An Optimization Rule for *In Silico* Identification of Targeted Overproduction in Metabolic Pathways

Mouli Das, C.A. Murthy, and Rajat K. De

Abstract—In an extension of previous work, here we introduce a second-order optimization method for determining optimal paths from the substrate to a target product of a metabolic network, through which the amount of the target is maximum. An objective function for the said purpose, along with certain linear constraints, is considered and minimized. The basis vectors spanning the null space of the stoichiometric matrix, depicting the metabolic network, are computed, and their convex combinations satisfying the constraints are considered as flux vectors. A set of other constraints, incorporating weighting coefficients corresponding to the enzymes in the pathway, are considered. These weighting coefficients appear in the objective function to be minimized. During minimization, the values of these weighting coefficients are estimated and learned. These values, on minimization, represent an optimal pathway, depicting optimal enzyme concentrations, leading to overproduction of the target. The results on various networks demonstrate the usefulness of the methodology in the domain of metabolic engineering. A comparison with the standard gradient descent and the extreme pathway analysis technique is also performed. Unlike the gradient descent method, the present method, being independent of the learning parameter, exhibits improved results.

Index Terms—Local minima, Newton-Raphson method, underdetermined problem, metabolic pathways, learning parameter

1 Introduction

It is well known that the enzyme-catalyzed biochemical reactions within the cell are grouped as metabolic pathways [1]. The importance of modeling these biochemical pathways has been extensively described in the literature. Consequently, mathematical (computational) modeling approaches for analyzing functionality and regulation of biochemical pathways are rapidly gaining importance. Various optimization algorithms such as the Levenberg-Marquardt method, genetic programming, simulated annealing and evolutionary algorithms have been applied to infer optimal pathways in biochemical models [2], [3], [4]. The flux balance analysis (FBA) technique, a constraint-based optimization approach applied to genome-scale metabolic models, can be used to make predictions of flux distributions and optimal pathways based on linear optimization [5].

Many learning algorithms find their roots in function minimization that can be classified into local minimization and global minimization [6]. Local minimization algorithms, such as gradient descent (GD), are fast but usually converge to local minima. In contrast, global minimization algorithms have heuristic strategies to help escape from local minima. Many techniques in data mining and machine learning follow a GD paradigm in the iterative process for optimization [7]. However, several drawbacks of the GD learning method have been observed:

1. its convergence speed is usually too low,
2. its convergence accuracy is hard to control,
3. it is easily stuck in bad local minima, and
4. the choice of proper learning constant largely depends on trial and error [8].

One common approach is to upgrade the normal GD learning, which is a first-order learning algorithm, to a second-order one. Since the second-order method is an optimization algorithm with quadratic convergence speed, it can be used to improve the learning speed and accuracy of the normal backpropagation (BP) [6].

We have recently presented a supervised second-order learning algorithm, a modification of the Newton-Raphson method that identifies some optimal metabolic pathways accurately and efficiently leading to the overproduction of a biochemical product of interest [9]. The learning method can be considered as a nonlinear global optimization problem in which the goal is to minimize a nonlinear objective function that involves the weights using heuristic strategies [10]. In applying our proposed second-order learning method to biochemical modeling, we implemented two subsidiary improvements: 1) Since it is difficult to model nonlinear biochemical systems, we adopted a second-order derivative transformation in the updated learning rule, thereby facilitating the optimization. 2) Our proposed method always converges to the global minimum experimentally (in contrast to local optima). However, for cases where the first-order derivative is zero, there may be a local minimum and also a point of inflection. The second-order method will not perform in such situations. 3) Our proposed

• *The authors are with the Machine Intelligence Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata, West Bengal 700 108, India. E-mail: mouli.das@gmail.com, {murthy, rajat}@isical.ac.in.*

Manuscript received 18 July 2012; revised 23 Apr. 2013; accepted 22 May 2013; published online 10 June 2013.
For information on obtaining reprints of this article, please send e-mail to: tcbb@computer.org, and reference IEEECS Log Number TCBB-2012-07-0172. Digital Object Identifier no. 10.1109/TCBB.2013.67.

1545-5963/13/$31.00 © 2013 IEEE Published by the IEEE CS, CI, and EMB Societies & the ACM

method is independent of the choice of the learning parameter whereas the GD method largely depends on the value of the learning parameter. Unlike the GD method, here we only need to adjust the starting point.

The remainder of the paper is organized as follows: The basic underlying FBA methodology has been described in Section 2. The extreme pathway analysis (EPA) method has also been described here. Section 3 provides a rigorous explanation of optimal metabolic pathways. Section 3.1 provides the theoretical formulation of the second-order analysis. The illustration of the proposed method that identifies optimal metabolic pathways by maximizing the amount of synthesis of the target metabolite has been provided in Section 3.2. Section 4 illustrates applications to both the synthetic data and the real data of various organisms belonging to different phylogeny. They include pentose phosphate pathways (PPP), glycolytic pathways, flavonoid and phenylpropanoid biosynthesis pathways. These metabolic pathway data has been collected from the KEGG database.[1] Using the above benchmark pathways, we compare our second-order optimization methodology against some of the other optimization algorithms such as GD (first-order optimization) method and the EPA method [11] in Section 4, and demonstrate its superior improvement in performance. The biological relevance of the results is provided in Section 5. The usefulness of our methodology in the field of metabolic engineering, has been discussed in Section 5.4. Finally, the paper is concluded in Section 6.

2 Background

In this section, we describe briefly FBA, the problem and the extreme pathway analysis.

2.1 Flux Balance Analysis

The majority of network studies has focused on topological properties and not on the rate of metabolic activity, which can vary significantly from reaction to reaction. This important function is not captured by standard topological approaches. It is necessary to include this information in the network description to develop an understanding of how the structure of a metabolic network affects metabolic activity. A meaningful understanding requires us to consider the intensity (i.e., enzyme concentration) and the temporal aspects of the interactions. Although much is still unknown about the temporal aspects of metabolic activity inside a cell, recent results have provided information about the incorporation of enzymatic concentrations in metabolic reactions in single-cell metabolism [12], [13]. A simple linear optimization approach, called FBA, has emerged as an important framework to assess the metabolic potential of an organism. By taking a complete inventory of all (known) metabolic capabilities of an organism, FBA can assess the maximum possible yield of a desired product for different substrates and growth levels. The FBA method [14] is based on the assumption that the concentration of all cellular metabolites, not subject to transport across the cell membrane, must satisfy the steady-state constraint. Any flux value satisfying the steady-state constraint corresponds to a stoichiometrically allowed state of the cell. To select flux values that are biologically relevant, we optimize the cellular growth. Experiments support this hypothesis under several conditions [15], [16].

2.2 Extreme Pathway Analysis Method

Metabolic pathway analysis based on extreme pathways [11] is an important tool for analyzing the properties of metabolic networks. The approach [1] is based on convex analysis for solving a homogeneous system of linear equations and inequality constraints that define the steady-state solution space for the reaction network. The convex solution corresponds to a convex polyhedral cone in an n-dimensional space (R^n). The extreme rays of the cone correspond to the extreme pathways of a metabolic network. Each extreme ray corresponds to a particular pathway or an active set of fluxes. Let us consider the cone as defined, in terms of flux vectors, and can be stated as

$$C = \left\{ \mathbf{v} : \mathbf{v} = \sum_{i=1}^{k} w_i \mathbf{p}_i, w_i \geq 0 \ \forall i \right\}. \quad (1)$$

Here, the vector $\mathbf{p_i}$ denotes the extreme pathways, $\mathbf{v}$ denotes the flux distribution where every point within this cone (C) can be written as a nonnegative linear combination of the extreme pathways and k denotes the total number of extreme pathways needed to generate the flux cone C. w indicates the weight assigned to each corresponding pathway, and can be interpreted as an indication of the importance of that pathway in the network. Thus, extreme pathways describe the full capabilities of the metabolic network in the simplest form possible. They are systemically independent and irreducible set of vectors. Each particular point within the flux cone corresponds to a different flux distribution representing a particular metabolic phenotype.

3 Method

Metabolic pathways can be represented in various manners depending on the required level of detail. There exist standard graph-based methods to model metabolic pathways. A metabolic pathway can be depicted as a directed hypergraph, where the reactants are represented by a set of vertices and the product metabolites are represented by another disjoint set of vertices. The hyperedges or hyperarcs connect these two sets of vertices thus representing distinct enzyme catalyzed reactions. A pathway is a connected sequence of the hyperedges where each reactant occurs only once. In contrary to ordinary graphs, a hypergraph considers reactions as complete entities and can connect groups of more than two nodes. Hypergraphs can be converted into directed graphs.

It is to be mentioned here that we have considered only the main metabolites in the reactions of a pathway. That is, we have ignored the presence of ADP, ATP, H_2O, NADP and so on either as reactants or products. In other words, entries corresponding to the main metabolites are only there in the stoichiometric matrix. This is just to minimize the order of the said matrix and size of the graph.

Here, we have represented a metabolic pathway as a graph in which metabolites are depicted by nodes and

1. http://www.genome.jp/kegg/pathway.html.

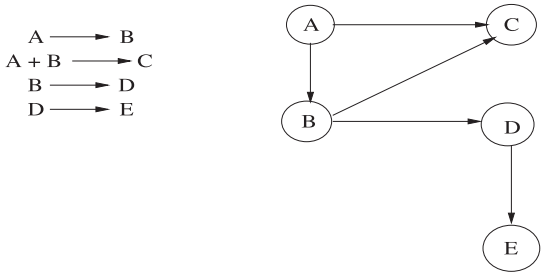


Fig. 1. A hypothetical metabolic network represented by a directed graph where there are five nodes corresponding to five metabolites A, B, C, D, and E, connected by the directed edges.

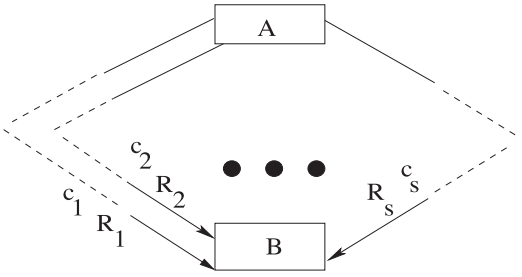


Fig. 2. A hypothetical metabolic pathway with A as the starting metabolite and B as the target metabolite. $R_1, \ldots, R_s$ represent the set of enzyme catalyzed biochemical reactions. $c_1, \ldots, c_s$ denote their corresponding enzyme concentration levels [17].

reactions by directed edges. For example, Fig. 1 depicts a hypothetical metabolic pathway in which five metabolites (A, B, C, D, and E) are involved, A being the starting substrate, and C or E being a target metabolite. Fig. 2 provides another hypothetical metabolic pathway. It constitutes a set of enzyme catalyzed biochemical reactions through which a substrate (starting metabolite A) gets converted to a target metabolite B. An enzyme acts as a catalyst for a reaction, i.e., presence of an enzyme c_1 to a higher level, corresponding to an edge R_1, represents the inclusion of the edge in the optimal pathway [17]. This has been explained below in this section.

The metabolic pathway in terms of graph representation is provided as follows: An edge e in a graph, $G = (V, E)$, is defined as $e = (p, q)$ where $p, q \epsilon V$, $e \epsilon E$. In ordinary words, p and q are joined by a line which is the edge. Two nodes $p, q \epsilon V$ are said to be connected if there exists $p_1, p_2, \ldots, p_{n-1}$ such that (p_i, p_{i+1}) is an edge $\forall i = 0, \ldots, n-1$, where $p_0 = p$ and $p_n = q$. We take this as a metabolic pathway. Note that in a graph for two nodes p, q, there may exist several paths joining p and q, i.e., there may exist several biochemical reactions in series that ultimately produce q from p. Let us assume that a metabolic pathway is starting at p and ending at q, and it has n edges/reactions namely, $(p_0, p_1), (p_1, p_2) \ldots (p_{n-1}, p_n)$ where $p_0 = p$ and $p_n = q$.

An edge/reaction here will become operative if the enzyme catalyzing the reaction is present at least at a required level. In other words, if the concentration of an enzyme catalyzing a reaction is high, the strength of the corresponding edge is high. There exist many paths from substrate to the target metabolite. On maximization of the target, we trace the path from the substrate to the target in such a way that the edge with higher enzyme concentration is included in the path. Thus the path, starting from the substrate to the target, contains a set of highly active reactions/edges (i.e., with higher enzyme concentration). This path leads to the production of the target to a maximum amount and we have called this path as an optimal path.

The differences in the definition of the path between the proposed and the extreme pathway analysis methods are mentioned in [17, p. 15]. For the sake of convenience of the readers, we again describe them as follows:

- Unlike the extreme pathway analysis, the present method considers the presence of enzymes.
- Extreme pathway analysis finds the flux vectors upon optimization, whereas the present method generates a set of some possible flux vectors and finds an optimal pathway in terms of weighting coefficients reflecting enzyme concentration.
- Extreme pathway analysis considers individual reactions in the pathway in a sequential manner, whereas the present method considers all the reactions in parallel.

3.1 The Second-Order Gradient Method

The second-order method is derived from Newton-Raphson method [7] whose principle is discussed here. Taylor expansion of a function $E(w)$ of a single variable w in the vicinity of a minimum w^* is given by

$$E(w) = E(w^*) + (w - w^*)(dE/dw)_{w=w^*} + 1/2(w - w^*)^2(d^2E/dw^2)_{w=w^*} + O(w^3). \quad (2)$$

The gradient of the cost function is zero at the minimum. Differentiating the above (2) with respect to w gives an approximation of the gradient of the cost function in the neighborhood of a minimum,

$$dE/dw = (w - w^*)(d^2E/dw^2)_{w=w^*}. \quad (3)$$

Therefore, if variable w is in the neighborhood of w^*, the minimum could be reached in a single iteration if the second derivative of the cost function at the minimum were known. w can simply be updated by an amount

$$\Delta w = -\frac{(dE/dw)}{(d^2E/dw^2)_{w=w^*}}. \quad (4)$$

Thus by contrast to simple GD, the direction of motion in parameter space, is not the direction of the gradient, but a linear transformation of the gradient. Our proposed methodology is an iterative technique that is based on the above formula (see (4)) with certain modifications in the second-order derivative term.

3.2 The Second-Order Modeling Technique

The key concept of our model is the development of a new learning rule based on Newton-Raphson method [18]. GD is the most widely used algorithm for supervised learning of neural networks [19]. The most popular training algorithm of this category is batch BP. It is the first-order method that minimizes the error function by updating the weights. The use of optimization tools such as first-order GD operating

on metabolic reconstructions to identify optimal pathways is becoming commonplace. Nevertheless, a number of shortcomings still exists [20]. To remedy these limitations, we introduce a new second-order computational framework that identifies some optimal metabolic pathways. The suggested second-order method can be shown to have much better convergence properties than the first-order GD system which will go on oscillating for steep functions and for functions that are constantly increasing or decreasing.

3.2.1 System Description

Here, we discuss how metabolic systems may be described mathematically. A cellular metabolic reaction network is a collection of enzymatic reactions and transport processes. A system boundary can be drawn around a metabolic network comprising internal fluxes inside the network and exchange fluxes that exist across the boundary. All reversible reactions are considered as two internal fluxes occurring in opposite directions. The theoretical formulation of the second-order optimization problem has been given here extensively in Section 3.2.4 and in Section 3.1, which was not considered in the previous work in [9].

3.2.2 Reaction Variables and Constraints

Our methodology considers a metabolic network with m metabolites and n reactions, i.e., n internal and exchange fluxes, where the final metabolite B can be reached from the substrate (starting metabolite) A through any one of s biochemical reactions/conversions $R_1, R_2, \ldots, R_s$ in different paths. Here a reversible reaction has been considered as two separate reactions corresponding to forward and backward reactions respectively [11]. For example, the reversible reaction A $\rightleftharpoons$ B is considered as A $\rightarrow$ B and B $\rightarrow$ A. The entries for A and B in the stoichiometric matrix, corresponding to the forward reaction (A $\rightarrow$ B) are -1 and $+1$ respectively. On the other hand, it is the reverse for B $\rightarrow$ A. In both the cases, the fluxes are positive. The number of internal and exchange fluxes is represented by n_I and n_E, respectively, i.e., $n = n_I + n_E$. The kth internal flux is denoted by v_k and the lth exchange flux is denoted by b_l. So there are $v_1, \ldots, v_{n_I}$ internal fluxes and $v_{n_I+1}, \ldots, v_n$ exchange fluxes where $v_{n_I+l} = b_l$. The rate of growth of the metabolite B on the substrate A, which needs to be maximized is obtained by taking linear algebraic sum of the weighted fluxes of reactions, and is given by

$$z = \sum_{k=1}^{s} c_k v_k. \tag{5}$$

z denotes the objective function for this problem. Here, v_k is the flux of the reaction R_k producing metabolite B. c_k denotes the weighting factor representing the level of concentration of the enzyme catalyzing the reaction R_k. c_k indicates how much kth reaction (such as the biomass reaction when simulating maximum growth) contributes to the objective function z. To completely describe the system we need to include the constraints on internal and exchange fluxes. All the internal fluxes are positive, yielding

$$v_i \geq 0, \forall i.$$

The exchange fluxes can operate in a bidirectional manner and is therefore unconstrained. These constraints can be expressed as

$$\alpha_j \leq b_j \leq \beta_j,$$

where α_j and β_j are either zero or negative, and positive infinity, respectively, based on the direction of the exchange flux, which allows a metabolite to enter or exit the system boundary.

Metabolic reactions are represented as a stoichiometric matrix $\mathbf{S}$ of size $m \times n$. Every row of $\mathbf{S}$ represents one unique metabolite (for a system with m metabolites) and every column represents one reaction (n reactions). The entries in each column are the stoichiometric coefficients of the metabolites participating in a reaction. There is a negative coefficient (-1) for every metabolite consumed and a positive coefficient ($+1$) for every metabolite that is produced. A stoichiometric coefficient of zero value is used for every metabolite that does not participate in a particular reaction. $\mathbf{S}$ is a sparse matrix because most biochemical reactions involve only a few metabolites.

3.2.3 Data Generation

The stoichiometric matrix $\mathbf{S}$ can be computed from a reaction database. The flux through all of the reactions in a network is represented by the vector $\mathbf{v}$. The concentrations of all metabolites are represented by the vector $\mathbf{x}$. The system of mass balance equations at steady state is

$$d\mathbf{x}/dt = \mathbf{0}. \tag{6}$$

That is, at steady state,

$$\mathbf{S}.\mathbf{v} \approx \mathbf{0}. \tag{7}$$

Equation (7) can be solved given a set of upper and lower bounds on $\mathbf{v}$ (mentioned in Section 3.2.2), and a linear combination of fluxes (as an objective function as shown in (5)). Any $\mathbf{v}$ that satisfies (7) is said to be in the null space of $\mathbf{S}$ [11]. In any realistic large-scale metabolic model, there are more reactions than there are metabolites ($n > m$). In other words, there are more unknown variables than equations, so there is no unique solution to this system of equations. So, (7) is under determined. We generate p number of basis vectors $\mathbf{v}_b$ by using standard routines and toolboxes available in MATLAB. Starting with the basis vectors we further generate p random numbers a_j, $j = 1, 2, \ldots, p$ and a vector

$$\mathbf{v} = \sum_{j=1}^{p} a_j \mathbf{v}_{bj}, \tag{8}$$

until certain inequality constraints on $\mathbf{v}$ are satisfied for all its components. In real-life systems, the genes that produce the enzymes may not be expressed at the required level. This imposes restrictions on the system, and for this purpose, we define another constraint as

$$\mathbf{S}.(\mathbf{C}.\mathbf{v}) = \mathbf{0}, \tag{9}$$

where $\mathbf{C}$ is an $n \times n$ diagonal matrix whose diagonal elements are the components of the vector $\mathbf{c}$. That is, if $\mathbf{C} = [\gamma_{kl}]_{n \times n}$, then $\gamma_{kl} = \delta_{kl} c_k$, where δ_{kl} is the Kronecker

delta. c_k denotes the weighting coefficient and is similar to the formulation in (5). The only difference that exists between the c_k values of (5) and (9) is that in the former case the z value denotes the weighted fluxes that are involved with the growth of the target metabolite B and the latter case includes all the reactions. Thus the optimization problem of determining a metabolic pathway yielding maximum rate of production of the target metabolite B starting from a substrate A, reduces to a maximization problem, where z is maximized with respect to $\mathbf{c}$, subject to satisfying the constraint given in (9) along with the inequality constraints discussed in Section 3.2.2.

3.2.4 Estimation of Weighting Coefficients c_k through Second-Order Learning Algorithm

Given the above particular set of constraints on internal and exchange fluxes (see Section 3.2.2) we have to maximize the growth rate (see (5)). Combining (5) and (9), we can reformulate the objective function as

$$y = 1/z + \mathbf{\Lambda}^T.(\mathbf{S}.(\mathbf{C}.\mathbf{v})), \tag{10}$$

which needs to be minimized with respect to the weighting factors c_k for all k. We are interested in determining the optimal pathway that maximizes the growth rate/objective function z and minimizes y. The term $\mathbf{\Lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_m]^T$ is the regularizing parameter. For the sake of simplicity, we have considered here $\lambda_1 = \cdots = \lambda_m = \lambda$ (say). Initially, a set of random values in $[0, 1]$ corresponding to c_k's are generated. The c_k's are then modified iteratively by the new learning algorithm incorporating modulus of the second-order derivative (based on (4)), where the amount of modification for c_k in each iteration is defined as

$$\Delta c_k = -\frac{\partial y}{\partial c_k} \Big/ \left|\frac{\partial^2 y}{\partial c_k^2}\right|. \tag{11}$$

This is a modified version of the Newton-Raphson method of weight updating. This method uses the second-order derivative in addition to the gradient to determine the next updating direction and step size [21]. The c_k's can also be modified by using the GD optimization technique which largely depends on the value of the learning parameter η as mentioned in the earlier work in [17]. In contrast to GD, the novelty of this new learning rule is that it is independent of the parameter η that indicates the rate of modification. The second-order method entirely depends on the proper choice of the initial value to reach the global optima.

If the first-order derivative is greater than zero, the functions are also increasing or shooting up nature, so we need to decrease the value of c to find the optimal solution which is handled by the minus sign in (11). The modulus of the second-order derivative in the denominator of (11) provides us an idea beforehand regarding the amount of updation necessary to reach the optima. For computing the values of Δc_k's, we use the following expression:

$$\Delta c_k = \frac{1}{z^2}\frac{\partial z}{\partial c_k} - \frac{\partial(\mathbf{\Lambda}^T.(\mathbf{S}.(\mathbf{C}.\mathbf{v})))}{\partial c_k} \Big/ \left|\frac{2}{z^3}\left(\frac{\partial z}{\partial c_k}\right)^2\right|. \tag{12}$$

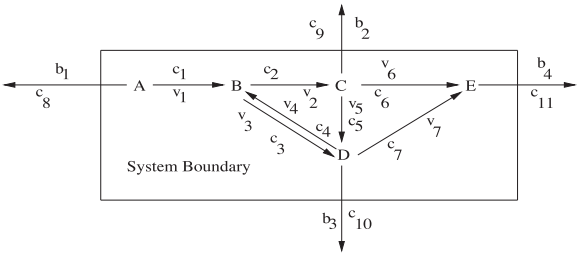


Fig. 3. A simple hypothetical pathway consisting of the five metabolites, seven internal fluxes (v_1-v_7), and four exchange fluxes (b_1-b_4), giving rise to a total of 11 fluxes.

Thus, the modified value of c_k is given by

$$c_k(t+1) = c_k(t) + \Delta c_k, \quad \forall i, \quad t = 0, 1, 2, \ldots$$

$c_k(t+1)$ is the value of c_k at iteration $(t+1)$, which is computed based on the c_k-value at iteration t.

For each value of the regularization parameter λ (chosen empirically from 0.1 to 1.0 in steps of 0.1), the c_k-values are observed for a minimum value of y. The concentration vector c_k attains values between 0 and 1, as mentioned above, corresponding to some values of v, and is negligible for other values of v. We take into account the values of c_k's that are close to 1, corresponding to the minimum value of y. This enables us to identify an optimal metabolic pathway that yields the maximum amount of the target metabolite B starting from the initial metabolite A. There may be several branches involved in the pathway and we may land at several intermediate points (other than the target metabolite) which may or may not lead to the target. By changing the value of the parameter λ, we are able to reach the target via the optimal path.

4 Results and Performance Comparison

To demonstrate the benefits of the new second-order learning algorithm, we benchmark it on some of the metabolic networks used in the previous study [9] and on some other networks. The networks considered are a synthetic network (see Fig. 3) [22] and two large complex metabolic networks of flavonoid and phenylpropanoid biosynthesis (see Figs. 4 and 6). The PPPs and the glycolytic pathways are considered in the supplemental material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TCBB.2013.67. The five real life pathways include the PPPs of *E. coli* K-12 MG1655 (see Fig. 8, which is available in the online supplemental material), *P. falciparum* (see Fig. 10, which is available in the online supplemental material), *T. cruzi* (see Fig. 12, which is available in the online supplemental material) and the glycolytic pathways of the two archae *H. salinarum R1* (see Fig. 14, which is available in the online supplemental material) and *N. pharaonis* (see Fig. 16, which is available in the online supplemental material), respectively. We use the scheme depicted in Figs. 4 and 6 to demonstrate the applicability of our approach to a system of reasonable complexity. We also demonstrate a comparative analysis of the performance of the proposed second-order learning algorithm to that of the GD rule (BP algorithm) and the EPA method [11] in Table 2. In the previous work [17], the results

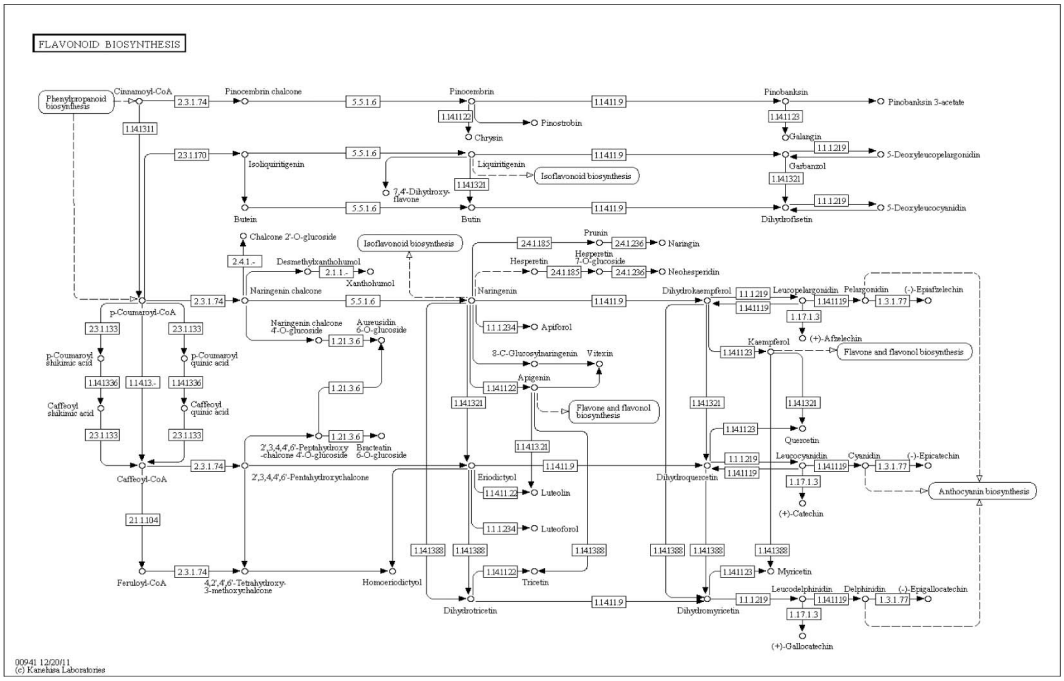


Fig. 4. Flavonoid biosynthesis pathway. There are 68 metabolites and 92 reactions. The enzymes with their EC numbers are depicted in boxes.

obtained after optimization by the GD method were entirely dependent on the proper choice of the learning parameter. This difficulty is removed in the present work where the optimal pathways obtained are independent of the learning parameter and they solely depend on the proper initialization condition of the weighting coefficient. The program files, along with a Read-me file, are also uploaded as additional files. The respective codes are available at the following link: https://sites.google.com/site/moulidas/home.

4.1 An Illustrative Hypothetical Reaction System

Prior to an application to more detailed biochemical models, we exemplify our approach using a simple hypothetical pathway. Suppose the reaction system depicted in Fig. 3, consisting of five metabolites and 11 fluxes (seven internal fluxes and four exchange fluxes), is designated for mathematical modeling [22]. The rate of yield ($z = c_6v_6 + c_7v_7 - c_{11}b_4$) of the target metabolite E from the starting substrate A is maximized. Within a reasonably realistic scenario, we can assume that the average concentrations of both the target and the starting metabolites have been determined experimentally. Applying the present second-order method, we have obtained the pathway as $R_1 \rightarrow R_2 \rightarrow R_3 \rightarrow R_8 \rightarrow R_{10} \rightarrow R_{11}$ as an optimal one, which is again conforming to [22]. Here, R_1: Ext $\rightarrow$ A, R_2: A $\rightarrow$ B, R_3: B $\rightarrow$ C, R_4: C $\rightarrow$ Ext, R_5: B $\rightarrow$ D, R_6: D $\rightarrow$ B, R_7: D $\rightarrow$ Ext, R_8: C $\rightarrow$ D, R_9: C $\rightarrow$ E, R_{10}: D $\rightarrow$ E, R_{11}: E $\rightarrow$ Ext. Optimal pathways obtained by EPA and GD analysis are the same as that obtained by the present method. It is to be mentioned here that 100 iterations were required for minimizing y by the GD method, whereas the present second-order analysis require only 80 iterations for convergence (see Table 2).

4.2 Flavonoid Biosynthetic Pathway

Flavonoids represent a major class of plant secondary metabolites that include the flavonols, anthocyanins, proanthocyanidins (condensed tannins) and isoflavonoids. They perform a variety of important functions in plant growth, reproduction and survival, and also serve as important micronutrients in human and animal diets. They impart various characteristics such as pigmentation and protection from ultraviolet-B (UV-B) ray. Flower color is predominantly due to three types of pigment: flavonoids, carotenoids, and betalains. The flavonoid biosynthetic pathway has been one of the most intensively studied metabolic systems in plants [23]. Flavonoids are synthesized via a well-characterized biosynthetic pathway that has been localized to the cytoplasm in many different plant species.

Considering the reference pathway from the KEGG database (see Fig. 4), there are three subpathways involved in flavonoid biosynthesis [23], [24] (see Fig. 5). Following our proposed second-order methodology, the optimal pathway for the first flavonoid biosynthetic subpathway is: *p-Coumaroyl-CoA* $\rightarrow$ *Naringeninchalcone* $\rightarrow$ *Naringenin* $\rightarrow$ *Dihydrokaempferol* $\rightarrow$ *Leucopelargonidin* $\rightarrow$ *Pelargonidin*. Similarly the optimal pathway for the second flavonoid biosynthetic subpathway is: *p-Coumaroyl-CoA* $\rightarrow$ *Naringeninchalcone* $\rightarrow$ *Naringenin* $\rightarrow$ *Eriodictyol* $\rightarrow$ *Dihydroquercetin* $\rightarrow$ *Leucocyanidin* $\rightarrow$ *Cyanidin*. Similarly, the optimal pathway for the third flavonoid biosynthetic subpathway is: *p-Coumaroyl-CoA* $\rightarrow$ *Naringeninchalcone* $\rightarrow$ *Naringenin* $\rightarrow$ *Dihydrotricetin* $\rightarrow$ *Dihydromyricetin* $\rightarrow$ *Leucodelphinidin* $\rightarrow$ *Delphinidin*. The three optimal subpathways derived by the present second-order method are depicted in Fig. 5 by three different arrows.

The first flavonoid biosynthetic subpathway on comparison with all the three methods are the same. On comparison with both the EPA and GD methods we have obtained the

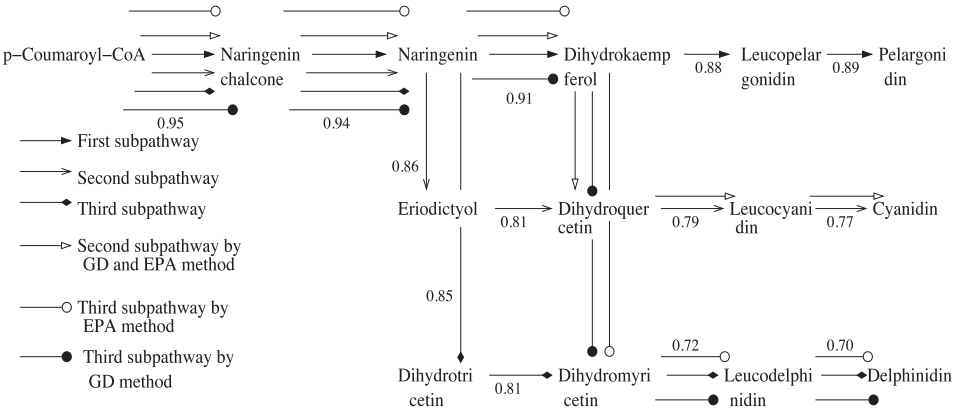


Fig. 5. Optimal flavonoid biosynthesis pathway. The starting metabolite for the three subpathways are p-Coumaroyl-CoA, and their corresponding target metabolites are pelargonidin, cyanidin, and delphinidin. The arrows indicate the respective pathways obtained by different methods.

second flavonoid subpathway as follows: *p-Coumaroyl-CoA → Naringeninchalcone → Naringenin → Dihydrokaempferol → Dihydroquercetin → Leucocyanidin → Cyanidin*. On comparison with the existing EPA we have obtained the third flavonoid subpathway as follows: *p-Coumaroyl-CoA → Naringeninchalcone → Naringenin → Dihydrokaempferol → Dihydromyricetin → Leucodelphinidin → Delphinidin*. On comparison with the proposed method that follows GD optimization the third flavonoid subpathway obtained is as follows: *p-Coumaroyl-CoA → Naringeninchalcone → Naringenin → Dihydrokaempferol → Dihydroquercetin → Dihydromyricetin → Leucodelphinidin → Delphinidin*. All these comparisons are shown in Fig. 5. The results are analyzed in Table 2 for all the three subpathways.

4.3 Phenylpropanoid Biosynthetic Pathway

The phenylpropanoids and the related plant polyketides are a group of plant secondary metabolites that have multiple biological functions. They serve to attract pollinators, support secondary cell-wall growth, provide protection against various plant diseases and interact with beneficial soil microbes. Their basic chemical properties also make them useful in the biofuel and biomaterial industries [25]. Phenylpropanoid metabolism begins with the amino acid phenylalanine, which feeds into various biosynthetic pathways that generate a wide range of structurally related polyphenolic compounds. Phenylalanine is first converted to cinnamic acid by deamination. It is followed by hydroxylation and frequent methylation to generate coumaric acid and other acids with a phenylpropane (C6-C3) unit. Reduction of the CoA-activated carboxyl groups of these acids results in the corresponding aldehydes and alcohols. The alcohols are called monolignols, the starting compounds for biosynthesis of lignin [26]. The phenylpropanoid pathway produces the majority of phenolic compounds found in nature.

We consider the reference pathway for phenylpropanoid biosynthesis from the Kegg database (see Fig. 6). Applying the present methodology, optimal pathway was found to be:

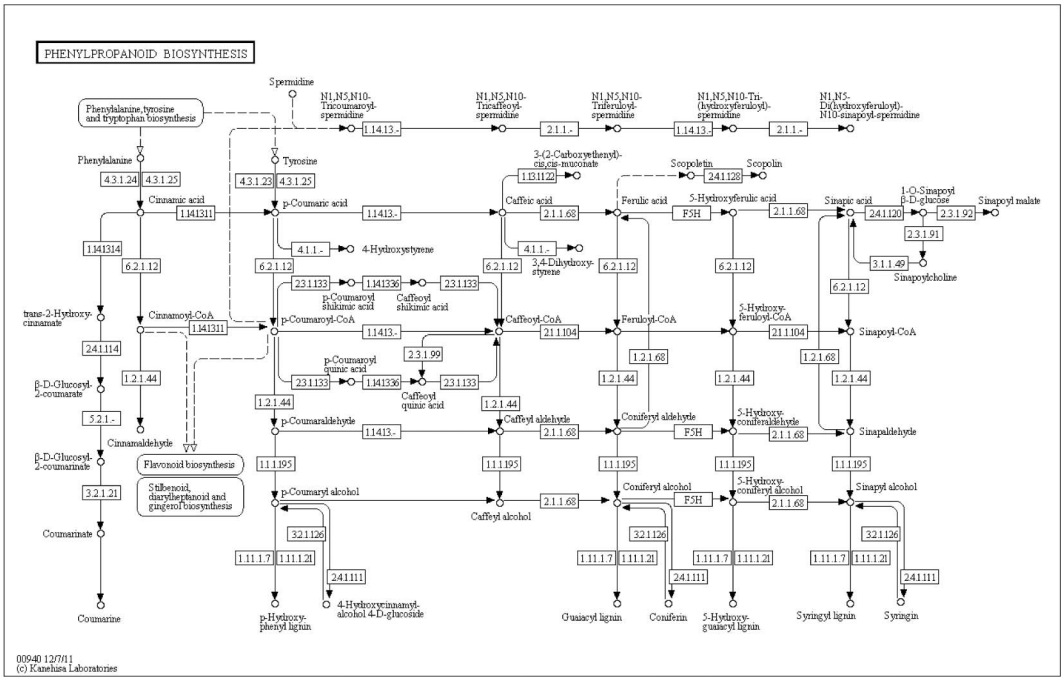


Fig. 6. Phenylpropanoid biosynthesis pathway. There are 55 metabolites and 73 reactions. The enzymes with their EC numbers are depicted in boxes.

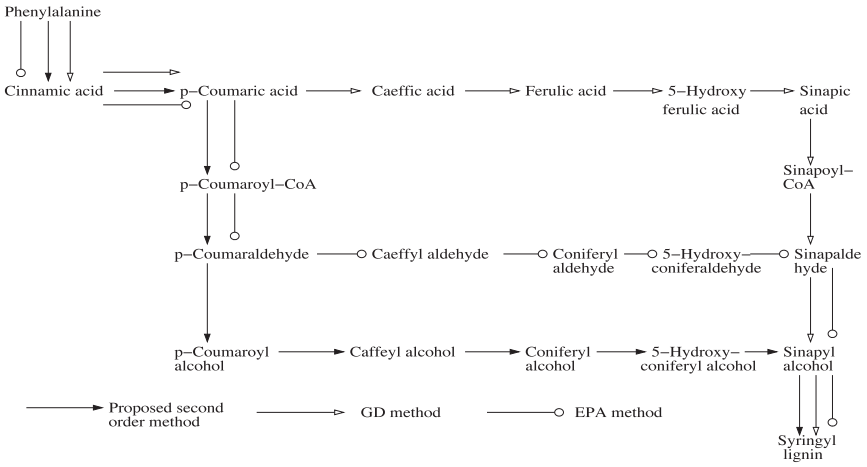


Fig. 7. Optimal phenylpropanoid biosynthesis pathway. The starting metabolite is phenylalanine and the target metabolite is syringyl lignin. The arrows indicate the respective pathways obtained by the different methods.

Phenylalanine → *Cinnamicacid* → *p-Coumaricacid* → *p-Coumaroyl-CoA* → *p-Coumaraldehyde* → *p-Coumarylalcohol* → *Caffeylalcohol* → *Coniferylalcohol* → 5-*Hydroxy-coniferyl alcohol* → *Sinapylalcohol* → *Syringyllignin* as shown in Fig. 7.

On comparison with the GD formulation we have obtained the phenylpropanoid pathway as follows: *Phenylalanine* → *Cinnamicacid* → *p-Coumaricacid* → *Caffeicacid* → *Ferulicacid* → 5-*Hydroxyferulic acid* → *Sinapicacid* → *Sinapoyl-CoA* → *Sinapaldehyde* → *Sinapylalcohol* → *Syringyl lignin*. Comparing with the EPA, the pathway obtained is: *Phenylalanine* → *Cinnamicacid* → *p-Coumaricacid* → *p-Coumaroyl-CoA* → *p-Coumaraldehyde* → *Caffeylaldehyde* → *Coniferyl aldehyde* → 5-*Hydroxy-coniferaldehyde* → *Sinapaldehyde* → *Sinapylalcohol* → *Syringyllignin*. Fig. 7 depicts the comparison with all the three methods. The results are mentioned in Table 2.

Table 1 provide a list of possible pathways from the starting metabolite phenylalanine to the target metabolite syringyl lignin with the average yield (z) of the target metabolite and their corresponding c-values. It can be inferred from Table 1 that the pathway corresponding to serial number 1 yields the highest average z and their corresponding c-values are larger compared to the other 14 instances, and hence it is the optimal pathway. Serial numbers 2 and 3 depict the paths obtained by the GD method and the EPA method, respectively. The z-values and their corresponding c-values for these two pathways in serial numbers 2 and 3 are smaller compared to the optimal pathway obtained by the proposed second-order method in serial number 1. Such tables can analogously be constructed for the other examples that has been considered in this work. They have not been presented here due to paucity in space. This will lead to similar observations for the c-values and the z-values as discussed above. The optimal pathways for the synthetic network, pentose phosphate metabolism and glycolysis derived by this second-order analysis as provided earlier in this section (Section 4.1) and also in the online supplementary file, can be confirmed from the c-values and the z-values obtained theoretically. Thus, it can be concluded that our second-order method correctly identifies the optimal pathways.

Table 2 demonstrate the computational results for all the networks considered here. The λ-value is varied from 0.1 to 1.0. The value of the learning parameter η is kept fixed at 0.5. It can be seen that for all the examples, the proposed method converges faster than the GD method. The number of iterations required by the GD method is much larger than our method. It can also be noted from the tables that the z-value obtained by the GD method is smaller than that obtained by our method. Thus, we can say that the proposed method reaches the global minima in contrast to the GD method.

5 Biological Validation

Here, we highlight that this second-order optimization method is more appealing from the biological point of view, than the GD and EPA methods considered earlier [17].

5.1 Pentose Phosphate and Glycolytic Pathways

The existence of the sugar phosphates Glyceraldehyde-3P, Ribulose-5P, Xylulose-5P, Fructose-6P, and Glucose-6P in the PPP are found in [27]. Ribose 5-phosphate is synthesized from glucose or glycolytic intermediates through two pathways: the oxidative branch of the PPP (catalyzed by glucose 6-P dehydrogenase and 6-P-gluconate dehydrogenase) and the nonoxidative branch of the PPP (catalyzed by transketolase and transaldolase). The PPP in *P. falciparum* (see Fig. 11 of the online supplementary file) as obtained by our proposed methodology has been observed in [28]. The PPP may be crucial in protecting the parasite from oxidative stress during red cell infection. The first three steps of the PPP, which has also been produced by the proposed 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 (see Figs. 8, 10, and 12) of the online supplementary file. 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 balance between oxidative and nonoxidative branches of the pentose phosphate cycle is necessary to maintain the metabolic efficiency of the cell for growth and proliferation [29].

Trypanosoma cruzi, a flagellated protozoan parasite, having a complex life cycle is the causative agent of the

TABLE 1
The Possible Pathways (Starting Substrate Phenylalanine and Target Metabolite Syringyl Lignin) with Their c-Values and z-Values for the System in Fig. 6

Sr. No.	Some possible paths	Optimal c-values	Average quantity (z) of target metabolite Syringyl lignin synthesis
1	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow p-Coumaric\ acid \rightarrow p-Coumaroyl-CoA \rightarrow p-Coumaraldehyde \rightarrow p-Coumaryl\ alcohol \rightarrow Caffeyl\ alcohol \rightarrow Coniferyl\ alcohol \rightarrow 5-Hydroxy-coniferyl\ alcohol \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.98, 0.95, 0.89, 0.91, 0.88, 0.85, 0.80, 0.86, 0.96, 0.92	60.94
2	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow p-Coumaric\ acid \rightarrow Caffeic\ acid \rightarrow Ferulic\ acid \rightarrow 5-Hydroxyferulicacid \rightarrow Sinapic\ acid \rightarrow Sinapoyl-CoA \rightarrow Sinapaldehyde \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.98, 0.95, 0.51, 0.55, 0.46, 0.09, 0.85, 0.80, 0.67, 0.92	50.25
3	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow p-Coumaric\ acid \rightarrow p-Coumaroyl-CoA \rightarrow p-Coumaraldehyde \rightarrow Caffeyl\ aldehyde \rightarrow Coniferyl\ aldehyde \rightarrow 5-Hydroxy-coniferaldehyde \rightarrow Sinapaldehyde \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.98, 0.95, 0.89, 0.91, 0.81, 0.43, 0.25, 0.94, 0.67, 0.92	20.75
4	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow p-Coumaric\ acid \rightarrow p-Coumaroyl-CoA \rightarrow p-Coumaroyl\ shikimic\ acid \rightarrow Caffeoyl\ shikimic\ acid \rightarrow Caffeoyl-CoA \rightarrow Caffeyl\ aldehyde \rightarrow Caffeyl\ alcohol \rightarrow Coniferyl\ alcohol \rightarrow 5-Hydroxy-coniferyl\ alcohol \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.98, 0.95, 0.89, 0.81, 0.61, 0.23, 0.65, 0.97, 0.80, 0.86, 0.96, 0.92	18.26
5	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow Cinnamoyl-CoA \rightarrow p-Coumaroyl-CoA \rightarrow p-Coumaroyl\ shikimic\ acid \rightarrow Caffeoyl\ shikimic\ acid \rightarrow Caffeoyl-CoA \rightarrow Caffeyl\ aldehyde \rightarrow Caffeyl\ alcohol \rightarrow Coniferyl\ alcohol \rightarrow 5-Hydroxy-coniferyl\ alcohol \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.98, 0.47, 0.50, 0.81, 0.61, 0.23, 0.65, 0.97, 0.80, 0.86, 0.96, 0.92	30.19
6	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow p-Coumaric\ acid \rightarrow p-Coumaroyl-CoA \rightarrow Caffeoyl-CoA \rightarrow Caffeyl\ aldehyde \rightarrow Caffeyl\ alcohol \rightarrow Coniferyl\ alcohol \rightarrow 5-Hydroxy-coniferyl\ alcohol \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.98, 0.95, 0.89, 0.80, 0.65, 0.97, 0.80, 0.86, 0.96, 0.92	19.93
7	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow Cinnamoyl-CoA \rightarrow p-Coumaroyl-CoA \rightarrow p-Coumaroyl\ quinic\ acid \rightarrow Caffeoyl\ quinic\ acid \rightarrow Caffeoyl-CoA \rightarrow Feruloyl-CoA \rightarrow 5-Hydroxy-feruloyl-CoA \rightarrow Sinapoyl-CoA \rightarrow Sinapaldehyde \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.98, 0.47, 0.50, 0.21 0.44, 0.68, 0.51, 0.29, 0.81, 0.80, 0.67, 0.92	20.73
8	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow Cinnamoyl-CoA \rightarrow p-Coumaroyl-CoA \rightarrow Caffeoyl-CoA \rightarrow Feruloyl-CoA \rightarrow 5-Hydroxy-feruloyl-CoA \rightarrow 5-Hydroxy-coniferaldehyde \rightarrow Sinapaldehyde \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.98, 0.47, 0.89, 0.80, 0.51, 0.29, 0.38, 0.94, 0.67, 0.92	27.21
9	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow p-Coumaric\ acid \rightarrow Caffeic\ acid \rightarrow Ferulic\ acid \rightarrow Feruloyl-CoA \rightarrow 5-Hydroxy-feruloyl-CoA \rightarrow Sinapoyl-CoA \rightarrow Sinapaldehyde \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.98, 0.95, 0.51, 0.55, 0.26, 0.29, 0.81, 0.80, 0.67, 0.92	33.67
10	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow p-Coumaric\ acid \rightarrow Caffeic\ acid \rightarrow Ferulic\ acid \rightarrow 5-Hydroxyferulicacid \rightarrow 5-Hydroxy-feruloyl-CoA \rightarrow Sinapoyl-CoA \rightarrow Sinapaldehyde \rightarrow Sinapylalcohol \rightarrow Syringyl\ lignin$	0.98, 0.95, 0.51, 0.55, 0.46, 0.15, 0.81, 0.80, 0.67, 0.92	25.56
11	$Phenylalanine \rightarrow Cinnamoyl-CoA \rightarrow p-Coumaroyl-CoA \rightarrow p-Coumaraldehyde \rightarrow p-Coumaryl\ alcohol \rightarrow Caffeyl\ alcohol \rightarrow Coniferyl\ alcohol \rightarrow 5-Hydroxy-coniferyl\ alcohol \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.43, 0.89, 0.91, 0.88, 0.85, 0.80, 0.86, 0.96, 0.92	17.58
12	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow Cinnamoyl-CoA \rightarrow p-Coumaroyl-CoA \rightarrow p-Coumaraldehyde \rightarrow p-Coumaryl\ alcohol \rightarrow Caffeyl\ alcohol \rightarrow Coniferyl\ alcohol \rightarrow 5-Hydroxy-coniferyl\ alcohol \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.98, 0.47, 0.89, 0.91, 0.88, 0.85, 0.80, 0.86, 0.96, 0.92	29.05
13	$Phenylalanine \rightarrow Cinnamic\ acid \rightarrow Cinnamoyl-CoA \rightarrow p-Coumaroyl-CoA \rightarrow p-Coumaroyl\ shikimic\ acid \rightarrow Caffeoyl\ shikimic\ acid \rightarrow Caffeoyl-CoA \rightarrow Caffeyl\ aldehyde \rightarrow Coniferyl\ aldehyde \rightarrow 5-Hydroxy-coniferaldehyde \rightarrow Sinapaldehyde \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.98, 0.47, 0.50, 0.81, 0.61, 0.23, 0.65, 0.43, 0.25, 0.94, 0.67, 0.92	36.24
14	$Phenylalanine \rightarrow p-Coumaric\ acid \rightarrow p-Coumaroyl-CoA \rightarrow p-Coumaroyl\ quinic\ acid \rightarrow Caffeoyl\ quinic\ acid \rightarrow Caffeoyl-CoA \rightarrow Feruloyl-CoA \rightarrow Coniferyl\ aldehyde \rightarrow Coniferyl\ alcohol \rightarrow 5-Hydroxy-coniferyl\ alcohol \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.35, 0.89, 0.21, 0.44, 0.68, 0.51, 0.17, 0.48, 0.86, 0.96, 0.92	32.55
15	$Phenylalanine \rightarrow p-Coumaric\ acid \rightarrow p-Coumaroyl-CoA \rightarrow p-Coumaroyl\ quinic\ acid \rightarrow Caffeoyl\ quinic\ acid \rightarrow Caffeoyl-CoA \rightarrow Feruloyl-CoA \rightarrow 5-Hydroxy-feruloyl-CoA \rightarrow 5-Hydroxy-coniferaldehyde \rightarrow Sinapaldehyde \rightarrow Sinapyl\ alcohol \rightarrow Syringyl\ lignin$	0.35, 0.89, 0.21, 0.44, 0.68, 0.51, 0.29, 0.38, 0.94, 0.67, 0.92	23.63

American trypanosomiasis, Chagas disease. The PPP has been shown to be functional in *T. cruzi* (see Fig. 12 of the online supplemental material) and the enzymes of the pathway have been expressed and are being fully characterized. The biochemical evidence obtained so far suggest that the oxidative branch of the PPP is essential for the protection of the parasite against oxidative stress. The existence of the respective metabolites in *T. cruzi* are explained in [30].

Glycolysis is a central metabolic pathway that is responsible for the production of numerous intermediary metabolites and energy in cells. Molecular phylogeny eventually revealed that archaea, like bacteria and eukaryotes, are a fundamentally distinct domain of life. The glycolytic pathways of *H. salinarum R1* and *N. pharaonis* (see Figs. 14 and 16 of the online supplementary file) have been reviewed in detail in [31]. The glycolytic pathway in *H. salinarum R1* as obtained by our proposed methodology has been observed in [32]. The glycolytic and PPPs are the primary sources of energy for all organisms and are of central importance for many biotechnological applications.

5.2 Flavonoid Biosynthesis Pathway

The starting metabolites for the three flavonoid biosynthetic subpathways are p-Coumaroyl-CoA (see Fig. 4). There are five paths emerging from p-Coumaroyl-CoA. Among them, only the path leading to Naringenin chalcone is followed as this leads to the desired target. Of the four paths from Naringenin chalcone, only the path leading to Naringenin is followed as it ultimately leads to the desired target. There are eight branches emerging from Naringenin. The path leading to Apiforol, 8-C-Glucosylnaringenin, Apigenin, Eriodictyol, Dihydrotricetin, Hesperetin and Naringin are not followed as they do not produce the required target. Of the four paths emerging from Dihydrokaempferol, only the path leading to Leucopelargonidin is followed to reach the

TABLE 2
Comparison of the Results for Different Pathways of the KEGG Database

Pathway/Species	λ	η	Number of Iterations		z-value	
			GD	Proposed Method	GD	Proposed Method
Synthetic reaction network	0.1	0.5	94	78	10.67	53.56
	0.2	0.5	94	76	13.56	51.11
	0.3	0.5	94	77	12.56	54.31
	0.4	0.5	95	77	14.45	53.89
	0.5	0.5	98	75	14.31	52.41
	0.6	0.5	99	78	13.69	53.61
	0.7	0.5	99	78	12.22	52.61
	0.8	0.5	99	79	13.61	54.28
	0.9	0.5	100	79	14.71	56.44
	1.0	0.5	100	80	13.75	55.39
PPP / *E. coli* K-12 MG1655	0.1	0.5	73	45	16.53	72.28
	0.2	0.5	74	46	15.78	73.56
	0.3	0.5	74	46	15.63	72.45
	0.4	0.5	76	47	14.43	71.42
	0.5	0.5	78	47	15.81	74.64
	0.6	0.5	78	48	14.64	68.91
	0.7	0.5	79	48	16.64	69.93
	0.8	0.5	79	49	17.72	69.53
	0.9	0.5	80	49	18.93	73.76
	1.0	0.5	80	50	15.79	74.61
PPP / *P. falciparum*	0.1	0.5	92	65	18.54	65.71
	0.2	0.5	92	66	17.65	67.53
	0.3	0.5	93	67	16.52	68.51
	0.4	0.5	93	67	15.78	68.51
	0.5	0.5	93	68	16.71	69.93
	0.6	0.5	94	68	17.84	65.82
	0.7	0.5	94	69	17.53	65.82
	0.8	0.5	95	69	16.83	68.32
	0.9	0.5	95	69	15.75	69.52
	1.0	0.5	95	69	16.49	67.52
PPP / *T. cruzi*	0.1	0.5	96	85	17.53	64.45
	0.2	0.5	96	85	15.56	63.21
	0.3	0.5	96	86	14.77	64.85
	0.4	0.5	97	87	16.63	66.53
	0.5	0.5	97	87	17.63	64.67
	0.6	0.5	98	88	16.73	65.86
	0.7	0.5	98	88	17.83	64.82
	0.8	0.5	98	88	15.73	66.51
	0.9	0.5	99	89	14.72	63.28
	1.0	0.5	99	89	13.67	61.56
Glycolytic Pathway / *H. salinarum R1*	0.1	0.5	86	45	16.62	64.62
	0.2	0.5	86	45	15.72	63.63
	0.3	0.5	87	45	15.54	63.93
	0.4	0.5	87	46	16.73	64.67
	0.5	0.5	88	46	15.56	63.59
	0.6	0.5	89	47	14.72	64.42
	0.7	0.5	89	47	14.72	61.71
	0.8	0.5	89	48	13.27	63.61
	0.9	0.5	90	48	14.62	64.72
	1.0	0.5	90	49	17.53	65.43
Glycolytic Pathway / *N. pharaonis*	0.1	0.5	76	55	20.77	66.89
	0.2	0.5	76	55	19.68	70.89
	0.3	0.5	77	56	18.97	68.95
	0.4	0.5	77	56	21.66	69.99
	0.5	0.5	78	57	18.88	70.56
	0.6	0.5	79	57	19.43	69.74
	0.7	0.5	79	58	20.51	71.58
	0.8	0.5	79	59	18.46	67.72
	0.9	0.5	80	59	19.56	68.65
	1.0	0.5	80	59	20.41	69.99
First Flavonoid biosynthesis pathway	0.1	0.5	83	55	17.66	62.33
	0.2	0.5	84	56	15.67	61.67
	0.3	0.5	86	56	16.78	64.43
	0.4	0.5	86	57	17.78	64.45
	0.5	0.5	88	57	18.89	63.75
	0.6	0.5	88	58	16.78	60.45
	0.7	0.5	89	58	18.66	64.74
	0.8	0.5	89	59	16.66	65.54
	0.9	0.5	90	59	15.57	61.78
	1.0	0.5	90	60	14.44	60.45
Second Flavonoid biosynthesis pathway	0.1	0.5	77	45	15.67	63.33
	0.2	0.5	77	45	15.77	62.57
	0.3	0.5	78	46	14.67	62.97
	0.4	0.5	78	47	13.56	64.20
	0.5	0.5	79	47	14.61	63.11
	0.6	0.5	79	48	13.48	61.18
	0.7	0.5	79	48	15.27	62.51
	0.8	0.5	79	49	13.45	61.11
	0.9	0.5	80	50	15.67	60.31
	1.0	0.5	80	50	14.67	63.41
Third Flavonoid biosynthesis pathway	0.1	0.5	97	64	16.34	63.45
	0.2	0.5	97	65	15.56	67.41
	0.3	0.5	98	65	14.42	65.41
	0.4	0.5	98	66	13.48	62.28
	0.5	0.5	99	67	17.52	66.61
	0.6	0.5	99	68	16.48	64.81
	0.7	0.5	98	68	13.26	61.24
	0.8	0.5	95	69	16.46	63.39
	0.9	0.5	99	69	15.62	65.41
	1.0	0.5	98	70	13.56	64.62
Phenylpropanoid biosynthesis pathway	0.1	0.5	78	55	15.53	64.51
	0.2	0.5	78	55	16.53	62.22
	0.3	0.5	77	57	17.53	67.41
	0.4	0.5	76	58	15.62	68.49
	0.5	0.5	77	56	14.65	66.53
	0.6	0.5	78	55	14.58	65.61
	0.7	0.5	79	58	16.67	64.44
	0.8	0.5	79	59	15.71	63.48
	0.9	0.5	80	59	14.57	64.62
	1.0	0.5	80	60	16.72	63.65

desired target Pelargonidin as the other paths terminate to some other target metabolites. The first optimal flavonoid biosynthetic pathway as obtained by our proposed method as explained above has been observed in [24].

The second flavonoid subpathway follows the same route till it reaches the intermediate metabolite Naringenin. The optimal path is through Eriodictyol and the other seven paths from Naringenin are not followed as they end up with some other intermediate metabolite. Of the five paths from Eriodictyol, the four paths leading to Luteolin, Luteoforol, Dihydrotricetin, and Homoeriodictyol are not followed as they terminate at some other metabolite other than the target, so the only path through Dihydroquercetin is followed to reach the target metabolite. Of the three paths starting from Dihydroquercetin, only the path leading to Leucocyanidin is followed to reach the target metabolite Cyanidin. The second flavonoid subpathway through the intermediate metabolites as obtained by our method, in contrary to the other two methods (GD and EPA), is observed in [23].

The third flavonoid subpathway follows the same route till it reaches the intermediate metabolite Naringenin. Of the eight paths emerging from Naringenin, only the path through Dihydrotricetin is followed as the optimal path and the other seven paths are not followed as they end up with some other intermediate metabolite. There are two paths emerging from Dihydrotricetin. The path that leads to Tricetin is not followed as this is not the desired target. The other path through Dihydromyricetin and then to Leucodelphinidin is followed as the optimal path. From Leucodelphinidin there are two paths leading to Delphinidin in one branch and Gallocatechin in another branch. The path that leads to Delphinidin is followed as this is the required target. The sequence of steps leading to the target metabolite Delphinidin as obtained by our proposed method, in contrary to the other two methods (GD and EPA), are cited in [33].

5.3 Phenylpropanoid Biosynthesis Pathway

The amino acid phenylalanine is the starting metabolite of the phenylpropanoid pathway in plant. Phenylalanine ammonia lyase (PAL; EC 4.3.1.24, 4.3.1.25), the first enzyme of this pathway, catalyzes the deamination of phenylalanine into cinnamic acid [34]. Oxidation of this aromatic acid by cinnamate 4-hydroxylase (C4H; EC 1.14.13.11), a cytochrome P450 enzyme, yields p-Coumaric acid. Next, p-Coumaroyl-CoA ligase (4CL; EC 6.2.1.12) attaches a CoA molecule to p-Coumaric acid, generating p-Coumaroyl-CoA. The three enzymes of the pathway (PAL, C4H and 4CL) (see Fig. 6) are highly conserved among plant species because they are important for normal growth and development [35]. The molecular foundations of phenylpropanoid synthesis in plants are well understood, as this pathway is arguably the best studied plant secondary metabolic pathway.

The pathway obtained by our proposed method starts from Phenylalanine and reaches Cinnamic acid as the intermediate product. Of the three emerging paths from Cinnamic acid, only the path leading to p-Coumaric acid is followed as it ultimately leads to the target metabolite. The other two paths are not followed as they terminate at Cinnamaldehyde and Coumarine, which are not the desired target metabolites. Of the three emerging paths from p-Coumaric acid, the path leading to p-Coumaroyl-CoA is followed as it ultimately leads to the target metabolite. There are four paths emanating from p-Coumaroyl-CoA. The path leading to p-Coumaraldehyde is followed as it terminates to the desired target. The other three paths terminates to the intermediate metabolite Caffeoyl-CoA. Of the two paths emerging from p-Coumaraldehyde, the path leading to p-Coumaryl alcohol is followed as it ends up with the desired target. The sequence of paths starting

from Phenylalanine till it reaches the intermediate metabolite p-Coumaryl alcohol has been observed extensively in [35] and [25]. The path obtained by our methodology coincides with the sequence of steps described above till p-Coumaryl alcohol, in contrary to the path derived by EPA and GD formalism. The path leading to Caffeyl alcohol from p-Coumaryl alcohol is followed till it reaches the target metabolite and the other two paths are not followed as they terminate at some other target metabolite. From Caffeyl alcohol, we arrive at Coniferyl alcohol, which has three paths emerging from it. The path leading to 5-Hydroxy-coniferyl alcohol is followed and the other two paths do not lead to the desired target. 5-Hydroxy-coniferyl alcohol leads to Sinapyl alcohol, which ultimately leads to the desired target Syringyl lignin. The sequence of steps from p-Coumaryl alcohol to the target metabolite Syringyl lignin is the same as obtained by our proposed methodology.

The target metabolite Syringyl lignin, the 3D polymer component of plant cell walls, is one of the most important products derived from phenylpropanoids. The intermediate metabolites p-Coumaric and Ferulic acids derived by our proposed method and the GD optimization formalism are the major precursors for the synthesis of the monolignols p-coumaryl, coniferyl and sinapyl alcohols that are the monomers for the subsequent synthesis of p-hydroxyphenyl, guaiacyl and syringyl lignins.

5.4 Impact on Metabolic Engineering

With increase in the understanding and documentation of the biological properties, and in many cases, potential beneficial effects of flavonoids and phenylpropanoids, there is increasing interest in the engineering of their metabolism [36]. An obvious approach toward this end is to exploit the structural genes encoding enzymes of the general phenylpropanoid metabolism, whose coordinate expression leads to its production. Transcription factors (TFs) may be exploited in the metabolic engineering of flavonoid and phenylpropanoid metabolism [37] as:

1. TFs typically control the expression of multiple genes encoding enzymes in a given pathway, allowing efficient manipulation of multienzyme pathways;
2. ectopic expression of specific TFs can be used as a tool for redirecting metabolic differentiation of the cells;
3. pathway-specific TFs could be used to modulate the production of specific secondary metabolites; and
4. inactivation of transcriptional repressors could be used to derepress metabolic channeling into a pathway leading to a specific metabolite.

By using TFs that are known to regulate flavonoid and phenylpropanoid metabolism, various metabolic engineering results can be achieved. It has been reported [38] that the production of naringenin, eriodictyol, dihydrokaempferol, dihydroquercetin, kaempferol, and quercetin can be done by the coexpression up to eight genes in the flavonoid metabolic pathway when the strains are provided in the precursor p-coumaric acid.

The PPPs and glycolysis pathways comprise the most central pathways in primary metabolism. The PPP is believed to be the major source of NADPH required for many biosynthetic and detoxification reactions. The flux through this pathway has been reported to increase at high NADPH requirements and to decrease when the need for production is decreased. The availability of NADPH in whole-cell systems might be increased by metabolic pathway engineering, i.e., by overproduction of enzymes in the PPP. NADPH is produced in two of the steps in the PPP, namely, the conversion of glucose 6-phosphate (G6P) to 6-phosphoglucono-d-lactone (6PGdL), catalyzed by glucose 6-phosphate dehydrogenase (G6PDH; EC 1.1.1.49), and conversion of 6-phosphogluconate (6PG) to ribulose 5-phosphate (Ru5P), catalyzed by 6-phosphogluconate dehydrogenase (6PGDH; EC 1.1.1.44). In the nonoxidative part of the PPP, two out of three reactions are catalyzed by transketolase (TKT; EC 2.2.1.1). Overproduction of these enzymes (G6PDH, 6PGDH, TKT) (see Figs. 8, 10, and 12 of the online supplemental material) might lead to increased flux through the PPP [39]. This will lead to enhanced yield of the target metabolite, which is the ultimate objective of our methodology.

Here we consider the synthetic reaction system of Fig. 3, where $c_1, c_2, \ldots, c_{11}$ denotes the enzyme concentration of the reactions $R_1, R_2, \ldots, R_{11}$, respectively. As mentioned in Section 4.1, the optimal pathway for this system is as follows $R_1 \rightarrow R_2 \rightarrow R_3 \rightarrow R_8 \rightarrow R_{10} \rightarrow R_{11}$. Subsequently, the enzyme concentration for this optimal path is $c_8 = 0.94,\ c_1 = 0.90,\ c_2 = 0.88,\ c_5 = 0.86,\ c_7 = 0.79,\ c_{11} = 0.78$. If the concentration of any enzyme of the optimal pathway becomes less (i.e., ~ 0), then the yield of the target metabolite will also decrease accordingly. To remedy this problem, we can apply our present second-order method. Thus, we can take adequate steps to activate the corresponding gene to enhance the concentration of that enzyme which ultimately leads to the overproduction of the desired target metabolite E.

Similarly, we again consider the flavonoid biosynthesis pathway of Fig. 4. The enzyme concentration corresponding to the 92 reactions for the pathway are $c_1, c_2, \ldots, c_{92}$. The enzyme concentration for the three optimal subpathways of Fig. 5 are 0.95, 0.94, 0.91, 0.88, 0.89 for the first optimal path; 0.95, 0.94, 0.86, 0.81, 0.79, 0.77 for the second optimal path; 0.95, 0.94, 0.85, 0.81, 0.72, 0.70 for the third optimal path. If we want to avoid any of these three pathways, we have to inhibit the genes producing some of the enzymes catalyzing the reactions in that pathway.

6 Conclusion and Discussion

A simple second-order learning algorithm has been developed in this paper, which represents an alternative to standard GD algorithm. Our approach relies crucially on the modified version of the Newton-Raphson method and is based on well-known flux balancing approach. One of the primary advantages of this method is that it is independent of the learning parameter. This global optimization technique for nonlinear optimization and neural network learning identifies an optimal metabolic pathway through which a metabolite attains a maximum rate of growth on a given substrate. The method incorporates weighting coefficients indicating the concentration levels of enzymes catalyzing biochemical reactions in the pathway.

The methodology presented here was tested on eight biologically significant networks. Extensive computer simulations and performance comparisons with the GD algorithm and the EPA algorithm have led to interesting insights. This new approach can model very efficiently even the big metabolic networks and provides means to perform *in silico* experiments on them. One of the foundations of our approach is the fact that it can identify the optimal metabolic pathways which conform to the results of some earlier studies. All the optimal metabolic pathways obtained by the present method in this work have greater biological significance as compared to that obtained by the GD and EPA method. From these applications, it is clear that this method is very efficient in modeling the dynamics of metabolic networks and can be used, to some extent, to understand the development of complex organisms. We emphasize that our approach is not restricted to only metabolic networks.

REFERENCES

[1] J.A. Papin, N.D. Price, S.J. Wiback, D.A. Fell, and B.O. Palsson, "Metabolic Pathways in the Post-Genome Era," *Trends in Biochemical Sciences,* vol. 28, pp. 250-258, 2003.

[2] K.Y. Tsai and F.S. Wang, "Evolutionary Optimization with Data Collocation for Reverse Engineering of Biological Networks," *Bioinformatics,* vol. 21, pp. 1180-1188, 2005.

[3] W.W. Chen, B. Schoeberl, and P.J. Jasper, "Input-Output Behavior of ErbB Signaling Pathways as Revealed by a Mass Action Model Trained Against Dynamic Data," *Molecular Systems Biology,* vol. 5, article 239, 2009.

[4] M. Sugimoto, S. Kikuchi, and M. Tomita, "Reverse Engineering of Biochemical Equations from Time-Course Data by Means of Genetic Programming," *Biosystems,* vol. 80, pp. 155-164, 2005.

[5] J.M. Lee, E.P. Gianchandani, and J.A. Papin, "Flux Balance Analysis in the Era of Metabolomics," *Briefings in Bioinformatics,* vol. 7, pp. 140-150, 2006.

[6] S. Becker and Y.L. Cun, "Improving the Convergence of Back-Propagation Learning with Second Order Methods," *Proc. Connectionist Models Summer School Conf.,* pp. 29-37, 1989.

[7] S. Haykin, *Neural Networks: A Comprehensive Foundation.* Prentice Hall, 1999.

[8] M. Gori and A. Tesi, "On the Problem of Local Minima in Backpropagation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 1, pp. 76-86, 1992.

[9] C.A. Murthy, M. Das, R.K. De, and S. Mukhopadhyay, "Determination of Optimal Metabolic Pathways through a New Learning Algorithm," *Proc. Int'l Conf. Pattern Recognition (ICPR '08),* pp. 1-4, Dec. 2008.

[10] J. Fliege, L.M.G. Drummond, and B.F. Svaiter, "Newton's Method for Multiobjective Optimization," *SIAM J. on Optimization,* vol. 20, pp. 602-626, 2009.

[11] C.H. Schilling, D. Letscher, and B.O. Palsson, "Theory for the Systemic Definition of Metabolic Pathways and Their Use in Interpreting Metabolic Function from a Pathway-Oriented Perspective," *J. Theoretical Biology,* vol. 203, pp. 229-248, 2000.

[12] K. Schallau and B.H. Junker, "Simulating Plant Metabolic Pathways with Enzyme-Kinetic Models," *Plant Physiology,* vol. 152, pp. 1763-1771, 2010.

[13] R. Adadi, B. Volkmer, R. Milo, M. Heinemann, and T. Shlomi, "Prediction of Microbial Growth Rate versus Biomass Yield by a Metabolic Network with Kinetic Parameters," *PLoS Computational Biology,* vol. 8, article e1002575, 2012.

[14] K.J. Kauffman, P. Prakash, and J.S. Edwards, "Advances in Flux Balance Analysis," *Current Opinion in Biotechnology,* vol. 14, pp. 491-496, 2003.

[15] B. Teusink, A. Wiersma, L. Jacobs, R.A. Notebaart, and E.J. Smid, "Understanding the Adaptive Growth Strategy of *Lactobacillus plantarum* by *In Silico* Optimisation," *PLoS Computational Biology,* vol. 5, no. 6, article e1000410, 2009.

[16] E.P. Gianchandani, M.A. Oberhardt, A.P. Burgard, C.D. Maranas, and J.A. Papin, "Predicting Biological System Objectives De Novo from Internal State Measurements," *BMC Bioinformatics,* vol. 9, article 43, 2008.

[17] R.K. De, M. Das, and S. Mukhopadhyay, "Incorporation of Enzyme Concentrations into FBA and Identification of Optimal Metabolic Pathways," *BMC Systems Biology,* vol. 2, article 65, 2008.

[18] E. Mizutani and S.E. Dreyfus, "Second-Order Stagewise Backpropagation for Hessian-Matrix Analyses and Investigation of Negative Curvature," *Neural Networks,* vol. 21, pp. 193-203, 2008.

[19] R. Pasti and L.N. Castro, "Bio-Inspired and Gradient-Based Algorithms to Train MLPs: The Influence of Diversity," *Information Sciences,* vol. 179, pp. 1441-1453, 2009.

[20] B.X. Chen, Q. Lin, S. Kim, J.G. Carbonell, and E.P. Xing, "Smoothing Proximal Gradient Method for General Structured Sparse Regression," *Annals of Applied Statistics,* vol. 6, pp. 719-752, 2012.

[21] N.N. Schraudolph, "Fast Curvature Matrix-Vector Products for Second-Order Gradient Descent," *Neural Computation,* vol. 14, pp. 1723-1738, 2002.

[22] C.H. Schilling and B.O. Palsson, "The Underlying Pathway Structure of Biochemical Reaction Networks," *Proc. Nat'l Academy of Sciences USA,* vol. 95, pp. 4193-4198, 1998.

[23] W.S. Brenda, "Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology," *Plant Physiology,* vol. 126, pp. 485-493, 2001.

[24] L. Ralston, S. Subramanian, M. Matsuno, and O. Yu, "Partial Reconstruction of Flavonoid and Isoflavonoid Biosynthesis in Yeast Using Soybean Type I and Type II Chalcone Isomerases," *Plant Physiology,* vol. 137, pp. 1375-1338, 2005.

[25] O. Yu and J.M. Jez, "Natures Assembly Line: Biosynthesis of Simple Phenylpropanoids and Polyketides," *Plant J.,* vol. 54, pp. 750-762, 2008.

[26] J. Edahiro and M. Seki, "Phenylpropanoid Metabolite Supports Cell Aggregate Formation in Strawberry Cell Suspension Culture," *J. Bioscience and Bioeng.,* vol. 102, pp. 8-13, 2006.

[27] A. Sillero, V.A. Selivanovy, and M. Cascante, "Pentose Phosphate and Calvin Cycles," *Biochemistry and Molecular Biology Education,* vol. 34, pp. 275-277, 2006.

[28] Z. Bozdech and H. Ginsburg, "Data Mining of the Transcriptome of *Plasmodium falciparum*: The Pentose Phosphate Pathway and Ancillary Processes," *Malaria J.,* vol. 4, pp. 1-12, 2005.

[29] A.R. Montoya, W.P. Lee, S. Bassilian, S. Lim, R.V. Trebukhina, M.V. Kazhyna, C.J. Ciudad, V. Noe, J.J. Centelles, and M. Cascante, "Pentose Phosphate Cycle Oxidative and Nonoxidative Balance: A New Vulnerable Target for Overcoming Drug Resistance in Cancer," *Int'l J. Cancer,* vol. 119, pp. 2733-2741, 2006.

[30] A.L. Stern, E. Burgos, L. Salmon, and J.J. Cazzulo, "Ribose 5-Phosphate Isomerase Type B from *Trypanosoma cruzi:* Kinetic Properties and Site-Directed Mutagenesis Reveal Information about the Reaction Mechanism," *Biochemical J.,* vol. 401, pp. 279-285, 2007.

[31] M. Falb, K. Muller, L. Konigsmaier, T. Oberwinkler, P. Horn, S. Gronau, O. Gonzalez, F. Pfeiffer, E. Bornberg-Bauer, and D. Oesterhelt, "Metabolism of Halophilic archaea," *Extremophiles,* vol. 12, pp. 177-196, 2008.

[32] O. Gonzalez, S. Gronau, M. Falb, F. Pfeiffer, E. Mendoza, R. Zimmer, and D. Oesterhelt, "Reconstruction, Modeling and Analysis of *Halobacterium salinarum R-1* Metabolism," *Molecular Biosystems,* vol. 4, pp. 148-159, 2008.

[33] U. Yukiko, K. Yukihisa, F. Yuko, F.M. Masako, O. Hideo, K. Takaaki, I. Takashi, and T. Yoshikazu, "Molecular Characterization of the Flavonoid Biosynthetic Pathway and Flower Color Modification of *Nierembergia* sp." *Plant Biotechnology,* vol. 23, pp. 19-24, 2006.

[34] M.J. MacDonald and G.B. DCunha, "A Modern View of Phenylalanine Ammonia Lyase," *Biochemistry and Cell Biology,* vol. 85, pp. 273-282, 2007.

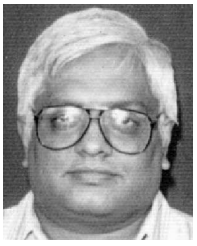
[35] D. Ro and C.J. Douglas, "Reconstitution of the Entry Point of Plant Phenylpropanoid Metabolism in Yeast (*Saccharomyces cerevisiae*): Implications for Control of Metabolic Flux into the Phenylpropanoid Pathway," *J. Biological Chemistry,* vol. 279, pp. 2600-2607, 2004.

[36] P. Gantet and J. Memelink, "Transcription Factors: Tools to Engineer the Production of Pharmacologically Active Plant Metabolites," *Trends in Pharmacological Sciences,* vol. 23, pp. 563-569, 2002.

[37] R. Koes, W. Verweij, and F. Quattrocchio, "Flavonoids: A Colorful Model for the Regulation and Evolution of Biochemical Pathways," *Trends in Plant Science,* vol. 10, pp. 236-242, 2005.

[38] E. Leonard, Y. Yan, and M.A.G. Koffas, "Functional Expression of a P450 Flavonoid Hydroxylase for the Biosynthesis of Plant-Specific Hydroxylated Flavonols in *Escherichia coli*," *Metabolic Eng.*, vol. 8, pp. 172-181, 2006.
[39] B.R. Poulsen, J. Nhr, S. Douthwaite, L.V. Hansen, J.J.L. Iversen, J. Visser, and G.J.G. Ruijter, "Increased NADPH Concentration Obtained by Metabolic Engineering of the Pentose Phosphate Pathway in *Aspergillus niger*," *FEBS J.*, vol. 272, pp. 1313-1325, 2005.

Mouli Das received the BSc and MSc degrees in physics from the University of Calcutta, India, in 2001 and 2003, respectively. Since 2004, she has been a research fellow with the Machine Intelligence Unit, Indian Statistical Institute, Kolkata, where she is currently working toward the PhD degree. Her current research interests include computational biology, pattern recognition, soft computing, and bioinformatics. Her work has been credited in about 10 publications in reputed journals and conferences. At the 2008 International Conference on Pattern Recognition, she received a travel support sponsorship from the Google India Diversity Team and Microsoft Research. She has acted as a reviewer for international journals and conferences.



C.A. Murthy received the BStat (Hons), MStat, and PhD degrees from the Indian Statistical Institute (ISI), Kolkata, India. In 1991-1992, he spent 6 months visiting Michigan State University, East Lansing, and in 1996-1997, he visited the Pennsylvania State University, University Park, for 18 months. He is a professor in the Machine Intelligence Unit of ISI. His fields of research interest include pattern recognition, image processing, machine learning, neural networks, fractals, genetic algorithms, wavelets, and data mining. He received the Best Paper Award in 1996 in computer science from the Institute of Engineers, India. In 1999, he and his two collegues received the Vasvik Award for electronic sciences and technology. He is a fellow of the National Academy of Engineering, India, and the National Academy of Sciences, India. He was the head of the Machine Intelligence Unit, ISI, from 2005 to 2010.

Rajat K. De received the bachelor's of technology degree in computer science and engineering and the master's of computer science and engineering degree from Calcutta University and Jadavpur University, India, in 1991 and 1993, respectively, and the PhD degree from the Indian Statistical Institute in 2000. He is a professor at the Indian Statistical Institute, Kolkata. During 2002-2003, he was a distinguished postdoctoral fellow at the Whitaker Biomedical Engineering Institute, Johns Hopkins University. He has about 60 research articles published in international journals, conference proceedings, and edited books. His research interests include bioinformatics, computational biology, systems biology, pattern recognition, and soft computing. He is a senior member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.