A Constraint Based Method for Optimization in Metabolic Pathways

Mouli Das[1], Subhasis Mukhopadhyay[2], and Rajat K. De[1]

[1] Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700 108, India
{mouli_r,rajat}@isical.ac.in
[2] Department of Bio-Physics, Molecular Biology and Bioinformatics, Calcutta
University, Kolkata 700 009, India
smbmbg@caluniv.ac.in

Abstract. The analysis of pathways in metabolic networks has become an important approach for understanding metabolic systems. Constraint-based approaches recently brought new insight into our understanding of metabolism. Here we introduce a method for identifying optimal metabolic pathways. The method, generates data on reaction fluxes based on biomass conservation constraint and formulates constraints incorporating weighting coefficients corresponding to concentration of enzymes catalyzing the reactions. The effectiveness of the present method is demonstrated on the pentose phosphate pathway (PPP) in *T. cruzi* and on a large network of the plant polyamine pathway existing in the literature. A comparative study with the existing extreme pathway analysis (EPA) also forms a part of this investigation.

Keywords: FBA, stoichiometric matrix, polyamine pathway.

1 Introduction

Cellular metabolism, the integrated interconversion of thousands of metabolic substrates into products through enzyme-catalysed biochemical reactions, is the most investigated complex intracellular web of molecular interactions. Metabolic pathways are defined by a series of chemical reactions that are interlinked with each other [1]. Metabolic pathway modeling is needed as metabolism is the "chemical engine" that drives the living process, and is a fundamental determinant of cell physiology. The flux balance analysis (FBA), a constraint-based approach applied to genome-scale metabolic models can be used to make predictions of flux distributions based on linear optimization [2]. Here we describe a method based on the framework of FBA for identifying optimal metabolic pathways that yields the maximum amount of the target metabolite from a given substrate. Results, comparison with the existing EPA [3] and their biological relevance have been provided for the PPP in *T. cruzi* and on one large polyamine biosynthesis pathway observed in plants.

2   A Constraint Based Method

A metabolic network consists of internal fluxes v within the network and exchange fluxes b that can either enter or exit the system. The rate of growth of the target metabolite B on the starting metabolite A, which needs to be maximized is the algebraic sum of the weighted fluxes of $R_1, R_2, \ldots, R_s$ reactions by s different paths and is given by

$$z = \sum_{k=1}^{s} c_k v_k \tag{1}$$

c_k indicates the enzyme concentration level. The flux vectors, $\mathbf{v}$ that satisfy approximately the quasi-steady state condition $\mathbf{S}.\mathbf{v} \approx \mathbf{0}$ and the inequalities described later are generated. $\mathbf{S}$ is the $m \times n$ stoichiometric matrix with m as the number of metabolites and n as the number of reactions. We generate l number of basis vectors $\mathbf{v}_b$ that form the null space of $\mathbf{S}$ and also generate l number of non-negative random numbers a_p, $p = 1, 2, \ldots, l$ to generate a vector $\mathbf{v} = \sum_{p=1}^{l} a_p \mathbf{v}_{bp}$ satisfying the following inequality constraints. All the internal fluxes are positive yielding: $v_i \geq 0, \forall i$. The constraints on the exchange fluxes depending on their direction can be expressed as $\alpha_j \leq b_j \leq \beta_j$ where $\alpha_j \epsilon \{-\infty, 0\}$ and $\beta_j \epsilon \{0, \infty\}$. In real systems the genes may not be expressed at the required level. So we define a new set of constraints as

$$\mathbf{S}(\mathbf{Cv}) = \mathbf{0} \tag{2}$$

where $\mathbf{C}$ is an $n \times n$ diagonal matrix whose i-th diagonal element is c_i for each i. That is, if $\mathbf{C} = [\gamma_{ij}]_{n \times n}$, then $\gamma_{ij} = \delta_{ij} c_i$, where δ_{ij} is the Kronecker delta. So the problem boils down to a maximization problem, where z is maximized with respect to $\mathbf{c}$, subject to satisfying the constraint given in equation(2) along with the inequality constraints. Combining equations (1) and (2), we can reformulate the objective function as

$$y = 1/z + \mathbf{\Lambda}^T.(\mathbf{S}.(\mathbf{C}.\mathbf{v})) \tag{3}$$

that has to be minimized with respect to the weighting factors c_i for all i. The term $\mathbf{\Lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_m]^T$ is the regularizing parameter. Initially, a set of random values in $[0, 1]$ corresponding to c_i's are generated which are then modified iteratively by gradient descent optimization algorithm, where $\Delta c_i = -\eta \partial y / \partial c_i$. The term η is a small positive quantity indicating the rate of modification. The modified value of c_i is given by $c_i(t + 1) = c_i(t) + \Delta c_i$, $\forall i$, $t = 0, 1, 2, \ldots$ λ is chosen empirically from 0.1 to 1.0 in steps of 0.1. For each value of λ as we are increasing the number of iterations, the value of y gradually decreases and the corresponding c_i-values indicate an optimal pathway.

3   Results and Comparison

When applied to various organisms, the constraint based method yields more significant and biologically relevant results as compared to that of EPA. Here,
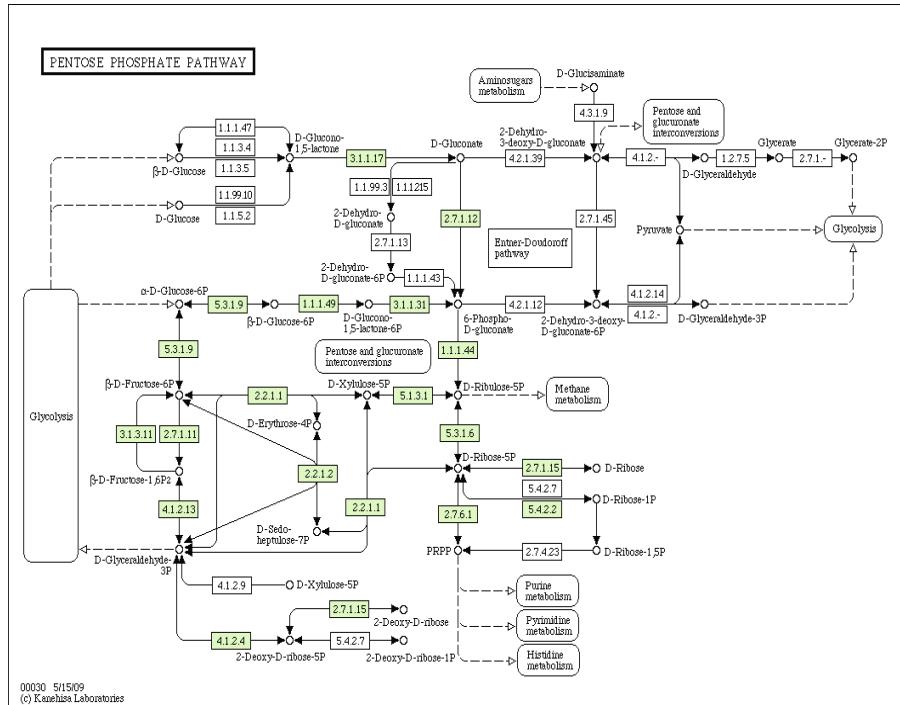
Fig. 1. Pentose Phosphate Pathway in *T. cruzi*

we apply our method to the PPP in the parasite *T. cruzi* and on one large plant polyamine pathway obtained from the KEGG database [1].

PPP is an anabolic pathway that utilizes 6 carbons of glucose to generate NADPH (reduced form of NADP+) and pentose(5-carbon) sugars. The PPP is a key metabolic process in the cell because it provides reductive power, ribose-5-phosphate and glycolytic intermediates. The pathway has an oxidative phase that generates NADPH, and the second non-oxidative phase synthesizes 5-carbon sugars [4]. The biochemical evidence obtained so far suggest that the oxidative branch is essential for protection of the parasite against oxidative stress. The PPP has been shown to be functional in T. cruzi and the seven enzymes of the pathway have been expressed and fully characterized. The balance between the 2 branches is necessary to maintain the metabolic efficiency of the cancer cell for growth and proliferation [5]. The pathway consists of 18 metabolites and 33 fluxes (Fig. 1) where the starting and target metabolites are α-D-Glucose-6P, D-Glyceraldehyde-3P and β-D-Fructose-6P respectively. The present method generates the optimal pathway as $\alpha-D-Glucose-6P \rightarrow \beta-D-Glucose-6P \rightarrow D-Glucono-1,5lactone-6P \rightarrow 6-Phospho-D-Gluconate \rightarrow D-Ribulose-5P \rightarrow D-Xylulose-5P + D-Erythrose-4P \rightarrow D-Glyceraldehyde-3P + \beta-D-Fructose-6P$. The EPA method generates a different pathway
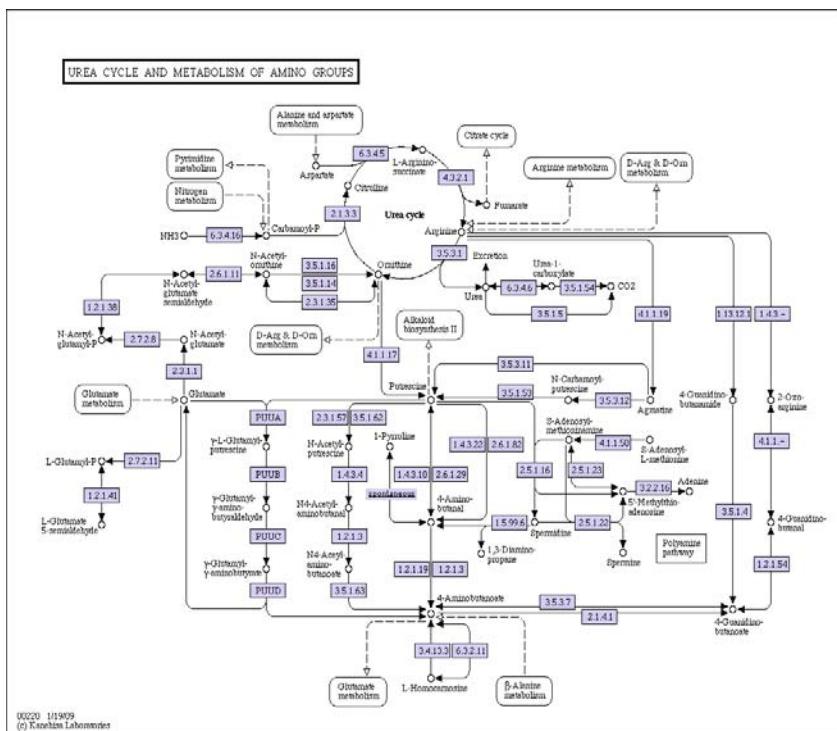
Fig. 2. Polyamine pathway

as $\alpha - D - Glucose - 6P \rightarrow \beta - D - Fructose - 6P \rightarrow D - Xylulose - 5P +$
$D - Erythrose - 4P \rightarrow D - Glyceraldehyde - 3P + \beta - D - Fructose - 6P.$

Polyamines (PAs) are naturally occurring low molecular weight, polycationic, aliphatic nitrogenous compounds found in all organisms and are essential for cellular proliferation and normal cellular function. The most common PAs in plants are spermidine, spermine and their precursor putrescine [6]. PAs take part in many important cellular processes such as cell division, protein synthesis, DNA replication and response to abiotic stress. The polycationic nature of PAs is one of the main properties believed to mediate their biological activity. They are, indeed, able to bind several negatively charged molecules, such as DNA, proteins, membrane phospholipids and proteins. Polyamine biosynthetic pathways in plants have been well elucidated [7]. The intermediate metabolite putrescine is formed either by direct decarboxylation of L-ornithine by the enzyme ornithine decarboxylase (ODC, EC 4.1.1.17), or by decarboxylation of arginine by arginine decarboxylase (ADC, EC 4.1.1.19) via agmatine [8]. Spermidine and the target metabolite spermine are synthesized by the sequential addition of an aminopropyl group to putrescine by spet-midine synthase and spermine synthase, respectively. The aminopropyl group is donated by decarboxylated S-adenosylmethionine (SAM), which is produced by S-adenosylmethionine decarboxylase (SAMDC, EC 4.1.1.50). There are 66 reactions and 42 metabolites

for the pathway (Fig. 2) where we are maximizing the yield of Spermine from the starting metabolite Arginine. The present method generates the optimal paths as follows: $Arginine \rightarrow Ornithine \rightarrow Putrescine \rightarrow Spermidine \rightarrow Spermine$ and $Arginine \rightarrow Agmatine \rightarrow Putrescine \rightarrow Spermidine \rightarrow Spermine$. The EPA method results in $Arginine \rightarrow Agmatine \rightarrow Putrescine \rightarrow Spermidine \rightarrow Spermine$ as an optimal pathway. It has been demonstrated in various plant tissues that the pathway obtained by our method (ornithine decarboxylase pathway) is active in cell proliferation and the the pathway obtained by EPA (arginine decarboxylase pathway) is involved in embryo and organ differentiation [9].

4   Biological Relevance and Validation

There are 2 paths starting from $\alpha - D - Glucose - 6P$ in the PPP in *T. cruzi* (Fig. 1) leading to formation of $\beta - D - Glucose - 6P$ and $\beta - D - Fructose - 6P$ in the 2 branches. The path through $\beta - D - Glucose - 6P$ is followed till the intermediate metabolite $D - Ribulose - 5P$ which gets divided into 2 paths of which the one leading to formation of $D - Xylulose - 5P + D - Erythrose - 4P$ is followed as it terminates in the desired target and the other path leading to $D - Ribose - 5P$ is not followed as it ends up in some other target. Of the 2 paths emerging from $D - Xylulose - 5P + D - Erythrose - 4P$, the path leading to $\beta - D - Fructose - 6P$ is not followed as it terminates in $2 - Deoxy - D - ribose$ which is not the desired target. The other path leads to formation of $D - Glyceraldehyde - 3P + \beta - D - Fructose - 6P$ which are the desired targets. The first 3 steps of the PPP which has been produced by the constraint based method, glucose 6-phosphate converted to ribulose 5-phosphate by the actions of the enzymes glucose 6-phosphate dehydrogenase (Glc6PD, EC 1.1.1.49), 6-phosphogluconolactonase (6PGL, EC 3.1.1.31) and 6-phosphogluconate dehydrogenase (6PGDH, EC 1.1.1.44) are crucial paths. These reactions are the only source of NADPH, which is needed to reduce peroxides and other oxidizing agents that may otherwise damage the cell. The sequence of steps that leads to formation of the optimal path has been cited in [10].

Starting from arginine there are five paths leading to formation of ornithine, urea, agmatine, 4-Guanidino-butanamide and 2-Oxo-arginine (Fig. 2). The path that leads to urea, 4-Guanidino-butanamide and 2-Oxo-arginine are not followed as they donot lead to the desired target. The paths leading to agmatine and ornithine are followed as they terminate in the desired target. Of the 3 paths emerging from ornithine the paths leading to formation of citrulline and N-acetyl-ornithine are not followed as they donot terminate in the desired target. The path that leads to putrescine is followed. Of the 4 paths emerging from Putrescine, the paths leading to γ-L-glutamyl-putrescine, N-Acetyl-putrescine and 4-Amino-butanal are not followed as they terminate in some other end products. The path leading to spermidine is followed as it leads to the desired target. From spermidine there are 3 paths forming 1,3-Diamino-propane, 4-Aminobutanal and spermine as the intermediate metabolites. The path leading to spermine is

followed which is the desired end product. The occurrence of the optimal path following the above mentioned methodology has been observed in [6, 9].

5   Conclusions and Discussions

We have developed a simple constraint based method for identifying an optimal metabolic pathway that involves weighting coefficients indicating the concentration levels of enzymes catalyzing biochemical reactions in the pathway. The method can suitably be used using reaction databases without going into complex mathematical calculations, and without using various kinetic parameters that are hard to estimate. It has been found that the method is able to produce biologically more relevant results than EPA and can be applied in metabolic engineering. This method could be of a great interest for the scientific community, as current pathway identification methods, e.g. elementary flux modes and extreme pathways cannot be applied to many real life models due to their numerical complexity.

References

1. Papin, J.A., Price, N.D., Wiback, S.J., Fell, D.A., Palsson, B.O.: Metabolic pathways in the post-genome era. Trends in Biochemical Sciences 28, 250–258 (2003)
2. Lee, J.M., Gianchandani, E.P., Papin, J.A.: Flux balance analysis in the era of metabolomics. Briefings in Bioinformatics 7, 1–11 (2006)
3. Schilling, C.H., Letscher, D., Palsson, B.O.: Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J. Theor. Biol. 203, 229–248 (2000)
4. Stryer, L., Berg, J.M., Tymoczko, J.L., Clarke, N.D. (eds.): Biochemistry. W H Freeman, New York (1998)
5. Montoya, A.R., Lee, W.P., Bassilian, S., Lim, S., Trebukhina, R.V., Kazhyna, M.V., Ciudad, C.J., Noe, V., Centelles, J.J., Cascante, M.: Pentose phosphate cycle oxidative and nonoxidative balance: A new vulnerable target for overcoming drug resistance in cancer. Int. J. Cancer 119, 2733–2741 (2006)
6. Mapelli, S., Brambilla, I.M., Radyukina, N.L., Ivanov, Y.V., Kartashov, A.V.: Free and bound polyamines changes in different plants as a consequence of uv-b light irradiation. Gen. Appl. Plant Physiology (special issue 34), 55–66 (2008)
7. Kumar, A., Altabella, T., Taylor, M.A., Tiburcio, A.F.: Recent advances in polyamine research. Trends Plant Sci. 2, 124–130 (1997)
8. Malmberg, R.L., Watson, M.B., Galloway, G.L., Yu, W.: Molecular genetic analyses of plant polyamines. Critical Reviews in Plant Sciences 17, 199–224 (1998)
9. Ponce, M., Martinez, L., Galmarini, C.: Influence of ccc, putrescine and gellam gum concentration on gynogenic embryo induction in allium cepa. Biologia Plantarum 50(3), 425–428 (2006)
10. Igoillo-Esteve, M., Maugeri, D., Stern, L., Beluardi, P., Cazzulo, M.J.: The pentose phosphate pathway in trypanosoma cruzi: a potential target for the chemotherapy of chagas disease. Annals of the Brazilian Academy of Sciences 79, 649–663 (2007)