Original Article

MODELLING AND FLUX BALANCE ANALYSIS OF THE HUMAN APOPTOTIC PATHWAY

UDAYAKUMAR MANI*a, AJITHAVALLI Ca, VIGNESH Kb, SAI MUKUND Ra

aDepartment of Bioinformatics, bDepartment of Biotechnology, Sastra University, Thanjavur 613401, Tamil Nadu, India
Email: uthay@bioinfo.sastra.edu

ABSTRACT

Objective: Flux balance analysis is one method to analyse genome-scale models for metabolism and transcriptional regulation. Her we use the apoptotic pathway, the most widely used FBA to perform a mathematical model also analyses the flux balance to study the relationship between the flux and the genes involved in the extrinsic pathway.

Methods: KEGG was used to obtain information of pathways and related ligands and genes. Biocyc was considered for its intricacy in explanation of the mechanisms of pathways. A specific wing of Biocyc called humancyc was used for human related pathways. SBML was used to sort the several pathways.

Results: The pathway was modelled using SBML version 2 level 1 specifications which are the most commonly used version making FBA one of the best methods for studying this process. Based on FBA, the genes like bid and cyc were classified as essential and non-essential. All these details suggest that our model is sensible scientifically. This technique can satisfy researchers to uncover several biologically complex pathways computationally. This is a more efficient and less time consuming process.

Conclusion: With the availability of experimental biochemical data, like the concentration or the ratio of these caspases in a cell, the accuracy of this method can be improved. In our method, the entire model is considered to be present in a single compartment. Separate compartments for the inclusion of inhibitors and activators can be added to represent the pathway more effectively. A more complex model including the entire apoptotic pathway can be modelled to determine all possible ways of inhibiting apoptosis.

Keywords: Apoptosis, Metabolic pathway modelling, Flux Profile, KEGG, System Biology Markup Language.

INTRODUCTION

Metabolic pathway modelling refers to the process of modelling the cellular metabolic network. Metabolism is the chemical engine that drives the living process. Through the utilization of a vast repertoire of enzymatic reactions and transport processes, unicellular and multicellular organisms can process and convert thousands of organic compounds into the various biomolecules necessary to support their existence [1].

Mathematical models are the most precise representation of knowledge. Even if the model is not used for simulation, it can help us to focus the attention on the essential parts of the system. The reproduction of experimental data by mathematical models is a well-established methodology in all scientific disciplines. However, it must be pointed out that in most cases this is merely a reproduction of measured data. It cannot be used to conclude whether the respective model is a "good" one or has any predictive power. Without further measures, the fitting of models to data tends to produce a consistent "interpretation" rather than an "explanation" of the empirical results. So, in order to obtain an explanation of the empirical results, we perform a series of procedures including Flux Balance Analysis to find the significant and non-significant genes in our modelled pathway. Even for systems with only five or six metabolites and a few modulators, the number of parameters is typically in the dozens and grows much faster than linearly with the number of variables involved. Metabolic modellers have been responding to this challenge in two ways [2]. Either they have limited the amount of necessary information to the stoichiometry of fluxes, thereby avoiding the need for kinetic details like regulatory feedback signals, or else fully kinetic models have been constructed with *in vivo* or *in vitro* information from different sources and, quite often, different organisms. In the absence of proper wet lab data, computational biologists can exclude the need of reaction kinetics and use only stoichiometric knowledge to gain information about the reaction fluxes. This is achieved by using computational methods like FBA and it is of immense advantage to computational biologists [3].
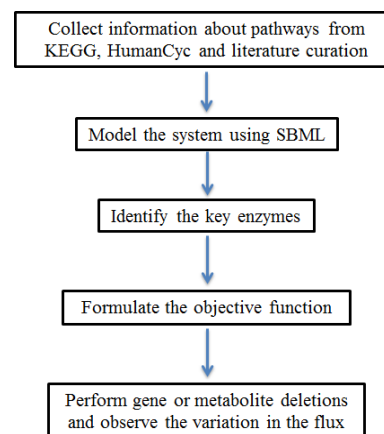

Fig. 1: Flow chart of the flux balance analysis

Apoptosis or programmed cell death is a normal component of the development and health of multicellular organisms [4]. Cells die in response to a variety of stimuli and during apoptosis they do so in a controlled, regulated fashion. This makes apoptosis distinct from another form of cell death called necrosis in which uncontrolled cell death leads to lysis of cells, inflammatory responses and, potentially, to serious health problems. Upon receiving specific signals, instructing the cells to undergo apoptosis, a number of distinctive changes occur in the cell. Apoptosis is a hugely complicated pathway. There are a lot of genes and processes involved in this pathway. Our main aim is to identify the essential (genes that contribute significantly to the pathway) and non-essential (genes that do not contribute significantly to the pathway) genes involved in the extrinsic pathway. Caspases–which are Cysteine-dependent aspartate-specific proteases, are the key proteins involved in carrying out the apoptosis. Though there is evidence for caspase-independent apoptosis [5], we deal only with the caspase-dependent apoptosis in our work.

A lot of work has already been carried out in the apoptotic pathway, but and our work comprises of mathematically modelling the pathway and performing FBA to identify essential and non-essential genes, which is purely computational and hasn't been done before. This method of mathematical modelling has been done for predicting drug targets in micro-organisms [6] but hasn't been employed for studying cellular signalling processes, especially apoptosis.

MATERIALS AND METHODS

KEGG [7], the Kyoto Encyclopedia for Genes and Genomes, is a knowledge base that was employed to extract information about pathways. KEGG is organized into three main categories, the PATHWAY database, the GENES database and the LIGAND database, among which the first two were exploited particularly for pathway modeling and for identifying the substrates for performing flux balance analysis. Bio Cyc [8], a collection of pathways, has an advantage over KEGG with its intricate explanations for reactions which includes details about their mechanisms (if available), cofactors, energy usage/discharge and the extensive commentary provided for the MetaCyc pathways. A wing of BioCyc called HumanCyc deals with pathways in Homo sapiens. It was employed for filling in missing information and enhancing the accuracy of the generated model. SBML refers to the Systems Biology Markup Language, a very powerful XML-based language. It is mainly used for storing and representing biological pathway data. Therefore, this language was utilized for modelling the pathway. The features of SBML are (a) SBML includes many simple tags and descriptors that help modellers define the pathway. b) Metabolites can be specified as either reactants or products and additional information about reversibility of the reaction, rates of the reaction, initial volume and concentration, inhibitory or accelerating nature of the reaction and ontologies can be specified. c) Metabolites or the entities involved in the pathway are represented as nodes while the reactions are represented by arrows or edges. SBML Version 2 Level 1 specifications were used to code the model. PHP, a hypertext preprocessing language, was used to develop images of flux profiles from the velocity data obtained. Flux Balance Analysis (FBA) is the method of analyzing a metabolic system of interactions for flux profiles. Flux refers to the rate of flow of a particular substance through a unit cross-sectional area. FBA aims at solving equations for fluxes in contrast to many other metabolic pathway analysis methods that use fluxes in the systems to find other parameters [9]. This is a huge advantage as computational biologists have a predominantly stronger dry lab perspective on problems which makes it difficult for them to calculate fluxes of various units in a pathway model via dry lab procedures (provided that KEGG and Biocyc do not include fluxes in their databases). A proper understanding of a metabolite's flux can help us provide a link to its availability in an environment and thereby its involvement in reactions pertaining to that area. An additional key concept involved in the analysis is that when a node (say A) in the metabolic pathway is deleted, the fluxes of metabolites falling in the cascade of reactions initiated by that node may vary. This fluctuation is mapped in our process which helps us explain the effect of each metabolite on others. Such an analysis will lead us towards the determination of essential (those which significantly alter the fluxes) and non-essential metabolites (those which don't significantly alter the fluxes). Further, identification of the genes corresponding to the metabolites can help us perform an *in silico* gene deletion study.

Solving of LPP using matlab

As mentioned in the flowchart, an objective function (c) has to be formed. The objective function aids in predicting the fluxes of altered metabolic pathways with respect to the optimum condition prevalent in the system. Examples of the objective function include Biomass production (which will be maximized) and nutrient uptake (which will be minimized). Imposing a constraint that ensures the steady state will bring about meaningful fluxes. For a pathway with a stoichiometric matrix S, and a flux vector, V, the steady state condition is given by:

$$S \cdot V = 0 \left(i.e. \frac{dx}{dt} = 0 \right)$$

Minimization/Maximization of the Objective function with the steady state constraint imposed, results in a Linear Programming

problem as the function has to be minimized or maximized with respect to a constraint, where both the constraint and the function are linearly related to the solution vector (here, it is our flux vector that is in a linear relation to both the objective function and the steady state constraint). It is solved using the linprog module present in MATLAB [10]. The work was split into two parts. The first step involved modelling of the extrinsic apoptotic pathway followed by performing a flux balance analysis on the network produced. The extrinsic pathway information was obtained from KEGG database (KEGG ID: hsa04210). Using SBML, the pathway was modelled and was validated for accuracy using HumanCyc. Modelled Pathway involved 24 reactions and 22 metabolites shown in fig. 2.
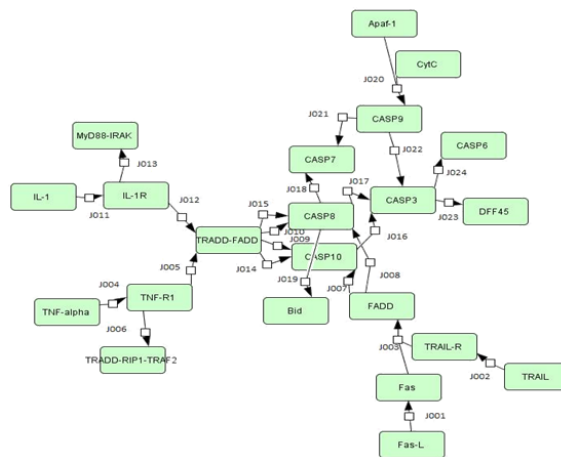
Fig. 2: Extrinsic apoptotic pathway model. This was done using SBML version 2 Level 1 specifications

It was observed and stoichiometric matrices were obtained for all the 24 reactions. A matrix having of order 24x22 was obtained, in which rows represented metabolites and the columns represented reactions which are shown in fig. 3.

Fig. 3: The stoichiometry matrix of our modeled pathway, having 22 rows and 24 columns

The objective function (equation 1.2) was formed based on the understanding that the three effector caspases (Caspases 3, 6, 7) are activated by three initiator caspases (Caspases 8, 9, 10).

$$c. v = 0.22Vcasp10 - casp3 + 0.22Vcasp8 - casp3 + 0.165Vcasp8 - casp7 + 0.22Vcasp9 - casp3 + 0.165Vcasp9 - casp7 \dots \rightarrow \text{Equation (1.2)}$$

This objective function represents the situation responsible for the production of the caspases that are involved as key players in the extrinsic pathway. A minimization of the objective function will be the aim of a natural biological system when it is undergoing apoptosis. If a cell should receive the signals properly, it has to be in a state of homeostasis. Therefore, a steady state is expected to prevail till apoptosis commences. To ensure this, a steady state

113

condition is imposed onto the system when it is minimized. Insilico gene deletion studies were performed with the gene listed in table 1.

Table 1: Some of the genes and the affected metabolites

| Gene	Affected metabolites
Bid	BID
Cyc	CytC
Fadd	FADD, TRADD-FADD
Tnfa	TNF-α
Fas	FAS

The first step involved deleting the corresponding node in the pathway and hence equating its stoichiometry's to zero. The resulting stoichiometric matrix was put into the linear programming problem and the problem was solved for the flux rate vector. Different flux profiles are depicted in fig. 4. While performing these gene deletion studies, a few reaction fluxes dropped down to zero upon deletion of certain genes while a few genes when deleted didn't alter the fluxes much. The fluxes obtained upon deletion of the genes were compared with the wild type fluxes. Based on these observations, the genes were classified as either essential or non-essential. The various flux profiles are displayed below.
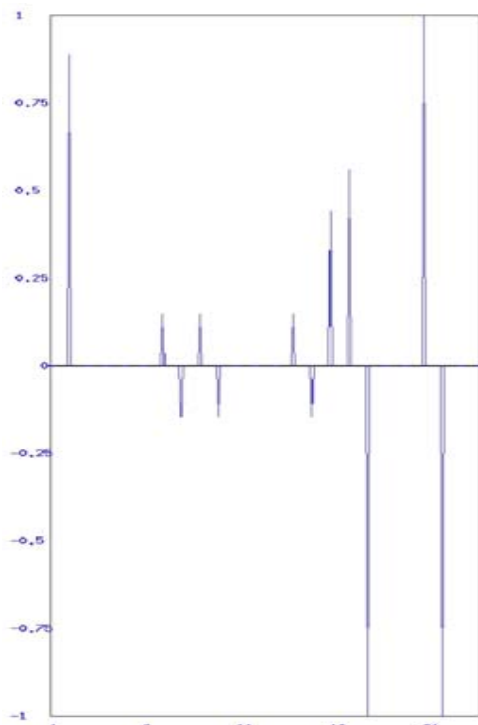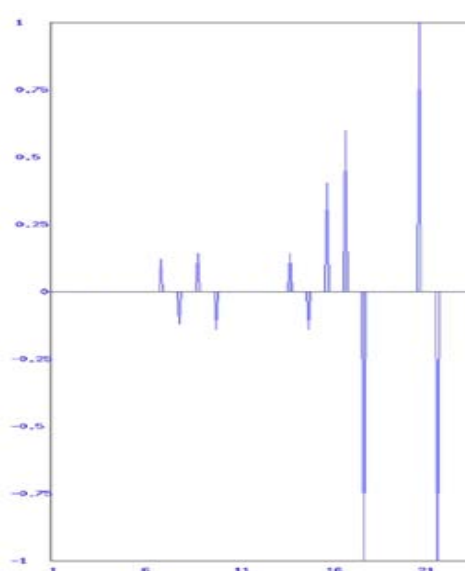

Fig. 4.1: Flux profile obtained by deleting trail gene
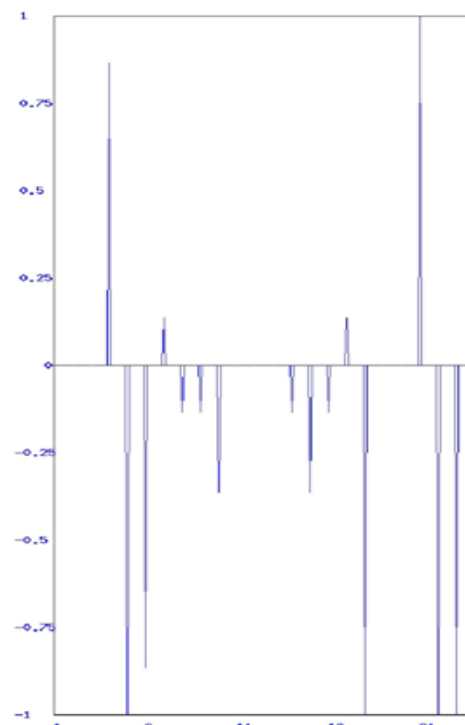
Fig. 4.2: Flux profile obtained by deleting TNF gene
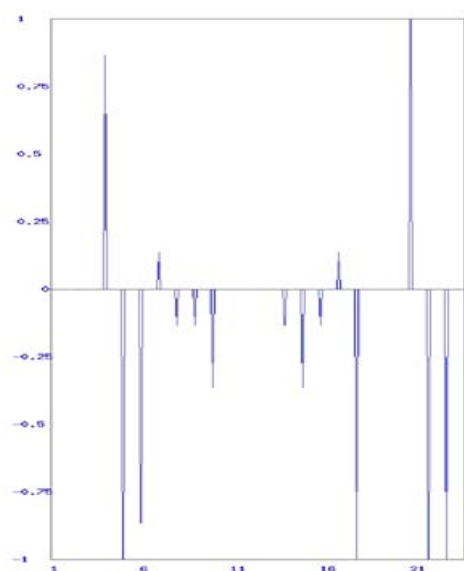
Fig. 4.3: Flux profile obtained by deleting TNF-r gene

Fig. 4.4: Flux profile obtained of the wild-type (without any deletion)

114
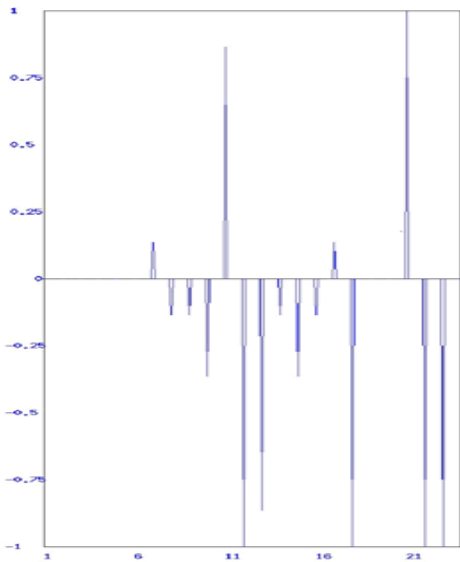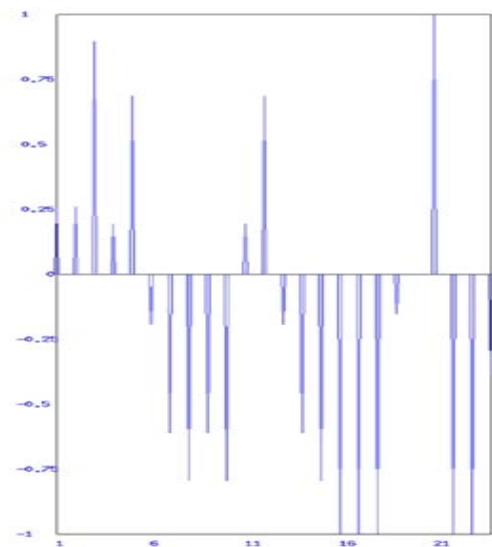
Fig. 4.5: Flux profile obtained by deleting il1r

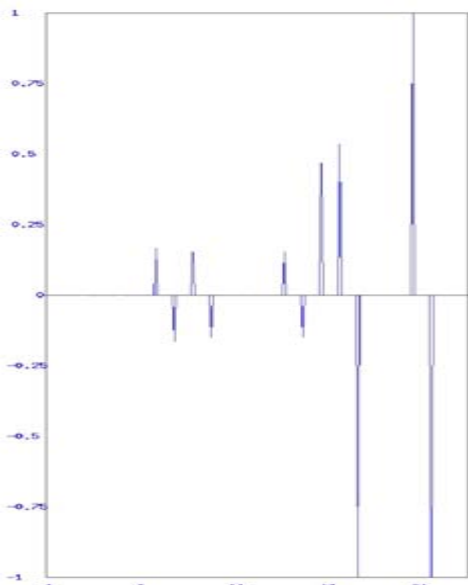**Fig. 4.6: Flux profile obtained by Deleting FADD-o.**

Fig. 4.7: Flux profile obtained by deleting cytc
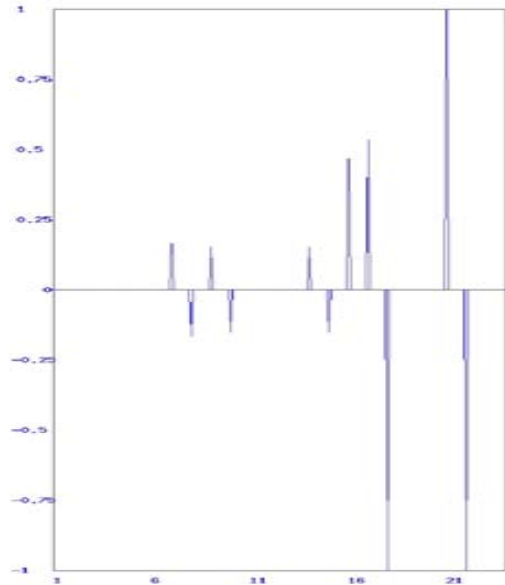

Fig. 4.8: Flux profile obtained by deleting apaf-1

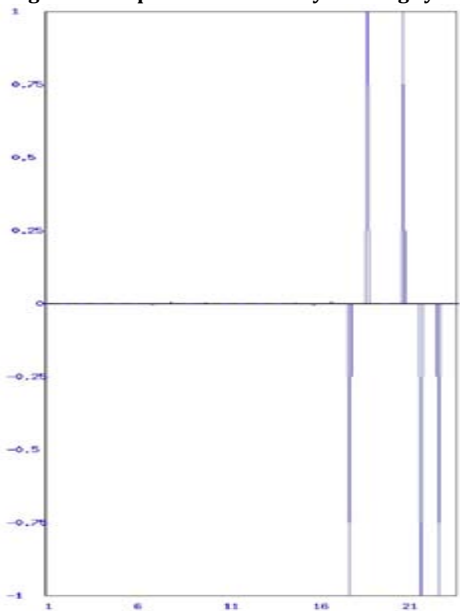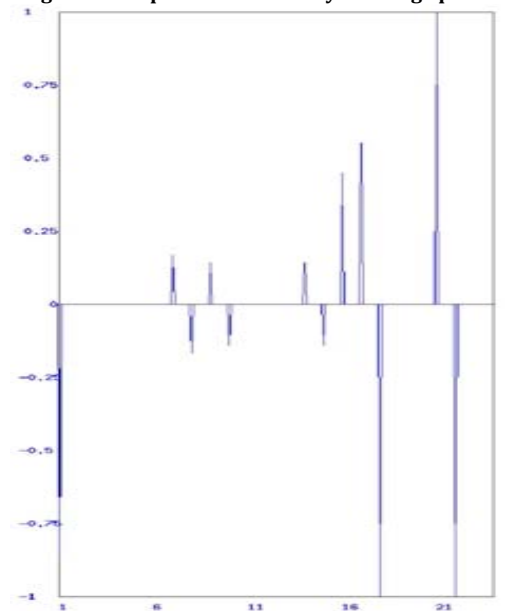**Fig. 4.9: Flux profile obtained by deleting bid**

Fig. 4.10: Flux profile obtained by deleting fasl

"Drugs from Nature: Plants as an important source of pharmaceutically important metabolites"
Guest Editor: Dr. Dhananjaya Bhadrapura Lakkappa

RESULTS AND DISCUSSION

The pathway was modelled using SBML version 2 level 1 specifications which is the most commonly used version. Since these pathways are very complex, it is not feasible to study them *in vivo*, thereby necessitating the use of mathematical modelling to study these pathways in-silico. Flux Balance Analysis (FBA) was done to determine the reaction fluxes. FBA was chosen because it doesn't require knowledge of the reaction kinetics and could be employed based on the stoichiometry's of the various metabolites alone. It is very comfortable for computational biologists as they don't have enough experience in a wet lab. These criteria make FBA one of the best methods for studying this process. Also, it is simple to be implemented and requires less time and memory. Based on FBA, the genes were classified as essential and non-essential. Some of the essential genes include bid, cyc etc. These genes code for Bid and CytC respectively. Bid is a BH3 interacting death domain. Experimental results suggest that it is essential for activating Bax or Bcl-2 associated X-protein. This results in the release of Cytochrome C from mitochondria, which is a very important step in apoptosis. So this gene is a very significant one. Similarly, Cytochrome C release from mitochondria results in apoptosome formation. Apoptosomes are the key factors which help in the conversion of procaspase 3 to caspase 3. This caspase is involved in the cleavage of actin filaments and thus disrupts the cellular cytoskeleton, resulting in the formation of cellular blebs. So CytC is also a potential target for inhibiting apoptosis. Few other genes like trail-r are comparatively less significant. Its products like TRAIL-R which is involved in the formation of FADD, which can also be obtained from other metabolites like Fas and tnf-α. Thus, this gene does not play a very critical role in carrying out an apoptosis. As expected, this gene was classified as non-essential by FBA conducted on our model. Thus, these details suggest that our model and hence our method is scientifically sensible. Our model is efficient in representing the biological condition more effectively. This method of mathematical modelling can be used to study complex biological pathways which are normally difficult to study in-vivo. In in-vitro studies, the concentrations of the various metabolites may be altered to variations in the environment. And some of the proteins may be stable only in a particular cellular environment. So these proteins cannot be studied in-vitro. Thus, mathematical modelling can be used as an alternative in these cases. Since it does not require any knowledge about the concentrations of the metabolites, there is no need for wet-lab work thereby making it cost and time efficient. With the availability of experimental biochemical data, like the concentration or the ratio of these caspases in a cell, the accuracy of this method can be improved. In our method, the entire model is considered to be present in a single compartment. Separate compartments for the inclusion of inhibitors and activators can be added to represent the pathway more effectively. A more complex model including the entire apoptotic pathway can be modelled to determine all possible ways of inhibiting apoptosis.

CONFLICT OF INTERESTS

Declare None

REFERENCES

1. Christophe HS, Stefan S, Bernhard OP, Reinhart H. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnol Prog 1999;15:296-303.
2. Eberhard OV, Jonas A. Decoupling dynamical systems for pathway identification from metabolic profiles. Bioinformatics 2003;20:1670–81.
3. Ivan VM. Systems Biology. In: Methods in Molecular Biology. 1st ed. New York: Humana Press; 2009.
4. Vanya IR, Pedro M. Cellular responses to endoplasmic reticulum stress and apoptosis. New York: Domingos Springer; 2009.
5. Potten CS, James WW. Apoptosis: the life and death of cells. New York: Cambridge University Press; 2004.
6. Markus WC, Christophe HS, Iman F, Jeremy SE, Igor IG, Evgeni S, *et al.* Metabolic modeling of microbial strains *in silico*. Trends Biochem Sci 2001;26:179-86.
7. Minoru K, Susumu G. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27-30.
8. Peter DK, Christos AO, Caroline MK, Leon G, Pallavi K, Dag A, *et al.* Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Res 2005;33:6083–9.
9. Karthik R, Preethi R, Nagasuma C. Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. PLoS Comput Biol 2005;5:349-58.
10. Dantzig GB, Orden A, Wolfe P. Generalized simplex method for minimizing a linear from under linear inequality constraints. Pacific J Mathematics 1955;5:183–95.