PATHWAY ANALYSER

A Systems Biology Tool for Flux Analysis of Metabolic Pathways

Karthik Raman and Nagasuma Chandra

Supercomputer Education and Research Centre, Bioinformatics Centre
Indian Institute of Science, Bangalore – 560 012, INDIA
E-mail: karthik@rishi.serc.iisc.ernet.in, nchandra@serc.iisc.ernet.in

Nature Precedings : doi:10.1038/npre.2008.1868.1 : Posted 8 May 2008

Abstract

Stoichiometric and constraint-based analyses of metabolic pathways has been gaining ground in the recent past with the increase in the quality and number of pathway databases available and the curation of genome-scale metabolic models. Genome-scale metabolic models of several organisms such as *Escherichia coli*, *Saccharomyces cerevisiae* and *Staphylococcus aureus* have already been constructed. Flux Balance Analysis (FBA) and Minimisation of Metabolic Adjustment (MoMA) are two of the popular techniques for the constraint-based analysis of metabolic pathways.

We have developed a computational tool, PathwayAnalyser, for the analysis of metabolic pathways, particularly by FBA and MoMA. PathwayAnalyser interfaces with the open-source GNU Linear Programming Toolkit (GLPK) for linear programming/ FBA and Object Oriented Quadratic Programming (OOQP) for quadratic programming/MoMA. It gives a comprehensive report on gene deletions from the Systems Biology Markup Language (SBML) Model and objective function input for FBA. PathwayAnalyser is open-source and is available at http://sourceforge.net/projects/pathwayanalyser.

Introduction

The availability of annotated genome sequences of several organisms has led to a fundamental change in the way biochemical systems are analysed. Based on this reductionist information, the metabolic networks of several organisms have been reconstructed [1,2]. The imposition of constraints on a reconstructed biochemical network leads to the definition of achievable cellular functions. Several computational methods have been developed to study the characteristics and capabilities of microorganisms using constraint-based methods. Flux Balance Analysis (FBA) is one such constraint-based method, for the analysis of biochemical systems [3,4]. Minimisation of Metabolic Adjustment (MoMA) is a similar method, that is based on quadratic programming [5]. FBA and MoMA have been extensively used for the analysis of microbial models for various applications, from metabolic engineering [6] to drug target identification [7]. In this paper, we describe PathwayAnalyser, which is a tool for performing FBA and MoMA on reconstructed biochemical networks.

FBA involves carrying out a steady state analysis, using the stoichiometric matrix for the system in question. The system is assumed to be optimised with respect to functions such as maximisation of biomass production or minimisation of nutrient utilisation, following which it is solved to obtain a steady state flux distribution. This flux distribution is then used to interpret the metabolic capabilities of the system. The dynamic mass balance of the metabolic system is described using the stoichiometric matrix, relating the flux rates of enzymatic reactions, $\mathbf{v}_{nx1}$ to time derivatives of metabolite concentrations, $\mathbf{x}_{mx1}$ as

$$d\mathbf{x}/dt = \mathbf{S}\cdot\mathbf{v} = \mathbf{0} \text{ (at steady state)}$$

Therefore, the required flux distribution belongs to the null space of **S**. Since m (metabolites) < n (reactions), the system of equations is under-determined. To obtain a single solution, further biological constraints are imposed on the system, followed by the specification of a biologically relevant objective function for optimisation. Typically, the system is assumed to be optimised with respect to metabolic functions such as biomass production (max) or nutrient utilisation (min), following which it is solved to obtain a steady state flux distribution. This flux distribution is then used to interpret the metabolic capabilities of the system. The problem now is one of linear programming:

$$\min_{\mathbf{v}} \langle \mathbf{c}\cdot\mathbf{v} \rangle \text{ s.t. } \mathbf{S}\cdot\mathbf{v} = \mathbf{0}$$

where **c** represents the objective function composition, in terms of the fluxes. Further, we can constrain the internal fluxes to be strictly positive and the external fluxes to be unbounded. Other constraints may also be added on the basis of the knowledge of cellular physiology.

The technique MoMA circumvents the use of an objective function for the analysis of mutants. Instead, MoMA provides an approximate solution for a sub-optimal growth flux state, which is nearest in flux distribution to the unperturbed state. The mathematical formulation of this yields a quadratic programming problem:

$$\min ||\mathbf{v}_w - \mathbf{v}_d||^2 \text{ s.t. } \mathbf{S}\cdot\mathbf{v} = \mathbf{0}$$

where $\mathbf{v}_w$ represents the wild-type (or unperturbed state) flux distribution and $\mathbf{v}_d$ represents the flux distribution on gene deletion that is to be solved for. This simplifies to a standard quadratic programming problem:

$$\min \ (\tfrac{1}{2})\mathbf{v}_d^T \mathbf{I}\, \mathbf{v}_d + (-\mathbf{v}_w)\cdot\mathbf{v}_d \text{ s.t. } \mathbf{S}\cdot\mathbf{v}_d = \mathbf{0}$$

where **I** is an identity matrix of size $n \times n$, n being the length of the vector $\mathbf{v}_d$. An important feature of MoMA is that the wild-type flux distribution used need not be obtained by performing an FBA; an experimentally determined flux distribution could serve better. Thus, objective functions for optimisation, which at times may not reflect the physiological situation very accurately, can be circumvented using MoMA.

Implementation

* Written in C++; can be compiled to run on any linux-based platform
* PathwayAnalyser accepts SBML files as input — Systems Biology Markup Language (SBML) [8] is the preferred medium for the encoding and exchange of systems biology models
* No accepted standard for specifying flux constraints and objective functions in SBML; hence, these are input as separate files
* Stoichiometric matrix is stored in memory-efficient co-ordinate sparse format
* FBA (linear programming problem) solved using the interior point method from the GNU Linear Programming Toolkit (GLPK)
* MoMA (quadratic programming problem) solved using Object-Oriented Quadratic Programming (OOQP) library [9]
* PathwayAnalyser outputs:
  * fluxes for the wild type and the various *in silico* deletion mutants
  * a gene deletion report based on the input objective function
* Criterion for essentiality: objective function evaluates to <5% of that for the wild-type
* For MoMA, the wild-type distribution is currently calculated by performing an FBA, although this can be altered if a suitable experimental distribution is available

Results and Discussion

We illustrate the capabilities of PathwayAnalyser with the example of two pathway models in *M. tuberculosis*. PathwayAnalyser can be used to perform FBA/MoMA on these models and subsequently analyse gene essentiality.

Model Composition	Carbohydrate Metabolism	Mycolic Acid Pathway [7]
Genes	59	28
Reactions	111	247
Metabolites	56	197
Objective function	*Various*	*Mycolate production*

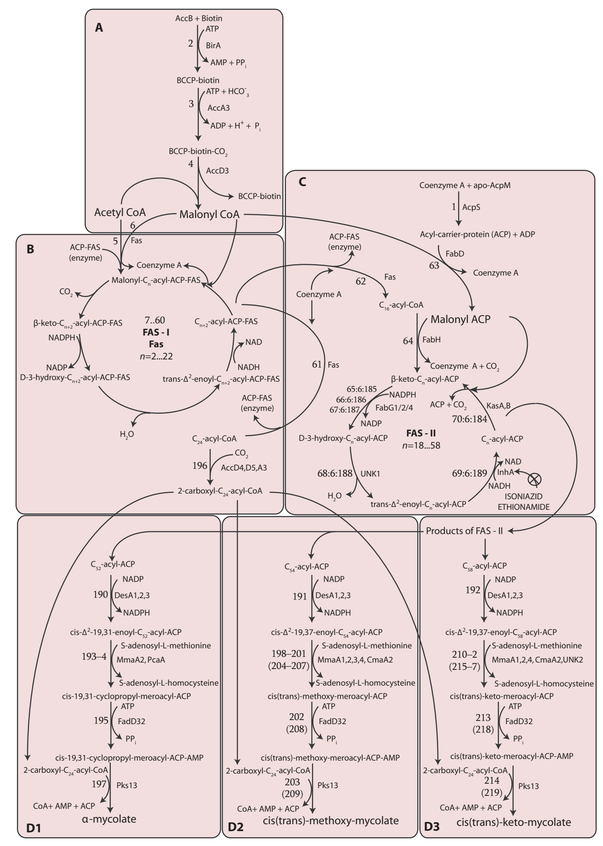
Carbohydrate Metabolism

The model of carbohydrate metabolism in *M. tuberculosis* was reconstructed from the KEGG database. The model includes most of the major carbohydrate metabolism pathways such as glycolysis, gluconeogenesis, citric acid cycle and the pentose phosphate pathway. For FBA, various objective functions were used, such as
* Minimisation of ATP utilisation
* Maximisation of pyruvate production
* Minimisation of glucose uptake

In all cases, the deletions were not found to be lethal, since most of the metabolites have an alternate route of production.

Mycolic Acid Pathway

This is based on the model published in [7]. The figure indicates a schematic of the mycolic acid pathway in *M. tuberculosis*.



Sample output from **PA_FBA** is indicated below:

```
Gene Deletion Report for MAP.sbml (MAP)

Mutant          Function Value  Essentiality

Wild            -0.012963       -
acpS            -0.012963       NON-ESSENTIAL
birA            -0.012963       NON-ESSENTIAL
accA3           -0.000000       ESSENTIAL
accD3           -0.000000       ESSENTIAL
fas             -0.000000       ESSENTIAL
fabD            -0.000000       ESSENTIAL
fabH            -0.000000       ESSENTIAL
fabG1           -0.012963       NON-ESSENTIAL
...
```

The proteins listed as essential in this output may be viable as drug targets.

Conclusions

* PathwayAnalyser performs FBA and MoMA based on the objective function for optimisation and other constraints on the system
* Produces a gene deletion report detailing gene essentiality
* Also outputs the flux distribution for the wild type and the *in silico* deletion mutants.
* It is an easy-to-use command-line tool for the analysis of reconstructed metabolic networks.

References

[1] Edwards JS, Palsson BØ (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A* **97**:5528–5533.
[2] Forster J, Famili I, Fu P, Palsson BØ, Nielsen J (2003) Genome-Scale Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network. *Genome Res* **13**:244–253.
[3] Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. *Curr Opin Biotechnol* **14**:491–496.
[4] Lee JM, Gianchandani EP, Papin JA (2006) Flux balance analysis in the era of metabolomics. *Brief Bioinform* **7**:140–150.
[5] Segre D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks *Proc Natl Acad Sci U S A* **99**:15112–15117.
[6] Alper H, Jin YS, Moxley J, Stephanopoulos G (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Engg* **7**:155–164.
[7] Raman K, Rajagopalan P, Chandra N (2005) Flux balance analysis of Mycolic Acid Pathway: targets for anti-tubercular drugs. *PLoS Comput Biol* **1**:e46.
[8] Hucka M, Finney A, Bornstein B, Keating S, Shapiro B *et al* (2004) Evolving a *lingua franca* and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *Syst Biol* **1**:41–53.
[9] Gertz EM, Wright SJ (2003) Object-oriented software for quadratic programming. *ACM Trans Math Softw* **29**:58–81.