG ^{1.53} , fmt ^{1.29} , rplQ ^{0.83} , prmA ^{0.81} , rplO ^{0.71} , rpsK ^{0.69} , rpsG ^{0.66} , rpsB ^{0.63} , yhdC ^{0.62} , rpsR ^{0.6} , rplS ^{0.59} , rpsQ ^{0.57} , rplB ^{0.56} , rpsC ^{0.56} , rpsJ ^{0.56} , rpsI ^{0.55} , rpsS ^{0.55} , rplN ^{0.52} , rpsU ^{0.52} , rplU ^{0.51} , rpmB ^{0.5} , rpsF ^{0.49} , rpsL ^{0.47} | RplJ ^{0.81} , RpsA ^{0.75} , KsgA ^{0.5} | RplJ ^{0.76} , RpsA ^{0.71} | RpmC ^{1.29} , RpsA ^{0.66} , RplL ^{0.55} |

| FUNCTIONAL CATEGORIES or sub-categories | TRANSCRIPTOME | | | PROTEOME | | |
|---|---|--|---|---|--------|---------------------|
| | 20 min | 1.34 h | 3.00 h | 20 min | 1.34 h | 3.00 h |
| | | | rplT ^{0.46} , rplM ^{0.44} , rplV ^{0.43} , rplM10 ⁻⁴ , rpsH ^{0.39} | | | |
| Translation factors | - (3.0x10 ⁻³) epf ^{0.75} , infA ^{0.67} , tsf ^{0.66} , fusA ^{0.65} , infC ^{0.59} | - (8.1x10 ⁻⁴) infA ^{0.65} , fusA ^{0.64} , tsf ^{0.63} , infC ^{0.52} | - (6.9x10 ⁻⁴) infB ^{0.78} , fusA ^{0.65} , infA ^{0.6} , tsf ^{0.59} , infC ^{0.5} | Tsf ^{0.83} , Frr ^{0.76} | | Tsf ^{0.76} |
| TRANSPORT AND BINDING PROTEINS | | - (3.3x10 ⁻²) | - (1.9x10 ⁻²) | | | |
| General | ymeB ^{5.58} , yjcA ^{3.92} , yijC ^{1.95} , yfcB ^{1.93} , yijD ^{1.91} , ysaB ^{1.85} , yxdG ^{1.79} , ydbA ^{1.67} , ytaB ^{1.56} , ydaG ^{1.46} , yxfA ^{1.46} , ysdA ^{1.44} , ykhF ^{0.78} , ynaC ^{0.7} | ymeB ^{4.26} , yjcA3 ⁹² , yijC ^{2.9} , yijD ^{2.72} , yfcB ^{1.93} , ydbA ^{1.86} , ysaB ^{1.83} , ydaG ^{1.69} , yrfE ^{1.65} , ytaB ^{1.64} , yxfA ^{1.62} , ysfB ^{1.6} , ypgD ^{1.51} , ygfB ^{1.42} , ysiB ^{1.36} , ysdA ^{1.31} , ychE ^{0.83} , ycfB ^{0.75} , ynaD ^{0.65} , yjif ^{0.62} , ykhF ^{0.6} , ynaC ^{0.6} | yjcA ^{4.08} , ymeB ^{3.21} , yijC ^{3.19} , yijD ³ , yrfE ^{1.99} , ysfB ^{1.97} , yfcB ^{1.91} , ytaB ^{1.89} , ydbA ^{1.76} , ysaB ^{1.75} , yxfA ^{1.72} , ypgD ^{1.67} , ydaG ^{1.66} , ygfB ^{1.4} , ysiB ^{1.37} , ysdA ^{1.34} , yifE ^{1.31} , yhcA ^{1.29} , ynaD ^{0.66} , ykhF ^{0.61} , yjif ^{0.58} , ynaC ^{0.55} , ydcF ^{0.39} | | | |
| Amino acids, peptides and amines | + (3.0x10 ⁻³) optS ^{3.27} , busAA ^{3.24} , busAB ^{2.99} , oppA ^{2.44} , optC ^{2.21} , oppD ^{2.11} , optD ^{2.11} , yjgC ^{2.1} , optB ^{1.83} , optA ^{1.82} , oppF ^{1.73} , ydgB ^{1.42} , yfcG ^{1.29} , glnP ^{0.76} , ylcA ^{0.62} , lysP ^{0.47} | + (2.9x10 ⁻²) busAA ^{4.84} , busAB ^{5.2} , optS ^{2.48} , oppA ^{2.46} , oppD ^{2.42} , yjgC ^{2.25} , optD ^{2.19} , optC ^{2.13} , oppC ^{2.07} , yvdF ^{2.05} , optF ^{1.99} , oppF ^{1.92} , optB ^{1.74} , optA ^{1.72} , yfcG ^{1.45} , ydgB ^{1.4} , glnP ^{0.66} , ylcA ^{0.58} , lysP ^{0.49} | busAB ^{5.42} , busAA ^{5.1} , yjgC ^{2.88} , oppD ^{2.42} , yvdF ^{2.29} , oppA ^{2.26} , oppC ^{2.16} , optC ^{2.1} , optD ^{2.03} , optS ^{1.91} , oppF ^{1.75} , optB ^{1.71} , optA ^{1.64} , ydgB ^{1.44} , yrfD ^{1.44} , yfcG ^{1.36} , choS ^{1.22} , arcD1 ^{1.14} , yshA ^{0.77} , glnP ^{0.64} , lysP ^{0.61} , ylcA ^{0.56} , phnA ^{0.63} | | | |
| Anions | | | | | | |
| Carbohydrates, organic alcohols and acids | - (2.4x10 ⁻⁴) glpF1 ^{1.26} , ypdA ^{0.41} , yphC ^{0.35} , ypcG ^{0.26} | - (2.6x10 ⁻⁴) y pbD ^{1.49} , yqgE ^{1.27} , yngF ^{0.63} , ypcH ^{0.35} , ypdA ^{0.34} , ypcG ^{0.27} | - (2.5x10 ⁻³) y pbD ^{1.54} , yqgE ^{1.41} , glpF1 ^{1.3} , rbsC ^{0.75} , yngE ^{0.61} , yngF ^{0.55} , ypcH ^{0.34} , ypdA ^{0.33} , ypcG ^{0.23} | | | |
| Cations | kupA ^{2.14} , ypbB ^{1.67} , yuiA ^{1.6} , yqgG ^{1.59} , cadA ^{1.45} , fhuD ^{0.6} , ysdE ^{0.48} | mtsA ^{2.26} , mtsB ^{2.11} , kupA ^{2.1} , qgG ^{1.91} , ypbB ^{1.74} , yuiA ^{1.62} , amtB ^{0.64} , ysdE ^{0.41} | kupA ^{2.27} , mtsB ^{2.25} , mtsA ^{2.18} , yqgG ^{2.16} , ypbB ^{2.01} , yuiA ^{1.87} , fhuB ^{1.33} , mgtA ^{0.85} , amtB ^{0.65} , zitS ^{0.52} | | | |
| Nucleosides, | | pbuX ^{4.09} | | | | |

| | | TRANSCRIPTOME | | | PROTEOME | | |
|---|--|--|--|--|---|----------------------|----------------------|
| FUNCTIONAL CATEGORIES or sub-categories | | 20 min | 1.34 h | 3.00 h | 20 min | 1.34 h | 3.00 h |
| purines and pyrimidines | | pbuX ^{4.89} | | pbuX ^{3.01} , pnuC2 ^{1.33} | | | |
| PTS system | | - (2.4x10 ⁻⁵) ptsI ^{0.73} , ptsH ^{0.57} , ptnD ^{0.43} , ptcA ^{0.42} , ptnAB ^{0.39} , ptnC ^{0.34} | - (4.9x10 ⁻⁴) ptsI ^{0.82} , celB ^{0.6} , ptsH ^{0.59} , ptcA ^{0.4} , ptnD ^{0.36} , ptnAB ^{0.32} , ptnC ^{0.28} | - (5.0x10 ⁻⁵) ptsI ^{0.79} , ptsH ^{0.58} , ptcB ^{0.48} , ptnD ^{0.44} , ptcA ^{0.41} , ptnC ^{0.28} , ptnAB ^{0.27} | PtnAB ^{0.65} | | PtsI ^{0.69} |
| Multidrug resistance | | pmrA ^{1.83} , yjdE ^{1.46} , lcnD ^{1.25} | pmrA ^{1.82} , yjfF ^{1.55} , lcnD ^{1.26} , yqiA ^{0.56} , blt ^{0.56} | pmrA ^{2.02} , yjfF ^{1.62} , napC ^{1.58} , ybfD ^{1.5} , lcnD ^{1.41} , cydD ^{1.4} , lcnC ^{1.16} , yqiA ^{0.54} , blt ^{0.52} | | | |
| OTHER CATEGORIES | | - (6.2x10 ⁻⁶) cpo ^{1.35} , hrcA ^{0.62} | + (4.8x10 ⁻²) clpE ^{1.58} , cpo ^{1.47} , tpx ^{1.39} , clpX ^{1.18} , hrcA ^{0.6} | clpE ^{1.67} , tpx ^{1.48} , cpo ^{1.42} , dinF ^{1.4} , clpX ^{1.3} , hrcA ^{0.67} | ClpE ^{3.13} , CspE ^{0.59} , GrpE ^{0.5} | GrpE ^{0.67} | GrpE ^{0.49} |
| Adaptations and atypical conditions | | | | | | | |
| Drug and analog sensitivity | | pacB ^{1.59} - (5.7x10 ⁻⁸) | bar ^{1.67} , pacB ^{1.59} | pacB ^{1.89} , bar ^{1.7} | | | |
| Phage related functions and prophages | | ps303 ^{2.46} , pi125 ^{1.92} , pi145 ^{1.51} , pi106 ^{1.38} , pi227 ^{1.38} , pi336 ^{1.35} , ps201 ^{1.26} , pi235 ^{0.77} , pi329 ^{0.73} , pi140 ^{0.6} | pi125 ^{2.66} , pi108 ^{2.52} , ps303 ^{2.19} , pi106 ^{2.11} , pi360 ^{1.79} , pip ^{1.53} , pi145 ^{1.46} , pi227 ^{1.45} , pi301 ^{1.43} , pi334 ^{1.42} , pi226 ^{1.36} , pi320 ^{1.24} , pi147 ^{0.58} | pi108 ^{3.45} , pi106 ^{3.02} , pi125 ^{2.76} , ps303 ^{2.51} , pi111 ^{2.17} , pi105 ^{2.03} , pi360 ^{1.92} , pi124 ^{1.76} , pip ^{1.7} , pi144 ^{1.64} , pi102 ^{1.62} , ps302 ^{1.6} , pi118 ^{1.59} , pi101 ^{1.53} , pi222 ^{1.45} , pi336 ^{1.45} , ps201 ^{1.43} , ps301 ^{1.43} , pi145 ^{1.41} , pi334 ^{1.41} , ps212 ^{1.41} , pi226 ^{1.4} , pi323 ^{1.4} , pi227 ^{1.39} , pi301 ^{1.35} , pi337 ^{1.31} , pi320 ^{1.23} , pi329 ^{0.79} , ps221 ^{0.71} , ps316 ^{0.71} , pi147 ^{0.48} | | | |
| Transposon related functions | | yajE ^{1.54} , tra981C ^{1.48} | + (3.1x10 ⁻²) tra981C ^{1.65} , yajE ^{1.59} | tra981C ^{1.91} , yajE ^{1.57} , ybdK ^{1.38} , tra983L ^{1.35} | | | |

| FUNCTIONAL CATEGORIES or sub-categories | TRANSCRIPTOME | | | PROTEOME | | |
|---|---|---|---|---|--------|--------|
| | 20 min | 1.34 h | 3.00 h | 20 min | 1.34 h | 3.00 h |
| Other | <p>crtK^{1.42}</p> <p>- (3.7x10⁻⁴²)</p> <p>yrgH^{4.55}, yrgI^{4.17}, yrbB^{2.93}, yuiB^{2.57}, yxdF^{2.44}, yecD^{2.42}, ybeF^{2.37}, yhjA^{2.36}, ysaD^{2.35}, yjgF^{2.22}, yjhA^{2.18}, ykjB^{2.16}, ymhC^{2.12}, yjfJ^{2.08}, ypbC^{2.07}, yfiE^{2.05}, yqeD^{2.03}, yuhH^{1.99}, yjaE^{1.94}, yfhB^{1.9}, ysgB^{1.9}, yqbA^{1.87}, yneB^{1.85}, ywal^{1.85}, ywfH^{1.85}, yphJ^{1.81}, ybiH^{1.79}, yqfC^{1.78}, yohD^{1.77}, ytgA^{1.76}, ywaD^{1.75}, yrfJ^{1.74}, yhci^{1.73}, ytcE^{1.73}, yteB^{1.73}, ygiC^{1.72}, ytfA^{1.67}, yuhl^{1.67}, yqfE^{1.65}, ypaA^{1.62}, ywfA^{1.62}, ysbB^{1.61}, yjih^{1.6}, ychC^{1.59}, yigC^{1.53}, ytfB^{1.53}, ysbC^{1.52}, yuhJ^{1.51}, yndD^{1.47}, ysbA^{1.46}, yudF^{1.46}, ybdH^{1.45}, yphI^{1.45}, ypjB^{1.45}, yvdC^{1.45}, yveG^{1.44}, yvhB^{1.44}, ypjI^{1.43}, ykjE^{1.42}, yncB^{1.41}, yojB^{1.41}, yijB^{1.38}, yqeA^{1.37}, ykca^{1.36}, yqgC^{1.36}, ywiB^{1.33}, yjie^{1.31}, yqeH^{1.31}, yvjA^{1.25}, yraE^{0.77}, yshB^{0.77}, ylid^{0.76}, ynhI^{0.75}, yciC^{0.74}, yheB^{0.74},</p> <p>ykjA^{0.73}, ykhE^{0.72}, ygiK^{0.7}, ybiC^{0.68}, yjcD^{0.65}, yahA^{0.65}, yleB^{0.63}, yfdB^{0.62}, yeiG^{0.61}, yrgE^{0.61}, ylaE^{0.56}, yejC^{0.52}, yqhA^{0.48}, ylhB^{0.41}, ypdC^{0.4}, ytcC^{0.39}, ygaJ^{0.38}, ypcB^{0.34}, yiaC^{0.31}, ypcC^{0.26}</p> | <p>- (1.4x10⁻⁵¹)</p> <p>yrgI^{5.27}, yrgH^{5.17}, yrbB^{4.58}, yjaE^{3.58}, yecD^{2.99}, yhjA^{2.54}, ybeF^{2.45}, yuiB^{2.41}, yqfC^{2.25}, yqfE^{2.25}, yfiE^{2.24}, ymhC^{2.19}, ykjB^{2.13}, yjfJ^{2.09}, yneB^{2.05}, ywal^{2.05}, ypbC^{2.02}, yuhH^{2.01}, yfhB², ysgB², yfhA^{1.98}, ygbB^{1.96}, ytgH^{1.89}, ytgB^{1.88}, yviA^{1.87}, yqeD^{1.83}, yjgF^{1.82}, yghC^{1.81}, yrfJ^{1.8}, yxdF^{1.79}, yjhA^{1.78}, ysaD^{1.76}, ytgA^{1.76}, yijB^{1.75}, yohD^{1.74}, ybiH^{1.7}, ypiI^{1.68}, ywfH^{1.68}, yteB^{1.66}, ychC^{1.62}, yhci^{1.62}, yigC^{1.58}, yogI^{1.57}, yfhl^{1.56}, ygbF^{1.56}, ybeA^{1.55}, yjih^{1.55}, ymbJ^{1.55}, yghG^{1.52}, yseF^{1.51}, yojB^{1.5}, ypaC^{1.49}, ypiL^{1.48}, ygiC^{1.47}, yncB^{1.47}, ytaA^{1.47}, yscA^{1.46}, ytfB^{1.46}, yvdC^{1.46}, ykca^{1.43}, yqaG^{1.43}, yfhL^{1.4}, yuhJ^{1.37}, ywfA^{1.36}, yphI^{1.34}, yggA^{1.33}, yjaB^{1.33}, yvhB^{1.32}, yndD^{1.3}, ytbC^{1.24}, ymcA^{1.23}, yoiB^{0.85}, ylgC^{0.81}, yscD^{0.81}, yuiE^{0.81}, ywaB^{0.81}, yfbJ^{0.8}, yujF^{0.79},</p> <p>yxaB^{0.74}, ynhI^{0.73}, yniH^{0.73}, yvaB^{0.73}, yccE^{0.71}, yibF^{0.71}, ykjA^{0.71}, ysjC^{0.71}, ytgF^{0.71}, ycjA^{0.7}, yvaA^{0.7}, yogG^{0.69}, yujA^{0.69}, ybiC^{0.67}, ycci^{0.67}, yraA^{0.67}, ytgE^{0.66}, ygiJ^{0.64}, yfdB^{0.63}, ygiK^{0.63},</p> | <p>crtK^{1.35}</p> <p>- (1.9x10⁻⁴⁸)</p> <p>yrgH^{5.21}, yrgI^{5.19}, yrbB^{4.47}, yjaE^{4.28}, yecD^{3.48}, yejJ^{3.24}, ynaE^{2.82}, yuiB^{2.69}, yhjA^{2.66}, ynaA^{2.56}, ybeF^{2.46}, ymhC^{2.4}, yjfJ^{2.39}, yneB^{2.36}, ykjB^{2.32}, yviA^{2.32}, yfiE^{2.31}, yqfE^{2.27}, ypbC^{2.26}, yfhB^{2.25}, ywal^{2.23}, yuhH^{2.22}, yfhA^{2.12}, yqfC^{2.07}, ytgH^{2.02}, yacC^{1.99}, ybiH^{1.98}, yjgF^{1.96}, yrfJ^{1.94}, yfhl^{1.93}, yqbA^{1.93}, ysgB^{1.93}, ytfA^{1.91}, ywfH^{1.9}, yijB^{1.88}, yqeD^{1.88}, yohD^{1.87}, ychC^{1.86}, ytgB^{1.86}, yfhJ^{1.85}, ygbB^{1.85}, yldE^{1.85}, yejE^{1.82}, yhci^{1.82}, ytgA^{1.82}, yjih^{1.8}, yghC^{1.79}, ysaD^{1.78}, ytaA^{1.78}, yacG^{1.77}, yiaD^{1.77}, yafD^{1.76}, ybfc^{1.76}, yjhA^{1.75}, yseF^{1.72}, ybeA^{1.7}, ycgJ^{1.7}, yxcA^{1.7}, yogI^{1.69}, yacB^{1.68}, ypiI^{1.68}, yuhJ^{1.68}, yscA^{1.66}, yedE^{1.64}, ygbF^{1.64}, ymbJ^{1.64}, yteB^{1.64}, yuhl^{1.64}, yghG^{1.63}, yigC^{1.6}, yxdF^{1.6}, ythA^{1.59}, ythB^{1.58}, yphJ^{1.57}, ygiC^{1.56}, yhcB^{1.56}, ypaC^{1.56}, ytfB^{1.56}, yuhD^{1.56}, yrgJ^{1.53}, yhfB^{1.52}, ynfC^{1.52}, yojB^{1.52}, yteG^{1.52}, yhjF^{1.51}, ynjH^{1.51}, ybfE^{1.5}, ysjF^{1.49}, yvel^{1.49}, ynhC^{1.48}, ypiL^{1.48}, yqaG^{1.48}, ytcD^{1.48}, yvdC^{1.48}, yvhB^{1.46}, yggA^{1.43}, yncB^{1.43}, yndD^{1.43}, yseA^{1.43}, ywfF^{1.43}, yabD^{1.42},</p> | <p>YtgG^{4.04}, YtgH^{3.77}, YtaA^{3.17}, YweD^{2.45}, YthC^{1.46}, YgdA^{1.43}, YteA^{0.49}, YhfA^{0.42}, YciH^{0.20}, YtaA^{7.65}, YahB^{4.07}, YtgH^{3.87}, YtgG^{3.18}, YweD^{2.99}, YeiG^{2.14}, YtdA^{1.77}, YtdH^{1.48}, YthC^{1.28}, YfgH^{0.2}</p> | | |
| UNKNOWN | | | | <p>YtgH^{2.99}, YthC^{1.16}</p> | | |

| FUNCTIONAL CATEGORIES or sub-categories | TRANSCRIPTOME | | | PROTEOME | | |
|---|---------------|---|--|----------|--------|--------|
| | 20 min | 1.34 h | 3.00 h | 20 min | 1.34 h | 3.00 h |
| | | <p>yngA^{0.63}, ybaD^{0.61}, yciC^{0.61}, yeiG^{0.57}, yнил^{0.54}, ysbD^{0.54}, yshB^{0.53}, yheB^{0.52}, ylaE^{0.5}, ytcC^{0.5}, yejC^{0.45}, ygaJ^{0.44}, yjcD^{0.41}, yiaC^{0.37}, ypdC^{0.36}, ylhB^{0.32}, ypcC^{0.32}, ypcB^{0.29}</p> | <p>ykbF^{1.42}, yueC^{1.42}, yudD^{1.4}, ybdH^{1.39}, yjaB^{1.39}, yqcF^{1.39}, ymcA^{1.38}, yjbB^{1.37}, ykiF^{1.37}, yudB^{1.37}, yeaC^{1.36}, yifA^{1.36}, ytgG^{1.36}, yudF^{1.36}, yphI^{1.35}, ycgl^{1.34}, ytcC^{1.34}, yhcH^{1.33}, yhfD^{1.32}, ywiB^{1.32}, ysaA^{1.3}, ylfD^{1.28}, ydbD^{1.27}, ybiG^{1.25}, yrgF^{1.25}, yuhA^{1.25}, yjjE^{1.23}, ykaE^{1.22}, yvja^{1.21}, ywiA^{1.21}, yihF^{1.2}, ytdB^{1.17}, ylfB1^{1.16}, yudK^{0.87}, yuiE^{0.84}, yahA^{0.83}, ymbD^{0.82}, ywaB^{0.82}, yfbJ^{0.81}, yweF^{0.81}, yhjB^{0.8}, yajH^{0.78}, ypaH^{0.77}, yibF^{0.76}, ynhl^{0.76}, yxaC^{0.76}, ycjA^{0.75}, yxaB^{0.75}, yliD^{0.74}, yniH^{0.73}, yviJ^{0.73}, ycfD^{0.72}, yheE^{0.72}, yogG^{0.72}, yvaB^{0.72}, ykjA^{0.7}, ytdA^{0.7}, ybaG^{0.69}, ycci^{0.69}, yjcE^{0.69}, yccE^{0.68}, yrbH^{0.67}, ybiC0^{0.66}, ywbA^{0.65}, ygiJ^{0.64}, yleB^{0.64}, yoaG^{0.64}, yreA^{0.63}, ysjE^{0.62}, yfdB^{0.61}, yngA^{0.61}, yhgB^{0.6}, ysbD^{0.59}, ywcC^{0.58}, yeiG^{0.57}, ylaE^{0.55}, ysel^{0.54}, yciC^{0.53}, yheB^{0.53}, yvgA^{0.53}, ydiB^{0.52}, ygiK^{0.52}, yqhA^{0.52}, yraA^{0.52}, yrgE^{0.52}, ybaD^{0.51}, yshB^{0.51}, ybhD^{0.49}, ylaD^{0.49}, ytcC^{0.49}, yjcD^{0.47}, ykiE^{0.47}, ynil^{0.46}, ygaJ^{0.45}, yejC^{0.38}, ypdC^{0.37}, ylhB^{0.35}, yiaC^{0.31}, ypcB^{0.3}, ypcC0^{0.29}, yxcC^{0.26}</p> | | | |

* genes differentially expressed but with p-value hardly higher than the significativity threshold.

Table 3: Comparison of *L. lactis* transcriptional response to isoleucine starvation with other responses available in the literature.

Signs + or - represents respectively up- or down-regulations. Numbers in the column entitled Ile represent genes involved in each (sub) category during isoleucine starvation response. Other columns give the number of genes that are common between isoleucine starvation response and the given response: growth rate decrease (μ), stringent response induction (SR), *codY* (CodY), *ccpA* or *glnR* (GlnR) deletion and aeration stress (O_2).

| FUNCTIONAL CATEGORIES or sub-categories | + | | | | | | | - | | | | | | |
|---|-----|-------|----|------|------|------|-------|-----|-------|----|------|------|------|-------|
| | Ile | μ | SR | CodY | CcpA | GlnR | O_2 | Ile | μ | SR | CodY | CcpA | GlnR | O_2 |
| AMINO ACID BIOSYNTHESIS | 25 | 2 | 8 | 17 | 0 | 0 | 0 | 4 | 2 | 1 | 0 | 0 | 1 | 0 |
| -- Aromatic amino acid family | 7 | 0 | 3 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| -- Aspartate family | 3 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Branched chain family | 7 | 0 | 3 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Glutamate family | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 3 | 1 | 1 | 0 | 0 | 1 | 0 |
| -- Histidine family | 4 | 2 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Serine family | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| BIOSYNTHESIS OF COFACTORS, PROSTHETIC GROUPS, AND CARRIERS | 6 | 1 | 3 | 0 | 1 | 0 | 1 | 11 | 0 | 2 | 0 | 1 | 0 | 0 |
| -- Folic acid | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Heme and porphyrin | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Menaquinone and ubiquinone | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 1 | 0 | 0 | 0 | 0 |
| -- Pantothenate | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| -- Riboflavin and cobalamin | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Thioredoxin, glutaredoxin and glutathione | 3 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Thiamin | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| -- Pyridine nucleotides | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CELL ENVELOPE | 12 | 3 | 4 | 0 | 0 | 0 | 2 | 10 | 4 | 2 | 0 | 1 | 0 | 1 |
| -- Membranes, lipoproteins and porins | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 1 | 1 | 0 | 1 | 0 | 1 |
| -- Murein sacculus and peptidoglycan | 6 | 0 | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| -- Surface polysaccharides and antigens | 5 | 3 | 2 | 0 | 0 | 0 | 2 | 4 | 3 | 0 | 0 | 0 | 0 | 0 |
| CELLULAR PROCESSES | 4 | 1 | 1 | 1 | 0 | 0 | 1 | 3 | 1 | 1 | 0 | 0 | 0 | 0 |
| -- Cell division | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Cell killing | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Chaperones | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 0 | 0 | 0 | 0 |
| -- Detoxification | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | | |
|--|----------|----|----|------|------|------|----------------|----------|---|----|------|------|------|----------------|
| -- Protein and peptide secretion | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Transformation | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CENTRAL INTERMEDIARY METABOLISM | 4 | 0 | 2 | 2 | 0 | 0 | 2 | 4 | 0 | 3 | 0 | 0 | 0 | 0 |
| -- General | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| FUNCTIONAL CATEGORIES or sub-categories | + | | | | | | | - | | | | | | |
| | Ile | μ | SR | CodY | CcpA | GlnR | O ₂ | Ile | μ | SR | CodY | CcpA | GlnR | O ₂ |
| -- Amino sugars | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| -- Degradation of polysaccharides | 2 | 0 | 1 | 1 | 0 | 0 | 1 | 2 | 0 | 2 | 0 | 0 | 0 | 0 |
| -- Phosphorus compounds | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Polyamine biosynthesis | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Other | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ENERGY METABOLISM | 35 | 10 | 16 | 4 | 5 | 0 | 6 | 25 | 6 | 9 | 0 | 1 | 1 | 1 |
| -- Aerobic | 12 | 5 | 6 | 1 | 0 | 0 | 1 | 3 | 2 | 1 | 0 | 1 | 0 | 0 |
| -- Amino acids and amines | 6 | 2 | 3 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| -- Anaerobic | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- ATP-PMF conversion | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Electron transport | 4 | 1 | 2 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Entner-Doudoroff | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Fermentation | 3 | 0 | 1 | 2 | 0 | 0 | 0 | 3 | 1 | 1 | 0 | 0 | 0 | 0 |
| -- Gluconeogenesis | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Glycolysis | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 2 | 1 | 0 | 0 | 0 | 0 |
| --Pentose phosphate pathway | 3 | 1 | 2 | 0 | 0 | 0 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| -- Pyruvate dehydrogenase | 4 | 0 | 1 | 0 | 3 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Sugars | 3 | 1 | 1 | 0 | 2 | 0 | 0 | 9 | 0 | 6 | 0 | 0 | 0 | 0 |
| -- TCA cycle | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| FATTY ACID AND PHOSPHOLIPID METABOLISM | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 9 | 2 | 0 | 1 | 0 | 0 |
| PURINES, PYRIMIDINES, NUCLEOSIDES AND NUCLEOTIDES | 13 | 3 | 4 | 0 | 2 | 0 | 2 | 10 | 3 | 6 | 0 | 1 | 0 | 0 |
| -- 2'-deoxyribonucleotide metabolism | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Nucleotide and nucleoside interconversion | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Purine ribonucleotide biosynthesis | 4 | 2 | 2 | 0 | 2 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 0 |
| -- Pyrimidine ribonucleotide biosynthesis | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 1 | 3 | 0 | 1 | 0 | 0 |
| -- Salvage of nucleosides and nucleotides | 4 | 1 | 0 | 0 | 0 | 0 | 1 | 3 | 1 | 1 | 0 | 0 | 0 | 0 |
| -- Sugar-nucleotide biosynthesis and interconversions | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| REGULATORY FUNCTIONS | 27 | 2 | 8 | 2 | 0 | 0 | 1 | 14 | 4 | 3 | 0 | 2 | 0 | 0 |
| -- General | 18 | 1 | 5 | 2 | 0 | 0 | 1 | 6 | 2 | 3 | 0 | 1 | 0 | 0 |

| | | | | | | | | | | | | | | |
|---|-----|----|----|------|------|------|----------------|-----|----|----|------|------|------|----------------|
| -- Two-component systems | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Lac-Ifamily regulators | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| -- LysR-family regulators | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- AraC-family regulators | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- GntR-family regulators | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- DeoR-family regulators | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- MarR-family regulators | 3 | 1 | 1 | 0 | 0 | 0 | 0 | 3 | 2 | 0 | 0 | 0 | 0 | 0 |
| FUNCTIONAL CATEGORIES or sub-categories | + | | | | | | | - | | | | | | |
| | Ile | μ | SR | CodY | CcpA | GlnR | O ₂ | Ile | μ | SR | CodY | CcpA | GlnR | O ₂ |
| -- BglG-family regulators | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- GTP-binding proteins | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| REPLICATION | 9 | 1 | 2 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| -- Degradation of DNA | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- DNA replication, restriction, modification, recombination and repair | 7 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| TRANSCRIPTION | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 3 | 0 | 0 | 0 | 0 | 0 |
| -- Degradation of RNA | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- RNA synthesis, modification and DNA transcription | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 3 | 0 | 0 | 0 | 0 | 0 |
| -- RNA processing | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| TRANSLATION | 14 | 3 | 5 | 0 | 0 | 0 | 1 | 43 | 33 | 11 | 0 | 5 | 0 | 0 |
| -- Amino acyl tRNA synthetases | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 8 | 4 | 2 | 0 | 3 | 0 | 0 |
| -- Degradation of proteins, peptides and glycopeptides | 4 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| -- Protein modification | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Ribosomal proteins: synthesis and modification | 6 | 1 | 1 | 0 | 0 | 0 | 1 | 27 | 26 | 7 | 0 | 2 | 0 | 0 |
| -- Translation factors | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 3 | 1 | 0 | 0 | 0 | 0 |
| TRANSPORT AND BINDING PROTEINS | 58 | 10 | 3 | 12 | 3 | 0 | 4 | 32 | 6 | 17 | 1 | 2 | 0 | 0 |
| -- General | 19 | 0 | 0 | 2 | 0 | 0 | 1 | 5 | 0 | 1 | 0 | 0 | 0 | 0 |
| -- Amino acids, peptides and amines | 18 | 4 | 0 | 10 | 3 | 0 | 2 | 5 | 1 | 3 | 1 | 0 | 0 | 0 |
| -- Anions | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| -- Carbohydrates, organic alcohols and acids | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 3 | 0 | 0 | 0 | 0 |
| -- Cations | 8 | 3 | 0 | 0 | 0 | 0 | 1 | 5 | 1 | 2 | 0 | 2 | 0 | 0 |
| -- Nucleosides, purines and pyrimidines | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- PTS system | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 3 | 6 | 0 | 0 | 0 | 0 |
| -- Multidrug resistance | 8 | 2 | 3 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 0 |
| OTHER CATEGORIES | 39 | 18 | 8 | 1 | 1 | 0 | 2 | 7 | 0 | 1 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | | | | |
|--|------------|-----------|------------|-----------|-----------|----------|-----------|------------|-----------|-----------|--------------|-----------|--------------|----------|
| -- Adaptations and atypical conditions | 5 | 1 | 3 | 0 | 1 | 0 | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| -- Drug and analog sensitivity | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Phage related functions and prophages | 27 | 12 | 4 | 0 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Transposon related functions | 4 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| -- Other | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| UNKNOWN | 152 | 27 | 56 | 4 | 0 | 0 | 12 | 79 | 25 | 16 | 1 | 0 | 0 | 2 |
| Total (in number of genes) | 405 | 81 | 120 | 43 | 12 | 0 | 34 | 262 | 96 | 75 | 2 | 14 | 2 | 4 |
| Global overlapping (in percentage of Ile regulon) | 61 | 12 | 22 | 6 | 2 | 0 | 5 | 39 | 14 | 11 | <1 | 2 | <1 | 1 |

**Chapitre III : Modélisation des efficacités
de traduction et de la stabilité des protéines
de *Lactococcus lactis***

Si le dogme central de la biologie moléculaire stipule que l'ADN code pour des ARN messagers codant eux même pour des protéines, il ne faut pas oublier que l'expression des gènes, bien que n'impliquant *a priori* que deux étapes, la transcription et la traduction, n'est pas un mécanisme linéaire et met également en jeu d'autres niveaux de régulations (Mata *et al.*, 2005). Au moins deux processus biologiques supplémentaires peuvent en effet affecter les niveaux intracellulaires de protéines : la dégradation des protéines et la dilution *via* la croissance cellulaire. En outre, il n'est pas exclu que des régulations de la traduction soient également à l'origine du manque de corrélation observé entre les niveaux d'ARN messagers et de protéines (Nie *et al.*, 2007).

L'étude présentée dans ce chapitre vise à considérer ces différents mécanismes susceptibles de moduler la concentration des protéines au travers de l'intégration des données de transcriptome et de protéome. Ces travaux s'inscrivent dans une démarche de biologie intégrative et sont plus spécifiquement abordés par une approche de type « bottom-up ». Des équations mathématiques décrivant les processus de traduction, dilution et dégradation des protéines sont en effet posées comme point de départ.

Nous avons travaillé avec des données expérimentales pour *Lactococcus lactis* provenant de régimes permanents correspondant, soit à des cultures continues en chémostat, soit à des cellules en phase exponentielle de croissance. L'utilisation d'états stables a autorisé la simplification des équations mathématiques. Parallèlement, d'autres équations ont pu être ajustées aux données expérimentales et ont permis leur résolution. La combinaison de ces approches de modélisation des processus biologiques et des données expérimentales nous a permis de quantifier les efficacités de traduction et les stabilités de chacune des protéines, à chacun des taux de croissance étudiés (figure 10). Les résultats montrent qu'efficacité de traduction et stabilité des protéines ne sont pas constantes mais inversement proportionnelles au taux de croissance (figure 10). Ainsi, à faible taux de croissance, la traduction est plus efficace qu'en croissance rapide mais, de façon antagoniste, la dégradation des protéines est plus importante. Les valeurs médianes de dégradation calculées à chacun des taux de croissance s'étendent de 23 à 224 minutes, ce qui souligne l'importance du phénomène de dégradation des protéines pourtant couramment considéré comme négligeable. D'autre part, l'efficacité de traduction, comme la dégradation des protéines, varient fortement selon la protéine étudiée. Nous avons cependant pu noter, parmi les plus efficacement traduites, un enrichissement en protéines impliquées dans les mécanismes de croissance. Similairement, les

protéines associées au stress sont apparues comme les plus stables. Les variations d'efficacité de traduction et de dégradation des protéines ont également pu être reliées à des paramètres génétiques (figure 10). Les gènes présentant les plus forts biais de codon et d'acides aminés sont logiquement les plus efficacement traduits alors qu'inversement, les gènes les plus longs sont moins efficacement traduits. La dégradation des protéines semble quant à elle uniquement affectée par la présence de résidus cystéine, probablement associés à la formation de ponts disulfures stabilisateurs.

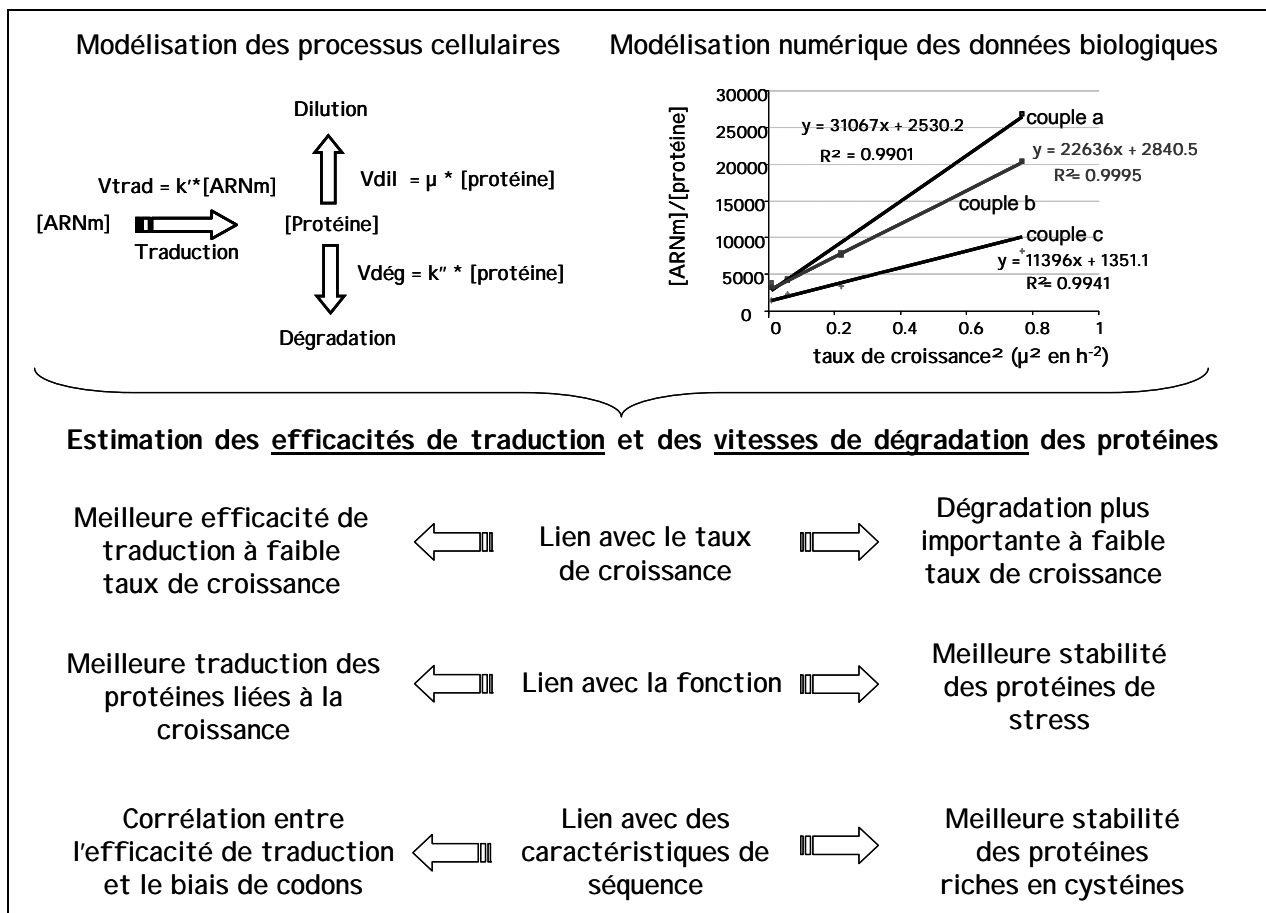


Figure 10 : Estimation des efficacités de traduction et de la dégradation des protéines de *L. lactis* en régime permanent.

Des coefficients de régulation ont pu être calculés en appliquant une méthode de modélisation des contrôles cellulaires initialement développée pour l'étude des régulations du flux glycolytique chez la levure *Saccharomyces cerevisiae* (Daran-Lapujade *et al.*, 2007). Ces coefficients ont révélé que les niveaux de protéines étaient majoritairement dépendants de la « disparition ». Ce terme combine les effets de la dégradation et de la dilution dont les influences ne peuvent pas être dissociées par cette approche. A faible taux de croissance, il

apparaît que le contrôle des niveaux de protéines est clairement dépendant soit de la traduction, soit de la disparition selon la protéine en question. A l'inverse, à des taux de croissance plus élevés la régulation se complexifie et le contrôle semble plutôt partagé.

Ces travaux nous ont permis d'estimer l'efficacité de traduction et la dégradation des protéines, deux paramètres difficilement mesurables par l'expérimentation biologique à l'échelle du génome et dont aucune étude bactérienne ne fait encore état à ce jour. Les résultats obtenus ont également montré la complexité du contrôle cellulaire des niveaux de protéines qui n'implique pas uniquement le processus de traduction mais fait aussi intervenir la dilution et la dégradation, trois mécanismes fortement dépendants du taux de croissance.

Version soumise à PLOS Computational Biology.

Transcriptome and Proteome Exploration to Model Translation Efficiency and Protein Stability in *Lactococcus lactis*

Clémentine Dressaire^{1,2,3}, Christophe Gitton⁴, Pascal Loubière^{1,2,3}, Véronique Monnet⁴, Isabelle Queinnec⁵, Muriel Cocaïgn-Bousquet^{1,2,3*}

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

(2) INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, F-31077 Toulouse, France

(3) CNRS, UMR5504, F-31077 Toulouse, France

(4) INRA, Unité de Biochimie Bactérienne, UR477, F-78350 Jouy en Josas, France

(5) Université de Toulouse; LAAS-CNRS, 7 Avenue du Colonel Roche, F-31077 Toulouse, France

*corresponding author

Abstract

Background: This genome scale study aimed at analysing the various parameters influencing protein levels in cells. To achieve this goal, the model bacterium *Lactococcus lactis* was grown at steady state in continuous cultures at different growth rates and proteomic and transcriptomic data were thoroughly compared.

Methodology/principal findings: Ratios of mRNA to protein were highly variable among proteins but also, for a given gene, between the different growth conditions. The modeling of cellular processes combined with a data fitting modeling approach allowed both translation efficiencies and degradation rates to be estimated for each protein in each growth condition. Estimated translation efficiencies and degradation rates strongly differed between proteins and were tested for their biological significance through statistical correlations with relevant parameters such as codon or amino acid bias. These efficiencies and degradation rates were not constant in all growth conditions and were inversely proportional to the growth rate, indicating a more efficient translation at low growth rate but an antagonistic high protein degradation rate. Estimated protein median half-lives ranged from 23 to 224 min, underlying the importance of protein degradation notably at low growth rates. The regulation of intracellular protein level was analysed through regulatory coefficient calculations revealing a complex control depending on protein and growth conditions.

Conclusions/significance: The modeling approach enabled translational efficiencies and protein degradation rates to be estimated, two biological parameters extremely difficult to determine experimentally and generally lacking in bacteria. This method is generic and can now be extended to other environments and/or other micro-organisms.

Authors Summary

This work is in the field of systems biology. From a profound comparison of proteomic and transcriptomic data in various culture conditions, our objective was to better understand the regulation of protein levels through translation, dilution rate and protein degradation. We have demonstrated that bacteria exert a tight control on intracellular protein level, through a multi-level regulation involving all three phenomena. We have estimated translational efficiencies and protein degradation rates. These two biological parameters are extremely difficult to measure experimentally and have not been previously determined or estimated in bacteria. We have found that they are growth rate dependent indicating a fine control of translation and degradation processes. We have worked with the small genome bacterium *Lactococcus lactis* on a limited number of mRNA - protein couples but keeping in mind that this approach could

be extended to other micro-organisms and biological phenomena. We have exhibited that mathematical modeling associated to experimental steady-states cultures is a powerful tool to learn about microorganisms' physiology.

Introduction

In the era of “omics”, systems biology has emerged with the availability of genome-wide data from different levels, *i.e.* genome, transcriptome, proteome, metabolome [1,2]. This approach aims at integrating omics data, mainly through computational and mathematical models [3,4] so as to decipher biological systems as a whole [5]. The integration of transcriptomic and proteomic results is a huge challenge by itself. The literature usually exploits these two approaches as complementary tools and does not often provide a correct confrontation of the two datasets. Until now, only a few researchers, mainly interested in yeast physiology [6,7], have been working on this aspect and the results typically revealed modest correlations between those two datasets [8-10]. These weak correlations between transcript and protein levels can be the consequence of the involvement of post-transcriptional regulations [11], such as translation control and protein degradation as evidenced by Brockmann *et al.* [12]. Translation regulations are believed to be involved in protein level control but are generally studied at the level of controlling specific molecular mechanisms and not at the genome scale [13-15]. Although polysome profile analysis allows translation efficiencies to be experimentally determined for the various transcripts simultaneously, this technique has been only rarely used and almost exclusively for *S. cerevisiae* [16]. Protein stability can also influence intracellular protein level and thus the correlation between transcript and protein [10,17,18]. However protein stability is rarely studied at the genome scale and data are only available for *S. cerevisiae* [19,20]. Finally, the rate of protein disappearance due to protein dilution by cellular growth is also potentially involved in protein level modifications but this physical phenomenon is generally neglected. More generally, even if translation efficiency, protein degradation and dilution rate can all influence protein levels, these parameters are not usually studied simultaneously. The role of each parameter in a whole cellular adaptation process has not been elucidated and it is not clearly known today which parameter is preponderant and if the control is constant or not when environmental conditions are modified.

The aim of this study was to analyse the control of intracellular protein level taking into account all the parameters of this control, in a prokaryotic organism, the model of lactic acid bacteria, *Lactococcus lactis*. To achieve this purpose, transcriptomic and proteomic analyses

were performed with cells from the same culture. Transcriptomic data were already available [21] and the corresponding proteome measurement was performed. The whole protein related processes including translation, dilution rate and protein degradation were modelled, and, since biological data were obtained at steady state, equations describing the protein levels equilibrium were solved. This modeling approach allowed translation efficiency and protein degradation to be estimated and the relative involvement of all the various parameters of protein control to be analysed.

Results

Transcriptomic - Proteomic data

L. lactis was grown in continuous culture at different growth rates in the conditions previously described [21] and samples were taken for both transcriptome and proteome analysis at three dilution rates, *i.e.* 0.09, 0.24 and 0.47 h⁻¹: the lowest growth rate ($\mu = 0.09$ h⁻¹) was chosen as reference. Despite the small size of the *L. lactis* genome (2310 genes [22]), a total of 346 different proteins were quantified corresponding to 308 different proteins measured in each repetition for each of the 3 steady states. Among these proteins, 194 showed differential profiles in response to a growth rate increase: 77 with reduced level and 117 with higher level. All the proteins displaying a significant level of modification for at least one of the dilution rates examined and are listed in Table 1. In accordance with what has previously been found with transcriptome analysis [21], increased levels of proteins related to biogenesis were observed when the growth rate was increased, *i.e.* proteins related to transcription (GreA, NusA, QueA, RpoA), translation and more specifically ribosomal proteins (GatA, GatB, RplE, RplI, RplJ, RplK, RplM, RplN, RpmE, RpsA, RpsF, RpsT), enzymes related to fatty acid and phospholipid metabolism (AccA, AccD, FabD, FabF, FabG1, FabH, FabZ1, HmcM, ThiL, YdiD, YscE), two proteins involved in cell division (FtsY, FtsZ), and some proteins associated with purine, pyrimidine, nucleoside and nucleotide metabolism (Add, Adk, Apt, DeoB, GuaA, GuaC, Hpt, NrdE, PydA, PyrC, PyrE, RmlA, RmlB, Upp).

Proteome profiles differed between the various stress-related proteins. On one hand, the two chaperones DnaK and GroEL, the superoxide dismutase associated to oxygen stress SodA, and DpsA, were found in higher quantity, while on the other hand, the cold shock associated protein CspE, ClpC and the adaptation related peroxidase Tpx, had decreased levels in response to growth rate increase. Besides those opposite punctual regulations, other proteins encoding important functions involved in stress protection such as ATPases or peptidases (excepting PepP), were present at constant levels, independently of the growth rate. No

general tendency was noticeable for this category neither at proteomic level nor at transcriptomic level [21]. In contrast, a wide down-regulation of genes involved in stress protection was observed in yeast when growth rate was increased [23,24]. Finally, one can notice that the two single phage-related proteins measured in those proteomics experiments showed significantly reduced levels at high growth rate. This last observation can be connected with the previously described massive down regulation of the expression of phage-related genes [21].

Table 1. List of proteins, ordered by functional category, and the nature of the change as steady-state growth rates increase from 0.09 to 0.24 and 0.47 h⁻¹ in *L. lactis*.

| Functional category | + | - |
|---|--|---|
| | Proteins displaying higher level in response to growth rate increase | Proteins displaying lower level in response to growth rate increase |
| AMINO ACID BIOSYNTHESIS | AroE, AroH, IlvA, IlvD, LeuA, LeuD, SerB, ThrC | AspB, AspC, GlnA, LysA, ProA |
| BIOSYNTHESIS OF COFACTORS, PROSTHETIC GROUPS, AND CARRIERS | CobQ, IspB | DfpA, GshR |
| CELL ENVELOPE | MurC, MurD | MurE |
| CELLULAR PROCESSES | DnaK, FtsY, FtsZ, GroEL, SodA | AhpC, SecA |
| CENTRAL INTERMEDIARY METABOLISM | GlmS, MetK | GlgD |
| ENERGY METABOLISM | ArcC2, CitC, CitE, DxsB, EnoB, GalE, Glk, GpdA, Mae, NdrI, Pkg, Pmg, Pyk, TpiA, YpjF, YpjH | AckA2, AldC, ArcA, ArcT, CitF, GadB, NifS, PdhA, PdhB, PdhC, Pfl, PycA, RpiA, ScrK, ybiE, YpdB, YpdD, YrbA, YrcA, YrjC |
| FATTY ACID AND PHOSPHOLIPID METABOLISM | AccA, AccD, FabD, FabF, FabG1, FabH, FabZ1, HmcM, ThiL, YdiD, YscE | PlsX |

| Functional category | + | - |
|---|---|--|
| | Proteins displaying higher level in response to growth rate increase | Proteins displaying lower level in response to growth rate increase |
| PURINES, PYRIMIDINES, NUCLEOSIDES AND NUCLEOTIDES | Add, Adk, Apt, DeoB, GuaA, GuaC, Hpt, NrdE, PydA, PyrC, PyrE, RmlA, RmlB, Upp | Pdp, PurB |
| REGULATORY FUNCTIONS | FhuR, LlrC, ObgL, PurR, PyrR, YsxL | CcpA, EraL, FhuR, LlrA |
| REPLICATION | SsbB | HslA, ParC |
| TRANSCRIPTION | GreA, NusA, QueA, RpoA | - |
| TRANSLATION | Frr, FusA, GatA, GatB, LeuS, PpiB, PrfA, RplE, RplI, RplJ, RplK, RplM, RplN, RpmE, RpsA, RpsF, RpsT, SerS, TrpS, Tsf, Tuf, TyrS | ArgS, GltX, KsgA, LeuS, LysS, PepP, PrfC, ProS, RplA, RpsB, SerS |
| TRANSPORT AND BINDING PROTEINS | GlnQ, OptD, PtsI, PtsK, YjgE, YsfB | BusAA, PtsH |
| OTHER CATEGORIES | DpsA | ClpC, CspE, Pi102, Pi125, Tpx |
| UNKNOWN | YbdD, YciC, YcjB, YejH, YgbD, YjgF, YjhD, YkhD, YlaC, YnhC, YraB, YseF, YshC, YtaA, YtfB, YtjH, YuhE | YahB, YcdB, YeiJ, YgdA, YgiI, YgiK, YhjA, YiiH, YlaF, YnfC, YpdC, YqfE, YrjD, YtgH, YtjA, YtjH, YwcC, YxbE |

Transcriptomic and proteomic analyses were performed with cells collected simultaneously from the same fermentor; thus data can be strictly compared. Proteins and their corresponding transcript levels were compared individually. Transcriptomic data were already available [21] but were nevertheless re-processed so as to obtain concentrations rather than abundances (see materials and methods). For proteomic data, concentration and abundance values are expected to be similar (see materials and methods). The mRNA/protein ratios were not constant for the different proteins since, at a given growth rate, data were spanned among five orders of

magnitude (Figure 1A). These ratios were compared between two conditions using the lowest growth rate (0.09 h^{-1}) as reference (Figure 1B) and they globally increased with the growth rate. This tendency was confirmed when we analyzed similarly data corresponding to the maximum growth rate of 0.88 h^{-1} . These last data, also available in our group (Dressaire *et al.* 2009, submitted), were obtained in batch culture during the exponential growth phase, since this high growth rate could not be reached in continuous culture without any wash out of the cells from the chemostat [25].

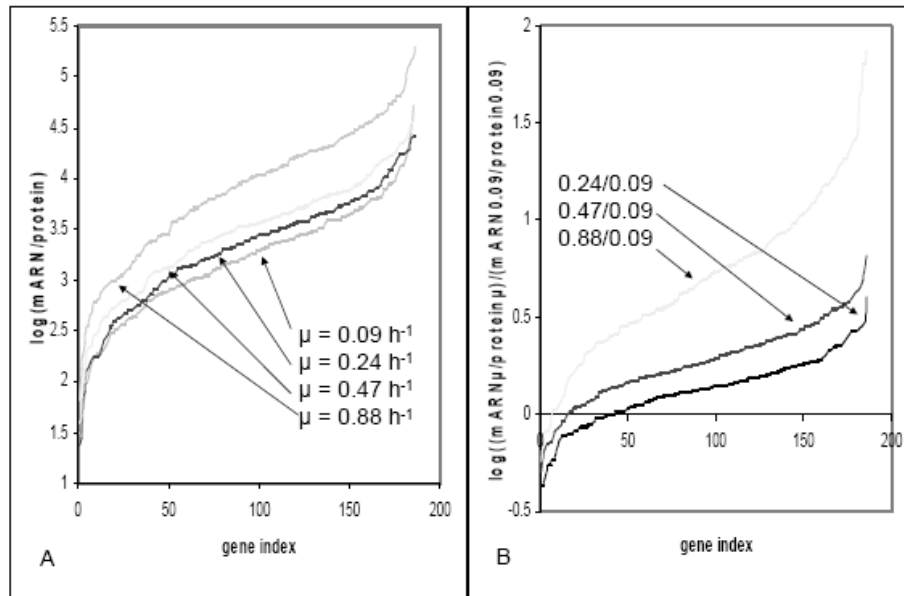


Figure 1. Distributions of mRNA/protein ratios (A) and of their ratios between two different growth conditions (B) in *L. lactis*. The 171 available ratios were ranked in increasing order.

Modeling of cellular process

What normally occurs in bacterial cells is the transcription of genes into mRNA, which are then translated into proteins that can be either diluted by growth or degraded (Figure 2). Hence, protein concentration is determined not only by translation rates (k') but also by dilution and degradation, therefore the following balance equation can be written:

$$\begin{aligned} \frac{d[\text{protein}]}{dt} &= \text{translation rate} - \text{dilution rate} - \text{degradation rate} \\ &= k'[\text{mRNA}] - \mu[\text{protein}] - k''[\text{protein}] \end{aligned} \quad (i)$$

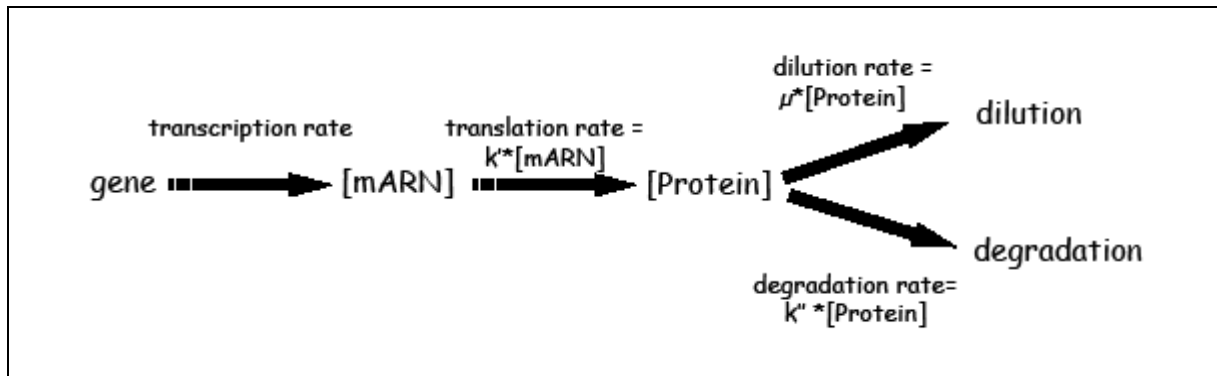


Figure 2. Modeling of the cellular process.

Genes are firstly transcribed into mRNA before being translated into proteins at a rate depending on transcript concentration ($[mRNA]$) and translation efficiency (k'). Proteins are then submitted to dilution by growth and degradation with rate constants equal to the growth rate μ and k'' respectively.

Biological rates were expressed as first order kinetics of their substrate concentration. Dilution constant corresponds to the growth rate (μ), degradation constant (k'') is proportional to protein half-life ($t_{1/2} = \ln(2)/k''$) and k' represents the translation efficiency. At steady state, the various concentrations are expected to remain constant, the time derivative of the protein concentration is equal to zero and the previous equation can be simplified and reorganized as follows:

$$\frac{[mRNA]}{[protein]} = \mu / k' + k'' / k' \quad (ii)$$

In the chemostat cultures at the various growth rates, cells are at steady state; similarly, during the exponential growth phase, cells are physiologically stable and are also considered to be at steady-state [26]. The previous observation, establishing a relationship between mRNA/protein ratios and the growth rate for these four steady states (see above), is in accordance with this last equation (ii). 171 different mRNA and protein couples were available in each repetition of the various steady states (intersection of 308 proteins in the 3 chemostat steady states and 191 proteins in the batch). For only a few proteins were probes missing on the microarray; hence it was not possible to rebuild these couples. In order to estimate translation efficiency and protein degradation rate, the best mathematical solution to the equation (ii) was sought, using numerical estimations performed on Matlab. The k' and k'' values were postulated to be positive, in accordance with biological reality. Different solutions with k' and/or k'' constant, directly or inversely proportional to μ were investigated. Estimation of the best fitting solution was based on the least square criterion [27]. For the 171

couples, the mean sum of the squared residuals (difference between a ratio and its estimation) associated to every combination are given in Table 2. Considering the lower mean sum of the squared residuals, the best solution was obtained when both k' and k'' were proportional to $1/\mu$ ($k' = \alpha/\mu$ and $k'' = \beta/\mu$). Hence the equation (ii) could be written as follows:

$$\frac{[mRNA]}{[protein]} = \mu^2 / \alpha + \beta / \alpha \quad (iii)$$

Table 2. Mean sum of squared residuals associated to different solutions that can be adopted to solve equation (ii).

The lowest mean sum of squared residuals, revealing the solution that best fits, is indicated in bold.

| $k' \backslash k''$ | β | $\beta * \mu$ | β / μ |
|---------------------|-----------------------|----------------------|-----------------------------|
| α | 1.80E+08 +/-1.21E+08 | 2.06E+08 +/-1.24E+08 | 1.79E+08 +/-1.21E+08 |
| $\alpha * \mu$ | 7.13E+08 +/- 4.21E+08 | 7.14E+08 +/-4.21E+08 | 7.12E+08 +/-4.21E+08 |
| α / μ | 7.80E+07 +/- 4.15E+07 | 1.13E+08 +/-5.28E+07 | 4.80E+07 +/-2.97E+07 |

The mRNA/protein ratios were thus linked to the growth rate (μ) through a polynomial function of order two (μ^2), which is consistent with the visual observation of the various curves (not shown). For each mRNA – protein couple the reliability of the two estimated constants α and β was evaluated by their associated R^2 . All regression coefficients are listed in Table 3. The mean linear coefficient (R^2) associated to this model was 0.83 ± 0.04 .

Finally, the consistency of our modeling approach was checked when removing the data of the batch exponential growth phase from the analysis. A high mean R^2 of 0.77 ± 0.12 was still obtained using chemostat data exclusively. On the contrary, data not at steady state coming from other growth phases in batch cultivation could not be included. Indeed when taking into account mRNA/protein values during growth deceleration or during stationary phase, R^2 was strongly affected and dropped to 0.21 ± 0.03 and 0.24 ± 0.04 respectively.

Model predictions and biological relevance

In order to carry on our modeling approach, the data were filtered and only the 130 couples with $R^2 \geq 0.90$ were retained for further analyses (Table 3). It could be noticed that among the

41 eliminated couples, 15 displayed non monotonous evolutions of their mRNA/protein ratios against μ^2 and 20 had very low mRNA/protein ratios, which were thus more sensitive to errors.

Table 3. Modeling of the mRNA/protein ratios by a linear relation of $\mu^2/\alpha + \beta/\alpha$. α and β estimations and the determination coefficient (R^2) are given for the 171 genes for which both transcriptomic and proteomic data were available. α and β are directly proportional to translation and degradation rates respectively ($k' = \alpha/\mu$ and $k'' = \beta/\mu$). Proteins that do not match the selection criteria ($R^2 \geq 0.90$) are grey-coloured.

| Protein | α | β | R^2 | Protein | α | β | R^2 |
|---------|-----------------------|---------|-------|---------|----------|---------|-------|
| ACCC | 2.58 ^E -05 | 0.0044 | 0.951 | GLTD | 1.86E-05 | 0.0024 | 0.972 |
| ACKA1 | 7.48 ^E -05 | 0.1044 | 0.986 | GLYA | 4.22E-05 | 0.0135 | 0.934 |
| ACKA2 | 3.87 ^E -05 | 0.0046 | 0.966 | GPDA | 2.89E-05 | 0.0136 | 0.953 |
| ADK | 1.35 ^E -04 | 0.6296 | 0.867 | GREA | 1.77E-04 | 0.1891 | 0.944 |
| AHPC | 5.11 ^E -05 | 0.0000 | 0.977 | GUAA | 1.55E-04 | 0.2359 | 0.969 |
| ALS | 3.63 ^E -05 | 0.1298 | 0.991 | HMCM | 4.85E-05 | 0.1263 | 0.916 |
| ARAT | 1.02 ^E -05 | 0.1763 | 0.989 | HOM | 1.68E-05 | 0.2862 | 0.867 |
| ARGS | 1.09 ^E -03 | 1.6172 | 0.083 | HPT | 1.08E-04 | 0.4998 | 0.999 |
| AROH | 4.51 ^E -05 | 0.1647 | 0.971 | HSLA | 3.87E-03 | 0.1440 | 0.988 |
| ASPS | 2.93E-04 | 0.7828 | 0.917 | ILVD | 7.93E-05 | 0.0000 | 0.935 |
| BCAT | 1.06E-04 | 0.1958 | 0.988 | LDH | 1.99E-03 | 0.1915 | 0.997 |
| CCPA | 4.06E-04 | 0.1217 | 0.962 | LEUS | 1.96E-04 | 0.8737 | 0.396 |
| CITE | 6.31E-04 | 1.9184 | 0.336 | LLRA | 4.63E-05 | 0.5233 | 0.922 |
| CITF | 1.42E-04 | 0.2135 | 0.894 | LLRC | 1.34E-02 | 15.5080 | 0.105 |
| CLPB | 4.13E-05 | 0.1174 | 0.994 | LPLL | 6.98E-05 | 0.3232 | 0.907 |
| CLPE | 1.98E-05 | 0.0000 | 0.959 | LYSA | A=0 | A=0 | 0.000 |
| CODY | 8.47E-05 | 0.1325 | 0.984 | LYSS | 1.91E-04 | 1.5935 | 0.128 |
| CSPE | 2.11E-04 | 0.8853 | 0.550 | MAE | 2.03E-04 | 0.1173 | 0.948 |
| DDL | 8.25E-05 | 0.1152 | 0.962 | MENB | 2.73E-05 | 0.0000 | 0.936 |
| DEF | 1.15E-05 | 0.0301 | 0.974 | METK | 3.24E-05 | 0.0826 | 0.991 |
| DEOD | 4.65E-04 | 1.6811 | 0.529 | METS | 4.67E-05 | 0.3397 | 0.835 |
| DNAK | 1.06E-03 | 0.3586 | 0.933 | MURC | 1.38E-03 | 2.6771 | 0.426 |
| DNAN | 6.62E-04 | 0.2719 | 0.997 | MURD | 1.85E-04 | 2.2192 | 0.320 |
| DPSA | 3.39E-04 | 0.1474 | 0.999 | MURF | 2.28E-05 | 0.0000 | 0.939 |
| DXSB | 2.23E-05 | 0.0131 | 0.963 | NADE | 1.16E-04 | 0.1545 | 0.996 |
| ENOA | 1.45E-03 | 0.5001 | 0.981 | NIFS | 1.02E-04 | 0.2124 | 0.946 |
| FABF | 1.84E-04 | 0.1183 | 0.964 | NUSA | 5.51E-04 | 0.2367 | 0.974 |
| FABG1 | 7.75E-03 | 19.3686 | 0.004 | PCRA | 7.04E-05 | 0.4776 | 0.699 |
| FABZ1 | 9.98E-04 | 1.0359 | 0.842 | PDHA | 2.09E-04 | 0.0734 | 0.991 |
| FBAA | 1.57E-03 | 0.5895 | 0.926 | PDHB | 4.80E-05 | 0.1629 | 0.994 |
| FMT | 3.34E-05 | 0.0000 | 0.942 | PDHD | 1.53E-04 | 0.1095 | 0.992 |
| FRR | 1.06E-03 | 2.1270 | 0.329 | PDP | 2.47E-05 | 0.0245 | 0.970 |
| FTSA | 4.57E-05 | 0.0520 | 0.961 | PEPC | 2.37E-05 | 0.0000 | 0.930 |
| FTSZ | 5.88E-05 | 0.0013 | 0.926 | PEPDB | 2.04E-04 | 0.2243 | 0.996 |
| FUSA | 1.77E-02 | 19.1277 | 0.018 | PEPN | A=0 | A=0 | 0.000 |
| GALE | 1.52E-05 | 0.0889 | 0.964 | PEPO | 1.48E-04 | 0.1551 | 0.996 |
| GAPA | 6.96E-05 | 0.0156 | 0.983 | PEPP | 1.72E-05 | 0.0000 | 0.914 |
| GAPB | 5.92E-03 | 0.3403 | 0.953 | PEPT | A=0 | A=0 | 0.000 |
| GATB | 1.54E-04 | 0.1645 | 0.979 | PEPV | 7.16E-04 | 0.2966 | 0.997 |
| GLK | 9.41E-05 | 3.2549 | 0.071 | PFL | 6.26E-05 | 0.0862 | 0.994 |
| GLMS | 1.08E-04 | 0.5803 | 0.703 | PGK | 7.25E-04 | 0.1807 | 0.997 |
| GLMU | 7.38E-05 | 0.0919 | 0.996 | PHET | 1.87E-04 | 0.7693 | 0.982 |

| Protein | α | β | R^2 | Protein | α | β | R^2 |
|---------|----------|---------|-------|---------|----------|---------|-------|
| PMG | 1.20E-03 | 0.5242 | 0.955 | SECA | 2.18E-05 | 0.0000 | 0.976 |
| PPIB | 5.00E-05 | 0.2441 | 0.980 | SERS | 8.83E-02 | 91.1164 | 0.000 |
| PRFA | 3.07E-05 | 0.5891 | 0.850 | SODA | 1.36E-04 | 0.1338 | 0.980 |
| PRFC | 5.61E-05 | 0.3069 | 0.951 | SSBB | 4.73E-05 | 0.0693 | 0.965 |
| PROS | 8.84E-05 | 0.1200 | 0.994 | THIL | 4.11E-06 | 0.0000 | 0.929 |
| PRSB | 1.12E-04 | 0.1547 | 0.919 | THRC | 2.85E-04 | 0.2982 | 0.988 |
| PTNAB | 1.73E-04 | 0.1310 | 0.972 | THYA | 5.59E-06 | 0.0000 | 0.952 |
| PTSH | A=0 | A=0 | 0.000 | TIG | 1.35E-03 | 0.2949 | 0.982 |
| PTSI | 4.57E-04 | 0.2365 | 0.907 | TKT | 1.63E-04 | 0.2476 | 0.981 |
| PTSK | 4.98E-05 | 0.0042 | 0.955 | TPIA | 1.19E-03 | 0.5620 | 0.992 |
| PURB | 8.04E-04 | 1.7234 | 0.243 | TRPS | 2.10E-05 | 0.0000 | 0.952 |
| PURR | 1.03E-04 | 0.0328 | 0.940 | TRXB1 | 8.20E-05 | 0.0000 | 0.960 |
| PYCA | 2.29E-04 | 0.6238 | 0.676 | TSF | 8.65E-04 | 0.3161 | 0.974 |
| PYK | 1.61E-03 | 0.2715 | 0.998 | TYPA | 2.30E-04 | 0.6947 | 0.907 |
| PYRB | A=0 | A=0 | 0.000 | TYRS | 1.33E-03 | 0.4962 | 0.941 |
| PYRC | 8.88E-05 | 0.2853 | 0.930 | YAHB | 3.18E-05 | 0.0000 | 0.974 |
| PYRE | A=0 | A=0 | 0.000 | YAHG | 6.36E-05 | 0.3240 | 0.943 |
| PYRH | 6.39E-06 | 0.0323 | 0.951 | YBJJ | 1.87E-05 | 0.0000 | 0.925 |
| PYRR | 2.83E-02 | 16.3996 | 0.007 | YCGE | 9.63E-05 | 0.2468 | 0.986 |
| QUEA | 3.73E-04 | 6.3452 | 0.087 | YCIC | 1.21E-04 | 0.3379 | 0.974 |
| RECA | 3.48E-04 | 0.3000 | 0.974 | YDJD | 2.17E-05 | 0.0441 | 0.957 |
| RMLA | 6.82E-05 | 0.3232 | 0.937 | YEIG | 5.27E-06 | 0.0000 | 0.920 |
| RMLB | 1.33E-03 | 0.9964 | 0.881 | YNGE | 4.06E-05 | 0.0285 | 0.955 |
| RMLC | 1.28E-05 | 0.0385 | 0.987 | YNIH | 5.58E-05 | 0.2482 | 0.935 |
| RPLA | 1.10E-03 | 0.2926 | 0.939 | YPDB | 1.10E-04 | 0.1722 | 0.938 |
| RPLC | 2.42E-03 | 0.8605 | 0.319 | YPDC | 2.62E-05 | 0.0553 | 0.975 |
| RPLF | 2.09E-02 | 0.5757 | 0.930 | YPDD | 1.21E-04 | 0.0639 | 0.995 |
| RPLI | 5.37E-04 | 0.0000 | 0.966 | YPJH | 2.02E-05 | 0.0000 | 0.906 |
| RPLJ | 6.45E-03 | 1.3131 | 0.349 | YRAB | 2.14E-04 | 0.0529 | 0.949 |
| RPLK | 7.19E-03 | 0.1591 | 0.984 | YRBA | 3.46E-05 | 0.4067 | 0.879 |
| RPLN | 2.58E-03 | 0.4285 | 0.959 | YRCA | 4.99E-05 | 0.2788 | 0.955 |
| RPLQ | 9.72E-04 | 0.1663 | 0.998 | YSEF | 1.46E-05 | 0.0000 | 0.973 |
| RPME | 3.92E-04 | 0.7395 | 0.777 | YSFB | 4.93E-05 | 0.1421 | 0.964 |
| RPOA | 3.43E-05 | 0.1314 | 0.987 | YTAA | 1.43E-05 | 0.0000 | 0.929 |
| RPSA | 1.60E-03 | 0.7394 | 0.855 | YTDDB | 4.45E-05 | 0.1270 | 1.000 |
| RPSB | 1.07E-03 | 3.1953 | 0.015 | YTGG | 7.14E-06 | 0.0706 | 0.998 |
| RPSC | 3.77E-04 | 0.0000 | 0.950 | YTGH | 1.27E-05 | 0.0000 | 0.925 |
| RPSD | 4.01E-04 | 0.0097 | 0.991 | YTHC | 6.60E-05 | 0.0000 | 0.950 |
| RPSE | 1.77E-03 | 0.4723 | 0.814 | YTJH | 4.18E-05 | 0.2331 | 0.983 |
| RPSF | 9.15E-05 | 0.4810 | 0.997 | YUHE | 1.12E-05 | 0.0000 | 0.915 |
| RPSG | 4.26E-03 | 0.7880 | 0.870 | YWCC | 1.20E-04 | 0.0455 | 0.967 |
| RPSH | 3.43E-04 | 0.1758 | 0.987 | YWED | 1.52E-05 | 0.0000 | 0.926 |
| RPSJ | 1.93E-04 | 0.0000 | 0.957 | ZWF | 4.68E-04 | 0.3305 | 0.999 |
| RPST | 1.48E-03 | 0.3613 | 0.976 | | | | |

The k' and k'' coefficients were numerically calculated for each protein and in each growth condition from the values of α and β (Table 2) by the relation $k' = \alpha/\mu$ and $k'' = \beta/\mu$. Value distributions (Figure 3) demonstrate the wide variability of k' and k'' among proteins but also between growth conditions. The k'' decreases when growth rate increases which is consistent with the general idea that protein degradation is high in stationary phases [28]. Protein half-lives were calculated and median values of $t_{1/2}$ were respectively 23, 61, 119 and 224 min for 0.09, 0.24, 0.47 and 0.88 h^{-1} . These values are in the same order of magnitude as those obtained recently for *S. cerevisiae* (mean 43 min at a growth rate of 0.1 h^{-1} [19]). Like k'' , the translation efficiency k' is also expressed as $1/\mu$ function, indicating that, when the degradation process increases at low growth rate, the translation efficiency is also increasing in order to attenuate this negative biological effect.

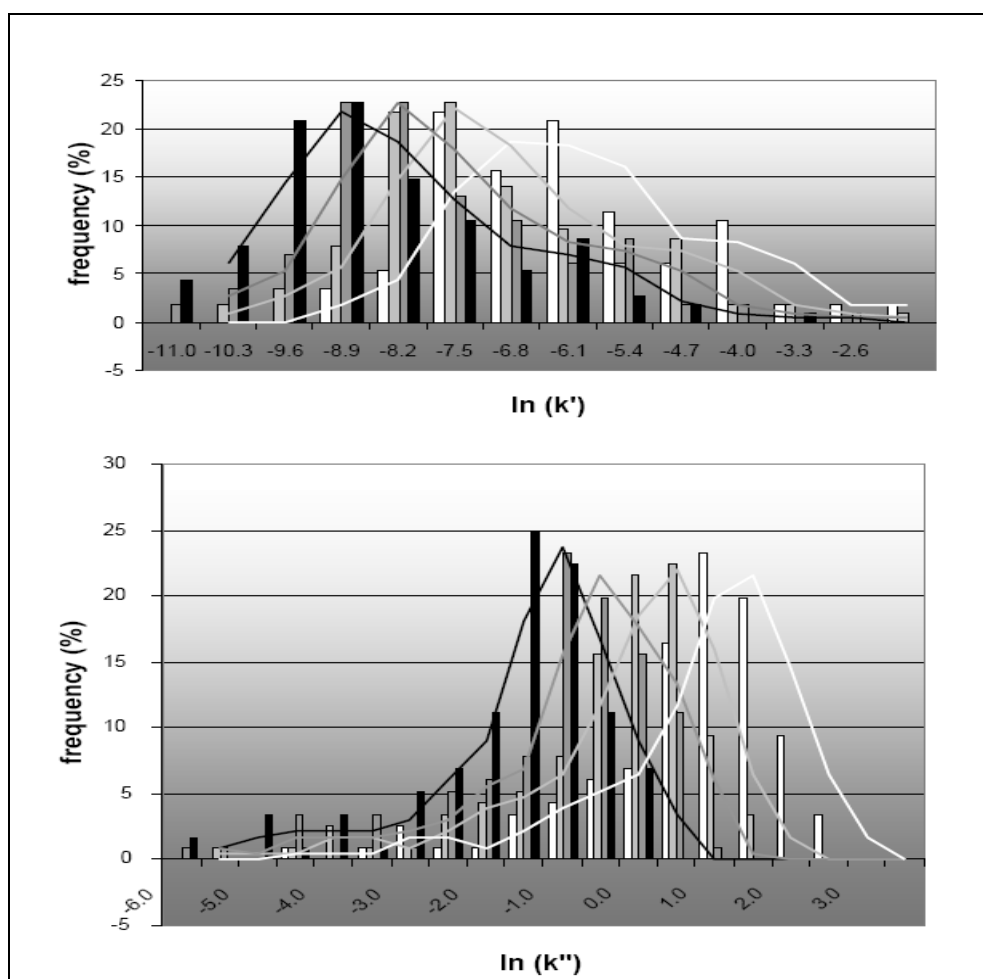


Figure 3. Distribution of translation efficiency, k' , (A) and degradation rate constant, k'' , (B) for different growth rates in *L. lactis*.

Black bars indicate the distribution at the higher growth rate. Different growth rates are represented by the different colours (dark grey for $\mu = 0.47 \text{ h}^{-1}$, grey for $\mu = 0.24 \text{ h}^{-1}$ and white for $\mu = 0.09 \text{ h}^{-1}$). Lines indicate the Gaussian tendency curves of k' and k'' .

Due to the restricted size of the dataset but also to the non-uniform distribution of detected proteins in the various functional categories, it was not possible to use statistical tests to rigorously determine functional enrichments in extreme values of k' or k'' . However, among the 15 genes that are translated the most efficiently (highest α values in Table 3), one can notice the over-representation of genes involved in major cellular processes: the Tig chaperone [29] and proteins involved in replication (HslA, which can unwind DNA and plays a role in its supercoiling, [30]), and translation (ribosomal proteins: RplA, RplF, RplK, RplN, RpsT). Carbon metabolism is also represented by 7 proteins (GapB, EnoA, FbaA, Pmg, Pyk, TpiA, Ldh), all belonging to the same metabolic pathway (glycolysis) which is the major energy metabolism pathway in *L. lactis*. The extremely stable proteins correspond to null values of β , and consequently k'' , were represented by a group of 26 proteins. Remarkably, half of them were related to stress responses: ClpE protease, PepC and PepP peptidases, three reductases that are usually linked to oxidative stress (AhpC, TrxB1, YpjH), but also MurF, involved in parietal structure, YtgH, which is homolog to *Staphylococcus aureus* alkaline stress protein [31], YtaA and YahB two hypothetical protein sharing homologies with *E. coli* universal stress protein Usp [32], YuhE, whose *E. coli* homologue is involved in copper resistance [33], and two cysteine desulfurases (YeiG and YseF) whose corresponding genes in *E. coli* are involved in oxygen and copper stress responses [34]. Moreover, those extremely stable proteins are rather in the last third for translation efficiency. Thus *L. lactis* may limit degradation of stress-related proteins so as to maintain a minimal pool ready to use in case of emergency, which is biologically relevant.

Biological determinants of translation efficiency and protein stability were investigated through correlation studies. Correlations providing a Spearman coefficient (R_{Spearman}) with associated p-values lower than 0.05 were considered as significant. The codon adaptation index (CAI) positively correlates with k' ($R_{\text{Spearman}} = 0.57$). Since CAI directly reflects translation efficiency during the elongation step [35], this result validates our translation efficiency estimations. Translation efficiency is also tightly related to the amino acid composition of proteins. A negative correlation of k' was obtained with tyrosine, cysteine, histidine, aspartic acid and isoleucine frequencies while lysine and alanine richness had a positive influence (Table 4). The amino acids the most used have a positive influence on k' whereas those with a negative effect are the less frequent ones (Table 4). The single exception is for isoleucine, but since it is the limiting nutrient it is not surprising to find it negatively correlated with translation efficiency, despite its high frequency in *L. lactis* proteins. This amino acid bias, together with the codon bias (CAI), shows that translation efficiency is

strongly dependent of the gene sequence. This optimized functioning state is probably the result of a long evolutionary process. Finally it was found that translation efficiency is affected by protein length: the longer the protein, the more k' decreases (R_{Pearson} of -0.18). This negative correlation with length has already been reported for yeast [19] and can possibly be explained by a decrease of the ribosome density on long mRNA as previously shown for *S. cerevisiae* [36]. The only apparent correlation emerging for protein degradation constants k'' is a negative influence of cysteine richness (Table 4).

Table 4. Correlation analysis between amino-acid usage in *L. lactis* proteins and translation efficiencies or degradation rates. These correlations were independent of the growth rate. Significant R_{Spearman} with p-value <0.05 are listed in the table.

| Amino-acid | Mean usage frequency (%/protein) | Correlation with translation efficiency (R_{Spearman}) | Correlation with degradation rate (R_{Spearman}) |
|---------------|----------------------------------|---|---|
| Cysteine | 0.5 | -0.29 | -0.19 |
| Tryptophan | 1.1 | | |
| Histidine | 1.8 | -0.22 | |
| Methionine | 2.6 | | |
| Proline | 3.0 | | |
| Tyrosine | 3.6 | -0.21 | |
| Glutamine | 3.6 | | |
| Arginine | 3.7 | | |
| Phenylalanine | 4.8 | | |
| Aspartic acid | 5.1 | -0.17 | |
| Asparagine | 5.1 | | |
| Threonine | 5.5 | | |
| Glycine | 6.2 | | |
| Serine | 6.4 | | |
| Valine | 6.5 | | |
| Alanine | 7.0 | 0.23 | |
| Glutamic acid | 7.1 | | |
| Isoleucine | 7.9 | -0.19 | |
| Lysine | 7.9 | 0.24 | |
| Leucine | 10 | | |

Controlling mechanisms

Degradation and dilution by growth are competitive reactions. The degradation and dilution constants, k'' and μ , can be directly compared. The k'' is higher than μ at low growth rate but becomes lower after a critical value of 0.39 h^{-1} (Figure 4). The role of the degradation may thus be major at low growth rate while dilution may become the main phenomenon at fast growth.

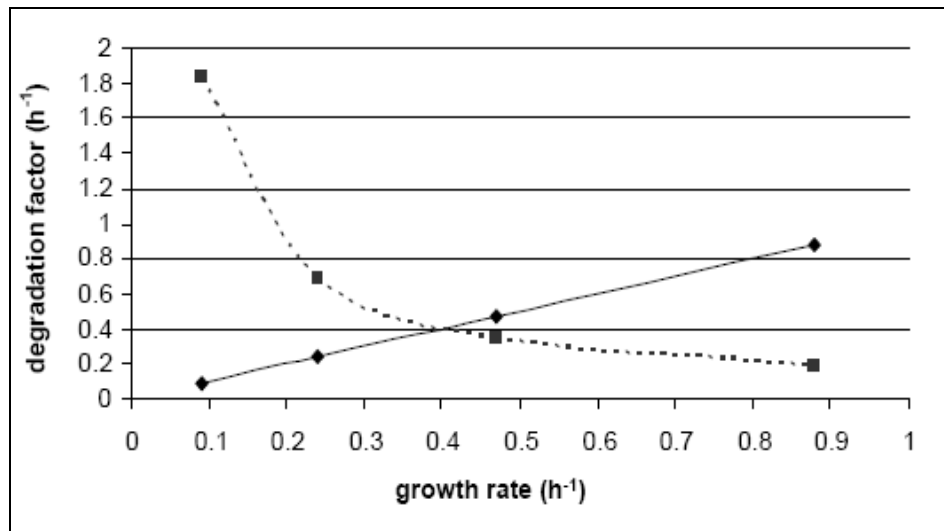


Figure 4. Comparison of growth rate (μ : straight line) and median value of the degradation constant (k'' : dotted line).

More generally, variations in protein concentration between two conditions can be related to changes in the three rates: protein synthesis, degradation and dilution. In order to better understand this regulatory node and identify which are the major controls, the quantitative involvement of the different actors was analyzed. Regulation coefficients corresponding to the protein level control were estimated with a method based on the one developed on *S. cerevisiae* [37,38]. Derivation of equation (ii) leads to the following relationship:

$$\frac{d(\ln(k'[\text{mRNA}]))}{d(\ln([\text{protein}]))} - \frac{d(\ln(\mu + k''))}{d(\ln([\text{protein}]))} = 1 \quad (\text{v})$$

The term $\frac{d(\ln(k'[\text{mRNA}]))}{d(\ln([\text{protein}]))}$ represents translation control on protein concentration and is

called ρ_t . The term $-\frac{d(\ln(\mu + k''))}{d(\ln([\text{protein}]))}$, named ρ_d , includes both the dilution and the

degradation and represents protein control by disappearance. ρ_t and ρ_d were estimated for each growth rate interval (between 0.09 and 0.24 h^{-1} ; between 0.24 and 0.47 h^{-1} ; between 0.47 and 0.88 h^{-1}). The values of ρ_t were used to elucidate the nature of the control. If $\rho_t \leq 0$, protein

disappearance is the major controlling mechanism; if $\rho_t \geq 1$, it is translation; and if $0 < \rho_t < 1$, the control of protein is shared. Independently of the growth rate, protein levels are mostly controlled by disappearance (Table 5). Translation control strongly decreases, and protein level control becomes less specific and more and more shared with increasing growth rates. Similar conclusions were valid when ρ_d was used for the control analysis instead of ρ_t (data not shown).

Table 5. Protein control analysis in the different ranges of growth rate based on ρ_t calculation.

| Growth rates intervals | Translation control of protein levels | Disappearance control of protein levels | Shared control of protein level |
|-----------------------------|---------------------------------------|---|---------------------------------|
| 0.09 – 0.24 h ⁻¹ | 38 % | 60 % | 2 % |
| 0.24 – 0.47 h ⁻¹ | 33 % | 56 % | 11 % |
| 0.47 – 0.88 h ⁻¹ | 14 % | 47 % | 39 % |

Discussion

The comparison of mRNA and protein ratios revealed a strong heterogeneity among genes but also for a given gene, at different growth conditions. Variability among genes has recently been reported for the model yeast *S. cerevisiae* but these ratios remained constant between the two studied conditions, *i.e.* a rich and a poor media [39]. Though lacking in this publication, the maximum growth rates of *S. cerevisiae* were estimated to be 0.46 and 0.35 h⁻¹ respectively in a rich and a poor media (Parrou J.L., personal communication). Thus it is postulated that the growth rate difference between these two conditions was too small to induce changes in mRNA/protein ratios.

The combination of two modeling approaches, one based on biological knowledge and the other on experimental data fitting, has enabled translation efficiency and protein degradation rate to be determined for each protein, phenomena which have been shown to be protein specific and growth rate dependant. The positive correlation of translation efficiency with codon bias in *L. lactis* is consistent with the results obtained for the yeast, though translation efficiencies have been calculated differently [12]. The presence of genes related to major cellular processes essential for growth were marked among the best translated. This finding corroborates with what was found in archaebacteria for ribosomal proteins [40]. In *L. lactis*, the growth-rate dependant variations in translation efficiency are probably not related to

changes in intracellular ribosome quantity. Indeed the mRNA/rRNA ratio estimated from transcriptomic analyses by the total intensity of membrane was independent of growth rate (mean value of 1504 ± 317 pixels). However one has to bear in mind that rRNA does not necessarily means assembled and/or active ribosomes. It is known for example that *E. coli* ribosome activity can be modulated by the inter-conversion between a functional 70S and a dimerized 100S inactive form [41]. To resolve this question, it will be necessary to investigate genome-wide ribosomal activity via polysome distribution which would provide key information to decipher the regulatory processes controlling translation. Polysome profile technology is already available for yeast but may be difficult to adapt to bacteria due to the co-localisation of transcription and translation in the cytoplasm.

Protein half-lives for the whole genome have never been determined nor estimated in any bacteria and data are only available in the literature for *S. cerevisiae*. However studies disagreed in terms of average half-life values: 31 h for Pratt *et al.* against 43 min for Belle *et al.* [19,20]. Those differences could be explained by methodological reasons since one study used pulse chase experiments [20] whereas the other one consisted in a direct measurement of each epitope-tagged proteins [19]. In our study, for *L. lactis*, protein median half-lives ranged from 23 to 224 min. These low values are in good agreement with most recent values obtained for *S. cerevisiae* [19] and indicate that protein degradation is considerably more rapid than was once believed. Degradation rates in *L. lactis* were negatively correlated to cysteine content in proteins. In yeast, stable proteins were previously found to have a higher valine density whereas unstable ones are enriched in serine [19]. It is difficult to strictly compare those results since amino acid bias may be species specific and reflect the particularities of proteases involved in protein degradation. The negative correlation with cysteine could nevertheless be related to the potential formation of disulfide bridges known for stabilizing proteins [42]. The current work also revealed the presence of stress related proteins among the most stable. This last observation differs from results obtained in yeast indicating that ribosomal proteins and enzymes from amino-acids metabolism have the higher half-life [19]. This high stability of stress protein together with the lack of global transcriptional stress response in *L. lactis* when the growth rate is changed clearly underlines differences of stress adaptation mechanisms between the two micro-organisms.

Protein degradation exerts a major role in the cellular adaptation process since protein half-life data depend on the growth rate ($1/\mu$ function). Moreover, the degradation rate is even higher than dilution rate at low growth rate (Figure 4). Considering that protein degradation is an ATP consuming process [44], high protein degradation at slow growth rate may contribute

to the increase of maintenance energy that is generally observed in such conditions [43]. Like protein degradation, translation efficiency is also increased at slower growth rates. Effects of translation efficiency and protein degradation are thus antagonist and this mode of regulation is probably dedicated to attenuate biological changes. Inversely, proteins with the lowest degradation rates also corresponded to low translation efficiencies. The analysis of the regulation involved in the control of protein concentrations demonstrated that it is not constant in the different ranges of growth rate. At low growth rates, disappearance seems to be the main controlling mechanism, which could be attributed to high degradation rate. At intermediary growth rates, the control is more balanced between protein synthesis and disappearance. At high growth rate, the control becomes more complicated with some proteins regulated at the level of synthesis, disappearance or both (shared control). This increased complexity is consistent with cells approaching their maximum growth performance.

With this modeling approach, we have estimated translational efficiencies and protein degradation rates. These two biological parameters are extremely difficult to measure experimentally and have even never been previously determined in bacteria. The method was based on an in depth comparison of proteome and transcriptome data and was developed with the small genome bacterium *L. lactis* on a limited number of mRNA - protein couples (171). It will be possible in the future to broaden these couples since other proteomic methodologies, such as the APEX technology [8], allow more proteins to be detected. The approach remains generic and can be applied to all microorganisms. Modelling equations were solved because steady-states cultures were used: chemostat fermentation technology enabling steady states to be studied has thus proved to be a powerful tool to understand microbial physiology. We have demonstrated that bacteria exert a sharp control on intracellular protein levels, through a multi-level regulation involving three growth rate dependant actors: translation, dilution and degradation. Here, the growth rate was changed *via* chemostat cultures, but such growth rate modifications are also encountered in nature when cells have to face new environments. In this case, the adaptation process involves growth rate adaptation as well as other specific metabolic adaptations. It remains to be determined how the protein control is exerted in such natural environment.

Materials and methods

Growth conditions

Lactococcus lactis ssp. *lactis* IL1403, whose genome has been entirely sequenced [22], was grown as previously described [21]. Briefly, three different growth rates have been studied, namely 0.09, 0.24 and 0.47 h⁻¹ during anaerobic chemostat cultures (under nitrogen atmosphere and regulated pH) on a chemically defined medium limited by isoleucine concentration. For each steady-state, samples have been harvested in at least quadruplicate with a minimum delay of five doubling time between each sampling.

Transcriptomic data reprocessing

Available transcriptomic raw data (geo platforms GSE10256 [21] for chemostat culture and GSE12962 for batch exponential phase) were reprocessed as previously described [45] in order to calculate mRNA concentrations rather than abundances. Raw data were standardized by the all spots' mean intensities for the reference membrane ($\mu = 0.09$ h⁻¹) and corrected by total RNA concentration for each growth condition (3.58 ± 0.39 , 4.92 ± 0.52 , and 7.34 ± 0.28 g for 100 g cell dry weight at $\mu = 0.09$, 0.24 and 0.47 h⁻¹ respectively).

Proteomic analyses

For each condition, three repetitions were performed with independent cultures, extractions and electrophoresis. Bacteria were harvested from the cultures and cell pellets were washed twice with ice-cold 200 mM Na-phosphate, pH 6.4 and re-suspended in 4 ml of 20 mM Na-phosphate buffer, pH 6.4, 1 mM EDTA, 10 mM tributylphosphine, a cocktail of protease inhibitors (P8465; Sigma Aldrich, St Louis, MO) 20-fold diluted and catalase 40 U/ml (C3155; Sigma Aldrich, St Louis, MO) to limit isoform formation. The cell suspension (approximately 35 units of optical density at 600 nm [OD600]/mL) was transferred to the pre-cooled chamber of a BASIC Z cell disrupter (Celld, Warwickshire, United Kingdom) and was subjected to a pressure of 2,500 bars. The suspension was centrifuged at 5,000 x g for 20 min at 4 °C to remove unbroken cells and large cellular debris. The supernatant was collected and centrifuged at 220,000 x g for 30 min at 4 °C. The total protein concentration in the resulting supernatant (cytosolic fraction) was determined with the Coomassie protein assay reagent (Pierce, Rockford, IL) using bovine serum albumin as standard and was included between 1 and 2 mg/mL. The cytosolic fraction was aliquoted and stored frozen at -20 °C.

2-Dimensional electrophoresis: A volume of cytosolic fraction corresponding to 350 µg or 500 µg (for basic gels) of proteins was incubated with nuclease (benzonase, Novagen 70664-

3; 25 U for 100 μ L of cytosolic fraction) for 30 min at 37 °C and then chilled on ice and precipitated with 75 % (vol/vol) methanol. The protein pellet was resuspended in 500 μ l (for pH 4.5-5.5 and 5-6 gels) or 100 μ l (for pH 6-11 gels) of isoelectric focusing (IEF) buffer 1, consisting of 7 M urea, 2 M thiourea, 4 % CHAPS{3-[(3-cholamidopropyl)-dimethylammonio]-1-propanesulfonate}, 100 mM dithiothreitol or 4 mM tributylphosphine and DeStreak (1.2 % v/v, Amersham Biosciences, GE Healthcare) (for basic gels), and 0.5 % pH 4.5 to 5.5 or 5 to 6 or 6 to 11 immobilized pH gradient (IPG) buffer (Amersham Biosciences, GE Healthcare). The sample was loaded on 24 cm pH 4.5 to 5.5 or 5 to 6 immobilized pH gradient (IPG) strip (Amersham Biosciences, GE Healthcare) which was previously rehydrated at 50 V for 11 h. IEF was carried out for 65,000 V.h at a maximum of 8,000 V, using the Protean II IEF cell (Bio-Rad, Hercules, CA). Analysis of basic proteins was performed with 18 cm pH 6-11 (IPG) strip (Amersham Biosciences, GE Healthcare). After passive rehydration of the strip in buffer 1, the protein sample was loaded on sample cups and IEF was carried out for 20,000 V.h at a maximum of 3,500 V using the IPGphor device (Amersham Biosciences, GE Healthcare). Before the second dimension, IPG strips were incubated for 15 min with shaking in 150 mM Tris-HCl pH 8.8, 0.1% w/v SDS. The IPG strip was then positioned on sodium dodecyl sulfate-polyacrylamide gels, using 1 % low-melting-point agarose in 150 mM Tris-HCl, pH 8.8. Second-dimension electrophoresis was performed on 12 % polyacrylamide gels (24 by 20 by 0.1 cm) in 25 mM Tris, 192 mM glycine, 0.1 % sodium dodecyl sulfate, pH 8.3, using the Ettan-Dalt II apparatus. Electrophoresis was run at 1 W/gel for 16 h at 15 °C. The gels were stained with BioSafe colloidal Coomassie blue (Bio-Rad) for 1 h and destained with three successive washes in deionized water.

Images files were recorded at 65536 gray levels (16 BitsPerPixel). Image manipulation and analysis were performed with Samespot V2 software (Nonlinear Dynamics). Protein abundances were given using arbitrary units which correspond to spot volumes and which were calculated as follows: spot area x spot pixel intensity - background intensity.

Protein identification: Protein identification was carried out at the PAPPSO platform (INRA, Jouy-en-Josas) using MALDI-TOF mass spectrometry (MS). Protein spots were excised from Coomassie blue-stained gels and in-gel digested with trypsin. Gel pieces were placed in Eppendorf tubes and washed with 30 μ L 25 mM ammonium carbonate, 50 % acetonitrile. The supernatants were discarded and gel pieces were dried at 37 °C for 15 min. The gels were rehydrated with 20 μ L 50 mM ammonium carbonate containing 100 ng of porcine trypsin (Promega, Madison, WI, USA). The solutions were incubated overnight at 37 °C. The

supernatants containing peptides were directly analyzed by MALDI-TOF Mass spectrometry on a Voyager DE STR Instrument (Applied Biosystems, Framingham, CA, USA). The α -cyano-4-hydroxycinnamic acid matrix was prepared at 4 mg/mL in 0.1 % TFA, 50 % acetonitrile. An equal volume (1 μ L) of matrix and sample were spotted onto the MALDI-TOF target plate. Spectra were acquired in the reflector mode with the following parameters: 2,000 laser intensity, 20 kV accelerating voltage, 62 % grid voltage, 120 ns delay. The mass gates used were 840–3500 Da. Internal calibration was performed by using the trypsin peptides at 842.5 and 2,211.1 Da. Database searches were conducted with the MS-Fit software (<http://prospector.ucsf.edu>) either on an *L. lactis*-specific protein database.

The few spots which could not be identified by MALDI-TOF were analysed by LC-MS/MS using an Ultimate 3000 LC system (Dionex, Voisins le Bretonneux, France) connected to a linear ion trap mass spectrometer (LTQ, Thermo Fisher, USA) by a nanoelectrospray interface to realize the separation, ionisation and fragmentation of peptides, respectively. The supernatant of trypsin hydrolysis was transferred to a new tube and the gel pieces were extracted with a) 25 μ l of buffer B (50 mM ammonium carbonate) and b) two times buffer C (Formic acid 0.1 % acetonitrile 50 %). For each extraction, the gel pieces were incubated for 15 min at room temperature while shaking. The supernatants of each extraction were pooled with the original trypsin digest supernatants and dried for 2h in a Speed-Vacuum concentrator. The peptides were then re-suspended in 25 μ l of precolumn loading buffer (0.08 % TFA and 2 % ACN in water). LC-MS/MS analysis was performed on an Ultimate 3000 LC system (Dionex, Voisins le Bretonneux, France) connected to linear ion trap mass spectrometer (LTQ, Thermo Fisher, USA) by nanoelectrospray interface for separation, ionisation and fragmentation of all peptides. Four μ l of tryptic peptide mixtures were loaded at flow rate 20 μ l/min onto precolumn Pepmap C18 (0.3 x 5 mm, 100 Å , 3 μ m; Dionex). After 4 min, the precolumn was connected with the separating nanocolumn Pepmap C18 (0.075 x 150 mm, 100 Å , 3 μ m, Dionex) and the gradient was started at 300 nL/min. All peptides were separated on the nanocolumn using a linear gradient from 2 to 36 % of buffer B, over 18 min. Eluting buffer A: 0.1 % Formic acid, 2 % acetonitrile and eluting buffer B : 0.1 % Formic acid, 80 % acetonitrile. Including the regeneration step, each run was 50 min in length. Ionization was performed on liquid junction with a spray voltage of 1.3 KV applied to non-coated capillary probe (PicoTip EMITER 10 μ m ID; New Objective, USA). Peptides ions were analysed by the Nth-dependent method as follows: (i) full MS scan (m/z 300-2000), (ii) ZoomScan (scan of the 3 major ions), (iii) MS/MS on these 3 ions with classical peptides fragmentation parameter: Qz = 0.25, activation time = 30 ms, collision energy = 40 %.

Proteins identifications were performed with Bioworks 3.3TM software. The raw data were converted and filtered in peak lists with default data generation parameters for LTQ mass spectrometer. All peak lists of precursor and fragment ions were matched automatically against a *Lactococcus lactis* IL 1403 protein database. The Bioworks search parameter included: trypsin specificity with one missed cleavage, variable oxydation of methionine and the mass tolerance was fixed to 1.4 Da for precursor ion and 0.5 Da for fragment ions. The search results were filtered using Bioworks 3.3. A multiple threshold filter applied at the peptide level consisted of the following criteria: Xcorr magnitude up to 1.7, 2.5 and 3.0 for respectively mono-, di- and tri-charged peptides; peptide probabilities lower than 0.01; ΔCn greater than 0.1 and only the first match result for each identified peptide.

Statistical treatment: Raw spot volumes were normalized by the mean intensity of the corresponding gel. A total of 542 spots corresponding to 352 different proteins were detected. Some of the spots corresponded to proteins mixture and were not considered. The intensities of spots corresponding to protein isoforms in a same gel were summed so as to represent the level of a single protein independently of post-transcriptional modifications. 15 proteins identified both on 4.5-5.5 and 5-6 pH ranges displayed very different amounts. We considered that the best protein level estimation was given by the highest signal.

Since total protein concentrations remain stable whatever the growth rate (42 ± 6 g protein per 100 g cell dry weight), the abundance data are considered to be equivalent to concentrations. Ratios were calculated using the slowest growth phase as a reference. The statistical significance of ratios were evaluated using Student test and False Discovery Rates (FDR, calculated according to Benjamini-Hochberg method [46]) calculated with R free statistical software. Proteins with ratio associated to a False Discovery Rate (FDR) lower than 20 % were considered as differentially regulated.

Mathematical treatments

R² calculations and equations resolution were perform with MATLAB software.

Correlation calculations

Correlations were estimated using R free statistical software to calculate Spearman rank correlation coefficient and the associated p-value.

Acknowledgments

We thank Nic Lindley for useful discussions.

References

1. Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet* 2: 343-372.
2. Kitano H (2002) Systems biology: a brief overview. *Science* 295: 1662-1664.
3. Kitano H (2002) Computational systems biology. *Nature* 420: 206-210.
4. Williamson MP (2005) Systems biology: will it work? *Biochem Soc Trans* 33: 503-506.
5. Joyce AR, Palsson BO (2006) The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* 7: 198-210.
6. Greenbaum D, Colangelo C, Williams K, Gerstein M (2003) Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* 4: 117.
7. Washburn MP, Koller A, Oshiro G, Ulaszek RR, Plouffe D, et al. (2003) Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 100: 3107-3112.
8. Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 25: 117-124.
9. Nie L, Wu G, Culley DE, Scholten JC, Zhang W (2007) Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. *Crit Rev Biotechnol* 27: 63-75.
10. Wu G, Nie L, Zhang W (2008) Integrative analyses of posttranscriptional regulation in the yeast *Saccharomyces cerevisiae* using transcriptomic and proteomic data. *Curr Microbiol* 57: 18-22.
11. Mata J, Marguerat S, Bahler J (2005) Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci* 30: 506-514.
12. Brockmann R, Beyer A, Heinisch JJ, Wilhelm T (2007) Posttranscriptional expression regulation: what determines translation rates? *PLoS Comput Biol* 3: e57.
13. El-Sharoud WM, Graumann PL (2007) Cold shock proteins aid coupling of transcription and translation in bacteria. *Sci Prog* 90: 15-27.
14. Kaberdin VR, Blasi U (2006) Translation initiation and the fate of bacterial mRNAs. *FEMS Microbiol Rev* 30: 967-979.
15. Proud CG (2007) Signalling to translation: how signal transduction pathways control the protein synthetic machinery. *Biochem J* 403: 217-234.
16. Arava Y, Boas FE, Brown PO, Herschlag D (2005) Dissecting eukaryotic translation and its control by ribosome density mapping. *Nucleic Acids Res* 33: 2421-2432.

17. Nie L, Wu G, Brockman FJ, Zhang W (2006) Integrated analysis of transcriptomic and proteomic data of *Desulfovibrio vulgaris*: zero-inflated Poisson regression models to predict abundance of undetected proteins. *Bioinformatics* 22: 1641-1647.
18. Nie L, Wu G, Zhang W (2006) Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in *Desulfovibrio vulgaris*: a quantitative analysis. *Genetics* 174: 2229-2243.
19. Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci U S A* 103: 13004-13009.
20. Pratt JM, Petty J, Riba-Garcia I, Robertson DH, Gaskell SJ, et al. (2002) Dynamics of protein turnover, a missing dimension in proteomics. *Mol Cell Proteomics* 1: 579-591.
21. Dressaire C, Redon E, Milhem H, Besse P, Loubiere P, et al. (2008) Growth rate regulated genes and their wide involvement in the *Lactococcus lactis* stress responses. *BMC Genomics* 9: 343.
22. Bolotin A, Wincker P, Mauger S, Jaillon O, Malarme K, et al. (2001) The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res* 11: 731-753.
23. Brauer MJ, Huttenhower C, Airoidi EM, Rosenstein R, Matese JC, et al. (2008) Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell* 19: 352-367.
24. Regenber B, Grotkjaer T, Winther O, Fausboll A, Akesson M, et al. (2006) Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*. *Genome Biol* 7: R107.
25. Calcott PH (1981) *Continuous culture of cells*; Data LoCCiP, editor: CRC Press. 191 p.
26. Fishov I, Zaritsky A, Grover NB (1995) On microbial states of growth. *Mol Microbiol* 15: 789-794.
27. Van Zandt T (2000) How to fit a response time distribution. *Psychon Bull Rev* 7: 424-465.
28. Goldberg AL, St John AC (1976) Intracellular protein degradation in mammalian and bacterial cells: Part 2. *Annu Rev Biochem* 45: 747-803.
29. Kaiser CM, Chang HC, Agashe VR, Lakshmipathy SK, Etchells SA, et al. (2006) Real-time observation of trigger factor function on translating ribosomes. *Nature* 444: 455-460.
30. Swinger KK, Rice PA (2004) IHF and HU: flexible architects of bent DNA. *Curr Opin Struct Biol* 14: 28-35.

31. Kuroda M, Ohta T, Hayashi H (1995) Isolation and the gene cloning of an alkaline shock protein in methicillin resistant *Staphylococcus aureus*. *Biochem Biophys Res Commun* 207: 978-984.
32. Nystrom T, Neidhardt FC (1994) Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. *Mol Microbiol* 11: 537-544.
33. Gupta SD, Lee BT, Camakaris J, Wu HC (1995) Identification of cutC and cutF (nlpE) genes involved in copper tolerance in *Escherichia coli*. *J Bacteriol* 177: 4207-4215.
34. Outten FW, Wood MJ, Munoz FM, Storz G (2003) The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in *Escherichia coli*. *J Biol Chem* 278: 45713-45719.
35. Sharp PM, Li WH (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 15: 1281-1295.
36. Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, et al. (2003) Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 100: 3889-3894.
37. Daran-Lapujade P, Rossell S, van Gulik WM, Luttik MA, de Groot MJ, et al. (2007) The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *Proc Natl Acad Sci U S A* 104: 15753-15758.
38. ter Kuile BH, Westerhoff HV (2001) Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett* 500: 169-171.
39. Tuller T, Kupiec M, Ruppin E (2007) Determinants of protein abundance and translation efficiency in *S. cerevisiae*. *PLoS Comput Biol* 3: e248.
40. Lange C, Zaigler A, Hammelmann M, Twellmeyer J, Raddatz G, et al. (2007) Genome-wide analysis of growth phase-dependent translational and transcriptional regulation in halophilic archaea. *BMC Genomics* 8: 415.
41. Wada A (1998) Growth phase coupled modulation of *Escherichia coli* ribosomes. *Genes Cells* 3: 203-208.
42. Vogl T, Brengelmann R, Hinz HJ, Scharf M, Lotzbeyer M, et al. (1995) Mechanism of protein stabilization by disulfide bridges: calorimetric unfolding studies on disulfide-deficient mutants of the alpha-amylase inhibitor tendamistat. *J Mol Biol* 254: 481-496.
43. Russell JB, Cook GM (1995) Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiol Rev* 59: 48-62.

44. Ogura T, Wilkinson AJ (2001) AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells* 6: 575-597.
45. Redon E, Loubiere P, Cocaign-Bousquet M (2005) Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. *J Biol Chem* 280: 36380-36385.
46. Benjamini YH, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B*: 289-300.

**Chapitre IV : Identification des
déterminants des concentrations protéiques
de *Lactococcus lactis* par modèles de
covariance**

Principalement menées sur la levure modèle *Saccharomyces cerevisiae*, les études portant sur le recoupement quantitatif des données de transcriptome et de protéome concluent à des corrélations relativement modestes avec des coefficients de corrélation de Pearson qui oscillent généralement entre 0.45 et 0.75 (Feder & Walser, 2005; Hack, 2004). En ce qui concerne les bactéries, des coefficients plus faibles, compris entre 0.45 et 0.53, ont été rapportés pour *Desulfovibrio vulgaris* cultivé dans plusieurs conditions différentes (Nie *et al.*, 2006a). L'analyse des corrélations globales entre les niveaux de transcrits et de protéines de la bactérie lactique *Lactococcus lactis* n'a pas permis d'obtenir de meilleures corrélations. Les coefficients de corrélation de Pearson associés s'élèvent seulement à 0.19 dans des conditions d'adaptation progressive à la carence en isoleucine (culture discontinue, batch) et 0.24 dans le cas d'une culture continue en chémostat limité par l'isoleucine où les cellules sont adaptées. La faiblesse de ces corrélations entre ARN messagers et protéines suggère que d'autres paramètres que la quantité de transcrits et de protéines doivent entrer en jeu dans ce mécanisme d'adaptation.

Afin d'identifier les déterminants biologiques des niveaux de protéines, une approche de biologie intégrative de type « top-down » a été adoptée (figure 11). Des modèles de covariance, qui intègrent à la fois des données qualitatives et quantitatives, ont été construits à partir de l'ensemble des données expérimentales de protéome et de transcriptome. Plusieurs paramètres génétiques susceptibles d'avoir une influence, tels que la longueur du gène, son biais de codon (CAI), sa position sur le chromosome, le nombre de codons qu'il utilise ou encore les indices d'hydrophobicité et d'aromaticité qui le caractérisent ont été introduits dans ces modèles. La catégorie fonctionnelle des gènes a également été prise en compte, tout comme la dimension dynamique des cultures symbolisée par le taux de croissance, considéré comme un paramètre qualitatif. Les variables quantitatives sont hiérarchisées en fonctions de la valeur absolue des coefficients qui y sont associés tandis que la comparaison des données qualitatives permet de comparer les niveaux de protéines des différents sous-groupes. Les modèles permettant de faire le meilleur compromis entre la qualité d'ajustement (mesurée par le R^2) et le nombre de paramètres inclus sont sélectionnés sur le critère d'information d'Akaike (AIC, (Burnham & Anderson, 2002)).

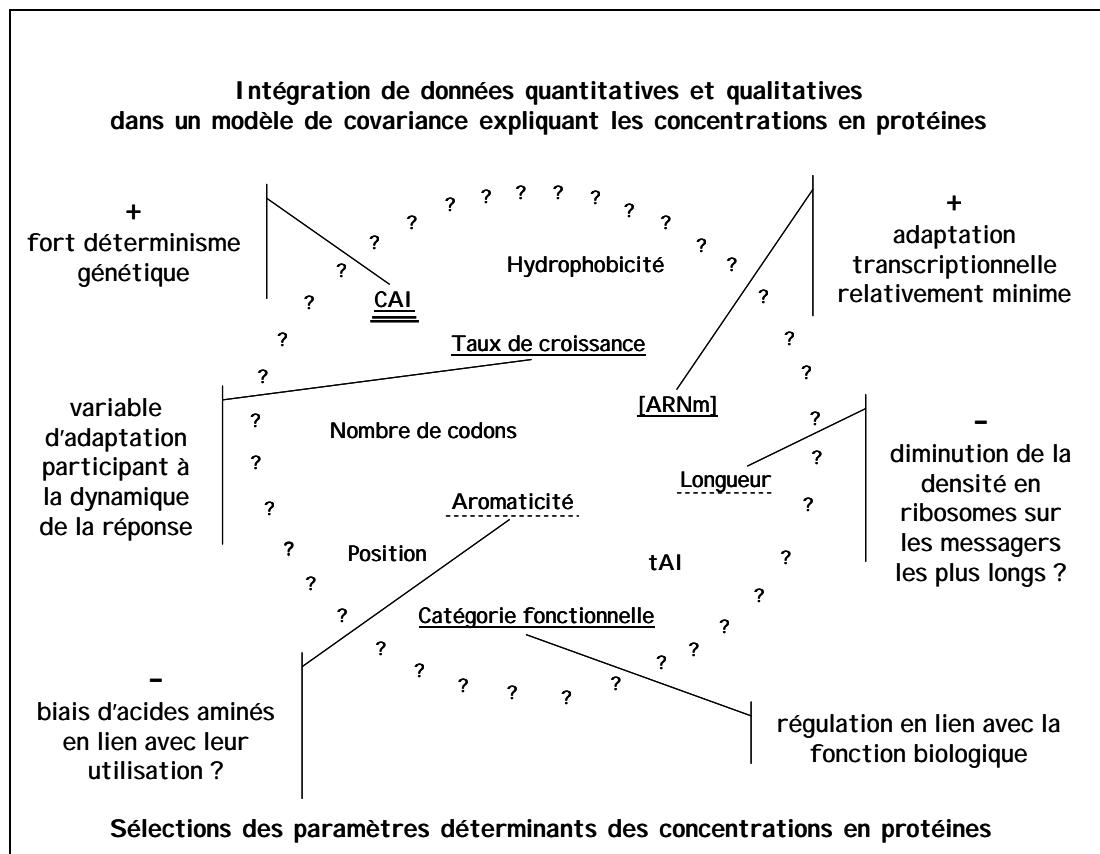


Figure 11 : Identification des déterminants de la traduction.

Les paramètres ayant une influence positive sur les concentrations en protéines sont signalés par des + et, inversement, les – représentent un effet négatif du paramètre. L’influence des paramètres quantitatifs (taux de croissance et catégorie fonctionnelle dépend du niveau du paramètre).

Dans le modèle d’adaptation dynamique (carence progressive en isoleucine en batch) comme dans le modèle adapté (chémostat au différents taux de croissance), le CAI a été sélectionné comme le paramètre le plus influent avec un effet sur la concentration en protéine plus important que celui des ARN messagers, suggérant un contrôle génétique plus influent que l’adaptation métabolique. L’aromaticité et la longueur des gènes sont à l’inverse des paramètres plutôt associés aux protéines les moins concentrées. La corrélation négative entre la longueur des gènes et les niveaux de protéines est probablement liée à la diminution de la densité de ribosomes sur les messagers les plus longs comme cela a pu être observé chez la levure (Arava *et al.*, 2003). L’effet négatif de l’aromaticité suggère quant à lui un biais d’utilisation des acides aminés vraisemblablement lié au tryptophane, un acide aminé rarement utilisé dans les protéines de *L. lactis* et donc peut être moins bien reconnu que les autres. La sélection du paramètre qualitatif catégorie fonctionnelle dans les deux modèles révèle un lien entre la fonction des protéines et la régulation de leur niveau d’expression. Les

protéines associées à la traduction et au métabolisme énergétique sont par exemple parmi les plus concentrées comme cela a pu être observé chez la bactérie *Escherichia coli* (Ishihama *et al.*, 2008).

Au delà de ces similarités, les deux modèles présentent des différences liées aux deux modes de culture analysés et correspondant à deux types d'adaptation différents. En premier lieu, le poids du CAI est bien plus important en batch qu'en chémostat, laissant supposer une grande influence génétique dans la dynamique d'adaptation. Les protéines de la catégorie régulation, dont la majorité sont des régulateurs transcriptionnels, ont tendance à être plus concentrées en batch et, *a contrario*, moins concentrées en chémostat. Cette dissemblance ne se répercute pas significativement sur les concentrations en ARN messagers qui, eux, conservent la même influence sur les niveaux de protéines dans les deux modèles. La dissimilitude des concentrations en protéines régulatrices entre les deux modes de culture doit donc provenir de l'activation transitoire des régulateurs qui reste visualisable lors de l'adaptation dynamique en batch mais pas dans des états adaptés provenant des chémostats. La différence majeure entre les deux modèles reste néanmoins la sélection du paramètre taux de croissance exclusivement en batch. Ce paramètre peut donc être considéré comme une variable d'adaptation participant à la régulation de la réponse transitoire.

L'intégration de nouveaux paramètres dans les algorithmes de sélection de ces modèles protéiques devrait dans l'avenir accroître encore la qualité des modèles de covariances sélectionnés. Néanmoins, s'ils ne permettent pas encore d'obtenir des résultats prédictifs, ces modèles sont importants d'un point de vue cognitif et ont permis d'identifier des déterminants majeurs des niveaux de protéines. En outre, l'approche développée dans cette étude est totalement transposable à des microorganismes autres que *L. lactis* tout comme à d'autres biomolécules.

Version à soumettre à Molecular Systems Biology.

**Linear covariance models to examine the determinants of protein levels in
*Lactococcus lactis***

**Clémentine Dressaire^{1,2}, Béatrice Laurent³, Pascal Loubière^{1,2}, Philippe Besse³, Muriel
Cocaign-Bousquet^{1,2*}**

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

(2) UMR5504, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, CNRS, INRA, 31400 Toulouse, France

(3) Institut de Mathématiques de Toulouse, UMR 5219, INSA de Toulouse, Université de Toulouse, 135 Avenue de Rangueil, F-31077 Toulouse, France

*correspondent foot-note: e-mail: cocaign@insa-toulouse.fr

ISBP, INSA 135 Avenue de Rangueil,

31077 Toulouse cedex 4, France.

Tel: (33) 561 559 438

Fax: (33) 561 559 400

ABSTRACT

Microorganisms simultaneously modify their protein levels and growth rate to adapt to new environments but regulations are not completely understood. This study aims at analysing, by statistical modelling, the determinants of protein levels in a model bacterium *Lactococcus lactis*, and at assessing whether they can change dynamically when the cellular growth is modified. For this purpose, two different adaptation processes, permanent and transient adaptations, were compared.

Protein levels were predicted by covariance models taking into account various quantitative and qualitative parameters. Best models were selected thanks to Akaike Information Criterion and allowed unexpected protein determinants to be identified. We found that Codon Adaptative Index was the most influent parameter (even more than transcript levels) and that length and aromaticity had negative influences on protein estimations. Protein determinants were for the first time demonstrated to be dependent on the adaptation process. Some functional categories had different profiles in both conditions and other parameters (dynamic parameter, gene location) were specifically selected during transient adaptation.

This statistical approach is adaptable to other microorganisms or even other biological variables.

INTRODUCTION

Deciphering biological systems as a whole is the huge challenge of systems biology (Joyce & Palsson, 2006). The increasing availability of proteomic data sheds more and more light on organism physiology and enables biological process elucidation at the functional level (Cordwell, 2004; Hecker *et al.*, 2008; Volker & Hecker, 2005). However, the cellular regulation of protein level itself is not yet completely understood. Protein level is primarily believed to depend on transcription and by extension on mRNA level. Only few comparisons of protein and transcript levels are available at the genome scale but the results typically reveal modest correlations between those two data sets (Feder & Walser, 2005; Hack, 2004; Nie *et al.*, 2006a). For the model yeast *Saccharomyces cerevisiae*, on which most of proteome-transcriptome comparisons were performed, Pearson correlation coefficient (R_p) ranged from 0.47 to 0.76 (Hack, 2004). The literature is even poorer when looking for mRNA-protein correlation in bacteria but a recent study estimated three determination coefficients R^2 of 0.28, 0.20 and 0.25 (respectively corresponding to R_p of 0.53, 0.45 and 0.50) for *Desulfovibrio vulgaris* grown under three different conditions (Nie *et al.*, 2006a). The best correlation coefficients ever obtained were reported with the APEX shotgun proteomics technology and interestingly suggested a lower correlation between mRNA and protein levels in the bacterium *E. coli* ($R^2=0.47$) than in yeast ($R^2=0.73$) (Lu *et al.*, 2007).

Altogether those modest correlations indicate that parameters other than transcript concentrations are involved in the regulation of protein levels. Gene sequence related features such as codon adaptation index (CAI) (Brockmann *et al.*, 2007), amino acid frequencies (Friberg *et al.*, 2004) and gene length (Beyer *et al.*, 2004) have for instance been correlated to *S. cerevisiae* proteome. CAI was also linked to protein expression in *E. coli* (Ishihama *et al.*, 2008), and was even found to correlate more significantly than Shine-Dalgarno sequence and stop codon identity (Lithwick & Margalit, 2003), other features that can influence translation

process. Interestingly, correlations were also found to increase within smaller categories related to cellular function (Nie *et al.*, 2006a) or localization (Greenbaum *et al.*, 2003). Whichever their nature, the influence of those various properties on protein levels have been almost systematically studied through basic linear correlation analyses. However such an approach can lead to misinterpretation since biological parameters are not necessarily independent as suggested by biological complexity (Williamson, 2005). The simultaneous analysis of the various protein level determinants together in an integrative model can overcome those overlapping effects. An approach based on multiple linear regressions has been developed by Zhang and co-workers, firstly on the bacterium *D. vulgaris* (Nie *et al.*, 2006a; Nie *et al.*, 2006b) and more recently on *S. cerevisiae* (Wu *et al.*, 2008). Though informative, this approach was incomplete since influences of parameters were determined from partial correlations and the various determinants contribution could not strictly be ordered.

The present study aims at analysing protein determinant in a firmicute model, *Lactococcus lactis*. A global and quantitative model taking into account all the potential protein determinants and allowing their hierarchical classification has been constructed. An integrated statistical approach based on covariance model regression establishment with an algorithm for parameter selection was developed. Protein levels were modelled with normalized data and without any *a priori* selection of the involved parameters. The adaptation of microorganisms to new environments involves changes of protein levels and growth rate simultaneously. For the first time, protein determinants were analysed dynamically through an adaptation process. Two different adaptation modes were investigated: one permanent adaptation corresponding to adapted cells in chemostat cultures at different growth rates; and a transient adaptation in batch cultures where the environment and thus the cellular growth rate are permanently changing. Application of this statistical approach enabled the parameters that have the greatest

influence on protein levels to be determined. Model dissimilarities between the two adaptation processes strongly suggested the involvement of different regulation modes.

With the growing accumulation of omics data this statistical method sounds as a valuable tool to understand the biological networks and their regulations. This approach was applied to study the translation process but can be extended to other metabolic nodes.

RESULTS

Description and specificities of the proteomic data

Proteomic data analysed in this paper were produced during the dynamic adaptation of *L. lactis* to isoleucine limitation in two different kinds of cultures. *L. lactis* proteome was first performed in discontinuous (batch) culture where isoleucine starvation is gradually imposed thanks to its natural consumption during cell growth (Dressaire *et al.* 2009a, submitted). Four time-points (exponential growth phase and after 20 min, 1.34 h and 3 h of amino acid starvation) corresponding to four different growth rates (respectively 0.88, 0.33, 0.06 and 0.04 h⁻¹) were available. This dynamic proteomic data set corresponds to a transient adaptation process since environment and consequently bacterial physiology and growth rate are changing during the culture. The second set of proteomic data, although performed in a similar isoleucine limited medium, was produced in chemostat cultures at three different growth rates (0.09, 0.24 and 0.47 h⁻¹) (Dressaire *et al.* 2009b, submitted). This second data set also integrates a dynamic dimension with growth rate modifications but corresponds to permanent adaptation to isoleucine limitation since cells are grown at steady state and are fully adapted.

The reliability of proteomic data was first estimated through the statistical exploration of the three independent biological repetitions of the two cultures. A total of seven samples in triplicates with approximately 250 proteins in each has been evaluated. Box-plots highlighted

the relative stability in the median with a constant variability in each repetition and histograms of log-transformed protein levels showed Gaussian distributions (graphs not shown). Scatter-plot indicated a good reproducibility in both continuous and discontinuous cultures although more dispersion was observed for batch cultivation probably due to the transient instability (figure 1).

Parameter selection for modeling

Like proteome, transcriptome was also available in both discontinuous and continuous fermentation conditions (Dressaire *et al.* 2009a, submitted, and Dressaire *et al.*, 2008). Although there is only a single biological step between transcripts and proteins (the translation), their levels were only weakly correlated together (figure 2). Pearson correlation coefficients were respectively of 0.190 and 0.235 for discontinuous and continuous cultivation (corresponding to R^2 of 0.04 and 0.06 respectively). Experimental noise in the data cannot explain alone this low correlation as shown by the good reproducibility of transcriptomic data in both cultures (figure 3). This observation is however consistent with the literature since only modest correlations were previously found in bacteria (R^2 lower than 0.3 for *D. vulgaris*, (Nie *et al.*, 2006a)). Hence protein levels can clearly not be inferred from transcript levels only.

In order to provide better estimations of protein levels, several parameters other than transcript levels were considered: features linked to codon bias (CAI, tAI, number of codon), amino-acid bias (aromaticity, hydrophobicity) and gene specificities (length, chromosomal position). The growth rate was also considered as a proper parameter to symbolise the dynamic dimension of both transient and permanent adaptations; it was considered as qualitative parameter due to its low dimension compared to proteomic data sets (3 or 4 depending on the culture mode compared to an average of 250 data). Moreover, in accordance with literature results (Greenbaum *et al.*, 2003; Nie *et al.*, 2006a), protein-messenger

correlations were found to depend on gene functions (table I). Correlations were indeed significantly better for the categories “transport and binding proteins”, “other” and “energy metabolism” independently of the culture mode and for “cell envelope” and “regulatory functions” in one culture exclusively. One can notice extremely opposite correlation results between the two cultures for the category “cellular process”. Considering these preliminary results, functional category was thus also considered as a potential interesting qualitative parameter.

Since parameters are not necessarily independent, correlations between the various quantitative parameters were first analysed. Spearman rank coefficients (R_S), which do not require any assumption on the distribution of the parameters, were used (figure 4). The highest correlation was found between the codon adaptation index (CAI) and the number of codon ($R_S = -0.827$). The link between those two parameters is not surprising since they both measure synonymous codon usage bias (Sharp & Li, 1987; Wright, 1990). They are nevertheless not completely similar because CAI is estimated from the sequences of the most abundant proteins (Sharp & Li, 1987) while the codon number is exclusively dependent on gene sequence (Wright, 1990). tRNA adaptation index (tAI), which similarly reflects codon bias but additionally takes into account the gene copy number of each tRNA (Brockmann *et al.*, 2007), is also well correlated both to CAI and number of codon ($R_S = 0.594$ and -0.433 respectively). As previously reported (dos Reis *et al.*, 2003; Fuglsang, 2004), mRNA levels correlated, though weakly, with CAI ($R_S = 0.115$), whereas the link with hydrophobicity ($R_S = -0.102$ with GRAVY) was rather unexpected.

Model establishment

In order to identify the major protein determinants, a linear analysis of covariance model of the protein levels was established from the various quantitative and qualitative selected parameters. The more the parameters included in the model, the higher the determination

coefficient R^2 will be, but also the more complex the model will become. As a consequence, coefficient estimations will be less precise, with higher variances. In order to select the best model making a compromise between fitting quality and complexity, the Akaike model selection criterion (AIC for Akaike Information Criterion) was used. This model selection process is derived from the classical least square minimization method implemented to estimate model coefficients but with an additional penalisation term which is increasing with the complexity of the model (Burnham & Anderson, 2002). It enables the most significant parameters to be selected without any *a priori* subjective selection that could bias the results. Parameters influencing protein levels may have multiplicative rather than additive effects. The log-transformed protein concentration was thus modelled as a function of log-transformed quantitative values so as to obtain a linear equation for the covariance model (see material and methods for details). Furthermore, every quantitative parameter introduced in the model was previously centred and reduced. This normalisation was applied in order to level all the features and to enable the comparison of the various model coefficients. It allows influences on protein concentration of all quantitative parameters of the model to be hierarchically classified. Different in their nature, estimated coefficients for qualitative parameters cannot be compared to those for quantitative parameters. Coefficients related to quantitative parameters indeed influence the slope of the model whereas those linked to qualitative parameters modify the intercept (see equations in material and methods). Estimated coefficients of qualitative parameters can however be compared to each others. The influence of quantitative parameters can thus be estimated by the ranking of the absolute values of associated coefficient. Coefficients linked to qualitative parameters have to be interpreted differently. A positive value of the coefficient means higher concentration of the corresponding proteins compared with the overall average protein level. A negative value is at inversely associated to lower protein levels.

Model results

The AIC model selection has been applied to the two data sets corresponding to the two cultivation modes. Residual normality and homoscedasticity were graphically controlled (see figure 5 for normality). The selected global models, the parameters they include and the associated coefficients are summarized in tables II and III. Drastic increases in correlation coefficients were obtained compared to simple linear regressions between protein and transcript levels: the determination coefficient R^2 respectively rose from 0.03 and 0.06 for discontinuous and continuous cultivation respectively up to 0.58 and 0.51.

In both conditions, quantitative and qualitative parameters are selected in the models justifying the use of analysis of covariance models rather than multiple linear regressions. CAI, [mARN], aromaticity, length and functional category were selected in both batch and chemostat models. More particularly, CAI clearly appeared as the most influent parameter. CAI and [mRNA] have positive influence on protein concentrations whereas aromaticity and length tend to decrease protein levels. The number of codon was also included in both models but was less influent (p-value not significant) in the discontinuous model.

Some important differences were however encountered between the two models. Although the parameter “functional category” was selected in the two models, the nature and relative influence of involved categories differed. Proteins linked to “translation”, “energy metabolism” and “amino-acids biosynthesis” seemed to be more concentrated than others whatever the cultivation mode. However, proteins classified into “transcription” and “purines/pyrimidines metabolism” were significantly less concentrated than others exclusively in the continuous fermentation, and protein related to “central intermediary metabolism” displayed significant lower levels only in discontinuous cultivation. In addition, proteins involved in “regulation” exhibited higher concentrations than the mean in batch but lower in chemostat. Furthermore, the two parameters, “gene position” and “growth rate”,

were selected in batch cultures only. The specific growth rate influence in batch fermentation seems to indicate that models are changing at the various growth rates in batch but significantly less in chemostat cultures. Partial models obtained at each growth rates in both conditions were compared (Table IV), suggesting messenger concentration is probably the main factor involved in this variability. The highest change in the estimated coefficient from 0.23 to 0.44 was indeed observed for mRNA concentration between 0.06 and 0.88 h⁻¹ in batch conditions.

Model improvement

Selected models in both cultures still remain not precise enough to be predictive with their R² lower than 0.6. In order to improve the correlation between predicted and measured protein concentrations, models allowing interactions between two components were computed (see material and methods for equation). Using the same AIC selection method increased the R² values to 0.76 and 0.74 for batch and chemostat data sets respectively. Those models were however extremely complicated, each including more than twenty different parameters. Moreover, some associated coefficients were not estimated by the models because identifiability conditions were not fulfilled. It is thus difficult to give a biological interpretation of those models. One can however notice that the qualitative parameter “functional category” seemed to interact with many of the quantitative parameters (CAI, position, aromaticity, number of codon, tAI and length) suggesting that functional roles rely on genetic bases.

In order to avoid such over-fitting models we have applied a more stringent statistical selection criterion, the Bayesian information criterion (BIC), which penalizes more the least square criterion than AIC (see material and methods for details). The selected models revealed major interaction between CAI and the number of codons but did not strongly

improve the fitting quality (R^2 of 0.64 in batch and 0.55 in chemostat). Moreover this information about parameter interaction was previously deduced from the colinearity between those two parameters (see above; figure 4).

DISCUSSION

In this paper, weak correlations between proteomic and transcriptomic data were observed in *L. lactis*, independently of its culture mode. Although protein and transcript levels are expected to be strictly balanced in continuous cultures because of the steady state while not in discontinuous cultures, the improvement in correlation was hardly significant ($R_p = 0.190$ in batch and 0.235 in chemostat). The weakness of the relationship between transcriptome and proteome data is consistent with the observation that genes with similar messenger levels can have very different protein levels (Gygi *et al.*, 1999). Correlations found in *L. lactis* were even lower than the modest correlations previously described in the literature (Nie *et al.*, 2007). Such comparisons between studies should however be done with caution because of the multiple differences in strain, cultivation, number of compared protein/mRNA couples and more particularly measurement techniques and data processing. Nevertheless, correlations in *L. lactis*, as well as those previously found for another bacterium *D. vulgaris* (R_p 0.45 to 0.50; (Nie *et al.*, 2006a)), seem lower than in the yeast *S. cerevisiae* (R_p 0.47 to 0.76; (Hack, 2004)). On the contrary, improvement of the correlation was expected in bacteria since the absence of compartmentalization may enhance the coupling between transcription and translation processes. The study of Lu *et al.* who found R^2 of 0.47 and 0.73 respectively for *E. coli* and *S. cerevisiae* using the same proteomic technique (Lu *et al.*, 2007) also corroborates this surprising observation. Interestingly, in this last work, correlations between proteome and transcriptome were higher than usually but this may be at least partially due to corrections introduced for protein quantification, notably by amino acid composition, here found to be determinant for protein estimation (see aromaticity factor).

mRNA concentrations only weakly reflect protein levels but the modelling approach presented here allowed additional protein determinants to be identified. The importance of the sequence-related feature CAI for protein level prediction was previously found for two bacteria (*E. coli* and *Haemophilus influenzae*) (Lithwick & Margalit, 2003). In *L. lactis*, CAI was proved to be even more important in protein level determination than messenger concentrations. This higher influence of a parameter connected to the genome sequence (CAI) compared to an adaptation parameter (mRNA) suggests the control by genetic determinism is stronger than by metabolic adaptation. The CAI was also more informative than the tAI found to have a wide influence on translation in the yeast (Brockmann *et al.*, 2007). Gene length was found to be negatively correlated to protein concentrations. This is biologically consistent with the decrease of the ribosome density observed on long mRNA by polysomal analyses in *S. cerevisiae* (Arava *et al.*, 2003). Moreover, long proteins may be less abundant in cells due to their high degradation rate since protein half-lives were negatively correlated to their length (Belle *et al.*, 2006). Though aromatic amino acids (tyrosine, phenylalanine and tryptophan) are all available in the fermentation medium, protein levels in *L. lactis* were negatively influenced by aromaticity. This clearly reveals amino acids bias in *L. lactis* which is corroborated by the particular low occurrence of tryptophan in proteins of this microorganism (0.7 % of proteins). This relation between aromaticity and protein level was not observed in *S. cerevisiae* (Tuller *et al.*, 2007) but, considering the phylogenetic distance between the two microorganisms, amino acid bias may be completely different.

The selection of the parameter “functional category” in the models indicates that translation and/or post-translational regulations are dependent of protein functions. Proteins related to “translation”, “energy metabolism” and “amino-acid biosynthesis” were more concentrated than others in both models. Ribosomal proteins, as well as enzymes linked to energy metabolism, were similarly found to be the most abundant in *E. coli* cytosol (Ishihama *et al.*,

2008). This general trait for “translation” and “energy metabolism” could be related to their function essentiality for bacterial adaption and survey. Here the higher levels of enzymes involved in amino-acid biosynthesis can be explained by their wide solicitation since the growth of *L. lactis* was limited in both cultures by an amino acid (isoleucine). The analysis of simple correlations between transcripts and proteins in the various functional categories did not allow such results to be underlined. The approach based on correlations with a single parameter can thus drive erroneous biological conclusions since parameters are not necessarily independent. This result highlights the interest of our approach based on a global analysis of covariance model.

Although similar in terms of media and amino acid limitation, there is a major biological difference between the discontinuous and the continuous fermentations: the first culture corresponds to transient evolution and the last to adapted steady states. This difference is reflected by the profound dissimilarity between both models. Protein determinants are indeed dependent on physiological state of the cells. CAI regulation is more important in batch than in chemostat cultures (Tables I and II), which could reveal a higher genetic influence when bacteria have to continuously revise their cellular control in an evolving environment. As a matter of fact, one can imagine that in discontinuous cultivation, cells do not have time to establish tight control networks and thus protein expression is mostly guided by natural selection. This hypothesis should however be considered with caution since the selection of both the number of codon and tAI in the continuous model should potentially increase the genetic influence in these conditions. The better correlation of CAI with protein concentration in batch compared to chemostat cultivations nevertheless explains the higher global R^2 of the batch model. Functional categories are neither regulated similarly in the two processes and the most striking example concerns the category “regulation”. Regulatory proteins are generally available at low quantities in cells as observed in chemostat cultures with the negative model

coefficient. The observation that regulatory proteins are more concentrated in batch cultures is more difficult to understand but may be due to transient up-regulations. Proteins ranked in “regulation” functional category are mainly transcriptional regulators and may have a deep impact on mRNA concentrations in cells. Coefficients associated to mRNA influence on protein levels are nevertheless identical in both cultures. Similar transient regulations may thus have occurred in chemostat but are not visualized since steady states are reached. Lastly, the most important difference between the two models is the selection of the growth rate as a discriminatory parameter in batch but not in chemostat cultures. This parameter, corresponding to the dynamic dimension, is tightly linked to cultivation conditions and can be considered as an adaptation feature. Its absence from the chemostat model thus suggests that in steady states, protein levels are regulated in a single way, independently of growth rate. On the contrary, in batch culture growth rate parameter participates to the transient adaptation and shows that the adaptation to isoleucine starvation is progressive.

The models presented in this paper allowed determinants of protein levels to be identified and ordered in the various studied conditions. Initial parameters were normalized in order to enable the comparison of the different estimated coefficients in the models. These models are however incomplete and not yet predictive since global R^2 range between 0.5 and 0.6. Other integrative models were previously constructed for *D. vulgaris* (Nie *et al.*, 2006a; Nie *et al.*, 2006b) or for *S. cerevisiae* (Tuller *et al.*, 2007) and gave similar or lower R^2 values. Differences between those models and the one for *L. lactis* lie in the parameters that were integrated. For the first time here, the dynamic aspect of the cultures reflecting the cellular adaptation was included through growth rate integration. Our model is also the single one allowing statistical selection of the parameters through the application of Akaike criterion without any initial assumptions. With such selection criterion only the most important

parameters are selected without any bias whereas other models automatically include all parameters that were *a priori* introduced. Now in order to improve our models and to further investigate protein concentration regulation, additional parameters have to be included, such as protein stability or ribosome occupancy data. Such biological data are currently available for yeast but not for any bacteria. AIC model selection method presented in this study is not limited by the number of parameters. Such a model selection program is not restricted to the analysis of protein determinants but can be extended to other biological variables (mRNA or metabolite). Furthermore, this approach has been developed for *L. lactis* but is fully adaptable to any other microorganism.

MATERIALS AND METHODS

Organism and growth conditions

Lactococcus lactis ssp. *lactis* IL1403, whose genome has been entirely sequenced (Bolotin *et al.*, 2001), was grown both continuously and discontinuously as previously described (respectively Dressaire *et al.*, 2008, and Dressaire *et al.* 2009a, submitted). For continuous cultivations, three different growth rates have been studied, namely 0.09, 0.24 and 0.47 h⁻¹ thanks to anaerobic chemostat cultures (under nitrogen atmosphere and regulated pH) on a chemically defined medium. Experiments were designed so as isoleucine to be the limiting nutrient. Batch cultivations were conducted in a similar medium but with lower glucose and bases initial concentrations (10 and 0.01 g/L respectively compared to 20 and 0.05 g/L in chemostat). Gradual stress was imposed through isoleucine natural consumption and samples for transcriptomic and proteomic analyses were harvested during exponential phase and after 20 min, 1.34 h and 3 h after isoleucine exhaustion.

Transcriptomic and proteomic data

Transcriptomic data previously produced in our group were available for growth rates (μ) of 0.09, 0.24 and 0.47 h⁻¹ in chemostat (Dressaire *et al.*, 2008) and exponential phase, 20 min, 1.34 h and 3 h after starvation (respectively corresponding to $\mu = 0.88, 0.33, 0.06$ and 0.04 h⁻¹) in batch cultivations (Dressaire *et al.* 2009a, submitted). The raw data available online on GEO database (<http://www.ncbi.nlm.nih.gov/geo/>, GSE10256 and GSE12962 for chemostat and batch respectively) were reprocessed as previously described (Dressaire *et al.*, 2008) in order to calculate mRNA concentrations rather than abundances. Briefly, raw data were standardized by the all spots' mean intensity of the reference membrane ($\mu = 0.09$ h⁻¹ for continuous cultures and exponential growth phase for batch cultures) and corrected by total RNA concentration in each growth condition (in g/100 g dried cells : 3.58 ± 0.39, 4.92 ± 0.52, 7.34 ± 0.28 for $\mu = 0.09, 0.24, 0.47$ h⁻¹ respectively in chemostat, and 11.06± 0.23, 8.75 ± 1.53, 6.95 ± 0.77, 6.95 ± 0.37 for exponential phase, 20 min, 1.34 h and 3 h after starvation in batch respectively).

Protein concentrations rigorously measured on the same experiments were obtained by multiple 2D proteomic analysis (3 pH range by samples) and were already available (Dressaire *et al.* 2009a and b, submitted).

Genes features

Gene's sequences and chromosomal position were retrieved from EntrezGenome database (<http://www.ncbi.nlm.nih.gov/sites/entrez?Db=genome&Cmd>ShowDetailView&TermToSearch=171>). Since *L. lactis* chromosome is circular, genes positions were converted into distances from replication origin: positions values associated to genes in the first half of the chromosome were conserved, whereas the position value of genes located in the second half of the chromosome were subtracted to the total chromosome length. Genes length, codon adaptation index (CAI), number of codon, hydrophobicity (GRAVY score) and aromaticity

were calculated using codonW on line software (<http://mobylye.pasteur.fr/cgi-bin/MobylyePortal/portal.py?form=codonw>). Genes functional category were used as initially defined by Bolotin *et al* (Bolotin *et al.*, 2001). tRNA adaptation index (tAI) were calculated using the definition given by Dos Reis *et al.* (dos Reis *et al.*, 2004; dos Reis *et al.*, 2003). All those parameters were log transformed except for the GRAVY score because it can take negative values.

Statistical methods

Analysis of covariance models can be written as follows:

$$\ln[prot_i] = \alpha + \sum_j \beta_j \ln(var_{i,j}) + \sum_k \lambda_{k(i)} + \varepsilon_i$$

where $[prot_i]$ is the measured level of the i^{th} protein, α is the intercept, $var_{i,j}$ the value of the j^{th} quantitative parameter for the i^{th} protein, β_j and $\lambda_{k(i)}$ the coefficients associated to the j^{th} quantitative parameter and the k^{th} qualitative parameter of the i^{th} protein respectively and ε_i the error term for the i^{th} protein. Identifiability conditions ensure the unicity of the coefficients in the above equation: we assume that for each qualitative parameter, the sum of the coefficients $\lambda_{k(i)}$ equals zero, the sum being taken over the set of all possible values for this qualitative parameter.

Models allowing dual interactions between parameters have a more complex mathematical formulation:

$$\ln[prot_i] = \alpha + \sum_j \beta_j \ln(var_{i,j}) + \sum_k \lambda_{k(i)} + \sum_l \sum_m \gamma_{l,m} (\ln(var_{i,l}) * \ln(var_{i,m})) + \sum_n \sum_p (\delta_{n(i),p} \ln(var_{i,p})) + \sum_q \sum_t \delta_{q(i),t(i)} + \varepsilon_i$$

$$\sum_l \sum_m \gamma_{l,m} (\ln(var_{i,l}) * \ln(var_{i,m})), \quad \sum_q \sum_t \delta_{q(i),t(i)} \quad \text{and} \quad \sum_n \sum_p (\delta_{n(i),p} \ln(var_{i,p})) \quad \text{respectively}$$

represent the interaction terms between two quantitative parameters, two qualitative parameters and a quantitative and a qualitative parameters. As previously, we have to impose identifiability conditions.

For each model, we can define $[prot_i]_{mod}$ as the level of the i^{th} protein fitted by the model. For example, for the model without interaction

$$\ln[prot_i]_{mod} = \hat{\alpha} + \sum_j \hat{\beta}_j \ln(\text{var}_{i,j}) + \sum_k \hat{\lambda}_{k(i)}$$

where the coefficients have been estimated by a least square procedure in order to minimize the residual sum of squares: $RSS = \sum_i (\ln[prot_i]_{mod} - \ln[prot_i])^2$.

We have many models in competition: any quantitative or qualitative parameter can be present or not in the model, the same holds for the interaction terms. Akaike selection procedure (AIC) looks for the model that minimizes the following criterion:

$$n \log(RSS / n) + 2N$$

where n is the number of proteins that were used to construct the model and N denotes the number of coefficients estimated by the model (Burnham & Anderson, 2002).

Bayesian information criterion (BIC) similarly looks for the model that minimizes:

$$n \log(RSS / n) + \ln(n)N$$

Correlation, modelling and selections were performed using R free statistical software (<http://www.r-project.org/>). Correlations were either estimated by Spearman rank correlation coefficient (RS) based on rank of paired observations, either by Pearson correlation coefficient (RP) calculated from the measured values. The linear regression coefficient R^2 equals the square of Pearson correlation coefficient between the values $\ln[prot_i]$ and the fitted values $\ln[prot_i]_{mod}$.

AKNOWLEDGMENTS

This work was supported by the INRA Integrative Biology Program (agroBI).

REFERENCES

- Arava Y., Wang Y., Storey J. D., Liu C. L., Brown P. O., and Herschlag D. (2003). Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **100**: 3889-94.
- Belle A., Tanay A., Bitincka L., Shamir R., and O'Shea E. K. (2006). Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci U S A* **103**: 13004-9.
- Beyer A., Hollunder J., Nasheuer H. P., and Wilhelm T. (2004). Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. *Mol Cell Proteomics* **3**: 1083-92.
- Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K., Weissenbach J., Ehrlich S. D., and Sorokin A. (2001). The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res* **11**: 731-53.
- Brockmann R., Beyer A., Heinisch J. J., and Wilhelm T. (2007). Posttranscriptional expression regulation: what determines translation rates? *PLoS Comput Biol* **3**: e57.
- Burnham K. P., and Anderson D. R. (2002). "Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach," Springer-Verlag, New-York.
- Cordwell S. J. (2004). Exploring and exploiting bacterial proteomes. *Methods Mol Biol* **266**: 115-35.
- dos Reis M., Savva R., and Wernisch L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res* **32**: 5036-44.
- dos Reis M., Wernisch L., and Savva R. (2003). Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res* **31**: 6976-85.
- Dressaire C., Redon E., Milhem H., Besse P., Loubiere P., and Coccagn-Bousquet M. (2008). Growth rate regulated genes and their wide involvement in the *Lactococcus lactis* stress responses. *BMC Genomics* **9**: 343.
- Feder M. E., and Walser J. C. (2005). The biological limitations of transcriptomics in elucidating stress and stress responses. *J Evol Biol* **18**: 901-10.
- Friberg M., von Rohr P., and Gonnet G. (2004). Limitations of codon adaptation index and other coding DNA-based features for prediction of protein expression in *Saccharomyces cerevisiae*. *Yeast* **21**: 1083-93.

- Fuglsang A. (2004). Bioinformatic analysis of the link between gene composition and expressivity in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Antonie Van Leeuwenhoek* **86**: 135-47.
- Greenbaum D., Colangelo C., Williams K., and Gerstein M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* **4**: 117.
- Gygi S. P., Rochon Y., Franza B. R., and Aebersold R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**: 1720-30.
- Hack C. J. (2004). Integrated transcriptome and proteome data: the challenges ahead. *Brief Funct Genomic Proteomic* **3**: 212-9.
- Hecker M., Antelmann H., Buttner K., and Bernhardt J. (2008). Gel-based proteomics of Gram-positive bacteria: a powerful tool to address physiological questions. *Proteomics* **8**: 4958-75.
- Ishihama Y., Schmidt T., Rappsilber J., Mann M., Hartl F. U., Kerner M. J., and Frishman D. (2008). Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* **9**: 102.
- Joyce A. R., and Palsson B. O. (2006). The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* **7**: 198-210.
- Lithwick G., and Margalit H. (2003). Hierarchy of sequence-dependent features associated with prokaryotic translation. *Genome Res* **13**: 2665-73.
- Lu P., Vogel C., Wang R., Yao X., and Marcotte E. M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25**: 117-24.
- Nie L., Wu G., Culley D. E., Scholten J. C., and Zhang W. (2007). Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. *Crit Rev Biotechnol* **27**: 63-75.
- Nie L., Wu G., and Zhang W. (2006a). Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: a multiple regression to identify sources of variations. *Biochem Biophys Res Commun* **339**: 603-10.
- Nie L., Wu G., and Zhang W. (2006b). Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in *Desulfovibrio vulgaris*: a quantitative analysis. *Genetics* **174**: 2229-43.
- Sharp P. M., and Li W. H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**: 1281-95.

- Tuller T., Kupiec M., and Ruppin E. (2007). Determinants of protein abundance and translation efficiency in *S. cerevisiae*. *PLoS Comput Biol* **3**: e248.
- Volker U., and Hecker M. (2005). From genomics via proteomics to cellular physiology of the Gram-positive model organism *Bacillus subtilis*. *Cell Microbiol* **7**: 1077-85.
- Williamson M. P. (2005). Systems biology: will it work? *Biochem Soc Trans* **33**: 503-6.
- Wright F. (1990). The 'effective number of codons' used in a gene. *Gene* **87**: 23-9.
- Wu G., Nie L., and Zhang W. (2008). Integrative analyses of posttranscriptional regulation in the yeast *Saccharomyces cerevisiae* using transcriptomic and proteomic data. *Curr Microbiol* **57**: 18-22.

FIGURES LEGENDS

Figure 1. Scatter plot showing proteomic experiments reproducibility in batch (A) and chemostat (B) cultures of *L. lactis*.

Plots correspond to one sample in each type of culture (growth rates of 0.88 h^{-1} and 0.09 h^{-1} in batch and chemostat respectively) but are representative of the rest of the culture samples.

Figure 2. Plot of protein and messenger concentrations in batch (A.) and chemostat (B.) cultures of *L. lactis*.

Figure 3. Scatter plot showing transcriptomic experiments reproducibility in batch (A) and chemostat (B) cultures of *L. lactis*.

Plots correspond to one point in each type of culture (growth rates of 0.88 h^{-1} and 0.09 h^{-1} in batch and chemostat respectively) but are representative of the rest of the culture samples.

Figure 4. Matrix scatter plot (left hand half) and correlation matrix (right hand half) of the parameters introduced in the current analyse.

Numbers indicate Pearson correlation coefficient (R_p). Significant correlation ($p\text{-value} < 0.05$) are in bold letters. mRNA stands for transcript concentration, CAI for codon adaptation index, Pos for distance from replication origin, Lgth for gene length, tAI for tRNA adaptation index, GRAVY for hydrophobicity, Aro for aromaticity and Nc for number of codon.

Figure 5. Histogram of residuals from models obtained in batch (A.) and chemostat (B.) cultivation of *L. lactis*.

TABLES

Table I. Pearson correlation coefficients of proteomic and transcriptomic data computed with the whole data set or on specific modules in *L. lactis* grown in batch and chemostat cultures. Functional categories were used as initially defined by Bolotin *et al.* (Bolotin *et al.*, 2001). Significant correlations (p-value lower than 0.05) are indicated in bold letters. The column entitled “available couples” indicates the number of genes for which both protein and messenger levels were measured. This number is multiplied by the number of samples available (4 in batch and 3 in chemostat corresponding to the different growth rates).

| Functional category | batch | | chemostat | |
|---------------------|----------------------|-------------------|----------------------|-------------------|
| | R _{Pearson} | available couples | R _{Pearson} | available couples |
| AMI | -0.036 | 8*4 | 0.155 | 18*3 |
| CEL | -0.412 | 7*4 | 0.397 | 9*3 |
| COF | 0.228 | 5*4 | 0.093 | 8*3 |
| ENV | 0.640 | 5*4 | 0.282 | 9*3 |
| FAT | 0.032 | 7*4 | 0.188 | 11*3 |
| INT | -0.304 | 3*4 | -0.020 | 5*3 |
| NRJ | 0.287 | 35*4 | 0.293 | 56*3 |
| OTH | 0.779 | 6*4 | 0.425 | 9*3 |
| PUR | 0.000 | 19*4 | 0.206 | 26*3 |
| REG | -0.133 | 8*4 | 0.462 | 15*3 |
| REP | 0.014 | 6*4 | 0.111 | 14*3 |
| TRD | 0.121 | 50*4 | 0.140 | 58*3 |
| TRS | 0.438 | 5*4 | 0.005 | 4*3 |
| TSP | 0.612 | 7*4 | 0.469 | 12*3 |
| All | 0.190 | 192*4 | 0.235 | 308*3 |

AMI = Amino acids biosynthesis, CEL = cellular process, COF = biosynthesis of cofactors, ENV = cell envelope, FAT = fatty acids metabolism, INT = central intermediary metabolism, NRJ = energy metabolism, OTH = other categories, PUR = purines, pyrimidines, nucleosides and nucleotides, REG = regulatory functions, REP = replication, TRD = translation, TRS = transcription, TSP = transport and binding proteins, All = Whole data set.

Table II. Selected model for batch data.

In the first column, all parameters of the model selected by the AIC procedure are given. The coefficients were estimated by the minimization of the least square criterion. The third column provides an estimation of the standard deviation associated to the estimated coefficients. The nullity test was performed for each coefficient and p-values are listed in the last column. Selected coefficients associated to non significant p-value (greater than 0.05) are coloured in grey.

| Selected parameter | Estimated coefficient | Standard error | p-value |
|----------------------|-----------------------|----------------|-----------|
| intercept | -5.16 | 0.10 | <2.00E-16 |
| CAI | 0.66 | 0.08 | 5.10E-16 |
| [mRNA] | 0.28 | 0.03 | <2.00E-16 |
| Aromaticity | -0.25 | 0.04 | 1.81E-9 |
| Length | -0.13 | 0.04 | 2.50E-3 |
| Position | 0.11 | 0.04 | 4.75E-3 |
| Functional category | | | |
| AMI | 0.61 | 0.18 | 5.41E-4 |
| CEL | 0.15 | 0.19 | 4.23E-1 |
| COF | -0.24 | 0.22 | 2.74E-1 |
| ENV | 0.11 | 0.21 | 6.21E-1 |
| FAT | -0.03 | 0.18 | 8.60E-1 |
| INT | -1.30 | 0.27 | 2.33E-6 |
| NRJ | 0.33 | 0.10 | 6.92E-4 |
| OTH | 0.01 | 0.20 | 9.79E-1 |
| PUR | 0.00 | 0.12 | 9.73E-1 |
| REG | 0.38 | 0.17 | 2.81E-2 |
| REP | 0.564 | 0.20 | 1.50E-3 |
| TRD | 0.43 | 0.09 | 1.28E-6 |
| TRS | -0.25 | 0.21 | 2.43E-1 |
| TSP | -0.11 | 0.19 | 5.49E-1 |
| Growth rate | | | |
| 0.88 h ⁻¹ | -0.96 | 0.11 | <2.00E-16 |
| 0.33 h ⁻¹ | 0.37 | 0.07 | 3.54E-7 |
| 0.06 h ⁻¹ | 0.32 | 0.07 | 2.72E-6 |
| 0.04 h ⁻¹ | 0.26 | 0.07 | 6.06E-5 |
| Number of codon | -0.12 | 0.07 | 9.70E-2 |

AMI = Amino acids biosynthesis, CEL = cellular process, COF = biosynthesis of cofactors, ENV = cell envelope, FAT = fatty acids metabolism, INT = central intermediary metabolism, NRJ = energy metabolism, OTH = other categories, PUR = purines, pyrimidines, nucleosides and nucleotides, REG = regulatory functions, REP = replication, TRD = translation, TRS = transcription, TSP = transport and binding proteins.

Table III. Selected model for chemostat data.

In the first column, all parameters of the model selected by the AIC procedure are given. The coefficients were estimated by the minimization of the least square criterion. The third column provides an estimation of the standard deviation associated to the estimated coefficients. The nullity test was performed for each coefficient and p-values are listed in the last column. Selected coefficients associated to non significant p-value (greater than 0.05) are coloured in grey.

| Selected parameter | Estimated coefficient | Standard error | p-value |
|---------------------|-----------------------|----------------|-----------|
| intercept | -6.12 | 0.04 | <2.00E-16 |
| CAI | 0.38 | 0.07 | 7.09E-19 |
| [mRNA] | 0.27 | 0.03 | <2.00E-16 |
| Aromaticity | -0.24 | 0.03 | 4.87E-12 |
| Number of codon | -0.23 | 0.06 | 4.82E-5 |
| tAI | 0.14 | 0.04 | 3.68E-4 |
| Length | -0.13 | 0.03 | 1.87E-4 |
| Functional category | | | |
| AMI | 0.34 | 0.13 | 8.43E-3 |
| CEL | 0.45 | 0.18 | 1.13E-2 |
| COF | 0.05 | 0.19 | 7.94E-1 |
| ENV | -0.00 | 0.17 | 9.99E-1 |
| FAT | -0.24 | 0.15 | 9.98E-2 |
| INT | -0.19 | 0.23 | 4.12E-1 |
| NRJ | 0.23 | 0.08 | 4.22E-3 |
| OTH | 0.23 | 0.18 | 1.87E-1 |
| PUR | -0.39 | 0.11 | 3.33E-4 |
| REG | -0.55 | 0.14 | 7.36E-5 |
| REP | 0.14 | 0.14 | 3.22E-1 |
| TRD | 0.36 | 0.09 | 3.91E-5 |
| TRS | -0.58 | 0.26 | 2.47E-2 |
| TSP | 0.13 | 0.15 | 4.24E-1 |

AMI = Amino acids biosynthesis, CEL = cellular process, COF = biosynthesis of cofactors, ENV = cell envelope, FAT = fatty acids metabolism, INT = central intermediary metabolism, NRJ = energy metabolism, OTH = other categories, PUR = purines, pyrimidines, nucleosides and nucleotides, REG = regulatory functions, REP = replication, TRD = translation, TRS = transcription, TSP = transport and binding proteins.

Table IV. Main parameters of the models and their estimated coefficients in batch and chemostat cultures at each growth rate.

| growth rate | Batch | | | | Chemostat | | |
|---|----------------------|----------------------|----------------------|----------------------|---------------------------|---------------------------|---------------------------|
| | 0.88 h ⁻¹ | 0.33 h ⁻¹ | 0.06 h ⁻¹ | 0.04 h ⁻¹ | 0.09 h ⁻¹ | 0.24 h ⁻¹ | 0.47 h ⁻¹ |
| major parameters selected by the model (estimated coefficients) | CAI (0,89) | CAI (0,72) | CAI (0,72) | CAI (0,72) | CAI (0,33) mRNA (0,29) | CAI (0,36) mRNA (0,29) | CAI (0,43) mRNA (0,29) |
| | mRNA (0,44) | Aromaticity (-0,34) | Aromaticity (-0,27) | mRNA (0,25) | Number of codons (-0,25) | Number of codons (-0,30) | Number of codons (-0,26) |
| | Aromaticity (-0,26) | mRNA (0,30) | mRNA (0,23) | Aromaticity (-0,24) | Aromaticity (-0,24) | Aromaticity (-0,25) | Aromaticity (-0,24) |
| R ² | 0,61 | 0,59 | 0,59 | 0,50 | 0,49 | 0,48 | 0,48 |

Conclusion et perspectives

DEMARCHE ET PRINCIPAUX RESULTATS

Afin d'obtenir une vision globale et compréhensive de la physiologie cellulaire, une démarche de biologie intégrative a été mise en place. Cette approche repose sur l'intégration de données biologiques correspondant à différents niveaux d'analyse. Le choix du microorganisme d'étude s'est porté vers *Lactococcus lactis*, le modèle des bactéries lactiques. Cette bactérie présente un intérêt industriel en raison de son utilisation dans les procédés de fabrication de produits laitiers fermentés comme certains fromages. Son innocuité en fait aussi un vecteur de choix pour la délivrance de molécules thérapeutiques. En tant que procaryote, son organisation cellulaire n'est pas compartimentée. Son génome est entièrement séquencé et présente l'avantage d'être relativement petit (2.3 Mb) par rapport à ceux des autres microorganismes modèles que sont *Escherichia coli* (4.6 Mb), *Bacillus subtilis* (4.2 Mb) et *Saccharomyces cerevisiae* (12 Mb). *L. lactis* est, de fait, doté d'un métabolisme simple ce qui constitue un atout en biologie des systèmes.

Le travail présenté dans ce manuscrit s'est effectué dans des conditions caractérisées et contrôlées grâce à l'utilisation d'outils de fermentation maîtrisés. La culture discontinue en batch revêt un aspect dynamique en lien avec l'évolution conjointe des différents paramètres de croissance ; elle a donc été utilisée pour étudier les cinétiques d'adaptation au stress. La culture continue en chémostat est, quant à elle, un bon outil d'analyse physiologique car elle maintient les cellules dans un état stable. D'autre part, cette technique permet un découplage et un contrôle des différents paramètres de culture. Elle a été mise en œuvre pour étudier spécifiquement l'effet d'un paramètre, indépendamment des autres.

Plusieurs techniques d'analyses ont été employées afin d'appréhender la physiologie de *L. lactis* de façon globale. Il est important de noter que l'ensemble des mesures biologiques ont été réalisées sur des échantillons cellulaires provenant d'une seule et même culture afin de préserver la cohérence des données en travaillant sur des états strictement identiques. Ces états physiologiques ont, en premier lieu, été caractérisés au travers de l'analyse des paramètres de fermentation (calcul des vitesses de croissance, de consommation et de production à partir du suivi cinétique des concentrations en biomasse, substrats et produits). Au niveau intracellulaire, le transcriptome a été mesuré de façon à accéder aux niveaux des ARN messagers, les premiers intermédiaires de l'expression du potentiel génétique. En aval

des ARN messagers, les véritables effecteurs cellulaires, les protéines, ont été quantifiés par une analyse protéomique.

Différents outils mathématiques, en lien avec la démarche de biologie intégrative mise en place, ont également été développés. Des outils statistiques ont d'abord permis la validation et le traitement des données recueillies avec chacune des techniques de mesure biologique. La recherche de liens entre les données et différents paramètres est passée par le calcul de corrélations. Enfin, l'intégration des données hétérogènes issues des différents niveaux d'analyse a fait appel à des techniques de modélisation.

Impact de la diminution du taux de croissance sur la physiologie cellulaire

La diminution du taux de croissance consécutive à l'imposition d'un stress est un phénomène qui peut se mesurer directement par le suivi cinétique de la croissance bactérienne. Souvent observé et décrit comme une conséquence du stress, son rôle dans l'adaptation n'est, par contre, pas clairement défini et n'est surtout pas distingué des autres mécanismes de réponse. Pourtant, deux études menées chez la levure modèle *S. cerevisiae* ont montré l'étendue de ce phénomène qui altère l'expression de 27 à 50 % des gènes de cet eucaryote (Castrillo *et al.*, 2007; Regenberg *et al.*, 2006). Chez les bactéries, l'influence de la réduction du taux de croissance sur la physiologie cellulaire est probablement aussi majeure, mais aucune étude de cet impact n'est encore publiée à ce jour. Dans cette thèse, nous avons donc choisi de caractériser la réponse de *L. lactis* à la diminution du taux de croissance.

La culture continue en chémostat a permis de faire varier spécifiquement le taux de croissance indépendamment des autres paramètres de culture. Le maintien d'un métabolisme homolactique et d'une énergie de maintenance constante quel que soit le taux de croissance a démontré la stabilité des états cellulaires analysés. La réponse apparaît aussi étendue que chez la levure puisqu'environ 30 % des gènes de *L. lactis* sont différentiellement exprimés en réponse à la diminution du taux de croissance. De la même façon, les niveaux de 194 protéines (sur les 346 mesurées) sont significativement modifiés lors de la baisse du taux de croissance.

Pour des raisons technologiques, les résultats obtenus en transcriptome sont plus exhaustifs que ceux obtenus en protéome. Néanmoins, d'une manière générale, les analyses issues de ces deux techniques sont cohérentes, ce qui renforce les conclusions établies. Les observations ont montré que la baisse du taux de croissance entraîne une répression globale des principales activités cellulaires associées à la biogénèse. En accord avec ce qui a été trouvé chez la levure (Castrillo *et al.*, 2007), ces effets se manifestent au niveau des fonctions

de transcription, traduction et division. Le métabolisme des bases, celui des acides gras et la glycolyse sont également ralentis en réponse à la diminution du taux de croissance. La surexpression massive des gènes de fonction inconnue, aussi observée chez *S. cerevisiae* (Regenberg *et al.*, 2006), redonne de l'intérêt à ces gènes souvent peu étudiés en raison de leur faible expression dans des conditions optimales de croissance. Enfin, l'activation des gènes de phages, prophages et transposons, des éléments génétiques mobiles connus pour leur rôle dans la diversification des génomes bactériens (Canchaya *et al.*, 2004), suggère un lien entre le taux de croissance et la capacité d'évolution de *L. lactis*.

La réponse stringente est un mécanisme d'adaptation bactérien régulé par l'alarmone ppGpp responsable du contrôle négatif de la transcription des ARN stables lors d'une carence en acide aminé. Le rôle pléiotropique de cette réponse fait qu'elle est potentiellement un mécanisme général permettant l'ajustement des principales fonctions bactériennes au ralentissement de croissance, indépendamment de son origine (Jain *et al.*, 2006). Afin de voir si la réponse de *L. lactis* à la diminution du taux de croissance correspondait à la réponse stringente, cette dernière a été induite et caractérisée par son transcriptome. L'induction s'est faite par ajout de norvaline dans le milieu de culture, car cet antibiotique est connu pour inhiber le chargement des ARN de transfert et déclencher la réponse stringente chez *B. subtilis* (Eymann *et al.*, 2002). Une analyse de transcriptomique comparative a révélé que le recoupement de la réponse à la baisse du taux de croissance et de la réponse stringente n'est que de 10 % chez *L. lactis*, montrant clairement que, chez cette bactérie, la réponse à la diminution du taux de croissance et la réponse stringente sont deux mécanismes distincts.

Caractérisation de la réponse de *L. lactis* à la carence en isoleucine

L. lactis rencontre de nombreux stress dans son environnement naturel mais également lors de son utilisation industrielle. De nature physico-chimique ou nutritionnelle, plusieurs stress ont déjà été étudiés mais pas la carence en azote alors même que c'est un élément crucial pour la survie et le développement des bactéries. Récemment, des profils transcriptomiques très complexes ont été observés chez les microorganismes modèles *E. coli* (Traxler *et al.*, 2008) et *B. subtilis* (Tam le *et al.*, 2007) en réponse à une carence en acide aminé. Chez *L. lactis* l'importance de l'élément azote apparaît d'autant plus forte que les bactéries lactiques sont amenées à croître dans des milieux laitiers souvent pauvres en acides aminés libres, mais surtout qu'elles présentent des auxotrophies pour de nombreux acides aminés (Cocaign-Bousquet *et al.*, 1995). Au cours de ces travaux, nous avons opté pour l'analyse de la carence en azote au travers de l'élément isoleucine. Avec les autres acides

aminés branchés (valine et leucine), l'isoleucine fait partie des acides aminés les plus abondants dans les protéines, notamment au niveau des domaines hydrophobes (Garault *et al.*, 2000).

L'imposition progressive de la carence en azote lors de la culture discontinuée, provoquée par la consommation naturelle de l'isoleucine initialement présente dans le milieu, se caractérise par une diminution graduelle de la vitesse de croissance de *L. lactis*. La croissance se poursuit néanmoins, au-delà de la disparition totale de l'isoleucine, à une vitesse extrêmement faible. Ces résultats révèlent que *L. lactis* n'est finalement pas auxotrophe pour cet acide aminé, mais que, contrairement à ce qui était attendu (Godon *et al.*, 1993), il possède une activité de biosynthèse d'isoleucine, certes très réduite.

L'augmentation du nombre de transcrits et de protéines dont l'expression varie significativement au fur et à mesure de l'adaptation à la carence en isoleucine indique une mise en place progressive de la réponse. Les processus cellulaires considérés comme affectés par la carence, en regard des profils transcriptomiques, ont généralement été confirmés par l'expression différentielle de protéines correspondant aux mêmes voies, bien qu'en nombre limité. Globalement, le processus d'adaptation à la carence en isoleucine se découpe en trois réponses. Tout d'abord, la répression de la traduction, du métabolisme des pyrimidines, de celui des acides gras ainsi que du transport et du métabolisme des sucres, est en accord avec la réduction des principales activités cellulaires liée à la diminution du taux de croissance. Le ralentissement de l'activité de transcription a, quant à lui, été mis en évidence par la chute de la vitesse d'incorporation d'uridine radioactive au cours de la cinétique de croissance.

Une réponse plus spécifique au stress imposé, visant à réduire l'impact de la carence subie, se met également en place. Cette réponse se manifeste par l'induction des systèmes de transport (*opp* et *opt*) et de dégradation des peptides (peptidases) de *L. lactis*. Ces inductions s'accompagnent d'une réorganisation complexe du métabolisme qui touche à la fois le métabolisme azoté et le métabolisme carboné, comme observé lors de la réponse à la carence en isoleucine chez *E. coli* (Traxler *et al.*, 2008). Au niveau du métabolisme des acides aminés l'induction des voies de biosynthèse des acides aminés branchés est connectée en amont au métabolisme d'autres acides aminés (sérine, alanine et aspartate par exemple) par le biais du métabolisme carboné (métabolisme du pyruvate en particulier). L'ensemble de ces inductions spécifiques à la carence imposée vise à mobiliser les différents processus cellulaires susceptibles de fournir de l'isoleucine. La comparaison transcriptomique de cette réponse avec la réponse stringente et le régulon CodY a montré l'implication conjointe de ces deux mécanismes dans la régulation des processus spécifiques de l'adaptation au stress. Ces

implications sont en accord avec la littérature puisque la réponse stringente est connue pour être mise en place lors d'une carence en acide aminé, en particulier chez *B. subtilis* (Tam le *et al.*, 2007), tandis que CodY est un répresseur de transcription associé au métabolisme azoté dont l'expression chez *L. lactis* est affectée par le pool d'acides aminés branchés (Guedon *et al.*, 2001).

Enfin, une réponse de lutte contre le stress oxydatif est également mise en place au travers de l'activation du métabolisme aérobie, du transport d'électrons, du métabolisme de la thiorédoxine et des différentes sous-unités de la pyruvate déshydrogénase. Bien qu'inattendu dans des conditions de culture strictement anaérobies, ce même type de lutte contre le stress oxydatif a déjà été observé chez *E. coli* en phase stationnaire de croissance, indépendamment de l'apport en oxygène dans le milieu (Dukan & Nystrom, 1999). Cette réponse au stress oxydatif est d'ailleurs totalement différente de la réponse au stress oxygène récemment caractérisée chez *L. lactis* (Pedersen *et al.*, 2008) comme le démontre le faible recoupement (moins de 6 %) des deux réponses transcriptomiques. Le contrôle de cette réponse au stress oxydatif semble assuré par plusieurs mécanismes révélés par transcriptomique comparative. Il implique les mécanismes de régulation de la réponse à la variation du taux de croissance et de la réponse stringente mais aussi l'agent de la répression catabolique CcpA. Habituellement relié au métabolisme carboné, ce dernier a aussi été impliqué dans la réponse à la carence en acide aminé chez *B. subtilis* (Tam le *et al.*, 2007).

Contrôle des concentrations intracellulaires en protéines chez *L. lactis*

Contrairement aux ARN messagers qui ne sont que des intermédiaires dans l'expression des potentialités d'un génome, les protéines sont les véritables effecteurs cellulaires. Elles vont donc jouer un rôle essentiel dans la physiologie de la cellule, au sein de laquelle elles sont directement reliées aux transcrits par le processus de traduction. Pourtant, les corrélations entre les niveaux de ces deux biomolécules rapportées dans la littérature sont relativement modestes avec des coefficients de corrélation de Pearson compris entre 0.45 et 0.75 pour la levure modèle *S. cerevisiae* (Hack, 2004) et du même ordre de grandeur (environ 0.5 selon les conditions de croissance) pour la bactérie *Desulfovibrio vulgaris* (Nie *et al.*, 2006a). Devant ce constat, se pose la question de l'existence de régulations au niveau de la traduction et de leur significativité. Deux autres phénomènes entrent également en jeu dans le contrôle des concentrations en protéines : la dégradation des protéines et leur dilution lors de la croissance cellulaire. Au cours de cette thèse, nous nous sommes intéressés aux différents facteurs susceptibles d'influencer les niveaux de protéines chez *L. lactis*, pour qui les corrélations avec

les niveaux de transcrits sont même plus faibles (coefficients de corrélation de Pearson voisins de 0.25) que pour les autres microorganismes.

La recherche des paramètres ayant le plus d'influence sur les concentrations en protéines a été abordée par une approche de modélisation statistique. Cette méthode consiste à introduire simultanément l'ensemble des paramètres susceptibles d'affecter la variable à expliquer (ici la concentration en protéine) dans un modèle de covariance. Ce type de modèle correspond à une régression linéaire multiple dont la particularité est de pouvoir prendre en compte indifféremment des paramètres quantitatifs et/ou qualitatifs. L'application d'un algorithme de sélection basé sur le critère d'information d'Akaike (Burnham & Anderson, 2002) incrémente successivement différents modèles de covariances pour ne retenir au final que celui s'ajustant au mieux aux données à modéliser tout en utilisant un minimum de paramètres. Ce compromis entre la qualité d'ajustement et la complexité du modèle permet donc de ne sélectionner que les paramètres les plus significatifs. D'autre part, les paramètres quantitatifs sont classés par ordre d'importance grâce à une normalisation préalable par centrage et réduction des valeurs associées à chacun d'entre eux.

Cette méthode de recherche de déterminants biologiques a été appliquée aux données de protéome obtenues lors de la culture continue de *L. lactis* en chémostat à différents taux de croissance et lors de l'adaptation progressive à la carence en isoleucine. Indépendamment du mode de culture et en accord avec les résultats obtenus pour les bactéries *E. coli* et *Haemophilus influenzae* (Lithwick & Margalit, 2003), l'indice de biais de codon (CAI) est clairement apparu comme le paramètre le plus influent. Sachant que le CAI traduit un biais de codon au niveau de la séquence du gène (Sharp & Li, 1987), le fait que son influence soit plus importante encore que celle des concentrations en ARN messagers suggère un contrôle génétique des niveaux de protéine plus important que l'adaptation transcriptionnelle. La sélection de l'aromaticité est révélatrice d'un biais d'acides aminés ; en particulier, la faible utilisation du tryptophane dans le protéome de *L. lactis* pourrait être à l'origine de l'influence négative de l'aromaticité sur la concentration en protéines. Aucun lien n'a été mis en évidence entre l'aromaticité et les niveaux de protéines chez *S. cerevisiae* mais cet eucaryote inférieur présente certainement un biais d'acides aminés différent de celui de la bactérie *L. lactis* en raison de la distance phylogénétique qui les sépare. L'impact négatif de la longueur des gènes sur les niveaux de protéines pourrait, quant à lui, résulter d'une baisse de la densité en ribosomes sur les messagers les plus longs. Un tel phénomène a en effet déjà pu être observé chez la levure grâce à une technique de séparation des transcrits en fonction de leur densité en ribosomes (Arava *et al.*, 2003). Parmi les paramètres qualitatifs, la catégorie fonctionnelle

dévoile un lien entre la fonction des protéines et la régulation de leur niveau d'expression. Les protéines associées à la traduction ou au métabolisme énergétique sont toujours les plus concentrées quel que soit le mode de culture, mais l'impact des différentes catégories fonctionnelles n'est pas le même pour les cellules adaptées en chémostat et celles devant répondre à la carence en isoleucine. Ce résultat implique donc une régulation de l'adaptation au travers de la fonction protéique. Le modèle sélectionné en batch est le seul à renfermer le paramètre taux de croissance. En chémostat au contraire, les modèles n'évoluent pas en fonction du taux de croissance. Ce dernier peut donc être considéré comme une variable d'adaptation participant à la régulation de la réponse dynamique à la carence en isoleucine.

La mise en équations des trois processus cellulaires qui influencent les niveaux de protéines se simplifie en régime permanent. En effet, à l'équilibre, la concentration en protéine résulte d'un équilibre entre les vitesses de traduction des messagers, de dilution par la croissance et de dégradation des protéines (cf. figure 3 présentée dans l'étude bibliographique). Couplée à la modélisation des mesures de transcriptome et de protéome, la résolution de cette équation a permis d'estimer l'efficacité de traduction (k') et la constante de dégradation (k'') pour chacune des protéines. Une grande variabilité des valeurs de k' et de k'' a été observée au sein de la population protéique de *L. lactis*. D'autre part, k' et k'' ne sont pas constantes mais inversement proportionnelles au taux de croissance. Ainsi, à faible taux de croissance, la traduction est plus efficace qu'en croissance rapide mais la dégradation des protéines est plus importante. Les processus de traduction et de dégradation des protéines exercent donc des contrôles antagonistes ; une telle régulation pourrait permettre d'exercer un contrôle plus fin des niveaux intracellulaires de protéines.

Parmi les protéines les plus efficacement traduites se trouvent essentiellement des protéines liées à la croissance bactérienne. Ce résultat est en accord avec le fait que les ARN messagers codant pour les protéines ribosomales sont ceux qui portent le plus grand nombre de ribosomes chez deux archaebactéries halophiles (Lange *et al.*, 2007). Ces observations sont également à relier avec la forte corrélation entre k' et le CAI puisque les protéines ribosomales sont connues pour avoir des CAI plus élevés que la moyenne. Ce lien entre l'efficacité de traduction et le CAI justifie la possibilité d'utiliser ce paramètre comme une mesure indirecte de la vitesse de traduction proposée par Brockmann *et al.* (2007).

Les valeurs de stabilité des protéines que nous avons pu calculer sont centrées autour de 23 minutes pour le taux de croissance le plus faible (0.09 h^{-1}) et 224 minutes pour le plus élevé (0.88 h^{-1}). Ces valeurs diffèrent totalement des 31 h de moyenne trouvées dans une première étude menée chez *S. cerevisiae* (Pratt *et al.*, 2002) mais sont en adéquation avec des

travaux plus récents, toujours menés chez la levure mais utilisant une technique de mesure différente, qui trouvent des valeurs de l'ordre de 43 minutes (Belle *et al.*, 2006). Cette dernière étude indique également que les protéines ribosomales ainsi que celles liées au métabolisme des acides aminés sont globalement plus stables que les autres. De façon différente, chez *L. lactis* nous avons relevé la très grande stabilité des protéines associées au stress. Alors que la stabilité des protéines de la levure a été corrélée à leur richesse en valine et en sérine (Belle *et al.*, 2006), chez *L. lactis*, k'' est inversement corrélé à la richesse en cystéine. Ces résultats sont difficilement comparables en raison du biais d'acide aminé et de la spécificité des protéases qui sont propres à chaque espèce. Le lien entre la stabilité et la richesse en cystéine d'une protéine est néanmoins fort probablement associé au potentiel stabilisateur des ponts disulfures qui peuvent se former entre deux cystéines.

QUESTIONNEMENTS ET PERSPECTIVES

Existe-t-il un mécanisme général de réponse au stress chez *L. lactis* ?

Chez les modèles des bactéries à Gram négatif et à Gram positif, *E. coli* et *B. subtilis*, respectivement, il existe un mécanisme général de réponse au stress (Hecker & Volker, 2001). Cette réponse est principalement coordonnée au niveau transcriptionnel par des facteurs sigma alternatifs, σ^S pour *E. coli* (Weber *et al.*, 2005) et σ^B pour *B. subtilis* (Hecker *et al.*, 2007). Ces facteurs sigma de stress modifient l'affinité de l'ARN polymérase, permettent la reconnaissance de promoteurs spécifiques et contrôlent par ce biais l'expression de plus de 150 gènes associés à la lutte contre différents types de stress (oxydatif, osmotique, thermique, antibiotique...). En particulier, la mise en place de la réponse générale au stress se caractérise par l'induction des chaperonnes (DnaK, GroES, GroEL, GrpE) et de leur régulateur HrcA, des protéases (Clp, HtrA et HflB), des ATPases, des protéines de réparation de l'ADN (RecA), des protéines de réponse au stress froid (CSP pour Cold Shock Protein), des transporteurs de type « multidrug resistance » (MDR), des protéines de détoxification (SodA, GshR) et des protéines générales de stress (YbjE, YtgH...).

La mise en place de ce mécanisme de protection croisée joue un rôle essentiel dans la survie et l'adaptation des bactéries en leur permettant de lutter contre le stress inducteur tout en les préparant à en affronter d'autres. La réponse générale au stress semble être assez répandue chez les bactéries mais son existence chez *L. lactis* reste hypothétique en

raison de l'absence de facteur sigma de stress. En effet, les deux seuls facteurs σ alternatifs présents sur le génome de *L. lactis* sont un facteur extracytoplasmique, SigX, dont le rôle est encore inconnu et un facteur, ComX, contrôlant l'expression des gènes de compétence (Wydau *et al.*, 2006).

Les résultats obtenus au cours de cette étude ne permettent pas de conclure à la mise en place d'une réponse générale au stress chez *L. lactis* telle qu'elle est définie chez *E. coli* ou *B. subtilis*. En effet, certains gènes ou protéines liées à cette réponse sont ponctuellement induits mais aucune tendance générale pour l'ensemble des protéines de stress n'a pu être dégagée. Par exemple, une réponse croisée au stress oxydatif se met bien en place lors de l'adaptation à la carence en isoleucine mais les chaperonnes, des éléments majeurs dans la réponse générale au stress, sont sous-exprimées. De plus, l'expression de la plupart des protéines de stress n'est pas modifiée en réponse à la diminution du taux de croissance alors que chez la levure une induction massive des protéines de stress avait été observée (Regenberg *et al.*, 2006). Enfin, une étude antérieure portant sur la carence carbone a montré l'induction de différents gènes conférant une résistance à divers stress mais l'absence d'induction des chaperonnes et des protéases ne permet pas non plus de conclure à l'activation d'une réponse générale au stress (Redon *et al.*, 2005b).

L'absence de réponse générale au stress semble donc être une originalité de *L. lactis* probablement en lien avec l'absence de facteur σ de stress. Il existe cependant des mécanismes tels que la réponse stringente et la réponse à la baisse du taux de croissance, que nous avons démontré comme étant distincts, qui pourraient être mis en place dans la réponse aux stress. D'après nos comparaisons transcriptomiques, la réponse stringente est impliquée dans l'adaptation à la carence en isoleucine à hauteur de 69 %. Elle régulerait en particulier le métabolisme carboné par son contrôle sur les gènes de la glycolyse et de la fermentation. Cependant, avec seulement 14 % de recoupement avec la réponse à la carence en glucose, son rôle semble bien plus modéré. L'hypothèse de la réponse stringente comme mécanisme commun de réponse au stress chez *L. lactis* semble donc peu probable. La réponse à la variation du taux de croissance permet, quant à elle, d'expliquer environ 30 % des réponses à la carence en glucose et à la carence en isoleucine. Son impact se matérialise essentiellement au niveau des gènes impliqués dans la biogénèse et la croissance en particulier pour ceux reliés à la traduction. Les gènes associés à ces fonctions font généralement partie de ceux qui répondent de façon identique à divers stress. Ils ont notamment été identifiés chez *E. coli* comme communs à l'adaptation à la carence en azote et en soufre (Gyaneshwar *et al.*, 2005a).

La mise en place d'une réponse commune à divers stress et liée à la variation du taux de croissance semble donc jouer un rôle important dans l'adaptation de *L. lactis*. Néanmoins, ce résultat s'étend bien au-delà d'un simple contrôle de la transcription des gènes par un effet taux de croissance puisque nous avons vu que ce dernier influence aussi l'efficacité de traduction, la dilution et la dégradation des protéines. Toutefois, les véritables mécanismes et régulateurs au travers desquels s'exerce la régulation par le taux de croissance restent encore à découvrir. En ce sens, l'étude des deux protéines identifiées comme des régulateurs potentiels de la réponse à la baisse du taux de croissance est nécessaire : LlrE est un régulateur de la famille OmpR, notamment connu pour être un régulateur global chez *E. coli* (Bipatnath *et al.*, 1998) tandis que YabB est un gène putatif présentant des homologues avec les régulateurs de la famille Cro/Ci. Deux types de mutation sont envisageables pour confirmer le rôle de ces gènes : l'inactivation, par délétion ou insertion d'une cassette de résistance, qui n'est possible que si les gènes visés ne sont pas essentiels, et la surexpression par le positionnement du gène d'intérêt sous contrôle d'un promoteur fort. Les caractéristiques de croissance des mutants et les profils transcriptomiques associés seront comparés à la souche sauvage. Si cette approche permet d'identifier les gènes impliqués dans la régulation de la réponse à la baisse du taux de croissance, l'aspect mécanistique de la régulation pourra alors être étudié, en particulier au travers de la recherche de motifs régulateurs. Enfin, un chémostat limité en isoleucine a été utilisé pour caractériser les gènes dont l'expression est modifiée avec le taux de croissance. Néanmoins, nous ne pouvons pas écarter un biais lié à cette limitation. Aussi, il serait intéressant d'affiner la définition du régulon. Des études similaires en chémostat mais utilisant des éléments limitants autres que l'isoleucine (souffre, phosphate, vitamine) pourront être réalisées. La limitation carbonée ne semble toutefois pas un bon choix en raison de la transition métabolique (entre le métabolisme homolactique et le métabolisme mixte) qu'elle risque d'engendrer.

Quels sont les facteurs prépondérants dans l'adaptation de *L. lactis* ?

Nous avons pu voir au cours de cette étude que l'adaptation de *L. lactis* est un mécanisme complexe qui met en jeu divers paramètres en plus des différents effets que peut avoir le taux de croissance. Divers facteurs sont en effet déterminants pour la régulation des concentrations en protéines et, par extension, pour la physiologie cellulaire qui en résulte. En particulier, le CAI et la concentration en ARN messager sont apparus comme les paramètres ayant le plus d'influence sur les niveaux de protéines. Comme nous avons démontré que le CAI est très fortement corrélé à l'efficacité de traduction, ces résultats suggèrent que le

processus de traduction jouerait un rôle majeur dans l'adaptation de *L. lactis*. Les régulations de la traduction vont donc également être importantes et se manifestent déjà au travers de la variabilité de l'efficacité de traduction en fonction du taux de croissance.

Une mesure biologique des efficacités de traduction dans différentes conditions de culture permettrait de confirmer l'implication du processus de traduction dans l'adaptation. Cette mesure peut se faire *via* la détermination des profils polysomiques. La mesure du nombre de ribosomes associés à chacun des transcrits est, en effet, une technique développée chez la levure modèle *S. cerevisiae* qui permet d'estimer l'état de traduction des ARN messagers (Arava *et al.*, 2003). Cette méthode est basée sur la séparation des polysomes sur gradient de saccharose et l'analyse sur puce à ADN des différentes fractions. Initialement développée pour la levure, cette technique pourrait être plus difficilement transposable aux procaryotes en raison du couplage et de la co-localisation des processus de transcription et de traduction au sein du cytoplasme. Néanmoins, elle a déjà pu être adaptée aux archaeobactéries avec succès (Lange *et al.*, 2007) et devrait donc pouvoir être appliquée à *L. lactis*.

Si la traduction et ses régulations apparaissent comme des phénomènes majeurs dans l'adaptation de *L. lactis* à son environnement, la dégradation des protéines et ses régulations ne peuvent pas non plus être négligées. Les valeurs des constantes de dégradation des protéines sont en effet variables entre protéines, dépendantes du taux de croissance et surtout plus élevées que ce qui était attendu. Enfin, une méthode de modélisation des contrôles cellulaires a révélé que l'adaptation des niveaux de protéines entre deux taux de croissance différents était davantage régulée par l'action conjointe de la dilution et de la dégradation protéique que par la traduction elle-même. Cette méthode est issue de l'approche développée par Westerhoff et ses collaborateurs sur d'autres types de données biologiques tels que les flux métaboliques (Daran-Lapujade *et al.*, 2007; ter Kuile & Westerhoff, 2001). Bien qu'à l'échelle individuelle des régulations différentes soient mises en place, les résultats indiquent que 50 à 60 % des variations de concentration protéique sont régulées par le phénomène de disparition qui regroupe les processus de dégradation et de dilution des protéines. La part contrôlée par la traduction oscille entre 15 et 40 % selon les conditions. Les mécanismes de régulation tendent néanmoins à se complexifier à fort taux de croissance et le contrôle devient de plus en plus partagé entre les différents processus.

Devant l'importance de la dilution et de la dégradation, deux processus biologiques « consommateurs » de protéines qui ont des effets antagonistes à celui de la traduction, une mesure biologique de la dégradation des protéines semble primordiale. Une telle détermination nous permettrait effectivement de valider les estimations réalisées par la

modélisation et de mieux quantifier le poids de chacun des processus cellulaires dans l'ajustement des niveaux de protéines chez *L. lactis*. Les deux seules études de stabilité des protéines à l'échelle globale publiées à ce jour portent sur la levure mais utilisent des techniques très différentes. Les travaux les plus anciens (Pratt *et al.*, 2002) se basent sur la cinétique de disparition d'un marquage isotopique des protéines de *S. cerevisiae* par la leucine deutériée lors d'une culture en régime permanent en chémostat. Dans la méthode la plus récente (Belle *et al.*, 2006) les protéines sont marquées par un épitope et leur constante de dégradation est mesurée individuellement après blocage de la traduction (cycloheximide). Similairement, il pourrait être envisagé de mesurer la dégradation protéique chez *L. lactis*, suite à l'inhibition de la traduction, par le suivi cinétique de la disparition de chacune des protéines simultanément par des techniques protéomiques. Cette méthode permettrait de s'affranchir de tout marquage et ne requiert pas une banque de protéines marquées. La protéomique shotgun (Pisitkun *et al.*, 2007), couplant la séparation par chromatographie liquide à l'identification par spectrométrie de masse, serait d'ailleurs un outil de choix puisqu'il couvrirait une plus large gamme de protéines que les techniques en deux dimensions plus classiques (Wolff *et al.*, 2006).

Le contrôle de l'adaptation peut-il s'effectuer à d'autres niveaux ?

Nous avons vu l'implication de divers processus dans la régulation des niveaux de protéines qui sont de fait déterminants pour la physiologie cellulaire. Cette vision, centrée sur les protéines, implique plusieurs mécanismes de régulation mais reste encore incomplète. Les niveaux de messagers qui interviennent dans le processus de traduction sont par exemple eux aussi influencés par d'autres processus cellulaires. Nous avons démontré l'implication de régulateurs de transcription dans les différentes réponses étudiées. Néanmoins, la transcription n'est pas le seul processus qui affecte les niveaux intracellulaires de transcrits. En effet, comme pour les protéines, la dilution par la croissance et la dégradation sont également impliquées (cf. figure 3 présentée dans l'étude bibliographique). Un élargissement de la fenêtre d'étude pour englober les différents processus cellulaires reliant l'expression des gènes aux quantités de protéines est donc une perspective attirante pour améliorer la compréhension globale de l'adaptation de *L. lactis* à son environnement.

Des résultats précédemment obtenus ont montré une variation de la stabilité des ARN messagers en réponse à la carence en carbone (Redon *et al.*, 2005a) et l'importance de ce mode de régulation sur les concentrations en ARN messagers. Les stabilités des transcrits ont également été mesurées en réponse à la variation du taux de croissance et lors de l'adaptation

à la carence en isoleucine (cf. chapitre annexe). Les premières analyses dévoilent une relation linéaire entre les valeurs de stabilité des ARN messagers et le taux de croissance, montrant une nouvelle fois, l'importance de ce paramètre. Contrairement aux protéines, la dilution des ARN messagers semble négligeable devant leur vitesse de dégradation.

L'utilisation de modèles de covariance semblables à ceux proposés pour modéliser les concentrations en protéines devrait maintenant permettre la recherche des paramètres déterminants des niveaux d'ARN messagers de *L. lactis* dans les différentes conditions environnementales étudiées. Des déterminants identiques à ceux proposés pour les protéines peuvent également s'avérer importants mais l'implication de nouveaux paramètres doit aussi être envisagée (nature de la séquence promotrice, pourcentage en GC...). L'étude de la stabilité des transcrits en réponse à la carence en carbone a, par exemple, permis d'observer la présence stabilisatrice d'un motif AGGAG dans la région 5' non traduite des ARN messagers (Redon *et al.*, 2005a). Cet élément est semblable au site de liaison des ribosomes ce qui suggère que la fixation de la machinerie de traduction sur les transcrits pourrait avoir un effet stabilisateur. Une telle hypothèse pourra être démontrée expérimentalement avec les données de polysome.

La multiplication des fourches de réplication, plus ou moins importante selon le taux de croissance, conduit à une différence du nombre de copies des gènes selon leur proximité par rapport à l'origine de réplication. Ainsi, le processus de transcription peut aussi être influencé par la quantité relative d'ADN disponible. Lors de l'étude en chémostat, nous avons en effet observé un lien entre l'expression des gènes reliés au taux de croissance et leur localisation chromosomique. Cet effet « dosage de gènes » apparaît donc comme potentiellement important dans l'adaptation de *L. lactis* et mérite donc d'être mesuré. La quantification du nombre de copies de chacun des gènes dans différentes conditions semble possible par hybridation directe de l'ADN génomique de *L. lactis* cultivé à différents taux de croissance sur puces à ADN. Elle devrait permettre une meilleure estimation de la vitesse de transcription et la modélisation des niveaux d'ARN messagers.

A plus long terme, la compilation mathématique des modèles obtenus lors de l'étude de la régulation des niveaux de protéines et de transcrits permettra d'obtenir une vision plus réaliste des mécanismes d'adaptation de *L. lactis*.

Comment caractériser au mieux l'adaptation ?

Comme décrit précédemment, chacune des techniques utilisées dans cette étude présente des avantages qui lui sont propres et a été utilisée dans un but bien spécifique. Le transcriptome en particulier présente l'avantage d'offrir une vision quasi-exhaustive de la cellule. La pertinence de l'étude transcriptomique pour relater l'adaptation est cependant remise en cause par la faible corrélation entre les niveaux de transcrits et de protéines. Malgré la ténuité de ce lien, les conclusions déduites de ces deux niveaux d'analyse se recourent. Ce constat souligne la nécessité d'appréhender les problématiques biologiques de manière intégrative.

La cellule est un système vivant dans lequel ces différents mécanismes s'imbriquent et interagissent. La mesure d'un paramètre biologique particulier renseigne sur un point précis de la physiologie cellulaire mais ne permet pas de comprendre son fonctionnement global. En ce sens, la biologie intégrative revêt tout son intérêt puisque, au travers de la modélisation des processus biologiques, elle a pour vocation d'apporter une vision globale et compréhensive de la physiologie cellulaire. La construction de modèles mathématiques nous a justement permis de réconcilier l'ensemble des données issues des différentes analyses effectuées et de mieux comprendre les mécanismes d'adaptation de *L. lactis*. La biologie intégrative demeure donc un outil de choix pour la caractérisation de l'adaptation et mérite absolument d'être étendue à d'autres processus cellulaires, comme à d'autres microorganismes.

Références bibliographiques

A

- Aderem A. (2005). Systems biology: its practice and challenges. *Cell* **121**: 511-3.
- Akashi H., and Gojobori T. (2002). Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **99**: 3695-700.
- Albers E., Larsson C., Andlid T., Walsh M. C., and Gustafsson L. (2007). Effect of nutrient starvation on the cellular composition and metabolic capacity of *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **73**: 4839-48.
- Amrani N., Sachs M. S., and Jacobson A. (2006). Early nonsense: mRNA decay solves a translational problem. *Nat Rev Mol Cell Biol* **7**: 415-25.
- Andersen A. Z., Carvalho A. L., Neves A. R., Santos H., Kummer U., and Olsen L. F. (2009). The metabolic pH response in *Lactococcus lactis*: an integrative experimental and modelling approach. *Comput Biol Chem* **33**: 71-83.
- Arava Y. (2003). Isolation of polysomal RNA for microarray analysis. *Methods Mol Biol* **224**: 79-87.
- Arava Y., Wang Y., Storey J. D., Liu C. L., Brown P. O., and Herschlag D. (2003). Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **100**: 3889-94.
- Ardo Y. (2006). Flavour formation by amino acid catabolism. *Biotechnol Adv* **24**: 238-42.
- Ayad E. H., Verheul A., Engels W. J., Wouters J. T., and Smit G. (2001). Enhanced flavour formation by combination of selected lactococci from industrial and artisanal origin with focus on completion of a metabolic pathway. *J Appl Microbiol* **90**: 59-67.

B

- Bachi A., and Bonaldi T. (2008). Quantitative proteomics as a new piece of the systems biology puzzle. *J Proteomics* **71**: 357-67.
- Bacon J., Dover L. G., Hatch K. A., Zhang Y., Gomes J. M., Kendall S., Wernisch L., Stoker N. G., Butcher P. D., Besra G. S., and Marsh P. D. (2007). Lipid composition and transcriptional response of *Mycobacterium tuberculosis* grown under iron-limitation in continuous culture: identification of a novel wax ester. *Microbiology* **153**: 1435-44.
- Baldi P., and Long A. D. (2001). A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes. *Bioinformatics* **17**: 509-19.
- Barriere C., Veiga-da-Cunha M., Pons N., Guedon E., van Hijum S. A., Kok J., Kuipers O. P., Ehrlich D. S., and Renault P. (2005). Fructose utilization in *Lactococcus lactis* as a model for low-GC gram-positive bacteria: its regulator, signal, and DNA-binding site. *J Bacteriol* **187**: 3752-61.
- Beilharz T. H., and Preiss T. (2004). Translational profiling: the genome-wide measure of the nascent proteome. *Brief Funct Genomic Proteomic* **3**: 103-11.
- Belle A., Tanay A., Bitincka L., Shamir R., and O'Shea E. K. (2006). Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci U S A* **103**: 13004-9.
- Benjamini Y. H., Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B.*: 289-300.

- Bernstein J. A., Khodursky A. B., Lin P. H., Lin-Chao S., and Cohen S. N. (2002). Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A* **99**: 9697-702.
- Berth M., Moser F. M., Kolbe M., and Bernhardt J. (2007). The state of the art in the analysis of two-dimensional gel electrophoresis images. *Appl Microbiol Biotechnol* **76**: 1223-43.
- Beyer A., Hollunder J., Nasheuer H. P., and Wilhelm T. (2004). Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. *Mol Cell Proteomics* **3**: 1083-92.
- Beyer N. H., Roepstorff P., Hammer K., and Kilstrop M. (2003). Proteome analysis of the purine stimulon from *Lactococcus lactis*. *Proteomics* **3**: 786-97.
- Bipatnath M., Dennis P. P., and Bremer H. (1998). Initiation and velocity of chromosome replication in *Escherichia coli* B/r and K-12. *J Bacteriol* **180**: 265-73.
- Blackstock W. P., and Weir M. P. (1999). Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol* **17**: 121-7.
- Bochner B. R. (2009). Global phenotypic characterization of bacteria. *FEMS Microbiol Rev* **33**: 191-205.
- Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K., Weissenbach J., Ehrlich S. D., and Sorokin A. (2001). The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res* **11**: 731-53.
- Bork P., and Serrano L. (2005). Towards cellular systems in 4D. *Cell* **121**: 507-9.
- Braisted J. C., Kuntumalla S., Vogel C., Marcotte E. M., Rodrigues A. R., Wang R., Huang S. T., Ferlanti E. S., Saeed A. I., Fleischmann R. D., Peterson S. N., and Pieper R. (2008). The APEX Quantitative Proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics* **9**: 529.
- Brantl S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol* **10**: 102-9.
- Braun R. J., Kinkl N., Beer M., and Ueffing M. (2007). Two-dimensional electrophoresis of membrane proteins. *Anal Bioanal Chem* **389**: 1033-45.
- Brazma A., Krestyaninova M., and Sarkans U. (2006). Standards for systems biology. *Nat Rev Genet* **7**: 593-605.
- Brockmann R., Beyer A., Heinisch J. J., and Wilhelm T. (2007). Posttranscriptional expression regulation: what determines translation rates? *PLoS Comput Biol* **3**: e57.
- Bruggeman F. J., and Westerhoff H. V. (2007). The nature of systems biology. *Trends Microbiol* **15**: 45-50.
- Budin-Verneuil A., Pichereau V., Auffray Y., Ehrlich D. S., and Maguin E. (2005). Proteomic characterization of the acid tolerance response in *Lactococcus lactis* MG1363. *Proteomics* **5**: 4794-807.
- Burnham K. P., and Anderson D. R. (2002). "Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach," Springer-Verlag, New-York.
- Burrage K., Hood L., and Ragan M. A. (2006). Advanced computing for systems biology. *Brief Bioinform* **7**: 390-8.

C

- Canas B., Lopez-Ferrer D., Ramos-Fernandez A., Camafeita E., and Calvo E. (2006). Mass spectrometry technologies for proteomics. *Brief Funct Genomic Proteomic* **4**: 295-320.
- Canchaya C., Fournous G., and Brussow H. (2004). The impact of prophages on bacterial chromosomes. *Mol Microbiol* **53**: 9-18.
- Carpousis A. J. (2007). The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E. *Annu Rev Microbiol* **61**: 71-87.
- Carrier T. A., and Keasling J. D. (1997). Mechanistic modeling of prokaryotic mRNA decay. *J Theor Biol* **189**: 195-209.
- Cassman M. (2005). Barriers to progress in systems biology. *Nature* **438**: 1079.
- Castrillo J. I., Zeef L. A., Hoyle D. C., Zhang N., Hayes A., Gardner D. C., Cornell M. J., Petty J., Hakes L., Wardleworth L., Rash B., Brown M., Dunn W. B., Broadhurst D., O'Donoghue K., Hester S. S., Dunkley T. P., Hart S. R., Swainston N., Li P., Gaskell S. J., Paton N. W., Lilley K. S., Kell D. B., and Oliver S. G. (2007). Growth control of the eukaryote cell: a systems biology study in yeast. *J Biol* **6**: 4.
- Chaussee M. A., Dmitriev A. V., Callegari E. A., and Chaussee M. S. (2008). Growth phase-associated changes in the transcriptome and proteome of Streptococcus pyogenes. *Arch Microbiol* **189**: 27-41.
- Cheung V. G., Morley M., Aguilar F., Massimi A., Kucherlapati R., and Childs G. (1999). Making and reading microarrays. *Nat Genet* **21**: 15-9.
- Chich J. F., David O., Villers F., Schaeffer B., Lutomski D., and Huet S. (2007). Statistics for proteomics: experimental design and 2-DE differential analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* **849**: 261-72.
- Claverys J. P., and Martin B. (2003). Bacterial "competence" genes: signatures of active transformation, or only remnants? *Trends Microbiol* **11**: 161-5.
- Cloonan N., and Grimmond S. M. (2008). Transcriptome content and dynamics at single-nucleotide resolution. *Genome Biol* **9**: 234.
- Cocaign-Bousquet M., Even S., Lindley N. D., and Loubiere P. (2002). Anaerobic sugar catabolism in Lactococcus lactis: genetic regulation and enzyme control over pathway flux. *Appl Microbiol Biotechnol* **60**: 24-32.
- Cocaign-Bousquet M., Garrigues C., Loubiere P., and Lindley N. D. (1996). Physiology of pyruvate metabolism in Lactococcus lactis. *Antonie Van Leeuwenhoek* **70**: 253-67.
- Cocaign-Bousquet M., Guarrigues C., Novak L., Lindley N., and Loubiere P. (1995). Rational development of a simple synthetic medium for the sustained growth of *Lactococcus lactis*. *J Appl Bacteriol* **79**: 108-116.
- Coleman S. A., Fischer E. R., Howe D., Mead D. J., and Heinzen R. A. (2004). Temporal analysis of Coxiella burnetii morphological differentiation. *J Bacteriol* **186**: 7344-52.
- Colombie V., Bideaux C., Goma G., and Uribe Larrea J. L. (2005). Effects of glucose limitation on biomass and spiramycin production by Streptomyces ambofaciens. *Bioprocess Biosyst Eng* **28**: 55-61.
- Condon C. (2003). RNA processing and degradation in Bacillus subtilis. *Microbiol Mol Biol Rev* **67**: 157-74, table of contents.
- Coveney P. V., and Fowler P. W. (2005). Modelling biological complexity: a physical scientist's perspective. *J R Soc Interface* **2**: 267-80.

- Cowley A. W., Jr. (2004). The elusive field of systems biology. *Physiol Genomics* **16**: 285-6.
- Cox B., Kislinger T., and Emili A. (2005). Integrating gene and protein expression data: pattern analysis and profile mining. *Methods* **35**: 303-14.
- Csonka L. N. (1989). Physiological and genetic responses of bacteria to osmotic stress. *Microbiol Rev* **53**: 121-47.

D

- Dabour N., Kheadr E., Benhamou N., Fliss I., and LaPointe G. (2006). Improvement of texture and structure of reduced-fat Cheddar cheese by exopolysaccharide-producing lactococci. *J Dairy Sci* **89**: 95-110.
- Daran-Lapujade P., Rossell S., van Gulik W. M., Luttik M. A., de Groot M. J., Slijper M., Heck A. J., Daran J. M., de Winde J. H., Westerhoff H. V., Pronk J. T., and Bakker B. M. (2007). The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *Proc Natl Acad Sci U S A* **104**: 15753-8.
- Dauner M., Storni T., and Sauer U. (2001). *Bacillus subtilis* metabolism and energetics in carbon-limited and excess-carbon chemostat culture. *J Bacteriol* **183**: 7308-17.
- de Lorenzo V., and Galperin M. (2009). Microbial systems biology: bottom up and top down. *FEMS Microbiol Rev* **33**: 1-2.
- Deana A., and Belasco J. G. (2005). Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev* **19**: 2526-33.
- Delorme C., Godon J. J., Ehrlich S. D., and Renault P. (1993). Gene inactivation in *Lactococcus lactis*: histidine biosynthesis. *J Bacteriol* **175**: 4391-9.
- Denich T. J., Beaudette L. A., Lee H., and Trevors J. T. (2003). Effect of selected environmental and physico-chemical factors on bacterial cytoplasmic membranes. *J Microbiol Methods* **52**: 149-82.
- Dennis P. P., Ehrenberg M., and Bremer H. (2004). Control of rRNA synthesis in *Escherichia coli*: a systems biology approach. *Microbiol Mol Biol Rev* **68**: 639-68.
- Dettmer K., Aronov P. A., and Hammock B. D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrom Rev* **26**: 51-78.
- Deutscher J., and Saier M. H., Jr. (2005). Ser/Thr/Tyr protein phosphorylation in bacteria - for long time neglected, now well established. *J Mol Microbiol Biotechnol* **9**: 125-31.
- Deutscher M. P. (2006). Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res* **34**: 659-66.
- Dhingra V., Gupta M., Andacht T., and Fu Z. F. (2005). New frontiers in proteomics research: a perspective. *Int J Pharm* **299**: 1-18.
- Do J. H., and Choi D. K. (2006). Normalization of microarray data: single-labeled and dual-labeled arrays. *Mol Cells* **22**: 254-61.
- Dobrogosz W. J. (1981). Metabolism - Enzymatic activity. In "Manual of methods for general bacteriology" (W. A. Wood, Ed.), pp. 366-392, American society for microbiology, Washington, DC.
- dos Reis M., Wernisch L., and Savva R. (2003). Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res* **31**: 6976-85.

- Dressaire C., Redon E., Milhem H., Besse P., Loubiere P., and Cocaign-Bousquet M. (2008). Growth rate regulated genes and their wide involvement in the *Lactococcus lactis* stress responses. *BMC Genomics* **9**: 343.
- Dukan S., and Nystrom T. (1999). Oxidative stress defense and deterioration of growth-arrested *Escherichia coli* cells. *J Biol Chem* **274**: 26027-32.
- Dunn W. B., Bailey N. J., and Johnson H. E. (2005). Measuring the metabolome: current analytical technologies. *Analyst* **130**: 606-25.
- Duwat P., Cesselin B., Sourice S., and Gruss A. (2000). *Lactococcus lactis*, a bacterial model for stress responses and survival. *Int J Food Microbiol* **55**: 83-6.
- Duwat P., Sourice S., Cesselin B., Lamberet G., Vido K., Gaudu P., Le Loir Y., Violet F., Loubiere P., and Gruss A. (2001). Respiration capacity of the fermenting bacterium *Lactococcus lactis* and its positive effects on growth and survival. *J Bacteriol* **183**: 4509-16.

E

- El-Sharoud W. M. (2004). Ribosome inactivation for preservation: concepts and reservations. *Sci Prog* **87**: 137-52.
- El-Sharoud W. M., and Graumann P. L. (2007). Cold shock proteins aid coupling of transcription and translation in bacteria. *Sci Prog* **90**: 15-27.
- Enfors S. O. (1991). Modes of operation in fermentation: batch, fed-batch and continuous processes. In "Bioreactors engineering course notes" (M. Bervovic, and T. Koloini, Eds.), pp. 1-40, Kridic, B.
- Even S., Lindley N. D., Loubiere P., and Cocaign-Bousquet M. (2002). Dynamic response of catabolic pathways to autoacidification in *Lactococcus lactis*: transcript profiling and stability in relation to metabolic and energetic constraints. *Mol Microbiol* **45**: 1143-52.
- Eymann C., Homuth G., Scharf C., and Hecker M. (2002). *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. *J Bacteriol* **184**: 2500-20.

F

- Fan R., Vermesh O., Srivastava A., Yen B. K., Qin L., Ahmad H., Kwong G. A., Liu C. C., Gould J., Hood L., and Heath J. R. (2008). Integrated barcode chips for rapid, multiplexed analysis of proteins in microliter quantities of blood. *Nat Biotechnol* **26**: 1373-8.
- Feder M. E., and Walser J. C. (2005). The biological limitations of transcriptomics in elucidating stress and stress responses. *J Evol Biol* **18**: 901-10.
- Ferenci T. (2008). Bacterial physiology, regulation and mutational adaptation in a chemostat environment. *Adv Microb Physiol* **53**: 169-229.
- Fernandez M., and Zuniga M. (2006). Amino acid catabolic pathways of lactic acid bacteria. *Crit Rev Microbiol* **32**: 155-83.
- Ferullo D. J., and Lovett S. T. (2008). The stringent response and cell cycle arrest in *Escherichia coli*. *PLoS Genet* **4**: e1000300.
- Fox A., Morgan S. L., Larsson L., and Odham G. (1990). "Chromatography and Mass Spectrometry," Plenum Press, New York.
- Franke T. A., and Wixforth A. (2008). Microfluidics for miniaturized laboratories on a chip. *Chemphyschem* **9**: 2140-56.

- Frees D., Vogensen F. K., and Ingmer H. (2003). Identification of proteins induced at low pH in *Lactococcus lactis*. *Int J Food Microbiol* **87**: 293-300.
- Freire S. L., and Wheeler A. R. (2006). Proteome-on-a-chip: mirage, or on the horizon? *Lab Chip* **6**: 1415-23.
- Friberg M., von Rohr P., and Gonnet G. (2004). Limitations of codon adaptation index and other coding DNA-based features for prediction of protein expression in *Saccharomyces cerevisiae*. *Yeast* **21**: 1083-93.
- Friboulet A., and Thomas D. (2005). Systems Biology-an interdisciplinary approach. *Biosens Bioelectron* **20**: 2404-7.

G

- Galvez A., Abriouel H., Lopez R. L., and Ben Omar N. (2007). Bacteriocin-based strategies for food biopreservation. *Int J Food Microbiol* **120**: 51-70.
- Ganesan B., Stuart M. R., and Weimer B. C. (2007). Carbohydrate starvation causes a metabolically active but nonculturable state in *Lactococcus lactis*. *Appl Environ Microbiol* **73**: 2498-512.
- Garault P., Letort C., Juillard V., and Monnet V. (2000). Branched-chain amino acid biosynthesis is essential for optimal growth of *Streptococcus thermophilus* in milk. *Appl Environ Microbiol* **66**: 5128-33.
- Garbis S., Lubec G., and Fountoulakis M. (2005). Limitations of current proteomics technologies. *J Chromatogr A* **1077**: 1-18.
- Garcia D. E., Baidoo E. E., Benke P. I., Pingitore F., Tang Y. J., Villa S., and Keasling J. D. (2008). Separation and mass spectrometry in microbial metabolomics. *Curr Opin Microbiol* **11**: 233-9.
- Garneau N. L., Wilusz J., and Wilusz C. J. (2007). The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* **8**: 113-26.
- Gasch A. P., and Werner-Washburne M. (2002). The genomics of yeast responses to environmental stress and starvation. *Funct Integr Genomics* **2**: 181-92.
- Gaudu P., Lamberet G., Poncet S., and Gruss A. (2003). CcpA regulation of aerobic and respiration growth in *Lactococcus lactis*. *Mol Microbiol* **50**: 183-92.
- Gelade R., Van de Velde S., Van Dijck P., and Thevelein J. M. (2003). Multi-level response of the yeast genome to glucose. *Genome Biol* **4**: 233.
- Gevaert K., Van Damme P., Ghesquiere B., Impens F., Martens L., Helsens K., and Vandekerckhove J. (2007). A la carte proteomics with an emphasis on gel-free techniques. *Proteomics* **7**: 2698-718.
- Gilchrist M. A., and Wagner A. (2006). A model of protein translation including codon bias, nonsense errors, and ribosome recycling. *J Theor Biol* **239**: 417-34.
- Giuliodori A. M., Gualerzi C. O., Soto S., Vila J., and Tavio M. M. (2007). Review on bacterial stress topics. *Ann N Y Acad Sci* **1113**: 95-104.
- Glisovic T., Bachorik J. L., Yong J., and Dreyfuss G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett* **582**: 1977-86.
- Godon J. J., Delorme C., Bardowski J., Chopin M. C., Ehrlich S. D., and Renault P. (1993). Gene inactivation in *Lactococcus lactis*: branched-chain amino acid biosynthesis. *J Bacteriol* **175**: 4383-90.
- Gomase V. S., Changbhale S. S., Patil S. A., and Kale K. V. (2008). Metabolomics. *Curr Drug Metab* **9**: 89-98.

- Gowrishankar J., and Harinarayanan R. (2004). Why is transcription coupled to translation in bacteria? *Mol Microbiol* **54**: 598-603.
- Gracey A. Y., and Cossins A. R. (2003). Application of microarray technology in environmental and comparative physiology. *Annu Rev Physiol* **65**: 231-59.
- Greenbaum D., Luscombe N. M., Jansen R., Qian J., and Gerstein M. (2001). Interrelating different types of genomic data, from proteome to secretome: 'oming in on function. *Genome Res* **11**: 1463-8.
- Griffin T. J., Gygi S. P., Ideker T., Rist B., Eng J., Hood L., and Aebersold R. (2002). Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Mol Cell Proteomics* **1**: 323-33.
- Grigull J., Mnaimneh S., Pootoolal J., Robinson M. D., and Hughes T. R. (2004). Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol Cell Biol* **24**: 5534-47.
- Gruber T. M., and Gross C. A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* **57**: 441-66.
- Grundy F. J., and Henkin T. M. (2006). From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Crit Rev Biochem Mol Biol* **41**: 329-38.
- Guedon E., Jamet E., and Renault P. (2002). Gene regulation in *Lactococcus lactis*: the gap between predicted and characterized regulators. *Antonie Van Leeuwenhoek* **82**: 93-112.
- Guedon E., Serror P., Ehrlich S. D., Renault P., and Delorme C. (2001). Pleiotropic transcriptional repressor CodY senses the intracellular pool of branched-chain amino acids in *Lactococcus lactis*. *Mol Microbiol* **40**: 1227-39.
- Guedon E., Sperandio B., Pons N., Ehrlich S. D., and Renault P. (2005). Overall control of nitrogen metabolism in *Lactococcus lactis* by CodY, and possible models for CodY regulation in Firmicutes. *Microbiology* **151**: 3895-909.
- Gyaneshwar P., Paliy O., McAuliffe J., Jones A., Jordan M. I., and Kustu S. (2005a). Lessons from *Escherichia coli* genes similarly regulated in response to nitrogen and sulfur limitation. *Proc Natl Acad Sci U S A* **102**: 3453-8.
- Gyaneshwar P., Paliy O., McAuliffe J., Popham D. L., Jordan M. I., and Kustu S. (2005b). Sulfur and nitrogen limitation in *Escherichia coli* K-12: specific homeostatic responses. *J Bacteriol* **187**: 1074-90.
- Gygi S. P., Rochon Y., Franza B. R., and Aebersold R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**: 1720-30.

H

- Hack C. J. (2004). Integrated transcriptome and proteome data: the challenges ahead. *Brief Funct Genomic Proteomic* **3**: 212-9.
- Hall H. K., Karem K. L., and Foster J. W. (1995). Molecular responses of microbes to environmental pH stress. *Adv Microb Physiol* **37**: 229-72.
- Hambraeus G., von Wachenfeldt C., and Hederstedt L. (2003). Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Mol Genet Genomics* **269**: 706-14.
- Harwood J. L. (2007). Temperature stress: reacting and adapting: lessons from poikilotherms. *Ann N Y Acad Sci* **1113**: 52-7.

- Haydock A. K., Porat I., Whitman W. B., and Leigh J. A. (2004). Continuous culture of *Methanococcus maripaludis* under defined nutrient conditions. *FEMS Microbiol Lett* **238**: 85-91.
- Hecker M., Antelmann H., Buttner K., and Bernhardt J. (2008). Gel-based proteomics of Gram-positive bacteria: a powerful tool to address physiological questions. *Proteomics* **8**: 4958-75.
- Hecker M., Pane-Farre J., and Volker U. (2007). SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria. *Annu Rev Microbiol* **61**: 215-36.
- Hecker M., and Volker U. (2001). General stress response of *Bacillus subtilis* and other bacteria. *Adv Microb Physiol* **44**: 35-91.
- Heller M. J. (2002). DNA microarray technology: devices, systems, and applications. *Annu Rev Biomed Eng* **4**: 129-53.
- Helmann J. D. (2002). The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* **46**: 47-110.
- Hengge-Aronis R. (2002). Recent insights into the general stress response regulatory network in *Escherichia coli*. *J Mol Microbiol Biotechnol* **4**: 341-6.
- Henkin T. M. (2008). Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev* **22**: 3383-90.
- Henriksen C. M., and Nilsson D. (2001). Redirection of pyruvate catabolism in *Lactococcus lactis* by selection of mutants with additional growth requirements. *Appl Microbiol Biotechnol* **56**: 767-75.
- Hoskisson P. A., and Hobbs G. (2005). Continuous culture--making a comeback? *Microbiology* **151**: 3153-9.

I

- Ideker T., Galitski T., and Hood L. (2001). A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet* **2**: 343-72.
- Inouye M., and Phadtare S. (2004). Cold shock response and adaptation at near-freezing temperature in microorganisms. *Sci STKE* **2004**: pe26.
- Irvine T. S. (1990). Fermentation: a practical approach. In "Laboratory fermentors" (B. McNeil, and L. M. Harvey, Eds.), pp. 17-38, Oxford University Press.
- Ishihama Y., Schmidt T., Rappsilber J., Mann M., Hartl F. U., Kerner M. J., and Frishman D. (2008). Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* **9**: 102.
- Ivanov Y. D., Govorun V. M., Bykov V. A., and Archakov A. I. (2006). Nanotechnologies in proteomics. *Proteomics* **6**: 1399-414.

J

- Jain V., Kumar M., and Chatterji D. (2006). ppGpp: stringent response and survival. *J Microbiol* **44**: 1-10.
- Jaluria P., Konstantopoulos K., Betenbaugh M., and Shiloach J. (2007). A perspective on microarrays: current applications, pitfalls, and potential uses. *Microb Cell Fact* **6**: 4.
- Jayapal K. P., Philp R. J., Kok Y. J., Yap M. G., Sherman D. H., Griffin T. J., and Hu W. S. (2008). Uncovering genes with divergent mRNA-protein dynamics in *Streptomyces coelicolor*. *PLoS ONE* **3**: e2097.

- Jensen O. N. (2006). Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol* **7**: 391-403.
- Johansson L., and Liden G. (2006). Transcriptome analysis of a shikimic acid producing strain of *Escherichia coli* W3110 grown under carbon- and phosphate-limited conditions. *J Biotechnol* **126**: 528-45.
- Joyce A. R., and Palsson B. O. (2006). The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* **7**: 198-210.
- Jurgen B., Tobisch S., Wumpelmann M., Gordes D., Koch A., Thurow K., Albrecht D., Hecker M., and Schweder T. (2005). Global expression profiling of *Bacillus subtilis* cells during industrial-close fed-batch fermentations with different nitrogen sources. *Biotechnol Bioeng* **92**: 277-98.

K

- Kaback H. R. (1990). Active transport: membrane vesicles, bioenergetics, molecules, and mechanisms. In "Bacterial energetics" (T. A. Krulwich, Ed.), pp. 151-202, Academic Press, INC., San Diego.
- Kaczanowska M., and Ryden-Aulin M. (2007). Ribosome biogenesis and the translation process in *Escherichia coli*. *Microbiol Mol Biol Rev* **71**: 477-94.
- Kalia A., and Gupta R. P. (2005). Proteomics: a paradigm shift. *Crit Rev Biotechnol* **25**: 173-98.
- Katagiri F., and Glazebrook J. (2009). Overview of mRNA expression profiling using DNA microarrays. *Curr Protoc Mol Biol* **Chapter 22**: Unit 22 4.
- Kaur I. P., Chopra K., and Saini A. (2002). Probiotics: potential pharmaceutical applications. *Eur J Pharm Sci* **15**: 1-9.
- Keller W. (1975). Determination of the number of superhelical turns in simian virus 40 DNA by gel electrophoresis. *Proc Natl Acad Sci U S A* **72**: 4876-80.
- Kilstrup M., Jacobsen S., Hammer K., and Vogensen F. K. (1997). Induction of heat shock proteins DnaK, GroEL, and GroES by salt stress in *Lactococcus lactis*. *Appl Environ Microbiol* **63**: 1826-37.
- Kilstrup M., and Martinussen J. (1998). A transcriptional activator, homologous to the *Bacillus subtilis* PurR repressor, is required for expression of purine biosynthetic genes in *Lactococcus lactis*. *J Bacteriol* **180**: 3907-16.
- Kirschner M. W. (2005). The meaning of systems biology. *Cell* **121**: 503-4.
- Kitano H. (2002a). Looking beyond the details: a rise in system-oriented approaches in genetics and molecular biology. *Curr Genet* **41**: 1-10.
- Kitano H. (2002b). Systems biology: a brief overview. *Science* **295**: 1662-4.
- Koebmann B. J., Nilsson D., Kuipers O. P., and Jensen P. R. (2000). The membrane-bound H(+)-ATPase complex is essential for growth of *Lactococcus lactis*. *J Bacteriol* **182**: 4738-43.
- Kok J., Buist G., Zomer A. L., van Hijum S. A., and Kuipers O. P. (2005). Comparative and functional genomics of lactococci. *FEMS Microbiol Rev* **29**: 411-33.
- Kolkman A., Daran-Lapujade P., Fullaondo A., Olsthoorn M. M., Pronk J. T., Slijper M., and Heck A. J. (2006). Proteome analysis of yeast response to various nutrient limitations. *Mol Syst Biol* **2**: 2006 0026.
- Kowalczyk M., and Bardowski J. (2007). Regulation of sugar catabolism in *Lactococcus lactis*. *Crit Rev Microbiol* **33**: 1-13.

- Kozak M. (2007). Some thoughts about translational regulation: forward and backward glances. *J Cell Biochem* **102**: 280-90.
- Kushner S. R. (2004). mRNA decay in prokaryotes and eukaryotes: different approaches to a similar problem. *IUBMB Life* **56**: 585-94.

L

- Lambert J. P., Ethier M., Smith J. C., and Figeys D. (2005). Proteomics: from gel based to gel free. *Anal Chem* **77**: 3771-87.
- Lange C., Zaigler A., Hammelmann M., Twellmeyer J., Raddatz G., Schuster S. C., Oesterhelt D., and Soppa J. (2007). Genome-wide analysis of growth phase-dependent translational and transcriptional regulation in halophilic archaea. *BMC Genomics* **8**: 415.
- Larsen R., Kloosterman T. G., Kok J., and Kuipers O. P. (2006). GlnR-mediated regulation of nitrogen metabolism in *Lactococcus lactis*. *J Bacteriol* **188**: 4978-82.
- Levesque M. P., and Benfey P. N. (2004). Systems biology. *Curr Biol* **14**: R179-80.
- Li Y., Hugenholtz J., Abee T., and Molenaar D. (2003). Glutathione protects *Lactococcus lactis* against oxidative stress. *Appl Environ Microbiol* **69**: 5739-45.
- Lithwick G., and Margalit H. (2003). Hierarchy of sequence-dependent features associated with prokaryotic translation. *Genome Res* **13**: 2665-73.
- Liu E. T. (2005). Systems biology, integrative biology, predictive biology. *Cell* **121**: 505-6.
- Lopez-Garcia P., and Forterre P. (2000). DNA topology and the thermal stress response, a tale from mesophiles and hyperthermophiles. *Bioessays* **22**: 738-46.
- Lovmar M., and Ehrenberg M. (2006). Rate, accuracy and cost of ribosomes in bacterial cells. *Biochimie* **88**: 951-61.
- Lovrien R., and Matulis D. (2005). Assays for total protein. *Curr Protoc Microbiol* **Appendix 3**: Appendix 3A.
- Lowry O. H., Rosebrough N. J., Farr A. L., and Randall R. J. (1951). Protein measurement with the Folin phenol reagent. *J Biol Chem* **193**: 265-75.
- Lu P., Vogel C., Wang R., Yao X., and Marcotte E. M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25**: 117-24.
- Luesink E. J., van Herpen R. E., Grossiord B. P., Kuipers O. P., and de Vos W. M. (1998). Transcriptional activation of the glycolytic *las* operon and catabolite repression of the *gal* operon in *Lactococcus lactis* are mediated by the catabolite control protein CcpA. *Mol Microbiol* **30**: 789-98.
- Lushchak V. I. (2001). Oxidative stress and mechanisms of protection against it in bacteria. *Biochemistry (Mosc)* **66**: 476-89.

M

- Mader U., Antelmann H., Buder T., Dahl M. K., Hecker M., and Homuth G. (2002a). *Bacillus subtilis* functional genomics: genome-wide analysis of the DegS-DegU regulon by transcriptomics and proteomics. *Mol Genet Genomics* **268**: 455-67.
- Mader U., Homuth G., Scharf C., Buttner K., Bode R., and Hecker M. (2002b). Transcriptome and proteome analysis of *Bacillus subtilis* gene expression modulated by amino acid availability. *J Bacteriol* **184**: 4288-95.

- Madrid R. E., and Felice C. J. (2005). Microbial biomass estimation. *Crit Rev Biotechnol* **25**: 97-112.
- Mandruzzato S. (2007). Technological platforms for microarray gene expression profiling. *Adv Exp Med Biol* **593**: 12-8.
- Mao C., Evans C., Jensen R. V., and Sobral B. W. (2008). Identification of new genes in *Sinorhizobium meliloti* using the Genome Sequencer FLX system. *BMC Microbiol* **8**: 72.
- Marles-Wright J., and Lewis R. J. (2007). Stress responses of bacteria. *Curr Opin Struct Biol* **17**: 755-60.
- Mata J., Marguerat S., and Bahler J. (2005). Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci* **30**: 506-14.
- McDougald D., Gong L., Srinivasan S., Hild E., Thompson L., Takayama K., Rice S. A., and Kjelleberg S. (2002). Defences against oxidative stress during starvation in bacteria. *Antonie Van Leeuwenhoek* **81**: 3-13.
- Mehra A., and Hatzimanikatis V. (2006). An algorithmic framework for genome-wide modeling and analysis of translation networks. *Biophys J* **90**: 1136-46.
- Mehra A., Lee K. H., and Hatzimanikatis V. (2003). Insights into the relation between mRNA and protein expression patterns: I. Theoretical considerations. *Biotechnol Bioeng* **84**: 822-33.
- Melamed D., and Arava Y. (2007). Genome-wide analysis of mRNA polysomal profiles with spotted DNA microarrays. *Methods Enzymol* **431**: 177-201.
- Melamed D., Pnueli L., and Arava Y. (2008). Yeast translational response to high salinity: global analysis reveals regulation at multiple levels. *Rna* **14**: 1337-51.
- Mercade M., Lindley N. D., and Loubiere P. (2000). Metabolism of *Lactococcus lactis* subsp. *cremoris* MG 1363 in acid stress conditions. *Int J Food Microbiol* **55**: 161-5.
- Meyer S., Temme C., and Wahle E. (2004). Messenger RNA turnover in eukaryotes: pathways and enzymes. *Crit Rev Biochem Mol Biol* **39**: 197-216.
- Mitarai N., Sneppen K., and Pedersen S. (2008). Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. *J Mol Biol* **382**: 236-45.
- Mitchell P., and Moyle J. (1967). Chemiosmotic hypothesis of oxidative phosphorylation. *Nature* **213**: 137-9.
- Miyoshi A., Rochat T., Gratadoux J. J., Le Loir Y., Oliveira S. C., Langella P., and Azevedo V. (2003). Oxidative stress in *Lactococcus lactis*. *Genet Mol Res* **2**: 348-59.
- Mogk A., Schmidt R., and Bukau B. (2007). The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. *Trends Cell Biol* **17**: 165-72.
- Mohanty B. K., and Kushner S. R. (1999). Analysis of the function of *Escherichia coli* poly(A) polymerase I in RNA metabolism. *Mol Microbiol* **34**: 1094-108.
- Moritz B., Striegel K., de Graaf A. A., and Sahn H. (2002). Changes of pentose phosphate pathway flux in vivo in *Corynebacterium glutamicum* during leucine-limited batch cultivation as determined from intracellular metabolite concentration measurements. *Metab Eng* **4**: 295-305.
- Mostertz J., Scharf C., Hecker M., and Homuth G. (2004). Transcriptome and proteome analysis of *Bacillus subtilis* gene expression in response to superoxide and peroxide stress. *Microbiology* **150**: 497-512.

Murata S., Yashiroda H., and Tanaka K. (2009). Molecular mechanisms of proteasome assembly. *Nat Rev Mol Cell Biol* **10**: 104-15.

Mykytczuk N. C., Trevors J. T., Leduc L. G., and Ferroni G. D. (2007). Fluorescence polarization in studies of bacterial cytoplasmic membrane fluidity under environmental stress. *Prog Biophys Mol Biol* **95**: 60-82.

N

Nie L., Wu G., Culley D. E., Scholten J. C., and Zhang W. (2007). Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. *Crit Rev Biotechnol* **27**: 63-75.

Nie L., Wu G., and Zhang W. (2006a). Correlation between mRNA and protein abundance in *Desulfovibrio vulgaris*: a multiple regression to identify sources of variations. *Biochem Biophys Res Commun* **339**: 603-10.

Nie L., Wu G., and Zhang W. (2006b). Correlation of mRNA expression and protein abundance affected by multiple sequence features related to translational efficiency in *Desulfovibrio vulgaris*: a quantitative analysis. *Genetics* **174**: 2229-43.

Nie L., Wu G., and Zhang W. (2008). Statistical application and challenges in global gel-free proteomic analysis by mass spectrometry. *Crit Rev Biotechnol* **28**: 297-307.

Nouaille S., Ribeiro L. A., Miyoshi A., Pontes D., Le Loir Y., Oliveira S. C., Langella P., and Azevedo V. (2003). Heterologous protein production and delivery systems for *Lactococcus lactis*. *Genet Mol Res* **2**: 102-11.

O

O'Connell-Motherway M., van Sinderen D., Morel-Deville F., Fitzgerald G. F., Ehrlich S. D., and Morel P. (2000). Six putative two-component regulatory systems isolated from *Lactococcus lactis* subsp. *cremoris* MG1363. *Microbiology* **146 (Pt 4)**: 935-47.

Obis D., Guillot A., and Mistou M. Y. (2001). Tolerance to high osmolality of *Lactococcus lactis* subsp. *lactis* and *cremoris* is related to the activity of a betaine transport system. *FEMS Microbiol Lett* **202**: 39-44.

Ogura T., and Wilkinson A. J. (2001). AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells* **6**: 575-97.

Ohara H., Owaki M., and Sonomoto K. (2007). Calculation of metabolic flow of xylose in *Lactococcus lactis*. *J Biosci Bioeng* **103**: 92-4.

Oliver J. D. (1993). Formation of viable but nonculturable cells. In "Starvation in bacteria" (S. Kjelleberg, Ed.), pp. 239-272, Plenum Press, New York.

Oliver S. G., Winson M. K., Kell D. B., and Baganz F. (1998). Systematic functional analysis of the yeast genome. *Trends Biotechnol* **16**: 373-8.

P

Pan Z., and Raftery D. (2007). Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal Bioanal Chem* **387**: 525-7.

Panoff J. M., Legrand F., Thammavongs B., and Boutibonnes P. (1994). The cold shock response in *Lactococcus lactis* subsp. *lactis*. *Current Microbiology* **29**: 213-216.

Pedersen M. B., Garrigues C., Tuphile K., Brun C., Vido K., Bennedsen M., Mollgaard H., Gaudu P., and Gruss A. (2008). Impact of aeration and heme-activated respiration on *Lactococcus lactis* gene expression: identification of a heme-responsive operon. *J Bacteriol* **190**: 4903-11.

- Peeters J. K., and Van der Spek P. J. (2005). Growing applications and advancements in microarray technology and analysis tools. *Cell Biochem Biophys* **43**: 149-66.
- Piper M. D., Daran-Lapujade P., Bro C., Regenber B., Knudsen S., Nielsen J., and Pronk J. T. (2002). Reproducibility of oligonucleotide microarray transcriptome analyses. An interlaboratory comparison using chemostat cultures of *Saccharomyces cerevisiae*. *J Biol Chem* **277**: 37001-8.
- Pisitkun T., Hoffert J. D., Yu M. J., and Knepper M. A. (2007). Tandem mass spectrometry in physiology. *Physiology (Bethesda)* **22**: 390-400.
- Pitt W. M., Harden T. J., and Hull R. R. (2000). Behavior of *Listeria monocytogenes* in pasteurized milk during fermentation with lactic acid bacteria. *J Food Prot* **63**: 916-20.
- Poetsch A., and Wolters D. (2008). Bacterial membrane proteomics. *Proteomics* **8**: 4100-22.
- Poole R. K. (2005). Nitric oxide and nitrosative stress tolerance in bacteria. *Biochem Soc Trans* **33**: 176-80.
- Poolman B., and Konings W. N. (1988). Relation of growth of *Streptococcus lactis* and *Streptococcus cremoris* to amino acid transport. *J Bacteriol* **170**: 700-7.
- Pratt J. M., Petty J., Riba-Garcia I., Robertson D. H., Gaskell S. J., Oliver S. G., and Beynon R. J. (2002). Dynamics of protein turnover, a missing dimension in proteomics. *Mol Cell Proteomics* **1**: 579-91.

R

- Ragoonanan V., Malsam J., Bond D. R., and Aksan A. (2008). Roles of membrane structure and phase transition on the hyperosmotic stress survival of *Geobacter sulfurreducens*. *Biochim Biophys Acta* **1778**: 2283-90.
- Rallu F., Gruss A., and Maguin E. (1996). *Lactococcus lactis* and stress. *Antonie Van Leeuwenhoek* **70**: 243-51.
- Ramanathan A., and Schreiber S. L. (2007). Multilevel regulation of growth rate in yeast revealed using systems biology. *J Biol* **6**: 3.
- Rauhut R., and Klug G. (1999). mRNA degradation in bacteria. *FEMS Microbiol Rev* **23**: 353-70.
- Raynaud S., Perrin R., Cocaign-Bousquet M., and Loubiere P. (2005). Metabolic and transcriptomic adaptation of *Lactococcus lactis* subsp. *lactis* Biovar *diacetylactis* in response to autoacidification and temperature downshift in skim milk. *Appl Environ Microbiol* **71**: 8016-23.
- Record M. T., Jr., Courtenay E. S., Cayley D. S., and Guttman H. J. (1998). Responses of *E. coli* to osmotic stress: large changes in amounts of cytoplasmic solutes and water. *Trends Biochem Sci* **23**: 143-8.
- Redon E., Loubiere P., and Cocaign-Bousquet M. (2005a). Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. *J Biol Chem* **280**: 36380-5.
- Redon E., Loubiere P., and Cocaign-Bousquet M. (2005b). Transcriptome analysis of the progressive adaptation of *Lactococcus lactis* to carbon starvation. *J Bacteriol* **187**: 3589-92.
- Regenber B., Grotkjaer T., Winther O., Fausboll A., Akesson M., Bro C., Hansen L. K., Brunak S., and Nielsen J. (2006). Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*. *Genome Biol* **7**: R107.

- Rezaiki L., Cesselin B., Yamamoto Y., Vido K., van West E., Gaudu P., and Gruss A. (2004). Respiration metabolism reduces oxidative and acid stress to improve long-term survival of *Lactococcus lactis*. *Mol Microbiol* **53**: 1331-42.
- Rocha E. P., Danchin A., and Viari A. (1999). Translation in *Bacillus subtilis*: roles and trends of initiation and termination, insights from a genome analysis. *Nucleic Acids Res* **27**: 3567-76.
- Roe M. R., and Griffin T. J. (2006). Gel-free mass spectrometry-based high throughput proteomics: tools for studying biological response of proteins and proteomes. *Proteomics* **6**: 4678-87.
- Romeo Y., Obis D., Bouvier J., Guillot A., Fourcans A., Bouvier I., Gutierrez C., and Mistou M. Y. (2003). Osmoregulation in *Lactococcus lactis*: BusR, a transcriptional repressor of the glycine betaine uptake system BusA. *Mol Microbiol* **47**: 1135-47.
- Ruas-Madiedo P., Hernandez-Barranco A., Margolles A., and de los Reyes-Gavilan C. G. (2005). A bile salt-resistant derivative of *Bifidobacterium animalis* has an altered fermentation pattern when grown on glucose and maltose. *Appl Environ Microbiol* **71**: 6564-70.
- Ryu Y. G., Kim E. S., Kim D. W., Kim S. K., and Lee K. J. (2007). Differential stringent responses of *Streptomyces coelicolor* M600 to starvation of specific nutrients. *J Microbiol Biotechnol* **17**: 305-12.

S

- Sampathkumar B., Napper S., Carrillo C. D., Willson P., Taboada E., Nash J. H., Potter A. A., Babiuk L. A., and Allan B. J. (2006). Transcriptional and translational expression patterns associated with immobilized growth of *Campylobacter jejuni*. *Microbiology* **152**: 567-77.
- Sanchez-Diaz P., and Penalva L. O. (2006). Post-transcription meets post-genomic: the saga of RNA binding proteins in a new era. *RNA Biol* **3**: 101-9.
- Savijoki K., Ingmer H., and Varmanen P. (2006). Proteolytic systems of lactic acid bacteria. *Appl Microbiol Biotechnol* **71**: 394-406.
- Schena M., Shalon D., Davis R. W., and Brown P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467-70.
- Schimel J., Balsler T. C., and Wallenstein M. (2007). Microbial stress-response physiology and its implications for ecosystem function. *Ecology* **88**: 1386-94.
- Schmid M., and Jensen T. H. (2008). The exosome: a multipurpose RNA-decay machine. *Trends Biochem Sci* **33**: 501-10.
- Schmidt M. W., Houseman A., Ivanov A. R., and Wolf D. A. (2007). Comparative proteomic and transcriptomic profiling of the fission yeast *Schizosaccharomyces pombe*. *Mol Syst Biol* **3**: 79.
- Schumann W. (2003). The *Bacillus subtilis* heat shock stimulon. *Cell Stress Chaperones* **8**: 207-17.
- Schweder T., and Hecker M. (2004). Monitoring of stress responses. *Adv Biochem Eng Biotechnol* **89**: 47-71.
- Selinger D. W., Saxena R. M., Cheung K. J., Church G. M., and Rosenow C. (2003). Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res* **13**: 216-23.

- Shalem O., Dahan O., Levo M., Martinez M. R., Furman I., Segal E., and Pilpel Y. (2008). Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol Syst Biol* **4**: 223.
- Sharp J. S., and Bechhofer D. H. (2005). Effect of 5'-proximal elements on decay of a model mRNA in *Bacillus subtilis*. *Mol Microbiol* **57**: 484-95.
- Sharp P. M., and Li W. H. (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**: 1281-95.
- Shinagawa H. (1996). SOS response as an adaptive response to DNA damage in prokaryotes. *Exs* **77**: 221-35.
- Shyu A. B., Wilkinson M. F., and van Hoof A. (2008). Messenger RNA regulation: to translate or to degrade. *Embo J* **27**: 471-81.
- Sikic Pogacar M., Rubesa Mihaljevic R., Klancnik A., Brumini G., Abram M., and Smole Mozina S. (2009). Survival of stress exposed *Campylobacter jejuni* in the murine macrophage J774 cell line. *Int J Food Microbiol* **129**: 68-73.
- Silberbach M., and Burkovski A. (2006). Application of global analysis techniques to *Corynebacterium glutamicum*: new insights into nitrogen regulation. *J Biotechnol* **126**: 101-10.
- Silberbach M., Schafer M., Huser A. T., Kalinowski J., Puhler A., Kramer R., and Burkovski A. (2005). Adaptation of *Corynebacterium glutamicum* to ammonium limitation: a global analysis using transcriptome and proteome techniques. *Appl Environ Microbiol* **71**: 2391-402.
- Simonetti A., Marzi S., Myasnikov A. G., Fabbretti A., Yusupov M., Gualerzi C. O., and Klaholz B. P. (2008). Structure of the 30S translation initiation complex. *Nature* **455**: 416-20.
- Soufi B., Gnad F., Jensen P. R., Petranovic D., Mann M., Mijakovic I., and Macek B. (2008). The Ser/Thr/Tyr phosphoproteome of *Lactococcus lactis* IL1403 reveals multiply phosphorylated proteins. *Proteomics* **8**: 3486-93.
- Spector M. P. (1998). The starvation-stress response (SSR) of *Salmonella*. *Adv Microb Physiol* **40**: 233-79.
- Srivatsan A., and Wang J. D. (2008). Control of bacterial transcription, translation and replication by (p)ppGpp. *Curr Opin Microbiol* **11**: 100-5.
- Storey J. D. (2002). A direct approach to false discovery rates. *J. R. Statist.* **64(B):Part3**: 479-498.
- Sugimoto S., Abdullah Al M., and Sonomoto K. (2008). Molecular chaperones in lactic acid bacteria: physiological consequences and biochemical properties. *J Biosci Bioeng* **106**: 324-36.
- Sunohara T., Jojima K., Tagami H., Inada T., and Aiba H. (2004). Ribosome stalling during translation elongation induces cleavage of mRNA being translated in *Escherichia coli*. *J Biol Chem* **279**: 15368-75.
- Swain J. F., and Gierasch L. M. (2006). The changing landscape of protein allostery. *Curr Opin Struct Biol* **16**: 102-8.
- Switzer R. L., Turner R. J., and Lu Y. (1999). Regulation of the *Bacillus subtilis* pyrimidine biosynthetic operon by transcriptional attenuation: control of gene expression by an mRNA-binding protein. *Prog Nucleic Acid Res Mol Biol* **62**: 329-67.

T

- Tai S. L., Daran-Lapujade P., Walsh M. C., Pronk J. T., and Daran J. M. (2007). Acclimation of *Saccharomyces cerevisiae* to low temperature: a chemostat-based transcriptome analysis. *Mol Biol Cell* **18**: 5100-12.
- Takayama K., and Kjelleberg S. (2000). The role of RNA stability during bacterial stress responses and starvation. *Environ Microbiol* **2**: 355-65.
- Tam le T., Eymann C., Antelmann H., Albrecht D., and Hecker M. (2007). Global gene expression profiling of *Bacillus subtilis* in response to ammonium and tryptophan starvation as revealed by transcriptome and proteome analysis. *J Mol Microbiol Biotechnol* **12**: 121-30.
- Tanaka K. (2009). The proteasome: overview of structure and functions. *Proc Jpn Acad Ser B Phys Biol Sci* **85**: 12-36.
- Tenenbaum S. A., Carson C. C., Lager P. J., and Keene J. D. (2000). Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proc Natl Acad Sci U S A* **97**: 14085-90.
- ter Kuile B. H., and Westerhoff H. V. (2001). Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett* **500**: 169-71.
- Titgemeyer F., and Hillen W. (2002). Global control of sugar metabolism: a gram-positive solution. *Antonie Van Leeuwenhoek* **82**: 59-71.
- Traxler M. F., Summers S. M., Nguyen H. T., Zacharia V. M., Hightower G. A., Smith J. T., and Conway T. (2008). The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol Microbiol* **68**: 1128-48.
- Tremaroli V., Fedi S., and Zannoni D. (2007). Evidence for a tellurite-dependent generation of reactive oxygen species and absence of a tellurite-mediated adaptive response to oxidative stress in cells of *Pseudomonas pseudoalcaligenes* KF707. *Arch Microbiol* **187**: 127-35.
- Tremaroli V., Workentine M. L., Weljie A. M., Vogel H. J., Ceri H., Viti C., Tatti E., Zhang P., Hynes A. P., Turner R. J., and Zannoni D. (2009). Metabolomic investigation of the bacterial response to a metal challenge. *Appl Environ Microbiol* **75**: 719-28.
- Tuller T., Kupiec M., and Ruppin E. (2007). Determinants of protein abundance and translation efficiency in *S. cerevisiae*. *PLoS Comput Biol* **3**: e248.
- Turner M. S., Tan Y. P., and Giffard P. M. (2007). Inactivation of an iron transporter in *Lactococcus lactis* results in resistance to tellurite and oxidative stress. *Appl Environ Microbiol* **73**: 6144-9.

V

- Valleron A. J. (2007). Chapitre 10. Analyse de la liaison entre deux variables. In "Biostatistique" (Flammarion, Ed.), pp. 120-142, Paris.
- van de Guchte M., Serror P., Chervaux C., Smokvina T., Ehrlich S. D., and Maguin E. (2002). Stress responses in lactic acid bacteria. *Antonie Van Leeuwenhoek* **82**: 187-216.
- van Gend C., and Snoep J. L. (2008). Systems biology model databases and resources. *Essays Biochem* **45**: 223-36.
- Varmanen P., Ingmer H., and Vogensen F. K. (2000). *ctsR* of *Lactococcus lactis* encodes a negative regulator of *clp* gene expression. *Microbiology* **146** (Pt 6): 1447-55.
- Varshavsky A. (1996). The N-end rule: functions, mysteries, uses. *Proc Natl Acad Sci U S A* **93**: 12142-9.

- Velculescu V. E., Zhang L., Vogelstein B., and Kinzler K. W. (1995). Serial analysis of gene expression. *Science* **270**: 484-7.
- Velculescu V. E., Zhang L., Zhou W., Vogelstein J., Basrai M. A., Bassett D. E., Jr., Hieter P., Vogelstein B., and Kinzler K. W. (1997). Characterization of the yeast transcriptome. *Cell* **88**: 243-51.
- Vido K., Diemer H., Van Dorsselaer A., Leize E., Juillard V., Gruss A., and Gaudu P. (2005). Roles of thioredoxin reductase during the aerobic life of *Lactococcus lactis*. *J Bacteriol* **187**: 601-10.

W

- Wang Y., Liu C. L., Storey J. D., Tibshirani R. J., Herschlag D., and Brown P. O. (2002). Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci U S A* **99**: 5860-5.
- Washburn M. P., Koller A., Oshiro G., Ulaszek R. R., Plouffe D., Deciu C., Winzeler E., and Yates J. R., 3rd (2003). Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **100**: 3107-12.
- Wasinger V. C., Cordwell S. J., Cerpa-Poljak A., Yan J. X., Gooley A. A., Wilkins M. R., Duncan M. W., Harris R., Williams K. L., and Humphery-Smith I. (1995). Progress with gene-product mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis* **16**: 1090-4.
- Wassarman K. M. (2002). Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell* **109**: 141-4.
- Weber H., Polen T., Heuveling J., Wendisch V. F., and Hengge R. (2005). Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. *J Bacteriol* **187**: 1591-603.
- Weber M. H., and Marahiel M. A. (2003). Bacterial cold shock responses. *Sci Prog* **86**: 9-75.
- Wheelan S. J., Martinez Murillo F., and Boeke J. D. (2008). The incredible shrinking world of DNA microarrays. *Mol Biosyst* **4**: 726-32.
- Whitaker R. D., and Batt C. A. (1991). Characterization of the Heat Shock Response in *Lactococcus lactis* subsp. *lactis*. *Appl Environ Microbiol* **57**: 1408-1412.
- Williamson M. P. (2005). Systems biology: will it work? *Biochem Soc Trans* **33**: 503-6.
- Wold B., and Myers R. M. (2008). Sequence census methods for functional genomics. *Nat Methods* **5**: 19-21.
- Wolff S., Otto A., Albrecht D., Zeng J. S., Buttner K., Gluckmann M., Hecker M., and Becher D. (2006). Gel-free and gel-based proteomics in *Bacillus subtilis*: a comparative study. *Mol Cell Proteomics* **5**: 1183-92.
- Wouters J. A., Frenkiel H., de Vos W. M., Kuipers O. P., and Abee T. (2001). Cold shock proteins of *Lactococcus lactis* MG1363 are involved in cryoprotection and in the production of cold-induced proteins. *Appl Environ Microbiol* **67**: 5171-8.
- Wouters J. A., Kamphuis H. H., Hugenholtz J., Kuipers O. P., de Vos W. M., and Abee T. (2000). Changes in glycolytic activity of *Lactococcus lactis* induced by low temperature. *Appl Environ Microbiol* **66**: 3686-91.
- Wouters J. A., Sanders J. W., Kok J., de Vos W. M., Kuipers O. P., and Abee T. (1998). Clustered organization and transcriptional analysis of a family of five csp genes of *Lactococcus lactis* MG1363. *Microbiology* **144** (Pt 10): 2885-93.

- Wu G., Nie L., and Zhang W. (2008). Integrative analyses of posttranscriptional regulation in the yeast *Saccharomyces cerevisiae* using transcriptomic and proteomic data. *Curr Microbiol* **57**: 18-22.
- Wydau S., Dervyn R., Anba J., Dusko Ehrlich S., and Maguin E. (2006). Conservation of key elements of natural competence in *Lactococcus lactis* ssp. *FEMS Microbiol Lett* **257**: 32-42.

X

- Xie Y., Chou L. S., Cutler A., and Weimer B. (2004). DNA Macroarray profiling of *Lactococcus lactis* subsp. *lactis* IL1403 gene expression during environmental stresses. *Appl Environ Microbiol* **70**: 6738-47.

Y

- Yarmush M. L., and Jayaraman A. (2002). Advances in proteomic technologies. *Annu Rev Biomed Eng* **4**: 349-73.
- Yoon S. H., Han M. J., Lee S. Y., Jeong K. J., and Yoo J. S. (2003). Combined transcriptome and proteome analysis of *Escherichia coli* during high cell density culture. *Biotechnol Bioeng* **81**: 753-67.
- Yoshida H., Yamamoto H., Uchiumi T., and Wada A. (2004). RMF inactivates ribosomes by covering the peptidyl transferase centre and entrance of peptide exit tunnel. *Genes Cells* **9**: 271-8.
- Yoshida K., Kobayashi K., Miwa Y., Kang C. M., Matsunaga M., Yamaguchi H., Tojo S., Yamamoto M., Nishi R., Ogasawara N., Nakayama T., and Fujita Y. (2001). Combined transcriptome and proteome analysis as a powerful approach to study genes under glucose repression in *Bacillus subtilis*. *Nucleic Acids Res* **29**: 683-92.

Z

- Zhang J., Fu R. Y., Hugenholtz J., Li Y., and Chen J. (2007). Glutathione protects *Lactococcus lactis* against acid stress. *Appl Environ Microbiol* **73**: 5268-75.
- Zhang Z., Qu Y., Zhang X., and Lin J. (2008). Effects of oxygen limitation on xylose fermentation, intracellular metabolites, and key enzymes of *Neurospora crassa* AS3.1602. *Appl Biochem Biotechnol* **145**: 39-51.
- Zomer A. L., Buist G., Larsen R., Kok J., and Kuipers O. P. (2007). Time-resolved determination of the CcpA regulon of *Lactococcus lactis* subsp. *cremoris* MG1363. *J Bacteriol* **189**: 1366-81.
- Zouridis H., and Hatzimanikatis V. (2007). A model for protein translation: polysome self-organization leads to maximum protein synthesis rates. *Biophys J* **92**: 717-30.
- Zouridis H., and Hatzimanikatis V. (2008). Effects of codon distributions and tRNA competition on protein translation. *Biophys J* **95**: 1018-33.

Publications et communications

Publications

Dressaire C., Redon E., Milhem H., Besse P., Loubière P., Coccagn-Bousquet M. (2008): Growth rate regulated genes and their wide involvement in the *Lactococcus lactis* stress responses *BMC Genomics* 2008, **9**:343

Communications par affiches

Dressaire C., Redon E., Raynaud S., Loubière P., Coccagn-Bousquet M. (2007) Etude de l'adaptation de *Lactococcus lactis* aux stress par une approche génomique intégrative. 7^{ième} congrès de la Société Française de Microbiologie, Nantes, mai 2007

Dressaire C., Gitton C., Milhem H., Laurent-Bonneau B., Besse P., Monnet V., Loubière P., Coccagn-Bousquet M. (2008) Proteome *versus* transcriptome: correlations study in *Lactococcus lactis*. 9th Symposium on Lactic Acid Bacteria, Zuiderduin, Pays-Bas, août 2008

Picard F., **Dressaire C.**, Loubière P., Girbal L., Coccagn-Bousquet M. (2009) : Études des régulations post-transcriptionnelles chez *Lactococcus lactis* en condition de stress. 16^{ième} Colloque du Club des Bactéries Lactiques, Toulouse, mai 2009

Communications orales

Dressaire C. (2007) : Effet du taux de croissance chez *Lactococcus lactis* : lien avec les réponses au stress. 15^{ième} Colloque du Club des Bactéries Lactiques, Rennes, novembre 2007

Dressaire C. (2008) : Transcriptome *versus* protéome, un examen des régulations traductionnelles par une approche de biologie intégrative chez *Lactococcus lactis*. Séminaire du pôle Physiologie et Métabolisme Microbien, LISBP, Toulouse, mars 2008

Dressaire C. (2009) : Intégration des données de protéome et de transcriptome par une approche de modélisation. 16^{ième} Colloque du Club des Bactéries Lactiques, Toulouse, mai 2009

Annexes

**Annexe I : Modulation de la stabilité des ARN messagers
de *Lactococcus lactis* par le taux de croissance**

Au cours des différents chapitres présentés dans ce manuscrit, nous nous sommes intéressés à la régulation des niveaux de protéines en lien avec l'adaptation de *L. lactis* à son environnement. Ainsi, nous avons démontré que les processus de traduction, de dilution et dégradation des protéines ainsi que leurs régulations sont impliqués dans le contrôle des concentrations intracellulaires en protéines. L'expression des potentialités du génome requiert néanmoins une étape de transcription des gènes en ARN messagers en amont de leur traduction en protéines. Les concentrations en transcrits ne sont pas apparues comme les paramètres les plus déterminants pour les niveaux de protéines mais leur influence n'est pas non plus négligeable et se place au second rang après le CAI (cf. chapitre IV). De fait, tout mécanisme qui régule les concentrations en ARN messager va indirectement influencer celles des protéines. Afin d'améliorer la compréhension globale de l'adaptation de *L. lactis* à son environnement, nous avons donc élargi notre approche à l'étude des mécanismes de contrôle des niveaux de transcrits.

La transcription n'est pas le seul processus cellulaire qui affecte les concentrations en ARN messagers car, comme pour les protéines, la dilution par la croissance et la dégradation sont également probablement impliquées (figure 12). D'une manière générale dans les analyses transcriptomiques, les variations des niveaux d'ARN messagers sont imputées à des régulations de la transcription et les processus de dilution et de dégradation sont négligés. Le processus de dégradation est d'ailleurs souvent abordé d'un point de vue mécanistique (Carpousis, 2007; Garneau *et al.*, 2007; Rauhut & Klug, 1999; Schmid & Jensen, 2008) et peu d'études de la stabilité des ARN messagers à l'échelle du génome sont disponibles dans la littérature. Chez la levure modèle *S. cerevisiae*, les temps de demi-vie varient très fortement d'un transcrit à l'autre et sont compris entre 3 et une centaine de minutes avec une valeur moyenne de 23 min (Wang *et al.*, 2002). Des valeurs moyennes plus courtes que pour la levure, ne dépassant pas les 10 min ont été reportées pour les modèles bactériens *E. coli* (Bernstein *et al.*, 2002; Selinger *et al.*, 2003) et *B. subtilis* (Hambraeus *et al.*, 2003).

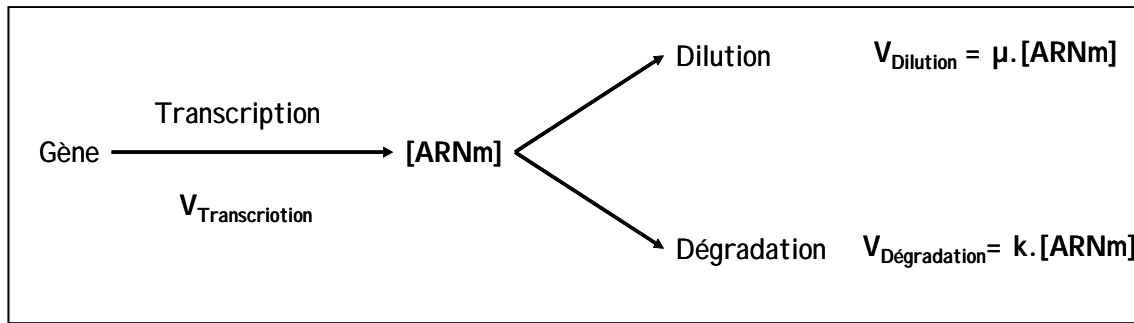


Figure 12 : Contrôle de la concentration intracellulaire en ARN messager par la transcription, la dégradation et la dilution.

μ : taux de croissance (h^{-1}), k : constante de dégradation (min^{-1})

La modulation de la stabilité des ARN messagers est potentiellement impliquée dans le contrôle de l'expression des gènes (Takayama & Kjelleberg, 2000). Cet aspect n'est que très peu étudié à l'échelle globale. Il a pourtant été précédemment démontré que la stabilité des transcrits de *L. lactis* était significativement augmentée en réponse à la carence en carbone (Redon *et al.*, 2005). Un temps de demi-vie moyen égal à 5.8 min, similaire à ceux de *B. subtilis* et *E. coli*, a en effet été obtenu en phase exponentielle de croissance et augmente à 19.4 min lors de la carence. Les travaux de thèse d'Emma Redon (2005) ont également mis en évidence une stabilisation globale des transcrits de *L. lactis*, mais de moindre amplitude, lors de son adaptation à la carence en isoleucine. Plus récemment, les travaux de Shalem *et al.* (2008) ont mis en évidence une modulation similaire de la dégradation des ARN messagers dans les réponses au stress de la levure *S. cerevisiae* : dans le cas d'un stress chimique causant des dommages à l'ADN, les messagers les plus induits sont les plus stabilisés tandis que, dans le cas d'un stress oxydatif, les auteurs ont montré que les gènes les plus induits sont déstabilisés, suggérant que la dégradation contrebalance l'effet de la transcription.

La modulation de la stabilité des ARN messagers semblent donc pleinement impliquée dans l'adaptation des microorganismes vis-à-vis de leur environnement. Dans le cas des travaux sur la vitesse de dégradation des transcrits de *L. lactis* dans des conditions de carence, les stress imposés sont toujours associés à un ralentissement de la croissance cellulaire. Nos précédents résultats ont clairement montré que le taux de croissance est impliqué dans la régulation de nombreux processus cellulaires dont la transcription, la traduction et la dégradation des protéines. La question est maintenant de déterminer si le taux de croissance influence aussi la stabilité des ARN messagers. Pour cela, nous avons mesuré les temps de demi-vie des transcrits de *L. lactis* lors de cultures en chémostat à deux taux de croissance différents (0.1 et 0.5 h^{-1}). L'emploi de cultures continues en chémostat permet de découpler les différents paramètres de culture et, dans ce cas précis, d'étudier spécifiquement l'effet du

taux de croissance sur la stabilité des ARN messagers en maintenant un état de limitation par l'isoleucine constant. La méthode de mesure des stabilités des transcrits, développée par Redon *et al.* (2005) est brièvement décrite dans la partie matériel et méthodes de cette thèse. Elle repose sur le suivi cinétique de la disparition des ARN messagers par puce à ADN après blocage de la transcription grâce à l'ajout de rifampicine (fixation sur la sous-unité β de l'ARN polymérase).

Détermination des temps de demi-vie des ARN messagers en chémostat

Pour chacun des taux de croissance quatre répétitions de mesure des stabilités des ARN messagers sont effectuées. Deux d'entre elles ont été réalisées sur des cultures indépendantes et correspondent donc à des répétitions biologiques. Les deux autres sont des répétitions méthodologiques car les étapes d'extraction, de marquage et de quantification sont indépendantes mais les échantillons, bien que prélevés à des instants différents, sont issus d'une même culture. Pour chacune de ces répétitions, seules les valeurs de constante de dégradation des transcrits (k) associées à des coefficients de détermination (R^2) supérieurs à 0.7 sont conservées. Enfin, après la mise en commun des différentes répétitions seuls les temps de demi-vie moyens pour lesquels le pourcentage d'erreur est inférieur au pourcentage d'erreur moyen sur l'ensemble des ARN messagers (58 %) ont été sélectionnés.

Les temps de demi-vie de 757 et 627 messagers ont pu être déterminés de façon reproductible pour les taux de croissance de 0.11 et 0.51 h^{-1} respectivement. Comme précédemment observé chez les bactéries (Bernstein *et al.*, 2002; Hambræus *et al.*, 2003; Selinger *et al.*, 2003), y compris chez *L. lactis* (Redon *et al.*, 2005), il existe une grande disparité des stabilités au sein des ARN messagers (figure 13). Les temps de demi-vie sont en effet compris entre 3 et 95 minutes. Le nombre de valeurs extrêmes correspondant à des gènes très stables est tout de même relativement faible puisque 90 % des valeurs restent inférieures à 20 min et 30 min pour des taux de croissance de 0.51 et 0.11 h^{-1} respectivement.

Sur la figure 13, l'histogramme de répartition des temps de demi-vies des ARN messagers se déplace vers des valeurs plus élevées lorsque le taux de croissance est plus faible (barres noires). Les valeurs moyennes et médianes des temps de demi-vie de l'ensemble des transcrits sont respectivement de 12.3 ± 1.0 min et 10.7 min pour le taux de croissance de 0.51 h^{-1} . Elles augmentent à 17.3 ± 1.2 et 14.8 min respectivement pour un taux de croissance de 0.11 h^{-1} .

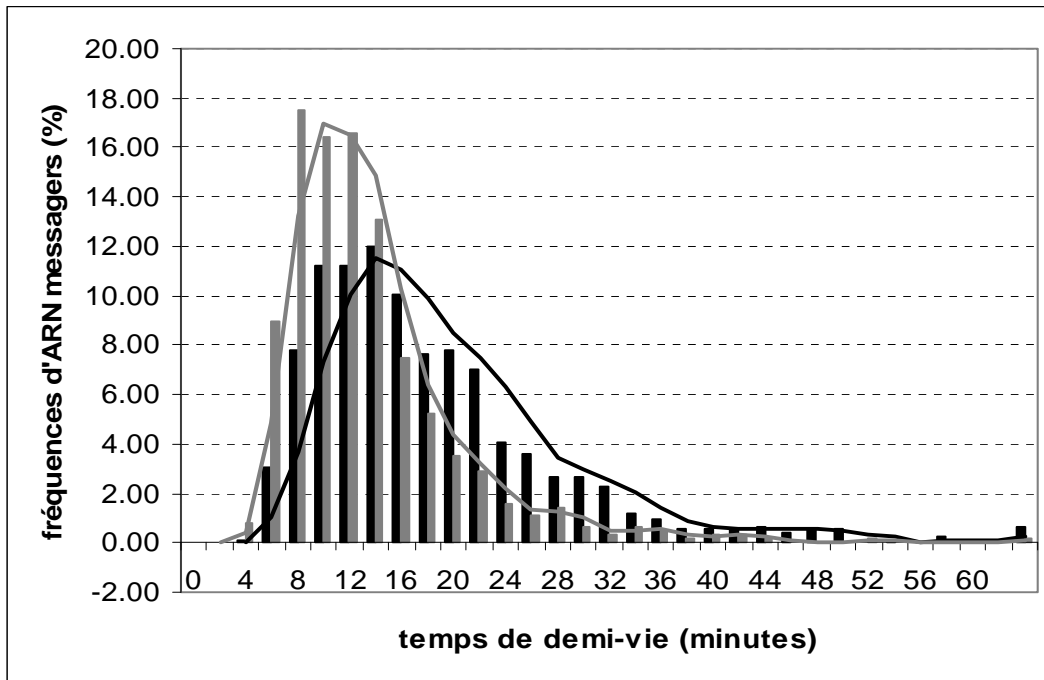


Figure 13 : Distribution des temps de demi-vie des ARN messagers de *L. lactis* cultivé en chimostat à deux taux de croissance différents. (Noir : 0.11 h⁻¹ ; gris : 0.51 h⁻¹).

Lien entre la stabilité des messagers et le taux de croissance

Les différences, en termes de moyenne, comme de médianes, entre les deux points de chimostat étudiés suggèrent un lien entre le taux de croissance et la stabilité des ARN messagers. Cette hypothèse est en opposition avec les résultats de Bernstein *et al.* (2002) qui ne relèvent aucune différence significative entre les temps de demi-vie moyens des transcrits d'*E. coli* se développant à 0.46 (sur milieu pauvre) ou 1.39 h⁻¹ (sur milieu riche). Chez cette bactérie modèle, la stabilité des transcrits semble donc indépendante du taux de croissance. Une légère différence peut cependant être notée sur les valeurs médianes qui sont de 4.7 min pour le taux de croissance le plus rapide et 5.4 pour le plus lent (Bernstein *et al.*, 2002). Une étude partielle de la stabilité de certains ARN messagers de *Streptococcus pyogenes* selon la phase de croissance suggère également un lien entre la stabilité des transcrits et le taux de croissance (Barnett *et al.*, 2007).

Afin de confirmer le lien potentiel entre la stabilité des ARN messagers de *L. lactis* et le taux de croissance, l'ensemble des mesures de temps de demi-vie disponibles et les valeurs des taux de croissance correspondant aux différents points de mesures ont été collectés. L'étude menée sur le rôle de la dégradation des transcrits lors de l'adaptation à la carence en

carbone, donne une stabilité de référence de 5.8 min, en phase exponentielle de croissance, lorsque le taux de croissance est de 0.78 h^{-1} mais, lors de l'accroissement du stress, qui s'accompagne d'un ralentissement de croissance, les transcrits sont globalement stabilisés et la moyenne se situe à 17.6 min en ralentissement puis 19.4 min en phase stationnaire (Redon *et al.*, 2005). Les données produites par Emma Redon au cours de sa thèse (2005) ne sont pas encore publiées mais rapportent de la même façon des valeurs de stabilité qui augmentent quand le taux de croissance diminue lors de la réponse de *L. lactis* à l'imposition progressive d'une carence en isoleucine. En phase exponentielle de croissance la moyenne est en effet de 7.4 min et elle augmente à 8.8 min lors du ralentissement puis jusqu'à 12.8 min lorsque la bactérie s'est adaptée à la carence et pousse extrêmement lentement (tableau 5).

Tableau 5 : Récapitulatif des données disponibles concernant la stabilité des ARN messagers de *L. lactis*

Les données issues des analyses de stabilité en réponse à la carence en isoleucine proviennent des travaux de thèse d'Emma Redon (2005).

| $\mu \text{ (h}^{-1}\text{)}$ | Analyse de l'effet du taux de croissance | | Analyse en réponse à la carence en isoleucine | | | Analyse en réponse à la carence en glucose (Redon <i>et al.</i> , 2005) | | |
|--------------------------------------|--|----------|---|---------|----------|---|----------|----------|
| | 0.11 | 0.51 | 0.85 | 0.38 | 0.04 | 0.78 | 0.28 | 0.00 |
| Nombre de temps de demi-vie calculés | 757 | 627 | 450 | 382 | 848 | 817 | 452 | 579 |
| Temps de demi-vie moyen (min) | 17.3±1.2 | 12.3±1.0 | 7.4±0.7 | 8.8±0.9 | 12.8±0.9 | 5.8±0.4 | 17.6±1.6 | 19.4±1.6 |
| Temps de demi-vie médian (min) | 14.8 | 10.7 | 6.4 | 7.3 | 10.9 | 4.2 | 14.4 | 17.3 |

Le tracé des temps de demi-vie moyens ou médians en fonction du taux de croissance démontre une relation linéaire négative entre la stabilité des ARN messagers et la vitesse de croissance de *L. lactis* (figure 14 A et B). Les coefficients de détermination (R^2) sont en effet de l'ordre de 0.7 et sont associées à des p-values significatives (0.01 que les droites soient tracées avec les valeurs moyennes ou médianes). La dégradation semble donc moins importante à faible taux de croissance et s'accroît avec la vitesse de croissance.

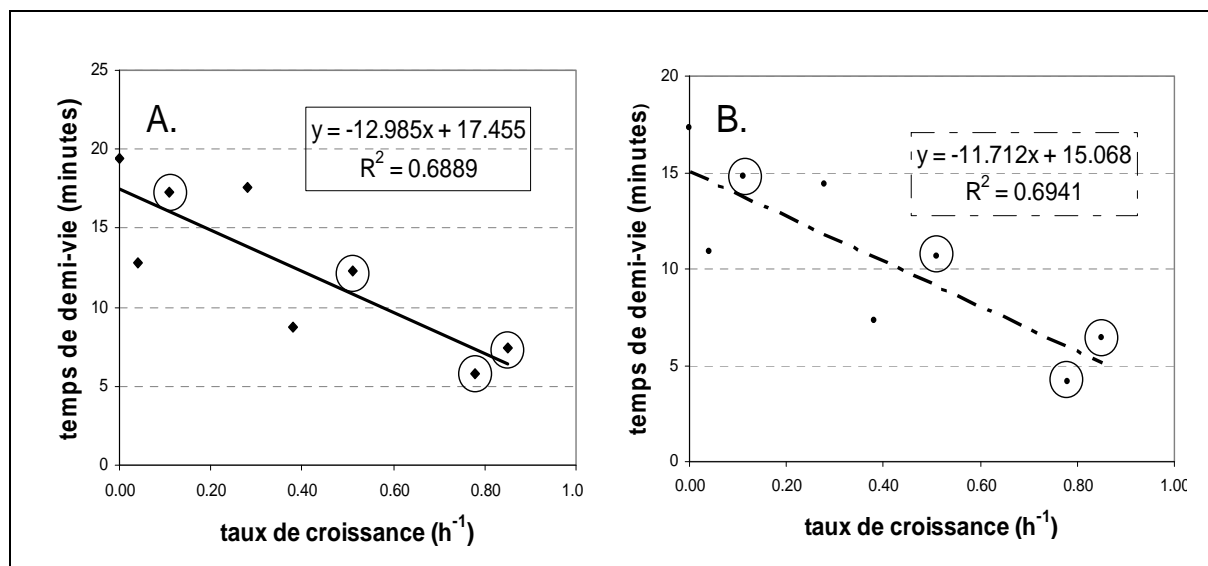


Figure 14 : Evolution des temps de demi-vie des messagers en fonction du taux de croissance de *L. lactis*.

A. : avec les valeurs moyennes, B. : avec les valeurs médianes. Les cerclent signalent les points correspondant à des régimes stables.

Les droites obtenues soulignent une fois de plus le rôle prépondérant du paramètre taux de croissance qui influence les divers processus cellulaires en lien avec l'adaptation de *L. lactis*. Nous noterons que cette linéarité est encore meilleure lorsque l'on s'intéresse aux états assimilables à des régimes permanents. Comme les deux régimes de chémostats, les deux points de croissance exponentielle étudiés lors des réponses à la carence en isoleucine et en glucose correspondent aussi à des états physiologiques stables qui peuvent être assimilés à des régimes permanents (Fishov *et al.*, 1995). En utilisant uniquement ces points, les coefficients de détermination correspondant aux droites avec les temps de demi-vie moyens et médians augmentent respectivement à 0.99 et 0.91, au lieu de 0.69 et 0.69 (résultats non présentés) et restent significatifs à un seuil de 5 % (p-value de 0.03 et 0.05 selon que les droites sont tracées avec les valeurs moyennes ou médianes respectivement).

Au niveau individuel, les mesures de stabilité sont disponibles à, au moins, trois taux de croissance différents pour 921 transcrits. Plus de 90 % des pentes associées aux droites tracées pour chacun des gènes sont négatives. Les R^2 correspondants sont néanmoins moins élevés avec 32 % des valeurs supérieures 0.7 mais une moyenne située aux alentours de 0.5. Ces résultats sont donc en accord avec les déductions faites à partir des profils moyens et médians.

Influence de la stabilité dans le contrôle de la concentration des ARN messagers

Sur l'ensemble des données disponibles, les temps de demi-vie des transcrits s'étendent en moyenne de 5.8 minutes en phase exponentielle de croissance lorsque le taux de croissance est maximal à 19.4 minutes en phase stationnaire, lorsque la croissance est nulle. Pour *L. lactis* cultivé dans les conditions présentées ici, le temps de génération minimum (lorsque la croissance est exponentielle) est de l'ordre de 50 min à comparer à des temps de demi-vie moyens compris entre 5 et 8 min. D'autre part, à l'échelle individuelle, quel que soit le taux de croissance (parmi l'ensemble des conditions étudiées), plus de 90 % des demi-vies mesurées sont inférieures à 5 temps de génération. La dégradation des ARN messagers est donc un phénomène beaucoup plus rapide que la dilution et l'effet de ce processus sur les niveaux de transcrits peut être négligé devant l'influence de la dégradation, contrairement à ce qui avait été démontré pour les protéines. Ainsi, nous pouvons négliger l'aspect dilution et considérer que les concentrations en transcrits résultent d'un simple équilibre entre la transcription et la dégradation. Toute modification de la concentration d'un ARN messager entre deux états physiologiques peut donc être le fruit de la variation d'une de ces vitesses, voire des deux.

Afin de mieux comprendre les mécanismes qui soutendent la régulation de l'expression des gènes en réponse à la variation du taux de croissance, nous avons calculé des coefficients de contrôle, selon une approche semblable à celle utilisée pour les protéines (cf. chapitre III) et précédemment utilisée par Redon *et al.* (2005). Cette démarche est basée sur la théorie du contrôle développée par Westerhoff et ses collaborateurs pour déterminer la régulation des flux métaboliques chez la levure modèle *S. cerevisiae* (Daran-Lapujade *et al.*, 2007; Rossell *et al.*, 2005; ter Kuile & Westerhoff, 2001). En se plaçant dans des conditions d'équilibre et en négligeant le phénomène de dilution par la croissance, il est possible d'écrire que les vitesses de transcription et de dégradation sont égales soit (cf. figure 12 pour la définition de chacun des signes et symboles) :

$$V_{transcription} = V_{dégradation} = k * [ARNm] \text{ soit } [ARNm] = \frac{V_{transcription}}{k}$$

Si k et la vitesse de transcription sont indépendants alors :

$$d[ARNm] = \frac{\partial[ARNm]}{\partial V_{transcription}} \times dV_{transcription} + \frac{\partial[ARNm]}{\partial k} \times dk = \frac{1}{k} \times dV_{transcription} - \frac{V_{transcription}}{k^2} \times dk$$

En divisant cette expression par la concentration en ARN messagers, il vient :

$$\frac{d[ARNm]}{[ARNm]} = \frac{dV_{transcription}}{V_{transcription}} - \frac{dk}{k}$$

Cette équation est équivalente à :

$$1 = \frac{d \ln(V_{transcription})}{d \ln([ARNm])} - \frac{d \ln(k)}{d \ln([ARNm])}$$

Des coefficients de régulation $\rho_{transcriptionnel} = \frac{d \ln(V_{transcription})}{d \ln([ARNm])}$ et $\rho_{dégradationnel} = -\frac{d \ln(k)}{d \ln([ARNm])}$,

peuvent donc être définis pour représenter les régulations des concentrations en ARN messagers par la transcription et la dégradation respectivement. Un $\rho_{dégradationnel}$ supérieur ou égal à 1 est révélateur d'une variation de vitesse de dégradation supérieure ou égale aux variations de concentrations en ARN messagers indiquant un contrôle majoritaire par le processus de dégradation. A l'inverse, un $\rho_{dégradationnel}$ inférieur ou égal à 0 signifie que le phénomène de transcription est prépondérant dans le contrôle des niveaux d'ARN messagers. Enfin, des valeurs comprises entre 0 et 1 ne permettent pas de déterminer la primauté d'un processus cellulaire sur l'autre et impliquent un contrôle partagé.

Le calcul du coefficient de contrôle dégradationnel est directement possible à partir des données expérimentales de concentrations en transcrit et de valeurs de k obtenues aux états d'équilibre. Ainsi, en utilisant toutes les données correspondant à des états stables (les deux régimes de chémostats et les deux phases exponentielles), il est possible de déterminer le mode de régulation de l'expression des gènes en réponse à la variation du taux de croissance. Pour chacun des messagers pour lesquels nous disposons de mesures correspondant au minimum à trois états stables, donc trois taux de croissance différents, nous avons tracé $\ln(k)$ en fonction de $\ln([ARNm])$. Les gènes pour lesquels la dilution ne peut être négligée ont été directement exclus des analyses car ils ne sont pas conformes aux hypothèses précédemment posées. Nous avons cependant pu remarquer qu'ils étaient, tous, sans exception, éliminés de l'analyse car les stabilités de ces messagers n'ont pas été mesurées de façon reproductible dans suffisamment de conditions (plus de deux) pour que les résultats soient exploitables. Ce constat remet en question la fiabilité des valeurs extrêmes trouvées pour ces quelques gènes, seulement dans certaines conditions. Dans la plupart des cas, le tracé $\ln(k)$ en fonction de $\ln([ARNm])$ correspond à des droites (66 % des données sont associées à des R^2 supérieurs à 0.7). Le fait d'obtenir majoritairement des droites signifie que le contrôle de l'expression des gènes par le taux de croissance s'effectue toujours par les mêmes mécanismes et avec une même intensité, indépendamment de la vitesse de croissance. Les résultats, présentés dans le

tableau 6, révèlent un contrôle massif de l'ajustement des concentrations intracellulaires en ARN messagers par le processus de transcription. Quelle que soit la stringence des critères de sélections appliqués pour ne conserver que les pentes significatives, l'expression de plus de 90 % des gènes demeure sous contrôle transcriptionnel, démontrant ainsi la robustesse des analyses effectuées (tableau 6). Très peu de gènes sont soumis à un contrôle partagé et seulement 1 à 2 % sont soumis à un contrôle dégradatif. Ainsi, malgré le lien entre la dégradation des transcrits et le taux de croissance et malgré la dispersion des valeurs individuelles de stabilité des ARN messagers, le contrôle des concentrations de chacun d'entre eux en réponse à la variation de la vitesse de croissance se situe essentiellement au niveau du processus de transcription. La dégradation existe donc et participe à l'adaptation mais ce n'est pas le phénomène prépondérant dans le contrôle suite au changement du taux de croissance.

Tableau 6 : Modes de régulation des concentrations en ARN messagers lors de la variation du taux de croissance de *L. lactis*

Seuls les messagers pour lesquels les mesures correspondant à au moins trois états stables ont été pris en compte.

| | Pas de sélection sur le R ² | R ² ≥ 0.7 | R ² ≥ 0.8 | R ² ≥ 0.9 |
|---------------------------------|---|----------------------|----------------------|----------------------|
| Nombre de coefficients calculés | 319 | 211 | 172 | 127 |
| Contrôle transcriptionnel (%) | 95 | 97 | 97 | 96 |
| Contrôle partagé (%) | 4 | 2 | 2 | 2 |
| Contrôle dégradatif (%) | 1 | 1 | 1 | 2 |

L'application d'une telle méthode de calcul de coefficients de contrôle est plus complexe lorsque les mécanismes de contrôle de l'adaptation à un stress nutritionnel tel que la carence en isoleucine ou en glucose sont recherchés. Si la dilution reste toujours un phénomène négligeable devant la dégradation, les différents points de prélèvement ne correspondent effectivement pas tous à des états stables et l'hypothèse selon laquelle la concentration en ARN messager est constante n'est plus valable. Dans ce cas, la vitesse de transcription s'écrit :

$$V_{transcription} = k * [ARNm] + \frac{\partial [ARNm]}{\partial t}$$

L'estimation des variations de concentrations en transcrits en fonction du temps entre deux points issus de la carence en carbone ou de la carence en isoleucine de *L. lactis* a néanmoins permis de démontrer que pour 72 à 95 % des ARN messagers, selon les transitions étudiées, $\frac{\Delta[ARNm]}{\Delta t}$ était inférieur à $5 * k * [ARNm]$. Pour ces transcrits, nous pouvons donc considérer que les calculs de coefficients de contrôle restent possibles sous les mêmes hypothèses que précédemment. Les ARN messagers dont les valeurs ne sont pas en accord avec ces hypothèses ont été exclus de l'analyse.

Les résultats, récapitulés dans le tableau 7, montrent que, dans les conditions de stress nutritionnel étudiées, les concentrations en ARN messagers restent majoritairement sous le contrôle de la transcription. Ce contrôle est cependant moins tranché que lors de la variation du taux de croissance. En effet, la part de régulation liée au processus de dégradation des transcrits est significative lors des carences où le pourcentage de gènes sous contrôle partagé est également plus important. Cette tendance souligne l'importance de la dégradation des transcrits en condition de stress et montre la complexité des processus d'adaptation de *L. lactis* à son environnement.

Tableau 7 : Modes de régulation des concentrations en ARN messagers lors de la réponse de *L. lactis* à une carence nutritionnelle

Les données issues des analyses de stabilité en réponse à la carence en isoleucine proviennent des travaux de thèse d'Emma Redon (2005). Les flèches indiquent les états entre lesquels les coefficients de contrôle ont été calculés.

| | Analyse en réponse à la carence en isoleucine (Redon, 2005) | | Analyse en réponse à la carence en glucose (Redon <i>et al.</i> , 2005) | |
|-------------------------------|--|-------------|--|-------------|
| | 0.85 ↔ 0.38 | 0.38 ↔ 0.04 | 0.78 ↔ 0.28 | 0.28 ↔ 0.00 |
| μ (h ⁻¹) | 0.85 ↔ 0.38 ↔ 0.04 | | 0.78 ↔ 0.28 ↔ 0.00 | |
| Contrôle transcriptionnel (%) | 57 | 79 | 93 | 50 |
| Contrôle partagé (%) | 20 | 6 | 4 | 23 |
| Contrôle dégradatif (%) | 22 | 15 | 3 | 27 |

Bilan et perspectives

Cette étude nous a permis de réévaluer le champ d'action du taux de croissance qui s'étend bien au-delà d'une simple adaptation de la vitesse de croissance mais affecte tous les processus physiologiques majeurs. Ces résultats sont cependant encore préliminaires et méritent absolument d'être complétés. Une approche par modèles de covariance, identique à celle utilisée pour l'identification des paramètres influençant les niveaux de protéines semble particulièrement adaptée pour la recherche de déterminants de la stabilité des ARN messagers, tout comme pour l'analyse des facteurs contrôlant les concentrations en transcrits. Une classification des gènes selon le mode de contrôle de leur expression lors de l'adaptation à différents stress pourrait également nous permettre de distinguer plusieurs groupes au sein desquels il conviendrait de rechercher la présence de motifs régulateurs dans les séquences géniques et intergéniques. Enfin, l'intégration des résultats concernant la régulation des concentrations en protéines et en ARN messagers devrait permettre d'élargir notre vision de *L. lactis* et d'améliorer la compréhension globale de ses processus d'adaptation.

- Barnett T. C., Bugrysheva J. V., and Scott J. R. (2007). Role of mRNA stability in growth phase regulation of gene expression in the group A streptococcus. *J Bacteriol* **189**: 1866-73.
- Bernstein J. A., Khodursky A. B., Lin P. H., Lin-Chao S., and Cohen S. N. (2002). Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A* **99**: 9697-702.
- Carpousis A. J. (2007). The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. *Annu Rev Microbiol* **61**: 71-87.
- Daran-Lapujade P., Rossell S., van Gulik W. M., Luttik M. A., de Groot M. J., Slijper M., Heck A. J., Daran J. M., de Winde J. H., Westerhoff H. V., Pronk J. T., and Bakker B. M. (2007). The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *Proc Natl Acad Sci U S A* **104**: 15753-8.
- Fishov I., Zaritsky A., and Grover N. B. (1995). On microbial states of growth. *Mol Microbiol* **15**: 789-94.
- Garneau N. L., Wilusz J., and Wilusz C. J. (2007). The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* **8**: 113-26.
- Hambraeus G., von Wachenfeldt C., and Hederstedt L. (2003). Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Mol Genet Genomics* **269**: 706-14.
- Rauhut R., and Klug G. (1999). mRNA degradation in bacteria. *FEMS Microbiol Rev* **23**: 353-70.
- Redon E. (2005). Identification des déterminants de l'expression génique lors d l'adaptation de *Lactococcus lactis* au stress: intégration des données de transcriptome et de stabilité des ARN messagers, INSA Toulouse, Toulouse.
- Redon E., Loubiere P., and Cacaïgn-Bousquet M. (2005). Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. *J Biol Chem* **280**: 36380-5.
- Rossell S., van der Weijden C. C., Kruckeberg A. L., Bakker B. M., and Westerhoff H. V. (2005). Hierarchical and metabolic regulation of glucose influx in starved *Saccharomyces cerevisiae*. *FEMS Yeast Res* **5**: 611-9.
- Schmid M., and Jensen T. H. (2008). The exosome: a multipurpose RNA-decay machine. *Trends Biochem Sci* **33**: 501-10.
- Selinger D. W., Saxena R. M., Cheung K. J., Church G. M., and Rosenow C. (2003). Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res* **13**: 216-23.
- Shalem O., Dahan O., Levo M., Martinez M. R., Furman I., Segal E., and Pilpel Y. (2008). Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol Syst Biol* **4**: 223.
- Takayama K., and Kjelleberg S. (2000). The role of RNA stability during bacterial stress responses and starvation. *Environ Microbiol* **2**: 355-65.
- ter Kuile B. H., and Westerhoff H. V. (2001). Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett* **500**: 169-71.
- Wang Y., Liu C. L., Storey J. D., Tibshirani R. J., Herschlag D., and Brown P. O. (2002). Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci U S A* **99**: 5860-5.

