JOURNAL OF PURE AND APPLIED MICROBIOLOGY, August 2014. Vol. **8**(4), p. 3029-3036

The Metabolic Network of *Bacillus subtilis* by Constraints Based Reconstruction and Analysis (COBRA)

Bin Rui[1#], Gang Li[2#], CongChong Wei[1], Yadong Fan[1], Yongkang Wang[1], Wenfeng Wang[1], Weichen Xin[1], Yi Yang[1] and Han Wen[1]*

[1]School of Life Science, Anhui Agricultural University, 230026, Hefei, China.
[2]School of Life Science, Guizhou Normal University, 550001, Guiyang, China.

(Received: 18 May 2014; accepted: 26 June 2014)

With the development of network theory and the metabolic network research tools, metabolic network has been gradually used to guide biomass production. This study analyzed the topology and interaction of the metabolic network of *Bacillus subtilis* based on computer simulation environment through Constraints Based Reconstruction and Analysis(COBRA) toolbox. By removing several reactions part and analyzing the flux distribution in FBA and FVA, we illustrate the correlation between one single reaction and system. We further verify the feasibility and applicability of the network theory and model method between reactions, so as to understand the topological and functional characteristics of *Bacillus subtilis* metabolism network.

Key words: System biology; COBRA Toolbox; FBA; FVA; Metabolic network; *Bacillus subtilis.*

With the rapid progress of the post genomic era (genome sequencing) and high throughput (high-throughput) technology[1], large amounts of original data of genome, proteome, transcriptome brought by the traditional biology began to keep recording and interpreting in digital form. Biology research enters a new stage, and uses mathematical models, computer simulation, and other bioinformatics tools to analysis all biological aspects, from which system biology derived[2-6].

Systems biology underlines the view of system science during studying biological process[5] Due to its mainly using the network to express, the research of metabolic networks become a hot area. Network theory combines biology, physics, sociology, mathematics, artificial intelligence, and other areas of research results. Meanwhile this theory benefits each discipline as well. Initially, the network theory only involves one substrate and one product. Actually it does not tally with the actual network. Later researchers update algorithm, but it requires to define four types of metabolites in advance which significantly makes it inconvenient to generalize. Fell et al began to focus on linear algebra method to search for more convenient way to calculate the metabolic pathway. The main idea is that first assuming the linearly independent solutions of the stoichiometric coefficient matrix of metabolic network is one of the ways of its metabolic network. By the linear combination of this pathway to form possible but not necessarily metabolic network pathway. Even if its theory assumes an unique network depth, but the disadvantages is obvious. First of all, the corresponding pathway is not sole. One is due to the fact that linear independence solution of its

* To whom all correspondence should be addressed.
School of Life Science, Anhui Agricultural University, 230026, Hefei, China
E-mail address:1485507359@qq.com
#: co-first author

forming coefficient matrix is not the only solution. Secondly the linearly independent solution can be positive or negative. And some reaction is irreversible in the actual network. So parts of the linearly independent solutions can not be metabolic network pathways in fact. Metabolic network theory was further developed after the convex analysis was used in metabolic pathway. Gradually two kinds of models which were approved internationally, the elementary flux mode pathways (EFMs) and extreme pathways (EPs). With the development of network theory, the research method of metabolic network slowly evolved into static analysis and dynamic analysis. Static analysis mainly includes FBA flux balance analysis proposed by Schilling and other researcher, elementary flux mode analysis, extreme pathway analysis and graph theory proposed by jeong and other researchers. Dynamic analysis which mainly based on dynamic information includes cybernetic modeling and metabolic control analysis. However, numerous parameters of reaction kinetics and metabolites concentration are unclear in the current large-scale metabolic network. Adopting dynamic analysis method to research is not well rounded. So the static analysis research method is the focus of the current metabolic network research[6-10].

A common method to study biological metabolic network is to knock out one or more reactions of the metabolic network and then testing its expected flux distribution by biomass function. Thus it is concluded that whether the reaction is necessary (this flux distribution has a substantial reduction in the biomass equations) or unnecessary (this flux distribution has little change in the biomass equations).

This method involves the relationship between reactions, genes and enzymes, while a reaction's relation to the corresponding genes and enzymes is built by Boolean expression which links the gene and the reaction catalyzed by enzyme/ protein to be a well-known Gene-Protein-Reaction expression, also known as GPR and corresponded to each reaction. Classic GPR express a complete protein which encoded by genes A and gene B together in the form of genes A and gene B. So missing any of them will lead to a termination of reaction. On the other hand, if the GPR gene is expressed in the form of gene A or gene B , it means that gene A and gene B are isoenzymes. Therefore, the impact after one or more gene being knocked out can be evaluated by Boolean expressions. In other words, if GPR 's estimate is FALSE, this reaction can be limited to zero and can be terminated by FBA.

This paper is based on the reconstruction limited, FBA (flux balance analysis) of analysis toolkit (COBRA Toolbox) and FVA (flux variability analysis) method to knock out the key nodes reaction enzymes of metabolism network of *Bacillus subtilis* on a single variable control. Thus its flow distribution is analyzed. We could learn that the importance of this enzyme in the metabolic networks as well as the correlation between enzymes. Consequently, we can verify the feasibility and applicability of the network theory and model method, and further getting to know the topological and functional characteristics of metabolism network of *Bacillus subtilis*.

MATERIALS AND METHODS

Network theory and method

FBA

FBA is a mathematical method that simulates the reconstruction metabolic network on genome-scale, mainly by adding constraints on the behavior of the system such as: chemical metering data, thermodynamic data, metabolic constraints and so on to limit it closed within a certain solution space, and then use linear programming methods to get the optimal solution.

It is also applied to identification of drug targets for pathogen assumes, the reasonable design of culture medium, reaction between host - pathogen. FBA assumed that in biological evolution one organism adapt itself to the various environment and optimize. Then simulated the process of the network's self-adjusting using mathematical model. Specifically, FBA defined all the reactions and metabolites of the system first, classified them by internal, external and reversible. Also interpreted the balance of metabolic network system as a dot product between stoichiometric coefficient matrix (S) and a large number of metabolic flux (vector v). If the system is in steady state, v is zero. Linear programming is used to calculated a solution set of a metabolic flux corresponds to the steady-state. Usually, the number of reactions is the number of redundant

metabolites, the number of unknowns is greater than the number of equations, thus we can not get all the metabolic flux solution theoretically. So it is necessary to add constraints to limit results in a certain range, such as: chemical metering data, thermodynamic data, metabolic constraints and so on. Finally, optimize the target, using linear programming method to obtain the optimal metabolic flux distribution[7].

FVA [12]

FVA is used to investigate metabolic state of the system. This program is used to calculate the minimum and maximum metabolic flux of reaction and its metabolic flux range, rather than the theoretical maximum growth rate. These values are specific to get the internal flow distribution. FVA is often used to detect robustness of the metabolic model in a variety of simulation environment. But compared with other modeling method based on limit, its disadvantage is long calculation time.

Tools, software

Cytoscape

Cytoscape is a open source bioinformatics software platform of virtual molecular interactions between network and integration of gene information and the data expressed. It obtains new functionality by adding a large number of plugins. These plug-ins can be used in analysing network and molecular outline, adding file format support, connecting database, searching in a vast network . Cytoscape was initially produced in System Biology Institute in Seattle in 2002. Now has developed into an international association, maintained and developed by open source developers

COBRA Toolbox

Since now we don't have enough detailed parameter data to realize a precise model of organism on genome-scale in biophysical sense. Based on the reconstruction of the limit and analysis of COBRA (Constraints-based Reconstruction and Analysis), it mainly use the physical and chemical constraints for a given condition state of biological networks to define a group of feasible set (the feasible solution space). These limiting conditions include: compartmentalization, mass conservation, molecular crowding, and thermodynamic directionality. Transcriptome data recently has also been used to reduce the size of the set of feasible state. Although the COBRA method does not provide a solution method, but it provides a set of simplified solution, which can be used to guide the development of the biological hypothesis.

COBRA Toolbox is developed by Palsson and Herrgard at the university of California, San Diego in the United States. Its functions only realized in the environment of MATLAB which is a set of MATLAB script based on constraint modeling. Those scripts are dependent on external library to read-write the model of SBML format. In addition, using some other functions may need to purchase a MATLAB toolbox of MATHWORKS. Its main functions are: reading and writing metabolic mode of SBML format, changing the content and parameters of the mode, using a variety of method based on restrictions such as: FBA (flux balance analysis), DFBA (dynamic flux balance analysis), RA (robustness analysis), GDA (gene deletion analysis), FVA (flux variability analysis)[1].

***Bacillus subtilis* metabolism network model**

This paper focused on *Bacillus subtilis* metabolism network. IBsu1103 was used as the model in the research. It contains 3306 metabolites, 7217 responses[9-10]. The virtual image can be seen in Fig. 1

Experimental steps [11]

The operation of metabolic network

The read of metabolic model

Run in the Matlab command line:

$ model=readCbModel('filename')

If not add path to Matlab current path.

Model modified

$ model = changeRxnBounds(model, 'rxnNameList', 'value', 'boundType')

Where rx NameList is the cell matrix of corresponding reaction of ID of model, value is a float, bound type having 'l', 'u', 'b' means that the three were lower, upper and both respectively.

Deletion of reaction

$ [model] = removeRxns(model, 'rxnRemoveList')

Where RxnRemoveList is the corresponding element in the model ID of the cell matrix response.

Get FBA value

Firstly modify LP solverÿ

$ changeCobraSolver('gurobi5', 'LP')

Then run the order of FBAÿ

$ [solution]=optimizeCbModel(model, [osenseStr],

[minNorm], [allowLoops])

osenseStr is 'max' or 'min', it represents the target value. The default of minNorm is 0,Try to find the way to minimize the loop. The default of allowloops is 'true'. If set to 'false', will use the loop rule algorithm to remove the loop, it will get a little time. Finally Gets flux values for each reactionÿ

$ flux=geometricFBA(model, [varargin])

Get FVA value

Runÿ

$ [minFlux,maxFlux]=fluxVariability(model, 90)ÿThe '90 'represents a percentage, is the ratio of the model can achieve the objective function with respect to the optimal value.

Saving and exporting results

Runÿ

$ path_xls=sprintf('%spress_%d.xls', 'filenameyouwanasave', 'datayouwanasave')

$ xlswrite(path_xls, 'datayouwanasave')

This command will export data for the XLS format file

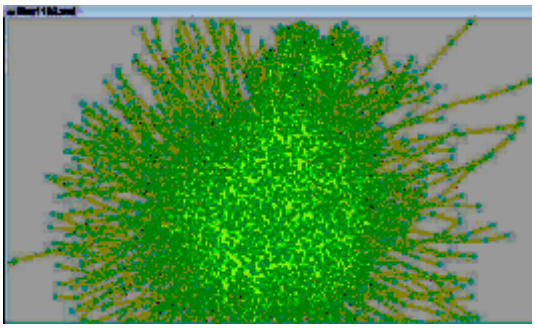

Fig. 1. The metabolic network of *Bacillus subtis*(part), including 3306 nodes and 7217 reactions

Data Processing

In the research, 21 reactions in total had been knocked out which corresponding with 21 maxflux value and 21 minflux and 21 flux values one by one. In addition, took the raw data which was recorded when no reaction was knocked out into consideration. A total of 223 groups data were sorted and summarized. We horizontally compared the flux of one particular reaction after deleting 21 different reactions and got different influences on this reaction. Then we horizontally compared several reactions selected from 7217 total reactions and longitudinally compared 21 deleting reactions with a total of 7217 reactions. And we could learn how a particular reaction affect system. Due to space limitations, only list several groups of data and charts.

RESULTS AND DISCUSSION

(1) Through the horizontal comparison, most reactions in metabolism system network of *Bacillus subtilis* have an obvious change response to the knock-out of 21 reactions. It is reflected in the metabolic flux distribution change. The knock-out of the 21 different reactions have different effects.

Fig.2 to Fig.7 show that no.0015 reaction almost has no change after 21 enzymes are knocked out respectively. Only one case has a significant inhibition. The no.0012 reaction not only can be inhibited but also can be activated after some enzymes being knocked out. And the direction of the no.0057 reaction of minfiux is also reversed after gene inactivation, resulting in a significant

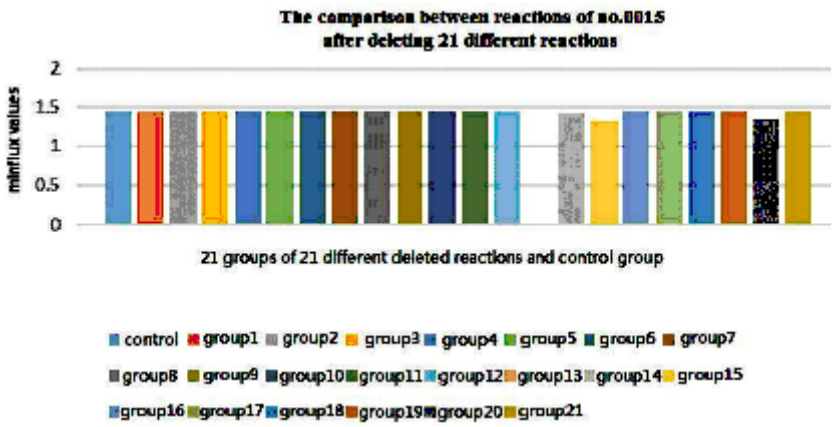

Fig. 2. The comparison between reactions of no.0015 after deleting 21 different reactions

change. Therefore different reaction made a different response of flux changes to different enzymes knock-out. This indicates that although metabolic network is an integral system, it has limitations on flux distribution. Particularly, there are interactions between fluxes and the correlation only showed in partial network.

(2) After being longitudinally compared, it is shown that each deleted reaction has an unique effect on the metabolic network of *Bacillus subtilis*. The deletion of udk reaction significantly inhibits on the whole reaction system. Most absolute flux values even reach zero. The flux value range of the whole network is greatly compressed almost zero (Fig.8). However, the deleting reaction of cmk have little impact on the reaction system. As shown above, the bar chart in blue is quite similar to the bar chart in red. Hardly we can distinguish the differences(Fig.9). However, the deleting reaction of pyrC facilitates the whole reaction system. The absolute value of the bar chart in red is greater than that in blue. This indicates an increase of flux range of the entire network and the carbon flux. Meanwhile several

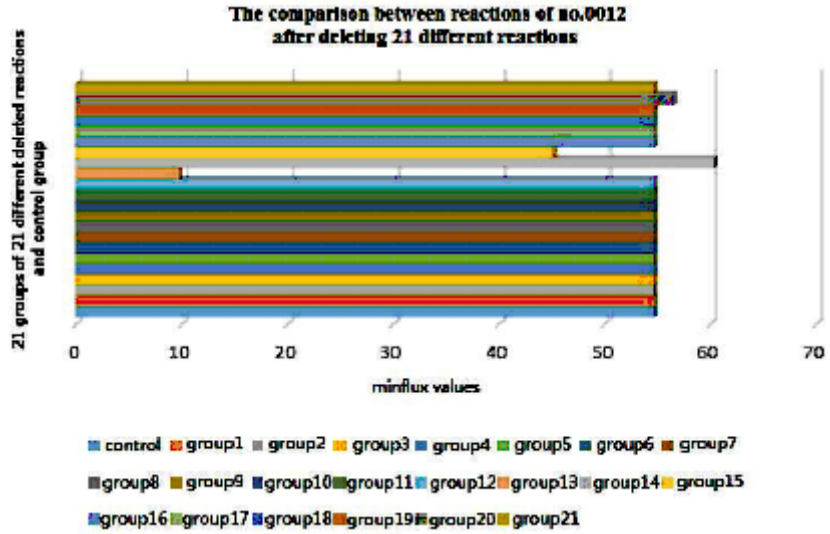

Fig.3. The comparison between reactions of no.0012 after deleting 21 different reactions

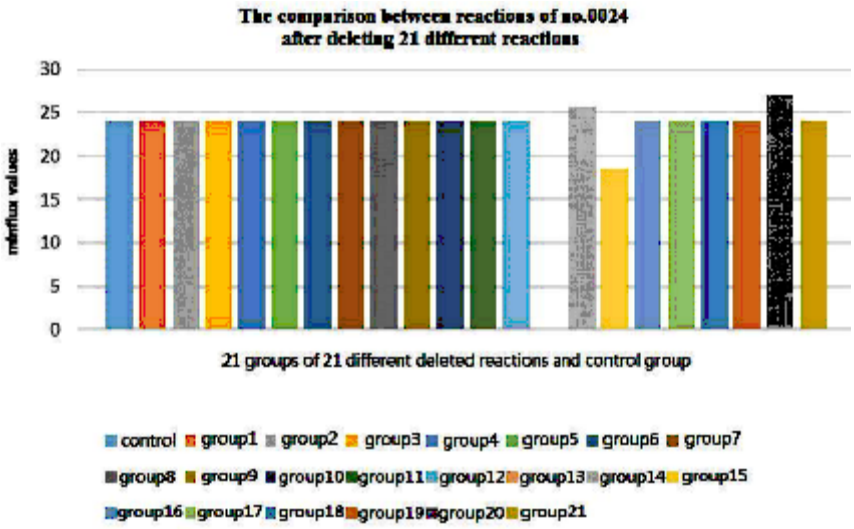

Fig.4. The comparison between reactions of no.0024 after deleting 21 different reactions

reaction fluxes listed in Fig.2 to Fig.7 have changed either positive or negative significantly after deleting enzyme 15 or enzyme 16. While the flux has no obvious change after deleting other enzymes.

Each reaction is not given the same weight in the formation of the network topology, although the metabolic network is composed of a number of metabolic reactions. Part of the metabolic reactions are the core modules of the network. For their metabolic modifications will result in a global change of metabolic network flux. While other more metabolic reactions as pathways increase the redundancy of function of metabolic network. It does not have obvious influence on most of other reactions due to the alternative pathways after being modified.

With the application of computer science and the support of other disciplines theory, the metabolic network analysis method develop very

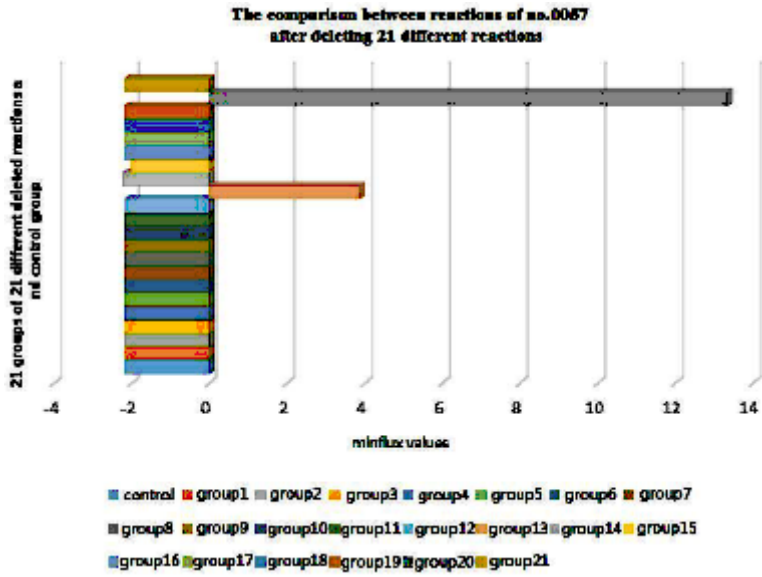

Fig.5. The comparison between reactions of no.0057 after deleting 21 different reactions

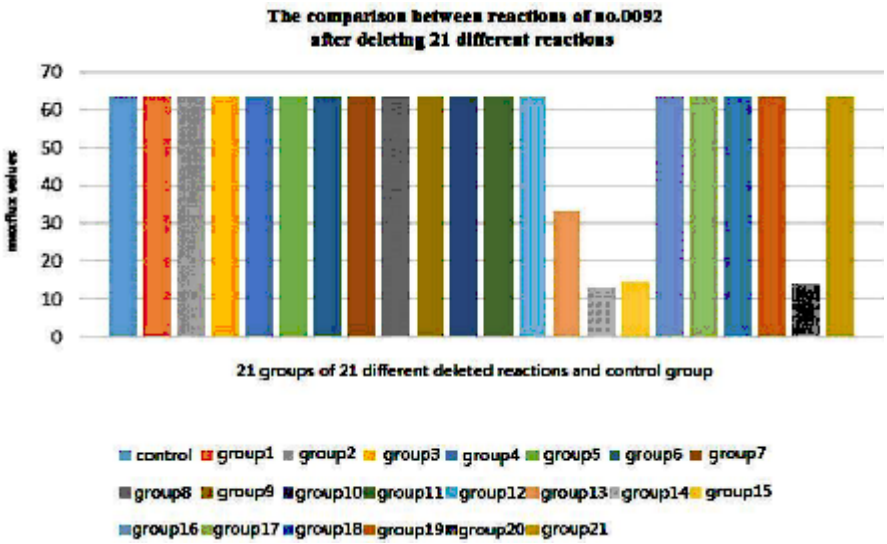

Fig.6. The comparison between reactions of no.0092 after deleting 21 different reactions

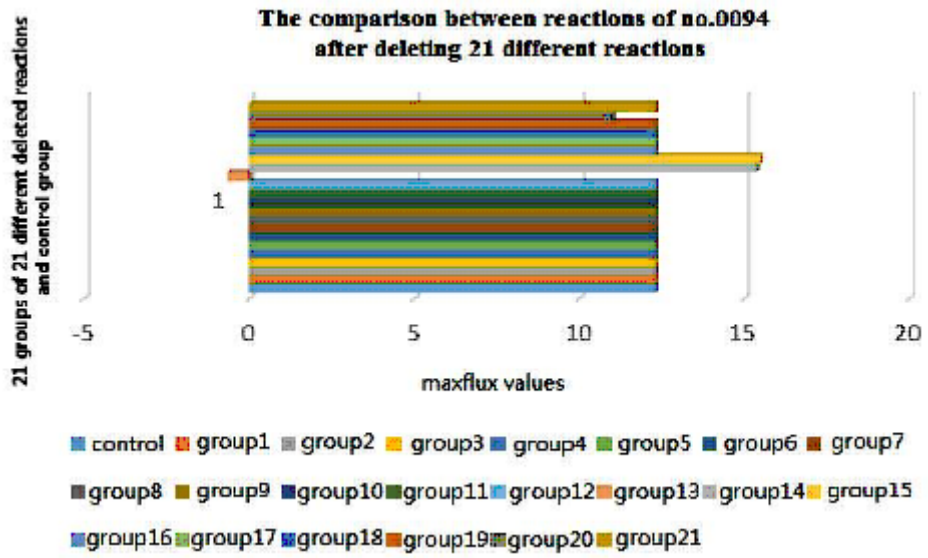

Fig.7. The comparison between reactions of no.0094 after deleting 21 different reactions

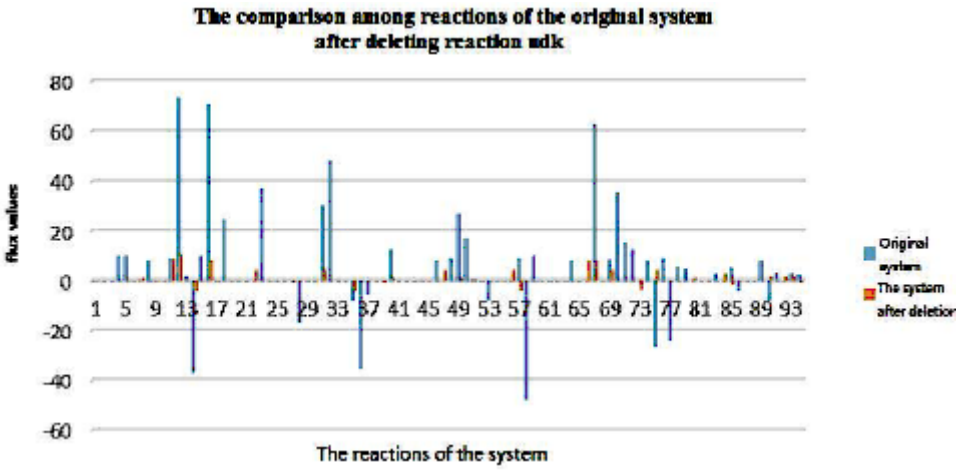

Fig.8. The comparison among reactions of the original system after deleting reaction udk

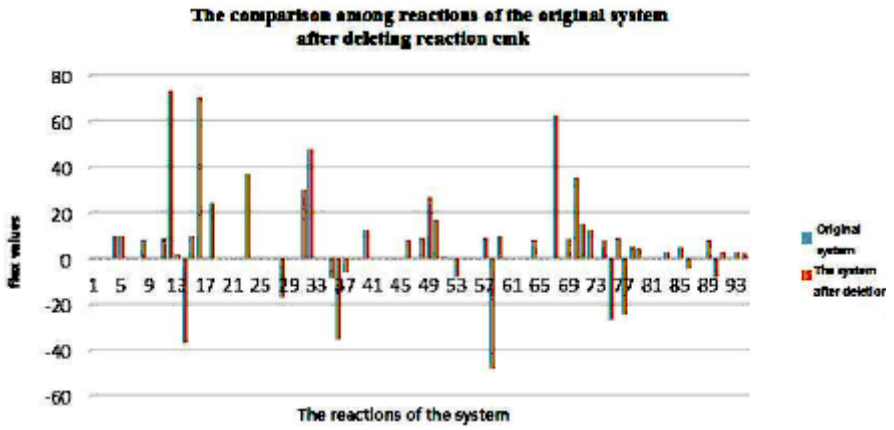

Fig. 9. The comparison among reactions of the original system after deleting reaction cmk

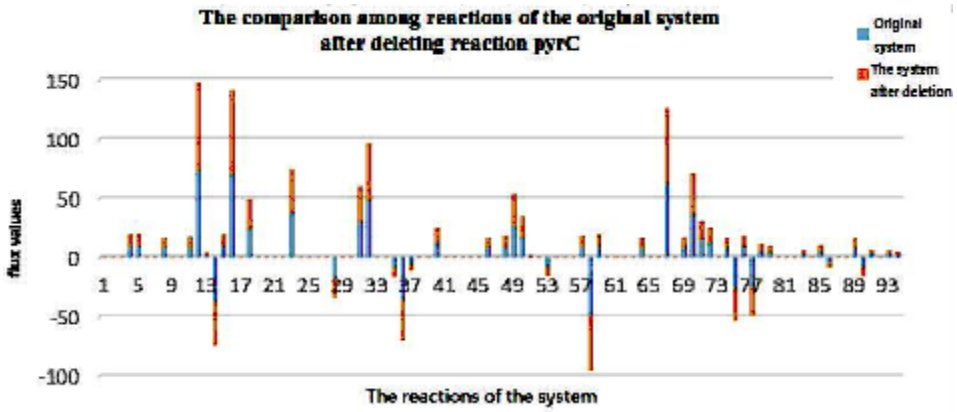

Fig. 10. The comparison among reactions of the original system after deleting reaction pyrC

fast. However, due to the complexity of the metabolic network, it still need support and validation of large number of traditional biological data to make metabolic network model and analysis method to be optimized.

Taken together, analysis of other enzymes and reactions in the metabolic system of *Bacillus subtilis* helps us to understand the metabolic system. The metabolic system model used here is complicated. This analysis program is quite suitable for other types of biological systems.

ACKNOWLEDGEMENTS

This work was supported by National Natural Science Foundation of China (31200626), Science and Technology Department of Guizhou Province Joint Fund (Guizhou Branch of J word LKS [2012] No. 22) and Dr. Guizhou Normal startup funds.

REFERENCES

1. Ding D.W., He X.Q., Lu K.Z., et al. Application of constrain-based method in *Escherichia coli* Metabolic network core model [J]. *Computers and Applied Chemistry*, 2010, **2**. DOI :10.3969/j.issn.1001-4160.2010.06.027
2. Jiang T.j., Xue Y.H., Xu T.,et al.System biology— The new field of life science *Prog. Biochem .Biophys*, 2004, **31**(11):957-964.
3. Zi Z.K., Sun Z.R. System biology:System-oriented biological research [J].*System Engineering—Theory&Practice*, 2005,**25**(2):47-55.
4. Jiang D. Network-based metabolic flux and structure analysis [D].*Dalian University of Technology*, 2006, DOI:10.7666/d.y865525.
5. Sun L.F., Jiang L. System biology:System science and biological system [J].*System Engineering—Theory &Practice*, 2005, **25**(10):67-72.
6. Ding D.W., Lu K.Z., Xu W.B.,ec al. Advance in metabolic pathway analysis and its application [J].*Computer engineering and application*, 2009,45(33):193-196.DOI:10.3778/j.issn. 1002-8331. 2009.**33**.062.
7. Chen q., Wang Z., Wei D.Q. Progress in the application of flux analysis of metabolic network [J].*Chinese Sci Bull*.2010.**55**(14):1302-1309.
8. Ding D.W. metabolic network reconstruction and its structural analysis of bacillus thuringiensis [D]. jiang nan university, 2008.DOI:10.7666/d.y1398263.
9. Wang B., Xiu Z.L., Tang H.W., ec al. Progress of metabolic network structure based on complex network theory [J]. China biotechnology, 2005, **25** (6) : 10-14. DOI: 10.3969/j.issn.1671-8135.2005.06.003.
10. Shen T., Rui B., Zhou H., ec al. Metabolic flux ratio analysis and multi-objective optimization revealed a globally conserved and coordinated metabolic response of *E. coli* to paraquat-induced oxidative stress [J].*Mol.BioSyst.*, 2013,**9**, 121-132. DOI: 10.1039/C2MB25285F.
11. Daniel H., Jan S.B., Richard Q.,ec al.Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *[J]. Nature protocols.* 2011 Sep; **6**(9):1290-307.
12. Steinn G.,Ines T. Computationally efficient flux variability analysis[J].*BMC Bioinformatics* 2010, **11**:489.doi:10.1186/1471-2105-11-489.