

动态分子网络的构建与分析*

刘伟** 谢红卫

(国防科学技术大学机电工程与自动化学院自动控制系, 长沙 410073)

摘要 分子网络研究是从全局角度揭示生物系统的结构和功能的重要手段, 现有的网络分析大部分是基于静态网络. 实际上, 在不同的环境条件、组织类型和疾病状态以及生长和分化的过程中, 分子网络时刻都在发生变化. 经过研究人员的努力, 人们已经提出了一些可用于分析分子网络动态的生物信息学方法, 如节点的动态性分类、动态蛋白质复合物的预测、条件特异子网的构建以及网络动态行为的模拟等. 本文综述了动态分子网络的构建与分析方法. 可以预见, 动态网络分析将成为未来网络研究的标准模式.

关键词 分子网络, 条件特异子网, 动态分析

学科分类号 Q61

DOI: 10.3724/SP.J.1206.2013.00055

分子网络是描述复杂生命系统的最直接、最有用的工具之一, 其中节点对应系统中的基因或者蛋白质, 两节点之间的连线则表示分子之间的相互作用. 很多生命活动都涉及多种分子的协同作用, 按照分子网络发挥的主要功能可以分为蛋白质相互作用网络、代谢途径、信号转导网络和基因调控网络等. 针对各种分子网络数据, 人们已经开展了大量的研究工作, 如采用复杂网络理论对生物网络的度分布、聚集系数、小世界特性的研究, 采用子图搜索算法和子图比较算法挖掘生物网络模块, 采用聚类方法挖掘生物网络模块, 多物种中生物网络的比较研究等. 但这些分析方法都是基于静态的分子网络模型, 即假定一对蛋白质能够发生相互作用, 那么在这两个节点之间存在一个连接, 网络的结构和特性不随着时间和条件的改变而改变. 在实际生物系统中, 分子网络时刻都在发生改变, 也正是这种改变才使得生物体能够对外界刺激快速作出响应, 完成各种复杂的生物学功能. 因此, 对分子网络进行动态的分析是揭示生物系统运行规律的关键^[1-3].

按照动态分子网络的研究历史, 可以分为三个阶段. 第一阶段: 在对单个节点属性的分析过程中, 人们发现部分节点展现出很强的动态性, 可依据节点在不同条件下的表达变化情况进行划分, 并且各类蛋白质因其动态性差异具有特定的功能. 第二阶

段: 通过将静态相互作用与基因表达或代谢流量相结合, 提取那些在不同实验条件下呈现活跃状态的节点和相互作用, 构建条件特异的相互作用网络, 如动态的蛋白质复合物、时间特异子网、组织特异子网等. 第三阶段: 以实验方法直接测定不同条件、不同物种以及不同时间对应的相互作用网络, 对网络的动态行为进行分析和模拟. 这些研究的主要目标是: 从表征系统的绝对属性过渡到分析在特定情况下的系统动态响应, 揭示生物系统的内部运行机制. 下面就按照这三个阶段对动态分子网络的研究予以介绍.

1 单个节点的动态属性分析

在分子网络中, 用于描述单个蛋白质拓扑属性的主要指标是连接度(degree), 即与某个节点发生相互作用的其他节点的数目. 网络中存在的少量连接度很高的节点被称为中心(hub)节点. 作为网络中的枢纽, 中心节点在生物的进化和维系相互作用网络的稳定性等方面有着不可替代的作用. 其他用

* 国家自然科学基金资助项目(31000591, 31000587, 31171266).

** 通讯联系人.

Tel: 0731-84573369, E-mail: angel_nudt@126.com

收稿日期: 2013-03-12, 接受日期: 2013-08-14

于描述网络拓扑属性的常用指标, 还有聚集系数 (clustering coefficient)、最短路径长度(shortest path length)和介度(betweenness), 可以衡量网络中节点的模块性、连通性和承载流量. 但这些指标主要针对静态网络进行设计, 很难刻画出网络中节点的动态特性.

为了更好地理解蛋白质相互作用网络和蛋白质复合物的动态组织规律, 人们在相互作用网络的范围内对重要蛋白质的瞬态行为开展了研究工作. 这些研究将蛋白质的基因表达谱与网络拓扑属性相结合, 揭示了一些有趣的发现. 其中最重要的研究是 Han 等^[4]发现, 中心蛋白可以划分成聚会蛋白(party hub)和约会蛋白(date hub)两类. 这两类蛋白在转录表达模式上有显著的差异, 在不同条件下, 聚会蛋白与其相互作用蛋白的转录共表达系数更高, 而约会蛋白与其相互作用蛋白的共表达系数则相对较低. 提示聚会蛋白能够同时与多个蛋白质发生相互作用, 而约会蛋白则在不同的地点和时间与不同的蛋白质发生相互作用. 进一步分析表明, ‘party’型中心蛋白处于功能模块的中心, 而‘date’型中心蛋白处于功能模块之间, 充当模块连接者(linker)的角色. 尽管这些发现受到了 Batada 等^[5-6]的质疑, 但这种划分方法在总体上已被学术界所认可^[7-11], 并且开拓了将蛋白质网络属性与基因表达谱结合研究的道路.

在 Han 等工作的基础上, 很多研究人员对网络中单个节点的动态属性进行了深入的分析. 在酵母蛋白质相互作用网络中, Yu 等^[12]研究了中心蛋白的拓扑属性, 发现‘date’型中心蛋白表现出较高的介度和内部模块性, 而‘party’型中心蛋白则表现出较高的聚集系数和模块间连接性. 类似于聚会蛋白和约会蛋白的划分方法, Taylor 等^[9]提出将中心蛋白分为模块内中心蛋白(intermodular hubs)和模块间中心蛋白(intramodular hubs). 这些研究可以认为是 Han 等研究工作的进一步验证和延伸. 还有一些研究人员对中心蛋白作了进一步细分. 例如, Komurov 等^[10]考察了酵母中各基因在 272 个实验条件下的表达情况, 计算了基因的表达变化方差(expression variance, *EV*). *EV* 越接近于 0, 说明该基因的动态性越弱; 而 *EV* 越接近于 1, 说明该基因的动态性越强. 他们在由 2 315 个基因组成的 5 456 对相互作用的网络中, 比较了各蛋白质与其邻居节点的 *EV* 值, 发现相互作用的蛋白质之间 *EV* 值具有很高的相关性, 说明能够发生相互作用

的蛋白质具有类似的动态特性. 进一步, Komurov 等将中心蛋白分为三类, 提出了‘family’型中心蛋白, 此类蛋白与其邻居节点协同表达组成静态模块, 而‘party’型中心蛋白则与其邻居节点组成动态模块, 静态模块和动态模块各自对应了特定的功能. 最近, Patil 等^[11]结合相互作用蛋白质的基因共表达系数和共表达稳定性, 对分子网络的中心蛋白进行了重新分类. 共表达稳定性能够度量一对蛋白质在本质上是共表达的程度. 根据这两个指标, Patil 等发现了两类中心蛋白: 第一类中心蛋白与其邻居节点间共表达系数和共表达稳定性都较高, 往往位于模块之间; 第二类中心蛋白与其邻居节点间的共表达系数较低但稳定性较高, 往往处于模块内部. 第二类蛋白类似于约会型中心蛋白, 多参与瞬时相互作用.

作为动态分子网络研究的初步尝试, 这些研究工作以中心蛋白作为突破口, 结合基因表达等动态信息将动态节点与静态节点进行区分, 有助于了解蛋白质的功能和分子网络的组织结构. 尽管这些研究工作仅对单个节点的动态属性进行分析, 提出的节点动态性划分方法也多种多样. 但是作为网络中起重要作用的中心蛋白, 这些具有不同动态特性的蛋白质从时间和空间等不同角度影响着整个生命体的活动, 反映了分子网络动态性的特点. 受这些研究工作的启发, 人们开始将大规模分子网络与动态的表达数据相结合, 提取网络中动态性较强的部分并对其属性进行分析.

2 条件特异子网的构建与分析

静态分子网络提供了对于细胞内系统行为的定性描述, 而分子表达数据可以提供分子在不同条件/时间/样本状态下的定量信息, 因此, 将这两种数据源结合起来可用于阐释细胞内系统的动态组织形式. 其基本思路是以静态的相互作用网络为骨架, 结合动态的分子表达数据发现在不同条件下具有明显改变的那部分特异子网, 从而研究系统的动态响应情况. 按照实验条件的不同, 条件特异的子网可分为时间特异(如进化上保守的模块)、空间特异(如依赖于亚细胞定位的蛋白质复合物、组织特异表达的基因)和研究内容相关(如疾病的生物标志物集合)几个大的类别. 下面对这几类条件特异子网的构建与分析方法作以介绍.

2.1 动态蛋白质复合物的发现

将蛋白质相互作用网络划分为网络模块, 对于

从网络角度理解细胞分子机制和结构组成具有重要意义. 目前, 人们已经提出了多种用于发现蛋白质复合物和功能模块划分的方法, 如 G-N^[12]、MCODE^[13]、RNSC^[14]、LCMA^[15]、DPCLUS^[16]、APcluster^[17]、MoNet^[18]、IPCA^[19]、COACH^[20]和SPICi^[21]等. 但传统的划分方法将蛋白质相互作用网络作为一个静态图, 忽略了网络中的动态信息. 实际上, 大部分的蛋白质复合物是动态单元. 一些亚基在特定的时间和亚细胞器中组装成复合物, 当行使完特定的功能, 该复合物就随之解体. 由于现有高通量相互作用数据集中缺乏复合物的瞬态信息, 因此很难通过计算方法研究和预测该复合物的动态行为. 例如, 部分蛋白质在某一时刻参与组成了复合物 A, 下一时刻又参与组成了复合物 B, 现有的基于蛋白质相互作用网络的复合物检测技术无法区分这两个复合物, 只能将它们融合成一个大的复合物 AB. 这严重地影响了蛋白质复合物预测的精度, 也妨碍了人们对细胞组织结构的正确理解^[22].

随着蛋白质相互作用和转录组数据的累计, 整合基因表达谱和蛋白质相互作用网络为动态的蛋白质复合物发现提供了新的途径^[23-30]. Jansen 等^[23]首先将蛋白质相互作用与 mRNA 表达水平相结合, 计算复合物的表达活性水平. Tornow 等^[24]利用超图方法评估了各基因的表达相关性, 构建了共表达基因网络用于发现功能模块. Hegde 等^[25]结合功能连接网络和基因表达数据, 分析了生物系统的动态结构. Luo 等^[26]通过整合转录调控数据、基因表达数据和蛋白质相互作用网络, 在系统生物学水平上对特定类型的蛋白质复合物进行了研究. 最近, Li 等^[29]提出了一种名为 TSN-PCD 的算法, 通过聚类算法从时间序列子网中识别蛋白质复合物. 他们将这种方法与已有的蛋白质复合物发现方法进行比较, 发现相比基于静态相互作用网络的方法, 将基因表达数据与蛋白质相互作用数据相结合的方法能够更加有效地发现蛋白质复合物, 复合物内部的各蛋白质在功能上更加接近. 2013 年, Wang 等^[30]根据表达曲线特征计算基因的动态阈值, 研究细胞循环中多种蛋白质的动态, 并分别基于静态网络和动态网络寻找蛋白质复合物. 他们的研究结果证明, 在敏感度、特异性和准确率上, 基于动态网络的方法都要优于基于静态网络的方法. 进一步, 他们发现在细胞循环过程中, 仅有 23%~45% 的蛋白质在同一个时间点处于激活状态, 说明了蛋白质复合物

具有高度的动态表达性. 这些预测方法为动态复合物的发现提供了重要的手段, 在总体性能上优于基于静态网络的方法, 将它们与实验方法相结合有助于更加准确地发现动态复合物, 有望成为动态复合物识别的主流方法.

2.2 组织特异子网的构建与分析

静态的蛋白质相互作用网络描述了在蛋白质之间可能发生的物理联系, 然而在特定的细胞或组织中, 仅有一部分蛋白质被表达. 理论上, 只有两个基因在一个细胞或组织中同时表达, 在某些条件下它们的产物才有可能发生相互作用. 根据基因在各组织中的表达情况, 可以定义组织特异蛋白和广泛表达的蛋白(看家蛋白质). 实际上, 因为相互作用需要同一组织中两个蛋白质都处于表达状态, 所以相互作用的组织特异性要比单个蛋白质的特异性更强. 基于组织特异蛋白和组织特异相互作用, 可以构建组织特异的生物分子网络.

目前, 已有一些研究人员将通路信息与基因表达数据相结合, 构建了组织特异的生物代谢网络^[31-32]和蛋白质相互作用网络^[33-35], 并比较了组织特异蛋白与其他蛋白在网络拓扑属性上的差别. Dezso 等^[33]测量了 31 个人体组织中的基因表达情况, 识别出 2 374 个看家基因和大量的组织特异基因. 经过功能分析发现, 看家基因在一些至关重要的生命过程中显著富集, 如氧化磷酸化、依赖性泛素蛋白质水解、翻译和能量代谢. 他们发现, 相比网络中所有节点的拓扑属性分布, 看家基因具有更高的连接度和更短的蛋白质间通路距离. 而组织特异的蛋白质则与该组织要实现的功能一致, 同一组织中的特异蛋白通常具有类似的基因表达模式, 相比看家基因更可能成为药物作用靶点. Bossi 等^[34]将一个大规模的蛋白质相互作用网络(包括 10 229 个蛋白质和 80 922 对相互作用)与 79 个人类细胞/组织的基因表达谱数据相结合, 考察了蛋白质的组织特异性与其相互作用数目之间的关系. 结果发现, 相比那些广泛表达的蛋白质, 组织特异性越强的蛋白质其相互作用的数目更少, 更可能是进化上比较年轻的蛋白质. 同时, 他们发现那些广泛表达的蛋白质与组织特异蛋白质之间也存在较多的相互作用. Zhu 等^[35]的研究得到了与前人基本一致的结论, 他们发现, 超过一半的中心蛋白属于广泛表达的单元, 因此, 相比组织特异性蛋白质, 具有广泛表达的蛋白质其连接度更高. 对这些研究进行总结, 可以发现看家基因与组织特异基因在连接模式

和执行的函数上都具有显著差异。通常，看家基因的连接度更高，连接模式更加丰富，它们不仅与看家基因发生相互作用，也与组织特异基因之间存在广泛的连接，用于完成各种组织所需要的重要的生物学功能；组织特异基因的连接度则相对较低，功能上与其对应的组织趋于一致。

将组织特异性网络与通用的静态网络进行比较研究，不仅有助于了解复杂网络的动态组织形式和运行机制，而且对于考察大规模相互作用网络的可靠性具有重要作用^[36-37]。Lopes 等^[36]将来自多个数据库的蛋白质相互作用网络与 84 个组织 / 细胞类型的基因表达数据相结合，构建了组织 / 细胞特异的相互作用子网。结果发现，组织特异子网的规模仅占原静态网络的 1%~25%(在不同的相互作用数据库中，所占比例不尽相同)，而且这些子网中的连接关系相比原网络更加松散。经研究发现，组织特异子网与该组织或细胞要实现的功能密切相关，并且其相互作用的可靠程度更高。这提示人们，由于没有考虑蛋白质的组织特异表达，现有的相互作用数据库中可能存在大量的假阳性，即，这些相互作用能够在体外发生，但在实际生物系统中，它们因不参与同一组织的细胞活动而没有发生相互作用。Schaefer 等^[37]的研究进一步证实了这一观点，他们推荐在相互作用数据库中包含更多的细胞类型、功能和疾病状态信息，以便筛选高可信度的蛋白质相互作用。

实际上，用于检测组织特异蛋白的方法不限于基因芯片数据。部分研究发现，高通量测序技术相比基因芯片方法的敏感度更高，在不同组织中发现的特异蛋白质的数目更多^[38-40]。但由于技术的成熟度和实验成本等问题，高通量测序技术在动态网络构建上的应用暂不如基因芯片数据广泛。

2.3 内容相关子网的识别

从基因表达谱中识别研究内容相关的特异子网具有重要的研究意义和实用价值，可帮助筛选疾病相关的生物标志物以及发现在不同表型之间通路的变化。为了对两组或多组基因芯片数据进行比较，人们已经提出了多种统计方法，如单基因差异分析^[41]、基因集差异分析^[42-43]以及基于聚类方法的基因共表达分析^[44]。这些方法的目标是找到一组具有条件特异性的基因集合，但它们仅考虑了网络节点(基因或蛋白质)在不同条件下的表达水平变化，而没有考虑节点间连接关系的改变。

为了更加精确地构建与样本 / 实验条件密切相关的动态分子网络，研究人员发展了基于图搜索的特异子网识别方法^[45-50]。这类方法的主要任务有两个：一是定义打分函数，用于度量不同条件下网络结构的改变程度；二是设计搜索算法，提取打分最高的条件特异子网。Ideker 等^[45]的工作具有一定的开创性意义，他们将节点变化的 z 值作为打分函数，采用模拟退火法搜索最优子网。Guo 等^[46]对打分函数的设计进行了改进，不再仅针对单个基因的改变，而是利用基因之间相关关系来定义打分函数。但由于最优连接子网的搜索是一个 NP 难问题，现有算法还只能通过启发式或近似方法来寻找条件特异的子网。除了模拟退火之外，很多研究工作采取了局部贪婪搜索算法、数学规划或者图论基础上的搜索算法^[47-49]，例如 Qiu 等^[47]使用了混合整数规划模型(mixed integer linear programming model)。

但是很少有方法同时考虑单个基因的差异表达和基因对之间的变化相关性。实际上，在疾病情况下不仅单个基因的表达水平会发生变化，在基因对之间的相关性也会发生改变，甚至是更高阶的拓扑属性改变。因此，Wang 等^[50]在打分函数中综合考虑了节点和边的改变，利用迭代算法进行局部最优化(图 1)。Ma 等^[51]建立了一种全局优化算法 COSINE，同时考虑网络中节点和边随条件的变化，利用遗传算法提取条件特异的子网。该方法被成功地用于两个仿真数据集和三个真实的芯片数据集，发现了一些具有生物学意义的合适规模的特异子网。经比较发现，相对于传统的单个基因表达差异分析(differential expression)和差异共表达性分析(differential correlation)，COSINE 能够发现更多与病理状态相关的基因和通路。由于问题的复杂性，条件特异子网的构建方法还不够成熟，如何综合考虑节点和边的动态变化，如何排除实验噪声的干扰，如何选择合适的分子表达数据集，都是有待解决的问题。

除此之外，RNA 干扰(RNAi)技术也是一种用于获得生物中扰动数据的有效手段。选择通路中一个感兴趣的基因作为报告基因，利用 RNA 干扰技术系统地敲除其他基因，测量对该报告基因的影响。对于那些对报告基因有影响的基因进行分析，并结合静态相互作用网络，可用于构建条件特异子网。

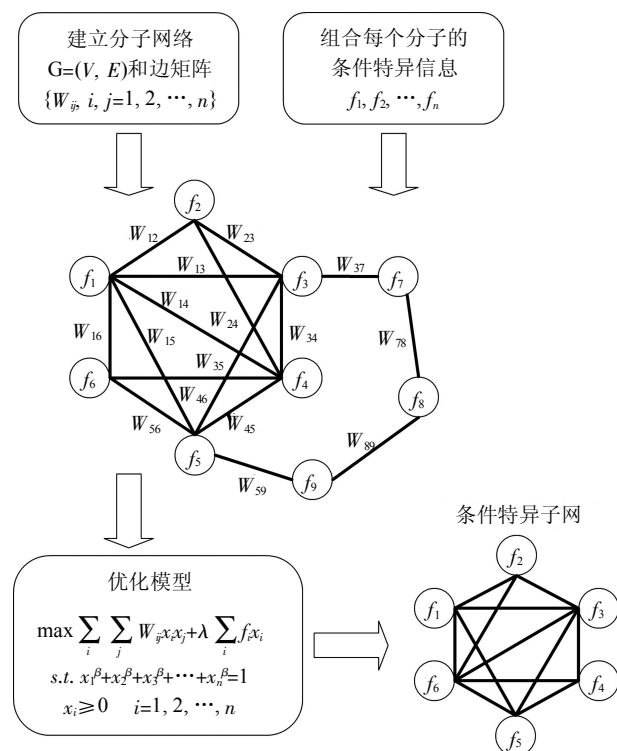


Fig. 1 The scheme of our subnetwork identification method^[50]

图 1 条件特异子网搜索算法^[50]

利用带权重的节点和边连接图表示分子间相互作用, 节点的权重代表其与条件之间的关联程度, 边的权重为相互作用的强度. 然后, 设计优化算法提取那些即紧密连接又具有条件特异性的子网.

通过将静态相互作用与基因表达或代谢流量相结合, 可以增进对于大规模网络动态的理解. 然而, 转录表达水平的变化无法完全地反映出蛋白质水平的变化, 这是因为转录表达和蛋白质表达并非完全相关, 而且对于某些瞬态的蛋白质相互作用, 在转录水平上还很难检测到. 同时, 这些方法以静态相互作用网络为基础, 无法识别那些条件特异的新的相互作用、复合物或者通路, 难以区分在不同条件下是网络状态发生了改变, 还是网络结构发生了改变. 因此, 这些方法还只能间接地描述一些网络的动态变化情况, 如果要更加精细地刻画相互作用网络的动态, 则有赖于新的实验技术和分析手段的发展.

3 网络动态的分析与模拟

动态相互作用是指那些随着时间、地点或条件

改变而发生变化的相互作用. 在网络中, 动态性最强的相互作用不是那些在静态网络中相互作用强度最大的那些, 而是最容易改变的那一部分. 相反的, 在两种条件下都存在的相互作用将会被淡化, 甚至从动态网络中删除. 因此, 动态的相互作用反映了在研究条件下, 哪一部分细胞进程对于细胞响应更为重要, 对于理解细胞功能具有重要意义. 例如, 网络动态可以描述细胞对环境刺激的响应过程, 以及分子网络随着发育或者分化的改变过程. 又如, 模拟疾病状态下的网络动态有助于揭示疾病发生机理, 发现特异的生物标志物和药物作用靶标.

人们对于生物系统的动态研究已经有很多年的历史. 但是, 传统的实验方法往往关注于特定条件下的少数几个基因、蛋白质或者相互作用. 而大部分用于检测蛋白质相互作用的高通量方法, 如酵母双杂交、串联亲和纯化结合质谱技术 (tandem affinity purification-mass spectrometry, TAP-MS) 还只能识别静态相互作用, 不能提供有关相互作用的动态信息. 随着实验技术的发展, 研究人员开始测定不同条件、不同物种或者不同时间对应的相互作用网络, 从蛋白质水平测定蛋白质的表达量和相互作用的强度, 以便更加直接和准确地描述分子网络的动态.

3.1 物理相互作用网络的动态研究

目前, 研究人员已发展了多种实验技术用于发现蛋白质之间动态的物理相互作用, 如荧光共振能量转移 (fluorescence resonance energy transfer, FRET)^[52-53]、蛋白质荧光标记与质谱鉴定技术^[54]、蛋白质片段互补法 (protein-fragment complementation assay)^[55]等. 但由于技术不够成熟、通量有限, 对不同条件下大规模动态相互作用网络的研究还较少.

2005年, Borrios等^[56]利用一种自动化的高通量技术 LUMIER (luminescence-based mammalian interactome mapping), 系统地构建了哺乳动物细胞中与转化生长因子 β (TGF β) 通路相关的动态相互作用网络. 这种方法不能直接给出不同条件下相互作用强度的变化, 但是通过定量测量诱饵 (它们与猎物之间能够发生相互作用) 的浓度变化, 可以估计相互作用的动态变化情况. 2011年, Bisson等^[57]提出了一种新的定量检测相互作用网络动态的质谱方法, 称为亲和纯化选择反应检测 (affinity purification-selected reaction monitoring, AP-SRM). 他们利用这种技术测量了在生长因子刺激下, 以

Grb2 为核心的相互作用网络在 6 个时间点的动态变化过程. Grb2 是一种参与多个蛋白质复合物的转换蛋白, 以其作为研究对象具有一定的代表性. 这种技术通过测量蛋白质复合物中每个肽段的整合峰强度, 计算在单个时间点或条件下相互作用的平均强度, 从而绘制出不同时刻对应的动态相互作用网络. 结果表明, Grb2 复合物的组成与用于刺激的生长因子有显著关联, 即, 在不同的刺激作用下, Grb2 复合物的组成情况发生了明显的改变. 进一步, 通过聚焦除 Grb2 之外的其他中心蛋白, 可以描绘出在生长因子刺激下细胞的整个响应过程. 作为动态分子网络的探索性研究, 这些方法为未来大规模的构建和分析动态相互作用网络提供了指导思路, 即不仅关心在某一条件下相互作用是否发生, 而且要测定相互作用发生的强弱, 以便定量地描述分子网络的动态变化过程.

除蛋白质相互作用研究之外, 也有一些研究人员构建了不同条件下的蛋白质 -DNA 相互作用谱. 利用染色质免疫共沉淀 - 芯片技术(chromatin immunoprecipitation-chip, ChIP-chip), Workman 等^[58]研究了在 DNA 损伤时酵母中的动态转录调控网络, 发现了 30 种不同的转录因子与基因的相互作用. 跨物种的转录因子研究则揭示, 蛋白质 -DNA 相互作用随着进化时间发生了快速的演化^[59].

3.2 遗传相互作用网络的动态研究

蛋白质通过相互作用来执行功能, 但是在某些信号通路或生物进程中, 关联的蛋白并不一定在物理学上发生相互作用, 这时可通过了解遗传水平的功能相互作用, 阐释基因的功能以及由基因型到表型的翻译过程. 利用遗传相互作用作图方法, 研究人员已经成功地建立了一些跨物种模型, 如比较多个萌芽和裂殖酵母在网络结构上的区别^[60]. 同时, 将遗传相互作用网络与蛋白质相互作用数据相结合, 能够帮助认识真核生物中相互作用结构的保守性^[61-63].

部分研究表明, 相比蛋白质复合物, 遗传相互作用的动态性更强^[60-63]. 也就是说, 在条件变化的情况下, 某些蛋白质复合物的组成没有改变, 但是各基因之间的功能关联却发生了明显的变化. 如 Roguev 等^[60]研究发现, 蛋白质复合物在不同的酵母中是高度保守的, 但是不同蛋白质复合物之间的遗传相互作用发生了显著改变. 进一步, 随着实验方法的进步, 人们发展了一种能够定量地测量遗传相互作用的实验技术, 即上位性微阵列剖面技术

(epistasis miniarray profile, E-MAP), 以测定相互作用的正负(正相互作用表示增强, 负相互作用表示减弱)和相互作用强度信息. 目前, 该技术已发展成动态的上位性微阵列剖面技术(differential epistasis miniarray profile, dE-MAP), 用于检测不同条件下基因相互作用的强度变化(图 2). 利用 dE-MAP, Bandyopadhyay 等^[61]检测了在 DNA 发生损伤时, 细胞内遗传相互作用网络的改变. 他们针对酵母中 418 个信号基因和调控基因, 共进行了 8 万多次双基因敲除实验. 结果发现, 在 DNA 损伤对应的相互作用网络中, 有 53% 的相互作用在静态(即标准实验状态, 没有损伤的情况)是无法检测到的, 说明遗传相互作用网络为了响应外界刺激发生了巨大的改变. 该研究也同样验证了蛋白质复合物的相对稳定性和遗传相互作用的重组特性. 对于物理相互作用网络(蛋白质 - 蛋白质相互作用、蛋白质复合物、蛋白质 -DNA 相互作用), 动态的相互作用意味着作用机制的改变; 而对于遗传相互作用网络, 动态的相互作用反映了突变对于功能关联的影响, 而非物理机制的改变. 这就提示人们, 为了响应外界刺激, 生物系统一方面在分子之间的相互作用模式上作出改变, 但更多的是在功能关系上进行了调整, 以尽可能少的代价来实现特定的生物学功能.

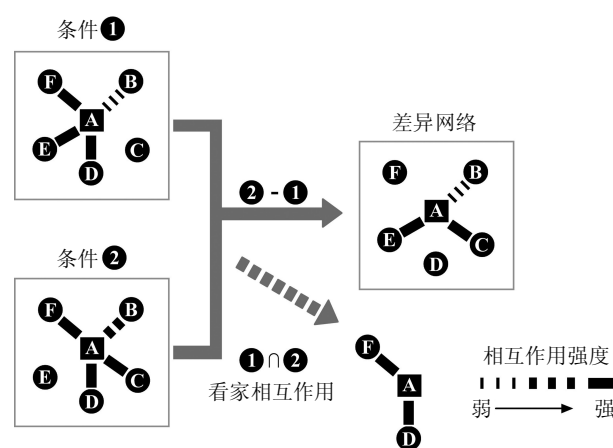


Fig. 2 Schematic showing principle of differential genetic interaction analysis^[61]

图 2 遗传相互作用的差异性分析示意图^[61]

在两种实验条件下, 采用 E-MAP 技术分别测量各遗传相互作用的连接关系和强度, 构建条件特异的相互作用网络. 将两个条件子网进行比较, 两个网络中共有的部分是看家相互作用, 而网络中不同的部分则代表了网络在外界刺激下发生的改变.

类似于看家基因(在不同条件下具有稳定的表达), 在不同条件下都能发生的相互作用被称为看家相互作用. 实际上, 看家基因不一定会导致看家相互作用的发生, 因为看家基因的相互作用模式相对比较丰富, 它们既与其他看家基因存在关联而导致看家相互作用的发生, 也与动态基因结合而导致动态相互作用的发生. 在静态相互作用网络的研究中, 人们关注于那些相对稳定的看家相互作用. 而在动态相互作用网络的研究中, 人们更关注那些随条件变化而发生改变的相互作用, 以便了解细胞对外界刺激的动态响应过程.

3.3 网络动态的建模与仿真

计算方法提供了一种用于推断和分析分子网络动态的有效手段. 其主要目的是建立生物过程的数学模型, 用于描述和预测生物系统在响应外界环境变化和内在遗传结构变化时的动态行为, 获得对于生物系统更加全面的认识. 相比实验测量技术, 计算方法对于动态网络分析具有一定优势. 例如蛋白质在一定条件下发生结构变化时, 通过测量技术只能获得变化前后两个稳态的数据, 而建立相应的模型可以了解到变化中间状态的全过程.

为了从计算角度描述相互作用网络的动态, 人们已开展了大量的研究工作, 主要包括: 发现网络中的动态通路信息、利用动态数据推断网络结构、识别网络拓扑和功能的大尺度变化以及理解网络对特定扰动(如 knock-outs 和 knock-downs)的响应. 同时, 人们也提出了多种用于模拟分子和网络动态的数学模型, 如布尔模型、逻辑模型、贝叶斯模型、Petri 网、随机模型和微分方程模型等^[64]. 每种模型都有一定的适用范围: 对于单个基因, 可以在分子细节上利用随机模拟的方法进行建模; 对于小规模基因开关回路, 利用微分方程模型表示基因的动态则更加合适; 对于中等规模的基因网络, 通过开/关转换行为近似基因动力学可用于模拟整个生物系统的动力学行为; 对于包含上千个基因的大规模网络, 还难以进行预测性模拟. 但利用简化动力学模型, 例如网络结构的流量分析, 可以帮助理解大规模网络的功能组成^[65]. 由于精确的动态模型需要已知描述系统的实验参数, 因此还仅适用于小规模分子网络, 无法用于基因组尺度的动态网络建模. 对于大规模网络的动态分析, 必须借助于一些不需要反应参数的不太精确的模型, 例如用于代谢网络分析的流量平衡模型, 常用于信号网络的布

尔逻辑模型等. 其中针对代谢网络的动态模拟发展得较为成熟, 如代谢平衡分析基于化学计量模型, 通过有约束的目标函数求解代谢流量分布, 能够系统地预测和估算出遗传以及环境影响给细胞带来的扰动^[66]. 进一步, 面对细胞内复杂的基因、蛋白质和代谢物网络, 普通微分方程模型常因为网络的复杂性和缺乏实验动力学参数而无法适用, 也可考虑基于人工智能的建模技术, 如细胞自动建模 (cellular automata)^[67] 和智能建模 (agent-based modelling)^[68].

大部分的分子网络建模方法都是针对单一网络类型, 然而, 在细胞内部各种网络是相互关联的, 任何一种网络的动态都会对其他网络的行为造成影响^[69]. 最近, 一些研究人员开始尝试将不同类型的网络整合起来进行建模和仿真^[70-74]. 例如, 为了获得代谢网络和调控网络的整合模型, Covert 等^[70-71] 利用流量分析模型来建模代谢网络并利用布尔网络来建模基因调控网络. Shlomi 等^[72] 利用整数规划方法(一种能够解决同时包含离散和连续参数问题的通用优化算法)来建立代谢网络和调控网络的混合模型. Wang 等^[73] 利用动态的基因表达数据建立了转录调控和蛋白质相互作用的混合动态模型(图 3), 其中, 目标基因的表达被认为是转录因子、该基因前一时刻的表达值和 mRNA 降解速率的函数, 而两个蛋白质发生相互作用的速率与它们分子浓度的乘积成正比. Buescher 等^[74] 结合转录物、蛋白质、代谢物丰度和启动子动态信息, 构建了枯草芽胞杆菌 (*Bacillus subtilis*) 中的整体代谢与调控网络, 研究细胞在应对环境变化时的响应机制. 这些方法为混合系统建模提供了一些解决方案, 但是将三种网络(代谢网络、转录调控、信号转导网络)进行整合的方法还相对缺乏, 而且对于基因组尺度的动态模型构建也存在较大难度.

可以发现, 现有的建模方法大部分是针对网络中分子的动态进行模拟, 即描述在不同时间/地点/条件下分子浓度的变化情况, 很少有研究能够模拟相互作用强度的变化过程. 分析原因, 一方面已测定的相互作用强度信息较少, 另外相互作用强度的变化过程及其发生机制尚不清楚. 可以猜想, 其不仅与单个分子的浓度有关, 而且与分子间的构象变化具有很大关联. 作为动态网络分析的一个重要方面, 对于分子间相互作用过程的精确模拟有可能成为动态网络研究中新的热点.

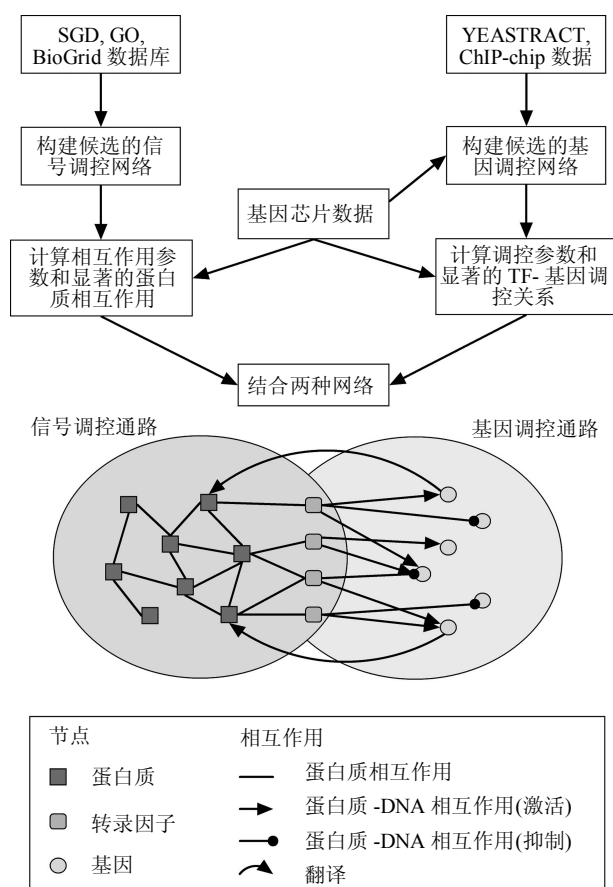


Fig. 3 Integrating transcriptional and signaling networks^[73]

图3 整合的基因调控与信号转导模型^[73]

首先, 提取候选的基因调控网络和信号调控网络, 然后结合时序的基因表达谱数据对两类网络进行推断, 最后得到与条件相关的整合网络. 转录因子作为媒介连接两种类型的网络.

4 结论与展望

本文综述了动态分子网络研究的现状和最新进展. 可以发现, 随着外界环境和内在条件的变化, 分子网络展现出了很强的动态性, 一方面体现在分子本身的浓度和位置变化, 另一方面体现在分子间相互作用关系的改变, 这不仅包括物理的蛋白质-蛋白质相互作用、蛋白质-DNA 相互作用, 也包括遗传上的蛋白质相互作用. 为了描述分子网络的动态特性, 研究人员进行了大量的工作, 通过将蛋白质相互作用网络与分子表达数据相结合以及对动态相互作用的测量, 构建了部分动态分子网络并对其进行了分析, 取得了一定的研究成果. 与静态网络相比, 条件特异子网包含的相互作用数目更少、可靠性更高、对应的功能与实验条件更加密切相关.

这些发现对于更加准确地识别蛋白质复合物、了解系统内部组织形式以及刺激响应过程都具有重要意义.

尽管如此, 动态网络的分析方法还不是很多, 这是由于相比静态网络, 研究动态的分子网络难度要大得多. 首先, 有些动态变化是缓慢发生的, 如细胞的生长和分化以及进化上的改变, 而有些则是瞬态现象, 现有的观测手段还只能从分子表达或蛋白质水平选择特定的时间间隔进行定量测量, 难以完全地表征分子的动态变化过程; 其次, 由分子表达数据来构建动态分子网络涉及复杂的算法设计, 一方面要考虑分子本身的变化, 另一方面要考虑分子间作用关系的改变, 而分子表达数据作为一种间接的动态信息, 想从中完全地还原出网络的结构和动态就成为一项几乎不可能完成的任务; 最后, 由于在细胞内部, 基因调控、信号转导和代谢途径没有明显的区分, 目前还难以建立统一的计算模型进行模拟分析, 而且对于分子相互作用的机制也没有完全揭示. 因此, 现有的研究工作还主要集中于动态子网的搜索和建立, 只有构建出准确可靠的动态网络, 在此基础上进行的分析和模拟工作才能更有意义. 可以预见, 随着实验手段的进步和研究的深入, 在条件特异性子网搜索算法、大规模动态相互作用的测量以及相互作用过程模拟等方面将可能取得重大突破. 与静态网络相比, 研究动态的分子网络将会带来全新的视野. 分子网络动态的分析与模拟对于了解生物体内的刺激响应过程, 生命体随着时间、地点和状态改变的分子作用机制, 甚至是生命的整个演化过程都将发挥重要的作用. 作为未来分子网络研究的发展方向, 分子网络的动态研究势在必行.

参 考 文 献

- [1] Ideker T, Krogan N J. Differential network biology. *Molecular Systems Biology*, 2012, **8**: 565
- [2] Przytycka T M, Singh M, Slonim D K. Toward the dynamic interactome: it's about time. *Brief Bioinformatics*, 2010, **11** (1): 15-29
- [3] Koyutürk M. Algorithmic and analytical methods in network biology. *Wiley Interdiscip Rev Syst Biol Med*, 2010, **2**(3): 277-292
- [4] Han J D, Bertin N, Hao T, *et al.* Evidence for dynamically organized modularity in the yeast protein interaction network. *Nature*, 2004, **430**(6995): 88-93
- [5] Batada N, Hurst L D, Tyers M. Evolutionary and physiological importance of hub proteins. *PLoS Computational Biology*, 2006, **2**(7): e88

- [6] Batada N, Reguly T, Breitkreutz A, *et al.* Still stratus not altocumulus: further evidence against the date/party hub distinction. *PLoS Computational Biology*, 2007, **5**(6): e154
- [7] Pereira-Leal J B, Levy E D, Teichmann S A. The origins and evolution of functional modules: lessons from protein complexes. *Phil Trans R Soc B*, 2006, **361**(1467): 507–517
- [8] Yu H, Kim P M, Sprecher E, *et al.* The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Computational Biology*, 2007, **3**(4): e59
- [9] Taylor I W, Linding R, Warde-Farley, *et al.* Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol*, 2009, **27**: 199–204
- [10] Komurov K, White M. Revealing static and dynamic modular architecture of the eukaryotic protein interaction network. *Molecular Systems Biology*, 2007, **3**(1): 110
- [11] Patil A, Nakai K, Kinoshita K. Assessing the utility of gene co-expression stability in combination with correlation in the analysis of protein-protein interaction networks. *BMC Genomics*, 2011, **12**(3): S19
- [12] Girvan M, Newman M E. Community structure in social and biological networks. *Proc Natl Acad Sci USA*, 2002, **99**(12): 7821–7826
- [13] Bader G D, Hogue C W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 2003, **4**: 2
- [14] King A D, Przulj N, Jurisica I. Protein complex prediction *via* cost-based clustering. *Bioinformatics*, 2004, **20**(17): 3013–3020
- [15] Li X L, Tan S, Foo C, *et al.* Interaction graph mining for protein complexes using local clique merging. *Genome Informatics*, 2005, **16**(2): 260–269
- [16] Altaf-Ul-Amin M, Shinbo Y, Mihara K, *et al.* Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC Bioinformatics*, 2006, **7**: 207–219
- [17] Frey B J, Dueck D. Clustering by passing messages between data points. *Science*, 2007, **315**(5814): 972–977
- [18] Luo F, Yang Y, Chen C F, *et al.* Modular organization of protein interaction networks. *Bioinformatics*, 2007, **23**(2): 207–214
- [19] Li M, Chen J, Wang J, *et al.* Modifying the DPCLUS algorithm for identifying protein complexes based on new topological structures. *BMC Bioinformatics*, 2008, **9**: 398
- [20] Wu M, Li X L, Kwok C, *et al.* A core-attachment based method to detect protein complexes in PPI networks. *BMC Bioinformatics*, 2009, **10**: 169
- [21] Peng J, Singh M. SPICi: a fast clustering algorithm for large biological networks. *Bioinformatics*, 2010, **26**(8): 1105–1111
- [22] Srihari S, Leong H W. Temporal dynamics of protein complexes in PPI Networks: a case study using yeast cell cycle dynamics. *BMC Bioinformatics*, 2012, **13**(Suppl 17): S16
- [23] Jansen R, Greenbaum D, Gerstein M. Relating whole-genome expression data with protein-protein interactions. *Genome Res*, 2002, **12**(1): 37–46
- [24] Tornow S, Mewes H W. Functional modules by relating protein interaction networks and gene expression. *Nucleic Acids Res*, 2003, **31**(21): 6283–6289
- [25] Hegde S R, Manimaran P, Mande S C. Dynamic changes in protein functional linkage networks revealed by integration with gene expression data. *PLoS Comput Biol*, 2008, **4**(11): e1000237
- [26] Luo F, Liu J, Li J. Discovering conditional co-regulated protein complexes by integrating diverse data source. *BMC Syst Biol*, 2010, **4**(Suppl 2): S4
- [27] Wang J, Li M, Chen J, *et al.* A fast hierarchical clustering algorithm for functional modules discovery in protein interaction networks. *IEEE/ACM Trans Comput Biol Bioinf*, 2011, **8**(3): 607–620
- [28] Tang X, Wang J, Liu B, *et al.* A comparison of the functional modules identified from time course and static PPI network data. *BMC bioinformatics*, 2011, **12**: 339
- [29] Li M, Wu X, Wang J. Towards the identification of protein complexes and functional modules by integrating PPI network and gene expression data. *BMC Bioinformatics*, 2012, **13**(1): 109
- [30] Wang J, Peng X, Li M, *et al.* Construction and application of dynamic protein interaction network based on time course gene expression data. *Proteomics*, 2013, **13**(2): 301–312
- [31] Sauer U. High-throughput phenomics: experimental methods for mapping fluxomes. *Curr Opin Biotechnol*, 2004, **15**(1): 58–63
- [32] Shlomi T, Cabili M N, Hergard M J, *et al.* Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol*, 2008, **26**(9): 1003–1010
- [33] Dezso Z, Nikolsky Y, Sviridov E, *et al.* A comprehensive functional analysis of tissue specificity of human gene expression. *BMC Biology*, 2008, **6**: 49
- [34] Bossi A, Lehner B. Tissue specificity and the human protein interaction network. *Mol Syst Biol*, 2009, **5**: 260
- [35] Zhu W, Yang L, Du Z. MicroRNA regulation and tissue-specific protein interaction network. *PLoS ONE*, 2011, **6**(9): e25394
- [36] Lopes T J, Schaefer M, Shoemaker J, *et al.* Tissue-specific subnetworks and characteristics of publicly available human protein interaction databases. *Bioinformatics*, 2011, **27**(17): 2414–2421
- [37] Schaefer M H, Lopes T J S, Mah N, *et al.* Adding protein context to the human protein-protein interaction network to reveal meaningful interactions. *PLoS Comput Biol*, 2013, **9**(1): e1002860
- [38] Lee J H, Park I H, Gao Y, *et al.* A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. *PLoS Genet*, 2009, **5**(11): e1000718
- [39] Emig D, Kacprowski T, Albrecht M. Measuring and analyzing tissue specificity of human genes and protein complexes. *EURASIP J Bioinform Syst Biol*, 2011, **2011**(1): 5
- [40] Emig D, Albrecht M. Tissue-specific proteins and functional implications. *J Proteome Res*, 2011, **10**(4): 1893–1903
- [41] Huang D W, Sherman B T, Lempicki R A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 2009, **4**(1): 44–57
- [42] Subramanian A, Tamayo P, Mootha V K, *et al.* Gene set enrichment

- analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*, 2005, **102**(43): 15545–15550
- [43] Ackermann M, Strimmer K. A general modular framework for gene set enrichment analysis. *BMC Bioinformatics*, 2009, **10**: 47
- [44] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 2008, **9**: 559
- [45] Ideker T, Ozier O, Schwikowski B, *et al.* Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, 2002, **18**(Suppl 1): S233–S240
- [46] Guo Z, Li Y, Gong X, *et al.* Edge-based scoring and searching method for identifying condition-responsive protein-protein interaction sub-network. *Bioinformatics*, 2007, **23**(16): 2121–2128
- [47] Qiu Y Q, Zhang S H, Zhang X S, *et al.* Identifying differentially expressed pathways *via* a mixed integer linear programming model. *IET Syst Biol*, 2009, **3**(6): 475–486
- [48] Qiu Y Q, Zhang S H, Zhang X S, *et al.* Detecting disease associated modules and prioritizing active genes based on high throughput data. *BMC Bioinformatics*, 2010, **11**: 26
- [49] De Lichtenberg U, Jensen L J, Brunak S, *et al.* Dynamic complex formation during the yeast cell cycle. *Science*, 2005, **307**(5710): 724–727
- [50] Wang Y, Xia Y. Condition specific subnetwork identification using an optimization model. *Proc Optim Syst Biol*, 2008, **9**: 333–340
- [51] Ma H S, Schadt E E, Kaplan L M, *et al.* COSINE: COndition-Specific sub-NEtwork identification using a global optimization method. *Bioinformatics*, 2011, **27**(9): 1290–1298
- [52] Piston D W, Kremers G J. Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem Sci*, 2007, **32**(9): 407–414
- [53] Kentner D, Sourjik V. Dynamic map of protein interactions in the *Escherichia coli* chemotaxis pathway. *Mol Syst Biol*, 2009, **5**: 238
- [54] Cristea I M, Carroll J W, Rout M P, *et al.* Tracking and elucidating alphavirus-host protein interactions. *J Biol Chem*, 2006, **281**(40): 30269–30278
- [55] Tarassov K, Messier V, Landry C R, *et al.* An *in vivo* map of the yeast protein interactome. *Science*, 2008, **320**(5882): 1465–1470
- [56] Barrios-Rodiles M, Brown K R, Ozdamar B, *et al.* High-throughput mapping of a dynamic signaling network in mammalian cells. *Science*, 2005, **307**(5715): 1621–1625
- [57] Bisson N, James D A, Ivosev G, *et al.* Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nat Biotechnol*, 2011, **29**(7): 653–658
- [58] Workman C T, Mak H C, McCuine S, *et al.* A systems approach to mapping DNA damage response pathways. *Science*, 2006, **312**(5776): 1054–1059
- [59] Schmidt D, Wilson M D, Ballester B, *et al.* Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*, 2010, **328**(5981): 1036–1040
- [60] Roguev A, Bandyopadhyay S, Zofall M, *et al.* Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science*, 2008, **322**(5900): 405–410
- [61] Bandyopadhyay S, Mehta M, Kuo D, *et al.* Rewiring of genetic networks in response to DNA damage. *Science*, 2010, **330**(6009): 1385–1389
- [62] Beltrao P, Cagney G, Krogan N J. Quantitative genetic interactions reveal biological modularity. *Cell*, 2010, **141**(5): 739–745
- [63] Ryan C J, Roguev A, Patrick K, *et al.* Hierarchical modularity and the evolution of genetic interactomes across species. *Mol Cell*, 2012, **46**(5): 691–704
- [64] Gitter A, Lu Y, Bar-Joseph Z. Computational methods for analyzing dynamic regulatory networks. *Methods Mol Biol*, 2010, **674**: 419–441
- [65] Bornholdt S. Systems biology. Less is more in modeling large genetic networks. *Science*, 2005, **310**(5747): 449–451
- [66] Orth J D, Thiele I, Palsson B O. What is flux balance analysis?. *Nature Biotechnology*, 2010, **28**: 245–248
- [67] Bonchev D, Thomas S, Apte A, *et al.* Cellular automata modelling of biomolecular networks dynamics. SAR and QSAR in Environmental Research, 2010, **21**(1–2): 77–102
- [68] Hinkelmann F, Brandon M, Guang B, *et al.* ADAM: analysis of discrete models of biological systems using computer algebra. *BMC Bioinformatics*, 2011, **12**: 295
- [69] Przytycka T M, Kim Y A. Network integration meets network dynamics. *BMC Biology*, 2010, **8**: 48
- [70] Covert M W, Schilling C H, Palsson B O. Regulation of gene expression in flux balance models of metabolism. *J Theor Biol*, 2001, **213**(1): 73–88
- [71] Covert M W, Palsson B O. Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J Theor Biol*, 2003, **221**(3): 309–325
- [72] Shlomi T, Eisenberg Y, Sharan R, *et al.* A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol*, 2007, **3**: 101
- [73] Wang Y C, Chen B S. Integrated cellular network of transcription regulations and protein-protein interactions. *BMC Syst Biol*, 2010, **4**: 20
- [74] Buescher J M, Liebermeister W, Jules M, *et al.* Global network reorganization during dynamic adaptations of *Bacillus subtilis* metabolism. *Science*, 2012, **335**(6072): 1099–1103

Construction and Analysis of Dynamic Molecular Networks*

LIU Wei, XIE Hong-Wei**

(*Department of Automatic Control, College of Mechanical & Electronic Engineering and Automatization,
National University of Defense Technology, Changsha 410073, China*)

Abstract Molecular network research is an important means to reveal the structure and function of biological systems from a global point of view. The existing analysis methods are mostly based on the static network. In fact, the biological networks change all the time with the environmental conditions, tissue types, disease status, growth and differentiation. Researchers have proposed some bioinformatics methods for the dynamic analysis of molecular networks, such as the dynamic classification of the hub nodes, the prediction of dynamic protein complexes, the construction of condition specific subnetworks, and the simulation of dynamic behavior of networks. In this paper, we reviewed these methods for the dynamic molecular network analysis. Predictable, dynamic network analysis will become the standard mode of the future network research.

Key words molecular network, condition specific subnetwork, dynamics analysis

DOI: 10.3724/SP.J.1206.2013.00055

* This work was supported by grants from The National Natural Science Foundation of China (31000591, 31000587, 31171266).

**Corresponding author.

Tel: 86-731-84573369, E-mail: angel_nudt@126.com

Received: March 12, 2013 Accepted: August 14, 2013