Proposal for PhD thesis
MAP-i: doctoral program in computer science

Thesis title:
Evaluation and Development of Algorithms and Computational Tools for Metabolic Pathway Optimization

Thematic area:
Bioinformatics

Supervisors:
Miguel Francisco de Almeida Pereira da Rocha
Associate Professor
Department of Informatics, School of Engineering, University of Minho

Co-supervisor:
Isabel Cristina de Almeida Pereira da Rocha
Auxiliar Professor
Department of Biological Engineering, School of Engineering, University of Minho

External researcher
Joel Perdiz Arrais
Auxiliar Professor
Department of Informatics Engineering, University of Coimbra

Research Unit:
CEB- Centre of Biological Engineering, School of Engineering, University of Minho

Brief description of the research theme:

Metabolic Engineering (ME) addresses the optimization of living organisms' metabolic processes, by genetically modifying them to achieve a certain goal, such as increasing production of a desired compound or to allow production of a new product. Over the last few years, computational tools have been used as a way to improve ME methods by computational screening of strains and strain optimization based on computer simulations. Indeed, in the past years, biological data increased at an exponential rate, information became even more accessible with the development of new technologies, making *in silico* simulation of living and synthetic systems more attractive [1], leading to the development of sophisticated mathematical models capable of predicting cellular behavior. Towards this end, recent efforts allowed the development of genome-scale metabolic models for several organisms, taking advantage on the growing number of sequenced genomes and the availability of Bioinformatics tools for automatic genome annotation [2]. These stoichiometric models allowed the study of the metabolism of different organisms through several methods for phenotype simulation and structural analysis.

A way to look at these models is to consider biochemical reactions performed inside of a living cell through a series of biological pathways that can be computationally modeled with graph-based representations, including variants such as bipartite graphs, hypergraphs [3] or process graphs [4]. Over these types of representation, a number of sophisticated algorithms related to pathway optimization have been proposed [3,4]. The use of such approaches to design novel pathways towards the production of interesting compounds in ME has not yet been fully explored in the literature. Indeed, although there are a few methods for the selection of new reactions to add to existing metabolic models [5], more sophisticated methods for ME using graph-based approaches are still lacking. Indeed, current approaches face a number of problems related to computational efficiency (both in terms of speed and memory consumption), scalability and data integration. These present major hurdles to the adoption of these algorithms by the ME community.

The problem to be addressed in this work is the validation and improvement of existing algorithms and tools, together with the development of novel ones when deemed appropriate, allowing the optimization of the best pathways to overproduce interesting compounds based on metabolic networks. Some of the underlying scientific / technological goals are the following:

- To build metabolic networks integrating distinct data sources including metabolic databases (e.g. KEGG or MetaCyc) or metabolic models, allowing flexible user defined filters to be applied and contemplating information related to the metabolic capabilities of each organism of interest. These should be represented using graph-based representations, or adequate variants.
- To implement, evaluate and improve existing optimization algorithms that allow searching over these metabolic networks for the best routes from sets of source metabolites to destination metabolites, given the specificities of the underlying representation and being able to optimize these paths according different criteria. When necessary novel algorithms should be designed and implemented. To address computational efficiency issues, the parallelization of these algorithms will be addressed.
- To design and implement algorithms that allow the selection of the most suitable organisms to perform specific biochemical transformations, according to different optimization goals.
- To integrate the algorithms implemented with available software under the OptFlux platform for metabolic engineering.
- To validate the proposed algorithms with several case studies.

[1] M. Tomita, "Whole-cell simulation: a grand challenge of the 21st century.," Trends in Biotechnology, pp. 19: 205-210, 2001.
[2] A.M. Feist et al (2009) Reconstruction of biochemical networks in microorganisms. Nat. Rev. Microbiol., 7:129-143.
[3] P. Carbonell et al. "Enumerating metabolic pathways for the production of heterogeneous target chemicals in chassis organisms". BMC Systems Biology. 6(1), 2012
[4] F. Friedler et al. A tool for consistent and complete decisions in process synthesis. Chemical Eng Science, 50(11): 1755-1778, 1995
[5] P. Pharkya et al, "OptStrain: A computational framework for redesign of microbial production systems", Genome Research, vol. 14, pp. 2367--2376, 2004.