Available online at www.sciencedirect.com
ScienceDirect
Journal of Biotechnology 131 (2007) 159–167
Journal of BIOTECHNOLOGY
www.elsevier.com/locate/jbiotec

On the local optimal solutions of metabolic regulatory networks using information guided genetic algorithm approach and clustering analysis

Ying Zheng [1], Chen-Wei Yeh, Chi-Da Yang, Shi-Shang Jang *, I-Ming Chu

Chemical Engineering Department, National Tsing-Hua University, Hsinchu 300, Taiwan

Received 24 January 2007; received in revised form 30 May 2007; accepted 14 June 2007

Abstract

Biological information generated by high-throughput technology has made systems approach feasible for many biological problems. By this approach, optimization of metabolic pathway has been successfully applied in the amino acid production. However, in this technique, gene modifications of metabolic control architecture as well as enzyme expression levels are coupled and result in a mixed integer nonlinear programming problem. Furthermore, the stoichiometric complexity of metabolic pathway, along with strong nonlinear behaviour of the regulatory kinetic models, directs a highly rugged contour in the whole optimization problem. There may exist local optimal solutions wherein the same level of production through different flux distributions compared with global optimum. The purpose of this work is to develop a novel stochastic optimization approach—information guided genetic algorithm (IGA) to discover the local optima with different levels of modification of the regulatory loop and production rates. The novelties of this work include the information theory, local search, and clustering analysis to discover the local optima which have physical meaning among the qualified solutions.

Keywords: Metabolic regulatory network; MINLP; Information theory; Cluster analysis; Genetic algorithm

1. Introduction

Complex systems such as biological systems are basically complicated with many integer and real variables, numerous reactions, and material/energy balance equations, especially in metabolic network problems. Although many discussions focused on the optimal design of the metabolic networks, one of the basic problems—the physical meanings of the feasible local optima, has not yet been studied. The objective of this study is to develop a systematic approach that allocates the feasible local optima, and discuss the rationality and feasibility of the network structure.

Almost every metabolic reaction network is subjected to a regulatory architecture built around it. In many works, the nonlinear biological systems are transferred to linear systems by taking logarithmic transformation. The S-system, for example, has been widely implemented in the problem formulation of metabolic engineering. It is necessary to develop mathematical tools to analyze and optimize integrated biochemical systems. Some literatures, e.g., Voit (1992), used linear programming approach to solve a simplified metabolic network, the other, e.g., Hatzimanikatis et al. (1996), implemented MILP to solve an optimal solution from super-structure of a metabolic network. In the work by Hatzimanikatis et al. (1996), a very complete study on the synthesis of pathway of XMP and GMP was conducted. Their solution is also general to solve different types of MINLP. However, in case of mixed integer programming problems, it had been well-known that the solutions of such a problem may not be unique and subject to rugged contour (Grossmann and Sargent, 1979). In this work, we further study the example studied by Hatzimanikatis et al. (1996) by taking a different view of the original problem to find the other qualified local optimal

* Corresponding author at: Department of Chemical Engineering, 101 Section 2, Kuang Fu Road, Hsinchu 30013, Taiwan. Tel.: +886 3 571 3697; fax: +886 3 571 5408.
E-mail address: ssjang@che.nthu.edu.tw (S.-S. Jang).
[1] Address: Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, PR China.

0168-1656/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2007.06.019

Nomenclature

D	the sum of the distance of all individuals in each population
D_t	a constant set as a threshold
m	acceptable original solutions are selected to construct clusters
m_r	the number of solutions in cluster r
m_s	the number of solutions in cluster s
P_l	the amount of manipulated variable l ($l = 1, \ldots, 6$)
q_l	the logarithm of P_l ($l = 1, \ldots, 6$)
w_l	binary variables ($l = 1, \ldots, 6$) represent whether the concentration of enzyme is changed or not
x_{ij}	the jth variable of the ith individual of the population
X_j	the concentration of the metabolite j ($j = 1, \ldots, 4$)
$y_{r,i}$	the ith solution in cluster r
Y_j	the logarithm of X_j
z_{ij}	binary variables represent the modification of regulatory loop

solutions of the problem, and study the physical meaning of these solutions.

Many researches have discussed alternate solutions in the optimization of metabolic models. Lee et al. (2000) implemented a recursive MILP algorithm to find all alternate optima, and explored a metabolic engineering problem in bacteriology. Mahadevan and Schilling (2003) used linear programming (LP) and quadratic programming (QP) methods to analyze the effect of the alternate optimal solutions on the predictions of the mutant growth rates. In recent works, dynamic behaviour has been an issue for optimization of metabolic pathways. Additional constraints are added for stability, robustness and uncertainty consideration, e.g., Chang and Sahinidis (2005).

Solution of optimal metabolic regulatory network problems can be formulated into a mixed integer nonlinear programming problem (MINLP). As reported in literature (see, e.g., Adjiman et al., 1997), it is well known that many local optima exist in these MINLP problems; and global optimum is very difficult to discover. In fact, the discovery of all alternative regulatory structures would be helpful to understand and improve the metabolic productivity. Furthermore, since only limited observations on metabolic network were available, understanding alternative regulatory networks will suggest us explanations for the limited observations (see e.g., Lee et al., 2000). Finding alternative regulatory networks will reveal the various ways in which biochemistry and physiology can give us explanation for the limited observations (see e.g., Lee et al., 2000). Besides, the impact of the qualified local optimal solutions on the biological conclusions drawn from a simulated flux distribution could range from being negligible to being highly significant (see e.g., Mahadevan and Schilling, 2003).

In solving general MINLP, some stochastic approaches attracted much attention in this area, e.g., Cardoso et al. (1998). In their work, stochastic optimization solver—genetic algorithm (GA) is implemented to solve this MINLP problem. Evolutionary algorithms, such as GA, have been widely applied to many engineering problem. Compared to gradient optimization techniques, one of the benefits of evolutionary algorithms is that only the information regarding the objective function is required. Genetic algorithm was developed by Holland (1975). It offers robust procedures that can exploit massively parallel architectures and can be applied to classifier systems. Many further modifications to the original GA have been investigated by numerous researchers, for instance, multi-agent approaches (e.g., Tomassini et al., 2003, 2004), collocation design of initial points (see, e.g., Zhong et al., 2004). In their work, information theory is applied to modify the original GA such that the efficiency of the algorithm can be improved. On the other hand, the local optima together with possible global optimum of the biological system can be conveniently observed.

Information entropy was proposed by Shannon (1948) for handling signal community problem. From then on, information entropy has been widely applied in many areas. It can be used as a measurement of the diversity of data. By incorporating information theory to GA, the feasible solutions can be chosen. Different types of solutions, which are of bio-engineers' interest, are given after the cluster analysis. The results in this work show that different regulatory networks at the cluster centres can be found without further mathematic treatments of these complicated system models.

The rest parts of this work are organized as the following: In Section 2, description of the example studied in this work and transformation techniques for S-system is introduced. It should be noted that although S-system is adopted in this work, the whole approach derived by this work is not limited to S-

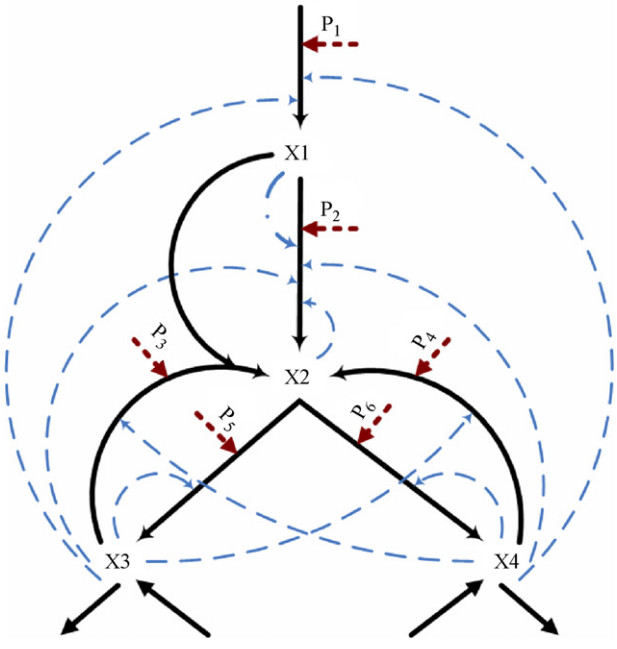

Fig. 1. Simplified XMP and GMP synthesis pathway.

system model. The Information-guided GA (IGA) approach can be conveniently implemented based on other types of models. In Section 3, the novel optimization approach—IGA approach and clustering analysis for solving qualified local optima of MINLP are derived. In Section 4, the solutions of the problem described in Section 2 are presented and discussed. Conclusive remarks are given in Section 6.

2. Background

2.1. Problem formulation

Savageau (1969) applied power-law approximation to evaluate biochemical processes, which combined enzyme kinetics and Bode analysis to linearize a nonlinear biochemical system. As shown in Fig. 1, Voit (1992) illustrated an example on yield optimization in xanthine monophosphate (XMP) and guanosine monophosphate (GMP) production.

There are two kinds of modifications involved in the above metabolic network: enzyme activity changes (P's) and regulation architectural changes (solid or dash lines). The former can be implemented through amplifying the expression level of the gene corresponding to the enzyme or through protein engineering that enhances the enzyme's activity. The regulatory architectural changes are much more involved in molecular biological manipulation. It may involve modification (or creation) of an entire regulatory pathway. Optimization of the possible metabolic network would suggest guidelines and ways to achieve the best results without too many repeating efforts in genetic engineering. Let X_j represent the concentration of the metabolite j ($j=1, \ldots, 4$); P_l represents the amount of manipulated variable l ($l=1, \ldots, 4$); and dashed lines denote inhibition, dashed-dotted lines activation.

Mass balance:

$$\begin{aligned}
\frac{dX_1}{dt} &= 900X_3^{-0.5}X_4^{-0.5}P_1 - 10X_1^{0.5}X_2^{-0.1}X_3^{-0.2}X_4^{-0.2}P_2^{0.6}P_3^{0.4},\\
\frac{dX_2}{dt} &= 7.34X_1^{0.308}X_2^{-0.062}X_3^{-0.162}X_4^{-0.1}P_2^{0.37}P_3^{0.245}P_4^{0.385}\\
&\quad - 43.8X_2^{0.42}X_3^{-0.339}X_4^{-0.5}P_5^{0.4}P_6^{0.6},\\
\frac{dX_3}{dt} &= 2.71X_2^{0.409}X_3^{-0.387}P_5^{0.455}\\
&\quad -0.036X_1^{0.041}X_3^{0.43}X_4^{-0.014}P_3^{0.28},\\
\frac{dX_4}{dt} &= 13.03X_2^{0.041}X_4^{-0.399}P_6^{0.405} - 0.143X_3^{-0.026}X_4^{0.40}P_4^{0.26}
\end{aligned} \tag{1}$$

The goal of optimization is to maximize the steady state concentration of X_4, while the constraints of X_1, X_2, and X_3 are restricted in 90%–110% of the steady state concentrations, and enzyme level changes are limited to the range within 20–500%. Variables z_{ij} are introduced to represent the modification of regulatory loop, and w_l are used to represent whether the concentration of enzyme is changed or not. The problem of consistency has to be considered when transforming the nonlinear model to MINLP, and some modifications was made by Hatzimanikatis et al. (1996).

Hatzimanikatis et al. (1996) discussed the modification in regulatory architectures for optimizing XMP and GMP yield. They transformed the nonlinear S-system into linear system and introduced binary variables as a set of key transformations to formulate the optimal manipulation of biochemical system to be a MILP problem. After logarithm transformation and modification of the steady state mass balance model as illustrated above, the optimization of maximizing X_4 in the metabolic network can be expressed as the following.

Let $y_i = \ln X_i$, the optimization problem becomes:

$$\begin{aligned}
&\text{Maximize } y_4\\
&\text{Subject to}\\
&\quad -0.5y_1 + 0.1y_2 - 0.3y_3 - 0.3y_4 - z_{13}\varepsilon_{13}y_3 - z_{14}\varepsilon_{14}y_4 + 0.6z_{21}\varepsilon_{21}y_1\\
&\quad + 0.6z_{22}\varepsilon_{22}y_2 + 0.6z_{23}\varepsilon_{23}y_3 + 0.6z_{24}\varepsilon_{24}y_4 + 0.4z_{34}\varepsilon_{34}y_4 + w_1q_1 - 0.6w_2q_2 - 0.4w_3q_3 = -4.4998\\
&\quad 0.308y_1 - 0.482y_2 + 0.177y_3 + 0.4y_4 - 0.37z_{21}\varepsilon_{21}y_1 - 0.37z_{22}\varepsilon_{22}y_2 - 0.37z_{23}\varepsilon_{23}y_3 - 0.37z_{24}\varepsilon_{24}y_4 - 0.245z_{34}\varepsilon_{34}y_4\\
&\quad - 0.385z_{43}\varepsilon_{43}y_3 + 0.4z_{53}\varepsilon_{53}y_3 + 0.6z_{64}\varepsilon_{64}y_4 + 0.37w_2q_2 + 0.245w_3q_3 + 0.385w_4q_4\\
&\quad - 0.4w_5q_5 - 0.6w_6q_6 = 1.7863\\
&\quad -0.14y_1 + 0.409y_2 - 0.817y_3 - 0.014y_4 - 0.455z_{53}\varepsilon_{53}y_3 + 0.287z_{34}\varepsilon_{34}y_4 - 0.28w_3q_3 + 0.455w_5q_5 = -4.3212\\
&\quad 0.041y_2 + 0.026y_3 - 0.799y_4 - 0.405z_{64}\varepsilon_{64}y_4 + 0.26z_{43}\varepsilon_{43}y_3 - 0.26w_4q_4 + 0.405w_6q_6 = -4.5122\\
&\quad \text{bounds on } y_j, \quad j = 1, 2, 3\\
&\quad \ln(4.9) \geq y_1 \geq \ln(6.0)\\
&\quad \ln(192) \geq y_2 \geq \ln(234)\\
&\quad \ln(2176) \geq y_3 \geq \ln(2660)\\
&\quad \text{bounds on } P_l, \quad l = 1, \ldots, 6\\
&\quad \ln(P_l^{\mathrm{L}}) \geq P_l \geq \ln(P_l^{\mathrm{U}})
\end{aligned} \tag{2}$$

where $w_l (l = 1, \ldots, 6)$ and $\{z_{13}, z_{14}, z_{21}, z_{22}, z_{23}, z_{24}, z_{34}, z_{34}, z_{43}, z_{53}, z_{64}\}$ are binary variables(1 or 0), $y_j (j = 1, \ldots, 4)$ and q_l $(l = 1, \ldots, 6)$ are continuous variables; L represents lower bound, U represents upper bound.

Note that Hatzimanikatis et al. (1996) further implemented linearization transformation techniques on the bilinear terms (z_{ij}, w_l, q_l and y_j) in Eq. (2) to solve this problem as a linear programming problem. After MINLP problem is transformed into an MILP problem, 16 variables and 64 constraints were added. In this work, genetic algorithm is used to solve this MINLP problem without any treatment of bilinear terms. Therefore, no additional variables and constraints are required, i.e., Eq. (2) is directly solved.

As indicated above, the objective of this work is to find all qualified local optima of Eq. (2) such that $y_4 \geq y_{\min}$. The method is introduced in the next section. It should be noted that the result derived here is not limited to S-system expression of metabolic pathways since the solution method is general for all MINLP's.

2.2. The basic genetic algorithm

In genetic algorithm theory (see, e.g., Gen and Chang, 2000; Vose, 1999), five basic operators should be considered in genetic algorithm: coding, fitness evaluation, selection, crossover, and mutation. Selection focusing on the best individuals will make genetic search go to narrow regions quickly, but the genetic search may be trapped on local optimal or to be terminated prematurely. Wild search is suggested at the beginning of a genetic search, and local search is suggested at the end of the genetic search. One of the linear crossover approaches—Intermediate recombination is adopted in this study. Mutation plays an important role in exploring solution space especially when genetic operation is trapped at local optima. A novel mutation approach is presented in the next section.

3. Information guided genetic algorithm

The basic idea of the proposed approach – information guided genetic algorithm (IGA) – is to detect the pre-maturity (i.e. the individuals of each generation is very close to each other) of the algorithm. Once the pre-maturity happens, an information guided mutation is performed. We also propose to implement a modified penalty function and local search for these particular problems as described below. Cluster analysis is proposed to discover the qualified local optima of the problem after implementation of IGA.

3.1. The pre-maturity detector

The detector of probable pre-maturity is introduced for detecting the situation of potential pre-maturity. In this work, we implement a premature detector D that is the sum of the distance of all individuals in each population. Suppose, $x_{i,j}$ is the jth variable of the ith individual of the population:

$$D = \sum_{j=1}^{N} \sum_{\substack{i=1 \\ i \neq d}}^{M} |x_{i,j} - x_{d,j}| \tag{3}$$

While $D \approx 0$ means probable premature situation occurs and all the individuals are the similar. In practice, it is necessary to set a threshold D_t such that $D > D_t$, otherwise information entropy guided mutation be will implemented to get ride of premature situation.

3.2. Information (IF) entropy mutation

Genetic operation may trap on local optima when probable pre-maturity happens. In order to get rid of probable premature, we introduce information entropy to refresh the probable premature population. According to Shannon's definition (1948), the information entropy of the set X is

$$E(X) = -\sum_{x \in X} p(x) \ln p(x) \tag{4}$$

where $p(x)$ is the probability of the event x occurring. Information entropy is a measure of how random a variable is distributed, and can be implemented to measure the diversity of the sampled data-set distributed in the solution space. The higher information entropy of a variable means the more diverse of the variable distributed in the solution space. The information guided mutation approach calculates the information entropy of each variable in MINLP. Variables have lowest information entropy are selected to perform information guided mutation to increase their information entropy. Yeh and Jang (2006) have discussed how to select the variables to be mutated. And the elements of new individuals will be generated by assigning random values in the range of the solution space of the selected variables.

3.3. Penalty function

Crossover and mutation may generate infeasible offspring during constrained optimization. Most of genetic algorithms apply penalty term to the original objective function, which makes infeasible solutions undesirable. The penalty function proposed by Barbosa and Lemonge (2004) is adopted and revised to handle infeasible solution.

3.4. Local search

Ombuki and Ventresca (2004) mentioned local search during mutation would accelerate convergent speed, but extend the computed time. A trade-off between convergent speed and computation time is necessary. In our study, local search is performed periodically after several generations, and only discrete variables are considered in local search to stress the influence in regulatory modification.

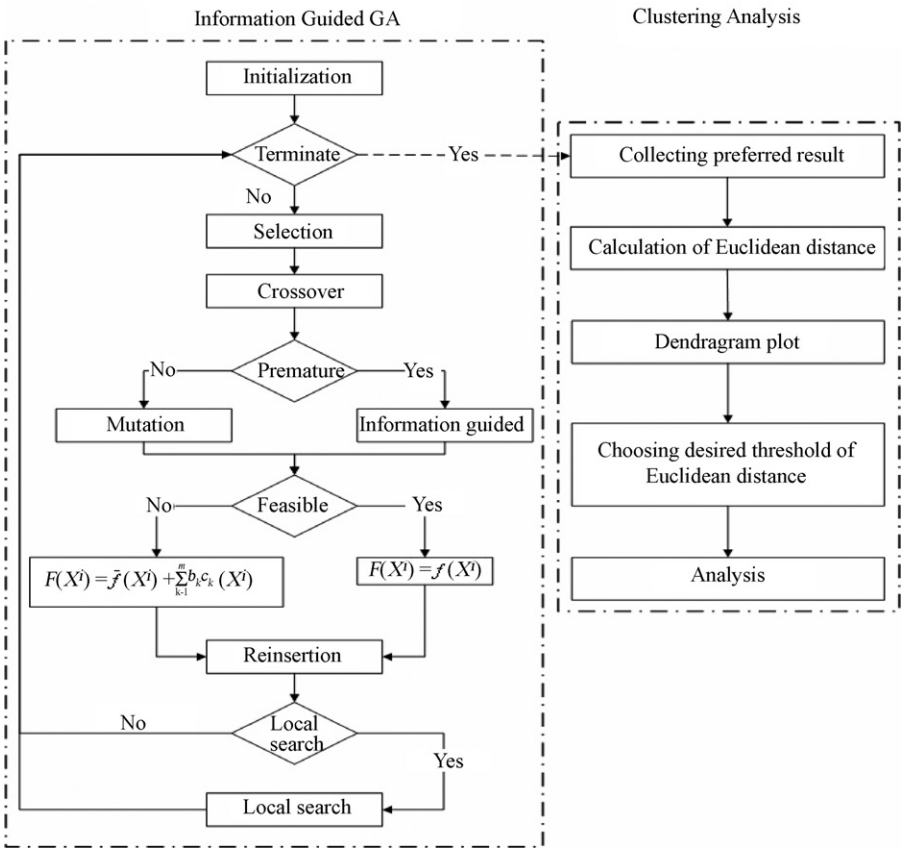

Fig. 2. The algorithm flowchart.

3.5. Clustering analysis for qualified local optima

Let m acceptable original solutions are selected to construct clusters. Euclidean distances between pairs of solutions are estimated for establishing a full-connected network, i.e., $m(m-1)/2$ linkages are generated in the network. The estimated distances are used to determine the proximity of solutions of each other.

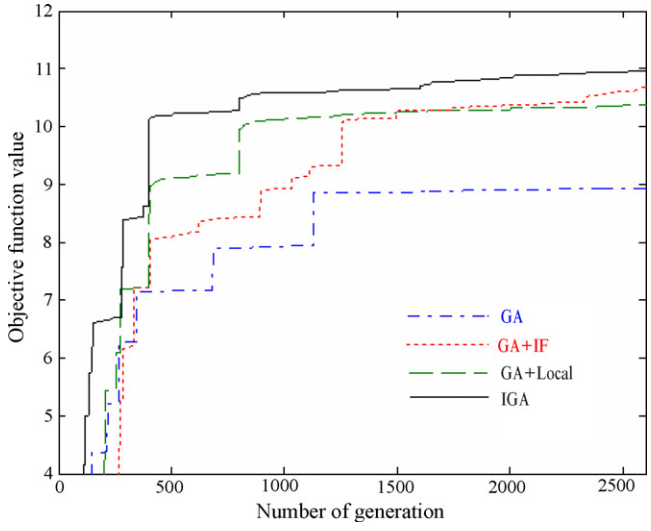

Fig. 3. Comparisons of the performance four different GA approaches.

The nearest neighbours of solutions are grouped into binary clusters, and the newly formed clusters are linked to each other or to other solutions to generate bigger clusters until all the solutions in the original solution set are linked together in a hierarchical tree. If m_r is the number of solutions in cluster r, m_s is the number of solutions in cluster s, and $x_{r,i}$ is the ith solution in cluster r, the definition of nearest neighbours is as follows:

$$d(r,s)=\min(d(x_{r,i}, x_{s,j})), \quad i\in(1,\ldots,m_r), \quad j\in(1,\ldots,m_s) \tag{5}$$

The $m-1$ nearest neighbours, with smallest distance between solutions in clusters, are selected to create a hierarchical clustering tree.

3.6. The algorithm

The overall flowchart is showed in Fig. 2. When prematurity happens, the information guided mutation scheme is performed. Local search will be performed under a proper probability after reinsertion. After the information guided genetic algorithm is performed, the qualified solutions are recorded. Clustering analysis is used on several batches of the records.

Table 1
Comparison of the statistics of four different GA's for the regulatory metabolic reaction

	GA	GA + IF	GA + local	IGA (GA + IF + local)
Average	8.9284	10.684	10.375	10.951
STD	3.4368	0.46561	1.3125	0.32403
Function call	67359	67358	79594	78670

4. Results

4.1. Local/global optima of the production of XMP and GMP

Fig. 3 gives the histories of the objective function value of Eq. (2) as a function of generation number based on the average of 50 simulation runs. As shown in Fig. 3, the performance of IGA is superior to other GA's. The comparison of the statistics of all four approaches is shown in Table 1. Compared with traditional GA, traditional GA + information guided mutation only and traditional GA + local search only, the proposed method (IGA) implement the slightly higher average number of function calls per run, the statistics of the objective function values are much superior to other GA's. In these four cases, maximum generation number is set to be 2600, local search is performed in every 400 generations, and the population size is set to be 14. The premature detector parameter D is set to 1. As shown in Fig. 3, it is obvious that by including information entropy and/or local search, the performance of GA can be drastically improved.

After 50 IGA batches, 463 feasible solutions are survived in the solution set. Among them we selected 84 high score solutions, which satisfy the following conditions: (1) $q_5 < 0$; (2) $q_6 < 0$; (3) $q_4 < 10.8$, i.e. $x_4 > 49021$. The integer variables $w_l (l = 1, \ldots, 6)$, $\{z_{13}, z_{14}, z_{21}, z_{22}, z_{23}, z_{24}, z_{34}, z_{43}, z_{53}, z_{64}\}$ and the continuous variables q_l ($l = 1, \ldots, 6$) were used for clustering analysis. The continuous variables were normalized and multiplied with w_l before clustering analysis. According to the 16 binary and the 6 continuous parameters, we computed Euclidean distance between these solutions and created the hierarchical clustering tree. The dendrogram graph is shown in Fig. 4.

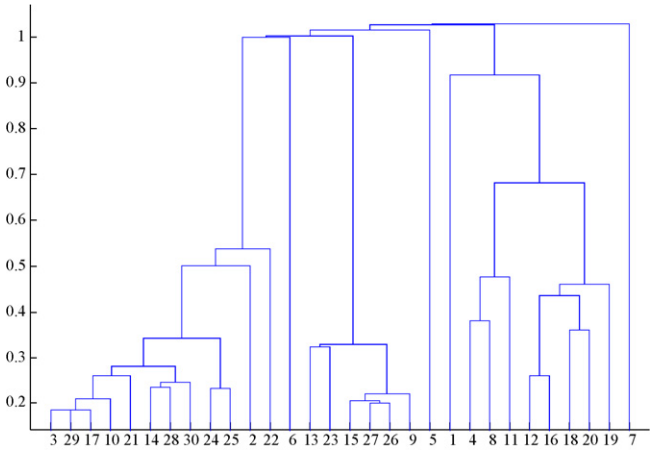

Fig. 4. The dendrogram graph.

As shown in Fig. 4, the level of Euclidean's distance is set equal to 1 for simplicity. In this case, only five clusters can be identified from the clustering tree. Solutions in the same cluster have the same binary variables. That means the solutions included in the same clusters are in the same regulatory architecture and same type of enzyme respectively. Of course, it is possible to analyze the solution set using lower level of Euclidean's distance as shown in Fig. 4, but it is not tested in this work.

5. Discussion

Table 2 compares the solutions of the above five cluster centres with the original solution obtained by Hatzimanikatis et al. (1996). The regulatory architectures of all cluster centres are demonstrated in Figs. 5–9 respectively. P_1+ and P_2+ are universally required for all the solutions. P_3- and P_5- are less than the reference enzyme level in original solution, solution #3 and solution #5. The two inhibitory loops (z_{13} and z_{53}) also appear in all solutions. Further comparison between the solutions obtained in this work with the original solution will help in elucidating the usefulness of this method. As shown in Fig. 5 and Table 2, solution #1 has the same regulatory architecture with the original one, but the enzyme activity changes of P_3, P_5 and P_6 are

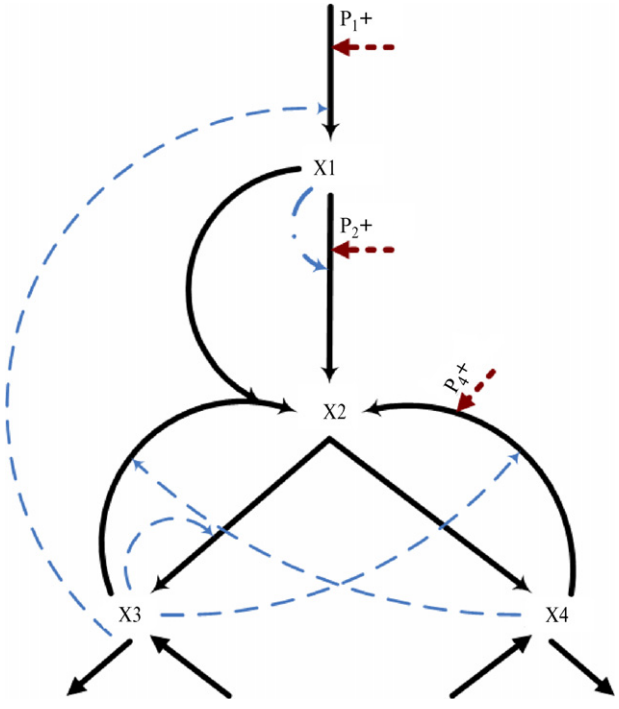

Fig. 5. The regulatory architecture of solution #1.

Table 2
The values of variables in original solution and IGA solutions #1–5

Node	w_l	q_l ($l = 1, \ldots, 6$), $q_r + w_l \times q_l = \ln(\mathrm{PI})$, $q_r = \ln(100) = 4.605$						$\{z_{13}, z_{14}, z_{21}, z_{22}, z_{23}, z_{24}, z_{34}, z_{43}, z_{53}, z_{64}\}$	y_4
Original solution (Hatzimanikatis et al., 1996)	111111	1.609	1.609	−0.363	1.609	−0.435	1.251	0101110001	55015.6
Solution #1 ($t = 21$)	110100	1.472	1.291	–	0.420	–	–	0101110001	52575
Solution #2 ($t = 7$)	110100	1.126	1.473	–	0.464	–	–	0100110001	49811
Solution #3 ($t = 13$)	111010	1.231	1.609	−1.609	–	−0.235	–	0101111001	65121
Solution #4 ($t = 5$)	110001	1.609	1.609	–	–	–	0.292	0111110101	56613
Solution #5 ($t = 12$)	111011	1.534	1.607	−0.455	–	−0.202	0.651	0101110101	69633

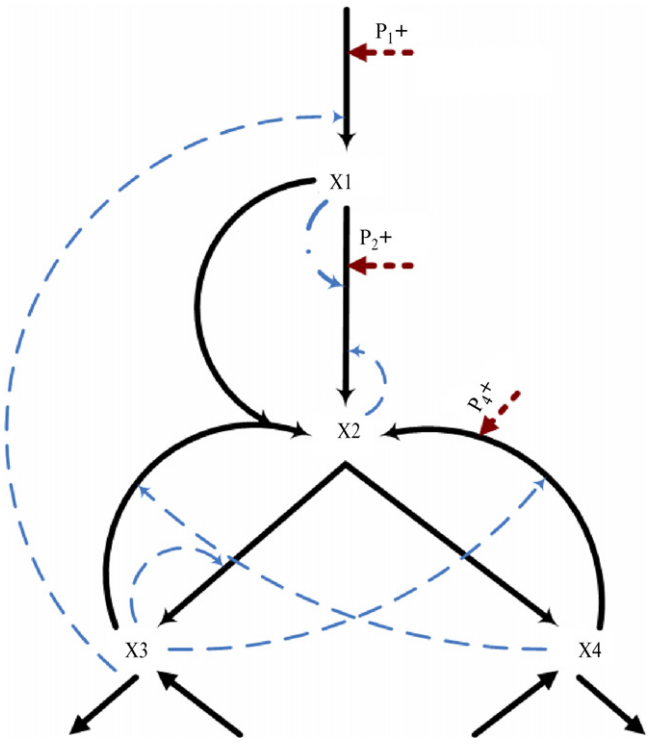

Fig. 6. The regulatory architecture of solution #2.

eliminated compared with the original architecture. Since the existence of change of P_6 will directly lead to a positive effect on X_4, it is not surprised that X_4 of solution #1 is lower than the original ones slightly.

Solution #2 (shown in Fig. 6) has a lower X_4 than the original solution too, but it needs less modification on the reg-

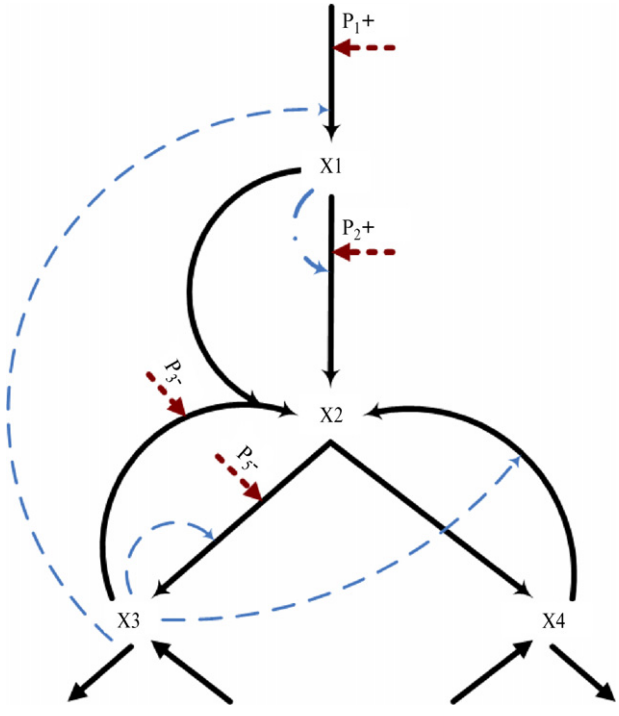

Fig. 7. The regulatory architecture of solution #3.

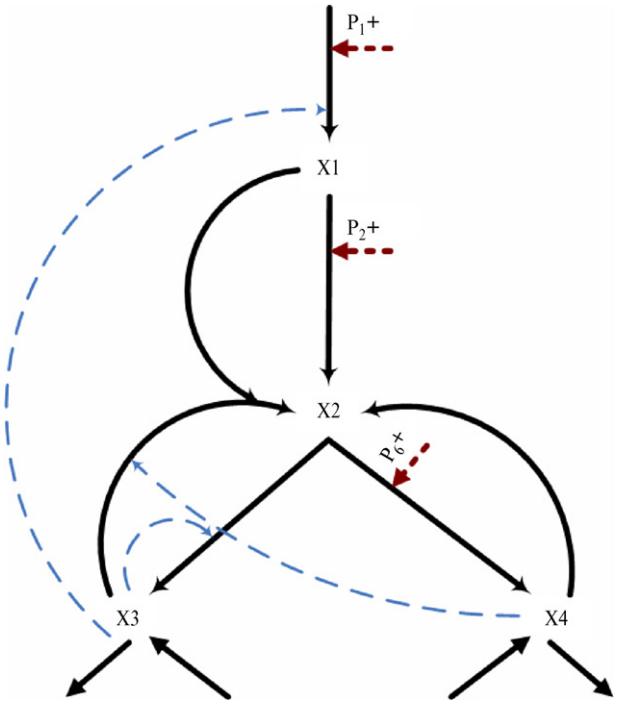

Fig. 8. The regulatory architecture of solution #4.

ulatory loops. According to the common sense of metabolic engineering, it is very important to achieve an improvement with less modification of regulatory loop. Figs. 7–9 demonstrate the regulatory structures of solution #3, #4 and #5. It should be noted that the solution #3, #4 and #5 result higher concentration of X_4 than original solution, but they need one more modification on the regulatory loops. Notably, all of the five solutions need less enzyme level changes than the original solution.

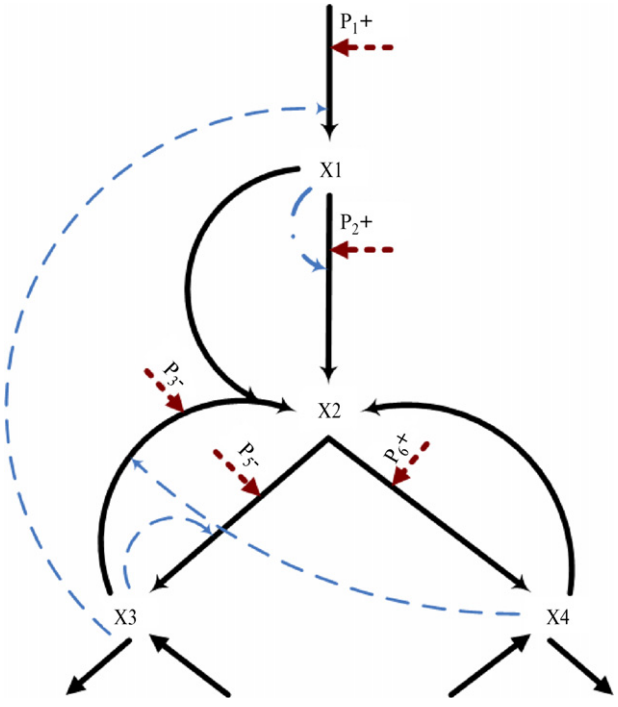

Fig. 9. The regulatory architecture of solution #5.

In solution #2, the inhibit loop (z_{22}) is kept. z_{22} will inhibit the metabolic path from X_1 to X_2. That will take a negative effect on increasing X_4. The differences between solution #2 and #4 are the existence of P_4, P_6, z_{21}, z_{22} and z_{43}. In solution #2, the positive effect P_6 for increasing X_4 is removed. So the objective function value in solution #2 is slightly lower than that in solution 4.

Comparing solution #4 with the original solution, the activated loop (z_{21}), the inhibited loop (z_{43}) and the three enzymes (P_3, P_4 and P_5) are removed. With this simpler structure, solution #4 still has superior result than the original one. Comparing solution #4 with solution #3 and #5, the activated loop (z_{21}) and enzymes (P_3- and P_5-) are removed in the former. The activated loop will help the metabolic path from X_1 to X_2. The metabolic path of X_2 to X_3 and X_2 to X_4 are competitive to each other. While the concentration of X_3 is increased, the concentration of X_4 will decrease. Decreasing the amount of P_5 may lead a decrease of metabolic flux from X_2 to X_3, therefore the concentration of X_4 may be increased. However, it is clear that the local optima at $t = 12$ (cluster center #5) and 13(cluster center #3) give the higher yields ($X_4 = 69633$ and $X_4 = 65121$ respectively) compared with other cases. Notably, the original yield by Hatzimanikatis et al. (1996) was $X_4 = 55015.6$.

6. Conclusion

Solution of optimal metabolic regulatory network problems can be formulated into a mixed integer nonlinear programming problem (MINLP). In this work, genetic algorithm is implemented to solve this MINLP problem. The information entropy and local search method are implemented to improve these solution approaches for MINLP problem. Unlike deterministic optimization, the novel approach takes advantage of stochastic optimization without further variable transformation. Furthermore, clustering analysis is implemented to allocate physically meaningful local optima. The example taken in this work is the Production of XMP and GMP. The solutions make sense by comparing to previous results. Useful local optima are discovered after clustering analysis. The results show that this approach is valid and efficient.

References

Adjiman, C.S., Androulakis, I.P., Floudas, C.A., 1997. Global optimization of MINLP problems in process synthesis and design. Comput. Chem. Eng. 21, S445–S450.

Barbosa, H.J.C., Lemonge, A.C.C., 2004. A new adaptive penalty scheme for genetic algorithms. Inform. Sci. 156 (3), 215.

Cardoso, M.F., Salcedo, R.L., de Azevedo, S.F., Barbosa, D., 1998. A simulated annealing approach to the solution of MINLP problems. Comput. Chem. Eng. 21 (12), 1349–1364.

Chang, Y.J., Sahinidis, N.V., 2005. Optimization of metabolic pathways under stability considerations. Comput. Chem. Eng. 29 (3), 467–479.

Gen, M., Cheng, R., 2000. Genetic Algorithms and Engineering Optimization. Wiley–Interscience, New York.

Grossmann, I.E., Sargent, W.H., 1979. Optimum design of multipurpose chemical plants. Ind. Eng. Chem. Process Des. Dev. 18 (2), 343–348.

Hatzimanikatis, V., Floudas, C.A., Bailey, J.E., 1996. Optimization of regulatory architectures in metabolic reaction networks. Biotechnol. Bioeng. 52 (4), 485–500.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. University of Michigan Press, Ann Arbor.

Lee, S., Phalakornkule, C., Domach, M.M., Grossmann, I.E., 2000. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. Comput. Chem. Eng. 24 (2), 711–716.

Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metabolic Eng. 5 (4), 264–276.

Ombuki, B.M., Ventresca, M., 2004. Local search genetic algorithms for the job shop scheduling problem. Appl. Intell. 21 (1), 99–109.

Savageau, M.A., 1969. Biochemical systems analysis.1. Some mathematical properties of rate law for component enzymatic reactions. J. Theor. Biol.: J. Theor. Biol. 25, 365–369.

Shannon, C.E., 1948. The mathematical theory of communication. MD Comput. 14 (4), 306–317 (Reprinted).

Tomassini, M., Vanneschi, L., Fernandez, F., Galeano, G., 2004. A study of diversity in multipopulation genetic programming. Artif. Evol. 2936, 243–255.

Voit, E.O., 1992. Optimization in integrated biochemical systems. Biotechnol. Bioeng. 40 (5), 572–582.

Vose, M.D., 1999. The Simple Genetic Algorithm: Foundations and Theory. MIT Press, Cambridge, MA.

Yeh, C.W., Jang, S.S., 2006. Chemical process model parameter estimation using an information guided genetic algorithm. J. Chem. Eng. Jpn. 39 (2), 197–210.

Zhong, W.C., Liu, J., Xue, M.Z., Jiao, L.C., 2004. A multiagent genetic algorithm for global numerical optimization. IEEE Trans. Syst. Man Cybern. Part B: Cybern. 34 (2), 1128–1141.