

应用复杂网络理论研究代谢网络的进展

赵 静 俞 鸿 骆建华 曹志伟* 李亦学*

(上海交通大学生命科学技术学院, 上海 200240; 上海生物信息技术研究中心, 上海 200235; 重庆后勤工程学院数学教研室, 重庆 400016. 中国科学院上海生命科学研究院, 上海 200031, * 联系人, E-mail: zwcao@scbit.org, yxli@sibs.ac.cn)

摘要 后基因组生物信息学研究的一个重要任务是系统地研究活细胞内所有分子和它们之间的相互作用, 从而了解这些分子及它们之间的相互作用对整个生物体功能的影响. 而网络则是对各种相互作用关系的恰当的抽象描述. 近年来, 复杂网络理论在揭示各种复杂的技术网络和社会网络的形成和演化机制方面取得了一些重要成果, 其方法和结果已对生物学研究产生影响. 本文评述了基于基因组的大规模代谢网络重建和分析的进展, 论述了利用复杂网络理论分析代谢网络结构的主要方法和结果.

关键词 生物信息学 系统生物学 代谢网络 网络结构 网络分解 网络稳健性

人类基因组计划的基本完成表明后基因组时代的到来. 生物学研究的热点由对细胞内个别基因或蛋白质功能的局部性研究, 转移到以细胞内全部的基因, mRNA, 蛋白质及代谢产物为研究对象的各种“组学”研究, 即整体性研究. 基因组学、转录组学、蛋白质组学、代谢组学等各种组学技术, 逐步把分子生物学推入系统生物学时代, 同时也把生物信息学带入了后基因组信息学时代. 由于基因与蛋白质倾向于成组地通过网状的相互作用而影响生物系统的功能, 因此对功能的研究必须分析其相互作用的网络. 后基因组信息学是通过对多种分子和基因相互作用网络(基因调控网络、信号转导网络、蛋白相互作用网络、代谢网络等)的研究来进行生物功能的分析, 目标是理解生物系统如何在单个构造模块的基础上组织起来, 是对包括从基因组信息到对生命基本规律理解的一系列生物学知识的综合, 同时它也有在生物化学领域应用的实际目的^[1].

代谢处于生命活动调控的末端, 是驱动生命过程的化学引擎, 产生能量来驱动各种细胞过程, 降解和合成许多不同的分子. 代谢网络把细胞内所有生化反应表示为一个网络, 反映了所有参与代谢过程的化合物之间以及所有催化酶之间的相互作用, 是对细胞代谢的抽象表达. 研究代谢网络能帮助我们更好地认识和利用细胞代谢过程, 从而促进发酵工程、制药工业等产业的发展. 另一方面, 网络的拓扑结构是网络形成和进化的反映, 研究代谢网络的结构特征, 能帮助我们认识代谢网络的形成演化机理, 从而更好地理解生命进化过程.

现实世界的大量复杂系统可以通过网络来描述.

在相当长的时间, 人们将真实复杂网络的拓扑结构看成是完全随机的, 用随机网络模型来描述它们^[2]. 然而, 近年来, 随着计算机技术的发展, 对一些大规模的复杂系统进行数据采集和统计分析变得可行, 人们发现真实网络的很多宏观性质不能被随机网络模型所刻画. 1998 年 Watts 和 Strogatz^[3]发现真实网络的小世界(small-world)性质及 1999 年 Barabasi 和 Albert^[4]发现真实网络的无标度(scale-free)性质, 在全世界范围内引发了复杂网络研究的浪潮, 标志着现代意义下复杂网络研究的开端. 复杂网络理论以社会网络(如疾病传播网络、科研论文引用网络、科学家合作网络等)、技术网络(如万维网等)、生物网络(如食物网、代谢网络、蛋白相互作用网络等)等真实网络为研究对象, 通过图论、统计学、统计物理、计算机模拟等方法, 研究网络的结构特征、结构与功能的关系、网络的生成机制及网络演化规律等一系列的问题. 复杂网络理论的发展, 为在系统水平上研究生物网络提供了新的平台. 自 2000 年 Jeong 等人^[5]在 *Nature* 上发表利用复杂网络理论研究代谢网络的拓扑特性的论文以来, 许多学者在代谢网络的结构与功能方面进行了大量的研究, 取得了大量有意义的结果, 代谢网络研究得到较大发展.

1 基于基因组的代谢网络重建

通过基因组数据重建针对特定物种的代谢网络, 是对代谢网络进行结构和功能分析的基础. 基于某一物种的基因组注释信息, 结合相关的酶反应数据库, 就可以确定包含在代谢网络中所有的代谢反应. 其主要步骤如下^[6]:

() 从基因组序列中识别所有可读框(ORF).

() 通过序列相似性比对, 预测该物种的所有酶基因。

() 查询酶命名数据库(Enzyme nomenclature database), 如 ENZYME data bank^[7] (<http://expasy.hcuge.ch/>), 将酶基因对应上酶号(EC number)。

() 查询酶与反应关系的数据库, 如KEGG的LIGAND/ENZYME^[6] (<ftp://ftp.genome.jp/pub/kegg/ligand/>), 将酶号对应上它所催化的代谢反应。

上述方法特别适用于重建新测序物种的代谢网络。在因特网上也存在着一些专门的代谢反应数据库, 包含了与代谢网络中的酶及其在不同物种中的编码基因相关的信息, 其中最著名的数据库有KEGG^[8,9] (<http://www.genome.ad.jp/kegg>), EcoCyc^[10,11] (<http://ecocyc.org>), WIT^[12] (<http://wit.mcs.anl.gov/WIT/>)等。由这些数据库可以方便地查到某一物种中所有的代谢反应。例如, 可以从KEGG下载催化某一物种代谢反应的所有酶, 以及所有物种的全部生化反应及其催化酶, 由这两个数据库就可以很容易地得出该物种中所有的代谢反应。

确定了所研究物种的所有代谢反应后, 就可以根据所研究问题的需要, 将代谢网络表示成不同类型的图。图是图论研究的对象, 它表示元素与元素之间的二元关系, 其中元素表示为图的顶点, 元素之间的关系表示为顶点之间的连线。按照图中顶点的不同含义, 代谢网络可表示为以下4种类型的图:

第一种, 化合物图(substrate graph)。每个顶点对应一种化合物, 若化合物A能通过反应生成化合物B, 则在这两个化合物对应的顶点连线^[13~18]。

第二种, 反应图(reaction graph)。每个顶点对应一个反应, 若反应A的一个产物是反应B的一个底物, 则在这两个反应对应的顶点连线^[14,19]。

第三种, 酶图(enzyme-centric graph)。每个顶点对应一个酶, 若酶A的一个产物是酶B的一个底物, 则在这两个反应对应的顶点连线^[20]。

第四种, 化合物-反应二部图(substrate-enzyme bipartite graph)。二部图的顶点有两类, 仅在不同类的顶点间有连线而同类的顶点间无连线。底物-反应二部图的一类顶点对应化合物, 另一类顶点对应反应, 在反应与它的底物及产物间连线^[5,21,22]。

另一方面, 按照图中的边是否有方向, 可以把图分为有向图和无向图, 有向图中的边称为弧。因为生化反应是有方向的, 即有一些反应是不可逆的, 在大多数情况下, 代谢网络被表示为有向图^[5,13,16~19,23]。

然而, 在研究代谢网络的某些结构特征时, 可以忽略反应的方向而将网络表示为无向图^[14,15,21]。

值得注意的是, 在细胞内, ATP, ADP, NADH及NAD⁺等通用代谢物(current metabolites)一般是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, 参与了大量的代谢反应。为了显示细胞中主要化合物的转化以及在代谢网络中确切地显示生化反应的步骤, 这些通用代谢物及一些小分子化合物(如H₂O, NH₃, O₂, CO₂等)不能作为反应之间的有效中间物, 否则将得出从葡萄糖到丙酮酸只需要两步反应的错误结论^[16]。基于此, Ma和Zeng^[16]在KEGG代谢反应数据库的基础上进行修正补充后得到一个新的数据库, 去掉了每个反应中的通用代谢物及小分子化合物, 并明确给出了每个反应的可逆性信息。根据这个数据库, 可以很容易地将代谢网络表示为一个有向图。

2 代谢网络的宏观结构

Ma和Zeng^[17]分析了他们所建的数据库中65个物种的代谢网络的连通特性, 发现代谢网络从宏观上与WWW网的结构^[24]类似, 呈现一种蝴蝶结结构(bow-tie)。网络的顶点集划分为4个不相交的子集: 庞大强连通体(giant strong component, GSC), 底物子集(substrate subset, S), 产物子集(product subset, P)和孤立子集(isolated subset, IS)。其中GSC是网络中最大的强连通体(所谓强连通体, 是指有向图的一部分子图, 其中任意两个顶点可通过有向路径彼此到达), 它比网络中其他强连通体大很多, 顶点约占整个网络的三分之一, 是网络中顶点间连接最紧密和最复杂的部分。GSC中任两种代谢物可以通过一系列反应彼此生成; 底物子集中的代谢物均可以转化成为GSC中的代谢物但反之不能; 产物子集中代谢物均可以由GSC中的代谢物经一系列反应转化得到; 而孤立子集中的代谢物既不能转化成GSC中的代谢物, 也不能由GSC中的代谢物转化得到。图1显示的是本研究用图论中压缩图的方法, 将网络中每个强连通体压缩成一个点(图中的菱形结点)而得到的*E. coli*代谢网络的一个粗粒化的框架图, 图中较大的菱形对应于顶点个数较多的强连通体, 较粗的弧代表对应的两个强连通体间有较多的弧, 中间最大的一个菱形对应GSC部分。此图清晰地显示出了代谢网络的蝴蝶结结构, 尤其显示出在GSC外的拓扑特征, 即代谢网络的GSC外的部分主要以一种分枝的树状

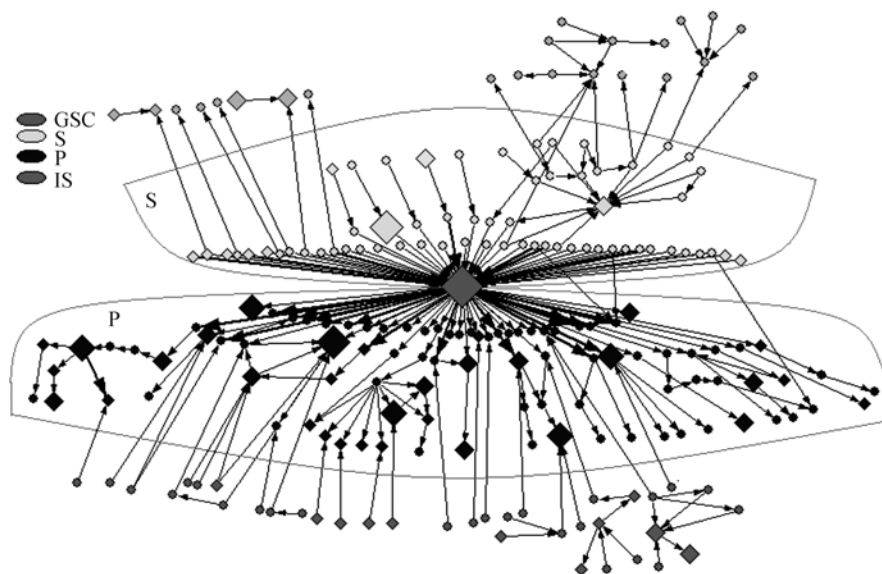


图1 *E. coli* 代谢网络蝴蝶结结构的粗粒化框架图

方式与 GSC 连接.

通过GSC部分的代谢物参与的反应,发现细胞最重要的代谢通路如糖酵解、三羧酸循环、氨基酸生物合成、核酸生物合成等,都包含在GSC部分,这充分说明GSC是代谢网络中最重要的部分. Ma等人^[19]还发现,不但代谢网络的化合物图具有蝴蝶结结构,其反应图也具有这种结构.

3 网络参数及其反映的网络拓扑特征

每个网络都存在其特有的拓扑特征,其中某些特征对网络功能有很大影响. 恰当定义网络参数,可以定量地描述网络在某方面的拓扑性质,从而对复杂网络进行分析和比较.

3.1 度分布与无标度网络

无向图中,与某顶点相连的边的数量称为该顶点的度. 有向图中,顶点的度是其出度和入度之和,其中出度指起始于该顶点的弧数,入度为指向该顶点的弧数. 网络的度分布 $P(k)$, 指网络中度数为 k 的顶点的出现频率, $k = 1, 2, \dots$.

随机图的顶点之间的连接是随机的,大多数顶点的度数都接近于图的平均顶点度,其顶点度分布服从泊松分布^[2]. 然而,通过对多种现实存在的复杂网络,如万维网、Internet及社会网络等的度分布统计,发现这些网络的度分布服从幂率^[4,24,25],即网络中大多数顶点的度很低,只有少数一些顶点(称为集散点

或Hub点)有较高的度,说明现实中的网络并非随机连接的. 度分布服从幂率的网络称为无标度网络^[4]. 随机网络与无标度网络的度分布比较见图2. Jeong等人^[3]统计了分别属于古细菌、细菌、真菌的43个物种的代谢网络的度分布,发现这些网络的出度和入度分布都服从幂率: $p(k) \sim k^{-r}$, 幂指数 r 约为2.2,说明代谢网络属于无标度网络.

顶点的度分布是与网络大小无关的参数,因此它是网络的某种内在特性的表现,可作为区分不同类型网络的指标.

Barabasi和Albert^[4]根据无标度网络在现实中的普遍存在及其结构特点,提出了无标度网络演化的BA模型. 该模型认为真实网络的无标度特性来自于它们内在的两个机制: () 网络规模的扩增: 不断有新结点加入网络; () 偏好连接: 新结点倾向于已有结点中度数较高的连接,即“富者越富”原则. 按照此模型,网络中连接度高的结点是在网络形成过程中出现较早的结点. 类似地,如果代谢网络的演化是由于新的催化酶的出现,网络中添加新的代谢物而扩大规模,那么集散点对应的代谢物应是系统发生学意义上最古老的代谢物. 糖酵解和三羧酸循环被认为也许是最早出现的代谢途径. Wagner和Fell^[14]列出了*E. coli*代谢网络中13个连接度最高的代谢物,其中包括糖酵解和三羧酸循环中间物如乙酰辅酶A、丙酮酸、3-磷酸甘油酸、琥珀酸、2-酮戊二酸等,以及

一些早期氨基酸如谷氨酸、谷氨酰胺、天冬氨酸、丝氨酸等。这说明,对代谢网络纯粹的拓扑结构分析,的确能在一定程度上揭示网络的演化历史和设计原理。

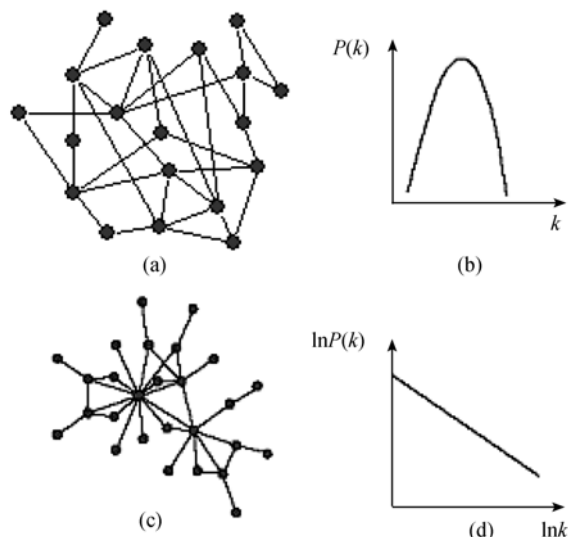


图2 随机网络与无标度网络的度分布

(a) 随机网络; (b) 随机网络的度分布图; (c) 无标度网络; (d) 无尺度网络度分布的双对数图

3.2 集聚系数(clustering coefficient)与层次模块化网络

在一个无向图中,顶点 v 的集聚系数描述了其邻接点之间的连接程度,即网络的局部集团化程度,其数学定义为:

$$CC(v) = \frac{2|N(v)|}{d(v)(d(v)-1)}.$$

其中 $d(v)$ 是顶点 v 的度, $|N(v)|$ 代表顶点 v 的 $d(v)$ 个邻接点之间的直接连线的数目^[3]. $CC(v)$ 的值介于0与1之间. 由此定义可知,若 v 的任两个邻居间都有连线, $CC(v)$ 等于1,若 v 的所有邻居间都无连线, $CC(v)$ 等于0.

Ravasz等人^[15,26]用 $C(k)$ 表示网络中所有度数为 k 的顶点的平均集聚系数,统计了43个物种的代谢网络中 $C(k)$ 的分布规律,发现 $C(k)$ 的分布服从幂率: $C(k) \sim k^{-1}$,与层次模块化网络中 $C(k)$ 的分布相同. 与网络度分布类似,这里的 $C(k)$ 分布与网络大小无关,因此可用来区分不同类型的网络. 由此推断代谢网络属于层次模块化网络,在其中小的模块之间相互连接、嵌套,组成较大的模块,然后组成更大的模块,

形成一种层次结构. Ravasz等人^[15]还提出一个算法将代谢网络分解成小的模块化子网络,发现同一模块内的大多数反应属于同一代谢通路,说明子模块具有相对独立的生物学功能,是生物功能模块. Ravasz等人的研究从代谢网络的拓扑结构上证实了,代谢网络和其他生物学系统一样^[27,28],呈现一种模块化的结构.

3.3 平均路长与小世界网络

网络中两顶点间的距离用最短路来度量,即从一个顶点到另一顶点所需要通过的最少边的数目(在有向图中为最少有向弧的数目). 网络直径指网络的所有顶点间最短路的最大值,网络的平均路长指网络中所有顶点间最短路的平均值,这两个参数是整个网络的顶点间信息交流速度的度量.

复杂网络的另一个重要特征就是“小世界特性”,即著名的“六度分离原理”,意指虽然复杂网络的规模很大,网络的平均路长却很小,而且平均集聚系数较高^[3]. Jeong等人^[5]统计了43个不同物种代谢网络的平均路长,发现无论是古细菌、原细菌还是真核生物,其代谢网络的平均路长都约为3.2,意味着任意两种代谢物之间平均只需要三步代谢反应就能彼此转化,从而得出代谢网络是小世界网络的结论. 但是,Jeong等人构建的代谢网络中包含了ATP及NADH等通用代谢物,由于这些在生化反应中到处出现的代谢物在网络中作为中间产物,缩短了大多数点对间的距离,因而得到了很小的平均路长,但这是与实际的生化反应不符的. Mat和Zeng^[16]去除了反应中的通用代谢物和一些小分子化合物,重建了80个物种的代谢网络,经统计发现真核生物、古细菌和原细菌代谢网络的平均路长分别是9.57, 8.50和7.22. 这些平均路长虽然远大于Jeong等人的网络的平均路长,但与代谢网络的规模相比还是很小,因此代谢网络的确属于小世界网络. 代谢网络的小世界特性意味着,代谢物浓度或酶的浓度的变化能很快地传递到整个网络,使得细胞能对外界环境的变化很快作出反应^[14].

4 代谢网络的分解

在代谢网络调控方面已建立了一些有效的代谢通路分析方法,如基元模式分析(elementary flux modes, EFMs)和极端途径分析(extreme pathways, EPs)^[13,29,30]. 然而,将这些方法应用到大规模的全基

因组代谢网络时,会出现组合爆炸问题.因此,在利用这些方法进行通路分析前,必须将网络分解成功能相对独立的子网络.一般有基于生物学意义或基于网络的拓扑结构两种分解方法.如生物学家按照反应属于不同的代谢通路:糖代谢、氨基酸代谢、脂代谢等,将代谢网络分解为子网络. Schilling与Palsson^[31]提出了按照直观的生物学标准将代谢网络分解为相对独立的反应簇的算法.然而,由于网络结构是功能的反应,而且已从拓扑结构上证实代谢网络是子模块按层次化的方式组织起来的^[15],更多的算法倾向于完全从网络拓扑出发的分解.即用算法将代谢网络分解成子模块,使得各个模块具有一定的结构独立性,模块内部节点连接紧密,而模块之间连接稀疏.基于拓扑结构的分解,能帮助我们更好地理解复杂生物系统的组织原理.

Schuster等人^[32]依据顶点的度对代谢网络进行分解.去掉度数大于某个设定的度阈值的高连结度顶点,由于代谢网络的无标度特性,网络被分割成一些彼此不连通的子网络. Schuster等人用此算法分解了肺炎支原体(*Mycoplasma pneumoniae*)的全基因组代谢网络,并证实了用此方法能得到一些具有相对独立生物学功能的模块,如精氨酸降解、四氢叶酸系统、核酸代谢等.

Guimera等人^[18]将网络分解看成一个优化问题,要求划分出的模块内部结点的连接尽可能紧密,而模块之间连接尽可能稀疏,并根据此要求建立了一个模块化参数modularity.通过模拟退火法解此优化问题,使modularity取得最大值,得到网络的最优分解,而且所得到的子模块都具有相对独立的生物学功能.

上述两种方法仅着眼于网络的分解,但不能体现模块之间的层次化连接,属于非层次式聚类方法(nonhierarchical clustering method).另一类分解算法使用层次式聚类方法(hierarchical clustering method),以树状构架呈现网络的模块结构,能体现网络的层次化模块组成方式.这种方法首先定义任意两个结点之间的相似指数(similarity index)或相异指数(dissimilarity index),用来量化表示两个结点位于同一子网络的可能性大小,然后用凝聚法(agglomerative clustering method)或分裂法(divisive method)对网络结点进行聚类.其中凝聚法在初始时先将每个点看作一个子集,依据结点间的相似程度作为合并条

件,每次将两个相似程度最高(即相似指数最大,或相异指数最小)的子集进行合并,直到所有结点合并成一个类为止.而分裂法在初始时把所有点视为一个集合,每次在相异度最高的地方分割网络,直到所有结点属于不同子集为止.不同算法的差别主要在于相似指数的不同.下面介绍代谢网络分解的几个层次式聚类算法.

Ravasz等人^[15]忽略生化反应的方向,将代谢网络表示成无向图,并通过“生化简化”和“拓扑简化”步骤,将代谢网络表示为一个更紧凑的形式,然后对此紧凑化的代谢网络进行分解.定义顶点*i*, *j*之间的相似指数为它们间的拓扑重叠度(topological overlap):

$$O_T(i, j) = \frac{J_n(i, j)}{\min(k_i, k_j)}.$$

其中 $J_n(i, j)$ 为顶点*i*, *j*的共同邻居点的个数(若*i*, *j*之间有直接连线,此数值再加1); k_i 和 k_j 分别为顶点*i*, *j*的度数.用层次凝聚法分解网络.

Ma等人^[19]根据代谢网络的蝴蝶结结构特征,先对网络的GSC部分进行分解.定义GSC中顶点*i*, *j*之间的相异指数为它们之间较短的有向距离:

$$\text{Dissimilarity}(i, j) = \min(d(i, j), d(j, i)).$$

其中, $d(i, j)$ 和 $d(j, i)$ 分别为顶点*i*到*j*和顶点*j*到*i*的最短有向路的长度.用层次凝聚法分解网络的GSC部分,然后根据S, P部分的结点与GSC内结点的连接关系,按照“少数服从多数”的原则,将S, P部分的结点划分到GSC相应的子模块中.

Holme等人^[33]将代谢网络表示为有向的化合物-反应二部图,定义化合物结点间的相异度为连接它们的反应结点*r*的介数:

$$C_B(r) = \frac{1}{k_{in}(r)} \sum_{s \neq t} \frac{\sigma_r(s, t)}{\sigma(s, t)}.$$

其中, $\sigma(s, t)$ 为顶点*s*到*t*的所有最短路的数目, $\sigma_r(s, t)$ 为顶点*s*到*t*的以*r*为中间点的所有最短路的数目, $k_{in}(r)$ 为顶点*r*的入度.用层次分裂法分解网络,即依照介数由大到小的顺序去掉相应的反应结点和与之相连的弧,逐步将网络分解成子网络.

5 代谢网络的稳健性

生物系统的一个重要特征就是具有稳健性(robustness),即它们能在一定范围内抵抗外界(如温度、食物变化)和内部(如自发的基因突变)干扰而保持相对正常功能.例如,微生物在剔除突变的情况下仍能生长,有时还保持与野生型几乎相同的生长率.代

谢网络结构与稳健性之间相互关系的研究,有助于了解细胞的功能组织原理,在疾病机制研究^[34,35]及药靶筛选方面都有重要应用^[36,37]。

酶突变对细胞生存的影响问题,从代谢网络结构稳健性的角度上看,就是网络上某些结点或边的删除对网络上信息交流的影响。代谢网络以下的结构特征决定了它具有高度稳健性:

() 模块化结构。模块化的结构能避免网络内执行不同功能的各模块之间不必要的相互影响,同时又能把外界有害的扰动控制在有限的范围内,防止其扩散到整个网络,从而从总体上提高了网络的稳健性,减小系统崩溃的风险^[38,39]。

() 蝴蝶结结构。代谢网络蝴蝶结的庞大强连通体(GSC)中包含了对细胞功能极为重要的代谢,如糖酵解、三羧酸循环等,是网络中顶点之间的连接最紧密的部分^[17]。它是代谢网络的一个稳健的核心,因为其中大多数顶点对之间都有多条通路。同时,蝴蝶结结构的S和P部分采用多分枝结构的方式与GSC相连通,这种多输入、多输出的方式,有助于对外界刺激产生协调的反映,进一步提高系统稳健性^[38]。

() 无标度特征。无标度网络的一个重要特征是具有高度的容错性,当随机地去掉网络中相当高比例的结点时,网络的连通性几乎不受影响^[40]。代谢网络的无标度特征使它对偶发干扰具有很强的抵抗力^[5]。由此可以解释代谢网络中为什么当编码某些酶的基因发生突变时,并不影响细胞的正常代谢和生理过程。另一方面,少数的集散点控制着整个无标度网络的连通性,因而对网络集散点的蓄意攻击会导致整个网络的瘫痪,即无标度网络兼具容错性和脆弱性^[40]。有趣的是,最近, Mahadevan和Palsson^[41]通过计算代谢网络中与每个代谢物关联的所有反应中致命反应所占比例发现,就整个网络的功能而言,低度数结点起着几乎与高度数结点同样关键的作用。Samal等人^[42]进一步发现几乎所有致命反应都与低度数的代谢物关联。这个发现与上述代谢网络的容错性和脆弱性的结论^[5,40]并不一致,其原因在于前面的研究^[5,40]中,代谢网络的结点对应着代谢物。然而,实际上代谢物是不能通过基因工程手段去除的,而催化酶却可以^[41]。说明真实网络的结构与功能之间的关系是很复杂的,仅通过一方面的拓扑特性来推断功能是很困难的。

Palumbo等人^[43]研究了酵母的致命酶在代谢网

络中的分布特征,发现如果在代谢网络中删除致命酶所对应的边,会破坏网络的连通性,使得网络被分裂成互不相通的几个子网络,或者破坏强连通体,使其失去强连通性。Lemke等人^[22]研究了大肠杆菌的代谢网络中酶的删除对网络的破坏性。某个酶的破坏性指数 d 定义为网络中由于该酶的缺失而不能合成的下游代谢物的数量,发现大多数酶的 d 值都很小(<5),只有少数酶(9%)的 d 值较大(≥ 5),且其中包含了50%以上的致命酶。该研究小组还将他们的方法用于预测6种支原体(mycoplasmas)的代谢网络中的重要酶,发现这6种支原体的预测的重要酶是大都相同的^[44]。所有这些方法在基因工程靶标和药物靶标的筛选方面都具有实际意义。

6 结论与展望

在过去几年中,代谢网络的研究取得了许多重要的结果,对了解细胞代谢及细胞功能有一定的指导意义。代谢网络所具有的无标度、小世界、层次模块化等结构特征,也存在于基因调控网络、蛋白相互作用网络等细胞网络中^[45,46]。这些细胞网络共有的特征,使我们能从更高的抽象的层次认识生物系统,揭示出用以前的方法不能发现的生物现象。

根据细胞网络的层次模块化结构特征,细胞的组织体系可以抽象地描绘成一个复杂金字塔^[47]。塔底是分别负责信息储存、加工和执行功能的细胞成分——基因、RNA、蛋白质、代谢物;这些基本的结构单元组成小的、重复出现的模式——代谢网络中的通路,基因调控网络中的基序(motif)(第二层);通路及基序又分别整合形成功能模块(functional module),执行独立的细胞功能(第三层);这些模块再以层次嵌套的方式构成细胞的大规模的功能组织(塔顶)——全基因组的基因调控网络、蛋白相互作用网络、代谢网络。金字塔自底向上是一个从特异性向普适性的逐渐过渡:从分子到基序再到模块,自顶向下则是从普遍性向特异性的过渡:从网络共同存在的无标度性、层次性特点,过渡到物种特异性的模块和分子。虽然通路、基序、模块与整个网络之间的拓扑联系,还有待更深入的研究,但这种对网络进行模块化的研究是具体认识复杂系统功能的一个重要途径。对复杂的生物系统进行更深入的功能和结构层次的研究,并识别与疾病和肿瘤相关的具体结构模体和功能模块,是后基因组信息学发展的重要方向之一。

目前对代谢网络的拓扑结构已有一定程度的了解,但对于为何形成这样的结构却了解较少,即对进化和自然选择在细胞网络的拓扑结构的形成方面所起的作用尚不清楚。例如,无标度特性是复杂网络的共性,其中少量的集散结点控制着网络的连通性,而集散结点之间的连接方式对网络的行为有重要的影响。研究发现,细胞网络和技术网络中,集散结点之间直接的连接很少,它们更多地与非集散结点连接,称为不相配混合(disassortative mixing);而社会网络正好相反,其中集散结点间有大量的连接,称为相配混合(assortative mixing)^[48]。怎样的机制使细胞网络演化成与社会网络截然相反的集散点连接模式,这种连接模式对网络功能有何影响,是需要研究的问题。

代谢网络以及细胞网络的研究还仅仅处于起步阶段^[46],有关网络的结构和性质、网络的进化机制等方面的一系列问题还有待人们进一步的研究。另外,由于目前在细胞代谢的动力学参数方面积累的信息极少,代谢网络的研究主要着重于网络静态特征的研究。可以预见,随着基因组学、蛋白组学、代谢组学的发展,以及细胞动力学信息的积累,在细胞网络的复杂动力学的研究方面,将取得更多的成果。这些研究将有助于对复杂生物系统的建模和动态模拟,并具有潜在的生物医学应用价值。

致谢 本工作为国家科技攻关计划(批准号:2004BA711A21)、国家重点基础研究发展计划(批准号:2004CB720103)、国家自然科学基金(编号:30500107)以及上海市科学技术委员会(批准号:04QMX1450, 04DZ19850, 04DZ14005)资助项目。

参 考 文 献

- 1 Kanehisa M, 著. 孙之荣, 等译. 后基因组信息学. 北京: 清华大学出版社, 2002
- 2 Erdős P, Rényi A. On random graphs. Publ Math Debrecen, 1959, 6: 290—297
- 3 Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks. Nature, 1998, 393: 440—442 [DOI]
- 4 Barabasi A L, Albert R. Emergence of scaling in random networks. Science, 1999, 286: 509—512 [DOI]
- 5 Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic networks. Nature, 2000, 407: 651—654 [DOI]
- 6 Goto S, Nishioka T, Kanehisa M. LIGAND: Chemical database for enzyme reactions. Bioinformatics, 1998, 14: 591—599 [DOI]
- 7 Bairoch A. The ENZYME data bank in 1995. Nucleic Acids Res, 1996, 24: 221—222 [DOI]
- 8 Nakao M, Bono H, Kawashima S, et al. Genome-scale gene expression analysis and pathway reconstruction in KEGG. Genome Informatics, 1999, 10: 94—103
- 9 Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res, 2000, 28: 27—30 [DOI]
- 10 Karp P D, Krummenacker M, Paley S, et al. Integrated pathway-genome databases and their role in drug discovery. Trends Biotechnol, 1999, 17: 275—281 [DOI]
- 11 Karp P D. Pathway databases: A case study in computational symbolic theories. Science, 2001, 293: 2040—2044 [DOI]
- 12 Overbeek R, Larsen N, Pusch G D, et al. WIT: Integrated system for high-throughput genome sequence analysis and metabolic reconstruction. Nucleic Acids Res, 2000, 28: 123—125 [DOI]
- 13 Schuster S, Fell D, Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnol, 2000, 18: 326—332 [DOI]
- 14 Wagner A, Fell DA. The small world inside large metabolic networks. Proc R Soc Lond B, 2001, 268: 1803—1810 [DOI]
- 15 Ravasz E, Somera A L, Mongru D A, et al. Hierarchical organization of modularity in metabolic networks. Science, 2002, 297: 1551—1556 [DOI]
- 16 Ma H W, Zeng A P. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics, 2003, 19: 270—277 [DOI]
- 17 Ma H W, Zeng A P. The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics, 2003, 19: 1423—1430 [DOI]
- 18 Guimera R, Amaral L A N. Functional cartography of complex metabolic networks. Nature, 2005, 433: 895—900 [DOI]
- 19 Ma H W, Zhao X M, Yuan Y J, et al. Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. Bioinformatics, 2004, 20: 1870—1876 [DOI]
- 20 Horne A B, Hodgman T C, Spence H D, et al. Constructing an enzyme-centric view of metabolism. Bioinformatics, 2004, 20: 2050—2055 [DOI]
- 21 Gagneur J, Jackson D B, Casari G. Hierarchical analysis of dependency in metabolic networks. Bioinformatics, 2003, 19: 1027—1034 [DOI]
- 22 Lemke N, Herédia F, Barcellos C K, et al. Essentiality and damage in metabolic networks. Bioinformatics, 2004, 20: 115—119 [DOI]
- 23 Arita M. The metabolic world of *Escherichia coli* is not small. Proc Natl Acad Sci USA, 2004, 101: 1543—1547 [DOI]
- 24 Broder A, Kumar R, Maghoul F, et al. Graph structure in the web. Comput Netw, 2000, 33: 309—320 [DOI]
- 25 Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the internet topology. Comp Comm Rev, 1999, 29: 251—262 [DOI]
- 26 Ravasz E, Barabási A L. Hierarchical organization in complex networks. Phys Rev E, 2003, 67: 026112 [DOI]
- 27 Hartwell L H, Hopfield J J, Leibler S, et al. From molecular to modular cell biology. Nature, 1999, 402: C47—C51 [DOI]
- 28 Lipson H, Pollack J B, Suh N P. On the origin of modular varia-

- tion. *Evolution*, 2002, 56: 1549—1556.
- 29 Papin J A, Price N D, Palsson B O. Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Res*, 2002, 12: 1889—1900 [DOI]
- 30 Stelling J, Klamt S, Bettenbrock K, et al. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 2002, 420: 190—193 [DOI]
- 31 Schilling C H, Palsson B O. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J Theor Biol*, 2000, 203: 249—283 [DOI]
- 32 Schuster S, Pfeiffer T, Moldenhauer F, et al. Exploring the pathway structure of metabolism: Decomposition into subnetworks and application to *Mycoplasma pneumoniae*. *Bioinformatics*, 2002, 18: 351—361 [DOI]
- 33 Holme P, Huss M, Jeong H. Subnetwork hierarchies of biochemical pathways. *Bioinformatics*, 2003, 19: 532—538 [DOI]
- 34 Schuster R, Holzhütter H G. Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations: Application to enzyme deficiencies of red blood cells. *Eur J Biochem*, 1995, 229: 403—418 [DOI]
- 35 Martinov M V, Plotnikov A G, Vitvitsky V M, et al. Deficiencies of glycolytic enzymes as a possible cause of hemolytic anemia. *Biochim Biophys Acta*, 2000, 1474: 75—87
- 36 Bakker B M, Mensonides F I C, Teusink B, et al. Compartmentation protects trypanosomes from the dangerous design of glycolysis. *Proc Natl Acad Sci USA*, 2000, 97: 2087—2092 [DOI]
- 37 Wagner C. Systembiologie gegen Parasiten. *BioWorld*, 2004, 9: 2—4
- 38 Kitano H. Biological robustness. *Nature Rev Genetic*, 2004, 5: 826—837 [DOI]
- 39 Stelling J, Sauer U, Szallasi Z, et al. Robustness of cellular functions. *Cell*, 2004, 118: 675—685 [DOI]
- 40 Albert R, Jeong H, Barabasi A L. Error and attack tolerance of complex networks. *Nature*, 2000, 406: 378—382 [DOI]
- 41 Mahadevan R, Palsson B O. Properties of metabolic networks: Structure versus function. *Biophysical J*, 2005, 88: L7—L9 [DOI]
- 42 Samal A, Singh S, Giri V, et al. Low degree metabolites explain essential reactions and enhance modularity in biological networks. *BMC Bioinform*, 2006, 7: 118 [DOI]
- 43 Palumbo M C, Colosimo A, Giuliani A, et al. Functional essentiality from topology features in metabolic networks: A case study in yeast. *FEBS Lett*, 2005, 579: 4642—4646 [DOI]
- 44 Mombach J C M, Lemke N, Silva N M, et al. Bioinformatics analysis of mycoplasma metabolism: Important enzymes, metabolic similarities, and redundancy. *Comput Biol Med*, 2006, 36(5): 542—52 [DOI]
- 45 Wuchty S, Ravasz E, Barabási A L. The architecture of biological networks. In: Deisboeck T S, Kresh J Y, Kepler T B, eds. *Complex Systems in Biomedicine*. New York: Kluwer Academic Publishing, 2003
- 46 Barabasi A L, Oltvai Z N. Network biology: Understanding the cells's functional organization. *Nature Rev Genetics*, 2004, 5: 101—113 [DOI]
- 47 Oltvai Z N, Barabási A L. Life's complexity pyramid. *Science*, 2002, 298: 763—764 [DOI]
- 48 Newman M E J. Mixing patterns in networks. *Phys Rev E*, 2003, 67: 026126 [DOI]

(2005-11-14 收稿, 2006-03-27 接受)