

热力学约束下代谢网络流量的蒙特卡洛采样方法

张夷, 谢璐, 袁子能

(中国科学技术大学生命科学学院, 安徽合肥 230027)

摘要: 发展了一种新的基于约束的代谢网络分析方法: 将代谢网络中可能的稳态类比成一个热力学系综进行蒙特卡洛采样, 定义以代谢流量为变量的势能函数, 既包括能使网络的生物量产出达到最优的目标量, 也包含网络需要满足的一些约束条件. 该方法可以对大肠杆菌稳态代谢流量空间在自动满足热力学约束和稳态约束的条件下进行采样, 比已有的方法更方便有效, 样本比完全随机采样具有更好的分布. 优化生物量产出得到的模拟结果与真实实验结果一致. 此外, 使用不同势能函数, 例如使乙醇产量最大化, 可获得不同目标下的流量分布, 对于代谢工程的遗传操作有指导作用.

关键词: 代谢网络; 大肠杆菌; 约束; 蒙特卡洛采样

中图分类号: Q811.4 **文献标识码:** A

Monte Carlo sampling of metabolic fluxes under thermodynamic constraints

ZHANG Yi, XIE Lu, YUAN Zi-neng

(School of Life Sciences, University of Science and Technology of China, Hefei 230027, China)

Abstract: A new constraint-based analysis method for metabolic networks has been developed. The possible steady-states in metabolic network were treated as a thermodynamic ensemble and a potential energy function enforcing additional constraints and virtual biomass was defined. The sampling in the steady-state flux space of the central metabolic network of *Escherichia coli* can avoid irrational fluxes violating thermodynamic constraints and mass balance and the results were consistent with the experimental data. The proposed method is more efficient than those reported, and the flux samples have better distribution than the random sampling method. Besides, other samples can be obtained, such as ethanol optimizing, via modifying the network and the potential function, which can be helpful to metabolic engineering.

Key words: metabolic network; *Escherichia coli*; constraint; Monte Carlo sampling

0 引言

近年来,随着分子间相互作用理论的完善以及分子水平检测技术的进步,人们构建了越来越多的全基因组水平上的代谢网络^[1~3]. 与此同时分析网络的方法的发展也相当迅速,对于代谢工程有很大的帮助和指导意义. 这些方法大体上分为三种: 基于

相互作用的, 基于约束的和基于机理的. 它们考虑的细节和准确度是递增的, 而所能研究网络的规模是递减的. 在这些方法中, 基于约束的方法由于其操作简便, 又可以研究中等规模的网络, 并且结果的信息含量也较高, 在三类方法中显得尤为重要^[4]. 基于约束的方法就是: 首先在代谢网络拓扑结构的基础上增加一些化学计量以及热力学约束, 即代谢网络中

反应物物质的量上的限制和网络状态的热力学可行性约束^[5,6],这样就可以避免对大量的反应参数的依赖;然后在生物体中各种代谢物质处于稳态状态的假设下,得到稳定状态下的各节点状态.例如,流量平衡分析(Flux balance analysis)^[7]就是基于质量平衡约束的.

基于约束的方法研究一个特定的代谢网络时,通常是先利用约束条件定义出流量组合的稳态空间,然后通过采样和样本分析得到具有生物学意义的结果.目前已经提出的一些采样方法,例如 hit and run sampling^[8], uniform random sampling^[9,10],都是完全随机的,因此在采样维度较高的情况下样本分布和采样效率可能会受到影响.我们发展了一种新的基于约束的分析方法并将之应用于大肠杆菌中心代谢网络的分析中:将代谢网络中可能的稳态类比成一个热力学系综进行蒙特卡洛采样,定义以代谢流量为变量的势能函数,既包括能使网络的生物量产出达到最优的目标量,也包含网络需要满足的约束条件,如质量平衡约束和热力学约束.之后的样本分析结果与真实实验相一致.并由此提出一个构想:通过优化大肠杆菌中某些特殊流量(例如乙醇排出)得到不同的流量分布和代谢模式,希望可以为代谢工程提供一些引导性的分析.

1 模型与方法

我们选择了大肠杆菌中心代谢网络为研究对象.首先利用流量平衡分析法^[7]构建出代谢流量稳态空间,然后加入两种热力学约束,最后使用 Metropolis 判据的蒙特卡洛方法^[11,12]实现稳态空间的采样.

1.1 研究体系和模型

我们的研究对象是 Stelling 等^[13]给出的大肠杆菌中心代谢网络.原网络中共计 90 种反应物,110 个反应:包括摄入反应,糖酵解,磷酸戊糖途径, TCA 循环及其旁路,副产品排泄(乙酸、甲酸、乳酸、乙醇),一些前体转化、大分子的合成以及一个虚拟的包含了大肠杆菌干重主要成分的生物量合成反应,如式(1)所示.



这里 M_i 代表对生物生长有贡献的代谢物, ω_i 是在大肠杆菌干重中测得的 M_i 的比例系数, M_{biomass} 代表虚拟的生物量这种代谢物, v_{growth} 是这个虚拟的

生物量合成反应的流量.

同时 Stelling 等也提出了最优生长假设^[13]:细胞倾向于选择优化的物质利用方式进行生长,即一定条件下,尽可能高效率地合成自身需要的生物量.这在后面采样的过程中将作为优化的目标量体现出来.

原网络中乙酸既可以作为碳源输入也可以作为发酵产物输出,是一个可逆反应.为了计算碳源的输入量和副产物输出量,必须区分输入和输出的乙酸,因此我们将乙酸的输入输出反应拆分成了两个不可逆反应.根据代谢网络结构可以构建出化学计量系数矩阵 S , 其行向量表示反应物,列向量表示反应,矩阵中的每个元素表示该行对应的反应物在该列对应的反应中的化学计量系数,负号表示消耗,正号表示生成.

1.2 流量平衡分析^[7]

选定体系之后,利用流量平衡分析法可以构建出代谢流量的原始稳态空间.流量平衡分析法基于质量平衡约束,并且假设系统是处于稳态的.当系统处于稳态时,各个代谢物的量和各个反应的流量都不变.

代谢物的变化量

$$\frac{dX}{dt} = S \cdot v \quad (2)$$

当系统处于稳态时,代谢物的量不会发生变化

$$\frac{dX}{dt} = 0 \Rightarrow S \cdot v = 0 \quad (3)$$

$$v = (v_1, \dots, v_n)' \quad (4)$$

$$X = (X_1, \dots, X_m)' \quad (5)$$

其中 X 代表代谢物的量, S 代表网络的化学计量系数矩阵($m \times n$), v 代表一组代谢流量.实际上,有很多 v 可以满足稳态约束,这些 v 构成了流量的稳态空间.为了更好的描述这个稳态空间,我们引入了 S 的零空间. S 的零空间是由式(6)定义的一组向量:

$$N(S) = \{x \mid S \cdot x = 0, x \in R^n\} \quad (6)$$

通过矩阵 S 的奇异值分解可以得到 $N(S)$ 的一组标准正交基,用矩阵 N 表示,在这里 $SN=0$ 并且 $N'N=I$.

N 中的每一个行向量都是稳态空间的一个基矢,可以通过对基矢的采样来替代直接对稳态空间的采样,这样可以简化采样过程.

1.3 热力学约束

在流量平衡分析的基础上,为了保证热力学可

行性,还需要引入热力学约束.这类约束主要体现在代谢网络中的反应可逆性约束和环流约束(Loop Law).可逆性约束就是网络中各个化学反应可以进行的方向约束.环流约束指的是代谢网络中封闭环的净流量应该等于零,因为没有化学势来驱动环流的产生^[14~16].环流约束包括以下两个方面:第一,沿着环,环中所有反应的化学势差应该等于零;第二,顺着反应方向,所有反应的化学势差为负值.

$$\sum_{i \in \text{loop}} \Delta\mu_i = 0 \quad (7)$$

$$v_i \cdot \Delta\mu_i \leq 0 \quad \forall i \quad (8)$$

这两个方程只有在生化反应环中的净流量为零时才能被满足.

我们提出了一种简单易行的方法来判断环流是否存在.首先敲除(Knock out)所有与外界有交换的输入输出反应,用来保证整个系统是封闭的.此类删除仅用于寻找环路,并不被记录到最终用于采样的网络中.然后使用线性规划的方法计算出每个反应流量可能的最大值和最小值^[17].

$$\max \& \min v_i; i = 1, 2, \dots, n$$

$$\text{st.} \quad \begin{cases} \mathbf{I} \cdot \mathbf{v} \geq \mathbf{b} & \text{下界约束} \\ \mathbf{S}_{kb} \cdot \mathbf{v} = 0 & \text{稳态约束} \\ v_i \leq 10 & \text{上界约束} \end{cases} \quad (9)$$

$$b_i = \begin{cases} -10; & \text{当第 } i \text{ 个反应可逆时} \\ 0; & \text{当第 } i \text{ 个反应不可逆时} \end{cases} \quad (10)$$

其中, \mathbf{I} 是一个 n 维单位矩阵; \mathbf{S}_{kb} 是敲除了输入输出反应的封闭系统计量系数矩阵; \mathbf{v} 代表一组代谢流量; b_i 对于可逆反应取 -10 , 不可逆反应取 0 .

为了闭合线性规划要求的凸空间,我们给每个反应定义了足够大的下界和上界:对于可逆反应取 $-10/10$, 不可逆反应取 $0/10$.实际上,在封闭网络中流量应该全部为零,上下界是无法达到的,但是如果网络存在闭合环路,则环路中的流量可以达到上界或下界.通过寻找非零流量就可以找出网络中的环路.结果显示仅有两个反应构成了环路: $\text{Fum} + \text{QuiH}_2 \rightarrow \text{Succ}$ 和 $\text{Succ} \rightarrow \text{Fum} + \text{QuiH}_2$. 将之合并为可逆反应 $\text{Fum} + \text{QuiH}_2 \leftrightarrow \text{Succ}$ 之后,这个反应网络就没有可能的环流存在.

1.4 代谢网络流量的蒙特卡洛采样

分析了最优生长假设与约束条件之后,我们使用统计物理中的模拟技术,将稳态空间类比成一个热力学系综,并定义了一个势能函数,如式(11)所示,用来引导采样在满足约束的同时更好地符合预

定的最优目标(最优生长假设).采样使用 Metropolis 判据的蒙特卡洛方法^[11,12],使各态分布符合玻耳兹曼分布,在这种分布下,能量越高的态出现的概率越小,可以更好地满足细菌的最优生长假设.这里借用了热力学系综的概念,但所采样本仅仅用来估计最优值和做流量间相关性分析,并非将稳态空间等价于一个热力学体系来拟合真实细菌内部流量的分布.

$$E = -v_{\text{output}} + \omega \cdot E_{\text{constraint}} \quad (11)$$

$$E_{\text{constraint}} = \sum_{\text{irreversible}}^i \left(\frac{|v_i| - v_i}{2} \right)^2 \quad (12)$$

这里 $v_{\text{output}} = v_{\text{growth}}$, 表示当前优化的流量为 v_{growth} , 当需要优化副产品产出时,可以改变势能函数中的 v_{output} 后进行采样,而这一项为零时,则不优化任何量,仅仅在满足约束的条件下在稳态空间中随机采样.式(11)的第二项为约束能量项, ω 为一个足够大的系数用来增强约束.式(12)表示如果某个流量 v_i 对应的反应是不可逆的并且值又为负时,会给你一个惩罚值加入到势能函数中去.

接下来使用 Metropolis 判据的蒙特卡洛方法实现对稳态空间的采样.

第一步:选取一组 0 到 1 的随机系数 \mathbf{A} , 得到一组初始流量 \mathbf{v} .

$$\mathbf{v} = \mathbf{A} \cdot \mathbf{N} \quad (13)$$

为了保证流量具有相对意义(每摄取 1 mmol 碳原子的相对值),每次产生一组流量都会做归一化处理.

$$\mathbf{v} = \frac{\mathbf{v}}{\sum_i n_{\text{subi}} \cdot v_{\text{subi}}} \quad (14)$$

这里 n_{subi} 表示第 i 种培养基物质的碳原子数, v_{subi} 表示第 i 种培养基物质摄取流量.

第二步:随机从 \mathbf{A} 中选取一个系数进行微扰,得到新的一组流量,并做归一化处理.

第三步:按照势能函数计算微扰前后的 ΔE , 如果 $\Delta E < 0$, 那么新的 \mathbf{A} 将被接受,其他情况按照

$$p = \exp(-\Delta E/T) \quad (15)$$

的概率接受新 \mathbf{A} . 由于 $\omega \gg 1$, 约束一定要被满足,否则将会引起能量陡增, \mathbf{A} 无法被接受.这里 T 是一个虚拟的温度参数,并不具备实际的物理意义,可以通过改变它的值来控制采样.

在模拟的初始阶段是不满足热力学约束的,能量很高,满足热力学约束后的开始阶段能量绝对值很小,由于我们关注的是最优解和次优解,因此舍去

前面的高能量点. 同时为了度量不同初始点下的样本空间的重叠程度, 即采样是否收敛, 可以计算不同采样空间各自流量相关系数矩阵特征向量之间的均方根内积(RMSIP)^[18].

$$\text{RMSIP} = \sqrt{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{C}_i^A \cdot \mathbf{C}_j^B)^2} \quad (16)$$

式中, \mathbf{C}_i^A 和 \mathbf{C}_j^B 分别代表样本系统 A 和样本系统 B 的各自的相关系数矩阵的第 i 个和第 j 个特征向量. 这里我们选取了不同样本空间各自最大的前 5 个特征值对应的特征向量进行计算. 计算表明, 在很短的采样时间内, 不同初始点下的样本的 RMSIP 就可以在 0.9 以上, 说明采样是很快收敛的.

在任意一初始点采样收敛后, 每隔 200 步记录一次系数和能量值, 共记录 10 000 次. 采样在 6 种环境下进行, 分别是不同的碳源、有无氧气环境, 其中第四组在不同温度下做了四次采样, 如表 1 所示.

表 1 样本详情
Tab. 1 The details of the samples

样本序号	碳源	氧气	温度
1	乙酸	有	0.000 05
2	甘油	有	0.000 05
3	琥珀酸	有	0.000 05
4	葡萄糖	有	0.001—0.000 05
5	葡萄糖	无	0.000 05
6	葡萄糖+乙酸+甘油+琥珀酸	有	0.000 05

2 结果与分析

2.1 碳源对模拟结果的影响

样本统计表明, 在不同的碳源条件下大肠杆菌的生长速度是不同的, 呈现出从甘油、葡萄糖到羧酸类递减的趋势, 如图 1 所示. 这和碳原子的氧化态有关: 甘油的氧化态最低, 葡萄糖和乙酸次之, 琥珀酸最高. 四种碳源的 O_2 摄取速率从高到低分别为: 乙酸、甘油、琥珀酸、葡萄糖, 如图 2 所示. 图 1 和图 2 的横轴为流量相对大小, 纵轴为样本计数.

在综合碳源的情况下大肠杆菌更倾向于吸收甘油和葡萄糖, 排出乙酸, 如图 3 所示(横轴为该碳源在总碳源摄取中所占比例, 纵轴为样本计数). 通过相关系数分析, 葡萄糖和甘油相关系数为 $-0.978 3$, 呈相当高的拮抗趋势, 如表 2 所示. 这是个很有趣的现象, 在代谢水平上甘油和葡萄糖摄取是互相抑制的, 而在基因调控的水平上编码甘油激

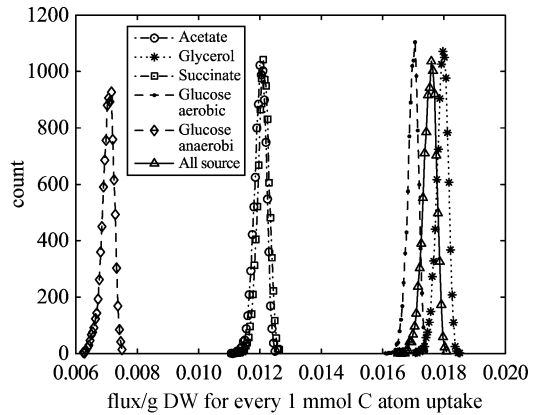


图 1 不同碳源下生物量合成流量分布图

Fig. 1 Biomass flux distributions under different carbon resources

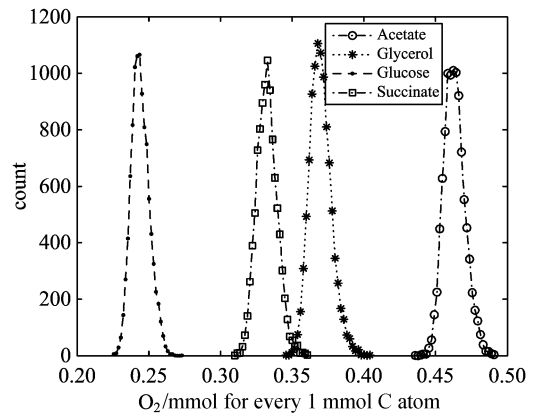


图 2 不同碳源下氧气的摄入分布图

Fig. 2 O_2 flux distributions under different carbon resources

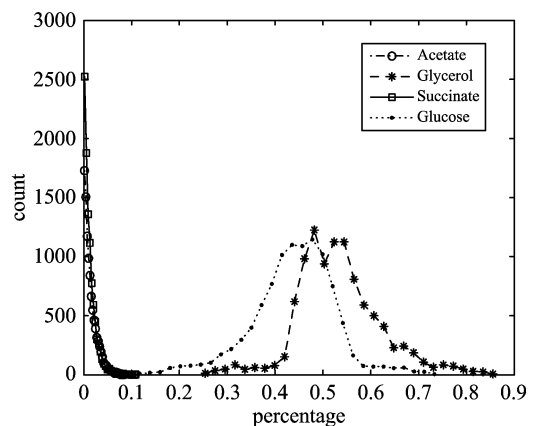


图 3 碳源摄取占总碳源百分比分布图

Fig. 3 Carbon resource uptake percentage distributions

酶的基因会被葡萄糖代谢中间物 1,6-二磷酸果糖抑制, 反过来葡萄糖的摄取则不受甘油的限制. 这样在实际的代谢过程中, 甘油的摄取就会受到强烈的抑制.

表 2 碳源间相关系数

Tab. 2 The correlation coefficients between different carbon resources

碳源间相关性	甘油	琥珀酸	葡萄糖
乙酸	-0.133 8	-0.028 1	-0.013 5
甘油	/	0.064 9	-0.978 3
琥珀酸	/	/	-0.078 1

2.2 氧气对模拟结果的影响

删除和加入氧气摄入反应可以用来控制有氧和无氧. 在有氧情况下每摄入 1 mmol 碳原子合成的生物量更多, 有机酸等有机副产品排出较少, 同时生成较多的二氧化碳; 在无氧的条件下正好相反. 同时有氧和无氧条件下葡萄糖吸收方式也有所改变: 由有氧状态下 ATP 介导的吸收方式(ATP 磷酸化葡萄糖)为主转变成以 PTS 介导的吸收方式(磷酸烯醇式丙酮酸磷酸化葡萄糖)为主, 如图 4,5 所示.

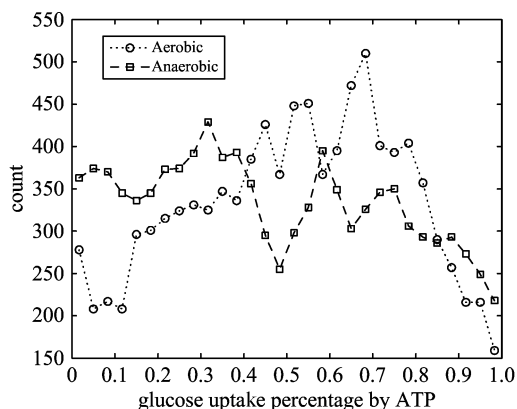


图 4 ATP 方式所占百分比分布

Fig. 4 ATP mode percentage distributions

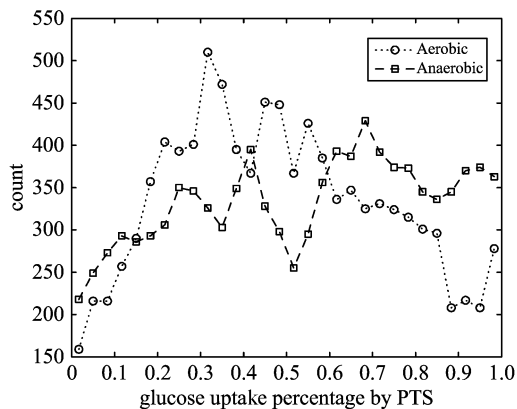


图 5 PTS 方式所占百分比分布

Fig. 5 PTS mode percentage distributions

流量均值的聚类分析可以找出有氧和无氧状态下不同的活跃反应, 如表 3 所示. 在有氧状态下,

ATP 泵, O₂ 输入和 CO₂ 输出及电子传递链比较活跃; 而在无氧状态下, 糖酵解和无氧代谢反应比较活跃.

表 3 有氧和无氧状态下的活跃反应

Tab. 3 Active reactions under aerobic and anaerobic conditions

有氧状态下的活跃反应	无氧状态下的活跃反应
3H \leftrightarrow ATP	DHAP \leftrightarrow G3P
CO ₂ \rightarrow 0	F16P \leftrightarrow DHAP+G3P
0 \rightarrow O ₂	G6P \leftrightarrow F6P
NADH \leftrightarrow 2H+QuiH ₂	PEP \rightarrow G6P
QuiH ₂ +1/2O ₂ \rightarrow 2H	PEP \rightarrow Pyr
	Pyr \rightarrow AcCoA+Form
	AcCoA+NADH \leftrightarrow Adh
	Adh+NADH \leftrightarrow Eth
	Eth \rightarrow 0
	Form \rightarrow 0

2.3 不同温度对模拟结果的影响

这里主要研究了葡萄糖为碳源的有氧代谢. 将温度这个虚拟参数从 0.001 调至 0.000 05, 温度越低, 采样越靠近最优解, 反之, 采样则比较分散. 在同一温度下, 有些流量的相对涨落很大, 远大于生物量的相对涨落. 这些流量对生物量贡献较少, 因此受生物量的约束较为宽松. 对比不同温度下的结果, 这些相对涨落较大的流量多数不受温度影响, 不随生物量的涨落变化而变化. 可以用聚类的方法将这些反应筛选出来: 无论温度高低涨落都很大的共有 18 个, 其中 12 个与无氧代谢有关, 2 个是碳源流入(乙酸和琥珀酸), 1 个是 ATP 流出, 以及其他 3 个反应; 标示这些反应与目标函数的相关性不强, 可以自由变化. 与此相反, 另外一些反应流量的相关性随温度变化非常明显. 如表 4 所示, 在越靠近最优生长的情况下, 糖酵解与三羧酸循环相关性越高, 三羧酸循环和 NADPH 的合成相关性越高. 这种反应流量刚性的不同标示着这些反应与目标量关系的密切程度以及反应与反应之间关系的密切程度, 可以为代谢工程调控提供非常有用的信息.

2.4 主成分分析

利用样本的相关系数做主成分的计算反映出两个主要成分. 第一主成分包含了蛋白质合成反应 23 个, 糖类脂类等合成反应 9 个, DNA 和 RNA 合成反应 9 个, 还有其他 4 个零散的合成反应. 第二主成分主要包含 5 个糖酵解反应, 9 个磷酸戊糖途径, 6 个 TCA 循环, 8 个包含电子传递链和其他一些零

表 4 不同温度下的相关性变化

Tab. 4 The variations of correlation coefficients under different temperatures

反应 1	反应 2	相关系数($T=0.000\ 05$)	相关系数($T=0.001$)
2PG \leftrightarrow PEP	Cit \leftrightarrow ICit	0.990 0	-0.088 4
2PG \leftrightarrow PEP	AcCoA+OxA \rightarrow Cit	0.990 0	-0.088 4
3PG \leftrightarrow 2PG	Cit \leftrightarrow ICit	0.990 0	-0.088 4
3PG \leftrightarrow 2PG	AcCoA+OxA \rightarrow Cit	0.990 0	-0.088 4
3H _{ex} \leftrightarrow ATP	Cit \leftrightarrow ICit	-0.881 7	0.727 9
H _{ex} +NADH \leftrightarrow NADPH	Cit \leftrightarrow ICit	0.935 6	0.369 1
3H _{ex} \leftrightarrow ATP	AcCoA+OxA \rightarrow Cit	-0.881 7	0.727 9
3H _{ex} \leftrightarrow ATP	Succ \rightarrow Fum+QuiH ₂	-0.880 7	0.717 4
H _{ex} +NADH \leftrightarrow NADPH	AcCoA+OxA \rightarrow Cit	0.935 6	0.369 1
H _{ex} +NADH \leftrightarrow NADPH	Succ \rightarrow Fum+QuiH ₂	0.935 5	0.372 0

表 5 不同碳源下模拟实验与真实实验的比较

Tab. 5 The comparisons between *in silico* and *in vivo* under different carbon resources

流量数据	模拟数据	实验数据
葡萄糖碳源 (有氧)	葡萄糖摄取速率: $2.17\text{ g}\cdot\text{L}^{-1}\cdot\text{h}^{-1}$ 细胞量增长速率: $(0.221\ 3\pm 0.002\ 2)\text{ g}\cdot\text{L}^{-1}\cdot\text{h}^{-1}$	葡萄糖摄取速率为 $2.17\text{ g}\cdot\text{L}^{-1}\cdot\text{h}^{-1}$ 时, 细胞量增长速率为 $0.213\text{ g}\cdot\text{L}^{-1}\cdot\text{h}^{-1}$ [19]
葡萄糖碳源 (无氧)	葡萄糖摄取速率: $21\text{ mmol}\cdot\text{g}^{-1}(\text{DW})\cdot\text{h}^{-1}$ 细胞倍增时间: $66.2\sim 70\text{ min}$	葡萄糖摄取速率: $21\text{ mmol}\cdot\text{g}^{-1}(\text{DW})\cdot\text{h}^{-1}$ 细胞的倍增时间: 70 min [20]
乙酸碳源 (有氧)	乙酸摄取速率: $41.4\text{ mmol}\cdot\text{g}^{-1}(\text{DW})\cdot\text{h}^{-1}$ 细胞倍增时间: $59.4\sim 61.4\text{ min}$ CO ₂ 排放: $(40.7\pm 0.6)\text{ mmol}\cdot\text{g}^{-1}(\text{DW})\cdot\text{h}^{-1}$	乙酸摄取速率: $41.4\text{ mmol}\cdot\text{g}^{-1}(\text{DW})\cdot\text{h}^{-1}$ 细胞的倍增时间: 70 min CO ₂ 排放: $45\text{ mmol}\cdot\text{g}^{-1}(\text{DW})\cdot\text{h}^{-1}$ [20]
甘油碳源 (有氧)	甘油摄取速率: $7.6\text{ mmol}\cdot\text{g}^{-1}(\text{DW})\cdot\text{h}^{-1}$ 生物量合成速率: $(0.41\pm 0.004\ 4)\text{ h}^{-1}$	甘油摄取速率: $7.6\text{ mmol}\cdot\text{g}^{-1}(\text{DW})\cdot\text{h}^{-1}$ 生物量合成速率: 0.22 h^{-1} [21]
琥珀酸碳源 (有氧)	生物量合成效率: $(0.048\ 4\pm 0.000\ 7)\text{ g}(\text{DW})\cdot\text{mmol}(\text{Succinate})^{-1}$	生物量合成效率: $0.051\text{ g}(\text{DW})\cdot\text{mmol}(\text{Succinate})^{-1}$ (线性规划)

散反应. 第一个主成分是由生物量合成所决定的一系列反应, 第二个则是代谢通路中的某些主要反应. 这刚好把整个网络划分为两个子结构, 一部分代谢, 一部分合成, 很符合中心代谢网络的结构特点.

3 比较与讨论

3.1 模拟实验(*in silico*)结果与真实实验(*in vivo*)结果的比较

我们分别将各种情况下以生物量为主的流量与真实实验数据进行了比对(琥珀酸单碳源由于缺乏真实实验室数据用他人线性规划的结果替代比较), 如表 5 所示. 可以发现, 大部分的模拟数据所计算出来的生物量与真实实验非常接近, 葡萄糖碳源有氧环境[19]和无氧环境[20]下结果近似, 乙酸碳源[20]稍高; 只有甘油碳源[21]有较大偏差, 可能由于代谢水平上的约束不够或基因水平的调控导致.

同时, 第 2 节中相关分析与真实实验观察到的宏观状况有着很好的一致性. 例如(2.1)节中所描述

的不同碳源由于氧化态的不同造成的碳原子利用效率的不同和氧气的摄入差别, 甘油和葡萄糖在代谢水平上的拮抗效应; (2.2)节中观察到的在有氧无氧不同条件下葡萄糖摄入模式的变更, 两种状况下不同的活跃反应群; 以及(2.4)节中主成分分析反映出的网络结构.

而在(2.3)节中不同温度下反应所呈现的相关性, 则是对某些关键调控通路和柔性通路很好的表征, 在代谢工程中可以做进一步的研究.

3.2 与其他方法的比较

目前基于约束的方法中, 线性规划的方法和稳态空间的均一随机采样是两种很重要的方法, 但是它们都有自身的局限性. 线性规划的方法虽然可以很快地找到最优解, 但是由于线性规划要求约束都是线性的, 因此无法处理非线性约束(例如环流约束), 而且最终只能找到一个最优解, 无法对稳态空间进行描述. 均一随机采样的方法[9,10]虽然可以描述稳态空间的分布状况以及研究流量之间的相关

性,但是如果采样维度较高,样本也很难在稳态空间内充分有效,很难收敛。

以葡萄糖碳源的有氧状况为例,运用线性规划优化生物量的合成。

$$\begin{aligned} & \max v_{\text{growth}}; \\ \text{st.} & \begin{cases} \mathbf{I} \cdot \mathbf{v} \geq \mathbf{b} & \text{下界约束} \\ \mathbf{S} \cdot \mathbf{v} = 0 & \text{稳态约束} \\ \sum_i v_{\text{sub}i} \cdot n_{\text{sub}i} = 1 & \text{碳原子归一化 (17)} \\ v_i \leq 10 & \text{上界约束} \end{cases} \\ & b_i = \begin{cases} -10; & \text{当第 } i \text{ 个反应可逆时} \\ 0; & \text{当第 } i \text{ 个反应不可逆时} \end{cases} \quad (18) \end{aligned}$$

与(1.3)节中的方法类似, \mathbf{I} 是一个单位矩阵; \mathbf{S} 是系统计量系数矩阵; \mathbf{v} 代表一组代谢流量; b_i 对于可逆反应取-10,不可逆反应取0。线性约束中加入了碳原子归一化(见式(14)),得到了一个可行解,其中生物量合成流量为 $0.0177 \text{ g(DW)} \cdot \text{mmol}^{-1}$ 。但这只是一个最优的可行解,真实实验中也无法达到这个最优解,也无从分析流量分布和流量间的关系。

接下来我们比较了随机采样和本文的优化采样。图6为葡萄糖碳源有氧情况下随机采样和优化采样的生物量合成流量的分布图。其中随机采样使用了 10^5 样本点,优化采样仅使用了 10^4 个样本点。可以看出虽然随机采样分布较为宽广,但是由于随机采样无法控制,收敛性较差,与真实实验数据也相差较远;而优化采样使得样本分布在最优生长速度的附近,与实验数据偏离很小,可靠性高。对随机样本做(2.1),(2.2)节中的分析,很多性质丢失。由于随机采样样本所有的流量变化范围都很大,无法做(2.3)节中提到的各种流量刚性的分析,仅(2.4)节

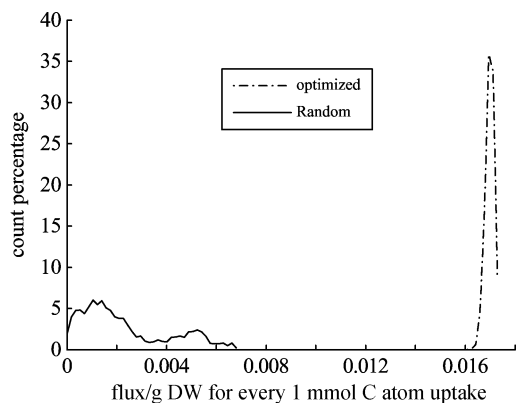


图6 随机采样和优化采样的生物量合成流量分布图

Fig. 6 Biomass flux distributions under random sampling and optimized sampling

中的主成分分析继续有效。

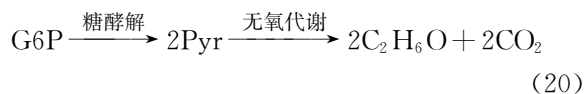
此外,如(1.4)节所述,运用本文的方法可以很简便地将优化采样转变为随机采样,只需令势能函数中的优化目标为零便可以在满足各类约束条件的情况下做随机采样。

3.3 乙醇发酵的讨论

前述实验的目标函数是使细胞生长最快,但是在很多实际应用中人们并不希望细胞将最多的物质用于自身生长,而是更关注那些副产物,例如乙醇。由此提出一个构想,通过修改目标函数来实现细菌的不同生长方式,得到一个新的流量分布,指导代谢工程的遗传操作。我们试探性地进行了优化乙醇的实验,结果表明其中大量与生物量合成反应相关的反应流量都很小,因为这些反应会影响到最优乙醇产出的碳原子利用效率。在以生长为方向的势能函数下,乙醇并不是主要的发酵产物,只有24.02%的碳原子转化为乙醇,占有所有发酵物的34.65%;而在新的势能函数的引导下,每摄取1 mmol葡萄糖生成1.98 mmol乙醇和1.95 mmol CO_2 ,葡萄糖几乎完全按照理论转化为乙醇和 CO_2 。



进行流量的对比可以发现流量发生了两极分化:一方面零流量增多;另一方面出现很多较大的且大小接近的流量,其目的是为了减少支路,最大效率地实现物质转化。这些较大的流量构成了如下的反应主干:



在实际中,可以应用遗传操作实现对流量的控制,从而提高乙醇产率。

4 结论

本文阐述的这种新方法不仅继承了基于约束的采样方法的优点,在网络结构已知的基础上,短时间内能得到大量的数据;而且还能保证在各种约束被满足的同时围绕最优目标进行采样,比现有的方法更方便有效。模拟数据与真实实验数据也具有较好的一致性,说明此方法是有效可行的。通过修改网络以及目标函数,可以改造出满足特殊要求的菌株,定量地验证和建议代谢工程对细胞的改造。但是由于实验数据的缺乏,现有算法只对碳源做出了限制,为了更好地拟合真实情况,可以加入对其他流量的最

大值和最小值的约束,例如氧气和氮源的输入是有上限的,生物量和其他的一些生物生长所必须的流量要有最低的保证等.此外,模拟所使用的网络是一个静态的网络,暂时没有考虑更深层次的、基因水平的调控关系.下一步的工作将围绕更大的全基因组代谢网络展开,并且尽可能地加入更多的约束和基因调控.这样会使得代谢分析更准确,在代谢工程中更有价值.

致谢 本文得到了刘海燕教授的精心指导,作者谨向他表示衷心的感谢!

参考文献(References)

- [1] Feist A M, Henry C S, Reed J L, et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information[J]. *Mol Syst Biol*, 2007, 3;121,doi:10.1038/msb4100155.
- [2] Duarte N C, Becker S A, Jamshidi N, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data[J]. *Proc Natl Acad Sci USA*, 2007, 104(6):1 777-1 782.
- [3] Thiele I, Vo T D, Price N D, et al. Expanded metabolic reconstruction of *Helicobacter pylori* (iT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants[J]. *J Bacteriol*, 2005, 187(16):5 818-5 830.
- [4] Stelling J. Mathematical models in microbial systems biology[J]. *Curr Opin Microbiol*, 2004,7(5):513-518.
- [5] Price N D, Papin J A, Schilling C H, et al. Genome-scale microbial in silico models: the constraints-based approach[J]. *Trends Biotechnol*, 2003,21(4):162-169.
- [6] Price N D, Reed J L, Palsson B O. Genome-scale models of microbial cells: evaluating the consequences of constraints[J]. *Nat Rev Microbiol*, 2004, 2(11): 886-897.
- [7] Varma A, Palsson B O. Metabolic flux balancing: Basic concepts, scientific and practical use[J]. *Bio/Technology*, 1994,12:994-998.
- [8] Almaas E, Kovacs B, Vicsek T, et al. Global organization of metabolic fluxes in the bacterium *Escherichia coli*[J]. *Nature*, 2004,427:839-843.
- [9] Price N D, Schellenberger J, Palsson B O. Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies [J]. *Biophys J*, 2004,87(4):2 172-2 186.
- [10] Price N D, Thiele I, Palsson B O. Candidate states of *Helicobacter pylori*'s genome-scale metabolic network upon application of "loop law" thermodynamic constraints[J]. *Biophys J*, 2006,90(11):3 919-3 928.
- [11] Metropolis N, Rosenbluth A W, Rosenbluth M N, et al. Equation of State Calculations by Fast Computing Machines[J]. 1953,21:1 087-1 092.
- [12] Wiback S J, Famili I, Greenberg H J, et al. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space[J]. *J Theor Biol*, 2004,228(4):437-447.
- [13] Stelling J, Klamt S, Bettenbrock K, et al. Metabolic network structure determines key aspects of functionality and regulation[J]. *Nature*, 2002, 420: 190-193.
- [14] Beard D A, Liang S D, Qian H. Energy balance for analysis of complex metabolic networks[J]. *Biophys J*, 2002,83(1):79-86.
- [15] Beard D A, Babson E, Curtis E, et al. Thermodynamic constraints for biochemical networks [J]. *J Theor Biol*, 2004, 228(3):327-333.
- [16] Heuett W J, Qian H. Combining flux and energy balance analysis to model large-scale biochemical networks[J]. *J Bioinform Comput Biol*, 2006, 4(6): 1 227-1 243.
- [17] Mahadevan R, Schilling C H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models [J]. *Metab Eng*, 2003, 5(4): 264-276.
- [18] de Groot B L, van Aalten D M, Amadei A, et al. The consistency of large concerted motions in proteins in molecular dynamics simulations[J]. *Biophys J*, 1996, 71(4):1 707-1 713.
- [19] Gokarn R R, Eiteman M A, Altman E. Expression of pyruvate carboxylase enhances succinate production in *Escherichia coli* without affecting glucose uptake[J]. *Biotechnology Letters*, 1998,20:795-798.
- [20] Pramanik J, Keasling J D. Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements [J]. *Biotechnology and Bioengineering*, 1997,56:398-421.
- [21] Fong S S, Joyce A R, Palsson B O. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states[J]. *Genome Res*, 2005,15(10):1 365-1 372.