

# 谷氨酸棒状杆菌集成细胞网络的构建与结构分析

姜金国<sup>1,2</sup>, 宋理富<sup>2,3</sup>, 郑平<sup>2,3</sup>, 贾士儒<sup>1</sup>, 孙际宾<sup>2,3</sup>

1 天津科技大学生物工程学院, 天津 300457

2 中国科学院天津工业生物技术研究所, 天津 300308

3 中国科学院系统微生物工程重点实验室, 天津 300308

姜金国, 宋理富, 郑平, 等. 谷氨酸棒状杆菌集成细胞网络的构建与结构分析. 生物工程学报, 2012, 28(5): 577-591.

Jiang JG, Song LF, Zheng P, et al. Construction and structural analysis of integrated cellular network of *Corynebacterium glutamicum*. Chin J Biotech, 2012, 28(5): 577-591.

**摘要:** 谷氨酸棒状杆菌是一种重要的传统工业微生物, 其基因组学和分子遗传操作工具的快速发展使得谷氨酸棒状杆菌具备了作为新型细胞工厂的潜力。但是, 相对于大肠杆菌等模式生物, 对于棒状杆菌的代谢调控研究较少, 特别是目前还缺乏谷氨酸棒状杆菌集成细胞网络的研究, 这一现状阻碍了谷氨酸棒状杆菌的系统生物学研究和大规模菌种理性设计优化。文中综合应用公共数据库、文献数据库资源, 首次构建了谷氨酸棒状杆菌的集成细胞网络, 包含 1 384 个反应, 1 276 个代谢物, 88 个调节子, 999 对转录调控关系。其转录调控可分为 5 层, 代谢网络呈现出清晰的 bow-tie 结构。文中还以赖氨酸的生物合成为例, 提出了一种提取代谢调控子网络的新方法, 这对氨基酸等产品高产生物机制的研究和工程菌株的重新设计具有指导意义。

**关键词:** 谷氨酸棒状杆菌, 转录调控网络, 代谢网络, 集成细胞网络, 代谢调控子网络, 赖氨酸生产

## Construction and structural analysis of integrated cellular network of *Corynebacterium glutamicum*

Jinguo Jiang<sup>1,2</sup>, Lifu Song<sup>2,3</sup>, Ping Zheng<sup>2,3</sup>, Shiru Jia<sup>1</sup>, and Jibin Sun<sup>2,3</sup>

1 School of Biological Engineering, Tianjin University of Science and Technology, Tianjin 300457, China

2 Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

3 Key Laboratory of Systems Microbial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

**Abstract:** *Corynebacterium glutamicum* is one of the most important traditional industrial microorganisms and receiving

**Received:** October 21, 2011; **Accepted:** January 9, 2012

**Supported by:** National Basic Research Program of China (973 Program) (Nos. 2009CB724703, 2011CBA00804), Knowledge Innovation Program of Chinese Academy of Sciences (No. KSCX2-YW-G-030).

**Corresponding author:** Jibin Sun. Tel: +86-022-84861949; Fax: +86-022-84861943; E-mail: sun\_jb@tib.cas.cn

国家重点基础研究发展计划 (973 计划) (Nos. 2009CB724703, 2011CBA00804), 中国科学院知识创新工程重要方向性项目 (No. KSCX2-YW-G-030) 资助。

more and more attention towards a novel cellular factory due to the recently rapid development in genomics and genetic operation toolboxes for *Corynebacterium*. However, compared to other model organisms such as *Escherichia coli*, there were few studies on its metabolic regulation, especially a genome-scale integrated cellular network model currently missing for *Corynebacterium*, which hindered the systematic study of *Corynebacterium glutamicum* and large-scale rational design and optimization for strains. Here, by gathering relevant information from a number of public databases, we successfully constructed an integrated cellular network, which was composed of 1 384 reactions, 1 276 metabolites, 88 transcriptional factors and 999 pairs of transcriptional regulatory relationships. The transcriptional regulatory sub-network could be arranged into five layers and the metabolic sub-network presented a clear bow-tie structure. We proposed a new method to extract complex metabolic and regulatory sub-network for product-orientated study taking lysine biosynthesis as an example. The metabolic and regulatory sub-network extracted by our method was more close to the real functional network than the simplex biochemical pathways. The results would be greatly helpful for understanding the high-yielding biomechanism for amino acids and the re-design of the industrial strains.

**Keywords:** *Corynebacterium glutamicum*, transcriptional regulatory network, metabolic network, integrated cellular network, metabolic and regulatory sub-network, lysine production

细胞的各种组成分子在一定的空间内相互作用,协调实现生命的机能。所有的细胞分子及其相互作用,就构成了细胞网络。细胞网络是系统生物学的重要研究内容之一,是从全局和整体上把握和理解复杂的代谢生理现象、实现对生命的有效调控、部分改造和从头设计创造的基础<sup>[1]</sup>。尽管生物分子发挥其机能是借助于同一个细胞网络,但是在传统研究中,根据研究者研究兴趣和生物分子参与的反应类型,往往将细胞网络人为分解成不同类型的子网络分别研究,如基因组规模代谢网络 (Genome-scale metabolic network, GSMN)、转录调控网络 (Transcriptional regulatory network, TRN) 和信号转导网络 (Signal transduction network) 等<sup>[2-3]</sup>。近年来,研究者已经逐步认识到这种分解所带来的问题,比如 Ma 等<sup>[4]</sup>研究大肠杆菌转录调控网络时,发现其中找不到生物学中最常见的反馈调控,进一步研究表明,这些转录反馈调控大部分是通过代谢网络与转录调控网络协同实现的,仅有 7 个是发生在转录调控这一单一网络层次上<sup>[5]</sup>。如今,以单一网络为基础的多网络集成,已成为系统生物

学领域的重要发展趋势,虽然相关的分析软件还相对缺乏,但是集成分析已经产生了一些重要的结果。Wang 等<sup>[6]</sup>集成了转录调控和蛋白质相互作用网络,认为网络集成后更接近于真实的生物网络。Yeang<sup>[7]</sup>、Herrgård<sup>[8]</sup>和 Chandrasekaran<sup>[9]</sup>等通过集成调控与代谢网络,认为集成网络比单一网络更能拟合生物的生长状态。Covert<sup>[10]</sup>和 Lee<sup>[11]</sup>等整合了信号转导、代谢和调控网络,发现多网络的集成在表型预测时更加准确。

谷氨酸棒状杆菌是最重要的工业微生物之一,被广泛应用于氨基酸 (如谷氨酸、赖氨酸等)、维生素 (如泛酸等) 和核酸的工业化生产<sup>[12]</sup>,所生产的产品总量达到每年数百万吨的规模。谷氨酸棒状杆菌细胞结构和调控关系相对简单,对于简单原料和复杂工业环境适应性好,生长快速,具有较好的分子改造基础,这些优势使得谷氨酸棒状杆菌具备了作为一个通用优秀细胞工厂的基础,研究者正在研究和改造棒状杆菌生产有机酸、醇等大宗化学品<sup>[13]</sup>。日本学者 Okino 等<sup>[14]</sup>通过改造谷氨酸棒状杆菌生产丁二酸,其产品浓度和发酵速度都达到了世界领先的水平,成为棒

杆菌细胞工厂改造的一个范例。对工业菌株的研究和改造离不开对其细胞网络的了解和认识,谷氨酸棒状杆菌的基因转录调控和代谢网络也一直是人们关注的焦点。Ikeda 等<sup>[15]</sup>和 Kalinowski 等<sup>[16]</sup>分别完成了谷氨酸棒状杆菌 *Corynebacterium glutamicum* ATCC13032 菌株的基因组测序,将谷氨酸棒状杆菌的研究带入了系统生物学新纪元。Kjeldsen 等<sup>[17]</sup>和 Shinfuku 等<sup>[18]</sup>利用基因组注释数据库分别构建了含有约 500 个反应的谷氨酸棒状杆菌基因组规模代谢网络。Brinkrolf 等<sup>[19]</sup>通过对文献报道数据的整理构建了首个谷氨酸棒状杆菌转录调控网络。但是目前关于谷氨酸棒状杆菌的研究大多只是针对单一网络,由于单一网络突现的信息割裂的局限性,限制了棒杆菌的研究和开发利用,急需构建基因组规模多网络集成的细胞网络。

本研究在前人工作的基础上,系统整理和分析了谷氨酸棒状杆菌的转录调控信息,对以往报道的代谢网络进行扩展,并借鉴 bow-tie 模型<sup>[20]</sup>的分析方法对谷氨酸棒状杆菌的代谢网络进行初步分析,最后以基因-酶-反应关系为接口对转录调控网络和代谢网络进行集成,为大规模的数据整合和模拟计算分析奠定了基础。同时文中还以赖氨酸的生物合成为例提出一种从集成细胞网络中提取代谢调控子网络的新方法,希望从系统水平上能更好地理解氨基酸等产品的生物合成与调控机制,对工程菌株的定向设计改造提出指导性建议。

## 1 方法

### 1.1 TRN 数据的采集与处理

从谷氨酸棒状杆菌调控网络数据库

CoryneRegNet<sup>[21]</sup> 5.0 提取转录调控信息,并结合文献<sup>[22-24]</sup>对数据补充和校正,去掉不确定的调控关系 *sugR* (cg2115)→*ramA*(cg2831) 和 *uriR* (cg1547)→*rbsR* (cg1410) (文献报道中无此调控关系的结论)。用转录调控因子 (Transcription factors, TF) 的编码基因来代替 TF 本身 (即所有节点均为基因),再根据基因的 cg 号进行排序,用 VBA 程序处理成可视化软件 Cytoscape<sup>[25]</sup>和 Pajek<sup>[26]</sup>可识别的文件格式。

### 1.2 TRN 的分层分解

我们借鉴 Ma 等<sup>[27]</sup>报道的分层思想对谷氨酸棒状杆菌的 TRN 进行分解,首先将出度为 0 的节点分配到网络第一层;然后移除这些节点,从剩余的节点中,将出度为 0 的节点分配到网络的第二层;依次类推,直到把所有的节点分配到不同的层中。

### 1.3 GSMN 数据采集与处理

#### 1.3.1 数据采集

用 Perl 程序,从 KEGG 数据库 (<http://www.genome.jp/kegg/>) 的 Ligand/reaction 文件中提取反应数据,结合谷氨酸棒状杆菌基因组注释信息 (cgb\_enzyme 和 cgb\_ko 文件),并保留自发反应,便可生成谷氨酸棒状杆菌 GSMN 的初始模型。

同时,我们根据文献数据<sup>[17-18]</sup>提供的化合物信息,用 VBA 程序从 KEGG/Ligand/compound 和 glycan 数据中搜索对应的化合物编号 (C number 和 G number),再由化合物编号从 KEGG/Ligand/reaction/reaction 数据中搜索到相应的反应号 (R number),然后从上述 GSMN 的初始模型中去掉与文献数据中反应号重叠的反应,但是保留先前反应的反应对 (Reactant pairs, 指底物与产物

的对应关系。例如,在反应 R00299: ATP+D-葡萄糖 $\rightleftharpoons$ ADP+D-葡萄糖 6-磷酸中, D-葡萄糖与 D-葡萄糖 6-磷酸就是一个反应对) 的信息。例如文献[18]中的反应: 丙酮酸+CoA+NAD<sup>+</sup> $\Rightarrow$ 乙酰-CoA+CO<sub>2</sub>+NADH+H<sup>+</sup>, 我们通过 VBA 程序搜索后发现它与 KEGG 中的反应 R00209: 丙酮酸+CoA+NAD<sup>+</sup> $\rightleftharpoons$ 乙酰-CoA+CO<sub>2</sub>+NADH+H<sup>+</sup> 基本匹配,但是 KEGG 中该反应的可逆性与文献中的不一致,鉴于文献中的数据有计算模拟和实验的支持,而 KEGG 中的数据仅供参考,所以最终我们把反应 R00209 定义为: 丙酮酸+CoA+NAD<sup>+</sup> $\Rightarrow$ 乙酰-CoA+CO<sub>2</sub>+NADH+H<sup>+</sup>, 反应对信息仍采用 KEGG 中 R00209 的反应对 (因为文献中的反应没有反应对的信息), 即 NAD<sup>+</sup>—NADH、CoA—乙酰-CoA、丙酮酸—乙酰-CoA 和丙酮酸—CO<sub>2</sub>。最后,通过将 KEGG 数据和文献代谢网络数据相整合,生成谷氨酸棒状杆菌 GSMN 模型。

### 1.3.2 反应方向的重新定义

反应的方向对于生物网络构建非常重要,但是 KEGG 中很多反应的方向并不可靠或没有定义,对此我们主要以文献报道数据中的反应方向为主,其余的采用 Ma 等<sup>[28]</sup>总结的 11 条规则重新定义。

### 1.3.3 反应对的拆分

KEGG 中的很多反应并未给出反应对信息,我们根据 KEGG 对已有反应对的拆分规律 RDM (Reaction center, R; Difference region, D; Matched region, M) 模式<sup>[29]</sup>对信息不全的反应进行拆分。

### 1.3.4 特殊反应对的删除

在 GSMN 中某些化合物出现的频率很高,

被称为通用代谢物 (Currency metabolites), 如 ATP、ADP、NADH、H<sub>2</sub>O、Pi 等,它们通常携带一些供转移的小分子化学基团 (如磷酸基、氨基、一碳单位、甲基等)<sup>[28]</sup>。这些通用代谢物所连接的一些特殊反应对必须去掉,否则整个网络的路径将会缩短,不利于后期拓扑结构的分析,而且与实际的生物学意义不符 (如 R00258: L-丙氨酸+2-酮戊二酸 $\rightleftharpoons$ 丙酮酸+L-谷氨酸中的反应对 L-丙氨酸—L-谷氨酸和 2-酮戊二酸—丙酮酸应该去掉,仅保留 L-丙氨酸—丙酮酸和 2-酮戊二酸—L-谷氨酸)。

## 1.4 GSMN 的 bow-tie 结构

为了降低网络复杂度,谷氨酸棒状杆菌的 GSMN 中节点均为代谢物。文中综合 Zhao 等<sup>[30]</sup>和 Ma 等<sup>[20]</sup>对代谢网络的分解方法,首先借助软件 Pajek 求出最大强连通体 (Giant strong component, GSC), 然后收缩 GSC 为单节点,再分别求出 GSC 的输入域 (S) 和输出域 (P), 剩余的节点归为孤立子集 (Isolated subset, IS) (同一个代谢物在细胞不同部位出现时,分别赋予不同的 ID, 例如细胞膜上和细胞内同一代谢物的 ID 不同)。

## 1.5 细胞网络的集成

以基因 $\rightarrow$ 酶的关系为接口可以将转录调控网络 (调节子 $\rightarrow$ 靶基因) 和代谢网络 (底物 $\rightarrow$ 反应 $\rightarrow$ 产物; 酶 $\rightarrow$ 反应) 整合起来,形成调节子 $\rightarrow$ 靶基因 $\rightarrow$ 酶 $\rightarrow$ 反应和底物 $\rightarrow$ 反应 $\rightarrow$ 产物的集成细胞网络模型 (Integrated cellular network, ICN)。

## 1.6 赖氨酸代谢调控子网络的提取

1) 在 ICN 数据中用酶的编码基因将酶替换,然后求葡萄糖的输出域,从生成的子网络中

提取赖氨酸的输入域得到节点集 A。

2) 将 A 收缩为单节点 (借鉴 Zhao 等<sup>[30]</sup>的方法), 并求其 K 输入节点集 (在有向图中, 如果某节点 j 在 K 步之内可以到达节点 i, 那么这些节点 j 组成的集合就是节点 i 的 K 输入节点集), 并从中去掉表示代谢物的节点, 仅保留基因节点, 组成 Genes 节点集合。

3) 将 Genes 收缩为单节点, 从其输出域中 去掉节点集 A, 得到节点集 B。

4) 求 A 的 1 近邻输入节点集 K1, 并从 K1 中提取仅是 A 的输入不是 B 的输入节点集记为 K11; K11 中调控关系已明确的节点集记为 C, 反之记为 H; K1 中既是 A 的输入又是 B 的输入的节点集记为 K12; K12 中调节基因已明确的节点集记为 F, 反之记为 G。

5) 求 A 的 2 近邻输入节点集 K2 (调节子); 从 K2 中提取既调控 F 又调控 C 的基因集记为 E, 仅调控 C 的记为 D。

以上步骤的分析均是借助 Pajek 软件和作者自己编写的 VBA 程序实现的。

## 2 结果

### 2.1 谷氨酸棒状杆菌 TRN 的分析

#### 2.1.1 TRN 的功能模块分析

我们最终构建的谷氨酸棒状杆菌 TRN 含有 660 个基因, 基于基因功能的生物学意义可以将其划分为 5 大功能模块<sup>[19]</sup>: A (碳水化合物代谢), B (细胞分裂与分隔), C (常量元素与金属离子的体内平衡), D (SOS 应激反应), E (特殊的生物合成与运输途径) (图 1)。

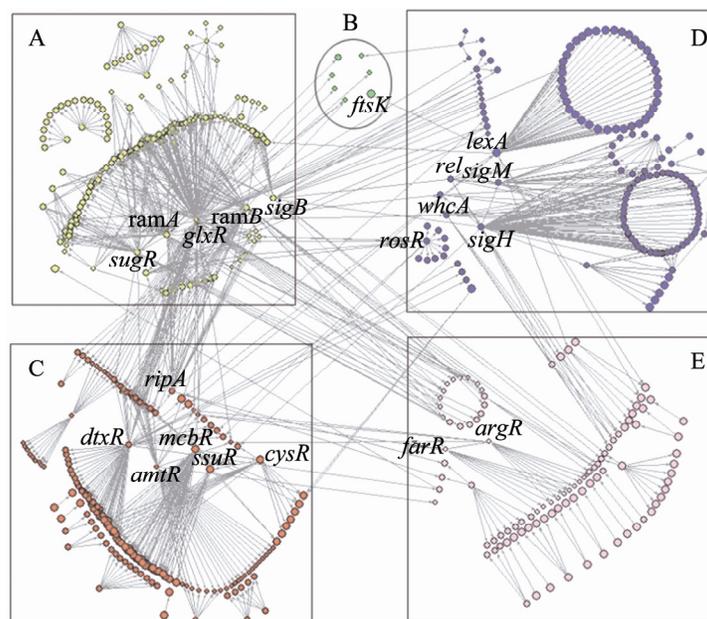


图 1 谷氨酸棒状杆菌 TRN 分解

Fig. 1 The decomposition of TRN of *Corynebacterium glutamicum*. (A) Carbohydrate metabolism. (B) Cell division and septation. (C) Macroelement and metal homeostasis. (D) SOS and stress response. (E) Specific biosynthesis and transport pathways.

模块 A 中的调控关系最为复杂 (212 个基因), 它控制着谷氨酸棒状杆菌的中心代谢。其中 *sugR*、*glxR*、*ramA*、*ramB* 和 *sigB* 还具有跨模块调节的作用, 如全局调节子 *glxR* 可以直接调控以上 5 个模块中的基因。此外, 位于模块 C 的 *dtxR* 调节子还调控模块 A 和模块 D 中的基因。模块 E 中的 *ArgR* 通过阻遏转氨酶的转录从而调控精氨酸代谢, 同时抑制谷氨酸脱氢酶的转录进而影响谷氨酸转化和 NADP 的体内平衡。模块 B 中未发现转录调节子, 基因均受到来自其他模块调节子的调控, 如细胞分裂蛋白基因 *fisK* 的转录要受到阻遏蛋白 LexA (模块 D) 的抑制。

### 2.1.2 全局调节子的重新定义

在细胞网络中, 某些调节子可调控 20 个以上属于不同转录单元的并且从属于多个功能模块的靶基因, 被称为全局调节子<sup>[31-32]</sup>。Brinkrolf 等<sup>[22]</sup>通过搜索调节子直接调控的靶基因和功能模块的数量, 认为在谷氨酸棒状杆菌 TRN 中只存在一个全局调节子 *glxR* (cg0350)。而事实上, 有些调节子的靶基因同样也是调节基因。因此, 我们引入调节子的输出域 (Output domain) 来替换其出度 (Out-degree) 重新分析, 并用自己编写的 VBA 程序广度搜索了调节子直接和间接调控的总模块数(表 1)。

表 1 谷氨酸棒状杆菌调节子调控的目标基因与模块的统计分析

**Table 1** Statistic analysis of target genes and modules which are regulated by regulons in *Corynebacterium glutamicum*

Gene ID	Gene name	Functional annotation	Number of DTG	Number of DRM	Number of DITG	Number of DIRM	Is global regulon?
cg0350	<i>glxR</i>	Repressor of acetate/gluconate metabolism and <i>dtxR</i> gene	151	5	196	5	True
cg1861	<i>rel</i>	(p)ppGpp synthetase involved in sigma factor competition	8	3	152	5	True
cg2831	<i>ramA</i>	Dual regulator of carbon metabolism	54	3	106	4	True
cg0876	<i>sigH</i>	ECF sigma factor involved in global stress response	67	4	99	4	True
cg2103	<i>dtxR</i>	Dual regulator of iron metabolism	63	4	80	4	True
cg0444	<i>ramB</i>	Repressor of acetate metabolism	53	4	54	4	True
cg2114	<i>lexA</i>	Dual regulator of SOS response	48	5	51	5	True
cg3253	<i>mcbR</i>	Repressor of sulfur metabolism	45	2	50	2	False
cg0986	<i>amtR</i>	Repressor exerting global nitrogen control	39	3	39	3	False
cg2115	<i>sugR</i>	Repressor involved in carbon metabolism	39	4	39	4	True
cg0156	<i>cysR</i>	Dual regulator of assimilatory sulfate reduction genes	17	1	26	1	False
cg1120	<i>ripA</i>	Repressor of iron protein expression	19	3	20	3	False

Note: only the regulons which regulate directly and indirectly no less than 20 target genes were listed above. DTG: direct target genes; DRM: direct regulatory modules; DITG: direct and indirect target genes; DIRM: direct and indirect regulatory modules.

结果我们发现了8个调节子调控影响范围广泛(直接和间接调控的目的基因总数不少于20,调控模块总数不少于4),可以认定为全局调节子:*glxR* (cg0350), *lexA* (cg2114), *ramA* (cg2831), *ramB* (cg0444), *sigH* (cg0876), *sugR* (cg2115), *rel* (cg1861) 和 *dtxR* (cg2103)。虽然有些调节子直接调控的功能模块并不多,但是它们还可以通过调控其他调节子间接调控一些模块。如 *rel* (cg1861) 通过调控调节子 *sigH* (cg0876) 实现对 cg2810 (Na<sup>+</sup>/H<sup>+</sup>泵转运系统, 模块 C) 的间接调控。这8个全局调节子共调控387个基因, 调控范围覆盖了整个 TRN 的59%。

### 2.1.3 TRN 的分层分解

借助网络分析软件 Pajek 我们发现谷氨酸棒状杆菌 TRN 中无强连通体存在, 这与枯草芽胞杆菌<sup>[33]</sup>和大肠杆菌<sup>[5]</sup>有所不同, 说明在 TRN 中无反馈环(基因 A 调控基因 B, 基因 B 也直接或间接调控基因 A) 的存在, 因此可以对 TRN 按层次分解, 如图2所示。

其中底层的基因最多(572个), 其次为第2层(68个)。分层后可以清楚地看到, *rel* (cg1861) 和 *dtxR* (cg2103) 分别位于顶层和第4层, 是重要的全局调节子。并且基因之间的调控关系还可以跨层实现, 如 *rel* (cg1861) 调节子对底层基因 cg0877 (*rshA*) 的调控不仅可以通过 *sigH* 间接实现, 还能直接完成, 这可以很好地解释谷氨酸棒状杆菌 TRN 的平均路径长度只有1.27, 即存在调控捷径。

## 2.2 GSMN 分析

### 2.2.1 GSMN 模型

经过对公共数据库和文献数据的收集和处理, 我们最终构建的 GSMN 模型含有1384个生化反应(其中978个为不可逆反应, 406个可逆反应)和1276个代谢物, 如图3A所示。去除经过通用代谢物的连接后, 连接数由原来的3783缩减为2803, 网络直径和平均路径<sup>[28]</sup>分别由原来的17和4.74变成了32和9.66, 网络明显得到了简化, 如图3B所示。

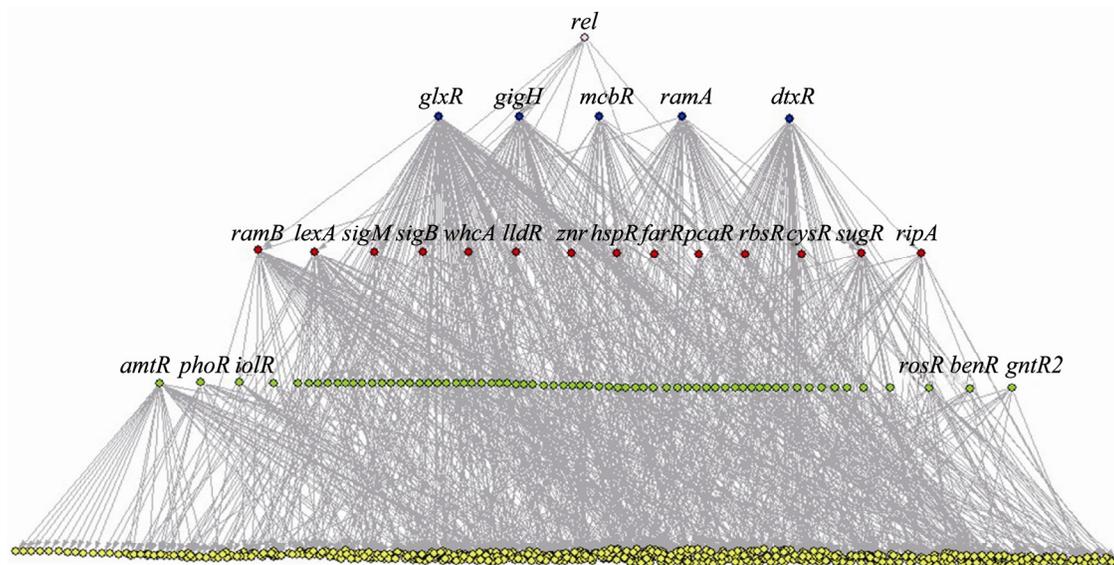


图2 谷氨酸棒状杆菌 TRN 的分层结构

Fig. 2 The hierarchical structure of TRN of *Corynebacterium glutamicum*.

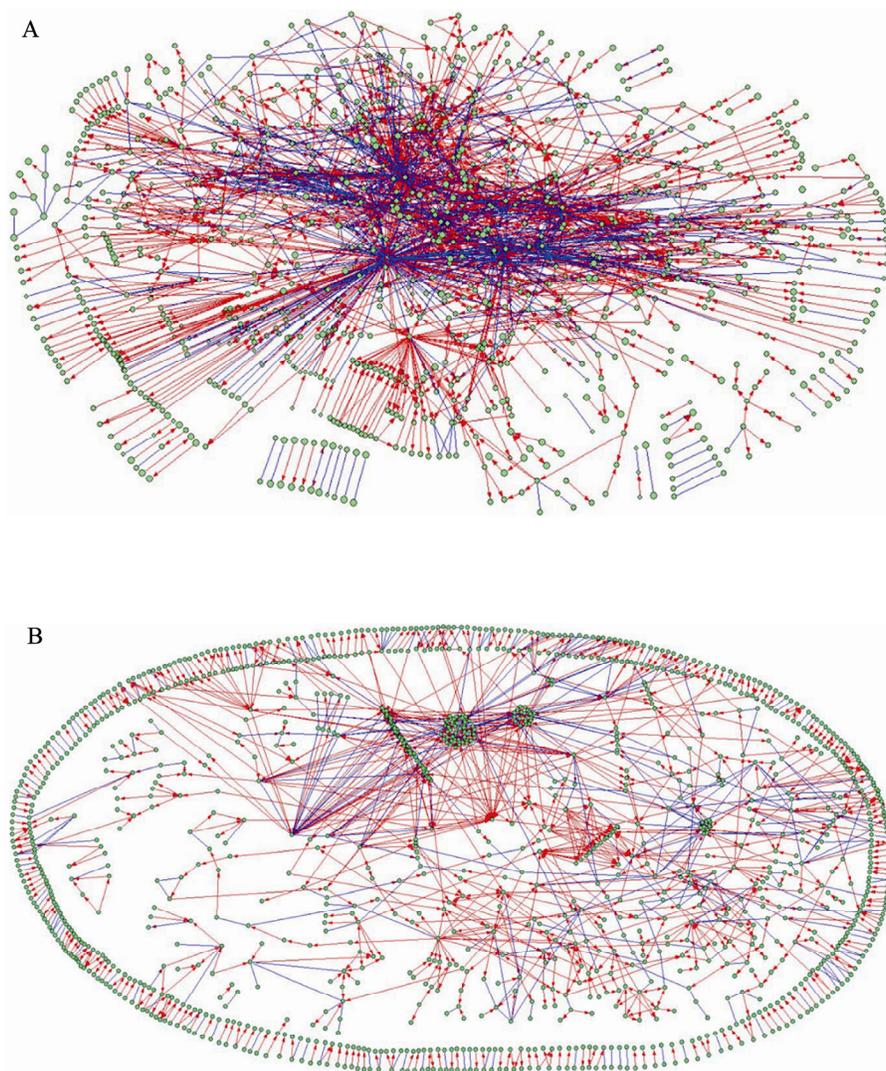


图3 谷氨酸棒状杆菌 GSMN 模型. (A) 含有经过通用代谢物连接的 GSMN. (B) 不含经过通用代谢物连接的 GSMN  
 Fig. 3 The GSMN model of *Corynebacterium glutamicum*. All the nodes represent metabolites. (A) GSMN including the connections through currency metabolites. (B) GSMN not including the connections through currency metabolites. The red lines with arrows correspond to irreversible reactions and the blue lines without arrows to reversible reactions.

### 2.2.2 Bow-tie 结构

经过分析,我们发现谷氨酸棒状杆菌的基因组规模代谢网络 GSMN 共 102 个强连通体 (节点数  $\geq 2$ ), 其中最大的强连通体 GSC 包含 456 个代谢物, 占整个代谢网络的 36% (图 4)。GSC 的平

均路径为 9.08, 而整个网络的平均路径为 9.66, 这说明与网络中的其他模块相比 GSC 中节点之间的联系更加紧密。按照 KEGG 对代谢反应的类别划分, 我们将 GSC 分解成 10 大功能模块 (图 5)。结果显示 GSC 几乎覆盖了所有的代谢功

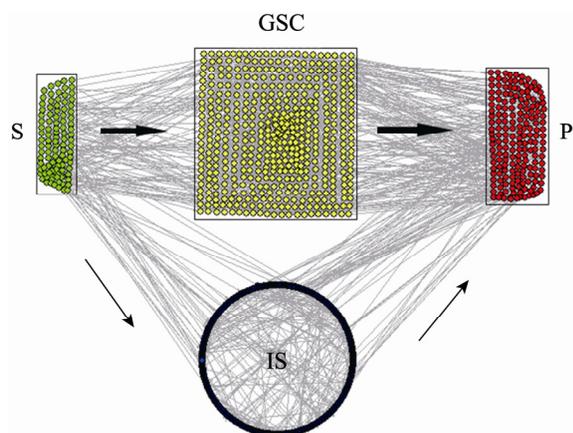


图4 谷氨酸棒状杆菌 GSMN 的 bow-tie 结构  
Fig. 4 The bow-tie structure of GSMN for *Corynebacterium glutamicum*. GSC: giant strong component; S: substrate subset; P: product subset; IS: isolated subset.

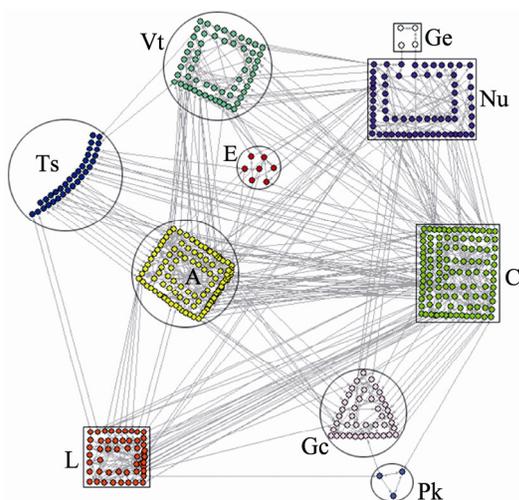


图5 GSC 的分解  
Fig. 5 The decomposition of GSC. A: amino acid metabolism; C: carbohydrate metabolism; E: energy metabolism; Ts: membrane transport or signal transduction; Gc: glycan biosynthesis and metabolism; Ge: genetic information processing; L: lipid metabolism; Nu: nucleotide metabolism; Pk: metabolism of terpenoids and polyketides; Vt: metabolism of cofactors and vitamins.

能,是整个代谢网络的核心,并且模块之间并非独立存在,而是通过“代理”(Broker)<sup>[27]</sup>或桥(Bridges)<sup>[34]</sup>(表2)相互联系。

### 2.2.3 GSC 的中心性分析

紧密度中心性 (Closeness centrality, CC)<sup>[27]</sup>和介数中心性 (Betweenness centrality, BC)<sup>[35-36]</sup>是用来衡量节点地位的重要参数,值越大节点就越重要,后者也常用来寻找网络的“代理”<sup>[27]</sup>或桥<sup>[34]</sup>以及分解网络<sup>[37]</sup>。经过计算分析,我们找到了29个中心代谢物 (Central metabolites)<sup>[20]</sup>(表2)。

很多化合物不仅仅参与了一个功能模块中的反应,如丙酮酸 (Pyruvate) 不仅是糖酵解和TCA的枢纽,而且连接着碳水化合物代谢和氨基酸代谢,并通过反应 R05553 (4-氨基-4脱氧分支酸 $\rightleftharpoons$ 4-氨基苯甲酸+丙酮酸)参与叶酸盐的生物合成。乙酰-CoA (Acetyl-CoA) 是网络的输出中心,连接着糖酵解、TCA、肽聚糖生物合成和脂类代谢以及氨基酸代谢。草酰乙酸 (Oxaloacetate) 是TCA中重要的代谢物,同时经反应 R00355 (草酰乙酸+L-谷氨酸 $\rightleftharpoons$ L-天冬氨酸+ $\alpha$ -酮戊二酸)使碳水化合物代谢与氨基酸代谢连接起来。

### 2.3 集成细胞网络的构建

我们以基因、酶、反应和代谢物为节点,相互作用关系为弧(边),去除通用代谢物参与的特殊反应对以及文献和数据库之间的冗余部分,最终所构建的谷氨酸棒状杆菌集成细胞网络包括1050个基因,785个酶,1384个反应,1276个代谢物,88个调节子,999对转录调控关系(图6)。

表 2 GSC 网络中中心性最大的几种中心代谢物

Table 2 The top central metabolites with higher centrality in GSC

Metabolites	Input CC ( $\times 10^{-2}$ )	Metabolites	Output CC ( $\times 10^{-2}$ )	Metabolites	BC ( $\times 10^{-2}$ )
Corynomycolic acid	18.19	Pyruvate	19.69	Pyruvate	37.77
C11437	16.34	Oxaloacetate	18.34	C00118	19.61
Formate	16.01	L-aspartate	18.02	D-ribose5-phosphate	15.58
C04465	15.96	Phosphoenolpyruvate	17.83	L-serine	14.64
Peptidoglycan	15.77	L-serine	17.78	Oxaloacetate	13.01
C11434	15.44	L-cysteine	17.49	D-fructose6-phosphate	12.53
C04691	15.40	Malate	17.19	D-glucose6-phosphate	12.01
N-formylkynurenine	15.34	L-alanine	17.14	L-tryptophan	11.48
C00118	15.22	N-acetylneuraminate	17.14	Acetyl-CoA	11.46
Acetate	15.20	Chorismate	16.98	C00119	10.95
D-ribose5-phosphate	15.10	Cystathionine	16.89	C03506	10.87
Pyruvate	15.06	Acetyl-CoA	16.87	Alpha, alpha-trehalose	10.53

Note: C11437 (1-deoxy-D-xylulose5-phosphate); C04465 (Trehalose Monocorynemycolate); C11434 (2-C-methyl-D-erythritol-4-phosphate); C04691 (3-deoxy-D-arabino-heptonate7-phosphate); C00118 (D-Glyceraldehyde3-phosphate); C00119 (5-Phos-pho-alpha-D-ribose1-diphosphate); C03506 (indole-3-glycerolphosphate).

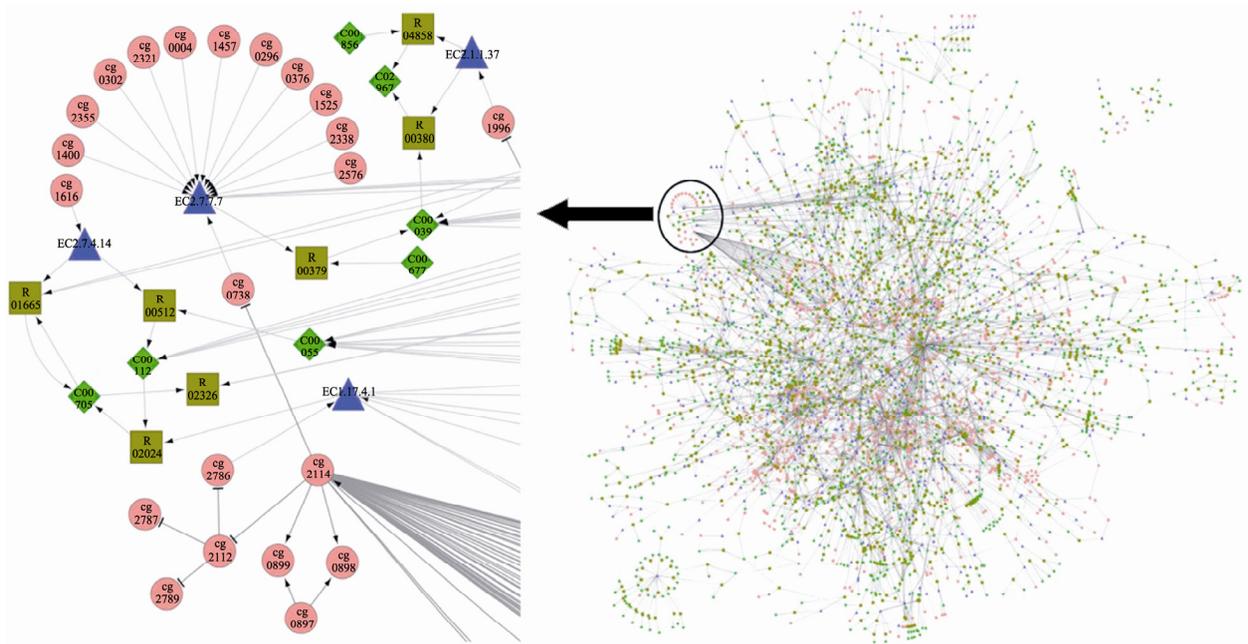


图 6 谷氨酸棒状杆菌集成细胞网络

Fig. 6 The integrated cellular network of *Corynebacterium glutamicum*. pink circle: gene; blue triangle: enzyme; green diamond: metabolite; green yellow square: reaction;  $\rightarrow$  activation;  $\dashv$  repression.

利用 Pajek 软件从构建的集成细胞网络中, 我们找到了 8 个基于转录与代谢调控的反馈环路。例如, *dtxR* 编码的转录因子 DtxR 可以抑制基因 *hrrA* (对 *hmuO* 的转录有激活作用) 和 *hmuO* 的转录, *hmuO* 编码的血红素氧化酶 (Heme oxygenase, HmuO) 可以催化亚铁血红素 Heme 生成铁离子, 而铁离子又是 DtxR 发挥作用时必需的辅因子, 这样就构成了一个反馈环路, 如图 7A 所示。该反馈回路对于维持 Heme 的体内平衡<sup>[38]</sup>和铁代谢<sup>[39]</sup>具有重要意义。此外, 在天冬氨酸家族 (如甲硫氨酸<sup>[40-42]</sup>) 和一些芳香族氨基酸的生物合成中也同样存在转录与代谢协同作用的复杂调控关系。例如, 在色氨酸的合成中, 几个关键酶 ANS (Anthranilate synthase)、PRT (Anthranilate phosphoribosyltransferase) 和 TS (Tryptophan synthase) 除了受终产物色氨酸的反

馈抑制外, 其编码基因的转录还受到色氨酸的衰减作用 (Attenuation) 的控制<sup>[43]</sup>, 如图 7B 所示。这些发生在细胞内的转录与代谢多层次复合调控关系在单一的代谢网络<sup>[17-18]</sup>或转录调控网络<sup>[19]</sup>中是无法体现的。

#### 2.4 L-赖氨酸生物合成代谢调控子网络的提取

分别以葡萄糖为底物约束和赖氨酸为产物约束, 从集成细胞网络中提取了赖氨酸生物合成的代谢调控子网络 (图 8)。它涉及的代谢功能模块有氨基酸代谢、碳水化合物代谢、能量代谢、膜运输与信号转导、多糖合成与代谢、遗传信息处理、脂类代谢、核酸代谢、萜类与聚酮代谢以及辅因子与维生素代谢, 而传统的赖氨酸合成途径<sup>[44]</sup>仅包括糖酵解、TCA 以及赖氨酸、苏氨酸等氨基酸合成途径。

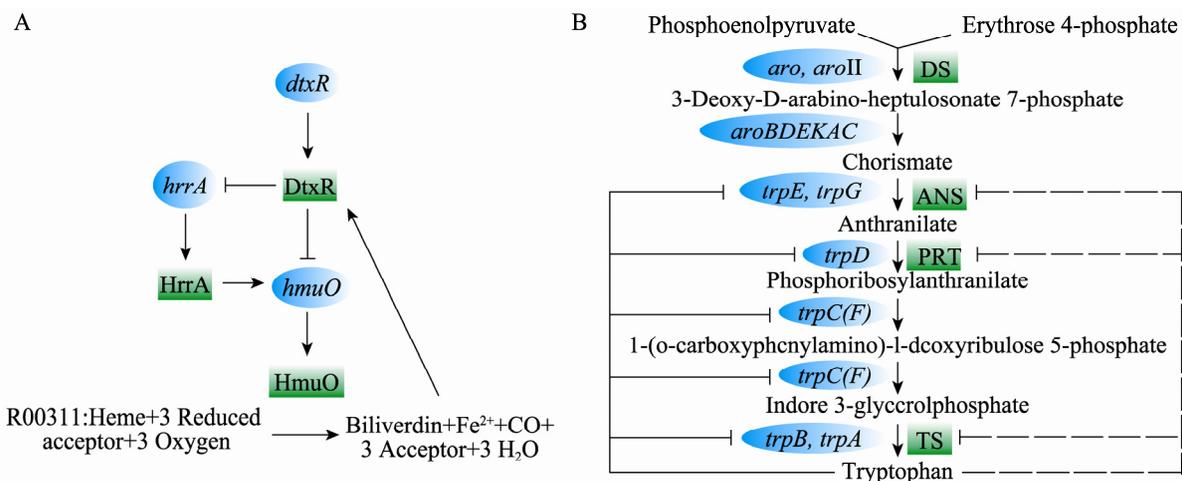


图 7 谷氨酸棒状杆菌中的复合调控回路。(A) 血红素铁代谢的转录与代谢调控反馈环。(B) 色氨酸的生物合成与调控

Fig. 7 Complex regulation loop in *Corynebacterium glutamicum*. (A) A feedback loop for transcriptional and metabolic regulation of Heme iron. (B) The biosynthetic pathway of tryptophan and its regulation. DS: 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase; blue ellipse: gene; green box: corresponding protein; → activation; —| repression. The dotted lines indicate feedback inhibition.

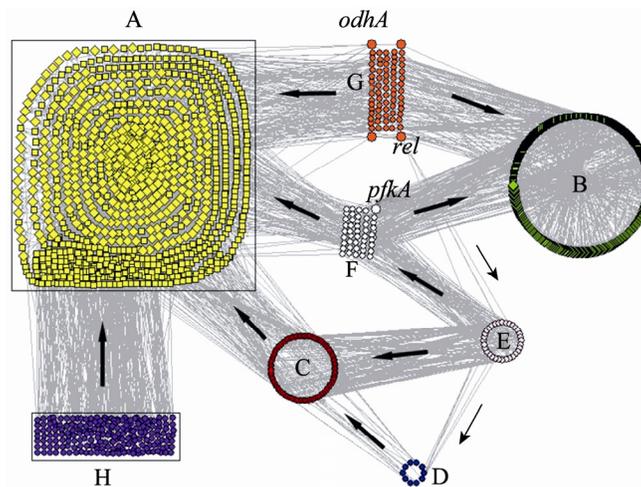


图 8 赖氨酸生物合成代谢调控子网络

Fig. 8 The metabolic and regulatory sub-network for lysine biosynthesis. (A) Lysine biosynthesis. (B) Other biosynthesis. C, F, G and H represent enzyme-coding genes. D and E represent regulons. Circle: gene; diamond: metabolite; square: reaction; arrow: the direction of material interaction.

模块 C 中的基因可以编码催化模块 A 中代谢反应的酶，同时受到模块 D 和模块 E 的调控。模块 F 和模块 G 中的基因所编码的酶不仅可以催化赖氨酸生物合成，还可以催化模块 B 中的反应，例如 *pfkA* (*cg1409*) 编码的磷酸果糖激酶既是中心代谢途径的重要限速酶 (R00756, 模块 A)，也是塔格糖酵解的重要酶 (R03236, 模块 B)，同时受到全局调节子 *glxR* 和 *sugR* 以及 *sigB* ( $\sigma$  因子) 的调控 (模块 E)。*rel* (*cg1861*, PPGPP 合成酶, 模块 E) 不仅可以编码催化 R00336 (模块 A) 和 R00429 (模块 B) 的酶，而且还是全局调节子。此外，模块 F 中的基因受到模块 E 中调节子的调控，而模块 G 和模块 H 所对应的调节基因，目前为止尚不清楚。

### 3 讨论

关于谷氨酸棒状杆菌转录调控网络和代谢

网络的研究已有相关文献介绍，但是对于集成细胞网络的分析尚未发现有人报道。文中从网络数据库和文献信息提取数据，成功构建了谷氨酸棒状杆菌的集成细胞网络模型，网络涉及 1 384 个生化反应和 999 对调控关系。通过后续的网络分析，发现了一些在单一网络中很难体现的转录与代谢多层次协同作用的复合调控关系，这为重新设计和优化改造谷氨酸棒状杆菌，发展新型细胞工厂奠定了基础。

通过对基因转录调控网络的分析，我们重新定义了 8 个全局调节子，其调控范围覆盖整个转录调控网络的 59%。转录调控网络可分为 5 层，并且有调控捷径的存在，这可能有利于生物体提高自身对环境变化的适应能力。

我们在谷氨酸棒状杆菌的代谢网络中找到了和大肠杆菌类似的 bow-tie 结构，其中 GSC 的强连通性可能有利于提高生物体的鲁棒性

(Robustness)。通过对 GSC 的进一步分析,我们找到了 29 个中心代谢物,它们像桥梁一样将一个个分开的反应模块紧密地连接在一起,对于微生物而言,这种结构可能有利于节省空间和资源,使生物代谢更加经济高效。

最后,基于集成细胞网络我们提出了一种代谢调控子网络的提取方法,与传统意义上的代谢途径相比,前者的代谢反应更加全面,物质间相互作用关系也更为复杂。应用这种提取方法,我们研究了从葡萄糖合成赖氨酸的代谢调控子网络,发现在这个复杂的子网络中,有些基因不仅编码赖氨酸合成代谢途径中的酶,同时还是其他代谢反应的酶编码基因(如 *cg1280 (odhA)* 和 *cg1409 (pfkA)*),更有甚者,如 *rel* 不仅编码赖氨酸合成途径中涉及的酶,同时还是全局调节子。这一结果告诉我们,在运用基因工程手段对菌株进行定向改造时,应该从系统生物学角度出发,通盘考虑转录调控和代谢反应,否则某些基因的改变可能会导致细胞整体代谢反应失去平衡,整个细胞生理机能发生改变。

本研究的工作,仅仅是谷氨酸棒状杆菌系统生物学工作的开始。在本研究提出的集成细胞网络的基础上,能够集成更多的实验数据,包括转录组、蛋白质组、代谢组、通量组数据,对发酵过程的动态数据进行全面地整合分析,将发现更多的代谢调控机制,进一步完善集成型细胞网络模型。同时,发展整合细胞网络的模拟计算方法,从代谢和调控的全局开发新型细胞工厂设计方法,提出最佳基因组改造策略,有可能大幅度改进谷氨酸棒状杆菌的性能,为新型生物制造产业的发展贡献力量。

致谢 感谢英国爱丁堡大学马红武博士的修改意见。

## REFERENCES

- [1] Kitano H. Systems biology: a brief overview. *Science*, 2002, 295(5560): 1662–1664.
- [2] Kim TY, Kim HU, Lee SY. Data integration and analysis of biological networks. *Curr Opin Biotechnol*, 2010, 21(1): 78–84.
- [3] Sik KJ, Yun H, Kim HU, et al. Resources for systems biology research. *J Microbiol Biotechnol*, 2006, 16(6): 832–848.
- [4] Ma HW, Buer J, Zeng AP. Hierarchical structure and modules in the *Escherichia coli* transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatics*, 2004, 5: 199.
- [5] Ma HW, Kumar B, Ditges U, et al. An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs. *Nucl Acids Res*, 2004, 32(22): 6643–6649.
- [6] Wang YC, Chen BS. Integrated cellular network of transcription regulations and protein-protein interactions. *BMC Syst Biol*, 2010, 4: 20.
- [7] Yeang CH, Vingron M. A joint model of regulatory and metabolic networks. *BMC Bioinformatics*, 2006, 7: 332.
- [8] Herrgård MJ, Lee BS, Portnoy V, et al. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res*, 2006, 16(5): 627–635.
- [9] Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA*, 2010, 107(41): 17845–17850.
- [10] Covert MW, Xiao N, Chen TJ, et al. Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics*, 2008, 24(18): 2044–2050.

- [11] Lee JM, Gianchandani EP, Eddy JA, et al. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput Biol*, 2008, 4(5): e1000086.
- [12] Eggeling L, Bott M. *Handbook of Corynebacterium glutamicum*. Boca Raton: CRC Press, 2005: 1–616.
- [13] Wendisch VF, Bott M, Eikmanns BJ. Metabolic engineering of *Escherichia coli* and *Corynebacterium glutamicum* for biotechnological production of organic acids and amino acids. *Curr Opin Microbiol*, 2006, 9(3): 268–274.
- [14] Okino S, Noburyu R, Suda M, et al. An efficient succinic acid production process in a metabolically engineered *Corynebacterium glutamicum* strain. *Appl Microbiol Biotechnol*, 2008, 81(3): 459–464.
- [15] Ikeda M, Nakagawa S. The *Corynebacterium glutamicum* genome: features and impacts on biotechnological processes. *Appl Microbiol Biotechnol*, 2003, 62(2/3): 99–109.
- [16] Kalinowski J, Bathe B, Bartels D, et al. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *J Biotechnol*, 2003, 104(1/3): 5–25.
- [17] Kjeldsen KR, Nielsen J. In silico genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnol Bioeng*, 2009, 102(2): 583–597.
- [18] Shinfuku Y, Sorpitiporn N, Sono M, et al. Development and experimental verification of a genome-scale metabolic model for *Corynebacterium glutamicum*. *Microb Cell Fact*, 2009, 8: 43.
- [19] Brinkrolf K, Brune I, Tauch A. The transcriptional regulatory network of the amino acid producer *Corynebacterium glutamicum*. *J Biotechnol*, 2007, 129(2): 191–211.
- [20] Ma HW, Zeng AP. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics*, 2003, 19(11): 1423–1430.
- [21] Baumbach J, Wittkop T, Kleindt CK, et al. Integrated analysis and reconstruction of microbial transcriptional gene regulatory networks using CoryneRegNet. *Nat Protoc*, 2009, 4(6): 992–1005.
- [22] Brinkrolf K, Schröder J, Pühler A, et al. The transcriptional regulatory repertoire of *Corynebacterium glutamicum*: reconstruction of the network controlling pathways involved in lysine and glutamate production. *J Biotechnol*, 2010, 149(3): 173–182.
- [23] Nentwich SS, Brinkrolf K, Gaigalat L, et al. Characterization of the LacI-type transcriptional repressor RbsR controlling ribose transport in *Corynebacterium glutamicum* ATCC 13032. *Microbiology*, 2009, 155(1): 150–164.
- [24] Engels V, Lindner SN, Wendisch VF. The global repressor SugR controls expression of genes of glycolysis and of the L-Lactate dehydrogenase LdhA in *Corynebacterium glutamicum*. *J Bacteriol*, 2008, 190(24): 8033–8044.
- [25] Cline MS, Smoot M, Cerami E, et al. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc*, 2007, 2(10): 2366–2382.
- [26] Batagelj V, Mrvar A. Pajek-analysis and visualization of large networks//Jünger M, Mutzel P. *Graph Drawing Software*. Berlin: Springer, 2003: 77–103.
- [27] Ma HW, Silva MR, Sun JB, et al. Reconstruction and structural analysis of metabolic and regulatory networks//Choi S. *Introduction to Systems Biology*. New Jersey: Humana Press, 2007: 124–146.
- [28] Ma HW, Zeng AP. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, 2003, 19(2): 270–277.
- [29] Kotera M, Okuno Y, Hattori M, et al. Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions. *J Am Chem Soc*, 2004, 126(50): 16487–16498.
- [30] Zhao J, Tao L, Yu H, et al. Bow-tie topological features of metabolic networks and the functional significance. *Chin Sci Bull*, 2007, 52(8): 1036–1045.

- [31] Rodionov DA. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev*, 2007, 107(8): 3467–3497.
- [32] Moreno-Campuzano S, Janga SC, Pérez-Rueda E. Identification and analysis of DNA-binding transcription factors in *Bacillus subtilis* and other Firmicutes—a genomic approach. *BMC Genomics*, 2006, 7: 147.
- [33] Liu Y, Zhao XM, Ma HW. The structure and decomposition of the transcriptional regulatory network of *Bacillus subtilis*. *Chin J Bioprocess Eng*, 2007, 5(3): 58–63.  
刘杨, 赵学明, 马红武. 枯草芽孢杆菌转录调控网络的连接结构及分解. *生物加工过程*, 2007, 5(3): 58–63.
- [34] Pinney JW, Westhead DR. Betweenness-based decomposition methods for social and biological networks. In: Barber S, Baxter PD, Mardia KV, et al. *Interdisciplinary Statistics and Bioinformatics*. Leeds, UK: Leeds University Press, 2006: 87–90.
- [35] Freeman LC. A set of measures of centrality based on betweenness. *Sociometry*, 1977, 40(1): 35–41.
- [36] Brandes U. A faster algorithm for betweenness centrality. *J Math Sociol*, 2001, 25(2): 163–177.
- [37] Holme P, Huss M, Jeong H. Subnetwork hierarchies of biochemical pathways. *Bioinformatics*, 2003, 19(4): 532–538.
- [38] Frunzke J, Gätgens C, Brocker M, et al. Control of heme homeostasis in *Corynebacterium glutamicum* by the two-component system HrrSA. *J Bacteriol*, 2011, 193(5): 1212–1221.
- [39] Brune I, Werner H, Hüser AT, et al. The DtxR protein acting as dual transcriptional regulator directs a global regulatory network involved in iron metabolism of *Corynebacterium glutamicum*. *BMC Genomics*, 2006, 7: 21.
- [40] Rey DA, Pühler A, Kalinowski J. The putative transcriptional repressor McbR, member of the TetR-family, is involved in the regulation of the metabolic network directing the synthesis of sulfur containing amino acids in *Corynebacterium glutamicum*. *J Biotechnol*, 2003, 103(1): 51–65.
- [41] Rey DA, Nentwich SS, Koch DJ, et al. The McbR repressor modulated by the effector substance S-adenosylhomocysteine controls directly the transcription of a regulon involved in sulphur metabolism of *Corynebacterium glutamicum* ATCC 13032. *Mol Microbiol*, 2005, 56(4): 871–887.
- [42] Krömer JO, Heinzle E, Schröder H, et al. Accumulation of homolanthionine and activation of a novel pathway for isoleucine biosynthesis in *Corynebacterium glutamicum* McbR deletion strains. *J Bacteriol*, 2006, 188(2): 609–618.
- [43] Ikeda M. Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering. *Appl Microbiol Biotechnol*, 2006, 69(6): 615–626.
- [44] Wittmann C. Analysis and engineering of metabolic pathway fluxes in *Corynebacterium glutamicum*. *Adv Biochem Eng Biotechnol*, 2010, 120: 21–49.