A Study of the Short and Long-term Regulation of *E. coli* Metabolic Pathways

Anália Lourenço[1]*, Sónia Carneiro[1]*, José P. Pinto[2], Miguel Rocha[2], Eugénio C. Ferreira[1] and Isabel Rocha[1]

[1]Institute for Biotechnology and Bioengineering, Centre of Biological, Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal, {analia,soniacarneiro,ecferreira,irocha}@deb.uminho.pt

[2]Department of Informatics/CCTC, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal, {josepedr,mrocha}@di.uminho.pt

Summary

The present study addresses the regulatory network of *Escherichia coli* and offers a global view of the short- and long-term regulation of its metabolic pathways. The regulatory mechanisms responsible for key metabolic activities and the structure behind such mechanisms are detailed. Most metabolic functions are dependent on the activity of transcriptional regulators over gene expression - the so-called long-term regulation. However, enzymatic regulation - the so-called short-term regulation - often overlays transcriptional regulation and even, in particular metabolic pathways, enzymatic regulation may prevail. As such, understanding the balance between these two types of regulation is necessary to be able to predict and control cell responses, specifically cell responses to the various environmental stresses.

1 Introduction

The analysis of biological networks aims at the understanding of metabolic capabilities of cells to adapt to and to maintain growth under different external and internal conditions [1, 2]. A particularly challenging task is the inference of the regulatory interactions commanding the activity of metabolic pathways. We know that different mechanisms are recruited for regulation, either on a long-term basis by changing the expression level of genes or a short-term basis by changing the activity of enzymes. Long-term regulation is crucial for cell survival and serves to adapt to environmental challenges that require extensive cellular alterations at the transcriptional level. In contrast, short-term regulation is used to control specific enzymatic activities and hence, balance metabolite pools to avoid metabolic and energetic losses.

Gene expression is mostly controlled by transcription factors (TFs) that are proteins able to bind to gene promoter regions, inducing or repressing the initiation of gene transcription. In turn, the activity of enzymes, proteins that catalyse biochemical reactions, can be controlled by some effector molecules binding at the allosteric site, by substrate inhibition or by alteration of some environmental condition (e.g. pH or ionic strength). In recent years, the interest in understanding the coordination of these regulatory mechanisms (i.e. co-regulation) to modulate the

---

* These authors equally contributed to this work.

enzymatic activity of specific biosynthetic steps in pathways has increased significantly. The manipulation of the activity of key enzymes, either by transcriptional regulation or allosteric control of the activity of the enzyme itself, can benefit the adaptation of cells to specific conditions or enhance the production of valuable metabolites.

In this work, we focus on the study of how transcriptional regulation couples with the regulation of the activity of metabolic pathways via enzymatic regulation, and how similar regulatory mechanisms are used across different metabolic pathways. Our goal is to obtain a global view of the regulatory interplay affecting the metabolism of the bacterium *Escherichia coli (E. coli)* K12, considering its common use as an industrial organism [3].

2   Methodology

This study required the construction of a genome-scale network involving known regulation on *E. coli* metabolic pathways. This network is intended to encompass various information:

- gene coding or genetic information, i.e. the genes that code for regulatory and non-regulatory proteins;

- gene regulation, i.e. the control of gene transcription by TFs encoded by regulatory genes (RGs), usually called long-term regulation;

- gene-reaction association, i.e. non-regulatory genes (nRGs) coding for enzymes catalysing metabolic reactions;

- enzymatic regulation, i.e. control of the activity of biochemical reactions by metabolites (Ms) that bind directly to enzymes, usually called short-term regulation.

Since *E. coli* is a model organism, there are a number of data resources that can be used for this purpose. We took a previously validated metabolic network and complemented it with regulatory information from publicly available repositories. For simplicity purposes, we reverted all information into nodes representing genes, i.e. TFs, enzymes and reactions nodes are replaced by nodes representing the corresponding coding genes

Next, we provide further details on the data integration and network analysis methodologies.

2.1   Data Integration

The genome-scale metabolic network of *E. coli* K12 (iAF1260) [4] was at the core of the data integration. This is a manually curated and experimentally validated model, which means that all its information is interconnected, i.e. there are no "loose ends" or isolated nodes and *in silico* simulations have shown that the set of reactions and enzymes described is sufficient to characterise the metabolism of *E. coli*. The model encompasses reaction stoichiometry (the identification and quantification of substrates and products), enzyme-reaction associations (grouping reactions into pathways) and enzyme-coding genes (identified by *bnumbers*, which are Locus identifiers specific for *E. coli*).

We then integrated the enzyme-coding genes with gene transcriptional regulation derived from the organism-specific EcoCyc database [5]. Specifically, we included TF-encoding or regulatory genes (RGs) that regulate the expression of metabolic or non-regulatory genes (nRGs), i.e. genes associated with a reaction from the iAF1260 model, and regulate other TF-encoding genes linked to metabolic genes. As shown in Figure 1, RGs can regulate metabolic genes (or nRGs) directly or indirectly through other RGs creating a cascade of regulatory interactions between genes. Information on the control of enzyme activity by metabolic regulators was also extracted from EcoCyc. However, we only considered metabolic regulators that intervene as substrates (Ms) or products (Mp) in our model.
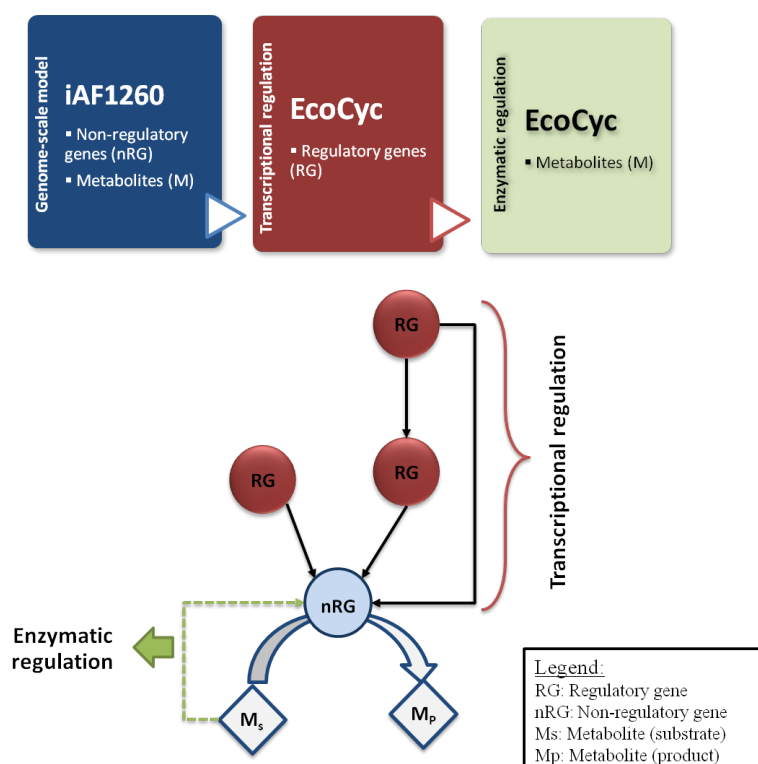

Figure 1: An integrated regulatory network for *E. coli*: combining the manually curated iAF1260 model with regulatory data from EcoCyc repository. Gene-reaction associations consisting on non-regulatory genes (nRGs), which code for enzymes that catalyse the conversion of metabolic substrates (Ms) into products (Mp), were obtained from the iAF1260 model. EcoCyc information on transcriptional and enzymatic regulation was further associated to non-regulatory genes (nRGs). All regulatory genes (RGs) that directly or indirectly affect the transcription of enzyme-coding genes were included in the integrated network. In the case of enzymatic regulation, all metabolites that affect the activity of an enzyme included in the metabolic model were associated with the respective enzyme-coding genes or non-regulatory genes (nRGs). In addition, metabolites that are not a substrate or a product in any reaction of the model were filtered.

Since the iAF1260 model was found to be incomplete in terms of enzyme-coding gene information, we implemented a semi-automatic integration process where manual curation was issued to complete enzyme-coding information and retrieve the corresponding regulatory information whenever needed.

2.2 Network Analysis

The graph that represents our network has two types of nodes, genes and metabolites, and two types of edges: one that connects metabolites (M) to the enzyme-coding genes (or nRGs) they regulate; and another that connects pairs of genes that are linked by transcriptional regulation.

The representation provided by the Java Universal Network/Graph Framework (JUNG) (`http://jung.sourceforge.net/`) supported the analysis of the properties of the graph. We analysed the frequency of occurrence of different types of edges associated with nRGs, specifically: RGs directly connected to nRGs, cascades of RGs that end up connected to nRGs, and metabolites (M) that are part of the cell's metabolism (i.e. are being processed by the cell or are somehow going into the cell) and act as regulators of enzymatic activities. Additionally, we looked into regulatory patterns (or motifs) commonly described in literature [6, 7] to get a deeper understanding on the interplays taking place on particular pathways (see details in Box 1).


Box 1
Network motifs

Single input module (SIM): two-gene pattern where a single regulatory gene (RG) is responsible for the regulation of more than one gene, either regulatory (RGs) or non-regulatory genes (nRGs).

Reversed Single input module (RSIM): a two-gene pattern where multiple regulatory genes (RGs) govern the transcriptional regulation of one gene, either a regulatory (RG) or a non-regulatory gene (nRG).

Feed-Forward Loop (FFL): three-gene pattern composed by two input regulatory genes (RGs), one of which regulates the other (X→Y) and jointly regulate a target gene (Z), which can be a regulatory (RG) or non-regulatory gene (nRG).

Bowtie: three-gene pattern that includes a central node that is highly regulated and, at the same time, regulates multiple gene nodes, either regulatory (RGs) or non-regulatory genes (nRGs).
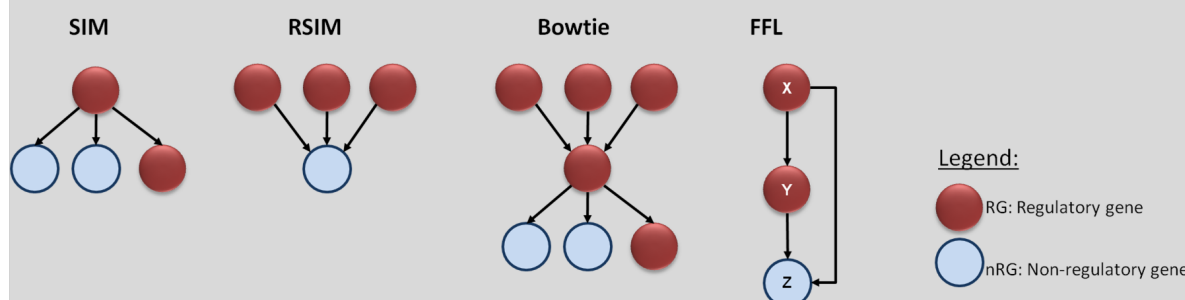
To be able to assess the importance of different regulatory motifs in particular metabolic pathways, we calculated the prevalence of motif types per pathway and the average number of genes in the pathway participating in such motifs as follows:

- the <u>absolute frequency</u> of each motif type T for every pathway P, i.e. we counted the number of motifs that <u>affect at least one gene G_P</u> in pathway P

$$\textbf{Abs\_Freq\_T\_P} = number\ of\ motifs\ of\ type\ \textbf{\textit{T}}\ in\ pathway\ \textbf{\textit{P}} \qquad (1)$$

- the <u>relative frequency</u> of motif type T, i.e. the number of times the motif type T occurs in the pathway (i.e. affects at least one of its genes) divided by the number of times it occurs in the network

$$\textbf{Rel\_Freq\_T\_P} = \frac{Abs\_Freq\_T\_P}{number\ of\ motifs\ of\ type\ \ \boldsymbol{T}\ in\ the\ network} \qquad (2)$$

- and the <u>relative frequency</u> of genes from pathway P (G_P) that are included in motifs of type T

$$\textbf{Rel\_Freq\_G}_P\textbf{\_T} = \frac{number\ of\ genes\ in\ motif\ type\ \boldsymbol{T}\ in\ pathway\boldsymbol{P}}{number\ of\ genes\ in\ \boldsymbol{P}} \qquad (3)$$

3 Results and discussion

3.1 Types of regulation

The distribution of different types of regulation over the metabolism of *E. coli* is illustrated in Figure 2 showing that 53% of genes are regulated by at least one TF, but co-regulation (i.e. the combination of transcriptional and enzymatic regulation) and enzymatic regulation are also present (16% and 29% of genes, respectively). This suggests that metabolic activities do not necessarily correlate proportionally with the gene expression levels of the corresponding enzymes, but are also dependent on immediate enzyme control by metabolites.
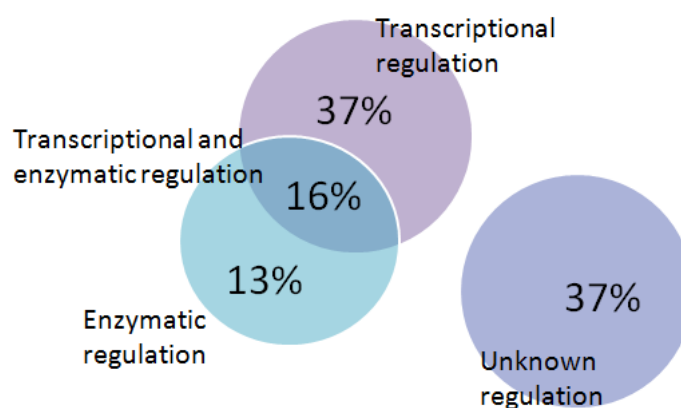

Figure 2: The different types of regulation in the *E. coli* network.

The high percentage of unknown regulation (37%) can be explained by the insufficiency of knowledge about enzyme activities (it has not yet been possible to assign a function to approximately one third of the proteins identified in *E.coli* and many details are still missing from its biochemical characterisation [8]) and the exclusion of some regulatory mechanisms of this study (e.g. post-translational modifications and ribosome-mediated transcriptional attenuations).

Some additional hypotheses about the mechanisms coordinating key metabolic processes can be formulated by detailing the regulation per pathway (Table 1). The **transcriptional regulation** is dominant in pathways like **"Nitrogen Metabolism", "Citric Acid Cycle", "Inorganic Ion Transport and Metabolism"** and other transport pathways (87%, 83%, 76% and 72-81%, respectively). These pathways require modifications at gene expression level to respond to environmental inputs, such as availability of nitrogen and carbon sources. In contrast, enzymatic activities that are only dependent on **enzymatic regulation** are not so representative within *E. coli* pathways. Most pathways have less than 40% of their gene products regulated at this level and 10 out of the 37 pathways considered in the model have no genes associated with this regulatory mechanism. The only exception is the pathway **"Folate Metabolism"**, which has approximately half of its genes regulated at the enzymatic level. This indicates that few pathways include metabolic functions governed exclusively by enzymatic regulation and only specific metabolic functions, such as energy-generating reactions, seem to be coupled with this type of regulation.

The existence of **co-regulation** in amino acids biosynthesis pathways, like **"Glutamate Metabolism", "Tyrosine, Tryptophan", and "Phenylalanine Metabolism"** and **"Arginine and Proline Metabolism"**, can be explained by the need to ensure the fine-tuning of metabolic activities and the rapid response to over-accumulation of end products. Co-regulated pathways present advantages in terms of functional changeability and reorganization. For instance, 64% of the genes coding for enzymes involved in the **"Glutamate Metabolism"** are simultaneously controlled at the transcriptional and enzymatic levels, while only 9% are exclusively regulated at the enzymatic level and none is exclusively dependent on the transcriptional control. The combination of regulatory mechanisms allows a tight control of glutamate biosynthesis (the principal amino donor to other amino acids), manipulating protein expression levels and restraining enzymatic activity whenever needed.

3.2   Regulatory motifs

Many metabolic pathways are dependent on the activity of transcriptional regulators, often organized into regulatory structures or motifs, such as SIMs, RSIMs, BowTies and FFLs, acting as specific functional modules (Table 2). Therefore, the association of certain structures with particular metabolic activities can be hypothesized as a consequence of specific information processing.

SIM and **RSIM motifs** are simple regulatory structures that coordinate the activity of multiple genes. While SIMs represent patterns where a set of genes are controlled by a single TF (i.e. a one-to-many relationship), RSIMs display structures where a single gene is being controlled by multiple regulators (i.e. a many-to-one relationship). The occurrence of these two-gene patterns, particularly SIMs, among the transcriptional regulatory network of *E. coli* has been previously evaluated [6]. These regulatory structures are commonly associated with the regulation of genes that belong to the same operon, and usually form a protein assembly (such as flagella), or are part of a metabolic pathway (such as the amino acid biosynthesis). Most genes in *E. coli* are regulated by several TFs, both negatively or positively.

This stringent control of gene expression ensures that cells are able to respond to several different stresses through the activation of rescue activities, such as anaplerotic reactions or, more commonly, by spreading this information from global regulators to more specialized regula-

Table 1: Effect of transcriptional, enzymatic and co-regulation in *E. coli* pathways. For each gene associated with a particular pathway it was determined if there are transcriptional regulations (i.e. a link between genes), enzymatic regulations (i.e. a link between a metabolite and a non-regulatory gene (nRG)) or both types of regulation (co-regulation). Unknown regulation denotes the absence of regulatory associations.

Pathway	Number of Genes	Transcriptional regulation	Metabolic regulation	Transcriptional and metabolic regulation	Unknown regulation
Alanine and Aspartate Metabolism	11	27%	9%	27%	36%
Alternate Carbon Metabolism	159	48%	6%	21%	26%
Anaplerotic Reactions	10	20%	40%	20%	20%
Arginine and Proline Metabolism	40	23%	18%	40%	20%
Cell Envelope Biosynthesis	53	17%	17%	11%	55%
Citric Acid Cycle	18	83%	0%	11%	6%
Cofactor and Prosthetic Group Biosynthesis	136	20%	16%	7%	57%
Cysteine Metabolism	20	35%	15%	20%	30%
Folate Metabolism	4	25%	50%	25%	0%
Glutamate Metabolism	11	0%	9%	64%	27%
Glycerophospholipid Metabolism	20	15%	35%	0%	50%
Glycine and Serine Metabolism	15	27%	7%	40%	27%
Glycolysis/Gluconeogenesis	32	31%	13%	28%	28%
Glyoxylate Metabolism	4	25%	0%	25%	50%
Histidine Metabolism	9	0%	22%	11%	67%
Information Transfer	167	56%	0%	0%	44%
Inorganic Ion Transport and Metabolism	67	76%	1%	0%	22%
Lipopolysaccharide Biosynthesis / Recycling	46	7%	9%	2%	83%
Membrane Lipid Metabolism	14	50%	14%	7%	29%
Methionine Metabolism	13	31%	0%	23%	46%
Methylglyoxal Metabolism	10	0%	20%	10%	70%
Murein Biosynthesis	10	20%	0%	0%	80%
Murein Recycling	34	29%	6%	0%	65%
Nitrogen Metabolism	15	87%	0%	7%	7%
Nucleotide Salvage Pathway	57	19%	26%	19%	35%
Oxidative Phosphorylation	77	65%	8%	5%	22%
Pentose Phosphate Pathway	13	15%	31%	23%	31%
Purine and Pyrimidine Biosynthesis	22	41%	18%	27%	14%
Pyruvate Metabolism	20	50%	15%	15%	20%
Threonine and Lysine Metabolism	19	11%	37%	21%	32%
Transport, Inner Membrane	216	72%	0%	1%	26%
Transport, Outer Membrane	21	81%	0%	0%	19%
Transport, Outer Membrane Porin	4	75%	0%	0%	25%
tRNA Charging	24	0%	17%	4%	79%
Tyrosine, Tryptophan, and Phenylalanine Metabolism	24	29%	4%	29%	38%
Unassigned	23	39%	4%	13%	43%
Valine, Leucine, and Isoleucine Metabolism	17	76%	0%	18%	6%

tory modules. Pathways like **"Alternate Carbon Metabolism"** and **"Transport, Inner Membrane"** have several genes whose expression is dependent on these forms of regulation to respond to external *stimuli* (in particular, to alterations in the nutrient carbon sources). As presented in Table 2, the **"Alternate Carbon Metabolism"** and **"Transport, Inner Membrane"** are coordinated by multiple regulatory structures encompassing about 46% and 66% SIMs and 17% and 22% RSIMs, respectively. RSIMs are also widely represented in **"Information Transfer"** with a percentage of almost 11%. Clearly, there is an overlap of regulatory

Table 2: Statistical evaluation of the occurrence of regulatory motifs per pathway. Abbreviations: # is the number of motifs of type T found in pathway P; T_P is the relative frequency of a motif type T in pathway P; and G_P_T is the frequency of genes from the pathway P (G_P) involved in motif type T (see section *Network Analysis* for details). The values of T_P indicate the prevalence of motif type T in metabolic pathway P, while the values of G_P_T indicate the number of genes in pathway P that are affected by the regulation of motif type T.

Pathway	FFL			SIM			RSIM			BowTie		
	#	T_P	Gp_T	#	T_P	Gp_T	#	T_P	Gp_T	#	T_P	Gp_T
Alanine and Aspartate Metabolism	3	0%	0.27	6	5%	1.00	4	1%	0.36	2	10%	0.18
Alternate Carbon Metabolism	144	13%	0.91	60	46%	1.74	82	17%	0.52	9	45%	0.21
Anaplerotic Reactions	6	1%	0.60	7	5%	1.70	3	1%	0.30		0%	
Arginine and Proline Metabolism	10	1%	0.25	15	12%	1.33	12	2%	0.30	3	15%	0.15
Cell Envelope Biosynthesis		0%		9	7%	0.38	4	1%	0.08	1	5%	0.02
Citric Acid Cycle	61	6%	3.39	17	13%	4.11	17	3%	0.94	6	30%	0.89
Cofactor and Prosthetic Group Biosynthesis	45	4%	0.33	27	21%	0.63	23	5%	0.17	5	25%	0.10
Cysteine Metabolism	6	1%	0.30	6	5%	0.85	4	1%	0.20	1	5%	0.05
Folate Metabolism	2	0%	0.50	5	4%	1.25	1	0%	0.25		0%	
Glutamate Metabolism	41	4%	3.73	13	10%	3.18	6	1%	0.55	5	25%	1.18
Glycerophospholipid Metabolism	7	1%	0.35	8	6%	0.45	2	0%	0.10	1	5%	0.05
Glycine and Serine Metabolism	7	1%	0.47	12	9%	1.40	6	1%	0.40	2	10%	0.13
Glycolysis/Gluconeogenesis	15	1%	0.47	10	8%	1.06	8	2%	0.25	3	15%	0.09
Glyoxylate Metabolism		0%		1	1%	0.50		0%			0%	
Histidine Metabolism		0%		1	1%	0.11		0%			0%	
Information Transfer	96	9%	0.57		0%		57	11%	0.34	16	80%	0.36
Inorganic Ion Transport and Metabolism	38	4%	0.57	28	22%	1.54	26	5%	0.39	5	25%	0.39
Lipopolysaccharide Biosynthesis / Recycling		0%		4	3%	0.13	2	0%	0.04	1	5%	0.04
Membrane Lipid Metabolism	2	0%	0.14	7	5%	1.21	7	1%	0.50	1	5%	0.14
Methionine Metabolism	3	0%	0.23	6	5%	1.00	6	1%	0.46	1	5%	0.08
Methylglyoxal Metabolism	1	0%	0.10	2	2%	0.20	1	0%	0.10		0%	
Murein Biosynthesis		0%		1	1%	0.10		0%			0%	
Murein Recycling	2	0%	0.06		0%		6	1%	0.18		0%	
Nitrogen Metabolism	49	5%	3.27	8	6%	0.59	11	2%	0.73	4	20%	1.53
Nucleotide Salvage Pathway	16	1%	0.28	15	12%	5.53	12	2%	0.21	2	10%	0.14
Oxidative Phosphorylation	218	20%	2.83	18	14%	0.82	52	10%	0.68	8	40%	0.65
Pentose Phosphate Pathway	3	0%	0.23	24	18%	3.35	3	1%	0.23	2	10%	0.15
Purine and Pyrimidine Biosynthesis	4	0%	0.18	8	6%	0.77	8	2%	0.36	3	15%	0.23
Pyruvate Metabolism	42	4%	2.10	8	6%	1.27	13	3%	0.65	3	15%	0.50
Threonine and Lysine Metabolism	5	0%	0.26	17	13%	2.50	4	1%	0.21	3	15%	0.16
Transport, Inner Membrane	229	21%	1.06	11	8%	0.63	111	22%	0.51	18	90%	0.39
Transport, Outer Membrane	12	1%	0.57	86	66%	1.93	7	1%	0.33	2	10%	0.33
Transport, Outer Membrane Porin	8	1%	2.00	18	14%	1.76	2	0%	0.50	1	5%	0.25
tRNA Charging		0%		9	7%	3.50		0%			0%	
Tyrosine, Tryptophan, and Phenylalanine Metabolism	2	0%	0.08	1	1%	0.04	4	1%	0.17		0%	
Unassigned	17	2%	0.74	7	5%	0.79	7	1%	0.30	4	20%	0.35
Valine, Leucine, and Isoleucine Metabolism	15	1%	0.88	16	12%	1.26	7	1%	0.41	2	10%	0.18

structures, which are essentially composed by RGs, repeated across pathways. Nevertheless, this information is still relevant to have an overview of the complexity of the regulatory mechanisms behind certain metabolic functions.

Bowtie motif can be interpreted as the coupling of SIM and RSIM motifs through a single central element, suggesting that these structures share a similar conceptual and architectural design. TFs that regulate more than one operon (i.e. genes involved in more than one metabolic function that are scattered over different operons) and are regulated by more than one TF are frequently part of these patterns as central nodes. Their main role is to spread information from different *stimuli* to the adequate functional components, behaving like decision points where a TF elects the most appropriate response to a certain *stimulus*. For example, the transcriptional activator GadE controls the transcription of genes involved in the maintenance of pH homeostasis, glutamate dependent (GAD) system and multidrug efflux, among others, in response to the transcriptional control of two-component systems, like EvgS/EvgA and PhoQ/PhoP, and other regulators involved in cellular response to acid resistance or in the catabolism of secondary carbon sources [11]. The capacity of central nodes to admit variability of input information (i.e. regulation from other genetic elements), confers high flexibility and robustness to the system, while supporting the modulation of multiple pathways simultaneously [9]. Pathways like **"Transport, Inner Membrane"** and **"Information Transfer"**, which support inherently complex information exchange processes, are in need of these regulatory structures to guarantee the adequate propagation of the information throughout the network (90% and 80%, respectively). However, the number of genes that are regulated by this type of structure is higher in pathways like **"Glutamate Metabolism"** and **"Nitrogen Metabolism"** (Gp_T values equal 1.18 and 1.53, respectively), probably as means to guarantee the activation of these metabolic functions in a more effective way.

FFL motif represents a simple form of regulation where the activities of two RGs regulate the expression of the target gene both directly and indirectly. These motifs are likely to occur when a rapid response to an external signal is required, such as shifts in carbon sources or availability of oxygen [10]. The abundance of these patterns suggests that many metabolic functions might respond favourably to persistent rather than transient signals. This means that activation of the target gene (node Z) via direct control is more effective if the signals from node X are transient, which would delay the conveyance of information via the indirect control (i.e. through node Y). For this reason, pathways like **"Oxidative Phosphorylation", "Transport, Inner Membrane"** and **"Alternate Carbon Metabolism"** present higher abundance of these motifs (20%, 21% and 13%, respectively). Also, most of the genes associated with these pathways are controlled by this regulatory pattern (Gp_T values of approximately 1 or larger), which allows a rapid functional switch in response to stimulus, e.g. deficiencies in the availability of exogenous electron acceptors like nitrate, nitrite, fumarate [12]. Although few FFL motifs were found in pathways like **"Glutamate Metabolism"** (4%), **"Citric Acid Cycle"** (6%) or **"Nitrogen Metabolism"** (5%), the number of genes that were repeatedly included in this type of structure was considerable (Gp_T values greater than 3). This indicates that the density of GRs overlapping is higher in these pathways and the flow of regulatory information crossing these GRs is decisive. This feature can be also observed in other regulatory structures for the same pathways, for example the regulation of **"Nitrogen Metabolism"** comprises 15 SIMs (around 12% of the SIMs found in the network), but the pathway only has 15 genes, which means that some genes must be included more than once in these structures explaining the Gp_T value larger than 5. These properties provide evidences that the entanglement of interactions and the density of regulatory structures coupled to metabolic networks define largely the complexity behind the coordination of cellular metabolism.

Next, we detail the regulatory mechanisms governing some of the metabolic activities in *E.*

coli.

3.2.1 Folate metabolism and anaplerotic reactions: tightly short-term regulated pathways

The interaction between enzymes and regulatory metabolites can decrease or increase the catalytic activity. As the binding of these metabolic effectors create an immediate effect over the enzymatic activity, it is considered that this type of control is a short-term regulatory mechanism. This provides for a powerful way to change the cell's phenotype, only by rearranging specific metabolic functions.

Some examples of short-term regulation can be found in the **"Folate Metabolism"** and **"Anaplerotic reactions"**. The **"Folate Metabolism"** is central to many cellular processes in *E. coli*, ranging from nucleotide and amino acid biosynthesis to the production of the starting amino acid residue in protein synthesis, i.e. N-formylmethionyl-tRNA(f). The enzyme-coding genes *pur*U and *fol*D are core of this metabolism and are only regulated enzymatically by the amino acids glycine and methionine, and coenzyme formyltetrahydrofolate (see Figure 3). This ensures the balance of the pools of tetrahydrofolate and one-carbon tetrahydrofolate metabolites to maintain the synthesis of glycine and methionine.
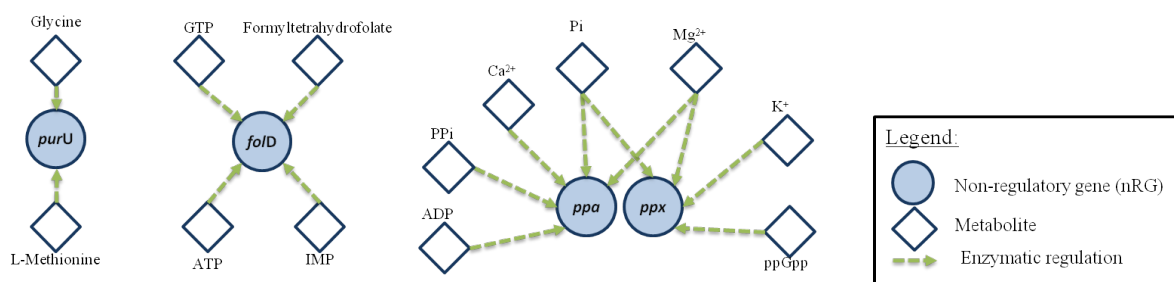
Figure 3: Examples of the short-term regulation affecting the "Folate Metabolism' (*pur*U and *fol*D genes) and "Anaplerotic reactions' (*ppa* and *ppx* genes). In the "Folate Metabolism', the enzymatic activity of formyltetrahydrofolate deformylase encoded by *pur*U is essentially controlled by two amino acids, glycine and L-methionine, while the *fol*D gene product is regulated by purine-related compounds (GTP, IMP and ATP) and formyltetrahydrofolate. Anaplerotic enzymes coded by *ppa* and *ppx* are phosphohydrolases playing an important role in energy metabolism and are mainly controlled by cations (Ca^{2+}, Mg^{2+} and K$^+$) and other phosphate-related compounds.

Likewise, the *ppa* and *ppx* genes that catalyse anaplerotic reactions responsible for the hydrolysis of inorganic pyrophosphate maintaining the pool of orthophosphates in the cytoplasm, are basically regulated by enzymatic regulators. The regulation of the PPase activity is dependent on the inhibitory effects of cations (Ca^{2+}) and other inorganic phosphorus-containing ions, while exopolyphosphatase (PPX) is further regulated by ppGpp, an alarmone that triggers the stringent response in bacteria.

Seemingly, metabolic functions that are mainly dependent on enzymatic regulation are preferred when cells need to deal with transient alteration in metabolite pools that are essential for other biosynthetic activities or to replenish intermediates in central metabolic pathways. Nevertheless, the expression of some gene products involved in these metabolic activities might still be controlled by TFs. The metabolic gene *met*F, which is associated to the reduction of folate coenzymes, is regulated by MetJ, a transcriptional repressor of genes involved in biosynthesis and transport of methionine. This can be explained by the fact that various reactions in this

pathway produce folates that are important for biosynthetic pathways leading to methionine, purine, and pyrimidine biosynthesis, and as such, some transcriptional regulation is needed to avoid the over-accumulation of these metabolites.

3.2.2 Citric Acid Cycle and Nitrogen Metabolism: two pathways depending on long-term regulation

Pathways such as **"Citric Acid Cycle"** and **"Nitrogen Metabolism"** have little incidence of enzymatic regulation. Both are related to respiratory catalytic pathways, generating energy and reducing power. The **"Citric Acid Cycle"** is a catabolic pathway of aerobic respiration, while **"Nitrogen Metabolism"** supports anaerobic respiration of various carbon sources.

In the **"Citric Acid Cycle"** the isocitrate dehydrogenase is the only enzyme with enzymatic regulation and is subjected to allosteric control by oxaloacetate, i.e. the end product of the **"Citric Acid Cycle"**, and glyoxylate, i.e. an anabolic intermediary that is present under specific physiological conditions (see Figure 4). This modulation enables cells to make a more rapid shift between **"Citric Acid Cycle"** and **"Glyoxylate Metabolism"** pathways and thus, adjust cellular growth to different carbon sources. However, the rest of the reactions included in this pathway are strongly dependent on transcriptional regulation, essentially by the action of global regulators like ArcA, FNR or CRP.
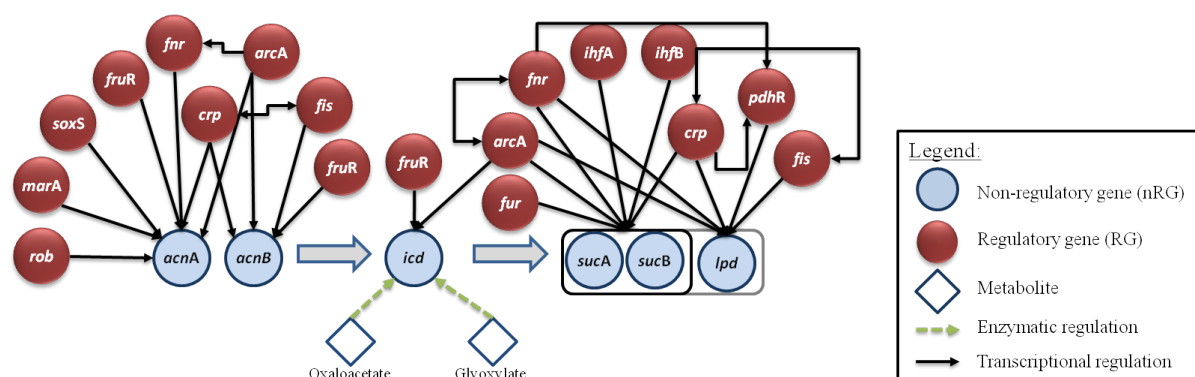

Figure 4: Examples of the long-term regulation affecting the "Citric Acid Cycle". From the three sequential reactions in this pathway, only isocitrate dehydrogenase encoded by the *icd* gene is both regulated at the transcriptional and enzymatic levels. In contrast, the two aconitases encoded by *acn*A and *acn*B genes, and the 2-oxoglutarate dehydrogenase complex (*suc*A-*suc*B-*lpd*) are only regulated at the transcriptional level.

Similarly, the regulation of the **"Nitrogen Metabolism"** relies on the action of the global regulators Fis, FNR and IHF, and the TFs NarP, NarL, FlhDC and Nsr (see Figure 5). NarL/NarP regulate the anaerobic electron transport and fermentation-related genes in response to the availability of high concentrations of nitrate or nitrite, and NsrR acts as a nitrite-sensitive repressor that regulates genes involved in cell protection against nitric oxide. FlhDC is the principal regulator of bacterial flagellum biogenesis and swarming migration, also involved in cellular response to the depletion of nitrogen sources.

In both pathways, we also observed that transcriptional regulation is structured in FFL and SIM circuits. The organization of most genes in the same operon allows the regulation of several genes by the same TF (SIM structure), while FFLs facilitate the transition between transient and persistent physiological states.
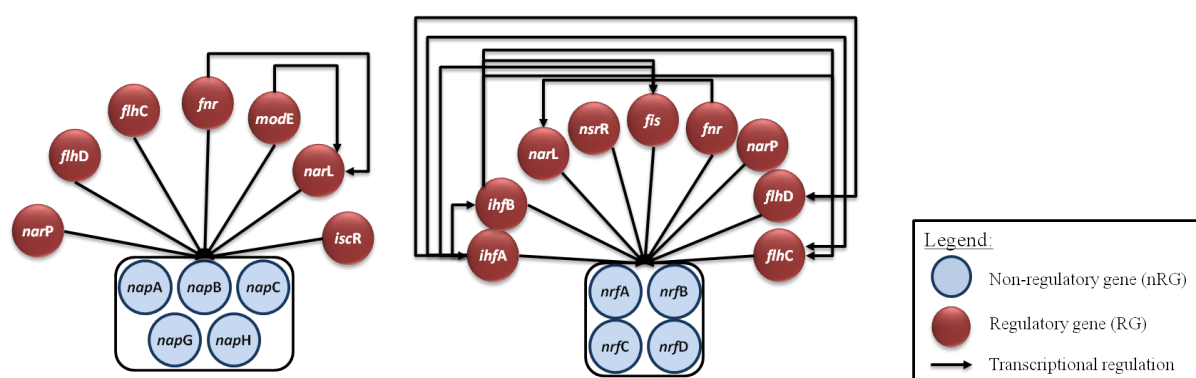
Figure 5: Examples of the long-term regulation affecting the "Nitrogen Metabolism". The enzymatic complexes nitrate and nitrite reductases encoded by the *nap*ABCGH and *nrf* ABCD operons, respectively, are exclusively controlled by the activity of transcription factors or regulatory genes (RGs).

The **"Nitrogen Metabolism"** genes from the *nap*ABCDGH operon are regulated by a SIM structure based on gene NarL, a nitrate/nitrite response regulator that represses genes involved in alternative respiratory pathways. NarL is also enclosed in a FNR-triggered FFL pattern. In the absence of FNR, NarL represses the expression of the *nap*ABCDGH operon and thus, the activity of the periplasmic nitrate reductase is not induced. In anaerobic conditions, FNR is induced mediating the transition from aerobic to anaerobic growth, and thus stimulating the activity of this alternative respiratory pathway.

In the **"Citric Acid Cycle"** pathway, the *acn*A gene presents a similar regulatory pattern where ArcA, a negative transcriptional regulator under anaerobic conditions, either represses the expression of the *acn*A gene or represses the FNR transcription factor that, in turn, represses the *acn*A gene.

3.2.3 Glutamate Metabolism: a pathway depending on co-regulation

Regulation of pathways like **"Glutamate Metabolism"** or other amino acids biosynthetic pathways, depends heavily on substrate availability and enzymatic regulation by inhibitors/activators, which are central in the short-term regulation of energy homeostasis or the synthesis and degradation of metabolic resources. However, transcriptional regulation of genes for key enzymes in such metabolic pathways is also important for long-term adaptation to specific nutritional states. The occurrence of both types of regulation within a pathway plays a major role in the maintenance of the metabolic balance under different physiological conditions.

The **"Glutamate metabolism"** was highlighted for discussion due to the significant co-occurrence of transcriptional and enzymatic regulation. This is a very important pathway, because glutamate is a major constituent of the proteins in *E. coli* and it is a major nitrogen donor for other biosynthesis activities. One possible example of the complexity of this pathway is the activity of the two glutamate decarboxylases that participate in the acid resistance system controlling responses to low pH. The coding genes *gad*A and *gad*B are regulated by multiple TFs and FFLs are the common regulatory structure (see Figure 6). The enzymes are also affected by various metabolic regulators. Intermediates of the **"Citric Acid Cycle"**, such as fumarate and succinate, act upon these enzymes to inhibit their activity when pH is back normal. These combined actions enable the cell to respond quickly to pH perturbations.
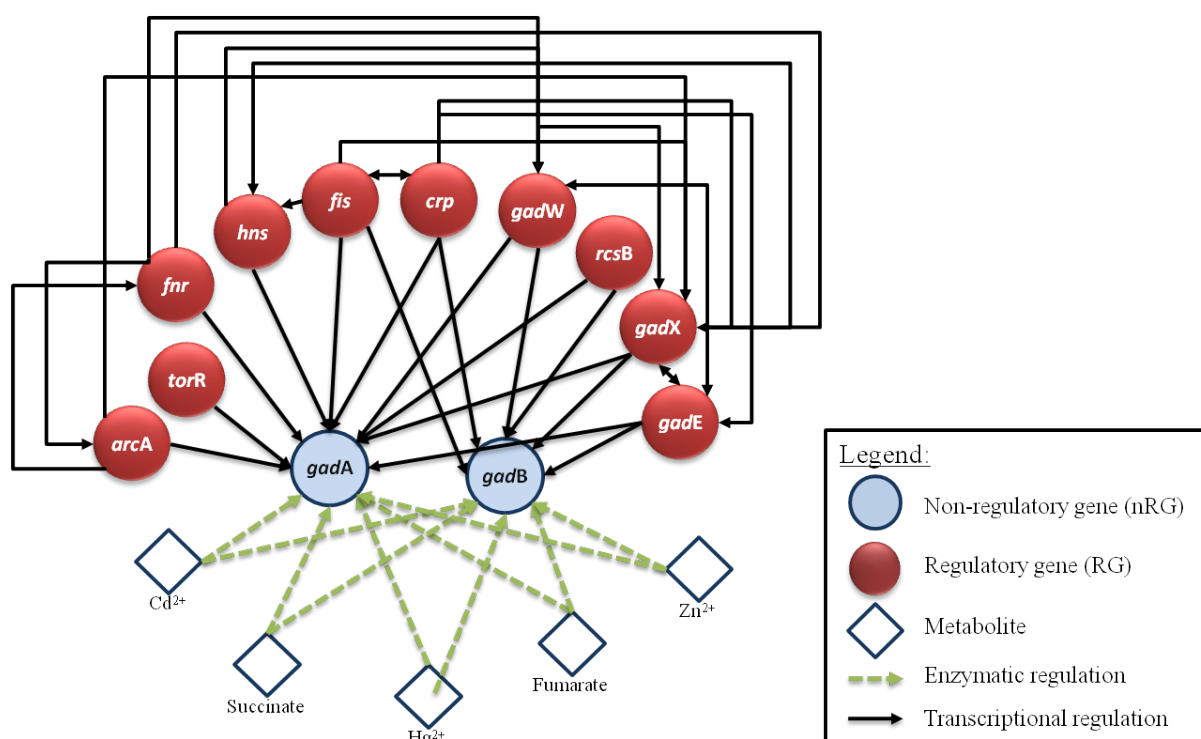
Figure 6: Examples of co-regulation affecting the "Glutamate Metabolism". Although with similar metabolic functions, the transcriptional control of the two glutamate decarboxylases coded by *gad*A and *gad*B genes is different, e.g. the TorR transcriptional regulator affects the expression of *gad*A, but not *gad*B. The enzymatic regulation of the two isoenzymes is identical, involving the inhibition of their activity via the binding of divalent cations (Zn^{2+}, Hg^{2+} and Cd^{2+}) and dicarboxylic acids (succinate and fumarate)

4 Conclusions

Considering that TFs and metabolic regulators have similar functional purpose, i.e. to ultimately regulate the activity of enzymatic reactions, the integrated analysis of their activities provides a new perspective over the capacity of modulation of *E. coli* metabolic pathways. TF-based regulation is meant to perform system adaptation whereas enzymatic regulation is chosen when a rapid shift in a given metabolic activity is needed. For example, the shift from aerobic respiration to anaerobic respiration requires gene expression adjustments as the cells have to adapt their entire metabolism to a new environment. In turn, enzymatic regulators are needed to balance certain metabolic pools and thus, maintaining the concentration of end products within acceptable ranges. Moreover, most of the observations gathered for one pathway and its regulation via a given TF or metabolic regulators show that the cooperation of both regulatory mechanisms is fundamental to coordinate the activity of the many enzymes working in a pathway.

The full networking between TFs by which cells can adapt their metabolic states is known to be supported by various structures of regulation capable of responding to one or more environmental/internal inputs. Each structure has a unique way to process information (it may receive multiple inputs and/or it may affect multiple gene targets) and its relevance over a pathway is given by the number of affected genes that belong to the pathway. This study discriminates pathways that are heavily regulated from those where TF regulation is scarce. Additionally, the study has shown that certain regulatory structures are characteristic of a subset of pathways.

In particular, the ability to accept a wide range of inputs and to convey information through a single node affecting several functional elements is seen in pathways where the transference of information is critical.

Acknowledgements

References

[1] Thiele I, Jamshidi N, Fleming RM, Palsson BO. Genome-scale reconstruction of *Escherichia coli's* transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. *PLoS Computational Biology*, 5(3):e1000312, 2009.

[2] Abdul Kadir TA, Mannan AA, Kierzek AM, McFadden J, Shimizu K. Modeling and simulation of the main metabolism in *Escherichia coli* and its several single-gene knockout mutants with experimental verification. *Microbiology Cell Factories*, 9:88, 2010.

[3] Feist AM, Palsson BO. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nature Biotechnology*, 26:659–667, 2008.

[4] Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3:121, 2007.

[5] Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muniz-Rascado L, Bonavides-Martinez C, Paley S, Krummenacker M, Altman T, Kaipa P, Spaulding A, Pacheco J, Latendresse M, Fulcher C, Sarker M, Shearer AG, Mackie A, Paulsen I, Gunsalus RP, Karp PD. EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Research*, 39(suppl 1):D583–D590, 2011.

[6] Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics*, 31:64–68, 2002.

[7] Alon U. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8:450–461, 2007.

[8] Hu P, Janga SC, Babu M, Diaz-Mejia JJ, Butland G, Yang W, Pogoutse O, Guo X, Phanse S, Wong P, Chandran S, Christopoulos C, Nazarians-Armavil A, Nasseri NK, Musso G, Ali M, Nazemof N, Eroukova V, Golshani A, Paccanaro A, Greenblatt JF, Moreno-Hagelsieb G, Emili A. Global functional atlas of *Escherichia coli* encompassing previously uncharacterized proteins. *PLoS Biology*, 7:e96, 2009.

[9] Ma HW, Zhao XM, Yuan YJ, Zeng AP. Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. *Bioinformatics*, 20:1870–1876, 2004.

[10] Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics*, 31:64–68, 2002.

[11] Hommais F, Krin E, Coppée JY, Lacroix C, Yeramian E, Danchin A, Bertin P. GadE (YhiE): a novel activator involved in the response to acid environment in *Escherichia coli*. *Microbiology*, 150(1):61–72, 2004.

[12] Stenmark P, Gurmu D, Nordlund P. Crystal structure of CaiB, a type-III CoA transferase in carnitine metabolism. *Biochemistry*,43(44):13996–14003, 2004.

[13] Gutiérrez-Ríos RM, Rosenblueth DA, Loza JA, Huerta AM, Glasner JD, Blattner FR, Collado-Vides J. Regulatory network of *Escherichia coli*: consistency between literature knowledge and microarray profiles. *Genome Research*,13(11):2435–2443, 2003.