

文章编号: 1001- 2486(2010) 01- 0107- 05

# 代谢网络结构特征的生物信息学研究\*

王正华, 周婷婷

(国防科技大学 并行与分布处理国家重点实验室, 湖南 长沙 410073)

**摘要:** 作为生物信息学的重要研究内容, 近年来代谢网络结构特征方面的研究取得了许多成果。回顾了代谢网络的重构和表示方法, 重点综述了代谢网络宏观结构特征所蕴含的生物学意义, 以及拓扑特征所揭示的网络演化过程和网络稳健性信息, 提出了目前研究中仍然存在的问题及可能的解决思路, 并对将来的研究趋势进行了阐述。

**关键词:** 生物信息学; 代谢网络; 网络重构; 宏观结构; 拓扑特征; 网络演化; 网络稳健性

**中图分类号:** TP301. 6      **文献标识码:** A

## Bioinformatics for Macro-structural and Topological Properties of Metabolic Networks

WANG Zheng-hua, ZHOU Ting-ting

(National Key Laboratory for Parallel and Distributed Processing, National Univ. of Defense Technology, Changsha 410073, China)

**Abstract:** In the field of bioinformatics, recently more and more researches are launched on the studies of the structure and topology of metabolic networks. We first review the progress for metabolic network reconstruction and representation. Then we look into the macro-structural and topological properties of metabolic networks which are meaningful for uncovering how metabolic networks evolve and why they are robust. Finally we expound our viewpoints on the current issues and the prospective studies in this field.

**Key words:** bioinformatics; metabolic network; network reconstruction; macro-structure; topology; network evolution; network robustness

生命既是遗传信息的表现, 也是化学信息的表现。生物体内众多生化反应组成了庞大而复杂的代谢网络, 构成生物体的基础并使其体现出生命的种种特征。代谢网络的结构特征不仅是网络形成和进化的反映, 也是网络分解和通路分析等相关研究的依据<sup>[1-2]</sup>。深入探究全基因组范围内大规模代谢网络的宏观结构和拓扑特征, 不仅有助于认识代谢网络的形成演化机理、指导生物网络的系统研究, 也是人们认识和利用生命活动规律的必要条件。

生物体的代谢网络主要由复杂的生化反应及参与反应的各种小分子代谢物和酶等构成。随着复杂网络理论的发展和广泛应用, 人们发现代谢网络具有典型的无尺度、小世界和分层模块化特性, 且在宏观结构上呈现出明显的“蝴蝶结”特征<sup>[3]</sup>。而根据其结构特征提出的网络演化模型以及网络稳健性分析, 不仅帮助人们认识生物体内代谢网络的演化和生命的遗传进化过程<sup>[4-6]</sup>, 也直接或间接指导了细胞代谢功能研究<sup>[7]</sup>、生物代谢过程控制<sup>[8]</sup>、突变菌株分析<sup>[9]</sup>、疾病诊断治疗<sup>[10]</sup>以及药物设计靶标选择<sup>[11]</sup>等各项研究的开展。

### 1 全基因组范围内代谢网络的重构和表示

代谢网络的重构和表示是分析其结构特征和演化规律的基础。一般根据数据库或文献中获得的反应物、产物、相应的酶及反应可逆性等信息实现其重构, 并抽象为图的形式。

对于新测序的物种, 主要根据其已知的基因组功能注释信息, 结合相关的酶-反应关系数据库信

\* 收稿日期: 2009- 07- 18

基金项目: 国家自然科学基金资助项目(60603054, 60773021)

作者简介: 王正华(1962—), 男, 教授, 博士。

息,确定其代谢网络中包含的所有代谢反应。首先从基因组序列中识别出所有可读框(ORF),然后通过相似性比对预测该物种的所有酶基因。通过查询酶命名数据库确定相应的酶号(EC Number),然后通过查询酶-反应关系数据库(如KEGG<sup>[12]</sup>等)根据酶号确定其催化的代谢反应,将酶和反应关联起来。

另一种较为常用的网络重构方法是先从数据库中下载得到基因组注释信息以及原始的代谢网络图,再人工校对其中每个基因的功能注释是否正确和全面,并根据数据库、文献及实验结果对错误和缺失的基因注释进行修正,即通过增补和修正的方法重构代谢网络<sup>[13]</sup>。

然而无论采用哪种方法,酶号信息缺失和功能注释错误都会影响网络重构。这一问题目前并无很好的解决方法,大多通过人工校对解决,费时费力。直到2006年,Notebaart等<sup>[14]</sup>提出一种半自动方法(AUTOGRAPH-method)来加速网络重构,其思路是充分利用已存在且经过人工修正的基因组信息,通过序列同源性分析来减少人工劳动量。

研究目的不同,代谢网络的表示方法也可不同,一般从以下4种图的形式中选择<sup>[3]</sup>:代谢物图,反应图,酶图和代谢物-酶二部图。4种表示法最大的区别在于顶点和边的定义不同。代谢物图以参与反应的代谢物为顶点,若两代谢物间存在反应直接关联,则相应两顶点间有边相连;反应图以生化反应为顶点,若一反应的生成物是另一反应的反应物,则对应两顶点间有边相连;酶图以酶为顶点,若酶A的生成物中包含有酶B的反应物,则对应顶点连线;而代谢物-酶图的顶点有两类,一类对应于参加反应的代谢物,另一类对应于催化反应的酶,仅在不同类顶点间有边相连。

## 2 代谢网络的宏观结构及生物学意义

基于基因组数据的代谢网络重建和表示使得针对大规模代谢网络的宏观结构研究和拓扑特征提取成为可能。从整体上把握代谢网络的宏观结构及其相应的生物学意义对于相关研究具有很强的指导作用。

Ma等<sup>[15]</sup>的研究结果表明,代谢网络在宏观上呈现出明显的蝴蝶结(Bow-tie)特征,且“蝴蝶结”的主体——GSC部分顶点连接最复杂,包含了糖酵解、三羧酸循环、氨基酸生物合成、核酸生物合成等重要代谢通路,充分说明GSC是代谢网络中最重要的功能构成。Zhao等<sup>[16]</sup>将蝴蝶结结构进一步细化为展开的蝴蝶结模型,如图1所示,图中节点粒度大小定性表示相应强联通体中顶点的多少,边的粗细定性表示两个强联通体间连接边数的多少,中间最大的菱形对应GSC部分。细化后的各部分具有更为合理的生物学意义,也进一步说明了蝴蝶结结构是代谢网络固有的内在特征。

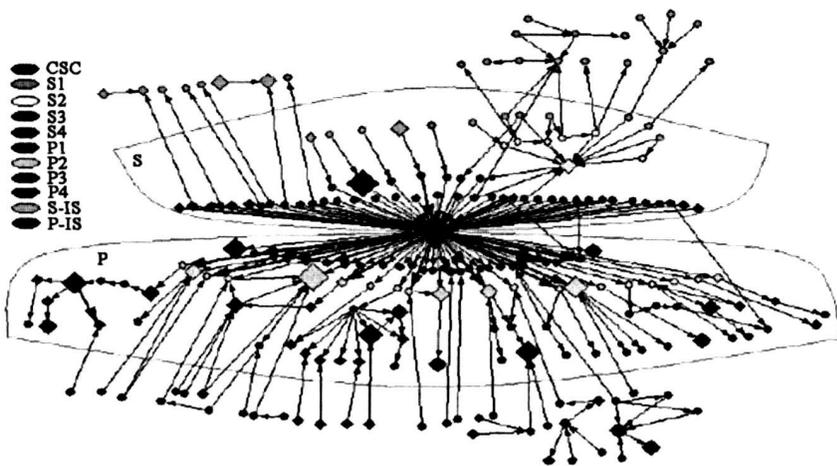


图1 *E. coli*代谢网络展开蝴蝶结模型的粗粒化示意图

Fig. 1 Spread bow-tie structure of the coarse grained for *E. coli* metabolic network

另外,代谢网络的蝴蝶结结构还呈现出一种分层嵌套(Hierarchical Nested)的特性<sup>[3]</sup>,即几乎所有的子网都与整个网络一样具有高度模块化的蝴蝶结拓扑模式。子网中小的蝴蝶结结构分层嵌套在上一层

子网稍大的蝴蝶结结构中,最后整合成完整的代谢网络。这种蝴蝶结结构也同样体现在特定化学功能的孤立模块中,而在随机混洗网络中却没有发现,说明该结构是进化的必然结果。

从生化观点来看,一小部分活性载体和前导物质是代谢系统最为重要和保守的部分,形成了蝴蝶结的“结”,营养素作为生化合成的输入而大面积“扇入”,产物则大面积“扇出”。GSC 涉及了中间代谢三条关键通路(糖酵解、TCA 循环和戊糖磷酸化)中的绝大多数碳水化合物反应,包含大部分生物合成常用的前体物质,说明作为生命本质和内在的过程,碳水化合物代谢,尤其是糖酵解、TCA 循环和戊糖磷酸化代谢通路的主要反应受到了很好的保护,整个网络具有很高的抗随机攻击性。另外,由于大部分代谢回路都存在于 GSC 中,GSC 的密集连接对于整个网络是一种保护机制,也是生物体调节自身代谢的核心所在。

### 3 代谢网络的演化机制及稳健性分析

代谢网络属于复杂网络范畴,具有典型的无尺度(Scale-free)、层次模块化(Hierarchical Modular)和小世界(Small-world)特征。这些特征在一定程度上反映网络的形成过程并体现其稳健性,即生物体代谢系统的抗攻击性和容错能力。

#### 3.1 无尺度特征

细胞的代谢网络具有无尺度特征,其形成和演化符合无尺度网络演化的 BA 模型,即不断有新节点出现(Growth)和新节点倾向于与连接度高的节点连接(Preferential Attachment)。此模型假设所有顶点的度增加遵循幂函数规律的时间相关性,表明顶点的度越高,该顶点加入网络的时间越早,进化速度也就越慢<sup>[4]</sup>。即如果代谢网络的演化和扩增是由于新物质的出现,则连接度最高的节点所对应的代谢物被认为是系统发生学意义上最古老的,GSC 中的三条关键通路也许是最早出现的代谢途径。由此可见,节点的连接度越高,对于代谢网络的演化越关键,其相应代谢物或酶对于代谢系统也越关键。

另外,度分布遵从幂律,使得一小部分高连接度的节点控制全局的网络连接。以代谢物图为例,说明代谢反应中的多数代谢物只参与极少数的反应,而少数几个代谢物则参与多数反应并发挥着代谢中枢的作用<sup>[17]</sup>。若是随机去除某些顶点,网络直径的变化并不明显,连通性也几乎不受影响;说明对于突变或其他原因造成的随机错误,代谢网络呈现出出人意料的稳健性。然而依次移除连接度最大的节点会使网络直径迅速增大,最后导致网络被分割成不再具有功能特性的独立簇,说明对于某些高连接度代谢物、酶或反应的“蓄意攻击”将造成整个代谢系统的瘫痪。而这也正是某些代谢类疾病潜在的致病原因和治疗机理之一。

#### 3.2 层次模块化特征

代谢网络属于复杂网络范畴,层次模块化是其固有特征。节点的聚集系数服从尺度法则,节点连接度越高,聚集系数越小<sup>[18]</sup>。根据尺度法则可以由连接度的大小将节点划分为不同层次,量化了不同模块中不同连接度节点的分层共存性。

从网络演化的角度来解释,代谢网络具有许多高度整合的小模块,这些小模块成组构成了大一些的模块,依次整合,形成越来越大的模块。从生物进化的角度讲,分层模块化的特征与进化同时作用于不同组织水平的观点相一致。另外,许多影响小规模高聚集度模块的局部改变,能够缓慢影响大规模低聚集度模块的性质。通过拷贝和再利用已有模块而形成的分级拓扑结构保留了基因复制的结果,因此进化过程中越早出现的模块作用越关键。

代谢网络的分层模块化结构特征避免了网络内不同层次不同功能模块之间不必要的相互影响,防止有害扰动扩散到整个网络,因此对于保持整个网络的稳健性具有积极的作用。然而,Huss 等<sup>[17]</sup>对模块性与鲁棒性相联系的说法提出了质疑,认为代谢网络有其特殊性,即有些致命的模块如氨基酸代谢或核苷酸代谢的整个模块遭到破坏,即使剩下的模块完好无损,生物体也无法幸存。与孤立考虑单个模块相比,整合考虑整个网络将更有意义,因为一个模块中的代谢遭到破坏或许可以通过其他模块进行合成。因此即使细胞的通常活动能够从功能模块方面理解,但若要考虑某些特殊或异常情况下网络的稳

健性,就需要从整个网络的全局角度来考虑。

### 3.3 小世界特征

代谢网络具有典型的小世界特性<sup>[19]</sup>,表示大多数代谢物或酶之间只需极少几步就可完成相互转化,且代谢物的浓度变化在很短的时间内就会传遍整个网络。

根据1998年Watts等<sup>[20]</sup>提出的WS模型,小世界网络是在规则网络的基础上随机删除某些边和节点而得到的。从某种意义上讲,小世界网络的演化过程一定程度地模拟了代谢网络的进化,随机断开连接边和删除顶点的操作相当于功能模块中某些代谢物/酶及反应的缺失。如Wagner等<sup>[21]</sup>发现多种化合物的催化酶(如赖氨酸或异戊二磷酸等)在不同生物体中的作用路线也不相同,且相当多的变异发生在三羧酸循环中存在或缺少某些特定反应的条件下,说明小世界特性允许代谢网络设计具有一定的弹性。

代谢网络需要应对各种频繁扰动,如酶或代谢物的浓度扰动,电中性及热力学参数的改变等。20世纪80年代,研究者就已经意识到了寻找途径减小代谢物之间转化时间这一问题的重要性并加以讨论,且得出了非常有意义的结论——小世界网络中微扰的传播时间接近于任何相同节点数网络的理论最小值。这意味着,代谢物或酶的浓度变化能很快传遍整个网络,使得细胞能在最短时间内对外界变化做出反应,从而增强了整个网络的容错性和抗攻击能力<sup>[21]</sup>。

## 4 存在的问题

在过去的几年里,代谢网络结构特征的分析取得了许多重要成果,直接和间接地指导着细胞代谢功能的研究、代谢过程的控制和设计。但不可否认,目前研究中仍存在许多问题亟待寻找更为合理的解决方案,如流通代谢物的判定问题以及动力学约束问题等。

### 4.1 流通代谢物(Current Metabolite)的判定

这里所说的流通代谢物一般指细胞内的ATP、ADP、NADH及NAD<sup>+</sup>等电子转移或某些功能基团(磷酸基、氨基、一碳单位、甲基等)转移的携带者,参与大量代谢反应,在网络中表现为连接度极高的节点。为了在代谢网络中确切显示细胞中主要化合物的转化及相应的生化反应步骤,这些流通代谢物及一些小分子化合物(如H<sub>2</sub>O、NH<sub>3</sub>、O<sub>2</sub>、CO<sub>2</sub>等)不能作为反应物之间的有效中间物,否则将会得出从葡萄糖到丙酮酸只需要两步反应的结论<sup>[22]</sup>。Huss等<sup>[17]</sup>根据Newman等提出的网络模块性Q函数,将流通代谢物定义为被删除后能使网络模块性增强的代谢物,并将这种方法扩展成一种较为有效的网络分解方案。这种基于数学的判别方法虽然大大降低了工作量,但难以从生物学角度解释其合理性。因此,寻找一种既兼顾代谢物生化意义又易于实现的合理的自动判别方法,便成为代谢网络研究亟待解决的问题之一。

### 4.2 大规模网络研究缺少动力学约束

生化反应中的膜渗透压、电中性和热力学参数及代谢物和酶的量等这些参数多作为化学计量参数用于通路分析,为了简化计算而很少在大规模代谢网络的表示和研究中出现。然而,无论将代谢网络表现为何种图的形式,都会因抽象而丧失大量生化信息。若要全面理解网络的结构和性质,探究其结构功能的生化意义,就需要将这些动力学参数以合理的形式整合到网络中去。而就目前的研究成果来看,网络构成与分析缺少动力学参数等约束条件这一问题还未得到较好的解决。

## 5 结论与展望

众多研究成果表明,代谢网络是一种宏观上具有明显的分层“蝴蝶结”结构的复杂生物学网络,同时具有无尺度、小世界和分层模块化的拓扑特征。这些特征从不同角度反映了代谢网络的形成和演化机理。且由于这些特征的存在,大规模代谢网络同时具有对于随机错误的容错性和对于“蓄意攻击”的脆弱性。

然而,包括网络结构特征在内的整个代谢网络的系统生物学研究还仅仅处于起步阶段。虽然目前对代谢网络的拓扑结构已有一定程度的了解,但对为何形成这样的结构,即对进化和自然选择在细胞网络拓扑结构的形成中所起的作用还不十分清楚;另外,由于目前在细胞代谢动力学参数方面积累的信息

比较少,代谢网络的研究在一定程度还停留在网络静态特征的研究上。因此,关于网络结构的功能作用、进化机制和动力学分析等一系列问题还有待进一步探讨。

## 参考文献:

- [1] 王正华,周婷婷.大规模代谢网络分解的生物信息学研究[J].国防科技大学学报,2008,30(3):48-53.
- [2] Takemoto K, Oosawa C. Modeling for Evolving Biological Networks with Scale-free Connectivity, Hierarchical Modularity, and Dissortativity [J]. *Mathematical Biosciences*, 2007, 208(2): 454-468.
- [3] Zhao J, et al. Hierarchical Modularity of Nested Bow-ties in Metabolic Networks [J]. *BMC Bioinformatics*, 2006(7): 386.
- [4] Vitkup D, Kharchenko P, Wager A. Influence of Metabolic Network Structure and Function on Enzyme Evolution [J]. *Genome Biology*, 2006, 7(R39).
- [5] Zhou T, et al. An Approach for Determining Evolutionary Distance in Network-based Phylogenetic Analysis [C]//4<sup>th</sup> International Symposium on Bioinformatics Research and Applications, Atlanta, GA, USA: ISBRA 2008, Springer, LNCS(LNBI), 2008, 4983: 45-56.
- [6] Zhou T, Chan C, Wang Z. TopEVM: Using Co-occurrence and Topology Patterns of Enzymes in Metabolic Networks to Construct Phylogenetic Trees [C]//Third IAPR International Conference on Pattern Recognition in Bioinformatics, Melbourne, Australia: PRIB 2008, Springer LNCS(LNBI), 2008, 5265: 225-236.
- [7] Steuer R, et al. Structural Kinetic Modeling of Metabolic Networks [C]//Proceedings of the National Academy of Sciences, 2006, 103(32): 11868.
- [8] Nagrath D, et al. Integrated Energy and Flux Balance Based Multiobjective Framework for Large-scale Metabolic Networks [J]. *Annals of Biomedical Engineering*, 2007, 35(6): 863-885.
- [9] Wunderlich Z, Mimy L A. Using the Topology of Metabolic Networks to Predict Viability of Mutant Strains [J]. *Biophysical Journal*, 2006, 91(6): 2304.
- [10] Asanuma K, et al. Network Modulation in the Treatment of Parkinson's Disease [J]. *Brain*, 2006, 129(10): 2667.
- [11] Kell D B. Systems Biology, Metabolic Modelling and Metabolomics in Drug Discovery and Development [J]. *Drug Discovery Today*, 2006, 11(23-24): 1085-1092.
- [12] Kanehisa M, et al. From Genomics to Chemical Genomics: New Developments in KEGG [J]. *Nucleic Acids Research*, 2006, 34(1): 354.
- [13] Duarte N C, et al. Global Reconstruction of the Human Metabolic Network Based on Genomic and Bibliomic data [C]//Proceedings of the National Academy of Sciences, 2007, 104(6): 1777.
- [14] Notelaert R A, et al. Accelerating the Reconstruction of Genome-scale Metabolic Networks [J]. *BMC Bioinformatics*, 2006, 7(1): 296.
- [15] Ma H W, Zeng A P. The Connectivity Structure, Giant Strong Component and Centrality of Metabolic Networks [J]. *Bioinformatics*, 2003, 19(11): 1423-1430.
- [16] Zhao J, et al. Bow-tie Topological Features of Metabolic Networks and the Functional Significance [J]. *Chinese Science Bulletin*, 2007, 52(8): 1036-1045.
- [17] Huss M, Holme P. Currency and Commodity Metabolites: Their Identification and Relation to the Modularity of Metabolic Networks [J]. *Systems Biology*, IET, 2007, 1(5): 280-285.
- [18] Ravasz E, et al. Hierarchical Organization of Modularity in Metabolic Networks [J]. *Science*, 2002, 297(5586): 1551-1555.
- [19] Jeong H, et al. The Large-scale Organization of Metabolic Networks [J]. *Nature*, 2000, 407(6804): 651-654.
- [20] Watts D J, Strogatz S H. Collective Dynamics of Small-world Networks [J]. *Nature*, 1998, 393(6684): 440-442.
- [21] Wagner A. The Small World inside Large Metabolic Networks [C]//Proceedings of the Royal Society B, Biological Sciences, 2001, 268(1478): 1803-1810.
- [22] Ma H W, Zeng A P. Reconstruction of Metabolic Networks from Genome Data and Analysis of Their Global Structure for Various Organisms [J]. *Bioinformatics*, 2003, 19(2): 270-277.