Predicting metabolic and proteomic variations in response to stress and protein overproduction in *B. subtilis*

Juan D. Tibocha-Bonilla[1], Eammon Riley[2], Cristal Zuñiga[3], Colton Lloyd[4], Kevin Rychel[4], Kit Pogliano[2], Karsten Zengler[3,4,5]*

[1]Department of Bioinformatics and Systems Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0760, USA

[2]Division of Biological Sciences, University of California San Diego, La Jolla, California, USA

[3]Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0760, USA

[4]Department of Bioengineering, University of California, San Diego, La Jolla CA 92093-0412, USA

[5]Center for Microbiome Innovation, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0403, USA

*Corresponding author: Karsten Zengler

Background

Bacillus subtilis is the most studied Gram-positive bacteria and it is the second-best characterized bacterium, only after *Escherichia coli*. *B. subtilis* has been widely used in industrial applications, including the synthesis of antibiotics, enzymes, and vitamins. The tremendous amount of omics and physiological data available for *B. subtilis* allowed reconstructing one of the first bacterial genome-scale metabolic models (M-models)[1]. M-models accurately predict metabolic responses to nutrient levels and gene-knockouts, however enzyme production costs and protein secretion simulations are beyond their scope.

A new generation of computational models was recently conceived by linking gene expression mechanisms to metabolic reactions[2]. The models of metabolism and gene expression (ME-models) leave protein production profiles as variables that depend on a source genome sequence associated with metabolic reaction fluxes with variable biosynthetic costs. ME-models can also simulate stress conditions that involve shifts in gene expression or biomass composition. Here, we describe the reconstruction and validation of the first ME-model of *B. subtilis*, *i*JT964-ME. This model achieved higher performance scores on the prediction of gene essentiality as compared to the M-model. We successfully validated the model by integrating physiological and omics data associated with gene expression responses to ethanol and salt stress, unraveling a hidden tryptophan mechanism that it is upregulated under stress conditions. We also used *i*JT964-ME to identified key metabolic pathways that permitted the increase in amylase production. All in all, we illustrate the potential of *i*JT964-ME in the study of proteomic response to stress and the optimization of protein production.

Methods

Model reconstruction. Reconstruction was performed in Python 3.6, using the reconstruction packages cobrapy 0.5.4 [3] and COBRAme [4]. Models were solved using the package SOLVEme [5]. In brief, every reaction in a template core metabolic network

(M-model) is coupled with the enzyme's consumption that catalyzes it.

Gene essentiality analysis. Single gene knockouts were modeled by closing their respective translation reactions. Genes were deemed essential when the single knockouts resulted in a growth rate of zero. Results were validated with a list of essential genes reported by Juhas et al. [6] for *B. subtilis* and EcoCyc [7] for *E. coli.*

Modeling flux distributions under stress. Ethanol is a small polar molecule that can readily diffuse through the cell membrane. Therefore, ethanol uptake was modeled with no enzymatic coupling. As opposed to ethanol, salt is transported through the membrane by a series of complexes. Salt stress was modeled by an artificial uptake of salt uncoupled to any transporter so that higher uptakes did not falsely trigger importer expression in the model.

Modeling and validating amylase secretion rates. Random sampling of the solution space was performed from 90% to 100% of the optimal growth rate at the simulation conditions to generate a robust distribution of biologically relevant fluxes [8,9]. Validation of secretion rates was performed in data collected from two previously reported experimental datasets, at a high [10] and low [11] growth rate. Amylase overexpression was performed by forcing it within the range from the base requirement at $0.17\ h^{-1}$ until the requirement, the model would predict for $0.195\ h^{-1}$.

Results

*i*JT964-ME improves gene essentiality accuracy by accounting for metabolism and gene expression

To reconstruct *i*JT964-ME we adapted the available metabolic modeling packages COBRAme, COBRApy, and ECOLIme. COBRAme that were originally built for *E. coli*. The resulting *B. subtilis*-specific pipeline expanded the existing M-model *i*YO844 with non-metabolic reactions, including translation, transcription, tRNA charging, and post-translational modification[4]. The final ME-model (*i*JT964-ME) contains 964 genes, 6282 reactions, and 4208 metabolites (Fig. 1a).

The addition of gene expression reactions into the network of *B. subtilis* resulted in a 14% increase in genome coverage, with 32% of them being essential as predicted by *i*JT964-ME. The extensive manual curation performed for *i*JT964-ME significantly increased the prediction scores of gene essentiality. In some cases, prediction scores surpass those predicted for *E. coli* M- and ME-models. Interestingly, just a 14% increase in gene content allowed *i*JT964-ME to predict essentiality with increases of 34% in the Matthews Correlation Coefficient (MCC) and 40% in Precision.

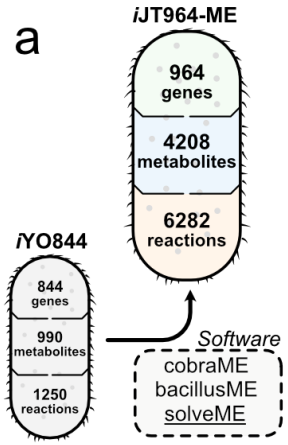

Fig. 1. Properties of *i*JT964-ME and gene essentiality prediction performance. a) Number of genes, metabolites, and reactions included in the model.

Predicting metabolic and proteomic shifts of B. subtilis under stress

Strategies adopted by living organisms to overcome stress are widely studied but not fully understood. In this section, we studied the predictive capability of *i*JT964-ME on salt and ethanol stress. Previous reports highlighted a group of significantly co-regulated genes associated with an outstanding upregulation of tryptophan synthesis under ethanol stress (4% v/v)[12]. We used these findings to assess the capability of *i*JT964-ME of identifying metabolic shifts of *B. subtilis* under ethanol stress.

First, we modeled ethanol stress by constraining experimentally observed uptake ethanol rates in the model. Our simulations show that the higher tryptophan demand can be caused solely by an increase in the demand for ethanol processing and acetate secretion enzymes. According to the resulting flux distributions (Fig. 2a), ethanol was converted to acetate through alcohol dehydrogenase (adhA) and aldehyde dehydrogenase (aldX), which was then secreted through a sodium-dependent acetate symporter (ywcA). The translation of ywcA caused approximately 65% of the total increment in tryptophan synthesis.

Second, we focused on the understanding of possible metabolic mechanisms to overcome osmotic stress. We used transcriptomics data of *B. subtilis* growing under salt stress[13] to evaluate the predicted flux distributions at the genome-scale. In accordance with this data, sodium uptake was modeled with no further expression of sodium importers. Interestingly, the regulation of 60% of differentially expressed genes was accurately captured by *i*JT964-ME. The model accurately captures the response of genes associated with main metabolic pathways, such as amino acid synthesis, ribosome formation, and nucleotide synthesis. These pathways are accurately captured since their activity is related to the organism's core metabolic response to stress. Furthermore, our model accurately predicts the arg operon's decrease when the excess sodium enters the cell.

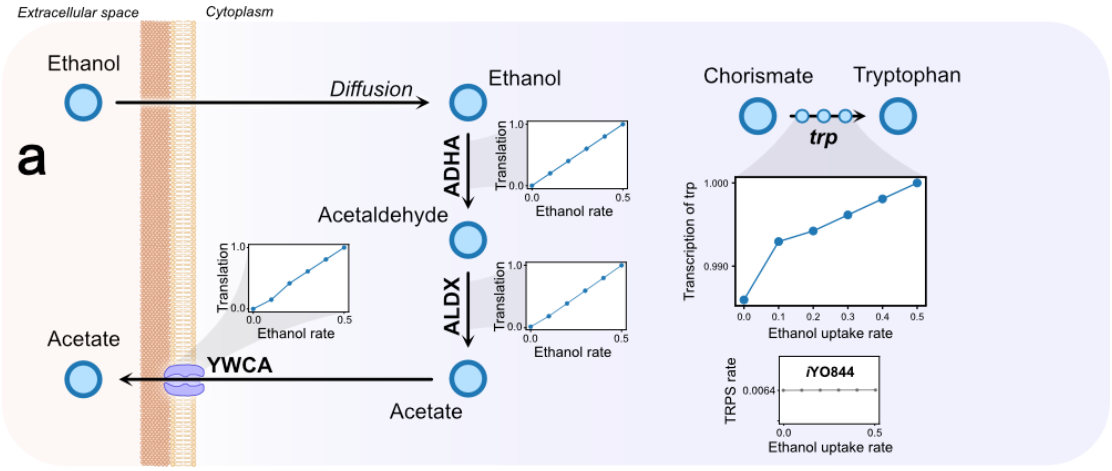

Fig. 2. Response of *B. subtilis* to stress. a Hypothesis for the cause of upregulation as predicted by simulations.

Optimization of protein secretion and identification of critical pathways

First, we tested whether the new features of *i*JT964-ME allowed it to capture amylase secretion shifts at two different growth conditions. While under low growth, the model slightly overestimates the amylase secretion rate, while *i*JT964-ME predicts secretion rates within the reported experimental ranges at a high growth rate (Fig. 3a). We then performed a Principal Component Analysis (PCA) on the simulated data to identify the underlying mechanism. Since both conditions occur at two very different metabolic activities, it is expected that the difference in the fluxes of biomass precursor synthesis, organic carbon assimilation, and energy production describes the most considerable portion of the variance.

We tested what groups of reactions would significantly drive the overexpression of amyE by fixing the growth rate at the lowest of the high growth conditions and forcing amyE overproduction until reaching the highest secretion rate. The two largest components described a strikingly higher portion of the variance (69%). Both components mostly consist of amino acid synthesis reactions. PC1 is described by valine, leucine, isoleucine, alanine, aspartate, arginine, and proline. On the other hand, PC2 consists of glutamate, glycine, serine, tyrosine, tryptophan, and phenylalanine. Interestingly, the weight of these does not directly correlate with the composition of the protein. This indicates that, although their composition in the protein is not nearly as high as alanine, their biosynthetic pathways might pose a bottleneck to target in the overexpression and industrial production of amylase.

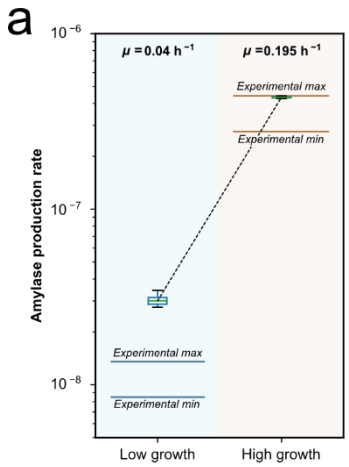

Fig. 3**. Prediction of amylase secretion. a** Prediction of amylase secretion rate at two different growth rates.

Conclusion

The iJT964-ME model represents a significant advancement in the metabolic modeling of *B.subtilis*. Its broad scope and ability to capture expression changes have improved gene essentiality predictions, shed light on recent hypotheses relating amino acid metabolism and stress, and explored the capacity to secrete industrially relevant proteins. This model can serve as the basis for unraveling further questions about metabolism and has the potential to be the foundation on which to optimize heterologous protein expression in this important model organism and cell factory.

Funding

References (No more than 15 references)

1. Oh, Y. K., Palsson, B. O., Park, S. M., Schilling, C. H. & Mahadevan, R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.* **282**, 28791–28799 (2007).
2. O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. & Palsson, B. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* **9**, (2013).
3. Ebrahim, A., Palsson, J. A. L. B. O. & Hyduke, D. R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol.* **7**, 1–6 (2013).
4. Lloyd, C. J. *et al.* COBRAme: A computational framework for building and manipulating models of metabolism and gene expression. *PLoS Comput. Biol.* **14**, 106559 (2018).
5. Yang, L. *et al.* solveME: Fast and reliable solution of nonlinear ME models. *BMC Bioinformatics* **17**, 1–10 (2016).
6. Juhas, M., Reuß, D. R., Zhu, B. & Commichau, F. M. *Bacillus subtilis* and *Escherichia coli* essential genes and minimal cell factories after one decade of genome engineering. *Microbiol. (United Kingdom)* **160**, 2341–2351 (2014).

7. Keseler, I. M. *et al.* The EcoCyc database: Reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* **45**, D543–D550 (2017).
8. Tibocha-bonilla, J. D. *et al.* Dynamic resource allocation drives growth under nitrogen starvation in eukaryotes. *npj Syst. Biol. Appl.* 1–9 (2020). doi:10.1038/s41540-020-0135-y
9. Armingol, E., Tobar, E. & Cabrera, R. Understanding the impact of the cofactor swapping of isocitrate dehydrogenase over the growth phenotype of *Escherichia coli* on acetate by using constraint-based modeling. *PLoS One* 1–21 (2018).
10. Chen, J. *et al.* Enhanced extracellular production of α-amylase in Bacillus subtilis by optimization of regulatory elements and over-expression of PrsA lipoprotein. *Biotechnol. Lett.* **37**, 899–906 (2015).
11. Ma, Y. *et al.* Significantly enhancing recombinant alkaline amylase production in Bacillus subtilis by integration of a novel mutagenesis-screening strategy with systems-level fermentation optimization. *J. Biol. Eng.* **10**, 1–11 (2016).
12. Rychel, K., Sastry, A. V & Palsson, B. O. Machine learning uncovers independently regulated modules in the *Bacillus subtilis* transcriptome. 1–30 (2020).
13. Zhu, B. & Stülke, J. SubtiWiki in 2018: From genes and proteins to functional network annotation of the model organism Bacillus subtilis. *Nucleic Acids Res.* **46**, D743–D748 (2018).