QuantPred

Quantitative Change Prediction for Reaction Networks

PhD Proposal 2022

Director	Joachim NIEHREN	directeur de recherche, Inria Lille, BioComputing
Co-encadrant	Cristian VERSARI	Maître de Conférences, Université de Lille, BioComputing
Co-encadrant	François Coutte	Maitre de Conférence, HDR, Université de Lille INRAE UMR transfrontalière BioEcoAgro (1158)

Abstract

We propose to develop novel bioinformacs methods in area of systems biology to improve the production of microbial biosurfactants with methods from biotechnology. Biosurfactants produced by microorganisms have the potential to replace chemical pesticides in agriculture, to serve for detergency industry and for cleaning petrol on the sea. These applications are relevant for the planet, so it is of highest interest to enable the production of biosurfactants of better quality, in higher quantities, and at lower prices.

Nonribosomal lipopeptides are a class of microbial biosurfactants that can be produced from amino and fatty acids by many bacteria such as Bacillus and Pseudomonas. In our previous interdisciplinary work we have developed successful methods for overproducing surfactine in the bioreactor based on artificial strains of Bacillus subtiis obtained by out-knocking dedicated genes. While the case of single gene knockouts has been largely studied, we believe that multiple gene knockouts are the key for better optimization methods in the future. Multiple gene knockout, however, raise serious difficulties for exhaustive tests in the wet lab, simply since the possibly number is too large.

The hope of model-based prediction methods from systems biology is to limit the number of candidates, so that only few of the many possible candidates need to be tested experimentally. The restriction to purely qualitative reasoning techniques, however, makes it difficult if not impossible to distinguish the good multiple knockouts from the others. Therefore, we propose to add quantitative aspects to existing prediction methods and to the existing biochemical reaction networks of the metabolism of B. subtilis and its control. Of course, this is not generally possible, since only much of the quantitative information on the regulartory control is unknown and will remain unknown in the near future. We therefore propose to integrate known aspects on the forms of the kinetic functions (mass-action, Michaelis-Menten, different kinds of repression, etc), both into the formal models of the metabolic and control networks and into model-based prediction methods.

Contents

A. Organisation

A.1. Interdisciplinary Research Teams

BioComputing team	Université de Lille	Cristal lab
Institut Charles Violette	Université de Lille	INRAE UMR transfrontalière BioEcoAgro (1158)

The INRAE UMR tranfrontalière BioEcoAgro headed by Philippe JACQUES, subsumes several labs of the Haut de France, including the Laboratory Charles Violettes of the University of Lille, and the TERRA lab of the Université de Liège in Gembloux.

- the bioinformatics group BioComputing works systems biology
- our partner Francois Coutte from BioEcoAgro works on biotechnology.

A.2. Funding Plans

Required 50%, Université de Lille

Co-funding 50% ANR Migad possible

A.3. Related Funded Projects

ANR Migad the topic subscribes to aspects the ANR Project Migad (2021-2025) coordinated by Cédric Lhoussaine (BioComputing). Cristian Versari and Joachim Niehren are leading the research on these aspects of Migad.

AI the research proposal promised to develop logical methods from classical artificial intelligence for repesenting the knowledge of biological systems and reasoning with it.

A.4. Priorities of Region Haut de France

CPER BiHautsEco the biotechnology aspect of QuantPred subsribes to the regions CPER on biotechnology for agroeconomics. This CPER is coordinated by the co-director of the proposed PhD thesis, Philippe JACQUES.

CPER CornellIA the computer science aspect of QuantPred subscribes to the regions CPER on artificial intelligence.

Both CPERs are supported by the **ISite** of the Université de Lille.

B. History of the Cooperation

This interdisciplinary cooperation was started by Joachim Niehren and F. Couttte via the postdoc of Mathias John in the BioComputing goup (2011-2013) see [2,3,4,6,7,8]. The cooperation was the continued with on the biotechnology side the PhD theses of Debarun Dhali (2014-2017) in cooperation with F. Coutte see [5], the PhD project of Ariane Theatre (2019-2021) see [9], and on the bioComputing side by the PhD thesis and postdoc of Emily Allart (2016-2021) [-1,0,1]. In 2022, it lead publication at CMSB, the international conference on systems biololyg [-3] and to an interdisciplinar publication at the biotechnology journal Metabolites [-2].
The cooperation was supported through ERA CoBioTech, H2020 project BestBioSurf (2018-2022). Parts of the project subscribe to the ANR project Migad that is coordidated by the head of BioComputing's C. Lhoussaine.

C. State of the Art

Microbial biosurfactants have the potential to replace chemical pesticides in agriculture, chemical counters part in detergency industry, and to serve for cleaning petrol on the sea. All three applications are relevant for the planet, so it is of highest interest to enable the production of biosurfactants of better quality, in higher quantities, and at lower prices. Surfac'n is a nonribosomal lipopeptide that can be produced from amino and fatty acids by many bacteria such as Bacillus and Pseudomonas [14, 15]. Surfactin is one of the most powerful biosurfactants known and it displays several biological activities of interest (antiviral, antimycoplasmic, elicitor, etc). It consists of a ring that carries 7 amino acids of which 4 are leucine and one fatty acid chain of different length and isomers. Is assembled by a non ribosomal mechanism involving multifunc- tional proteins called Non Ribosomal Peptide Synthesis (NRPS). The overproduction of surfactin on [13] was obtained by replacing the native promoter of the surfactin operon (srfA) by a constitu- tive one. This lead to new artifical strains of B. Subtilis, which then are overproducing surfactin in the bioreactor.

Leucine is a branched chain amino acid produced by a metabolic pathway with complex regulatory mechanisms. The case of knockouts of single genes that are relevant for the production of leucine has been studies in own previous work. For this, we formally modelled the reaction network of leu- cine production, and developed novel qualitative reasoning methods based formal methods, that on abstract interpretation [-1,0-4] and constraint programming. We also combined this approach with techniques from flux balance analysis [10, 11, 12]. This way we obtained 12 single knockout predictions for leucine overproduction of which 6 were confirmed by wet lab experiments [5-9]. Of course, these techniques for modelling and prediction are sufficiently general so that they should be useful for quite different applications. But still there also remains much space for improvements concerning he quantify, quality, and price of the surfactin production, that we would like to explore.

D. Resaerch Project

Challenge. We believe that multiple gene knockouts are the key to fetter optimization me- thods in the future. Multiple gene knockout, however, are very difficult for exhaustive tests in the wet lab, simply since the possibly number too large. The hope of model-based prediction methods from systems biology is to limit the number of candidates, so that only few of the many possible candidates need to be tested experimentally. The restriction to purely qualitative reasoning techniques, however, makes it difficult if not impossible to distinguish the good multiple knockouts from the others.

Approach. Therefore, we propose to add quantitative aspects to existing prediction methods and to the existing biochemical reaction networks of the metabolism of B. subtilis and its control. Of course, this not generally possible, since only much of the quantitative information on the regulato- ry control is unknown and will remain unknown in the near future. We therefore propose to integrate known aspects on the forms of the kinetic functions (mass-action, Michaelis-Menten, different kinds of transcriptomic induction or repression, etc), both into the formal models of the metabolic and control networks and into model-based prediction methods. Our approach to make quantitative predictions will be to interval based methods instead of purely qualitative reasoning on increases and decreases.The intervals will abstract from size of changes. In particular, when it comes to Michaelis-Menten kinetics, one needs to know whether one is close to saturation or far away, in order to estimate the size of increase or decreases. How this kind of knowledge can be incorporated into the models and analysis is the topic of the present PhD pro- ject.

Working program

Year 1

In the first 6 month, the student has to learn the state of the art, methodologically and technologically. Therefore

A. we will revisit and improve the existing qualitative prediction methods based on qualitative reasoning, so that that they become more precise. This will be tested on concrete challenges, where the models are already existing in the biocomputing group (leucine model, aspartate mo- del). In the next 6 month we will work on parallel on modelling and analysis methods in parallel

B. We will study small parts of reaction network of leucine production, to see, which kind of quantitative information on the kinetics of the metabolic control could be added to the existing qualitative models, talking about activation and inhibition only

C. We will develop our ideas towards intervals-based methods for qualitative prediction that can be applied to the example models from Year 1.B).

Year 2

We will ontinue with the 2 lines of the second half of the first year in parallel:

A. During a visit of the TERRA lab in Gembloux Belgium (with Philippe Jacques), the student will learn about biological modelling problems in the context of surfactine overproduction.

B. We will define a general modeling language that supports reaction networks with partial kinetic information and sufficient quantitative information to captures the example studied in year 1B).

C. We will develop and implement a general interval-based method for qualitative prediction of single knockouts and apply the to larger metabolic networks, that can be applied to all models of the language from Year 2.B).

Year 3

A. We will extend the prediction methods for single knockouts from Year 2c) to multiple gene kno- ckouts and its application to a model of leucine and asparate overproduction.

B. the student will write up his PhD thesis.

E. References

E.1. On previous work on bioinformatics and systems biology

[-3] **J .Niehren**, A. Vaginay, **C. Versari**.: Abstract Simulation of Reaction Networks via Boolean Networks. CMSB: 20th International Conference on Computational Methods in Systems Biology. Springer. 2022
[-2] J.S. Guez, **J. Niehren**, **F. Coutte**, P. Jacques et al.: Bioinformatics modelling and metabolic engineering of the branched chain amino acid pathway for specific production of mycosubtilin isoforms in Bacillus subtilis. Metabolites, 2022.
[-1] E. Allart, **J. Niehren, C. Versari**: Computing Difference Abstractions of Linear Equation Systems. Theoretical Computer Science, Elsevier, volume 893: 72-104, 2021.
[0] Allart E, **Niehren J, Versari C**. Exact Boolean Abstraction of Linear Equation Systems. MDPI Computation. 9(11):113, 2021.
[1] E. Allart, **C. Versari, J. Niehren**: Computing Difference Abstractions of Metabolic Networks Un- der Kinetic Constraints. 17th International Conference on Computational Methods in Systems Bio- logy (CMSB), Springer, 2019.
[2] **J. Niehren, C. Versari**, M. John, F. Coutte, **P. Jacques**: Predicting Changes of Reaction Net- works with Partial Kinetic Information. BioSystems, 2016.
[3] **J. Niehren**, M. John, **C. Versari**, F. Coutte, **P. Jacques**: Qualitative Reasoning about Reac- tion Networks with Partial Kinetic Information. Computational Methods for Systems Biology (CMSB), Springer, 2015
[4] M. John, M. Nebut, **J. Niehren**: Knockout Prediction for Reaction Networks with Partial Kinetic Information. 14th International Conference on Verification, Model Checking, and Abstract Interpre- tation, Springer, 2013.

E.2. Own previous work on biochemistry and biotechnology

[5] D. Dhali, F. Coutte, A. Argüelles Arias, S. Auger, V. Bidnenko, G. Chataigné, M. Lalk, **J. Niehren**, J. de Sousa, **C. Versari, P. Jacques**: Genetic engineering of the branched fatty acid metabolic pa- thway of Bacillus subtilis for the overproduction of surfactin C14 isoform. Biotechnology Journal, 2017.
[6] F. Coutte, **J. Niehren**, D. Dhali, M. John, **C. Versari, P. Jacques**: Modeling Leucine's Metabolic Pathway and Knockout Prediction Improving the Production of Surfactin, a Biosurfactant from Ba- cillus Subtilis. Biotechnology Journal, 2015.
[7] F. Coutte, M. John, M. Béchet, M. Nebut, **J. Niehren**, V. Leclère, P. **Jacques**: Synthetic Enginee- ring of Bacillus subtilis to Overproduce Lipopeptide Biosurfactants. 9th European Symposium on Biochemical Engineering Science, 2012.
[8] M. John, F. Coutte, M. Nebut, **P. Jacques, J. Niehren**: Knockout Prediction for Reaction Net- works with Partial Kinetic Information: Application to Surfactin Overproduction in Bacillus subtilis. 3Ird International Symposium on Antimicrobial Peptides, 2012.
[9] A. Théatre, C. Cano-Prieto, M. Bartolini, Y. Laurin, M. Deleu, **J. Niehren**, T. Fida, S. Gerbinet, M. Alanjary, M. Medema, A. Léonard, L. Lins, A. Arabolaza, H. Gramajo, H. Gross, **P. Jacques**: The surfactin-like lipopeptides from Bacillus spp.: natural biodiversity and synthetic biology for a broader application range. Frontiers in Bioengineering and Biotechnology, 2021.

E.3. External previous work on systems biology

[10] E. Almaas, B. Kovacs, T. Vicsek, Z. N. Oltvai, and A. L. Barabasi. Global organization of meta- bolic fluxes in the bacterium Escherichia coli. Nature, 427:839–843, 2004.
[11] Markus W Covert and Bernhard O Palsson. Constraints-based models: regulation of gene ex- pression reduces the steady-state solution space. J. Theor. Biol., 221(3):309–325, 2003.
[12] Lope A Flórez, Katrin Gunka, Rafael Polanía, Stefan Tholen, and Jörg Stülke. Spabbats: A pathway-discovery method based on boolean satisfiability that facilitates the characterization of suppressor mutants. BMC Syst. Biol., 5(1):5, 2011. Previous publications of our biotechnology partner
[13] 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 expres- sion of srfa on surfactin productivity, spreading and antagonistic properties of Bacillus subtilis 168 derivatives. J. Appl. Microbiol., 109(2):480–491, 2010.
[14] P. Jacques. Surfactin and other lipopeptides from Bacillus spp. In Biosurfactants, pages 57– 91. Springer, 2011.s for the efficient production of a lipopeptide pseudofactin by Pseudomonas fluorescens BD5. Microbial cell factories, 17(1), 121, 2018.
[15] P Biniarz, F Coutte, F Gancel and M Łukaszewicz. High-throughput optimization of medium components and culture conditions.