

文章编号 :1009 - 038X(2000)05 - 0458 - 03

MATLAB 在代谢网络优化中的应用

鲁 晋 , 马红武 , 赵学明
(天津大学化工学院 , 天津 300072)

摘 要 :生物代谢网络是一个具有高度非线性的复杂系统 , 如何对其进行适当的优化 , 一直是代谢工程领域的研究热点 . MATLAB 是一个应用广泛的工程计算软件 , 应用于代谢网络的优化 , 能方便、快捷得到系统的优化结果 . 作者以典型的代谢网络为模型 , 通过对目标代谢物浓度的优化来说明 MATLAB 在代谢网络优化中的应用 .

关键词 :代谢网络 ; 网络优化 ; MATLAB

中图分类号 : Q493

文献标识码 : A

The Application of MATLAB in the Metabolic Network Optimization

LU Jin , MA Hong-wu , ZHAO Xue-ming
(School of Chemical Engineering , Tianjian University , Tianjian 300072)

Abstract : Biological metabolic network is a system with high nonlinear characteristic . How to optimize it properly is a difficulty in the field of metabolic engineering . MATLAB is a software of technical computing , which has been used widely in many areas . Applying it to the optimization of metabolic network , the result was obtained easily and quickly .

Key words : metabolic network ; network optimization ; MATLAB

生物化学系统是一个非常复杂的网络系统 , 理解完整生化系统功能的中心问题是分子间相互作用的大量非线性特性 . 传统的直觉和试差法难以解决 , 必须要用系统的数学模型和分析方法 .

现在常用的网络分析方法有代谢控制分析 (MCA) 和生化系统理论 (BST)^[1] . 在 MCA 中定义了弹性系数、控制系数等重要的表征参数 . 由控制系数的大小比较即可确定代谢网络中的限制性步骤和关键酶 , 从而为代谢工程方法提供了目的位点 . 生化系统理论 (BST) 则是基于模型的方法 , 将代谢物浓度的变化用其生成速率和消耗速率两部分表示 , 并以幂律函数的形式对各速率与其影响因素

(包括中间代谢物和催化各步反应的酶) 进行线性化处理 , 通过在一定约束空间范围内搜索模型的最优解 , 可以找出系统的最佳影响因素浓度分布 . 以上这两种方法的缺点也是很明显的^[2] , 在 MCA 中 , 控制系数是系统在稳态附近的局部特征 ; BST 中幂律函数的动力学级次数是定义在原始稳态基础之上的 , 当系统偏离原始稳态点时 , 二者的准确性都会显著降低 . 完整生化系统的各种调控机理只有用非线性模型本身才能表达出来 , 因此需要用非线性模型方法对代谢网络进行优化 .

收稿日期 : 1999 - 12 - 23 ; 修订日期 : 2000 - 06 - 06 .

作者简介 : 鲁晋 (1975 -) , 男 , 天津人 , 硕士研究生 .

万方数据

1 MATLAB 的非线性优化功能

MATLAB 是适用于科学和工程计算的数学软件系统,20 世纪 80 年代中期由 Mathworks 公司推出.随后对其进行了不断的改进和升级,使得 MATLAB 成为一个多领域、多学科、多功能的优秀科技应用软件.作为数值软件的 MATLAB,对计算数学领域的特定问题类型(如数值线性代数,微分方程数值解,最优化方法等),用 MATLAB 系统求解,可以得到问题的离散解或近似的解析表达式.针对每一类问题的求解,MATLAB 都给出了该类问题的各种高效算法.因此,MATLAB 已经成为科学计算和教学实验的强大工具^[3].

选用 MATLAB 工具箱中自带的工程优化软件包,其中的函数命令“Constr”是专门为工程的非线性优化设计的.该函数命令应用变尺度 BFGS 法对目标函数进行优化,具有数值稳定性好,能找到全局最优等优点,特别适用于解决具有高度非线性的优化问题,是常用的非线性优化方法.

2 模型的建立与优化

首先,对代谢网络中的分支路径进行优化.之所以选择分支路径,是因为在代谢网络中,分支路径是最基本的非线性结构,同时也是构成整个复杂代谢网络的基础.此外,前人对分支路径的研究相对比较透彻,动力学方程比较完备.本研究所用到的模型如图 1 所示.

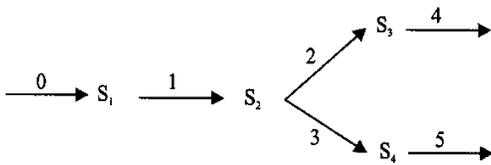


图 1 分支代谢路径

Fig.1 The branch metabolic pathway

$$r_0 = 3$$

$$r_1 = 10 \frac{s_1}{0.3333 + s_1}$$

$$r_2 = 4.1667 \frac{s_2}{0.6667(1 + \frac{s_3}{1 + 0.5s_1}) + s_2}$$

$$r_3 = 7.5 \frac{s_2}{0.6429(1 + \frac{s_4}{(1 + 0.6667s_3)})} + s_2(1 +$$

$$r_4 = 6.25 \frac{s_3}{0.1875 + s_3}$$

$$r_5 = 3.516 \frac{s_4}{0.3333 + s_4}$$

各代谢物浓度变化速率方程可写为:

$$\frac{ds_1}{dt} = r_0 - r_1$$

$$\frac{ds_2}{dt} = r_1 - r_2 - r_3$$

$$\frac{ds_3}{dt} = r_2 - r_4$$

$$\frac{ds_4}{dt} = r_3 - r_5$$

通过对 s 解非线性方程组 $\frac{ds}{dt} = 0$ 可以得到系统的定解:

$$0.1428 \quad 0.1781 \quad 0.2999 \quad 0.5187$$

假设对该分支路径进行优化的目的是使代谢中间物 S_3 得到积累,即优化目标函数是 $\max s_3 / (s_3 + s_4)$.在模型中引入催化各步反应的酶的浓度 $x(5) \sim x(9)$,假设酶浓度与速率方程成简单的正比关系,酶浓度的变化范围是 $(0.2 \sim 5.0)$,中间物浓度变化范围是 $(-10\% \sim +10\%)$ ^[5].这样,整个优化函数可写为:

$$\begin{aligned} &\max \quad \frac{s_3}{s_3 + s_4} \\ &s.t. \quad r_0 - r_1 * x(5) = 0 \\ &\quad r_1 * x(5) - r_2 * x(6) - r_3 * x(7) = 0 \\ &\quad r_2 * x(6) - r_4 * x(9) = 0 \\ &\quad r_3 * x(7) - r_5 * x(9) = 0 \\ &\quad 0.1428 * 0.9 \leq s_1 \leq 0.1428 * 1.1 \\ &\quad 0.1781 * 0.9 \leq s_2 \leq 0.1781 * 1.1 \\ &\quad 0.5187 * 0.9 \leq s_4 \leq 0.5187 * 1.1 \end{aligned}$$

$0.2 \leq x(5), x(6), x(7), x(8), x(9) \leq 5$

编写简单的 MATLAB 程序,得到运行结果为:
 中间物浓度: 0.1571 0.1959 4.1267
 0.4668
 酶浓度: 0.9366 5.0000 0.7369
 0.2000 0.8796

此时目标函数为: $\frac{s_3}{s_3 + s_4} = 0.8984$

从计算结果可以看出,模型的目标产物 s_3 浓度达到 4.1267,与初始态 0.0299 相比提高了近 138 倍.

直接调用 MATLAB 中用于计算方程组雅可比矩阵的函数命令“Jacobian”,计算优化态下目标函数的雅可比矩阵得到:

$$\begin{bmatrix} -4.31661150659885 & 0 \\ 3.93522595807563 & -15.60464002812114 \\ 0.38138554852323 & 6.39510967069901 \\ 0 & 9.20953035742244 \\ 0 & 0 \\ 0.22843661426318 & 0.05092227688155 \\ -0.24525769787475 & 0 \\ 0.00422511574526 & -1.66112612401033 \end{bmatrix}$$

计算矩阵的特征值有：

$$\begin{cases} -1.62998921018393 \\ -15.73229122628845 \\ -0.14874341353415 \\ -4.31661150659885 \end{cases}$$

由于特征值均为负值,说明该优化态是稳态,从理论上讲是可以达到的。

应用 MATLAB 还可以很容易地考察不同酶对系统优化结果的影响,以及可变酶个数的不同对优化结果的影响。

当只允许有一个酶的浓度发生变化时,优化结果见表 1。

表 1 中间代谢物浓度变化情况
Tab.1 The variation of intermediates

条件	S_1	S_2	S_3	S_4
只允许 $x(5)$ 发生变化	0.1429	0.1781	0.299	0.5187
只允许 $x(6)$ 发生变化	0.1428	0.1698	0.0336	0.4668
只允许 $x(7)$ 发生变化	0.1428	0.1959	0.0326	0.4802
只允许 $x(8)$ 发生变化	0.1428	0.1845	0.2147	0.5706
只允许 $x(9)$ 发生变化	0.1428	0.1774	0.0298	0.4668

可以看出,酶 $x(8)$ 对优化结果影响最大,其次为酶 $x(6)$ 和 $x(7)$ 。根据正交实验的原则,从对结果影响大的因素出发,分别考察允许两个、三个和四

个酶浓度发生变化时,优化结果见表 2。

表 2 中间代谢物浓度变化情况

Tab.2 The variation of intermediates

条件	S_1	S_2	S_3	S_4
只允许 $x(6), x(8)$ 发生变化	0.1428	0.1678	0.5902	0.4668
只允许 $x(7), x(8)$ 发生变化	0.1428	0.1959	0.2945	0.5706
允许 $x(5), x(6), x(8)$ 发生变化	0.1378	0.1678	0.5902	0.4668
允许 $x(6), x(7), x(8)$ 发生变化	0.1428	0.1959	0.5902	0.4668
只允许 $x(5), x(6), x(7), x(8)$ 发生变化	0.1534	0.1959	0.5902	0.4668

从以上的结果可以看出,同时改变 2 个、3 个或 4 个酶浓度时,目标产物的优化态浓度没有发生变化,始终保持在 0.5902 的水平,仅比初始态提高了 $0.5902/0.0299 = 19.74$ 倍,而同时改变 5 个酶浓度时,目标产物的浓度可以提高 138 倍。这种结果很好地反映了实际情况,即在用基因工程的方法改造菌种时,要想较大幅度地提高目标产物,仅仅改变一、两个基因序列是远远不够的,需要同时改变多个基因才有可能达到目的,而这也正是的困难所在。

3 结 论

以分支路径为模型, MATLAB 的优化工具箱为手段,对典型的代谢途径进行了较完整的优化研究。从优化过程和结果都可以看出,作为一个功能强大的数值计算软件, MATLAB 在对复杂的非线性生化系统进行优化处理方面具有广阔的应用前景。

参考文献

- [1] 赵学明, 马洪武. 代谢工程: 历史、现状和未来 [A]. 第八届全国化学工程论文报告会论文集 [C]. 天津: 1996, 1319 ~ 1326.
- [2] PISSARA N, NIELSEN J, BAZIN M J. Pathway Kinetics and Metabolic Control analysis of a High-Yielding Strain of penicillium chrysogenum During Fed Batch Cultivation [J]. **Biotechnology and Bioengineering**, 1996, 51: 168 ~ 176.
- [3] 杜藏, 骆源. 科学计算语言 MATLAB 简明教程 [M]. 天津: 南开大学出版社, 1998.
- [4] HATZIMANIKATIS V, BAILEY J E. Effect of Spatiotemporal Variation on Metabolic Control: Approximate Analysis Using (Log) Linear Kinetic Models [J]. **Biotechnology and Bioengineering**, 1996, 54: 91 ~ 104.
- [5] VOIT E O. Optimization in Integrated Biochemical Systems [J]. **Biotechnology and Bioengineering**, 1992, 40: 572 ~ 582.

(责任编辑 朱 明)