



HAL
open science

Modeling Leucine's Metabolic Pathway and Knockout Prediction Improving the Production of Surfactin, a Biosurfactant from *Bacillus Subtilis*

François Coutte, Joachim Niehren, Debarun Dhali, Mathias John, Cristian Versari, Philippe Jacques

► **To cite this version:**

François Coutte, Joachim Niehren, Debarun Dhali, Mathias John, Cristian Versari, et al.. Modeling Leucine's Metabolic Pathway and Knockout Prediction Improving the Production of Surfactin, a Biosurfactant from *Bacillus Subtilis*. *Biotechnology Journal*, 2015, 10 (8), pp.1216-34. 10.1002/biot.201400541 . hal-01153704

HAL Id: hal-01153704

<https://inria.hal.science/hal-01153704v1>

Submitted on 9 Jul 2015

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

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

Modeling Leucine’s Metabolic Pathway and Knockout Prediction Improving the Production of Surfactin, a Biosurfactant from *Bacillus Subtilis*

François Coutte^{*,1,4}, Joachim Niehren^{*,2,3}, Debarun Dhali^{1,4}, Mathias John^{2,4}, Cristian Versari^{2,4}, Philippe Jacques^{1,4}

¹ ProBioGEM team, Research Institute for Food and Biotechnology - Charles Viollette (EA1026)

² BioComputing team, Laboratory CRISTAL (CNRS UMR9189)

³ INRIA Lille, France

⁴ Université de Lille, France

Abstract. A *Bacillus subtilis* mutant strain overexpressing surfactin biosynthetic genes was previously constructed. In order to further increase the production of this biosurfactant, our hypothesis is that the surfactin precursors, especially leucine, must be overproduced. We present a three step approach for leucine overproduction directed by methods from computational biology. Firstly, we develop a new algorithm for gene knockout prediction based on abstract interpretation, which applies to a recent modeling language for reaction networks with partial kinetic information. Secondly, we model the leucine metabolic pathway as a reaction network in this language, and apply the knockout prediction algorithm with the target of leucine overproduction. Out of the 21 reactions corresponding to potential gene knockouts, the prediction algorithm selects 12 reactions. Six knockouts were introduced in *B. subtilis* 168 derivatives strains to verify their effects on surfactin production. For all generated mutants, the specific surfactin production is increased from 1.6 to 20.9 fold during the exponential growth phase, depending on the medium composition. These results show the effectiveness of the knockout prediction approach based on formal models for metabolic reaction networks with partial kinetic information, and confirms our hypothesis that precursors supply is one of the main parameter to optimize surfactin overproduction.

Keywords : Abstract interpretation, *Bacillus subtilis*, Modeling language, Knockout prediction, Surfactin.

1 Introduction

In this paper we develop and implement genetic engineering methods with the objective to overproduce surfactin by *B. subtilis*. Surfactin is one of the most powerful biosurfactant known and it displays several biological activities of interest (antiviral, antimycoplasmic, elicitor, etc) [19]. This promising molecule

* These two authors contributed equally to this work.

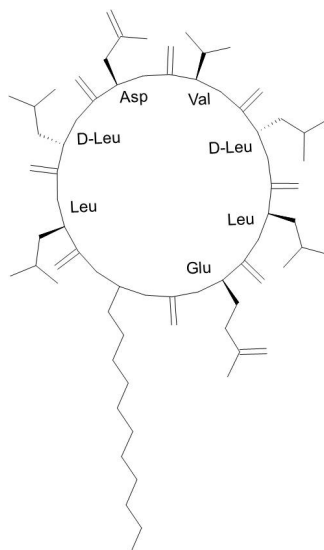


Figure 1: Schematic representation of surfactin [20]

is assembled by a non ribosomal mechanism involving multifunctional proteins called Non Ribosomal Peptide Synthesis (NRPS). The substrates of these enzymes are amino acids or fatty acids residues present in the cytoplasm of the cell. The surfactin is composed of a ring of seven amino acid residues connected to a β -hydroxylated fatty acid chain of different length and isomery [19, 20]. The peptide moiety contains four leucines (Figure 1). Genetic engineering of *B. subtilis* have already been made in order to increase the lipopeptide production. In previous work [8], the overproduction of surfactin was obtained by replacing the native promoter of the surfactin operon (*surfA*) by a constitutive one and disrupted the plipastatin operon (*ppsA*) to save the precursor availability. The same approach was recently developed for the mycosubtilin production [2]. Sun et al. [40] replaced the *surfA* promoter by *Pspac*. Liu et al. [26] explore feeding strategies of amino acid to modify the peptide moiety of the lipopeptide. However, there is no work directly related to the orientation of the metabolism to increase the supply precursor without modifying the extracellular medium. The hypothesis of the present work is that any further increase of surfactin production can only be achieved by increasing the production of these precursors, and in particular leucine.

The branched chain amino acids, i.e. leucine, valine, and isoleucine are produced by a metabolic pathway with complex regulatory mechanisms. The question is how to adapt this pathway of *B. subtilis* by gene knockouts, in order to overproduce leucine. Our hope is that one can predict such gene knockouts from models of the pathway by methods from computational biology, in order

to narrow down the testing duties in wet lab. Model-based knockout prediction for metabolic pathways, however, requires to overcome a number of difficulties.

Flux balance analysis is the most prominent approach for model-based prediction [1, 9, 13]. The idea is to infer flux balance equations from the models [34], and to analyze the space of their solutions. Flux balance equations are linear equations with variables for fluxes, i.e., real numbers that design the speed of a reaction in a steady state. The problem is that the number of variables exceeds the number of equations in the system by far, so that the spaces of solutions are huge. Therefore, particular solutions must be filtered, typically by imposing optimization criteria in various manners. One can try to optimize the biomass [24, 35, 6], that is the production of metabolites that are needed for an optimal growth of the cells. Or else, one may apply metabolic flux analysis [45] in order to measure experimentally the sizes of a sufficient number of metabolic flows. However, the specific growth conditions used in the experiments strictly limit the applicability of the results. The advantage of flux balance analysis is that it can be applied to models of reaction networks without any kinetic information. But this is also problematic, since flux balance equations cannot express any kinetic information. Therefore, they cannot distinguish the various kinds of modifiers such as enzymes and inhibitors, that is positive from negative regulatory effects, which are available in more recent models of metabolic networks [15]. On the other hand, these models lack any clear formal semantics which makes them useless for formal reasoning.

Constraint-based approaches such as [35, 6] derive boolean constraints from reaction networks, which mainly state the dependencies of where a reaction is on or off from the presence of its reactants or modifiers. Such boolean constraints can be seen as an abstraction of the steady state equations used by flux balance analysis. But the underconstrainedness problem remains as with flux balance analysis, so that one needs to impose further restrictions on the solution space. This can be done for instance by adding negative constraints stating that the reaction is off if one of its inhibitors is present. Note that such negative constraints are often too strong, since an inhibitor often slows down a reaction in a steady state, but does not shut it down totally at least not in average.

Alternatively, pathway analysis techniques may help to discover genetically independent pathways [33, 17]. Among these, elementary mode analysis is also based on flux balance equations. The idea is change the basis of the flux balance equations, so that they constraint the speeds of pathways, i.e., of paths from an inflow to an outflow in a reaction graph, rather than the speeds of the reactions themselves. When adding boolean constraints in addition, one can express information about inhibitors of reactions too [22]. This method is known to be problematic for networks with many pathways, that is for metabolic networks with many paths from inflows to outflows, such as in our case.

A more systematic analysis of reaction networks with partial kinetic information was proposed in [21]. Steady state equations are inferred from the reaction networks, which besides the usual flux balance equations contain equations for the reaction speeds. These equations contain constrained variables for the par-

tially known kinetic functions. Abstract interpretation is then applied in order to abstract away the variables on the kinetic knowledge. It produces difference constraints, that can state that the speed of a reaction will increase if the concentration of one of its inhibitors decreases, and conversely. This is contrast to the previous treatments of inhibitors by boolean constraints, where one can only reason about reactions being on or off. Difference constraints are particular finite domain constraints. This makes it possible to use finite domain constraint solving to predict gene knockouts that will speed up a target reaction. The approach of [21], however, is too restrictive since it is limited to networks of reactions with mass action kinetics, for which the rate constants are unknown. In order to overcome this restriction, Niehren et al. [31] proposed a modeling language for reaction networks with partial kinetic information, which supports unknown kinetic functions up to a similarity constraint that captures activating or inhibiting effects. Reaction networks modeled in this language can still be analyzed by abstract interpretation while generalizing on [21]. And the finite domain constraints inferred from the models can be used to predict flux changes based on finite domain constraint solving. However, no knockout prediction algorithm along these lines have been proposed yet.

As a first contribution in this article, we develop a knockout prediction algorithm for reaction networks in this modeling language. Basically, the algorithm is obtained by the combination of the ideas of the two previous papers [21, 31]. For the sake of self-containedness, we elaborate these ideas from scratch, and illustrate the algorithm obtained at the example reaction network modeling the regulation of the *ilv-leu* operon promoter (**PIIv-Leu**). As we argue, its correctness can be reduced to the previous results. Our algorithm is based on finite-domain constraint solving to enumerate the solutions of a difference constraint obtained by abstract interpretation. The maximal number of changes in a solution must be fixed by the user. From the set of solutions obtained, the algorithm then filters those that correspond to single knockouts in a straightforward manner. The result is a list of a reactions that are predicted for single knockout. All this can be done fully automatically, when given the reaction network, the knockout candidates, the overproduction target, and the maximal number of changes in a solution.

The second contribution is an implementation of all these algorithms. This is obtained by extending the finite domain constraint solver for difference constraints from [21], which is written in SCALA, as well as the tools for the modeling language from [31], which are written in XSLT. More precisely, these tools allow the validation of reaction networks in XML, convert them into a nice graphical output (see e.g. Figure 3), compute the steady state equations, and infer the difference constraints by abstract interpretation. Finally, the knockout predictions are obtained by filtering the solutions of the difference constraints corresponding to single knockouts by yet another XSLT script.

Our third contribution is a model of the quite complex leucine metabolic pathway [38, 43, 44, 5, 4] in the modeling language from [31]. This is obtained by rewriting the model from SubtiWiki [15, 28] into this formal modeling lan-

guage, while adding few missing reactions. We then applied the new prediction algorithm to this model. It predicts 14 reaction knockouts from the 21 given candidates, but only for 12 of these reactions, the effect of a knockout is plausible with respect to the model. It might be possible to rule out the two unplausible predictions based on the model too, but for now we do not know how to do this automatically.

Our fourth contribution is a successful experimental verification *in-vivo* of the 6 predicted knockouts with the most direct effects by genetic engineering of *B. subtilis*. It turns out that for all generated mutants the specific surfactin production is increased by a factor ranging from 1.6 to 20.9 during the exponential growth phase, depending on the gene knockout and on the medium composition. These results show the effectiveness of the knockout prediction and confirms our hypothesis that precursors supply is one of the main parameter to optimize surfactin overproduction.

2 Materials and Methods

2.1 *In-vivo* experiments and genetic engineering of *B. subtilis*

2.1.1 Metabolic effect of leucine feeding In order to test the working hypothesis that an increase in the intracellular pool of leucine led to an increase of the surfactin production, the strain *B. subtilis* BBG111 [8] was cultivated in presence of leucine. Two different cultures were carried-out in Erlenmeyer flask at 37°C with 160 rpm of agitation. The first medium used was Landy [25] with 3.6 g/L (NH₄)₂SO₄ (instead of glutamate) and the second one Landy with 2.3 g/L (NH₄)₂SO₄ and 2.5 g/L of leucine (instead of glutamate). Medium were supplemented with tryptophan (16 mg/L) and buffered at pH 7.0 with MOPS 0.1M. Sample were withdrawn after 6 hours of culture. Biomass and optical density measurement were carried out as described previously [8]. Surfactin production was measured by RP-HPLC in the centrifugated and filtered samples, according to Coutte et al. [8] and using pure surfactin as the standard (Lipofabrik, Villeeneuve d’Ascq, France). Cultures were performed in triplicate and results are expressed as mean value with standard deviation. The results given in Section 3.5 confirmed the working hypothesis.

2.1.2 Genetic engineering of *B. subtilis* 168 derivative strains

We introduced gene knockout in the leucine metabolic pathway using the pop-in pop-out technique [41]. Construction of all deletion mutant strains were carried out as described in Table 1. A modified *B. subtilis* BSB1 strain [30] was constructed using the genomic DNA of TF8A strain by introducing a neomycin-resistance gene under the control of the Lambda Pr promoter ($\lambda Pr-neo$). All deletions were introduced in the master strain by homologous replacement of the targeted chromosome region by a DNA fragment called “cassette *upp-phleo-cI*”, carrying the phleomycin-resistance gene for positive selection of cassette integration and *Psak- λcI* genes for counterselection. After deletion of a particular

Table 1: Strains used or constructed in this work.

Strains of <i>B. subtilis</i>	Genotype before	Genotype modification	knock-out	Reference/source
BBG111	<i>trpC2, amyE::sfp-cat</i>			derived from 168 [8]
BSB1 ¹				[30]
SRB94	<i>ilvBp4</i>			derived from SMY ¹ [5]
SRB87	<i>ilvBpΔT2</i>			derived from SMY ¹ [5]
BBG251	<i>IlvBp4</i>	<i>amyE::sfp-cat</i>	r ₁	newly derived from SRB94
BBG252	<i>IlvBpΔT2</i>	<i>amyE::sfp-cat</i>	r ₄₀	newly derived from SRB87
BBG253	<i>trpC2, amyE::sfp-cat</i>	Δbcd	r ₁₀	newly derived from BBG111
BBG254	<i>trpC2, amyE::sfp-cat</i>	$\Delta codY$	r ₁₅	newly derived from BBG111
BBG255	<i>trpC2, amyE::sfp-cat</i>	ΔbkL ²	r ₁₁	newly derived from BBG111
BBG256	<i>trpC2, amyE::sfp-cat</i>	$\Delta tnrA$	r ₁₆	newly derived from BBG111

¹ The strains SMY and BSB1 are propototroph and 168 lineage.

² Whenever we talk about a knockout for the *bkL* operon, we mean the *lpdV* gene of this operon.

fragment the chromosomal arrangement was confirmed by replica plate method as well as through PCR analysis and gel electrophoresis on 0.8% agarose gel.

Then, two sets of primer pairs were designed for upstream and downstream region of the gene to be deleted (using the sequence from SubtiList World-Wide Web Server and supplied by Eurofins Genomics, Ebersberg, Germany). These two sets of primers were named (forward₁,reverse₁) for the upstream gene and (forward₂,reverse₂) for the downstream gene respect to the gene of interest. During the design of the primers sets, direct repeat sequence are either added to the primer sequence reverse₁ or forward₂ as well as complement sequence of 5' and 3' region of the cassette *upp-phleo-cl* are also added to the primer sequence of reverse₁ and forward₂ (see Table 10 in supplementary materials). Polymerase chain reaction was carried out using specific primer pair (forward₁,reverse₁) and (forward₂,reverse₂) (0.5 μM final each) and 200 ng of master strain chromosomal DNA under standard condition to produce DNA fragments of least 1.0 kb long. The primers used for amplification are listed in Table 10 in supplementary materials. The PCR was carried out under following condition: 5 min at 94°C; (30s at 94°C, 30s at 55°C, 2 min at 72°C) for 25 cycles; 5 min at 72°C. The PCR products were purified using gel extraction kit (Qiagen, Hilden, Germany). The DNA fragments generated using (forward₁,reverse₁) and (forward₂,reverse₂) was mixed equally with higher proportion of *upp-cl-phleo* cassette and joining PCR reaction was conducted under the following conditions: 5 min at 94°C; (15 s at 94°C, 15 s at 55°C, 12 min at 65°C) for 12 cycles; (15 s at 94°C, 15 s at 55°C, 12 min at 65°C) for 25 cycles; 10 min at 72°C in the presence of forward₁ and reverse₂ (0.1 μM final). The resultant product was confirmed through electrophoresis on 0.8% agarose gel and further used to transform competent cells of the master strain.

Positive selection of deletion mutants was carried out through phleomycin resistance on LB medium plates, which were incubated at 37°C up to 24 hours. Colonies, with proper cell morphology were streaked twice on the same medium along with on neomycin-LB medium plates in order to purify the colony. Each clone was checked for the absence of the deleted chromosomal region through PCR using primer pair forward₁ and reverse₂ respectively. Strains constituting the expected deleted chromosomal structure were preserved in glycerol stock at -80°C.

All deletion obtained in modified *B. subtilis* **BSB1** strain were introduced in *B. subtilis* **BBG111** [8] through transformation of genomic DNA of deletion strain into *B. subtilis* **BBG111** through natural competence transformation [37]. The deletion was confirmed through replica plate method using phleomycin, neomycin and phleomycin/neomycin antibiotic plates (Phleomycin was supplied by Euromedex, Souffelweyersheim, France, and Neomycin by Sigma-Aldrich, St Louis, USA). The constructions of the mutant strains are provided Table 1. This procedure was followed for the knockout of the genes *bcd*, *codY*, *bkL*² and *tnrA*, giving rise to mutants strains named *B. subtilis* **BBG253**, **BBG254**, **BBG255** and **BBG256**.

Two remaining strains were obtained from the strains constructed in the Sonenshein Lab [5]. The first modified one was **SRB94**, in which site directed mutagenesis have been carried to generate A10C and C7G mutation in the region II of **CodY** binding site at *ilv-leu* promoter. The second modified one was **SRB87**, in which the author implemented overlapping PCR technique to develop a mutant strain devoid of stem-loop terminator structure responsible for leucine-dependent transcriptional regulation. These strains were converted into surfactin producer by the introduction of *sfp* gene of *B. subtilis* ATCC 21332 in the *amyE* locus by homologous recombination using pBG129 plasmid [8]. The strain were specified as *B. subtilis* **BBG251** and **BBG252**, respectively. All the strains used in this paper are listed in the Table 1.

2.1.3 Culture conditions and analysis

Each mutant obtained was cultivated in high-throughput system of fermentation Biolector (Mp2-labs GmbH; Baesweiler; Germany) available on RealCat platform of the University of Lille. Culture were carried-out in 48 wells flower plate designed for the Biolector (containing pH, dissolved oxygen and biomass probes). Cultures were performed at 37°C, under shaking (1100 rpm). Landy medium with 3.6 g/L (NH₄)₂SO₄ and TSS medium supplemented by 16 amino-acids were used in these experiments [5]. These media were supplemented with tryptophan at 16 mg/L and buffered at pH 7.0 with MOPS 0.1 M. Samples were withdrawn after 6 hours of culture. Biomass measurement for Biolector calibration and surfactin production were carried out as described previously (see Section 2.1.1). Experiments were carried out in triplicate and two biological replicate were performed. Results are expressed as mean value with standard

² Whenever we talk about a knockout for the *bkL* operon, we mean more precisely the *lpdV* gene of this operon.

deviation and statistical analysis of these data is made using one-way ANOVA using a Student-Newman-Keuls post hoc test (p value = 0.05). Different letters are used to indicate significant differences between groups.

2.2 A Modeling Language for Reaction Networks

We first recall the modeling language for reaction networks with partial kinetic information from [31], and illustrate it at the regulation network of the promoter **PIIv–Leu**, which is central for leucine production.

2.2.1 Graphical Syntax

Most aspects of modeling language for reaction networks from [31] are quite standard in systems biology. The exceptions are the treatments of partially known kinetic functions and of exchange flows by which the network interacts with its context.

A *reaction network with partial kinetic knowledge* consists of a finite set of species, a subset of inflow species, a subset of outflow species, and a finite set of reactions. A reaction is a multiset of pairs between species and roles. In the present paper, we fix the following set of *roles*: *reactant*, *product*, *activator*, *accelerator*, or *inhibitor*. A species is called a *modifier* of a reaction if its role is equal to *activator*, *accelerator*, or *inhibitor*. Modifiers are neither consumed nor produced, but the presence of a modifier affects the rate of the reaction. Note that the same species can be used several times in the same reaction and with different roles.

The partial knowledge on the kinetics of a reaction is given by the roles that are assigned to its species. The intuition is as follows. The rate of a reaction is increased by reactants, activators, and accelerators, and it is decreased by inhibitors. A reaction is enabled only if all its reactants and activators are present, i.e. have nonzero concentrations, while some or the accelerators may be absent. We next explain the steady state semantics of a reaction network with partial kinetic information informally, before giving a formal definition in Section 2.2.3.

As usual in deterministic semantics, we consider chemical solutions as functions from species to nonnegative real numbers including zero. For most kinds of species, this number represents the concentration of the species, but in other cases it may also denote the activity of a promoter or the activity of a promoter binding site. Reactions can be applied to chemical solutions. In this case, its reactants are consumed and its products are produced, respectively as many times as they occur in the reaction. A reaction can only be applied if all its activators are present (have a nonzero concentration), while its accelerators may be absent. The higher the concentration of an accelerator or activators, the higher will be the rate of the reaction. A reaction is applicable even if some inhibitors are present, but the higher the concentration the slower will be the rate of the reaction.

Reaction networks are usually situated in a context, which may be adjacent reaction networks, or the “chemical medium” from which some species may inflow

and to which some species may outflow during wet lab experimentation. The exchange with the context is modeled by the set of inflow and outflow species of the network: inflow species may inflow from the context, while outflow species may outflow into the context. The precise rate laws of inflows and outflows are unknown, except that we assume that the outflow rate must increase with the concentration of the outflow species.

Reaction networks in our modeling language can be represented as graphs similar to Petri nets¹. These graphs contain two kinds of nodes, round nodes for representing its species and boxed nodes for representing its reactions. More precisely, any species S is represented by a round node \textcircled{S} and any reaction with name r by a boxed node \boxed{r} . Solid edges either link a reactant to its reaction $\textcircled{S} \rightarrow \boxed{r}$, or a reaction to one of its products $\boxed{r} \rightarrow \textcircled{S}$. There are three kind of dashed edges, which start at the three kinds of modifiers. An accelerator edge links an accelerator to a reaction $\textcircled{S} -\circ \boxed{r}$, an activator edge links an activator to a reaction $\textcircled{S} -\bullet \boxed{r}$, and an inhibitor edge links an inhibitor to a reaction $\textcircled{S} -\vdash \boxed{r}$. Whenever a species plays the same role several times in the same reaction, the multiplicity will be annotated to the corresponding edge. But this will not be the case in any of our examples.

An inflow edge $\textcircled{S} \leftarrow \dots$ points from the context to an inflow species S , while an outflow edge $\textcircled{S} \rightarrow \dots$ points from an outflow species S to the context. For convenience, we have introduced a last kind of edges $\boxed{r} \rightsquigarrow \textcircled{S}$ as a shortcut for a product that is degraded by a hidden reaction, i.e., as a shortcut for: $\boxed{r} \rightarrow \textcircled{S} \rightarrow \boxed{r'}$.

An example network is given in Figure 2. There we use species nodes with three different colors, which indicate their biological roles. The dynamics of the network is not concerned by the colors though. In yellow, we draw metabolites such $\textcircled{\text{Leu}}$ and in blue proteins such $\textcircled{\text{CodY}}$. There is a third color for “artificial species” that serve for modeling regulation, such as for instance the promoter of the *ilv-leu* operon $\textcircled{\text{P}ilv\text{-Leu}}$. The real number that a chemical solution will assign to this artificial species will correspond to the activity of promoter $\text{P}ilv\text{-Leu}$ in the solution (and not to its concentration).

In our graphs, we will annotate in orange, reactions that are potential candidates for knockouts. Dark orange indicates candidates that have been selected by our knockout prediction, while light orange indicates candidates that have not. We note again that node colors are irrelevant for the dynamics of the network.

2.2.2 Modeling the Regulation of $\text{P}ilv\text{-Leu}$ Promoter

The *ilv-leu* operon contains the main genes involved in the production of branched chain amino acids and specially the leucine as shown in simplified

¹ The concrete syntax of our reaction networks is based on XML, from which the graphs are computed. The XML representation is also the input for the prediction algorithm. See the complementary material for examples.

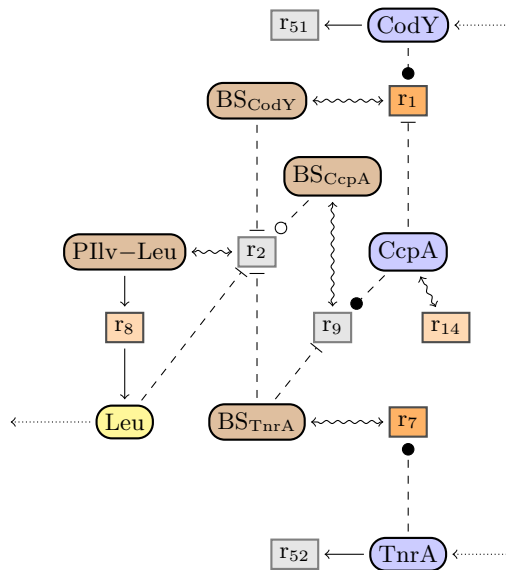


Figure 2: Reaction network of PIIv-Leu regulation.

manner in Figure 2. The *ilv-leu* operon is subject to multiple forms of regulation at its promoter **PIIv-Leu**, as studied in a series of articles. Here we recall the model from [31] which considers the inhibition and activation events of the promoter, without explaining in detail how they are raised by binding to different promoter sites. The same abstraction was adopted earlier in one of the models for leucine production from the SubtiWiki [15], but these models lack a formal semantics as required for reasoning approaches.

We use the species **Leu** modeling the concentration of the metabolite leucine, that we want to overproduce and let outflow to the context. We also introduce the species **PIIv-Leu** modeling the activity of the *ilv-leu* promoter. Reaction **r₈** states that the expression of **PIIv-Leu** will finally lead to the expression of **Leu** in any case. This is an over-simplification adopted by this model, whose purpose is precisely to ignore the complex metabolic network behind the leucine production and its regulation. However, the promoter **PIIv-Leu** is also down-regulated by **Leu** in terms of ribosome-mediated attenuation mechanism T-box [5, 16], which is captured by that **Leu** is an inhibitor of **r₂**, which activates **PIIv-Leu**.

The regulation of **PIIv-Leu** is modeled by reaction **r₂** that activates it. The degree of activation can also be degraded by the implicit reaction **r_{2'}**. Following [43, 5, 14], there are two proteins that down-regulate the activity of **PIIv-Leu** by binding to different sites upstream region of the promoter and inhibiting the RNA polymerase [3]. These are **CodY** (transcriptional pleiotropic regulator, which regulates the transcription of early-stationary-phase genes [36]) and **TnrA** (nitrogen pleiotropic transcriptional regulator), that we model by the actors **BS_{CodY}** and **BS_{TnrA}** respectively. The values of the corresponding variables

$z_{\text{BS}_{\text{CodY}}}$ and $z_{\text{BS}_{\text{TnrA}}}$ represent the activity degrees of these binding sites. It can be increased by applying reactions r_1 and r_7 respectively, and decreased by applying the implicit degradation reactions $r_{1'}$ and $r_{7'}$. The degradation reactions r_{51} and r_{52} are necessary to balance the constant influxes of **CodY** or **TnrA** in steady states.

There is a further protein **CcpA** (Carbon catabolite control protein A), which is expressed by reaction r_{14} and can be degraded by the implicit reaction $r_{14'}$. Protein **CcpA** can bind to a site BS_{CcpA} of **Pilv–Leu** by reaction r_9 . This binding site overlaps with that of **CodY** [38]. Therefore, both proteins cannot bind simultaneously to the promoter, which we model by letting the reaction r_1 that binds **CodY** be subject to inhibition by **CcpA**. In addition, the binding of **CcpA** to the promoter **Pilv–Leu** accelerates the promoter [38, 43, 14]. Therefore, BS_{CcpA} is an accelerator of reaction r_2 . The catabolite acceleration by BS_{CcpA} can be inhibited when BS_{TnrA} is active, since then the active binding sites BS_{TnrA} and BS_{CcpA} can bind to each other while forming a DNA loop, which inhibits the acceleration. Therefore, BS_{TnrA} inhibits reaction r_9 .

The reactions of the network that can be disabled by gene knockouts are drawn in orange color. These are all reactions modelling gene expressions, that is r_8 and r_{14} , and the reactions modeling promoter binding of the inhibitors, that is r_1 and r_7 .

2.2.3 Steady-State Equations

We are interested in the deterministic steady state semantics of a reaction network, but projected to variables for exchange fluxes. Let \mathbb{R}_+ be the set of nonnegative real number including zero. For networks with n inflow species and m outflow species, the deterministic semantics of the network will be an exchange relation in $(\mathbb{R}_+)^{n+m}$. In our example, we have $n = 2$ and $m = 1$. However, it is never possible to determine the steady state semantics completely, given that we have only partial kinetic knowledge. Furthermore, note that we don't know about any initial concentrations, so it may be possible that different steady states are reached with different initial concentration, even if we had complete kinetic knowledge otherwise.

In addition, we are usually given a set of k reactions that are knockout candidates, but do not know which of them will be knocked out. In our example, we had $k = 4$ knockout candidates. In this case, the steady state semantics will be enriched with unknowns for knockout candidates, and the projection will be to variables for both knockout candidates and exchange fluxes. We thus obtain a richer knockout-exchange relation in $(\mathbb{R}_+)^{k+n+m}$.

Given a species s of the network, we will introduce a variable z_s for its concentration (or activity) in a steady state. For any inflow species we will introduce an influx variable x_s for the rate of its inflows, and for any outflow variable we will introduce an outflux variable y_s for the rate of its outflux. Furthermore, we will use variables v_{r_i} for the rate of reaction r_i in a steady state, and knockout variables o_{r_i} with values in $\{0, 1\}$, stating whether reaction r_i is off or on respectively.

One possible kinetic functions $kin_{S,Act,Acc,I}$ for a reaction, with set of with substrates S , activators Act , accelerators Acc and inhibitors I is equal to:

$$kin_{S,Act,Acc,I}((z_s)_{s \in S \cup Act \cup Acc \cup I}) = \prod_{s \in S} z_s \prod_{a \in Act} z_a \frac{(1 + \sum_{a' \in Acc} z_{a'})}{1 + \sum_{i \in I} z_i}$$

The function $kin_{S,Act,Acc,I}$ satisfies the basic requirements, which beside its well definedness for all possible concentrations, are that:

- it is monotonic in the concentrations of all substrates, activators, and accelerators, and
- it is anti-monotonic in the concentrations of all inhibitors, and
- it is 0 if the concentration of one of the activators or substrates becomes 0.

However, all the constants in the definition of $kin_{S,Act,Acc,I}$ all equal to 1. This is clearly randomly chosen, so there is no reason to belief that $kin_{S,Act,Acc,I}$ will be to the true kinetic function of any chemical reaction in our example or elsewhere. In order to face this situation, we will assume that the kinetic functions of reactions with substrates S , activators Act , accelerators Acc , and inhibitors I are similar to $kin_{S,Act,Acc,I}$. The precise definition of similarity in Definition 1 will be such that the above properties are satisfied.

So let $kin^{(i)}$ be a variable for the kinetic function of reaction r_i . Any steady state of the **Pllv–Leu** regulation network must then satisfy the equations in Table 2. The first equation states that the speed v_{r_1} of reaction r_1 is equal to 0 if the reaction is knocked out, so that $o_{r_1} = 0$, and equal to $kin^{(1)}(i : z_{CcpA}, act : z_{CodY})$ otherwise. In addition, the model imposes the constraint that unknown function $kin^{(1)}$ is similar to the known function $kin_{\emptyset, \{CodY\}, \emptyset, \{CcpA\}}$. The equations for the speeds of all other reactions are similar. Note that the roles of arguments are always indicated by attributes (i:inhibitor in I, act:activator in Act, acc:accelerator in Acc, s: substrate in S).

In a (limit) steady state, we need that all species are consumed and produced at equilibrium. For **CcpA** for instance, this requires that $v_{r_{14'}} = v_{r_{14}}$ and for **CodY** that $v_{r_{51}} = x_{CodY}$. Finally, we have to impose that the rate of an outflow species increase monotonically with its concentrations. For **Leu** this means that $y_{Leu} = kin^{(o)}(s : z_{Leu})$ for some kinetic function $kin^{(o)}$ that is similar to $kin_{\{Leu\}, \emptyset, \emptyset, \emptyset}$. Without such an outflow model, we would lack any relation between the concentration of the species z_{Leu} and the outflux of y_{Leu} . This would spoil our knockout predictions for leucine overproduction in particular.

The flux balance equations can be used to simplify the remaining equations by replacing equal by equal, while removing some local variables (those for reaction speeds v_{r_i} and species concentrations z_S). The simplified equations are given in Table 3. It should be noticed that such simplifications before abstract interpretation improve the precision of this analysis.

2.2.4 Similarity

Table 2: Steady state equations.

Reaction Speeds	Flux Balance Equations
$v_{r_1} = o_{r_1} \text{kin}^{(1)}(i : z_{\text{CcpA}}, \text{act} : z_{\text{CodY}})$	$y_{\text{Leu}} = v_{r_8}$
$v_{r_{1'}} = \text{kin}^{(1')}(s : z_{\text{BS}_{\text{CodY}}})$	$v_{r_{14'}} = v_{r_{14}}$
$v_{r_2} = \text{kin}^{(2)}(i : z_{\text{BS}_{\text{CodY}}}, \text{acc} : z_{\text{BS}_{\text{CcpA}}}, i : z_{\text{Leu}}, i : z_{\text{BS}_{\text{ThrA}}})$	$v_{r_{51}} = x_{\text{CodY}}$
$v_{r_{2'}} = \text{kin}^{(2')}(s : z_{\text{Pllv-Leu}})$	$v_{r_{52}} = x_{\text{ThrA}}$
$v_{r_7} = o_{r_7} \text{kin}^{(7)}(\text{act} : z_{\text{ThrA}})$	$v_{r_{1'}} = v_{r_1}$
$v_{r_{7'}} = \text{kin}^{(7')}(s : z_{\text{BS}_{\text{ThrA}}})$	$v_{r_{2'}} + v_{r_8} = v_{r_2}$
$v_{r_9} = \text{kin}^{(9)}(\text{act} : z_{\text{CcpA}}, i : z_{\text{BS}_{\text{ThrA}}})$	$v_{r_{7'}} = v_{r_7}$
$v_{r_{9'}} = \text{kin}^{(9')}(s : z_{\text{BS}_{\text{CcpA}}})$	$v_{r_{9'}} = v_{r_9}$
$v_{r_{51}} = \text{kin}^{(51)}(s : z_{\text{CodY}})$	
$v_{r_{52}} = \text{kin}^{(52)}(s : z_{\text{ThrA}})$	
$v_{r_{14}} = o_{r_{14}} \text{kin}^{(14)}()$	
$v_{r_{14'}} = \text{kin}^{(14')}(s : z_{\text{CcpA}})$	
$v_{r_8} = o_{r_8} \text{kin}^{(8)}(s : z_{\text{Pllv-Leu}})$	
	Outflow Model
	$y_{\text{Leu}} = \text{kin}^{(o)}(z_{\text{Leu}})$

When doing flux analysis or knockout prediction, we are interested in what may happens with the steady states, when the model is changed, either by changing some influx or outflux, or by knocking out some of the reactions, or both.

The state space for the change of a concentrations or of a reaction speed $\mathbb{R}_+ \times \mathbb{R}_+$. We partition this space into the 6 difference relations in $\Delta = \{\uparrow, \downarrow, \uparrow\downarrow, \sim, \approx\}$. The increase relation $\uparrow = \{(x, y) \mid 0 < x < y\}$ contains all increases but not from zero, while the increase-from-zero relation $\uparrow\downarrow = \{(0, y) \mid 0 < y\}$ contains all other increases. The relation \downarrow captures all decreases but not to zero, while $\uparrow\downarrow$ covers all decreases to zero. There are also two ways of no-change, those at a value different to zero $\sim = \{(x, x) \mid x \neq 0\}$ and the no-change at zero $\approx = \{(0, 0)\}$.

Let R be a relation in $(\mathbb{R}_+)^p$ for some $p \geq 0$. We define the abstract interpretation of R over Δ as follows:

$$R^\Delta = \{(\delta_1 \dots, \delta_p) \in \Delta^p \mid \exists (x_1, \dots, x_p) \in R, (y_1, \dots, y_p) \in R, (x_1, y_1) \in \delta_1 \dots, (x_p, y_p) \in \delta_p\}$$

Since kinetic functions $\kappa : (\mathbb{R}_+)^{p-1} \rightarrow \mathbb{R}_+$ can be seen as p -ary relations, their abstract interpretations κ^Δ are well defined.

Definition 1. We call two kinetic functions $\kappa_1, \kappa_2 : (\mathbb{R}_+)^{p-1} \rightarrow \mathbb{R}_+$ similar if and only if $\kappa_1^\Delta = \kappa_2^\Delta$.

It can now be verified that all kinetic functions similar to $\text{kin}_{S, \text{Act}, \text{Acc}, I}$ satisfy the three properties required above. Furthermore, similarity is preserved when changing the parameter of mass action kinetics. It can even be seen, that the binary mass action kinetics is similar to the Michaelis-Menten kinetics.

Table 3: Simplified steady state equations.

$v_{r_1} = o_{r_1} kin^{(1)}(i : z_{CcpA}, act : z_{CodY})$	$v_{r_9} = kin^{(9')}(s : z_{BS_{CcpA}})$
$v_{r_1} = kin^{(1')}(s : z_{BS_{CodY}})$	$x_{CodY} = kin^{(51)}(s : z_{CodY})$
$v_{r_2} = kin^{(2)}(i : z_{BS_{CodY}}, acc : z_{BS_{CcpA}},$	$x_{TnrA} = kin^{(52)}(s : z_{TnrA})$
$i : z_{Leu}, i : z_{BS_{TnrA}})$	$v_{r_{14}} = o_{r_{14}} kin^{(14)}()$
$v_{r_2} = kin^{(2')}(s : z_{Pllv-Leu}) + v_{r_8}$	$v_{r_{14}} = kin^{(14')}(s : z_{CcpA})$
$v_{r_7} = o_{r_7} kin^{(7)}(act : z_{TnrA})$	$y_{Leu} = o_{r_8} kin^{(8)}(s : z_{Pllv-Leu})$
$v_{r_7} = kin^{(7')}(s : z_{BS_{TnrA}})$	$y_{Leu} = kin^{(o)}(s : z_{Leu})$
$v_{r_9} = kin^{(9)}(act : z_{CcpA}, i : z_{BS_{TnrA}})$	

3 Results

3.1 Knockout prediction algorithm

We now generalize the knockout prediction algorithm from [21] so that it applies to all reaction networks modeled in the language recalled in Section 2.2. The difficulty compared to there is to overcome the limitation to reactions with mass action kinetics. This will be done by extending the reasoning methods from [31] for our modeling language so that we can predict reaction knockouts, instead of changes of influxes or outfluxes only.

We have already prepared this extension by adding the following two aspects from [21] to our presentation of the modeling language from [31]. Firstly, we insert knockout variables o_{r_i} into steady state equations in Section 2.2.3. And secondly, we have chosen the domain with the six elements $\Delta = \{\uparrow, \downarrow, \uparrow\downarrow, \sim, \approx\}$ for abstract interpretation in Section 2.2.4, rather than the three element domain $\Delta' = \{<, =, >\}$. The intuition is that the change performed by the knockout of a reaction r_i can be reflected by a difference constraint $o_{r_i} = \downarrow$, meaning that r_i was on before and off after the change.

3.1.1 Abstract Interpretation

Abstract interpretation can be lifted to the steady state equations as obtained from a reaction network. For this, all variables get interpreted over Δ (instead of \mathbb{R}_+), and all arithmetics functions are interpreted abstractly over Δ (instead of \mathbb{R}_+) as defined in Section 2.2.4: Addition is interpreted as $+\Delta$, multiplication as $\cdot\Delta$ and unknown kinetic functions $kin^{(i)}$ as the fully known functions $(kin^{(i)})^\Delta = (kin_{S,Act,Acc,I})^\Delta$. It should be noticed that functions can become set valued by abstract interpretation. For instance, $+\Delta(\uparrow, \downarrow) = \{\uparrow, \downarrow, \sim\}$, so that abstract interpretation will produce difference constraints of the form $z_1 \in z_2 + z_3$.

The abstract interpretation of the simplified steady state equations in Table 3, for instance, are given in Table 4. Most importantly, we could replace the variables $kin^{(i)}$ by the definition of $kin_{S,Act,Acc,I}$, since $(kin^{(i)})^\Delta = (kin_{S,Act,Acc,I})^\Delta$. In this way, we successfully abstracted away the lacking kinetic knowledge.

Table 4: Difference constraints obtained by abstract interpretation, where $inh(x) = 1/acc(x)$ and $acc(x) = 1 + x$.

$v_{r_1} \in o_{r_1} \cdot inh(z_{CcpA}) \cdot z_{CodY}$	$x_{CodY} = z_{CodY}$
$v_{r_1} = z_{BS_{CodY}}$	$x_{TnrA} = z_{TnrA}$
$v_{r_2} \in inh(z_{BS_{CodY}} + z_{Leu} + z_{BS_{TnrA}}) \cdot acc(z_{BS_{CcpA}})$	$v_{r_{14}} = o_{r_{14}}$
$v_{r_2} \in z_{Pllv-Leu} + v_{r_8}$	$v_{r_{14}} = z_{CcpA}$
$v_{r_7} \in o_{r_7} \cdot z_{TnrA}$	$y_{Leu} \in o_{r_8} \cdot z_{Pllv-Leu}$
$v_{r_7} = z_{BS_{TnrA}}$	$y_{Leu} = z_{Leu}$
$v_{r_9} \in z_{CcpA} \cdot inh(z_{BS_{TnrA}})$	
$v_{r_9} = z_{BS_{CcpA}}$	

We next recall the main insight from John et al. [21], which states that the difference assignment between two steady states of a given the reaction network must satisfy the abstract interpretation of its steady state equations. Indeed, the proposition is slightly more general, so that it remains independent of the choice of the modeling language. For any two variable assignments ν, ν' into \mathbb{R}_+ and any variable x they consider the unique difference relation $\delta(x) \in \Delta$ such that $(\nu(x), \nu'(x)) \in \delta(x)$.

Proposition 1 (John et al. [21]). *If ν and ν' are two solutions of the arithmetic equations ϕ over \mathbb{R}_+ then their difference assignment δ satisfies the abstract interpretation of ϕ over Δ .*

The solutions ν and ν' of interest are the steady states of a reaction network with simplified steady state equations ϕ . More precisely, ν should be a solution of ϕ without any knockout and ν' another solution of ϕ with some knockout. The abstract interpretation of ϕ must then be satisfied by all possible changes δ from ν to ν' .

3.1.2 Knockout prediction by abstract interpretation

A knockout prediction problem receives as inputs a reaction network, a subset of reactions that are candidates for knockouts, and an overproduction target. In the case of the **Pllv-Leu** network, the knockout candidates are r_1 , r_7 , r_8 , and r_{14} , and the overproduction target is $y_{Leu} = \uparrow$. The questions then is which of the knockout candidates can be knocked out in the network, while satisfying the overproduction target.

Let ϕ be the simplified steady state equations of the network. The question then is whether there exists two variable assignments ν and ν' solving ϕ such that $\nu(o_{r_i}) = 1$ for all knockout variables in ϕ , and such that the difference assignment of ν and ν' satisfies the overproduction target. In this case, a solution for the knockout problem is the subset of all reactions r_i such that $\nu'(o_{r_i}) = 0$.

In order to search for two variable assignments satisfying ϕ and the overproduction target, we can apply Proposition 1. It shows that it is sufficient to

Table 5: The top-4 solutions of the difference constraints in Table 4 and the overproduction target $y_{\text{Leu}} = \uparrow$.

solution	o_{r_1}	o_{r_7}	o_{r_8}	$o_{r_{14}}$	x_{CodY}	x_{TnrA}	y_{Leu}	penalty
1.	\sim	\sim	\sim	\sim	\downarrow	\sim	\uparrow	2
2.	\sim	\sim	\sim	\sim	\sim	\downarrow	\uparrow	2
3.	\Downarrow	\sim	\sim	\sim	\sim	\sim	\uparrow	2
4.	\sim	\Downarrow	\sim	\sim	\sim	\sim	\uparrow	2

For any i and S , $o_{r_i} = \Downarrow$ means a knockout of reaction r_i , $x_S = \downarrow$ a decrease of the influx of species S (but not to zero), and $y_S = \uparrow$ an increase of the outflux of S (but not from zero).

search for a variable assignment satisfying the abstract interpretation of ϕ over Δ . Furthermore, we require to be satisfied the overproduction target, the constraints $o_{r_i} = \{\Downarrow, \sim\}$ for all reactions r_i that can be knocked out, and $o_{r_i} = \{\sim\}$ for all others, and the flux restrictions $x_S, y_S \in \{\uparrow, \downarrow, \sim\}$ for all species S .

The number of solutions is often huge. Therefore, we are interested only in those solutions where the overall number of knockouts, influx changes, and outflux changes is as small as possible. The hope is that the solutions with the minimal number of changes contain exactly those changes that are really needed to satisfy the overproduction target. In order to restrict the search space further, we will rule out all solutions in $o_{r_i} = \Downarrow$ for more than one i .

3.2 Application to PIlv-Leu example network

We define the penalty of a solution as the number of global variables mapped to \uparrow , \downarrow , or \Downarrow . As argued above, the lower is the penalty, the better is the solution, since no irrelevant changes were added besides those that are really needed.

The top- n solutions of a difference constraint can be computed by a finite domain constraint solver with branch and bound optimization. The top-4 solutions of the difference constraints for the **PIlv-Leu** network are given in Table 5; these are exactly all solutions with penalty 2. Of course, there are many further solutions with higher penalties, but the top-4 solutions indeed correspond to the most interesting predictions in this case:

1. Decrease influx of **CodY**, or
2. Decrease influx of **TnrA**, or
3. Knockout reaction **r₁**, or
4. Knockout reaction **r₇**.

What is predicted for this small network are 2 flux changes and 2 single knockouts. However, a decrease of the fluxes x_{CodY} and x_{TnrA} can be obtained by knocking out the respective genes in the context, i.e., reaction r_{15} and r_{16} in the larger leucine network in Figure 3. Therefore, the 4 predictions for the small

PIIv–Leu network correspond to 4 single knockout predictions for the larger leucine model. This illustrates the modularity of our approach. It is obtained by admitting inflow and outflow species for the interaction of a module of a network with its context, without need to model the whole network.

3.3 Modeling of leucine metabolic pathway

A larger and more realistic model of leucine production is presented in Figure 3. It describes the metabolic transformations of **Thr**, **Pyr** and **Akb** into **Ile**, **Val**, and **Leu**, concomitantly with its regulation. Our model mainly follows the abstraction level chosen by the previous work in [15, 28], but the regulation is expressed more cleanly within our formal modeling language. In the supplementary material, we list the metabolites, proteins and actors of this model in Table 11, and all its reactions in Table 12.

Generally, we will assume that the precise kinetics are unknown, but hope that we still have good knowledge on inhibitors and activators of the reactions. We do not have to know all inhibitors and activators, but the more we know the better predictions we can expect. This is an important departure from previous methods purely based on flux balance equations.

In order to group several proteins that all participate in the same reactions, we introduce artificial species that we call protein clusters, and draw them as for instance **BkL**. This protein cluster presents all the proteins generated by the *bkL* operon.

We will slightly enrich our graphical syntax with some syntactic sugar, in order to keep bigger models readable. This sugar can be compiled away in a preprocessing step. First, we use split-points \bigcirc that link different copies of the same species. For instance, we have one copy of **Leu**, which are drawn as **Leu** and linked together with the original **Leu** over a split point. Logically, there is no difference between a species and its copies; it just allows nicer graphs with fewer intersections of edges. Second, we use edge-clusters that group several with the same sources or targets (for instance the activator edges from **BkL** and **YwaA** to r_{35} , r_{32} , and r_{29} are clustered).

Production of Isoleucine, Valine, and Leucine. For clarity, we here only consider those intermediates that produce the outflow species of the pathway, i.e. **Ket_a** for **Ile** in r_{26} , **Ket_b** for **Val** in r_{36} , and **Ket_c** for **Leu** in r_{45} . All three reactions must be activated by **YwaA+YbgE**, that is either by protein **YwaA** (r_6) or protein **YbgE** (r_5) which are exchangeable catalysts following [42]. There are also inverse reactions r_{29} , r_{32} , r_{35} , but with different activators, the proteins **Bcd** and **YwaA**. The intermediates **Ket_a**, **Ket_b**, and **Ket_c** can be transformed alternatively into **Acyl–CoA** by reactions r_{28} , r_{31} , and r_{34} , which can then outflow to the context for fatty acid biosynthesis. This is regulated by protein **BkL**.

The creation of **Ile**, **Val**, and **Leu** from **Ket_a**, **Ket_b**, **Ket_c** is always coupled with a parallel transformation of **Glu** into **OxoGlu**, and the inverse decompositions

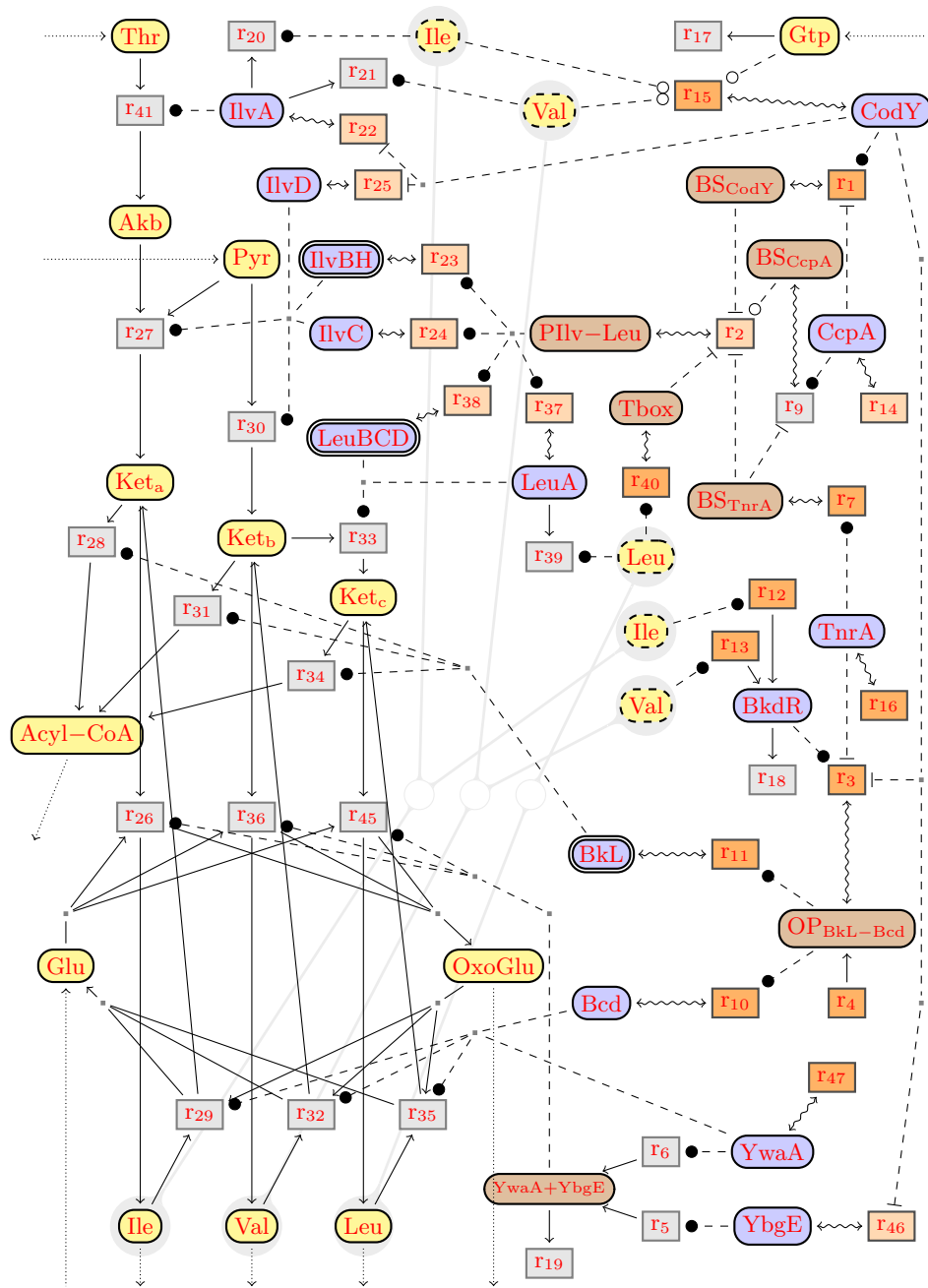


Figure 3: Leucine network.

are coupled with the inverse transformation of **OxoGlu** into **Glu**. The required **Glu** for the production of **Ile**, **Val**, and **Leu**, can inflow from the glutamate biosynthesis cycle of the context (derived from **OxoGlu**) produced **OxoGlu** can outflow from this pathway to the context (as for example used in the TCA cycle).

The intermediates are produced as follows: **Ket_a** is produced from **Akb** by **r₂₇**, **Ket_b** from **Pyr** by **r₃₀**, and **Ket_c** from **Ket_b** by **r₃₃**. **Akb** can be produced from **Thr** by reaction **r₄₁**, which in turn can also inflow from the threonine biosynthesis pathway of the context (which is linked to the glycolysis pathway via the aspartate biosynthesis). **Pyr** can inflow from the glycolysis pathway of the context. **IlvA** activates the conversion of **Thr** into **Akb**. **IlvBH** (a cluster of **IlvB** and **IlvH**), **IlvC** and **IlvD** activates the transformation of **Akb** to **Ket_a**. The proteins **LeuA** and **LeuBCD** (a cluster of **LeuA**, **LeuB**, and **LeuC**) perform a sequence of reactions justifying the transformation of **Ket_b** to **Ket_c** [28]. The genes of the proteins **IlvBH**, **IlvC**, **LeuA**, and **LeuBCD** are co-located in the same operon (*ilv-leu*), so they are under the dependence of one and the same promoter **PIlv-Leu** [43]. This specific regulation is detailed below.

Regulation. The expression of the different enzymes involved in the production of branched chain amino acids (explained above) are denoted by the following reactions: **r₂₂** for **IlvA**; **r₂₅** for **IlvD**; **r₂₃** for **IlvBH**; **r₂₄** for **IlvC**; **r₃₈** for **LeuBCD**; **r₃₇** for **LeuA**; **r₁₁** for **BkL**; **r₁₀** for **Bcd**; **r₄₇** for **YwaA** and **r₄₆** for **YbgE**. All these expression reactions are controlled by a complex regulation [5, 29, 27]. Indeed, we consider proteins **CcpA**, **CodY**, **TnrA**, and **BkdR** that have regulatory functions and are expressed respectively by the reactions **r₁₄**, **r₁₅**, **r₁₆** and **r₁₂**, **r₁₃**. The expression reactions of these regulators can be modulated by different factors from outside the context, which are not modelled here. **Gtp** come from the context and can be degraded through the reaction **r₁₇**. This metabolite increase the affinity of **CodY** for its binding sites, as well as **Ile** and **Val** [38, 43, 44, 29, 27].

Expression of **IlvA**, **IlvD**, and **YbgE** are dependent on their own promoters and their expressions are down-regulated by **CodY** at the transcriptional level [29, 27], which is represented by reactions **r₂₂**, **r₂₅**, and **r₄₆**, respectively. In [18] it was further shown that **IlvA** is deactivated by both **Ile** and **Val** (reactions **r₂₀**, **r₂₁**). The transcription of all proteins **IlvC**, **IlvBH**, **LeuA**, and **LeuBCD** starts through the activation of their promoter **PIlv-Leu**, which controls the reactions **r₂₃**, **r₂₄**, **r₃₇** and **r₃₈**, respectively [5, 27]. Its regulation is captured by the reaction **r₂** and is dependent on several mechanisms, which are described before. In addition to what has been described above we introduced here the down-regulation of the promoter **PIlv-Leu** by **Leu** in terms of ribosome-mediated attenuation mechanism **Tbox** (**r₄₀**) [5, 16] and the deactivation of **LeuA** by **Leu** which is represented by the reaction **r₃₉**. Proteins **BkL** and **Bcd** lie under the action of two common but independent promoters (**Prm_b** and **Prm_c**), whose activity level is represented by actor **OP_{BkL-Bcd}**. Among them one is constitutive (no regulation, represented by the reaction **r₄**), and the other is positively impacted by **BkdR** and down-regulated by both **TnrA** and **CodY**, as reflected by reaction **r₃** [10]. Although not regulated on the transcriptional level, protein **BkdR** is activated by **Ile** and **Val**, which is modeled by introducing the two **BkdR** producing

Table 6: Top-2 solutions: no knockouts and 4 context changes.

<i>solution</i>	x_{Glu}	x_{Pyr}	$y_{\text{Acyl-CoA}}$	y_{Leu}	y_{OxoGlu}	<i>penalty</i>
1.	\uparrow	\uparrow	\sim	\uparrow	\uparrow	4
2.	\uparrow	\sim	\downarrow	\uparrow	\uparrow	4

For any i and S , $x_S = \downarrow$ means a decrease of the influx of species S (but not to zero), and $y_S = \uparrow$ an increase of the outflux of S (but not from zero).

reactions r_{12} and r_{13} . Proteins **YbgE** (expressed by the reaction r_{46}) is down regulated by **CodY**. However, no regulation is known for protein **YwaA**, which is expressed by the reaction r_{47} .

3.4 Knockout Prediction for leucine overproduction

As for the small model, our target for knockout prediction is leucine overproduction, i.e., $y_{\text{Leu}} = \uparrow$. But in the big model, there are also some unwanted possibilities to do so, which are to decrease the outfluxes of valine or isoleucine, since the remaining valine and isoleucine could then be used to produce more leucine. In order to rule out this, we add to our knockout prediction target that the outfluxes of **Val** and **Ile** cannot decrease, i.e., $y_{\text{Val}}, y_{\text{Ile}} \in \{\sim, \uparrow\}$.

We first present the solutions of the solver for difference constraints, and then show how they can be interpreted as knockout predictions. In all the solutions we observed it holds that $x_{\text{Glu}} = y_{\text{OxoGlu}} = y_{\text{Leu}} = \uparrow$, so this seems to be always the case. This shows that any increase of y_{Leu} requires to increase the influx of **Glu** and the outflux of **OxoGlu**. It implies for all solutions, that their minimal overall penalty is at least 3.

The top-2 solutions presented in Tables 6 have penalty 4. These do not require any reaction knockout. Besides the 3 necessary context changes discussed above, we either have that $y_{\text{Acyl-CoA}} = \downarrow$, i.e. the outflux of **Acyl-CoA** is decreased, or else that $x_{\text{Pyr}} = \uparrow$, so that the influx of **Pyr** is increased. Both possibilities are plausible, since for overproducing **Leu**, either the production of **Leu** by r_{45} must be increased, and this requires more **Pyr**, or else the outflux of **Leu** into the fatty acid pathway (via reactions r_{35} and r_{34}) must be decreased, and this imposes a higher outflux of **Acyl-CoA**.

All the single knockout solutions with penalty 5 are given in Table 7, with numbers 3-28. Indeed, it turns out that they all extend on either of the top-2 solutions. This is a rather particular and surprising situation. It means that none of these knockouts is strictly necessary to justify **Leu** overproduction, but that they are all compatible with it. In other words, all other candidates which were not selected are incompatible with a single knockout and 4 context changes. A further single knockout if found with penalty 6 in the solutions with number 142 and 143.

Table 7: Next best solutions: a single knockout and 4 context changes.

solution	knockouts							context changes					penalty
	o_{r_1}	$o_{r_{10}}$	$o_{r_{12}}$	$o_{r_{13}}$	o_{r_3}	o_{r_4}	o_{r_7}	x_{Glu}	x_{Pyr}	$y_{Acyl-CoA}$	y_{Leu}	y_{OxoGlu}	
3.	↓	≈	≈	≈	≈	≈	≈	↑	↑	≈	↑	↑	5
4.	↓	≈	≈	≈	≈	≈	≈	↑	≈	↓	↑	↑	5
5.	≈	≈	≈	≈	↓	≈	≈	↑	↑	≈	↑	↑	5
6.	≈	≈	≈	≈	↓	≈	≈	↑	≈	↓	↑	↑	5
7.	≈	≈	≈	≈	≈	↓	≈	↑	↑	≈	↑	↑	5
8.	≈	≈	≈	≈	≈	↓	≈	↑	≈	↓	↑	↑	5
9.	≈	≈	≈	≈	≈	≈	↓	↑	↑	≈	↑	↑	5
10.	≈	≈	≈	≈	≈	≈	↓	↑	≈	↓	↑	↑	5
11.	≈	↓	≈	≈	≈	≈	≈	↑	↑	≈	↑	↑	5
12.	≈	↓	≈	≈	≈	≈	≈	↑	≈	↓	↑	↑	5
13.	≈	≈	↓	≈	≈	≈	≈	↑	↑	≈	↑	↑	5
14.	≈	≈	↓	≈	≈	≈	≈	↑	≈	↓	↑	↑	5
15.	≈	≈	≈	↓	≈	≈	≈	↑	↑	≈	↑	↑	5
16.	≈	≈	≈	↓	≈	≈	≈	↑	≈	↓	↑	↑	5

solution	knockouts						context changes					penalty
	$o_{r_{47}}$	$o_{r_{15}}$	$o_{r_{16}}$	$o_{r_{40}}$	$o_{r_{46}}$	$o_{r_{14}}$	x_{Glu}	x_{Pyr}	$y_{Acyl-CoA}$	y_{Leu}	y_{OxoGlu}	
17.	≈	↓	≈	≈	≈	≈	↑	↑	≈	↑	↑	5
18.	≈	↓	≈	≈	≈	≈	↑	≈	↓	↑	↑	5
19.	≈	≈	↓	≈	≈	≈	↑	↑	≈	↑	↑	5
20.	≈	≈	↓	≈	≈	≈	↑	≈	↓	↑	↑	5
21.	≈	≈	≈	↓	≈	≈	↑	↑	≈	↑	↑	5
22.	≈	≈	≈	↓	≈	≈	↑	≈	↓	↑	↑	5
23.	↓	≈	≈	≈	≈	≈	↑	↑	≈	↑	↑	5
24.	↓	≈	≈	≈	≈	≈	↑	≈	↓	↑	↑	5
25.	≈	≈	≈	≈	↓	≈	↑	↑	≈	↑	↑	5
26.	≈	≈	≈	≈	↓	≈	↑	≈	↓	↑	↑	5
27.	≈	≈	≈	≈	≈	↓	↑	↑	≈	↑	↑	5
28.	≈	≈	≈	≈	≈	↓	↑	≈	↓	↑	↑	5

solution	knockouts	context changes					penalty	
	$o_{r_{11}}$	x_{Glu}	x_{Thr}	$y_{Acyl-CoA}$	y_{Ile}	y_{Leu}		y_{OxoGlu}
142.	↓	↑	↓	↓	≈	↑	↑	6
143.	↓	↑	≈	↓	↑	↑	↑	6

For any i and S , $o_i = \downarrow$ means a knockout of reaction r_i , $x_S = \downarrow$ a decrease of the influx of species S (but not to zero), and $y_S = \uparrow$ an increase of the outflux of S (but not from zero).

There are 38 further solutions with a penalty of 5 which do not require any knockouts. Of these 12 are extensions of the top-2 solutions by further context changes, and the others remove assign one knockout variable to \approx , meaning that one reaction is removed from the model. There are many further solution with

Table 8: Single knockout predictions for leucine overproduction.

knock-out	description of reaction	knockout effect	mutant
r_1	CodY binding to BS_{CodY}	increases Pllv-Leu activity	BBG251
r_3	bind BkdR to BkL-Bcd promoter	decrease Bcd and thus r_{35}	
r_4	constitutive expression of BkL-Bcd operon	decreases Bcd and thus r_{35}	
r_7	TnrA binding to BS_{TnrA}	increases Pllv-Leu activity	
r_{10}	Bcd expression	deactivates reaction r_{35}	BBG253
r_{11}	BkL expression	deactivates Ket_a , Ket_b , Ket_c outflow via Acyl-CoA by r_{28} , r_{31} , r_{34}	BBG255
r_{12}	activate BkdR by Ile	decreases speed of r_3 and thus Bcd	BBG254
r_{13}	activate BkdR by Val	decreases speed of r_3 and thus Bcd	
r_{15}	CodY expression	increase Pllv-Leu activity	
r_{16}	TnrA expression	increase Pllv-Leu activity	
r_{40}	Tbox of Leu attenuation	increase Pllv-Leu activity	BBG256
r_{47}	expression of YwaA	increase speed of r_{35}	BBG252
r_{14}	expression of CcpA	unclear	
r_{46}	expressions of YbgE	unclear	

penalty 6 ranging from 67-784 and many more solutions with penalty 7 ranging from 785-4249. But these do not add further single knockout predictions. The computation of all solutions up to penalty 7 requires about 6:21 minutes on a MacBook Pro laptop (2.4 GHz Intel core i7). For higher penalties, our algorithm is running out of memory.

All solutions corresponding to single knockouts are collected in Table 8. This can be done automatically by our tools. In addition, we have added an informal explanation of the positive effect of these single knockouts with respect to leucine overproduction. These effect can be seen in the model rather easily. It turns out, however that knockout of reactions r_{14} and r_{46} do not have clear positive effect. The only effect of a knockout of reaction r_{46} is a deactivation of r_{45} producing Leu, which is unwanted. The knockout of reaction r_{14} has a negative effect on the activity of Pllv-Leu . This is not very plausible for leucine overproduction: it could still be argued that a decrease of the activity of Pllv-Leu will decrease the speed of r_{27} which in turn will increase the speed of r_{30} , but at the same time, it also has a negative effect on the activation of r_{30} , which counter balances this positive effect on the Leu production. Moreover, it have been demonstrated previously that a knockout of *ccpA* causes a decrease of *ilv-leu* expression [5].

² Whenever we talk about a knockout for the *bkL* operon, we mean the *lpdV* gene of this operon.

3.5 Effect of leucine feeding in surfactin overproduction

In order to check the working hypothesis that an increase in the intracellular pool of leucine led to an increase of the surfactin production, a preliminary experiment was performed by adding leucine in the Landy medium (instead of $(\text{NH}_4)_2\text{SO}_4$). In the medium with only $(\text{NH}_4)_2\text{SO}_4$ as the nitrogen source the specific surfactin production reached 117.09 ± 19.62 mg/g of DW. In contrast, the presence of leucine in the medium led to 3-fold increase of the specific surfactin production of **BBG111** (328.41 ± 31.94 mg/g of DW) (data not shown). This experiment shows that an increase in the extracellular concentration of leucine certainly produces an intracellular increase, which was beneficial for the production of surfactin. This confirms that intracellular concentration of the precursors are one of the most limiting factor for surfactin overproduction. Therefore, the computer modeling of the metabolic pathway of leucine and the knockout prediction obtained appear consistent and can be checked by genetic engineering.

3.6 Genetic engineering of *B. subtilis* for the overproduction of surfactin and one of its precursor

In this part, we present the different phenotypic profiles of the mutant strains obtained through genetic engineering of *B. subtilis* in terms of growth and specific surfactin production. These results were obtained in different culture media when the cells were in the exponential growth phase (about 6 hours of culture). Indeed, at this time the *sfA* operon's promoter reaches its maximum activity [8, 11], as well as the *ilvB* operon is highly expressed [28, 30].

3.6.1 Genetic engineering of *B. subtilis* and growth analysis

According to the prediction results based on the model, various genes were deleted. Firstly, to inhibit the **CodY** binding to its high affinity site on the **PIIv–Leu** promoter (r_1) and secondly, to delete the *codY* expression (r_{15}), the strains **BBG251** and **BBG254** were constructed. **TnrA** also negatively regulates the *ilv-leu* expression under nitrogen-limited condition, the knockout predictions giving as the result the deletion of the **TnrA** binding on the **PIIv–Leu** promoter (r_7) and the expression of *tnrA* (r_{16}). To achieve the same effect, we have chosen to completely suppress the expression of *tnrA* (r_{16}) in the mutant strain **BBG256**. Leucine intracellular concentration can inhibits the expression of *ilv-leu* operon via **Tbox** attenuator. In the mutant strain **BBG252**, **Tbox** was suppressed (r_{40}), as recommended by the predictions. Lastly, prediction have suggested to avoid leucine degradation by the knockout of the reactions r_3 , r_4 , r_{12} , r_{13} and r_{47} . To get this, the expression of *bcd* (r_{10}) and *bkL²* (r_{11}) were suppressed, giving the strains **BBG253** and **BBG255**, respectively. Growth kinetics study was carried out during the exponential growth phase in Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$ and in TSS medium with 16 amino acids (Table 9). The specific growth rate of the control strain **BBG111** was found to be 0.47 ± 0.06 h^{-1} and 0.55 ± 0.05 h^{-1} in the former and later media respectively. The specific growth rate of various mutant were not significantly different from the control strain

Table 9: Specific growth rate and specific surfactin production of the different mutants strains derived from *B. subtilis* 168.

Strains of <i>B. subtilis</i>	Knock-out	In TSS medium		In Landy medium	
		Specific growth rate (h^{-1})	Specific surfactin production (mg/g of DW)	Specific growth rate (h^{-1})	Specific surfactin production (mg/g of DW)
BBG111		0.55±0.05	25.37±13.03	0.47±0.06	134.39±32.23
BBG251	r₁	0.58±0.06	235.36±7.41	0.51±0.01	366.94±5.12
BBG252	r₄₀	0.55±0.02	132.88±15.41	0.51±0.02	310.40±16.49
BBG253	r₁₀	0.38±0.08	147.76±3.04	0.39±0.05	634.13±35.62
BBG254	r₁₅	0.39±0.03	529.29±79.93	0.48±0.03	1300.21±177.76
BBG255	r₁₁	0.58±0.07	63.78±7.89	0.54±0.04	580.46±58.22
BBG256	r₁₆	0.59±0.01	42.03±12.14	0.47±0.12	817.99±59.01

BBG111 in the Landy medium supplemented with $(NH_4)_2SO_4$. However in TSS medium with 16 amino acids, only the specific growth rates of **BBG253** and **BBG254** were lower and significantly different from that of the control strain.

3.6.2 Surfactin yield as an indicator for increase in intracellular leucine

Specific surfactin production for **BBG111** along with the mutant strains are provided in Figure 4A for TSS medium with 16 amino acids and in Figure 4B for Landy medium supplemented with $(NH_4)_2SO_4$ and summarized in Table 9. In TSS medium with 16 amino acids, the specific surfactin production of the different mutant strains is better than the control strain. Statistical analysis reveals that the results obtained for **BBG256** (*tnrA* deletion) and **BBG255** (*bkL²* deletion) mutants was not significantly different from the control strain. Moreover, the specific surfactin production of **BBG251** (*BS_{CodY}* deletion) and **BBG253** (*bcd* deletion) can be considered in the same statistical group. In Landy medium supplemented with $(NH_4)_2SO_4$, the specific surfactin production of the different mutant strains is also higher than the control strain. Statistical analysis reveals that these results were significantly different from those obtained with the control strain, except for the strain **BBG251**, in which the *CodY* high affinity binding site was deleted (*BS_{CodY}*). The specific surfactin production of this strain can be also considered as close to those obtained with the strain **BBG252** (*Tbox* deletion). Moreover, the specific surfactin production of **BBG253** (*bcd* deletion) and **BBG255** (*bkL²* deletion), which are the two strains modified in the branched chain amino acids degradation pathway can be considered in the same statistical group.

4 Discussion

The metabolic flux analysis is a central method of cell engineering to increase the productivity of host microorganisms such as bacteria, yeast and fungi. In

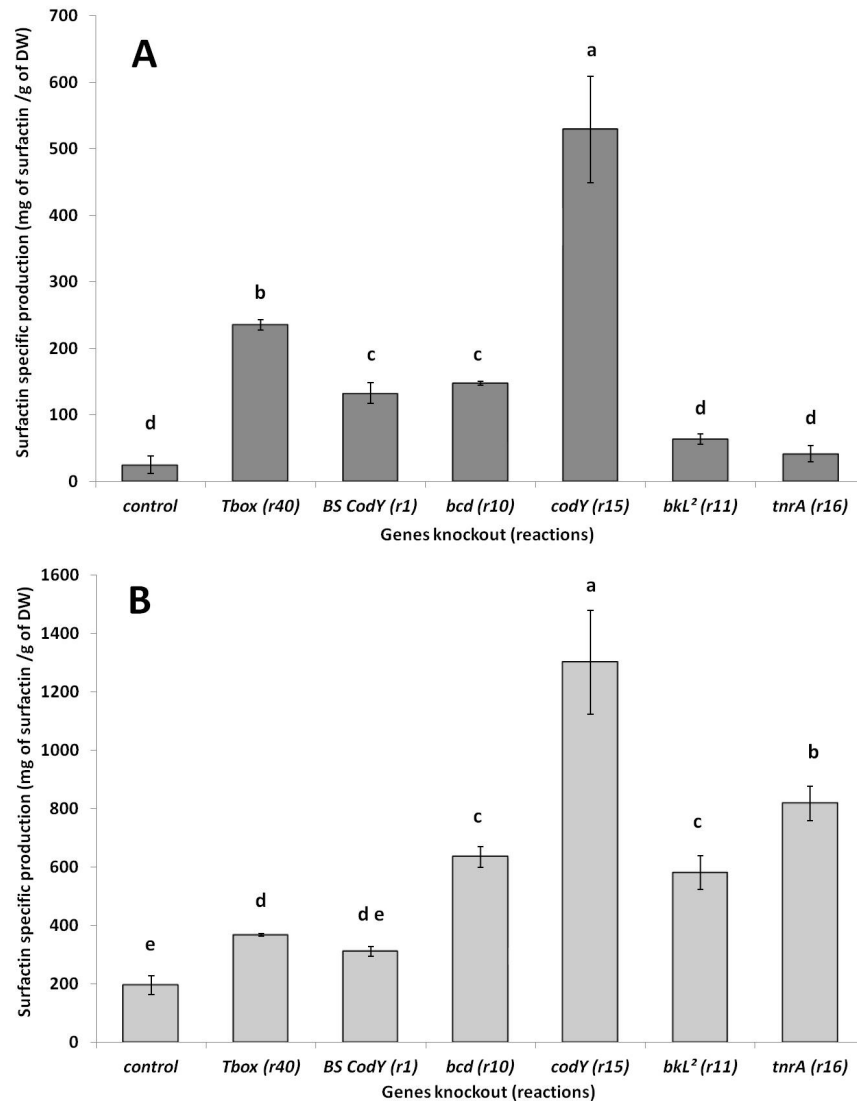


Figure 4: Specific surfactin production of the different knockout mutants strains after 6 hours of culture in TSS medium containing mix of 16 amino acids as nitrogen source (A) and in Landy medium containing $(\text{NH}_4)_2\text{SO}_4$ as nitrogen source (B). Results are expressed as mean value with standard deviation and statistical analysis of these data was made using one-way ANOVA using a Student-Newman-Keuls post hoc test (p value = 0.05). Different letters were used to indicate significant differences between groups.

this area, many fundamental studies have been conducted on the pathways regulations and metabolic flux analysis [32]. The use of bioinformatics in recent years helped to develop tools for reconstruction of metabolic pathways and genome scale metabolic modeling in order to overexpress metabolites, but most of them are only applied to primary metabolites and not on secondary metabolites [39, 15, 1, 24, 35]. In this work we are interested to overproduction of a secondary metabolite, i.e., the surfactin biosynthesized by NRPS. In a previous work, the problem of lipopeptide synthetase expression was avoided by replacing the surfactin operon promoter by a constitutive one, and this revealed only a small enhancement of surfactin production suggesting that the precursor supply is the problem for the surfactin overproduction in *B. subtilis* 168 derivative strains and not the synthetase expression [8]. To overcome this limitation, we have developed here for the first time an original and fully integrated approach of synthetic biology to overproduce the secondary metabolite surfactin and its precursors. This approach was focused on a modeling language for reaction networks with partial information and on a new knockout prediction algorithm based on abstract interpretation, that applies to all models written in this language. We have applied this new knockout prediction algorithm to the overproduction of intracellular leucine production, which in turn was conjectured to increase the production of surfactin in different mutants of *B. subtilis* 168.

To achieve our result, we have completely modeled one of the most complex and regulated metabolic pathway of *B. subtilis*, i.e. the branched chain amino acid metabolic pathway, in the new modeling language. When applied to the regulation subnetwork of the *ilv-leu* operon, our prediction algorithm returned 4 solutions corresponding to 4 single knockout predictions with respect to the larger model. And when applied to the complete model, 12 plausible single knockout predictions were returned, which included the 4 previous predictions. From these 12 predictions, we carried-out the 6 with most direct effect and verified them *in-vivo*. Since, the peptide moiety of surfactin contains four leucine molecules; we have measured the surfactin concentration as an indicator to detect the intracellular level of leucine.

The regulation of the *ilv-leu* operon is dependent on the availability and nature of carbon and nitrogen sources in the chemical medium (the context), since these are involved in various global regulators i.e. **CcpA** (carbon), **TnrA** (nitrogen) and **CodY** (amino acids and **Gtp**) and other regulators which depends upon the influence of these global regulators and change in concentration of various intracellular metabolites [38, 44, 5, 4, 14]. Therefore, our results also depend on the medium that is chosen. To analyse the specific surfactin production the mutants strains were cultivated in two different media, which contain a different composition in carbon and nitrogen sources. Indeed, TSS medium with 16 amino acid is low concentrated in glucose (5 g/L), but very rich in organic nitrogen (4.5 g/L), unlike the Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$ is very rich in glucose (20 g/L) and mainly contains inorganic nitrogen (3.6 g/L) and a small amount of organic nitrogen (1 g/L yeast extract). As it shown on the Figure 4 the difference in medium composition led to difference in specific

surfactin production. The production is increased for all the mutants in glucose-rich medium (Landy), and this is also true for the control strain. This finding confirms the prediction made without knockout (penalties 4; shown in Tables 6) where an increase in pyruvate influx leads to an increase of leucine production. Pyruvate is a direct result of glycolysis, we could consider that an increase in glucose concentration in the medium may result in an increase of pyruvate intracellular concentration. Otherwise, the difference in nitrogen supply reveals a very different behavior of the mutants.

Most of the regulatory genes involved in the branched chain amino acid metabolic pathway being pleiotropic regulators in *B. subtilis*, so the effect on the growth of the strains in which these genes were deleted has also been studied. Growth analysis of the different mutant strains shows that the deletion of *tnrA* and *Tbox* have no significant impact on the specific growth rate of the strains. These data suggest that no significant influence was contributed by these regulators on the growth of the mutants regardless of the carbon and nitrogen composition of the medium. The disruption of high affinity *CodY* binding site has no effect on the growth of the strain but the *codY* deletion mutant has a detrimental effect on the growth in comparison to **BBG111** in TSS medium with 16 amino acids. The delay in the growth of **BBG254** in this medium may be due to the impact of *CodY* on the growth of the strain in carbon limited condition. Regarding the genes involved in the branched chain amino acid degradation *bcd* and *bkL*², which encodes for branched chain amino acid dehydrogenase and E3 lipoamide dehydrogenase (*lpdV*) [10] the effects shown are different. Thus, a significant reduction in the growth rate of the strain deleted for *bcd* (**BBG253**) was observed in comparison to the control strain **BBG111** in the TSS medium with 16 amino acids. This has also occurred to a lesser extent in the Landy medium. *bcd* deletion may lead to a disturbance in the synthesis of branched-chain fatty acids, which plays an important role in maintaining the fluidity of the membrane lipids [23]. In contrast, in absence of *lpdV*, the production of α -keto acids occurs, which suggests that *lpdV* is not essential as two other homologs of lipoamide dehydrogenase exist in *B. subtilis* complementing its function [10]. Although, for some of these genes knocked out an effect on growth is observed, all the obtained mutants are viable because these genes are not listed as essential for the cell [7].

The expression of *ilv-leu* operon is regulated by *CodY* in the presence of branched chain amino acid and is known to respond to high content of *Gtp* present in exponentially growing phase [36]. In the Landy medium supplemented with $(\text{NH}_4)_2\text{SO}_4$, there was an increase of 7 times in the specific surfactin production with **BBG254** (*codY* deleted; **r₁₅**), which reached an exceptional value of 1.3 g of surfactin produced per gram of dried biomass. This provides evidence for more expression of *ilv-leu* in the mutant strains. Thus, results in an increase of intracellular pool of leucine. The production of the strain **BBG251** (**BS_{CodY}** deleted; **r₁**) was not significantly different in this condition than the control strain. But, in TSS medium with 16 amino acids, the difference between *CodY* mutants and the control strain is more important. An enhancement of 21 times in the specific surfactin production was shown when *codY* is totally

deleted (BBG254) and 5 times when the **CodY** high affinity binding site is removed (**BS_{CodY}**; BBG251), in comparison to BBG111. These results highlight the strong negative regulation provided by **CodY** on the branched chain amino acids metabolism and specially when the medium is rich in amino acids. Shivers and Sonenshein [38] have been reported that the presence of **Ile** and **Val** increase the affinity of **CodY** for its binding site. In the case where **CodY** remains active, we suppose that it could be able to repress the *ilv-leu* operon via its others binding sites. Indeed, the **Pilv-Leu** promoter contains four binding site for **CodY**, in the promoter-proximal CodY-I+II (nt -84/-32) site is a high-affinity binding site of **CodY**, while the promoter-distal CodY-III (nt-154/-107) and CodY-IV (-185/-168) sites are low-affinity ones [38]. This shows that modification of **CodY** regulation in the cell led to an enhancement of specific surfactin production by increasing the production of leucine via the overexpression of *ilv-leu* operon. This observation have been perfectly predicted by our modeling of the **Pilv-Leu** regulation (**r₁** and decreasing the influx of **CodY**, see Table 5 and also in our modeling of the complete metabolic pathway (**r₁₅**)). Moreover the comparison between the specific surfactin production of the strain BBG254 and the strain BBG251 are consistent regarding the results previously published by Brinsmade et al. [5]. In this study, the authors show that the copies of *ilv-leu* transcript for **CodY** null mutant strain was 80 times more as well as for high-affinity **CodY** binding site it was 28 times higher in TSS medium with 16 amino acids.

The **Pilv-Leu** is also subject to a negative regulation by **TnrA** through its binding to the specific sites. This stimulates DNA binding inhibiting transcription initiation and repressing the *ilv-leu* expression [14]. To avoid this regulation, the prediction suggest to delete the inflow of **TnrA** (**r₁₆** or to suppress its binding on the **Pilv-Leu**: **r₇**). This repression was disrupted by the deletion of *tnrA* in the strain BBG256. In TSS medium with 16 amino acids, the specific surfactin production was increased twice, but this result was not considered as significant reported to that of the control strain. But in Landy medium supplemented with (NH₄)₂SO₄, where the concentration of organic nitrogen was less than in the latter medium, a 4 times increase in specific surfactin production was observed. These results are quite logical since the impact of this regulator is bestowed only in organic nitrogen limited conditions [12].

Intracellular concentration of leucine also downregulated the *ilv-leu* operon using a feedback loop through an attenuator [5, 16]. This 80-bp leucine-dependent stem-loop structure (**Tbox**) has been identified by the predictions as a target to be deleted (**r₄₀**). The effect of its deletion on specific surfactin production of the BBG252 strain was clearly better in TSS medium with 16 amino acids (9 fold increasing) than in the Landy medium supplemented with (NH₄)₂SO₄ (1.5 fold increasing) in comparison to BBG111. Presence of leucine in the TSS medium probably has the effect of inhibiting the expression of *ilv-leu* operon in BBG111, regulation which does not occur in the Landy medium containing only (NH₄)₂SO₄ as nitrogen source. Brinsmade et al. [5] also found that the copies of *ilv-leu* transcript were 19 times higher in **Tbox** mutant (**SRB87**) in comparison to the wild type strain in TSS medium with 16 amino acids.

Most of the predictions newly obtained from the complete leucine network were related to the genes involved in the leucine degradation: *ywaA* via r_{47} , *bkL* via r_{11} and *bcd* via r_{10} , and the mechanism which activates the expression of these two latter genes, i.e., expression of *BkdR* via *Ile* and *Val* activation via r_{12} and r_{13} and its binding on the promoter via r_3 or the constitutive expression of the promoter by r_4 . All these knock-out predictions have same effect of interrupting the two stages of reactions that allow the degradation of branched chain amino acids to make fatty acid precursors (Figure 3). To test all these predictions, the *bcd* and *bkL*² were deleted in the strains **BBG253** and **BBG255**, respectively. Here, we have chosen not to interrupt *ywaA* because it is also involved in leucine production through the reaction r_{45} . However, *YwaA* action in this step might be able to be offset by *YbgE* as these two enzymes are exchangeable and have the same function [42]. In TSS medium with 16 amino acids, an increment of 6 times in the specific surfactin production was observed with **BBG253** strain, but no significant increase was detected with **BBG255** strain. This small change observed is perhaps due to the presence of its homologs in *B. subtilis*. Nevertheless, in Landy medium supplemented with (NH₄)₂SO₄, the effect of knockout on *bcd* and *bkL*² led to the same result for both strains with an enhancement of 3 times in specific surfactin production. Finally, *lpdV* deletion seems to be the right strategy because few effects on growth were observed and the impact on the surfactin production was good.

All these results discussed above confirm the efficiency and the relevance of the modeling language and the abstract interpretation algorithm developed for gene knockout prediction. Indeed regarding the medium conditions all the predicted knockouts led to an increase in surfactin production. There are many questions for future work when it comes to the knockout prediction algorithm. First, they could be extended to the whole metabolism of *B. subtilis*, but before this, it would be very helpful to fully automatize the interpretation of the solutions of the constraint solver. This includes better tools that can automatize the presentation of the results. Second, we would like to point out that the quality of the approximation obtained by abstract interpretation depends on the strength of the simplification of the steady state equations, before applying the abstract interpretation. In order to obtain reasonable guarantees for the quality of these simplifications, the algorithm should be investigated in greater depth. The question is whether it can be formulated as a confluent term rewrite system. Third and most importantly, the present prediction algorithm is purely qualitative, so that it cannot predict anything quantitative on the size of the changes. Finding more accurate quantitative abstract interpretations would be an important improvement. New *in-vivo* metabolic engineering experiments should be performed with the different mutants to determine the missing values of model parameters. This in-depth metabolic study will be implemented in the models to make it consistent. In summary, this work opens numerous perspectives, firstly to improve the models and secondly to apply this original approach to more complex secondary metabolites biosynthesis or more complex metabolic pathways such as may be found in eukaryotic organisms.

5 Acknowledgements

This research received funding from the European Union, Marie Curie ITN AMBER, 317338 and partly funded by the University of Lille through the BQR project *Biologie synthétique pour la synthèse dirigée de peptides microbiens bioactifs*. The authors thank Dr Max Béchet from the ProBioGEM team, Dr Vladimir Bidnenko and Dr Sandrine Auger from INRA Jouy-en-Josas for their contributions in the mutants construction. We are also grateful to the Re-alCat platform of the University Lille that allowed the mutants screening and to Pr Sonenshein from the Tufts University of Boston for donations of the *B. subtilis* SRB87 and SRB94 strains. And most importantly, we thank our masters student Thibault Etienne for helping us with the time consuming experiments with the constraint solver.

References

- [1] E. Almaas, B. Kovacs, T. Vicsek, Z. N. Oltvai, and A. L. Barabasi. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*, 427:839–843, 2004.
- [2] Max Béchet, Joany Castéra-Guy, Jean-Sébastien Guez, Nour-Eddine Chihib, Françoise Coucheney, François Coutte, Patrick Fickers, Valérie Leclère, Bernard Wathelet, and Philippe Jacques. Production of a novel mixture of mycosubtilins by mutants of *Bacillus subtilis*. *Bioresource Technol.*, 145: 264–270, 2013.
- [3] Boris R Belitsky and Abraham L Sonenshein. Contributions of multiple binding sites and effector-independent binding to cody-mediated regulation in *Bacillus subtilis*. *J. Bacteriol.*, 193(2):473–484, 2011.
- [4] Shaun R Brinsmade and Abraham L Sonenshein. Dissecting complex metabolic integration provides direct genetic evidence for cody activation by guanine nucleotides. *J. Bacteriol.*, 193(20):5637–5648, 2011.
- [5] Shaun R Brinsmade, Roelco J Kleijn, Uwe Sauer, and Abraham L Sonenshein. Regulation of cody activity through modulation of intracellular branched-chain amino acid pools. *J. Bacteriol.*, 192(24):6357–6368, 2010.
- [6] Anthony P Burgard, Priti Pharkya, and Costas D Maranas. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.*, 84(6):647–657, 2003.
- [7] Fabian M Commichau, Nico Pietack, and Jörg Stülke. Essential genes in *Bacillus subtilis*: a re-evaluation after ten years. *Mol. Biosyst.*, 9(6):1068–1075, 2013.
- [8] F Coutte, V Leclère, M Béchet, J-S Guez, D Lecouturier, M Chollet-Imbert, P Dhulster, and P Jacques. Effect of pps disruption and constitutive expression of srfa on surfactin productivity, spreading and antagonistic properties of *Bacillus subtilis* 168 derivatives. *J. Appl. Microbiol.*, 109(2):480–491, 2010.
- [9] Markus W Covert and Bernhard O Palsson. Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J. Theor. Biol.*, 221(3):309–325, 2003.
- [10] Michel Debarbouille, Rozenn Gardan, Maryvonne Arnaud, and George Rapoport. Role of bkdr, a transcriptional activator of the sigI-dependent isoleucine and valine degradation pathway in *Bacillus subtilis*. *J. Bacteriol.*, 181(7):2059–2066, 1999.
- [11] Erwin Hans Duitman. *Nonribosomal peptide synthesis in Bacillus subtilis*. PhD thesis, University Library Groningen, 2003.
- [12] Susan H Fisher. Regulation of nitrogen metabolism in *Bacillus subtilis*: vive la difference! *Mol. Microbiol.*, 32(2):223–232, 1999.
- [13] Lope A Flórez, Katrin Gunka, Rafael Polanía, Stefan Tholen, and Jörg Stülke. Spabbats: A pathway-discovery method based on boolean satisfi-

- ability that facilitates the characterization of suppressor mutants. *BMC Syst. Biol.*, 5(1):5, 2011.
- [14] Yasutaro Fujita, Takenori Satomura, Shigeo Tojo, and Kazutake Hirooka. Ccpa-mediated catabolite activation of the *Bacillus subtilis* *ilv-leu* operon and its negation by either cody-or tnra-mediated negative regulation. *J. Bacteriol.*, 196(21):3793–3806, 2014.
- [15] Anne Goelzer, Fadia Bekkal Brikci, Isabelle Martin-Verstraete, Philippe Noirot, Philippe Bessières, Stéphane Aymerich, and Vincent Fromion. Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst. Biol.*, 2(1):20, 2008.
- [16] J. A. Grandoni, S. A. Zahler, and J. M. Calvo. Transcriptional regulation of the *ilv-leu* operon of *Bacillus subtilis*. *J. Bacteriol.*, 174(10):3212–3219, May 1992.
- [17] Evamaria Gruchattka, Oliver Hädicke, Steffen Klamt, Verena Schütz, and Oliver Kayser. In silico profiling of escherichia coli and saccharomyces cerevisiae as terpenoid factories. *Microbial cell factories*, 12(1):1–18, 2013.
- [18] G. W. Hatfield and H. E. Umbarger. Threonine deaminase from *Bacillus subtilis*. II. The steady state kinetic properties. *J. Biol. Chem.*, 245(7):1742–1747, April 1970.
- [19] Philippe Jacques. Surfactin and other lipopeptides from *Bacillus* spp. In *Biosurfactants*, pages 57–91. Springer, 2011.
- [20] Paula Jauregi, François Coutte, Lucie Catiau, Didier Lecouturier, and Philippe Jacques. Micelle size characterization of lipopeptides produced by *B. subtilis* and their recovery by the two-step ultrafiltration process. *Sep. Purif. Technol.*, 104:175–182, 2013.
- [21] Mathias John, Mirabelle Nebut, and Joachim Niehren. Knockout prediction for reaction networks with partial kinetic information. In *Verification, Model Checking, and Abstract Interpretation*, pages 355–374. Springer, 2013.
- [22] Christian Jungreuthmayer and Jürgen Zanghellini. Designing optimal cell factories: integer programming couples elementary mode analysis with regulation. *BMC systems biology*, 6(1):103, 2012.
- [23] TOSHI Kaneda. Iso-and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol. Rev.*, 55(2):288–302, 1991.
- [24] Joonhoon Kim and Jennifer L Reed. Optorf: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst. Biol.*, 4(1):53, 2010.
- [25] M Landy, GH Warren, SB RosenmanM, and LG Colio. Bacillomycin an antibiotic from *Bacillus subtilis* active against pathogenic fungi. *Exp. Biol. Med.*, 67(4):539–541, 1948.
- [26] Jin-Feng Liu, Juan Yang, Shi-Zhong Yang, Ru-Qiang Ye, and Bo-Zhong Mu. Effects of different amino acids in culture media on surfactin variants produced by *Bacillus subtilis* td7. *Appl. Biochem. Biotech.*, 166(8):2091–2100, 2012.
- [27] Ulrike Mäder, Susanne Hennig, Michael Hecker, and Georg Homuth. Transcriptional organization and posttranscriptional regulation of the *Bacil-*

- lus subtilis* branched-chain amino acid biosynthesis genes. *J. Bacteriol.*, 186(8):2240–2252, April 2004.
- [28] Ulrike Mäder, Arne G. Schmeisky, Lope A. Flórez, and Jörg Stülke. SubtiWiki—a comprehensive community resource for the model organism *Bacillus subtilis*. *Nucleic Acids Res.*, 40:1278–1287, January 2012.
- [29] Virginie Molle, Yoshiko Nakaura, Robert P. Shivers, Hirotake Yamaguchi, Richard Losick, Yasutaro Fujita, and Abraham L. Sonenshein. Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J. Bacteriol.*, 185(6):1911–1922, March 2003.
- [30] Pierre Nicolas, Ulrike Mäder, Etienne Dervyn, Tatiana Rochat, Aurélie Leduc, Nathalie Pigeonneau, Elena Bidnenko, Elodie Marchadier, Mark Hoebeke, Stéphane Aymerich, et al. Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science*, 335(6072):1103–1106, 2012.
- [31] Joachim Niehren, Mathias John, Cristian Versari, Francois Coutte, and Philippe Jacques. Qualitative reasoning for reaction networks with partial kinetic information. In *Computational Methods in Systems Biology, International Conference (CMSB)*, 2015. To appear.
- [32] José Manuel Otero and Jens Nielsen. Industrial systems biology. *Industrial Biotechnology: Sustainable Growth and Economic Success*, 2010.
- [33] Jason A Papin, Joerg Stelling, Nathan D Price, Steffen Klamt, Stefan Schuster, and Bernhard O Palsson. Comparison of network-based pathway analysis methods. *Trends in biotechnology*, 22(8):400–405, 2004.
- [34] Nathan D Price, Jennifer L Reed, and Bernhard Ø Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.*, 2(11):886–897, 2004.
- [35] Sridhar Ranganathan, Patrick F Suthers, and Costas D Maranas. Optforce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.*, 6(4):e1000744, 2010.
- [36] Manoja Ratnayake-Lecamwasam, Pascale Serror, Ka-Wing Wong, and Abraham L Sonenshein. *Bacillus subtilis* cody represses early-stationary-phase genes by sensing gtp levels. *Gene Dev.*, 15(9):1093–1103, 2001.
- [37] J Russel Sambrook and DW Russel. D. w (2001). *Molecular Cloning: A Laboratory Manual*, 2001.
- [38] Robert P. Shivers and Abraham L. Sonenshein. Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids. *Mol. Microbiol.*, 53(2):599–611, July 2004.
- [39] Seung Bum Sohn, Tae Yong Kim, Jong Myoung Park, and Sang Yup Lee. In silico genome-scale metabolic analysis of *Pseudomonas putida* kt2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival. *Biotechnol. J.*, 5(7):739–750, 2010.
- [40] Huigang Sun, Xiaomei Bie, Fengxia Lu, Yaping Lu, Yundailai Wu, and Zhaoxin Lu. Enhancement of surfactin production of *Bacillus subtilis* fimb by replacement of the native promoter with the *Pspac* promoter. *Canad. J. Microbiol.*, 55(8):1003–1006, 2009.

- [41] Kosei Tanaka, Christopher S Henry, Jenifer F Zinner, Edmond Jolivet, Matthew P Cohoon, Fangfang Xia, Vladimir Bidnenko, S Dusko Ehrlich, Rick L Stevens, and Philippe Noirot. Building the repertoire of dispensable chromosome regions in *Bacillus subtilis* entails major refinement of cognate large-scale metabolic model. *Nucleic Acids Res.*, pages 1–13, 2012.
- [42] Helena B. Thomaidis, Ella J. Davison, Lisa Burston, Hazel Johnson, David R. Brown, Alison C. Hunt, Jeffery Errington, and Lloyd Czaplewski. Essential bacterial functions encoded by gene pairs. *J. Bacteriol.*, 189(2): 591–602, January 2007.
- [43] Shigeo Tojo, Takenori Satomura, Kaori Morisaki, Josef Deutscher, Kazutake Hirooka, and Yasutaro Fujita. Elaborate transcription regulation of the *Bacillus subtilis ilv-leu* operon involved in the biosynthesis of branched-chain amino acids through global regulators of CcpA, CodY and ThrA. *Mol. Microbiol.*, 56(6):1560–1573, June 2005.
- [44] Anuradha C. Villapakkam, Luke D. Handke, Boris R. Belitsky, Vladimir M. Levdikov, Anthony J. Wilkinson, and Abraham L. Sonenshein. Genetic and Biochemical Analysis of the Interaction of *Bacillus subtilis* CodY with Branched-Chain Amino Acids. *J. Bacteriol.*, 191(22):6865–6876, November 2009.
- [45] W. Wiechert. ¹³C metabolic flux analysis. *Metabolic engineering*, 3(3): 195–206, July 2001. ISSN 1096-7176. doi: 10.1006/mben.2001.0187. URL <http://dx.doi.org/10.1006/mben.2001.0187>.

6 Supplementary materials

Table 10: PCR primers used for gene deletion.

Gene	Primer	DNA-Sequence
<i>bcd</i>	forward ₁	5'-ACAGCACCCCTTAAGAGCTGGC-3'
	reverse ₁	5'-CGACCTGCAGGCATGCAAGCTATCCAGATTG TTCATCCTGGC-3'
	forward ₂	5'-GCTCGAATTCACTGGCCGTCGCCAGGATGAA CAATCTGGATAACGGCCACAGTGTATTAAGC-3'
	reverse ₂	5'-AGCACAATCGCTTCAACTTCGC-3'
<i>codY</i>	forward ₁	5'-GAGGCAATTACGCTTTGGCAG-3'
	reverse ₁	5'-GCTCGAATTCACTGGCCGTCGAGAAAATC TAAAATCTCATTAA-3'
	forward ₂	5'-CGACCTGCAGGCATGCAAGCTTAATGAGATTTT AGATTTTCTGATAAATAATCCTCCTAAACATTC CTCGCTCGAATTCACTGGCCGTCGAGAAAATCT-3'
	reverse ₂	5'-GCTGCAGTTAGAGAGATGCTAG-3'
<i>bkL</i> ²	forward ₁	5'-AGAGCTGCAGCGATTTGACCG-3'
	reverse ₁	5'-CGACCTGCAGGCATGCAAGCTGACTACGTCA TACTCAGTTGC-3'
	forward ₂	5'-GCTCGAATTCACTGGCCGTCGCAACTGAGTAT GACGTAGTCATTTACCCGCATCCAACGC-3'
	reverse ₂	5'-CTATCTCATCGGACAGCAGGC-3'
<i>tnrA</i>	forward ₁	5'-CAGTACAGCAAATTCAGTGG-3'
	reverse ₁	5'-CGACCTGCAGGCATGCAAGCTCGGACTTTTA TTATTTAACGGTCATTTTCCACCCCTGGATG-3'
	forward ₂	5'-GCTCGAATTCACTGGCCGTCGTAAATAATAA AAGTCCGGC-3'
	reverse ₂	5'-AGCATTACACGGTAAAAGACG-3'

² Whenever we talk about a knockout for the *bkL* operon, we mean more precisely the *lpdV* gene of this operon.

Table 11: Molecules of leucine network.

Role	Short name	Chemical Species
Metabolites	Ile	Isoleucine
	Leu	Leucine
	Val	Valine
	Akb	L-2-amino-acetoacetate
	Glu	Glutamate
	OxoGlu	Oxogluterate
	Gtp	Guanosine triphosphate
	Ket _a	2-keto-3-methylvalerate
	Acyl-CoA	Acyl Coenzyme A
	Ket _b	2-keto-isovalerate
	Ket _c	2-keto-isocaproate
	Pyr	Pyruvate
	Thr	Threonine deshydratase
	Proteines	Bcd
BkL		2-oxoisovalerate dehydrogenase
BkdR		Transcriptional activator of BkL
CcpA		Carbon catabolite control protein A
CodY		Transcriptional pleiotropic regulator
IlvA		Threonine deshydratase
IlvBH		Acetolactate synthase
IlvC		Ketol-acid reductoisomerase
IlvD		Dihydroxy-acid dehydratase
LeuA		2-isopropylmalate synthase
LeuBCD		3-isopropylmalate dehydratase
TnrA		Nitrogen pleiotropic transcriptional regulator
YbgE		Branched chain amino-acid aminotransferase
YwaA		branched chain amino-acid aminotransferase
Actors	Pilv-Leu	Activity of promoter of IlvBH IlvC LeuA LeuBCD operon
	BS _{CodY}	Activity of CodY binding to promoter Pilv-Leu
	BS _{TnrA}	Activity of TnrA binding to promoter Pilv-Leu
	BS _{CcpA}	Activity of CcpA binding to promoter Pilv-Leu without BS _{TnrA} loop
	OP _{BkL-Bcd}	Activity of promoter of BkL Bcd operon
	YwaA+YbgE	Joint activity of YbgE and YwaA
Tbox	Activity of tryptophan attenuation	

Table 12: Reactions of leucine network.

Name	Function
r ₁	bind CodY to Pilv–Leu for inhibition
r ₂	activate Pilv–Leu promoter
r ₃	bind BkdR to BkL Bcd promoter
r ₄	constitutive expression of BkL Bcd operon
r ₅	express YbgE
r ₆	express YwaA
r ₇	bind TnrA to Pilv–Leu promoter for inhibition
r ₉	bind CcpA to Pilv–Leu promoter without BS_{TnrA} loop
r ₁₀	expression of Bcd , activated by OP_{BkL–Bcd}
r ₁₁	expression of BkL , activated by OP_{BkL–Bcd}
r ₁₂	activate BkdR by Ile
r ₁₃	activate BkdR by Val
r ₁₄	expression of CcpA
r ₁₅	express and accelerate CodY (to be explained acceleration)
r ₁₆	expression of TnrA
r ₁₇	Gtp degradation
r ₁₈	BkdR degradation
r ₁₉	YwaA+YbgE degradation
r ₂₀	disabeling of IlvA by Ile
r ₂₁	disabeling of IlvA by Val
r ₂₂	expression of IlvA inhibited by binding of CodY
r ₂₃	expression of IlvBH
r ₂₄	expression of IlvC
r ₂₅	expression of IlvD , inhibited by binding of CodY to promotor
r ₂₆	metabolic transformation of Ket_a to Ile activated by YwaA+YbgE , after an amino addition taken from Glu , which becomes OxoGlu .
r ₂₇	metabolic transformation of Akb and Pyr into Ket_a
r ₂₈	prepare outflow of Ket_a activated by BkL
r ₂₉	degradation of Ile into Ket_a activated by YwaA and Bcd , with an amino transfer from OxoGlu which becomes Glu .
r ₃₀	metabolic transformation from Pyr to Ket_b , activated by IlvD
r ₃₁	prepare outflow of Ket_b activated by BkL
r ₃₂	degradation of Val into Ket_b activated by YwaA and Bcd with an amino transfer from OxoGlu which becomes Glu .
r ₃₃	metabolic transformation of Ket_b into Ket_c , activated by LeuBCD and LeuA
r ₃₄	prepare outflow of Ket_c activated by BkL
r ₃₅	degradation of Leu into Ket_c ctivated by YwaA and Bcd with an amino transfer from OxoGlu which becomes Glu .
r ₃₆	metabolic transformation of Ket_b to Val activated by YwaA+YbgE , after an amino addition from Glu which becomes OxoGlu
r ₃₇	expression of LeuA
r ₃₈	expression of LeuBCD
r ₃₉	deactivation of LeuA by Leu
r ₄₀	Leucine attenuation
r ₄₁	metabolic transformation of Thr into Akb using IlvA
r ₄₅	metabolic transformation of Ket_c to Leu activated by YwaA+YbgE after an amino addition from Glu , which becomes OxoGlu
r ₄₆	expression of YbgE , inhibited by binding of CodY to promoter
r ₄₇	expression of YwaA