



RESEARCH ARTICLE

Open Access

# Capturing the response of *Clostridium acetobutylicum* to chemical stressors using a regulated genome-scale metabolic model

Satyakam Dash<sup>1</sup>, Thomas J Mueller<sup>1</sup>, Keerthi P Venkataramanan<sup>2,3</sup>, Eleftherios T Papoutsakis<sup>2,3</sup> and Costas D Maranas<sup>1\*</sup>

## Abstract

**Background:** Clostridia are anaerobic Gram-positive Firmicutes containing broad and flexible systems for substrate utilization, which have been used successfully to produce a range of industrial compounds. In particular, *Clostridium acetobutylicum* has been used to produce butanol on an industrial scale through acetone-butanol-ethanol (ABE) fermentation. A genome-scale metabolic (GSM) model is a powerful tool for understanding the metabolic capacities of an organism and developing metabolic engineering strategies for strain development. The integration of stress-related specific transcriptomics information with the GSM model provides opportunities for elucidating the focal points of regulation.

**Results:** We describe here the construction and validation of a GSM model for *C. acetobutylicum* ATCC 824, *iCac802*. *iCac802* spans 802 genes and includes 1,137 metabolites and 1,462 reactions, along with gene-protein-reaction associations. Both <sup>13</sup>C-MFA and gene deletion data in the ABE fermentation pathway were used to test the predicted flux ranges allowed by the model. We also describe the CoreReg method, introduced in this paper, to integrate transcriptomic data and identify core sets of reactions that, when their flux was selectively restricted, reproduced flux and biomass-formation ranges seen under all regulatory constraints. CoreReg was used in response to butanol and butyrate stress to tighten bounds for 50 reactions within the *iCac802* model. These bounds affected the flux of tens of reactions in core metabolism. The model, incorporating the regulatory restrictions from CoreReg under chemical stress, exhibited an approximate 70% reduction in biomass yield for most stress conditions.

**Conclusions:** The regulation placed on the model for the two stresses using CoreReg identified differences in the respective responses, including distinct core sets and the restriction of biomass production similar to experimental observations. Given the core sets predicted by the CoreReg method, remedial actions can be taken to counteract the effect of stress on metabolism. For less well-known systems, plausible regulatory loops can be suggested around the affected metabolic reactions, and the hypotheses can be tested experimentally.

**Keywords:** *Clostridium acetobutylicum*, CoreReg, Regulation, Genome-scale metabolic model

\* Correspondence: costas@psu.edu

<sup>1</sup>Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania, USA

Full list of author information is available at the end of the article

## Background

The organisms of the genus and class *Clostridium*, anaerobic Gram-positive Firmicutes, contain broad and flexible systems for substrate utilization [1]. Their inherent ability to use simple and complex carbohydrates, gases, and many other chemicals as substances to produce a wide range of products, such as carboxylic acids and various alcohols, underscores their unique potential as platform organisms for the production of chemicals and fuels [2]. In particular, *Clostridium acetobutylicum* has been the model organism for the production of butanol on an industrial scale through the acetone-butanol-ethanol (ABE) fermentation [1].

ABE fermentation is biphasic in nature; the acidogenic, exponential growth phase is characterized by the production of butyric and acetic acids, while the solventogenic stationary phase is characterized by the production of the ABE solvents. Production of acids and the resulting drop in the culture pH during the acidogenic phase drives the transition towards solventogenesis [1,2]. These metabolites, notably butyric acid and butanol, are toxic to the cells that produce them and affect their ability to function and eventually to survive. While several studies have been carried out to understand the changes during stress at various levels such as transcription [3-7] and translation [8], the impact of stress remains poorly understood at the systems levels in the context of the detailed cellular metabolism.

An important asset for understanding the metabolic capacity of an organism and deciding on metabolic engineering interventions is a genome-scale metabolic (GSM) model [9]. These models are network representations of the metabolic repertoire of an organism and are derived from genome-annotation information, metabolomic/fluxomic data, and biochemical characterizations. Advanced GSM models account for reaction stoichiometry and directionality, gene to protein to reaction (GPR) associations, reaction localization, transporter information, and biomass composition. They form a structured, multilayered framework for the integration and interpretation of experimental data and computational studies. These models computationally can direct engineering interventions in microbial strains for targeted overproduction of chemicals [10-13] and for elucidating the organizing principles of metabolism [14-17].

One of the earliest metabolic reconstructions was, in fact, a model of *C. acetobutylicum* [18]. A small stoichiometric model including core glycolytic, acidogenic, and solventogenic pathways was later generated [19]. These early models were used to examine how *C. acetobutylicum* produces butanol and byproducts such as acetate and butyrate. More recently, two GSM models of *C. acetobutylicum* ATCC 824 have been developed [20,21]. These models contain approximately 450 genes (that is, one-sixth of the number

of genes coded on its genome). The Senger and Papoutsakis model [21] has recently been updated to include 242 additional reactions and contains a total of 490 genes along with thermodynamic constraints on the reversibility of reactions [22]. A larger, automatically generated model containing reactions associated with nearly 1,000 genes was constructed as part of the Model SEED effort [23]. However, all these models include only metabolic pathways without any information regarding metabolic changes in response to stressors. It is important to note that the activity and directionality of metabolic pathways under different conditions continue to be unraveled for *C. acetobutylicum*. The tricarboxylic acid cycle (TCA cycle), known to operate in a non-cyclic bifurcated manner, was recently shown to use *Re*-citrate synthase to produce  $\alpha$ -ketoglutarate via citrate [24]. More recently, it has been shown based on  $^{13}\text{C}$ -metabolic flux analysis ( $^{13}\text{C}$ -MFA) data that both  $\alpha$ -ketoglutarate dehydrogenase ( $\alpha$ -KGDH) and succinate dehydrogenase (SDH) are inactive during the acidogenic phase [25]. In contrast, the reaction that converts succinate to succinyl-CoA can carry flux in both directions [25]. While GSM models alone are quite useful for determining the metabolic potential of an organism, determination of the metabolic phenotype under various stress conditions requires the incorporation of additional information, such as transcriptomic data, which for now at least, are the most comprehensive, and genomically complete sets of genomic data that can be acquired.

A number of approaches have been proposed to incorporate regulatory information into GSM models. Regulatory flux balance analysis (rFBA) introduces Boolean constraints for gene expression into flux balance analysis (FBA) by linking the regulators to their targets in an iterative fashion [26,27]. The approach termed steady-state regulatory flux balance analysis (SR-FBA) combines the regulatory and metabolic models and solves the problem as a mixed-integer linear program [28]. GeneForce identifies incorrect regulatory rules and GPR associations in integrated metabolic and regulatory models [29]. PROM uses a probabilistic description of gene states and gene-transcription factor interactions while integrating heterogeneous high-throughput data [30]. The GIM3E method penalizes the flux for reactions whose associated genes have low expression levels in the transcriptome [31]. The recently proposed MTA method identifies minimal transformation rules from one metabolic state to another based on transcriptomic data [32] as in OptForce [33]. E-Flux modifies the maximum and minimum flux bounds of reactions as a function of the associated gene expression values [34]. All the aforementioned methods attempt to throttle back the flux in reactions associated with genes that are differentially expressed at a lower level. They differ in the use of penalty terms or bound contractions, threshold values for

down-regulation, and the use of the parsimony criterion. CoreReg is fundamentally different, as it aims to explain the observed flux redirections as the consequence of a bound contraction of a small set of reactions (the core set). A hierarchy of core sets is identified (primary, secondary, tertiary, and so forth) by eliminating from consideration the dominant focal point of regulation and looking for additional modalities. This is analogous to the FORCE sets in the OptForce procedure [33]. For each one of the stress conditions we identify the minimal number of reaction fluxes (core set) whose change is sufficient to reproduce the flux ranges seen in the model when all regulatory constraints are imposed. The regulatory effect by the core set is propagated through stoichiometry throughout the model, recapitulating the experimentally observed changes. The method is described in detail in the Methods section.

In this paper, we describe the construction of a second generation genome-scale reconstruction of *C. acetobutylicum* ATCC 824, *iCac802*, validation with experimental data. New reactions and pathways absent in earlier models include an updated TCA cycle, a completed fatty acid synthesis pathway, and additions to the purine, pyrimidine, and cobalamin biosynthetic pathways. The *iCac802* model along with the corresponding GPRs and metabolite information is available as SBML and excel files in Additional files 1 and 2 respectively. We also describe the use of the CoreReg method to integrate gene expression data into *iCac802* and predict nexus points of regulation, that underlie cellular response to the physiological stressors butanol and butyrate [3].

## Results

### Model comparisons

The GSM model *iCac802* for *C. acetobutylicum* ATCC 824 spans 802 genes and includes 1,137 metabolites participating in 1,462 reactions. All reactions present are elementally and charge balanced. GPR associations were determined from the available functional annotation information and homology predictions accounting for monofunctional proteins, multifunctional proteins, isozymes, and protein complexes. The model was curated to remove any thermodynamically infeasible cycles, as detailed in the Methods section. The *iCac802* model statistics and those of all other published models for *C. acetobutylicum* are shown in Table 1. *iCac802* has 64% more genes and 84% more reactions than the McAnulty *et al.* model [22]. *iCac802* contains a citrate synthase leading to a partial and bifurcated TCA cycle (based on the findings by Au *et al.* [25]), which is absent in the GSM by Lee *et al.* [20]. The latter model also does not predict the change from acidogenic phase to solventogenic phase under CO gassing conditions due to lack of internal protons [35] as reaction participants. This

**Table 1 Genome-scale model comparison**

Model statistics	Lee <i>et al.</i> [20]	Senger <i>et al.</i> [21]	McAnulty <i>et al.</i> [22]	<i>iCac802</i>
Genes	432	473	490	802
Reactions	502	522	794	1,462
Metabolites	479	422	707	1,137

The number of genes, reactions, and metabolites present in three previous genome-scale models of *C. acetobutylicum* and *iCac802*.

change is correctly predicted by *iCac802* as described in the model testing section. In addition, the GSM model by Lee *et al.* suggests that  $\Delta adc$  is lethal for cell growth due to coupling of succinate production with acetoacetyl-CoA production, contrary to experimental observations [36] and *iCac802* predictions. While all previous models contained an aggregate reaction for the production of hexadecanoyl-*acp* and hexadecanoyl-CoA from acetyl-*acp* and crotonyl-CoA, respectively, *iCac802* includes all participating reactions in fatty acid synthesis and metabolism pathways building up to these metabolites. *iCac802* also contains additional reactions from purine, pyrimidine metabolism, and cobalamin biosynthesis pathways (Additional file 3).

### Model testing

The model was extensively tested to ascertain that it is capable of replicating flux ranges and phenotypes that have been documented for the wild-type (WT) organism and its mutants. The model predicted flux ranges were compared with experimental flux values from  $^{13}\text{C}$ -metabolic flux analysis ( $^{13}\text{C}$ -MFA) [25]. The  $^{13}\text{C}$ -MFA data revealed that four reactions (pyruvate carboxylase (PC), fumarate hydratase (FH), pyridoxal phosphate synthase (PLPS), and alanine-glyoxylate (AGT)), which were originally removed to eliminate thermodynamically infeasible cycles, carried flux in the organism, and therefore, they were reinserted in the model. The cycles were instead eliminated by removing three alternate reactions (malate synthase (MS), succinate dehydrogenase (SDH), and malate dehydrogenase (MDH)) and by modifying the directionality of two others: succinyl-coenzyme A synthetase (SCS) was made reversible, and aspartate ammonia-lyase (ASPA) was restricted to the production of fumarate from aspartate. Figure 1C shows one of these cycles. ASPA was initially removed to fix the cycle due to a lack of literature evidence (Figure 1D), however subsequently MFA results indicated that this reaction carried flux whereas the MDH did not. Figure 1E shows how the addition of ASPA (directionally restricted) and removal of MDH avoids the formation of thermodynamically infeasible cycles while agreeing with experimental data.

After these changes in the model, flux variability analysis (FVA) was performed on the core carbon metabolism reactions, and the flux ranges were compared to the values

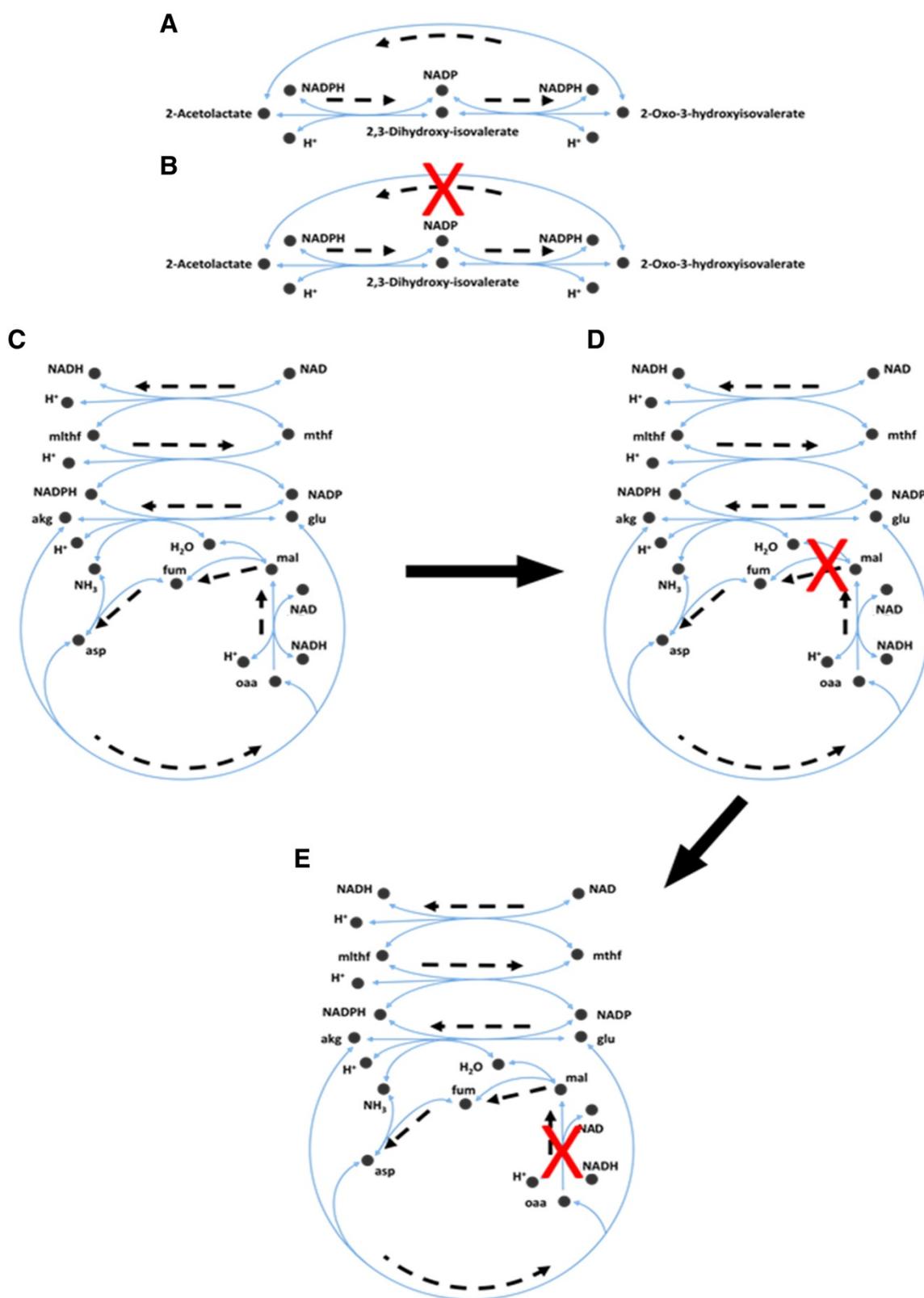


Figure 1 (See legend on next page.)

(See figure on previous page.)

**Figure 1 Examples of thermodynamically infeasible cycles and their resolution.** Dashed lines indicate the direction of flux through the cycle forming reactions. **A)** A cycle containing three reactions. **B)** The fixed cycle from 1A after the removal of a single reaction with minimal literature evidence. **C)** A cycle containing seven reactions. **D)** The original cycle correction for 1C, involving the removal of a reaction with minimal literature evidence. **E)** The final cycle correction upon examination of  $^{13}\text{C}$ -MFA data [25].

obtained by the  $^{13}\text{C}$ -MFA analysis [25]. These experiments were carried out with a chemically defined medium that results in slower growth and lower biomass yields. First, all fluxes were normalized for a glucose uptake of  $10 \text{ mmol gDW}^{-1} \text{ h}^{-1}$ . FVA was performed while constraining the growth rate to the WT value of  $0.32 \text{ h}^{-1}$  for *C. acetobutylicum* grown in complex media [37]. The comparison showed that the flux ranges of only four reactions (catalyzed by enolase (ENO), hexokinase (HK), pyruvate kinase (PYK), phosphotransacetylase (PTA), and phosphofructokinase (PFK)) encompassed the reported experimental values, as shown by Figure 2A. The reason for this is that *C. acetobutylicum* was grown in defined media by Au et al. [25], exhibiting significantly slower growth than in complex media [37]. In addition, the  $^{13}\text{C}$ -MFA data [25] was collected during the late growth phase with small amounts of solvents being produced, resulting in a reduced growth rate. Matching the FVA results with MFA data, we identified a growth rate value of  $0.07 \text{ h}^{-1}$ . Upon reapplying FVA with the biomass yield constrained to  $0.07 \text{ h}^{-1}$  (see Figure 2B), all reactions except for HK and PC had flux ranges that encompassed experimental values. The two reaction experimental flux values differed from the model predicted range by only a value of  $0.02 \text{ mmol gDW}^{-1} \text{ h}^{-1}$ . It can be observed that, under these slow growing conditions, the TCA cycle reactions carry less flux and lie near the lowest end of the predicted flux range in Figure 2. The remaining flux is directed towards production of acids and solvents through pyruvate. This causes the flux of glycolytic reactions to lie near the high end of the predicted flux ranges (as shown by FVA predictions in Figure 2).

Following the model updates and comparisons with  $^{13}\text{C}$ -MFA data, the model's responses to gene knockouts and varying environmental conditions were also tested. The model was used to analyze the effect of increasing the size of the NADH pool on the production of various acids and solvents. It has been shown experimentally that an increase in the level of NADH leads to a concomitant increase in butyrate, solvents, and hydrogen production (Figure 3) [38]. Allowing for the free conversion between NAD and NADH resulted in an increase in their production with the exception of acetate, whose production was, as expected, found to be independent of reducing equivalent availability.

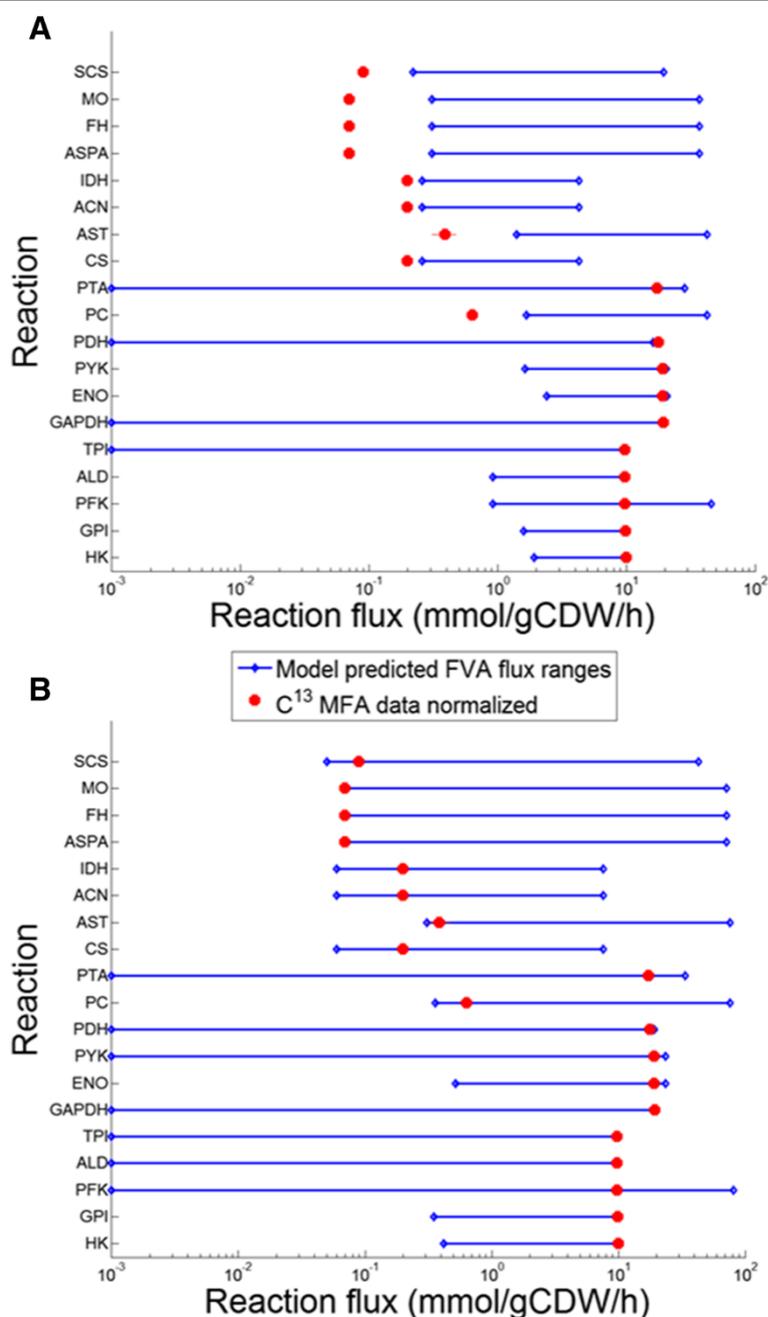
The model was also queried with respect to the ability to co-utilize glycerol. Glycerol as a highly reduced carbon

source (its degree of reduction is 4.67 compared to 4.0 for glucose) allows for the generation of more reducing equivalents which drive the production of butyrate and alcohols (that is, butanol and ethanol). While *C. acetobutylicum* does not have the inherent ability to grow on glycerol as the sole carbon source, co-utilization of glycerol with glucose has been shown to result in a largely homo-butanol fermentation (that is, a fermentation where butanol is the predominant solvent produced) in *C. acetobutylicum* [39]. It is interesting to note that the glycerol uptake (CAC1319) and utilization (CAC1322) genes have been found to be up-regulated under butanol stress [3,5]. Based on this information, a glycerol uptake reaction was added to *iCac802* in order to test the impact of glycerol as a carbon source. The increased availability of reducing equivalents showed a similar affect, as having no redox constraint in the model by allowing for free interconversion between  $\text{NAD}^+$  and  $\text{NADH}$  or  $\text{NADP}^+$  and  $\text{NADPH}$  resulted in an increase in butyrate, solvents, and hydrogen production, as seen in Table 2.

The model was further tested by showing that it can predict results from experiments examining the impact of  $\text{CO}$  gassing on product formation and cell growth [35].  $\text{CO}$  gassing affects the cellular metabolism of *C. acetobutylicum* by forcing the transition from acidogenic to solventogenic fermentation (that is, initiating the uptake of butyrate and leading to the production of butanol and ethanol).  $\text{CO}$  inhibits the hydrogenase arresting  $\text{H}_2$  production (Figure 3) [35]. Therefore, the hydrogenase reaction flux was set equal to zero in the presence of  $\text{CO}$ . Since the organism has been shown to uptake butyrate during the  $\text{CO}$  sparging period [35], butyrate was supplied as an additional nutrient for the model. Using these additional constraints, the model predicted alcohol production (Table 3) during the acidogenic phase in accordance with experimental findings [35].

Experimental data from fermentations using cell recycle were also examined using the GSM model. Cell-recycle conditions result in limited ammonia and phosphate uptake by the cells and an increase in overall alcohol production along with a reduction in biomass yield [40,41]. These conditions were simulated by restricting the flux bounds of ammonia and phosphate uptake reactions to an assumed 80% of their maximum allowable ranges determined by FVA [40,41]. The model showed a reduction in biomass yield and an increase in solvents yield, as shown in Table 4.

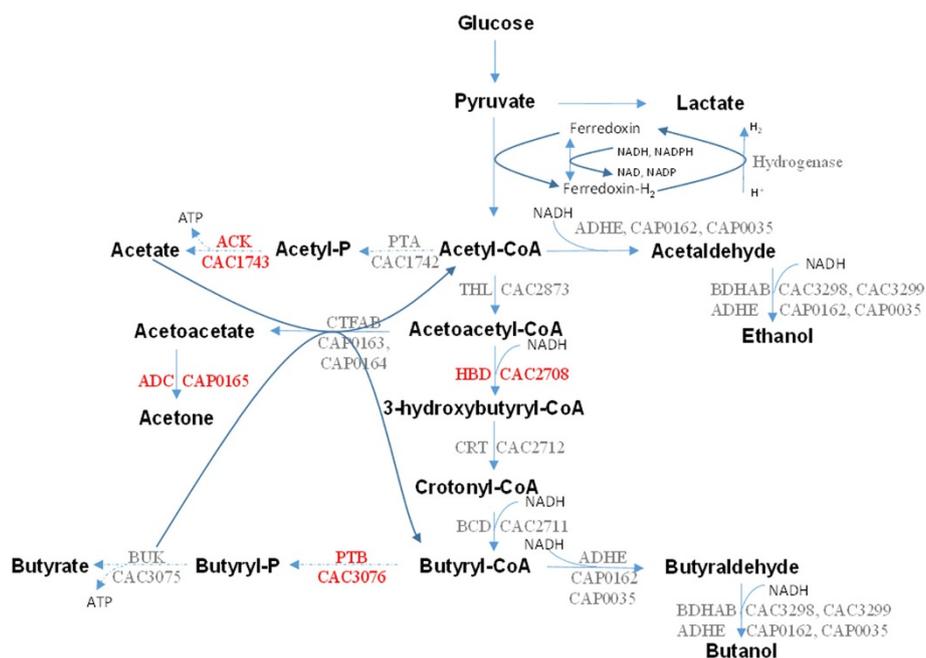
Further model testing was performed by comparing experimental data of solvent yields for a number of *C. acetobutylicum* mutants [42] with *in silico* results.



**Figure 2** Comparison of *in silico* and experimentally measured  $^{13}C$ -MFA flux ranges for *C. acetobutylicum* [25]. **A**) under wild-type biomass constraint ( $0.32 \text{ mmol gDW}^{-1} \text{ h}^{-1}$ , grown in complex media) [37]. **B**) under reduced biomass constraint ( $0.07 \text{ mmol gDW}^{-1} \text{ h}^{-1}$ , grown in defined media) given that the data were also collected during the transition to the solventogenic phase [25]. Hexokinase (HK) and pyruvate carboxylase (PC) had their experimental values outside the model predicted ranges under the reduced biomass constraint. FVA was performed with a glucose uptake rate of  $10 \text{ mmol gDW}^{-1} \text{ h}^{-1}$ . Full reaction names can be found in the list of abbreviations.

Mutants involving gene deletions affecting acid and solvent production in the ABE pathway were used to test *iCac802*. Biomass was constrained to the reported growth rate values for the respective experiments. Reaction fluxes associated with a deleted gene were set to zero. FVA was performed to determine the possible range of solvent production. FVA

was first performed with biomass constrained to the reported growth rate to evaluate the flux ranges for the produced acids and solvents. The identified flux ranges of solventogenic nutrients (glucose, acetate, butyrate, and carbon dioxide) were subsequently calculated by fixing both the growth rate and restricting the acids/solvents to



**Figure 3** The butyrate (butanoate) metabolism in *C. acetobutylicum* summarizing the formation of acids (acetic and butyric acid) and ABE solvents. The acid formation pathways are represented by dotted lines. The mutants that were used to validate the GSM model are represented in red. (ACK - acetate kinase; PTA - phosphotransacetylase; ADHE - alcohol/aldehyde dehydrogenase; THL - thiolase; ADC - acetoacetate decarboxylase; CTFAB - CoA-transferase; HBD - hydroxybutyryl-CoA dehydrogenase; CRT - crotonase; BCD - butyryl-CoA dehydrogenase; BDHAB - butanol dehydrogenase; PTB - phosphotransbutyrylase; BUK - butyrate kinase).

the FVA calculated maximum and minimum values. Yield ranges were determined by evaluating the ratio of acids/solvents to the corresponding minimum nutrient flux. Mutants  $\Delta ack$  and  $\Delta ptb$  reduce (but do not eliminate) acetate and butyrate production by removing acetate kinase (ACK) and phosphotransbutyrylase (PTB) activities, respectively (Figure 3) [37,43]. For the two mutant strains, as well as for the WT strain, the model predicts a broad

**Table 2** Reducing equivalent dependence analysis of various acids, solvents, and hydrogen

Metabolites	Glucose		Glycerol
	With redox constraint	No redox constraint	
Carbon source	10.00	10.00	20.00
Acetate	23.74	23.74	25.16
Butyrate	0.96	8.59	10.79
Ethanol	12.17	17.68	16.32
Acetone	7.34	17.01	17.04
Butanol	5.75	9.65	12.65
Hydrogen	31.18	35.85	61.73

Comparisons made between glucose with and without constraints on production of reducing equivalents and glycerol. Equivalent carbon flux values were chosen for both glucose and glycerol. All values except for carbon source uptake represent production fluxes with units of  $\text{mmol gDW}^{-1} \text{h}^{-1}$ . Increasing availability of reducing equivalents led to increased product formation for all cases except acetate.

range of yields for the three solvents (butanol, acetone, and ethanol), as shown in the three-dimensional phenotypic solution space (Figure 4). This increased solvent production is also observed in experimental work by Jang *et al.* along with a reduction in acetate and butyrate production [44]. The study by Jang *et al.* also demonstrates that the butanol molar yield per glucose mole fed increases by 55% [44]. *iCac802* predicts that incorporation of these two knockouts results in an increase in butanol production by 63.6%. An earlier GSM model by Lee *et al.* [20] predicts an increase in butanol production but by a larger value of 86% due to a lack of internal protons in the model. In the case of mutant strains  $\Delta adc$  and  $\Delta hbd$ , acetone and butanol production is impaired [36,45] by knocking out acetoacetate decarboxylase (ADC) and hydroxybutyryl-CoA dehydrogenase (HBD), respectively (see Figure 5). In all cases the experimental yield is within the model-based calculated allowable yield for mutant phenotypes.

#### Modeling metabolic stressors using the CoreReg

*iCac802* is a metabolic model and does not include any regulatory information. This section describes model regulation under various conditions by using transcriptomic data. Regulation was implemented in order to better describe the metabolism of *C. acetobutylicum* under butyrate and butanol stress and to pinpoint the reactions where flux changes are needed to explain the

**Table 3 CO gassing analysis during acidogenic phase**

Metabolites	Wild type (0.52 h <sup>-1</sup> )		H inhibited (0.47 h <sup>-1</sup> )	
	Lower flux bound	Upper flux bound	Lower flux bound	Upper flux bound
Acetate	20.56	20.56	22.88	24.75
Butyrate	0.01	0.01	-1.99	-0.12
Acetone	0.00	0.00	0.00	0.00
Ethanol	0.00	0.00	0.00	1.87
Butanol	0.00	0.00	0.00	1.87
Hydrogen	13.02	13.02	0.00	0.00

Under CO gassing conditions the model shows inhibition of hydrogen production and butyrate uptake with alcohol but no acetone production. All the values represent production flux ranges with units of mmol gDW<sup>-1</sup> h<sup>-1</sup>. A negative value indicates consumption instead of production. The numbers in parentheses indicate the maximum growth rate values.

observed impact on biomass formation (that is, growth inhibition). Regulatory constraints on the *iCac802* model were imposed using the transcriptomic data from Wang *et al.* [3] in the form of modified reaction flux bounds for each of the stress conditions using the CoreReg method (see Methods section for full description).

When regulation was imposed on the model, the biomass yield decreased by approximately 70% for all stress conditions except for the low-level butyrate stress, where the biomass yield was unaffected. For each one of the stress conditions we identify the reactions (core set) for which the application of regulatory constraints is sufficient to reproduce the flux ranges seen in the model when all regulatory constraints are imposed. Core sets of reactions were identified for each of the stress conditions by comparing flux bounds of the regulated model with the imposed regulatory constraints (Step 4 in Figure 6). The core sets represent likely nexus points of regulation under stress conditions, as they can broadly propagate the regulatory effect to the stress affected pathways through model stoichiometry. When regulatory bounds

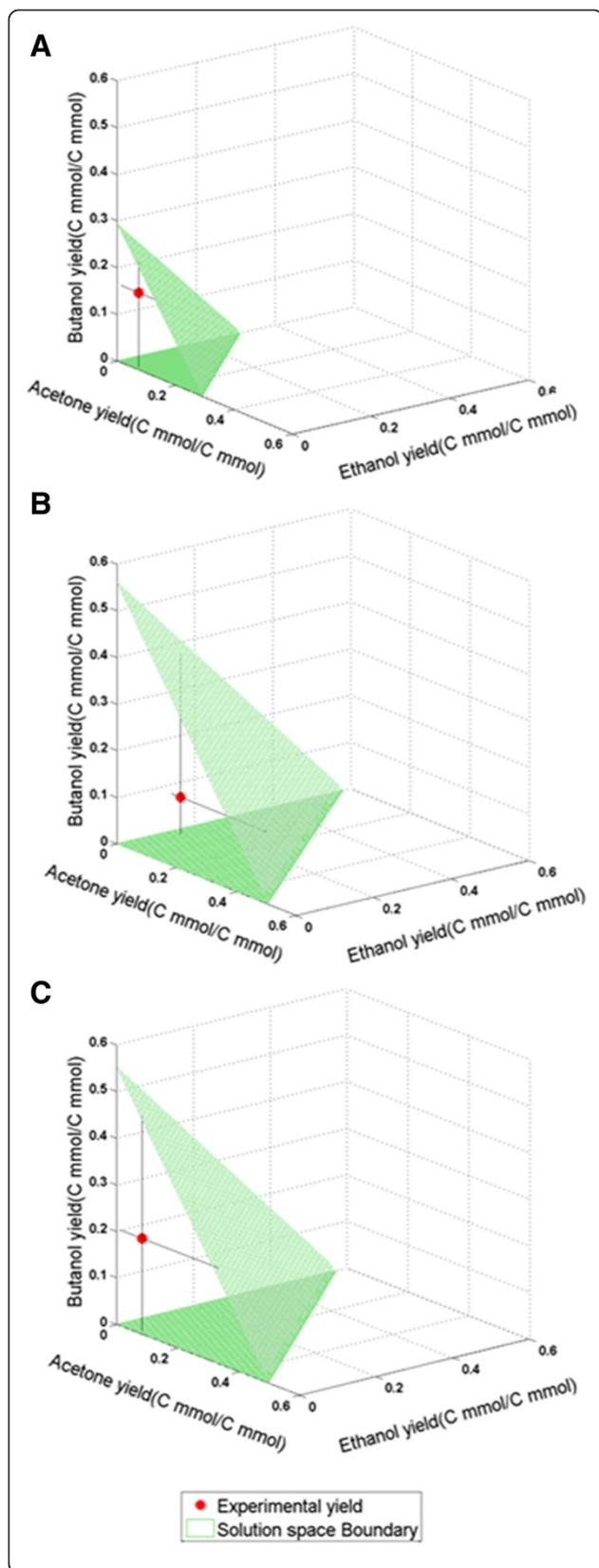
**Table 4 Cell recycling analysis during solventogenic phase showing lowering of biomass yield and increased solvent yield**

Metabolites	Wild type (0.32 h <sup>-1</sup> )		Cell recycle (0.17 h <sup>-1</sup> )	
	Lower flux bound	Upper flux bound	Lower flux bound	Upper flux bound
Ammonia	-2.94	-2.94	-2.35	-2.35
Phosphate	-0.37	-0.37	-0.30	-0.30
Ethanol	8.39	8.73	6.92	10.00
Butanol	0.00	0.52	0.00	4.04
Acetone	0.00	0.20	0.00	1.52

All the values represent production flux ranges with units of mmol gDW<sup>-1</sup> h<sup>-1</sup>. Negative values indicate consumption instead of production. The numbers in parentheses indicate the maximum growth rate values. The solvent fluxes were converted to C mmol units to compare the overall solvent yields.

were imposed exclusively on the core set of reactions, the flux ranges of all reactions matched those of the model with all regulatory constraints. Subsequent core sets were found for the various stress levels by excluding the regulatory constraints on previously identified core sets (primary, secondary, tertiary sets, and so on). These subsequent core sets consisted of reactions whose regulatory constraints restrict the fluxes from wild-type distribution, and represent additional reactions that may be focal points of regulation. All these core sets are listed in Tables 5 and 6. In most cases, the same core sets of reactions were shared among the different levels of butanol stress. Three of the four reactions that made up these core sets (ornithine carbamoyltransferase (OCBT), arginosuccinate lyase (ARGSL), and arginosuccinate synthase (ARGSS)) belonged to arginine metabolism. OCBT was present in the primary core set of all levels of butanol stress. The final reaction, N-acetyl-gamma-glutamyl-phosphate reductase (AGPR), which is associated with amino acid metabolism, was present in the primary high-level butanol stress core set. The arginine metabolism genes identified in the core set for butanol stress are regulated by ArgR, the arginine repressor [3]. Expression of the genes corresponding to these identified arginine metabolic reactions was strongly down-regulated under butanol stress [3,5] with a corresponding effect on biomass formation (growth inhibition). Identification of reactions in the arginine metabolism using the regulated model and its corroborative evidence from transcriptomic studies confirms the key role of arginine metabolism in response to stress and its subsequent effect of growth and metabolism. Furthermore, apart from the arginine metabolism, these genes are also involved in the biosynthesis of proline and lysine, which further emphasizes their role in regulating the amino acid metabolism and hence growth and biomass formation.

The primary medium level butyrate stress core set contained a reaction from pyrimidine metabolism, sulfate adenylyltransferase (SAT). The presence of this reaction can be related to the regulation of the DNA replication and repair mechanism which is initiated to protect the DNA from any damage owing to the oxidative stress component of the butyrate stress [3]. However, the subsequent core sets contained reactions involved in arginine metabolism, such as ARGSS in the high-level butyrate stress core set. Under butyrate stress, the effect that the regulatory constraints had on biomass yield was small. In comparison to butanol stress, under butyrate stress (low and medium), there is a strong up-regulation of genes in the arginine metabolism [3]. The addition of butyrate has a direct effect on the induction of solventogenesis, as the formation of solvents is directly related to reassimilation of butyrate from the medium (Figure 3). Jones *et al.* [46] have reported induction of the genes in



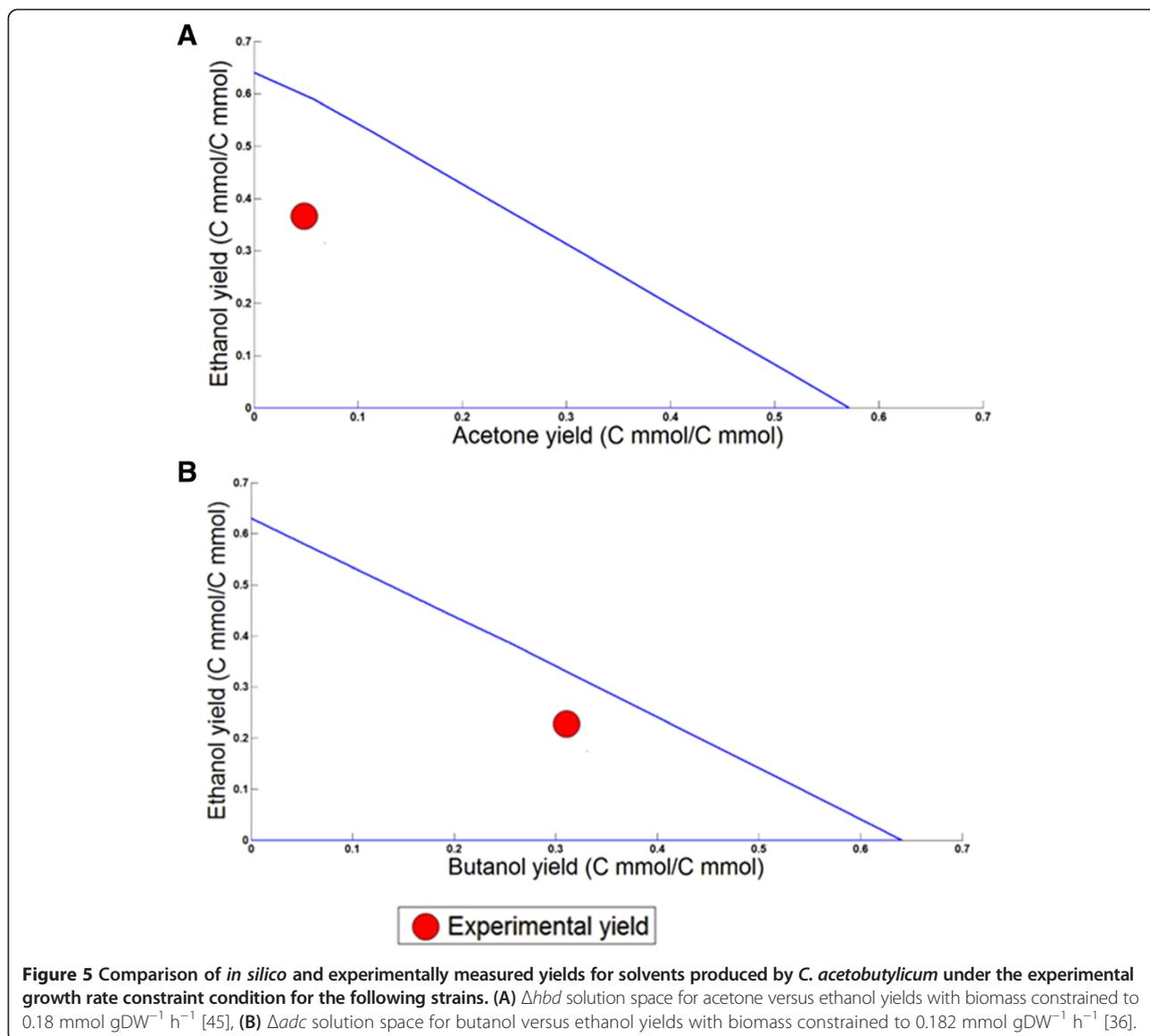
**Figure 4** Comparison of *in silico* and experimentally measured yields for solvents produced by *C. acetobutylicum* under the experimental growth rate constraints. **(A)** Wild-type solution space with biomass constrained to  $0.32 \text{ mmol gDW}^{-1} \text{ h}^{-1}$  [37]. **(B)**  $\Delta ptb$  solution space with biomass constrained to  $0.18 \text{ mmol gDW}^{-1} \text{ h}^{-1}$  [37]. **(C)**  $\Delta ack$  solution space with max biomass constraint of  $0.184 \text{ mmol gDW}^{-1} \text{ h}^{-1}$  [43].

arginine metabolism during the onset of solventogenesis, and this suggests that up-regulation of arginine genes under low and medium butyrate stress is associated with the induction of solventogenesis. The observation of arginine metabolism in the core set of high butyrate stress can be linked to the role of arginine metabolism as the acid response (AR3) system [3]. This ability of the model and regulatory modeling framework CoreReg to explicitly delineate the effect of two different metabolite stresses (at various levels) demonstrates the robustness and discriminatory capabilities of the model.

The addition of butyrate in the media leads to earlier onset of solventogenesis with higher butanol production [47,48], which is due to the corresponding up-regulation of the genes involved in solvent production and notably those of the *sol* operon (*CAP0162-CAP0164*, *adhe2-ctfA-ctfB*) [3,5]. The CoreReg method was able to simulate increased flux ranges for these reactions involved in solvent production during butyrate stress (Additional file 4).

## Discussion

A GSM model is a powerful tool that serves as a framework to visualize the changes captured from transcriptomic or proteomic data at the global metabolic level. The strength of such a model relies on the inherent characteristic capabilities to predict phenotypes from genotype. The proposed CoreReg method managed to elucidate focal points of regulation (core sets) on metabolic pathways. The core sets represent likely nexus points of regulation under stress conditions, as they can broadly propagate the regulatory effect to the stress affected pathways. Interestingly, different stressors and levels elicited different metabolic responses, as also corroborated by the DNA microarray data. The prediction of phenotypes and the corresponding regulation that leads to the phenotype along with model performance can be greatly enhanced by the development of a whole cell model. This would include the integration of regulatory knowledge derived from gene expression, transcription factors and their binding sites, regulation, and post-transcriptional regulation in the form of small non-coding regulatory RNA (sRNA) into GSM models. With the recent reconstruction of a transcriptional regulatory network [3] and the identification of the small RNome [4], the development of an integrated whole cell metabolic and regulatory model for *C. acetobutylicum* could provide superior insight into predicting phenotypes



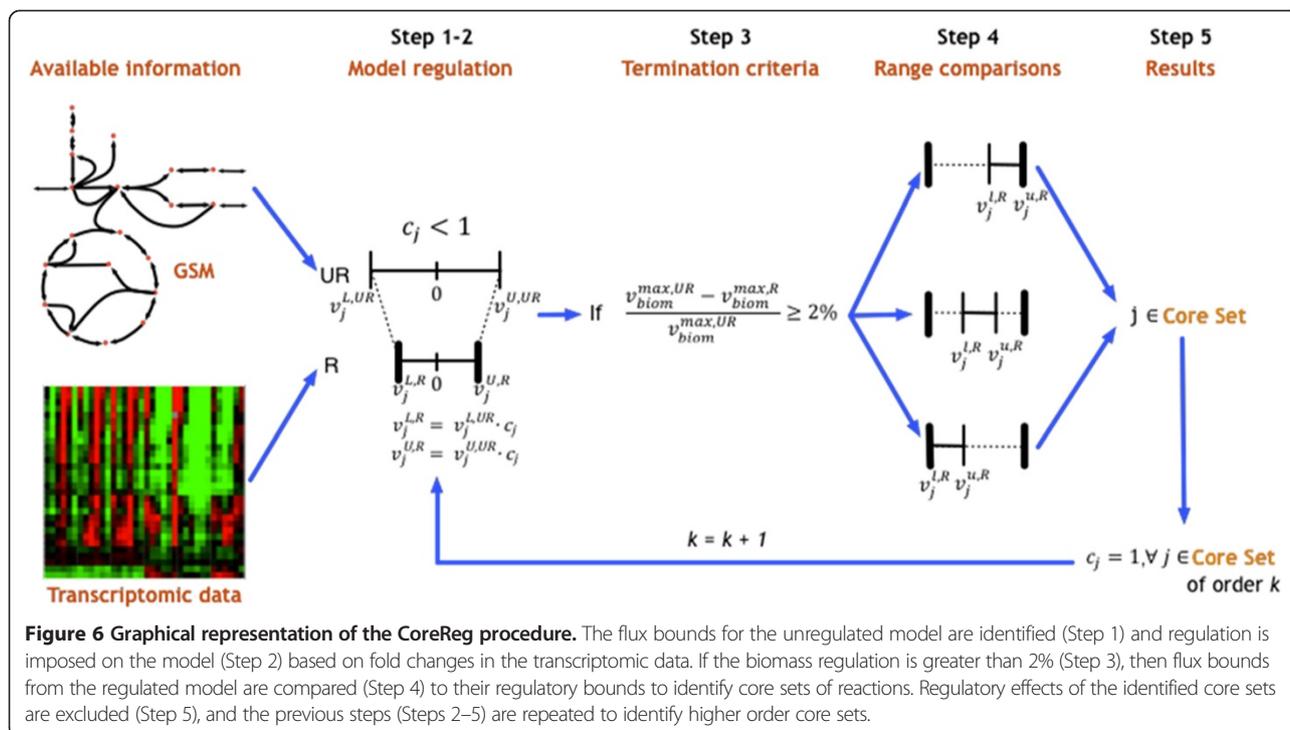
for the development of strains with higher tolerance to stressors and higher production of desired products. Thus, a more stress resilient strain of *C. acetobutylicum* may be engineered by improving these cellular functions.

## Conclusions

In this paper we have described the creation of a second-generation genome-scale metabolic model for *C. acetobutylicum* ATCC 824, *iCac802*, and the use of transcriptomic data to apply additional constraints on reaction flux bounds using the CoreReg method. These constraints were calculated for varying levels of butyrate and butanol stress and were used to identify core sets of reactions whose changes in flux values can explain broadly all observed changes in metabolism.

CoreReg was able to differentiate between the two stressors, with a larger restriction on biomass for butanol

stress. The core sets for butanol stress contain reactions in arginine and amino acid metabolism, while the butyrate stress core sets contain reactions in arginine and pyrimidine metabolism. These results corroborate previous findings concerning the down-regulation of arginine metabolism and regulation of DNA replication under stress conditions. Given transcriptomic data for other stressors or environmental conditions, the CoreReg method can be used to predict both the metabolic response and candidate focal points of regulation in terms of core sets. If there exists an available mechanistic description of the regulation, a remedial action can be taken to counteract the effect of stress on metabolism (for example, an up-regulating alternate pathway or a blocking regulator protein). In cases where the regulation mechanism is less well known, CoreReg results could be used to design plausible regulatory loops around the affected metabolic



reactions. These regulatory hypotheses can then be tested experimentally.

## Methods

### Model construction

The general principles of the metabolic model reconstruction process have been previously described [49-51]. Construction of *iCac802* entailed the following processes: 1) identification of biotransformations using previous models and homology searches; 2) assembly of reaction sets into a genome-scale metabolic model and subsequent conversion into a computations-ready format; 3) identification and removal of thermodynamically infeasible cycles; and 4) evaluation and improvement of model performance when compared to *in vivo* growth phenotypes.

We began metabolic model reconstruction by parsing the existing genome-scale models for *C. acetobutylicum* ATCC 824 [20-22] and the automated model generated on the Model SEED website [23]. We also made use of MetRxn, a knowledgebase that includes standardized metabolite and reaction descriptions drawn from multiple databases and genome-scale models [52], to parse and compare the contents of these models. We converted the gene associations in the Model SEED to the same open reading frame (ORF) naming scheme as the other models and available experimental data (CACxxxx or CAPxxxx) using the start and stop sites and in the genome annotation at the Model SEED website and those in the TIGR gene annotation [53]. We examined and updated the elemental and charge balancing of all reactions

by making use of the chemical formulas and charges provided by the SEED database (molecular charge values calculated at neutral pH) [23]. All reactions were checked to verify that the reactants and products shared the same total numbers of different atoms and the same total charge. Reactions that did not pass these tests were replaced with equivalent balanced reactions from Model SEED (Additional file 3). In the case of no alternate Model SEED reaction, the reactions underwent rebalancing, where protons, hydroxide, and/or water molecules were added to balance atoms and/or charges. Metabolites that were not fully specified due to the use of generic R groups as side chains, or that were oligomeric with a non-specified number of repeat units, were flagged for manual examination and balancing. One-to-one equivalency of metabolites with such generic R groups and metabolites without generic R groups was verified to ensure balanced reactions. The biomass equation was adapted from the McAnulty model [22], the only modifications being the use of charged and uncharged tRNAs to account for amino acid utilization, the doubling of the coefficients for solutes in the equation, and a slight decrease in the coefficient for protein use from 3.1 to 3.

Model curation began with the removal of thermodynamically infeasible cycles present in the model. Flux variability analysis (FVA) was performed with no biomass constraints to identify unbounded reaction fluxes. The number of unbounded reactions was reduced by restricting the directionality of certain reactions using Model SEED's calculated values of the reaction's Gibbs free energy.

**Table 5 Genes associated with core set of reactions under butanol stress conditions**

	Stress level	ORF	Gene
Primary (K=1)	Low	CAC0316	Ornithine carbamoyltransferase (OCBT)
	Med	CAC0316	Ornithine carbamoyltransferase
	High	CAC0316	Ornithine carbamoyltransferase
Secondary (K=2)	Low	CAC2390	N-acetyl-gamma-glutamyl-phosphate reductase (AGPR)
		CAC0974	Argininosuccinate lyase (ARGSL)
	Med	CAC2390	N-acetyl-gamma-glutamyl-phosphate reductase
		CAC0974	Argininosuccinate lyase
	High	CAC2390	N-acetyl-gamma-glutamyl-phosphate reductase
		CAC0974	Argininosuccinate lyase
Tertiary (K=3)	Low	CAC2389	Acetylglutamate kinase (ACGK)
		CAC0973	Argininosuccinate synthase (ARGSS)
		CAC2389	Acetylglutamate kinase
	Med	CAC0973	Argininosuccinate synthase
		CAC2389	Acetylglutamate kinase
		CAC0973	Argininosuccinate synthase
Quaternary (K=4) [minimal effect on biomass]	Low	CAC2391 or CAC3020	Bifunctional ornithine acetyltransferase/N-acetylglutamate synthase (OACT/AGS)
	Med	CAC2391 or CAC3020	Bifunctional ornithine acetyltransferase/N-acetylglutamate synthase
	High	CAC2388 or CAC0368	4-aminobutyrate aminotransferase acetylornithine/ N-succinyldiaminopimelate aminotransferase (GABAT/ARGD)

If regulatory bounds were imposed exclusively on the reactions in the core set, the flux ranges would match those of the model with all regulatory constraints. Regulation was removed from the primary core sets to obtain secondary cores sets, and so on. Biomass regulation reduced with increasing order of core sets.

If the entire range, including error, was more than 4 kcal/mol removed from zero, the reaction was restricted to the direction specified by the free energy [54]. The method described by Schellenberger *et al.* [55] was then used to identify the core set of thermodynamically

infeasible cycles which form the basis of all such possible cycles. This method relies on the observation that all possible thermodynamically infeasible cycles form the null space of the stoichiometric matrix [55]. During the identification of the basis of the null space, only the

**Table 6 Genes associated with core set of reactions under butyrate stress conditions**

	Stress level	ORF	Gene
Primary (K=1)	Low [minimal effect on biomass]	CAC2113	Bifunctional pyrimidine regulatory protein PyrR/uracil phosphoribosyltransferase (UPRT)
	Med [minimal effect on biomass]	CAC0109 and CAC0110	Sulfate adenylyltransferase (SAT)
	High	CAC0973	Argininosuccinate synthase
Secondary (K=2)	Low		No core set
	Med		No core set
	High	CAC0316 CAC2389	Ornithine carbamoyltransferase Acetylglutamate kinase
Tertiary (K=3)	Low		No core set
	Med		No core set
	High	CAC0974	Argininosuccinate lyase
Quaternary (K=4)	Low		No core set
	Med		No core set
	High [minimal effect on biomass]	CAC2391 or CAC3020	4-aminobutyrate aminotransferase acetylornithine/ N-succinyldiaminopimelate aminotransferase

If regulatory bounds were imposed exclusively on the reactions in the core set, the flux ranges would match those of the model with all regulatory constraints. Regulation was removed from the primary core sets to obtain secondary cores sets, and so on. Biomass regulation reduced with increasing order of core sets.

set of reactions whose fluxes hit the bounds during FVA were used to decrease processing time. The null space basis of this set was evaluated by LU decomposition of the stoichiometric matrix [55].

Of the cycles found, those involving only two reactions were examined first. These were composed of equivalent reactions, that is, reactions which both converted a given set of reactants to products or vice versa using the same co-factors, varying only in their directionality. One of the two identified equivalent reactions was subsequently removed based on a preponderance of literature or annotation information concerning its directionality. Cycles containing more than two reactions were formed due to the presence of pathways that recycled metabolites consumed or produced by an alternate route with a zero net metabolic cost. Reactions in these cycles were reviewed, and those with no literature or annotation reference were either removed or had their directionality restricted. Figure 1A is an example of such a cycle. The 2-acetolactate methylmutase (ACLM) reaction is more accurately represented by two separate reactions, (R)-2, 3-dihydroxy-3-methylbutanoate: NADP + oxidoreductase (R-DMBO) and 2, 3-dihydroxy-3-methylbutanoate: NADP + oxidoreductase (DMBO). The removal of this aggregate reaction eliminates the thermodynamically infeasible cycle, as shown in Figure 1B. Figure 1 (C-E) represents another cycle and its subsequent removal using experimental <sup>13</sup>C-metabolic flux analysis (<sup>13</sup>C-MFA) data.

After reviewing all of the cycles in the null basis with up to five participating reactions, the null basis was reevaluated. After fixing the smaller cycles, the same review process was extended to larger cycles. This reevaluation process was iterated until no more cycles were detected using FVA. A total of 101 reactions were removed from the model, and 58 had their directionality restricted to stop the involvement of 404 reactions in thermodynamically infeasible cycles.

### Model simulations and analysis

Flux balance analysis (FBA) was performed to obtain the maximum attainable growth rate under the given constraints [56].

$$\begin{aligned} & \text{Max } v_{\text{Biomass}} \\ & \text{Subject to} \\ & \sum_{j=1}^M S_{ij}v_j = 0, \forall i \in 1, \dots, N \quad (1) \\ & v_j^{\min} \leq v_j \leq v_j^{\max}, \forall j \in 1, \dots, M \quad (2) \end{aligned}$$

where  $S_{ij}$  is the stoichiometric coefficient for metabolite  $i$  in reaction  $j$ .  $v_j$  represents the flux of reaction  $j$ , while  $v_j^{\min}$  and  $v_j^{\max}$  denote the minimum and maximum flux bounds on reaction  $j$ .  $N$  and  $M$  represent the overall sets of metabolites and reactions, respectively.

FVA was used to identify reactions present in thermodynamically infeasible cycles.

$$\left. \begin{aligned} & \text{Max/Min } v_{j^*} \\ & \text{Subject to} \\ & \sum_{j=1}^M S_{ij}v_j = 0, \forall i \in 1, \dots, N \quad (3) \\ & v_j^{\min} \leq v_j \leq v_j^{\max}, \forall j \in 1, \dots, M \quad (4) \end{aligned} \right\} \forall j^* \in 1, \dots, M$$

No constraints were placed on the biomass in order to identify all thermodynamically infeasible cycles. This analysis was performed iteratively for all metabolites.

### Model testing

The model predictions were tested against experimental <sup>13</sup>C flux data [25], experimental fermentation data [3,35,38,40,41], and *in vivo* gene knockout data from the literature [37,43,45,57].

While the acidogenic phase was simulated in the model solely through the inclusion of required nutrients, additional constraints were required for the solventogenic phase. The solventogenic phase of Clostridia is characterized by the uptake of acetate and butyrate along with the reduction of carbon flux towards amino acids [58]. Thus, the solventogenic phase was simulated with acetate and butyrate as additional nutrients, and constraints on the export of acids and excess amino acids.

### Incorporation of regulation using CoreReg

Regulation was incorporated into the model by a stepwise procedure that modified the minimum and maximum flux bounds of reactions based on fold change values of corresponding gene expression values by a new method called CoreReg. Gene expression data for each stress condition (GSE48031 and GSE48039), obtained from Wang *et al.* [3], were used to calculate the fold change from unstressed conditions using significance analysis of microarrays (SAM) [3,59]. Transcriptomic data were collected for three concentrations each of butyrate and butanol. We refer to these as low (30 mM butyrate, 30 mM butanol), medium (40 mM butyrate, 60 mM butanol), and high (50 mM butyrate, 90 mM butanol) stress conditions. The fold change for each reaction under each stress condition was calculated from gene expression fold changes under stressed conditions by using GPRs. In the case of multiple enzyme subunits, the minimum expression value for the genes associated with the subunits was considered for calculating the reaction fold change. In the case of isozymes the total transcript level, obtained by summation of all isozyme transcripts, was considered. The unregulated model reaction bounds without any biomass constraint represent the minimum and maximum possible bounds of each reaction. Thus, a further increase in these bounds does not affect any

maximum yield calculations, as the bounds are not active. Thus, only down-regulated genes were considered for evaluating the regulated model. The procedure for implementing regulation can be divided into five steps as explained below (see Figure 6).

**Step 1:** FVA is performed on the unregulated (UR) model to obtain lower  $v_j^{L,UR}$  and upper  $v_j^{U,UR}$  reaction flux bounds. The value of  $k$  is set equal to one to indicate primary core set.

**Step 2:** FVA upper and lower bounds for the unregulated model are multiplied by the fold change value ( $c_j$ ) obtained using the transcriptomic data. The lower  $v_j^{L,R}$  and upper  $v_j^{U,R}$  bounds for the regulated (R) model are evaluated as follows:

$$\forall j \text{ such that } c_j < 1 \rightarrow \begin{cases} v_j^{U,R} = v_j^{U,UR} \cdot c_j & (5) \\ v_j^{L,R} = v_j^{L,UR} \cdot c_j & (6) \end{cases}$$

Note that the updated lower bound is non-zero only for reversible reactions, effectively lowering the maximum possible flux value in the reverse direction.

**Step 3:** FBA is performed on the unregulated and the regulated model to obtain maximum biomass yields  $v_{biom}^{max,UR}$  and  $v_{biom}^{max,R}$ , respectively. If  $v_{biom}^{max,R}$  varies from  $v_{biom}^{max,UR}$  by less than 2%, then the process is terminated, because the effect of the remaining regulation in the model is too small to cause any significant changes in metabolism as exemplified by the max biomass yield. Therefore, no additional regulatory core sets are extracted.

**Step 4:** FVA is performed at max biomass  $v_{biom}^{max,R}$  on the regulated (R) model to obtain lower  $v_j^{L,R}$  and upper  $v_j^{U,R}$  reaction flux bounds. These bounds are next compared with the imposed regulatory constraints from Step 2.

**Step 5:** Reactions  $j$  whose flux bounds are equal to the regulatory constraints (that is,  $v_j^{u,R} \equiv v_j^{u,R}$  or  $v_j^{l,R} \equiv v_j^{l,R}$ ) are assembled into the core set of reactions (of order  $k$ ). To identify secondary, tertiary, and higher order core sets, the fold change values ( $c_j$ ) for the previously determined  $k$  core sets are set to one, thus removing their regulatory role in the model. The process is repeated from Step 2 with the value of  $k$  increased by one.

## Additional files

**Additional file 1:** The genome-scale model *iCac802* with GPRs and metabolite information in SBML format.

**Additional file 2:** The genome-scale model *iCac802* with GPRs and metabolite information.

**Additional file 3:** *iCac802* model comparisons with other *C. acetobutylicum* models.

**Additional file 4:** Set of modified reaction bounds and analysis for all six stress conditions.

## Abbreviations

$^{13}\text{C}$ -MFA:  $^{13}\text{C}$ -metabolic flux analysis; AADC: acetoacetate decarboxylase; ABE: acetone-butanol-ethanol; ACGK: acetylglutamate kinase; ACK: acetate kinase; ACLM: 2-acetolactate methylmutase; ACN: aconitase; AGPR: N-acetyl-gamma-glutamyl-phosphate reductase; AGS: N-acetylglutamate synthase; AGT: alanine-glyoxylate; ALD: aldolase; ARGD: N-succinylidiaminopimelate aminotransferase; ARGSL: arginosuccinate lyase; ARGSS: arginosuccinate synthase; ASPA: aspartate ammonia-lyase; AST: aspartate transaminase; CS: citrate synthase; DMBO: 2,3-dihydroxy-3-methylbutanoate:NADP + oxidoreductase; ENO: enolase; FBA: flux balance analysis; FH: fumarate hydratase; FVA: flux variability analysis; GABAT: 4-aminobutyrate aminotransferase acetylmithine; GAPDH: glyceraldehyde 3-phosphate dehydrogenase; GPI: glucose-6-phosphate isomerase; GPR: gene-to-protein-to-reaction; GSM: genome-scale metabolic (model); HBDH: hydroxybutyryl-CoA dehydrogenase; HK: hexokinase; IDH: isocitrate dehydrogenase;  $\alpha$ -KGDH:  $\alpha$ -ketoglutarate dehydrogenase; MDH: malate dehydrogenase; MS: malate synthase; OACT: ornithine acetyltransferase; OCBT: ornithine carbamoyltransferase; PC: pyruvate carboxylase; PDH: pyruvate dehydrogenase; PFK: phosphofructokinase; PLPS: pyridoxal phosphate synthase; PTA: phosphotransacetylase; PTB: Phosphotransbutyrylase; PYK: pyruvate kinase; R-DMBO: (R)-2,3-dihydroxy-3-methylbutanoate:NADP + oxidoreductase; rFBA: regulatory flux balance analysis; SAM: significance analysis of microarrays; SAT: sulfate adenylyltransferase; SCS: succinyl coenzyme A synthetase; SDH: succinate dehydrogenase; SR-FBA: steady-state regulatory flux balance analysis; sRNAs: small non-coding regulatory RNAs; TPI: triosephosphate isomerase; UPRT: uracil phosphoribosyltransferase.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

SD developed the method. SD, TJM, and KPV analyzed the data. SD, TJM, and KPV prepared the manuscript. ETP and CDM conceived and supervised the study. All authors read and approved the final manuscript.

## Acknowledgements

This work was supported by the US Department of Energy (DE-SC0007092).

## Author details

<sup>1</sup>Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania, USA. <sup>2</sup>Delaware Biotechnology Institute, 15 Innovation Way, Newark 19711, Delaware, USA. <sup>3</sup>Department of Chemical Engineering, University of Delaware, Newark, Delaware, USA.

Received: 30 May 2014 Accepted: 22 September 2014

Published online: 14 October 2014

## References

- Tracy BP, Jones SW, Fast AG, Indurthi DC, Papoutsakis ET: **Clostridia: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications.** *Curr Opin Biotechnol* 2012, **23**:364–381.
- Papoutsakis ET: **Engineering solventogenic clostridia.** *Curr Opin Biotechnol* 2008, **19**:420–429.
- Wang Q, Venkataramanan KP, Huang H, Papoutsakis ET, Wu CH: **Transcription factors and genetic circuits orchestrating the complex, multilayered response of *Clostridium acetobutylicum* to butanol and butyrate stress.** *BMC Syst Biol* 2013, **7**:120.
- Venkataramanan KP, Jones SW, McCormick KP, Kunjeti SG, Ralston MT, Meyers BC, Papoutsakis ET: **The *Clostridium* small RNome that responds to stress: the paradigm and importance of toxic metabolite stress in *C. acetobutylicum*.** *BMC Genomics* 2013, **14**:849.
- Alsaker KV, Paredes C, Papoutsakis ET: **Metabolite stress and tolerance in the production of biofuels and chemicals: gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe *Clostridium acetobutylicum*.** *Biotechnol Bioeng* 2010, **105**:1131–1147.
- Schwarz KM, Kuit W, Grimmer C, Ehrenreich A, Kengen SWM: **A transcriptional study of acidogenic chemostat cells of *Clostridium acetobutylicum*- cellular behavior in adaptation to n-butanol.** *J Biotechnol* 2012, **161**:366–377.
- Janssen H, Grimmer C, Ehrenreich A, Bahl H, Fischer RJ: **A transcriptional study of acidogenic chemostat cells of *Clostridium acetobutylicum*-solvent stress caused by a transient n-butanol pulse.** *J Biotechnol* 2012, **161**:354–365.

8. Hou S, Jones SW, Choe LH, Papoutsakis ET, Lee KH: **Workflow for quantitative proteomic analysis of *Clostridium acetobutylicum* ATCC 824 using iTRAQ tags.** *Methods* 2013, **61**:269–276.
9. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO: **Reconstruction of biochemical networks in microorganisms.** *Nat Rev Microbiol* 2009, **7**:129–143.
10. Burgard AP, Pharkya P, Maranas CD: **OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization.** *Biotechnol Bioeng* 2003, **84**:647–657.
11. Oliveira AP, Nielsen J, Forster J: **Modeling *Lactococcus lactis* using a genome-scale flux model.** *BMC Microbiol* 2005, **5**:39.
12. Alper H, Jin YS, Moxley JF, Stephanopoulos G: **Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*.** *Metab Eng* 2005, **7**:155–164.
13. Pharkya P, Maranas CD: **An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems.** *Metab Eng* 2006, **8**:1–13.
14. Almaas E, Kovacs B, Vicssek T, Oltvai ZN, Barabasi AL: **Global organization of metabolic fluxes in the bacterium *Escherichia coli*.** *Nature* 2004, **427**:839–843.
15. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD: **Flux coupling analysis of genome-scale metabolic network reconstructions.** *Genome Res* 2004, **14**:301–312.
16. Motter AE, Gulbahce N, Almaas E, Barabasi AL: **Predicting synthetic rescues in metabolic networks.** *Mol Syst Biol* 2008, **4**:168.
17. Jin YS, Jeffries TW: **Stoichiometric network constraints on xylose metabolism by recombinant *Saccharomyces cerevisiae*.** *Metab Eng* 2004, **6**:229–238.
18. Papoutsakis ET: **Equations and calculations for fermentations of butyric acid bacteria.** *Biotechnol Bioeng* 1984, **26**:174–187.
19. Desai RP, Harris LM, Welker NE, Papoutsakis ET: **Metabolic flux analysis elucidates the importance of the acid-formation pathways in regulating solvent production by *Clostridium acetobutylicum*.** *Metab Eng* 1999, **1**:206–213.
20. Lee J, Yun H, Feist AM, Palsson BO, Lee SY: **Genome-scale reconstruction and in silico analysis of the *Clostridium acetobutylicum* ATCC 824 metabolic network.** *Appl Microbiol Biotechnol* 2008, **80**:849–862.
21. Senger RS, Papoutsakis ET: **Genome-scale model for *Clostridium acetobutylicum*: Part I. Metabolic network resolution and analysis.** *Biotechnol Bioeng* 2008, **101**:1036–1052.
22. McNulty MJ, Yen JY, Freedman BG, Senger RS: **Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism in silico.** *BMC Syst Biol* 2012, **6**:42.
23. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL: **High-throughput generation, optimization and analysis of genome-scale metabolic models.** *Nat Biotechnol* 2010, **28**:977–982.
24. Crown SB, Indurthi DC, Ahn WS, Choi J, Papoutsakis ET, Antoniewicz MR: **Resolving the TCA cycle and pentose-phosphate pathway of *Clostridium acetobutylicum* ATCC 824: Isotopomer analysis, in vitro activities and expression analysis.** *Biotechnol J* 2011, **6**:300–305.
25. Au J, Choi J, Jones SW, Venkataramanan KP, Antoniewicz MR: **Parallel labeling experiments validate *Clostridium acetobutylicum* metabolic network model for <sup>13</sup>C metabolic flux analysis.** *Metabolic engineering*, **26**:23–33.
26. Covert MW, Palsson BO: **Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*.** *J Biol Chem* 2002, **277**:28058–28064.
27. Covert MW, Schilling CH, Palsson B: **Regulation of gene expression in flux balance models of metabolism.** *J Theor Biol* 2001, **213**:73–88.
28. Shlomi T, Eisenberg Y, Sharan R, Ruppin E: **A genome-scale computational study of the interplay between transcriptional regulation and metabolism.** *Mol Syst Biol* 2007, **3**:101.
29. Barua D, Kim J, Reed JL: **An automated phenotype-driven approach (GeneForce) for refining metabolic and regulatory models.** *PLoS Comput Biol* 2010, **6**:e1000970.
30. Chandrasekaran S, Price ND: **Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*.** *Proc Natl Acad Sci U S A* 2010, **107**:17845–17850.
31. Schmidt BJ, Ebrahim A, Metz TO, Adkins JN, Palsson BO, Hyduke DR: **GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data.** *Bioinformatics* 2013, **29**:2900–2908.
32. Yizhak K, Gabay O, Cohen H, Ruppin E: **Model-based identification of drug targets that revert disrupted metabolism and its application to ageing.** *Nat Commun* 2013, **4**:2632.
33. Ranganathan S, Suthers PF, Maranas CD: **OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions.** *PLoS Comput Biol* 2010, **6**:e1000744.
34. Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, Cheng TY, Moody DB, Murray M, Galagan JE: **Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production.** *PLoS Comput Biol* 2009, **5**:e1000489.
35. Meyer CL, Roos JW, Papoutsakis ET: **Carbon-monoxide gasing leads to alcohol production and butyrate uptake without acetone formation in continuous cultures of *Clostridium acetobutylicum*.** *Appl Microbiol Biotechnol* 1986, **24**:159–167.
36. Jiang Y, Xu C, Dong F, Yang Y, Jiang W, Yang S: **Disruption of the acetoacetate decarboxylase gene in solvent-producing *Clostridium acetobutylicum* increases the butanol ratio.** *Metab Eng* 2009, **11**:284–291.
37. Lehmann D, Radomski N, Lutke-Eversloh T: **New insights into the butyric acid metabolism of *Clostridium acetobutylicum*.** *Appl Microbiol Biotechnol* 2012, **96**:1325–1339.
38. Meyer CL, Papoutsakis ET: **Increased levels of ATP and NADH are associated with increased solvent production in continuous cultures of *Clostridium acetobutylicum*.** *Appl Microbiol Biotechnol* 1989, **30**:450–459.
39. Vasconcelos I, Girbal L, Soucaille P: **Regulation of carbon and electron flow in *Clostridium acetobutylicum* grown in chemostat culture at neutral pH on mixtures of glucose and glycerol.** *J Bacteriol* 1994, **176**:1443–1450.
40. Meyer CL, Papoutsakis ET: **Continuous and biomass recycle fermentations of *Clostridium acetobutylicum*.2. Novel patterns in energetics and product-formation kinetics.** *Bioprocess Eng* 1989, **4**:49–55.
41. Meyer CL, Papoutsakis ET: **Continuous and biomass recycle fermentations of *Clostridium acetobutylicum*.1. ATP supply and demand determines product selectivity.** *Bioprocess Eng* 1989, **4**:1–10.
42. Cooksley CM, Zhang Y, Wang H, Redl S, Winzer K, Minton NP: **Targeted mutagenesis of the *Clostridium acetobutylicum* acetone-butanol-ethanol fermentation pathway.** *Metab Eng* 2012, **14**:630–641.
43. Kuit W, Minton NP, Lopez-Contreras AM, Eggink G: **Disruption of the acetate kinase (*ack*) gene of *Clostridium acetobutylicum* results in delayed acetate production.** *Appl Microbiol Biotechnol* 2012, **94**:729–741.
44. Jang YS, Lee JY, Lee J, Park JH, Im JA, Eom MH, Lee J, Lee SH, Song H, Cho JH, Seung do Y, Lee SY: **Enhanced butanol production obtained by reinforcing the direct butanol-forming route in *Clostridium acetobutylicum*.** *MBio* 2012, **3**:e00314–12.
45. Lehmann D, Lutke-Eversloh T: **Switching *Clostridium acetobutylicum* to an ethanol producer by disruption of the butyrate/butanol fermentative pathway.** *Metab Eng* 2011, **13**:464–473.
46. Jones SW, Paredes CJ, Tracy B, Cheng N, Sillers R, Senger RS, Papoutsakis ET: **The transcriptional program underlying the physiology of clostridial sporulation.** *Genome Biology* 2008, **9**:R114.
47. Huesemann M, Papoutsakis ET: **Effect of acetoacetate, butyrate, and uncoupling ionophores on growth and product formation of *Clostridium acetobutylicum*.** *Biotechnol Lett* 1986, **8**:37–42.
48. Huesemann MHW, Papoutsakis ET: **Solventogenesis in *Clostridium acetobutylicum* fermentations related to carboxylic-acid and proton concentrations.** *Biotechnol Bioeng* 1988, **32**:843–852.
49. Feist AM, Scholten JC, Palsson BO, Brockman FJ, Ideker T: **Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*.** *Mol Syst Biol* 2006, **2**:2006 0004.
50. Reed JL, Famili I, Thiele I, Palsson BO: **Towards multidimensional genome annotation.** *Nat Rev Genet* 2006, **7**:130–141.
51. Thiele I, Palsson BO: **A protocol for generating a high-quality genome-scale metabolic reconstruction.** *Nat Protoc* 2010, **5**:93–121.
52. Kumar A, Suthers PF, Maranas CD: **MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases.** *BMC Bioinformatics* 2012, **13**:6.
53. Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, Wu H, Rabinowicz PD, Town CD, Buell CR, Chan AP: **The TIGR Plant Transcript Assemblies database.** *Nucleic Acids Res* 2007, **35**:D846–D851.
54. Mueller TJ, Berla BM, Pakrasi HB, Maranas CD: **Rapid construction of metabolic models for a family of Cyanobacteria using a multiple source annotation workflow.** *BMC Syst Biol* 2013, **7**:142.

55. Schellenberger J, Lewis NE, Palsson BO: **Elimination of thermodynamically infeasible loops in steady-state metabolic models.** *Biophys J* 2011, **100**:544–553.
56. Orth JD, Thiele I, Palsson BO: **What is flux balance analysis?** *Nat Biotechnol* 2010, **28**:245–248.
57. Hou XH, Peng WF, Xiong L, Huang C, Chen XF, Chen XD, Zhang WG: **Engineering *Clostridium acetobutylicum* for alcohol production.** *J Biotechnol* 2013, **166**:25–33.
58. Amador-Noguez D, Brasg IA, Feng XJ, Roquet N, Rabinowitz JD: **Metabolome remodeling during the acidogenic-solventogenic transition in *Clostridium acetobutylicum*.** *Appl Environ Microbiol* 2011, **77**:7984–7997.
59. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci U S A* 2001, **98**:5116–5121.

doi:10.1186/s13068-014-0144-4

**Cite this article as:** Dash *et al.*: Capturing the response of *Clostridium acetobutylicum* to chemical stressors using a regulated genome-scale metabolic model. *Biotechnology for Biofuels* 2014 **7**:144.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

