Université de Toulouse

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

Cotutelle internationale avec :

Présentée et soutenue par :
Pierre MILLARD

Le lundi 17 décembre 2012

Titre :

Role of the Csr system in carbon nutrition and in the control of central metabolism in *Escherichia coli* K-12 MG1655 and Nissle 1917

ED SEVAB : Ingénieries microbienne et enzymatique

Unité de recherche :
Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés

Directeur(s) de Thèse :

Jean-Charles PORTAIS
Jean-Philippe NOUGAYREDE

Rapporteurs :

Erick DENAMUR
Guillaume TCHERKEZ

Autre(s) membre(s) du jury :

Fabien LETISSE
Patrick KIEFER
Nic LINDLEY

Abstract

The implantation of *Escherichia coli* in the gut results from global adaptive strategies that allow the bacteria to survive in the changing environment of the intestine. At the metabolic level, recent findings indicate that colonisation is mainly related to the utilization of sugars and sugar derivatives through glycolytic pathways. In contrast, persistence of *E. coli* in the gut is supported by less favorable substrates, including small organic acids. The use of the latter compounds requires activation of gluconeogenic pathways, and efficient switching between glycolytic and gluconeogenic carbon sources is likely to be a major feature of successful adaptation to life in the intestine. These adaptive processes are controlled by highly sophisticated regulatory networks, such as the Csr (carbon storage regulator) system which is the main post-transcriptional regulator in *E. coli*. Csr was found to control a broad range of phenotypes allowing *E. coli* to successfully implant and persist in the gut, such as biofilm formation, motility as well as many functions involved in carbon nutrition, including glycolysis, gluconeogenesis, acetate and glycogen metabolism. Although Csr is likely to play an important role in the adaptation of bacteria to the nutritional context of the host, it is poorly understood sofar.

In this work, we investigate the role of the Csr system in the control of *E. coli* metabolism on physiologically-relevant carbon sources representative of the main glycolytic (Entner-Doudoroff pathway, pentose phosphate pathway, glycolysis) and gluconeogenic pathways of *E. coli*. This work was carried out on two *E. coli* strains with distinct implantation capabilities: the K12 MG1655 laboratory strain and the Nissle 1917 strain, an efficient colonizer of the gut belonging to the highly competitive B2 phylogenetic group.

First, we designed a complete methodology (metabolomics and ^{13}C-metabolic flux analysis) for quantitative, system-level investigations of the actual operation of *E. coli* metabolism. Then, we performed detailed, system-level investigations of wild-type strains and Csr mutants. This work provides valuable information regarding systemic properties of *E. coli* metabolism, and identifies metabolic specificities of the Nissle 1917 strain likely involved in its competitiveness in the gut. The role of Csr appears to be qualitatively and quantitatively the same in both K12 MG1655 and Nissle 1917 strains. We show that i) Csr enhances the utilisation of a broad spectrum of glycolytic and gluconeogenic carbon sources, ii) Csr controls a range of metabolic pathways wider than expected from its known target enzymes, and iii) the actual impact of the Csr system on the central metabolism of *E. coli* depends on the carbon source. We also demonstrate that Csr controls energy and redox metabolism in *E. coli*. Csr enhances the production of ATP and of reduced cofactors (NADH and NADPH), and we suggest that it also may control the catabolism-anabolism balance in *E. coli*. Finally, our results reinforce the potential role of the Csr system in the global adaptation of the bacterium to the gut environment.

Résumé

L'implantation d'*Escherichia coli* dans l'intestin résulte de stratégies adaptatives globales permettant à la bactérie de survivre face aux changements de conditions environnementales. Sur le plan métabolique, la colonisation de l'intestin par *E. coli* parait associée à la disponibilité de sources de carbone préférentielles glycolytiques, alors que la maintenance et la persistance reposent sur sa capacité à utiliser différents substrats alternatifs (principalement gluconéogéniques) lorsque les substrats préférentiels deviennent limitants. Cette activation des processus gluconéogéniques, qui implique une réorganisation fonctionnelle complète du métabolisme, est associée *in situ* à de profonds remaniements physiologiques (perte de motilité, formation de biofilms, etc) nécessaires à la persistance d'*E. coli* dans l'intestin. Ces différents processus sont coordonnés par des réseaux de régulation particulièrement complexes, dont le système Csr (Carbon storage regulator), un des principaux régulateur post-transcriptionnel d'*E. coli*.

Lors de ces travaux, nous avons analysé le rôle du système Csr dans la nutrition carbonée et le contrôle du métabolisme central d'*E. coli* sur des sources de carbone supportant sa croissance dans l'intestin et représentatives des principales voies de son métabolisme central (glycolyse, voie des pentoses phosphate, voie d'Entner-Doudoroff, cycle de Krebs). Cette étude a été réalisée chez deux souches d'*E. coli* présentant des capacités d'implantation distinctes: la souche de laboratoire K12 MG1655, et la souche Nissle 1917, un excellent colonisateur appartenant au groupe phylogénétique B2.

Une analyse détaillée du fonctionnement métabolique par des approches systémiques quantitatives haut-débit (métabolomique et analyse des flux métaboliques par marquage isotopique) a été mise en place. Elle a été exploitée pour caractériser finement le comportement métabolique de souches sauvages et de mutants du système Csr d'*E. coli* sur différentes sources de carbone, identifier des caractères métaboliques propres à chaque souche, et étendre le rôle du système Csr dans la nutrition carbonée et dans le contrôle du métabolisme central d'*E. coli*. Nos résultats démontrent que i) Csr favorise l'utilisation d'un large spectre de sources de carbone aussi bien glycolytiques que gluconéogéniques, ii) Csr contrôle un nombre de voies métaboliques plus important que ce que l'on pourrait attendre à partir de ses cibles identifiées, et iii) que le contrôle global exercé par le système Csr sur le fonctionnement du métabolisme central dépend de la source de carbone. Un rôle du système Csr dans le contrôle du métabolisme redox (production de NADH et de NADPH) et énergétique (production d'ATP), non reporté à ce jour, est également démontré. Enfin, nos résultats suggèrent un rôle de Csr dans le contrôle de la balance anabolisme-catabolisme d'*E. coli*. Ces travaux renforcent le rôle potentiel du système Csr dans l'adaptation d'*E. coli* face aux changements de conditions environnementales.

Remerciements

Cette thèse n'aurait pu voir le jour sans le soutien scientifique et moral de nombreuses personnes que je tiens à remercier.

A Jean-Charles Portais, mon directeur de thèse, à Fabien Letisse, qui a co-encadré ces travaux, et à Jean-Philippe Nougayrède, qui a co-dirigé cette thèse, sincères remerciements pour m'avoir guidé, encouragé et conseillé au cours de ces trois années, tout en me laissant une grande liberté. J'ai pris énormément de plaisir à collaborer avec eux, dans une ambiance qu'ils ont su rendre particulièrement propice aux échanges scientifiques.

A Erick Denamur et Guillaume Tcherkez pour avoir accepté d'être les rapporteurs de ce travail, ainsi qu'à Patrick Kiefer pour avoir accepté de participer à ce jury.

A Nic Lindley, directeur du Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP, Toulouse), pour m'avoir accueilli au sein de son unité et aiguillé pour la suite de mon parcours professionnel.

A l'INRA, et notamment au département MICA, pour m'avoir permis de réaliser ces travaux et de m'évader deux ans de plus vers de nouveaux cieux scientifiques grâce au CJS dont je bénéficie.

A la plateforme MetaToul (LISBP, Toulouse) pour m'avoir permis d'accéder à ses ressources analytiques (spectrométrie de masse qu'en RMN) et à la plateforme automatisée de culture microbienne.

A tous les membres de l'équipe MetaSys et de la plateforme MetaToul, qui ont rendu ces travaux possibles dans un cadre particulièrement agréable. Olga, pour notre collaboration durant une bonne partie de cette thèse, sans qui les travaux présentés dans le chapitre 7 n'auraient pas vu le jour. B. Enjalbert, roi de la transition, pour nos discussions sur le système Csr chez *E. coli*, et pour sa facétie. Juliette et Stéphanie, qui ont été particulièrement disponibles pour m'épauler dans l'utilisation de "Nono", la plateforme automatisée de culture microbienne, aussi bien de jour que de nuit. Serguei, sobre "mathémagicien", pour sa patience et ses précieux conseils en modélisation et en programmation. Lara et Lindsay, pour leur gentillesse, leur soutien constant et la flexibilité dont elles ont su faire preuve sur les plannings des systèmes analytiques. Stéphane, pour avoir partagé son expertise en RMN et réussi à m'inculqué quelques notions rugbystiques. Matthieu, informaticien et cuisinier anglais émérite à ses heures, sans qui les données du robot seraient restées brutes. Gilles, qui croit toujours que les fruits sont des desserts, pour nos discussions autour de la modélisation. Edern, pour tous les bons moments partagés

durant les quatre dernières années. Stéphanie, Jane, Rémi et Vincent, croisés pendant cette thèse. Et enfin la boite à meuh qui a rythmé ce voyage.

A Muriel, Laurence, Thomas et Brice E., de l'équipe "Génie du métabolisme des procaryotes" du LISBP, pour avoir entre autre construit le mutant *csrA51* chez la souche K12 MG1655, et pour les bons moments partagés autours des chemostats.

A mes amis, dont Aude, Fanny, Frédéric et Vanessa, trop nombreux pour que je les cite tous, à qui je tiens énormément et qui me le rendent bien.

A ma famille, et particulièrement ma mère, pour leur soutien indéfectible et m'avoir toujours encouragé dans les démarches que j'ai entreprises.

Enfin, à Jean-Marie, pour me supporter sans faillir depuis tant d'années.

Publications

Sokol S., Millard P., Portais J.C. (2012). influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments. *Bioinformatics*, 28(5):687-693.

Millard P., Letisse F., Sokol S., Portais J.C. (2012). IsoCor: Correcting MS data in isotope labeling experiments. *Bioinformatics*, 28(9):1294-1296.

Publications in preparation

Revelles O., Millard P., Nougayrède J.P., Oswald E., Letisse F. and Portais J.C. The Csr system exerts global control over central metabolism in *Escherichia coli* strain Nissle 1917 in a nutrient-specific manner. (for Plos One)

Millard P., Letisse F., Sokol S., Portais J.C. IsoDesign: rationalizing ^{13}C-metabolic flux analysis by in-silico optimisation of the experimental design. (for Bioinformatics)

Millard P., Massou S., Bolten C., Wittmann C., Portais J.C. and Letisse F. Performing a reliable sampling of metabolic intermediates for stationary and instationary ^{13}C-metabolic flux analysis: Practical guidelines for Escherichia coli. (for Analytical Biochemistry)

Millard P., Massou S., Portais J.C. and Letisse F. Isotopic studies of metabolic systems by mass spectrometry: using the Pascal's triangle to produce biological standards with fully controlled labelling patterns - Theoretical aspects and applications. (for Analytical Chemistry)

Enjalbert B., Durand S., Millard P., Letisse F., and Portais J.C. Molecular and phenotypic characterization reveals the dynamic and complexity of the glucose to acetate transition in *Escherichia coli*. (for Journal of Bacteriology)

Oral presentations

Mass spectrometry studies of metabolic systems: using the Pascal's triangle to probe the isotopologue space. (23-25 May 2011, 5th scientific days of the French Network for Metabolomics and Fluxomics, Paris, France)

IsoCor and influx_s : Two bioinformatic tools for high-throughput ^{13}C-fluxomics. (21-23 May 2012, 6th scientific days of the French Network for Metabolomics and Fluxomics, Nantes, France)

Posters

Massou S., Millard P., Portais J.C., Letisse F. (2009). High-accuracy LC-MS/MS quantification of isotopomer distribution in intracellular metabolites for ^{13}C-fluxomics. (11 – 14 June, Fifth International Conference of the Metabolomics Society, Edmonton, Canada)

Revelles O., Millard P., Heux S., Oswald E., Nougayrède J.P., Letisse F., Portais J.C. (2010). Role of Csr system in the metabolism and fitness of *Escherichia coli* Nissle 1917. (Metabolism meets virulence

Millard P., Revelles O., Sokol S., Oswald E., Nougayrède J.P., Letisse F., Portais J.C. (2010). Systems biology of commensal and pathogenic *Escherichia coli* strains. (5-7 September, Current Opinion in Cellular Host-Pathogen Interactions, Amsterdam, The Netherlands)

Dubois M., Millard P., Letisse F., Jourdan F. (2011). MetMS : identification software for global metabolomics. (23-25 May, 5th scientific days of the French Network for Metabolomics and Fluxomics, Paris, France)

Millard P., Sokol S., Portais J.C., Letisse F. (2012). IsoCor & influx_s: Bioinfomatics tools for high-throughput ^{13}C-fluxomics. (25-28 June, Metabolomics Society 8th Annual Meeting, Washington, DC, USA)

Contents

Abbreviations

2PG: 2-phosphoglycerate
3PG: 3-phosphoglycerate
6PG: 6-phosphogluconate
AKG: Alpha-ketoglutarate
CID: Carbon isotopologue distribution
Cit: Citrate
DNA: Deoxyribonucleic acid
ED: Entner-Doudoroff pathway
EMP: Embden-Meyerhof-Parnas pathway (glycolysis)
Ery4P: Erythrose-4-phosphate
FAD(H): (Reduced) Flavin adenine dinucleotide cofactor
Fru6P: Fructose-6-phosphate
FruBP: Fructose-1,6-bisphosphate
Fum: Fumarate
GC: Gas chromatography
Glc6P: Glucose-6-phosphate
HSQC: Heteronuclear single quantum correlation
ID: Isotopologue distribution
IDMS: Isotope dilution - mass spectrometry
LB: Luria Bertani
LC: Liquid chromatography
Mal: Malate
MAR: Mass action ratio
MFA: Metabolic flux analysis
MRM: Multiple reaction monitoring
mRNA: Messenger RNA
MS(/MS): (Tandem) Mass spectrometry
NAD(H): (Reduced) Nicotinamide adenine dinucleotide cofactor
NADP(H): (Reduced) Nicotinamide adenine dinucleotide phosphate cofactor
NMR: Nuclear magnetic resonance
OAA: Oxaloacetate
OD: Optical density
PEP: Phosphoenolpyruvate
PP: Pentose phosphate pathway
ppGpp: Guanosine-3',5'-bisdiphosphate
Pyr: Pyruvate
Rib5P: Ribose-5-phosphate
RNA: Ribonucleic acid
Sed7P: Sedoheptulose-7-phosphate
SIM: Selected ion monitoring
Suc: Succinate
TCA: Tricarboxylic acid cycle (Krebs cycle)
ZQF-TOCSY: Zero-quantum filtered total correlated spectroscopy
WT: Wild type

List of figures

Chapter 1 - Introduction

Chapter 2 - IsoDesign: rationalizing ^{13}C-metabolic flux analysis by *in silico* optimisation of the experimental design

Chapter 3 - Reliable sampling of ^{13}C-labelled metabolic intermediates for stationary and instationary ^{13}C-metabolic flux analysis: Practical guidelines for *Escherichia coli*

Chapter 4 - Isotopic studies of metabolic systems by mass spectrometry: using the Pascal's triangle to produce biological standards with fully controlled labelling patterns - Theoretical aspects and applications

Chapter 5 - Correcting MS data in isotope labeling experiments

Chapter 6 - influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments

Chapter 7 - The post-transcriptional regulator CsrA exerts global control over central metabolism in *Escherichia coli* strain Nissle 1917 in a nutrient-specific manner

Chapter 8 – The Csr system coordinates carbon, energy and redox metabolism in *Escherichia coli* K12 MG1655 and Nissle 1917

List of tables

Chapter 4 - Isotopic studies of metabolic systems by mass spectrometry: using the Pascal's triangle to produce biological standards with fully controlled labelling patterns - Theoretical aspects and applications

Chapter 7 - The post-transcriptional regulator CsrA exerts global control over central metabolism in *Escherichia coli* strain Nissle 1917 in a nutrient-specific manner

Chapter 8 – The Csr system coordinates carbon, energy and redox metabolism in *Escherichia coli* K12 MG1655 and Nissle 1917

General context

This work is included in a collaboration between the Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (UMR5504, UMR792, CNRS, INRA, INSA, Toulouse) and the Laboratoire des Interactions Hôtes - Agents Pathogènes (UMR1225, INRA, ENVT, Toulouse), which aims at the comprehensive understanding of the intestine-*Escherichia coli* relationship. This thesis is funded by the program CJS (Contrat Jeune Scientifique) of INRA.

Escherichia coli is a normal inhabitant of the gut microbiota but is also a highly versatile bacterial species with considerable pathogenic potential. Certain *E. coli* strains are responsible of enteric and extraintestinal infections such as neonatal meningitis and septicemia. *E. coli* is also a contaminant of the food chain, as shown by recent examples in Germany (during the 2011 outbreak, 3 950 people were affected by *E. coli* O104:H4 and 53 died). Comprehensive understanding of the intestine-*E. coli* interaction is a major issue for human or animal health and nutrition, as well as for the basic knowledge of the functioning of gut microflora.

To cope with the highly changing environment of the gut, enteric bacteria use a wide variety of regulatory mechanisms. Highly sophisticated global regulatory networks modulate the expression of genes involved in numerous cellular processes, including metabolism. Regulation of gene expression is mediated in large part through the activation or repression of mRNA transcription. However, transcriptional regulations are necessary but insufficient for the quantitative description of biological system adaptations, and to explain the actual phenotype. Indeed, post-transcriptional regulatory mechanisms – at the level of mRNA degradation and translation – are also critical determinants of gene expression. In *E. coli*, the main type of post-transcriptional control is mediated by the Csr (carbon storage regulator) system. Csr was found to control a broad range of physiological processes allowing *E. coli* to successfully implant and persist in the gut, such as biofilm formation, motility as well as many functions involved in carbon nutrition, including glycolysis, gluconeogenesis, acetate and glycogen metabolism. Although the Csr system is likely to play an important role in the adaptation of bacteria to the nutritional context of the host, it is poorly understood sofar.

In that context, this thesis mainly focuses on the comprehensive understanding of the role of the Csr system in carbon nutrition and in control of central metabolism in *E. coli*.

Chapter 1

Introduction

1. The gut microbiota

The human gastro-intestinal tract harbours a complex and dense community of micro-organisms, the gut microbiota. In numerical terms, the microbiota community exceeds the number of human somatic and germ cells by a factor of 10 (Savage 1977; Ley, Peterson et al. 2006), and the combined microbial metagenome has a coding capacity that exceeds the human genome by a factor of 100, making human in large part prokaryotic. The gut microbiota is separated from the internal medium by a single-layer of epithelial cells, which permits bi-directional host-microbiota exchanges on a total surface area of almost 200 m^2. Hence, the gut microbiota can be considered as an organ (O'Hara and Shanahan 2006) located in the gastro-intestinal tract, with a set of biochemical pathways having a greater metabolic diversity than the liver (Egert, de Graaf et al. 2006; Gill, Pop et al. 2006), which is considered to be the most versatile human organ. It has been shown that the gut microbiota plays a key role in health and disease because it is involved in developmental [gastro-intestinal tract, brain (Heijtz, Wang et al. 2011), cardiovascular system], metabolic [conversion of indigestible compounds (Chassard, Scott et al. 2008; Flint, Bayer et al. 2008; Bernalier-Donadille 2010), regulation of host fat storage (Turnbaugh, Ley et al. 2006), vitamin synthesis], and immunological [inflammation processes, mucosal/systemic immunity maturation, tolerance at the gastro-intestinal tract mucosa] functions of the host (Fujimura, Slusher et al. 2010; Sekirov, Russell et al. 2010). Finally, the microbiota has a critical protective function by providing a natural defense barrier against invading pathogenic micro-organisms.

The gut microbiota comprises more than 1000 bacterial species (Sekirov, Russell et al. 2010), with strict anaerobes dominating the facultative anaerobes and aerobes by 2 to 3 decades. The microbiota is not homogeneous within the gut and is characterized by important spatiotemporal gradients. The number of bacteria per gram of content shows a continuum from 10^1 in the stomach to 10^{12} in the colon, with preferential niches for most of the species (Sekirov, Russell et al. 2010). Human born germ-free and are immediately colonised by micro-organisms at birth (mainly coming from the mother's microbiota), then the bacterial population widely evolves until it starts to resemble that of an adult after the first year. During life, several factors strongly reshape the microbiota composition, such as host food diet, implantation of other micro-organisms,

diseases or drugs (Leitch, Walker et al. 2007; Zhang, DiBaise et al. 2009; Muegge, Kuczynski et al. 2011; Wu, Chen et al. 2011; Costello, Stagaman et al. 2012). Consistently, strong differences in the composition of the gut microbiota have been reported among individuals according to their lifestyle or to their geographic localization (Duriez, Clermont et al. 2001; Escobar-Paramo, Grenet et al. 2004; Bailey, Pinyon et al. 2010).

Thus, a better knowledge of microbial functioning is vital for unravelling host-microbial interactions and for understanding the roles that intestinal microorganisms play in health and disease. Such knowledge should also permit the development of novel strategies for the prevention or treatment of diseases that are caused by microbiota dysfunction.

2. Colonisation and persistence of *Escherichia coli* in the gut

Escherichia coli (figure 1) is a Gram-negative bacterium of the Enterobacteriacae species and has been isolated from human faeces in 1885 by Theodor Escherich. *E. coli* is a normal inhabitant of the human gut (with a prevalence higher than 90%) and is the main facultative anaerobe of the microbiota. It is mainly found in the large intestine, especially in the caecum and the colon, where it resides in the mucus layer covering the epithelial cells. *E. coli* is also a highly versatile bacterium with considerable pathogenic potential that is implicated in a broad spectrum of diseases and is responsible of 2.5 million deads per year (Russo and Johnson 2003).

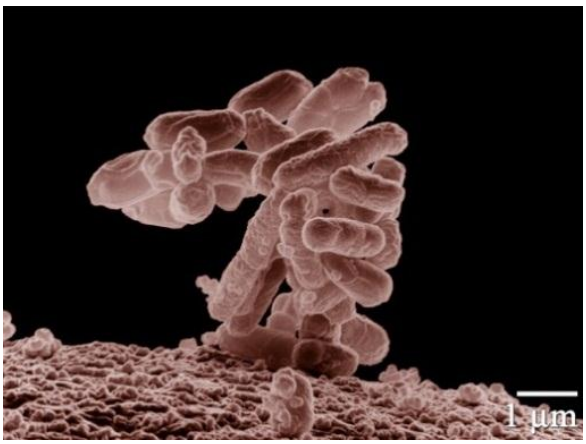


Figure 1. *Escherichia coli* observed by transmission electron microscopy (x 10 000).

(source: Agricultural Research Service, image number K11077-1)

Escherichia coli: a versatile bacteria

Mutations, gene duplications and horizontal gene transfers are frequent in *E. coli*, which make its genome highly flexible and dynamic (Touchon, Hoede et al. 2009; Wielgoss, Barrick et al. 2011). These evolutionary mechanisms may result in changes in the physiology or lifestyle of the bacterium (Chaudhuri and Henderson 2012). For example, a strain may gain pathogenic capacity, the ability to use a carbon source (Shiloach, Reshamwala et al. 2010), the ability to take upon a particular ecological niche (Moulin-Schouleur, Schouler et al. 2006; Moulin-Schouleur, Reperant et al. 2007; Tivendale, Logue et al. 2010; Chanteloup, Porcheron et al. 2011) or the ability to resist antimicrobial agents (Graziani, Luzzi et al. 2009; Giufre, Graziani et al. 2012).

Most *E. coli* strains are non-pathogenic commensal strains of humans. However, several strains can cause diseases by expressing a set of virulence factors such as adhesins, capsule, toxins, etc. Based on their virulence traits, two main groups of pathogenic *E. coli* responsible for enteric and extraintestinal infections have been defined, intestinal pathogenic *E. coli* (InPEC) and extra-intestinal pathogenic *E. coli* (ExPEC) respectively (Figure 2) (Kaper, Nataro et al. 2004; Croxen and Finlay 2010). InPEC strains are classified in 6 well-described pathotypes according to their pathogenicity scheme (Kaper, Nataro et al. 2004): enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli*

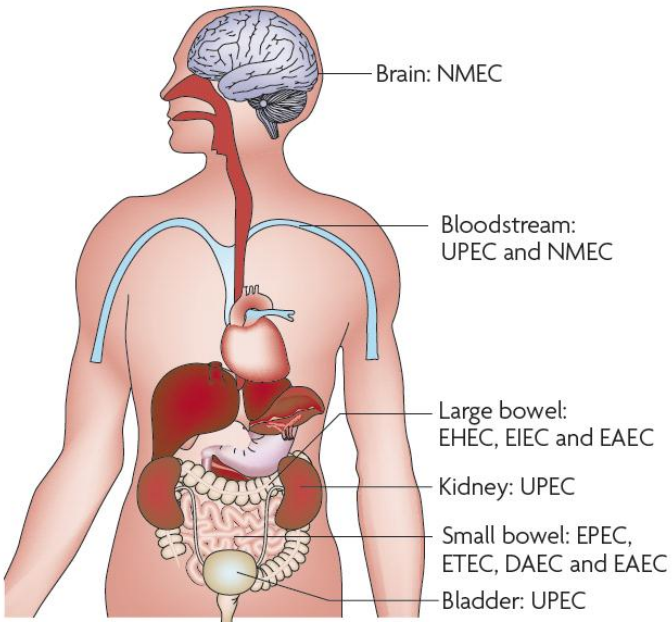


Figure 2. Pathogenic *Escherichia coli* and their site of colonisation (from (Croxen and Finlay 2010)).

(EIEC) and diffusely adherent *E. coli* (DAEC). InPEC are implicated in intestinal diseases such as diarrhoea and gastro-enteritis. In contrast, ExPEC colonise the gut without inducing clinical symptoms (an ExPEC strain is the predominant *E. coli* strain of 20% of healthy people) but can cause opportunistic infections when introduced into extraintestinal sites. ExPEC are classified in 2 pathotypes: uropathogenic *E. coli* (UPEC) and meningitis-associated *E. coli* (NMEC) responsible for meningitis and for sepsis.

Based on whole genome sequences of human-isolated *E. coli* strains – including both commensal and pathogenic strains –, phylogenetic analysis has revealed four main phylogenetic groups (A, B1, B2, D1) and two accessory groups (D2, E) (Figure 3) (Chaudhuri and Henderson 2012). These phylogroups differ in their ecological niches, life-history, metabolic capabilities and show some characteristic traits, such as their antibiotic-resistance profiles, their growth-rates or their genome sizes (Carlos, Pires et al. 2010).

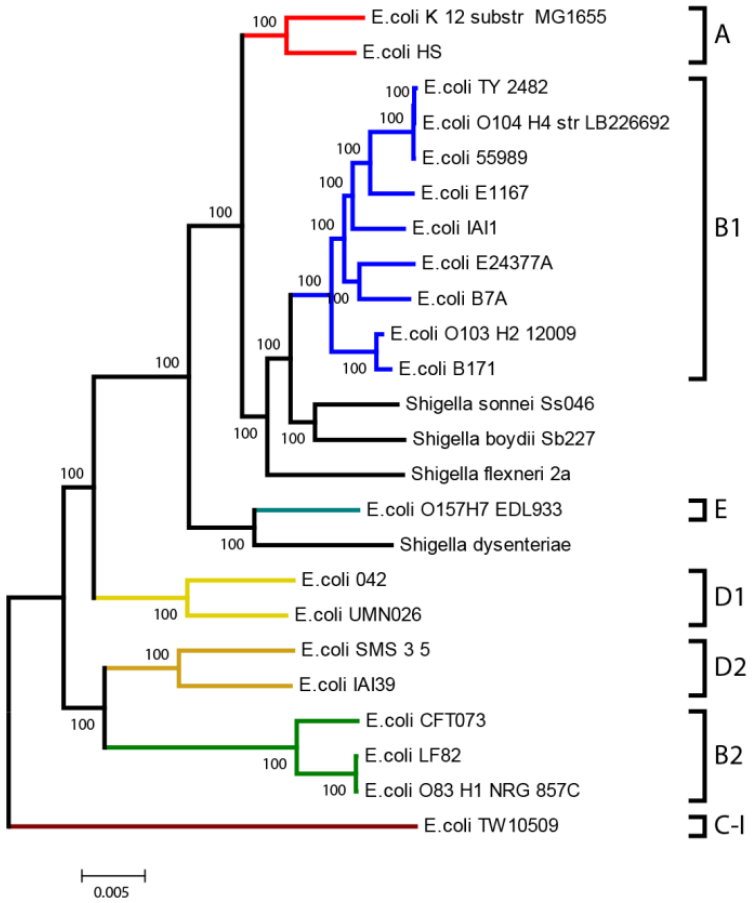


Figure 3. Phylogenetic relationships amongst selected *E. coli* strains, as determined by phylogenetic analysis of complete and draft genome sequences available in GenBank (from (Chaudhuri and Henderson 2012)).

For instance, strong differences in average genome sizes (from 4.5 up to 5.5 million base pairs) exist between groups, A and B1 strains having smaller genomes compared to B2 and D1 strains. A and B1 groups contain predominantly commensal strains and some intestinal pathogenic strains, while B2 and D1 groups encompass a high number of extraintestinal pathogenic strains and a low number of commensal strains. Strains of the B2 group also contain more virulence factors than strains of other groups. For instance, it is the only group containing the *pks* genomic island (Nowrouzian and Oswald 2012) which encodes for the colibactin, a recently discovered genotoxin (Nougayrede, Homburg et al. 2006) shown to confer a significant advantage for long-term persistence (Nowrouzian and Oswald 2012). The B2 group is of special interest because it contains numerous highly versatile and highly competitive pathogenic and commensal strains (Nowrouzian, Wold et al. 2005; Nowrouzian, Adlerberth et al. 2006), and its prevalence is increasing in developed countries (Escobar-Paramo, Grenet et al. 2004; Bailey, Pinyon et al. 2010).

In particular, comparative genome analyses have highlighted the important role of carbon nutrition in adaptation of commensal and pathogenic *E. coli* isolates to their ecological niches. By reconstructing metabolic networks of 29 *E. coli* strains from their genomic sequences, (Vieira, Sabarly et al. 2011) found that the core-metabolism (i.e. the set of reactions common to all strains) represents the major part (57%) of the panmetabolism (the set of all reactions of all strains). The central metabolism (the key supplier of precursors and energy and redox cofactors needed for growth) and synthesis processes are found in the core-metabolism, which indicate a selective pressure common to all strains. In contrast, the variable part of the metabolism mainly contains the degradation processes which are directly linked to environment and lifestyle (e.g. the *dsdXA* operon responsible for D-serine utilization, a carbon source found in urine, is mainly present in UPEC strains). Consistently, (Grasselli, Francois et al. 2008) have shown that a large proportion of hyperdivergent genes between 20 *E. coli* strains are involved in carbohydrates transport and metabolism. The differences in metabolic capabilities of these strains (i.e. their growth capacities on a set of carbon sources) predicted by these genomic studies have been experimentally confirmed by comparative analysis of their metabolic phenotype (Bernier-Febreau, du Merle et al. 2004; Sabarly, Bouvet et al. 2011). (Sabarly, Bouvet et al. 2011) have analysed the capacity of oxidation (indicative of a metabolic activity) of 95 carbon sources by 168 *E. coli* using the

BioLog GN2 system. Over the 48 carbon sources oxidised by at least one strain, only 7 carbon sources are oxidised by all strains. Although metabolic pathways are distributed according to the species phylogeny, metabolic phenotypes (e.g. oxidation rate) and metabolomes (i.e. the total pool of metabolites) are weakly correlated with strain phylogeny, lifestyle or pathogenicity (Sabarly, Bouvet et al. 2011; Maharjan and Ferenci 2005).

Hence, *E. coli* encompasses a high number of strains that exhibit a very important degree of genetic and phenotypic diversity. The discrepancies in terms of metabolic phenotypes originate from two distinct levels: i) the available enzymatic material encoded into the genome and ii) the operation of metabolism (e.g. enzyme activities). The latter is controlled by strain-specific regulatory mechanisms (Sabarly, Bouvet et al. 2011; Waegeman, Beauprez et al. 2011; Aubron, Glodt et al. 2012) which coordinate physiological and metabolic responses of the cell to adapt to its environment.

So far, few studies focused on the strain specificities in terms of metabolic functioning, mainly because of the lack of high-throughput tools allowing accurate and detailed metabolic and phenotypic characterization. For instance, some carbon sources (e.g. acetate) are considered as not oxidised by *E. coli* using the BioLog system (Sabarly, Bouvet et al. 2011) while it is well-established that they support the growth of several *E. coli* strains (Oh, Rohlin et al. 2002; Zhao and Shimizu 2003). A close examination of the functional metabolic diversity of distinct *E. coli* strains is critical to identify specific metabolic features involved in their competitiveness in the host.

Role of the metabolism in colonisation and persistence of *E. coli*

The intestine defines a nutritional ecological niche to which *E. coli* has adapted from a metabolic point of view. The nutrients allowing growth and survival of gut microbiota are of three different origins (figure 4). Human lacks enzymatic material needed for degradation of some polysaccharide-rich compounds (such as plant cell-wall, starch, inulin, pectins, gums, mucilages, fructooligosaccharides, etc), which pass into the distal portion of the gastro-intestinal tract. Together with endogenous epithelial-derived mucus, these polysaccharides are degraded into monomers by glycosyl-hydrolases produced by anaerobic communities of the microbiota (Hoskins and Boulding 1981; Backhed, Ley et al. 2005; Gill, Pop et al. 2006). The subsequently released

carbohydrates (Figure 4) support the growth of other bacteria. Fermentation of these glycolytic carbon sources by the microbiota results in a variety of metabolites, mainly small organic acids (acetate, lactate, succinate, etc) which can also serve as nutrients for other bacterial species.

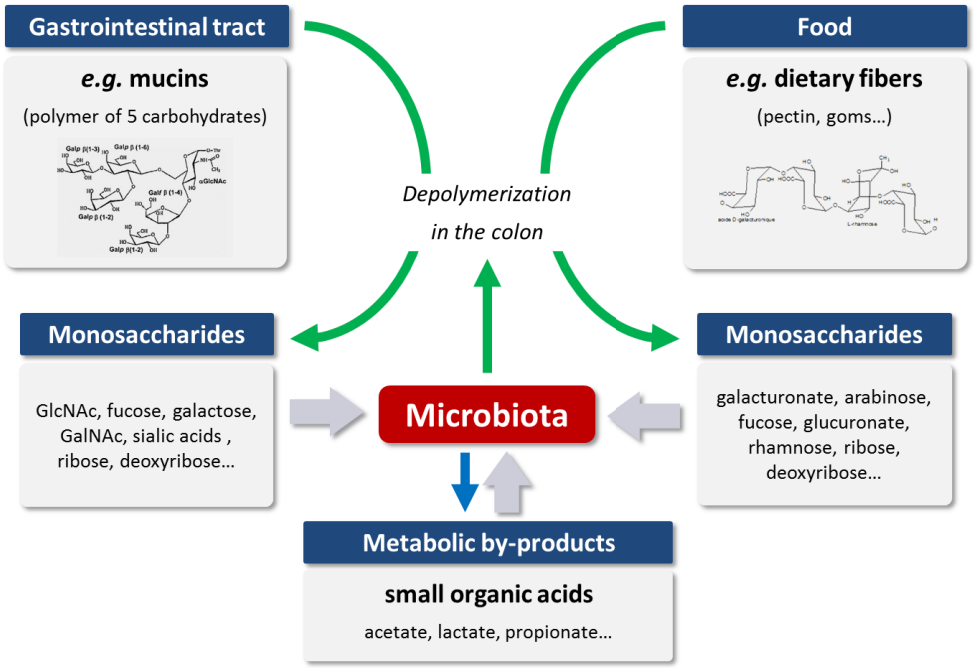


Figure 4. Origin of nutrients found in the gut, including monosaccharides released from hydrolysis of undigested dietary food or of mucus degradation by anaerobes and metabolites produced during fermentation of glycolytic substrates by the microbiota. Grey arrows represent nutrients utilisation by the microbiota, blue and green arrows refer to the production of nutrients by the microbiota via direct and indirect mechanisms, respectively.

According to Freter's nutrient niche theory, the ability to compete for carbon nutrition is a critical factor for gut colonisation, and is part of the arsenal of strategies employed by pathogenic *E. coli* strains to outcompete the gut microbiotia (Miranda, Conway et al. 2004; Kamada, Kim et al. 2012). Consistently, the genomic content of *E. coli* strains reflects its nutritional environment. For instance, the *deoK* operon confers to some ExPEC strains a significant colonisation advantage in the gut by allowing the use of deoxyribose (Bernier-Febreau, du Merle et al. 2004; Martinez-Jehanne, du Merle et al. 2009), a sugar derived from DNA degradation due to colonic epithelial turn-over. A *fos* locus involved in the metabolism of short-chain fructooligosaccharides was found to provide a fitness advantage to some pathogenic *E. coli* strains in the gut (Schouler, Taki et al. 2009). A *frz* operon, highly associated with ExPEC strains and encoding for

transport and catabolism of an unidentified carbohydrate, was also found to promote colonisation of the gut by the ExPEC strain BEN2908 (Rouquet, Porcheron et al. 2009).

Although the operation of metabolism varies between different *E. coli* strains both *in vitro* (Shiloach, Reshamwala et al. 2010; Sabarly, Bouvet et al. 2011; Aubron, Glodt et al. 2012) and *in vivo* (Chang, Smalley et al. 2004; Miranda, Conway et al. 2004; Fabich, Jones et al. 2008; Snider, Fabich et al. 2009), a common pattern of carbon nutrition is observed in *E. coli* to colonise and survive within the gut. Colonisation is mainly related to the utilisation of sugars and sugar derivatives resulting from the degradation of mucus and dietary polysaccharides (e.g. gluconate, N-acetylglucosamine, fucose, deoxyribose, ribose, galactose, maltose, etc) (figure 5) (Chang, Smalley et al. 2004; Miranda, Conway et al. 2004; Fabich, Jones et al. 2008; Snider, Fabich et al. 2009). These carbon sources are metabolized via glycolytic pathways such as the Embden-Meyerhof-Parnas (EMP), the Entner-Doudoroff (ED), and the Pentose Phosphates (PP) pathways (figure 5).

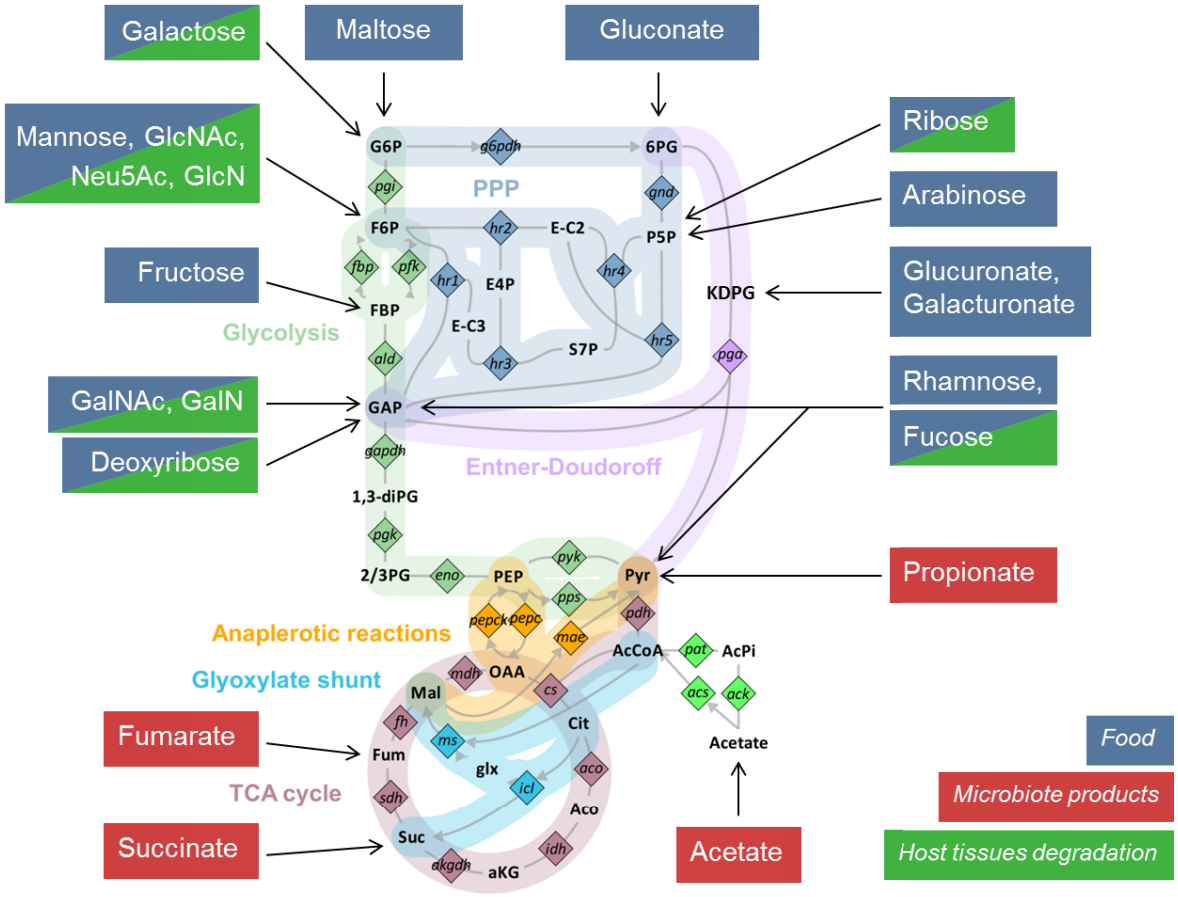


Figure 5. Central carbon metabolic network of *E. coli*, and carbon sources supporting its growth in the gut. The colour refers to the origin of the carbon source, which is indicated in the legend.

Persistence of *E. coli* in the gut is supported by less favourable substrates, including both sugars (glucuronate, mannose, fucose, and ribose, etc) and non-sugar compounds such as the small organic acids (e.g. acetate, fumarate, succinate, etc) produced by the microbiota (Chang, Smalley et al. 2004; Miranda, Conway et al. 2004) (figure 5). The use of the latter compounds requires activation of gluconeogenic pathways, and efficient switching between glycolytic and gluconeogenic carbon sources is likely to be a major feature for successful adaptation to life in the intestine (Pernestig, Georgellis et al. 2003; Miranda, Conway et al. 2004).

However, little is known on the actual operation of *E. coli* metabolism on physiologically-relevant carbon sources. This knowledge is needed for understanding how *E. coli* successfully implants and survives in the gut, and may also pave the way towards new therapeutic and/or preventive strategies for diseases caused by pathogenic *E. coli* or by microbiota dysfunction.

3. The Csr system

To cope with the changing environment of the intestine, *E. coli* has developed a variety of adaptation mechanisms. Highly sophisticated global regulatory networks coordinate physiological and metabolic responses by controlling the functional expression of relevant sets of genes. Different kind of regulatory mechanisms modulate their expression by controlling DNA-RNA transcription and RNA-protein translation, the main steps of gene expression. Transcriptional regulation controls when transcription occurs and how much RNA is created. Post-transcriptional regulation controls how much the mRNA is translated into proteins by regulating RNA stability and initiation of translation.

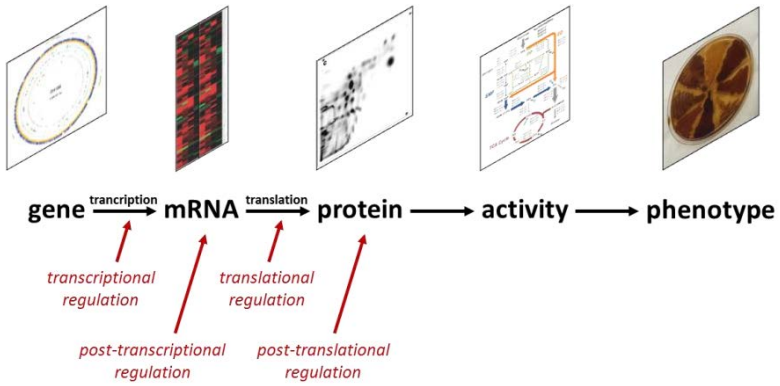


Figure 6. Different kinds of regulatory mechanisms control the functional expression of the genome to the actual phenotype.

Translational and post-translational regulations control the levels of translation and the levels of active protein, respectively (figure 6). To date, transcriptional regulation is generally well-characterised in *E. coli*, while post-transcriptional regulation has been less studied although it is also a critical determinant of functional gene expression.

The carbon storage regulator (Csr) system is a post-transcriptional regulation system that controls a broad range of physiological adaptative mechanisms and is a global regulator of central metabolism. Although originally discovered in *E. coli*, homologous systems have been then identified in more than 176 bacteria (Lucchetti-Miganeh, Burrowes et al. 2008). The Csr system is one of the main post-transcriptional regulator in *E. coli*.

The CsrA protein

The main component of the Csr system is the post-transcriptional regulator CsrA, an mRNA-binding protein of 61 amino-acids that influences both translation and stability of different mRNA targets. CsrA homodimerizes and regulates gene expression by binding mainly to the 5' untranslated region of its target mRNAs on a highly conserved GGA sequence often located in the loop of hairpin structures. In most cases, CsrA binds multiple sites in the target mRNA (Mercante, Edwards et al. 2009), which strengthens the regulatory effect. CsrA can also bind to two separate RNAs (Schubert, Lapouge et al. 2007).

Opposite effects of CsrA on the stability of target transcripts have been described. In negative CsrA-mediated regulation, CsrA binds at or near the Shine-Dalgarno sequence, thus competes with 30S ribosomal subunits for mRNA binding. This situation leads to rapid degradation of transcripts both actively by increasing ribonucleolytic cleavage and passively by reducing translation. Recent studies also found that CsrA binds to intergenic regions to repress the expression of a specific gene within the transcript (Pannuri, Yakhnin et al. 2012). The mechanism of positive CsrA-mediated regulation has been less studied, but (Wei, Brun-Zinkernagel et al. 2001) found that CsrA stabilizes the flhDC transcript (encoded by the *flhDC* operon responsible of flagellum biosynthesis) by protecting from endonucleolytic cleavage by RNase E.

The Csr regulatory network

CsrA level and activity are controlled by other elements as shown in figure 7. CsrA is negatively regulated by two small non-coding RNAs, csrB (Liu, Gui et al. 1997) and csrC (Weilbacher, Suzuki et al. 2003), which contain multiple CsrA-binding sites (respectively ~18-22 and ~9-10) (Mercante, Edwards et al. 2009). Together, these RNAs antagonize CsrA activity by sequestering multiple CsrA dimers, thus reduce its availability to interact with its target mRNAs. The CsrD protein positively controls CsrA activity by driving the RNAs csrB and csrC to RNAse E degradation (Suzuki, Babitzke et al. 2006). CsrA represses expression of *csrD* (Jonas, Tomenius et al. 2006; Suzuki, Babitzke et al. 2006; Jonas and Melefors 2009) and indirectly activates the transcription of *csrB* and *csrC* via the BarA/UvrY two-component system (Gudapaty, Suzuki et al. 2001; Suzuki, Wang et al. 2002). CsrA also controls its own expression via autoregulatory loops, either negatively by binding and destabilizing its own transcript or positively by indirectly activating its own transcription from the P3 promoter (Yakhnin, Yakhnin et al. 2011). All the components of the Csr system are densely connected via several feedback loops, which allow the cell to finely tune CsrA level.

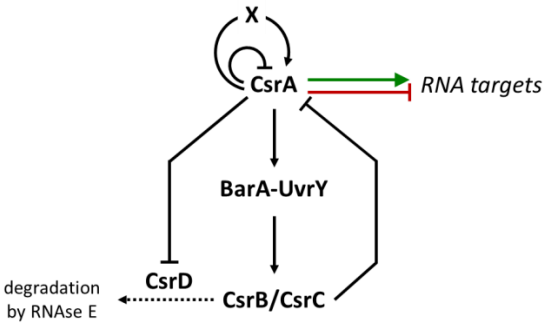


Figure 7. Schematic view of the CsrA regulatory network (adapted from (Romeo, Vakulskas et al. 2012)) including the central post-transcriptional regulator CsrA, the two-component system BarA/UvrY, the two small non-coding RNAs csrB and csrC and the protein CsrD. Arrows denote positive control, and arrows with bar represent negative control.

Role of the Csr system

The Csr system is a pleiotropic regulator that controls a broad range of physiological adaptive mechanisms in *E. coli*, including motility, biofilm formation, stringent response, virulence and metabolism (figure 8).

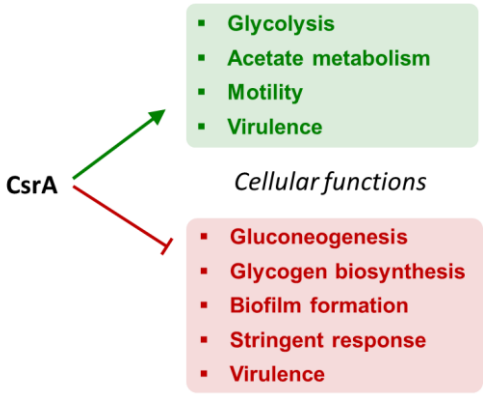


Figure 8. Physiological and metabolic targets of the CsrA post-transcriptional regulator. Green and red colors denote positive and negative control by CsrA, respectively.

A graphical representation of the global Csr network, compiling bibliographic knowledge on the interaction of the Csr network with other molecular and functional components of *E. coli*, has been performed from literature mining by Dr. Brice Enjalbert (Metasys team, LISBP) and is presented figure 9. The main functional targets of the Csr system are detailed hereafter.

Biofilm formation. CsrA inhibits biofilm formation by i) inhibiting *pgaA* translation and destabilizing the *pgaABCD* mRNA required for biofilm synthesis in *E. coli* (Wang, Dubey et al. 2005), ii) inhibiting expression of *nhaR* (Pannuri, Yakhnin et al. 2012) which activates *pgaABCD* transcription, and iii) repressing genes involved in the synthesis of c-di-GMP (Jonas, Edwards et al. 2008), which positively regulates biofilm formation. Interestingly, CsrA also directly stimulates motility by activating *flhDC* expression, the master operon for flagellum biosynthesis (Wei, Brun-Zinkernagel et al. 2001).

Stringent response. CsrA represses *relA* (Edwards, Patterson-Fortin et al. 2011) which mediates the stringent response to amino acid starvation by activating the production of guanosine tetraphosphate (ppGpp). The latter compound is a secondary messenger that inhibits macromolecule biosynthesis (DNA, RNA, proteins, fatty acids) and controls expression of a set of genes involved in a broad range of physiological processes in *E. coli* (such as *flhDC* operon for flagellum biosynthesis and the stationary phase sigma factor RpoS (Brown, Gentry et al. 2002)). During the stringent response, ppGpp also activates the synthesis of csrB and csrC RNAs (Jonas and Melefors 2009; Edwards, Patterson-Fortin et al. 2011). This feedback regulatory loop decreases CsrA level and thus increases the ppGpp level.

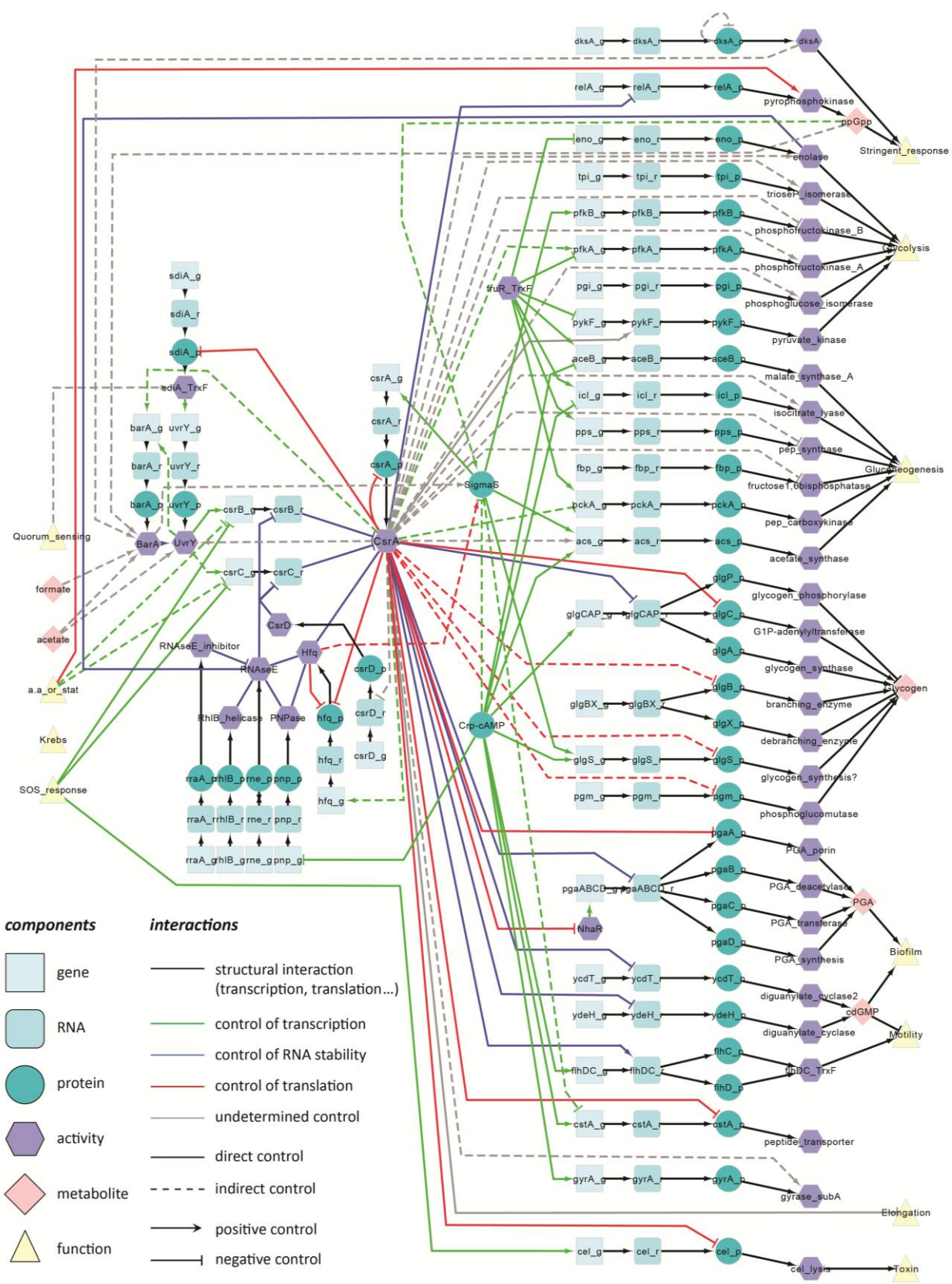


Figure 9. Schematic representation of the CsrA interaction network in *E. coli*. From left to right: the sensor and trigger elements, the regulatory cascade, the CsrA protein in the central part, the CsrA targets, their activities and finally their functions.

Virulence. The Csr system and its homologues are involved in bacterial pathogenesis. Both *in vivo* and *in vitro* studies have revealed that the Csr system controls expression of virulence in *E. coli*. The BarA-UvrY two-component system regulates, likely via CsrA, pathogenicity of the avian pathogenic *E. coli* strain O78:K80:H9 (Herren, Mitra et al. 2006). (Palaniyandi, Mitra et al. 2012) observed that a reduced CsrA activity increases virulence of the UPEC strain CFT073. Interestingly, in the EPEC strain 2348/69 serotype O127:H6, both increase and decrease in CsrA activity negatively regulates virulence by controlling gene expression from the LEE pathogenicity island via 2 distinct mechanisms (Bhatt, Edwards et al. 2009; Bhatt, Anyanful et al. 2011), as previously observed in the control of pathogenicity in *Salmonella enterica* serovar *Typhimurium* (Altier, Suyemoto et al. 2000). These data demonstrate that CsrA is both an activator and an inhibitor of virulence, and highlight the importance for the cell to finely tune CsrA activity.

Metabolism. Biochemical studies have shown that CsrA exerts a control over several enzymes of the central carbon metabolism of *E. coli*, either directly or indirectly. However, the regulation mechanisms by CsrA are not known for most of its targets. CsrA positively control enzymes of the EMP pathway (phosphoglucoisomerase, phosphofructokinase A, triose-phosphate isomerase, enolase, and pyruvate kinase F) (Sabnis, Yang et al. 1995). An unidentified and indirect positive control by CsrA of an enzyme of the ED pathway (2-keto-3-deoxygluconate 6-phosphate aldolase) has been reported by (Murray and Conway 2005). In addition, CsrA negatively controls gluconeogenic activities (fructose-1,6-biphosphatase, PEP synthase, PEP carboxykinase A) and glycogen biosynthesis (Romeo, Gong et al. 1993; Baker, Morozov et al. 2002), and positively regulates acetate utilisation (acetyl-coA synthetase and isocitrate lyase) (Wei, Shin et al. 2000). CsrA also regulates transportation and utilisation of amino acids through repression of *cstA* (Dubey, Baker et al. 2003), which encodes a peptide transporter induced by carbon starvation. Several studies concluded that neither enzymes of the pentose phosphate pathway nor enzymes of the TCA cycle are controlled by CsrA (Sabnis, Yang et al. 1995; Wei, Shin et al. 2000). Consistently, the CsrA protein is essential for growth on glycolytic carbon sources such as glucose, but not on a gluconeogenic compound like pyruvate (Timmermans and Van Melderen 2009). Because CsrA regulates glycolysis and gluconeogenesis in opposite ways, the Csr system is assumed to play the role of master switch between glycolytic and gluconeogenic metabolisms. Interestingly, the BarA/UvrY system – which controls CsrA – is needed for

efficient switching between glycolytic and gluconeogenic carbon sources (Pernestig, Georgellis et al. 2003) although the mechanism by which this is achieved is not clear.

The Csr system also controls other global regulators (e.g. Sigma S, Crp), some of them acting in turn on the Csr system. For instance, Sigma S regulates expression of *csrA*, which controls expression of *hfq*, which indirectly controls expression of Sigma S (Baker, Eory et al. 2007; Yakhnin, Yakhnin et al. 2011). These regulatory loops complicate the understanding of the role of the Csr system.

Recently, (Edwards, Patterson-Fortin et al. 2011) purified, sequenced and identified 721 different transcripts non-covalently bounded with CsrA. Hence, there are likely a high number of unidentified targets of the Csr system, making the regulon of CsrA far to be fully identified. Since this set of transcripts represent more than 15% of *E. coli* genes and is part of 20 *E. coli* COGs (clusters of orthologous groups of proteins), CsrA potentially controls all of the major physiological and molecular processes of *E. coli* (Edwards, Patterson-Fortin et al. 2011).

The control exerted by Csr on metabolic and physiological functions shown to be critical for colonisation and persistence of *E. coli* in the gut makes potentially CsrA a key regulator for physiological and metabolic switches *in vivo*. Despite advances in our understanding of the molecular basis of Csr-mediated regulation, its metabolic effects have not yet been fully investigated. The role of Csr on *E. coli* metabolism has been characterized based on quantification of mRNA and protein levels, and of protein activities. Although these data are of great importance for understanding the mechanisms by which Csr regulates its targets, the actual effect of CsrA on the operation of *E. coli* metabolism in terms of metabolic pools and of metabolic fluxes remain unknown. Indeed, the final flux profile of the cell may substantially differ from its transcriptome and proteome profiles because the different metabolic pathways are closely interconnected within the cell, and changes in the Csr-targeted pathways may lead to modifications in the activity of the entire central metabolism or may be counterbalanced by flux regulatory mechanisms. That makes functional system-level investigations promising to characterise and understand the role of the Csr system in control of *E. coli* metabolism. Among these approaches, ^{13}C-metabolic flux analysis, which gives access to metabolic fluxes – i.e. the actual rates of biochemical reactions in

living organisms –, is particularly suited to understand the role of global regulators on the metabolism (Perrenoud and Sauer 2005; Yao, Hirose et al. 2011).

4. ^{13}C-Metabolic Flux Analysis

Metabolic fluxes are the actual rates of biochemical reactions occurring in living cells. Metabolic flux analysis (MFA) aims to quantify metabolic fluxes, i.e. the actual operation of metabolism under given physiological conditions. The result is a flux map which represents the distribution of carbon over catabolic and anabolic pathways of the metabolic network. Energetic and redox fluxes can also be determined from fluxes which consume or produce ATP and redox cofactors, respectively. MFA provides the most accurate picture of the actual metabolic behaviour and is of special interest in functional genomics and systems biology to establish the genotype-phenotype relationship. During the last years, MFA has also revealed new insights on microbial metabolism (Fischer and Sauer 2003; Shimizu 2004; Ishii, Nakahigashi et al. 2007; Nicolas, Kiefer et al. 2007).

Principle

MFA requires prior knowledge of the topology of the metabolic network, i.e. the set of biochemical reactions in the cell (figure 10a). The network stoichiometry can be represented by a matrix S of m columns (representing all reactions) and n rows (representing the metabolites). At metabolic steady-state, i.e. when the concentrations of all intracellular metabolites are constant, the metabolic fluxes can be calculated from the following equation, which expresses the mass conservation law:

$$S.v = \frac{dC}{dt} = 0$$

where S is the stoichiometric matrix of the system, C is the vector of metabolite concentrations, and v is the vector of fluxes to determine (figure 10b).
For most metabolic systems the matrix S is underdetermined ($n<m$). Some fluxes can be experimentally determined, e.g. extracellular fluxes can be calculated from the time-course measurement of extracellular concentrations (figure 10c), but it is generally not enough to determine the remaining intracellular fluxes. In addition, some particular biologically-relevant metabolic structures (parallel, cyclic or bidirectional reactions) cannot be determined. Additional equations must be added to fix the degrees of

freedom, such as cofactors and energetic balances. However, uncertainties on the assumptions made on these balances (such as the ATP requirements or the occurrence of substrate cycles) raise questions on the validity of the calculated fluxes. To overcome this drawback, an alternative method incorporates growth or product formation as optimization criterion and predicts the intracellular fluxes via minimization or maximization of the objective function (Varma, Boesch et al. 1993; Orth, Thiele et al. 2010). However, metabolism is often suboptimal and differences may be observed between experimental and predicted behaviours (Schuetz, Kuepfer et al. 2007).

^{13}C-MFA has been developed to overcome these limitations. In this approach, an organism is fed with a ^{13}C-labeled substrate, and the label distributes into the metabolites as function of pathways activities (i.e. metabolic fluxes). Since the isotopic content of metabolites can be measured using appropriate analytical tools, equations describing the mass balance of the isotopes can be established and provide additional constraints for calculation of intracellular fluxes (figure 10d).

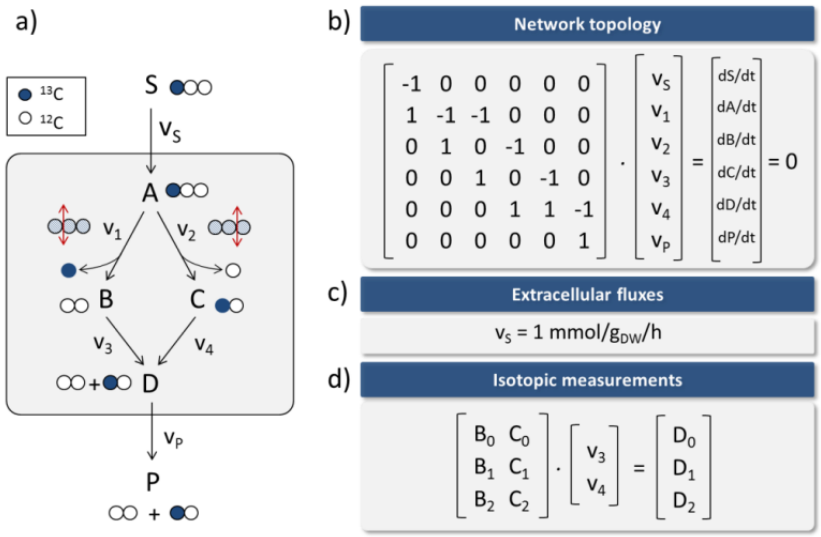


Figure 10. General principle of ^{13}C-metabolic flux analysis. a) Representation of a simple metabolic network consuming the substrate S to form the product P in 4 steps. Two parallel pathways v_1/v_3 and v_2/v_4 result in the loss of C_1 and C_3 of the metabolite A, respectively. b) A stoichiometric matrix S can be calculated based on the network topology. At metabolic steady-state $S.v=dc/dt=0$, where v is the flux vector. c) Extracellular fluxes can be determined from time-course measurement of extracellular metabolites concentrations to fix some degrees of freedom. d) Additional equations can be obtained from mass balance of the isotopes to fix the remaining degrees of freedom. X_i denotes the relative proportion of the metabolite X containing i ^{13}C atoms, e.g. B_1 represents the proportion of the metabolite B with 1 ^{13}C atom relative to the total pool of B.

Workflow of a ^{13}C-MFA experiment

In practice, metabolic flux analysis is a highly time-consuming and complex process requiring the combination of a set of wet and dry methods. The schematic workflow of a ^{13}C-MFA experiment is shown figure 11, and the main steps are briefly detailed hereafter.

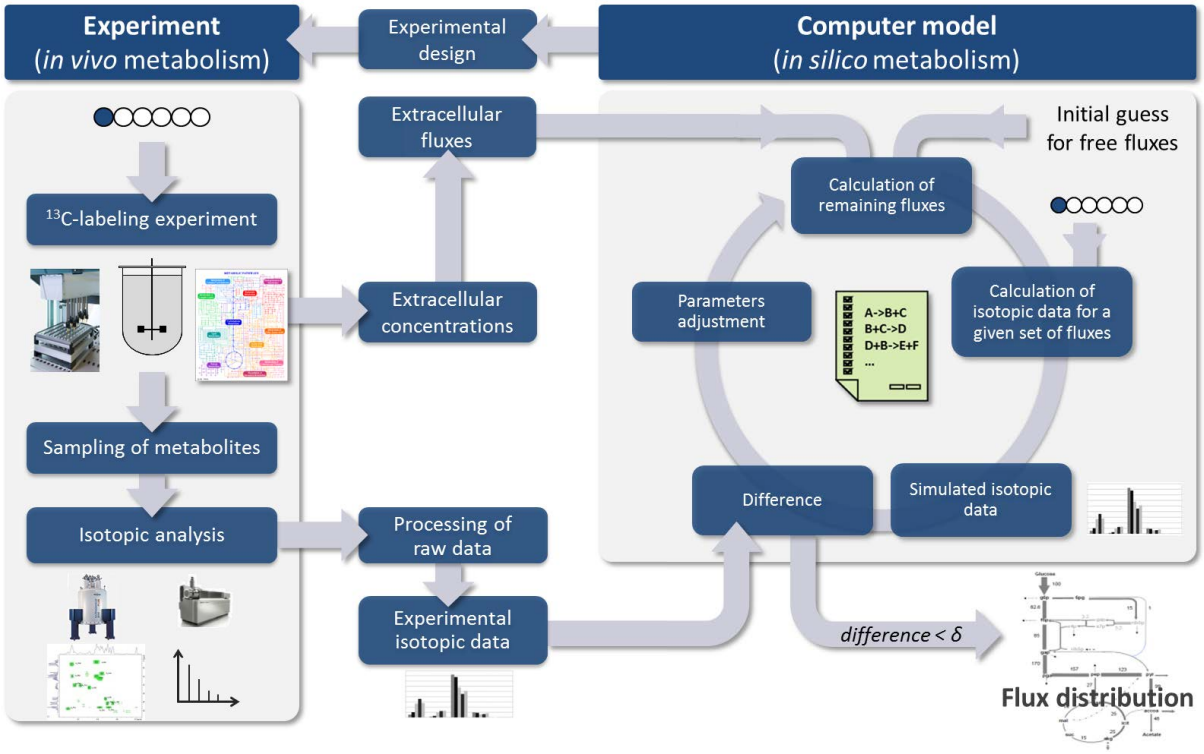


Figure 11. Schematic workflow of a ^{13}C-MFA experiment with i) the experimental design, ii) the ^{13}C-labelling experiment which includes the growth of the microorganism on the labelled carbon source(s), the measurement of label incorporation in metabolites and the determination of extracellular fluxes, iii) the processing of raw isotopic data, iv) the flux calculation based on a mathematical model of the metabolic network and v) the sensitivity analysis to determine the confidence on calculated fluxes.

Experimental design. The precision on fluxes to be determined strongly depends on both the isotopic composition of the labelled substrate and the measurable isotopic information. Obviously either pure unlabelled or pure fully labelled substrate will yield no flux information at all. These isotopic forms must be mixed with one or several other isotopic forms to improve the flux resolution. Thus, a set of tools and methods (Mollney, Wiechert et al. 1999; Quek, Wittmann et al. 2009; Crown and Antoniewicz 2012; Walther, Metallo et al. 2012) have been developed to determine the optimal label input, which highly depends on the biological question to be addressed.

Carbon labelling experiment. Once the experimental conditions have been set up, the organism must be cultivated with the labelled carbon source. Metabolic and isotopic stationary states must be reached and maintained during the entire experiment (Wiechert 2001). In addition, cells must be grown in chemically defined medium (i.e. minimal medium) since all carbon inputs must be known and controlled. Experiment can become very expensive due to the price of the labelled substrate(s) (e.g. the price of U-$^{13}C_6$-rhamnose is over 120 000 €/g), which requires minimizing the cultivation volume. Growth experiments are usually performed in shake flask or bioreactor (~100 mL – 1 L) (Perrenoud and Sauer 2005; Nicolas, Kiefer et al. 2007) or more recently in deep-well microtiter plate (~1 mL) (Fischer, Zamboni et al. 2004; Haverkorn van Rijsewijk, Nanchen et al. 2011). Small-scale chemostat or bioreactor systems (~10 mL) ensuring well-controlled and optimal growth conditions (in terms of temperature, pH, pO_2, etc) have been recently developed (Puskeiler, Kaufmann et al. 2005; Betts and Baganz 2006; Nanchen, Schicker et al. 2006; Kusterer, Krause et al. 2008) and are becoming commercially available. These cultivation systems allow highly reproducible growth experiments with small volumes and open the way to more high-throughput methods.

Sampling of metabolites. The metabolite sampling method must be carefully designed to avoid alteration of the labelling patterns of metabolites. The most routine methods for steady-state flux analysis are based on the quantification of isotopes in proteinogenic amino acids for which sampling protocols are easy to set up and well-documented (Zamboni, Fendt et al. 2009). These metabolic end-products provide isotopic information on 8 metabolic precursors of the central carbon metabolism. Because proteinogenic amino acids accumulate during growth, long experimental durations are necessary to reach the isotopic steady-state, and the metabolic steady-state must be maintained during the entire experiment. That constitutes a major limitation to the applicability of these methods to investigate metabolism of non-growing cells or short transient biological changes. The analysis of metabolic intermediates is of special interest because they provide the richest source of information to improve flux precision (Zamboni, Fendt et al. 2009), allowing the analysis of dynamic metabolic systems (Noh and Wiechert 2011), and paving the way toward the resolution of cell-scale metabolic networks. Since most metabolic intermediates have turnover times in the order of seconds and cells rapidly adapt their metabolism in response to environmental changes (Taymaz-Nikerel, de Mey et al. 2009), obtaining a reliable snapshot of their labelling patterns at the time

of collection is not trivial (Zamboni, Fendt et al. 2009). In that case, sampling requires specific procedures (Schaub, Schiesling et al. 2006; Bolten, Kiefer et al. 2007; Hiller, Franco-Lara et al. 2007; Toya, Ishii et al. 2007; van Gulik 2010), e.g. the rapid blockage of metabolic activities in a so-called quenching step.

Quantitative analysis of labelling patterns. Initially based on the analysis of labelling in proteinogenic amino-acids by NMR (Szyperski 1995; Szyperski, Glaser et al. 1999; Massou, Nicolas et al. 2007), ^{13}C-MFA has evolved to their analysis by GC-MS (Christensen and Nielsen 1999; Wittmann 2007), and more recently to the analysis of intracellular intermediates from very small samples using highly sensitive LC-MS/MS methods (in the femtomolar range) (van Winden, van Dam et al. 2005; Kiefer, Nicolas et al. 2007). MS instruments give access to global isotopic information of intact or fragmented metabolites (van Winden, van Dam et al. 2005; Kiefer, Nicolas et al. 2007; Choi and Antoniewicz 2011; Ruhl, Rupp et al. 2012), while NMR gives access to diverse detailed positional information according to the pulse sequence applied (Szyperski, Glaser et al. 1999; Massou, Nicolas et al. 2007; Massou, Nicolas et al. 2007). MS and NMR can also be combined to improve flux precision by increasing the number and nature of isotopic data (Kleijn, Geertman et al. 2007; Nicolas, Kiefer et al. 2007). Because metabolic fluxes are calculated from these data, both high accuracy and high precision of these measurements are needed to avoid mistaken biological conclusions (Antoniewicz, Kelleher et al. 2007).

Data processing. The raw isotopic measurements must be processed to determine the meaningful isotopic information, i.e. the distribution of ^{13}C-tracer atoms into metabolites. In that aim, bioinformatics tools and methods have been developed for both MS- (Wahl, Dauner et al. 2004; Moseley 2010) and NMR-based (Szyperski, Glaser et al. 1999; Tiainen, Maaheimo et al. 2008) analyses.

Flux calculation. Metabolic fluxes are calculated from the isotopic data by a computer-based fitting procedure (Wiechert, Mollney et al. 2001) (Figure 11). From the measured extracellular fluxes and an initial random set of free fluxes, a mathematical model computes the remaining fluxes, simulates the isotopic data for the particular flux distribution, and compares the simulated isotopic data with the experimental ones to calculate the sum of their weighted differences (residuum). The model iteratively adjusts the free fluxes to minimize the residuum until an optimum fit is obtained. Because of the

size and of the non-linearity of the equation system linking metabolic fluxes to isotopic data, this fitting procedure is a slow and unstable process. Recently, new mathematical methods with variable changes have been developed to reduce the dimension of the problem and/or make simulation of isotopic data faster, such as cumomers (Wiechert, Möllney et al. 1999), EMUs (Antoniewicz, Kelleher et al. 2007), fluxomers (Srour, Young et al. 2011), and SCCs (Weitzel, Wiechert et al. 2007).

Sensitivity analysis. Finally, confidence intervals on the calculated fluxes have to be reliably estimated to avoid mistaken biological interpretations. In that aim, several methods have been developed. The first approaches were based on linear methods (Mollney, Wiechert et al. 1999), but recently the more rigorous Monte-Carlo (Quek, Wittmann et al. 2009) or Chi2-inversion (Antoniewicz, Kelleher et al. 2006) methods were introduced, allowed by the increase of calculation power capabilities.

Although recent advances described in this section have addressed some experimental or mathematical limitations and issues, ^{13}C-MFA is still an expensive, complex, and low-throughput tool. Few conditions are usually investigated, in most cases without biological replicates. The growing demand for comprehensive phenotyping in systems biology and synthetic biology is driving the need for more high-throughput approaches allowing the detailed investigation of large sets of organisms, strains, or physiological conditions.

5. General aim and scope of this thesis

The general aim of the work carried out in the team MetaSys of the LISBP is the comprehensive, system-level understanding of metabolic adaptation and its role in the behaviour of micro-organisms. The main challenge is to understand how the bacterium develops an integrated response to changes in its environment or in its genome. Typically such a response involves both metabolic regulation and gene regulation, and our objective is to develop a scalable methodological framework for precisely understanding how metabolic networks and regulatory networks cooperate in the biological response. To address this question, methodologies for the detailed, functional investigation of microbial metabolic networks are developed, including both experimental and modelling approaches. The comprehensive understanding of the system properties and its response to genetic and environmental modifications will allow to understand how environmental conditions can favour the development of bacteria in their ecological

niches as well as the expression of the specific traits, including virulence in pathogenic strains.

The specific aim of this thesis is to understand the role of post-transcriptional regulation in the control of metabolism in *E. coli*. This was made by investigating the role of the Csr system in carbon nutrition and in the control of *E. coli* metabolism. Csr was found to control a broad range of physiological processes allowing *E. coli* to successfully implant and persist in the gut, such as biofilm formation, motility as well as many functions involved in carbon nutrition, including glycolysis, gluconeogenesis, acetate and glycogen metabolism. Thus, the Csr system appears as a key regulator for adaptation of *E. coli* in the gut environment.

However, current knowledge about the control of metabolic processes by Csr is too limited to get comprehensive understanding of its role in the adaptation to the nutritional context of the intestine. So far, the role of the Csr system on *E. coli* metabolism has been investigated with molecular approaches. The actual effect of Csr on the operation of *E. coli* metabolism – i.e. the actual phenotype – remains unknown because i) CsrA likely controls most of the enzymes of its central metabolism and ii) metabolic networks are highly flexible, i.e. changes in the Csr-targeted pathways may be counterbalanced by flux regulatory mechanisms or may lead to modifications in the activity of the entire central metabolism. To obtain a detailed picture of the actual operation of *E. coli* metabolism, a set of quantitative system-level tools have been developed during last decades, such as metabolomics – which aims to identify and quantify metabolites – and ^{13}C-MFA – which aims to quantify metabolic fluxes –. ^{13}C-MFA is particularly well-adapted to understand the role of global regulators on the metabolism (Perrenoud and Sauer 2005; Yao, Hirose et al. 2011), but it is a cumbersome, low-throughput tool.

This thesis encompasses two parts as detailed hereafter. My personal contribution to the different chapters is detailed in annex.

Part 1 addresses some issues of ^{13}C-MFA, with three major aims:

- Bring ^{13}C-MFA into the high-throughput era to allow investigations of large sets of mutants and conditions
- Improve precision and accuracy on metabolic fluxes
- Make easier ^{13}C-MFA experiments for biologists

The methodological developments made in this work to address these goals are described in Chapters 2 to 6. They deal with the development and validation of a set of tools and methods regarding experimental, analytical and mathematical aspects of ^{13}C-MFA in *E. coli*. These methodological developments have been applied for the investigation of *E. coli* metabolism described in part 2.

Part 2 of this thesis (Chapters 7 and 8) addresses fundamental questions regarding the physiology and metabolism of *E. coli*, with three major aims:

- Obtaining metabolic knowledge on *E. coli* on physiologically-relevant carbon sources

The metabolism of *E. coli* on nutrients that support its growth in the gut has been characterised at the molecular level, but its actual operation on these carbon sources is not known. In this thesis, we have performed in-depth investigations of the metabolism of *E. coli* on such carbon sources using a combination of quantitative system-level approaches (high-throughput growth characterization, metabolomics, ^{13}C-MFA).

- Obtaining metabolic knowledge on B2 strains

The metabolism of B2 strains has not been well-documented, although it would provide fundamental knowledge for understanding metabolic specificities of B2 strains that could favour their competitiveness in the gut. Detailed functional investigations of the metabolism of two distinct *E. coli* strains, K12 MG1655 and Nissle 1917, have been performed in this work. The K12 MG1655 laboratory strain, which is the most-studied and the best-known *E. coli* strain, is a poor colonizer of the gut belonging to the phylogenetic group A. In contrast, the Nissle 1917 strain is an efficient colonizer of the gut belonging to the highly competitive B2 phylogenetic group. The differences in the operation of metabolism between the two strains were analysed and interpreted in light of their implantation capabilities in the gut.

- Understanding the role of the Csr system in carbon nutrition and in control of *E. coli* metabolism

Little is known about the actual control exerted by the Csr system over the entire metabolism of *E. coli*, i.e. about the actual phenotypic effects at the metabolic level. To

our knowledge, system-level investigations have not been carried out to address these questions. In addition, the role of Csr has been investigated almost exclusively on glucose, a scarce carbon source in the gut which is not representative of the nutritional environment of *E. coli*. In this work, a detailed metabolic characterisation of a set of Csr mutants (constructed by Olga Revelles and Thomas Esquerre, LISBP) and wild-type strains has been achieved by combining metabolomics and ^{13}C-MFA to provide a complete picture of the actual operation of their metabolism. These analyses have been carried out on a broad range of carbon sources shown to support the growth of *E. coli* in the gut.

6. References

Altier, C., M. Suyemoto, et al. (2000). "Regulation of Salmonella enterica serovar typhimurium invasion genes by csrA." Infect Immun **68**(12): 6790-6797.

Antoniewicz, M. R., J. K. Kelleher, et al. (2006). "Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements." Metab Eng **8**(4): 324-337.

Antoniewicz, M. R., J. K. Kelleher, et al. (2007). "Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis." Anal Chem **79**(19): 7554-7559.

Antoniewicz, M. R., J. K. Kelleher, et al. (2007). "Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions." Metab Eng **9**(1): 68-86.

Aubron, C., J. Glodt, et al. (2012). "Variation in endogenous oxidative stress in Escherichia coli natural isolates during growth in urine." BMC Microbiol **12**(1): 120.

Backhed, F., R. E. Ley, et al. (2005). "Host-bacterial mutualism in the human intestine." Science **307**(5717): 1915-1920.

Bailey, J. K., J. L. Pinyon, et al. (2010). "Distribution of human commensal Escherichia coli phylogenetic groups." J Clin Microbiol **48**(9): 3455-3456.

Baker, C. S., L. A. Eory, et al. (2007). "CsrA inhibits translation initiation of Escherichia coli hfq by binding to a single site overlapping the Shine-Dalgarno sequence." J Bacteriol **189**(15): 5472-5481.

Baker, C. S., I. Morozov, et al. (2002). "CsrA regulates glycogen biosynthesis by preventing translation of glgC in Escherichia coli." Mol Microbiol **44**(6): 1599-1610.

Bernalier-Donadille, A. (2010). "[Fermentative metabolism by the human gut microbiota]." Gastroenterol Clin Biol **34 Suppl 1**: S16-22.

Bernier-Febreau, C., L. du Merle, et al. (2004). "Use of deoxyribose by intestinal and extraintestinal pathogenic Escherichia coli strains: a metabolic adaptation involved in competitiveness." Infect Immun **72**(10): 6151-6156.

Betts, J. I. and F. Baganz (2006). "Miniature bioreactors: current practices and future opportunities." Microb Cell Fact **5**: 21.

Bhatt, S., A. Anyanful, et al. (2011). "CsrA and TnaB Coregulate Tryptophanase Activity To Promote Exotoxin-Induced Killing of Caenorhabditis elegansby Enteropathogenic Escherichia coli." J Bacteriol **193**(17): 4516-4522.

Bhatt, S., A. N. Edwards, et al. (2009). "The RNA binding protein CsrA is a pleiotropic regulator of the locus of enterocyte effacement pathogenicity island of enteropathogenic Escherichia coli." Infect Immun **77**(9): 3552-3568.

Bolten, C. J., P. Kiefer, et al. (2007). "Sampling for metabolome analysis of microorganisms." Anal Chem **79**(10): 3843-3849.

Brown, L., D. Gentry, et al. (2002). "DksA affects ppGpp induction of RpoS at a translational level." J Bacteriol **184**(16): 4455-4465.

Carlos, C., M. M. Pires, et al. (2010). "Escherichia coli phylogenetic group determination and its application in the identification of the major animal source of fecal contamination." BMC Microbiol **10**: 161.

Chang, D. E., D. J. Smalley, et al. (2004). "Carbon nutrition of Escherichia coli in the mouse intestine." Proc Natl Acad Sci U S A **101**(19): 7427-7432.

Chanteloup, N. K., G. Porcheron, et al. (2011). "The extra-intestinal avian pathogenic Escherichia coli strain BEN2908 invades avian and human epithelial cells and survives intracellularly." Vet Microbiol **147**(3-4): 435-439.

Chassard, C., K. P. Scott, et al. (2008). "Assessment of metabolic diversity within the intestinal microbiota from healthy humans using combined molecular and cultural approaches." FEMS Microbiol Ecol **66**(3): 496-504.

Chaudhuri, R. R. and I. R. Henderson (2012). "The evolution of the Escherichia coli phylogeny." Infect Genet Evol **12**(2): 214-226.

Choi, J. and M. R. Antoniewicz (2011). "Tandem mass spectrometry: A novel approach for metabolic flux analysis." Metab Eng **13**(2): 225-233.

Christensen, B. and J. Nielsen (1999). "Isotopomer analysis using GC-MS." Metab Eng **1**(4): 282-290.

Costello, E. K., K. Stagaman, et al. (2012). "The application of ecological theory toward an understanding of the human microbiome." Science **336**(6086): 1255-1262.

Crown, S. B. and M. R. Antoniewicz (2012). "Selection of tracers for 13C-metabolic flux analysis using elementary metabolite units (EMU) basis vector methodology." Metab Eng **14**(2): 150-161.

Croxen, M. A. and B. B. Finlay (2010). "Molecular mechanisms of Escherichia coli pathogenicity." Nat Rev Microbiol **8**(1): 26-38.

Dubey, A. K., C. S. Baker, et al. (2003). "CsrA regulates translation of the Escherichia coli carbon starvation gene, cstA, by blocking ribosome access to the cstA transcript." J Bacteriol **185**(15): 4450-4460.

Duriez, P., O. Clermont, et al. (2001). "Commensal Escherichia coli isolates are phylogenetically distributed among geographically distinct human populations." Microbiology **147**(Pt 6): 1671-1676.

Edwards, A. N., L. M. Patterson-Fortin, et al. (2011). "Circuitry linking the Csr and stringent response global regulatory systems." Mol Microbiol.

Egert, M., A. A. de Graaf, et al. (2006). "Beyond diversity: functional microbiomics of the human colon." Trends Microbiol **14**(2): 86-91.

Escobar-Paramo, P., K. Grenet, et al. (2004). "Large-scale population structure of human commensal Escherichia coli isolates." Appl Environ Microbiol **70**(9): 5698-5700.

Fabich, A. J., S. A. Jones, et al. (2008). "Comparison of carbon nutrition for pathogenic and commensal Escherichia coli strains in the mouse intestine." Infect Immun **76**(3): 1143-1152.

Fischer, E. and U. Sauer (2003). "Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS." Eur J Biochem **270**(5): 880-891.

Fischer, E., N. Zamboni, et al. (2004). "High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints." Anal Biochem **325**(2): 308-316.

Flint, H. J., E. A. Bayer, et al. (2008). "Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis." Nat Rev Microbiol **6**(2): 121-131.

Fujimura, K. E., N. A. Slusher, et al. (2010). "Role of the gut microbiota in defining human health." Expert Rev Anti Infect Ther **8**(4): 435-454.

Gill, S. R., M. Pop, et al. (2006). "Metagenomic analysis of the human distal gut microbiome." Science **312**(5778): 1355-1359.

Giufre, M., C. Graziani, et al. (2012). "Escherichia coli of human and avian origin: detection of clonal groups associated with fluoroquinolone and multidrug resistance in Italy." J Antimicrob Chemother **67**(4): 860-867.

Grasselli, E., P. Francois, et al. (2008). "Evidence of horizontal gene transfer between human and animal commensal Escherichia coli strains identified by microarray." FEMS Immunol Med Microbiol **53**(3): 351-358.

Graziani, C., I. Luzzi, et al. (2009). "Phylogenetic background and virulence genotype of ciprofloxacin-susceptible and ciprofloxacin-resistant Escherichia coli strains of human and avian origin." J Infect Dis **199**(8): 1209-1217.

Gudapaty, S., K. Suzuki, et al. (2001). "Regulatory interactions of Csr components: the RNA binding protein CsrA activates csrB transcription in Escherichia coli." J Bacteriol **183**(20): 6017-6027.

Haverkorn van Rijsewijk, B. R., A. Nanchen, et al. (2011). "Large-scale 13C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in Escherichia coli." Mol Syst Biol **7**: 477.

Heijtz, R. D., S. Wang, et al. (2011). "Normal gut microbiota modulates brain development and behavior." Proc Natl Acad Sci U S A **108**(7): 3047-3052.

Herren, C. D., A. Mitra, et al. (2006). "The BarA-UvrY two-component system regulates virulence in avian pathogenic Escherichia coli O78:K80:H9." Infect Immun **74**(8): 4900-4909.

Hiller, J., E. Franco-Lara, et al. (2007). "Fast sampling and quenching procedures for microbial metabolic profiling." Biotechnol Lett **29**(8): 1161-1167.

Hoskins, L. C. and E. T. Boulding (1981). "Mucin degradation in human colon ecosystems. Evidence for the existence and role of bacterial subpopulations producing glycosidases as extracellular enzymes." J Clin Invest **67**(1): 163-172.

Ishii, N., K. Nakahigashi, et al. (2007). "Multiple high-throughput analyses monitor the response of E. coli to perturbations." Science **316**(5824): 593-597.

Jonas, K., A. N. Edwards, et al. (2008). "The RNA binding protein CsrA controls cyclic di-GMP metabolism by directly regulating the expression of GGDEF proteins." Mol Microbiol **70**(1): 236-257.

Jonas, K. and O. Melefors (2009). "The Escherichia coli csrB and csrC small RNAs are strongly induced during growth in nutrient-poor medium." FEMS Microbiol Lett **297**(1): 80-86.

Jonas, K., H. Tomenius, et al. (2006). "Identification of YhdA as a regulator of the Escherichia coli carbon storage regulation system." FEMS Microbiol Lett **264**(2): 232-237.

Kamada, N., Y. G. Kim, et al. (2012). "Regulated virulence controls the ability of a pathogen to compete with the gut microbiota." Science **336**(6086): 1325-1329.

Kaper, J. B., J. P. Nataro, et al. (2004). "Pathogenic Escherichia coli." Nat Rev Microbiol **2**(2): 123-140.

Kiefer, P., C. Nicolas, et al. (2007). "Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry." Anal Biochem **360**(2): 182-188.

Kleijn, R. J., J. M. Geertman, et al. (2007). "Metabolic flux analysis of a glycerol-overproducing Saccharomyces cerevisiae strain based on GC-MS, LC-MS and NMR-derived C-labelling data." FEMS Yeast Res **7**(2): 216-231.

Kusterer, A., C. Krause, et al. (2008). "Fully automated single-use stirred-tank bioreactors for parallel microbial cultivations." Bioprocess Biosyst Eng **31**(3): 207-215.

Leitch, E. C., A. W. Walker, et al. (2007). "Selective colonization of insoluble substrates by human faecal bacteria." Environ Microbiol **9**(3): 667-679.

Ley, R. E., D. A. Peterson, et al. (2006). "Ecological and evolutionary forces shaping microbial diversity in the human intestine." Cell **124**(4): 837-848.

Liu, M. Y., G. Gui, et al. (1997). "The RNA molecule csrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli." J Biol Chem **272**(28): 17502-17510.

Lucchetti-Miganeh, C., E. Burrowes, et al. (2008). "The post-transcriptional regulator CsrA plays a central role in the adaptation of bacterial pathogens to different stages of infection in animal hosts." Microbiology **154**(Pt 1): 16-29.

Maharjan, R. P. and T. Ferenci (2005). "Metabolomic diversity in the species Escherichia coli and its relationship to genetic population structure." Metabolomics **1**(3): 235-242.

Martinez-Jehanne, V., L. du Merle, et al. (2009). "Role of deoxyribose catabolism in colonization of the murine intestine by pathogenic Escherichia coli strains." Infect Immun **77**(4): 1442-1450.

Massou, S., C. Nicolas, et al. (2007). "Application of 2D-TOCSY NMR to the measurement of specific(13C-enrichments in complex mixtures of 13C-labeled metabolites." Metab Eng **9**(3): 252-257.

Massou, S., C. Nicolas, et al. (2007). "NMR-based fluxomics: quantitative 2D NMR methods for isotopomers analysis." Phytochemistry **68**(16-18): 2330-2340.

Mercante, J., A. N. Edwards, et al. (2009). "Molecular geometry of CsrA (RsmA) binding to RNA and its implications for regulated expression." J Mol Biol **392**(2): 511-528.

Miranda, R. L., T. Conway, et al. (2004). "Glycolytic and gluconeogenic growth of Escherichia coli O157:H7 (EDL933) and E. coli K-12 (MG1655) in the mouse intestine." Infect Immun **72**(3): 1666-1676.

Mollney, M., W. Wiechert, et al. (1999). "Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments." Biotechnol Bioeng **66**(2): 86-103.

Moseley, H. N. (2010). "Correcting for the effects of natural abundance in stable isotope resolved metabolomics experiments involving ultra-high resolution mass spectrometry." BMC Bioinformatics **11**: 139.

Moulin-Schouleur, M., M. Reperant, et al. (2007). "Extraintestinal pathogenic Escherichia coli strains of avian and human origin: link between phylogenetic relationships and common virulence patterns." J Clin Microbiol **45**(10): 3366-3376.

Moulin-Schouleur, M., C. Schouler, et al. (2006). "Common virulence factors and genetic relationships between O18:K1:H7 Escherichia coli isolates of human and avian origin." J Clin Microbiol **44**(10): 3484-3492.

Muegge, B. D., J. Kuczynski, et al. (2011). "Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans." Science **332**(6032): 970-974.

Murray, E. L. and T. Conway (2005). "Multiple regulators control expression of the Entner-Doudoroff aldolase (Eda) of Escherichia coli." J Bacteriol **187**(3): 991-1000.

Nanchen, A., A. Schicker, et al. (2006). "Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of Escherichia coli." Appl Environ Microbiol **72**(2): 1164-1172.

Nicolas, C., P. Kiefer, et al. (2007). "Response of the central metabolism of Escherichia coli to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase." FEBS Lett **581**(20): 3771-3776.

Noh, K. and W. Wiechert (2011). "The benefits of being transient: isotope-based metabolic flux analysis at the short time scale." Appl Microbiol Biotechnol **91**(5): 1247-1265.

Nougayrede, J. P., S. Homburg, et al. (2006). "Escherichia coli induces DNA double-strand breaks in eukaryotic cells." Science **313**(5788): 848-851.

Nowrouzian, F. L., I. Adlerberth, et al. (2006). "Enhanced persistence in the colonic microbiota of Escherichia coli strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells." Microbes Infect **8**(3): 834-840.

Nowrouzian, F. L. and E. Oswald (2012). "Escherichia coli strains with the capacity for long-term persistence in the bowel microbiota carry the potentially genotoxic pks island." Microb Pathog.

Nowrouzian, F. L., A. E. Wold, et al. (2005). "Escherichia coli strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants." J Infect Dis **191**(7): 1078-1083.

O'Hara, A. M. and F. Shanahan (2006). "The gut flora as a forgotten organ." EMBO Rep **7**(7): 688-693.

Oh, M. K., L. Rohlin, et al. (2002). "Global expression profiling of acetate-grown Escherichia coli." J Biol Chem **277**(15): 13175-13183.

Orth, J. D., I. Thiele, et al. (2010). "What is flux balance analysis?" Nature Biotechnology **28**(3): 245-248.

Palaniyandi, S., A. Mitra, et al. (2012). "BarA-UvrY two-component system regulates virulence of uropathogenic E. coli CFT073." PLoS One **7**(2): e31348.

Pannuri, A., H. Yakhnin, et al. (2012). "Translational repression of NhaR, a novel pathway for multi-tier regulation of biofilm circuitry by CsrA." J Bacteriol **194**(1): 79-89.

Pernestig, A. K., D. Georgellis, et al. (2003). "The Escherichia coli BarA-UvrY two-component system is needed for efficient switching between glycolytic and gluconeogenic carbon sources." J Bacteriol **185**(3): 843-853.

Perrenoud, A. and U. Sauer (2005). "Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli." J Bacteriol **187**(9): 3171-3179.

Puskeiler, R., K. Kaufmann, et al. (2005). "Development, parallelization, and automation of a gas-inducing milliliter-scale bioreactor for high-throughput bioprocess design (HTBD)." Biotechnol Bioeng **89**(5): 512-523.

Quek, L. E., C. Wittmann, et al. (2009). "OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis." Microb Cell Fact **8**: 25.

Romeo, T., M. Gong, et al. (1993). "Identification and molecular characterization of csrA, a pleiotropic gene from Escherichia coli that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties." J Bacteriol **175**(15): 4744-4755.

Romeo, T., C. A. Vakulskas, et al. (2012). "Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems." Environ Microbiol.

Rouquet, G., G. Porcheron, et al. (2009). "A metabolic operon in extraintestinal pathogenic Escherichia coli promotes fitness under stressful conditions and invasion of eukaryotic cells." J Bacteriol **191**(13): 4427-4440.

Ruhl, M., B. Rupp, et al. (2012). "Collisional fragmentation of central carbon metabolites in LC-MS/MS increases precision of (1)(3)C metabolic flux analysis." Biotechnol Bioeng **109**(3): 763-771.

Russo, T. A. and J. R. Johnson (2003). "Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem." Microbes Infect **5**(5): 449-456.

Sabarly, V., O. Bouvet, et al. (2011). "The decoupling between genetic structure and metabolic phenotypes in Escherichia coli leads to continuous phenotypic diversity." J Evol Biol **24**(7): 1559-1571.

Sabnis, N. A., H. Yang, et al. (1995). "Pleiotropic regulation of central carbohydrate metabolism in Escherichia coli via the gene csrA." J Biol Chem **270**(49): 29096-29104.

Savage, D. C. (1977). "Microbial ecology of the gastrointestinal tract." Annu Rev Microbiol **31**: 107-133.

Schaub, J., C. Schiesling, et al. (2006). "Integrated sampling procedure for metabolome analysis." Biotechnol Prog **22**(5): 1434-1442.

Schouler, C., A. Taki, et al. (2009). "A genomic island of an extraintestinal pathogenic Escherichia coli Strain enables the metabolism of fructooligosaccharides, which improves intestinal colonization." J Bacteriol **191**(1): 388-393.

Schubert, M., K. Lapouge, et al. (2007). "Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA." Nat Struct Mol Biol **14**(9): 807-813.

Schuetz, R., L. Kuepfer, et al. (2007). "Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli." Mol Syst Biol **3**: 119.

Sekirov, I., S. L. Russell, et al. (2010). "Gut microbiota in health and disease." Physiol Rev **90**(3): 859-904.

Shiloach, J., S. Reshamwala, et al. (2010). "Analyzing metabolic variations in different bacterial strains, historical perspectives and current trends--example E. coli." Curr Opin Biotechnol **21**(1): 21-26.

Shimizu, K. (2004). "Metabolic flux analysis based on 13C-labeling experiments and integration of the information with gene and protein expression patterns." Adv Biochem Eng Biotechnol **91**: 1-49.

Snider, T. A., A. J. Fabich, et al. (2009). "E. coli O157:H7 catabolism of intestinal mucin-derived carbohydrates and colonization." Vet Microbiol **136**(1-2): 150-154.

Srour, O., J. D. Young, et al. (2011). "Fluxomers: A new approach for 13C metabolic flux analysis." BMC Syst Biol **5**(1): 129.

Suzuki, K., P. Babitzke, et al. (2006). "Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs csrB and csrC for degradation by RNase E." Genes Dev **20**(18): 2605-2617.

Suzuki, K., X. Wang, et al. (2002). "Regulatory circuitry of the CsrA/csrB and BarA/UvrY systems of Escherichia coli." J Bacteriol **184**(18): 5130-5140.

Szyperski, T. (1995). "Biosynthetically directed fractional 13C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism." Eur J Biochem **232**(2): 433-448.

Szyperski, T., R. W. Glaser, et al. (1999). "Bioreaction network topology and metabolic flux ratio analysis by biosynthetic fractional 13C labeling and two-dimensional NMR spectroscopy." Metab Eng **1**(3): 189-197.

Taymaz-Nikerel, H., M. de Mey, et al. (2009). "Development and application of a differential method for reliable metabolome analysis in Escherichia coli." Anal Biochem **386**(1): 9-19.

Tiainen, M., H. Maaheimo, et al. (2008). "Spectral analysis of 1H coupled 13C spectra of the amino acids: adaptive spectral library of amino acid 13C isotopomers and positional fractional 13C enrichments." Magn Reson Chem **46**(2): 125-137.

Timmermans, J. and L. Van Melderen (2009). "Conditional essentiality of the csrA gene in Escherichia coli." J Bacteriol **191**(5): 1722-1724.

Tivendale, K. A., C. M. Logue, et al. (2010). "Avian-pathogenic Escherichia coli strains are similar to neonatal meningitis E. coli strains and are able to cause meningitis in the rat model of human disease." Infect Immun **78**(8): 3412-3419.

Touchon, M., C. Hoede, et al. (2009). "Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths." PLoS Genet **5**(1): e1000344.
Toya, Y., N. Ishii, et al. (2007). "Direct measurement of isotopomer of intracellular metabolites using capillary electrophoresis time-of-flight mass spectrometry for efficient metabolic flux analysis." J Chromatogr A **1159**(1-2): 134-141.
Turnbaugh, P. J., R. E. Ley, et al. (2006). "An obesity-associated gut microbiome with increased capacity for energy harvest." Nature **444**(7122): 1027-1031.
van Gulik, W. M. (2010). "Fast sampling for quantitative microbial metabolomics." Curr Opin Biotechnol **21**(1): 27-34.
van Winden, W. A., J. C. van Dam, et al. (2005). "Metabolic-flux analysis of Saccharomyces cerevisiae CEN.PK113-7D based on mass isotopomer measurements of (13)C-labeled primary metabolites." FEMS Yeast Res **5**(6-7): 559-568.
Varma, A., B. W. Boesch, et al. (1993). "Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates." Appl Environ Microbiol **59**(8): 2465-2473.
Vieira, G., V. Sabarly, et al. (2011). "Core and panmetabolism in Escherichia coli." J Bacteriol **193**(6): 1461-1472.
Waegeman, H., J. Beauprez, et al. (2011). "Effect of iclR and arcA knockouts on biomass formation and metabolic fluxes in Escherichia coli K12 and its implications on understanding the metabolism of Escherichia coli BL21 (DE3)." BMC Microbiol **11**: 70.
Wahl, S. A., M. Dauner, et al. (2004). "New tools for mass isotopomer data evaluation in (13)C flux analysis: mass isotope correction, data consistency checking, and precursor relationships." Biotechnol Bioeng **85**(3): 259-268.
Walther, J. L., C. M. Metallo, et al. (2012). "Optimization of 13C isotopic tracers for metabolic flux analysis in mammalian cells." Metab Eng **14**(2): 162-171.
Wang, X., A. K. Dubey, et al. (2005). "CsrA post-transcriptionally represses pgaABCD, responsible for synthesis of a biofilm polysaccharide adhesin of Escherichia coli." Mol Microbiol **56**(6): 1648-1663.
Wei, B., S. Shin, et al. (2000). "Global regulatory mutations in csrA and rpoS cause severe central carbon stress in Escherichia coli in the presence of acetate." J Bacteriol **182**(6): 1632-1640.
Wei, B. L., A. M. Brun-Zinkernagel, et al. (2001). "Positive regulation of motility and flhDC expression by the RNA-binding protein CsrA of Escherichia coli." Mol Microbiol **40**(1): 245-256.
Weilbacher, T., K. Suzuki, et al. (2003). "A novel sRNA component of the carbon storage regulatory system of Escherichia coli." Mol Microbiol **48**(3): 657-670.
Weitzel, M., W. Wiechert, et al. (2007). "The topology of metabolic isotope labeling networks." BMC Bioinformatics **8**: 315.
Wiechert, W. (2001). "13C metabolic flux analysis." Metab Eng **3**(3): 195-206.
Wiechert, W., M. Möllney, et al. (1999). "Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems." Biotechnol Bioeng **66**(2): 69-85.
Wiechert, W., M. Mollney, et al. (2001). "A universal framework for 13C metabolic flux analysis." Metab Eng **3**(3): 265-283.
Wielgoss, S., J. E. Barrick, et al. (2011). "Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With Escherichia coli." G3 (Bethesda) **1**(3): 183-186.
Wittmann, C. (2007). "Fluxome analysis using GC-MS." Microb Cell Fact **6**: 6.
Wu, G. D., J. Chen, et al. (2011). "Linking long-term dietary patterns with gut microbial enterotypes." Science **334**(6052): 105-108.
Yakhnin, H., A. V. Yakhnin, et al. (2011). "Complex regulation of the global regulatory gene csrA: CsrA-mediated translational repression, transcription from five promoters by Esigma(7)(0) and Esigma(S), and indirect transcriptional activation by CsrA." Mol Microbiol **81**(3): 689-704.
Yao, R., Y. Hirose, et al. (2011). "Catabolic regulation analysis of Escherichia coli and its crp, mlc, mgsA, pgi and ptsG mutants." Microb Cell Fact **10**(1): 67.
Zamboni, N., S. M. Fendt, et al. (2009). "(13)C-based metabolic flux analysis." Nat Protoc **4**(6): 878-892.
Zhang, H., J. K. DiBaise, et al. (2009). "Human gut microbiota in obesity and after gastric bypass." Proc Natl Acad Sci U S A **106**(7): 2365-2370.
Zhao, J. and K. Shimizu (2003). "Metabolic flux analysis of Escherichia coli K12 grown on 13C-labeled acetate and glucose using GC-MS and powerful flux calculation method." J Biotechnol **101**(2): 101-117.

Part 1

Methodological developments for ^{13}C-metabolic flux analysis

To investigate the role of Csr in the control of *E. coli* metabolism, detailed investigations of different strains and mutants in various physiological conditions have to be performed. ^{13}C-Metabolic Flux Analysis (^{13}C-MFA) provides the most accurate picture of the actual behaviour of microbial metabolism and is of special interest in functional genomics and systems biology to establish the genotype-phenotype relationship. However, ^{13}C-MFA is still an expensive, cumbersome, and low-throughput tool. In this section, we present experimental, analytical and computational developments for high-throughput ^{13}C-MFA. A special care has also been taken to make ^{13}C-MFA more accessible to a wider biological community.

Chapter 2 is related to the experimental design of ^{13}C-MFA experiments. This step allows optimizing the flux information obtained from a single experiment and thus is of special value for high-throughput ^{13}C-MFA. We have developed the software IsoDesign to maximise both the number of fluxes that can be calculated and their precision. IsoDesign optimises the isotopic composition of the label input (for mixtures up to 6 isotopic forms), and can exploit quantitative isotopic measurements obtained from a broad panel of analytical tools (MS, MS/MS, ^{1}H-NMR, ^{13}C-NMR, etc). It provides a visualisation module to intuitively select the optimal label input according to the biological question to be addressed by the investigator.

Chapter 3 is devoted to the development of a robust method for collecting reliable isotopic data from intracellular metabolites. The analysis of metabolic intermediates is of special interest for high-throughput ^{13}C-MFA because they provide the richest source of information from very small samples (< 1 mL) without requiring complicated sample preparation (such as hydrolysis and derivatization). However, obtaining a reliable measurement of their labelling patterns at the time of collection is not trivial and requires specific procedures. In this chapter we rigorously evaluate five distinct methods to collect intracellular metabolites of *E. coli* for ^{13}C-MFA. We show that the fast filtration procedure does not provide reliable isotopic information, while methods based on cold or boiling solvents are appropriated for efficient quenching of metabolism. In addition, the simplest sampling procedure is fully suitable with high-throughput ^{13}C-MFA and can be automated for high-throughput platforms. We also propose a strategy to evaluate sampling methods for future ^{13}C-MFA approaches based on time course measurements of label incorporation in metabolic intermediates.

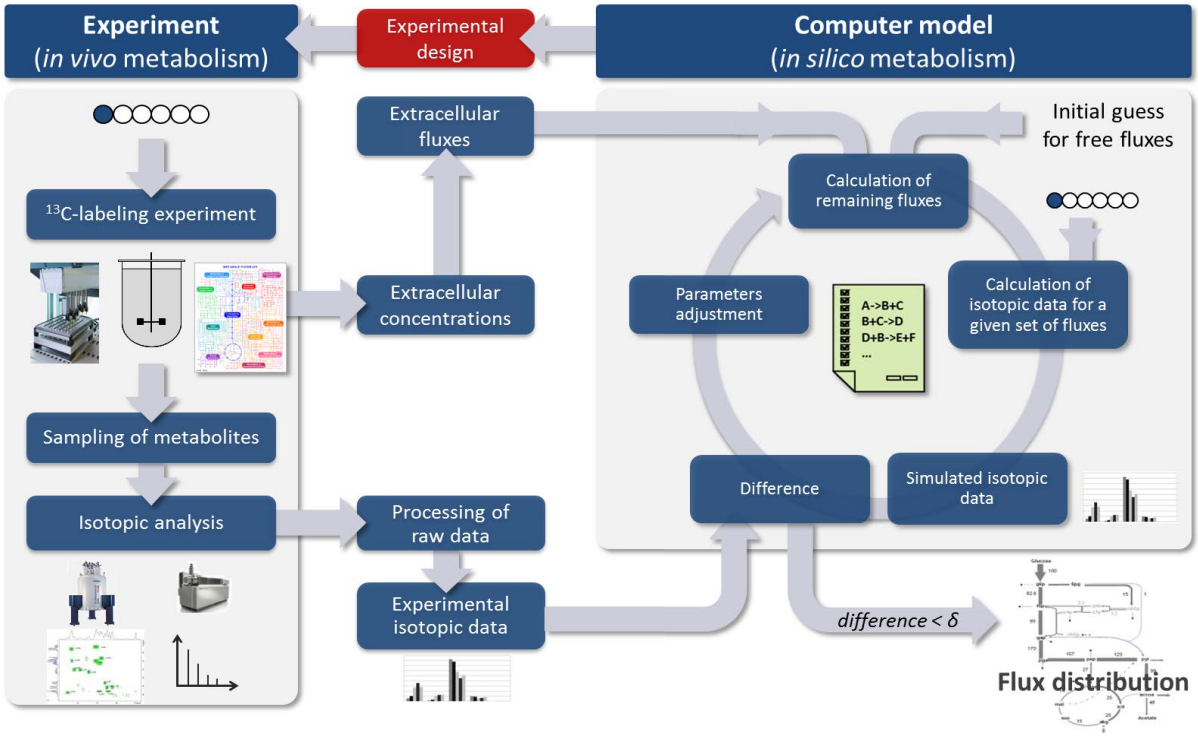
Chapter 4 addresses questions related to the quality of the isotopic data collected by mass spectrometry. The reliability of a ^{13}C-MFA experiment strongly depends on the quality of isotopic measurements, but there is a lack of methods to investigate their precision and accuracy. In this chapter, we provide a complete conceptual and practical framework for full control of the labelling patterns of biologically-produced metabolites. Such cellular samples have been produced using *P. augusta* and *E. coli*, and analysed on our analytical platform (ionic chromatography coupled to a triple quadrupole instrument) for which the precision and accuracy of quantitative isotopic measurements have been investigated in great detail.

Chapter 5 deals with the processing of isotopic data collected by mass spectrometry. The raw MS data must be corrected for the contribution of all naturally abundant isotopes to extract the meaningful isotopic information. We have implemented this correction process in IsoCor, a software that allows such correction to be applied to any chemical species and provides new features – e.g. correction for the isotopic purity of the tracer – to improve the accuracy of isotope labelling studies. IsoCor is a flexible user-friendly tool designed to process the large isotopic datasets that can be acquired with modern high-throughput MS technologies, and is therefore fully suitable for high-throughput ^{13}C-MFA.

Chapter 6 focuses on the development of a fast and precise algorithm for flux calculation. Several algorithms and software packages are available for solving this problem, but they require a lot of time to achieve acceptable precision and often lack of convergence stability. In this chapter, we present a new deterministic algorithm (NLSIC) with significantly increased numerical stability and accuracy of flux estimation compared with commonly used algorithms. It requires relatively short CPU time (from several seconds to several minutes with a standard PC architecture) to estimate fluxes in the central carbon metabolism network of *E. coli*, and thus represents an additional and valuable step towards high-throughput ^{13}C-MFA.

Chapter 2

IsoDesign: rationalizing ^{13}C-metabolic flux analysis by *in silico* optimisation of the experimental design



NOTE: *this chapter contains the first draft of a manuscript in preparation for Bioinformatics.*

1. Abstract

The growing demand for ^{13}C-metabolic flux analysis (^{13}C-MFA) in the field of systems biology and metabolic engineering is driving the need for rationalizing the expensive and time-consuming ^{13}C-labelling experiments. Experimental design is a key step to improve both the number of fluxes that are calculable and their precision. IsoDesign is a software that allows to maximise these parameters by optimising isotopic composition of the label input (for mixtures up to 6 isotopic forms), and can exploit quantitative isotopic measurements obtained from a broad panel of analytical tools (MS, MS/MS, ^{1}H-NMR, ^{13}C-NMR, etc). It provides a visualisation module to intuitively select the optimal label input according to the biological question to be addressed and to the labelling data that can be made available. Its user-friendly interface makes experimental design of ^{13}C-MFA experiments more accessible to a wider biological community.

Availability: IsoDesign is distributed under OpenSource licence at http://metasys.insa-toulouse.fr/software/isodesign/

2. Introduction

^{13}C-Metabolic flux analysis (^{13}C-MFA) aims to quantify metabolic fluxes, i.e. the actual rates of biochemical reactions under given physiological conditions. During the last decade, this tool has been increasingly used to provide novel biological insights in systems biology (Ishii, Nakahigashi et al. 2007; Nicolas, Kiefer et al. 2007) and to improve industrial processes in biotechnology (van Gulik, de Laat et al. 2000; Iwatani, Yamada et al. 2008). In practice, ^{13}C-MFA relies on complex, expensive, and time-consuming investigations requiring the combination of ^{13}C-labelling experiments and mathematical tools to calculate fluxes from the labelling patterns of metabolites (Wiechert 2001). The experimental design (Mollney, Wiechert et al. 1999) consists in optimising the isotopic composition of the label input in the ^{13}C-labelling experiments to:

- Maximise the number of fluxes that can be calculated from a single experiment
- Maximise the precision on flux values
- Maximise both parameters in the same time

The design step plays a key role in maximizing the flux information with minimal costs and efforts. A set of tools and methods have been recently developed to determine the optimal label input for a given ^{13}C-MFA experiment (Mollney, Wiechert et al. 1999; Wiechert, Mollney et al. 2001; Chen, Zheng et al. 2007; Libourel, Gehan et al. 2007; Quek, Wittmann et al. 2009; Crown, Ahn et al. 2012; Walther, Metallo et al. 2012). These methods require flux values, and most of them are restricted to MS-based ^{13}C-MFA. Recently, a new approach based on EMU (elementary metabolite units) basis vectors has been proposed to decouple input labelling from flux values, but it is also restricted to MS-based ^{13}C-MFA (Crown and Antoniewicz 2012).

Here, we present IsoDesign, a flexible tool for the experimental design of ^{13}C-MFA, which includes the following features:

- IsoDesign can determine the optimal composition of label inputs containing up to 6 different isotopic forms.
- The user can easily define lower and upper bounds for each isotopic form to mix in input. This can be useful for taking into account cost considerations (e.g. by fixing upper bounds for the proportions of expensive isotopic forms).
- All kinds of isotopic data can be included (MS, MS/MS, ^{1}H-NMR, ^{13}C-NMR, etc) to cover the current and future panel of analytical tools dedicated to quantitative isotopic analysis.
- Linear and non-linear statistical methods are available to make the calculations faster and improve the reliability of the results, respectively.
- Several scoring criteria can be applied to compare label inputs according to the biological question to be addressed.
- The output of IsoDesign is a sensitivity landscape which displays in an intuitive way information on flux(es) of interest yielded by label inputs. This makes the choice of the optimal input easier for the investigator.

3. Method and implementation

The general strategy for selection of optimal label input is presented figure 1. It consists in calculating for a particular metabolic network the sensitivity of flux values towards the composition of the label input. Briefly, a series of models of the same metabolic network with different label inputs is generated. The labelling patterns of metabolites that can be measured experimentally are simulated for each model, and a set of statistical tools is applied on the completed set of simulated data to determine the precision on each flux. These results are used to calculate a score for each input according to criteria defined by the investigator. The optimal input is selected in a visual way on the basis of this score.

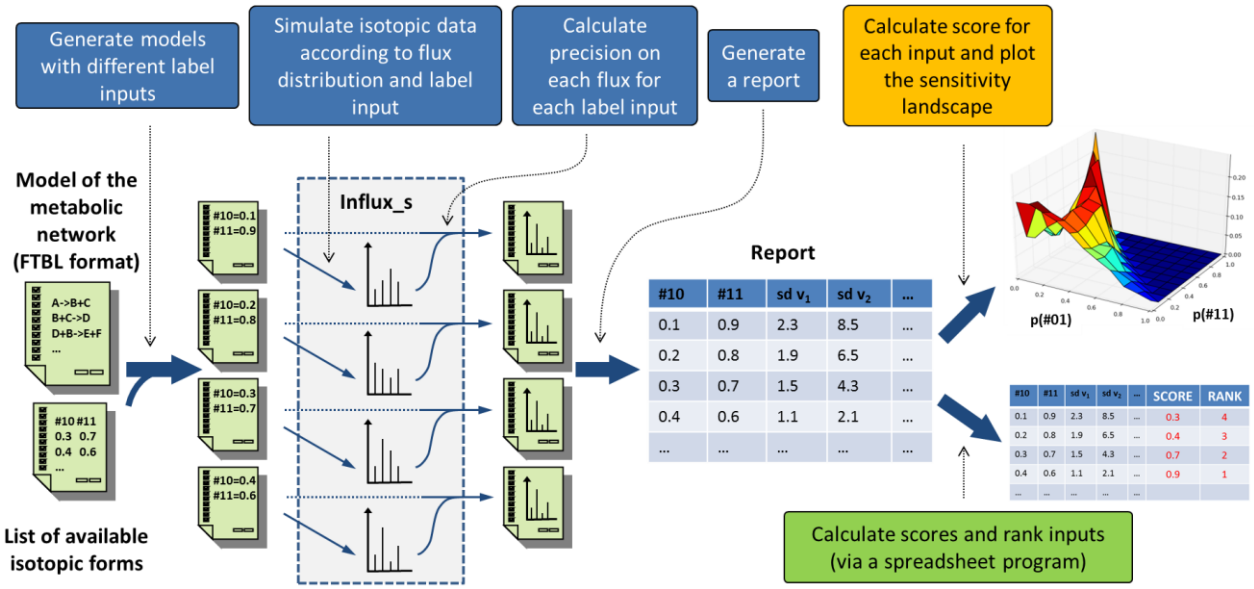


Figure 1. Workflow implemented in IsoDesign to determine the optimal label input in a ^{13}C-MFA experiment. Blue, orange and green boxes detail steps performed by the calculation module, the visualisation module and via a spreadsheet program, respectively.

The tracer evaluation process requires three inputs:

- The definition of the metabolic network – including the set of reaction and the carbon atom transitions of each reaction – to be investigated.
- The assumed flux distribution in the physiological conditions to be examined. It may be obtained from the literature for similar conditions, strains or organisms, or computed with relevant tools such as Flux Balance Analysis (Orth, Thiele et al.

2010). Large differences between assumptions and actual flux distribution may affect the optimal label input.

- The nature of labelling data – and their precision – that can be made available experimentally. Since the optimal label input can vary according to the nature of the isotopic data that are measured, the investigator must define which measurements are available to him/her.

The IsoDesign workflow has been implemented in two distinct modules dedicated to calculation and to data visualisation, respectively, both developed in Python programing language (http://python.org). In the following section a brief overview of the software usage, including important practical considerations, is presented. Complete usage of IsoDesign is detailed in the tutorial provided with the software.

4. Usage and practical considerations

Calculation module

A model of the metabolic network to be investigated, containing the network reactions, the various isotopic forms to evaluate as input, and the measurable isotopic data (including measurement precision) must be loaded in FTBL format (http://www.13cflux.net/13cflux/). The user can adjust the lower bound, upper bound, and increment step for each isotopic form to be mixed in label input (figure 2). IsoDesign generates a list of label inputs and calculates the precision on each flux for each input using influx_s (Sokol, Millard et al. 2011), for which options can be adapted by the user. Thus, IsoDesign benefits of all influx_s' features, including the highly stable, accurate and fast NLSIC algorithm. After calculation, IsoDesign generates an output file containing for each label input i) the precision on each flux that can be calculated and ii) the list of fluxes that cannot be calculated from the defined experimental conditions. This file can be easily edited and is the input of the visualisation module.

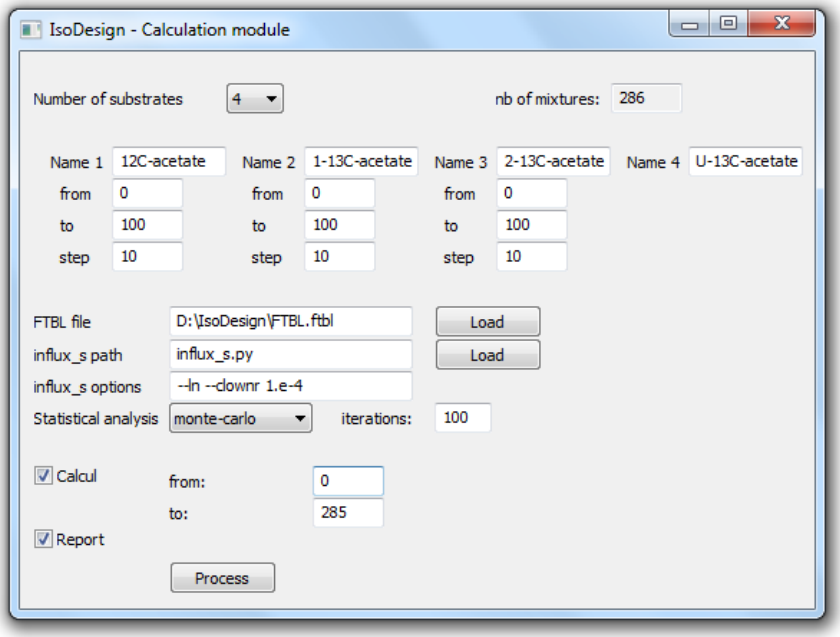


Figure 2. Screenshot of the calculation module of IsoDesign.

Two calculation methods are available to determine the precision on individual flux values: i) a linear method, based on covariance between labelling data and fluxes, which runs fast (from minutes to hour depending on the number of label inputs) but provides overestimated confidence intervals on the fluxes (Antoniewicz, Kelleher et al. 2006), and ii) a Monte-Carlo method, which gives reliable confidence intervals but requires more computational time (from hours to day) (Antoniewicz, Kelleher et al. 2006; Quek, Wittmann et al. 2009). However, Monte-Carlo and linear methods should provide the same optimal label input (Mollney, Wiechert et al. 1999).

Visualisation module

The visualisation module (figure 3) has been developed to make easier and more intuitive the selection of the optimal label input.

The user must load the output file generated by the calculation module, select the axis on which each isotopic form will be plotted, and select a criterion for which a score is calculated for each label input. The available scoring criteria include i) the number of identified fluxes (i.e. fluxes that can be calculated), ii) the sum of flux standard deviations, and the number of fluxes having a precision iii) higher or iv) lower than a threshold defined by the user. The score can be calculated for the entire set of fluxes, or for a selected subset of fluxes. The choice of the criteria and of the (set of) fluxes for

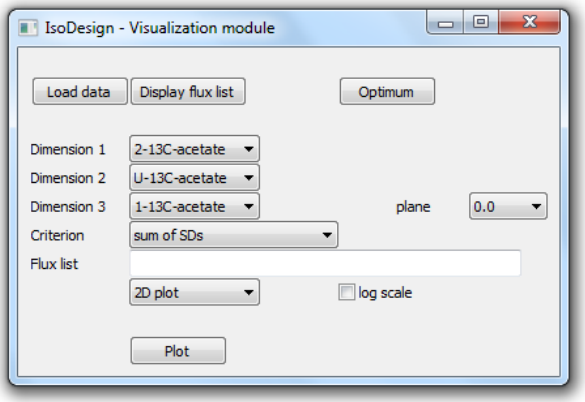


Figure 3. Screenshot of IsoDesign visualisation module.

which the score is calculated mainly depends on the biological question to be addressed (e.g. to maximize precision on a pathway of interest). The user can click on the "Optimum" button to identify inputs having the higher and lower scores. Visual comparison of input can be performed by this module for input mixtures containing up to 4 isotopic forms. The output is a 2D- or 3D-sensitivity landscape – a plot of the score value as function of the isotopic composition of the label input – from which the user can easily identify the optimal input. Comparative analyses of sensitivity landscapes based on different scoring criterion are strongly encouraged to improve the precision of ^{13}C-MFA experiments (see example in figure 4).

5. Application for the optimal design of label inputs

For illustration purpose, IsoDesign was used to design a ^{13}C-MFA experiment aiming to determine the flux distribution in *Escherichia coli* K12 MG1655 grown on acetate, a C_2-carbon source for which the 4 existing isotopic forms are all commercially available. The metabolic network and the assumed flux distribution were taken from literature (Zhao and Shimitzu, 2002). The isotopic data considered for the simulations were the isotopologue abundances of intracellular metabolites that can be measured using the sensitive LC-MS/MS of (Kiefer, Nicolas et al. 2007). The actual content in intracellular metabolites was verified from the LC-MS/MS analysis of samples collected from *E. coli* cells grown on unlabelled acetate. The precision on the isotopic data was determined according to the methodology developed in chapter 4. The proportions of each of the four isotopic forms of acetate in the input were varied from 0 to 100% by step of 10%, which represents a total of 286 label inputs. The precision on fluxes was selected as

scoring parameter, and was calculated for the entire flux dataset using Monte-Carlo method (100 iterations).

Figure 4 shows examples of sensitivity landscapes generated by IsoDesign. As intuitively found, label inputs containing only fully labelled or only unlabelled acetate yield no flux information. By comparing the various sensitivity landscapes, it clearly appears that both the number of calculable fluxes and precision on flux values are maximal when the label input consists in pure 2-^{13}C-acetate. IsoDesign predictions are in good agreements with the flux precision experimentally obtained by performing the ^{13}C-MFA experiment with 2-^{13}C-acetate as label input (data not shown).

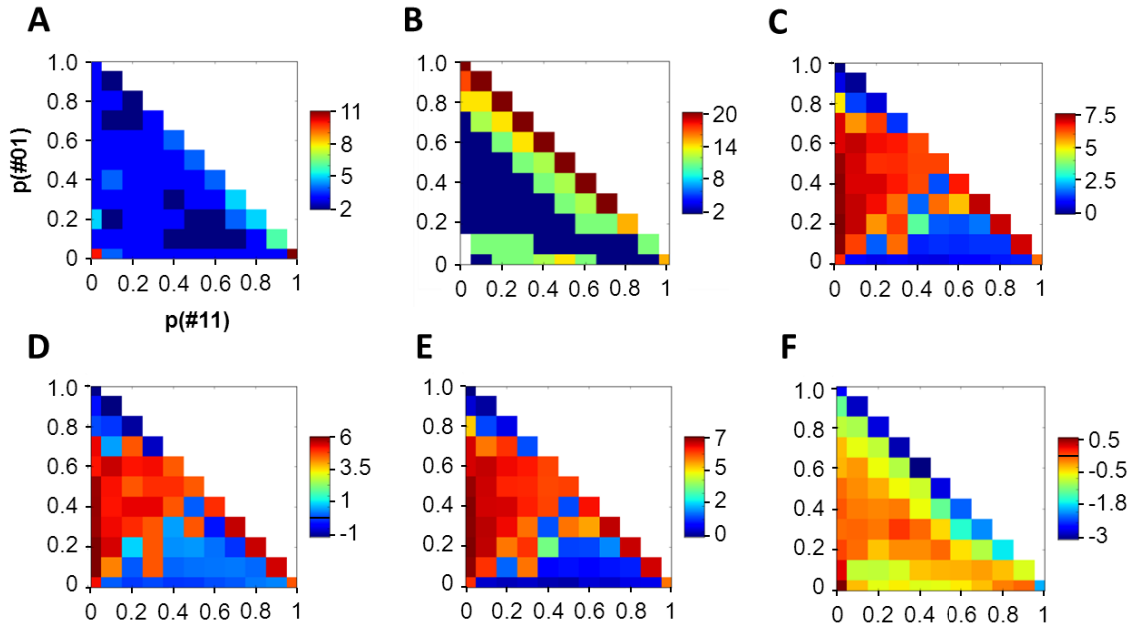


Figure 4. Sensitivity landscapes: number of unidentifiable fluxes (A), number of fluxes having a standard deviation (SD) lower than 0.05 (B), sum of SDs of net fluxes (C), sum of SDs of fluxes of particular pathways – gluconeogenesis (D), TCA cycle (E), pentose phosphate pathway (F). Sum of SDs are in logarithmic scale. p(#01) and p(#11) refer to the relative proportions of 2-^{13}C-acetate and U-^{13}C-acetate in the label input, respectively. Proportion of ^{12}C-acetate is 1-p(#01)-p(#11), and there is no 1-^{13}C-acetate.

In addition, the relationship between the input and scoring criterion is usually considered to be convex (Mollney, Wiechert et al. 1999; Quek, Wittmann et al. 2009), which makes possible the use of linear optimisation algorithm to find the (global) optimum. In this work, we observe that the solution space is sometimes non-convex (e.g. Figure 4E). This finding must be considered to avoid erroneous optimal input when optimum search algorithms are implemented.

6. Other applications

Besides the optimal design of label inputs IsoDesign can be also applied to optimizing other aspects of ^{13}C-labelling experiments. For instance, IsoDesign can be used to evaluate the potential benefit of additional isotopic measurements. An example is given in Figure 5 for acetate-grown cells. Sensitivity landscapes have been calculated for a particular flux (pyruvate kinase) according to the dataset considered above without (Figure 5A) or with (Figure 5B) addition of MS data for pyruvate, oxalo-acetate and glyoxylate. The data showed that the sensitivity landscape is substantially modified when the additional data are considered, with a significant increase in the number of label inputs giving a good precision on pyruvate kinase. Such type of simulation is of special value to improve the design of the label input when biologically-important fluxes are not initially well determined, and to make decision before engaging novel – and costly – methodological developments aiming to access new isotopic data.

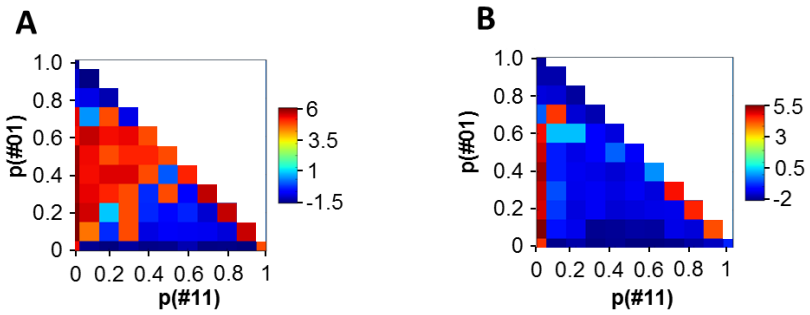


Figure 5. Impact of additional isotopic data on the standard deviation of the flux through the pyruvate kinase. A) Sensitivity landscape calculated with 50 MS data from 9 metabolites. B) Sensitivity landscape calculated with 62 MS data from 12 metabolites. Standard deviation is in logarithmic scale. Proportion of ^{12}C-acetate is 1-p(#01)-p(#11), and there is no 1-^{13}C-acetate.

IsoDesign can be also used to compare the flux precision brought by different analytical tools. An example is shown in figure 6 for acetate-grown cells. Sensitivity landscapes have been calculated for the entire set of fluxes according to the dataset that can be measured by MS – on intracellular metabolites – (Figure 6A) or by NMR – on proteinogenic amino acids – with an HSQC experiment (Figure 6B). The data showed that the sensitivity landscape is substantially modified when different analytical tools are considered, with a strong difference in the label inputs giving a good precision on the

entire set of fluxes. Such type of simulation is of special value to select the analytical platform to be used for isotopic analysis.

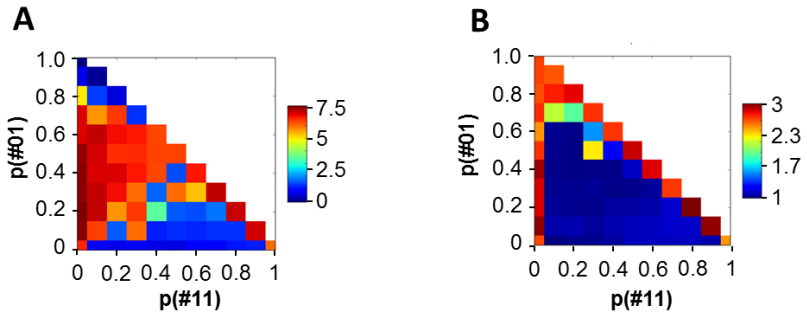


Figure 6. Impact of analytical platform on the sum of SDs of net fluxes (logarithmic scale).

A) Sensitivity landscapes calculated from an MS dataset on intracellular metabolites. B) Sensitivity landscapes calculated from an NMR dataset (HSQC pulse sequence) on proteinogenic amino acids. Proportion of ^{12}C-acetate is 1-p(#01)-p(#11), and there is no 1-^{13}C-acetate.

7. Conclusion

IsoDesign is a user-friendly software dedicated to experimental design for ^{13}C-MFA. This tool has been successfully applied to optimise label input for maximising the flux information with regard to the biological question to be addressed by the investigator. In addition, some other applications have been briefly presented. We expect that IsoDesign will help defining more effective label inputs in future ^{13}C-MFA experiments.

8. References

Antoniewicz, M. R., J. K. Kelleher, et al. (2006). "Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements." Metab Eng **8**(4): 324-337.

Chen, J., H. Zheng, et al. (2007). "Improving metabolic flux estimation via evolutionary optimization for convex solution space." Bioinformatics **23**(9): 1115-1123.

Crown, S. B., W. S. Ahn, et al. (2012). "Rational design of 13C-labeling experiments for metabolic flux analysis in mammalian cells." BMC Syst Biol **6**(1): 43.

Crown, S. B. and M. R. Antoniewicz (2012). "Selection of tracers for 13C-metabolic flux analysis using elementary metabolite units (EMU) basis vector methodology." Metab Eng **14**(2): 150-161.

Ishii, N., K. Nakahigashi, et al. (2007). "Multiple high-throughput analyses monitor the response of E. coli to perturbations." Science **316**(5824): 593-597.

Iwatani, S., Y. Yamada, et al. (2008). "Metabolic flux analysis in biotechnology processes." Biotechnol Lett **30**(5): 791-799.

Kiefer, P., C. Nicolas, et al. (2007). "Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry." Anal Biochem **360**(2): 182-188.

Libourel, I. G., J. P. Gehan, et al. (2007). "Design of substrate label for steady state flux measurements in plant systems using the metabolic network of Brassica napus embryos." Phytochemistry **68**(16-18): 2211-2221.

Mollney, M., W. Wiechert, et al. (1999). "Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments." Biotechnol Bioeng **66**(2): 86-103.

Mollney, M., W. Wiechert, et al. (1999). "Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments." Biotechnol Bioeng **66**(2): 86-103.

Nicolas, C., P. Kiefer, et al. (2007). "Response of the central metabolism of Escherichia coli to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase." FEBS Lett **581**(20): 3771-3776.

Orth, J. D., I. Thiele, et al. (2010). "What is flux balance analysis?" Nature Biotechnology **28**(3): 245-248.

Quek, L. E., C. Wittmann, et al. (2009). "OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis." Microb Cell Fact **8**: 25.

Sokol, S., P. Millard, et al. (2011). "influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labeling experiments." Bioinformatics.

van Gulik, W. M., W. T. de Laat, et al. (2000). "Application of metabolic flux analysis for the identification of metabolic bottlenecks in the biosynthesis of penicillin-G." Biotechnol Bioeng **68**(6): 602-618.

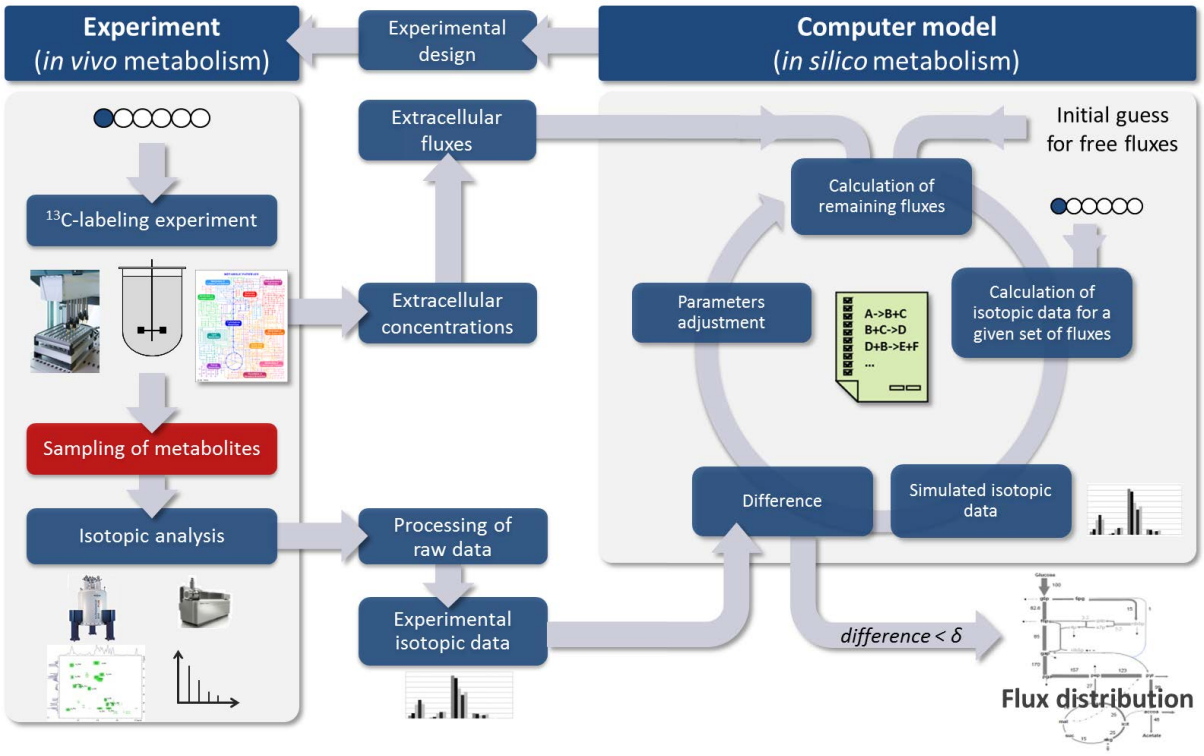
Walther, J. L., C. M. Metallo, et al. (2012). "Optimization of 13C isotopic tracers for metabolic flux analysis in mammalian cells." Metab Eng **14**(2): 162-171.

Wiechert, W. (2001). "^{13}C metabolic flux analysis." Metab Eng **3**(3): 195-206.

Wiechert, W., M. Mollney, et al. (2001). "A universal framework for 13C metabolic flux analysis." Metab Eng **3**(3): 265-283.

Chapter 3

Reliable sampling of ^{13}C-labelled metabolic intermediates for stationary and instationary ^{13}C-metabolic flux analysis: Practical guidelines for *Escherichia coli*



NOTE: *this chapter contains the first draft of a manuscript in preparation for Analytical Biochemistry.*

1. Abstract

The reliable sampling of metabolic intermediates remains a serious challenge in ^{13}C-metabolic flux analysis (^{13}C-MFA) which is commonly circumvented by the analysis of end-products of the metabolism. This constrained to determine the steady-state metabolic fluxes in isotopic stationary state and to leave aside the richest source of labelling information provided by metabolic intermediates. The current study aims to provide guidelines for the evaluation of sampling procedures of metabolic intermediates for stationary and instationary ^{13}C-MFA. We first revisited the efficiency of different methods for quenching the metabolism of *Escherichia coli* and it was found that cold solvents are suitable quenching solutions unlike fast-filtration and, that in stationary state, the separation of cells from the broth is not absolutely required despite of the presence of metabolic intermediates outside the cells. To evaluate such a need in instationary state, we designed an original ^{13}C-labelling pulse experiment. It leads to the conclusion that pools of intra- and extracellular metabolites must be discriminated in instationary state. We demonstrated that it could be achieved by centrifugation in cold aqueous solvent (glycerol at -20°C) or without separation provided that the fraction of intracellular pool in the sample is determined.

2. Introduction

Metabolic flux analysis (MFA) is a powerful tool that aims to provide a quantitative snapshot of the actual operation of metabolic networks and plays an important role in the comprehensive understanding of the cellular response to changes in their surrounding or their genome. For these reasons, MFA is extensively applied in the field metabolic engineering (Iwatani, Yamada et al. 2008). It is also a key *omics* technology for systems biology as an integrative and quantitative view of all the processes occurring within a cellular system (Dauner 2010; Kohlstedt, Becker et al. 2010). Briefly, MFA includes several types of model-based approaches to determine the metabolic fluxes since they are not directly measurable. Besides pure *in silico* techniques, numerous

approaches rely on an experimental basis for which the quality of the data is of the paramount importance.

^{13}C-MFA refers to the use of ^{13}C-labelling strategies to access to metabolic fluxes within metabolic networks. During a carbon labelling experiment (CLE), a biological system is fed with a ^{13}C-enriched substrate. ^{13}C atoms are distributed in the metabolic intermediates. The labelling patterns of ^{13}C-metabolites depend both on the topology of the network and activities of metabolic pathways. Thanks to computational and mathematical tools based on stoichiometric balances (Wiechert 2001; Sokol, Millard et al. 2011), the labelling patterns of metabolites allow to calculate the fluxes.

Because of their extremely high turnover rates of metabolic intermediates, the sampling of intracellular metabolites is a major challenge. A large body of papers reports sampling procedures to qualitatively and/or quantitatively analyse the metabolome of a wide variety of organisms (Bolten, Kiefer et al. 2007; Villas-Boas and Bruheim 2007; Canelas, ten Pierick et al. 2009; Taymaz-Nikerel, de Mey et al. 2009). In contrast, very little attention has been paid to metabolites sampling in the context of ^{13}C-MFA. As for metabolomics, it requires also an immediate blockage of all metabolic activities (designated as "quenching of the metabolism"), although the absolute titration or the relative quantification of the metabolome is not directly wanted in labelling experiment. Indeed, isotopic analyses aims at the fine characterization of the labelling pattern of metabolites that could be properly achieved even if it concerns a partial amount of the total pool of the examined metabolites. To circumvent the sampling issue in ^{13}C-MFA, end-products of the metabolism may be examined instead of the metabolic intermediates. These compounds accumulate all along the experiment and have lower turnover rates as compared to metabolic intermediates. A large number of methods have been developed relying on the analysis of amino-acids forming the protein ("proteinogenic amino acids") of which the biosynthesis derivates from metabolic intermediates. Hence the labelling pattern of metabolic intermediates can be deduced from the labelling patterns of proteinogenic amino acids from which they are the precursors. The proteins – of which the content represents a significant proportion of the total dry weight – are then cleaved to liberate amino acids (mostly by acidic hydrolyzation) that are subsequently analysed by MS or NMR. This type of approach is well-established, widely applied and very well reviewed in literature (Wittmann 2007;

Zamboni, Fendt et al. 2009). However, it is limited to isotopic and metabolic steady states, which is practically complicated to achieve in certain cases if not impossible. In addition, although the pathways of the central metabolism are well covered, some parts of the metabolic networks remain not accessible by the unique use of metabolic end-products. This limitation is all the more regrettable that progress on metabolomics allows us to access to an extremely high numbers of metabolic intermediates of which the isotopic analysis after a ^{13}C-MFA experiment would enlighten the functioning of some parts of the metabolic network that would remain hidden otherwise. Besides stationary state, labelling experiments may also be carried out in instationary state. They offer significant advantages as they allow i) to calculate metabolic fluxes for methylotrophic and autotrophic organisms, ii) to multiply the numbers of biological replicates that can be performed in a time span of minute instead of several hours in stationary state while limiting the amount of ^{13}C-labelled substrate used, iii) to elucidate or clarify the topology and the operation of metabolic networks. Advances concerning this approach have been reviewed quite recently (Noh and Wiechert 2011).

To our knowledge, there is no study dealing specifically with the sampling of metabolic intermediates for ^{13}C-MFA. Obviously, it can be though that the sampling procedures reliable for quantitative metabolomics may be straightforwardly applied for ^{13}C-MFA and this is logically what is commonly done. However, the presence of metabolic intermediates outside the cells (a phenomenon that has been widely reported in literature for microorganisms) and the leakage of metabolites induced mainly by quenching solvents have led to the set-up of sophisticated procedures for quantitative metabolomics. Some of them are even not applicable in the framework of ^{13}C-MFA. For instance, the method developed by (Taymaz-Nikerel, de Mey et al. 2009) based on a differential titration of metabolites in the total broth and in the filtrate of culture sample of *Escherichia coli*, is a very powerful tool for quantitative metabolomics that could not be directly applied for ^{13}C-MFA. Indeed, labelling patterns of intracellular and extracellular metabolites cannot be subtracted without additional measurements. In opposite, it is reported in literature that some quenching procedures based on cold solvents (e.g. methanol, ethanol, etc) are not applicable for quantitative metabolomics because of release of a major part of the metabolites from the cells (Bolten, Kiefer et al. 2007; Villas-Boas and Bruheim 2007). For ^{13}C-MFA, since absolute quantification is not required, the use of such solvents may be suitable for setting up reliable sampling

methods unlike the labelling pattern is disturbed during the process. It can be thus concluded that the sampling of metabolic intermediates for ^{13}C-MFA may be consider in a different way than for quantitative metabolomics.

To date, for ^{13}C-MFA studies based on ^{13}C-metabolites analysis reported in literature, the fast quenching of all metabolic activity is performed by direct sampling in cold solvent as aqueous methanol prior to the removal of the extracellular compounds by centrifugation (van Winden, van Dam et al. 2005; Kleijn, Geertman et al. 2007; Noh, Gronke et al. 2007; Zhao, Kuijvenhoven et al. 2008; Ruhl, Rupp et al. 2012). Alternatively, the cells are first separated from the broth by fast-filtration and subsequently quenched in cold liquid like liquid nitrogen (Toya, Ishii et al. 2007; Nakahigashi, Toya et al. 2009; Toya, Ishii et al. 2010; Chen, Alonso et al. 2011). A derivative method developed by Rabinowitz and coworkers consists in growing the cells directly on a membrane filter setting on the top of agarose plate and to operate the fast-quenching by placing the filters into cold organic solvent mixture (Yuan, Bennett et al. 2008). The metabolism is then quenched simultaneously with ^{13}C-metabolites extraction while ensuring a separation with the extracellular metabolites. Finally, one other method is applied without removal of extracellular compounds thanks to a rapid heating of the broth that provokes the simultaneous blockage of metabolic activity and the release of intracellular metabolites subsequently to cell membrane disruption (Schaub, Schiesling et al. 2006). These studies have been performed with a wide range of microorganisms (yeast, gram positive and negative bacteria, fungi) irrespective of whether they are performed in isotopic stationary or instationary state.

The main aim of the current work was to define a methodology for assessing the reliability of sampling procedures relevant for ^{13}C-MFA in stationary and instationary states. An additional objective of the study was to give guidelines to execute properly the sampling of labelled metabolic intermediates of the bacteria *Escherichia coli* which is extensively used in metabolic engineering and for system biology approach. First, we investigated the efficiency of five different sampling procedures to obtain a rapid quenching of all metabolic activity according to methods that are mainly reported in literature for *E. coli*. To evaluate the efficiency of each individual procedure, flux maps calculated from labelling pattern of ^{13}C-metabolites so-obtained and recovered from an *E. coli* batch culture under isotopic and metabolic stationary states with ^{13}C-labelled

glucose, have been compared thoroughly to a flux map established from proteinogenic amino acids-based approach. Having been widely demonstrated in literature that large amount of metabolic intermediates are present outside the cells, we tested sampling methods without or with separation of the cells from broth by centrifugation or fast-filtration. Having shown that cells separation is not a crucial requirement in stationary state, we addressed the same question in instationary state. Thanks to an original method, we concluded that extracellular metabolites induced isotopic dilution implying that both pools of metabolites have to be discriminated in isotopic instationary state. Finally, we also demonstrated that the isotopic pulse experiment may be a very promising practical approach to quantify pools of metabolites according to their cellular compartments.

3. Material and methods

Strain and cultivations

Escherichia coli K12 MG1655 was grown on minimal synthetic medium containing 5 mM KH_2PO_4, 10 mM Na_2HPO_4, 9 mM NaCl, 40 mM NH_4Cl, 0.8 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.1 g/L of thiamine and 3 g/L of glucose. Glucose and thiamine were sterilized by filtration (Minisart polyamide 0.2 µm, Sartorius, Germany), other solutions were autoclaved separately. All stock cultures were stored at -80°C in LB medium containing glycerol (40% v/v). For cultivation of *E. coli* , 1 mL of LB overnight cultures were used as inoculum, and then sub-cultured in cultivation tubes containing 50 mL of minimal medium with 3 g/L glucose starting with an OD_{600nm} = 0.05 and incubated at 37 °C, 200 rpm in an orbital shaker (G25, New Brunswick Scientific, Inc., New Brunswick, NJ). Cells were harvested during the exponential growing phase by centrifugation (10 000 g, at room temperature, 10 min) with a Sigma 3-18K centrifuge (Sigma, Germany), washed in the same volume of fresh medium (lacking glucose and thiamine) and used to inoculate a 1 L baffled shake flask containing 150 mL of minimal medium with 3 g/L glucose at an OD_{600nm} = 0.01, and then incubated at 37 °C, 200 rpm. Cell growth was monitored by optical density at 600 nm with a Genesys 6 spectrophotometer (Thermo, Carlsbad, USA). Extracellular fluxes were determined from the rate of disappearance of glucose and the rate of appearance of acetate (no other by-products being detected), which were

quantified in culture supernatants by HPLC as previously described (Cocaign-Bousquet, Garrigues et al. 1995).

For ^{13}C-labelling experiments in metabolic and isotopic steady-state, unlabeled glucose was substituted by a mixture of 80% [$^{13}C_1$]glucose and 20% [U-^{13}C]glucose (Eurisotop, France).

For ^{13}C-labelling experiments in isotopic instationary state, isotopic pulses were performed by replenishing with 20 mL of a culture of *E. coli* MG1655 (collected in mid-exponential growth phase on glucose at natural abundance) in a syringe prefilled with 500 µL of [U-^{13}C]glucose (150g/L) and pre-heated at 37°C. After rapid mixing, approximately 200 µL of the whole broth were collected during the first 90 s every 2 s for the first 30 s and then every 15 s during the last 60 s according to sampling procedures described below. Filtrate samples were collected every 6 s during an additional ^{13}C-pulse experiment, filtering the broth with a 0.2 µm syringe filter (Minisart polyamide 0.2 µm, Sartorius), which was substituted for each individual sample by a new filter.

Sampling procedures

Sampling of proteinogenic amino acids. Respectively 3 and 60 mg of biomass were collected for GC-MS and NMR analyses, respectively, at the mid-exponential growth phase to ensure both isotopic and metabolic steady-states. Samples were prepared as described previously (Nicolas, Kiefer et al. 2007).

Sampling of intracellular metabolites. Five different protocols for quenching the metabolism and to extract metabolites were tested in this study. They were performed in triplicate from the same cultivation – from which the samples for GC-MS and NMR analysis have been collected as well – to avoid biological variations.

The **T procedure** consisted in a simultaneous quenching and extraction of the total amount of metabolites by spraying 120 µL of whole broth in a 10 mL glass tube containing 5 mL of an ethanol/H_20 (75/25) solution at 95°C during 5 min. Tubes were then placed in a cooling bath of ethanol pre-cooled at -80°C and subsequently centrifuged at -20°C (12 000g, 5 min) with a Sigma 3-18K centrifuge (Sigma, Germany). Supernatants were evaporated to dryness in a SpeedVac (SC110A SpeedVac Plus, ThermoSavant, USA) under vacuum for 4 hours and then stored at -80°C until further treatment.

For the four additional procedures, quenching and extraction were performed in individual steps. Cells were separated from the broth by centrifugation in cold quenching solutions (M, E and G procedures) or by fast-filtration prior to quenching in liquid nitrogen (F procedure). The quenching solutions that were used, were i) 99.9 % of MS grade methanol (Sigma-Aldricht Chemie, France) precooled at -80°C for the **M procedure**, ii) 99.9 % of MS grade ethanol (Sigma-Aldricht Chemie) precooled at -80°C for the **E procedure** and, iii) glycerol/water (80/20) + 0.9% NaCl solution precooled at -23°C for the **G procedure**. After quenching 120 µL of broth in 500 µL of the quenching solution, the vortexed mixtures were centrifuged (3-18K centrifuge, Sigma, Germany) at -20°C and at 12 000 g during 5 min for methanolic and ethanolic mixtures, and at 16 000 g during 20 min for the glycerolic mixture. The pellets were stored at -80°C until further metabolite extraction. For the **F procedure**, 120 µL of broth were harvested by vacuum filtration (Sartolon polyamide 0.2 µm, Sartorius, Germany). Cells were washed with 1 mL of medium with reduced contents in phosphate and sulphate salts (0.5 mM KH_2PO_4, 1 mM Na_2HPO_4, 0.08 mM $MgSO_4$) to avoid ion suppression effects during mass spectrometry analysis; all other components, including the labelled substrate, were kept unchanged in the washing solution. Filters were rapidly transferred in liquid nitrogen and stored at -80°C until further metabolite extraction. The whole filtration procedure (including washing step) took place in less than 5 s.

For metabolite extraction, cell pellets and filters were incubated for 5 min in closed glass tubes containing 5 mL of an ethanol/H_20 (75/25) solution at 95°C. The extraction was terminated by placing the tubes in a cooling bath of ethanol precooled at -80°C. Subsequently, the extracts were centrifuged at -20°C (12 000g, 5 min), supernatants were evaporated to dryness in a SpeedVac (SC110A SpeedVac Plus, ThermoSavant, USA) under vacuum for 4 hours and then stored at -80°C until further treatment.

Isotopic analyses

NMR analysis of extracellular metabolites. 500 µL of broth were filtered (0.2 µm, Sartorius) and analysed by quantitative ^{1}H 1D-NMR at 286 K with a 30° pulse and a relaxation delay of 20 s, with an Avance 500 MHz spectrometer (Bruker, Switzerland) equipped with a BBI 5 mm. For steady-state ^{13}C-labelling experiments, the four isotopomers of acetate accumulating in the medium during growth were resolved and quantified. For each ^{13}C-labelling experiment carried out in instationary state conditions,

the ^{13}C-enrichment of the C1 (anomeric carbon) of glucose was measured after the pulse.

IC-MS/MS analysis of intracellular metabolites. After resuspension of cellular extracts stored at -80°C in 200 µL of ultrapure water, cell debris were removed by centrifugation 10 000 g for 10 min. Supernatants were subsequently analysed as described by (Kiefer, Nicolas et al. 2007) with a Dionex ICS 2000 system (Dionex, Sunnyvale, USA) coupled to a triple quadrupole QTrap 4000 (Applied Biosystems, Foster City, USA) mass spectrometer. Isotopic clusters of malate (MAL), glucose-6-phosphate (G6P), fructose-6-phosphate (F6P), fructose-1,6-bisphosphate (FBP), 6-phosphogluconate (6PG), ribose-5-phosphate (R5P), sedoheptulose-7-phosphate (S7P), phosphoenolpyruvate (PEP) and the combined pool of 2- and 3-phosphoglycerate (2/3PG) were determined as described by (Kiefer, Nicolas et al. 2007). The injection volume was 15 µL, originating from approximately 3 µg of biomass.

NMR and GC-MS analysis of proteinogenic amino acids. The specific enrichments and positional isotopomers of proteinogenic amino acids were quantified by 2D-NMR (using ZQF-TOCSY (Massou, Nicolas et al. 2007) and HSQC (Szyperski 1995) experiments, respectively) with an Avance 500 MHz spectrometer (Bruker, Germany) as previously described (Nicolas, Kiefer et al. 2007). The isotopic clusters of proteinogenic amino acids were measured by GC-MS as described in (Wittmann 2007), mainly monitoring by $[M-57]^+$ fragments in the SIM mode.

Computational part

Correction of mass spectrometry data for naturally occurring isotopes. Isotopologues distributions of intracellular metabolites and of proteinogenic amino acids were calculated from their measured isotopic clusters after correction for naturally occurring isotopes of elements other than carbon, performed with the software IsoCor (Millard, Letisse et al. 2012).

Flux calculations. Flux calculations were performed with influx_s (Sokol, Millard et al. 2011), in which both mass balances and carbon atom transitions describing the biochemical reaction network were implemented. The metabolic network implemented in the FTBL model includes glycolysis (EMP), Entner-Doudoroff pathway (ED), pentose phosphate pathway (PP) (Kleijn, van Winden et al. 2005), tricarboxylic acid cycle (TCA)

and amino-acid biosynthesis pathways. In total, the model was made up of 25 reactions for central carbon metabolism and 63 biosynthetic reactions, for a total of 103 fluxes (73 unidirectional and 15 reversible reactions). Precursor requirements for biomass formation were determined according to the biomass composition of *E. coli* (Neidhart 1996) and measured growth rate.

Statistical methods. For datasets comparison, hierarchical clustering was performed with the warp method implemented in the 'hclust' package of R, and the significance of differences between datasets were assessed using Student's paired t-tests with a two tailed distribution. 95% confidence intervals of fluxes were estimated from standard deviations of IDs – with a threshold of 1% (chapter 4) – applying a Monte-Carlo approach with 100 iterations/dataset.

4. Results and discussion

In contrast to metabolomics (Bolten, Kiefer et al. 2007; Villas-Boas and Bruheim 2007; Taymaz-Nikerel, de Mey et al. 2009; Douma, de Jonge et al. 2010), few attentions are paid to the sampling of ^{13}C-labelled intracellular metabolites (^{13}C-metabolites) of which isotopic profiles are used for metabolic flux quantification (^{13}C-MFA). Although these ^{13}C-metabolites are not absolutely tittered, ^{13}C-labelling experiments require an accurate and precise quantification of their isotopic content. Therefore, it is crucial to obtain an immediate blockage of the metabolism whilst cells are harvested and to ensure that all steps from the recovering of the ^{13}C-metabolites to the sample preparation and their analyses do not disturb the isotopic pattern of metabolites of interest.

Sampling of ^{13}C-metabolites at isotopic steady-state

First, our work was focused on the optimization of experimental procedures for the preparation of ^{13}C-metabolites from cells of *Escherichia coli* cultivated in metabolic and isotopic steady-state. In particular, we evaluated the ability of sampling procedures to obtain a rapid (immediate) quenching of all metabolic activity during the sampling procedure. In that aim, we assessed classical approaches of rapid inactivation of the metabolism by quenching the cells in cold or boiling liquid and investigated the impact of steps of cells separation on the quenching efficiency. ^{13}C-metabolites were systematically extracted with boiling ethanol, which has been shown to be highly reproducible and to

not provoke extensive degradation or interconversion of metabolites (Canelas, ten Pierick et al. 2009). All sampling procedures were carried out from a cultivation of *E. coli* grown on a mixture of 20% 1-^{13}Cglucose and 80% U-^{13}C-glucose as sole carbon source. All the samples were harvested in the mid-exponential growth phase after 6 generations, ensuring the isotopic steady-state has been reached. The isotopologue distributions (IDs) of ^{13}C-metabolites were analysed by IC-MS/MS using the highly sensitive method of (Kiefer, Nicolas et al. 2007). For each sampling procedure, the isotopic dataset was used to calculate flux maps. In addition, the biomass produced during the cultivation was collected to analyse the labelling patterns of proteinogenic amino acids using state-of-art methods. Because proteinogenic amino acids accumulate into the biomass as end-products, their recovering are not as subjected to immediate quenching of metabolism as ^{13}C-metabolites are. For these reasons, the isotopic dataset obtained and the related flux map are considered in this paper as a benchmark for evaluating the reliability and the efficiency of the different sampling procedures of the ^{13}C-metabolites.

Metabolic flux map from proteinogenic amino acids

The flux distribution within the central metabolism of *E. coli* obtained for the labelling patterns of proteinogenic amino acids is shown figure 1. Samples of biomass hydrolysates were analysed by NMR spectroscopy, from HSQC and ZQF-TOCSY experiments (Massou, Nicolas et al. 2007; Massou, Nicolas et al. 2007), and by GC-MS after trimethylsilyl derivatization (Wittmann 2007). The flux map was obtained from a dataset containing a total of 184 isotopic measurements. The correlation factor (r^2) between the experimental *versus* the simulated ID was 0.998 (Figure 1, inserted graph). This demonstrates that the metabolic model used to resolve the flux distribution did not introduce any experimental artefacts. Furthermore, the flux distribution so-obtained was in well agreement with previous published results (Fuhrer, Fischer et al. 2005; Nicolas, Kiefer et al. 2007). Finally, an *in silico* isotopic dataset for ^{13}C-metabolites was calculated from the obtained flux map by reverse simulation of the isotopic profiles.

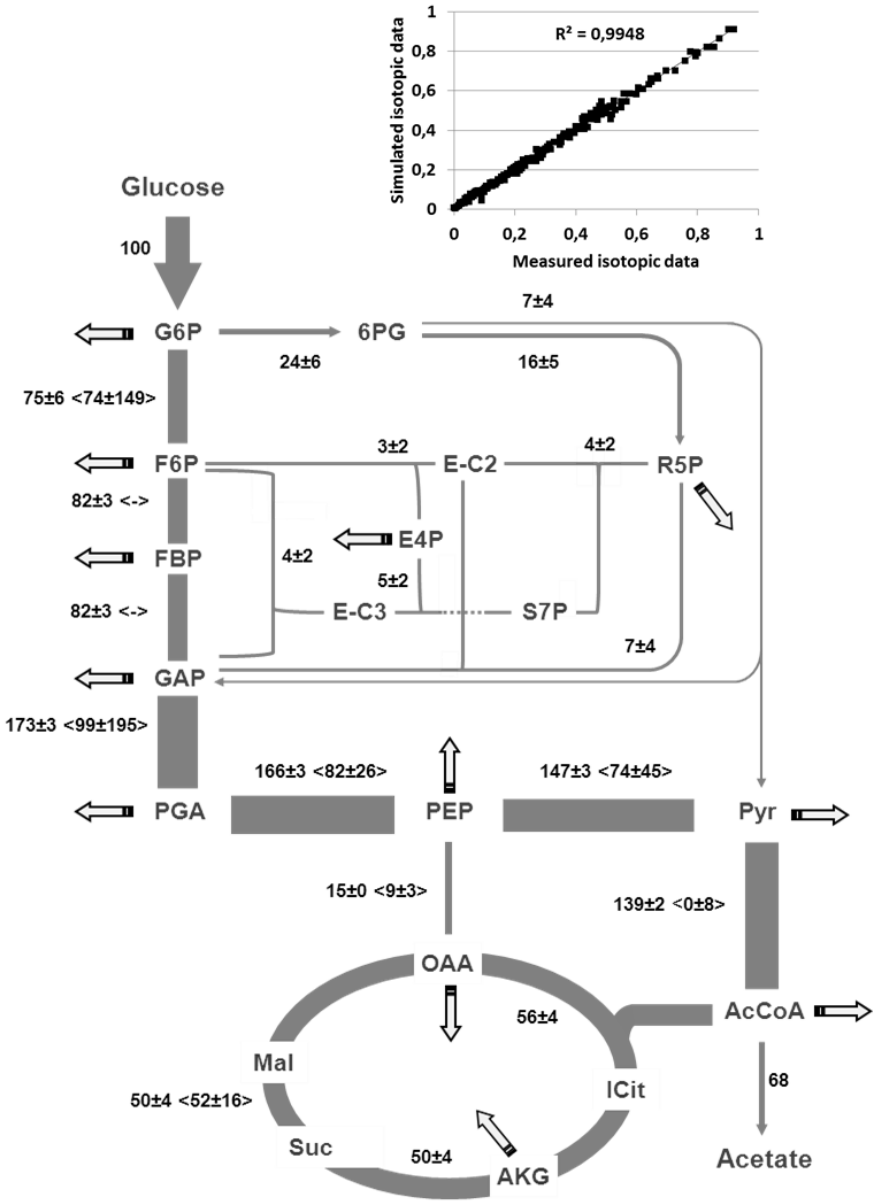


Figure 1. Flux distribution in the central carbon metabolism of *E. coli* K12 MG1655 grown on glucose, calculated from GC-MS and 2D-NMR (HSQC and ZQF-TOCSY experiments) measurements of ^{13}C-labeling patterns in proteinogenic amino acids. Inserted graph is the correlation plot between measured and simulated isotopic data.

Comparison of labelling data obtained from ^{13}C-metabolites

The presence of intermediates of central metabolism out of the cells is generally observed for micro-organisms (Bolten, Kiefer et al. 2007; Winder, Dunn et al. 2008; Taymaz-Nikerel, de Mey et al. 2009), this phenomenon being emphasized for gram-negative bacteria. Therefore, we evaluated sampling procedures separating cells from the cultivation medium either by fast-filtration (F) prior to nitrogen quenching or by centrifugation in cold solvent (-80 °C Methanol, M, -80 °C Ethanol, E or -20 °C Glycerol,

G) as commonly applied in metabolomics (Bolten, Kiefer et al. 2007; Villas-Boas and Bruheim 2007; Taymaz-Nikerel, de Mey et al. 2009) and fluxomics (van Winden, van Dam et al. 2005; Kleijn, Geertman et al. 2007; Toya, Ishii et al. 2007). Intracellular metabolites were extracted by boiling ethanol (95°C, 5 min) (Taymaz-Nikerel, de Mey et al. 2009). In addition, we tested a procedure in which quenching and extraction of ^{13}C-metabolites were performed in a single step without cells separation (T). In this procedure both quenching and extraction are carried out with boiling ethanol (95°C, 5 minutes). All sampling procedures were carried out in triplicate. For each procedure, the measurements of IDs by IC-MS/MS (Kiefer, Nicolas et al. 2007) generated 54 isotopic data from 9 ^{13}C-metabolites located at several nodes of the *E. coli*'s central metabolic network.

Hierarchical clustering (Figure 2A) was applied to the five isotopologue datasets (F, M, E, G and T) *plus* the isotopologue dataset simulated from the flux map described on Figure 1. This analysis shows clearly that the data are strongly clustered all together excepted when fast-filtration is performed. It should be noticed that the sampling procedure without separation (T) led to labelling patterns close to simulated ones. A Z-score analysis (Figure 2B) of the data confirmed these results and indicated that the main differences between fast-filtration and the 5 other datasets are mainly related to metabolites FBP, 2/3PG, R5P and S7P.

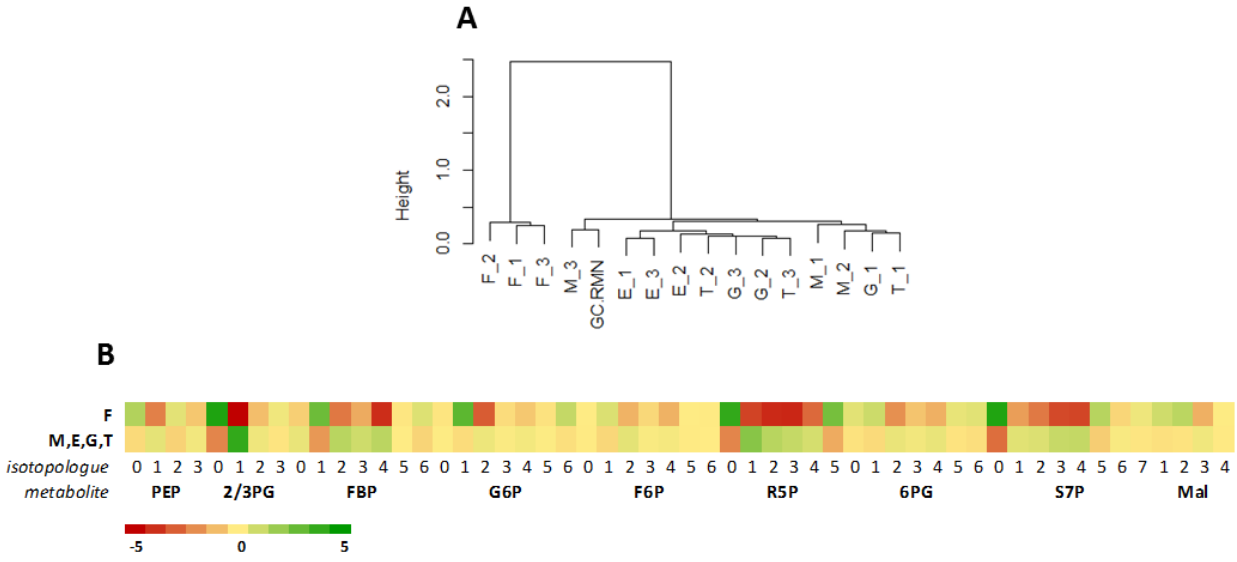


Figure 2. (A) Hierarchical clustering of isotopic data from samples obtained with each of the 5 evaluated sampling methods (F, M, E, G, T) – 3 independent samples per method – and of the dataset simulated from the flux distribution used as benchmark (GC+RMN). (B) Z-score comparison of IDs observed for the F procedure vs M,E,G and T procedures.

Comparison of metabolic flux maps

To further evaluate the sampling procedures, metabolic flux maps were calculated for all single sampling procedure and IDs resulting from the flux calculation (simulated IDs) were compared to the measured IDs. As shown in Figure 3, the correlation between measured and simulated IDs was high ($r^2 \geq 0.99$) for sampling procedures with cold solvents (E, M, G) and for the extraction in boiling ethanol (T). The fast filtration method (F) resulted in significant differences between measured and simulated IDs ($r^2 = 0.88$), meaning that the flux calculation cannot be performed successfully. The main differences were observed for metabolites belonging to PP (R5P, S7P) and glycolysis (FBP, 2/3PG).

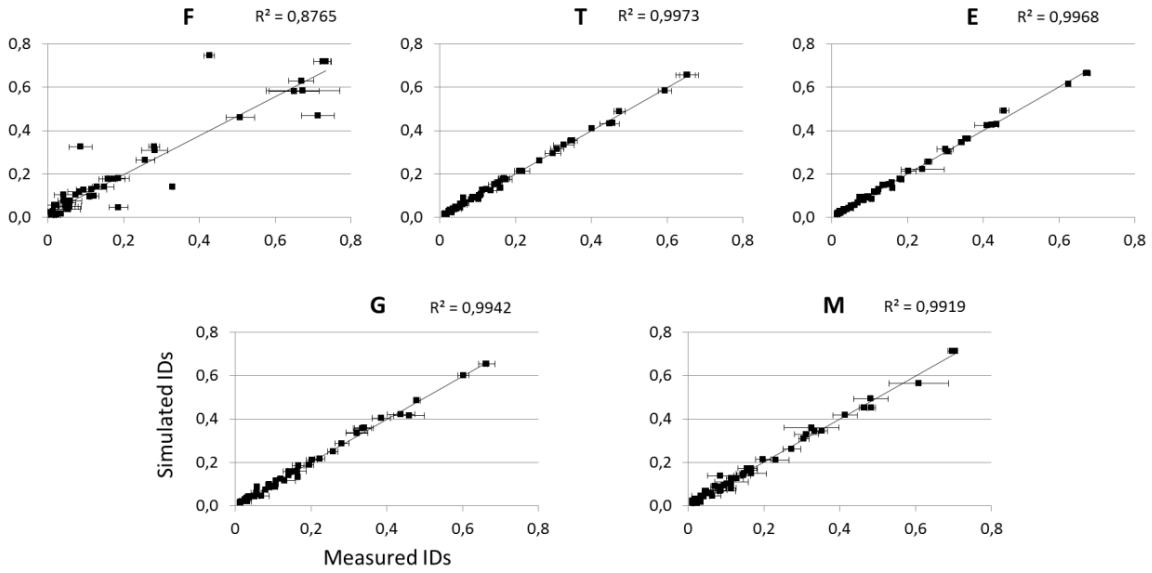


Figure 3. Correlation between IDs simulated from the optimal flux distribution and measured IDs for the five evaluated protocols.

The heat map on Figure 4A allows to visualize the flux distributions – relative to the proteinogenic amino acids based flux map – through the entire central metabolism. We observed no substantial difference in metabolic fluxes in glycolysis and TCA cycle across the sampling procedures. However, the flux in Entner-Doudoroff pathway (edd) appeared to be lower (from 0 to 4% of the glucose uptake rate according to sampling procedures) compared to those determined from the proteinogenic amino acids (8% of the glucose uptake rate). Nevertheless, sensitivity analysis – 95% confidence interval are displayed on Figure 4B as gray bars – showed that the differences are not significant. Better resolution could be reached by completing the current isotopic dataset with additional ^{13}C-metabolites (e.g. pyruvate) or by using other isotopic forms of labelled glucose.

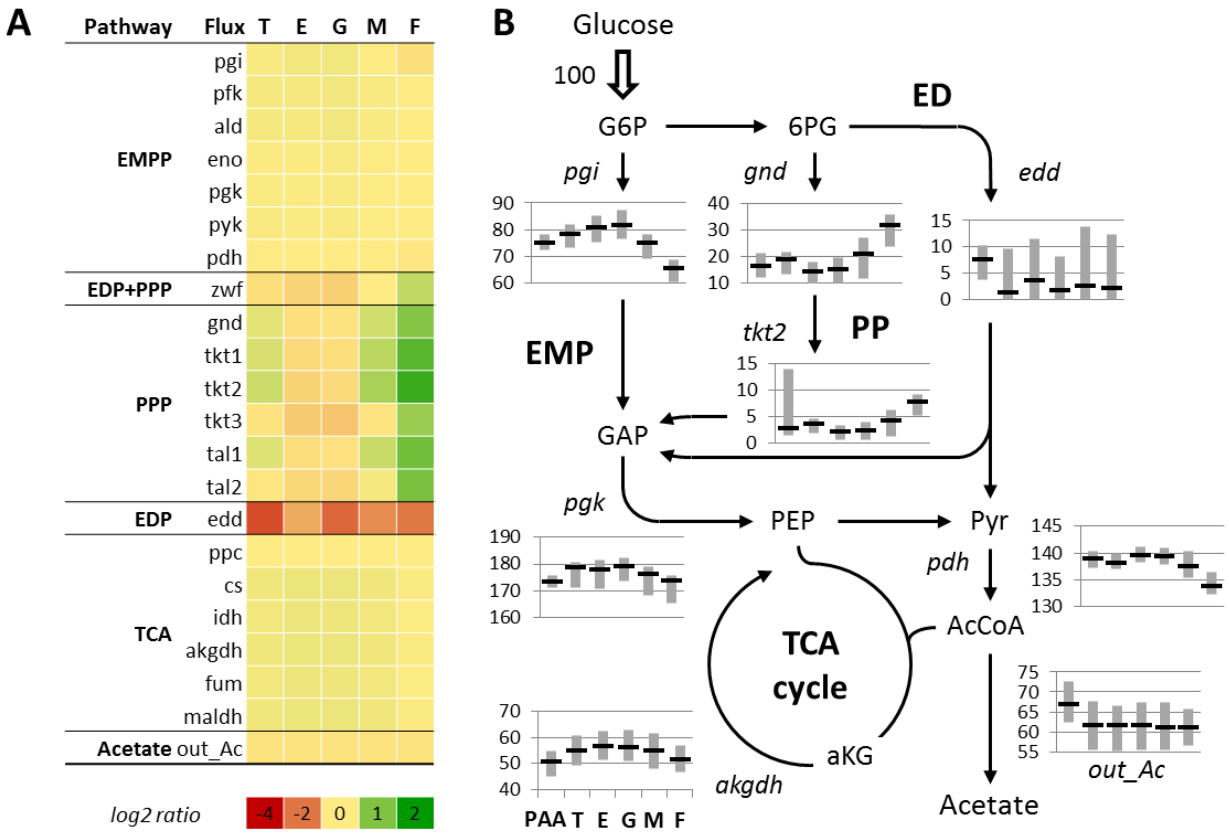


Figure 4. (A) Heatmap of central metabolic fluxes in *E. coli* K12 MG1655 on glucose. Values correspond to the ratio relative to the proteinogenic amino acids flux distribution (log2 scale). (B) Measured fluxes (and their respective 95% confidence intervals) applying the five sampling methods and the proteinogenic amino acids based method (PAA). EMP: Embden-Meyerhof-Parnas pathway ; PP: Pentoses Phosphates pathway ; ED: Entner-Doudoroff pathway.

In addition, the fast-filtration sampling lead to an overestimation of the fluxes related to the PP pathway, which are twofold to fourfold higher compared with the other sampling methods. As examples, it can be shown on Figure 4B that fluxes through 6-phosphogluconate deshydrogenase (gnd) and through transketolase (tkt2) obtained with the fast-filtration procedure are significantly higher than those obtained with the other sampling methods. Despite of the handling precautions taken during the process of fast-filtration, this procedure does not permit to get an immediate blockage of the metabolic activity and results in erroneous metabolic fluxes. This resulted in an overestimation of the absolute fluxes for some reactions of the PP pathway as reported in literature (Toya, Ishii et al. 2007) that are in this work, twice higher than those that could be determined with the others sampling methods (see annex). Interestingly, the single step procedure – consisting in a simultaneous quenching of the metabolism and the extraction of ^{13}C-

metabolites – gave similar results as procedures with centrifugation of cells in cold solvent prior to the extraction of metabolites. For all these methods, no significant difference in terms of IDs and flux quantification can be brought out with results related to proteinogenic amino acids method.

We conclude that ^{13}C-metabolites-based flux analysis can be carried out without requiring complicated sample preparation (quenching and extraction simultaneously) provided cells are growing under isotopic and metabolic steady-state conditions.

Sampling of ^{13}C-metabolites for instationary isotope labelling experiments

We focused then our work on the assessment of sampling procedures for ^{13}C-metabolites collected from *Escherichia coli* cells in instationary isotope labelling experiments. The question here is related to the presence of metabolic intermediates – which support the isotopic information – outside the cells. As mentioned above, there is a large body of papers reporting their detection in cultivation medium during growth of the micro-organisms (Bolten, Kiefer et al. 2007; Winder, Dunn et al. 2008; Taymaz-Nikerel, de Mey et al. 2009; Paczia, Nilgen et al. 2012) without a clear explanation of this phenomenon. Although we can deal with them in steady-state (see above), we investigated the need to eliminate these metabolites in ^{13}C-labelling experiments carried out in isotopic instationary state. This was achieved thanks to a methodology we developed in this work, which evaluates such a need as well as the efficiency of the removal. The basic principles of the methodology that are described below rely on a labelling pulse experiment with fully ^{13}C-labelled substrate. Additionally, the latter methodology permits to quantify the fractions of metabolites inside and outside the cells and the results were compared to direct quantification through LC-MS/MS approach (section *'Quantification of intracellular and extracellular ratios of metabolites'*).

Basic principle of the ^{13}C pulse experiment

The general principle consists in feeding the organism with a ^{13}C-labelled substrate in which all carbon positions contain the same proportion of ^{13}C isotopes (same ^{13}C abundance). This can be easily achieved by using a mixture of the uniformly labelled form of the substrate with the unlabelled form.

The global ^{13}C abundance A_M of a metabolite M can be calculated from the specific ^{13}C specific abundance by using equation 1:

$$A_M = \frac{\sum_{i=1}^{n} A(C_i^M)}{n} \quad \text{eq (1)}$$

where $A(C_i^M)$ refers to the specific ^{13}C abundance of the *i*-th carbon atom of the metabolite M, which contains *n* carbon atoms in total. Because i) the specific ^{13}C abundance for all the carbon of the substrate are the same [i.e. $A(C_i^S) = A(C_{i+1}^S) = A_S$ where *S* is the substrate], and ii) each carbon of each metabolite originates from a carbon of the substrate, equation 1 can be reduced to $A_M = A(C_i^S) = A_S$, $\forall\, i$. Hence, the ^{13}C abundance in both the intermediates and the end-products of the metabolism are identical to that of the substrate independently of the flux distribution through the metabolic network provided i) the isotopic steady-state has been reached, ii) isotopic fractionation phenomenon are negligible, and iii) no exogenous CO_2 is incorporated within the metabolic network.

Inner (M_{in}) and outer (M_{out}) pools of metabolites may be discriminated using a ^{13}C pulse experiment. This relies on the assumption that M_{in} first incorporated the label before being exported and – eventually – exchanged between both compartments. If there is no exchange or if the latter is too low to achieve fast isotopic equilibrium between M_{in} and M_{out}, the ^{13}C abundance of M_{in} would reach the theoretical A_S value prior than that of M_{out}. In contrast, if the exchange between intracellular and extracellular metabolites is fast enough, the A_S value would be reached at the same time by both pools (exemplified in Figure 5). If so, no separation between cells and medium is required. M_{in} and M_{out} may thus be considered as a unique pool in the metabolic flux model. In opposite, when equilibrium exchange flux is null or too low to reach a fast isotopic equilibrium, the separation becomes essential to distinguish the incorporation of the labelling within the cell from that in its surrounding (Figure 5).

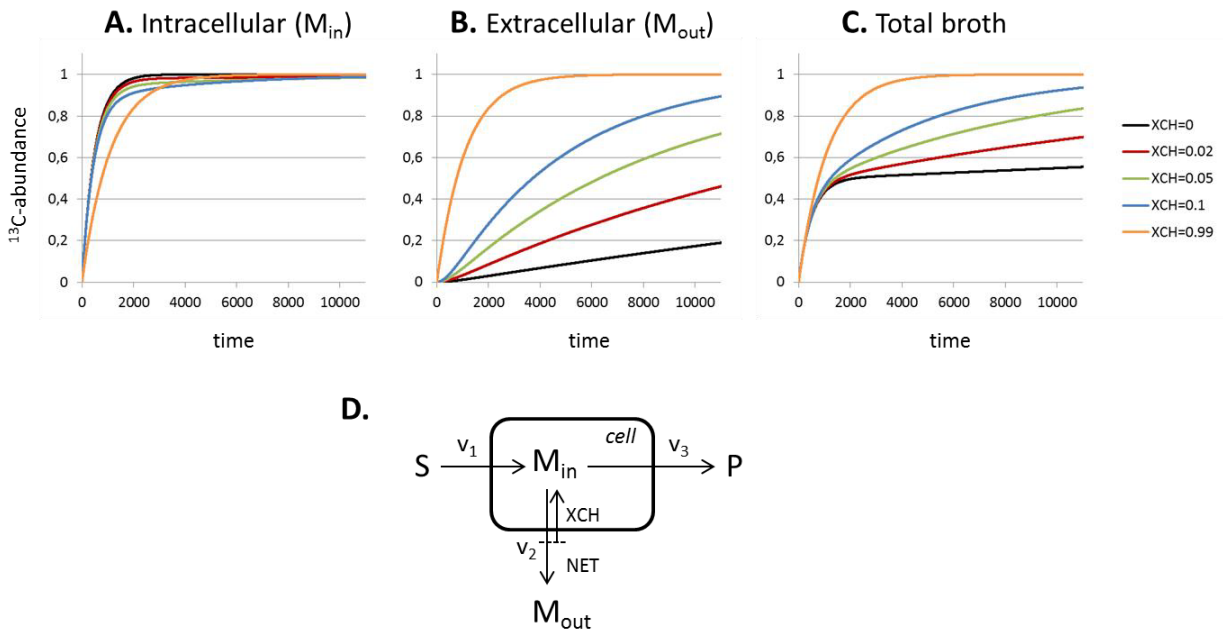


Figure 5. Simulations of labelling dynamics of the metabolite M inside (A) and outside (B) the cell, and in total broth (C). The pulse consists in switching the substrate S from fully unlabelled to fully labelled at t=0, i.e. the isotopic steady state is reached when all metabolites are fully labelled (^{13}C-abundance=1). These data have been simulated considering the network (D) with the following parameters (arbitrary units): v_1=1; v_2^{NET}=0.01; v_3=0.99; [M_{in}]=10; [M_{out}]=0.01 at t=0; $volume_{in}$=1; $volume_{out}$=1000; and for different values of exchange coefficient of v_2 (0, 0.02, 0.05, 0.1, 0.99). While the concentration of M_{in} is constant over time (the "cell" is in metabolic steady state), M_{out} is accumulated in the extracellular medium as observed in batch experiments.

Sampling without separation of cells and medium

The pulse experiment was carried out by adding [U-^{13}C] glucose during an exponential phase of a growth of *E. coli* on glucose as sole carbon source. The ^{13}C abundance of the extracellular glucose in the medium (A_S) was measured from the determination of the ^{13}C abundance of the anomeric carbon of the glucose by NMR, from three filtrate samples collected just at the end of the pulse. In order to capture the kinetics of label incorporation, IDs of metabolic intermediates were examined during 90 s after the isotopic pulse collecting samples every 2 s during the 30 first seconds and then, every 15 s for the last 60 seconds. ^{13}C abundances of these metabolic intermediates have been calculated from their IDs using the following equation:

$$A_M = \sum_{i=0}^{n} \frac{i.M_i}{n} \qquad \text{eq (2)}$$

where M_i denotes the proportion of isotopologue having i ^{13}C atoms for a molecule containing n carbon atom.

First, we applied the procedure T consisting in simultaneously quenching and extracting the metabolites in boiling ethanol – without separation step – for sampling the ^{13}C-metabolites during the pulse. The isotopic data (Figure 6A) showed i) that examined metabolites reached a plateau after about 30 s since no significant variation was observed between 30 and 90 s and ii) that the values at the plateau depend on metabolites (given in brackets for each metabolite in the legend of Figure 6A) and are significantly lower than the ^{13}C abundance of the substrate (0.61±0.01). From these results, it can be concluded that the exchange fluxes between pools of metabolites inside and outside the cells are very low – and even negligible in respect of the pulse duration –; otherwise the ^{13}C abundance of metabolites would be identical and close to the ^{13}C abundance value of the substrate, as exemplified in Figure 5.

Simulations (Figure 5) suggest that the lower ^{13}C-abundances measured for metabolites originate from the label dilution brought by the unlabelled extracellular pools of metabolic intermediates. In order to test this hypothesis, we applied the same protocol on supernatant samples obtained by direct filtration of the broth medium through a 0.2 µm filter and collected every 6 seconds during the pulse. We found that the ^{13}C abundances of the metabolic intermediates present in the extracellular compartment were stable during the first 70 seconds (Figure 6B). Although the ^{13}C abundance of the glucose was 0.65±0.01, the values for the metabolites (given in brackets in the legend of Figure 6B) are similar for all of them and close to the natural abundance of ^{13}C isotope (0.0107). Thereafter, we observe that the ^{13}C abundance tended to increase during the last 10 seconds of the isotopic pulse, likely related to the excretion of metabolites from the cells to their surroundings according to the simulation displayed on Figure 5.

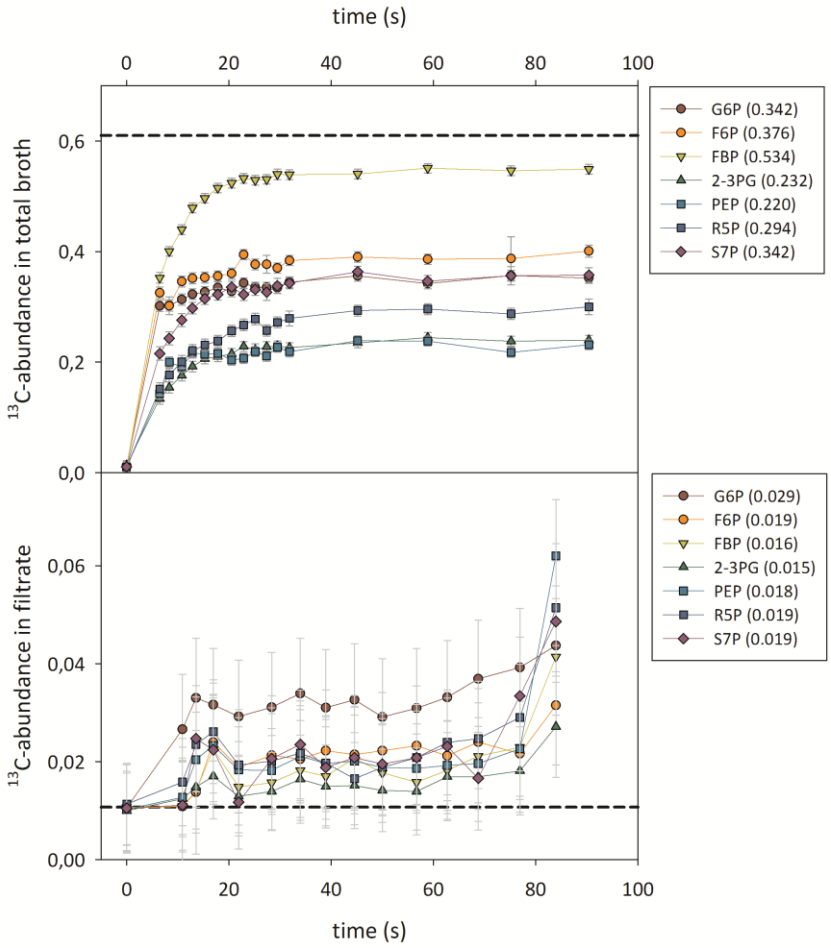


Figure 6. Labelling dynamics of metabolic intermediates from total (upper graph) and filtrate (lower graph) samples after the addition of [U-^{13}C]glucose at t=0s. Numbers between brackets are the average of enrichments observed between 30 and 70s, which are not significantly different. Dotted line denotes ^{13}C-abundance of extracellular glucose measured by NMR to be 0.61 (upper graph), or natural abundance of ^{13}C isotopes (0.0107) (lower graph).

These results confirmed our previous findings and lead to the conclusion that appropriate sampling procedures for isotopic experiments in instationary state require to discriminate between both pools either by a differential analysis of the ^{13}C-metabolites in the broth and those in the supernatant as this is described here or by physical separation of the cells from the broth (e.g. centrifugation, see below).

Sampling with separation of cells and medium

The isotopic pulse methodology has been applied with the same sampling frequency as described above quenching the metabolism in cold ethanol (E) or glycerol (G) (being precooled at -80 °C and -20 °C, respectively) in respect with their ability to rapidly block the metabolism activity. The samples were centrifuged and the ^{13}C-metabolites were recovered with boiling ethanol extraction prior to their IC-MS/MS analysis. Time-courses

of ^{13}C abundances displayed plateaus which were reached in a time span of 15 s and 25 s after the isotopic pulse depending on the metabolites, meaning that a pseudo isotopic steady-state is achieved after 30 s (Figure 7).

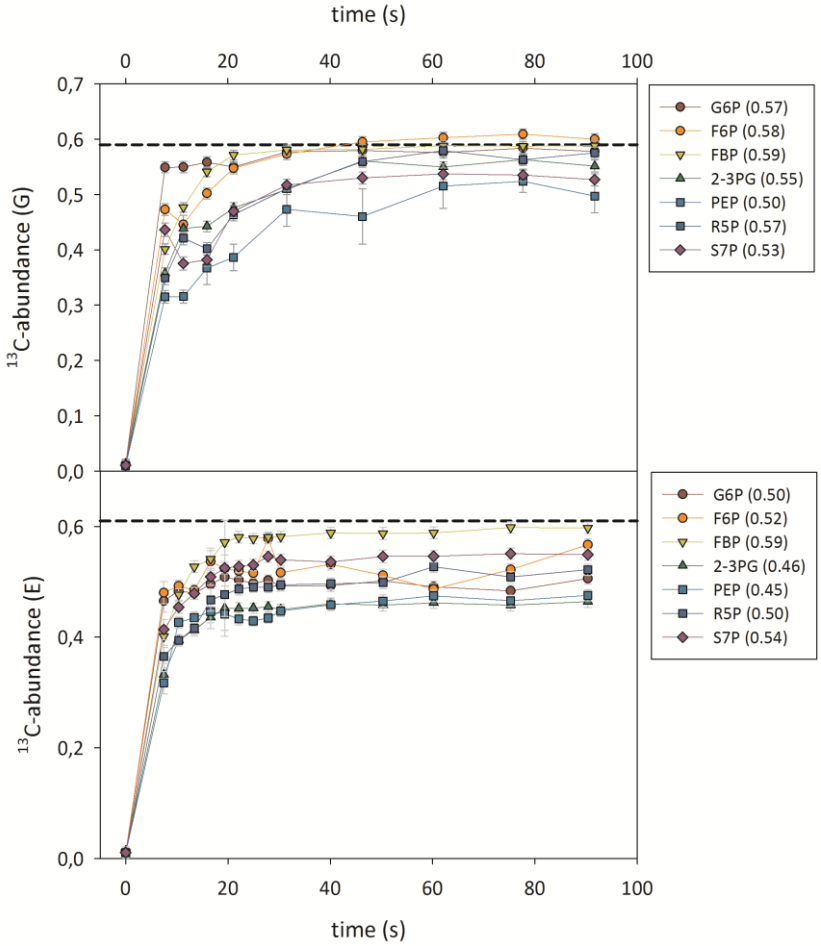


Figure 7. Labelling dynamics of metabolic intermediates from samples produced with the G (upper graph) and E (lower graph) procedures. The pulse is performed by addition of [U-^{13}C]glucose at t=0s. Numbers between brackets are the average of enrichments observed between 30 and 70s. Dotted lines represents ^{13}C-abundance of extracellular glucose measured by NMR to be 0.59 (upper graph) and 0.61 (lower graph).

According to the duration of the plateaus (45 s), mean value of ^{13}C abundance for each ^{13}C-metabolite was calculated and quoted in Figure 7 where the theoretical values – determined by NMR analysis for each pulse – are given in legend for each sampling procedures. As would be expected, ^{13}C abundance of metabolites extracted from the cell pellets are closer to the predicted ones compared to those measured from the total procedure. Moreover, we observed that the centrifugation in cold glycerol allows to reach ^{13}C abundance values close to the expected ones although it is not achieved in cold ethanol. It has been shown in this work that both procedures lead to an efficient

quenching of metabolism. We therefore hypothesized that significant amounts of extracellular metabolites remained in the interstitial space within the pellets or precipitated or co-precipitated with the salts of the cultivation medium as a consequence of the high ethanol content (about 80% v/v). Complete removal of extracellular metabolites should require thus additional steps of rinsing the cell pellets for improving the separation (Douma, de Jonge et al. 2010).

Quantification of intracellular and extracellular fractions of metabolites

Once the pseudo isotopic steady-state has been reached after the pulse, the measured ^{13}C abundances of each metabolite correspond to the sum of the ^{13}C abundance of extra- and intracellular pools weighted by their relative fraction inside and outside the cells, as described by equation (3) and exemplified in Figure 5.

$$A_{total} = (1 - f_{in}).A_{out} + f_{in}.A_{in} \qquad \text{eq (3)}$$

where A_{total} refers to the ^{13}C abundance measured in the sample, f_{in} the relative fraction of the metabolite inside and outside the cells, A_{out} the ^{13}C abundance of the extracellular pool, and A_{in} the ^{13}C abundance of the intracellular pool. The proportion f_{in} can be calculated by the equation 4 knowing A_{out} – which is measurable or presumed to be at the natural abundance – and A_{in} which is calculated from the enrichment of the substrate.

$$f_{in} = \frac{A_{total} - A_{out}}{A_{in} - A_{out}} \qquad \text{eq (4)}$$

As a proof of concept, we determined the fractions of different central metabolites inside and outside the cell in broth sample. They were compared to ratios determined by state-of-art quantitative metabolomics (Mashego, Wu et al. 2004; Wu, Mashego et al. 2005). Figure 6 presents the results of the measurements of a number of different central metabolites by differential titrations in the broth and in the filtrate (black box) (Taymaz-Nikerel, de Mey et al. 2009) and by ^{13}C-labelled pulse experiments (CLE) for the three different sampling procedures (T, white box ; E, gray box and G, dark gray box). These data are averages from three replicate samples for metabolomics quantification or result from the use of equation 4 (A_{out} is the natural ^{13}C abundance and A_{total} is a mean value of the plateaus reached during the ^{13}C labelling experiments). From these measurements, it can be first seen that fractions of intracellular metabolites measured

by quantitative metabolomics (black box) and by the T procedure (white box) matched very well and that CVs are in same order of magnitude (mean of CVs of 5.1% and 1.6% respectively). This demonstrates that fractions of intra- and extracellular metabolites may be estimated thanks to ^{13}C labelling pulse experiments, with accuracy and precision similar to that of well-established quantitative metabolomics methods.

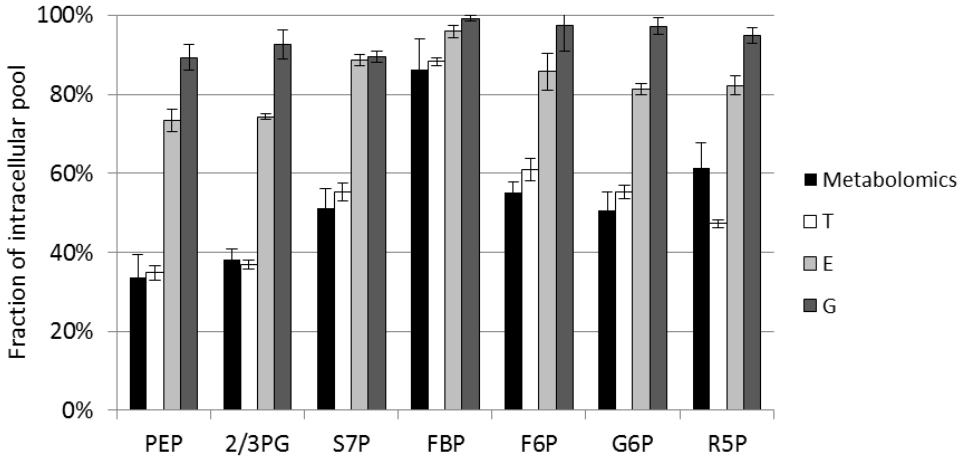


Figure 8. Relative fractions of intracellular pools of central intermediates in total broth determined by metabolomics (black bars) and with the fast-heating procedure (T, white bars). Fractions of intracellular pools have also been determined in samples produced using E and G protocols.

As previously published (Bolten, Kiefer et al. 2007; Taymaz-Nikerel, de Mey et al. 2009), the extracellular fractions depends on metabolites (FBP pool is almost intracellular although the fraction of intracellular pools of both PEP and 2/3PG is approximately of 1/3). That indicates that the presence of metabolic intermediates outside the cells cannot be solely related to cellular lyses and confirms earlier published work (Taymaz-Nikerel, de Mey et al. 2009). Finally, as it was expected, the centrifugation in cold solvents improved significantly the fractions of intracellular metabolites even though it was still incomplete. Centrifugation in cold glycerol allows better recovery of intracellular metabolites, reaching a mean value of 94% compared to 83% with cold ethanol. This result is in agreement with previous report (Villas-Boas and Bruheim 2007).

Sampling in practice

Compared to stationary isotope labelling experiments, isotopic instationary experiments offers several benefits as the access to particular configurations of metabolic networks, the establishment of metabolic flux maps for autotroph and methylotroph organisms, etc. (Wiechert and Noh 2005; Noh and Wiechert 2011). To measure the distribution of

metabolic fluxes in isotopic instationary and metabolic stationary states, both the titration of intracellular pools of metabolite and the quantification of their IDs have to be achieved. Since no significant changes in intracellular pool of metabolite occur within the time span of the isotopic pulse – that should be verified beforehand –, one unique sampling for quantitative metabolomics is sufficient. To capture the dynamic of labelling incorporation in intracellular metabolites, the sampling of ^{13}C-metabolites must be performed with sufficiently high frequency. This is highly dependent on the organism, the condition of its cultivation, the portion of metabolism examined, etc. In this work, where the bacterium *E. coli* is cultivated on glucose, it can be seen from the time course data (Figure 6 and 7), that G6P, the first metabolic intermediate resulting from the glucose uptake *via* the phosphotransferase system (PTS), reaches isotopic steady state before 10 seconds. The time span is significantly wider for R5P of which the labelling is stable after 20 seconds as shown on Figure 6. This means that the frequency of sampling should be in the order of the second. Although this remains challenging, it is manageable within such short timescales to perform the whole broth sampling procedure (T) or the cold glycerol sampling procedure (G). In the first case, the IDs of intracellular metabolites (ID_{in}) can be calculated from measured IDs (ID_{total}) taking into account the IDs of extracellular pools of metabolite (ID_{out}) by the following equation:

$$ID_{in} = \frac{ID_{total} - (1 - f_{in}).ID_{out}}{f_{in}}$$

where f_{in}, which refers to the fraction of intracellular pool, can be measured by quantitative metabolomics methods. As shown in this work and due to the large turnover of metabolites present in the supernatant, the ^{13}C abundance of the later can be assumed to be at the natural abundance, something that can be verified beforehand by withdrawing supernatant samples once the labelling reaches the plateaus.

5. Conclusion

In this work, it is shown that the choice of the most suitable procedure for sampling ^{13}C-metabolites from *E. coli* cultivations highly depends on the type of labelling experiments carried out. In particular, the presence of significant amount of metabolite outside the cells addressed the question of the need for separating them from the inner pools according to whether the experiment is performed under isotopic stationary or isotopic

instationary states. Obviously, this operation has to be realized ensuring the complete blockage of the all metabolic activities. In that respect, from our benchmarking of quenching procedures for subsequent ^{13}C-metabolites analysis to establish metabolic flux maps in isotopic steady-state, it can be conclude that i) cold solvent tested here are appropriated for efficient quenching of metabolism and ii) fast-filtration and subsequent quenching of cell pellets with liquid nitrogen does not permit an immediate blockage of the metabolism, resulting in erroneous metabolic fluxes especially at the level of the PP pathway. Therefore, this procedure cannot be applied. We believe that despite the handling care, the metabolism of the cells on filters is enough perturbed to induce changes in flux distribution that are then captured by ^{13}C-metabolites analyses. Furthermore, the data reported here demonstrated that in isotopic steady-state, the separation of the cells from the broth is not necessary. This offers the ability to perform the quenching and the extraction simultaneously with boiling ethanol as it has been done here or with appropriate cold solvent mixtures like acidic acetonitrile/methanol/water mix developed for *E. coli* metabolome analysis by Rabinowitz and coworkers (Bennett, Kimball et al. 2009) or to realize first the quenching of the whole broth with liquid nitrogen for example, prior to the ^{13}C-metabolite extraction from the frozen broth by appropriate methods. Finally, simple sampling procedures may be yielded that could easily be automated for high-throughput platforms.

From the methodology developed in this work, which is based on an isotopic pulse experiment under metabolic steady-state, our findings showed that during an instationary isotope labelling experiment, intra- and extracellular pools of metabolites must be discriminated in order to obtain specifically the IDs of the intracellular metabolites. This is due to the exchange fluxes of the metabolites between the inside and the outside of the cells which is too low to obtain a rapid isotopic equilibrium. Nevertheless, this exchange flux might depend on the kind of metabolite and it would be relevant to extend this observation to metabolites belonging to other chemical classes. The intracellular metabolites can be recovered by centrifugation in cold solvent prior to their extraction. The data reported here showed that centrifugation in cold glycerol allows to achieve the recovery of metabolites with the highest purities compared to centrifugation in cold ethanol. An another way to obtain reliable data is to apply a differential analysis of IDs whereby the IDs measured for the metabolites of the whole broth are corrected with the IDs of the metabolites solely present in the supernatant.

Our observations lead to the conclusion that the IDs of extracellular metabolites are only related to the natural abundance of the ^{13}C isotope during the isotopic pulse, and can thus be predicted. However, the correction of the IDs measurements implies also to know the fraction of intracellular metabolites in the whole broth. This fraction can be determined before the isotopic pulse thanks to a differential titration of metabolites in the cells from those present in the supernatant.

Finally, The methodology presented in this work may also be used to estimate the fraction of intracellular metabolites in a sample as well as the purity of the recovery of intracellular metabolites after sample preparation procedure. This methodology appears to be as robust as quantitative metabolomics approach in terms of both accuracy and precision. Because fractions of intracellular metabolites depend on the cellular-type (Bolten, Kiefer et al. 2007; Villas-Boas and Bruheim 2007), it might be applicable for others organisms than *E. coli*. We also believe that this approach would be a reliable tool to estimate exchange fluxes of metabolites between two compartments at the cellular level (between inside and outside the cells) or even at subcellular level (between cytosol and organelles) in eukaryotes.

6. References

Bennett, B. D., E. H. Kimball, et al. (2009). "Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli." Nat Chem Biol **5**(8): 593-599.

Bolten, C. J., P. Kiefer, et al. (2007). "Sampling for metabolome analysis of microorganisms." Anal Chem **79**(10): 3843-3849.

Canelas, A. B., A. ten Pierick, et al. (2009). "Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics." Anal Chem **81**(17): 7379-7389.

Chen, X., A. P. Alonso, et al. (2011). "Synergy between (13)C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in E. coli." Metab Eng **13**(1): 38-48.

Cocaign-Bousquet, M., C. Garrigues, et al. (1995). "Rational Development of a Simple Synthetic Medium for the Sustained Growth of Lactococcus-Lactis." Journal of Applied Bacteriology **79**(1): 108-116.

Dauner, M. (2010). "From fluxes and isotope labeling patterns towards in silico cells." Curr Opin Biotechnol **21**(1): 55-62.

Douma, R. D., L. P. de Jonge, et al. (2010). "Intracellular metabolite determination in the presence of extracellular abundance: Application to the penicillin biosynthesis pathway in Penicillium chrysogenum." Biotechnol Bioeng **107**(1): 105-115.

Fuhrer, T., E. Fischer, et al. (2005). "Experimental identification and quantification of glucose metabolism in seven bacterial species." J Bacteriol **187**(5): 1581-1590.

Iwatani, S., Y. Yamada, et al. (2008). "Metabolic flux analysis in biotechnology processes." Biotechnol Lett **30**(5): 791-799.

Kiefer, P., C. Nicolas, et al. (2007). "Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry." Anal Biochem **360**(2): 182-188.

Kleijn, R. J., J. M. Geertman, et al. (2007). "Metabolic flux analysis of a glycerol-overproducing Saccharomyces cerevisiae strain based on GC-MS, LC-MS and NMR-derived C-labelling data." FEMS Yeast Res **7**(2): 216-231.

Kleijn, R. J., W. A. van Winden, et al. (2005). "Revisiting the 13C-label distribution of the non-oxidative branch of the pentose phosphate pathway based upon kinetic and genetic evidence." FEBS J **272**(19): 4970-4982.

Kohlstedt, M., J. Becker, et al. (2010). "Metabolic fluxes and beyond-systems biology understanding and engineering of microbial metabolism." Appl Microbiol Biotechnol **88**(5): 1065-1075.

Mashego, M. R., L. Wu, et al. (2004). "MIRACLE: mass isotopomer ratio analysis of U-13C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites." Biotechnol Bioeng **85**(6): 620-628.

Massou, S., C. Nicolas, et al. (2007). "Application of 2D-TOCSY NMR to the measurement of specific(13C-enrichments in complex mixtures of 13C-labeled metabolites." Metab Eng **9**(3): 252-257.

Massou, S., C. Nicolas, et al. (2007). "NMR-based fluxomics: quantitative 2D NMR methods for isotopomers analysis." Phytochemistry **68**(16-18): 2330-2340.

Millard, P., F. Letisse, et al. (2012). "IsoCor: Correcting MS data in isotope labeling experiments." Bioinformatics.

Nakahigashi, K., Y. Toya, et al. (2009). "Systematic phenome analysis of Escherichia coli multiple-knockout mutants reveals hidden reactions in central carbon metabolism." Mol Syst Biol **5**: 306.

Nicolas, C., P. Kiefer, et al. (2007). "Response of the central metabolism of Escherichia coli to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase." FEBS Lett **581**(20): 3771-3776.

Noh, K., K. Gronke, et al. (2007). "Metabolic flux analysis at ultra short time scale: isotopically non-stationary 13C labeling experiments." J Biotechnol **129**(2): 249-267.

Noh, K. and W. Wiechert (2011). "The benefits of being transient: isotope-based metabolic flux analysis at the short time scale." Appl Microbiol Biotechnol **91**(5): 1247-1265.

Paczia, N., A. Nilgen, et al. (2012). "Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms." Microb Cell Fact **11**(1): 122.

Ruhl, M., B. Rupp, et al. (2012). "Collisional fragmentation of central carbon metabolites in LC-MS/MS increases precision of (1)(3)C metabolic flux analysis." Biotechnol Bioeng **109**(3): 763-771.

Schaub, J., C. Schiesling, et al. (2006). "Integrated sampling procedure for metabolome analysis." Biotechnol Prog **22**(5): 1434-1442.

Sokol, S., P. Millard, et al. (2011). "influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labeling experiments." Bioinformatics.

Szyperski, T. (1995). "Biosynthetically directed fractional 13C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism." Eur J Biochem **232**(2): 433-448.

Taymaz-Nikerel, H., M. de Mey, et al. (2009). "Development and application of a differential method for reliable metabolome analysis in Escherichia coli." Anal Biochem **386**(1): 9-19.

Toya, Y., N. Ishii, et al. (2007). "Direct measurement of isotopomer of intracellular metabolites using capillary electrophoresis time-of-flight mass spectrometry for efficient metabolic flux analysis." J Chromatogr A **1159**(1-2): 134-141.

Toya, Y., N. Ishii, et al. (2010). "13C-metabolic flux analysis for batch culture of Escherichia coli and its Pyk and Pgi gene knockout mutants based on mass isotopomer distribution of intracellular metabolites." Biotechnol Prog **26**(4): 975-992.

van Winden, W. A., J. C. van Dam, et al. (2005). "Metabolic-flux analysis of Saccharomyces cerevisiae CEN.PK113-7D based on mass isotopomer measurements of (13)C-labeled primary metabolites." FEMS Yeast Res **5**(6-7): 559-568.

Villas-Boas, S. G. and P. Bruheim (2007). "Cold glycerol-saline: the promising quenching solution for accurate intracellular metabolite analysis of microbial cells." Anal Biochem **370**(1): 87-97.

Wiechert, W. (2001). "13C metabolic flux analysis." Metab Eng **3**(3): 195-206.

Wiechert, W., M. Mollney, et al. (2001). "A universal framework for 13C metabolic flux analysis." Metab Eng **3**(3): 265-283.

Wiechert, W. and K. Noh (2005). "From stationary to instationary metabolic flux analysis." Adv Biochem Eng Biotechnol **92**: 145-172.

Winder, C. L., W. B. Dunn, et al. (2008). "Global metabolic profiling of Escherichia coli cultures: an evaluation of methods for quenching and extraction of intracellular metabolites." Anal Chem **80**(8): 2939-2948.

Wittmann, C. (2007). "Fluxome analysis using GC-MS." Microb Cell Fact **6**: 6.

Wu, L., M. R. Mashego, et al. (2005). "Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards." Anal Biochem **336**(2): 164-171.

Yuan, J., B. D. Bennett, et al. (2008). "Kinetic flux profiling for quantitation of cellular metabolic fluxes." Nat Protoc **3**(8): 1328-1340.

Zamboni, N., S. M. Fendt, et al. (2009). "(13)C-based metabolic flux analysis." Nat Protoc **4**(6): 878-892.

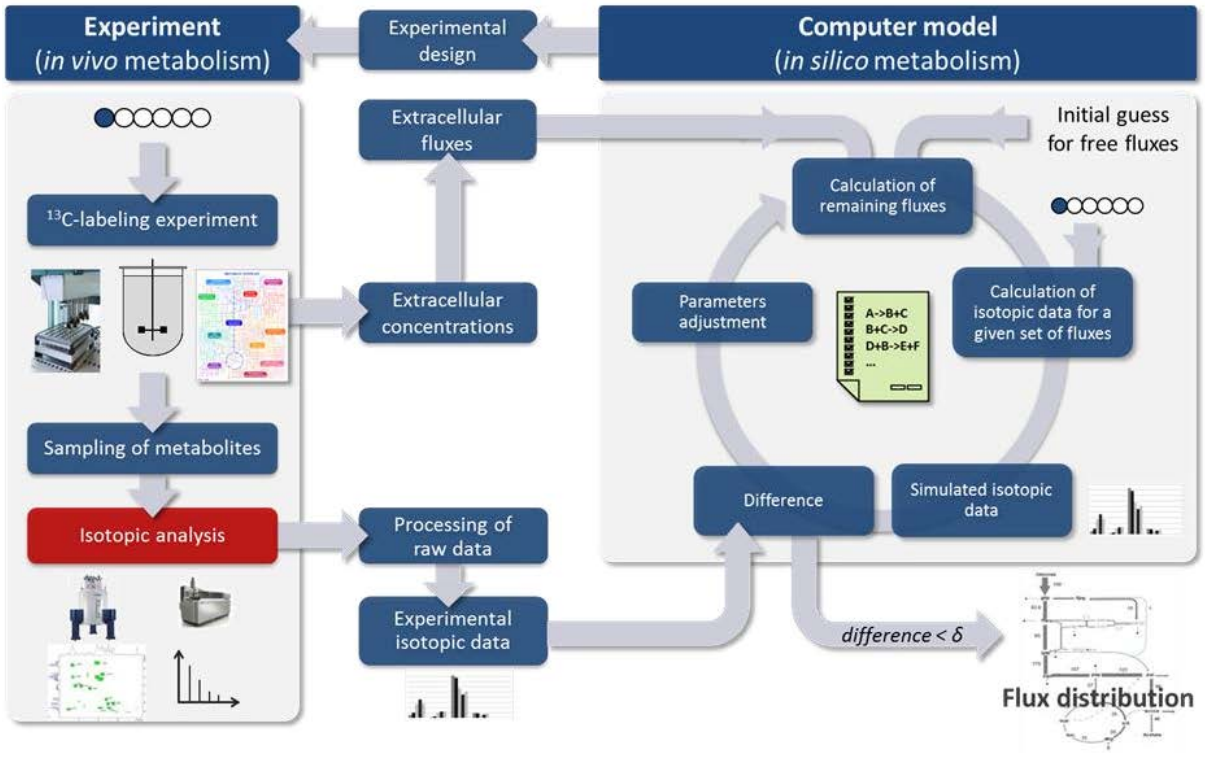
Zhao, Z., K. Kuijvenhoven, et al. (2008). "Isotopic non-stationary 13C gluconate tracer method for accurate determination of the pentose phosphate pathway split-ratio in Penicillium chrysogenum." Metab Eng **10**(3-4): 178-186.

Chapter 4

Isotopic studies of metabolic systems by mass spectrometry: using the Pascal's triangle to produce biological standards with fully controlled labelling patterns

-

Theoretical aspects and applications



NOTE: *this chapter contains the first draft of a manuscript in preparation for Analytical chemistry.*

1. Abstract

Mass spectrometry (MS) is widely used for isotopic studies of metabolism in which the incorporation of isotopic tracers into metabolites is monitored. The reliability of such approaches relies on the accurate quantification of each isotopologue – that refers to molecules of the same molecular entity differing only in the number of isotopes – within isotopic cluster. However, the experimental workflow and/or MS instruments may provoke fractionation or distortion effects, prejudicial for the correctness of biological interpretation. The lack of relevant standards restricts the investigations on the accuracy and precision of the measurement of isotopologue distributions. Here, we present a strategy to quantify both the accuracy and the precision on the measurements of the complete isotopic cluster and ^{13}C-enrichment by MS. In that aim, intracellular metabolites having a predictable distribution in isotopologues, have been biosynthesized through the feeding of micro-organisms with a mixture of labelled substrate in which ^{13}C atoms are binomially distributed. Two model microorganisms, *Pichia augusta* and *Escherichia coli*, have been cultivated respectively on methanol or acetate as sole carbon source and at appropriate proportions of all existing isotopic species. These proportions have been chosen to produce metabolites in which the isotopologue distribution should remarkably give the binomial coefficients found in the Pascal's triangle. The experimental IDs were in close agreement with predictions for entire isotopologue distributions of intracellular metabolites with carbon skeletons ranging from 3 to 10 carbon elements. Statistical analyses showed that the isotopologue distributions by our LC-MS/MS procedures were quantified within 1% accuracy and of ±1% precision.

2. Introduction

Mass spectrometry (MS) is extensively used for the purpose of isotopic studies of metabolism, either to unravel the nature and rate of biochemical reactions or to identify or quantify metabolites (Wiechert, Mollney et al. 2001; Zamboni and Sauer 2004; Bajad, Lu et al. 2006; Wiechert 2007; Wittmann 2007; Peyraud, Kiefer et al. 2009; Quek, Wittmann et al. 2009; Taymaz-Nikerel, de Mey et al. 2009). MS allows the discrimination

of molecules according to the number of label atoms incorporated, i.e. allows the discrimination of isotopologues (IUPAC 1997). The distribution of the various isotopologues of a compound is obtained from the quantitative analysis of the isotopic cluster in the MS spectrum (Hellerstein and Neese 1999). The reliability of the MS measurements determines the quality of the isotopic information and hence the biological value of ^{13}C-labelling experiments (Antoniewicz, Kelleher et al. 2006). The potential sources of errors in such investigations are diverse and stand at all levels of the experimental workflow. These errors have been detailed in various publications (Hellerstein and Neese 1999; Dauner and Sauer 2000; Vogt, Wachter et al. 2003; Antoniewicz, Kelleher et al. 2007) and their description is outside the scope of this work. The reliability – in terms of accuracy and precision – of isotopic measurements made by MS can be assessed by analyzing standards with known or predictable isotopic contents. Classically this is done using standards or biological samples at natural isotopic abundance, or using commercially available labeled compounds (Hellerstein and Neese 1999; Antoniewicz, Kelleher et al. 2007; Kiefer, Nicolas et al. 2007; Zamboni, Fendt et al. 2009; Ruhl, Rupp et al. 2012). However, a major issue in the MS analysis of isotopologue distributions is the potential occurrence of isotopic effects all along the analytical procedure, which can induce significant differences in the behavior of analytes according to their isotopic content. Hence, the various isotopologues of a given compound do not have necessarily the same analytical behavior. This means that the reliability of MS measurement should be evaluated for each individual isotopologue of the compounds of interest. This is not possible from standards at natural isotopic abundance for which only the lightest isotopologues can be measured. The same limitations exist with fully labeled metabolites – which are now routinely produced for the purpose of quantitative metabolomics (Mashego, Wu et al. 2004; Wu, Mashego et al. 2005) – in which only the heaviest isotopologue is usually detected. Defined mixtures of commercially available labeled compounds can be used for the evaluation process, but most compounds are not commercially available (in all relevant labeled forms) or at prohibitive costs.

The biological production of isotopically-labelled compounds as standard can be a valuable alternative to the above-mentioned approaches. It provides several advantages, including the capability to obtain a broad spectrum of labeled compounds, the capability to generate different types of standard samples (e.g. extracts, biofluids, etc) that have

similar composition and analytical behaviors (e.g. comparable matrix effects, etc) as the samples of interest, the capability to evaluate different types of analytical methods or instruments, etc. To use such materials as analytical standards, the isotopic composition of the biologically-produced compounds must be known or predicted with a high precision. Carbon-13 labeled compounds with predicted isotopologue compositions can be produced using organisms able to grow on methanol (methylotrophs) or other C1 compounds as sole carbon source (Shen, Shen et al. 2009). In this situation all the carbons of all metabolites are derived from the same unique source, and their labeling state is fully determined by the ^{13}C content of the C1 compound used as carbon source. Using this property, ^{13}C-labeled amino-acids with predicted isotopologue content were produced and used to evaluate the reliability of GC-MS measurements, thought only the first 2-3 isotopologues per metabolite were truly evaluated (Shen, Shen et al. 2009).

There is a current need to produce biological standards from any type of organisms and covering all metabolites (hence all analytical methods and instruments), and for all isotopes of interest in metabolic studies. In the present work, we provide a complete conceptual and practical framework for full control of the labelling patterns of biologically-produced metabolites. We first show that the isotopic labeling of metabolites can be not only predicted but also fully controlled if the isotopic composition of the label input is defined appropriately. The proposed strategy was first applied to the production of ^{13}C-labeled intracellular metabolites with fully pre-determined IDs by a methylotrophic organism, which can be used to evaluate the reliability of LC-MS(/MS) measurements. Then, the approach was extended to multicarbon label inputs, which extends the biological production of labeled standards to various types of metabolisms and various types of organisms. The approach was validated for carbon-labelled compounds, but is generally valid for any type isotope tracer.

3. Experimental section

Organisms and cultivations

Pichia augusta was grown on YNB medium without amino acids (DIFCO, BD, UK) and supplemented with 15 $g.L^{-1}$ of methanol as unique carbon source. Methanol (HPLC grade) was obtained from Sigma-Aldrich Chemie (France) and ^{13}C-methanol from Eurisotop (St Aubin, France). Batch cultivations were carried out at 40 °C in a 500 mL

Multifors bioreactor (Infors, Switzerland). The pH was regulated at 4.0 by the addition of NaOH (1 M) and the dissolved oxygen content was maintained above 20% saturation by aeration with synthetic gas 80%N_2/20%O_2 (Air Liquide, Paris, France) containing less than 1 ppmv CO_2.

Escherichia coli K12 MG1655 was grown on minimal medium containing 5 mM KH_2PO_4, 10 mM Na_2HPO_4, 9 mM NaCl, 40 mM NH_4Cl, 0.8 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.3 mM of thiamine and 45 mM of acetate as unique carbon source. ^{12}C-acetate was obtained from Sigma (St. Louis, MO, USA) and 1-^{13}C-, 2-^{13}C- and U-^{13}C-acetate from Eurisotop. Acetate and thiamine were sterilized by filtration, other compounds were autoclaved separately. Batch cultivations were carried out in 1 L baffled flask (37 °C, 200 rpm) with 100 mL of medium sparged with synthetic gas 80%N_2/20%O_2 containing less than 1 ppmv CO_2 (Air Liquide). The pH was adjusted to 7.0 every hour by adding appropriate volumes of HCl 2M.

Cell growth was monitored by measurement of optical density at 600nm with a Genesys 6 spectrophotometer (Thermo, Carlsbad, USA). Samples of intracellular metabolites were collected in mid-exponential growth phase for both *P. augusta* and *E. coli* cultivations with the procedures described below.

Sample preparation

All solvents were LC-MS grade and obtained from Sigma-Aldrich Chemie (France). For sampling intracellular metabolites of *P. augusta*, 1 mL of broth were filtered (Sartolon polyamide 0.2 µm, Sartorius, Goettingen, Germany). Cells were washed with 1 mL of fresh cultivation medium containing the same mix of labeled substrate as the one used for ^{13}C-labelling experiment. Filters were rapidly plunged into liquid nitrogen and stored at -80 °C. The intracellular metabolites were extracted by incubating the filters in closed glass tubes containing 5mL of an ethanol/$H_2$0 (75/25) mixture at 95 °C for 5 min. The cellular extracts were cooled on ice, and the cell debris were removed by centrifugation (5 min, 6 500 g, 4 °C) (Bolten, Kiefer et al. 2007).

For sampling intracellular metabolites of *E. coli*, 120 µL of broth were rapidly sprayed into precooled centrifuged tubes maintained at -80 °C and containing 500 µL of ethanol, homogenized using a vortex and centrifuged at -20 °C (12 000 g, 5 min) (Sigma 3-18K,

Sigma, Germany). Extraction was performed by pouring on the pellets 5mL of an ethanol/H_20 (75/25) solution at 95 °C and incubated in closed tubes for 2min.

Cellular extracts were cooled on ice and evaporated 4 h (SC110A SpeedVac Plus, ThermoSavant, USA). The remaining aqueous extract was freeze-dried, resuspended in 200 µL of milliQ water, and stored at -80 °C (Taymaz-Nikerel, de Mey et al. 2009).

Preparation and quality of label inputs

Special care was paid to the preparation of the isotopic mixtures to be used as label inputs. The actual label content of all isotopically-labeled compounds and all isotopic mixtures considered in this work were controlled by quantitative 1H 1D-NMR before being used. NMR spectra were recorded with an Avance 500 MHz spectrometer (Bruker, Switzerland) at 298 K, using a 30° pulse and a relaxation delay of 20 s. Using NMR, the two isotopic species of methanol and the four isotopic species of acetate could be fully resolved and quantified.

IC-MS/MS analysis.

Metabolite separation was performed by high performance anion exchanged chromatography as described by (Kiefer, Nicolas et al. 2007). Briefly, the separation was performed using an ICS 2000 system (Dionex, Sunnyvale, CA, USA) equipped with an EG5 eluant generator (Dionex), an AG11 pre-column (50 × 2 mm, Dionex), and an IonPac AS11 column (250 × 2 mm, Dionex). The MS/MS analysis were performed with a QTrap 4000 triple quadrupole mass spectrometer (Applied Biosystems, Foster City, CA, USA) equipped with a Turbo V source for electrospray ionization. All samples were analyzed in the negative mode giving $[M - H]^-$ ions, which were monitored in the multiple reaction monitoring (MRM) mode (with a mass resolution of 0.5 amu at half peak height) for organic acids (Cit, Suc, Fum, Mal) and phosphorylated compounds (AMP, ADP, CDP, G6P, F1P, F6P, M6P, FBP, 6PG, R5P, PEP and 2PG/3PG). The isotopic clusters of each compound was measured from relevant sets of MRM transitions in which the daughter ion was a phosphate group (PO_3^- m/z=79 or $H_2PO_4^-$ m/z= 97) for phosphorylated metabolites, and a carboxylic group ($[M-44]^-$ or $[M-45]^-$) for organic acids, according to the matrix of MRM transitions given by (Kiefer, Nicolas et al. 2007).

Processing of raw data

The Carbon Isotopologue Distributions (CIDs), which represent the relative distribution of the isotopologues of a particular molecular entity differing only in the carbon isotopes, were calculated as indicated in the Results section from the relevant isotopic clusters, after correction for naturally-occurring isotopes of elements other than carbon. This correction was performed using IsoCor (Millard, Letisse et al. 2012) considering the following natural isotopic abundances: ^{1}H 99.9885%, ^{2}H 0.0115%; ^{14}N 99.632%, ^{15}N 0.368%; ^{16}O 99.757%, ^{17}O 0.038%, and ^{18}O 0.205% (Rosman and Taylor 1998).

The ^{13}C-enrichment (p) of each metabolite was determined from its CIDs, assuming fully random distribution of ^{13}C atoms in the molecule. For each compound, the p value was approximated by iteratively minimizing the sum of the squared weighted differences between experimentally-measured CIDs and CIDs predicted from Equation (2). Minimization was performed using Brent's method, as implemented in the *optimize.fminbound()* function provided in the SciPy package of Python (http://python.org).

The standard deviations of experimental CID measurements and of experimental ^{13}C-enrichments were determined from the analysis of three independent biological samples.

Accuracy of CID prediction and sensitivity analysis

The proposed strategy for producing metabolites with fully controlled labeling patterns was evaluated by comparing experimental CIDs to CIDs predicted from equation (2). The *accuracy* of the strategy was assessed from the closeness of experimental and predicted CIDs, and was formally defined as the average of absolute differences between measured and predicted values for each isotopologue. Predicted CIDs were calculated according to the actual composition of label inputs, as measured by NMR. Confidence intervals on predicted CIDs were calculated to account for the errors in the preparation of label inputs and for the precision of NMR measurements. For each predicted CID, confidence intervals were estimated by a Monte Carlo approach in which the CID value was calculated from isotopic compositions of the label input randomly generated within the margin of the NMR measurement errors (1000 iterations/CID).

Similar approach was used to estimate the values and confidence intervals on predicted ^{13}C-enrichments.

4. Results and discussion

Basic concepts and definitions

In mass spectrometry, the information regarding the isotopic content of a molecule (chemical entity) is extracted from the analysis of the isotopic cluster (Hellerstein and Neese 1999). To properly describe this isotopic cluster and the information it contains, two different aspects have to be considered: the number of the different isotopic forms of a molecule, which depends only on its chemical structure (i.e. elemental formula), and the abundance of these isotopic forms, which depends on the synthetic origin (isotopic composition of precursors, mechanism of formation) of the metabolite and therefore is sample-dependent.

A given compound (chemical entity) exists as a mixture of several isotopic forms because of the occurrence of natural isotopes for most elements. The various isotopic forms differ one from another by the nature (e.g. ^{13}C, ^{14}C, ^{15}N, etc), number, and position(s) of the isotope(s) within the molecule. According to its elemental formula, a single molecule can therefore exist as tens or hundreds of different isotopic forms. In this paper, the term '*isotopic space*' will refer to the set of all isotopic forms that can possibly exist for a molecule or for a set of molecules.

By mass spectrometry, only isotopic forms having different molecular masses can be discriminated. For a given compound, the molecular mass varies with the isotopic composition. In this paper, the term '*isotopologue*' will be further employed, according to its IUPAC definition, to refer to molecular entities that differ only in their isotopic composition (number of isotopic substitutions), e.g. CH_4, CH_3D, CH_2D_2. This term is meaningful in MS to describe the isotopic cluster. Three different situations can be considered:

i) isotopic forms having exactly the same isotopic composition but differing in the position of the isotopes in the molecule are not isotopologues. They cannot be separated by MS and are included in the same peak of the isotopic cluster.

ii) isotopic forms not having the same isotopic composition but having the same nominal mass – e.g. $^{13}CH_3NH_2$ and CH_2DNH_2 – are isotopologues. They can be separated when the MS detector operates at high mass resolution but cannot be discriminated at a unit mass resolution. Their separation is therefore dependent on the MS detector and its operating conditions

iii) isotopic forms having neither the same isotopic composition nor the same nominal mass – e.g. CH_4, CH_3D, CH_2D_2 – are isotopologues and are separated by MS whatever the operating mass resolution of the MS detector. They are part of distinct peaks of the isotopic cluster.

It follows from the above considerations that a given peak of the isotopic cluster can contain the contribution of one or more isotopologues, and each isotopologue can contain itself one or several isotopic forms. This is determined only from the elemental formula of the considered molecule. For the purpose of this paper, only nominal masses will be considered. In such condition, each peak of the isotopic cluster will contain all the isotopologues that have the same nominal molecular mass (i.e. situation *i* and *ii*).

Distribution of isotopic forms in the isotopologue space: Pascal's triangle

The term *isotopologue space* refers to the complete set of all isotopologues that can possibly exist for a molecule or for a set of molecules. Because one particular isotopologue can contain several isotopic forms, the dimension of the isotopologue space is lower than the dimension of the isotopic space. For example, a molecule with the elemental formula $C_2H_4O_2$ – e.g. acetate – has 576 different isotopic forms but only 11 isotopologues. The distribution of the various isotopic forms into the isotopoloque space can be described using a polynomial law. For clarity, this is further described for the simplifying case of one single element, carbon, but is generally valid for any element and combinations of elements.

The element carbon has three natural isotopes, namely ^{12}C, ^{13}C and ^{14}C. The natural abundance of the radioisotope ^{14}C is very low ($<10^{-12}$) compared to that of the stable isotopes ^{12}C (0.9893) and ^{13}C (0.0107) (Rosman and Taylor 1998), and will reasonably be neglected in this work. Considering solely carbon, a molecular entity containing *n* carbon atoms has 2^n isotopic forms (2^n is the dimension of the isotopic space describing this molecule). These 2^n isotopic forms distribute in n+1 isotopologue groups (n+1 is the

dimension of the isotopologue space), which corresponds also to the maximum number of peaks in the isotopic cluster due to the contribution of carbon isotopes. The distribution of the 2^n isotopic forms in the isotopologue space follows a binomial distribution. The number of possible isotopic combinations of a molecular entity having *k* ^{13}C atoms ($0 \leq k \leq n$) is equal to the binomial coefficients $\binom{n}{k}$. The latter coefficients can be arranged into the so-called Pascal's Triangle (Figure 1) in which a particular line gives the distribution of the various isotopic forms into each isotopologue group (corresponding here to each single peak of the isotopic cluster). The sum of the binomial coefficients of a line gives the total number of isotopic forms for the considered molecular entity. For instance, according to the third line of the Pascal triangle (1-3-3-1), a molecule with 3 carbons has $2^3=8$ isotopic forms which distribute into 3+1=4 isotopologue groups: 1 isotopic form with 0 ^{13}C atom, 3 isotopic forms with 1 ^{13}C atom, 3 isotopic forms with 2 ^{13}C atoms, and 1 isotopic form with 3 ^{13}C atoms.

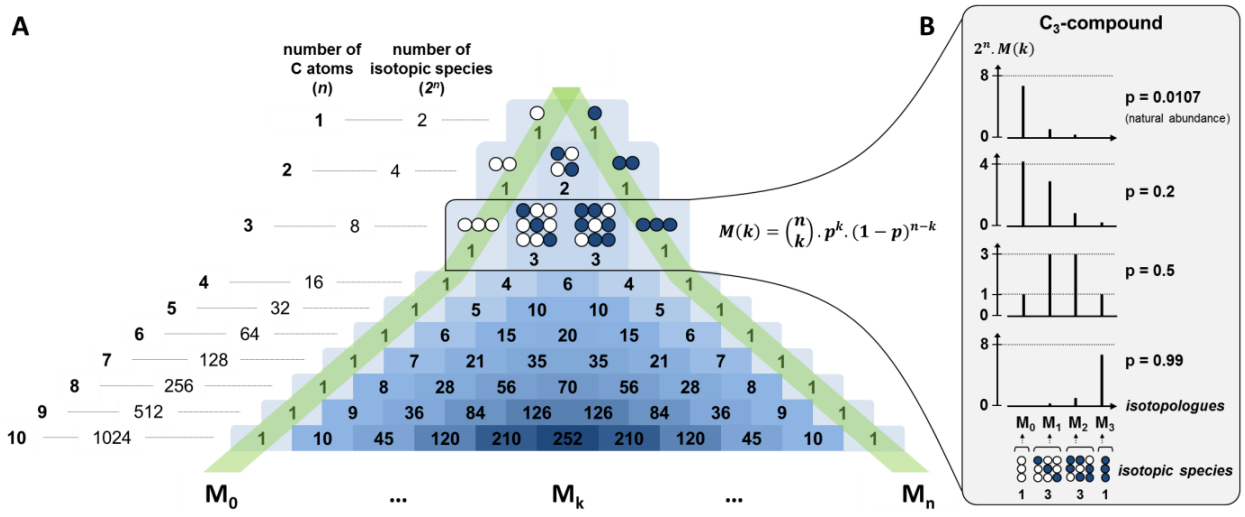


Figure 1. (A) Representation of the Pascal's triangle containing the binomial coefficient used to predict CIDs as function of the abundance of ^{13}C (p). The number of isotopic species for each isotopologue is detailed for compounds having from 1 to 3 carbon atoms. (B) CIDs of a C_3-compound at natural abundance for different p varied from natural abundance (p=0.0107) to p=0.99. When the distribution of ^{13}C is random with p=0.5, the CID weighted by the total number of isotopic species of the molecule – *i.e.* 2^n for a molecule having *n* atoms of carbon – gives the *n*th line of the Pascal's triangle. White and blue circles represent ^{12}C and ^{13}C atoms, respectively.

Distribution of isotopologue abundances

If the number and nature of possible isotopic forms of a molecule depends only on its elemental formula, the abundance of each isotopic form depends uniquely on its

synthetic origin (i.e. isotopic composition of synthetic precursors, synthesis reaction mechanism). The isotopologue abundances can be extracted from the quantitative analysis of the isotopic cluster, i.e. from the intensities of cluster peaks. Indeed, the data collected by MS represent directly the isotopologue abundances. In mass spectrometry, the intensity of the isotopic cluster peaks are commonly expressed relative to the intensity of the monoisotopic peak, itself arbitrarily set at 100%. In ^{13}C-labelling experiments, they are preferably expressed relative to the total intensity of the cluster, so that they represent the relative contribution of each peak to the total cluster. The relative intensity m(k) of the *k*th peak of the isotopic cluster is given by the following equation:

$$m(k) = \frac{I_k}{\sum_{i=0}^{n} I_i} \qquad \text{eq (1)}$$

Where *i* ranges the nominal masses from the lightest (m_0) to the heaviest (m_n), and I is their respective intensity.

Prediction of isotopologue abundances in tracer experiments

The isotopologue abundances of a molecule can be predicted provided i) the abundance of all isotopes is known and ii) no isotopic effects interfere with the distribution of the isotope into (the various positions of) the molecules. In isotopic-labeling experiments, the abundance of the tracer isotope is artificially increased meanwhile all other isotopes occur at natural abundances. In this situation, the isotopologue abundances can be predicted provided the abundance of the tracer isotope is known and assuming the distribution of the other isotopes in the molecule is not modified by the presence of the tracer. In this work, the concept of isotopologue predictability is further developed for the particular case of carbon-labeling experiments, but is generally valid for any isotopic tracer. Because the ^{14}C nucleus can be neglected, carbon is representative of the general case of an element for which the distribution of two isotopes (^{12}C and ^{13}C) has to be considered. In that case, the relative contribution of the two isotopes follows a binomial distribution. The distribution of carbon isotopologues, or Carbon Isotopologue Distribution (CID), is the vector of isotopologue abundances CID(M)= [M_0, M_1,...M_k, …, M_n] of a metabolite M, where the relative distribution M(k) of the carbon isotopologue having incorporated *k* ^{13}C atoms is given by the following equation:

$$M(k) = \binom{n}{k}.p^k.(1-p)^{n-k} \qquad \text{eq (2)}$$

where *n* is the total number of carbon atoms in the molecular entity, and *p* is the abundance of the ^{13}C isotope. As discussed previously, $\binom{n}{k}$ are the binomial coefficients. The term $p^k.(1-p)^{n-k}$ is the proportion of each isotopic form composing the isotopologue. When the values of *n* and *p* are known, the CIDs can be calculated from equation 2.

Biosynthesis of ^{13}C-molecules with predictable isotopologue abundances

The experimental strategy for the production of metabolites with fully predictable isotopic patterns is presented figure 2.

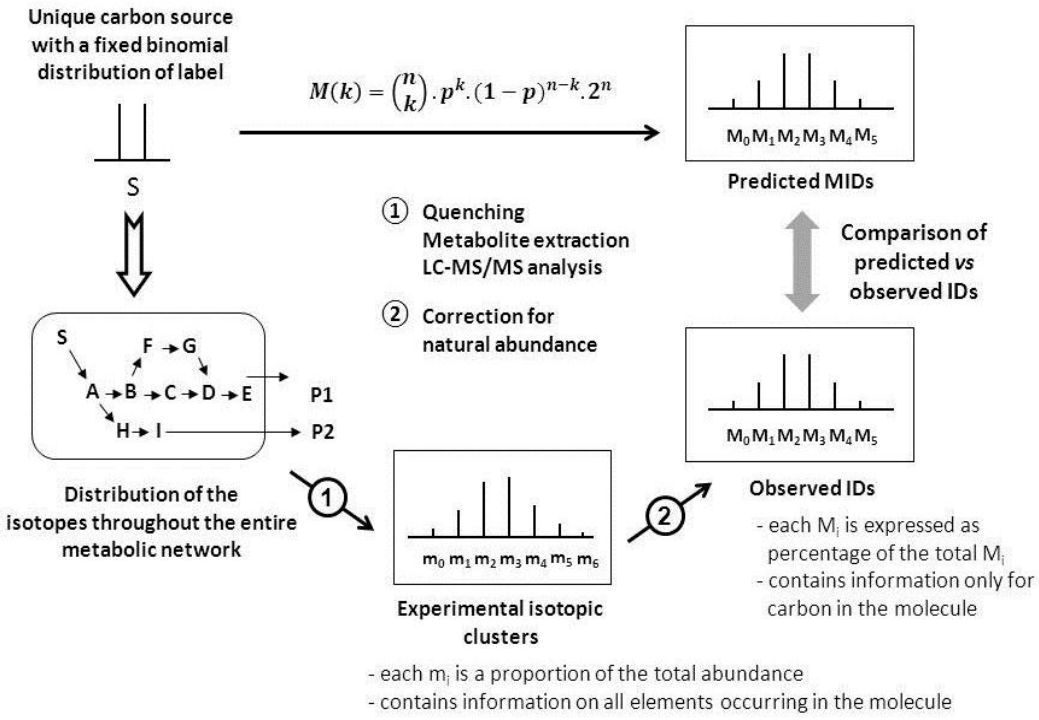


Figure 2. Experimental strategy applied to evaluate the accuracy and the precision of the whole procedure. Metabolites with predictable IDs were synthesized by growing *P. augusta* and *E. coli* on carbon sources with a fixed random distribution of ^{13}C isotopes. Intracellular metabolites were sampled and analyzed by LC-MS/MS (1), processed (2) and finally their IDs were compared to the ones predicted from the distribution of isotopes into the ^{13}C-labeled carbon source.

The basic principle is to design the labeling experiments in such a way that the isotopic patterns become fully independent of the metabolic network (biochemical reactions) and its operation (reaction rates), but depend uniquely on the label input. Assuming no isotopic effect interfere with the propagation of the label in the biochemical network, the

labelling patterns of metabolites become independent of the metabolic network provided the isotopic composition of the label input complies with two basic rules:

i) all the isotopologues composing the label input must be mixed in proportions according to the binomial law, and

ii) all the isotopic forms composing a given isotopologue of the label input must be provided in the same proportions.

In such conditions, the CIDs of all metabolites in the considered metabolic system are determined uniquely by the isotopic composition of the label but no longer by the metabolic network. Moreover, the values of metabolite CIDs can be controlled and modulated according to needs by controlling the isotopic composition of the label input.
The best standard for the evaluation of analytical procedures aiming at measuring the isotopic content of molecules should ideally contain all isotopic species of a particular metabolite in sufficient amounts, in order to evaluate the measurement reliability over the entire isotopic cluster and to check for any bias in the measurement of each individual isotopic form. If labelled metabolites are produced biologically at *p* values tending to extreme values; i.e. 0 or 1, the CID of all metabolites are imbalanced in favour of the lightest or the heaviest isotopologues, respectively. The best situation is obtained when p=0.5. In such case, $p^k.(1-p)^{n-k}=0.5^n$ whatever the value of *k*. This means that each isotopic form of the label input must be provided in the same proportion. Equation 2 is reduced to M(k) = $\binom{n}{k}$ after multiplication by the total number of isotopic species (i.e. 2^n), and experimental CIDs should match exactly the binomial coefficients found in the relevant row of the Pascal's triangle (Figure 1). This means that for p=0.5 all isotopic forms of a particular metabolite are obtained in the same proportions in a unique sample, which represents a valuable standard mixture for the evaluation of MS instruments over the entire range of isotopologues.

Production of intracellular metabolites with controlled labelling patterns

The above strategy was applied to two organisms, i.e. the eukaryotic methylotroph *Pichia augusta* and the multi-carbon utilizer prokaryote *Escherichia coli*. Both organisms are extensively used for biotechnological purposes and can be grown at high cellular densities under tightly controlled conditions. Hence, they represent highly valuable

producers of biologically-labelled standards, even at large-scale. They have also different metabolisms – and hence metabolomes (metabolite contents) –, so that combining labelled standards from both organisms can extend the coverage of the metabolome that can be investigated.

We first showed that CID predictability can be applied to intracellular metabolites and to LC-MS/MS analysis. The metabolome encompasses a broad range of species with highly different chemical properties. Currently, LC-MS represents the most appropriate analytical tool for analyzing such chemical diversity since the separation power of LC is much higher than that of GC. Therefore much broader coverage of the metabolome can be achieved by LC-MS. A major difficulty of LC-MS is the strong matrix effects that have significant impact on the detection and quantification of chemical species. Moreover these effects are compound dependent and isotopologue-dependent, hence must be assessed for each individual isotopologue and each analyte of interest.

To demonstrate that intracellular metabolites with predictable CIDs could be produced, the methylotroph organism *Pichia augusta* was grown under steady-state conditions on an equimolar mixture of ^{12}C-methanol and ^{13}C-methanol, and the CIDs of intracellular metabolites were analyzed by LC-MS/MS. During the ^{13}C-labelling experiment, a critical issue is the undesired contribution of carbon sources other than the label input. To avoid such problems, *P. augusta* was grown on mineral medium and cultivations were oxygenated with CO_2-depleted synthetic air. The actual ^{13}C content of the label input was measured by NMR to be 50.4% (Table 1).

Carbon source	Isotopic composition (%)				^{13}C-enrichment[a] (%)
	^{12}C	U-^{13}C	1-^{13}C	2-^{13}C	
Methanol	49.6 ± 0.5	50.4 ± 0.5	-	-	50.4 ± 0.5
Acetate	24.7 ± 0.5	24.9 ± 0.5	25.7 ± 0.5	24.7 ± 0.5	50.2 ± 0.7

a: percentage of ^{13}C atoms incorporated in the molecules

Table 1. Isotopic composition of ^{13}C-labeled mixtures of methanol and acetate used as unique carbon sources by *P. augusta* and *E. coli*, respectively. The ^{13}C-enrichment refers to the proportion of ^{13}C isotopes relative to the total amount of carbon atoms into the molecules.

A second experimental issue is the sampling of intracellular metabolites. Compared to metabolic end-products, intracellular metabolites have much smaller pool sizes and much higher turn-over rates. Fast and reliable sampling procedures are required to avoid any

alteration in the content and isotopic patterns of these metabolites. In this work, *P. augusta* metabolites were sampled using the fast-filtration method developed by (Bolten, Kiefer et al. 2007). The CIDs of intracellular metabolites were measured by the highly sensitive LC-MS/MS method developed by (Kiefer, Nicolas et al. 2007), using appropriate correction of the natural abundances of isotopes other than carbon (Millard, Letisse et al. 2012). For demonstration purpose, the method was applied to metabolites belonging to the most classical metabolic pathways: glycolysis (glucose-6P, fructose-6P, phosphoenol pyruvate, 2- and 3-phosphoglycerate), pentose-phosphate pathway (6-phosphogluconate, ribose-5P), nucleotides (AMP, ADP) and TCA cycle (succinate, fumarate, citrate). These compounds cover a broad range of chemical species, and their carbon skeletons range from 3 (PEP, combined pools of 2- and 3-phosphoglycerate) to 10 (AMP, ADP), which represent a valuable situation for evaluating the quality of CID data, as shown further.

Values and quality of predicted CIDs

According to equation 2, the expected distribution of CIDs in the above experiment where p=0.5 is (1:3:3:1) for C3 compounds and (1:10:45:120:210:252:210:120:45:10:1) for C10 compounds. The dynamic range of CIDs, which corresponds to the ratio between the most abundant and the less abundant isotopologue, is ranging from 3 to 252 according to the carbon skeleton. This dynamic range spans over 2.5 decades, which is lower than the dynamic range of current mass spectrometers (3-5). Consequently, all isotopologues from the various metabolites that have been analyzed can be properly detected and quantified.

Predicted CIDs were calculated according to the actual ^{13}C-enrichment of the labeled methanol input, as measured by NMR (50.4±0.5%, Table 1). Because this value is slightly higher than the theoretical 50%, predicted CIDs were slightly drifted toward the heaviest isotopologues, e.g. (0.98:2.98:3.02:1.02) instead of (1:3:3:1) for C3 compounds. In addition, standard deviations of predicted CIDs were calculated to account for errors in the preparation of the label input given by the NMR measurements of the label input mixture (0.5%). Predicted CIDs showed standard deviations (~0.2%) much lower than experimental CIDs (~1%), meaning that the former offered valuable reference data to evaluate the accuracy and precision of the latter.

Values and quality of experimentally measured CIDs

Intracellular metabolites accumulate in amounts much smaller than metabolic end-products. Hence highly sensitive analytical techniques have to be used to collect CIDs on such compounds. Here, rather than single MS measurements, CIDs were measured using MS/MS measurements, – multiple reaction monitoring (MRM) mode – in which parent ions are fragmented into daughter ions by collision-activated dissociation (CAD). This approach provides highly specific and highly sensitive measurements of labeling data (Kiefer, Nicolas et al. 2007; Ruhl, Rupp et al. 2012). For some compounds, the fragmentation process results in the breakage of carbon-carbon bounds. It has been hypothesized that the efficiency of fragmentation could depend on the isotopic content of the carbon positions that are disrupted, leading to biases or uncertainties in CID measurements (Shen, Shen et al. 2009). However, the CIDs measured for the most frequent fragments – i.e. C1 fragments due to the, release of carboxylic groups in the considered analytical method (Kiefer, Nicolas et al. 2007) – closely matched the expected values (Figure 3), indicating that no bias resulted from the application of the MRM method. Similar conclusion could be drawn from the analysis of C3-fragments (Figure 3). Hence, highly sensitive and reliable CID measurement of intracellular metabolites can be obtained by LC-MS/MS.

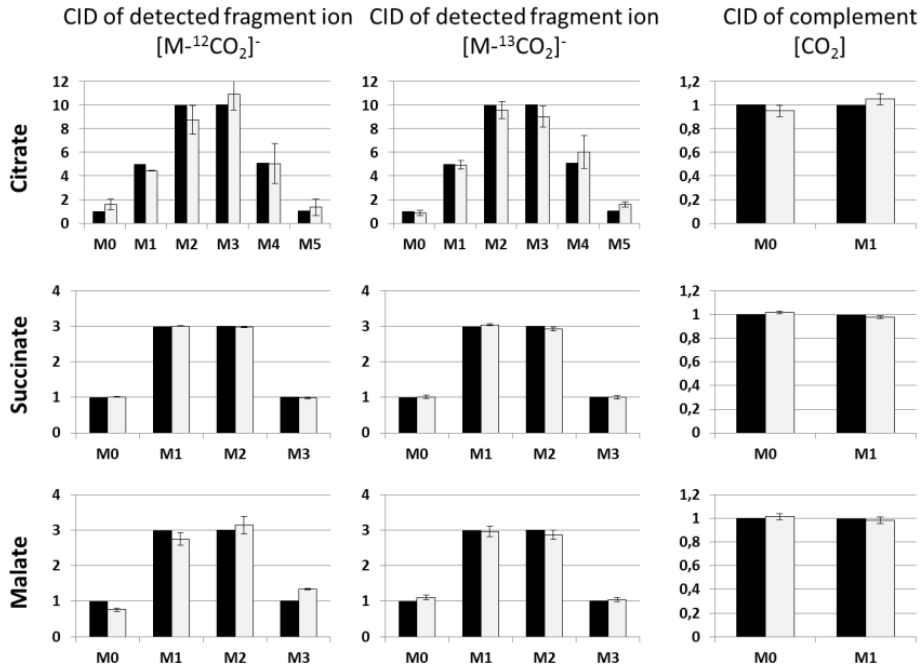


Figure 3. Reliability of MS/MS measurements.

The CIDs measured for *P. augusta* metabolites are given in Annex and are represented in Figure 4. The standard deviations of measured CIDs were calculated for each individual isotopologue from the analysis of 3 independent biological samples. The average standard deviation was 0.9%, indicating a very good precision of MS/MS measurements.

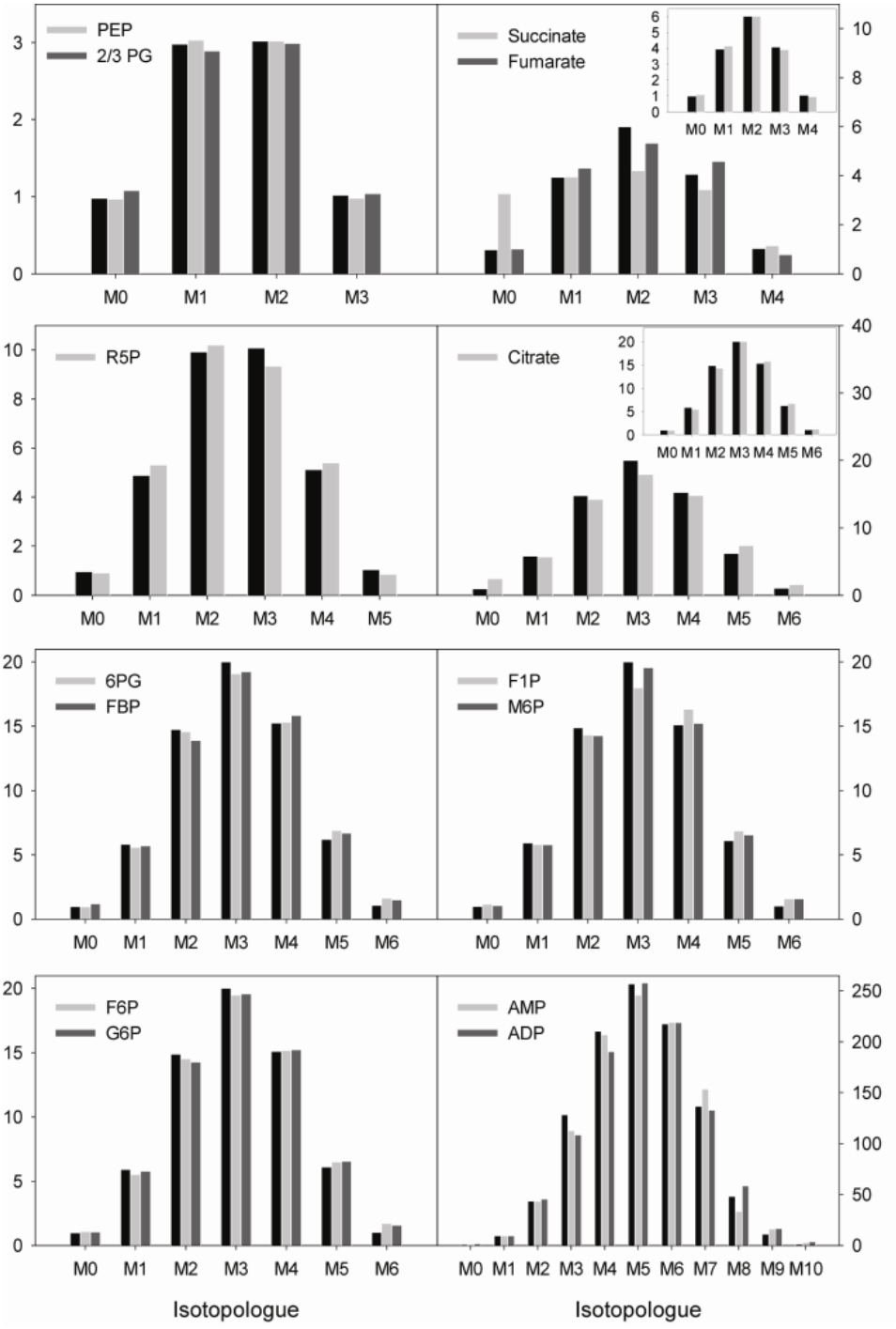


Figure 4. Comparison of predicted (black bars) and observed CIDs of intracellular metabolites collected from *P. augusta* cells grown on an equimolar mixture of ^{13}C- and ^{12}C-methanol. Insets show, where appropriate, the data corrected for the occurrence of unlabelled contaminants

Comparison of experimental vs predicted CIDs

To validate the strategy proposed for producing metabolites with fully controlled labeling patterns, the experimental CIDs were compared to the predicted CIDs. The CID closeness, which was defined as the average of differences between the experimental CIDs and their predicted value, was 0.8% for *P. augusta* metabolites, showing a very good consistency between experimental values and predictions. The plot of the experimental *vs.* predicted CIDs for each individual isotopologue (Figure 5) followed a linear regression with a correlation coefficient r^2 equal to 0.993, indicating that the experimental values closely matched the predicted values.

Metabolites with a high number of carbon atoms are expected to be more sensitive to errors than compounds with a small number of carbon atoms. The data showed that CID closeness for compounds with 10 carbon atoms (ATP, ADP) was similar to that obtained for C3 compounds. This result is of particular value since it indicates that all the 11 isotopologues of these compounds can be produced at the same with accurately predicted proportions.

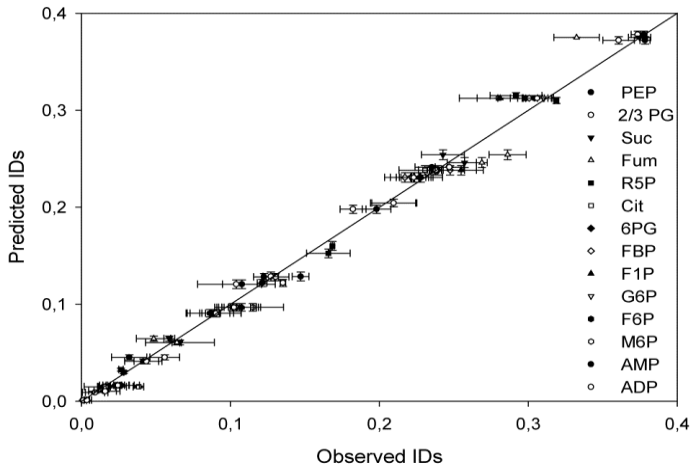


Figure 5. Correlation between observed and predicted CIDs of intracellular metabolites from *P. augusta*.

The experimental CIDs did not fit with the predicted CIDs for only two compounds, *i.e.* succinate and citrate. For both compounds, the experimental data showed an overestimation of the M_0 peak. This result is explained by a contamination with unlabelled compounds. Such contamination is commonly observed in LC-MS/MS and was confirmed from the detection of unlabeled succinate and citrate in blank samples (data not shown). These results indicate that the labeled samples can be also used to reveal the occurrence of contaminations from unlabeled pools in ^{13}C-labeling experiments.

Interestingly, the experimental CIDs of succinate and citrate could be corrected from the contribution of the contaminating, unlabeled pools (estimated to be 16.6% and 3.3% of total pools, respectively), and the corrected data matched closely with the predicted values.

The proposed strategy for the production of metabolites with controlled labeling patterns is valid only if no isotopic effects interfere with the propagation of the label within the considered reaction network. Isotopic fractionation associated to enzymatic activities has been reported in literature, but such effects are considered to be negligible in ^{13}C-labelling experiments (Wiechert 2001; van Winden, van Dam et al. 2005; Shen, Shen et al. 2009). The data presented here, where the experimental CIDs closely match the predicted values, are consistent with this expectation. They indicate that the isotopic effects, if any, are within the precision of CID measurements. In addition, the distribution of differences between predicted and experimental CIDs was centered on zero for all individual CIDs (Figure 6), indicating that no systematic biases – related or not to isotopic effects - occurred during the entire experimental workflow.

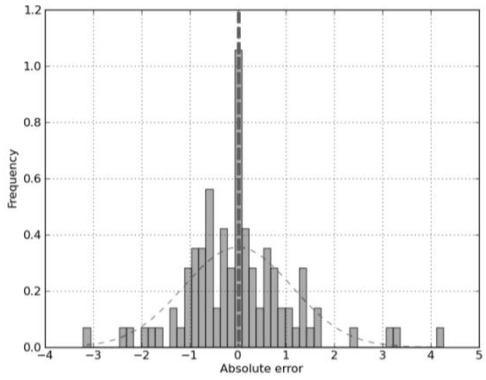


Figure 6. Distribution of differences between measured and predicted CIDs from intracellular metabolites from *P. augusta*. Gaussian curve has been fitted on this distribution (dotted curve). The vertical dotted line corresponds to the maximum of the Gaussian curves.

Extension to multicarbon label inputs

The strategy for the biological production of samples with controlled labeling patterns was further extended to multicarbon label inputs, in order to extend the range of metabolisms and organisms which can be investigated or used to produce labeled standards. A basic difficulty is the (commercial) availability of all required isotopic forms of the label input. This difficulty dramatically increases with the number of carbon atoms

in the considered substrate. For example, 8 different isotopic forms must be available to use a C3 compound as label input, and 16 forms for a C4 compounds. Currently, the complete set of all required isotopic forms is commercially available only for C2 compounds. Hence we validated the strategy by using a C2 compound - acetate - as multicarbon label input. This allowed the utilization of a non-methylotroph organism such as *E. coli* as producer. As for *P. augusta*, the *E. coli* isotopologue space was investigated for the particular condition where the ^{13}C-abundance in the label input is p=0.5. As shown in figure 7, that condition corresponds to using as label input an equimolar mixture of the 4 isotopic forms (^{12}C,1-^{13}C,2-^{13}C, U-^{13}C) of acetate.

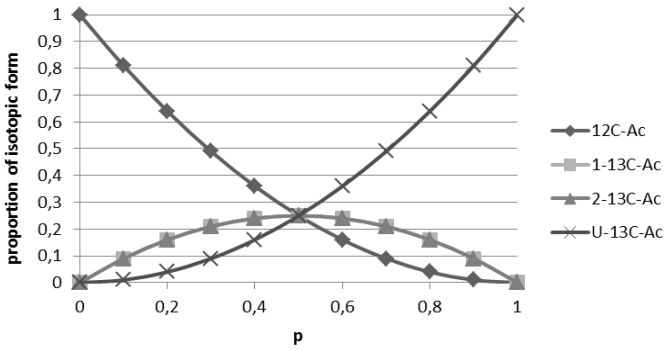


Figure 7. Proportions of each isotopic forms of acetate to be mixed in the label input for a given ^{13}C-abundance *p* with a random distribution of isotopes.

From the practical point of view, the error in the preparation of the label input mixture increases with the number of isotopic species to mix. Hence special care was taken to mix the 4 isotopic forms of acetate as close as possible of the required (1:1:1:1) distribution, and the actual isotopic content of the labeled acetate input mixture was checked by NMR to be close to the target values (Table 1).

E. coli cells were grown with the labeled acetate mixture. As with *P. augusta,* cultivations were inoculated at very low cellular density and aerated with CO_2-depleted synthetic air to avoid the contribution of undesired unlabelled carbon source. Intracellular metabolites were sampled at isotopic steady-state using cold ethanol quenching and hot ethanol extraction, and analyzed by IC-MS/MS. As expected from the different metabolisms and growth conditions, the metabolic profiles of *E. coli* cells differed significantly, both qualitatively and quantitatively, from that of *P. augusta*. Accordingly, the list of *E. coli* metabolites for which CIDs were measured was slightly different from that in *P. augusta*, but still covered central metabolic pathways: glycolysis (glucose-6P, fructose-6P,

mannose-6P, phosphoenol pyruvate, 2- and 3-phosphoglycerate), pentose-phosphate pathway (sedoheptulose-7P, ribose-5P), TCA cycle (succinate, malate, citrate), nucleotides (ADP, CDP). This list includes compounds with a number of carbon atoms ranging from 3 to 10. The average standard deviation of experimental CIDs collected for *E. coli* metabolites was 0.5%, again indicating a very high quality of the MS/MS measurements.

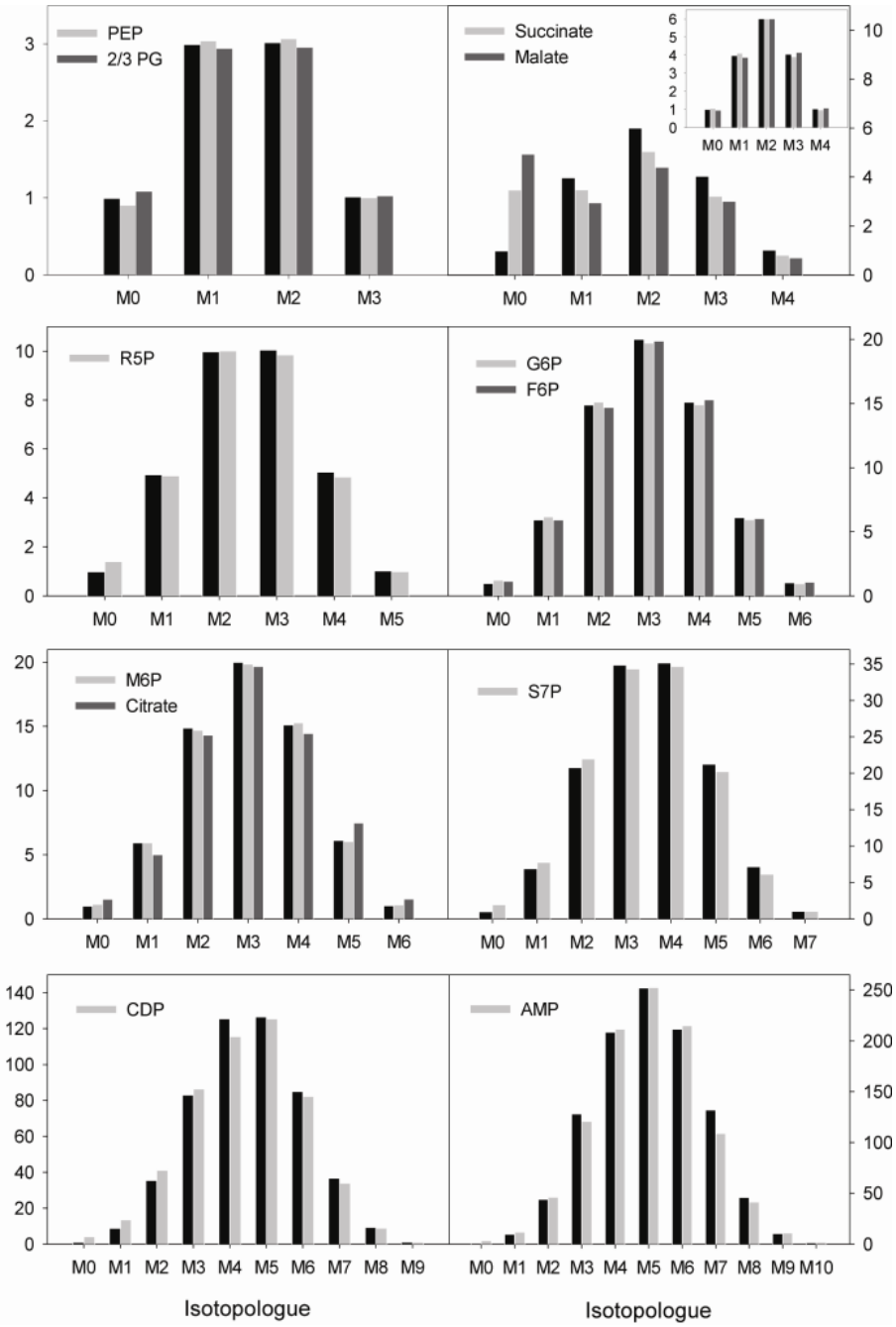


Figure 8. Comparison of predicted (black bars) and observed CIDs of intracellular metabolites collected from *E. coli*.cells grown on an equimolar mixture of all isotopic forms of acetate. Insets show, where appropriate, the data corrected for the occurrence of unlabelled contaminants

The experimental CIDs measured for *E. coli* metabolites were compared to the predicted CIDs to evaluate the applicability of the strategy to multi-carbon label inputs. The values and standard deviations of CIDs predicted for *E. coli* metabolites were calculated according to the actual composition of the acetate label input given by NMR (Table 1). As obtained with *P. augusta*, the experimental CIDs collected on *E. coli* metabolites matched closely the predicted values over the entire range of carbon skeletons (figure 8). A contamination with unlabelled compounds was observed for two organic acids, namely malate and succinate. The contamination levels could be calculated to represent 16.9% and 27.7% for succinate and malate, respectively, and the CIDs of the two metabolites matched the predicted values after correction for the contamination.

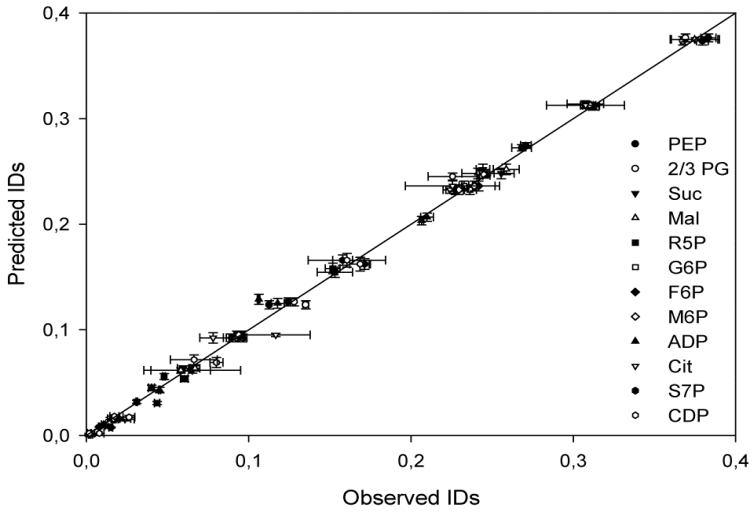


Figure 9. Correlation between observed and predicted CIDs of intracellular metabolites from *E. coli*.

The plot of experimental vs predicted isotopologue abundances for *E. coli* metabolites is shown in Figure 9. The plot regressed linearly with a correlation coefficient r^2 equal to 0.997, indicating a very good correlation between experimental and predicted CIDs. The average CID closeness was calculated to be 0.5% for *E. coli* metabolites, which is lower than that observed for *P. augusta* (0.8%). As observed with the latter organism, the distribution of differences between predicted and experimental CIDs was centered on zero (Figure 10) indicating that no systematic biases occurred during the entire experimental workflow. All these data indicate that the CIDs obtained for *E. coli* are closely consistent to predictions, validating the application of the proposed strategy to multicarbon substrates.

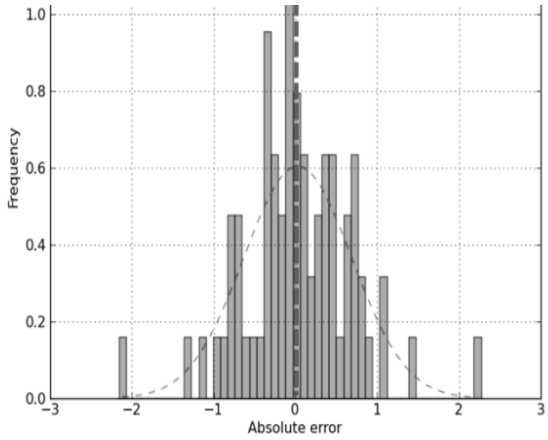


Figure 10. Distribution of differences between measured and predicted CIDs from intracellular metabolites from *E. coli*. Gaussian curve has been fitted on this distribution. The vertical dotted line corresponds to the maximum of the Gaussian curves.

Determination of ^{13}C-enrichments

In the proposed strategy to obtain metabolites with predictable CIDs, their labelling patterns should depend exclusively on the label input. If so, the ^{13}C-enrichment (*p)* of metabolites should be identical to that of the labeled substrate, and identical among all compounds. To further evaluate the strategy, the ^{13}C-enrichments of metabolites were calculated and compared to that of the label inputs for the two organisms investigated. The ^{13}C-enrichments of metabolites were calculated from experimental CIDs using equation (2) and assuming fully random distribution of ^{13}C atoms into the metabolite. The average *p* value for metabolites extracted from *P. augusta* and from *E. coli* were 50.7±0.7% and 50.0±0.6%, respectively. These values were in closed agreement with the enrichments of the ^{13}C-labeled substrates measured by NMR (50.4±0.5% for labelled methanol and 50.2±0.7% for labelled acetate), and included in the gray zone displaying the NMR experimental error (figure 11).

The precisions on metabolite *p* values were calculated from the standard deviations of experimental CIDs. Considering that CIDs were measured with precisions of 0.5% and 1% for *E. coli* and *P. augusta*, 95% confidence intervals on *p* are about 1.5% and 2.25% respectively. The results are displayed in figure 11. Error bar for every metabolite covers the mean value of *p* for *P. augusta* dataset. For *E. coli* dataset, we also observed that the individual values of *p* match well with its mean value for almost all metabolites. As expected, the ^{13}C-enrichments calculated for the two contaminated species - citrate and malate – showed errors higher than the other compounds, but were within the expected

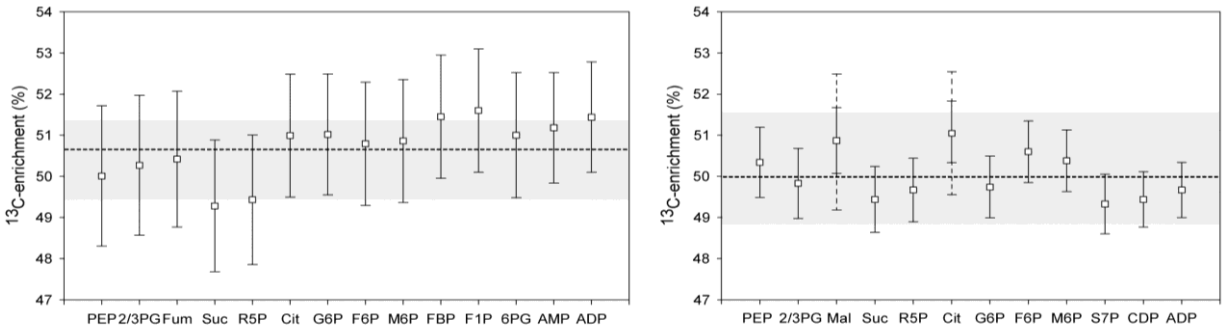


Figure 11. ^{13}C-enrichments of intracellular metabolites *of P. augusta* (left panel) and *E. coli* (right panel) calculated from the CIDs measured by LC-MS/MS. The dotted lines represent the average ^{13}C-enrichments determined from CIDs measured for each organism. The grey area represents the 95 % confidence interval of the enrichment determined by NMR analysis of their respective substrates. 95 % confidence intervals on the enrichments were estimated with Monte-Carlo analyses from 1000 artificially noised CIDs assigning a precision of 1 % or 0.5 % on the observed proportion of each isotopologue for *P. augusta* and *E. coli* respectively. Dotted error bars of malate and citrate represents 95 % confidence intervals estimated assigning a precision of 1 % on their CIDs.

values when a precision of 1% was applied. These results indicated that ^{13}C-enrichments of intracellular metabolites can be statistically considered as identical assigning a precision within 1% on isotopologue measurements, and below this limit for most metabolites.

Effects of correction for naturally occurring isotopes

The calculation of CIDs from isotopic clusters integrates a correction step to subtract the contribution of isotopes other than the tracer. The natural abundances of isotopes are subjected to variations or uncertainties that may introduce biases during the correction step in both CIDs and ^{13}C-enrichments. The impact of these uncertainties is expected to be more significant for molecules containing a high number of atoms. Therefore, we calculated the variations in CIDs and *p* for ADP ($C_{10}H_{15}N_5O_{10}P_2$) generated by variations in the isotopic abundance of ^{2}H, ^{18}O and ^{15}N (Garcia Alonso, Rodriguez-Gonzalez et al. 2010). We observed that the variations in CID and in *p* are below 0.03 % and 0.05 % respectively (Figure 12). These values are much lower than the experimental errors measured in this work. It can be therefore concluded that the uncertainties in the natural abundances did not significantly impact CIDs and ^{13}C-enrichments and can be thus neglected.

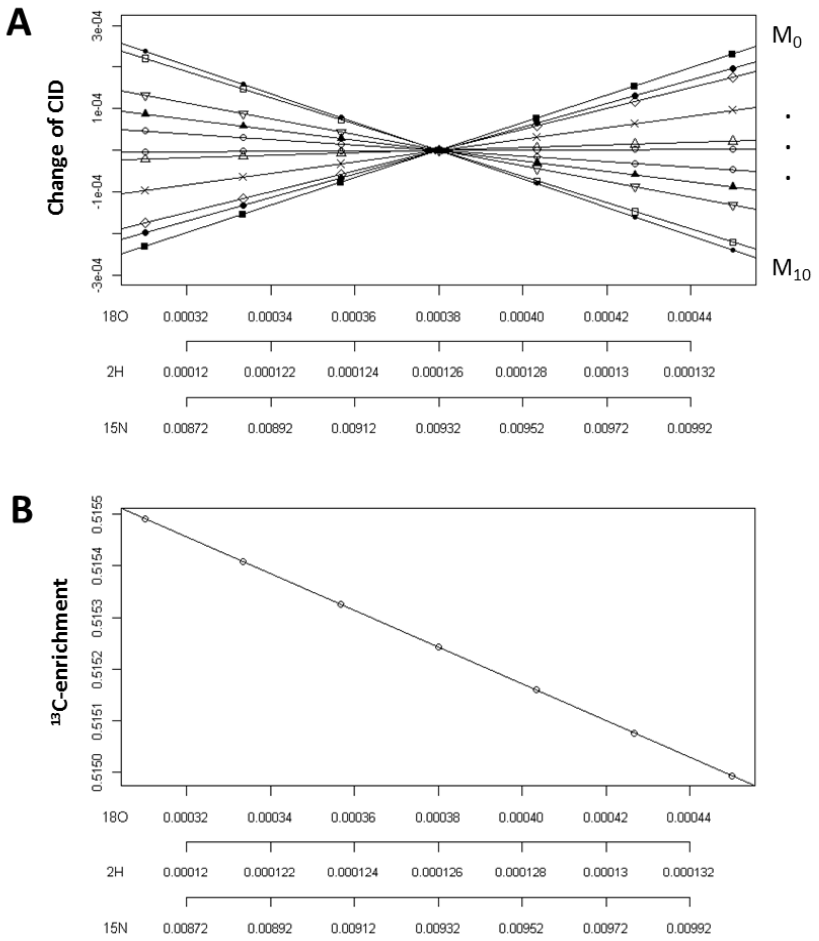


Figure 12. Changes in CID of ADP (A) when the proportions at natural abundance of ^{2}H, ^{15}N and ^{18}O are varied in their specific ranges, and propagation of these uncertainties on the ^{13}C-enrichment (B).

5. Conclusion

The data reported here show that the complete theoretical and experimental framework for the biological production of metabolites with controlled labelling patterns proposed in this work can be successfully applied to produce ^{13}C-labelled metabolites with isotopologue abundances matching closely the predicted values. Samples in which all isotopologues of each metabolite could be produced with a high accuracy and precision in the isotopologue abundances. The approach was validated for different organisms, carbon sources, and can be applied to the production of a broad range of metabolites (intracellular metabolites, metabolic end-products, etc), in order to cover the entire metabolome or to evaluate different analytical platforms. Such biologically-produced labeled metabolites can be used for different purposes, including evaluation or

calibration of MS instruments for isotopic measurements, evaluation of analytical methods (e.g. different acquisition modes in mass spectrometry) or of complete analytical strategies (from sampling to data processing). It can be also used to provide evidence of isotopic effects, provided these effects are larger than the precision of both the considered MS instrument and metabolite CIDs. In addition, the isotopic composition of the labeled metabolites in the standard can be modulated according to specific purposes. Moreover, the strategy was applied here to the production of ^{13}C-labeled compounds, but it is generally valid for any isotope that can be potentially incorporated in living matter.

6. References

Antoniewicz, M. R., J. K. Kelleher, et al. (2006). "Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements." Metab Eng **8**(4): 324-337.

Antoniewicz, M. R., J. K. Kelleher, et al. (2007). "Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis." Anal Chem **79**(19): 7554-7559.

Bajad, S. U., W. Lu, et al. (2006). "Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry." J Chromatogr A **1125**(1): 76-88.

Bolten, C. J., P. Kiefer, et al. (2007). "Sampling for metabolome analysis of microorganisms." Anal Chem **79**(10): 3843-3849.

Dauner, M. and U. Sauer (2000). "GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing." Biotechnol Prog **16**(4): 642-649.

Garcia Alonso, J. I., P. Rodriguez-Gonzalez, et al. (2010). "Determination of the uncertainties in the theoretical mass isotopomer distribution of molecules." Anal Chim Acta **664**(1): 68-76.

Hellerstein, M. K. and R. A. Neese (1999). "Mass isotopomer distribution analysis at eight years: theoretical, analytic, and experimental considerations." Am J Physiol **276**(6 Pt 1): E1146-1170.

IUPAC (1997). "Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: http://goldbook.iupac.org (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins.".

Kiefer, P., C. Nicolas, et al. (2007). "Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry." Anal Biochem **360**(2): 182-188.

Mashego, M. R., L. Wu, et al. (2004). "MIRACLE: mass isotopomer ratio analysis of U-13C-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites." Biotechnol Bioeng **85**(6): 620-628.

Millard, P., F. Letisse, et al. (2012). "IsoCor: Correcting MS data in isotope labeling experiments." Bioinformatics.

Peyraud, R., P. Kiefer, et al. (2009). "Demonstration of the ethylmalonyl-CoA pathway by using 13C metabolomics." Proc Natl Acad Sci U S A **106**(12): 4846-4851.

Quek, L. E., C. Wittmann, et al. (2009). "OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis." Microb Cell Fact **8**: 25.

Rosman, K. J. R. and P. D. P. Taylor (1998). "Isotopic compositions of the elements 1997." Pure Appl Chem **70**(1): 217-235.

Ruhl, M., B. Rupp, et al. (2012). "Collisional fragmentation of central carbon metabolites in LC-MS/MS increases precision of (1)(3)C metabolic flux analysis." Biotechnol Bioeng **109**(3): 763-771.

Shen, T., W. Shen, et al. (2009). "Increasing the accuracy of mass isotopomer analysis through calibration curves constructed using biologically synthesized compounds." J Mass Spectrom **44**(7): 1066-1080.

Taymaz-Nikerel, H., M. de Mey, et al. (2009). "Development and application of a differential method for reliable metabolome analysis in Escherichia coli." Anal Biochem **386**(1): 9-19.

van Winden, W. A., J. C. van Dam, et al. (2005). "Metabolic-flux analysis of Saccharomyces cerevisiae CEN.PK113-7D based on mass isotopomer measurements of (13)C-labeled primary metabolites." FEMS Yeast Res **5**(6-7): 559-568.

Vogt, J. A., U. Wachter, et al. (2003). "Non-linearity in the quadrupole detector system: implications for the determination of the 13C mass distribution of an ion fragment." J Mass Spectrom **38**(2): 222-230.

Wiechert, W. (2001). "13C metabolic flux analysis." Metab Eng **3**(3): 195-206.

Wiechert, W. (2007). "The thermodynamic meaning of metabolic exchange fluxes." Biophys J **93**(6): 2255-2264.

Wiechert, W., M. Mollney, et al. (2001). "A universal framework for 13C metabolic flux analysis." Metab Eng **3**(3): 265-283.

Wittmann, C. (2007). "Fluxome analysis using GC-MS." Microb Cell Fact **6**: 6.

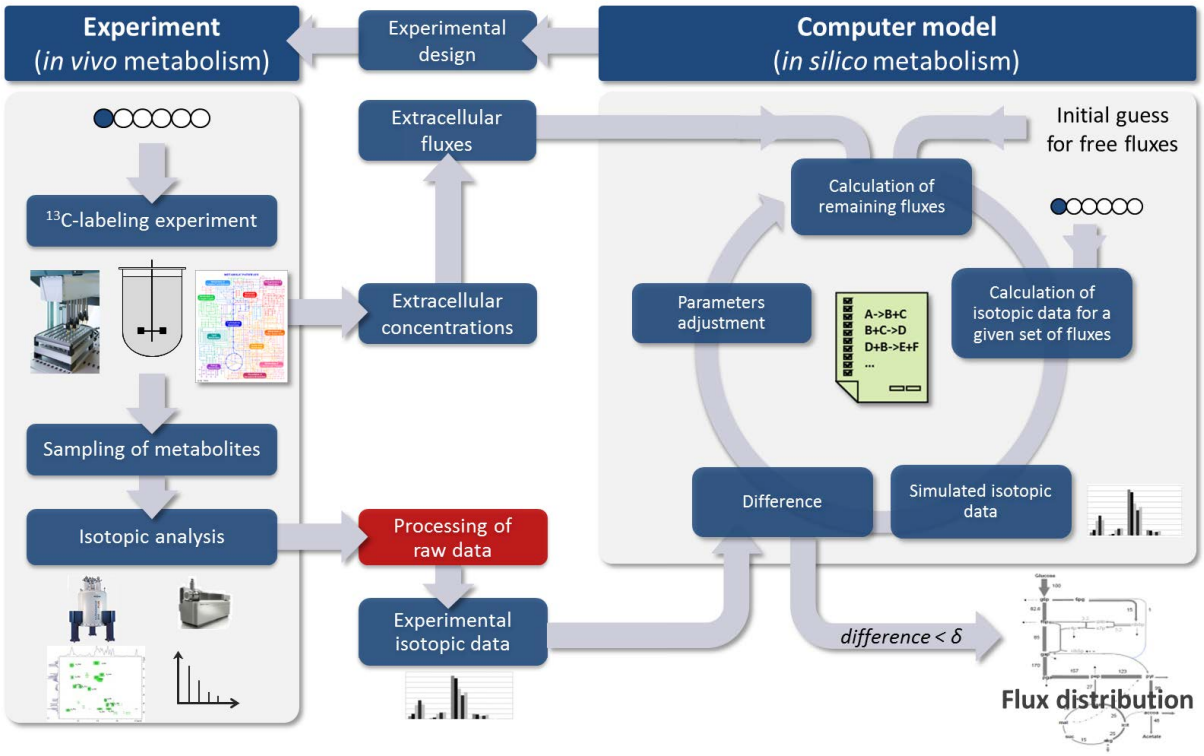
Wu, L., M. R. Mashego, et al. (2005). "Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards." Anal Biochem **336**(2): 164-171.

Zamboni, N., S. M. Fendt, et al. (2009). "(13)C-based metabolic flux analysis." Nat Protoc **4**(6): 878-892.

Zamboni, N. and U. Sauer (2004). "Model-independent fluxome profiling from 2H and 13C experiments for metabolic variant discrimination." Genome Biol **5**(12): R99.

Chapter 5

IsoCor: Correcting MS data in isotope labeling experiments



NOTE: *this chapter contains a manuscript published in Bioinformatics:*

Millard P., Letisse F., Sokol S., Portais J.C. (2012). IsoCor: Correcting MS data in isotope labeling experiments. *Bioinformatics,* 28(9):1294-1296.

1. Abstract

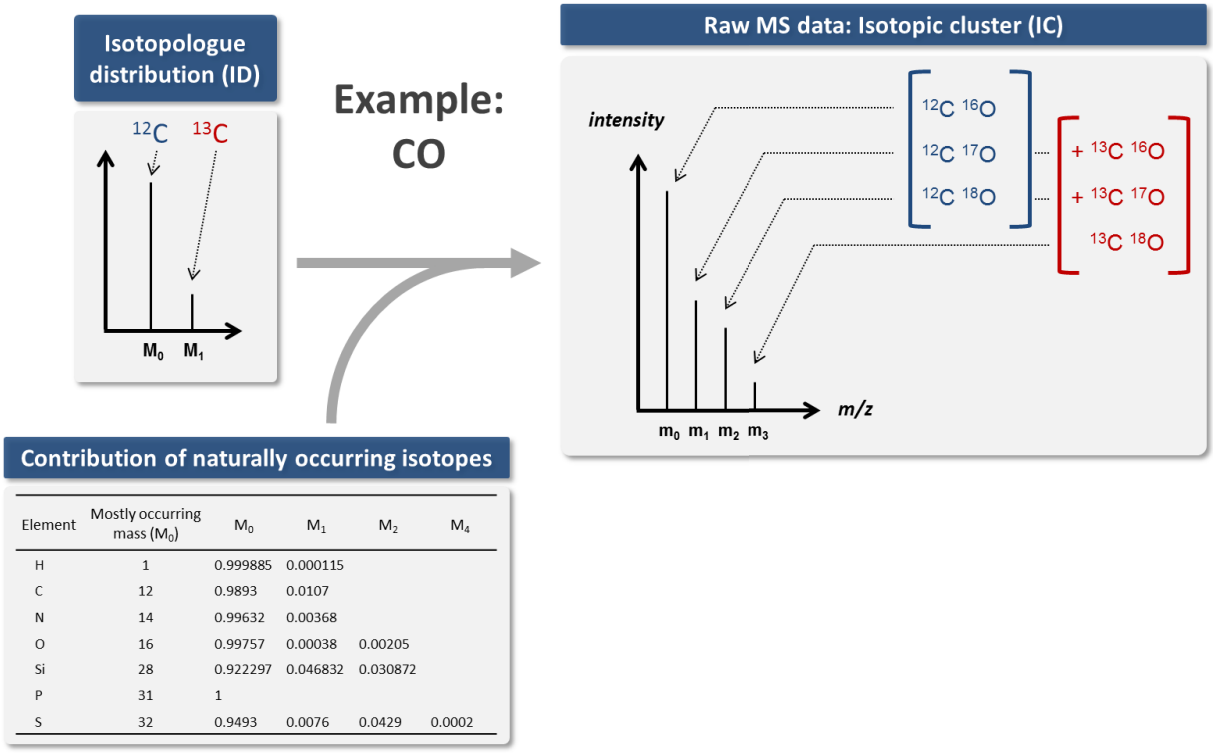
Mass spectrometry (MS) is widely used for isotopic labeling studies of metabolism and other biological processes. Quantitative applications – e.g. metabolic flux analysis – require tools to correct the raw MS data for the contribution of all naturally abundant isotopes. IsoCor is a software that allows such correction to be applied to any chemical species. Hence it can be used to exploit any isotopic tracer, from well-known (^{13}C, ^{15}N, ^{18}O, etc) to unusual (^{57}Fe, ^{77}Se, etc) isotopes. It also provides new features – e.g. correction for the isotopic purity of the tracer – to improve the accuracy of quantitative isotopic studies, and implements an efficient algorithm to process large datasets. Its user-friendly interface makes isotope labeling experiments more accessible to a wider biological community.

Availability: IsoCor is distributed under OpenSource licence at http://metasys.insa-toulouse.fr/software/isocor/

2. Introduction

Mass spectrometry (MS) is extensively used for stable isotopic studies of metabolism. During the last decade, quantitative approaches – such as ^{13}C metabolic flux analysis – have been increasingly used in the fields of systems biology and biotechnology to provide novel biological insights and improve industrial processes (Sauer 2006; Nicolas, Kiefer et al. 2007). In such experiments, the different labeled forms – or isotopologues – of metabolites are quantified by the fine exploitation of isotopic clusters in MS spectra (Kiefer, Nicolas et al. 2007). These clusters also contain information on all other isotopes that occur naturally in the molecules. To extract meaningful labeling information – i.e. isotopologue distribution – the contribution of naturally-occurring isotopes first has to be subtracted (Figure 1). This can be achieved using publicly available software such as MSCorr (Wahl, Dauner et al. 2004) developed under Matlab or the algorithm developed in Perl by (Moseley 2010). However, the former software is only applicable to ^{13}C-

labeling experiments and the latter is restricted to the correction of ultra-high resolution MS data for natural abundance of ^{13}C or ^{15}N.



Element	Mostly occurring mass (M_0)	M_0	M_1	M_2	M_4
H	1	0.999885	0.000115		
C	12	0.9893	0.0107		
N	14	0.99632	0.00368		
O	16	0.99757	0.00038	0.00205	
Si	28	0.922297	0.046832	0.030872	
P	31	1			
S	32	0.9493	0.0076	0.0429	0.0002

Figure 1. Raw MS data (isotopic clusters) contain information on all isotopes that occur naturally in the molecules. Their contribution first has to be subtracted to extract meaningful labeling information (isotopologue distribution).

Here, we present a novel software to correct MS data, IsoCor, which includes the following features:

- Correction for any chemical element, thereby extending the range of isotopic tracers – from well-known (^{13}C, ^{15}N) to unusual (^{57}Fe, ^{77}Se, etc) isotopes – and chemical species that can be investigated. Tracer elements with more than two isotopes (e.g. ^{16}O, ^{17}O, ^{18}O) can be considered.

- When metabolite derivatization is required for analytical purposes (e.g. derivatization of amino acids for GC-MS analysis (Wittmann 2007)), IsoCor performs appropriate correction by taking into account the contribution of naturally-occurring isotopes brought by the derivatization reagent.

- IsoCor provides two options that allow flexible correction of isotopic clusters to account for (or not) the composition of the label input including: i) correction for the

isotopic purity of the label input and ii) correction for the occurrence of unlabeled positions in the label input. These options are of particular value for specific applications ((Wittmann and Heinzle 1999); (Rodriguez-Castrillon, Moldovan et al. 2008)).

- IsoCor calculates the mean isotopic enrichment of molecules, which refers to the molecular content in the isotope. This information is particularly useful for targeted metabolic investigations, such as the quantification of split ratios between two metabolic pathways.

- IsoCor can be applied to large datasets and can deal with the increasing number of data that are generated with modern methods in a single experiment.

3. Method and implementation

The correction is performed with the matrix-based method introduced by (van Winden, Wittmann et al. 2002), which requires solving the following equation:

$$IC_{cor} = CM^{-1}.IC_{meas} \qquad eq\,(1)$$

where IC_{cor} denotes the corrected isotopic cluster, CM is a correction matrix, and IC_{meas} is the measured isotopic cluster.
This formalism has been implemented in Python programing language (http://python.org) which enables seamless usage of IsoCor on Windows, MacOS, Linux and other platforms supporting Python.

Construction of the correction matrix

The size of the correction matrix is $m \times n$, with $m \geq n$, where m is the length of the measured isotopic cluster and n is the number of isotopologues, i.e. $a+1$ where a is the number of atoms of the tracer element in the molecule. The i-th column of the correction matrix ($1 \leq i \leq n$) is the isotopic cluster of the molecule containing $i-1$ atoms of the tracer. It is commonly calculated using combinatorial probabilities (Hellerstein and Neese 1999). For molecules containing a high number of atoms, the number of possible combinations to be calculated is high which can cause a bottleneck in the correction process. In probability theory, the probability distribution of the sum of independent variables is the

convolution of their individual distributions. Therefore, the calculation can be performed faster by iteratively convolving the isotopic vectors v_A=[p(A(1)),…,p(A(x))] of all the atoms – except those of the label tracer –, where p(A(x)) denotes the natural abundance of the *x*-th isotope of the atom *A*. Finally, the contribution of the tracer is added by convolving the previous isotopic cluster *i-1* times by the purity vector of the tracer (e.g. [0.01, 0.99] for a purity of 99%), and *n-i* times by its natural abundance (e.g. [0.9893, 0.0107] in a ^{13}C-labeling experiment) when it has to be corrected.

Optimization algorithm

Because the measurement of isotopic clusters contains noise, equation 1 has to be solved using a least squares method. This process can result in biases, such as the calculation of negative isotopic fractions sometimes observed with MSCorr. Because negative fractions cannot exist, they are just flattened to zero in (Moseley 2010), which can lead to errors in the estimated parameters and in their interpretation. To avoid such biases, the optimization method used in IsoCor includes constraints on the isotopic fraction values. We applied the L-BFGS-B algorithm described in (Byrd, Lu et al. 1995) and implemented in the *optimize.l_bfgs_g()* function of the SciPy module. The cost function of the minimization process is defined as the sum of the squared weighted errors. The lower boundary of corrected isotopic fractions is constrained to 0, and the stopping criterion is fixed to 10^{-10} of relative reduction in cost value. The fitted isotopologue distribution is normalized to 1. The residuum vector may be used for an independent quality control (Moseley 2010).

The calculation speed and correction accuracy of IsoCor were evaluated by processing several sets of simulated isotopologues with varying numbers of tracer atoms (up to 100 to stress the algorithm) and levels of labeling. The correction was performed after addition of the theoretical contributions of other naturally abundant elements to these isotopologues, with or without addition of measurement noise. The implementation was quite robust and highly accurate in both cases (see the IsoCor tutorial), showing that IsoCor provides high calculation speed together with high precision. Hence IsoCor is capable to efficiently process the large datasets that can now be acquired with modern, high-throughput MS technologies. This is critical for high-throughput metabolomics and even more for high-throughput fluxomics, which is a current major challenge for many biological applications.

4. Usage

IsoCor is provided with plain text files that contain the information required to perform the correction. The file Isotopes.dat contains the values of the natural abundances of isotopes commonly found in biological compounds (Rosman and Taylor 1998). The files Metabolites.dat and Derivatives.dat contain the elemental formula of most common metabolites and derivative residues, respectively. These files can be easily edited and implemented according to the user's needs.

Complete usage of IsoCor is detailed in the tutorial provided with the software. Figure 2 shows its graphic user interface. Briefly, the user can load datasets containing one or more isotopic clusters to be corrected (Panel C).

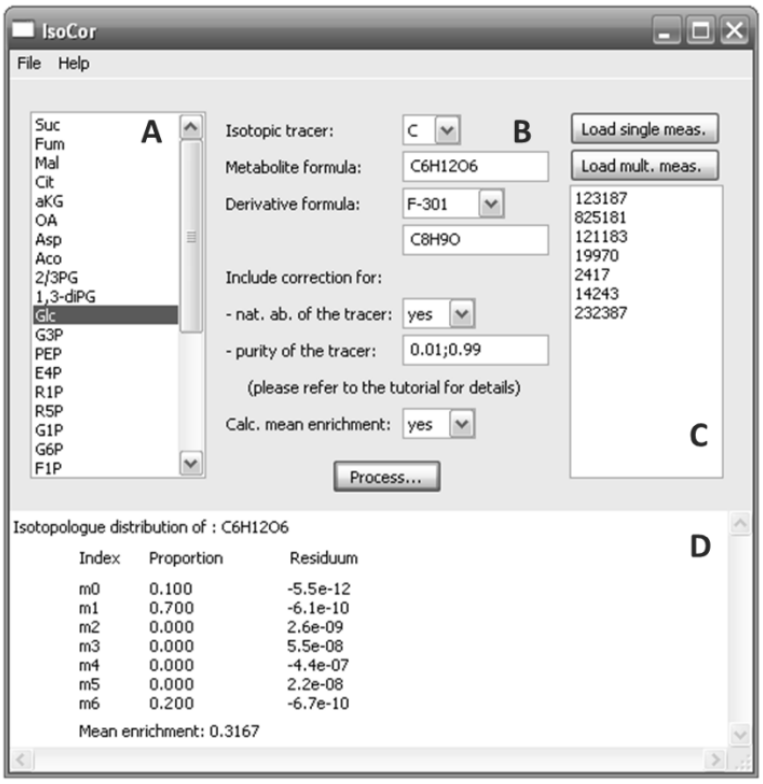


Figure 2. Screenshot of IsoCor after correction of a single isotopic cluster.
(A) List of metabolites. (B) Correction parameters panel.
(C) Selection of experimental data files. (D) Calculation results.

These data files should be generated according to the format described in the tutorial. The options – i.e. selection of the isotopic tracer, correction for the isotopic purity of the label input, correction for unlabeled positions in the label input, and calculation of mean enrichment – of the correction process have to be selected (Panel B) before running the

calculation. When correction is applied to single isotopic clusters, the calculated data – corrected isotopic fractions and minimization residuum – are displayed in panel D. When correction is applied to a dataset containing several measurements, the corrected data are written in a text file, and other information – e.g. errors that occur during correction – is written in a log file and displayed in panel D.

5. References

Byrd, R. H., P. H. Lu, et al. (1995). "A Limited Memory Algorithm for Bound Constrained Optimization." Siam Journal on Scientific Computing **16**(5): 1190-1208.

Hellerstein, M. K. and R. A. Neese (1999). "Mass isotopomer distribution analysis at eight years: theoretical, analytic, and experimental considerations." Am J Physiol **276**(6): 1146-1170.

Kiefer, P., C. Nicolas, et al. (2007). "Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry." Anal Biochem **360**(2): 182-188.

Moseley, H. N. (2010). "Correcting for the effects of natural abundance in stable isotope resolved metabolomics experiments involving ultra-high resolution mass spectrometry." BMC Bioinformatics **11**: 139.

Nicolas, C., P. Kiefer, et al. (2007). "Response of the central metabolism of Escherichia coli to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase." FEBS Lett **581**(20): 3771-3776.

Rodriguez-Castrillon, J. A., M. Moldovan, et al. (2008). "Isotope pattern deconvolution as a tool to study iron metabolism in plants." Anal Bioanal Chem **390**(2): 579-590.

Rosman, K. J. R. and P. D. P. Taylor (1998). "Isotopic compositions of the elements 1997." Pure Appl Chem **70**(1): 217-235.

Sauer, U. (2006). "Metabolic networks in motion: 13C-based flux analysis." Mol Syst Biol **2**: 62.

van Winden, W. A., C. Wittmann, et al. (2002). "Correcting mass isotopomer distributions for naturally occurring isotopes." Biotechnol Bioeng **80**(4): 477-479.

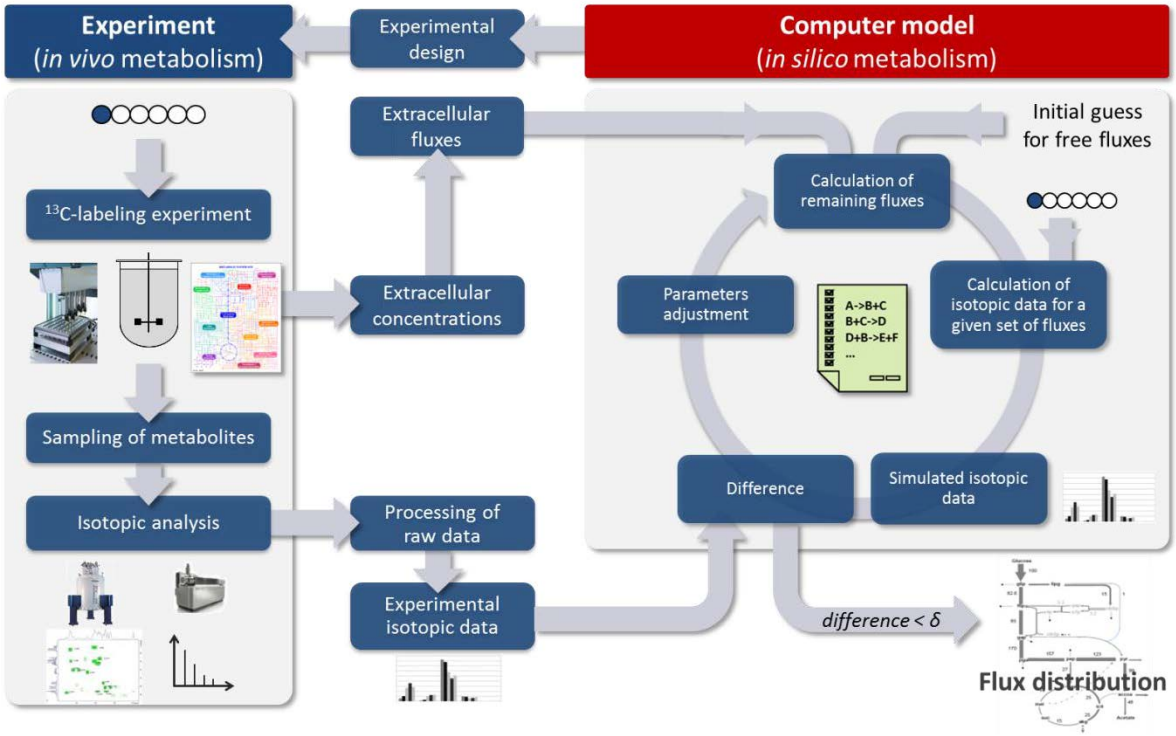
Wahl, S. A., M. Dauner, et al. (2004). "New tools for mass isotopomer data evaluation in (13)C flux analysis: mass isotope correction, data consistency checking, and precursor relationships." Biotechnol Bioeng **85**(3): 259-268.

Wittmann, C. (2007). "Fluxome analysis using GC-MS." Microb Cell Fact **6**: 6.

Wittmann, C. and E. Heinzle (1999). "Mass spectrometry for metabolic flux analysis." Biotechnol Bioeng **62**(6): 739-750.

Chapter 6

influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments



NOTE: *this chapter contains a manuscript published in Bioinformatics:*

Sokol S., Millard P., Portais J.C. (2012). influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments. *Bioinformatics*, 28(5):687-693.

1. Abstract

The problem of stationary metabolic flux analysis based on isotope labelling experiments first appeared in the early 1950s and was basically solved in early 2000s. Several algorithms and software packages are available for this problem. However, the generic stochastic algorithms (simulated annealing or evolution algorithms) currently used in these software require a lot of time to achieve acceptable precision. For deterministic algorithms, a common drawback is the lack of convergence stability for illconditioned systems or when started from a random point. In this article, we present a new deterministic algorithm with significantly increased numerical stability and accuracy of flux estimation compared with commonly used algorithms. It requires relatively short CPU time (from several seconds to several minutes with a standard PC architecture) to estimate fluxes in the central carbon metabolism network of *Escherichia coli*.

Availability: The software package influx_s implementing this algorithm is distributed under an OpenSource licence at http://metasys.insa-toulouse.fr/software/influx/

2. Introduction

Metabolic flux analysis (MFA) aims at quantifying the actual rates of biochemical reactions occurring in living cells. In recent decades, MFA has been increasingly used to identify novel metabolic pathways (Fischer and Sauer 2003; Peyraud, Kiefer et al. 2009), for in depth understanding of metabolism (Sauer, Canonaco et al. 2004; Perrenoud and Sauer 2005; Nicolas, Kiefer et al. 2007). It is extensively used in biotechnology to improve the metabolic properties of industrially relevant organisms (van Gulik, de Laat et al. 2000; Becker, Klopprogge et al. 2007). More recently, MFA has been successfully integrated with other omics tools (transcriptomics, proteomics, metabolomics, etc.) to obtain novel biological insights through systems biology (Shimizu 2004; Ishii, Nakahigashi et al. 2007; Lemuth, Hardiman et al. 2008).

The growing interest in MFA underlines the importance of developing reliable tools. The present contribution particularly addresses the need for accurate and stable algorithms for solving the least-squares problem that underlies the calculation of fluxes in MFA.
In a stationary metabolic system, the biochemical reactions which occur in a cell can be described by the following stoichiometric linear equation:

$$Sv = 0$$

where S is m×n stoichiometric matrix, m rows and n columns correspond to the number of metabolites and reactions, respectively, v is the vector of all net fluxes. Each component of the vector v expresses a net flux, i.e. the net quantity of material converted by a particular reaction per time unit. The whole equation system expresses the mass conservation law in the metabolic system. At metabolic (quasi-)steady-state, the intracellular concentrations of metabolites are kept constant.

For most metabolic systems, the stoichiometry matrix S is underdetermined, i.e. the number of equations m is lower than the number of fluxes n. Some fluxes can be measured experimentally. This is generally true of input and output fluxes, but is usually not enough to allow the calculations of all fluxes in the system. The remaining degrees of freedom, so-called free fluxes, need additional equations to be calculated. This can be achieved using different approaches. For example, flux balance analysis (FBA) requires maximization of some linear cost function like biomass yield (Edwards, Ibarra et al. 2001). In the approaches using isotope labelling experiments (ILE) discussed in this article, additional relationships between fluxes come from the measurement of the labelling patterns (or isotopomer distributions) of selected metabolites. Currently, these measurements can be made by mass spectrometry (mass isotopomers) or by Nuclear magnetic resonance (NMR) (positional isotopomers).

The MFA-ILE approach was developed in the 1950s when ^{14}C radioactive isotopes were used to elucidate fragments of carbon metabolism in rat liver (Weinman, Chaikoff et al. 1950; Strisower, Chaikoff et al. 1951). Since the 1980s–1990s, a stable isotope ^{13}C has preferably been used instead of the radioactive ^{14}C. For many years, the equations describing the label distribution in a given metabolic network and their solution were derived by hand (Heath 1968). In the early 1990s, general mathematical descriptions of the labelling problem were introduced (Schuster, Schuster et al. 1992; Wiechert 1994;

Zupke and Stephanopoulos 1994). This generalization led to a need to solve algebraic systems of high dimensions (often ill-conditioned) to find the labelling state of a given metabolic network. This paved the way for the intensive use of applied mathematics in the MFA field.

(Wiechert, Möllney et al. 1999) introduced a variable change based on the cumomer (cumulated isotopomer) concept. This approach allowed a large non-linear system to be decomposed into a cascade of smaller linear problems, one per cumomer weight and thus greatly simplified the computational needs for labelling state calculations. A general software exploiting the cumomer approach 13CFlux was then published, which provides flux estimation based on all types of labelling data (mass and positional isotopomers) (Wiechert, Mollney et al. 2001).

(Antoniewicz, Kelleher et al. 2007) proposed another variable change based on elementary metabolite unit (EMU) concept. The EMU framework is based on an efficient decomposition of an atom transition network, which identifies the minimum amount of information needed to simulate the experimental data without any loss of information. This led to a significant reduction in the dimension of the labelling problem. As a consequence, the computation of labelling variables now requires much less time and memory. A software based on EMUs, called OpenFlux, is now available (Quek, Wittmann et al. 2009).

A comparable gain in the computation of cumomers was announced by (Weitzel, Wiechert et al. 2007). These authors showed that often a highly dimensional linear system describing cumomers of a given weight can be decomposed into series of smaller linear subsystems corresponding to so-called strongly connected components (SCCs). To the best of our knowledge, there is currently no publicly available software exploiting this approach. For a more complete review of recent progress in MFA-ILE, including both experimental and mathematical improvements, see (Tang, Martin et al. 2009).

The EMU and SCC concepts led to computational gains by improving the mathematical formulation of the labelling data problem involved in residue calculation. Another potential source of improvement is the optimization process, i.e. the iterative process by which the simulated labelling data are fitted to the experimental data.

The two steps - i.e. residue calculation and optimization - of the flux calculation process are independent, and can be improved separately. It is likely that future software will combine the most efficient solutions for each calculation step.

Here, we present an original deterministic algorithm called non-linear least squares with inequality constraints (NLSIC) with its improved optimization process. It solves MFA-ILE problems with good convergence robustness without sacrificing convergence speed. NLSIC also enhances the numerical accuracy of a final solution. Moreover, it did not suffer from the local minima problem in the cases we had at hand. We discuss this aspect in the Section 5.

In this article, the performance of the NLSIC algorithm is illustrated by solving a MFA-ILE problem. But this algorithm can be advantageously applied to a larger framework of non-linear least squares problems with inequality constraints. The NLSIC algorithm is particularly suited to solve constrained least-squares problems with ill-conditioned or even rank-deficient Jacobian.

We start with the formulation of the mathematical problem in MFA-ILE. We then describe the NLSIC algorithm and its practical implementation for flux estimation. The influx_s software is further validated in the numerical section by comparing its performance to that of several other algorithms available in the widely used 13CFlux software package. For this purpose, the different algorithms were applied to publicly available data from (Zamboni, Fendt et al. 2009).

3. Problem formulation

In this section, we use the same conventions and notations as in (Mollney, Wiechert et al. 1999). The free fluxes and free scaling parameters in a given metabolic system can be estimated using a least-squares problem that can be written as follows:

$$\arg min_{\theta,\omega}\ T(\theta,\omega) = \|F_w(\theta) - w\|^2_{\sum w} + \left\|F_y(\theta,\omega) - y\right\|^2_{\sum y} \qquad \text{eq (1)}$$

Here T is a cost function representing the sum of squared weighted errors. Its arguments, θ a free flux vector and ω a free scale vector, are the free parameters that are adjusted during the minimization process. Vectors w and y are the vectors of

measured fluxes and labelling data, respectively, whereas vector functions F_w and F_y represent the data simulations matching measured values w and y.

Matrices Σ_w and Σ_y are covariance matrices characterizing the experimental noise in flux and labelling data, respectively. They are often assumed to be diagonal as the noise is expected to be uncorrelated.

The solution of (1) must satisfy linear inequality constraints:

$$U\binom{\theta}{\omega} \geq c \qquad \text{eq (2)}$$

where U is an inequality matrix which is multiplied by a compound vector of free parameters θ and ω, c is a right-hand side vector.

Inequalities express some biological conditions which might not be naturally respected by solving the unconstrained problem (1). These conditions can include the unidirectionality of some reactions, or the existence of lower or upper bounds for the flux values, and so on.

Compared with the original cumomer formulation (Wiechert, Möllney et al. 1999), the main difference consists in using a reduced cumomer set in simulations $F_y(\theta,\omega)$. Usually, simulations are performed with the entire set of isotopomers allowed by the network. But in most cases, the full set of cumomers is not needed to simulate the experimental data. To reduce the number of cumomers to be calculated, the network is read starting from cumomers involved in simulated data. Then, upstream the fluxes, all the precursor cumomers, up to the input ones, are added to the minimum set. This technique can result in a significant reduction in the number of cumomers to be simulated and, hence, in the calculation time. A similar idea was proposed in the EMU framework (Antoniewicz, Kelleher et al. 2007). For example, for a network of *Escherichia coli* described in Section 5, the total number of cumomers in the complete model was 3183, whereas it was only 2271 in the reduced model. The number of cumomers of weight 1 remains unchanged but a significant reduction is obtained for all other weights (Supplementary Material 1). The complexity of solving a dense linear system is $O(n^3)$ where n is the problem size. The theoretical speedup in the cumomer calculations for the above-mentioned example is $\sum n_{full}^3 / \sum n_{reduced}^3 \approx 2.8$ where the sum involves the problem sizes of all cumomer weights present in the system. The savings are highly dependent on the network and on

the measurements concerned. The greatest reduction in the total number of cumomers is expected for the simulation of positional labelling data where only cumomers of weight 1 are used. So all cumomers of higher weights can be excluded from the simulations without loss of information.

For the sake of simplicity, and without loss of generality, we omit writing vector ω in the free parameter vector. We also omit the matrices Σ_w and Σ_y. So, from now on, θ denotes the whole vector of free parameters. The vector functions F_w and F_y are combined in one vector function F, and vectors w and y in one vector u. Using these new terms, the problem (1, 2) takes the form of a classical non-linear least-squares problem with inequality constraints:

$$\begin{cases} arg\ min_\theta\ T(\theta) = \|F(\theta) - u\|^2 \\ U\theta \geq c \end{cases}$$

A Jacobian matrix at a point θ defined as $J(\theta) = \partial F(\theta)/\partial\theta$ and a residual vector $r(\theta) = F(\theta) - u$, which also depends on θ, are introduced. Taking into account the first-order Taylor development:

$$\|r(\theta + p\| = \|r(\theta) + J(\theta)p\| + O(\|p\|^2)$$

and searching for a vector p that minimizes the norm $\|r(\theta + p)\|$, a linearized incremental form of the problem in the point θ is obtained:

$$\begin{cases} Jp = -r \\ Up \geq c - U\theta \end{cases} \qquad \text{eq (3)}$$

where the new variable vector p is an unknown correction to the current free parameter vector θ and the equality sign must be interpreted in the least-squares sense. This linear problem is hereafter referred to as least-squares with inequalities (LSI).

4. NLSIC algorithm

Generally, non-linear, unconstrained optimization problems like (1) can be solved using a wide range of methods which are well described in text books. These methods can be divided into two classes: stochastic and deterministic algorithms. Among the most frequently cited deterministic algorithms are gradient descent, conjugate gradient, Broyden-Fletcher-Goldfarb-Shanno optimization algorithm (BFGS). To deal with

inequality constraints, the deterministic algorithms have to be used in conjunction with appropriate techniques, e.g. interior point methods like barrier function or active set. Non-linearity is most often treated using a classical line search approach, sequential quadratic programming (SQP) or a combination of the two approaches. Finally, techniques which are widely used to make these algorithms globally converge, i.e. independent of the distance between the starting point and the convergence point, are trust region and backtracking. All these algorithm aspects - non-linearity, inequalities, globalization and stopping criterion - are detailed hereafter.

Non linearity

The non-linearity approach in the NLSIC algorithm is close to the classical SQP method. The latter solves a normal (square) linear system at each iteration:

$$Hp = -J^T r$$

where H is a matrix proportional to Hessian, i.e. the matrix of second partial derivatives of the cost function T:

$$H_{ij} = (J^T J)_{ij} + k\frac{\partial^2 r_k}{\partial\theta_i \partial\theta_j} r_k$$

or more frequently H is taken to be just J^TJ; while in the NLSIC algorithm, a least-squares problem with rectangular matrix J is solved:

$$Jp = -r \qquad \text{eq (4)}$$

This point is crucial for numerical stability and precision even if its numerical cost is higher than solving the normal system. In real-world problems, like in MFA-ILE, the Jacobian J is often ill-conditioned, i.e. $\kappa(J)>>1$ (here κ is a condition number) so that the condition of J^TJ will be even worse. For example, in l^2-norm $\kappa(J^TJ)=\kappa(J)^2$. If $\kappa(J)$ is say 10^7, the problem (4) can still be solved in double precision arithmetics with satisfactory precision; whereas the corresponding normal system has $\kappa(J^TJ)$ as high as 10^{14} and will be considered as numerically singular. It is also well known that the matrix condition is a determining factor for convergence stability and for the precision of many numerical methods. By solving a system with matrix J instead of H, condition deterioration is avoided and hence numerical stability and precision are preserved

Inequalities

The presence of inequalities makes the optimization problem more difficult from a numerical point of view. Like for SQP (Liu 2005), for NLSIC, an active set method was chosen to deal with inequalities.

As an overdetermined linear system has to be solved at each non-linear iteration, one particular method called non-negative least squares (NNLS) suits this purpose well (Lawson and Hanson 1974). To reduce the LSI problem to an NNLS problem, an intermediate problem called LDP has to be formulated as detailed hereafter.
Let us start by formulating an NNLS problem. Given a rectangular m×n matrix A and right-hand side vector b of size m, find a vector x of size n such that $x_i \geq 0, \forall i$. In short:

$$\begin{cases} Ax=b \\ x_i \geq 0, \quad \forall i \end{cases} \qquad \text{eq (5)}$$

here again, the equality sign must be interpreted in the least-squares sense.
The NNLS algorithm solves this problem by combining QR decomposition (Björck 1996) of the matrix A with an active set method. One attractive feature of this algorithm is that QR decomposition is not recalculated each time a new active set is tried.

Only those elements of matrices Q and R that are affected by changes in the active set, are updated. Another feature worth mentioning is that an appropriate active set is rapidly found in only a few iterations, at least in the problems we had to deal with. An interface to this algorithm in R language is available under OpenSource licence (http://cran.r-project.org/web/packages/nnls/).

It can be shown that an NNLS problem can be set as a reformulation of another so-called least distance programming (LDP) problem (Lawson and Hanson 1974). An LDP problem is formulated in the following way: find a vector z of smallest norm l^2 satisfying a set of inequalities:

$$Ez \geq f \qquad \text{eq (6)}$$

where E and f are any real rectangular matrix and vector of appropriate sizes, respectively. Given an LDP problem, an equivalent NNLS problem can be obtained by setting:

$$A = \begin{pmatrix} E^T \\ f^T \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Once NNLS is solved for x, a residual vector $v = Ax - b$ is calculated and the solution for LDP is given by $z_i = -v_i/(v_{n+1})$, for $i = 1, \ldots, n$. The case $v_{n+1} = 0$ corresponds to infeasible inequality constraints (6).

In turn, the LDP solution can be used to address the LSI problem (3). To reduce the problem (3) to (6), the following variable change is performed:

$$z = Rp + Q^T r$$

where Q and R are components of QR decomposition of the Jacobian $J = QR$. We then put:

$$E = UR^{-1}$$
$$f = c - U(\theta - R^{-1}Q^T r)$$

Once LDP with E and f is solved for z, a vector p can be calculated as $p = R^{-1}(z - Q^T r)$. It can happen that during iterations the Jacobian becomes rank deficient. In this case, we provide an option to continue iterations with an approximate least norm solution. A strict least norm solution always exists and is unique if the inequalities are feasible. A description of the method chosen to obtain an approximate least norm solution for p satisfying inequality constrains is beyond the scope of this article. Interested readers can refer to the source code of the function lsi_ln() in the file nlsic.R distributed with the influx_s software.

To summarize, starting from the LSI problem, an equivalent LDP problem is formulated which in turn is reformulated as an NNLS problem. Having solved NNLS, we go backwards in the same way: first the solution for LDP and then the solution for the original LSI problem are calculated.

Globalization

The last but not least component of the NLSIC algorithm is the use of a backtracking algorithm to ensure the global convergence of the non-linear iterations. The backtracking

was derived and successfully used to solve systems of non-linear equations and later for non-linear unconstrained optimization (Dennis and Schnabel 1996). The key idea behind this method is to iteratively shorten a candidate vector p if at some point θ the norm $\|r(\theta + p)\|$ is much higher than predicted by the linear model norm $\|r(\theta) + Jp\|$. If r is differentiable in the neighbourhood of θ, a sufficiently small vector p can always be found such that two norms are close enough. Backtracking ensures that the increment vector p remains in the validity domain of linear approximation (3), and hence ensures the norm $\|r(\theta)\|$ is reduced during non-linear iterations. Near the solution, where the first value of p found is sufficiently small to satisfy backtracking conditions, a very rapid convergence of Newton-like methods is often observed.

Algorithm 1. NLSIC

1. Set initial values $i:=0$, Θ_0
2. Solve LSI problem $J_i p_i = -r_i$ subject to $Up_i \geq c - U\Theta_i$
3. If $||p_i||$ is sufficiently small or the limit of iteration number
is reached, stop iterations with $\Theta := \Theta_i + p_i$
4. While backtracking condition (7) is not satisfied or maximal number of backtrack iterations is reached, iteratively
set $p_i := \alpha p_i$ with $0 < \alpha < 1$
5. Set $\Theta_{i+1} = \Theta_i + p_i$
6. Set $i := i+1$ and restart with p. 2.

More formally, the backtracking method consists in iteratively setting p:=αp with some positive constant α<1 until the following backtrack condition is fulfilled

$$\|r(\theta + p\| \leq \|r(\theta)\| + \beta r^T Jp \qquad \text{eq (7)}$$

with some constant 0<β<1. Parameters α and β are chosen a priori by the user.

As the original backtracking method was designed to be used in an unconstrained context, it is necessary to check that it can be used in the context of inequality constraints. This is easy to do by observing that if two vectors θ and p are such that θ and $\theta + p$ satisfy the inequalities, i.e. $U\theta \geq c$ and $U(\theta + p) \geq c$ then by multiplying both inequalities by positive numbers (1−α) and α respectively and summing them, we obtain $U(\theta + \alpha p) \geq c$. This means that for any α, the intermediate vector of backtracking iterations $\theta + \alpha p$ cannot leave the feasibility domain delimited by the inequalities. As

mentioned above, due to backtracking, the sequence of norms $\|r(\theta_i)\|$ monotonously decreases. This sequence has a low bound 0, and is hence necessarily convergent. This guarantees that the whole NLSIC algorithm is globally convergent under the assumption that there is no significant problem due to finite precision arithmetics. Strictly speaking, the convergence point may differ from the global minimum but achieving convergence independently of the starting point is already appreciable.

Stopping criterion

The stopping criterion in NLSIC is based on the norm of the correction vector p and not on the variation of the cost function T as it is usually the case. This point is important for ill-conditioned systems, as relatively big correction vectors sometimes produce very small variations in cost function. In this situation, the error control based on T variations could stop iterations relatively far from the solution. In practice, this could be mistakenly interpreted as falling in a local minimum. In a nutshell, the NLSIC algorithm can be classified as a sequential LSI algorithm (based on NNLS) with backtracking globalization.

5. Implementation

Input data for the MFA-ILE problem, i.e. metabolic network, carbon transitions in reactions and measurements are all supplied in a plain text file. The input file must be in the FTBL format developed for the wide-spread 13CFlux software (http://www.13cflux.net) (Wiechert, Mollney et al. 2001). A module of influx_s package parsing input file and generating an executable code is written in Python (http://python.org). The executable program calculating flux estimations (specific to each network) is automatically generated in R language (http://www.r-project.org).

The NLSIC algorithm is also programmed in R. As it is an interpreted language, for the sake of calculation speed, the most time consuming linear algebra operations: QR and LU decompositions, NNLS algorithm and so on, are programmed in FORTRAN and C in several third part software distributed independently of influx_s. A small part of the influx_s package concerning sparse matrix updates is written in FORTRAN. This part of the executable code is compiled only once and is not regenerated for every new network

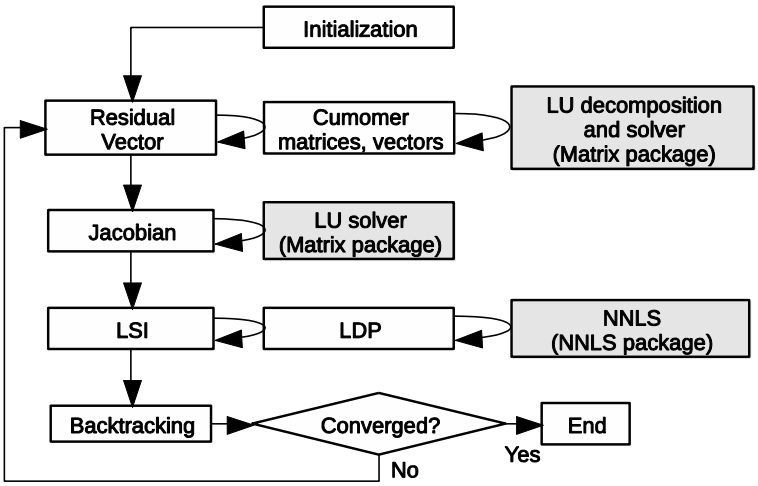


Figure 1. Flow diagram of the NLSIC algorithm. The third part software packages are in the grey boxes. Curved arrows mean 'need a call to …', straight arrows mean 'go to the next block'.

input. The use of sparse matrices (http://cran.r-project.org/web/packages/Matrix/) greatly reduces memory and CPU requirements, especially in the case of large networks which are transformed into huge matrices. Third part software used by NLSIC is indicated in Figure 1 in grey boxes.

For the sake of rigorous simulations, an option to include the growth flux μM into the stoichiometric balance is provided. Even if its inclusion does not have a major impact on the final results in most common situations (data not shown), users can themselves check that this is true in their own case. It should be noted that this option requires absolute concentrations for intracellular metabolites, which may be not very easy to measure.

Confidence intervals of the free and dependent fluxes can be assessed using two methods in influx_s: by linearized statistics as proposed in (Mollney, Wiechert et al. 1999) or by Monte Carlo simulations. The latter is programmed for parallel execution on multi-core architecture where a third part R package multicore is available.

The influx_s software was developed and tested as a command line tool on a Linux platform but potentially it can be run on any platform where Python and R are installed. At the time of writing, the multicore package is not stable enough on Windows platforms and consequently, until a stable version is available, users will have to run such simulations on a single core on their Windows platform. The package influx_s is available

at http://metasys.insa-toulouse.fr/software/influx/ under OpenSource licence. All third part libraries used in our software are also freely available at their respective sources.

6. Numerical results

In this section, a publicly available FTBL file Ecoli.ftbl from Supplementary Material 3 of (Zamboni, Fendt et al. 2009) was used to compare the NLSIC algorithm implemented in the influx_s software with the three optimization algorithms used in the 13CFlux software (v20050329): Evolution, BFGS and donlp2 (Spellucci 1993). The 13CFlux software is widely distributed throughout the biological community concerned with MFA and, in the last decade, has become *de facto* reference method for flux calculations.
The file Ecoli.ftbl describes a central metabolic network of the bacterium *E. coli*. The network includes central carbon metabolism (glycolysis, gluconeogenic reactions, pentose phosphate pathways, the tricarboxylic acid cycle, glyoxylate shunt and anaplerotic reactions) and reactions for amino acid biosynthesis. It contains 35 internal metabolites, 68 reactions (of which 16 are reversible and 52 are not reversible). The labelling data includes 193 isotopomer measurements obtained only by gas chromatography–mass spectrometry (GC–MS). Free parameters were composed of 27 free fluxes and 35 scale parameters.

The GC-MS data provided in the FTBL file are not sufficient to determine all free fluxes. So the network is structurally undefined. This leads to rank-deficient Jacobian. So the least norm solution (available in influx_s with an option --ln) that we discussed in Section 3.2, was revealed to be indispensable in this situation. Another problem encountered with this FTBL file was that cumomer balance matrix could become singular when some fluxes vanished to 0. To prevent this happening, the net fluxes over non-reversible reactions were constrained to be over 10^{-4} with an option --clownr=1e-4.

To compare various algorithms, a suite of 10 FTBL files was generated. These files differed only in their random starting points while the network and measurement sections remained the same. Initial flux values were uniformly drawn from [0,1] interval and the resulting vector was projected on the feasibility domain by solving an LDP problem.

As the Evolution algorithm implemented in 13CFlux does not include a stopping criterion, it was run for a fixed time of 3 h on each input file and the results achieved at this time (and corresponding to the best fit) are shown. The stopping criterion for BFGS and donlp2 algorithms were those set by default in 13CFlux software. The running time for these two algorithms were limited to 15 min. The BFGS algorithms was often stalled at this time, and donlp2 stopped before reaching the time limit. The NLSIC algorithm was stopped when the norm ||p|| was lower than 10^{-5} or the upper limit on the non-linear iteration number (50) was reached. All numerical experiments were run on a laptop with 1 GHz bi-core processor (actually only one core was used) and 2 GB of RAM.

Figure 2 shows box-plots representing the spread of the final cost values for the four methods applied to 10 different starting points and the three deterministic methods: BFGS, donlp2 and NLSIC when they followed the results of the Evolution method. The double algorithm such as a stochastic algorithm followed by a deterministic algorithm is sometimes recommended (Zamboni, Fendt et al. 2009) to avoid local minima while avoiding prohibitive calculation time. The final cost values for the Evolution algorithm ranged between 156 and 268. The BFGS methods achieved the minima between 138 and 4409 when run alone and between 156 and 268 when it followed the Evolution results. In fact, BFGS rarely improved the Evolution results and most often stagnated at the same level of the cost value. This behaviour was also observed in (Zamboni, Fendt et al. 2009). The algorithm donlp2 produced better results. Its final cost was spread between 138 and 291 when run alone and between 131 and 211 when preceded by the Evolution algorithm. Finally, the NLSIC algorithm converged to the same value (127) from all starting points and achieved the lowest cost of all the results obtained whether preceded or not by the Evolution algorithm. While multiple convergence points for the BFGS and donlp2 methods could be interpreted as local minima in which the convergence was trapped, in fact this was not the case. Since using the final results of, for example, the Evolution or BFGS algorithms as starting point for NLSIC, it continued convergence until it reached the previously found solution, 127.

The network was not well defined by the data that were provided, so almost all fluxes are statistically undetermined. (Zamboni, Fendt et al. 2009) had to constrain 21 of 27 free fluxes to be able to evaluate the confidence intervals of the remaining fluxes. Of all the free fluxes in the original (not constrained) FTBL file, only the uptake flux was

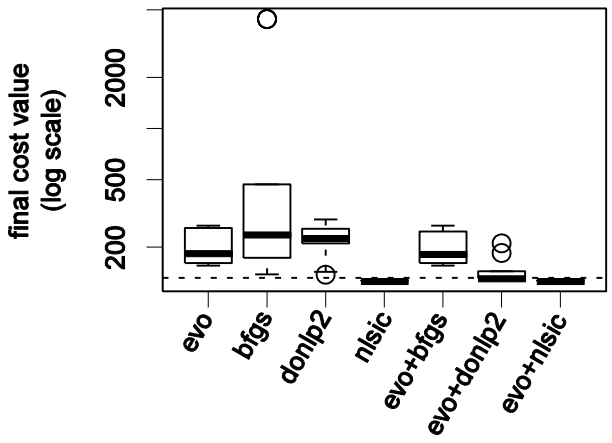


Figure 2. The spread of final cost values from 10 random starting points for four algorithms: evolution algorithm, BFGS, donlp2 and NLSIC. NLSIC was shown to be the most stable, it converged to the same point in all cases. It was also the most accurate as its final cost value was the lowest among the four algorithms tested. The dashed line corresponds to a level of 132 found in Zamboni et al. (2009).

statistically determined. In Figure 3, the box-plots corresponding to this flux are shown for all algorithms tested. Ideally, this value as well as the final cost should not depend on the starting points. When this happens, for example with the Evolution, BFGS and donlp2 algorithms, this numerical instability adds to already present experimental noise and reduces the quality of flux assessment. The NLSIC algorithm was almost free of this drawback as can be seen in Figure 3.

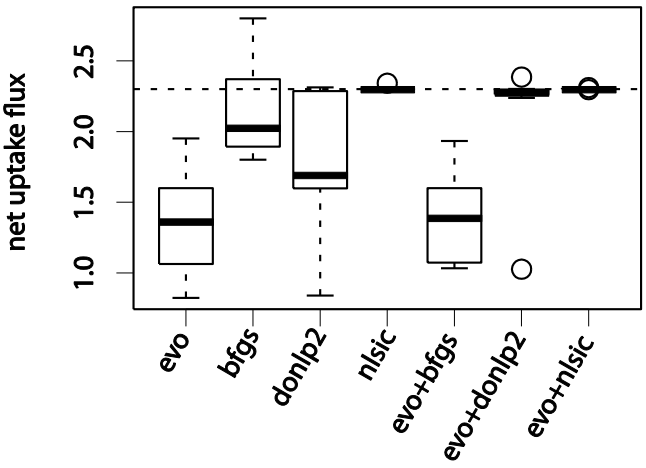


Figure 3. Spread of the uptake flux estimated by the tested algorithms for 10 random starting points. The NLSIC algorithm produced the most stable results and the closest to the measured value 2.3 represented by the dashed line.

A word about calculation times. (Zamboni, Fendt et al. 2009) estimated the time necessary for flux calculation using 13CFlux software to be 1 day. This time included

several runs of the Evolution algorithm, optionally followed by some deterministic algorithm(s) and choosing the lowest final cost value of all the runs. In the example given here, NLSIC ran only for several minutes and provided a better minimum, 127, than the 132 found by the authors of the above-cited paper. By considerably shortening the calculation time, our software represents an additional step in the direction of high-throughput flux calculation.

7. Conclusion

We developed an algorithm for solving NLSIC, which was used to solve the MFA-ILE problem. In this field, it outperformed widely used algorithms such as BFGS, donlp2 or Evolution algorithms not only in numerical stability but also in the accuracy of the solution. The increased numerical accuracy led us to conclude that a problem of local minima, as often mentioned in the literature dedicated to MFA-ILE, could in some cases be a false problem. Probably, it was lack of precision in convergence which was interpreted as trapping in a local minimum.

The significantly improved computer efficiency, accuracy and reliability of influx_s makes the MFA-ILE approach more accessible for a wide biological community interested in fluxomics. We can reasonably expect that the NLSIC algorithm will also provide the same benefits in future MFA-ILE developments dealing with the dynamics of labelling propagation in metabolically stable networks.

The software influx_s implementing NLSIC in the MFA-ILE context is distributed under OpenSource licence. It has the same general character and input format as 13CFlux software as it can take into account all types of labelling data coming both from MS and NMR.

8. References

Antoniewicz, M. R., J. K. Kelleher, et al. (2007). "Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions." Metab Eng **9**(1): 68-86.

Becker, J., C. Klopprogge, et al. (2007). "Metabolic flux engineering of L-lysine production in C*orynebacterium glutamicum*-over expression and modification of G6P dehydrogenase." J Biotechnol **132**(2): 99-109.

Björck, A. (1996). Numerical Methods for Least Squares Problems. Philadelphia, SIAM.

Dennis, J. E. and R. B. Schnabel (1996). Numerical methods for unconstrained optimization and nonlinear equations. Philadelphia, SIAM.

Edwards, J. S., R. U. Ibarra, et al. (2001). "In silico predictions of E*scherichia coli* metabolic capabilities are consistent with experimental data." Nat Biotechnol **19**(2): 125-130.

Fischer, E. and U. Sauer (2003). "Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS." Eur J Biochem **270**(5): 880-891.

Heath, D. F. (1968). "The redistribution of carbon label by the reactions involved in glycolysis, gluconeogenesis and the tricarboxylic acid cycle in rat liver." Biochem J **110**(2): 313-335.

Ishii, N., K. Nakahigashi, et al. (2007). "Multiple high-throughput analyses monitor the response of E. coli to perturbations." Science **316**(5824): 593-597.

Lawson, C. L. and R. J. Hanson (1974). Solving least squares problem, Prentice-Hall.

Lemuth, K., T. Hardiman, et al. (2008). "Global transcription and metabolic flux analysis of Escherichia coli in glucose-limited fed-batch cultivations." Appl Environ Microbiol **74**(22): 7002-7015.

Liu, X.-W. (2005). "Global convergence on an active set SQP for inequality constrained optimization." Journal of Computational and Applied Mathematics **180**(1): 201 - 211.

Mollney, M., W. Wiechert, et al. (1999). "Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments." Biotechnol Bioeng **66**(2): 86-103.

Nicolas, C., P. Kiefer, et al. (2007). "Response of the central metabolism of Escherichia coli to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase." FEBS Lett **581**(20): 3771-3776.

Perrenoud, A. and U. Sauer (2005). "Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli." J Bacteriol **187**(9): 3171-3179.

Peyraud, R., P. Kiefer, et al. (2009). "Demonstration of the ethylmalonyl-CoA pathway by using 13C metabolomics." Proc Natl Acad Sci U S A **106**(12): 4846-4851.

Quek, L. E., C. Wittmann, et al. (2009). "OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis." Microb Cell Fact **8**: 25.

Sauer, U., F. Canonaco, et al. (2004). "The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of Escherichia coli." J Biol Chem **279**(8): 6613-6619.

Schuster, R., S. Schuster, et al. (1992). "Simplification of complex kinetic models used for the quantitative analysis of nuclear magnetic resonance or radioactive tracer studies." J. Chem. Soc., Faraday Trans. **88**: 2837-2844.

Shimizu, K. (2004). "Metabolic flux analysis based on 13C-labeling experiments and integration of the information with gene and protein expression patterns." Adv Biochem Eng Biotechnol **91**: 1-49.

Spellucci, P. (1993). "A SQP Method For General Nonlinear Programs Using Only Equality Constrained Subproblems." Mathematical Programming **82**: 413-448.

Strisower, E. H., I. L. Chaikoff, et al. (1951). "Conversion of C14-palmitic acid to glucose. I. Normal and diabetic rats." J Biol Chem **192**(2): 453-463.

Tang, Y. J., H. G. Martin, et al. (2009). "Advances in analysis of microbial metabolic fluxes via (13)C isotopic labeling." Mass Spectrom Rev **28**(2): 362-375.

van Gulik, W. M., W. T. de Laat, et al. (2000). "Application of metabolic flux analysis for the identification of metabolic bottlenecks in the biosynthesis of penicillin-G." Biotechnol Bioeng **68**(6): 602-618.

Weinman, E. O., I. L. Chaikoff, et al. (1950). "Relative rates of conversion of the various carbon atoms of palmitic acid to carbon dioxide by the intact rat." J Biol Chem **184**(2): 735-744.

Weitzel, M., W. Wiechert, et al. (2007). "The topology of metabolic isotope labeling networks." BMC Bioinformatics **8**: 315.

Wiechert, W. (1994). Design of a software framework for flux determination by ^{13}C NMR isotope labelling experiments I. What is Controlling Life?, Innsbruck University Press. **3:** 305-310.

Wiechert, W., M. Möllney, et al. (1999). "Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems." Biotechnol Bioeng **66**(2): 69-85.

Wiechert, W., M. Mollney, et al. (2001). "A universal framework for 13C metabolic flux analysis." Metab Eng **3**(3): 265-283.

Zamboni, N., S. M. Fendt, et al. (2009). "(13)C-based metabolic flux analysis." Nat Protoc **4**(6): 878-892.

Zupke, C. and G. Stephanopoulos (1994). "Modeling of Isotope Distributions and Intracellular Fluxes in Metabolic Networks Using Atom Mapping Matrixes." Biotechnology Progress **10**(5): 489-498.

Part 2

Role of the Csr system in carbon nutrition and in control of central metabolism of *Escherichia coli* K12 MG1655 and Nissle 1917

This part of the thesis addresses fundamental questions regarding the physiology and metabolism of *Escherichia coli* strains, with three main objectives:

- Obtaining detailed metabolic knowledge on *E. coli* on physiologically-relevant carbon sources.
- Obtaining metabolic knowledge on B2 strains.
- Understanding the role of the Csr system in carbon nutrition and in control of *E. coli* metabolism.

Chapter 7 aims at the understanding on the role of the Csr system in control of the central metabolism of *E. coli* Nissle 1917, a highly competitive strain of the B2 group. First, we investigate the role of each component of the Csr system (CsrA, csrB, csrC, and CsrD) in nutrient utilisation by analysing the growth capabilities of different Csr mutants on a broad range of physiologically-relevant carbon sources. Then, we investigate the control of central metabolism exerted by Csr with detailed analyses of the metabolomes and fluxomes of mutants and wild-type cells grown on carbon sources representative of glycolysis and of the Entner-Doudoroff pathway (glucose and gluconate, respectively). This work demonstrates that i) Csr enhances the use of a wide range of carbon sources, ii) Csr controls a range of metabolic pathways wider than expected from its known target enzymes, and iii) the actual impact of the Csr system on the central metabolism of *E. coli* depends on the carbon source.

In **Chapter 8**, we extend the investigation on the role of the Csr system in the control of *E. coli* metabolism on carbon sources (gluconate, ribose, galactose, acetate) representative of the main glycolytic (Entner-Doudoroff pathway, pentose phosphate pathway, glycolysis) and gluconeogenic pathways of *E. coli*. This work is carried out on both the Nissle 1917 strain and the laboratory strain K12 MG1655 using high-throughput quantitative system-level tools (growth characterisation, ^{13}C-MFA, and metabolomics). First, this work provides fundamental knowledge on the actual operation of the metabolism of their wild type strains, and identifies metabolic specificities of the Nissle 1917 strain likely involved in its competitiveness in the gut. Then, analyses of mutant strains demonstrate that the Csr system controls not only carbon metabolism in *E. coli*, but also energy and redox metabolisms, and reinforce the potential role of the Csr system in adaptation to the gut environment.

These studies, which have required ^{13}C-MFA of different strains and mutants on various carbon sources, have been made possible thanks to the methods and tools developed in the first part of this thesis.

Chapter 7

The Csr system exerts global control over central metabolism in *Escherichia coli* strain Nissle 1917 in a nutrient-specific manner

NOTE: *this chapter contains the first draft of a manuscript in preparation for PLoS One. Olga Revelles (LISBP) has equally contributed to this work.*

1. Abstract

The role of the post-transcriptional carbon storage regulator (Csr) system in nutrient use and in the control of central metabolism in *E. coli* Nissle 1917 was investigated. Analysis of the growth capabilities of different Csr mutants showed that the protein CsrA - the key component of the system - increases the growth efficiency of a broad range of physiologically relevant carbon sources, including compounds utilized by the Entner-Doudoroff (ED) pathway. Detailed analyses of the metabolomes and fluxomes of mutants and wild-type cells grown on carbon sources representative of glycolysis and of the ED pathway (glucose and gluconate, respectively), revealed that CsrA exerts significant control over central carbon metabolism for both types of compounds. However, the metabolic reorganization actually observed in the CsrA mutant on gluconate was strikingly different from that observed on glucose, indicating different control mechanisms. This work demonstrates the actual impact of CsrA on the central metabolism of *E. coli* and suggests a significant role for the Csr system in adaptation to the gut environment.

2. Introduction

Escherichia coli is a normal inhabitant of the intestine and the predominant facultative anaerobe in the gastrointestinal tract of mammals (Todar 2007). The intestine is a highly complex and changing environment in which *E. coli* experiences constantly shifting growth conditions. Recent findings indicate that the ability to compete for carbon nutrition is a critical factor for gut colonization, and is part of the arsenal of strategies employed by pathogenic *E. coli* strains to outcompete the gut microbiotia (Miranda, Conway et al. 2004; Kamada, Kim et al. 2012). Colonization is mainly related to the utilization of sugars and sugar derivatives resulting from strain-dependent degradation of mucus and of dietary fibers (Chang, Smalley et al. 2004; Snider, Fabich et al. 2009). Glycolytic pathways such as the Embden-Meyerhof-Parnas (EMP) and Entner-Doudoroff (ED) pathways play an important role in colonization (Peekhaus and Conway 1998; Chang, Smalley et al. 2004; Leatham, Stevenson et al. 2005; Fabich, Jones et al. 2008).

Persistence of *E. coli* in the gut is supported by less favorable substrates, including both sugars and non-sugar compounds such as the small organic acids resulting from the degradation of mucus by anaerobes of the microflora (Chang, Smalley et al. 2004; Snider, Fabich et al. 2009). The use of the latter compounds requires activation of gluconeogenic pathways, and efficient switching between glycolytic and gluconeogenic carbon sources is likely to be a major feature of successful adaptation to life in the intestine (Pernestig, Georgellis et al. 2003).

To cope with the changing environment of the intestine, *E. coli* has developed a variety of adaptation mechanisms. Highly sophisticated global regulatory networks coordinate physiological and metabolic responses by controlling the functional expression of relevant sets of genes. The carbon storage regulator (Csr) system (Romeo, Gong et al. 1993) is a post-transcriptional regulation system that controls a broad range of physiological adaptative mechanisms (including formation of biofilm, motility, and virulence) and is a global regulator of central metabolism (Romeo, Vakulskas et al. 2012). The Csr system has four molecular components. The main component is the post-transcriptional regulator CsrA, an mRNA-binding protein that influences both translation and degradation of different mRNA targets (Mercante, Edwards et al. 2009). CsrA is negatively regulated by two small non-coding RNAs, csrB and csrC, which antagonize CsrA activity by sequestering the protein (Liu, Gui et al. 1997; Weilbacher, Suzuki et al. 2003). Lastly, the CsrD protein positively controls CsrA activity by driving the RNAs csrB and csrC to RNAse E degradation (Suzuki, Babitzke et al. 2006). Their role in the global control of metabolism suggests that this system plays a critical role in the adaptation of *E. coli* to changes in nutrient availability. The main component of the Csr system, the CsrA protein, is essential for growth on glucose – a glycolytic carbon source –, but not on a gluconeogenic compound like pyruvate (Timmermans and Van Melderen 2009). Biochemical studies have shown that CsrA exerts a positive control over several glycolytic enzymes (phosphoglucoisomerase, phosphofructokinase A, triose-phosphate isomerase, enolase, and pyruvate kinase F). In addition, it negatively controls gluconeogenic activities (fructose-1,6-biphosphatase, PEP synthase, PEP carboxykinase A), glycogen biosynthesis (glgABC) and the use of acetate (Liu, Yang et al. 1995; Sabnis, Yang et al. 1995; Wei, Shin et al. 2000). Because it regulates glycolysis and gluconeogenesis in opposite ways, the Csr system is assumed to play the role of master switch between glycolytic and gluconeogenic metabolisms (Sabnis, Yang et al. 1995; Pernestig,

Georgellis et al. 2003). Despite advances in our understanding of the molecular basis of Csr-associated regulation, its metabolic effects have not yet been fully investigated. The actual impact on the target metabolic pathways, in terms of metabolic pools and carbon fluxes, are not known. In addition, the different metabolic pathways are closely interconnected within the cell, and changes in the Csr-targeted pathways should lead to modifications in the activity of the entire central metabolism. More generally, the role of post-transcriptional regulation in the control of cellular metabolism is not yet fully understood.

In this work, we investigated the role of the Csr system in the control of carbon metabolism in *E. coli*. First, we analyzed the capability of mutants altered for the various components of the Csr system to grow on a range of physiologically relevant carbon sources. Next, we studied the actual impact of Csr on the entire central carbon metabolism by analyzing the metabolomes and fluxomes of selected Csr mutants on two carbon sources, glucose and gluconate, metabolized via the EMP and ED pathways, respectively. The study was carried out with the strain Nissle 1917, a non-pathogenic efficient colonizer of the gut (Grozdanov, Raasch et al. 2004), rather than with the classical domesticated *E. coli* such as K12. This was done to ensure that all the physiological and metabolic mechanisms necessary for gut colonization and survival were in place and had not been affected by the prolonged lab use of K12-like strains (Mira, Ochman et al. 2001; Hobman, Penn et al. 2007). Furthermore, Nissle 1917 belongs to the phylogenetic group B2, which is over-represented among *E. coli* strains persisting in the microbiota of humans from Europe, Australia, Japan and the USA (Bailey, Pinyon et al. 2010).

3. Experimental procedures

Bacterial strains and genetic methods

E. coli Nissle 1917 mutants were constructed using the allele replacement method described in (Datsenko and Wanner 2000). The primers used for the constructions are listed in annex (supplementary Table S1). Target genes were deleted and replaced with a chloramphenicol-resistance cassette. Strains containing multiple mutations were constructed by sequential allelic replacement; the first inserted cassette was removed with the FLP recombinase, followed by allelic replacement. The Nissle *ΔcsrA51* mutant

was kindly provided by Prof. U Dobrindt (University of Muenster, Germany). All constructions were verified by PCR. For complementation experiments, the *csrA* gene was amplified from chromosomal DNA and cloned into pGEM-T Easy (pCsrA). The resulting plasmid was transformed into *E. coli* Nissle *ΔcsrA51* by electroporation. *E. coli* Nissle 1917 and its mutant derivatives are listed in Table 1.

Cultivation conditions

Bacteria were grown at 37 ºC either in Lysogeny Broth (LB) or in synthetic minimal medium supplemented with 15 mM of carbon source, as described elsewhere (Nicolas, Kiefer et al. 2007).

Microreader plate and shake flask cultures

Detailed growth phenotyping of the strains was performed using multiplex measurement of online growth parameters based on fluorescent dyes sensitive to pH or oxygen and optical density, according to the method developed by (Heux, Philippe et al. 2011). Cells were grown in 48-well plates at 37 ºC in a volume of 500 µL of minimal medium, on a plate reader equipped with a shaker (Fluostar OPTIMA, BMG LABTECH).

For in-depth metabolic characterization, cells were cultivated in 500 mL baffled flasks (37 °C, 200 rpm) with 40 mL of medium. For ^{13}C-labelling experiments, the unlabeled substrate (glucose or gluconate) was replaced by a mixture of 80% [1-^{13}C]- and 20% [U-^{13}C]- substrate (Eurisotop, France). Inoculation was performed after centrifugation and washing with the same medium deprived of carbon source.

Glycogen staining & assay

To compare their glycogen content, the different strains, were grown overnight at 37 °C on LB plates containing 2% glucose and stained with iodine vapor (Chester 1968). Three independent experiments were carried out for each condition; the data shown are representative of one such experiment. To determine the rate of glycogen production upon exponential growth on glucose and on gluconate, glycogen content was quantified along the growth curve using the procedure of (Parrou and Francois 1997). Briefly, glycogen was extracted and hydrolyzed to glucose monomers, which were quantified using an enzymatic method. All experiments were performed in triplicate.

Motility assay

A colony from a fresh LB plate was streaked on LB 0.3% agar plate and incubated upright at 37 ºC overnight. To determine the extent of bacterial motility, the diameter of the diffusion halo was measured after 24 h. Motility assays were determined in three independent experiments.

Biofilm assay

Cultures were grown statically overnight in 500 µL of LB, in a 48-well polystyrene microtiter plate. Biofilm formation was measured by discarding the medium, rinsing the wells (three times) with water, and staining bound cells with 20% crystal violet in methanol (Hancock, Dahl et al. 2010). Absorbance was determined at 595 nm using a microtiter plate reader (Fluostar OPTIMA, BMG LABTECH). For each experiment, background staining was corrected by subtracting the crystal violet bound to an empty well.

Growth characterization

The growth rate (µ) was determined from log-linear regression of time-dependent changes in optical density at 600 nm (OD_{600}), which was detected with a spectrophotometer (Genesys 6, Thermo Electron Corporation, USA) with appropriate dilutions when needed. Extracellular fluxes were determined from the rates of disappearance (or appearance) of substrates and products in the culture supernatants. The different compounds were quantified by HPLC (Agilent Series 1100, USA on an Aminex HPX-87H column, Biorad, Hercules, CA) as previously described (Cocaign-Bousquet, Garrigues et al. 1995). To calculate specific biomass yields, correlation factors between cell dry weights and optical density (g_{CDW}/OD_{600}) were established for each strain.

Sampling of intracellular metabolites for metabolome analysis

Samples were taken at the mid-exponential phase using the differential method of (Taymaz-Nikerel, de Mey et al. 2009). Briefly, 120 µL of broth or filtered extracellular medium (Sartolon polyamid 0.2 µm, Sartorius, Goettingen, Germany) were plunged with 120 µL of fully ^{13}C-labeled cellular extract (used as internal standard) in 5 mL of an

ethanol/water (75/25) solution at 95 °C, incubated for 2 minutes, cooled on ice and stored at -80 °C. Three samples of broth and filtrate were taken and analyzed for each biological replicate.

Sampling of intracellular metabolites for ^{13}C-metabolic flux analysis

Sampling was performed at the mid-exponential phase in two steps; i) rapid quenching of metabolism followed by ii) metabolite extraction. For quenching, 120 µL of broth were rapidly sprayed into precooled centrifuge tubes maintained at -80 °C and containing 500 µL of cold ethanol, homogenized using a vortex and centrifuged (12 000 g for 5 min at -20 °C) with a Sigma 3-18K centrifuge (Sigma, Germany). Metabolites were extracted by pouring 5 mL of an ethanol/water (75/25) solution at 95 °C onto the cell pellets. After incubation for 2 min in closed tubes, the cellular extracts were cooled on ice and stored at -80 °C. For each biological replicate, three metabolite samples were collected and analyzed.

Preparation of cellular extracts and IC-MS(/MS) analysis of intracellular metabolites

Cellular extracts were evaporated for 4 h (SC110A SpeedVac Plus, ThermoSavant, USA). The remaining aqueous extracts were freeze-dried, resuspended in 200 µL of milliQ water, and stored at -80 °C. Intracellular metabolites were analyzed as previously described (Bolten, Kiefer et al. 2007; Kiefer, Nicolas et al. 2007). Briefly, analysis was performed by high performance anion exchange chromatography (Dionex ICS 2000 system, Sunnyvale, USA) coupled to a triple quadrupole QTrap 4000 (AB Sciex, CA) mass spectrometer. All samples were analyzed in the negative mode by multiple reaction monitoring (MRM). The injection volume was 15 µL, originating from approximately 2 µg of biomass. For metabolomics experiments, the amounts of metabolites of glycolysis (G6P, F6P, FBP, PEP, 1,3-diPG and combined pools of 2- and 3-PG), ED and pentose phosphate (PP) pathways (6PG, R5P, S7P), TCA cycle (Mal, Suc, Cit, Fum, Aco), as well as nucleotides (ADP, AMP, ATP, cAMP, CDP, CMP, CTP, dADP, dATP, dCDP, dGTP, dTDP, dTTP, GDP, GMP, UDP, UDP-Glucose, UMP and UTP) were determined. To ensure highly accurate quantification, the isotope dilution mass spectrometry (IDMS) method was used (Wu, Mashego et al. 2005). For fluxomics experiments, the ^{13}C-labeling patterns of central metabolites, including organic acids (Mal, Cit) and phosphorylated compounds

(G6P, F6P, FBP, PEP, 6PG, R5P, S7P, combined pools of 2- and 3-PG) were determined as described in (Kiefer, Nicolas et al. 2007). The labeling patterns (isotopologue distributions) were calculated from the isotopic clusters after correction for naturally occurring isotopes with IsoCor (Millard, Letisse et al. 2012).

NMR analysis of extracellular medium

The labeling patterns of the main metabolic products – e.g. acetate, pyruvate, etc – which accumulated during growth in the extracellular medium were measured by NMR spectroscopy (Massou, Nicolas et al. 2007; Massou, Nicolas et al. 2007). A total of 500 µL of broth was filtered and analyzed by quantitative ^{1}H 1D-NMR at 298 °K, using a 30° pulse and a relaxation delay of 20 s, with an Avance 500 MHz spectrometer (Bruker, Germany). The four acetate isotopomers and the eight pyruvate isotopomers were fully resolved and quantified.

Fluxome analysis

Fluxes were calculated using influx_s software (Sokol, Millard et al. 2011), in which both mass balances and carbon atom transitions describing the biochemical reaction network were implemented. The metabolic network contained the main pathways of *E. coli* central metabolism: glycolysis (EMP), pentose phosphate pathway (PPP), Entner Doudoroff (ED) pathway, the tricarboxylic acid cycle (TCA), and anaplerotic reactions. The precursors required for biomass synthesis were estimated according to biomass composition (Neidhardt, Curtiss et al. 1996) and measured growth rates. Intracellular fluxes were estimated from measurement of extracellular fluxes and from the ^{13}C-labelling patterns of metabolites using appropriate mathematical models of glucose or gluconate metabolisms in *E. coli* (Nicolas, Kiefer et al. 2007). Labeling data were collected from intracellular metabolites by IC-MS/MS and from metabolic end-products by 1D ^{1}H NMR, as detailed above. The fluxes were normalized to the rate of substrate uptake, which was arbitrarily set at 100. To determine the confidence intervals of the calculated fluxes, we used a Monte-Carlo approach where 100 independent optimization runs were performed from datasets in which noise was added according to the standard deviations of the measurements.

Calculation of ATP and redox balances

The adenylate energy charge (AEC) was calculated from the intracellular pools of adenosine mono-, di- and tri-phosphate according to (Atkinson 1968):

$$AEC=([ATP]+0.5\times[ADP])/([AMP]+[ADP]+[ATP])$$

NADH, NADPH and $FADH_2$: for each reduced cofactor, total production was calculated by summing the fluxes of all reactions which produced the reduced cofactor. The NADH that was synthesized in biosynthetic pathways, as well as NADPH requirements for anabolism, were estimated from biomass composition (Neidhardt, Curtiss et al. 1996) and measured growth rates. The remaining NADPH was assumed to be converted into NADH via transhydrogenase activities.

ATP: production of ATP via substrate-level phosphorylation was calculated by adding all the fluxes measured for all ATP-producing reactions in carbon metabolism, and subtracting the fluxes of all ATP-consuming reactions. The ATP produced from NADH and $FADH_2$ via oxidative phosphorylation was estimated assuming a maximum value for the P:O ratio of 1 for both redox cofactors (Feist, Henry et al. 2007). ATP requirements were calculated by summing the requirements for anabolism, which were estimated from biomass composition, measured growth rates, and maintenance needs (including both growth associated and non-growth associated maintenance (Feist, Henry et al. 2007)).

Statistical analysis

Standard deviations of metabolic fluxes and intracellular pool sizes were determined from three biological replicates. Statistical differences in metabolite concentrations between the wild type and mutant strains were calculated using Student's paired *t*-tests with two tailed distributions.

4. Results

Construction and phenotypic characterization of Csr mutants

To analyze the role of the different components of the Csr system, we first constructed null mutations in the *csrB, csrC, csrBC* and *csrD* genes of the *E. coli* strain Nissle 1917 by allelic exchange (Table 1, primers used in that study are listed in Table S1).

Due to the fact that the *csrA* gene is essential for growth on LB and glycolytic carbon sources (Timmermans and Van Melderen 2009), all attempts to generate a *csrA* deletion mutant on glucose failed. A mutant of the laboratory *E. coli* strain MG1655 strain encoding a truncated CsrA protein was shown to be viable on glycolytic compounds, but was significantly impaired for processes controlled by CsrA (Timmermans and Van Melderen 2009).

Strains	Relevant genotype	Source
Wild-type	Wild-type *E. coli* strain Nissle 1917	This study
Δ*csrC*	*E. coli* Nissle 1917 *csrC*::Cm^R	This study
Δ*csrB*	*E. coli* Nissle 1917 *csrB*::Tc^R	This study
Δ*csrBC*	*E. coli* Nissle 1917 *csrC csrB*::Tc^R	This study
Δ*csrD*	*E. coli* Nissle 1917 *csrC*::Cm^R	This study
Δ*csrA51*	*E. coli* Nissle 1917 *csrA51*::Km^R	Prof. Dobrindt
Δ*csrA51* pCsrA	*E. coli* Nissle 1917 *csrA51*::Km^R pGEM-T *csrA*	This study

Table 1. *Escherichia coli* strains used in this study.

A similar mutant (*ΔcsrA51*) was successfully generated in the Nissle strain to analyze the effects of attenuated CsrA activity. Mutants of the Csr system are expected to be altered for glycogen storage, biofilm formation, and motility (Romeo, Gong et al. 1993; Wei, Brun-Zinkernagel et al. 2001; Wang, Dubey et al. 2005). The different mutants obtained in this work were therefore phenotyped against these criteria. The glycogen content of the Csr mutants was measured by iodine staining. The *ΔcsrB, ΔcsrC, ΔcsrBC*, and *ΔcsrD* mutants accumulated amounts of glycogen similar or slightly lower than the wild-type strain. In contrast, the *ΔcsrA51* mutant accumulated substantially more glycogen than all other strains (Table 2). Biofilm formation was estimated by measuring the crystal violet staining of the cells attached to the microtiter plate wells. The *ΔcsrA51* mutant formed dense biofilms (at twice the ratio of the wild type) while all other strains showed the same density as the wild type. Growth measurements on semi-solid tryptone agar plate revealed that the motility of *ΔcsrA51* was reduced whereas that of the other mutants was not. All these results showed that the phenotype of the Nissle *ΔcsrA51* is in close agreement with the phenotype described in the laboratory strain MG1655, and with expectations for the other Csr mutants.

Strains	Glycogen	Motility	Biofilm
wild-type	++	++	+
Δ*csrA51*	++++	+	+++
Δ*csrB*	+	++	+
Δ*csrC*	+	+++	+
Δ*csrBC*	+	+++	++
Δ*csrD*	++	+	+

Table 2. Phenotypic characterization of Nissle 1917 and its isogenic Csr mutants. Each experiment was performed with three independent biological replicates. The data shown here are semi-quantitative.

Metabolic capabilities of *E. coli* Nissle and its Csr mutants on a variety of carbon substrates

To determine the metabolic capabilities of *E. coli* Nissle 1917 wild-type strain and its Csr mutants, a large-scale physiological analysis was performed in which growth phenotypes were monitored on 15 carbon sources found in the gastrointestinal tract (Chang, Smalley et al. 2004; Fabich, Jones et al. 2008) (Figure 1). Detailed information on growth was obtained using a multiplex monitoring system which allows the simultaneous follow-up of growth, pH and oxygen profiles in plates (Heux, Philippe et al. 2011). The growth phenotypes of the *ΔcsrB, ΔcsrC, ΔcsrBC* and *ΔcsrD* mutants were very similar to that of the wild-type strain (Figure 1). In contrast, the growth of the *ΔcsrA51* mutant was significantly altered. This mutant showed longer lag phases for all carbon sources tested (Figure 1A). The growth rates were also altered in a compound-dependent manner (Figure 1B). As expected, growth of the *ΔcsrA51* mutant was significantly reduced on most carbon sources utilized via the EMP pathway (glucose, mannose, lactose, N-acetyl-glucosamine or maltose), although no significant difference was observed for fructose and fucose. In agreement with the positive control exerted by CsrA on acetyl-CoA synthetase (Wei, Shin et al. 2000), the *ΔcsrA51* mutant grew poorly on acetate.

Furthermore, the gastrointestinal tract contains compounds that are used by the Entner-Doudoroff (ED) – i.e. gluconate, glucuronate - and Pentose-Phosphate (PP) – i.e. xylose, ribose - pathways. Although all strains grew similarly on xylose, growth on ribose was reduced by 50% in the *ΔcsrA51* mutant. This was unexpected as the Csr system is not assumed to control the PP pathway (Romeo 1998). The role of the Csr system in the use of ED-utilized compounds has not been investigated so far. Our data show that the growth of the *ΔcsrA51* mutant on compounds degraded via this pathway was impaired. The growth rate was reduced by 10% and 20% on galacturonate and glucuronate,

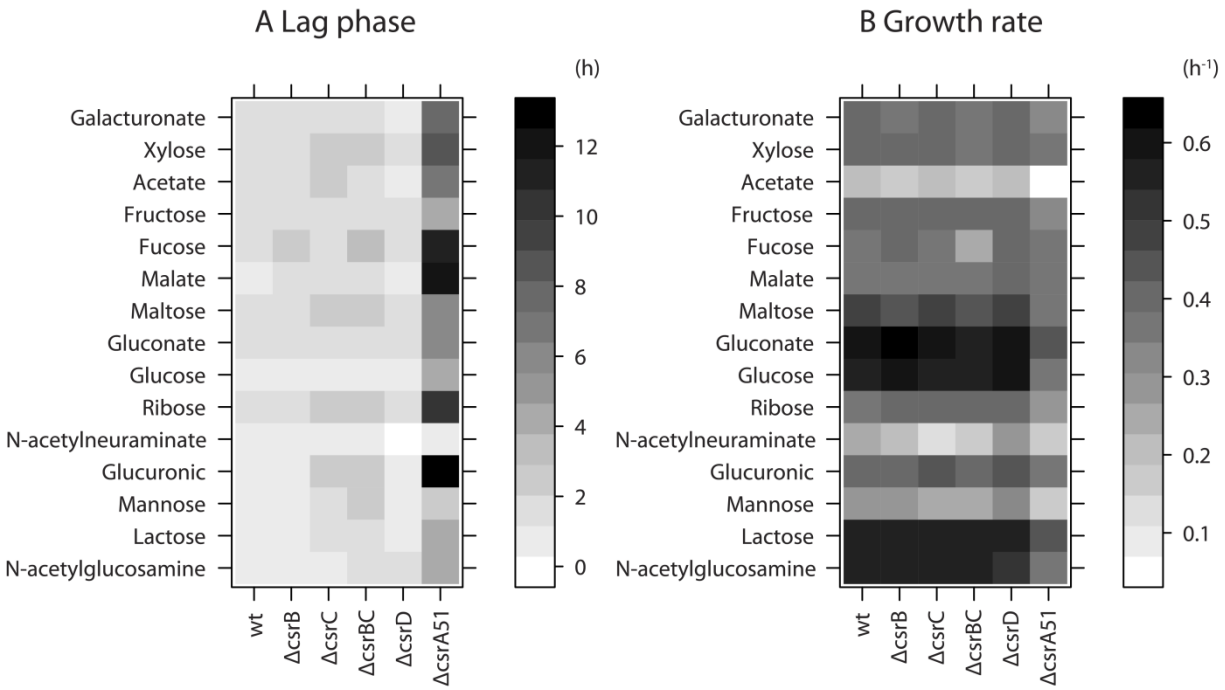


Figure 1. Growth phenotypes of *E. coli* Nissle 1917 and its Csr mutants. Heat maps representing the lag times (A) and growth rates (B) upon growth on 15 carbon sources representative of the gut environment. Rows represent different carbon sources and columns represent strains. The gray scale indicates high (darker) and low (paler) growth rate.

respectively, and by 30% on gluconate, compared to the wild-type strain. This indicates that the Csr system is involved in the utilization of the three compounds degraded by the ED pathway.

Growth parameters of *E. coli* Nissle and its Csr mutants on glucose and gluconate

Physiological analysis of the *ΔcsrBC* and *ΔcsrA51* mutants was performed on growth on glucose and gluconate, as valuable representatives of nutrients utilized via the EMP and ED pathways (Table 3). As control, the *ΔcsrA51* mutant was complemented with a plasmid carrying the intact *csrA* gene (pCsrA). The complemented mutant displayed physiological parameters similar to those of the wild-type strain (Table 3).

The wild-type Nissle strain grew on glucose at a rate of 0.79 h^{-1} with a glucose uptake rate of 12.5 $mmol.g^{-1}.h^{-1}$. The growth parameters of the *ΔcsrBC* mutant were very similar to those of the wild-type strain, despite its putative inability to regulate CsrA. In agreement with the data reported above, the physiology of the mutant *ΔcsrA51* was significantly altered, showing a longer lag phase, lower growth rate and lower substrate

Strain	Glucose				Gluconate					
	μ (h^{-1})	q_{Glc} (-)	v_{Ac}	Y_{Ac}	μ (h^{-1})	q_{Gnt} (-)	v_{Ac}	Y_{Ac}	v_{Pyr}	Y_{Pyr}
Wild-type	0.79 ± 0.02	12.5 ± 1.0	6.25 ± 0.09	0.50 ± 0.02	0.74 ± 0.02	20.0 ± 1.7	10.4 ± 0.1	0.52 ± 0.03	3.2 ± 0.1	0.16 ± 0.02
Δ*csrBC*	0.75 ± 0.03	12.0 ±0.3	5.16 ± 0.19	0.43 ± 0.07	0.76 ± 0.01	20.4 ± 0.9	11.0 ± 0.1	0.54 ± 0.07	2.8 ± 0.1	0.14 ± 0.01
Δ*csrA51*	0.63 ± 0.02	10.8 ± 0.7	7.24 ± 0.09	0.67 ± 0.05	0.57 ± 0.01	12.6 ± 1.3	6.3 ± 0.1	0.50 ± 0.05	0.4 ± 0.3	0.03 ± 0.01
Δ*csrA51* pCsrA	0.72 ± 0.02	12.9 ± 0.5	6.06 ± 0.11	0.47 ± 0.05	0.74 ± 0.04	19.8 ± 2.0	9.3 ± 0.1	0.47 ± 0.03	4.5 ± 0.1	0.23 ± 0.01

Table 3. Growth parameters of Nissle 1917 and Csr mutants with glucose and gluconate as carbon source: μ (h^{-1}), specific growth rate; q (mmol.g^{-1}.h^{-1}), carbon source uptake rate; v_{Ac} (mmol.g^{-1}.h^{-1}), acetate production rate; Y_{Ac}, acetate production yield ($= v_{Ac}/q$); v_{Pyr} (mmol.g^{-1}.h^{-1}), pyruvate production rate; Y_{Pyr}, pyruvate production yield ($= v_{Pyr}/q$). Values represent the mean ± standard deviation of three independent biological replicates.

uptake compared to its parental strain. The conversion of glucose into acetate in the *ΔcsrA51* mutant was significantly higher than in the wild-type strain (67% vs. 50%), indicating less efficient use of the sugar.

Gluconate is one of the most abundant carbon sources in the intestine of mammals and is assumed to play a critical role in the colonization of the large intestine by *E. coli* (Peekhaus and Conway 1998; Chang, Smalley et al. 2004). Because the growth of the *ΔcsrA51* mutant was impaired on gluconate (Figure 1), and due to the importance of this carbon source in colon colonization, we investigated the metabolome of *E. coli* Nissle 1917 wild-type strain and its *csr* mutants on gluconate. The wild-type strain grew on gluconate at a rate similar to that observed on glucose (0.74 h^{-1} vs. 0.79 h^{-1}), although much more gluconate than glucose (20.0 vs. 12.5 mmol.g^{-1}.h^{-1}) was needed to achieve the same growth rate, indicating different metabolic efficiencies. Gluconate was converted into two by-products, pyruvate and acetate, with molar yields of 0.16 and 0.52, respectively. Like on glucose, the growth of the *ΔcsrBC* mutant did not significantly differ from that of the wild-type strain, whereas the mutant *ΔcsrA51* showed marked differences (Table 3): (i) the growth rate of the *ΔcsrA51* mutant was 20% lower than that of the two other strains, (ii) the lower growth rate was correlated with a 40% reduction in gluconate consumption, (iii) acetate production was reduced but acetate yield remained unchanged, and (iv) the production of pyruvate was drastically reduced and was almost negligible in the *ΔcsrA51* mutant.

Metabolic profiles of *E. coli* strains grown on glucose

To increase our understanding of the role of the Csr system in glucose metabolism, the metabolomes of the *ΔcsrA51* and *ΔcsrBC* mutants were compared to those of the wild-type strain. Absolute quantification of central and energetic metabolites was obtained from cells growing exponentially in minimal medium with glucose as the sole source of carbon. The results of the most representative metabolites are listed in table 4, while more details are given in Table S2 in supplementary materials.

	Concentration (µmol/g CDW)		
	wild-type	***ΔcsrBC***	***ΔcsrA51***
1,3-diPG[a]	2.28±1.06	3.89±3.76	1.90±0.15
2/3-PG	1.42±0.11	1.67±0.52	1.12±0.23
6PG	0.38±0.09	0.38±0.07	0.77±0.20
ADP	0.71±0.06	0.82±0.25	0.91±0.27
AMP	0.16±0.03	0.11±0.09	0.21±0.06
ATP	3.50±0.29	3.21±0.03	2.86±0.48
cAMP	0.01±0.02	0.00±0.02	0.01±0.01
Cit	1.01±0.90	0.81±0.33	0.40±0.14
F1P	0.03±0.01	0.04±0.01	0.02±0.01
F6P	0.44±0.03	0.41±0.03	0.90±0.08
FBP	2.01±0.55	2.30±0.12	1.44±0.32
Fum	0.59±0.34	0.56±0.16	0.44±0.09
G6P	1.78±0.19	1.93±0.61	2.00±0.40
M6P	0.26±0.04	0.26±0.04	0.56±0.17
Mal	1.82±0.45	0.93±0.30	0.95±0.30
PEP	0.19±0.05	0.21±0.08	0.21±0.05
R1P	0.02±0.01	0.01±0.001	0.01±0.01
R5P	0.75±0.06	0.73±0.28	0.71±0.15
S7P[a]	4.63±0.31	4.17±1.19	12.8±1.5
Suc	4.44±0.27	5.34±1.52	1.98±0.46
UDP-Glucose	4.65±0.14	4.86±0.47	5.20±0.47

[a]For these metabolites only the $^{12}C/^{13}C$ ratio is given. This ratio represents the concentration of the metabolite in samples relative to the IDMS standard, which added at the same amounts in all samples.

Table 4. Metabolite concentrations in the wild-type *E. coli* Nissle 1917, *ΔcsrBC* and *ΔcsrA51* mutants growing on glucose.

The metabolite content of the *ΔcsrBC* mutant generally did not significantly differ from that of the wild-type strain. In contrast, the mutant with a truncated CsrA protein underwent a dramatic increase in the levels of metabolites from the upper part of the EMP and PP pathways. The levels of F6P, 6PG and S7P, were two to three times higher than in the wild type. Interestingly, a significant drop in the intracellular levels of FBP was observed in the *ΔcsrA51* strain, indicating a reduction in phosphofructokinase activity, a known target of CsrA. The level of M6P in the *ΔcsrA51* mutant was twice as

high as in the wild type. No changes were detected in metabolites of the lower part of glycolysis, such as 2- and 3-PG and PEP. Surprisingly, the pools of metabolites related to the TCA cycle (Fum, Mal and Suc) were lower than in the wild type.

The adenylate energy charge (AEC) was measured from the ATP, ADP and AMP contents listed in Table S3. No significant differences were observed between the wild type and the mutants, providing evidence that all strains were able to maintain a stable energetic status.

Metabolic Flux Responses in the Csr mutants upon growth on glucose

To better understand the effect of the Csr system on central metabolism, the distribution of metabolic fluxes was analyzed in the wild-type strains and in *ΔcsrBC* and *ΔcsrA51* mutants. For this purpose, steady-state ^{13}C-labeling experiments were carried out in which the strains were fed with a mixture of 80% 1-^{13}C and 20% U-^{13}C-glucose as sole carbon source.

We first wondered about the flux toward glycogen production as this is assumed to be responsible for the growth deficiency of a CsrA deletion mutant (Timmermans and Van Melderen 2009). The rate of glycogen production and glycogen accumulation was calculated during the exponential phase of growth on glucose. The flux values obtained for glycogen synthesis in both the wild type and in the *ΔcsrA51* mutant were in the range of µmoles.g^{-1}.h^{-1} (Table 5). This is 3 to 4 orders of magnitude lower than the rates of substrate uptake (in the range of 10 to 20 mmol.g^{-1}.h^{-1}) or the fluxes observed in the main metabolic pathways.

Strains	**Glycogen content** (nmolGlc/OD_{600})		**Glycogen production rate** (mmol.g^{-1}.h^{-1})	
	Glucose	Gluconate	Glucose	Gluconate
wild-type	1.55 ± 0.15	1.93 ± 0.16	0.003 ± 0.001	0.001 ± 0.001
Δ*csrA51*	22.20 ± 0.05	4.01 ± 0.83	0.057 ± 0.014	0.007 ± 0.001

Table 5. Glycogen contents and glycogen production rates in exponentially-growing cells of *E. coli* Nissle 1917 and its *ΔcsrA51* mutant during growth on glucose and gluconate.

Thus, the flux of carbon diverted towards glycogen formation in the *ΔcsrA51* mutant is much too low to significantly alter the overall carbon flux, and is likely not the main reason for the flux redistribution observed in exponentially growing cells for this mutant.

However, as the *ΔcsrA51* mutant retains partial CsrA activity, it cannot be excluded that the effect on glycogen metabolism in this mutant is not as significant as in a mutant that has lost complete activity.

So far, no investigations have been undertaken of flux distribution in *E. coli* strains belonging to the phylogenetic B2 group. For this reason, the flux data obtained for the wild-type Nissle 1917 strain (Figure 2) were first compared to those of the K12 laboratory strain MG1655 (Nicolas, Kiefer et al. 2007).

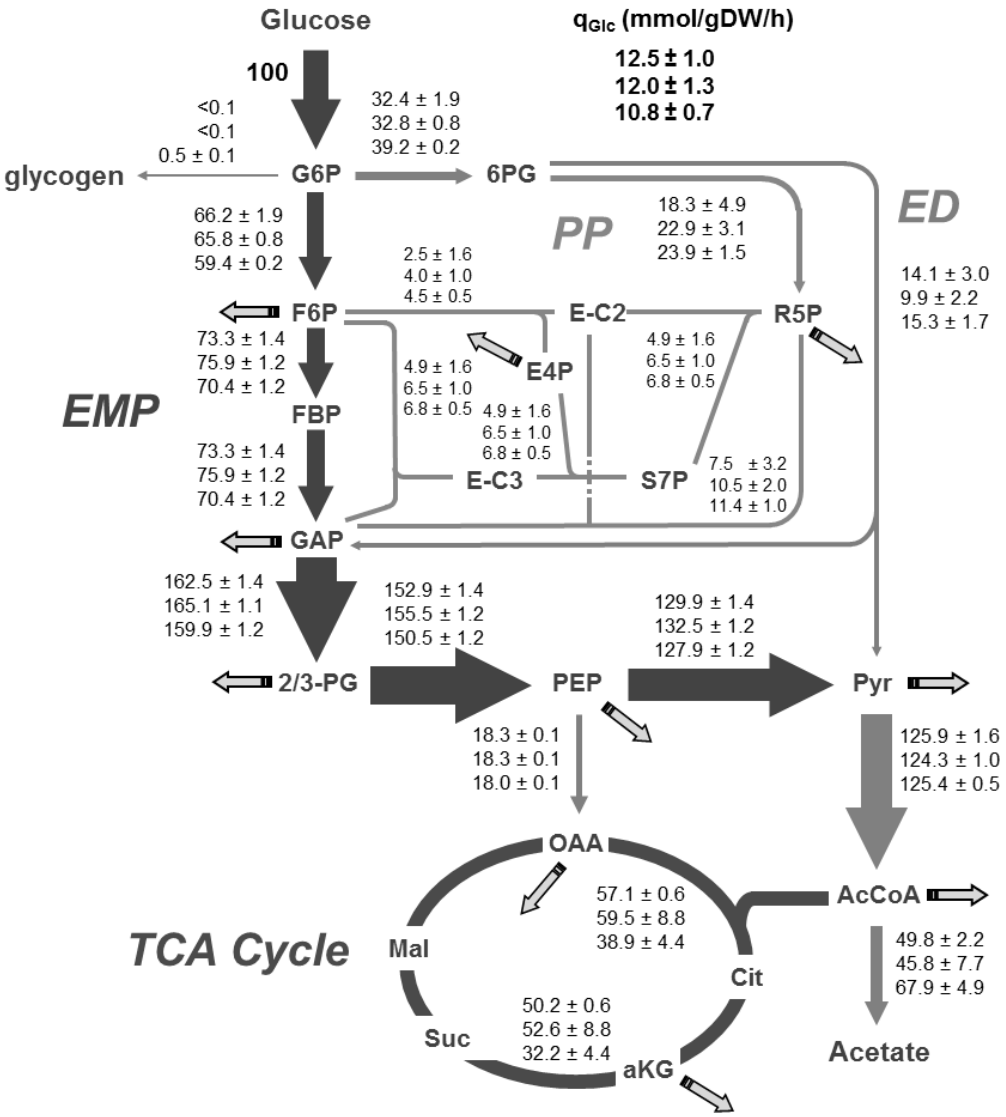


Figure 2. Metabolic flux distributions in *E. coli* Nissle 1917 wild-type strain and its derivates Δ*csrBC* and Δ*csrA51* mutants. Values from top to bottom: wild-type strain, Δ*csrBC* and Δ*csrA51* in glucose as carbon source. Flux values are normalized to the specific glucose uptake rate of each strain, which was arbitrarily given the value of 100. Values represent the mean of at least three biological replicates for each strain. Arrow thicknesses represent the flux values in the wild type strain.

In K12 strains, glucose is predominantly catabolized via the EMP pathway, with a 20% contribution by the PP pathway, while the ED pathway displays negligible activity. In the Nissle 1917 strain, the flux in the PP pathway is similar to that in the K12 strains, but the ED pathway contributes significantly (14%) to glucose catabolism. The total flux through the two dehydrogenating pathways (PP+ED) is significantly higher than that reported in K12 strains (32% vs. 20%) These data are consistent with the higher growth rate of the former strain on glucose compared to K12 strains, since these pathways produce the NADPH required for anabolic needs. The contribution of the TCA cycle to NADPH production in the Nissle 1917 strain was higher than that observed in K12 strains under similar conditions (50% vs. 25% of glucose uptake, respectively). Our data suggest that total NADPH production in the carbon pathways is sufficient to fulfill all growth requirements in the Nissle 1917 strain (Figure 3). This is not the case in K12 strains, in which transhydrogenase activity is necessary to fulfill NADPH requirements (Sauer, Canonaco et al. 2004). Such a difference between the two types of strains could partly explain the lower growth efficiency of K12 strains.

The flux distribution in the *ΔcsrBC* mutant (Figure 2) did not differ significantly from that in the wild-type strain, although the split ratio between the PP and the ED pathway was slightly higher. The flux distribution in the *ΔcsrA51* mutant showed more differences. The relative flux through the EMP in the mutant was lower than in the wild-type. As mentioned above, this was not due to increased flux towards glycogen synthesis, but to a redirection of the carbon flow towards the PP pathway. Indeed, the EMP/PP split ratio was decreased from 2 in the wild-type to 1.5 in the *ΔcsrA51* mutant. The major differences between the Δ*csrA51* mutant and the wild-type strain were strikingly lower fluxes via the TCA cycle for the mutant than in the wild type. The fluxes carried out by citrate synthase and succinate dehydrogenase were 30% and almost 40% lower, respectively. The decrease in the TCA cycle in the Δ*csrA51* mutant was accompanied by a 40% increase in acetate production. These differences were shown to be statistically significant ($p<0.01$) by the sensitivity analysis applied to each flux data set, and reproducible in three biological replicates.

To further evaluate the impact of the Csr mutations on *E. coli* Nissle 1917 metabolism, the production of high-energy molecules NADPH and ATP was calculated from the ^{13}C flux data. In the wild-type strain and the *ΔcsrBC* mutant, the PP pathway and the TCA

cycle contributed almost equally to the production of NADPH. The contribution of the TCA cycle was significantly (around 30%) lower in the Δ*csrA51* mutant and was compensated for by the PP pathway. In all strains, the production of NADPH was sufficient to fulfill the anabolic needs (Figure 3). To establish the ATP balance, the ATP requirements for growth and maintenance energy were calculated according to (Feist, Henry et al. 2007), and the ATP production was calculated in the different strains. In all strains, total ATP production was higher than the predicted requirements (Figure 3). The absolute production of ATP in the Δ*csrA51* mutant was 20% lower than in the two other strains, which was consistent with the lower growth rate. Substrate-level phosphorylation and oxidative phosphorylation represented 30% and 70% of total ATP production, respectively. This percentage remained remarkably constant in all strains, including in the Δ*csrA51* mutant.

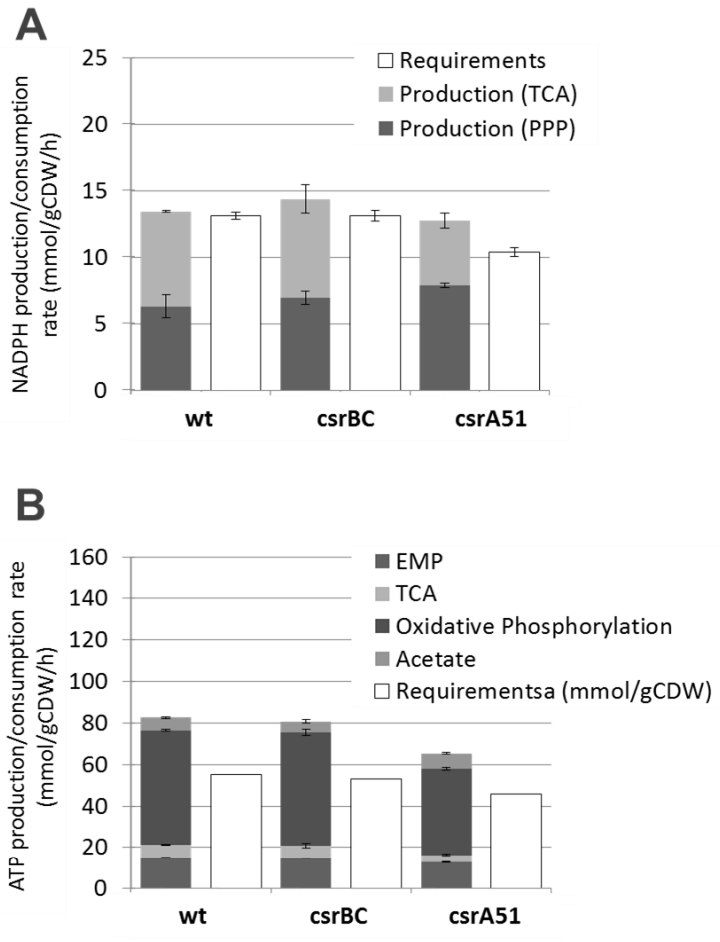


Figure 3. NADPH (A) and ATP (B) balances in *E. coli* Nissle 1917 and Δ*csrA51* mutant upon growth on glucose. Productions of cofactors were calculated from carbon fluxes. The requirements were calculated on a theoretical basis from biomass composition and growth rates.

Metabolic profiles of cells grown on gluconate

Like on glucose, on gluconate, there were no significant differences between the metabolome of *ΔcsrBC* and the wild-type strain. In contrast, significant differences ($p < 0.01$) were observed in the *ΔcsrA51* mutant in the upper part of carbohydrate metabolism (Table 6 & Table S2).

	Concentration (µmol/ g CDW)		
	wild-type	***ΔcsrBC***	***ΔcsrA51***
1,3-diPG[a]	4.45±0.56	3.30±3.00	3.68±0.73
2/3-PG	4.20±0.33	4.66±0.45	2.66±0.20
6PG	3.59±0.21	2.33±0.55	0.65±0.79
ADP	0.95±0.28	1.01±0.35	0.80±0.19
AMP	0.20±0.05	0.40±0.10	0.17±0.02
ATP	3.51±0.15	4.08±0.15	2.95±0.24
cAMP	0.01±0.01	0.01±0.01	0.01±0.03
Cit	0.27±0.10	0.35±0.25	0.57±1.05
F1P	0.19±0.11	0.21±0.05	0.14±0.05
F6P	0.26±0.02	0.19±0.03	0.19±0.01
FBP	2.95±0.47	2.51±0.14	1.82±0.09
Fum	0.75±0.26	0.72±0.10	0.63±0.15
G6P	0.97±0.70	0.48±0.05	0.39±0.05
M6P	0.05±0.03	0.12±0.01	0.15±0.01
Mal	1.64±0.57	2.29±0.29	2.35±0.67
PEP	0.49±0.05	0.54±0.12	0.49±0.11
R1P	7.81±0.81	42.15±9.23	17.85±4.33
R5P	1.93±0.45	2.60±0.16	3.38±0.41
S7P[a]	19.9±1.5	27.8±2.5	52.9±2.7
Suc	3.34±0.38	5.58±0.85	6.69±0.56
UDP-Glucose	2.86±0.31	3.79±0.57	2.14±0.56

[a]For these metabolites only the $^{12}C/^{13}C$ ratio is given. This ratio represents the concentration of the metabolite in samples relative to the IDMS standard, which added at the same amounts in all samples.

Table 6. Metabolite concentrations in the wild-type *E. coli* Nissle 1917, *csrBC* and *ΔcsrA51* mutants growing on gluconate.

A drastic 5-fold reduction was observed in the intracellular level of 6PG, the first intermediary of gluconate utilization. Conversely, the intracellular levels of the PP pathway metabolites R5P, R1P and S7P, were 2 to 3-fold higher in the *ΔcsrA51* strain. The EMP metabolites in this mutant were 2 times lower than in the wild type, including F6P, FBP and 2- and 3-PG. The levels of the TCA cycle metabolites Suc and Cit increased significantly, while the concentration of Fum remained almost stable. The level of M6P increased dramatically. As observed on glucose, the AEC was similar in all the strains analyzed, indicating a stable energetic status.

Metabolic Flux Responses in Csr mutants on growth on gluconate

The metabolic flux distribution in cells grown on gluconate was determined from steady-state ^{13}C-labeling experiments (Figure 4) with ^{13}C- a mixture of 80% [1-^{13}C]- and 20% [U-^{13}C]-gluconate as sole carbon source, similarly to the experiments carried out with labeled glucose. Here again, we found that the flux from glycolysis to glycogen was negligible in all three strains growing on gluconate.

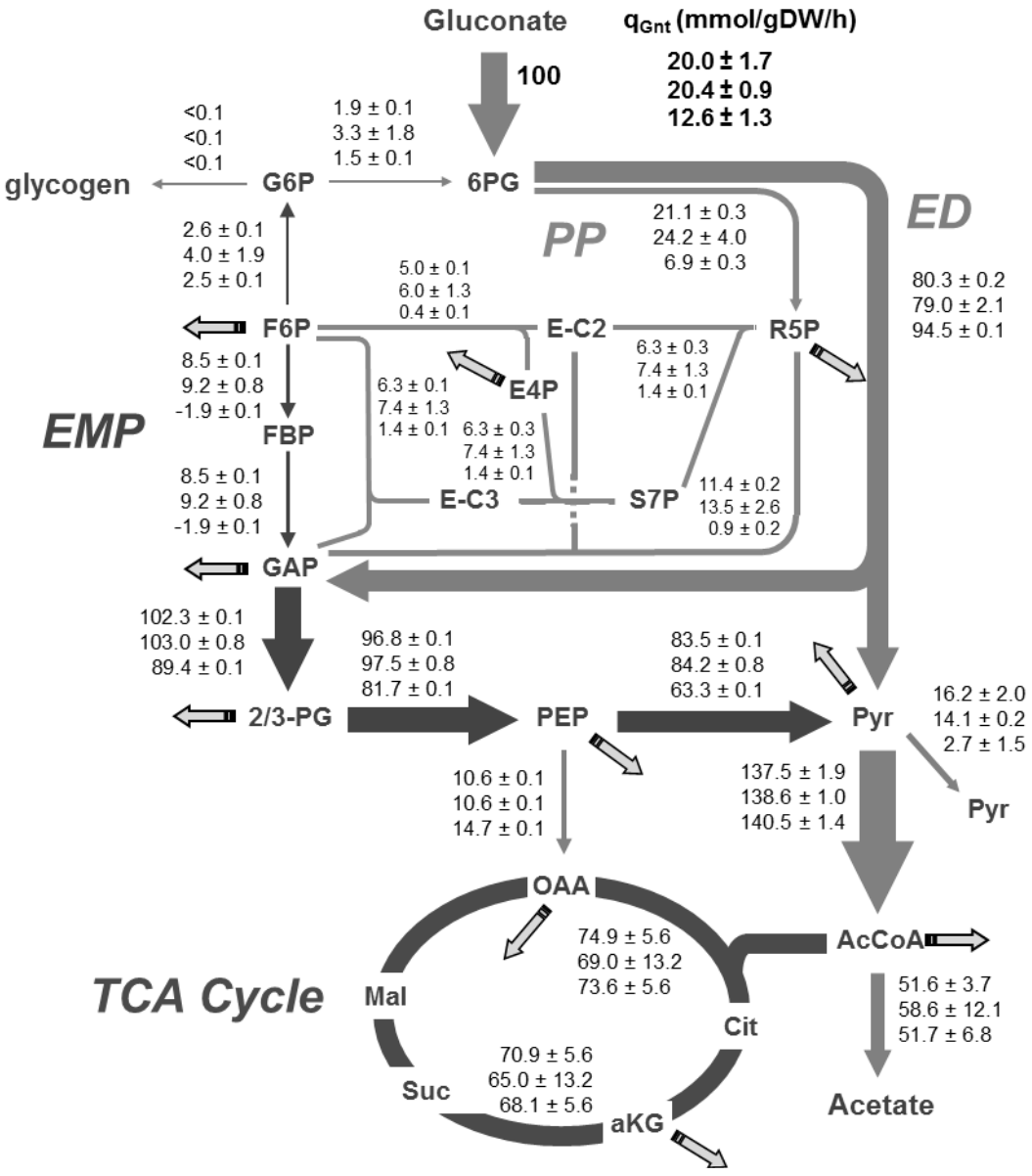


Figure 4. Metabolic flux distributions in *E. coli* Nissle 1917 wild-type strain and its derivates Δ*csrBC* and Δ*csrA51* mutants. Values from top to bottom: wild-type strain, Δ*csrBC* and Δ*csrA51* in gluconate as carbon source. Flux values are normalized to the specific gluconate uptake rate of each strain, which was arbitrarily given the value of 100. Values represent the mean of at least three biological replicates for each strain. Arrow thicknesses represent the flux values in the wild type strain.

In the wild-type Nissle 1917 strain, gluconate was mainly used via the ED pathway (80% of gluconate uptake). A significant proportion was also used via the PP pathway, and the PP/ED flux ratio was 20:80. A small proportion of hexose 6-phosphate was also recycled in the ED and PP pathways. The conversion of GAP into pyruvate in the lower part of glycolysis was much lower than observed on glucose (80% and 130% of substrate uptake, respectively). However, on gluconate, pyruvate is also generated via the ED pathway. The total flux of pyruvate formation by the two EMP & ED glycolytic pathways was higher on gluconate than on glucose (164% vs. 144% of substrate uptake, respectively). Flux data indicated that 10% of the pyruvate produced was released into the medium as a fermentation by-product on gluconate, and that 84% was converted into acetyl-CoA via pyruvate dehydrogenase. The TCA cycle used 54% of acetyl-CoA, while 38% was released as acetate. From flux data and the data listed in Table 3, it was possible to calculate that 35% of the gluconate molecules entering the cells were fully oxidized to CO_2 (via the combined action of the glycolytic pathways and the TCA cycle), while 34% were converted into by-products (pyruvate & acetate). This is significantly different from what happened on glucose, where 25% of the molecules were fully oxidized by the combined action of glycolytic pathways + the TCA cycle, and 25% were released as a by-product (acetate).

Here again, the relative flux distribution in the *ΔcsrBC* mutant was no different from that in the wild-type strain. Not surprisingly, significant differences were observed in the *ΔcsrA51* mutant. In this strain, the PP flux (i.e. flux through 6-phosphogluconate dehydrogenase) fell from 21% to 7% of the substrate uptake rate. While the ED flux increased from 80% in the wild-type to 95% in *ΔcsrA51*, indicating almost pure ED metabolism. The formation of hexose 6-phosphate via the PP pathway stopped. Instead, hexose phosphates were obtained by recycling of the triose phosphate generated by the ED pathway, as shown by the negative flux in the series of reactions from F6P to GAP. Such recycling of triose phosphate has been already observed in ED-using species (Portais, Tavernier et al. 1999; Fuhrer, Fischer et al. 2005).

The flux in the lower part of glycolysis in the *ΔcsrA51* mutant was reduced compared to that in the wild-type strain. The reduced formation of pyruvate by lower glycolysis was compensated for by an increase in direct production of the latter compound via the ED pathway. Indeed, the total pyruvate flux was almost the same in the two strains (158%

vs. 164% of substrate uptake, respectively). However, pyruvate use differed significantly in the two strains, since the excretion of pyruvate was almost abolished in the *ΔcsrA51* mutant. Unlike growth on glucose, the relative fluxes through the TCA cycle and acetate production were not altered in the *ΔcsrA51* mutant for growth on gluconate. These data showed that the reduced activity of the CsrA protein in this mutant did not modify the fraction of gluconate molecules that were fully oxidized to CO_2 (34% vs. 35% in the wild-type strain).

The NADPH and ATP balances in the different strains were calculated for growth on gluconate (Figure 5). Compared to glucose metabolism, whatever the strain, the production of NADPH for growth on gluconate relied more on the TCA cycle than on the PP pathway. This phenomenon was amplified in the *ΔcsrA51* mutant, in which 90% of NADPH was produced in the TCA cycle, compared with 76% in the wild-type strain. In that respect, the effect of the *csrA51* mutation was the reverse of the situation observed for growth on glucose, when the contribution of the TCA cycle decreased.

The ATP balance for growth on gluconate was determined in the wild-type, *ΔcsrA51*, and *ΔcsrBC* strains (Figure 5). As observed on glucose, total ATP production calculated from flux data was consistently higher than predicted requirements. The maximum amount of ATP produced on gluconate in the wild-type strain was significantly higher than on glucose (138 vs. 82.6 mmol/gCDW). This is consistent with the fact that much more gluconate than glucose is consumed to obtain similar growth rates. As expected, there were no significant differences between the *ΔcsrBC* mutant and the wild-type strain. In the *ΔcsrA51* mutant, ATP production was drastically reduced (77.7 mmol/gCDW). Nevertheless, the respective contributions of substrate-level phosphorylation and oxidative phosphorylation to ATP production were constant in all three strains, as observed on glucose.

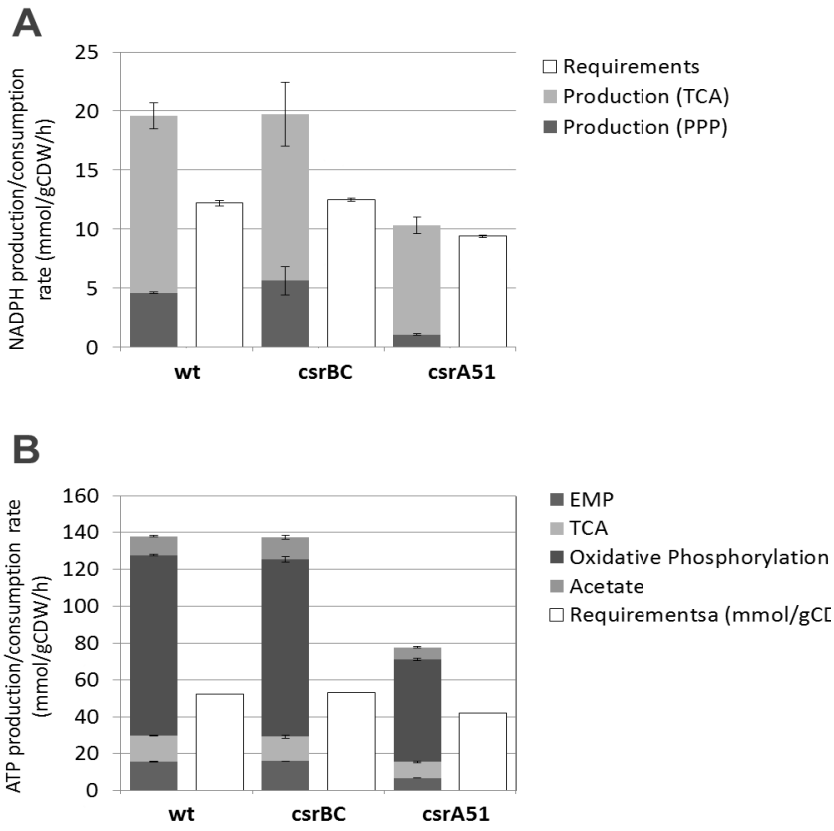


Figure 5. NADPH (A) and ATP (B) balances in *E. coli* Nissle 1917 and Δ*csrA51* mutant upon growth on glucose. Productions of cofactors were calculated from carbon fluxes. The requirements were calculated on a theoretical basis from biomass composition and growth rates.

5. Discussion

The Csr system is thought to reinforce the competitiveness of *E. coli* in the use of carbon sources metabolized through the upper part of the EMP such as glucose (Timmermans and Van Melderen 2009; Timmermans and Van Melderen 2010). The data reported here reinforce and extend this role by showing that CsrA enhances the use of a wide range of carbon sources. The Csr system plays no role in nutrient selection (i.e. there was no alteration in Csr mutants in the range of carbon sources used) but enhances growth efficiency. These effects were observed for compounds metabolized by different metabolic pathways. These results indicate that Csr either directly or indirectly controls a wider range of metabolic pathways than expected from its known target enzymes. Specifically, the work reported here demonstrates a role for Csr in the use of substrates metabolized via the ED and PP pathways. This is of particular physiological importance due to large number of such compounds in the colon and the reported role of the ED

pathway in gut colonization (Chang, Smalley et al. 2004; Leatham, Stevenson et al. 2005).

Central metabolism is under the control of a complex regulatory network that involves multiple transcriptional regulators (Perrenoud and Sauer 2005; Ishii, Nakahigashi et al. 2007). The actual role of post-transcriptional regulation in the control of metabolism is not well understood. Our detailed investigations – combining both metabolomics and fluxomics experiments – of the metabolic alterations occurring in Csr mutants show that CsrA indeed plays a significant role in the control of cellular metabolism. The metabolic effects depend on the nature of the carbon source, and striking differences were observed between the two carbon sources investigated, i.e. glucose and gluconate. The effects observed on glucose metabolism are generally consistent with the known role of the Csr system. Less efficient use of glucose in the CsrA mutant, with a reduced glycolytic flux, oxidative metabolism (reduced TCA cycle activity) and increased acetate production were all observed. These effects can be explained by the known targets of CsrA, since the positive control of glycolysis and the negative control of acetate metabolism triggers the carbon flow towards the TCA cycle. However, control of CsrA via the latter pathway, though not yet described, cannot be excluded. All these data suggest that CsrA plays a significant global role in the control of central metabolism on glucose by triggering carbon flow towards complete substrate oxidation (glycolysis + oxidative metabolism). These effects are likely to explain how the Csr system contributes to a higher metabolic efficiency on glucose.

Loss of function by CsrA in the laboratory *E. coli* strain MG1655 was reported to be lethal for growth on glucose due to a drastic diversion of the glycolytic flow towards glycogen synthesis (Liu, Yang et al. 1995; Sabnis, Yang et al. 1995). Accordingly, we could not generate a deletion mutant viable for growth on glucose in the Nissle 1917 strain. The attenuated *ΔcsrA51* mutant studied here accumulated significant amounts of glycogen during the stationary phase of growth, but the carbon flux towards glycogen synthesis measured during the exponential growth phase was much less than the glycolytic flux. Hence, the diversion of the carbon flux towards glycogen synthesis in this mutant is not responsible for either the decreased growth efficiency or for the observed metabolic rearrangements. Due to the residual partial activity of CsrA a lethal phenotype in a CsrA-

deletion mutant for growth on glucose cannot be excluded. In addition, glycogen synthesis and its regulation are strain dependent (Phue, Noronha et al. 2005), and actual effects in the Nissle 1917 strain may differ from those in the MG1655 strain. It is worth noting that due to its evolution in the laboratory, MG1655 remains a poor colonizer compared to Nissle 1917. Furthermore, in this work, we showed that the actual metabolic effect of CsrA is not only restricted to carbon sources that enter metabolism at the level of G6P, hence it cannot be explained by an increase in glycogen metabolism.

The detailed investigations of gluconate metabolism revealed surprising features. No Csr targets have been identified in the enzymes of the PP pathway, and only an indirect – as yet unidentified – action of CsrA on one of the ED enzyme (ED aldolase) has been proposed (Murray and Conway 2005). Due to the positive action of CsrA on ED aldolase and the absence of action on PP enzymes, a reduced contribution of the ED pathway to gluconate metabolism and a higher contribution of the PP pathway were expected in the *ΔcsrA51* mutant. However, the opposite effects were observed. Both the metabolome and fluxome data showed that gluconate was metabolized almost exclusively via the ED pathway, while the contribution of the PP pathway was strongly reduced. Contrary to the observations made on glucose, the relative fluxes in the TCA cycle and acetate metabolism in the mutant did not differ from that in the wild-type strain. Moreover, the wild-type strain released significant amounts of pyruvate as a by-product, but this process was almost absent in the *ΔcsrA51* mutant. The complete metabolisation of gluconate by the ED pathway is also accompanied by the concomitant inversion of the carbon flux in upper glycolysis, which operates in a gluconeogenetic direction, resulting in the recycling of 6PG. All these observations show that i) the Csr system exerts significant and global control over metabolism during growth on gluconate, and ii) the impact of the Csr system on the central metabolism upon growth on gluconate differs significantly from that observed on glucose. The regulatory mechanisms by which the Csr system exerts global control over central metabolism upon growth on gluconate remain unclear. Trioses-phosphate recycling in the CsrA mutant is consistent with the known antagonistic effects of CsrA on glycolysis and gluconeogenesis. The other metabolic reorganizations observed in the mutant ("pure" ED metabolism, decreased pyruvate production) cannot be explained from current knowledge on Csr targets, while most expected effects (glycogen accumulation, decreased acetate metabolism,

decreased TCA cycle) were not observed. A first explanation could be the occurrence of yet unidentified Csr targets in gluconate utilization pathways. The Csr system is also embedded in a complex regulatory network, and global control of metabolism on gluconate could be triggered through indirect mechanisms resulting from the interplay between the various regulation systems. To identify these regulatory mechanisms, further molecular investigations combining proteomics and transcriptomics should be performed, in light of the metabolome and fluxome data obtained here.

Whatever the regulatory mechanisms involved, this work show that Csr is involved in the utilization of ED compounds and other physiologically-relevant carbon sources. Hence, CsrA appears as a key regulator for establishment and persistence of *E. coli* in the changing environment of the intestine.

6. References

Atkinson, D. E. (1968). "The energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers." Biochemistry **7**(11): 4030-4034.

Bailey, J. K., J. L. Pinyon, et al. (2010). "Distribution of human commensal Escherichia coli phylogenetic groups." J Clin Microbiol **48**(9): 3455-3456.

Bolten, C. J., P. Kiefer, et al. (2007). "Sampling for metabolome analysis of microorganisms." Anal Chem **79**(10): 3843-3849.

Chang, D. E., D. J. Smalley, et al. (2004). "Carbon nutrition of Escherichia coli in the mouse intestine." Proc Natl Acad Sci U S A **101**(19): 7427-7432.

Chester, V. E. (1968). "Heritable glycogen-storage deficiency in yeast and its induction by ultra-violet light." J Gen Microbiol **51**(1): 49-56.

Cocaign-Bousquet, M., C. Garrigues, et al. (1995). "Rational Development of a Simple Synthetic Medium for the Sustained Growth of Lactococcus-Lactis." Journal of Applied Bacteriology **79**(1): 108-116.

Datsenko, K. A. and B. L. Wanner (2000). "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products." Proc Natl Acad Sci U S A **97**(12): 6640-6645.

Fabich, A. J., S. A. Jones, et al. (2008). "Comparison of carbon nutrition for pathogenic and commensal Escherichia coli strains in the mouse intestine." Infect Immun **76**(3): 1143-1152.

Feist, A. M., C. S. Henry, et al. (2007). "A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information." Mol Syst Biol **3**: 121.

Fuhrer, T., E. Fischer, et al. (2005). "Experimental identification and quantification of glucose metabolism in seven bacterial species." J Bacteriol **187**(5): 1581-1590.

Grozdanov, L., C. Raasch, et al. (2004). "Analysis of the genome structure of the nonpathogenic probiotic Escherichia coli strain Nissle 1917." J Bacteriol **186**(16): 5432-5441.

Hancock, V., M. Dahl, et al. (2010). "Probiotic Escherichia coli strain Nissle 1917 outcompetes intestinal pathogens during biofilm formation." J Med Microbiol **59**(Pt 4): 392-399.

Heux, S., B. Philippe, et al. (2011). "High-throughput workflow for monitoring and mining bioprocess data and its application to inferring the physiological response of Escherichia coli to perturbations." Appl Environ Microbiol **77**(19): 7040-7049.

Hobman, J. L., C. W. Penn, et al. (2007). "Laboratory strains of Escherichia coli: model citizens or deceitful delinquents growing old disgracefully?" Mol Microbiol **64**(4): 881-885.

Ishii, N., K. Nakahigashi, et al. (2007). "Multiple high-throughput analyses monitor the response of E. coli to perturbations." Science **316**(5824): 593-597.

Kamada, N., Y. G. Kim, et al. (2012). "Regulated virulence controls the ability of a pathogen to compete with the gut microbiota." Science **336**(6086): 1325-1329.
Kiefer, P., C. Nicolas, et al. (2007). "Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry." Anal Biochem **360**(2): 182-188.
Leatham, M. P., S. J. Stevenson, et al. (2005). "Mouse intestine selects nonmotile flhDC mutants of Escherichia coli MG1655 with increased colonizing ability and better utilization of carbon sources." Infect Immun **73**(12): 8039-8049.
Liu, M. Y., G. Gui, et al. (1997). "The RNA molecule csrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli." J Biol Chem **272**(28): 17502-17510.
Liu, M. Y., H. Yang, et al. (1995). "The product of the pleiotropic Escherichia coli gene csrA modulates glycogen biosynthesis via effects on mRNA stability." J Bacteriol **177**(10): 2663-2672.
Massou, S., C. Nicolas, et al. (2007). "Application of 2D-TOCSY NMR to the measurement of specific(13C-enrichments in complex mixtures of 13C-labeled metabolites." Metab Eng **9**(3): 252-257.
Massou, S., C. Nicolas, et al. (2007). "NMR-based fluxomics: quantitative 2D NMR methods for isotopomers analysis." Phytochemistry **68**(16-18): 2330-2340.
Mercante, J., A. N. Edwards, et al. (2009). "Molecular geometry of CsrA (RsmA) binding to RNA and its implications for regulated expression." J Mol Biol **392**(2): 511-528.
Millard, P., F. Letisse, et al. (2012). "IsoCor: Correcting MS data in isotope labeling experiments." Bioinformatics.
Mira, A., H. Ochman, et al. (2001). "Deletional bias and the evolution of bacterial genomes." Trends Genet **17**(10): 589-596.
Miranda, R. L., T. Conway, et al. (2004). "Glycolytic and gluconeogenic growth of Escherichia coli O157:H7 (EDL933) and E. coli K-12 (MG1655) in the mouse intestine." Infect Immun **72**(3): 1666-1676.
Murray, E. L. and T. Conway (2005). "Multiple regulators control expression of the Entner-Doudoroff aldolase (Eda) of Escherichia coli." J Bacteriol **187**(3): 991-1000.
Neidhardt, F. C., R. Curtiss, et al. (1996). "Escherichia coli and Salmonella." Washington D.C., ASM press.
Nicolas, C., P. Kiefer, et al. (2007). "Response of the central metabolism of Escherichia coli to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase." FEBS Lett **581**(20): 3771-3776.
Parrou, J. L. and J. Francois (1997). "A simplified procedure for a rapid and reliable assay of both glycogen and trehalose in whole yeast cells." Anal Biochem **248**(1): 186-188.
Peekhaus, N. and T. Conway (1998). "What's for dinner?: Entner-Doudoroff metabolism in Escherichia coli." J Bacteriol **180**(14): 3495-3502.
Pernestig, A. K., D. Georgellis, et al. (2003). "The Escherichia coli BarA-UvrY two-component system is needed for efficient switching between glycolytic and gluconeogenic carbon sources." J Bacteriol **185**(3): 843-853.
Perrenoud, A. and U. Sauer (2005). "Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli." J Bacteriol **187**(9): 3171-3179.
Phue, J. N., S. B. Noronha, et al. (2005). "Glucose metabolism at high density growth of E. coli B and E. coli K: differences in metabolic pathways are responsible for efficient glucose utilization in E. coli B as determined by microarrays and Northern blot analyses." Biotechnol Bioeng **90**(7): 805-820.
Portais, J. C., P. Tavernier, et al. (1999). "Cyclic organization of the carbohydrate metabolism in Sinorhizobium meliloti." Eur J Biochem **265**(1): 473-480.
Romeo, T. (1998). "Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule csrB." Mol Microbiol **29**(6): 1321-1330.
Romeo, T., M. Gong, et al. (1993). "Identification and molecular characterization of csrA, a pleiotropic gene from Escherichia coli that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties." J Bacteriol **175**(15): 4744-4755.
Romeo, T., C. A. Vakulskas, et al. (2012). "Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems." Environ Microbiol.
Sabnis, N. A., H. Yang, et al. (1995). "Pleiotropic regulation of central carbohydrate metabolism in Escherichia coli via the gene csrA." J Biol Chem **270**(49): 29096-29104.

Sauer, U., F. Canonaco, et al. (2004). "The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of Escherichia coli." J Biol Chem **279**(8): 6613-6619.
Snider, T. A., A. J. Fabich, et al. (2009). "E. coli O157:H7 catabolism of intestinal mucin-derived carbohydrates and colonization." Vet Microbiol **136**(1-2): 150-154.
Sokol, S., P. Millard, et al. (2011). "influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labeling experiments." Bioinformatics.
Suzuki, K., P. Babitzke, et al. (2006). "Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs csrB and csrC for degradation by RNase E." Genes Dev **20**(18): 2605-2617.
Taymaz-Nikerel, H., M. de Mey, et al. (2009). "Development and application of a differential method for reliable metabolome analysis in Escherichia coli." Anal Biochem **386**(1): 9-19.
Timmermans, J. and L. Van Melderen (2009). "Conditional essentiality of the csrA gene in Escherichia coli." J Bacteriol **191**(5): 1722-1724.
Timmermans, J. and L. Van Melderen (2010). "Post-transcriptional global regulation by CsrA in bacteria." Cell Mol Life Sci **67**(17): 2897-2908.
Todar, K. (2007). "Pathogenic E. coli." Online textbook of bacteriology. University of Wisconsin-Madison, Department of Bacteriology.
Wang, X., A. K. Dubey, et al. (2005). "CsrA post-transcriptionally represses pgaABCD, responsible for synthesis of a biofilm polysaccharide adhesin of Escherichia coli." Mol Microbiol **56**(6): 1648-1663.
Wei, B., S. Shin, et al. (2000). "Global regulatory mutations in csrA and rpoS cause severe central carbon stress in Escherichia coli in the presence of acetate." J Bacteriol **182**(6): 1632-1640.
Wei, B. L., A. M. Brun-Zinkernagel, et al. (2001). "Positive regulation of motility and flhDC expression by the RNA-binding protein CsrA of Escherichia coli." Mol Microbiol **40**(1): 245-256.
Weilbacher, T., K. Suzuki, et al. (2003). "A novel sRNA component of the carbon storage regulatory system of Escherichia coli." Mol Microbiol **48**(3): 657-670.
Wu, L., M. R. Mashego, et al. (2005). "Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards." Anal Biochem **336**(2): 164-171.

Chapter 8

Impact of Csr on carbon, energy and redox metabolism in *Escherichia coli* K12 MG1655 and Nissle 1917

1. Résumé

Dans ce chapitre, nous avons poursuivi l'étude sur le rôle du système Csr dans le contrôle du métabolisme central d'*Escherichia coli*. Nous avons finement caractérisé le comportement métabolique de souches sauvages d'*E. coli* sur différentes sources de carbone capables de supporter leur implantation dans l'intestin (gluconate, ribose, galactose, acétate) et représentative de ses principales voies du métabolisme central (voie d'Entner-Doudoroff, voie des pentose phosphate, glycolyse, néoglucogénèse). Cette étude, réalisée chez la souche K12 MG1655 et chez la souche Nissle 1917, a permis de mettre en lumière des propriétés fonctionnelles de leur métabolisme et d'identifier des spécificités métaboliques propres à la souches Nissle 1917. L'analyse comparative de mutants i) démontre que le système Csr ne contrôle pas seulement le métabolisme carboné d'*E. coli* mais également son métabolisme énergétique et rédox, et ii) suggère qu'il joue un rôle central dans la coordination entre anabolisme et catabolisme.

2. Introduction

Escherichia coli constitue un groupe très divers de bactéries entériques contenant des souches commensales, qui jouent un rôle essentiel dans le fonctionnement de la flore intestinale, et des souches pathogènes particulièrement versatiles. La compétition pour les ressources nutritionnelles disponibles semble jouer un rôle majeur dans la colonisation de l'intestin par *E. coli*, et la capacité à utiliser des ressources spécifiques ou à utiliser de manière plus efficace les nutriments disponibles fait partie des stratégies mises en place par certaines bactéries pathogènes pour supplanter le microbiote (Miranda, Conway et al. 2004; Kamada, Kim et al. 2012).

La colonisation de l'intestin par *E. coli* parait associée à la disponibilité de sources de carbone préférentielles, parmi lesquelles les sucres (mannose, ribose, galactose, etc) et leurs dérivés (principalement des acides osidiques – gluconate, acide sialique, glucuronate, etc –, des déoxysucres – fucose, déoxyribose –, et des osamines – N-acétylglucosamine, glucosamine, N-acétylgalactosamine, etc –) (Chang, Smalley et al. 2004; Miranda, Conway et al. 2004; Snider, Fabich et al. 2009). Les voies glycolytiques [glycolyse (EMP), voie d'Entner-Doudoroff (ED), voie des pentoses phosphate (PP)] (Figure 1) semblent donc jouer un rôle déterminant dans les processus de colonisation

(Peekhaus and Conway 1998; Chang, Smalley et al. 2004; Leatham, Stevenson et al. 2005). La maintenance et la persistance d'*E. coli* dans l'intestin reposent quant à elles sur sa capacité à utiliser différents substrats alternatifs lorsque les substrats préférentiels deviennent limitant. Ces sources de carbone alternatives sont principalement gluconéogéniques (acétate, propionate, etc) (Figure 1), ce qui implique une réorganisation fonctionnelle complète du métabolisme (Pernestig, Georgellis et al. 2003; Snider, Fabich et al. 2009). Cette activation des processus néoglucogéniques liée à la consommation des acides organiques est associée *in situ* à des profonds remaniements physiologiques (perte de motilité, modifications pariétales, production de polysaccharides, etc) nécessaires à la persistance d'*E. coli* dans l'intestin.

L'implantation d'*E. coli* dans l'intestin résulte donc de stratégies adaptatives globales permettant à la bactérie de survivre dans un environnement très variable. Ces processus sont coordonnés par des réseaux de régulation particulièrement complexes, dont le système Csr (Carbon storage regulator), un des principaux systèmes de régulation post-transcriptionnel d'*E. coli* (Romeo, Vakulskas et al. 2012). Ce système de régulation contrôle l'expression fonctionnelle de gènes impliqués dans les principaux processus physiologiques et moléculaires indispensables à la colonisation et la survie d'*E. coli* dans l'intestin, parmi lesquels la motilité, la formation de biofilms, la réponse stringente, et le métabolisme. Historiquement identifié comme un régulateur de la synthèse de glycogène (Romeo and Gong 1993; Romeo, Gong et al. 1993), le contrôle exercé par le système Csr a été étendu à l'ensemble du métabolisme carboné central au cours des dernières années (Figure 1). CsrA active de nombreuses enzymes glycolytiques (glucose-6-phosphate isomérase, phosphofructokinase A, triose-phosphate isomérase, enolase, pyruvate kinase F, Entner-Doudoroff aldolase) et l'utilisation d'acétate (acétylcoA synthétase, isocitrate lyase), et inhibe des enzymes gluconéogéniques (fructose-1,6-biphosphatase, PEP synthase, PEP carboxykinase A) et les enzymes impliquées de la biosynthèse du glycogène (phosphoglucomutase, ADP-glucose pyrophosphorylase, glycogène synthase, glycogène phosphorylase) (Liu, Yang et al. 1995; Sabnis, Yang et al. 1995; Wei, Shin et al. 2000; Tatarko and Romeo 2001; Baker, Morozov et al. 2002; Murray and Conway 2005). Sur le plan métabolique, ce système de régulation est actuellement considéré comme un activateur de la glycolyse et un inhibiteur de la gluconéogénèse.

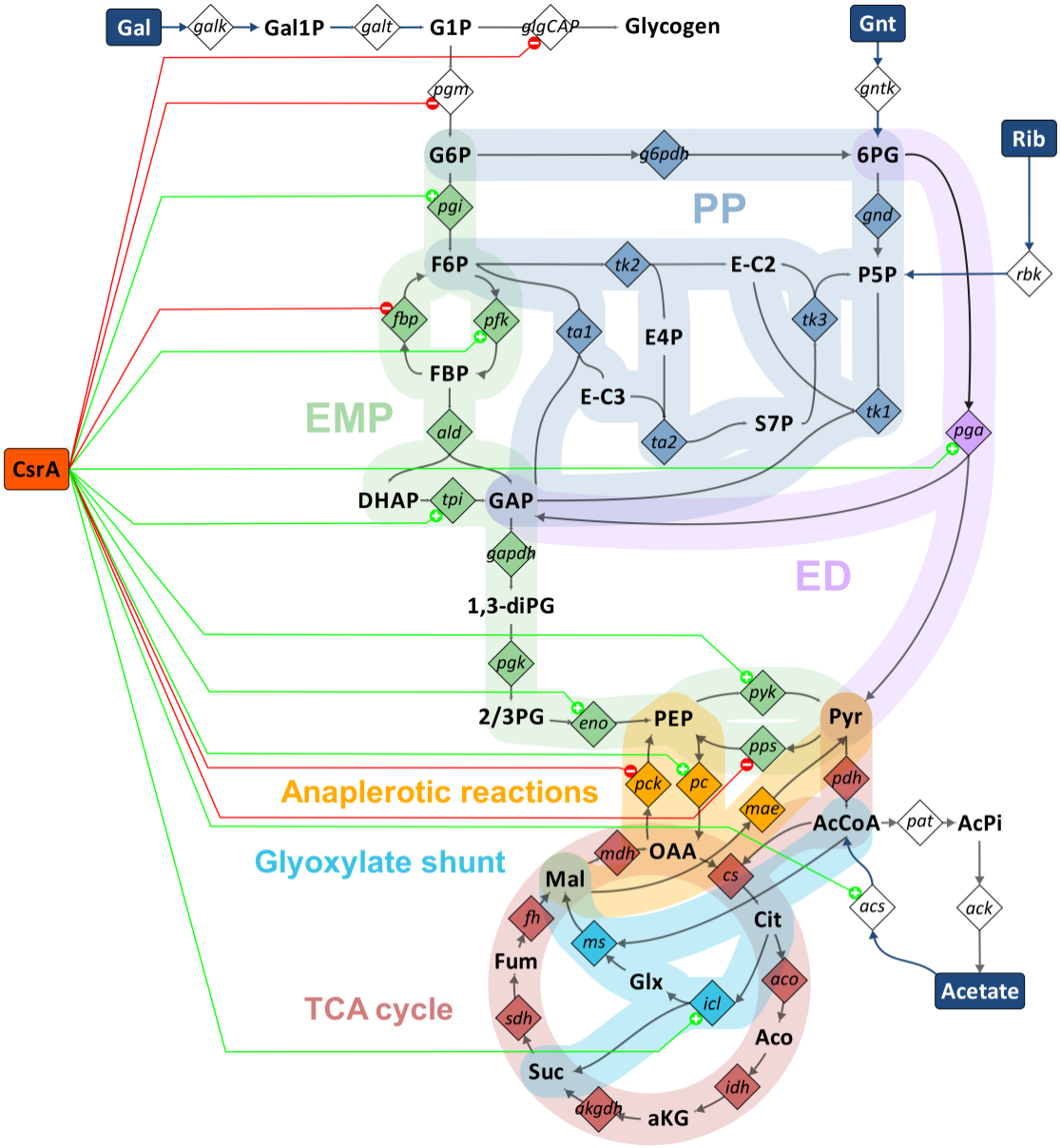


Figure 1. Réseau du métabolisme carboné central d'*Escherichia coli* incluant les principales voies métaboliques (EMP: glycolyse, ED: voie d'Entner-Doudoroff, PP: voie des pentoses phosphates, TCA: cycle de Krebs, shunt glyoxylate et voies anaplérotiques). Les effets positifs et négatifs de csrA sur différentes enzymes du métabolisme central sont représentés par les flèches vertes et rouges, respectivement. Les sources de carbones considérées dans ce travail (gluconate, galactose, ribose, acétate) entrent dans différentes voies du réseau central.

Le système Csr est composé de quatre éléments principaux (Romeo, Vakulskas et al. 2012). L'élément central est la protéine CsrA qui se fixe sur la partie 5' non traduite ou à proximité de la séquence Shine-Dalgarno des ARNs messagers cibles. Par ce mécanisme, CsrA inhibe leur traduction de manière directe (compétition avec les ribosomes) et indirecte (déstabilisation de l'ARNm). Il a été récemment démontré que CsrA active

l'expression de certains transcrits (e.g. flhDC, impliqué dans la motilité), ce qui souligne la complexité mécanistique du contrôle exercé par ce système. L'activité de CsrA est inhibée par deux petits ARNs non codants, csrB et csrC, qui contiennent plusieurs sites de fixation de CsrA et séquestrent la protéine (Figure 2). Enfin, la protéine CsrD active CsrA en facilitant la dégradation de csrB et csrC par la RNAse E. De nombreuses boucles de régulation assurent un contrôle précis du niveau d'activité de CsrA, qui inhibe (directement) et active (indirectement) sa propre expression, inhibe l'expression de *csrD*, et active la transcription de *csrB* et *csrC* via le système à deux composants BarA/UvrY.

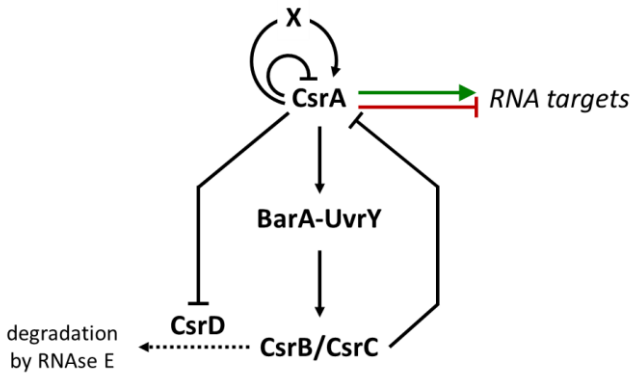


Figure 2. Représentation schématique (adapté de (Romeo, Vakulskas et al. 2012)) du réseau de régulation de CsrA, l'élément central du système Csr. Le contrôle par les différents éléments peut être positif (→) ou négatif (⊢).

A ce jour, la quasi-totalité des études portant sur le système Csr d'*E. coli* ont été réalisées lors de croissances sur glucose, une source de carbone absente du colon et donc peu représentative du métabolisme d'*E. coli in vivo*. Puisque les résultats précédents (chapitre 7) indiquent que le contrôle exercé par Csr sur le métabolisme dépend fortement de la source de carbone, il parait indispensable d'étendre ces travaux à d'autres sources de carbone physiologiques afin d'appréhender son rôle. De plus, l'analyse du système Csr a été réalisée principalement par des approches moléculaires, alors que la flexibilité importante des réseaux métaboliques implique qu'un changement sur une voie régulée par Csr peut engendrer des modifications dans tout le métabolisme. Le nombre importants de transcrits co-purifiés avec la protéine CsrA renforce également la nécessité d'analyser le rôle de ce régulateur à l'échelle du système biologique. Des approches quantitatives globales telles que la métabolomique (permettant de quantifier les intermédiaires métaboliques) et l'analyse des flux métaboliques par marquage

isotopique (^{13}C-MFA, permettant de quantifier les flux métaboliques) ont été récemment développées dans le but d'obtenir une image détaillée du fonctionnement métabolique d'*E. coli in vivo*. Bien que la ^{13}C-MFA soit particulièrement bien adaptée pour comprendre le rôle des régulateurs globaux sur le métabolisme (Perrenoud and Sauer 2005), cette approche n'a jamais été utilisée dans le cadre d'études sur le système Csr.

Dans ce chapitre, nous avons analysé le rôle du système Csr dans le contrôle du métabolisme central d'*E. coli* sur des sources de carbones utilisées *in vivo* par cette bactérie lors des phases de colonisation et de maintenance dans l'intestin (galactose, ribose, gluconate, acétate). Ces sources de carbones – dont le métabolisme est bien caractérisé sur le plan moléculaire mais peu sur le plan fonctionnel – sont représentatives des principales voies du métabolisme central (EMP, PP, ED, et gluconéogénèse, respectivement) (Figure 1). Cette étude a été réalisée chez deux souches distinctes d'*E. coli*, la souche de laboratoire K12 MG1655 (la souche d'*E. coli* la plus étudiée depuis le milieu du 20ème siècle, et de ce fait la mieux connue) (Mira, Ochman et al. 2001; Hobman, Penn et al. 2007) et la souche Nissle 1917 (Grozdanov, Raasch et al. 2004) appartenant au groupe phylogénétique B2. Ce groupe, dont la prévalence augmente dans les pays occidentaux (Escobar-Paramo, Grenet et al. 2004; Bailey, Pinyon et al. 2010), est d'un intérêt particulier puisqu'il contient des souches caractérisées par une capacité de persistance supérieure à celles des autres groupes (Nowrouzian, Wold et al. 2005; Nowrouzian and Oswald 2012). Le mutant de délétion de *csrA* n'étant pas viable sur substrats glycoliques et sur milieu riche, nous avons utilisé lors de cette étude des mutants (*csrA51*) codant pour une protéine CsrA partiellement active (Timmermans and Van Melderen 2009). L'analyse des souches sauvages et des mutants *csrA51*, par des approches systémiques quantitatives (^{13}C-MFA et métabolomique), a permis i) de caractériser en détail le fonctionnement métabolique d'*E. coli* sur ces sources de carbone physiologiques, ii) d'identifier des caractères métaboliques propres à chaque souche, et iii) d'étendre le rôle du système Csr dans le contrôle du métabolisme central d'*E. coli*.

L'analyse par ^{13}C-MFA du nombre relativement important de conditions (2 souches sauvages, 2 mutants dont le niveau d'activité de CsrA est diminué, et 4 sources de carbone, soit 16 conditions au total, avec plusieurs réplicats biologiques) a été rendue

possible par les développements réalisés dans la première partie de cette thèse (chapitres 2 à 6).

3. Matériel et méthodes

Souches. Les souches sauvages d'*Escherichia coli* utilisées dans ces travaux sont la souche K12 MG1655 et la souche Nissle 1917. Les mutants *csrA51* de ces souches ont été construits par Olga Revelles et Thomas Esquerre (LISBP, Toulouse) en utilisant la méthode de remplacement allélique mise au point par (Datsenko and Wanner 2000). Le gène inséré code pour une protéine CsrA tronquée après le 51ème acide aminé et dont le niveau d'activité est atténué (Timmermans and Van Melderen 2009).

Milieux et conditions de culture. Les précultures ont été réalisées en microplaques 48 puits (CytoOne non traitées, USA Scientific, USA) dans 1 mL de milieu minimum M9 (37 °C sur la nuit) ensemencés à une DO_{600nm} de 0,05 à partir d'un inoculum obtenu sur milieu riche LB (37 °C, 8 heures). Les cultures ont été réalisées dans 15 mL de milieu M9 (Nicolas, Kiefer et al. 2007) et supplémenté en source de carbone (gluconate, galactose, ribose ou acétate) à une concentration finale de 90 Cmmol.L^{-1}. Ces expériences ont été réalisées sur une plateforme automatisée haut-débit développée par la plateforme MetaToul (LISBP, Toulouse) en collaboration avec Tecan (Suisse). Le système de culture implémenté est un bloc de 48 bioréacteurs développé par (Puskeiler, Kaufmann et al. 2005). La pO_2 a été maintenue au-dessus de 20% pendant la croissance, l'agitation à 2200 rpm, la température à 37 °C, et le pH à 7,0±0,1 par ajout automatisé de soude (NaOH 0,4M) ou d'acide (HCl 0,4M). Pour les expériences d'analyse des flux métaboliques par marquage isotopique (^{13}C-MFA), la source de carbone non marquée a été remplacée par une source de carbone marquée au ^{13}C (1-^{13}C-ribose/U-^{13}C-ribose 80/20, 1-^{13}C-galactose/U-^{13}C-galactose 80/20, 1-^{13}C-gluconate/U-^{13}C-gluconate 80/20, et 2-^{13}C-acétate).

Paramètres de croissance. Le suivi de croissance a été effectué de manière automatisée par mesure de l'absorbance à 600 nm dans des microplaques 96 puits (CytoOne non traitées, USA Scientific) à l'aide d'un lecteur de microplaques (Sunrise Absorbance reader, Tecan). Des dilutions ont été réalisées si nécessaire afin de rester dans la gamme de linéarité du lecteur. Pour quantifier les substrats et produits du métabolisme, 200 µL de milieu de culture ont été filtrés manuellement (Minisart 0,2 µm, Sartorius,

Allemagne) et analysés par RMN après ajout de 200 µL de D_2O et de 100 µL d'une solution de TSP-d4 (utilisé comme standard interne) à une concentration de 0,74 g/L. Les analyses ont été réalisées à 286 K sur un spectromètre RMN 800 MHz (Bruker, Germany) équipé d'un passeur d'échantillons (SampleJet, Bruker) et d'une cryosonde QCI H/C/N/P 5 mm, avec un pulse de 30° et un délai de relaxation de 20 s. En phase exponentielle de croissance, les taux de croissance µ (h^{-1}) ont été déterminés à partir des pentes calculées par régression linéaire sur les droites ln(DO)=f(t), les rendements en biomasse Y_X ($g_{DW}.mmol_{Substrat}^{-1}$) et en co-produits $Y_{co\text{-}produit}$ ($mmol_{co\text{-}produit}.mmol_{Substrat}^{-1}$) ont été déterminés respectivement à partir des pentes des droites [biomasse]=f([substrat]) et [co-produit]=f([substrat]), et les vitesses de consommation de substrat q_S ($mmol_{Substrat}.g_{DW}^{-1}.h^{-1}$) ont été obtenues par calcul des ratios μ/Y_X.

Echantillonnage des métabolites intracellulaires pour la métabolomique. Les cellules ont été prélevées en milieu de phase exponentielle de croissance en utilisant la méthode différentielle développée par (Taymaz-Nikerel, de Mey et al. 2009). 120 µL de milieu+cellules ou de milieu filtré (Sartolon 0,2 µm) ont été plongés dans 1,25 mL d'une solution de méthanol/acétonitrile/eau (40/40/20) maintenue à -20 °C et contenant 120 µL d'un extrait cellulaire uniformément marqué au ^{13}C (utilisé comme standard interne) (Wu, Mashego et al. 2005). Cette solution a été acidifiée par ajout d'acide formique (0,1%) afin d'éviter la dégradation des nucléotides (Bennett, Kimball et al. 2009). Les extraits cellulaires ont été stockés à -80 °C jusqu'à leur analyse.

Echantillonnage des métabolites intracellulaire pour la ^{13}C-MFA. En phase exponentielle de croissance, 500 µL de milieu+cellules ont été plongés dans 3 mL d'une solution méthanol/acétonitrile/eau (40/40/20) maintenue à -20 °C (procédure adaptée des travaux présentés chapitre 3). Les extraits ont été stockés à -80 °C.

Préparation des échantillons et analyses LC-MS/MS. Les extraits cellulaires ont été évaporés 6 h (SC110A SpeedVac Plus, ThermoSavant, USA), lyophilisés, et repris dans 200 µL d'eau milliQ pour ceux dédiés à l'analyse du métabolome et dans 500 µL d'eau milliQ pour ceux dédiés à l'analyse des flux métaboliques. Les débris cellulaires ont été éliminés par centrifugation (10000 g, 5 min), et les analyses ont été réalisées par chromatographie liquide échangeuse d'anions (Dionex ICS 2000, Dionex, USA) couplée à un spectromètre de masse triple quadrupole (QTrap 4000, AB Sciex, Canada). Tous les échantillons ont été analysés en mode négatif par MRM. Le volume d'injection est de 15

µL, correspondant à approximativement 3 µg de biomasse. Pour la métabolomique, les métabolites de la glycolyse (EMP) (Glc6P, Fru6P, FBP, pools combinés de 2- et 3-PG, PEP), des voies d'Entner-Doudoroff (ED) et des pentoses phosphate (PP) (6PG, Sed7P, Rib5P), du cycle de Krebs (TCA) (Mal, Fum, Suc, Cit), de voies périphériques (M6P, GlcN6P, GalN6P, GalN1P, GlcNAc6P, GlcNAc1P, TagBP, UDP-Glc) et du métabolisme des nucléotides (AMP, ADP, ATP, CMP, CDP, CTP, UMP, UDP, UTP) ont été quantifiés. La méthode de dilution isotopique généralisée (IDMS) (Wu, Mashego et al. 2005) a été utilisée afin d'améliorer la précision et la justesse des résultats. Pour les expériences de ^{13}C-MFA, les massifs isotopiques des métabolites centraux (Glc6P, Fru6P, FBP, 6PG, Rib5P, Sed7P, PEP, Mal, Suc, Fum, Cit) ont été mesurés par la méthode décrite dans le chapitre 4. Les distributions des isotopologues ont été obtenues après correction des massifs isotopiques pour l'abondance naturelle des différents éléments en utilisant IsoCor (chapitre 5).

Quantification du marquage des co-produits par RMN. Les différentes formes isotopiques des co-produits (acétate et pyruvate) ont été quantifiées par RMN. 500 µL de milieu ont été filtrés en phase exponentielle de croissance et, après ajout de 100 µL de D_2O, analysés à 286 K en utilisant un pulse de 30° avec un délai de relaxation de 30 s.

Construction des modèles du réseau métabolique. La structure du réseau métabolique initial contient les principales voies du métabolisme carboné central d'*E. coli*: la glycolyse (EMP), la voie des pentoses phosphates (PP), la voie d'Entner-Doudoroff (ED), le cycle de Krebs (TCA), le shunt glyoxylique (GS), les voies anaplérotiques, et les voies de sorties des précurseurs vers la formation de biomasse. Ce réseau a été adapté pour chaque source de carbone en combinant i) les informations obtenues dans les différentes bases de données (Kegg, EcoCyc), ii) les informations quantitatives sur les niveaux d'expression d'ARNm et les activités des différentes enzymes disponibles dans la littérature, et iii) les résultats des calculs de flux réalisés à partir des données isotopiques mesurées.

Optimisation du marquage d'entrée. Dans les expériences de ^{13}C-MFA, le marquage d'entrée (i.e. la proportion de différentes formes isotopiques utilisées pour chaque source de carbone) a été optimisé à l'aide d'IsoDesign (outil et méthodologie détaillés dans le chapitre 2) afin de maximiser le nombre de flux calculables et la précision de chaque flux.

Calcul des flux de carbone. Les distributions des flux de carbone ont été calculées en utilisant influx_s (chapitre 6). Les besoins en précurseurs pour la formation de biomasse ont été estimés à partir de la composition de la biomasse et des taux de croissance (Neidhardt, Curtiss et al. 1996). Les flux intracellulaires ont été estimés à partir des flux extracellulaires et des données isotopiques mesurées sur les métabolites intracellulaires par LC-MS/MS et sur les co-produits par RMN, en utilisant les modèles de réseaux métaboliques construits pour chaque source de carbone. Les flux ont été normalisés par rapport à la vitesse de consommation de substrat fixée arbitrairement à 100. La précision sur les flux a été estimée par la méthode de Monte-Carlo (100 itérations).

Calcul des bilans carbone. Les bilans carbone ont été calculés en divisant la somme des flux de production de co-produits, de CO_2 et de biomasse par la vitesse de consommation de substrat, tous exprimés en $Cmol.g_{DW}^{-1}.h^{-1}$.

Calcul des flux de production des co-facteurs rédox. Les vitesses de production de NADH, de $FADH_2$, et de NADPH ont été calculées en sommant les flux de toutes les réactions produisant ces co-facteurs réduits. Le NADH produit par l'anabolisme, ainsi que les besoins anaboliques en NADPH, ont été calculés à partir de la composition de la biomasse et du taux de croissance. Le NADPH excédentaire est converti en NADH en considérant l'activité des transhydrogénases (Fischer and Sauer 2003; Hua, Yang et al. 2003; Sauer, Canonaco et al. 2004; Fong, Nanchen et al. 2006).

Calculs des flux de production d'ATP et de la charge adénylate. La production totale d'ATP a été calculée en sommant les flux des réactions du métabolisme central produisant de l'ATP, et en soustrayant les flux des réactions consommant de l'ATP. L'ATP produit à partir du NADH et du $FADH_2$ par phosphorylation oxydative a été estimé en considérant un ratio P:O de 1 (Feist, Henry et al. 2007; Taymaz-Nikerel, Borujeni et al. 2010). Les besoins en ATP ont été estimés à partir des besoins théoriques pour la croissance (59,8 $mmol_{ATP}.g_{DW}^{-1}$) et pour la maintenance indépendante de la croissance (8,4 $mmol.g_{DW}^{-1}.h^{-1}$) (Feist, Henry et al. 2007). La charge adénylate (AEC) a été calculée à partir des concentrations intracellulaires en AMP, ADP et ATP selon la formule établie par (Atkinson 1968):

$$AEC = ([ATP]+0,5\times[ADP]) / ([AMP]+[ADP]+[ATP])$$

Analyses statistiques et analyses de corrélations. Les moyennes et écarts types des paramètres de croissance, des concentrations intracellulaires en métabolites, et des flux métaboliques ont été déterminés à partir d'au minimum 3, 3 et 2 réplicats biologiques indépendants, respectivement. Les corrélations de Pearson entre les jeux de données ont été calculées en utilisant la fonction *cor.test()* de R (http://www.r-project.org), et la significativité des résultats a été évaluée sur la base du coefficient de Pearson et de la p-value (test de Student bilatéral).

4. Résultats

Ces résultats ayant été obtenus en fin de thèse, l'analyse des données n'a pu être menée à son terme. Seuls les principaux résultats sont donc présentés et discutés dans ce chapitre.

Paramètres de croissance

Le métabolisme de souches sauvages d'*Escherichia coli* K12 MG1655 et Nissle 1917 et de leurs mutants *csrA51* a été caractérisé sur des sources de carbone capables de supporter leur croissance dans l'intestin (galactose, gluconate, ribose, acétate) et représentatives des principales voies du métabolisme carboné central [glycolyse (EMP), voie d'Entner-Doudoroff (ED), voie des pentoses phosphate (PP), et cycle de Krebs (TCA)]. Les cultures ont été réalisées sur une plateforme automatisée haut-débit assurant un contrôle de différents paramètres (pH, pO_2, température) et un suivi de la croissance pour un grand nombre de conditions en parallèle (jusqu'à 48). Cette dimension haut-débit a été rendue indispensable par le nombre important de conditions analysées et la nécessité de répliquer les différentes expériences sur le plan biologique. Des prélèvements ont été effectués au cours de la phase exponentielle de croissance pour suivre l'évolution des concentrations en biomasse, en substrat et en co-produits. Les surnageants ont été analysés par RMN, une technique permettant la détection, l'identification et la quantification des composés protonés présents dans le milieu (limite de détection de l'ordre de 10 µM dans le milieu de culture). Seule la production de deux petits acides organiques, l'acétate et le pyruvate, a été détectée lors de ces analyses.

A partir de l'ensemble de ces données, 4 paramètres physiologiques ont été calculés pour chaque condition expérimentale: le taux de croissance (µ, en h^{-1}), la vitesse de

consommation de substrat (q_S, en $mmol.g_{DW}^{-1}.h^{-1}$), le rendement en biomasse (Y_X, en $g_{DW}.mol_{Substrat}^{-1}$), et le rendement en co-produits ($Y_{co\text{-}produits}$, en $mmol_{Produit}.mmol_{Substrat}^{-1}$). Les résultats sont présentés dans le tableau 1.

Source de carbone	Souche	μ (h^{-1})	Y_X (g_{DW}/molS)	q_S (mmol/g_{DW}/h)	Y_{ac} (mol_{ac}/molS)	Y_{pyr} (mol_{pyr}/molS)
Gluconate	K12 WT	0,55 ± 0,04	30,3 ± 5,0	17,9 ± 2,7	0,48 ± 0,03	0,16 ± 0,04
	K12 *csrA51*	0,38 ± 0,05	55,2 ± 6,3	7,1 ± 0,7	0,11 ± 0,06	0,00 ± 0,01
	Nissle WT	0,99 ± 0,04	31,3 ± 5,4	32,4 ± 4,7	0,57 ± 0,03	0,14 ± 0,04
	Nissle *csrA51*	0,68 ± 0,01	42,8 ± 6,0	16,1 ± 2,4	0,43 ± 0,06	0,07 ± 0,01
Galactose	K12 WT	0,43 ± 0,01	39,3 ± 2,4	10,6 ± 0,6	0,00 ± 0,01	-
	K12 *csrA51*	0,30 ± 0,03	57,6 ± 2,8	5,4 ± 1,5	0,00 ± 0,01	-
	Nissle WT	0,61 ± 0,05	43,5 ± 3,9	14,1 ± 1,2	0,01 ± 0,01	-
	Nissle *csrA51*	0,40 ± 0,01	48,8 ± 3,2	8,2 ± 0,4	0,00 ± 0,01	-
Ribose	K12 WT	0,40 ± 0,04	31,2 ± 5,7	13,1 ± 1,4	0,19 ± 0,07	-
	K12 *csrA51*	0,37 ± 0,01	54,9 ± 8,2	6,5 ± 1,2	0,24 ± 0,03	-
	Nissle WT	0,55 ± 0,01	29,6 ± 0,8	18,8 ± 0,7	0,10 ± 0,04	-
	Nissle *csrA51*	0,37 ± 0,02	54,4 ± 8,6	6,9 ± 1,4	0,20 ± 0,06	-
Acetate	K12 WT	0,20 ± 0,01	9,4 ± 2,7	21,7 ± 5,6	-	-
	K12 *csrA51*	0,18 ± 0,01	12,1 ± 2,1	14,9 ± 2,6	-	-
	Nissle WT	0,30 ± 0,03	13,7 ± 2,3	21,1 ± 4,3	-	-
	Nissle *csrA51*	0,19 ± 0,01	15,8 ± 2,3	12,7 ± 1,3	-	-

Tableau 1. Paramètres de croissance mesurés pour chacune des souches sur les différentes sources carbonées. Les valeurs correspondent à la moyenne ± écart-type calculés à partir de 3 cultures indépendantes au minimum.

L'analyse de ces résultats est rendue complexe par le fait que les paramètres mesurés sont la résultante de 3 facteurs différents: la souche, la source de carbone, et le niveau d'activité de CsrA. Afin d'avoir une vision claire de l'effet de chacun sur le comportement métabolique d'*E. coli*, il est nécessaire de les dissocier. Pour deux conditions *A* et *B* où un seul facteur *f* diffère, l'effet E_f de ce facteur sur le paramètre *p* peut être estimé par:

$$E_f = \frac{p(condition\ A)}{p(condition\ B)}$$

La condition A est donc normalisée par rapport à la condition B qui est la condition de référence. Une valeur de E_f supérieure à 1 dénote un effet positif du facteur *f* sur le paramètre *p*, et une valeur inférieure à 1 dénote un effet négatif. Les conditions A et B

définies pour calculer l'effet des facteurs sur les différents paramètres pour chaque condition sont présentées tableau 2. Le rôle de CsrA pouvant différer entre la souche K12 et la souche Nissle, son effet a été analysé indépendamment pour chaque souche.

Facteur évalué	Condition A	Condition B (référence)	Facteurs communs entre A et B
Souche	Nissle WT	K12 WT	Source de carbone
Source de carbone	Source de carbone *i*	Moyenne du paramètre *p* sur l'ensemble des sources de carbone	Souche
Activité de CsrA (K12)	K12 *csrA51*	K12 WT	Source de carbone
Activité de CsrA (Nissle)	Nissle *csrA51*	Nissle WT	Source de carbone

Tableau 2. Conditions utilisées pour quantifier l'effet de chaque facteur sur les différents paramètres de croissance.

A partir de ces résultats les effets minimum, maximum et moyen de chaque facteur sur chaque paramètre ont été déterminés (Figure 3).

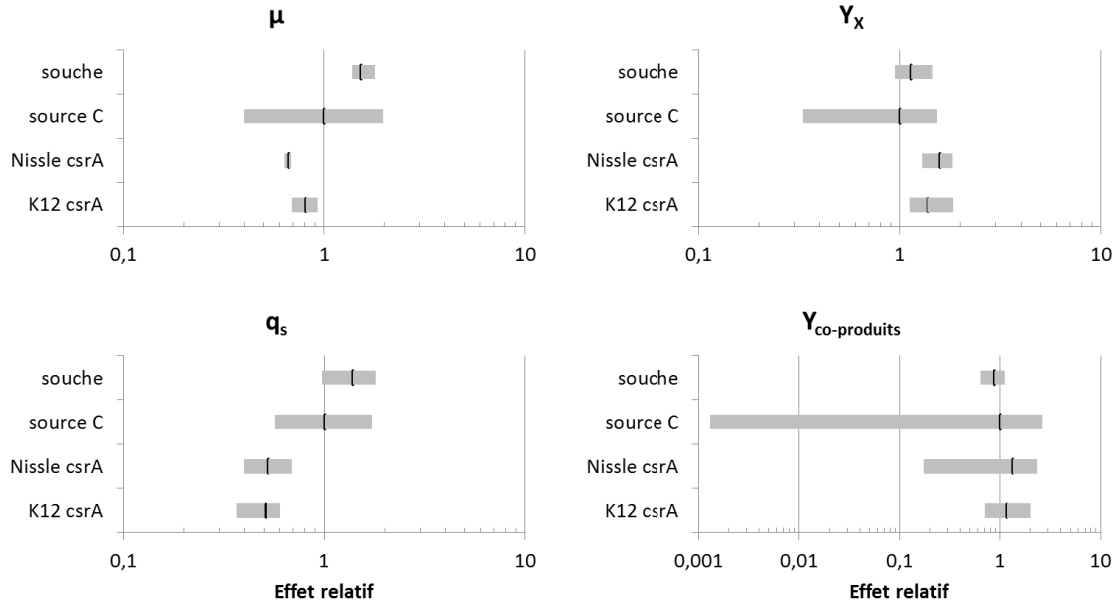


Figure 3. Effet des différents facteurs (souche, source de carbone, et activité de CsrA) sur les paramètres physiologiques mesurés. L'effet de la diminution d'activité de CsrA a été analysé indépendamment chez la souche K12 MG1655 et chez la souche Nissle 1917. L'effet moyen est représenté par la barre noire, les effets maximum et minimum correspondent aux extrémités des barres grises. La largeur de chaque barre représente donc la gamme de variation du paramètre considéré pour un facteur donné.

Le rendement en co-produits est le paramètre le plus variable (> 4 décades) en comparaison au taux de croissance, à la vitesse de consommation de substrat et au rendement en biomasse (< 1 décade).

Source de carbone. Ce facteur est le principal déterminant du comportement métabolique d'*E. coli*, affectant tous les paramètres mesurés de manière extrêmement importante et variable. Une grande gamme de taux de croissance est observée (facteur 4) en fonction des sources de carbone. Le taux de croissance le plus élevé est mesuré sur gluconate (0,55 h^{-1} pour K12, 0,99 h^{-1} pour Nissle), suivi par le galactose (0,43 h^{-1} pour K12, 0,61 h^{-1} pour Nissle), le ribose (0,40 h^{-1} pour K12, 0,55 h^{-1} pour Nissle), et enfin l'acétate (0,20 h^{-1} pour K12, 0,30 h^{-1} pour Nissle). La vitesse de consommation des substrats varie d'un facteur égal ou supérieur à 2 (de 10,6 à 21,7 $mmol.g_{DW}^{-1}.h^{-1}$ pour K12 MG1655 et de 14,1 à 32,4 $mmol.g_{DW}^{-1}.h^{-1}$ pour Nissle 1917). La nature des co-produits détectés n'est influencée que par la source de carbone. De l'acétate est produit sur galactose, ribose et gluconate, et du pyruvate est produit uniquement sur gluconate. Les rendements de production varient dans une gamme très large (> 4 décades) et dépendent également de la source de carbone. Enfin les rendements de production de biomasse varient d'un facteur 4, allant de 9,4 g_{DW}/mol d'acétate à 39,3 g_{DW}/mol de galactose pour la souche K12 et de 13,7 g_{DW}/mol d'acétate à 43,5 g_{DW}/mol de galactose pour la souche K12.

Activité de CsrA. L'effet de la diminution d'activité de CsrA est similaire chez la souche K12 et chez la souche Nissle, aussi bien sur le plan qualitatif que quantitatif, avec principalement une diminution du taux de croissance (en moyenne -22% pour K12 et -34% pour Nissle) liée à la diminution de la vitesse de consommation de substrat (-48% pour K12 et -49% pour Nissle), et une augmentation du rendement en biomasse (+58% pour K12 et +37% pour Nissle). Ces résultats sont en accord avec ceux présentés dans le chapitre 7. L'effet de CsrA sur le rendement en co-produits dépend de la source de carbone. Une augmentation du rendement de production d'acétate est observée sur ribose alors qu'il n'est pas affecté significativement sur galactose et est diminué sur gluconate.

Souche. La souche Nissle 1917 présente clairement des capacités métaboliques supérieures à celles de la souche K12 MG1655. Le taux de croissance de la souche Nissle est supérieur de 51% en moyenne à celui de la souche K12, l'ordre de préférence des nutriments (gluconate > galactose > ribose > acétate) étant le même pour les deux souches. La vitesse de consommation de substrat est plus élevée chez Nissle que chez K12 (+39% en moyenne), ce qui explique les différences de taux de croissance. De

même, le rendement en biomasse moyen de Nissle est légèrement supérieur à celui de K12 (+13%). Les différences de rendement en co-produits liées au facteur souche sont quant à elles très faibles (< 4% du carbone consommé).

Les différences importantes de comportement métabolique basées sur ces paramètres physiologiques reflètent des différences aux niveaux moléculaire et fonctionnel. Afin de caractériser de manière détaillée le fonctionnement du métabolisme central pour chaque condition, deux approches systémiques quantitatives ont été utilisées: la métabolomique et l'analyse des flux métaboliques par marquage isotopique.

Profils métaboliques

La métabolomique, dont le but est de quantifier les métabolites intracellulaires, permet d'obtenir une image de l'état physiologique de la cellule. Les concentrations de métabolites intracellulaires appartenant à différentes voies du métabolisme central (EMP, PP, ED, TCA) et énergétique (métabolisme des nucléotides) ont donc été mesurées pour chaque souche et chaque condition de croissance. Les différents échantillons ont été prélevés en phase exponentielle de croissance. Les moyennes de 3 réplicats biologiques indépendants sont représentées sous forme de heatmaps dans la figure 4 (moyennes±écart types en Annexe).

Les concentrations intracellulaires en métabolites de K12 et Nissle sont représentatives de la source de carbone supportant leur croissance. La taille des pools de la partie haute de la glycolyse est globalement plus faible sur substrat néoglucogénique (acétate) que sur substrats glycolytiques (gluconate, galactose, ribose). Les concentrations en 6PG et en Rib5P sont multipliées en moyenne par 65 et par 4,5 respectivement sur gluconate et sur ribose par rapport aux autres sources de carbone. De même, bien que le galactose-1-phosphate n'ait pu être quantifié de manière distincte (co-élution avec le glucose-1-phosphate et le mannose-1-phosphate lors de l'analyse LC-MS/MS), une augmentation de l'intensité de ce signal d'un facteur 10 est observée lors de croissance sur galactose, ce qui peut s'expliquer par une forte augmentation de la concentration en galactose-1-phosphate sur ce substrat. La diminution d'activité de CsrA entraine une réorganisation globale du métabolome, avec notamment une diminution significative des pools de métabolites d'entrée.

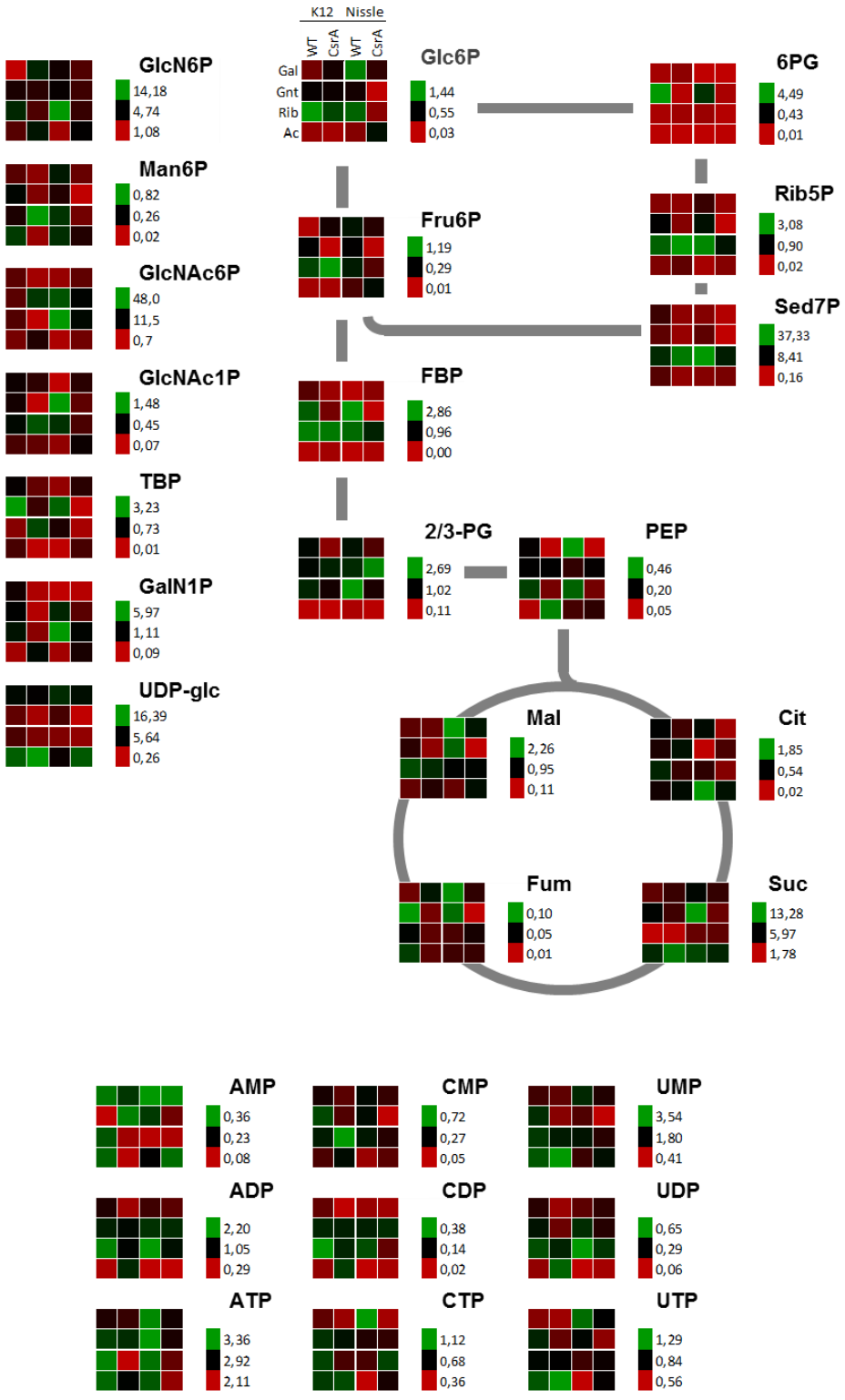


Figure 4. Profils métaboliques de souches sauvages d'*Escherichia coli* K12 MG1655 et Nissle 1917 et de leur mutant *csrA51* en phase exponentielle de croissance sur gluconate, galactose, ribose et acétate. Les concentrations sont exprimées en µmol.g_{DW}^{-1}, excepté pour le Sed7P, GalN1P, GlcN6P, GlcNAc6P, GlcNAc1P, UDP-Glc et TBP qui sont données en unité arbitraires.

Sur gluconate, la concentration en 6PG chez les mutants *csrA51* est divisée en moyenne par 65 par rapport aux souches sauvages, et retombe à un niveau similaire à ceux mesurés sur les autres sources de carbone. Une tendance identique est également observée sur ribose, bien que plus légère (non significatif pour K12, -60% pour Nissle). De même, le signal LC-MS correspondant aux pools d'hexoses-1-phosphate est diminué chez les mutants *csrA51* sur galactose. Cela suggère un contrôle par le système Csr des voies métaboliques en amont du nœud d'entrée dans le métabolisme carboné central, et donc potentiellement au niveau du transport ou des toutes premières voies cataboliques empruntées. Cette hypothèse est en accord avec les résultats d'(Edwards, Patterson-Fortin et al. 2011) qui ont observé une interaction entre CsrA et plusieurs transcrits codant pour des transporteurs de sucres.

La charge adénylate, un indicateur du statut énergétique de la cellule dont la valeur est comprise entre 0,7 et 0,9 en absence de limitation ou de stress, a été calculée à partir des concentrations intracellulaires en AMP, ADP et ATP. La moyenne entre les différentes conditions (0,82±0,03) et le faible écart type indiquent qu'*E. coli* maintient son statut énergétique quel que soit la source de carbone ou le niveau d'activité de CsrA.

Les profils métaboliques obtenus sont propres à chaque souche et constituent un exemple de la diversité métabolique fonctionnelle des *Escherichia coli*. Ils dépendent également de la source de carbone et du niveau d'activité de CsrA, et restent à ce jour difficilement interprétables sans l'utilisation d'outils permettant d'intégrer ces données.

Flux de carbone

Afin d'accéder aux distributions de flux intracellulaires pour les différentes souches et leur mutant, des cultures ont été réalisées sur gluconate, galactose, ribose et acétate marqués au ^{13}C. La composition isotopique des sources de carbone a été optimisée à l'aide d'IsoDesign dans le but de maximiser le nombre de flux calculables et la précision des flux calculés (chapitre 2). Le calcul des flux a été réalisé à partir des données isotopiques mesurées sur les métabolites intracellulaires par MS et sur les co-produits (acétate, pyruvate) par RMN. Les résultats, présentés figure 5, sont issus d'au moins 2 réplicats biologiques indépendants. Les sources de carbones sélectionnées entrent à différents points du métabolisme carboné central, c'est donc principalement selon ce facteur que diffèrent la direction et l'intensité des flux métaboliques.

E. coli possède deux systèmes de transport du **galactose** (codés par *galP* et *mglBAC*) qui est phosphorylé par la galactokinase (EC 2.7.1.6) et entre dans le métabolisme central au niveau du Glc6P, un carrefour métabolique entre les trois principales voies glycolytiques (EMP, ED, PP). La partition observée entre la glycolyse (70%), la partie oxydative de la voie des pentoses phosphates (13%) et la voie d'Entner-Doudoroff (15%) est conservée entre les différentes souches et leur mutant *csrA51*. La diminution d'activité de CsrA entraine une légère (mais significative) diminution de la part de carbone dirigée vers le cycle de Krebs (-16% pour K12, -7% pour Nissle).

Le **gluconate** peut entrer dans la cellule via quatre transporteurs différents (codés par *gntT*, *gntU*, *gntW* et *gntP* (Peekhaus and Conway 1998)) avant d'être phosphorylé en 6PG par la glucokinase (EC 2.7.1.12). Il est métabolisé principalement par la voie d'Entner-Doudoroff et la partie oxydative de la voie des pentoses phosphates dans des proportions similaires pour Nissle WT et K12 MG1655 WT (85-15% et 80-20%, respectivement). La diminution d'activité de CsrA modifie la partition du flux de carbone au niveau de ce nœud chez les deux souches, en faveur de la voie d'Entner-Doudoroff, alors que la contribution de la voie des pentoses phosphates diminue de 50% (partition ED-PP: 95-5% pour K12 et 90-10% pour Nissle). Les cartes de flux de Nissle WT et *csrA51* sont similaires à celles obtenues lors de croissance en erlenmeyer (chapitre 7). La diminution d'activité de CsrA se traduit également par une diminution de la production de pyruvate et d'acétate (-78% d'acétate formé chez K12, -25% chez Nissle), et une augmentation des flux relatifs dans le cycle de Krebs (+28% pour K12, +10% pour Nissle). Enfin, le flux relatif au travers de la PEP carboxylase augmente chez les mutants *csrA51* (+67% pour K12, +33% pour Nissle).

Le **ribose** peut entrer dans la cellule grâce à des perméases (codées par *rbsABC* et *xylABC*) et est phosphorylé en Rib5P par la ribokinase (EC 2.7.1.15) avant d'entrer dans le métabolisme central au niveau de la voie des pentoses phosphates. Le flux au travers de cette voie est particulièrement important et converge vers la glycolyse. Le flux néoglucogénique du Fru6P vers le Glc6P est faible et fournit seulement les besoins en précurseurs pour la formation de biomasse et de glycogène, aucun recyclage par la partie oxydative de la voie des pentoses phosphates n'étant observé. Le flux au travers de la partie basse de la glycolyse est orienté vers le cycle de Krebs, dont l'activité est plus élevée chez Nissle que chez K12 (+85%). La diminution d'activité de CsrA entraine

une diminution des flux du TCA, cet effet étant plus prononcé pour Nissle (-37%) que pour K12 (-14%), au profit de la production d'acétate (+41% pour K12, +100% pour Nissle). Comme observé sur gluconate, le flux relatif au travers de la PEP carboxylase est augmenté chez les mutants *csrA51* (+88%).

L'**acétate** entre dans la cellule par diffusion et via une perméase (codée par *actp*). Il est converti en acétylcoA (réaction catalysée par l'acétylcoA synthétase, EC 6.2.1.1), qui alimente le cycle de Krebs. La partition du flux de carbone entre l'isocitrate déshydrogénase (permettant d'assurer la production d'énergie) et l'isocitrate lyase (permettant de fournir le carbone nécessaire à la formation de biomasse) est hautement conservée entre les souches et les mutants (respectivement 75% et 25% pour l'isocitrate déshydrogénase et l'isocitrate lyase), malgré leur comportement de croissance différent. Le PEP et le pyruvate sont respectivement formés par la PEP carboxykinase et l'enzyme malique, le flux gluconéogénique observé est relativement faible et fournit les précurseurs pour la formation de biomasse. Aucun recyclage du carbone n'est observé au travers des voies des pentoses phosphates et d'Entner-Doudorroff. Enfin, alors qu'un contrôle négatif de CsrA sur l'activité de la PEP carboxykinase sur glucose est reporté dans la littérature, le flux au travers de cette voie diminue chez les mutants *csrA51* (-28%).

Bien que la quantité de CO_2 produit au cours de la croissance n'ait pas été mesurée, il a été possible d'estimer sa production sur la base des flux du métabolisme carboné central. A partir de ces données et des paramètres de croissance mesurés, les bilans carbone (ratio entre la quantité de carbone récupéré – biomasse et produits métaboliques – et consommé) ont été calculés pour l'ensemble des conditions (Tableau 3). Ils sont proches de 100%, ce qui confirme la cohérence des différents jeux de données.

Souche	**Source de carbone**											
	Gluconate			**Galactose**			**Ribose**			**Acetate**		
K12 MG1655 WT	0,99	±	0,04	0,98	±	0,04	0,90	±	0,09	0,99	±	0,06
K12 MG1655 *csrA51*	1,00	±	0,04	1,00	±	0,05	1,03	±	0,07	0,99	±	0,04
Nissle 1917 WT	1,00	±	0,04	1,00	±	0,07	1,03	±	0,02	0,96	±	0,05
Nissle 1917 *csrA51*	1,00	±	0,04	0,99	±	0,02	1,06	±	0,07	1,03	±	0,05

Tableau 3. Bilans carbone calculés pour les souches K12 MG1655 et Nissle 1917 sauvages et leur mutant *csrA51* sur gluconate, galactose, ribose et acétate.

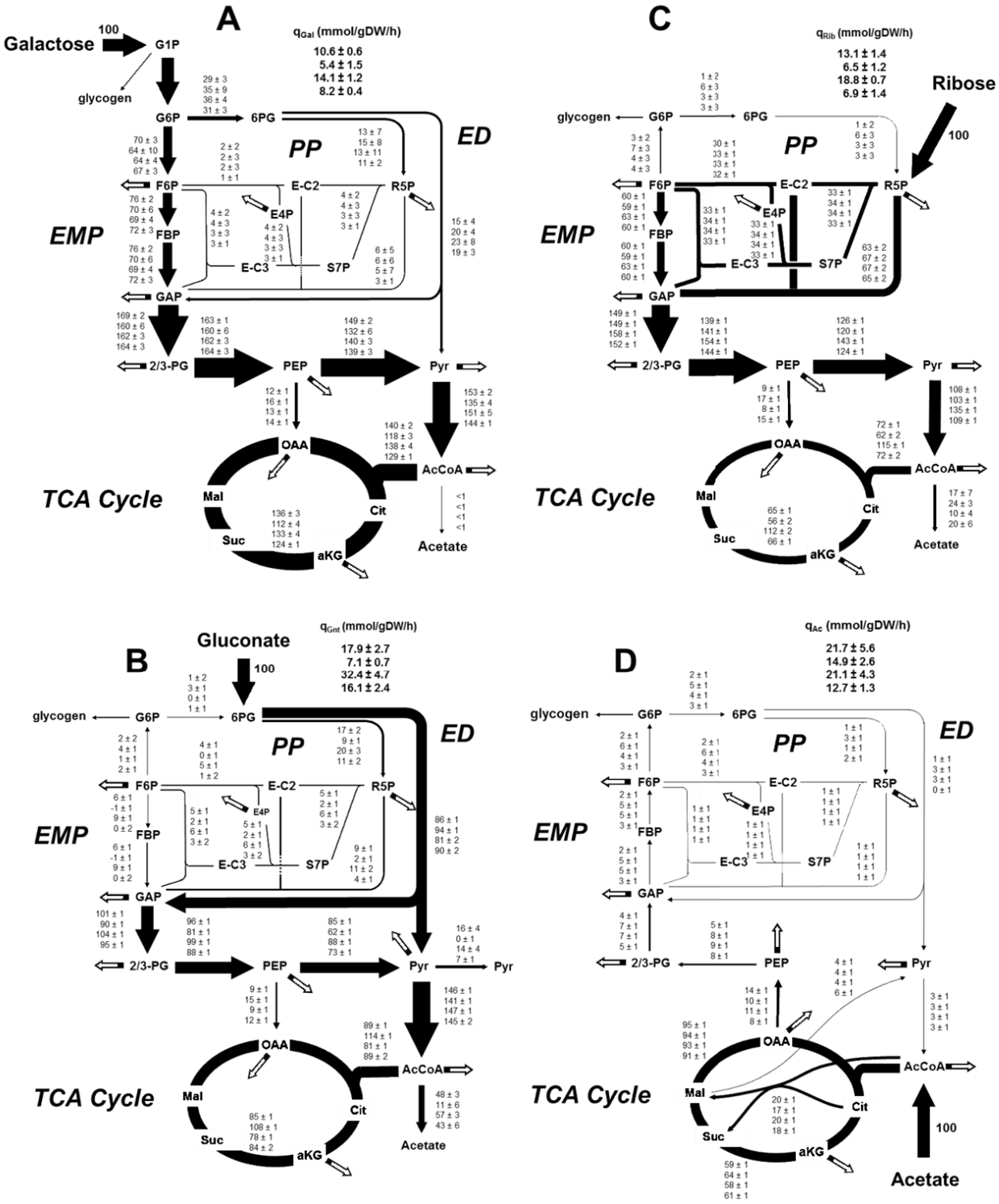


Figure 5. Distributions des flux de carbone chez *Escherichia coli* K12 MG1655 (WT et mutant *csrA51*) et Nissle 1917 (WT et *csrA51*) en phase exponentielle de croissance sur galactose (A), gluconate (B), ribose (C) et acétate (D). Les flux sont exprimés relativement à la vitesse d'entrée de carbone dont les valeurs sont données au-dessus de la carte de flux. De haut en bas: K12 MG1655 WT, K12 MG1655 *csrA51*, Nissle 1917 WT, Nissle 1917 *csrA51*.

Flux énergétique et rédox

Afin de caractériser le métabolisme énergétique et rédox des souches sauvages de K12 MG1655 et Nissle 1917 sur les sources de carbone sélectionnées, puis d'évaluer l'impact de la diminution d'activité de CsrA sur cette partie du métabolisme, les vitesses de production de NADPH et d'ATP ont été calculées à partir des distributions de flux de carbone et comparées aux besoins théoriques. Cette approche offre l'avantage majeur de ne pas reposer sur des hypothèses fortes sur l'utilisation des cofacteurs. Les résultats sont présentés figure 6.

Pour les souches sauvages, les flux de production d'ATP et de NADPH sont supérieurs aux besoins (anabolisme + maintenance) sur toutes les sources de carbone, comme observé lors de précédentes études sur glucose (Kayser, Weber et al. 2005; Nanchen, Schicker et al. 2006) et sur gluconate (chapitre 7). L'excès de production de NADPH (i.e. différence entre production et besoins théoriques) est plus important chez la souche K12 MG1655 que chez la souche Nissle 1917 (+47% chez Nissle, +61% chez K12). Concernant l'ATP, l'excès est plus important chez Nissle que chez K12 (en moyenne +61% chez Nissle et +44% chez K12). La part des différentes voies dans la production de NADPH (PP, TCA) et d'ATP (TCA, EMP, production d'acétate et phosphorylation oxydative) dépend principalement de la source de carbone.

La diminution d'activité de CsrA entraine une diminution globale des flux de production d'ATP et de NADPH, qui restent cependant suffisants pour satisfaire les besoins anaboliques. L'excès de production est largement inférieur à celui observé chez les souches sauvages pour toutes les sources de carbone. La diminution de l'activité de CsrA modifie également la part des différentes voies impliquées dans la production d'ATP et de NADPH, ces réorganisations dépendant là encore de la source de carbone.

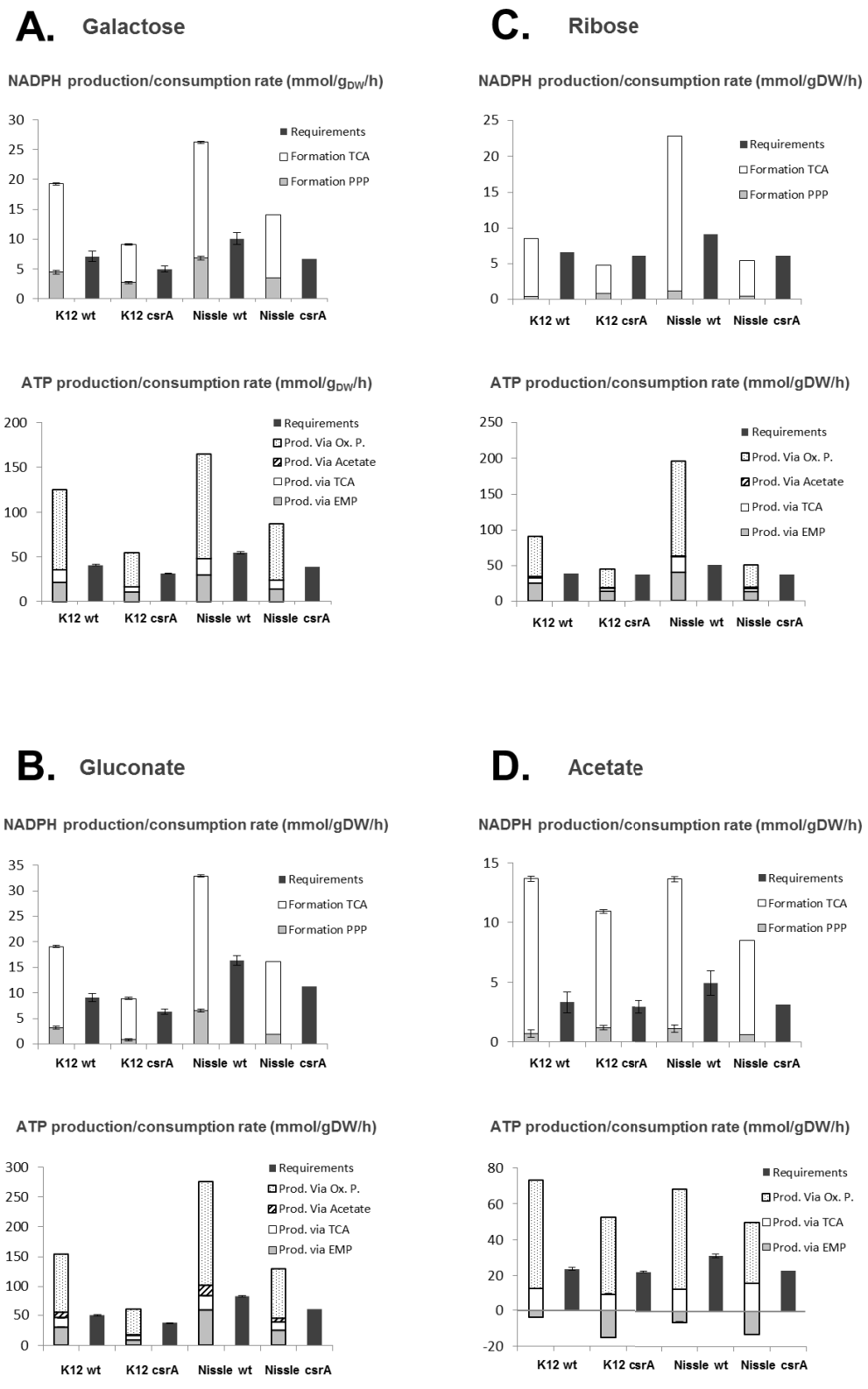


Figure 6. Bilans NADPH et ATP d'*E. coli* K12 MG1655 (WT et mutant *csrA51*) et Nissle 1917 (WT et mutant *csrA51*) lors de croissance sur galactose (A), gluconate (B), ribose (C) et acétate (D). Les vitesses de production ont été calculées à partir des flux de carbone produisant ces métabolites, les besoins théoriques ont été calculés à partir de la composition en biomasse, des taux de croissance et de la maintenance.

Intégration des données

Le manque de modèles mathématiques permettant d'intégrer les différents jeux de données quantitatives multi-échelles rend leur analyse particulièrement complexe. Cependant, des outils statistiques tels que l'analyse des corrélations ont été récemment appliqués avec succès pour l'analyse du métabolome, mettant en lumière des informations sur la structure et le fonctionnement des réseaux métaboliques (Camacho, de la Fuente et al. 2005; Steuer 2006; Boer, Crutchfield et al. 2010).

La majorité des études de ce type a été réalisée dans des conditions proches (e.g. plusieurs chemostats avec différentes limitations de la même source de carbone, le plus souvent du glucose) ou sur des temps courts lors de perturbations environnementales (e.g. un pulse de substrat lors de croissances en conditions de limitation), ce qui génère une variabilité relativement faible du contenu moléculaire et du fonctionnement métabolique. Une forte corrélation est couramment observée entre un nombre non négligeable de métabolites (Camacho, de la Fuente et al. 2005; Boer, Crutchfield et al. 2010). Les données présentées dans les sections précédentes ont été obtenues dans une large gamme de conditions (source de carbone, génome de la souche, et niveau d'activité de CsrA) pour lesquelles une diversité métabolique importante est observée aussi bien en termes de fonctionnement (paramètres de croissance, fluxome) que de contenu moléculaire (métabolome). L'analyse de ces différents jeux de données peut donc permettre de mettre en exergue des corrélations hautement conservées qui peuvent être considérées comme des propriétés du système (Camacho, de la Fuente et al. 2005). Cette approche ne permet pas d'établir de mécanisme ni de lien de causalité *ab initio* mais offre cependant des pistes qui peuvent être explorées par des études biochimiques.

Les coefficients de corrélations (R) ont été calculés entre les jeux de données en utilisant la méthode de Pearson, et la significativité des résultats a été évaluée sur la base de la p-value (seuil de significativité fixé à 0,05). L'analyse effectuée permet donc de capturer les relations linéaires entre les jeux de données, qui ne sont cependant qu'une partie des corrélations observables.

Intégration du métabolome

La nature de la source de carbone et le niveau d'activité de CsrA modifient fortement le réseau de régulation opérationnel et donc la quantité (et le niveau d'activité) des enzymes mais n'influent pas à priori sur leurs propriétés cinétiques (K_{cat}, K_M, K_i,...). Ces propriétés, qui dépendent des paramètres intracellulaires – pH, température, activité… – et de la séquence protéique codée dans le génome, peuvent différer entre deux souches. Une comparaison des séquences codantes des enzymes du métabolisme central des souches K12 MG1655 et Nissle 1917 permettrait de révéler des différences de séquence primaire de la protéine (et donc de possibles différences de propriétés cinétiques), cette analyse n'a cependant pas pu être réalisée car la séquence génomique de la souche Nissle 1917 n'est pas disponible à ce jour. Afin de ne pas biaiser les analyses en considérant des systèmes biologiques ayant potentiellement des propriétés cinétiques différentes, ce qui peut générer des différences de corrélations, ces analyses ont été réalisées indépendamment pour chaque souche dans un premier temps (Nissle 1917 WT + *csrA51* et K12 MG1655 WT + *csrA51*). Les matrices de corrélation obtenues sont présentées figure 7.

Une corrélation forte entre métabolites peut indiquer i) que les réactions entre ces métabolites fonctionnent proche de l'équilibre thermodynamique de la réaction, ii) que des facteurs communs contrôlent les concentrations de ces métabolites, ou iii) que ces métabolites font partie d'une structure métabolique particulière (e.g. un cycle métabolique).

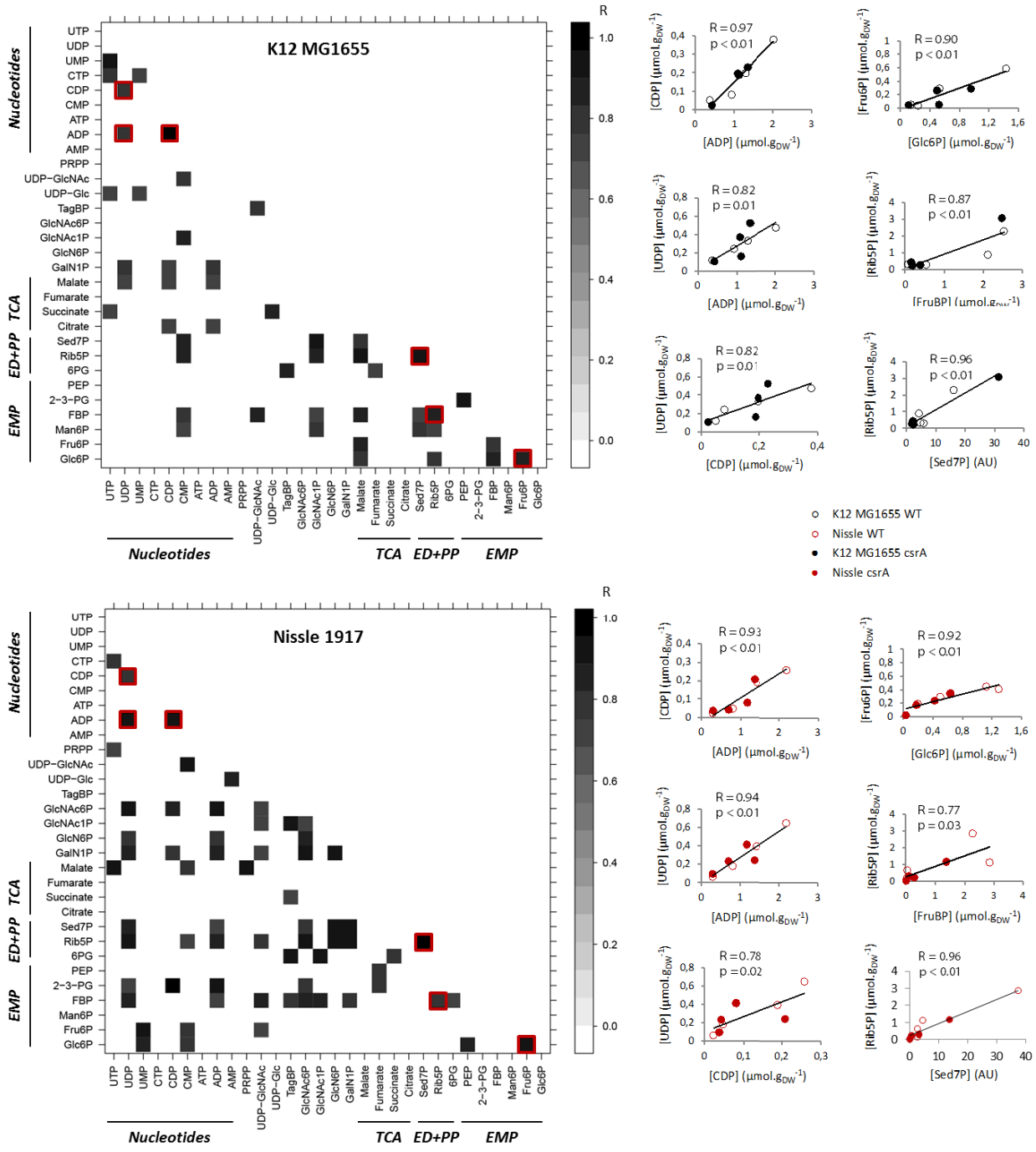


Figure 7. Matrices de corrélation entre les concentrations intracellulaires de différents intermédiaires métaboliques d'*Escherichia coli* K12 MG1655 (WT + *csrA51*) et Nissle 1917 (WT + *csrA51*) cultivés sur différentes sources de carbone. Les valeurs non-significatives (p > 0.05) ne sont pas représentées. Les corrélations encadrées sont représentées sur la droite.

Sur les 420 corrélations calculées pour chaque souche, 66 sont statistiquement significatives pour Nissle 1917, et 52 pour K12 MG1655. Un jeu consistant de 15

corrélations communes aux deux souches est observé (Tableau 4). Bien que le contenu métabolique de K12 diffère fortement de celui de Nissle, ces corrélations communes sont conservées lors de l'analyse effectuée sur les jeux de données combinés (Tableau 4 et Annexe), ce qui augmente la significativité de ces résultats et suggère que des propriétés communes aux deux systèmes (i.e. aux deux souches) en sont responsables.

Métabolites corrélés		K12		Nissle		K12 + Nissle	
		R	p-value	R	p-value	R	p-value
Glc6P	Fru6P	0,91	0,002	0,92	0,001	0,83	< 0,001
FBP	Rib5P	0,87	0,004	0,77	0,026	0,82	< 0,001
FBP	GlcNAc1P	0,71	0,048	0,88	0,004	0,78	< 0,001
6PG	TagBP	0,86	0,006	0,95	< 0,001	0,86	< 0,001
Rib5P	Sed7P	0,96	< 0,001	0,96	< 0,001	0,94	< 0,001
Rib5P	CMP	0,85	0,008	0,75	0,032	0,77	< 0,001
GalN1P	ADP	0,81	0,015	0,82	0,012	0,73	< 0,001
GalN1P	CDP	0,71	0,049	0,76	0,027	0,77	< 0,001
GalN1P	UDP	0,82	0,013	0,87	0,005	0,77	< 0,001
FBP	UDP-GlcNAc	0,85	0,008	0,87	0,005	0,81	< 0,001
ADP	CDP	0,97	< 0,001	0,93	0,001	0,66	< 0,001
CMP	UDP-GlcNAc	0,83	0,011	0,91	0,002	0,87	< 0,001
CTP	UTP	0,78	0,022	0,80	0,018	0,76	< 0,001
UDP	ADP	0,82	0,012	0,94	0,001	0,89	< 0,001
UDP	CDP	0,82	0,014	0,78	0,022	0,76	< 0,001

Tableau 4. Métabolites dont les concentrations sont significativement corrélées ($p < 0.05$) dans tous les jeux de données. Les coefficients de corrélation de Pearson (R) ont été calculés à partir des profils métaboliques de la souche K12 MG1655, de la souche Nissle 1917, et des profils combinés des deux souches. La combinaison des jeux de données entraine une augmentation importante de la significativité des corrélations, vraisemblablement liée à l'augmentation de la taille du jeu de données.

Une corrélation forte est observée entre des métabolites appartenant à la même voie métabolique ou à des voies connectées. A titre d'exemple, plusieurs métabolites de la partie haute de la glycolyse et de la voie des pentoses phosphates (Glc6P ↔ Fru6P, Rib5P ↔ Sed7P, Rib5P ↔ FBP) sont significativement corrélés, ce qui suggère que ces deux voies métaboliques interconnectées fonctionnent à l'équilibre de manière générale. De même, une corrélation est observée entre plusieurs nucléotides densément connectés (ADP ↔ CDP, UDP ↔ CDP, UDP ↔ ADP, CTP ↔ UTP). Certaines de ces corrélations sont cependant difficilement interprétables sur le plan biologique (e.g. TagBP ↔ 6PG) et

pourraient se révéler être de faux positifs, qu'une approche de type "false discovery rate" permettait d'identifier.

Lorsqu'une enzyme fonctionne à l'équilibre, son ratio d'action de masse (MAR, égal au ratio entre les concentrations des produits et des substrats) est proche de la constante d'équilibre thermodynamique K_{eq} de la réaction. Cette constante dépend des paramètres thermodynamiques des réactants à l'état standard (ΔrG_0) et du milieu environnant (température, activité, pH…), et elle est indépendante de la concentration et des propriétés de l'enzyme. Il est classiquement admis que la réaction d'isomérisation entre le Glc6P et le Fru6P, catalysée par la glucose-6-phosphate isomerase (pgi, EC 5.3.1.9), fonctionne proche de son équilibre thermodynamique lorsque les flux glycolytiques sont importants (Camacho, de la Fuente et al. 2005; Taymaz-Nikerel, de Mey et al. 2009; Canelas, Ras et al. 2011). Cela se traduit par une forte corrélation entre les concentrations en Glc6P et Fru6P, cette corrélation semblant disparaitre lorsque les flux glycolytiques sont faibles. Les concentrations de ces deux métabolites sont fortement corrélées dans notre jeu de données (Tableau 4). Nous avons donc calculé le MAR moyen de cette enzyme dans les différentes conditions à partir de la pente *a* de la droite de régression [Fru6P] = a×[Glc6P]. Sa valeur (0,32±0,05) est du même ordre de grandeur que le K_{eq} déterminé expérimentalement lors de précédentes études (Tableau 5). Ces résultats suggèrent donc que la réaction catalysée par la pgi a un fonctionnement proche de l'équilibre thermodynamique bien que le flux varie dans une gamme extrêmement large (0,4 à 9 $mmol.g_{DW}^{-1}.h^{-1}$) et qu'un fonctionnement gluconéogénique soit observé dans la majorité des conditions (12/16).

Valeur de K_{eq}					
Taymaz-Nikerel *et al.* (2009)	Canelas *et al.* (2011)	Goldberg *et al.* (2004)	Alberty (2006)	Chassagnole *et al.* (2002)	$[F6P]/[G6P]_{moy}$ dans ces travaux
0.27 ± 0.03	0.259 ± 0.003	0.32 ± 0.08	0.28	0.17	0.32 ± 0.05

Tableau 5. Constantes d'équilibre de la réaction catalysée par la glucose-6-phosphate isomérase issues de la littérature et déterminé dans les travaux présentés.

L'analyse des corrélations entre les concentrations en intermédiaires métaboliques met donc en évidence des corrélations souche-dépendantes difficilement interprétables à ce jour, mais également un jeu commun de données corrélées cohérentes avec la topologie

du réseau métabolique central et informatives sur le fonctionnement du métabolisme d'*E. coli*.

Intégration métabolome-fluxome

Dans le but d'identifier des métabolites potentiellement rapporteurs de l'intensité des flux métaboliques, nous avons analysé les corrélations entre les concentrations en métabolites et les vitesses des différentes réactions (Figure 8).

Sur les 780 corrélations calculées pour chaque souche, 116 sont statistiquement significatives pour Nissle 1917, et 67 pour K12 MG1655. Environ la moitié des 30 métabolites quantifiés (16 pour K12 et 19 pour Nissle) sont de potentiels indicateurs de l'intensité d'un ou de plusieurs flux. Dans la majorité des corrélations observées, les métabolites sont les produits ou les substrats des réactions des voies dont font partie les flux. Le Rib5P et le Sed7P corrèlent avec les flux de la partie non oxydative de la voie des pentoses phosphates, le fumarate avec les flux du TCA, les pools de 2- et 3-PG avec la partie basse de la glycolyse, et le 6PG avec les flux de la voie d'Entner-Doudoroff et de la partie oxydative de la voie des pentoses phosphates. La concentration en ATP corrèle avec les flux du TCA – une des principales voies de production d'énergie – chez la souche Nissle 1917, une tendance similaire étant également observée chez la souche K12 MG1655.

De manière plus surprenante, la concentration intracellulaire en ADP corrèle positivement avec les flux au travers de la partie non oxydative de la voie des pentoses phosphates, ce qui suggère que l'ADP est un régulateur de cette voie *in vivo* chez *E. coli*. Ce métabolite est en effet un inhibiteur compétitif des transcétolases *in vitro* chez d'autres organismes (Ashihara and Komamine 1974; Hosomi, Tara et al. 1989). Le signe des corrélations que nous observons peut être expliqué par l'augmentation des ratios [R5P]/[ADP] et [S7P]/[ADP] avec l'intensité de ces flux ([R5P]/[ADP] ↔ tk2: R=0,57, p=0,14 pour K12 MG1655 ; R=0,68, p=0,06 pour Nissle 1917), la compétition entre substrats et inhibiteur pour l'enzyme basculant ainsi en faveur des substrats.

Les corrélations observées semblent résulter des relations thermodynamiques directes existant entre les flux et les métabolites impliqués dans les réactions. L'impact de ces résultats est discuté plus largement dans la section 5.

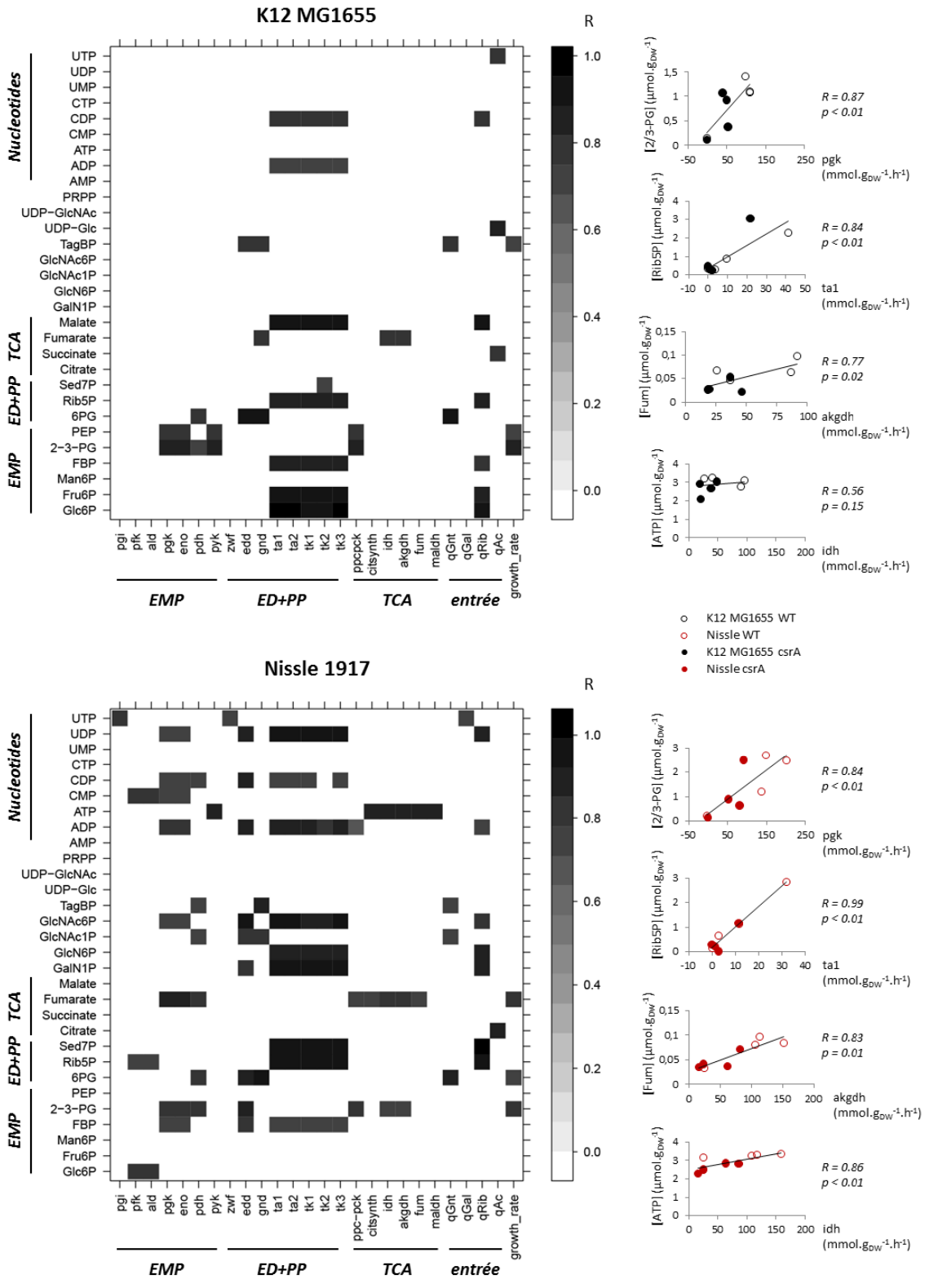


Figure 8. Matrices de corrélation entre les concentrations intracellulaires de différents intermédiaires métaboliques et les flux de carbone chez les souches K12 MG1655 (WT + *csrA51*) et Nissle 1917 (WT + *csrA51*) en phase exponentielle de croissance sur différentes sources de carbone. Les valeurs non-significatives ($p > 0.05$) ne sont pas représentées. Des exemples de corrélations significatives sont représentés sur la droite de chaque matrice.

Intégration fluxome-phénotype

Les besoins anaboliques en précurseurs, en énergie, et en cofacteurs réduits doivent être assurés par les processus cataboliques afin de soutenir la croissance. Nous avons donc analysé les relations entre le taux de croissance et les flux de consommation de carbone, de production d'énergie (ATP) et de cofacteurs réduits (NADH et NADPH).

Il apparait que la vitesse d'entrée de carbone (exprimée en mmolC.g_{DW}^{-1}.h^{-1}) et les flux énergétique et rédox sont fortement corrélés au taux de croissance pour les différentes souches sur les sources de carbone sélectionnées. Bien qu'il soit classiquement admis que ces flux cataboliques et anaboliques sont corrélés (Taymaz-Nikerel, Borujeni et al. 2010), cette étude est à notre connaissance la première à le démontrer dans des conditions radicalement différentes en termes de fonctionnement métabolique. Ces résultats sont renforcés par l'approche utilisée (^{13}C-MFA) qui ne repose pas sur des hypothèses fortes telles que les besoins théoriques en cofacteurs ou en ATP. Les paramètres de régression linéaire entre les souches sauvages ne diffèrent pas significativement, ce qui est également observé entre les mutants *csrA51*. En revanche, les paramètres calculés pour les souches sauvages diffèrent de ceux déterminés pour les mutants. Les résultats présentés (tableau 6 et figure 9) séparent donc les souches sauvages (K12 + Nissle WT) de leur mutant (K12 + Nissle *csrA51*).

		souches sauvages				**mutants *csrA51***			
y	x	a	y_0	R	p-value	a	y_0	R	p-value
q_S[a]	μ[c]	198	-12,0	0,96	< 0,001	137	-6,2	0,93	< 0,001
Total ATP prod.[b]	μ[c]	286	-6	0,96	< 0,001	186	-6	0,93	< 0,001
ATP prod. via OP[b]	μ[c]	161	12,7	0,94	< 0,001	93	10,3	0,67	0,122
ATP prod. via CM[b]	μ[c]	125	-19	0,97	< 0,001	93	16,4	0,97	< 0,001
Excess ATP[b]	μ[c]	210	-23	0,93	< 0,001	110	-14	0,82	0,012
NADPH prod.[b]	μ[c]	29	5,0	0,90	0,003	13	5,0	0,50	0,206
NADH prod.[b]	μ[c]	140	1,00	0,96	< 0,001	93	0,70	0,91	0,002
TCA flux[a,d]	μ[c]	206	-12,0	0,95	< 0,001	159	12,0	0,88	0,004

[a] mmolC.g_{DW}^{-1}.h^{-1}; [b] mmol.g_{DW}^{-1}.h^{-1}; [c] h^{-1}; [d] flux entrants - flux sortants du TCA

Tableau 6. Corrélation observées entre certains flux (carbone, énergie et cofacteurs réduits) et le taux de croissance de souches sauvages d'*E. coli* (K12 MG1655 et Nissle 1917) et de leur mutant *csrA51*. Les paramètres de régression linéaire (pente *a* et ordonnée à l'origine y_0) sont donnés pour chaque corrélation, qui sont évaluées sur le plan statistique à partir de coefficient de Pearson (R) et de la p-value.

La vitesse de consommation de carbone est fortement corrélée au taux de croissance, de même que les flux intracellulaires de carbone au travers du TCA (qui produit une part significative des acides aminés indispensables à la synthèse de protéines et de l'énergie nécessaire à la croissance) et les flux de formation d'ATP et de cofacteurs réduits. La concentration intracellulaire en ATP reste en revanche stable quel que soit le taux de croissance alors que sa vitesse de production augmente de manière linéaire, ce qui indique une augmentation du turn-over de l'ATP avec le taux de croissance. Cela permet à la cellule de maintenir sa charge adénylate, et donc son statut énergétique, stables. L'excès d'ATP produit est également proportionnel au taux de croissance, il alimente donc vraisemblablement des processus liés à la croissance (pouvant inclure une partie de la maintenance) dont les besoins sont actuellement largement sous-estimés. La balance entre l'anabolisme (i.e. le taux de croissance) et le métabolisme énergétique, rédox et carboné semble donc similaire pour les souches Nissle et K12, bien que des différences importantes soient observées aux niveaux moléculaire (métabolome), fonctionnel (fluxome) et macroscopique (q_S, μ, $Y_{co\text{-}produits}$, Y_X).

La diminution d'activité de CsrA modifie significativement la balance entre anabolisme et catabolisme. La production de NADPH n'est plus corrélée au taux de croissance, de même que la production d'ATP via la phosphorylation oxydative. En revanche, les vitesses de production d'ATP (totale) et de NADH restent corrélées au taux de croissance. Outre ces différences qualitatives, des différences quantitatives sont également observées entre les souches sauvages et leur mutant *csrA51*. Les pentes des droites de régression sont significativement plus faibles chez les mutants (-35% en moyenne), ce qui révèle une production énergétique et rédox et une consommation de carbone plus faibles pour un même taux de croissance. Le statut énergétique du mutant *csrA51* (concentration intracellulaire en ATP et charge adénylate) n'est cependant pas significativement différent de celui de la souche sauvage. Au vue de ces résultats, il est clair que les différences observées entre les souches sauvages et les mutants *csrA51* ne sont pas dues à la différence de taux de croissance mais bien à une réorganisation complète du métabolisme. Afin de déterminer si la diminution des différents paramètres est liée à la diminution de la consommation de carbone, nous avons calculé les rendements de production de biomasse, de co-produits+CO_2, d'ATP, de NADH et de NADPH pour chaque condition. Les ratios des rendements entre les mutants *csrA51* et les souches sauvages (Figure 10) soulignent une diminution du rendement de production

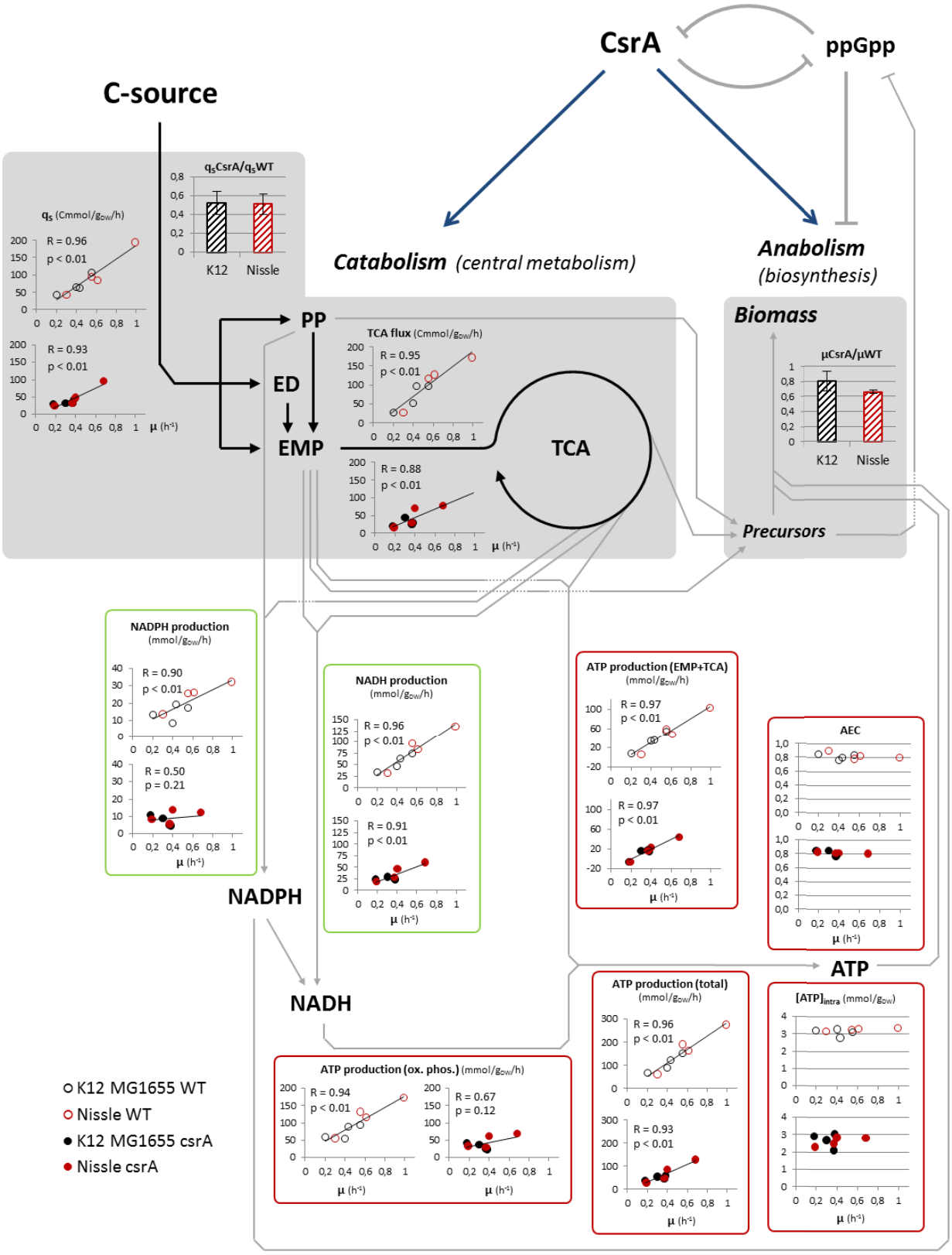


Figure 9. Corrélation observées entre le taux de croissance et certains flux anaboliques et cataboliques de souches sauvages d'*E. coli* (K12 MG1655 et Nissle 1917) et de leur mutant *csrA51*. Les graphes encadrés de vert et de rouge correspondent respectivement au métabolisme rédox et énergétique.

d'ATP chez les mutants *csrA51* (-12%), alors que les rendements de production de cofacteurs réduits sont comparables entre les souches sauvages et leur mutant *csrA51*. La diminution de la vitesse de production de NADH et de NADPH résulte donc en partie de la diminution de la vitesse de consommation de carbone. De plus, la part de carbone utilisé pour la formation de biomasse est supérieure chez le mutant *csrA51* (+47%), au détriment des co-produits et du CO_2 (-14%), ce qui implique une redirection des flux de carbone, de NADPH et d'ATP en faveur de l'anabolisme. Le système Csr exerce donc un contrôle fort sur le métabolisme carboné, mais vraissemblablement également rédox et énergétique. Ce rôle est qualitativement et quantitativement similaire chez la souche K12 MG1655 et chez la souche Nissle 1917.

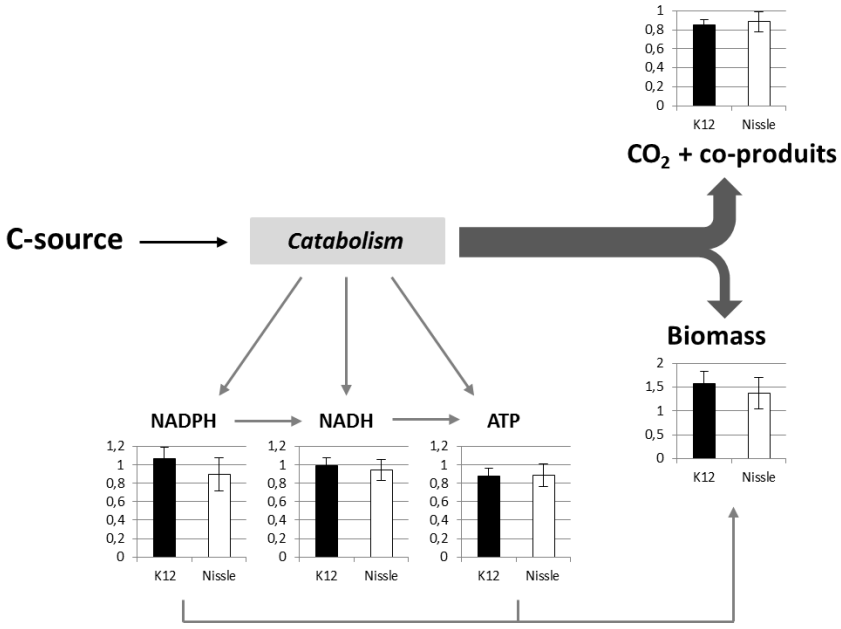


Figure 10. Ratios des rendements de production de différents produits (NADPH, NADH, ATP, biomasse, CO_2+co-produits) entre les mutants *csrA51* et les souches sauvages d'*E. coli* K12 MG1655 et Nissle 1917.

5. Discussion

Diversité métabolique fonctionnelle des *Escherichia coli*: K12 MG1655 vs Nissle 1917

Les résultats obtenus constituent tout d'abord un exemple la diversité métabolique fonctionnelle des *E. coli*, en révélant d'importantes différences entre les capacités métaboliques de la souche de laboratoire K12 MG1655 et de la souche Nissle 1917, un excellent colonisateur de l'intestin. La forte capacité de colonisation et de persistance de

la souche Nissle 1917 résulte non seulement de la présence de facteurs d'interactions avec l'hôte (adhésines, capsule, etc) et de compétition avec le microbiote (microcines, sidérophores, etc) (Grozdanov, Raasch et al. 2004; Sun, Gunzer et al. 2005), mais également de voies de dégradations spécifiques (e.g. catabolisme du déoxyribose) (Sun, Gunzer et al. 2005; Martinez-Jehanne, du Merle et al. 2009). Nos résultats démontrent que le fonctionnement du métabolisme de Nissle 1917 est particulièrement efficace, avec des taux de croissance, des vitesses de consommation de carbone, et des rendements en biomasse supérieurs à ceux de la souche K12 MG1655 pour les différentes sources de carbone considérées. L'efficacité métabolique supérieure de la souche Nissle 1917 résulte donc de flux cataboliques et anaboliques plus rapides. Le fonctionnement de son métabolisme constitue ainsi un facteur supplémentaire favorisant son implantation et sa persistance dans l'intestin, aussi bien de manière directe (multiplication plus rapide) qu'indirecte (compétition avec le microbiote pour les ressources nutritionnelles disponibles). Il est indispensable d'étendre ce type d'approches à d'autres souches d'*E. coli* (et notamment d'autres *E. coli* du groupe phylogénétique B2) afin d'analyser plus largement le lien entre l'efficacité métabolique des souches et leur capacité de colonisation et de persistance chez l'hôte.

Métabolisme des *Escherichia coli*

Bien que les résultats présentés ne constituent qu'une première étape d'analyse, ils révèlent de précieuses informations sur le fonctionnement du métabolisme central d'*E. coli*.

L'analyse des corrélations entre fluxomes et métabolomes a permis d'identifier de métabolites potentiellement rapporteurs des intensités de certains flux métaboliques. Ces métabolites appartiennent à la voie dont le flux est rapporté (e.g. Rib5P ↔ PP) ou sont des régulateurs des enzymes catalysant ces réactions (e.g. ADP ↔ transcétolases). Cette approche est donc prometteuse pour identifier de nouveaux régulateurs enzymatiques, des études biochimiques permettant de valider *in vitro* le rôle de ces métabolites candidats. De plus, certains flux pourraient être estimés par la mesure des concentrations de leurs métabolites rapporteurs. Cela nécessite de construire des modèles stœchiométriques sur lesquels des contraintes basées sur la concentration absolue des métabolites et/ou de ratios de ces concentrations pourraient être

appliquées. Il est nécessaire d'élargir ces travaux à d'autres conditions afin de confirmer ces résultats et d'évaluer les approches proposées.

L'intégration des données physiologiques et des flux métaboliques révèle que les vitesses de consommation de carbone, de production d'énergie (ATP), et de production de cofacteurs réduits (NADH et NADPH) évoluent linéairement avec le taux de croissance, et ce indépendamment de la source de carbone. Cela semble donc être une propriété fondamentale du fonctionnement métabolique d'*E. coli*, conservée dans diverses conditions nutritionnelles. De plus, bien que d'importantes différences de fonctionnement métabolique soient observées entre les souches K12 MG1655 et Nissle 1917, les vitesses de ces flux sont identiques pour les deux souches pour un même taux de croissance, ce qui pourrait suggérer que cette propriété ne dépend pas de la souche et serait donc conservée lors de l'évolution. Bien entendu, ces hypothèses demandent à être validées par la caractérisation du métabolisme d'autres souches d'*E. coli* et d'autres sources de carbone.

Rôle du système Csr

Les résultats présentés démontrent que le système Csr ne contrôle pas seulement le métabolisme carboné d'*E. coli* mais est un régulateur majeur de son métabolisme énergétique et rédox. Ces travaux étendent donc le rôle du système Csr dans la nutrition carbonée et dans le contrôle du métabolisme central d'*E. coli*, ce rôle étant qualitativement et quantitativement similaire chez les souches K12 MG1655 et Nissle 1917.

Contrôle du métabolisme carboné, énergétique et rédox par le système Csr

Le système Csr permet d'augmenter la compétitivité d'*E. coli* pour l'utilisation de sources de carbone glycolytiques (galactose, gluconate, ribose) – et ce quelle que soit la voie métabolique empruntée (EMP, ED, PP) –, mais également pour l'utilisation de l'acétate, une source de carbone gluconéogénique. Le système Csr augmente le taux de croissance et la vitesse de consommation de substrat, au prix d'une diminution du rendement en biomasse. Les réorganisations du métabolisme central observées chez les mutants *csrA51* dépendent fortement de la source de carbone, étendant ainsi les résultats obtenus dans le chapitre 7. Tout d'abord, la diminution d'activité de CsrA entraîne une diminution globale des flux du métabolisme central. En termes de distribution de

carbone, les modifications des flux relatifs de la partie haute de la glycolyse observées sur gluconate sont cohérentes avec l'effet antagoniste de CsrA sur la glycolyse et la gluconéogénèse. En revanche, aucune différence n'est observée dans cette partie du métabolisme entre les mutants et les souches sauvages sur acétate, ribose ou galactose. Le système Csr exerce également un contrôle sur les flux au travers de la voie des pentoses phosphates sur gluconate, ainsi que sur les flux du cycle de Krebs dont la contribution augmente (sur gluconate) ou diminue (sur galactose et ribose). Enfin, les changements observés sur les flux au travers de la PEP carboxykinase vont dans le sens opposé à celui attendu à partir de la mesure des activités enzymatiques lors de croissances sur glucose (Sabnis, Yang et al. 1995; Wei, Shin et al. 2000). Il est clair que les réorganisations observées ne peuvent être prédites à partir des seuls effets de CsrA sur ses cibles, ce qui confirme l'intérêt des approches systémiques pour appréhender ce type de problématique. Au vue de ces résultats, considérer le système Csr comme un simple activateur de la glycolyse et inhibiteur de la gluconéogenèse semble être une vision réductrice et simpliste de son rôle sur le métabolisme central d'*E. coli*, qui est bien plus global.

Etant donné que le niveau d'activité de CsrA affecte le taux de croissance, et que le métabolome et le fluxome sont également affectés par ce paramètre (Tweeddale, Notley-McRobb et al. 1998; Nanchen, Schicker et al. 2006; Schaub and Reuss 2008), on peut se demander si les réorganisations métaboliques observées ne résultent pas simplement de la diminution du taux de croissance. L'intégration des données physiologiques et des flux métaboliques indique clairement que ce n'est pas le cas, et suggèrent que le système Csr pourrait être impliqué dans le contrôle de la balance anabolisme-catabolisme d'*E. coli*. En effet, à taux de croissance identique, la diminution d'activité de CsrA abouti à:

- une diminution de la vitesse de consommation des sources de carbone
- une diminution de la production d'ATP et de cofacteurs réduits (NADH et NADPH), liée aux réorganisations des flux intracellulaires et à la diminution de la consommation de carbone
- une augmentation de la part de carbone, d'énergie et de cofacteurs réduits dirigée vers la formation de biomasse. Cette observation peut en partie s'expliquer par les changements physiologiques importants observés chez le mutant *csrA51* (e.g. la diminution de la motilité qui consomme une part

significative de l'énergie cellulaire), et suggère une réorganisation du métabolisme secondaire.

Le système Csr contrôle le métabolisme d'*E. coli* par des mécanismes aussi bien directs qu'indirects. Ce système est en effet interconnecté à d'autres réseaux de régulation globaux (e.g. la réponse stringente via le ppGpp, l'entrée en phase stationnaire via sigma S, la répression catabolique via Crp), ce qui permet à *E. coli* d'assurer un ajustement précis et coordonné de son métabolisme et de sa physiologie en fonction de divers stimuli environnementaux (densité bactérienne, disponibilité et nature des nutriments, stress oxydatif, pH, température, etc). Des investigations complémentaires par des approches de transcriptomique et de protéomique, à la lumière des données métabolomiques et fluxomiques présentées ici, pourraient permettre d'identifier et de quantifier le contrôle exercé par chacun de ces systèmes, qui peut être coopératif ou antagoniste. Il est également important d'analyser plus en détail les paramètres environnementaux affectant le système Csr.

Rôle du système Csr dans la réponse aux stress oxydatifs

Le système Csr contrôle donc non seulement le métabolisme carboné et énergétique, mais également le métabolisme rédox en augmentant la production de NADH et de NADPH. Ce résultat apporte un éclairage nouveau sur certaines études récemment publiées.

En effet, il a été montré que CsrA favorise *in vitro* et *in vivo* la résistance aux stress oxydatifs chez *Helicobacter pylori* (Barnard, Loughlin et al. 2004) et *Campylobacter jejuni* (Fields and Thompson 2008), via un mécanisme encore inconnu. Le rôle de Csr dans la réponse aux stress oxydatifs n'a pas été étudié à ce jour chez *E. coli*. De récents travaux ont cependant démontré que DksA, une protéine contrôlant CsrA en régulant l'expression de *csrB* et *csrC* (Edwards, Patterson-Fortin et al. 2011), favorise la résistance de *Salmonella enterica* aux mécanismes de défenses radicalaires mis en place par les cellules hôtes lors de processus d'invasion en coordonnant le niveau de NADPH et l'expression des gènes de biosynthèse du glutathion (*gshA* et *gshB*), tous deux assurant la détoxication et l'élimination d'espèces réactives de l'oxygène (Henard, Bourret et al. 2010). Un contrôle indirect par DksA de la voie des pentoses phosphates et du TCA est également observé (Henard, Bourret et al. 2010). Cependant, le mécanisme par lequel

DksA contrôle le métabolisme carboné et rédox n'est pas établi. Le niveau d'activité de CsrA a un effet négligeable sur l'expression de DksA (Edwards, Patterson-Fortin et al. 2011), le contrôle du métabolisme rédox que nous observons ne fait donc pas intervenir DksA. Puisque DksA régule le niveau d'expression de *csrA*, nos résultats suggèrent que le contrôle du métabolisme carboné et de la production de NADPH exercé par DksA se fait par l'intermédiaire de CsrA. Le système Csr pourrait donc jouer un rôle majeur dans la réponse au stress oxydatif chez *E. coli* en augmentant notamment le flux de production de cofacteurs réduits.

Rôle du système Csr dans l'adaptation d'E. coli aux conditions nutritionnelles

L'intestin est un environnement très variable où se succèdent des périodes d'abondance de nutriments et de famine. La guanosine tétraphosphate (ppGpp), un indicateur du statut nutritionnel de la cellule, est produite par la ppGpp synthétase (codée par *relA*) lorsque les concentrations intracellulaires en acides aminés sont faibles (Traxler, Zacharia et al. 2011). Le ppGpp active la réponse stringente, qui aboutit à une diminution de la synthèse de macromolécules (ADN, ARN, protéines, acides gras), à de profondes modifications physiologiques (diminution de la motilité, formation de biofilms), et à l'inhibition de CsrA via un contrôle positif de l'expression de *csrB* et *csrC* (Edwards, Patterson-Fortin et al. 2011; Traxler, Zacharia et al. 2011). CsrA inhibe quant à lui la synthèse de ppGpp via un contrôle négatif de l'expression *relA*. Ces deux systèmes sont donc connectés par plusieurs boucles de rétrocontrôle, et 40% des gènes co-purifiés avec CsrA sont également contrôlés par le ppGpp et/ou DksA lors la réponse stringente (Edwards, Patterson-Fortin et al. 2011). L'effet opposé de ces deux systèmes sur l'expression de gènes cibles communs (e.g. *flhDC*, *glgCAP*) assure un contrôle précis et rapide de leur niveau d'expression en réponse à des changements environnementaux, à la fois au niveau transcriptionnel et post-transcriptionnel.

En période d'abondance nutritionnelle, l'absence de limitation carbone et la quantité importante d'acides aminés dans la cellule inhibe la synthèse de ppGpp, ce qui augmente le niveau d'activité de CsrA. Nos résultats suggèrent que ce dernier pourrait alors coordonner le métabolisme central (carboné, énergétique et rédox) et l'anabolisme afin de maximiser le taux de croissance, un facteur de compétitivité majeur dans l'intestin. En période de limitation en nutriments, l'augmentation de la concentration en ppGpp diminue le niveau d'activité de CsrA. Cela pourrait se traduire par une réorganisation

profonde du métabolisme visant à économiser les ressources nutritionnelles, i.e. une part plus importante du carbone consommé est utilisée pour la formation de biomasse, au détriment du taux de croissance. Nos résultats, bien qu'ayant été obtenus lors d'états métaboliques stables, suggèrent que le système Csr assure un ajustement permanent de l'utilisation du carbone, de la production d'énergie, de la production de cofacteurs réduits, et du taux de croissance en réponse aux ressources nutritionnelles disponibles. Le système Csr favoriserait ainsi l'adaptation d'*E. coli* dans l'intestin, et donc sa persistance.

6. Conclusion

Les différentes approches systémiques quantitatives utilisées lors de ces travaux ont permis de caractériser le fonctionnement du métabolisme central d'*E. coli* sur des sources de carbone capables de supporter son implantation dans l'intestin et représentatives de ses principales voies cataboliques (PP, ED, EMP, TCA). Cette étude, menée en parallèle chez la souche K12 MG1655 et chez la souche Nissle 1917, a mis en avant des différences importantes en termes de comportement physiologique. En effet, Nissle 1917 présente un métabolisme particulièrement efficace comparé à celui de K12 MG1655, ce qui explique en partie sa meilleure capacité de colonisation et de persistance. Malgré ces différences, des propriétés fondamentales de leur fonctionnement métaboliques, communes aux deux souches, ont été identifiées. Un des résultats majeurs est le lien observé entre l'anabolisme et les processus cataboliques, énergétiques, et rédox d'*E. coli*, qui est indépendant de la source de carbone. L'analyse des données générées n'ayant pas été menée à son terme, il est possible que d'autres aspects du fonctionnement métabolique d'*E. coli* émergent de ces travaux. L'utilisation de modèles cinétiques du métabolisme (que ces données peuvent être à même de calibrer) pourrait notamment fournir un éclairage nouveau sur les propriétés de ces systèmes métaboliques.

Nos observations étendent le rôle du système Csr dans le contrôle du métabolisme central d'*E. coli*, qui est similaire chez les deux souches étudiées. Le système Csr n'est vraisemblablement pas seulement un acteur du basculement glycolyse-néoglucogénèse mais régule l'ensemble du métabolisme central, le contrôle exercé sur chaque voie dépendant hautement de la source de carbone. Un contrôle du métabolisme rédox et

énergétique par le système Csr, non reporté à ce jour, est également démontré. Ce résultat s'ajoute au faisceau d'arguments suggérant un rôle du système Csr dans la réponse aux stress oxydatifs chez *E. coli*. De manière globale, Csr pourrait assurer un ajustement permanent du catabolisme, de l'anabolisme et des processus énergétiques et rédox en réponse aux conditions environnementales, et en particulier aux ressources nutritionnelles disponibles.

Cela renforce l'hypothèse d'un rôle du système Csr dans la dynamique adaptative d'*E. coli* lors de changements de sources de carbone. Il a notamment été montré que le système Csr affecte l'expression de nombreux gènes en phase diauxique, et le système à deux composantes BarA/UvrY, qui active l'expression de *csrB* et *csrC*, est nécessaire dans l'adaptation du métabolisme lors de certains changements nutritionnels (Pernestig, Georgellis et al. 2003). Le système Csr est également impliqué dans l'expression de la virulence chez de nombreux organismes (Barnard, Loughlin et al. 2004; Heroven, Bohme et al. 2008; Lucchetti-Miganeh, Burrowes et al. 2008; Bhatt, Edwards et al. 2009; Sze and Li 2011; Palaniyandi, Mitra et al. 2012), dont *E. coli*, et un faisceau d'arguments semble relier l'expression de toxines à des changements de nutrition carbonée (Homburg, Oswald et al. 2007; Li, Granat et al. 2008; Le Bouguenec and Schouler 2011). L'analyse du rôle du système Csr dans la dynamique adaptative d'*E. coli* (capacité à changer de substrat, à alterner phase de carence et phase d'abondance nutritionnelle) revêt donc un intérêt majeur pour mieux comprendre i) les mécanismes permettant l'implantation d'*E. coli* chez l'hôte et ii) les bases physiologiques sur lesquelles s'expriment les caractères propres à chaque souche, et en particulier les facteurs de virulence chez les pathogènes.

7. Références

Ashihara, H. and A. Komamine (1974). "Regulation of the activities of some enzymes of the pentose phosphate pathway in *Phaseolus mungo*." Zeitschrift für Pflanzenphysiologie **74**(2): 130-142.

Atkinson, D. E. (1968). "The energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers." Biochemistry **7**(11): 4030-4034.

Bailey, J. K., J. L. Pinyon, et al. (2010). "Distribution of human commensal Escherichia coli phylogenetic groups." J Clin Microbiol **48**(9): 3455-3456.

Baker, C. S., I. Morozov, et al. (2002). "CsrA regulates glycogen biosynthesis by preventing translation of glgC in Escherichia coli." Mol Microbiol **44**(6): 1599-1610.

Barnard, F. M., M. F. Loughlin, et al. (2004). "Global regulation of virulence and the stress response by CsrA in the highly adapted human gastric pathogen Helicobacter pylori." Mol Microbiol **51**(1): 15-32.

Bennett, B. D., E. H. Kimball, et al. (2009). "Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli." Nat Chem Biol **5**(8): 593-599.
Bhatt, S., A. N. Edwards, et al. (2009). "The RNA binding protein CsrA is a pleiotropic regulator of the locus of enterocyte effacement pathogenicity island of enteropathogenic Escherichia coli." Infect Immun **77**(9): 3552-3568.
Boer, V. M., C. A. Crutchfield, et al. (2010). "Growth-limiting intracellular metabolites in yeast growing under diverse nutrient limitations." Mol Biol Cell **21**(1): 198-211.
Camacho, D., A. de la Fuente, et al. (2005). "The origin of correlations in metabolomics data." Metabolomics **1**(1): 53-63.
Canelas, A. B., C. Ras, et al. (2011). "An in vivo data-driven framework for classification and quantification of enzyme kinetics and determination of apparent thermodynamic data." Metab Eng.
Chang, D. E., D. J. Smalley, et al. (2004). "Carbon nutrition of Escherichia coli in the mouse intestine." Proc Natl Acad Sci U S A **101**(19): 7427-7432.
Datsenko, K. A. and B. L. Wanner (2000). "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products." Proc Natl Acad Sci U S A **97**(12): 6640-6645.
Edwards, A. N., L. M. Patterson-Fortin, et al. (2011). "Circuitry linking the Csr and stringent response global regulatory systems." Mol Microbiol.
Escobar-Paramo, P., K. Grenet, et al. (2004). "Large-scale population structure of human commensal Escherichia coli isolates." Appl Environ Microbiol **70**(9): 5698-5700.
Feist, A. M., C. S. Henry, et al. (2007). "A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information." Mol Syst Biol **3**: 121.
Fields, J. A. and S. A. Thompson (2008). "Campylobacter jejuni CsrA mediates oxidative stress responses, biofilm formation, and host cell invasion." J Bacteriol **190**(9): 3411-3416.
Fischer, E. and U. Sauer (2003). "Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS." Eur J Biochem **270**(5): 880-891.
Fong, S. S., A. Nanchen, et al. (2006). "Latent pathway activation and increased pathway capacity enable Escherichia coli adaptation to loss of key metabolic enzymes." J Biol Chem **281**(12): 8024-8033.
Grozdanov, L., C. Raasch, et al. (2004). "Analysis of the genome structure of the nonpathogenic probiotic Escherichia coli strain Nissle 1917." J Bacteriol **186**(16): 5432-5441.
Henard, C. A., T. J. Bourret, et al. (2010). "Control of redox balance by the stringent response regulatory protein promotes antioxidant defenses of Salmonella." J Biol Chem **285**(47): 36785-36793.
Heroven, A. K., K. Bohme, et al. (2008). "A Csr-type regulatory system, including small non-coding RNAs, regulates the global virulence regulator RovA of Yersinia pseudotuberculosis through RovM." Mol Microbiol **68**(5): 1179-1195.
Hobman, J. L., C. W. Penn, et al. (2007). "Laboratory strains of Escherichia coli: model citizens or deceitful delinquents growing old disgracefully?" Mol Microbiol **64**(4): 881-885.
Homburg, S., E. Oswald, et al. (2007). "Expression analysis of the colibactin gene cluster coding for a novel polyketide in Escherichia coli." FEMS Microbiol Lett **275**(2): 255-262.
Hosomi, S., H. Tara, et al. (1989). "Inhibitory effect of 5-phosphoribosyl 1-pyrophosphate and ADP on the nonoxidative pentose phosphate pathway activity." Biochem Med Metab Biol **42**(1): 52-59.
Hua, Q., C. Yang, et al. (2003). "Responses of the central metabolism in Escherichia coli to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts." J Bacteriol **185**(24): 7053-7067.
Kamada, N., Y. G. Kim, et al. (2012). "Regulated virulence controls the ability of a pathogen to compete with the gut microbiota." Science **336**(6086): 1325-1329.
Kayser, A., J. Weber, et al. (2005). "Metabolic flux analysis of Escherichia coli in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state." Microbiology **151**(Pt 3): 693-706.
Le Bouguenec, C. and C. Schouler (2011). "Sugar metabolism, an additional virulence factor in enterobacteria." Int J Med Microbiol **301**(1): 1-6.
Leatham, M. P., S. J. Stevenson, et al. (2005). "Mouse intestine selects nonmotile flhDC mutants of Escherichia coli MG1655 with increased colonizing ability and better utilization of carbon sources." Infect Immun **73**(12): 8039-8049.

Li, H., A. Granat, et al. (2008). "RpoS, H-NS, and DsrA influence EHEC hemolysin operon (ehxCABD) transcription in Escherichia coli O157:H7 strain EDL933." FEMS Microbiol Lett **285**(2): 257-262.
Liu, M. Y., H. Yang, et al. (1995). "The product of the pleiotropic Escherichia coli gene csrA modulates glycogen biosynthesis via effects on mRNA stability." J Bacteriol **177**(10): 2663-2672.
Lucchetti-Miganeh, C., E. Burrowes, et al. (2008). "The post-transcriptional regulator CsrA plays a central role in the adaptation of bacterial pathogens to different stages of infection in animal hosts." Microbiology **154**(Pt 1): 16-29.
Martinez-Jehanne, V., L. du Merle, et al. (2009). "Role of deoxyribose catabolism in colonization of the murine intestine by pathogenic Escherichia coli strains." Infect Immun **77**(4): 1442-1450.
Mira, A., H. Ochman, et al. (2001). "Deletional bias and the evolution of bacterial genomes." Trends Genet **17**(10): 589-596.
Miranda, R. L., T. Conway, et al. (2004). "Glycolytic and gluconeogenic growth of Escherichia coli O157:H7 (EDL933) and E. coli K-12 (MG1655) in the mouse intestine." Infect Immun **72**(3): 1666-1676.
Murray, E. L. and T. Conway (2005). "Multiple regulators control expression of the Entner-Doudoroff aldolase (Eda) of Escherichia coli." J Bacteriol **187**(3): 991-1000.
Nanchen, A., A. Schicker, et al. (2006). "Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of Escherichia coli." Appl Environ Microbiol **72**(2): 1164-1172.
Neidhardt, F. C., R. Curtiss, et al. (1996). "Escherichia coli and Salmonella." Washington D.C., ASM press.
Nicolas, C., P. Kiefer, et al. (2007). "Response of the central metabolism of Escherichia coli to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase." FEBS Lett **581**(20): 3771-3776.
Nowrouzian, F. L. and E. Oswald (2012). "Escherichia coli strains with the capacity for long-term persistence in the bowel microbiota carry the potentially genotoxic pks island." Microb Pathog.
Nowrouzian, F. L., A. E. Wold, et al. (2005). "Escherichia coli strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants." J Infect Dis **191**(7): 1078-1083.
Palaniyandi, S., A. Mitra, et al. (2012). "BarA-UvrY two-component system regulates virulence of uropathogenic E. coli CFT073." PLoS One **7**(2): e31348.
Peekhaus, N. and T. Conway (1998). "Positive and negative transcriptional regulation of the Escherichia coli gluconate regulon gene gntT by GntR and the cyclic AMP (cAMP)-cAMP receptor protein complex." J Bacteriol **180**(7): 1777-1785.
Peekhaus, N. and T. Conway (1998). "What's for dinner?: Entner-Doudoroff metabolism in Escherichia coli." J Bacteriol **180**(14): 3495-3502.
Pernestig, A. K., D. Georgellis, et al. (2003). "The Escherichia coli BarA-UvrY two-component system is needed for efficient switching between glycolytic and gluconeogenic carbon sources." J Bacteriol **185**(3): 843-853.
Perrenoud, A. and U. Sauer (2005). "Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli." J Bacteriol **187**(9): 3171-3179.
Puskeiler, R., K. Kaufmann, et al. (2005). "Development, parallelization, and automation of a gas-inducing milliliter-scale bioreactor for high-throughput bioprocess design (HTBD)." Biotechnol Bioeng **89**(5): 512-523.
Romeo, T. and M. Gong (1993). "Genetic and physical mapping of the regulatory gene csrA on the Escherichia coli K-12 chromosome." J Bacteriol **175**(17): 5740-5741.
Romeo, T., M. Gong, et al. (1993). "Identification and molecular characterization of csrA, a pleiotropic gene from Escherichia coli that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties." J Bacteriol **175**(15): 4744-4755.
Romeo, T., C. A. Vakulskas, et al. (2012). "Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems." Environ Microbiol.
Sabnis, N. A., H. Yang, et al. (1995). "Pleiotropic regulation of central carbohydrate metabolism in Escherichia coli via the gene csrA." J Biol Chem **270**(49): 29096-29104.

Sauer, U., F. Canonaco, et al. (2004). "The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of Escherichia coli." J Biol Chem **279**(8): 6613-6619.
Schaub, J. and M. Reuss (2008). "In vivo dynamics of glycolysis in Escherichia coli shows need for growth-rate dependent metabolome analysis." Biotechnol Prog **24**(6): 1402-1407.
Snider, T. A., A. J. Fabich, et al. (2009). "E. coli O157:H7 catabolism of intestinal mucin-derived carbohydrates and colonization." Vet Microbiol **136**(1-2): 150-154.
Steuer, R. (2006). "Review: on the analysis and interpretation of correlations in metabolomic data." Brief Bioinform **7**(2): 151-158.
Sun, J., F. Gunzer, et al. (2005). "Genomic peculiarity of coding sequences and metabolic potential of probiotic Escherichia coli strain Nissle 1917 inferred from raw genome data." J Biotechnol **117**(2): 147-161.
Sze, C. W. and C. Li (2011). "Inactivation of bb0184, Which Encodes Carbon Storage Regulator A, Represses the Infectivity of Borrelia burgdorferi." Infect Immun **79**(3): 1270-1279.
Tatarko, M. and T. Romeo (2001). "Disruption of a global regulatory gene to enhance central carbon flux into phenylalanine biosynthesis in Escherichia coli." Curr Microbiol **43**(1): 26-32.
Taymaz-Nikerel, H., A. E. Borujeni, et al. (2010). "Genome-derived minimal metabolic models for Escherichia coli MG1655 with estimated in vivo respiratory ATP stoichiometry." Biotechnol Bioeng **107**(2): 369-381.
Taymaz-Nikerel, H., M. de Mey, et al. (2009). "Development and application of a differential method for reliable metabolome analysis in Escherichia coli." Anal Biochem **386**(1): 9-19.
Timmermans, J. and L. Van Melderen (2009). "Conditional essentiality of the csrA gene in Escherichia coli." J Bacteriol **191**(5): 1722-1724.
Traxler, M. F., V. M. Zacharia, et al. (2011). "Discretely calibrated regulatory loops controlled by ppGpp partition gene induction across the 'feast to famine' gradient in Escherichia coli." Mol Microbiol **79**(4): 830-845.
Tweeddale, H., L. Notley-McRobb, et al. (1998). "Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ("metabolome") analysis." J Bacteriol **180**(19): 5109-5116.
Wei, B., S. Shin, et al. (2000). "Global regulatory mutations in csrA and rpoS cause severe central carbon stress in Escherichia coli in the presence of acetate." J Bacteriol **182**(6): 1632-1640.
Wu, L., M. R. Mashego, et al. (2005). "Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards." Anal Biochem **336**(2): 164-171.

Conclusions

This work aimed at getting a comprehensive understanding of the system-level operation of *Escherichia coli* metabolism on physiologically-relevant carbon sources, and of the role of the post-transcriptional regulator Csr in carbon nutrition and control of central metabolism. We have first developed and validated a complete methodology for the quantitative, high-throughput, and system-level investigations of *Escherichia coli* metabolism.

In Part 1 of this thesis, we have designed and rigorously evaluated new experimental and computational techniques for ^{13}C-MFA. These developments include i) experimental design of ^{13}C-label inputs to rationalize labelling experiments, ii) sampling procedures to collect reliable isotopic information on true metabolic intermediates, iii) a general strategy to evaluate the reliability of MS instruments to provide quantitative isotopic measurements, as well as the complete experimental workflow in MS-based ^{13}C-MFA, iv) raw data processing to extract meaningful isotopic information, and v) a new algorithm (NLSIC) for fast, stable and precise flux calculations. Altogether they represent additional and valuable steps towards high-throughput ^{13}C-MFA and are expected to make ^{13}C-MFA more accessible to a wider biological community.

In Part 2 of this thesis, we have applied this methodology to address fundamental questions regarding the physiology and the metabolism of *E. coli*.

Detailed system-level investigations on the actual operation of the metabolism of *E. coli* were performed on a set of physiologically-relevant carbon sources representative of the major metabolic pathways (ED, PP, EMP, TCA). The investigations were carried out on two distinct *E. coli* strains (K12 MG1655 and Nissle 1917) differing in their implantation capabilities. Strong metabolic differences in terms of both molecular content (metabolome) and metabolism operation (growth parameters and fluxome) were observed. A significantly higher metabolic efficiency of the Nissle 1917 strain was observed, which likely contributes to the superior implantation capabilities of this strain, either directly (faster cell division) or indirectly (outcompete the microbiota for nutrients availability). Multi-level data integration has provided valuable information regarding systemic properties of *E. coli* metabolism. Linear relationships between catabolic, energy, redox and anabolic processes of *E. coli* were observed independently of the considered nutrient. These same relationships were observed for the two strains K12 MG1655 and Nissle 1917, suggesting that they represent generic characteristics of *E. coli*, perhaps as

a result of evolutionary pressure. The analysis of correlations between metabolite levels and fluxes may also allow the identification of metabolic regulators, and a novel method for metabolic flux analysis is proposed, which briefly consists in stoichiometric modelling with metabolomics-derived constraints. Deeper understanding of the operation of *E. coli* metabolism may emerge from these data by using computational modelling approaches.

The Csr system, one of the major post-transcriptional regulators in *E. coli*, is thought to reinforce the competitiveness of the bacterium in the utilisation of carbon sources metabolized through the upper part of the EMP such as glucose. Detailed system-level investigations of a set of mutants altered for the various components of the Csr system and their parent strains were performed. Our results reinforce and extend the role of Csr in carbon nutrition of *E. coli* by showing that it enhances the use of physiologically-relevant carbon sources metabolized by all of the major metabolic pathways of this bacterium. We also show that Csr either directly or indirectly controls, in a nutrient-dependent manner, a wider range of metabolic pathways than expected from its known target enzymes. CsrA does not act only as a master switch between glycolytic and gluconeogenic processes in *E. coli* but appears to exert a more global control on all of its major metabolic pathways. This is of particular physiological relevance due to the critical role of metabolism (and of the ED pathway in particular) in gut colonization.

The role of Csr is qualitatively and quantitatively the same in K12 MG1655 and Nissle 1917 *E. coli* strains, and we show that the reorganizations observed in *csrA51* mutants are not due to the reduced growth-rate. The relationships observed between catabolic, energy and anabolic processes in wild type *E. coli* strains remain linear when CsrA activity is reduced, but their correlation parameters strikingly differ from those observed in wild-type strains. In addition, correlations between anabolism and redox metabolism are no longer significant in *csrA51* mutants. This demonstrates for the first time that Csr controls energy and redox metabolism in *E. coli*. Its action enhances the production of ATP and of reduced cofactors (NADH and NADPH), suggesting a role of Csr in the oxidative stress response of *E. coli*.

Altogether, our results may suggest that Csr finely tunes the catabolic and anabolic processes of *E. coli* in response to changes in its nutritional environment. Accordingly, the BarA/UvrY is needed for efficient switching between glycolytic and gluconeogenic carbon sources, and *csrA51* mutants show a longer lag phase for all carbon sources

tested (chapter 7). The implantation of *E. coli* results from global adaptive strategies that allow the bacteria to survive in the changing environment of the intestine. A close examination of these adaptive processes (e.g. during changes in carbon nutrition) would permit to get comprehensive understanding of the conditions that favour the implantation of gut bacteria and, also, of the physiological basis that determines the expression of the specific traits of each strain, including virulence factors in pathogenic strains.

Annexes

Annexes are provided in electronic format and can be downloaded at:
http://metasys.insa-toulouse.fr/manuscrit_pm/annex/

My personal contribution to the different chapters presented in this manuscript is detailed in the following table.

Analysis of the role of Csr on the operation of *Escherichia coli* metabolism and methodological developments								
Study	**Chapter**	**Conception and design of experiments**	**Construction of mutants**	**Performed the experiment**	**Sample preparation and analysis**	**Data and statistical analysis**	**Interpretation**
Sampling of ^{13}C-labelled metabolites for ^{13}C-MFA	3	T	*n.a.*	T	T	T	T
Evaluation of isotopic analyses by mass spectrometry	4	C (SM)	*n.a.*	C (SM)	T	T	T
Role of Csr in *Escherichia coli* Nissle 1917 on glucose and gluconate	7	C (OR)	- (OR)	C (OR)	C (OR)	C (OR)	C (OR)
Role of Csr in *Escherichia coli* K-12 MG1655 and Nissle 1917 on galactose, gluconate, ribose and acetate	8	T	- (OR, TE)	C (JP)	T	T	T

Software development							
Software	**Chapter**	**Problem formulation**	**Code writting**	**Documentation writting**	**Validation**	**Bug tracking**	
IsoDesign	2	C (SS)	T	T	T	T	
IsoCor	5	C (SS)	T	T	T	C (SS)	
influx_s	6	- (SS)	- (SS)	- (SS)	C (SS)	C (SS)	

"-": no contribution (name of contributors in brackets), "C": performed in collaboration (name of other contributors in brackets), "T": performed completely, "n.a.": not applicable
JP: Juliette Poinot, OR: Olga Revelles, SM: Stéphane Massou, SS: Serguei Sokol, TE: Thomas Esquerre

The totality of these investigations were performed under the supervision of Jean-Charles Portais (LISBP, Toulouse), Fabien Letisse (LISBP), and Jean-Philippe Nougayrède (ENVT, Toulouse).