PIBB 生物化学与生物物理进展
Progress in Biochemistry and Biophysics
2009, 36(1): 13~24
www.pibb.ac.cn

蛋白质相互作用网络进化分析研究进展 *

刘中扬[1,2] 李 栋[1,2] 朱云平[1,2)**] 贺福初[1,2)**]
([1]蛋白质组学国家重点实验室，北京蛋白质组研究中心，北京 102206；[2]北京放射医学研究所，北京 100850)

摘要 近年来，随着高通量实验技术的发展和广泛应用，越来越多可利用的蛋白质相互作用网络数据开始出现．这些数据为进化研究提供了新的视角．从蛋白质、蛋白质相互作用、模体、模块直到整个网络五个层次，综述了近年来蛋白质相互作用网络进化研究领域的主要进展，侧重于探讨蛋白质相互作用、模体、模块直到整个网络对蛋白质进化的约束作用，以及蛋白质相互作用网络不同于随机网络特性的起源和进化等问题．总结了前人工作给学术界的启示，探讨了该领域未来可能的发展方向．

关键词 蛋白质相互作用网络，进化，生物信息学
学科分类号 Q51，Q11
DOI: 10.3724/SP.J.1206.2008.00393

正像 Dobzhansky 的著名观点所述[1]，“除了进化的光亮外，没有其他东西能使你明白生物学．”自 20 世纪中叶沃森和克里克提出 DNA 双螺旋模型以来，分子生物学的迅速发展使人们能够从复杂细胞表型下的微观分子层次探索生物进化过程．1990 年启动人类基因组计划，海量基因组数据涌现，生物信息学应运而生．世纪之交，人类基因组计划宣告完成，后基因组时代到来．与此同时，随着高通量实验技术的发展，转录组、蛋白质组、代谢组以及相互作用组等组学数据也开始大量出现，系统生物学被适时提出．系统生物学的目的在于从系统水平来理解生物学系统，即在细胞、组织、器官和生物体整体水平上研究结构和功能各异的生物大分子及其相互作用，并通过生物信息学和计算生物学来定量阐明并预测生物功能、表型和行为[2]．系统生物学的产生和发展，为进化研究提供了新的视角．

近年来，随着酵母双杂交、基于质谱的串联亲和纯化等高通量实验技术的发展和生物信息学在蛋白质相互作用预测领域的广泛应用，人们得到了越来越多可利用的蛋白质相互作用网络数据．作为系统生物学的一个重要研究内容，蛋白质相互作用网络被越来越多地用于探讨生物学系统的进化问题．这些研究工作涉及了蛋白质相互作用网络从小到大的所有五个层次：蛋白质个体、蛋白质之间的相互作用、模体(motif)、模块(module)直到整个网络．其中，模体和模块都是由多个蛋白质及其相互作用组成的．前者一般包含 3～5 个蛋白质，而后者更大一些，它可能包含模体作为其结构成分．本文试图从以上五个层次对近年来出现的蛋白质相互作用网络进化研究工作做一综述，并对相关问题展开讨论．侧重于探讨两个基本问题：a．蛋白质相互作用、模体、模块直到整个网络对蛋白质进化有无约束作用，如果有，又是怎样的约束作用? 从进化的角度，相互作用蛋白质或模体、模块成员蛋白质是否趋向具有相同的性质? b．为什么蛋白质相互作用网络存在不同于随机网络的无尺度、层次模块化等特性，这些特性是如何起源和进化的?

另外，目前可被称为蛋白质相互作用的蛋白质间关系至少有两种．第一种就是被研究最多的直接物理相互作用．第二种是间接相互作用，也即功能关联，比如 STRING 库就收录了这种间接相互作用[3]．本文的综述范围主要集中在第一种．

* 国家重点基础研究发展计划(973)(2006CB910803, 2006CB910700)，国家高技术研究发展计划(863)(2006AA02A312)和国家自然科学基金(30621063)资助项目．
** 通讯联系人．
朱云平. Tel: 010-80705225, E-mail: zhuyp@hupo.org.cn
贺福初. Tel: 010-66931246, E-mail: hefc@nic.bmi.ac.cn
收稿日期：2008-08-07，接受日期：2008-10-24

1 蛋白质相互作用网络和蛋白质进化

研究蛋白质相互作用网络进化，在蛋白质层次上，主要是研究蛋白质相互作用网络对蛋白质个体进化性质的影响．其中，颇具争议并且被研究最多的问题便是探讨蛋白质相互作用是否会减慢蛋白质进化速率．

早在2002年，通过评估酵母同线虫中的直系同源蛋白的进化距离，Fraser等[4]发现，在酵母蛋白质相互作用网络中，蛋白连接度(与该蛋白质发生相互作用的蛋白质数目)与其进化速率呈显著负相关．他们进一步指出，那些具有更多相互作用对象的蛋白质之所以进化更慢是因为它们承受了更强烈的选择压力．但这些蛋白质之所以承受更强烈的选择压力不是因为它们对生物体更重要，而是因为它们的氨基酸序列中有更多的位点参与了更多的相互作用．在随后几年，蛋白连接度同其进化速率的关系受到了学术界的广泛关注．2003年，使用另一种进化距离计算方法，利用在进化上更近的参考物种和可信度更高的相互作用数据集，Jordan等[5]发现，蛋白连接度同其进化速率之间仅呈十分微弱的负相关．他们认为，这种负相关仅仅是由于一小部分具有最高连接度的蛋白质(hub蛋白)进化速率慢于其他蛋白质导致的．他们进一步推测，最高连接度蛋白质进化速率慢于其他蛋白质，可能是蛋白质可移除性(dispensability)对进化速率的间接影响．同年，Jordan等的观点就遭到了Fraser等的反驳，Fraser等[6]认为，之所以没有检测到明显的负相关关系，是因为使用了过小的相互作用数据集．当使用一个大得多的蛋白质相互作用数据集和更完备的直系同源数据时，蛋白连接度同其进化速率之间确实呈显著的单调负相关．他们强调，这种关系既适用于高连接度蛋白质也适用于低连接度蛋白质．

2003年底，Bloom等[7]将蛋白质表达水平引入到这个争论的课题中．早在2001年，就有研究发现，酵母中存在高表达基因进化较慢的现象[8]．另外，一些高通量等实验技术，如酵母双杂交、质谱等，存在的弊端之一是它们偏向于检测到那些高丰度蛋白之间的相互作用[7]．通过对多个高通量酵母蛋白质相互作用数据集进行分别考察，Bloom等发现，这种进化速率同相互作用数目之间的负相关关系是直接随着数据集包含高丰度蛋白的偏性而变化的．以没有丰度偏性的相互作用数据集为考察对象，作者并不能得出两者负相关的结论．这也就意味着蛋白连接度同其进化速率之间表现出来的负相关关系是由蛋白质表达水平同其进化速率之间的关系导致的．2004年，Fraser等[9]对此又提出了反对意见．他们首先肯定了Bloom等的观点——蛋白质表达水平同进化速率的关系显著增强了连接度同进化速率的负相关关系．而后在控制了蛋白质表达水平这一因素后，通过重新分析一个质谱相互作用数据集，Fraser等仍然坚持蛋白连接度和进化速率确实呈负相关的结论．不到两个月后，同一杂志又发表了Bloom等的反驳文章[10]．Bloom等的理由是这次Fraser等选择的重新分析的数据集正好是包含高表达蛋白质相互作用偏性最明显的一个，并且他们使用了不恰当的“去偏性”的方法和统计分析方法．不过Bloom等也认为，在数据集的噪声和偏性的背后，确实可能存在着蛋白连接度同进化速率之间的一个真正的关系，但到目前为止，蛋白质进化速率同连接度之间不存在关系还没有被令人信服地反驳．两年来的争论给学术界的告诫是：今后研究蛋白质进化速率的影响因素时，蛋白质表达水平应该成为必须控制的因素之一．

由于发表时间较近，2004年Wuchty[11]发表的文章没有考虑控制蛋白质表达水平这个因素．他们针对这样一个问题——蛋白连接度同其进化速率的关系强烈地依赖于蛋白质相互作用数据集的质量和直系同源蛋白数据的质量[6]，发展了新的方法——Excess Retention(ER)来衡量蛋白质的保守性．结果发现，蛋白质必要性和连接度与ER的相关性要比与进化速率的相关性高得多，并且这个结果对数据质量不敏感．

2005年，通过对高质量的酵母和线虫相互作用数据集分别作分析，Agrafioti等[12]再次肯定了Bloom等关于蛋白质表达水平在蛋白连接度同其进化速率负相关中起重要作用的观点．他们又进一步将蛋白质功能、参与的生物学过程和细胞组分等因素引入了分析．通过利用AIC(akaike information criterion)准则衡量所考虑的几个因素对进化速率提供的信息量得出结论：蛋白质表达水平、蛋白质功能以及其参与的生物学过程才是影响蛋白质进化的主要因素．而蛋白质在相互作用网络中的位置和连接度可能只对蛋白质进化起到“调节作用”，因为它们只对蛋白质进化速率起到十分微弱的影响．同年，学术界又出现了一篇研究蛋白质于网络中所处的位置与其进化速率关系的文章[13]．其作者认为，在酵母、线虫和果蝇三个真核物种的蛋白质相互作

用网络中，蛋白质中心性越高，其进化越慢，成为生存必需的蛋白质的可能性越高．并且在控制蛋白质表达水平后，此结论仍然成立．而蛋白连接度与进化速率和成为生存必需的蛋白质的可能性就没有这样的关系．其中，蛋白质中心性由介数来衡量．介数定义为网络中经过该节点的最短路径的数量．Drummond 等[14]的工作则更进一步，他们联合分析了包括基因表达水平、基因可移除性、连接度、中心性、蛋白质丰度、密码子适应指数(codon adaptation index，CAI)、基因长度等 7 个可能影响酵母蛋白进化速率的因素．考虑到这 7 个因素之间并不互相独立，因而不适于使用一般回归分析．Drummond 等改用了主成分回归分析，结果发现，存在一个对蛋白质进化速率起决定性影响的主成分．该主成分几乎完全决定于基因表达水平、蛋白质丰度和密码子适应指数，而这 3 个因素都与翻译事件的量相关，所以文章支持这样的假设：翻译选择主导了酵母蛋白质的进化．

2006 年，BMC Bioinformatics 上又发表了一篇讨论蛋白质相互作用、进化速率和丰度的文章[15]．先后使用两种不同的进化距离计算方法，采用小鼠和古巴沟齿鼩(*S. paradoxus*)两个物种作为参考物种，Saeed 等分别对六大目前被广泛使用的酵母蛋白质相互作用数据集进行了分析．他们肯定了前人的观点：a．蛋白连接度同进化速率之间的负相关关系确实存在；b．当控制表达水平后，这种关系被明显削弱；c．与连接度相比，表达水平可以更好地预测进化速率．同时，作者发现，蛋白连接度同进化速率之间的相关性只表现在高准确性的数据集上，所以据此认为，前人研究争论的原因在于所使用的相互作用数据集的准确性不同．

2007 年，Kim 等[16]将研究视野扩展到整个网络，研究蛋白质在整个网络中的位置与其自身所受选择压力的关系．通过衡量种内单核苷酸多态性、种间序列分歧以及检测基因组结构变量来衡量蛋白质所受的选择压力，发现处于正向选择的蛋白质趋向于定位在相互作用网络的外围．由此推测，这种现象的出现是由于网络外围蛋白暴露面较多或者有更多的对外界环境变化的适应性事件发生．可以预见，作者观察到的这种现象对网络数据的完整性和质量应该很敏感．

自 2002 年 Fraser 等提出蛋白连接度同其进化速率呈负相关的观点以来，在随后的四五年里，讨论这一问题的文章不断出现．虽然他们所选择的研究物种大都为酵母，但所使用的具体的相互作用数据集、进化速率的评估方法、寻找直系同源蛋白的方法以及所选用的统计分析方法不尽相同．这就为横向比较这些研究带来了困难．但从现有的研究成果可以基本得出这样的结论：蛋白连接度同其进化速率之间可能存在着负相关关系，并且这种关系应是相对较弱的．目前，这些因素之间相关性背后的机制还没有被清楚地揭示．我们从多年来的争论中得到的启发是：a．影响蛋白质进化速率的因素很多，但它们之间并非互相独立，而是存在复杂的相互依赖关系的．在研究它们同进化速率的关系时，千万不能忽视对它们之间的依赖关系进行控制．b．高通量实验技术的发展带来了越来越多可利用的蛋白质相互作用数据，同时也带来了数据质量不高的问题[17]．并且，当前网络数据仍然不完整的问题也不容忽视．在蛋白质相互作用网络(进化)研究领域，应时刻不忘讨论数据的质量和规模对结论造成的影响．c．应重视使用正确的统计分析方法．只有经得起正确统计方法检验的结果才是能站得住脚的，而无统计学意义的结果很可能是随机波动造成的假象而并非真实的规律．

2 蛋白质相互作用层次的相关问题

蛋白质相互作用对蛋白质序列进化有无约束作用，有怎样的约束作用? 发生相互作用的两蛋白质在分子水平上是否趋向共进化，甚至他们是否趋向具有相似的性质? 多年来，这些问题引起了许多学者的关注．

回答蛋白质相互作用对蛋白质序列进化有无约束作用，有怎样的约束作用这个问题，最可靠的方式可能就是利用蛋白质和蛋白质复合物结构数据来研究相互作用接触面处残基的进化情况．最早在 1994 年，学者们就开始研究，相对于整个蛋白质序列，相互作用接触面处的氨基酸残基是否更加保守，其中不乏有争议的观点[18~20]．特别是，Caffrey 等[19]虽然对这一问题持不肯定的观点，但他们敏感地指出：瞬时接触面与专一性(obligate)接触面在氨基酸残基保守性上显著不同．而相互作用的两个蛋白质接触面处的残基共进化问题，也从 1994 年开始被广泛研究[21, 22]．甚至很早就出现了基于接触面处残基共进化对相互作用接触面进行预测的工作[22]．虽然这些研究工作起步较早也为数不少，但受制于十分有限的蛋白质和蛋白质复合物结构数据，考察范围往往很窄．这些工作并非生物信息学家擅

长和感兴趣的研究范畴．直到 2005 年，Mintseris 和 Weng[23]考察了较大的蛋白质复合物结构数据集后，才开始了对相关问题较大规模的研究．研究表明：专一性复合物(如多亚基酶)接触面处的残基以相对慢的速率进化，这就允许它们共进化，而瞬时相互作用(单个蛋白质之间短暂的相遇，如酶 - 底物)接触面处的残基有相对高的替换率，这几乎不支持它们有相互关系的突变．随着蛋白质和蛋白质复合物结构数据的不断增加，以及未来高通量的结构检测方法的产生和发展，与相互作用接触面相关的序列进化研究将可能成为很多生物信息学家感兴趣的问题．

与相互作用接触面处序列进化研究不同，相互作用蛋白质全序列水平的相关研究引起了许多生物信息学家的兴趣．2002 年，Teichmann[24]研究发现，相对于裂殖酵母(*S. pombe*)中的直系同源蛋白，参与不同类型相互作用的酿酒酵母(*S. cerevisiae*)蛋白有不同的平均序列一致比例．其中，参与稳定复合物的酵母蛋白序列一致比例最高，参与瞬时相互作用的蛋白质次之，而其他蛋白质最低，并且这一趋势独立于蛋白质功能和可移除性．蛋白质平均序列一致比例可以反映该蛋白质的进化保守性，所以此结论支持了蛋白质相互作用对蛋白质序列进化有附加的约束作用这一观点．同年，Fraser 等[4]通过对酵母中相互作用两蛋白的进化速率差异进行考察，发现相互作用两蛋白进化速率的平均差异要显著低于随机期望．这暗示相互作用蛋白倾向于具有更相似的进化速率．但在 2004 年，Agrafioti 等[12]发现，虽然发生相互作用两蛋白的进化速率在统计学上呈正相关，但相关性不强．并且，考虑了蛋白质自身的表达水平、功能和参与的生物学过程等其他影响蛋白质进化速率的因素后，一个蛋白质的进化速率对与其发生相互作用的另一个蛋白质的进化速率的影响就变得很小．

除了在分子水平上共进化，近几年有多项研究表明，发生相互作用的两蛋白在表达水平等层次上也可能存在共进化的现象．2004 年，Fraser 等[25]利用四种酵母物种的序列数据，使用密码子适应指数来评估基因表达水平，发现发生相互作用的两蛋白对应的基因表达水平在不同物种中表现出显著的协同变化．这就暗示发生相互作用的两蛋白在表达水平上也存在共进化的现象．作者认为，表达水平共进化比序列共进化能更有效地预测蛋白质相互作用．同年，Lemos 等[26]在酵母和果蝇中都发现，发生相互作用两蛋白对应的基因表达水平的相关程度高于随机期望．同时，在酵母中相互作用两蛋白的基因表达水平多态性在数量上也比随机期望更相近，并且在控制了基因表达水平后，这个结论仍然成立．作者推测，蛋白质相互作用可能对调控进化有约束作用，并且这种约束可能起到了维持蛋白质与其相互作用对象剂量平衡的作用．相反，也有人证明酵母中具有相似基因表达谱的基因更可能编码发生相互作用的蛋白质[27].

多年来，学者们发展的多种用于预测蛋白质相互作用的方法大多是基于相互作用蛋白共进化的思想．这些方法包括比较基因组学方法[3]、利用系统发育树相似性进行预测的方法[28, 29]、利用基因表达水平相关性进行预测的方法[25]和同源预测方法[30]等．这些预测算法的成功，也从相反的角度为相互作用蛋白共进化提供了有力证据．

总的来说，学者们普遍认为，相互作用对蛋白质序列进化有一定的约束作用，相互作用两蛋白在多个层面上共进化，虽然这种共进化趋势可能比较微弱[9]．学者们提出了两种普遍的假设来解释这种相互作用两蛋白在多个层面上共进化的现象．一种假设认为，这种共进化是施加在相互作用两蛋白上相似的进化压力的结果．而这种相似的进化压力可能来源于作用在这两个相互作用蛋白上的相似的控制机制，比如协同转录和调控．这种角度的解释不仅适用于发生直接物理相互作用的两蛋白的共进化，而且对共享一个生物学关系(比如相同的生化通路)的一组蛋白质的共进化也是适用的(本文第 **3** 部分的讨论内容)[31]．而另一种假设则认为，这种共进化直接与相互作用蛋白的共适应相关．比如 Fraser 等[4]提出，当蛋白质序列上直接参与或间接通过影响蛋白质折叠而参与相互作用的位点发生有害突变时，与其发生相互作用的蛋白质通过发生互补的改变来维持两蛋白的相互作用关系，进而保持功能．目前，支持两种假设的研究成果都有一些．正像 Juan 等[31]总结的那样，两种共进化推动力正在不同程度地、在不同水平上和不同情况中发挥作用.

3 模体、模块层次的相关问题

类比于相互作用层次，在模体、模块层次等方面，学者们也围绕着模体或者功能模块是否对其成员蛋白质进化有约束作用，蛋白质是否趋于共进化等问题展开了讨论．

蛋白质相互作用网络等许多类型的生物学网络

中存在模体结构[32]，并且也都具有层次模块化特性[33~35]．网络模体(network motif)定义为复杂网络中在不同位置重复出现的特定的相互连接模式，并且在数量上显著地高于随机期望[32]．模体中一般含有 3～5 个节点．在系统生物学领域，网络模体在基因调控网络中被研究得最多，而在蛋白质相互作用网络中，也有些学者从进化角度对其进行研究[36, 37]．细胞网络能够被分割成功能模块．功能模块(functional module)一般定义为一组相互作用的成分，它们在一起发挥作用以能够在相对空间、时间和化学隔离下完成某种离散的生物学功能．功能模块最显著的特点是其往往表现出内部更可能在功能和拓扑上互相联系[38, 39]，这也是许多模块检测算法的理论依据．功能模块是比模体更大的生物学网络的组成单元，模块可能包含模体作为其结构成分．功能模块的定义侧重于功能，所以明显不同于纯拓扑学意义上的模体结构，同时也不能像模体那样可以被精确确定和计算．功能模块在不同的生物学网络中以不同的形式存在，包括在蛋白质相互作用网络中的蛋白质复合物、代谢网路中的代谢通路、信号转导网络中的信号通路等[40]．其中，蛋白质复合物是最明确定义的，可被实验直接检测到的功能模块[41, 42]．

2003 年，Wuchty 等[36]对酵母蛋白质相互作用网络中模体成员蛋白的保守性进行研究，发现模体成员蛋白要比非模体成员蛋白在进化上更具有保守性，并且不同拓扑结构模体中，成员蛋白的保守性不同．作者认为他们的研究暗示模体可能是网络中进化保守的拓扑单元．2006 年，Lee 等[37]在前人工作的基础上，利用分子功能的 Gene Ontology 注释，进一步探究了具有相同拓扑结构但不同分子功能组成的不同模体模式(motif mode)的进化保守性．结果发现，不同的模体模式所承受的进化约束显著不同．因而，相对于模体，这种综合了拓扑和功能的模体模式可能更好地代表了蛋白质相互作用网络中进化保守的拓扑单元．不过，正像 Wuchty 等指出的，对于模体、模体模式甚至是模块进化保守性的更深入研究，应该是同时研究成员蛋白和它们之间相互作用的保守性．但是，目前不同物种的相互作用数据交叉很小、不适宜直接比较的现状限制了对相互作用保守性的研究[43, 44]．随着相互作用数据的进一步完善，成员蛋白和它们之间相互作用的联合进化研究将成为网络模体和模块进化研究中非常有意义的发展方向之一．

生物网络模块化是否与进化约束有关系呢? 有些学者从研究高连接度蛋白，即 hub 蛋白所经历的进化约束来回答这个问题．2004 年，Nature 上发表的一篇文章报道，在酵母蛋白质相互作用网络中[45]，高连接度蛋白可以根据其同相互作用蛋白之间的协同表达情况划分为 date hub 和 party hub 两类，并且后者比前者倾向于具有更多的亚细胞定位．同时文章揭示，酵母蛋白相互作用网络具有模块化结构．在这种结构中，data hub 负责连接不同的模块或生物学过程，而 party hub 在模块内部发挥作用．因此，data hub 也被称为模块间 hub，而 party hub 被称为模块内 hub．2005 年，Fraser 等[46]通过对酵母网络中这两类 hub 的某些进化性质进行比较研究，证实模块性对蛋白质进化有抑制作用．他们发现，模块内 hub 受到更高的进化约束，而模块间 hub 更易变化．Hub 蛋白受到更高的进化约束表现为其进化速率更慢，能在更多的真核物种中找到直系同源蛋白以及具有更少的遗传相互作用．这个结论暗示，网络进化变革趋向于通过改变模块间而非模块内蛋白质和相互作用而实现．但在 2006 年，上述两篇文章的结论受到了 Batada 等[47]的正面质疑．他们指出，前人之所以得出上述结论是因为：a．使用了不完整的蛋白质相互作用和遗传相互作用数据集；b．没有使用或者是使用了不恰当的统计分析方法；c．没有控制如蛋白质丰度等因素的影响．这次，Batada 等克服前人不足，从 5 个方面否定了两类 hub 存在差异的结论．这 5 个方面包括：hub 蛋白同其相互作用蛋白共表达强度是否展示有统计学意义的双峰分布、亚细胞定位的数量、去除后对网络拓扑的影响、遗传连接度以及进化速率．到了 2007 年，随着酵母蛋白质相互作用数据集的进一步发展，Bertin 等[48]联合了 Han 和 Fraser 等两篇文章的作者，开始又证实，两类 hub 蛋白在上述 5 方面中的后三方面确实有区别．但他们承认，区分两类 hub 的主要指标——hub 蛋白同其相互作用蛋白的共表达强度确实在统计学上不满足双峰分布．但这并不影响他们使用一个共表达强度阈值作为区分标准．在相同杂志(PLoS Biology)的同一期上紧接着刊登了 Batada 等[49]针对 Bertin 等[48]文章的反驳意见．Batada 等有意使用与 Bertin 文章完全相同的相互作用数据集，但运用了更加恰当的统计分析方法和严密的逻辑推理，最终得出：两类 hub 性质上的不同几乎都是由其定义方法造成的，即依据共表达强度不同区分两类 hub．总之，

无论真实的情况如何，以上工作都不能否认模块化对蛋白质进化可能有约束作用，即使不能从模块间hub和模块内hub存在进化性质上的差异这个角度对其提出更有力的证据.

类似于相互作用层次，在模体、模块层次，也有多篇文章支持模体或模块成员蛋白趋向具有共进化、共表达的特性. 比如，Vergassola等[50]通过将酿酒酵母蛋白与其他4种酵母物种中的直系同源蛋白比较分析发现，全连通子图中成员蛋白的进化速率呈现强烈的多点相关性. Chen等[51]也在酵母相互作用网络中发现，相同模块内蛋白之间比不同模块间蛋白之间，在进化速率和表达水平上更相似. 这一结果暗示了模块成员蛋白在序列和表达水平上的共进化. 另外，类似于蛋白质相互作用预测领域，近年来，功能模块预测领域发展的许多方法也都是基于模块成员蛋白共进化的思想，如比较基因组学方法[52, 53]. 而这些方法的成功也反过来支持了功能模块成员蛋白的共进化. 尽管如此，Snel和Huynen[54]通过对蛋白质复合物、代谢通路等多种功能模块的进化模块化进行考察，却发现大部分功能模块只表现出有限的进化模块化. 功能模块的进化模块化是指模块成员在不同物种基因组中同时出现和消失. 但有大约一半的模块显著地比随机一组基因表现出更强的进化模块化. 并且不同类型、不同数据库来源甚至是相同来源但不同功能的模块表现出不同的进化模块化.

总的来说，与相互作用层次对应，在模体、模块层次，学者们也普遍支持模体或者模块对其成员蛋白进化有约束作用，并且其成员蛋白之间在进化速率、表达水平等方面表现出共进化特性. 另外，从上文的讨论中还可以得到启示：即使是同为物理相互作用，也应在某些情况下区别对待. 比如目前已有研究表明，比起二元瞬时相互作用，蛋白质相互作用网络中的特有功能模块——稳定的蛋白质复合物对其成员蛋白质有更强的约束作用[23, 24]，并且其在不同基因组间趋向更加保守[55].

4 蛋白质相互作用网络进化

研究蛋白质相互作用网络进化的一个最原始和基本的问题是探索蛋白质相互作用网络的起源. 蛋白质相互作用网络不同于随机网络的无尺度分布、小世界性质和模块化结构等是如何起源和进化的? 这些特性的存在是生物体长期进化过程中自然选择的结果，还是存在着某些内在约束机制使其不可避免，例如被中性选择所驱使的自组织机制[56]? 为了回答这些问题，多年来，学者们从多个方面做了很多努力.

4.1 蛋白质相互作用网络无尺度和小世界性质的起源和进化

众所周知，蛋白质相互作用网络拥有无尺度分布和小世界性质[57]. 前者是指网络中连接度为 k 的节点出现的概率 $P(k)$满足幂律分布，即 $P(k)\propto k^{-\gamma}$. 对于生物学网络，一般 $2<\gamma<3$[58]. 而当网络具有较短的平均最短路径长度和较高的平均聚集系数时，此网络就满足小世界性质. 节点 i 的聚集系数定义为 $C_i=2n_i/(k_i(k_i-1))$，其中 n_i 表示节点 i 的 k_i 个邻居节点之间边的数目. 早在2000年，就有文章指出，无尺度网络结构对网络中随机节点的去除表现出很好的鲁棒性(robustness)[59]，但不能抵抗hub节点的去除. 而较快的扰动传播速度和较小的反应时间与小世界性质有关[42]. 这些在功能上存在一定优势的特性可能是在自然选择的作用下产生的，但是理论模拟的方法已经揭示，具有与真实网络可比的拓扑特性的网络完全能够通过包含简单规则的网络生长模型得到. 多年来，学者们先后提出了多个无尺度和小世界网络的进化模型.

最早提出且最简单的模型是1999年Barabasi和Albert等[60]提出的优先连接模型(preferential attachment model). 该模型描述网络的生长是新添加的节点与现存节点的连接度成比例地连接到网络中的现存节点上. 利用此模型产生的网络具有无尺度性质. 2003年和2005年，Eisenberg等[61]和Joy等[62]分别利用了不同的酵母蛋白质相互作用数据集检测了该模型. 他们利用在其他物种中寻找直系同源蛋白的方法评估蛋白质起源年代，得出：a. 蛋白质起源越早，其连接度越高；b. 在进化过程中，一个蛋白质获得的相互作用的数目同其连接度成正比. 这些结论支持了蛋白质相互作用网络进化的优先连接模型. 不过2004年，利用更精确的GeneTrace算法[63]评估酵母蛋白的起源年代，Kunin等[64]并没有得到期望的结果. 他们发现，起源于真核物种开始分化时期的蛋白连接度最高，而更古老的蛋白质反而连接度稍低一些. 进一步，他们又考察了蛋白质功能同连接度和起源年代的关系，发现之所以没有得到蛋白质年龄同连接度之间令人期望的关系是因为蛋白质功能的影响. 这暗示，考察无尺度网络的产生机制还应考虑蛋白质功能. 2006年，Saeed等[15]使用了Excess Retention(ER)方法来

评估蛋白质年龄．结果也发现，蛋白质年龄与连接度之间存在强烈而显著的关系，即蛋白质起源越早，其连接度越高．并且当控制表达水平后，这种关系并没有被显著地削弱．这一结果同样支持了网络生长过程中优先连接机制的存在．

2002 年到 2003 年，提出蛋白质相互作用网络的复制 - 分歧模型(duplication-divergence model, DD model)，并在多篇文献中讨论[65~69]．在该模型中，现存网络中的蛋白质被随机选择并复制，且伴随着该蛋白质参与的所有相互作用．接着，基因突变导致副本和原蛋白逐渐发生分歧，表现为它们参与的相互作用发生改变．复制 - 分歧模型实际上可以理解成发生于基因组上的变化在网络拓扑结构变化上的体现．在选择适当参数的情况下，由复制 - 分歧模型进化来的网络满足无尺度和小世界特性．这一时期研究复制 - 分歧进化模型的文章都以酵母蛋白质相互作用网络作为参考网络，所描述的模型差别主要在于基因复制后相互作用的改变规则．非对称模型假设副本蛋白保留部分原蛋白的相互作用[65, 67, 69]，而在对称模型中，副本和原蛋白都可能丢失相互作用[66]．还有一些模型也考虑了副本蛋白与网络中已存在蛋白质特别是原蛋白之间出现新的相互作用[65, 66, 69]．2005 年，Ispolatov 等[70]又提出了一种简单的完全非对称的复制 - 分歧模型，该模型中只包含一个参数——相互作用保持概率．

2003 年，利用酵母基因组数据和相互作用网络数据，Wagner 等[71]首次定量地评估了影响蛋白质相互作用网络进化的两个过程——基因复制和相互作用改变的速率．结果发现，在酵母网络中，虽然这两个速率都足够高，以至于能够在相对短的时间内影响网络结构，但相互作用获得和丢失的平均速率要比基因复制速率至少高一个数量级．并且，高连接蛋白有更高的相互作用改变速率，这与优先连接模型一致．2004 年，Wagner 等[72]基于之前得到的快的链路动力学和慢的复制动力学，提出了一种不同于传统的复制 - 分歧模型的定量的网络进化模型．在该模型中，链路动力学的要点是非对称的优先连接规则，即相互作用连接的速率只随参与的两个节点之一的连接度的增加而增加． 该模型可以只依靠链路动力学预测蛋白质相互作用网络重要的结构特性，而较慢的基因复制主要影响网络尺寸．该模型可以预测出真实的酵母蛋白质相互作用网络的连接度分布以及网络中相互作用蛋白之间的连接度关系．

2003 年，Chung 等[67]指出，大网络无尺度性质的产生仅决定于网络生长过程而与最初的种子网络无关．然而连接度满足无尺度分布并不能确定唯一的网络拓扑．具有相同连接度分布的网络可能具有非常不同的拓扑结构[73]．2007 年，Hormozdiari 等[74]发现，除无尺度性质外，真实的蛋白质相互作用网络所具有的其他一些关键拓扑学特性的产生，强烈地依赖于所选择的种子网络和进化模型．作者进一步发现，当选择包含两个大小相当的全连通子图的网络作为种子网络时，利用复制 - 分歧模型进化得到的网络除了满足无尺度性质外，还具有真实网络所具有的紧密度(closeness)分布和介数(betweenness)分布等，而利用优先连接模型是得不到这些的．

虽然优先连接模型被最早提出，并且之后也有几篇文章从研究蛋白质年龄与连接度关系的角度支持它，但从近年来发表的文献看，该模型并非当今学术界认可的主流．其中一个重要原因是很多学者认为，这种连接过程并不能与真正的生物学过程对应起来[59]．相反，复制 - 分歧模型越来越受到认可，并且其可能确实是真实的蛋白质相互作用网络进化所遵循的规则．因为已经有研究证明，在酵母中至少有 40%的蛋白质相互作用来源于复制事件[41]．

4.2 蛋白质相互作用网络模块化结构的起源和进化

蛋白质相互作用网络等许多类型的生物学网络具有层次模块化结构．当无尺度网络中连接度为 k 的所有节点的平均聚集系数 $C(k)$满足 $C(k) \propto k^{-1}$ 时，网络具有层次模块化结构[17]．模块化被认为能够增加演化性(evolvability)，既能通过提供一些可重复使用的部分用于组成新的功能，又可通过减小基因多向性(指单基因影响多性状的程度)促使性状能够被自然选择，以个体为单位优化．同时功能模块的形成能够提供鲁棒性以抵御突变和化学攻击[38]．

上文说到很多研究证明，由复制 - 分歧模型进化来的网络满足无尺度分布和小世界性质．近年来，有人证明，在控制好某些参数后，由该模型进化产生的网络也具有模块化结构[34, 35]．但优先连接模型做不到这一点[34]．例如，Sole 和 Fernandez[35]证明，存在这样一个明显分开高度连接图和分离图的突变点，在突变点附近，复制 - 分歧模型进化来的网络满足上述三性质．这里所用的复制 - 分歧模型为非对称模型，蛋白质伴随着其所有相互作用被复制后，模型只考虑副本蛋白的分歧导致的相互作

用重连.

2005 年，Takemoto 和 Oosawa[75]完全从理论上提出了一种通过合并全连通图来进化网络的模型. 即，网络进化以一个具有 m 个节点的全连通图为种子网络. 在每个时间点，新的具有 m 个节点的全连通图被加入，并与现存网络共用 n 个节点($n < m$). 使用优先选择模型来选择这 n 个节点. 共用的边依然被计数，用于下一步的优先选择中. 经分析，该模型生成的网络满足无尺度分布和层次模块化特性. 2008 年，Takemoto 和 Oosawa 两位学者对此模型进行了进一步改进[76]. 他们改用适应性驱使的优先连接模型来选择共用的 n 个节点，结果证明，改进模型产生的网络在上述性质的基础上还具有不对称(assortative)的连接度关系. 不对称的连接度关系是指网络中具有相似连接度的节点趋向互相避免. 有文章支持此性质，认为也是蛋白质相互作用网络所具有的性质之一[77, 78]. 不过从理论上说，该网络进化模型和实际的生物学过程很难对应，只能给生物网络进化研究提供有限的借鉴作用.

2007 年，Fernández[73]从实际的酵母蛋白质相互作用网络数据出发，发现现存网络的不对称结构是从祖先网络的对称结构逐渐进化而来的. 在这一过程中，网络一直满足无尺度分布特性. 这一过程在理论上可通过冷淡相似度连接的网络生长模型而实现. 作者认为，现存网络非对称结构的选择优势在于其最小化了 hub 节点而非随机节点去除的网络脆弱性，因为非对称结构尽量减少了一个 hub 节点的去除对其他 hub 的影响.

近两年也有文章专门探讨最典型的功能模块——蛋白质相互作用复合物的起源和进化. 比如，Pereira-Leal 等[79]通过对已知复合物的观察和复合物进化的理论模拟，揭示了许多蛋白质复合物的起源和进化，是通过最初的自相互作用的建立和紧随其后的自相互作用蛋白的复制而实现的. 为了研究网络中在模块建立后是否进一步发生模块复制，2005 年，同一研究小组发展了一个简单的分析模块复制的框架[41]. 结果发现，在酵母中，至少 6%～20%的蛋白质复合物与其他复合物有强烈的相似性. 因而推断，有相当一部分复合物是通过复制而进化来的. 进一步研究表明，许多复合物是通过逐步的部分复制而进化来的. 并且被复制的复合物仍保持原复合物的核心功能，但有了不同的绑定特异性和规则. 例如，原复合物和副本复合物的核心接触反应活性或受体绑定活性保持一致，但它们的底物特异性发生了改变. 这暗示，模块的复制与功能特化(functional specialization)有关. 值得注意的是：Pereira-Leal 等两个研究工作以及 2007 年 Fernandez 的工作不同于之前讨论的网络进化的理论模拟，他们是从实际的相互作用数据出发来探讨网络可能的进化过程和机制.

生物学网络不同于随机网络的特点可被视作基本的表型特征(phenotypic traits). 进化机制造就了现今生物学网络表型特征的存在. 目前，学术界存在两种不同的进化机制. 第一种是达尔文的自然选择观点. 该观点认为，首先突变引起表型特征的随机变化，然后环境选择适者，使得本无方向性的表型特征根据功能向着最优化方向发展. 第二种包含了很多可选机制，如遗传漂变、自组织等. 它们被 Huang[56]统称为内在约束(intrinsic constraints)，共同特点是都包含一组规则，规则施加约束使得表型特征不可避免. 这里的约束是一种创造力而非抑制力，它防止系统稳定在一个随机的、无差别的状态. 第二种机制强调不需要外力.

正如上文所述，蛋白质相互作用网络所具有的无尺度分布、小世界性质、层次模块化结构甚至是不对称结构等确实在增强对环境改变的鲁棒性、提高对环境改变的反应速度和增加演化性等方面存在某些优势. 可以说，这些特性是“适应环境的”. 但是，这并不意味着它们的存在一定是自然选择造就的，或者说并不一定是自然选择一手造就的. 正如有些学者所言，自然选择也许并没有强大到能够从混乱中，通过在表型空间中的许多可选的、简单但缺乏功能的形式中，逐步探索并铸造出现今的表型特征[56]. 功能适应也许是基于内在设计规则，创造结构的内在约束过程的副产品. 就是说，内在约束也许才是铸造表型特征的“主力”，而自然选择只起到了“微调”的作用. 比如上文说到的各种进化模型，特别是复制 - 分歧模型，它们完全不需要自然选择的参与，就能够依据简单的规则，产生现今蛋白质相互作用网络所具有的某些结构特性. 再如，无尺度性质在许多非生物网络中也有发现，这表明自然选择可能并非是这种结构的创造者. 不过，无疑，大部分情况下内在约束并不能独自完美地解释一切. 我们并不偏向两种机制中的任何一种. “自然选择学说”在学术界占有很重要的地位甚至是统治一时的地位，绝对有其充分的道理. 只是，我们要意识到除了自然选择，内在约束也是游戏的一部分，并且它所起的作用也许比自然选择更

关键．况且这两种机制虽本质上不同，但并不互相排斥，它们是互相补充，甚至是相互协同的．正如Huang所言，在大部分情况下，两种机制以协同方式促成某一特征形成．

5 总结和展望

本文从蛋白质、蛋白质相互作用、模体、模块直到整个网络等五个层次，综述了近年来蛋白质相互作用网络进化研究领域的最新进展．侧重探讨蛋白质相互作用网络对蛋白质进化的约束作用和网络特性的起源及进化机制两个基本问题．目前，虽有一些争议，但学术界普遍支持网络中蛋白质的进化受到其参与的相互作用、模体、模块甚至是整个网络中其他蛋白质的影响．而对于网络特性的起源和进化机制，学术界也已初步形成两种本质上不同但又并非互相排斥的观点．网络进化机制的完全揭示还需学者们更多的努力．另外，由于篇幅所限，还有一些问题我们没能详细讨论，如网络比对等．

近年来，虽然该领域取得了许多研究成果，但应该意识到：这些研究的基础是实验所得数据，因此也受限于这些数据．目前可利用的蛋白质相互作用网络数据的质量和规模仍在不断发展当中，网络进化研究的很多结论也可能并非一成不变．另外还应警惕，当前的相互作用网络是一个高度平均和理想化的结构，其中包含了细胞中不同条件、不同时间、甚至不同空间的各种相互作用．也许这样的网络数据能够反映一些生物体内真实的细胞网络的特点，但其肯定不能反映全部，或者其中一些被反映出的特点可能并非是真实的．

展望未来，在本领域，我们预计可能有以下几个发展方向：

a．进化研究的对象将会更加细化．目前，学者们已经意识到网络中不同类型节点、不同类型相互作用甚至是不同类型模体或模块在进化性质、进化过程中的地位以及对细胞功能实现的贡献等方面有所不同．在未来，这种区分研究将会继续细化和深入．

b．理论模拟与实际网络数据研究将会更加紧密结合．一方面，需要进一步提高网络进化模型，发展更加精确、更加符合实际网络数据的网络进化模型；另一方面，应更多地从实际网络数据出发，更加深入地揭示网络进化机制和规律．

c．网络比对将会在进化研究中发挥重要作用并成为研究热点．网络比对类比于序列比对．正像序列比对对进化生物学发展起到的巨大作用一样，网络层次的比对将在检测保守的功能模块、揭示生物网络的进化路径、预测未注释子网功能等诸多方面发挥重要作用．目前，虽然已发展了一些网络比对算法[80, 81]，但限于当前远不能与基因组序列数据相比的生物网络数据，这些算法的应用范围还十分有限．相信，随着网络数据的继续发展和学术研究的需要，网络比对的继续发展和广泛应用将成为必然趋势．

d．与动态蛋白质相互作用网络相关的进化研究将成为热点．动态蛋白质相互作用网络更接近细胞的真实情况．同时，不断增加的基因表达等数据也为动态网络研究提供了一定的数据基础．

e．不同学科研究方法相结合．目前虽已有学者利用电子学知识来研究网络模块进化[82, 83]，但由于其与真实的生物学过程还不能够很好地对应，利用其他学科方法研究网络进化等生物学问题还处于起步阶段．未来，这可能会是本领域的研究方向之一．这需要生物学家和其他领域学者的共同努力．

参 考 文 献

1 Dobzhansky T. Nothing in biology makes sense except in light of evolution. Am Biol Teacher, 1973, **35**: 125～129

2 石铁流, 李亦学. 系统生物学的现状与展望. 中国科学基金, 2005, **19**(5): 282～286
Shi T L, Li Y X. Bulletin of National Natural Science Foundation of China, 2005, **19**(5): 282～286

3 Von Mering C, Jensen L J, Kuhn M, *et al*. STRING 7—recent developments in the integration and prediction of protein interactions. Nucleic Acids Res, 2007, **35**(Database issue): D358～D362

4 Fraser H B, Hirsh A E, Steinmetz L M, *et al*. Evolutionary rate in the protein interaction network. Science, 2002, **296**(5568): 750～752

5 Jordan I K, Wolf Y I, Koonin E V. No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. BMC Evol Biol, 2003, **3**(1): 1

6 Fraser H B, Wall D P, Hirsh A E. A simple dependence between protein evolution rate and the number of protein-protein interactions. BMC Evol Biol, 2003, **3**(1): 11

7 Bloom J, Adami C. Apparent dependence of protein evolutionary rate on number of interactions is linked to biases in protein-protein interactions data sets. BMC Evol Biol, 2003, **3**(1): 21

8 Pal C, Papp B, Hurst L D. Highly expressed genes in yeast evolve more slowly. Genetics, 2001, **158**(2): 927～931

9 Fraser H B, Hirsh A. Evolutionary rate depends on number of protein-protein interactions independently of gene expression level. BMC Evol Biol, 2004, **4**(1): 13

10 Bloom J D, Adami C. Evolutionary rate depends on number of

protein-protein interactions independently of gene expression level: Response. BMC Evol Biol, 2004, **4**(1): 14

11 Wuchty S. Evolution and topology in the yeast protein interaction network. Genome Res, 2004, **14**(7): 1310～1314

12 Agrafioti I, Swire J, Abbott J, *et al*. Comparative analysis of the *Saccharomyces cerevisiae* and *Caenorhabditis elegans* protein interaction networks. BMC Evol Biol, 2005, **5**(1): 23

13 Hahn M W, Kern A D. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol, 2005, **22**(4): 803～806

14 Drummond D A, Raval A, Wike C O. A single determinant dominates the rate of yeast protein evolution. Mol Biol Evol, 2006, **23**(2): 327～337

15 Saeed R, Deane C M. Protein protein interactions, evolutionary rate, abundance and age. BMC Bioinformatics, 2006, **7**(1): 128

16 Kim P M, Korbel J O, Gerstein M B. Positive selection at the protein network periphery: Evaluation in terms of structural constraints and cellular context. Proc Natl Acad Sci USA, 2007, **104**(51): 20274～20279

17 Von Mering C, Krause R, Snel B, *et al*. Comparative assessment of large-scale data sets of protein-protein interactions. Nature, 2002, **417**(6887): 399～403

18 Grishin N V, Phillips M A. The subunit interfaces of oligomeric enzymes are conserved to a similar extent to the overall protein sequences. Protein Sci, 1994, **3**(12): 2455～2458

19 Caffrey D R, Somaroo S, Hughes J D, *et al*. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface?. Protein Sci, 2004, **13**(1): 190～202

20 Valdar W S, Thornton J M. Protein-protein interfaces: analysis of amino acid conservation in homodimers. Proteins, 2001, **42** (1): 108～124

21 Gobel U, Sander C, Schneider R, *et al*. Correlated mutations and residue contacts in proteins. Proteins, 1994, **18**(4): 309～317

22 Pazos F, Helmer-Citterich M, Ausiello G, *et al*. Correlated mutations contain information about protein-protein interaction. J Mol Biol, 1997, **271**(4): 511～523

23 Mintseris J, Weng Z. Structure, function, and evolution of transient and obligate protein-protein interactions. Proc Natl Acad Sci USA, 2005, **102**(31): 10930～10935

24 Teichmann S A. The constraints protein-protein interactions place on sequence divergence. J Mol Biol, 2002, **324**(3): 399～407

25 Fraser H B, Hirsh A E, Wall D P, *et al*. Coevolution of gene expression among interacting proteins. Proc Natl Acad Sci USA, 2004, **101**(24): 9033～9038

26 Lemos B, Meiklejohn C D, Hartl D L. Regulatory evolution across the protein interaction network. Nat Genet, 2004, **36**(10): 1059～1060

27 Ge H, Liu Z, Church G M, *et al*. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. Nat Genet, 2001, **29**(4): 482～426

28 Pazos F, Valencia A. Similarity of phylogenetic trees as indicator of protein-protein interaction. Protein Eng, 2001, **14**(9): 609～614

29 Waddell P J, Kishino H, Ota R. Phylogenetic methodology for detecting protein interactions. Mol Biol Evol, 2007, **24**(3), 650～659

30 Huang T W, Lin C Y, Kao C Y. Reconstruction of human protein interolog network using evolutionary conserved network. BMC Bioinformatics, 2007, **8**(1): 125

31 Juan D, Pazos F, Valencia A, *et al*. Co-evolution and co-adaptation in protein networks. FEBS Lett, 2008, **582**(8): 1225～1230

32 Milo R, Shen-Orr S, Itzkovitz S, *et al*. Network motifs: simple building blocks of complex networks. Science, 2002, **298** (5594): 824～827

33 Ravasz E, Somera A L, Mongru D A, *et al*. Hierarchical organization of modularity in metabolic networks. Science, 2002, **297** (5586): 1551～1555

34 Hallinan J. Gene duplication and hierarchical modularity in intracellular interaction networks. Biosystems, 2004, **74**(1～3): 51～62

35 Sole R V, Fernandez P. Modularity "for free" in genome architecture. arXiv: q-bio.GN/0312032v1, 2003-12-19

36 Wuchty S, Oltvai A N, Barabasi S L. Evolutionary conservation of motif constituents in the yeast protein interaction network. Nat Genet, 2003, **35**(2): 176～179

37 Lee W P, Jeng B C, Pai T W, *et al*. Differential evolutionary conservation of motif modes in the yeast protein interaction network. BMC Genomics, 2006, **7**(1): 89

38 Hartwell L H, Hopfield J J, Leibler S, *et al*. From molecular to modular cell biology. Nature, 1999, **42**(Suppl): C47～52

39 Wagner G P, Pavlicev M, Cheverud J M. The road to modularity. Nat Rev Genet, 2007, **8**(12): 921～931

40 Qi Y, Ge H. Modularity and dynamics of cellular networks. Plos Comput Biol, 2006, **2**(12): e174

41 Pereira-Leal J B, Teichmann S A. Novel specificities emerge by stepwise duplication of functional modules. Genome Res, 2005, **15** (4): 552～559

42 Pereira-Leal J B, Levy E D, Teichmann S A. The origins and evolution of functional modules: lessons from protein complexes. Philos Trans R Soc Lond B Biol Sci, 2006, **361**(1467): 507～517

43 Cesareni G, Geol A, Gavrila C, *et al*. Comparative interactomics. FEBS Lett, 2005, **579**(8): 1828～1833

44 Gandhi T K, Zhong J, Mathivanan S, *et al*. Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nat Genet, 2006, **38**(3): 285～293

45 Han J D, Bertin N, Hao T, *et al*. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature, 2004, **430**(6995): 88～93

46 Fraser H B. Modularity and evolutionary constraint on proteins. Nat Genet, 2005, **37**(4): 351～352

47 Batada N N, Reguly Teresa, Treitkreutz A, *et al*. Stratus not altocumulus: a new view of the yeast protein interaction network. PLOS Biol, 2006, **4**(10): e317

48 Bertin N, Simonis N, Dupuy D, *et al*. Confirmation of organized modularity in the yeast interactome. PLoS Biol, 2007, **5**(6): e153

49 Batada N N, Reguly T, Breitkreutz A, *et al*. Still not Altocumulus: Further evidence against the data/party hub distinction. PLoS Biol, 2007, **5**(6): e154

50 Vergassola M, Vespignani A, Dujon B. Cooperative evolution in protein complexes of yeast from comparative analysis of its interaction network. Proteomics, 2005, **5**(12): 3116～3119

51 Chen Y, Dokholyan N V. The coordinated evolution of yeast proteins is constrained by functional modularity. Trends Genet, 2006, **22**(8): 416～419

52 Von Mering C, Zdobnov E M, Tsoka S, *et al*. Genome evolution reveals biochemical networks and functional modules. Proc Natl Acad Sci USA, 2003, **100**(26): 15428～15433

53 Compillos M, Von Mering C, Jensen L J, *et al*. Identification and analysis of evolutionarily cohesive functional modules in protein networks. Genome Res, 2006, **16**(3): 374～382

54 Snel B, Huynen M A. Quantifying modularity in the evolution of biomolecular systems. Genome Res, 2004, **14**(3): 391～397

55 Brown K R, Jurisica I. Unequal evolutionary conservation of human protein interactions in interologous networks. Genome Biol, 2007, **8** (5): R95

56 Huang S. Back to the biology in systems biology: what can we learn from biomolecular networks?. Brief Funct Genomic Proteomic, 2004, **2**(4): 279～297

57 Li D, Li J, Ouyang S, *et al*. Protein interaction networks of saccharomyces cerevisiae, *Caenorhabditis elegans* and *Drosophila melanogaster*: large-Scale organization and robustness. Proteomics, 2006, **6**(2): 456～461

58 Stumpf M P, Kelly W P, Thorne T, *et al*. Evolution at the systems level: the natural history of protein interaction networks. Trends Ecol Evol, 2007, **22**(7): 366～373

59 Albert R, Jeong H, Barabasi A L. Error and attack tolerance of complex networks. Nature, 2000, **406**(6794): 378～382

60 Barabasi A L, Albert R. Emergence of scaling in random networks. Science, 1999, **286**(5439): 509～512

61 Eisenberg E, Levanon E Y. Preferential attachment in the protein network evolution. Phys Rev Lett, 2003, **91**(13): 138701

62 Joy M P, Brock A, Ingber D E, *et al*. High-betweenness proteins in the yeast protein interaction network. J Biomed Biotechnol, 2005, **2005**(2): 96～103

63 Kunin V, Ouzounis C A. GeneTRACE-reconstruction of gene content of ancestral species. Bioinformatics, 2003, **19**(11): 1412～1416

64 Kunin V, Jacq B, Brun C. Functional evolution of the yeast protein interaction network. Mol Biol Evol, 2004, **21**(7): 1171～1176

65 Sole R V, Pastor-Satorras R, Smith E, *et al*. A model of large-scale proteome evolution. Adv Compl Syst, 2002, **5**(1): 43～54

66 Vazquez A, Flammini A, Maritan A, *et al*. Modeling of protein interaction networks. Complexus, 2003, **1**(1): 38～44

67 Chung F, Lu L, Dewey T G, *et al*. Duplication models for biological networks. J Comput Biol, 2003, **10**(5): 677～687

68 Pastor-Satorras R, Smith E, Sole R V. Evolving protein interaction networks through gene duplication. J Theor Biol, 2003, **222** (2): 199～210

69 Kim J, Krapivsky P L, Kahng B, *et al*. Infinite-order percolation and giant fluctuations in a protein interaction network. Phys Rev E Stat Nonlin Soft Matter Phys, 2002, **66**(5 Pt 2): 055101

70 Ispolatov I, Krapivsky P L, Yuryev A. Duplication-divergence model of protein interaction network. Phys Rev E Stat Nonlin Soft Matter Phys, 2005, **71**(6 Pt 1): 061911

71 Wagner A. How the global structure of protein interactrion networks evolves. Proc R Soc Lond B, 2003, **270**(1514): 457～466

72 Berg J, Lässig M, Wagner A. Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. BMC Evol Biol, 2004, **4**(1): 51

73 Fernández A. Molecular basis for evolving modularity in the yeast protein interaction network. PLoS Comput Biol, 2007, **3**(11): e226

74 Hormozdiari F, Berenbrink P, Pržulj N, *et al*. Not all scale-free networks are born equal: the role of the seed graph in PPI network evolution. PLoS Comput Biol, 2007, **3**(7): e118

75 Takemoto K, Oosawa C. Evolving networks by merging cliques. Phys Rev E Stat Nonlin Soft Matter Phys, 2005, **72**(4 Pt 2): 046116

76 Takemoto K, Oosawa C. Modeling for evolving biological networks with scale-free connectivity, hierarchical modularity, and disassortativity. Math Biosci, 2007, **208**(2): 454～468

77 Colizza V, Flammini A, Maritan A, *et al*. Characterization and modeling of protein-protein interaction networks. Physica A, 2005, **352**(1): 1～27

78 Goh K I, Kahng B, Kim D. Graph theoretic analysis of protein interaction networks of eukaryotes. Physica A, 2005, **357** (3～4): 501～512

79 Pereira-Leal J B, Levy E D, Kamp C, *et al*. Evolution of protein complex by duplication of homomeric interactions. Genome Biol, 2007, **8**(4): R51

80 Flannick J, Novak A, Srinivasan B S, *et al*. Græmlin: General and robust alignment of multiple large interaction networks. Genome Res, 2006, **16**(9): 1169～1181

81 Zhang S, Zhang X S, Chen L. Biomolecular network querying: a promising approach in systems biology. BMC Syst Biol, 2008, **2**(1): 5

82 Paladugu S R, Chickarmane V, Deckard A, *et al*. In silico evolution of functional modules in biochemical networks. Syst Biol (Stevenage), 2006, **153**(4): 223～235

83 Deckard A, Sauro H M. Preliminary studies on the in silico evolution of biochemical networks. Chembiochem, 2004, **5**(10): 1423～1431

Progress in The Evolutionary Analysis of Protein Interaction Networks*

LIU Zhong-Yang[1,2], LI Dong[1,2], ZHU Yun-Ping[1,2)**], HE Fu-Chu[1,2)**]

([1)]*State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing* 102206, China;

[2)]*Beijing Institute of Radiation Medicine, Beijing* 100850, China)

Abstract Recently, advances in high-throughput experimental technologies enable an ever-increasing amount of data on protein interaction networks available. These data provide new insights into the evolutionary processes of protein interaction networks. The researches associated with analyzing such data from an evolutionary perspective was reviewed at five different levels: from proteins to protein interactions, motifs, modules and the whole network. Two aspects were focused on: 1) the constraints of the network organization on protein evolution, 2) origins and evolution of the topological features of protein interaction networks which are different from those of random networks. In addition, the enlightenments from the former studies were presented and the development trends in this field were discussed.

Key words protein interaction network, evolution, bioinformatics

---

*This work was supported by grants from The State Key Development Program for Basic Research of China (2006CB910803, 2006CB910700), The National High Technology Research and Development Program of China (2006AA02A312) and The National Natural Science Foundation of China (30621063).

**Corresponding author.

ZHU Yun-Ping. Tel: 86-10-80705225, E-mail: zhuyp@hupo.org.cn

HE Fu-Chu. Tel: 86-10-66931246, E-mail: hefc@nic.bmi.ac.cn

Received: August 7, 2008 Accepted: October 24, 2008