OptGene – a framework for *in silico* metabolic engineering

Isabel Rocha, Paulo Maia, Miguel Rocha, Eugénio C. Ferreira
[1]IBB – Institute for Biotechnology and Bioengineering, Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710–057 Braga, Portugal
[2]Department of Informatics/CCTC, University of Minho, Campus Gualtar, 4710-057 Braga, Portugal

Keywords: Metabolic Engineering, Evolutionary Algorithms, Flux Balance Analysis, Genome-scale stoichiometric models

Introduction

In metabolic engineering problems, due to the complexity of metabolic networks, it is often difficult to identify *a priori* which genetic manipulations will originate a given desired phenotype. Genome-scale metabolic models (Patil et al., 2004), available for several microorganisms, can be used to simulate the metabolic phenotype and therefore help the tasks of metabolic engineering. This simulation can be performed by calculating the fluxes through all metabolic reactions using techniques like the Flux Balance Analysis (FBA) (Edwards et al, 2002) or the MOMA (Segre et al., 2002 approaches, among others.
Several algorithms, like the framework OptKnock (Burgard et al., 2003) have been developed that use metabolic models to enable the identification of gene knockout strategies for obtaining improved phenotypes. However, the problem of finding optimal gene deletion strategy is combinatorial and consequently the computational time increases exponentially with the size of the problem.

OptGene Algorithms and Software

We have proposed the use of bio-inspired algorithms, namely Evolutionary Algorithms (EAs) and Simulated Annealing for the optimisation of the set of gene deletions to apply to a microorganism, in order to maximize a given objective function, associated with a given industrial objective (Patil et al., 2005) and (Rocha et al., 2007). The proposed algorithms enable solving large gene knockout problems in relatively short computational time and also allow the optimization of non-linear objective functions and additionally provide a family of close to optimal solutions.
Each mutant strain is evaluated by resorting to the simulation of its phenotype using the Flux-Balance Analysis approach, together with the premise that microorganisms have maximized their growth along natural evolution (Ibarra et al., 2002). The application of this methodology to several case-studies allowed to identify directed genetic modifications for the construction of improved bacterial and yeast strains for the production of valuable chemical compounds such as succinate. The results suggest that non-intuitive genetic modifications spanning several different pathways may be necessary for solving challenging metabolic engineering problems.
Based on these validated approaches, the user-friendly software framework *OptGene* was designed and developed that includes a number of tools to support *in silico* metabolic engineering. The application allows the user to load a genome-scale stoichiometric model of a given organism. This serves as the basis to simulate the phenotype of the wild type and of mutant strains. The simulation of these strains is conducted using a number of approaches (e.g. Flux-Balance Analysis or MOMA) that allow the set of fluxes in the organism's metabolism to be determined, given a set of environmental constraints. The core of the software tool is, however, the identification of metabolic engineering targets, performed by a number of optimization methods that allow to reach the best set of gene deletions given an objective function.
This application is complemented by an independent visualization tool, named *BioVisualizer*, that allows the visualization of graphs representing metabolic networks. These graphs have a number of distinct node types (e.g. metabolites, enzymes,

reactions) and connections. One of the major features of this tool is the ability to associate numerical values to the different types of nodes and edges in the graph.
The integration of the two applications allows the visualization of the metabolic network, superimposed by the values of the fluxes of a given simulation. The values of the fluxes for the wild type and different mutants can, therefore, be adequately visualized in this way, providing useful analysis tools for the researchers.
Additionally, this software is compatible with several SBML (Systems Biology Markup Language) standards (Hucka et al., 2003), allowing the use of models stored in public databases or built in other software tools both for simulation and visualization.
The *OptGene* application is being developed taking as a basis the *AIBench* framework. This is an environment for the development of Data Mining / Bioinformatics tools, using the Java programming language. The details of this project, a collaboration between the universities of Minho (Portugal) and Vigo (Spain), as well as updated documentation can be found at the web site www.aibench.org.

References

Burgard,A.P., Pharkya,P., and Maranas,C.D. (2003) OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering* **84** (6), 647-657.

Edwards,J.S., Covert,M., and Palsson,B. (2002) Metabolic modelling of microbes: the flux-balance approach. *Environmental Microbiology* **4** (3), 133-140.

Hucka,M. et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19** (4), 524-531.

Ibarra,R.U., Edwards,J.S., and Palsson,B.O. (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420** (6912), 186-189.

Patil,K.R., Akesson,M., and Nielsen,J. (2004) Use of genome-scale microbial models for metabolic engineering. *Current Opinion in Biotechnology* **15** 1-6.

Patil,K.R., Rocha,I., Forster,J., and Nielsen,J. (2005) Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics* **6**.

Rocha,M., Mendes,R., Maia,P., Pinto,J.P., Rocha,I., and Ferreira,E.C. (2007) Evaluating simulated annealing algorithms in the optimization of bacterial strains. *Lecture Notes in Artificial Intelligence* **4874** 473-484.

Segre,D., Vitkup,D., and Church,G.M. (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America* **99** (23), 15112-15117.