

基因组尺度集成细胞网络模型研究进展

武敏^{1,2} 马红武^{2*}

(1. 代谢工程与系统生物学重点实验室 天津大学化工学院 天津 300072)

(2. 中国科学院系统微生物工程重点实验室 中国科学院天津工业生物技术研究所 天津 300308)

摘要: 细胞网络研究是系统生物学的一个研究热点, 通过结合计算机模型和实验技术, 从系统角度分析复杂的生物系统, 可以为生物实验提供指导和预测。近十年来, 国内外许多研究团队致力于基因组规模代谢网络、基因调控网络和信号转导网络模型的构建和分析, 并取得了一定成果。而不同类型网络的集成和分析是当前生物网络研究中一个新的方向, 并带来了诸多新的挑战。在本文中, 主要对基因组尺度集成细胞网络模型的研究进展, 特别是对代谢网络和转录网络的集成进行了详细论述, 着重于集成网络的构建和分析方法, 最后对该领域研究前景进行了展望。

关键词: 基因组尺度, 细胞网络模型, 集成方法, 集成网络

The progress of integrated genome-scale cellular networks

WU Min^{1,2} MA Hong-Wu^{2*}

(1. Key Laboratory for Metabolic Engineering and Systems Biology, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China)

(2. Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China)

Abstract: The study of cellular networks is one of the hot research topics in systems biology. Combining computational network models with experiment techniques, researchers can analyze complex biological systems from a systemic perspective, and provide new hypothesis and guidance for biological discovery. In the last ten years, many methods for the reconstruction and analysis of genome-scale metabolic networks, gene regulatory networks and signal transduction networks have been developed. The integration of these different types of biological networks is becoming a new research hotspot recently. The study of integrated network raises many new challenges. In this paper, we reviewed the methods for the reconstruction and analysis of integrated networks, with a particular focus on the integrative models of metabolic and transcriptional regulatory networks as most researches are on this type of integrated networks. We first gave a brief introduction of the different cellular networks, and then discussed the reconstruction methods and the application of integrated networks. In the end, the problems and future development directions were also discussed.

基金项目: 国家 973 计划项目(No. 2012CB725203, 2011CBA00804); 国家 863 计划项目(No. 2012AA022103); 天津市应用基础及前沿技术研究计划项目(No. 12JCYBJC33000)

*通讯作者: Tel: 86-22-24828735; 信箱: ma_hw@tib.cas.cn

收稿日期: 2013-03-18; 接受日期: 2013-05-14; 优先数字出版日期(www.cnki.net): 2013-10-11

Keywords: Genome-scale, Cellular networks, Integration methods, Integrated network

系统生物学是一个在整体水平上整合不同分子水平信息以理解生物系统如何运作的领域。通过研究某生物系统各不同组成成分之间的相互关系和相互作用(例如,与细胞信号传导、代谢通路、细胞器、细胞、生理系统与生物等相关的基因和蛋白网络),系统生物学希望最终能够建立整个系统的可被理解的模型。随着基因组计划、生物信息学的发展,代谢网络、细胞信号传导与基因调控网络的研究,高通量组学分析技术、生物计算机软件设计技术的应用,系统生物学研究已经在不断趋近这一最终目标,但也面临着许多新的挑战和困难。以细胞网络研究为例,目前一个研究难点就是不同类型细胞网络的集成,并通过集成网络获得对生物系统更全面的认识,由此获得新的发现。在本文中我们将讨论基因组规模集成细胞网络的研究进展,其中包括代谢网络、基因调控网络和信号转导网络等,重点是代谢网络和转录调控网络的集成。

1 细胞网络分类

细胞网络是系统生物学的一个研究热点,其通过结合计算机模型和实验数据,从整体系统角度分析复杂的生物系统。研究者一般根据生物分子间相互作用的类型将细胞网络分解成不同类型的生物网络分别进行研究,如代谢网络、转录调控网络和信号转导网络等^[1]。

1.1 基因组尺度代谢网络

基因组尺度代谢网络从基因组序列出发,结合基因、酶反应和代谢物数据库及相关实验数据,从系统的角度定量研究生命体的代谢过程^[2-4]。基因组尺度代谢网络是目前研究最广泛和最深入的网络,模型构建方法已较完善。其构建方法主要包含四大部分,即数据收集、关系模型建立、数据整理修正和转化为数学模型。数据收集的主要来源是各种生物信息数据库(如 KEGG、BRENDA、UniProt 等)和文献。从数据库和文献中获得的原始数据主要包括:物种特异的基因、蛋白质、反应和代谢物

信息;其次,将这些数据关联在一起,构建了关系模型,如基因与酶、酶与反应、反应与代谢物之间的关联;再次,进行数据整理和修正,这是构建基因组尺度代谢网络过程中工作最繁重的一个环节,需要大量时间进行人工校正。通过以上过程后,得到的结果是包括了物种的基因-蛋白质-反应对应信息以及反应的详细信息(包括反应方程式、反应方向、反应所属途径和所属分室等)的一个反应列表;最后,要将其转化为数学模型才可以在计算机上进行相应的模拟运算,而基因组尺度代谢网络模型的核心就是计量系数矩阵。计量系数矩阵可以通过编程,直接将上面整理得到的反应方程式转化为 SBML (Systems biology markup language) 格式的计量系数矩阵;SBML 的数据文件是通用的系统生物学语言格式的文件,可以被大多数生物模拟软件识别并加载^[5]。近年来,比较基因组学的发展也促进了代谢网络重构的发展,该方法通过尽量发掘、综合基因组蕴藏的相关重要信息,如基因簇、调节位点共享等,准确预测基因的功能,有效解决缺失基因问题;然后在此基础上,通过实验验证基因功能预测,最终实现代谢网络的重构;不少研究表明,该方法能够显著提高基因功能注释和代谢网络重构的质量,有助于发现新的代谢功能基因、调控元件、甚至新的代谢途径,从而为优化设计细胞内从原料到产品的生物合成路线提供指导^[6]。具体的基因组尺度代谢网络构建方法可参考文献^[5-7],其中 Palsson 等在 2010 年发表的文章中详细介绍了基因组尺度代谢网络的构建方法,共有 96 步,可供进行基因组尺度代谢网络研究的生物工作者参考和使用。另外,王晖等构建了一个基因组尺度代谢网络模型数据库(GSMNDB),本文作者也参与了该数据库的信息收集、整理和更新,目前该数据库中已经包含了 135 个已发表的基因组尺度代谢网络模型,包括原核生物、真核生物和古细菌。

在基因组尺度代谢网络模型构建方面,作为原核常用模式菌,大肠杆菌已经有了多个模型。目前

最新的大肠杆菌基因组尺度代谢网络模型是 2011 年 Palsson 研究组发表的 *iJO1366* 模型^[8]。*iJO1366* 模型是迄今为止所有微生物中较准确全面的代谢网络模型, 可以用作其他细菌代谢网络构建的基础。另外, 在此基础上还可以进行基因组尺度代谢网络与其他细胞系统(如转录和翻译模型或转录调控)的集成。

1.2 基因组尺度转录调控网络

当前的高通量功能基因组实验技术, 主要是 DNA 微阵列(DNA microarray)技术的发展, 产生了海量的基因组表达谱数据。表达谱数据不仅可用于分析基因表达的时空规律, 研究基因的功能, 而且还可用于分析基因之间的相互制约关系, 研究基因转录调控网络。基因转录调控网络构建原理如图 1 所示, 在一个基因的转录过程中, 一组转录因子作用于该基因的启动子区域, 控制该基因转录; 而转录因子本身又是其它基因的表达产物; 当一个基因通过转录、翻译形成功能基因产物后, 它将改变细胞的生化状态, 从而直接或间接地影响其它基因的表达, 甚至影响自身的表达; 由此, 多个基因的表达状态不断变化, 就使得细胞的生化状态不断变化, 便构成了复杂的基因转录调控网络^[9]; 目前, 图论模型是最简单的而且是基因组规模调控网络研究中用的最多的模型。类似于代谢网络的构建方法, 转录调控网络的构建包括数据收集、整理和转化为计算模型, 其数据来源主要包括两种, 一种是自动构建, 这需要理想的 ChIP-chip 实验数据, 该实验是在典型实验条件下在转录因子敲除菌株中获得所有主要的转录因子与基因表达谱。第二种是基于基因组和文献数据获得。然后, 利用自动构建或基因组和文献数据构建的转录调控网络以两种方式呈现: (1) 以图的方式, 每个转录因子节点与它的目标基因节点通过有向边连接; (2) 以共调控基因模块的方式, 其中包括转录因子和与每个模块相关的环境(如碳源)调控子。为了预测对环境或基因扰动应答, 这种模型要转化为计算模型, 大尺度调控网络计算模型多采用布尔网络方法。另外, 随

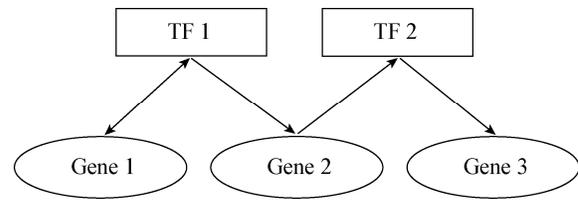


图 1 一个简单的基因转录调控网络

Figure 1 A simple gene transcriptional regulatory network

注: 基因 1 表达生成转录因子 1; 基因 2 表达生成转录因子 2, 该转录因子又可以调控基因 3 的表达; 同时, 转录因子 1 又可以调控自身基因 1 的表达; 根据这种关系, 所有基因和转录因子的关系便构成了完整的基因转录调控网络。

Note: Gene 1 expressed and produces TF 1; Gene 2 expresses and produces TF 2 which regulates Gene 3's expression; meanwhile, TF 1 can regulate Gene 1's expression; All the relations between genes and TFs form a complete transcriptional regulatory network.

着大量原核微生物基因组被破译, 基因序列信息利用比较基因组学新技术发掘和分析基因上游调控区序列, 准确预测转录因子结合位点和 RNA 元件, 可以重构各个调节单元, 即由一个转录因子或调节 RNA 调控的所有操纵子, 在此基础上可构建细胞内的转录调控网络, 这可以为基因功能预测提供重要依据, 并阐明对代谢途径起重要调控作用的转录因子以及代谢物和转录因子之间、不同转录因子之间的相互作用关系^[7]。基因转录调控网络的详细构建方法可参考文献[10]。基因调控网络是一个非常复杂的非线性系统, 因此用数学模型进行描述有一定的难度。迄今为止, 基因调控网络的研究在许多方面尚处于探索阶段, 相关数据库的建设和高效率算法的开发是一项紧迫而艰巨的任务^[11]。在转录调控网络的应用方面, 转录调控网络分析可以用于鉴别新的调控关系以及预测细胞行为。

1.3 信号转导网络

生物需要复杂而精密的信号系统感受、传导和响应外界环境的变化; 而细胞对外界刺激的感受和反应都是通过信号转导系统的介导实现的。该系统由受体、酶、通道和调节蛋白等构成; 通过该系统, 细胞能感受、放大和整合各种外界信号。其过程是

细胞识别与之相接触的细胞,或者识别周围环境中存在的各种化学和物理信号,对特定的外源信号产生反应,并将信号传递到细胞内转变为细胞内各种分子活性的变化,最终产生相应的调控信号以调节靶细胞的基因表达,从而改变细胞的某些代谢过程,影响细胞的生长速度,甚至诱导细胞凋亡,这种针对外源信息所发生的细胞应答反应全过程称为信号转导。由于细胞内各种信号通路之间存在着紧密的联系和交叉调控,因而形成了非常复杂的信号转导网络。信号转导网络构建原理如图2所示,存在两种传导方式,第一种是胞外信号分子B由膜上接收蛋白直接传递到细胞内,影响相应的转录因子的活性,改变它对目标基因的调控状态;第二种即胞外的信号分子A由膜上接收蛋白传递给胞内的蛋白激酶或其他分子,然后再活化相应的转录因子,使其对目标基因的调控发生改变,从而影响基因的表达。这样根据已知的细胞信号转导途径以及信号转导分子之间的生化作用机制的研究结果,可以构建出复杂细胞信号转导网络的模型。在该模

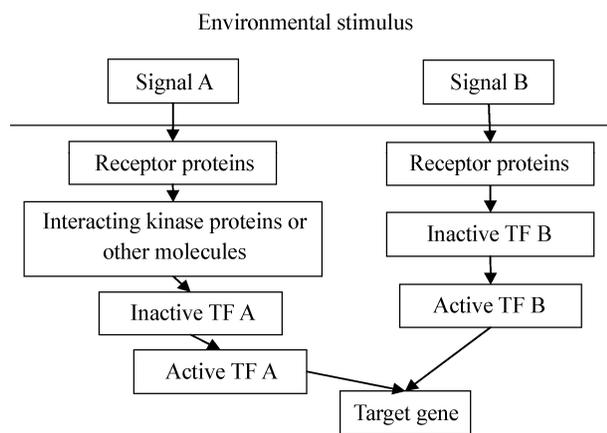


图2 信号转导网络简图

Figure 2 A simple signal transduction network

注:胞外信号分子由膜上接收蛋白直接或者间接传递给转录因子,引起转录因子活性改变,从而改变对目标基因的调控状态。

Note: Extracellular signals are received by receptor protein directly or indirectly and passed to transcription factors, whose activities are changed, thus changing regulation state for target genes.

型中,各种信号转导分子构成了信号转导网络的节点,而信号转导分子之间的生物化学反应则形成了节点之间的连结。信号转导网络是外部环境、基因组和代谢系统的连接纽带,因此构建基因组尺度信号网络对理解复杂疾病和探索新治疗方法是必需的。不同于基因组尺度代谢网络,基因组尺度信号转导网络的构建研究才刚刚开始,其挑战在于目前我们仍缺乏信号转导系统中的所有组分列表以及良好的构建原则,而合成生物学方法在推动信号网络构建方面会越来越重要,通过设计新途径和重构已有途径会为理解信号途径的规则提供支持。

2 集成网络进展

系统生物学的目的在于通过研究不同生物分子和其特性之间的相互关系来定量模拟细胞行为。生物网络的建立模拟和验证了系统生物学中的关键活动。在此领域里,围绕着不同的网络类型,国际上形成了许多大型研究团队,他们应用不同工具对不同网络进行分析。随着对各单一网络的深入研究,研究者们认识到了单一网络的局限性和预测的不准确性,因而以单一网络为基础的多网络集成已成为系统生物学领域新的研究热点。这主要是人们认识到生物中各组成成分(基因、蛋白质、代谢物等)的功能都是通过与其他部分的相互作用体现出来的,因此只有在系统水平上对所有相关成分进行整体分析才能对生物的生理功能有全面正确的了解^[12]。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^[1]。其中,代谢网络与转录网络的集成研究最广泛、也最深入,因此本文重点介绍它们的研究进展。

2.1 代谢网络与转录网络集成进展

基因调控和代谢反应是生命的两个主要活动。一些实验表明,代谢反应的底物影响转录因子或信号转导途径的活性;而更多的研究表明,在不同环境条件下,催化代谢反应的酶是差异表达的,因此基因调控和代谢反应之间的相互关系不容忽视。然而,虽然许多工作用于研究每个系统,这两者之间

的耦合以前被较少研究和理解。近十年来, 调控-代谢网络的集成建模成为生物网络建模的一个重要方向^[13]。通过集成代谢-调控网络, 使我们更好地预测基因变异和转录扰动是如何在代谢水平上转变为通量应答的^[13]。转录调控与代谢网络的关系如图 3 所示, 转录因子调控目标基因的表达, 目标基因表达合成蛋白质(如酶), 而酶催化反应生成代谢物。而酶、代谢物的浓度又可以影响转录因子的活性, 从而影响目标基因的表达水平。其集成分析方法主要有 RFBA 和 PROM 等, 如下详述。

2.1.1 集成方法 1 (RFBA): 目前, 将转录调控与代谢网络联系在一起的最普遍的基因组尺度分析方法是调控通量平衡分析(RFBA)^[14-15]。RFBA 将通量平衡分析和调控布尔逻辑结合起来构建了代谢调控网络, 即将调控约束条件结合到代谢网络中。具体来说, 在代谢网络中, 代谢通量分析根据代谢路径中各反应的计量关系以及实验中所测得的数据来确定整个代谢反应网络的代谢通量分布。

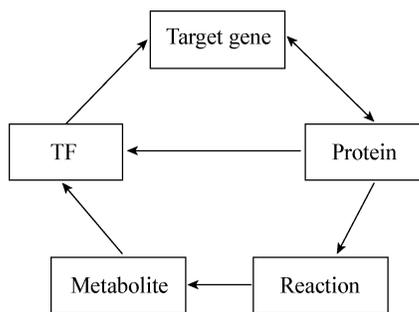


图 3 代谢-转录调控关系图

Figure 3 Association between metabolism and transcriptional regulation

注: 转录因子(TF)调控目标基因(Target gene)的表达, 目标基因表达合成蛋白质(Protein), 如酶(Enzyme), 而酶催化反应(Reaction)生成代谢物(Metabolite); 而酶、代谢物的浓度又可以影响转录因子的活性, 从而影响目标基因的表达水平。

Note: Transcriptional factors regulate the expression of target genes; target genes express and produce proteins, such as enzymes; enzymes catalyze reactions which produce metabolites; the concentration of enzymes and metabolites affect the activity of transcriptional factors, which indirectly affect the expression level of target genes.

然后在稳态条件下, 把调控约束条件结合到代谢网络中, 评价基因在某一特定条件下的表达或抑制, 例如若一个基因被抑制, 则把被抑制的基因产物的通量设定为零; 若一个基因表达, 则其产物实际通量值由 FBA 确定(FBA 即通量平衡分析, 该算法假定系统处于拟稳态, 即中间代谢物的生成与消耗相同, 在这种假设下设置一定的约束条件和目标函数, 来研究胞内相应状态下的通量分布情况), 由此改变了原有解空间的形状。因此, 在 RFBA 中, 代谢网络不仅受到质量、热动力学和能量的限制, 也受基因调控网络的限制。与单一代谢网络模型(只采用 FBA 方法进行分析)相比较, RFBA 的优点在于: 由于调控影响在代谢通量平衡模型中起到重要作用, 调控约束条件的加入极大地提高了代谢通量平衡模型的预测能力和精确性, 可在多种环境条件下模拟动态定量底物吸收、细胞生长和副产物分泌情况^[16]。另外, 调控约束条件的加入对于解释细胞培养基中的动态行为也是有用的, 例如在某一特定条件下 RFBA 能预测某个特定同工酶の利用情况, 而 FBA 分析不能预测两个同工酶中哪个对于催化更有活性。最后, 该模型还可以定性预测酶产物的上调或下调, 以及研究转录调控的系统效应, 这些都是优于 FBA 分析的。

但是在用 RFBA 方法构建集成模型时, 有几个缺点。RFBA 模型把转录组与代谢组的关系简化为二元过程, 基因只有两种状态(表达或不表达), 而没有中间值; 这种把转录调控的效果当作“开或关”来处理过于简单, 比如负调控的基因并不是完全不表达, 而是表达量下降^[13], 因而单纯的“开或关”会导致模拟不准确。其次, 用 RFBA 构建集成网络的最大障碍可能是缺乏一个将调控子和目标基因联系在一起的确定的布尔规则的自动算法。虽然在构建代谢调控中, 手动过程是精确的, 但是手动构建极大地限制了建模的相互关系数量, 因而不能构建基因组尺度的代谢-调控网络。考虑到相互关系数量巨大, 在基因组尺度上设定布尔规则并证实重要的相互关系是极其困难的。

2.1.2 集成方法 2 (PROM): 2010 年 Sriram Chandrasekaran 和 Nathan D. Price 提出了代谢系统的概率调控方法(PROM)^[17]。用 PROM 构建集成的代谢-调控网络需要以下几方面:(1) 基因组尺度代谢网络;(2) 调控网络结构,由转录因子和其目标基因组成;(3) 充分的基因表达数据,其转录组在各种环境和遗传扰动中被测量;(4) 由代谢物和蛋白质引起的其他涉及酶调控的相互关系。PROM 引入概率来代表基因状态和基因-转录因子相互关系。例如,当调控的转录因子 TF B 不表达(Off)时,基因 A 表达(On)的概率被表示为 $P(A=1|B=0)$ 。类似的, $P(A=1|B=1)$ 表示当 B 表达(On)时, A 表达(On)的概率。转录因子和目标基因的关系由微阵列数据进行量化,具体方法可参见文献[17]。利用这些相互关系数据,用 PROM 模拟了扰动对调控网络的影响。这些调控关系的概率值用于限制目标基因控制的反应通量。例如,在代谢网络中,如果基因 A 表达(On),由基因 A 产物催化的反应通量不会超过该反应的最大可能通量(最大速率 V_{max}),当基因 A 不表达(Off)时,其为 0。因此,当加入调控关系的概率值后,也就是当 P 代表基因表达(On)的概率时,其代谢网络反应通量的上限值为 $P \cdot V_{max}$ 。系统反应 V_{max} 在未受调控的代谢模型上由通量变化分析(FVA)估量。当约束条件设定后,被调控网络的最优化生长由和 FBA 相同的线性最优化问题决定。该 PROM 构建方法为在尽可能多的环境条件下,利用微阵列数据获得转录因子与目标基因的开/关状态的概率值;然后用这些调控关系中的概率值来约束通过代谢网络的通量,即用概率值乘以代谢网络原有的反应通量,这样就获得了转录调控和代谢网络的集成。在 RFBA 中基因是完全的开或关,而在 PROM 中,基因状态可以取中间值,如图 4 所示。因此,对于有多个转录因子的基因,可以有多个概率值。该概率的应用,优点在于可使模型能够区分强调控子和弱调控子,而且每个转录因子对于此目标基因的相互影响都可以进行模拟。与 RFBA 相比, PROM 能够直接集成

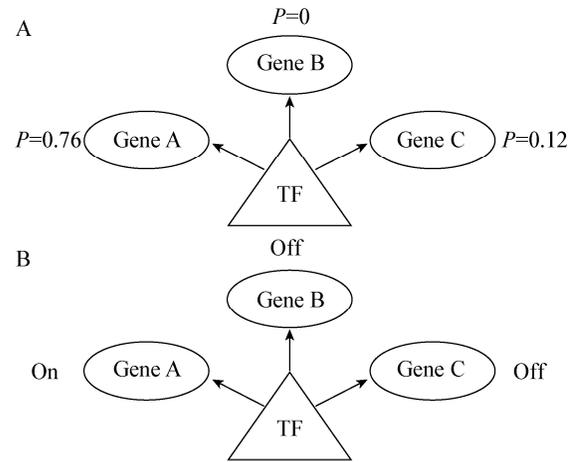


图 4 PROM 与 RFBA 的比较

Figure 4 Comparison between PROM (A) and RFBA (B)

注: PROM 中的基因状态用概率表示,存在中间值;而 RFBA 中的基因只有两种状态,开或关。

Note: In PROM the gene state can take intermediate values; in RFBA, the gene can only be completely on or off.

转录和代谢网络,并自动量化来自高通量数据的相互作用关系,克服了手动编写布尔规则的需要,能极大地增加基因组尺度集成模型的数量。由此 PROM 能帮助构建全面的基因组尺度细胞网络模型。但是我们认为该方法也存在一定的缺陷,因为它通过对高通量数据量化,得到了综合环境条件下基因状态的概率值,却不是精确值,即不能得到某一特定条件下的基因状态,因此不能精确地获得某一特定环境条件下的细胞状态。这对于我们想研究和精确预测某一特定环境条件下细胞生长状态是不适用的。

2.1.3 集成方法 3 (关系图法): 2012 年姜金国等^[1]结合公共数据库、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。集成方法是以基因→酶的关系为接口将转录调控网络(调节子→靶基因)和代谢网络(底物→反应→产物;酶→反应)整合起来,形成调节子→靶基因→酶→反应和底物→反应→产物的集成细胞网络模型。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,最终构建了谷氨酸棒状杆菌集成细胞网络。该集成网络的优点在于发现了一些细胞内的转录与代谢双层复合调控关系,而这在单一的代谢网络或转录调控网络中是无法获得的。

2.1.4 集成工具(TIGER 工具箱): 2011 年, Jason A Papin 等提出了 TIGER 工具箱,该工具箱用于集成基因组尺度代谢模型、表达数据和转录调控网络^[18]。TIGER 将一系列普遍规则、布尔规则或多层次规则转变为一套混合整数不等式。该程序包也包括了现有算法,来集成高通量表达数据和代谢-转录调控的基因组尺度模型。研究者展示了 TIGER 是如何自动耦合基因组尺度代谢模型和 GPR (Gene-protein reaction)逻辑,以及转录调控模型,并把它作为一个算法发展平台进行大尺度代谢网络分析。另外,用酿酒酵母已构建的转录调控网

络作例子,研究者展示了 TIGER 算法是如何用于确定不一致性和改进转录调控的现有模型的。TIGER 程序包为算法发展、扩增调控网络和现有的基因组尺度代谢网络,以及高通量数据提供了平台。

2.2 转录调控网络与信号调控途径集成

2010 年, Wang 等^[19]将转录调控和蛋白质-蛋白质相互作用(主要是信号调控途径)结合起来构建了酵母的集成细胞网络(图 5)。他们从数据库集成了几种组学数据信息,包括微阵列基因表达数据、转录因子和基因的调控关系、蛋白质-蛋白质相互关系,并提出了一个系统方法构建了基于耦合动态模型和统计评价的集成细胞网络。转录调控和蛋白质-蛋白质相互作用之间的关系构成了集成细胞网络的框架。应用微阵列数据,研究者可以构建基因调控网络和信号调控途径,然后基于基因调控网络和信号调控途径的连接点(即转录因子),这两

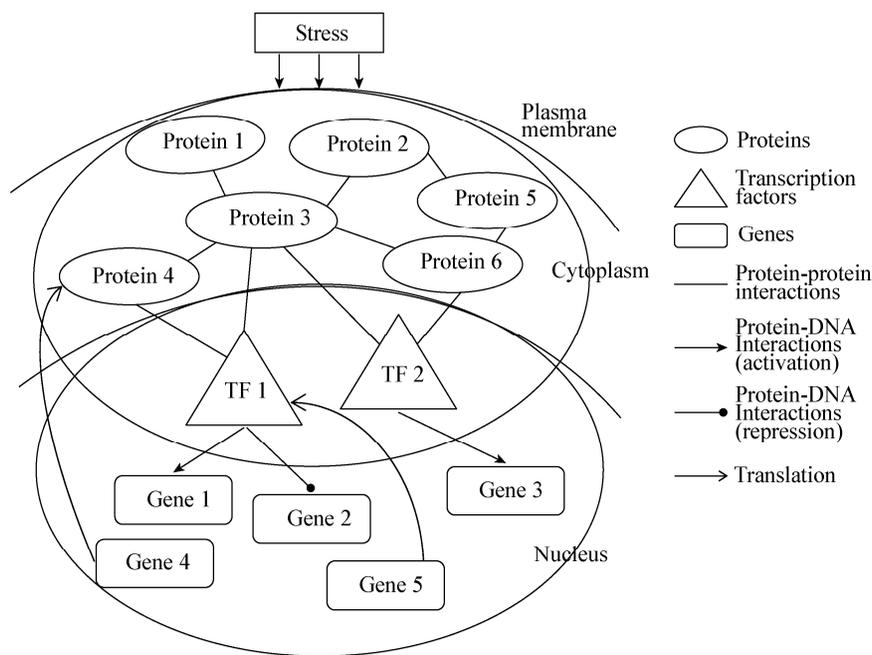


图 5 集成的细胞网络图示

Figure 5 Schematic diagram of the integrated cellular network

注:集成的细胞网络由两个子网络构成。一个是信号调控途径(上方椭圆);另一个是基因调控网络(下方椭圆);其中,转录因子是连接信号调控途径和基因调控网络的枢纽。

Note: The integrated cellular network consists of two subnetworks; one is the signaling regulatory pathway (the oval above) in which protein-protein interactions are presented and the other is the gene regulatory network (the oval below) in which transcription regulations exist; the transcription factors serve as the interface between the signaling regulatory pathway and the gene regulatory network.

个网络被耦合集成细胞网络。该方法可应用于酿酒酵母的应激反应,阐明酵母的应激反应机制。信号调控途径高效表达以感受和传递应激信号,基因调控网络表达以编码保护蛋白来应答应激反应,因此应激反应机制可以充分验证集成细胞网络。最后经过文献证实,不同应激条件下酿酒酵母的集成细胞网络所采用的构建方法是可行的。该集成方法的提出,为我们研究转录调控网络和信号转导网络的集成提供了参考,然而该方法只是将转录调控数据与信号调控途径中的蛋白质-蛋白质相互关系集成在一起,还不能算是转录调控网络与信号转导网络的完整集成。要想获得更加完整的集成,我们还需要获得信号转导网络中的所有组分列表信息,然后与转录调控网络中的数据对应起来以实现集成化构建。

2.3 代谢网络、转录调控网络和信号转导网络的集成

关于3个网络的集成,可能因为缺少合适的集成方法,目前文献报道还很少。小尺度的集成网络已报道过,如2008年Markus W. Covert等在RFBA基础上提出了iFBA(集成FBA)^[20],他们结合通量平衡分析(FBA)、调控布尔逻辑和常微分方程(ODE)构建了大肠杆菌代谢、调控和信号集成动态网络模型iFBA。该集成模型通过验证两个模型(即RFBA模型和ODE模型)的数值(两个模型共有的代谢物和通量)来集成两个模型。iFBA集成模型包括了基于通量平衡的中心碳代谢过程和转录调控模型,以及基于常微分方程的碳水化合物吸收控制模型。其中,通量平衡模型(FBA)优于传统常微分方程(ODE)模型的方面在于他们可以分析整个代谢和调控网络,ODE模型的优势在于他们可以获得细胞内浓度和短时间水平上的动态行为,而这对于信号转导是必需的。iFBA模型结合了以上两种模型的优势,整合了不同来源的信息和不同精度的数据,从而使得多种建模方法能够取长补短,同时由于加入了信号模型,能够发现一些酶表达并且具有活性,但这些酶并不参与到代谢工程中,而是和信号转导有关

系^[13];但是由于基于随机模型或微分方程的方法受建立小系统模型的限制,目前还无法扩展到基因组尺度。同年,Jason A. Papin等提出了idFBA(集成动态FBA)^[21]。idFBA将信号转导、代谢和转录过程通过化学计量矩阵形式进行构建,动态模拟细胞表型行为。它假定了假稳态条件作为快反应,并把慢反应结合到化学计量形式中。该方法除采用与RFBA相似的迭代策略外,还增加了一个接口,用于接收每个慢速反应在每个时间点是否发生的信息,这样可以通过时间延迟的方式将慢速反应集成进来。将idFBA结构用于单细胞真菌生物酿酒酵母的典型模块中,最终发现idFBA使定量动态分析细胞外信号对细胞表型的系统影响更加容易,并在与一个同等的动态模型比较中产生了可比较的时间进程上的预测。该模型的优势在于能够刻画三种网络在统一模型下的相互作用,同时这种相互作用对理解细胞过程,诊断和治疗疾病都有重要意义;但目前信号网络的数据较少,研究受到了限制^[13]。idFBA解决了线性规划问题,并且不需要详细的动态参数,如果我们可以获得足够的信号网络数据,或许它可以被有效地放大到包含了基因组尺度的信号、代谢和调控过程的集成细胞内系统。

3 细胞网络研究展望及应用价值

细胞网络(基因调控网络、代谢网络、信号转导网络等)的分析和重构是后基因组学研究中的一个崭新的、极其重要的研究领域。其研究方法主要是利用大量的生物实验数据,运用数据挖掘技术来反向分析和挖掘特定生物元件之间的关联信息,并试图以复杂系统的观点为出发点,从生物元件之间关系的角度来揭示和再现它们之间相互作用的网络拓扑结构,揭示其复杂的作用机理及其功能信息。其中代谢网络、转录网络和信号转导网络的构建、集成和分析成为系统生物学领域的研究热点和难点。从2001年代谢和转录调控网络集成到现在,其发展历史才仅仅十年,可以说细胞网络仍处于发展的初始阶段,虽然取得了一定进展,但仍有很多

待解决的问题, 有很大的发展空间。在我国, 生物网络的研究才刚刚起步, 但我们相信网络生物学必然会帮助取得对生物系统更深入的认识。网络生物学也是系统生物学的组成部分, 可以为实验提供指导和预测。但只是靠生物网络对生物学的研究有其局限性, 必须和生物学自身的生化规律相结合, 与实验研究及各种新技术相结合, 探索前进。

参考文献

- [1] 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[J]. 生物工程学报, 2012, 28(5): 577-591.
- [2] Ravasz E, Somera AL, Mongru DA, et al. Hierarchical organization of modularity in metabolic networks[J]. Science, 2002, 297: 1551-1555.
- [3] Stelling J, Klamt S, Bettenbrok K, et al. Metabolic network structure determines key aspects of functionality and regulation[J]. Nature, 2002, 420: 190-193.
- [4] Famili I, Forster J, Nielsen J, et al. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network[J]. Proceedings of the National Academy of Sciences, 2003, 100: 13134-13139.
- [5] 王晖, 马红武, 赵学明. 基因组尺度代谢网络研究进展[J]. 生物工程学报, 2010, 26(10): 1340-1348.
- [6] 杨琛. 基于比较基因组学重构细菌的代谢途径和调控网络[J]. 生物工程学报, 2010, 26(10): 1349-1356.
- [7] Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction[J]. Nature Protocol, 2010, 5: 93-121.
- [8] Orth JD, Conrad TM, Jessica Na, et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011[J]. Molecular Systems Biology, 2011, 7: 535.
- [9] 刘化锋, 王文燕. 基因转录调控网络模型[J]. 山东大学学报, 2006, 41(6): 103-108.
- [10] Feist AM, Herrgard MJ, et al. Reconstruction of biochemical networks in microorganisms[J]. Nature Review Microbiol, 2009, 7(2): 129-143.
- [11] 雷耀山, 史定华, 王翼飞. 基因调控网络的生物信息学研究[J]. 自然杂志, 2004, 26(1): 7-12.
- [12] Kitano H. Systems biology: a brief overview[J]. Science, 2002, 295(5560): 1662-1664.
- [13] 赵月, 奚燕萍, 王飞, 等. 调控-代谢网络的集成建模与分析方法研究[J]. 计算机应用与软件, 2011, 28(7): 1-4,15.
- [14] Covert MW, Knight EM, Reed JL, et al. Integrating high-throughput and computational data elucidates bacterial networks[J]. Nature, 2004, 429(6987): 92-96.
- [15] Covert MW, Schilling CH, Palsson B. Regulation of gene expression in flux balance models of metabolism[J]. Journal of Theoretical Biology, 2001, 213: 73-88.
- [16] Covert, MW, Palsson BO. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*[J]. Journal of Biological Chemistry, 2002, 277(31): 28058-28064.
- [17] Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(41): 17845-17850.
- [18] Jensen PA, Lutz KA, Papin JA. TIGER: toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks[J]. BMC Systems Biology, 2011, 5: 147.
- [19] Wang YC, Chen BS. Integrated cellular network of transcription regulations and protein-protein interactions[J]. BMC Systems Biology, 2010, 4: 20.
- [20] Covert MW, Xian N, Chen TJ, et al. Integrating metabolic transcriptional regulatory and signal transduction models in *Escherichia coli*[J]. Bioinformatics, 2008, 24(18): 2044-2050.
- [21] Lee JM, Gianchandani EP, Eddy JA, et al. Dynamic analysis of integrated signaling, metabolic and regulatory networks[J]. PLoS Computational Biology, 2008, 4(5): e1000086.