Analysis of regulatory architectures in BST*

BIAN Fuping[1,2**], LIANG Min[1,2] and ZHAO Xueming[3]

(1. Department of Mathematics, School of Science, Tianjin University; 2. Liu Hui Center for Applied Mathematics, Nankai University & Tianjin University; 3. School of Chemical Engineering, Tianjin University, Tianjin 300072, China)

Received April 11, 2003; revised June 12, 2003

Abstract Based on the data envelopment analysis (DEA) theory, the optimal pathway of metabolic reaction networks in biochemical systems is studied. After calculating the mixed-integer linear programming (MILP) model given by Bailey et al. twice, the decision making units (DMU) and the prediction model of DEA are constructed, where the inputs are levels of manipulated parameters (enzyme) and outputs are concentrations of metabolites. When the metabolic networks are reconstructed, the data are obtained by calculating MILP framework twice and the optimal levels of the manipulated parameter at different regular loops are predicted, thus simplifying the calculations of Bailey's.

Keywords: metabolic reaction networks; MILP; the prediction model of DEA; efficient DMU.

There are over one thousand kinds of enzymes in cells which catalyze various reactions and form a complex reaction network. Due to the development of biochemistry and cellular physiology, researchers get to know the decomposition and synthesis pathways of various components in cells and have a full understanding of controlling and regulation of these enzymes and their regulators. A large number of data of enzyme's dynamics have been accumulated through *in vitro* measurements. On this basis, through the quantity analysis of the metabolic networks and description of the flux distribution of various metabolic pathways at different statuses in cells, we can take some improved measures and regulate their distributions and get more interesting products. Commonly, the used methods include metabolic flux analysis (MFA), metabolic control analysis (MCA) and biochemical system theory (BST)[1].

BST is an analytical method for metabolic network developed in the 1970's. On the basis of the optimization theory, Voit et al.[2,3] found out one network architecture which optimizes the object function. First of all, the relation between the reaction rates and their parameters, e.g. concentrations of enzyme, substrate, and reagents, is set up. Then the variation range of the constraints is specified. The objective function is the maximization of rate of production. By resolving the optimization problem (S-system model) we can get the network architecture with its optimal object.

Introducing the constraints containing binary variables in S-system model we can get the mixed-integer linear programming (MILP) model[4]. These newly introduced constraints contain changes of a variety of enzyme regulatory architectures, reducing the amount of calculations. However, when reaction pathways become more complex, we need to solve the MILP model many times. And the model becomes very complicated as a result of introducing the new constraints. It is very difficult to solve the problem using linear programming method, and the optimal production rate cannot be predicted. In order to get the optimal network architecture, this article predicts the production rate under different enzyme regular structures through utilizing the data envelopment analysis (DEA) and objective programming. It is based on the result of resolving the MILP twice and the optimization of the metabolic network.

1 Mixed-integer linear model

We will consider that every reaction can be modulated by any of the two metabolites, X_1 and X_2, which will either inhibit or activate a reaction. This consideration results in the postulation of 12 regulatory loops (Fig. 1).

Four manipulated variables are considered: the amount of the enzymes, P_1, P_2, and P_3, that cat-

* Supported by the National Natural Science Foundation of China (Grant No. 20036010)

** To whom correspondence should be addressed. E-mail: fpbian@public.tpt.tj.cn

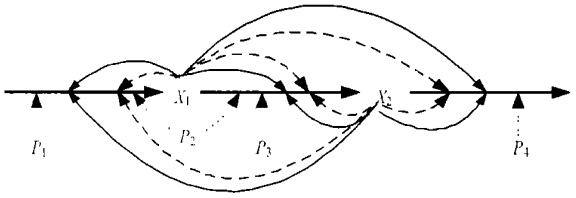

Fig. 1. A linear pathway with feedback inhibition. Bold solid lines denote reaction steps, dotted lines denote dependency on the corresponding parameters P_L, dashed lines denote inhibition, and thin solid lines denote activation.

alyze the three reactions, and the amount of the effector, P_4, that activates the first and the second reactions of the pathway. Moreover, for each loop we will consider $N_{reg}=6$ alternative levels of regulatory strength and type of regulation: $\{-0.5, 0.5, -0.1, 0.1, -0.01, 0.01\}$. We will allow only two regulatory loops to be active in the pathway.

Consider the manipulated parameter levels and the regulatory structures that should be changed to maximize the final product concentration X_2 when the following conditions are satisfied:

(i) The system is at a steady state;

(ii) $X_{1,ss}\leqslant 500$;

(iii) $V_{3,ss}\leqslant 10$; and

(iv) for the three enzymes, only overexpression is considered—that is, $P_L\geqslant 1$ ($L=1, 2, 3, 4$), and up to 10 times of their reference value.

We will introduce the binary variables z_{ijm} and the parameters ε_{ijm} with

$$m=1, \cdots, N_{reg},$$
$$i=1, \cdots, N_{rxn},$$
$$j=1, \cdots, N_{met},$$

where N_{reg} is the number of the alternative strength and types of regulation for each regulatory loop in the superstructure, and N_{rxn} and N_{met} are the numbers of the reactions and metabolites, respectively, in the metabolic network. In this example, we have $N_{reg}=6$, $N_{rxn}=3$ and $N_{met}=2$. And for the binary variables z_{ijm} and the parameters ε_{ijm} we let

(i) z_{ij1} and z_{ij2} be equal to 1, if reaction i is inhibited with strength $\varepsilon_{ij1}=-0.5$ or activated with strength $\varepsilon_{ij2}=0.5$, from metabolite j;

(ii) z_{ij3} and z_{ij4} be equal to 1, if reaction i is inhibited with strength $\varepsilon_{ij3}=-0.1$ or activated with strength $\varepsilon_{ij4}=0.1$, from metabolite j;

(iii) z_{ij5} and z_{ij6} be equal to 1, if reaction i is inhibited with strength $\varepsilon_{ij5}=-0.01$ or activated with strength $\varepsilon_{ij6}=0.01$, from metabolite j.

Then the S-system representation of the pathway is obtained[4].

We can introduce a set of variables:

$$q_L^r+q_L=\ln(P_L),\quad L=1, 2, 3, 4,$$

where q_L^r denotes the logarithm of the reference value of the parameter L, and q_L denotes the logarithm of the factor by which the reference value is multiplied to give the value p_L. In the example studied here

$$q_L^r=\ln 1=0$$

and

$$q_L=\ln 1=0.$$

Moreover, we introduce a set of binary variables, w_L, for which we will have

$$q_L^r+w_Lq_L=\ln(P_L),\quad L=1, 2, 3, 4$$

and z_{12}:

$$v_1^+=X_2^{z_{12}\cdot g_{12}}P_1P_4^2,$$

where v_1^+ is net rate laws describing the processes that increase the concentration of metabolite 1, and g_{12} is kinetic orders.

At last, we introduce variables y_j, t_L, s_{ijm}, for which we will have

$$y_j=\ln x_j,$$
$$t_L=w_Lq_L,$$
$$s_{ijm}=z_{ijm}\varepsilon_{ijm}v_j,$$

respectively. These variables will be used in the description of the steady-state equations after the logarithmic transformation. Thus we can write the MILP model[4]:

$$\max y_2$$

s.t.

$$\begin{cases}
-\sum_{m=1}^{6}\sum_{j=1}^{2}s_{1jm}+0.5y_1+\sum_{m=1}^{6}\sum_{j=1}^{2}s_{2jm}-q_1^r-t_1+q_2^r+t_2=\ln(1/0.02),\\
0.5y_1+\sum_{m=1}^{6}\sum_{j=1}^{2}s_{2jm}-y_2-\sum_{m=1}^{6}\sum_{j=1}^{2}s_{3jm}+q_2^r+t_2+2q_4^r+2t_4-q_3^r-t_3=\ln(2/0.02),\\
y_1\leqslant\ln(500),\\
-s_{12}+q_1^r+t_1+2q_4^r+2t_4\leqslant\ln(10),
\end{cases}$$

$$
\begin{cases}
P_1 \geqslant 1, \\
P_2 \geqslant 1, \\
P_3 \geqslant 1, \\
\left.\begin{aligned}
&\epsilon_{ijm} y_j - s_{ijm} + \min(y_j^{\mathrm{L}}\epsilon_{ijm}, y_j^{\mathrm{U}}\epsilon_{ijm}) z_{ijm} \\
&\quad \geqslant \min(y_j^{\mathrm{L}}\epsilon_{ijm}, y_j^{\mathrm{U}}\epsilon_{ijm}) \\
&\epsilon_{ijm} y_j - s_{ijm} + \max(y_j^{\mathrm{L}}\epsilon_{ijm}, y_j^{\mathrm{U}}\epsilon_{ijm}) z_{ijm} \\
&\quad \leqslant \max(y_j^{\mathrm{L}}\epsilon_{ijm}, y_j^{\mathrm{U}}\epsilon_{ijm}) \\
&z_{ijm} \min(y_i^{\mathrm{L}}\epsilon_{ijm}, y_j^{\mathrm{U}}\epsilon_{ijm}) - s_{ijm} \leqslant 0 \\
&s_{ijm} - z_{ijm} \max(y_j^{\mathrm{L}}\epsilon_{ijm}, y_j^{\mathrm{U}}\epsilon_{ijm}) \leqslant 0
\end{aligned}\right\} \begin{aligned} &i = 1, 2, 3; \\ &j = 1, 2; \\ &m = 1, \cdots, 6 \end{aligned} \\
\left.\begin{aligned}
&q_l - t_l + w_l q_l^{\mathrm{L}} \geqslant q_l^{\mathrm{L}} \\
&q_l - t_l + w_l q_l^{\mathrm{U}} \leqslant q_l^{\mathrm{U}} \\
&w_l q_l^{\mathrm{L}} - t_l \leqslant 0 \\
&t_l - w_l q_l^{\mathrm{U}} \leqslant 0
\end{aligned}\right\} l = 1, \cdots, 4 \\
w_1 + w_2 + w_3 + w_4 \leqslant 1, \\
\sum_{m=1}^{6} \sum_{j=1}^{2} \sum_{i=1}^{2} z_{ijm} = 2, \\
\sum_{m=1}^{6} z_{ijm} \leqslant 1 \ (i = 1, 2, 3; j = 1, 2).
\end{cases}
$$

The best solution is found for this problem:

$x_1 = 500$, $x_2 = 125$;
$P_L = 1 \ (L = 1, 3, 4)$, $P_2 = 2.236$;
$v_1 = 1$;
$z_{311} = z_{321} = 1$.

Including the additional constraint in the model and solving the problem again, we find the second best solution:

$x_1 = 500$, $x_2 = 111.8$;
$P_L = 1 \ (L = 1, 3, 4)$, $P_2 = 10$;
$v_1 = 10$;
$z_{311} = z_{212} = 1$.

As indicated by the results above, after the constraint containing binary variable is introduced, we include eight more linear constraints that will guarantee the consistency between $w_L q_L$ and t_L, $z_{ijm}\epsilon_{ijm} v_j$ and s_{ijm}, and thus the model becomes more complicated and the degree of difficulty in solving the model is very high. Therefore, after getting the two results above, we do not introduce the MILP model again, but analyze the new regulatory structure and predict the optimal concentration based on the given data for the purpose of reducing the steps of iteration and constraint conditions.

2 DEA model and its efficiency

2.1 Fundamental definition

Definition 1. The production possible set is $\{(x, y) \mid$ output vector y can be obtained from input $x\}$.

In this paper, when four enzyme expression levels are P_1, P_2, P_3, P_4, respectively, the obtainable intermediate product concentration is x_1 and the final product concentration is x_2.

Definition 2. If (x_j, y_j) is an observed activity, then the reference set is $T = \{(x_1, y_1), \cdots, (x_n, y_n)\}$.

Definition 3. When the relative increment percentage of input is more than that of the corresponding output, the corresponding DMU of (x, y) is decreasing returns-to-scale; when the relative increment percentage of input is less than that of corresponding output, the corresponding DMU of (x, y) is increasing returns-to-scale; when the relative increment percentage of input is equal to that of corresponding output, the corresponding DMU of (x, y) is constant returns-to-scale.

2.2 Fundamental DEA model

Consider n DMU_j $(1 \leqslant j \leqslant n)$. Their corresponding input vectors and output vectors are

$$\boldsymbol{x}_j = (x_{1j}, \cdots, x_{mj})^{\mathrm{T}} > 0, \quad j = 1, \cdots, n,$$

$$\boldsymbol{y}_j = (y_{1j}, \cdots, y_{sj})^{\mathrm{T}} > 0, \quad j = 1, \cdots, n$$

respectively. The intensity level of inputs and outputs are

$$v = (v_1, \cdots, v_m)^{\mathrm{T}}$$

and

$$\boldsymbol{u} = (u_1, \cdots, u_s)^{\mathrm{T}}$$

respectively. The calculating of the intensity level of inputs and outputs is based on certain laws.

Definition 4. $h_j = \dfrac{\boldsymbol{u}^{\mathrm{T}} \boldsymbol{y}_j}{v^{\mathrm{T}} \boldsymbol{x}_j} = \dfrac{\sum_{k=1}^{s} u_k y_{kj}}{\sum_{i=1}^{m} v_j x_{ij}}$, $j = 1, \cdots, n$, is called the evaluation factor of efficiency of the jth DMU.

The larger value of h_{j_0} indicates that more outputs are obtainable from fewer inputs.

Using Charnes-Cooper transformation, based on the principle of duality of linear programming, we can construct a dual model [5] that has the non-

Archimedes infinite small value ε:

$$\min[\theta - \varepsilon(e^{\mathrm{T}}s^{-} + e^{\mathrm{T}}s^{+})] = V_{D_\varepsilon},$$

$$(D_\varepsilon)\text{s.t.}\begin{cases}\sum_{j=1}^{n}\lambda_j x_j + s^{-} = \theta x_0,\\ \sum_{j=1}^{n}\lambda_j y_j - s^{+} = y_0,\\ \lambda_j \geqslant 0,\ j = 1, \cdots, n,\\ s^{-} \geqslant 0,\ s^{+} \geqslant 0.\end{cases}$$

The purpose of model (D_ε) is to get the maximum outputs with minimum inputs. When $\theta = 1$, $s^{-} = 0$, $s^{+} = 0$, (x_0, y_0) is efficient DMU; when $\theta < 1$, (x_0, y_0) is inefficient DMU of the DEA model. We can still produce the same outputs y_0 while consuming fewer inputs.

The limitation of this model is that it is not suitable for the production process that regulates the inputs and outputs proportionally. To solve this problem, Bian et al.[6] have constructed a non-radial DEA model whose inputs and outputs can be regulated proportionally.

2.3 Non-radial DEA model

Now we introduce the non-radial DEA model containing non-Archimedes infinite small value ε:

$$\min \frac{1}{m}\sum_{i=1}^{m}\theta_i - \frac{1}{p}\sum_{i=1}^{p}\beta_i - \varepsilon(e_{IP}^{\mathrm{T}}S_{IP} + e^{\mathrm{T}}S),$$

$$\text{s.t.}\begin{cases}\sum_{j=1}^{n}\lambda_j x_{ij} + S_{IP} = \theta_i x_{ij}, & i = 1, 2, \cdots, m,\\ \sum_{j=1}^{n}\lambda_j y_{lij} - S = \beta_l y_{lj}, & l = 1, 2, \cdots, s,\\ \sum_{j=1}^{n}\lambda_j = 1,\\ \lambda_j \geqslant 0, & j = 1, 2, \cdots, n,\\ S_{IP},\ S \geqslant 0,\\ 0 \leqslant \theta_i \leqslant 1, & i = 1, 2, \cdots, m,\\ \beta_l \geqslant 1, & l = 1, 2, \cdots, s,\\ e_{IP} = (1, \cdots, 1)^{\mathrm{T}} \in R^m,\\ e = (1, \cdots, 1)^{\mathrm{T}} \in R^s.\end{cases}$$

Theorem 1[1]. The optimal solutions of this programming problem are θ_i^{*}, $i=1, \cdots, m$, β_j^{*}, $j=1, 2, \cdots, s$, S_{IP}^{*}, S^{*}, and when $\theta_i^{*} = 1$, $i=1, \cdots, m$, $\beta_j^{*} = 1$, $j=1, \cdots, s$, $S_{IP}^{*} = S^{*} = 0$, DMU_j is correction efficient for DEA.

Proof. See the footnote.

Theorem 1 gives a method for determining whether DMU is effective when the inputs and outputs are regulated proportionally.

2.4 The prediction model of DEA

Consider how to predict the effective outputs of the new DMU when there are a group of inputs and outputs of n DMU and an input of a new DMU. Its algorithm is as follows:

Let $\boldsymbol{X}_{m\times n}$, $\boldsymbol{Y}_{s\times n}$ be the matrix composed of inputs and outputs, respectively, $\boldsymbol{X}_0$ be the input of the new DMU, $\boldsymbol{Y}_0$ be the unknown outputs to be predicted. First of all, we construct the following s programming problems:

$$\max y_{i0} \qquad i = 1, 2, \cdots, s$$

$$\text{s.t.}\begin{cases}\sum_{j=1}^{n}\lambda_j y_{ij} - y_{i0} = 0, & i = 1, 2, \cdots, s,\\ \sum_{j=1}^{n}\lambda_j x_{ij} \leqslant x_{i0}, & i = 1, 2, \cdots, m\\ \sum_{j=1}^{n}\lambda_j = 1,\\ \lambda_j \geqslant 0, & j = 1, 2, \cdots, n.\end{cases}$$

Resolving these problems respectively, we can get the ideal point of Y_0:

$$\boldsymbol{Y}_0^{*} = (y_{10}^{*}, y_{20}^{*}, \cdots, y_{s0}^{*}).$$

Secondly, we can set up a model containing weight:

$$\max \sum_{i=1}^{s}\frac{y_{i0}}{\overline{Y}_i} \qquad \text{here } \overline{Y}_i = \frac{1}{n}\sum_{k=1}^{n}y_{ik}$$

$$(D)\text{s.t.}\begin{cases}\sum_{j=1}^{n}\lambda_j y_{ij} - y_{i0} = 0, & i = 1, 2, \cdots, s,\\ \sum_{j=1}^{n}\lambda_j x_{ij} \leqslant x_{i0}, & i = 1, 2, \cdots, m,\\ \sum_{j=1}^{n}\lambda_j = 1,\\ \lambda_j \geqslant 0, & j = 1, 2, \cdots, n.\end{cases}$$

The optimal solutions of the linear programming model (D) are $y_{1A}, y_{2A}, \cdots, y_{sA}$. Combining these two steps we get $s+1$ groups of outputs, which is the output vector of the jth DMU.

1) Bian, F. P. et al. The prediction model of DEA with undesirable outputs. Systems engineering-theory applications.

3 The DEA analysis in metabolic reaction networks

For the metabolic reaction networks in Fig. 1, we analyzed the optimal network regulatory architectures by using the prediction model of DEA, considering four enzyme expression levels P_1, P_2, P_3, P_4 as inputs and intermediate product concentration x_1 and final product concentration x_2 as outputs, and constructed DMU by using prediction model of DEA. We determined if the reaction pathway in Fig. 1 is optimal according to the efficiency of DMU and then constructed 4×2 input matrix and 2×2 output matrix:

$$\boldsymbol{P}_{4\times 2}=\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \\ P_{31} & P_{32} \\ P_{41} & P_{42} \end{bmatrix}, \quad \boldsymbol{X}_{2\times 2}=\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}.$$

The elements of the ith column in $P_{4\times 2}$ are the level of enzyme of the ith reaction pathway, $i=1, 2$. The first row and the second row elements of the ith column in $X_{2\times 2}$ are the intermediate product concentration and final product concentration of the ith reaction pathway, respectively.

Substituting the data of P_1, P_2, P_3, P_4 and X_1, X_2 into the matrixes above after solving the MILP model twice, we have

$$\boldsymbol{P}_{4\times 2}=\begin{bmatrix} 1 & 10 \\ 2.236 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{X}_{2\times 2}=\begin{bmatrix} 500 & 500 \\ 125 & 111.8 \end{bmatrix}.$$

Now, we predict the optimization of the third reaction pathway by constructing DMU_0, letting the level of 4 kinds of enzymes $(P_1, P_2, P_3, P_4)=(1, 1, 1, 5\sqrt{2})$ be the inputs, and predicting its optimal outputs.

According to the steps of the predicted algorithm, the first step is setting up s programming problems as follows:

$$(D_1)\text{s.t.}\begin{cases} \max y_{10} \\ 500\lambda_1+500\lambda_2-y_{10}=0, \\ 125\lambda_1+111.8\lambda_2-y_{20}=0, \\ \lambda_1+10\lambda_2\leqslant 1, \\ 2.236\lambda_1+\lambda_2\leqslant 1, \\ \lambda_1+\lambda_2=1, \\ \lambda_1\geqslant 0,\ \lambda_2\geqslant 0, \end{cases}$$

$$(D_2)\text{s.t.}\begin{cases} \max y_{20} \\ 500\lambda_1+500\lambda_2-y_{10}=0, \\ 125\lambda_1+111.8\lambda_2-y_{20}=0, \\ \lambda_1+10\lambda_2\leqslant 1, \\ 2.236\lambda_1+\lambda_2\leqslant 1, \\ \lambda_1+\lambda_2=1, \\ \lambda_1\geqslant 0,\ \lambda_2\geqslant 0. \end{cases}$$

Solving the model (D_1) and model (D_2) (calculate them by using the linprog function in Matlab), we get the optimal point of $\boldsymbol{X}_0$: $\boldsymbol{X}_0^*=(500, 125)$.

The second step is constructing the model containing weight:

$$\max\left(\frac{y_{10}}{\frac{1}{2}(500+500)}+\frac{y_{20}}{\frac{1}{2}(125+111.8)}\right)$$

$$(D_3)\text{s.t.}\begin{cases} 500\lambda_1+500\lambda_2-y_{10}=0, \\ 125\lambda_1+111.8\lambda_2-y_{20}=0, \\ \lambda_1+10\lambda_2\leqslant 1, \\ 2.236\lambda_1+\lambda_2\leqslant 1, \\ \lambda_1+\lambda_2=1, \\ \lambda_1\geqslant 0,\ \lambda_2\geqslant 0. \end{cases}$$

The optimal solution of the model (D_3) is $(500, 125)$.

In summary, it is clear that the optimal production of the third reaction pathway is $(500, 125)$. This conclusion is the same as the optimal production with multiple construction MILP in Ref. [4]. The production calculated from S-system model is, however, $(100, 5)$. Obviously, this pathway is not optimal. The prediction DEA model calculation used in this paper is simpler than MILP. For the reaction in which more than three pathways exist, we can reset the enzyme expression levels and regulatory architectures, list the prediction DEA model several times, and predict the new reaction pathway, and iterate like this until we obtain an optimal reaction pathway.

4 Conclusion

This article constructs the efficient DMU by using given inputs (levels of enzyme) and predicted outputs (concentrations of products). It resolves the predicted optimal outputs and compares them with the actual value of the outputs. Through observing the optimization of the reaction, we confirmed whether or not to regulate manipulated parameters and reconstructed predicted DEA model. Due to the simplicity

of predicting DEA model and the possibility of calculating them using the given procedure, the amount of calculation is reduced dramatically, indicating that this method is feasible.

References

1 Zhao, X. M. et al. Metabolic engineering: its history, present and future. Chemical Engineering (in Chinese), 1997: 1319.

2 Torres N. V. et al. Optimization of nonlinear biotechnological processes with linear programming: application to citric acid production by Aspergillus niger. Biotechnol. Bioeng., 1996, 49(3): 247.

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

4 Vassily, H. et al. Optimization of regulatory architectures in metabolic reaction networks. Biotechnology and Bioengineering, 1996, 52: 485.

5 Sheng, Z. H. et al. Theory, Method and Applying of DEA (in Chinese). Beijing: Science Press, 1996: 30~36.

6 Bian, F. P. et al. The efficiency valuation of DMUs with undesirable outputs. Transaction of Tianjin University (in Chinese), 2002, 8(4): 295~298.