

大规模高通量方法在蛋白质相互作用研究中的应用*

符庆璞 高钰琪**

(第三军医大学高原军事医学系病理生理学与高原生理学教研室, 重庆 400038)

摘要 随着后基因组时代的到来, 阐明蛋白质间相互作用关系成为蛋白质研究的又一热点, 促进了相关技术的不断产生、发展和完善。其中涉及到诸多大规模高通量的方法, 如双杂交系统、噬菌体展示、质谱、蛋白质芯片以及生物信息学等, 这为系统分析蛋白质相互作用提供视点, 有望在蛋白质组学研究中发挥重要作用。每种方法各有其优缺点且适用范围不同, 在一定程度上各方法的实验结果互为补充。现拟就这些大规模高通量方法的研究进展及其在蛋白质相互作用研究中的应用作一综述。

关键词 蛋白质相互作用, 大规模, 高通量筛选

学科分类号 Q51, Q503

随着人类大规模基因组测序工作的完成, 生命科学步入后基因组时代, 即功能基因组时代。由于全基因组序列信息并不足以解释机体的各种生命现象, 于是研究重点从揭示生命的遗传信息转向对生物功能的研究。蛋白质是生物功能的最终执行者, 体内蛋白质并不是独立发挥功能, 而是通过与其他分子相互作用完成, 其中蛋白质-蛋白质间的相互作用是蛋白质发挥功能的重要途径。然而, 简单的两两相互作用的研究并不符合生命的动态性特征, 应从整体角度系统地考察错综复杂的蛋白质相互作用调控网络, 从而为最终研究蛋白质功能及其细胞全局特征打下基础。目前, 研究蛋白质相互作用的方法较多, 各有其优势和局限, 采用哪种方法来研究应根据具体的研究目的、要求和对象而定。本文拟就一些大规模高通量方法在蛋白质相互作用研究中的应用作一综述。

1 双杂交系统

1.1 酵母双杂交

酵母双杂交系统于 1989 年由 Fields 和 Song 首先创建^[1], 原理如图 1 所示, 利用了真核细胞转录因子组件式结构的性质。转录因子发挥功能所必需的 2 个结构域是 DNA 结合结构域(binding domain, BD)和转录激活结构域(activation domain, AD), 二

者在结构和功能上相对独立, 分别与诱饵蛋白 X 的基因及另一蛋白 Y 的基因融合形成杂交体 BD-X 和 AD-Y。当这 2 个杂交体在同一酵母细胞中表达时, 只要 2 种蛋白质发生相互作用, BD 和 AD 就能在空间上充分接近, 形成完整有活性的转录因子形式, 从而激活报告基因的转录。如果 Y 是基因文库, 那么即可直接从基因文库中筛到与蛋白 X 相互作用的结合蛋白基因。

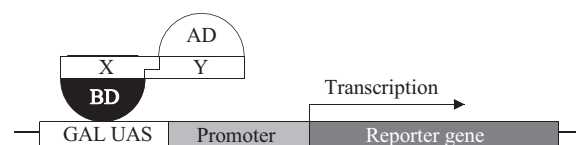


Fig. 1 Schematic diagram of the yeast two-hybrid system
图 1 酵母双杂交系统示意图

UAS: 上游活化序列; Promoter: 启动子; Reporter gene: 报告基因; Transcription: 转录。

该系统的优点在于: a. 双杂交筛选中用作诱饵的真核生物蛋白在酵母这种真核生物中更有可能正

* 国家自然科学基金“高原低氧高寒的损伤与适应机制的研究”重大项目(30393131)。

** 通讯联系人。

Tel: 023-68752331, E-mail: gyq@mail.tmmu.com.cn

收稿日期: 2007-06-20, 接受日期: 2007-08-08

确折叠和进行翻译后修饰; b. 蛋白质之间相互作用的研究是在酵母细胞体内进行, 无需蛋白质的分离、纯化等步骤, 所证实的蛋白质间相互作用在一定程度上代表了细胞内的真实情况; c. 可检测存在于蛋白质之间微弱或暂时的相互作用; d. 整个过程仅需基因操作, 方法简便、灵敏、高效, 很容易实现蛋白质相互作用的高通量自动化分析. 因此, 酵母双杂交系统自建立以来, 成为蛋白质相互作用大规模研究方面应用较为广泛的技术.

将酵母双杂交用于蛋白质相互作用网络研究的第一个具有开创性的工作是 Bartel 等^[2]对 T7 噬菌体进行了大规模的双杂交分析, 结果在 55 个噬菌体中发现了 25 个相互作用蛋白. 随后, 双杂交研究在更大范围内展开. Uetz 等^[3]利用阵列筛选模式, 将 192 个诱饵菌株与 6 000 个猎物菌株一一结合, 形成二倍体, 结果发现有 87 个诱饵与猎物之间存在 281 个相互作用. 另外, 将约 6 000 个诱饵菌株分别与含有 6 000 种猎物的文库接合, 结果从 817 个诱饵中筛选到 692 个相互作用. Ito 等^[4]用类似方法也完成了酵母细胞中几乎所有蛋白质之间相互作用的高通量筛选, 结果在 3 278 个蛋白质中发现了 4 549 个相互作用, 其中 797 个蛋白质中的 841 个相互作用最可靠, 具有 3 次以上的可重复性. 与前面 Uetz 等^[3]的结果比较, 只有 141 个蛋白质相互作用对是共同的, 其原因尚不清楚, 但提示用一种方法筛选出的蛋白质相互作用, 需要与其他方法结合才能对实验结果作出完整和准确的判断.

此外, 在果蝇^[5]、线虫^[6]、病毒^[7,8]、幽门螺旋菌^[9]、*E. coli* 噬菌体 T7^[10]中也都有了大规模酵母双杂交分析的报道. 并且在热休克蛋白(heat shock protein, HSP) 方面也有应用, 如筛选 HSP105 α 的结合蛋白^[11], 以及研究酵母 HSP90 分子伴侣系统的蛋白质-蛋白质相互作用^[12].

最近, Stelzl 等^[13]采用酵母双杂交方法通过 4 456 个诱饵蛋白和 5 632 个猎物蛋白分析了人脑组织中蛋白质相互作用的情况, 发现在 1 705 种蛋白质间存在 3 186 种新的相互作用关系. 同时用免疫共沉淀、Pull-down 以及生物信息学进行分析验证, 最终在 401 种蛋白质间确认了 911 对高可信度的相互作用, 该研究是首次大规模研究人类蛋白质间的相互作用. 随后, Raul 等^[14]分析了大约 8 100 个人类已知开放阅读框编码产物, 发现了约 2 800 个相互作用, 联合亲和纯化证实符合率达 78%. 该研究结果补充了约 70% 的相互作用关系, 揭示了 300

多种新的相互作用, 涉及 100 多种疾病相关蛋白. 上述对人脑组织和人开放阅读框的研究有助于系统全面地认识人类蛋白质相互作用图谱.

尽管酵母双杂交已经得到了广泛应用, 但其本身也存在一定的局限性: a. 所研究的相互作用必须定位于核内, 发生在胞浆和胞膜的相互作用无法检测, 导致“假阴性”; b. 由于酵母核内其他因素的介导, 可能使 2 个原本无直接相互作用的蛋白质拉到一起, 激活报告基因转录, 造成“假阳性”; c. 能与酵母细胞内进化保守的蛋白质相互作用的蛋白质也不能用此方法研究; d. 不能研究具有自激活特性的蛋白质. 为了克服酵母双杂交的某些缺陷, 初步建立了哺乳动物双杂交系统作为酵母双杂交系统的辅助手段, 另外, 还逐步发展了 SOS 招募系统和泛素专一性蛋白酶系统, 分别针对膜受体、细胞外分泌蛋白质进行研究. 目前这些系统尚无应用于大规模探测蛋白质相互作用的实例.

1.2 细菌双杂交

细菌双杂交是继酵母双杂交之后, 由 Dove 等^[15]建立的另一种新方法, 同样也是基于转录激活的原理. 如图 2 所示, 目的蛋白(诱饵)与全长 λ 噬菌体 cI 蛋白(λ cI)融合, 其中 λ cI 氨基端具有 DNA 结合域; 靶蛋白则是与 RNA 聚合酶 α 亚基的 N 端结合域融合. 诱饵通过 λ cI 的 DNA 结合域与报告基因启动子上游的 λ 操纵区连接. 当诱饵与靶蛋白相互作用时, 彼此相互接近, 使 RNA 聚合酶结合到报告基因的启动子上, 启动报告基因的转录和表达. 如果与 RNA 聚合酶 α 亚基 N 端结合域融合的是基因文库, 那么通过筛选可以直接得到与诱饵蛋白相互作用蛋白的 DNA 序列, 实现蛋白质相互作用的高通量分析. 对于报告基因, 最初用的是编码 β -内酰胺酶的 *bla* 基因和编码 β -半乳糖苷酶的 *lacZ* 基因. 在此基础上, 将参与组氨酸生物合成途径的 *HIS3* 基因和链霉素抗性基因 *aadA* 作为新一代报告基因. 这种改变有利于筛选容量更大的文库, 以及降低假阳性率, 因此成为目前细菌双杂交中最具优势的系统之一.

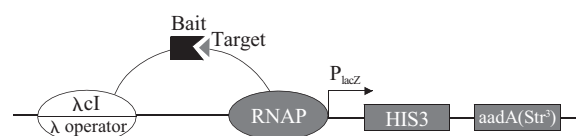


Fig. 2 Schematic diagram of the bacterial two-hybrid system

图 2 细菌双杂交系统示意图

Bait: 诱饵; Target: 靶蛋白; λ operator: λ 操纵子; RNAP: RNA 聚合酶; P_{lacZ} : *lacZ* 启动子.

该系统继承了酵母双杂交系统的部分优点,同时也克服了酵母双杂交系统存在的一些局限性,在以下几方面具有独特的优势:a.研究周期短,操作更加简单、快速;b.细菌的转化效率显著高于酵母细胞,更有利于筛选大文库,发现可能的稀有蛋白质作用对;c.诱饵和靶蛋白不需要核定位信号;d.细菌与高等真核生物之间具有更大的进化距离,可以避免由酵母内源蛋白所引起的假阳性和假阴性现象,同时也显著降低真核生物蛋白在报告菌株中的毒性作用,以及减少自激活阳性率.细菌双杂交系统也存在不足,其缺点在于表达的蛋白质缺乏翻译后修饰.

最近有报道,利用细菌双杂交筛选人肺 cDNA 文库获得了 7 种与人转录生长因子 β 抑制性膜相关蛋白相互作用的蛋白质,它们参与肌动蛋白细胞骨架的组建、细胞粘附以及细胞转录和翻译的调节^[16]. 尽管目前细菌双杂交的应用远不及酵母双杂交广泛,但其作为一种筛选更为高通量、操作更为简便的方法,可以预见将在未来研究中发挥更大的作用.

1.3 酵母/细菌相结合的双杂交

无论是细菌双杂交还是酵母双杂交都存在一定的假阳性和假阴性,鉴于此, Serebriiskii 等^[17]对双杂交诱饵表达质粒和报告菌株进行改造,使得诱饵表达质粒既能在酵母中表达,也能在细菌中表达,同时具备能够在酵母和细菌双杂交系统之间快速转换的能力,即酵母/细菌相结合的双杂交系统(yeast/bacterial two-hybrid system, YBTHS). YBTHS 中的酵母系统具有比细菌系统更宽泛的动力学范围,而细菌系统出现诱饵自激活现象的情况比酵母系统少,从而大大增加了双杂交筛选的覆盖率和准确率.利用 YBTHS 筛选 HeLa cDNA 文库寻找 Ras 蛋白的生理性相互作用子,结果经酵母和细菌系统均筛选到了相当数量的初筛转化子,有趣的是,二者间并无重复克隆,提示,在不同生物系统中可能探测到不同的特异性相互作用.这也再次证明,一种方法不可能筛选到全部的相互作用,方法与方法之间应该互补.

2 噬菌体展示

噬菌体展示技术是一种噬菌体表面表达筛选技术,其原理是以噬菌体为载体,将外源核酸片段(一组随机多肽编码序列或基因群)克隆至噬菌体外壳蛋白基因中形成融合,并以融合蛋白的形式表达

于噬菌体表面,众多融合型噬菌体便组成了噬菌体展示库.随后,将噬菌体过柱,用固定化的靶分子去筛选与之亲和的噬菌体.通过与特定的靶标反应,不能结合的噬菌体被洗脱掉,而能结合的噬菌体被保留下来并通过感染大肠杆菌得以扩增和富集,实现高通量筛选.在噬菌体展示技术中,展示库的每一个噬菌体只展示一种序列的外源肽,于是融合表达的外源肽与其基因之间建立了直接联系,通过测定噬菌体 DNA 的序列便可推知被展示的多肽或蛋白质的氨基酸序列.

已利用噬菌体展示技术从 cDNA 展示库中直接分离出与靶蛋白相互作用的结合蛋白及基因^[18~20],这为研究细胞内蛋白质间的识别、细胞信号转导通路及其调节机制提供了线索.近年来随着噬菌体展示肽库的建立,噬菌体展示技术在蛋白质相互作用研究中得到了更为广泛的应用.用于筛选的靶分子包括抗体、膜受体、酶以及其他功能蛋白质,其中最为常见的是,以单克隆抗体为靶分子筛选抗原表位模拟肽.筛选出的具有较高亲和力和特异性的目的肽段可作为候选药物进行开发,尤其在阻断天然抗体促进异种器官移植方面具有重要价值^[21].不仅如此,噬菌体展示技术还常与光散射分析相结合进行蛋白质相互作用的研究^[22,23].

与酵母双杂交相比,噬菌体展示系统在筛选蛋白质间相互作用时更为简便、高通量,更适用于不能用酵母双杂交研究的膜蛋白和转录因子,并且无需了解被研究蛋白质的结构,在筛选过程中通过适当改变条件可以直接评价相互结合的特异性.但该方法也存在一些缺点,如蛋白质在宿主菌中有时无法正确折叠或修饰,构建的融合蛋白也可能改变天然蛋白质的结构和功能,影响蛋白质的活性等.现已尝试用噬菌体以外的其他方式展示肽库以克服噬菌体展示系统的不足,其中切实可行又具开发潜力的主要有质粒展示肽库和核糖体展示肽库,但目前还有待进一步的改进.

3 质 谱

3.1 串联亲和纯化

近年来,质谱技术已用来鉴定蛋白质复合体或复合体亚基的相互作用,但实际工作中往往难以获得足量纯化的蛋白质复合体,因而应用受到极大限制.后来基于质谱发展了一种串联亲和纯化(tandem affinity purification, TAP)方法,并最先应用于酵母蛋白复合体的研究中.该方法的原理

如图 3 所示, 其巧妙之处在于设计了一个双重分子标签, 包括蛋白 A、钙调素结合蛋白以及二者间连接的烟草蚀纹病毒(tobacco etch virus, TEV)蛋白酶切割位点. 将 TAP 标签构建到靶蛋白上, 在宿主细胞内表达融合蛋白, 并与内源相互作用蛋白形成复合体. 细胞裂解物经 IgG 偶联的层析柱纯化, 通过蛋白 A, 复合体与 IgG 特异结合. 接着经 TEV 蛋白酶酶解、洗脱, 分离蛋白 A 标签, 使含有靶蛋白的复合体与层析柱分离. 在钙离子存在的情况下, 将第一步纯化得到的洗脱物再经钙调素偶联的层析柱纯化, 并洗涤去除杂质和 TEV 蛋白酶. 然后加入 EGTA 螯合, 使钙调素结合蛋白与亲和层析柱分离, 于是含有靶蛋白的复合体得以纯化. 最后, 经 SDS-PAGE 分离复合体各组分以及胰酶酶切后, 进行质谱分析等操作.

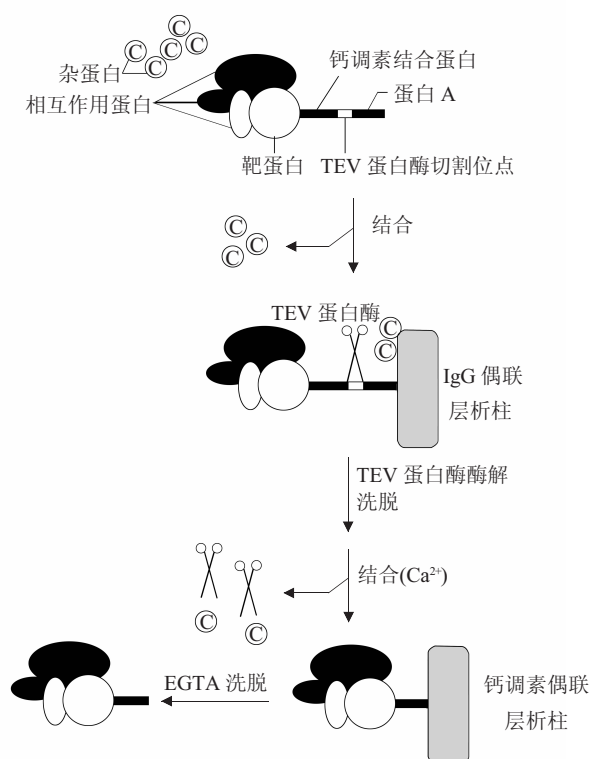


Fig. 3 Principle of the TAP

图 3 TAP 技术基本原理

该方法特别适用于探测蛋白质在生理条件下的相互作用. Gavin 等^[24]采用 TAP 方法对酿酒酵母多蛋白复合体的 1 739 个基因进行研究, 其中有 1 143 个与人类基因相关, 并且纯化了 589 种蛋白质, 经生物信息学分析, 它们分属于 232 个不同的多蛋白

复合体, 共有 344 个蛋白质具有新的细胞功能, 其中 231 个以前未见功能报道. 随后, Forler 等^[25]在 TAP 方法中引入 RNA 干扰技术, 探测果蝇 S2 细胞中的蛋白质间相互作用, 该策略避免了真核细胞中表达的内源蛋白与标签蛋白发生竞争作用, 从而使得该方法能更好地应用于高级整合生物细胞中蛋白质相互作用的研究. 最近, 还利用质谱对人类蛋白质相互作用图谱进行了大规模研究, 共分析了与疾病(已知或可疑)及其功能相关的 338 个诱饵蛋白, 发现在 2 235 个不同蛋白质之间存在 6 463 个相互作用. 该研究结果揭示了许多新的蛋白质相互作用关系及相关通路, 尤其对于人类疾病的研究具有重要价值^[26].

3.2 高通量质谱蛋白质鉴定

高通量质谱蛋白质鉴定 (high-throughput mass-spectrometric protein complex identification, HMS-PCI) 是又一种基于质谱的高通量方法, 它与 TAP 的最大不同在于融合蛋白在细胞内的表达水平不同. 如果天然状态下靶蛋白在细胞内的浓度过低不足以分析时, 可采用 HMS-PCI, 但当靶蛋白过表达时若仍采用 HMS-PCI, 则会产生假阳性. 因此, 应根据具体实验对象选用合适的分析方法.

利用 HMS-PCI 对酵母的多蛋白质复合物进行大规模分析, 用 10% 已被预测的酵母蛋白作为诱饵, 发现了 3 617 个相关蛋白质, 约占酵母蛋白质组的 25%. 同时, 还检测到许多蛋白质复合物, 它们涉及不同信号通路以及 DNA 损伤反应中众多新的相互作用^[27]. 前面提到 Gavin 等^[24]采用 TAP 方法研究酵母蛋白质相互作用网络, 比较这 2 项研究中部分重叠的诱饵, 有一些有趣的发现: a. 通过 HMS-PCI 获取的蛋白质复合物的复杂程度要高于 TAP 方法得到的复合物; b. 2 项研究中相同诱饵所得到的复合物鉴定结果只有少量重叠, 并且质谱法得到的结果与以前酵母双杂交法得到的结果偏差较大, 到底哪种方法得到的结果更为可信, 还有待新的证据验证.

4 蛋白质芯片

继基因芯片之后发展起来的蛋白质芯片技术, 是在固相载体表面按照预先设计的方法固定大量的蛋白质探针所形成的高密度蛋白质阵列. 将带有特殊标记的蛋白质与该芯片孵育反应, 探针可以捕获蛋白质分子并与之结合, 然后用相应的检测系统进行检测, 即可得到蛋白质相互作用的信息. 如果样

品中的蛋白质预先未被标记, 则可采用以质谱技术为基础的直接检测法。蛋白质芯片根据用途不同分为蛋白质检测芯片和蛋白质功能芯片2种。在蛋白质功能芯片中许多蛋白质甚至是整个蛋白质组被排列, 从而能够对蛋白质间的相互作用进行大规模分析。

Uetz 等^[2]首次把蛋白质芯片的概念用于全基因组范围内的蛋白质研究, 他们将蛋白质芯片技术与酵母双杂交系统结合, 通过在芯片上进行酵母双杂交反应, 系统地研究了所有 6 000 种酵母蛋白之间的相互作用。通过检测 192 种蛋白质, 得到了 281 种蛋白质 - 蛋白质间交互作用关系。此外, Zhu 等^[28]克隆了 5 800 个酵母菌的开放读码框, 并过量表达和纯化了相应的蛋白质, 将这些纯化了的蛋白质高密度固定在芯片上形成了酵母蛋白质组芯片, 该芯片包含了来自酵母基因组的所有表达蛋白质。用钙调素筛选该蛋白质芯片, 结果发现 39 种与钙调素相互作用的蛋白质, 其中有 33 种是新发现的。基于这些实验结果, 他们识别了一个多种钙调素结合蛋白的通用结合位点, 并由此推测这种包含完整真核蛋白质组的芯片可以制备并用来检测筛选具有不同生物活性的蛋白质分子。

目前, 蛋白质功能芯片还用于检测激酶和底物的相互作用。Zhu 等^[29]将 119 种酵母蛋白激酶制作成蛋白质芯片, 运用 17 种特异性底物与之作用。最初认为这些酵母的蛋白激酶中, 除了 2 种属组氨酸家族外, 其余的都属于丝氨酸 - 苏氨酸家族。但研究发现, 有 27 种激酶都能有效磷酸化具有聚“谷氨酸 - 酪氨酸”结构的底物, 提示酵母中至少存在 27 种以前不为人们所知的酪氨酸激酶, 并且它们在催化区附近具有共同的氨基酸残基。该研究揭示了蛋白激酶许多新的特征, 证明蛋白质芯片在高通量筛选蛋白质相互作用和生化活性方面具有重要价值。随后, Ptacek 等^[30]利用酵母蛋白质组芯片研究 87 种酵母蛋白激酶的相互作用蛋白, 发现了大约 4 200 个磷酸化反应, 涉及 1 325 种不同的蛋白质。筛选结果描绘了酵母中的蛋白质磷酸化图谱, 揭示了蛋白激酶可能的新作用以及新的调节分子, 这些将有助于研究真核生物蛋白质磷酸化的机制和作用。此外, 最近研究者还利用人类蛋白质组芯片来确定蛋白质之间的相互作用^[31]。

影响蛋白质芯片检测结果的因素较多, 需要发展新的数据处理和资料分析的软硬件, 对检测结果进行标准化以确认结果的有效性。尽管如此, 蛋白

质芯片技术有着许多其他方法无法比拟的优势。蛋白质芯片以其高通量、高灵敏度、微型化、集成化、快速准确以及大规模平行检测等特点, 为研究蛋白质相互作用提供了许多传统技术难以得到的信息, 成为研究蛋白质相互作用的有力工具。

5 生物信息学分析方法

利用生物信息学方法预测蛋白质相互作用现已成为研究的一大热点。对于蛋白质相互作用网络的生物信息学研究大都是基于对大量蛋白质相互作用数据的研究, 这就需要有强大的专业数据库予以支撑。已知的蛋白质与蛋白质相互作用的生物公共数据库主要包括: DIP (the Database of Interacting Proteins, 网址为 <http://dip.doe-mbi.ucla.edu/>)、BIND (the Biomolecular Interaction Network Database, 网址为 <http://bind.ca/>)和 Predictome(网址为 <http://predictome.bu.edu/>)。此外, PROTEOME (网址为 <http://www.proteome.com/>)、MIPS(网址为 <http://www.mips.biochem.mpg.de/>)和 PIM (网址为 <http://proteome.wayne.edu/PIMdb.html>)等数据库也都各有特点。目前已有的诸多预测蛋白质相互作用的生物信息学方法中, 应用较为广泛的主要有同源预测、比较基因组和基于结构域的方法等。

5.1 同源预测

同源预测的理论基础是进化保守的蛋白质之间倾向于具有保守的蛋白质相互作用。由于不同物种间许多相互作用是保守的, 因此利用模式生物蛋白质相互作用数据推测人类蛋白质的相互作用成为可能。具体说就是, 将模式生物的蛋白质相互作用网络贮存在一个参考数据集中, 然后在目标蛋白质组中依据直系同源关系搜索同源蛋白相互作用。该预测方法的准确性依赖于同源蛋白的判断标准以及模式生物蛋白质相互作用数据的准确度。利用此法建立了第一个人类蛋白质相互作用图谱, 其中预测了 1/3 人类基因(包括 448 个人类疾病基因和 1 482 个未知功能基因)的 70 000 个相互作用, 为生物医学研究提供了丰富的资源^[32]。

5.2 比较基因组

基因在基因组中规律排列, 功能相近或相关的基因往往具有特定的上下文关系。因此, 通过基因组比较可预测蛋白质间的相互作用。基于基因组信息的预测方法包括系统发育谱、基因邻接、基因融合等。

功能相关基因的物种分布相同或相似, 同时存

在或不存在某一基因组中的模式, 称之为系统发育谱. 系统发育谱法的依据是, 如果一对基因在一组基因组中的系统发育谱相同, 提示这对基因在功能上是相关的, 由此预测它们编码的 2 个蛋白质之间存在相互作用. 可见, 系统发育谱法是一种通过进化信息推测蛋白质功能的大规模比较基因组方法.

但这种方法不能确定功能相关蛋白是否是直接的物理相互作用, 并且其准确性有赖于完成测序基因组的数量及系统发育谱方法的可靠性.

基因邻接基于的假设是, 在原核生物基因组中, 功能相关的基因倾向于紧密连在一起存在于一个特定区域, 构成一个操纵子. 这种基因之间的邻接关系具有保守性, 可以作为基因产物之间的功能提示. 如果基因邻接关系在多基因组中仍是保守的, 就可以用来预测蛋白质的相互作用.

另外, 在基因的进化过程中可能发生基因融合事件, 即一个物种的 2 个或多个基因, 在另一个物种中融合成为一个基因, 发生融合基因必然存在功能上的联系. 利用基因融合事件预测蛋白质相互作用是基于下面的发现, 一些(2 个或多个)相互作用的蛋白质在另一个物种中融合形成一条蛋白质多肽链. 因此, 基因融合事件可作为蛋白质功能相关或相互作用的指示. 尽管这种方法也获得了一定的应用, 但基因融合事件发生的频率较低限制了其使用, 并且该法也不能判断发生融合的蛋白质是否一定是直接的物理性相互作用.

由上可见, 比较基因组方法的优势在于能够预测非同源蛋白质之间的相互作用, 随着全基因组测序的完善, 其预测能力会更强. 其主要缺点是不能判断蛋白质间所谓的“相互作用”是直接的物理相互作用, 还是间接的功能相关. 上述 3 种比较基因组学方法预测得到的众多蛋白质相互作用数据收录于 STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) 数据库(网址为 <http://string.embl.de>), 关于预测结果的准确性也给出了相应的权重. 这成为建立大规模蛋白质相互作用网络的重要资源, 同时对于采用实验方法研究蛋白质的功能及生物学意义也具有重大意义^[33].

5.3 基于结构域的方法

蛋白质的相互作用并不需要全长蛋白质的参与, 而是通过蛋白质结构域来完成的, 因此, 从已知的蛋白质相互作用数据归纳得到的结构域相互作用规律, 也可以预测另一种生物的蛋白质相互作用图谱. 此外, 这种基于结构域的方法还可作为蛋白

质间相互作用的旁证. 例如, 收录相互作用蛋白质结构域的 InterDom(interacting domains)数据库(网址为 <http://InterDom.lit.org.sg>), 就是以蛋白质结构域相互作用为基础, 为酵母双杂交实验方法检测到的蛋白质相互作用进行评估^[34].

目前利用生物信息学方法对已存在的 8 个人类蛋白质相互作用的大规模图谱进行比较分析, 发现仅一小部分重叠, 并且分析结果明确提示, 所有的相互作用图谱都暗含了相当大的选择性和检测偏向性. 于是, 研究者根据分析结果绘制了一个综合性的人类相互作用网络, 称之为“统一的人类互作组”, 它将对未来人类互作组的研究起到指导作用^[35]. 现今高通量蛋白质-蛋白质相互作用探测方法的局限性已经造成相互作用数据含有相当大比例的假阳性和假阴性, 数据质量不高. 尽管已采用小范围的针对性实验对高通量方法进行补充, 然而大量的互作组要求更多规模化的方法以提取出真正的蛋白质相互作用. 现开发了一种称为“IRAP”的新方法^[36], 该法通过去除被看作是假阳性的相互作用, 增加被看作是假阴性的相互作用, 对实验获得的蛋白质互作组结果进行“纯化”, 以提高可信度. 曾用酵母双杂交方法大规模分析酵母、果蝇、蠕虫所得到的蛋白质相互作用的数据现用 IRAP 进行分析, 结果显示, 经纯化后的相互作用数据所含假阳性和假阴性的比例降低, 蛋白质互作组的可信度增加. 最近还发展了一种分析蛋白质相互作用数据的网络工具 APID (Agile Protein Interaction DataAnalyzer, 网址为 <http://bioinfo.dep.usal.es/apid>), 它以统一、比较的平台模式整合分析目前已被实验证实的蛋白质间相互作用信息^[37].

预测蛋白质相互作用的生物信息学方法, 不仅是对传统实验方法有价值的补充, 而且在很大程度上弥补了原有数据的缺陷, 扩展了实验方法的预测范围, 为建立蛋白质间相互作用网络提供了一条快速、有效、节省的途径, 同时也是注释蛋白质功能的重要工具. 高通量的生物信息学方法虽然可以处理大量数据, 但是如何评估和校验预测得到的结果仍是目前遇到的共同问题. 此外, 生物信息学还需要开发出更多高通量的数据分析方法和可视化分析工具, 以满足不断增加的数据量要求, 从而更准确地预测和验证更多的相互作用蛋白质.

6 实验技术的最新研究进展

用传统实验方法如亲和层析、免疫沉淀等研究

蛋白质 - 蛋白质相互作用时，往往会对细胞造成损伤，不能维持细胞内蛋白质相互作用时的正常生理条件。近来发展的荧光共振能量转移(fluorescence resonance energy transfer, FRET)技术为在活体细胞生理条件下对蛋白质间的相互作用进行实时、动态、定量研究提供了有利条件。FRET 是指足够靠近(1~10 nm)的 2 种不同的荧光生色基团，其中一种生色基团(供体)的发射谱与另一种生色基团(受体)的激发谱有一定的重叠，当供体被激发时，供体和受体之间通过偶极 - 偶极耦合作用以非辐射方式将供体激发态能量转移到受体激发态的过程。这是一种非辐射能量跃迁，表现为：当 FRET 发生时，供体的荧光强度降低，而受体发射的荧光却大大增强，强于本身的特征荧光。随着荧光成像技术的发展，FRET 与其相互结合，联合应用于细胞内蛋白质相互作用的研究中^[38]。将蛋白质标记上荧光探针，当蛋白质间不发生相互作用时，其相对距离较大，无 FRET 现象，而当蛋白质发生相互作用时，则 FRET 发生。利用激光共焦扫描显微镜或宽场荧光显微镜，通过多色荧光成像，追踪同一活细胞中不同的蛋白质，有利于从空间上研究蛋白质的相互作用；利用荧光寿命成像显微技术，通过测量荧光寿命，确定荧光寿命的分布，有利于从时间上研究蛋白质的相互作用。从而最终实现在活体细胞生理条件下对细胞内蛋白质间的相互作用进行实时的动态研究。已有报道，研究者通过将一种靶标受

体与 FRET 接收器、一个大的候选肽配基库与一种 FRET 供体进行基因融合，实现了在大肠杆菌细胞质中进行最佳相互作用伴侣的高通量光学筛选。此外，FRET 杂交相互作用筛选也为发现细胞中可能具有重要意义的蛋白质配基提供了一种强大的工具^[39]。

另外，研究者还开发出 Pull-down 技术检测蛋白质 - 蛋白质相互作用，该技术是将标签融合蛋白和细胞裂解液在谷胱甘肽亲和树脂存在下混合孵育，标签融合蛋白作为诱饵，从细胞裂解液中捕获与之相互作用的靶蛋白并结合在亲和树脂上。接着洗脱结合物，通过 SDS-PAGE 电泳分析，筛选并证实蛋白质间的相互作用。Pull-down 这种体外试验技术通常用于对筛选结果进行验证。最近，Arifuzzaman 等^[40]将此技术用于大规模分析 *E.coli* 中的蛋白质相互作用。在被检测的 4 339 个诱饵蛋白中，找到了 2 667 个相互作用蛋白质，其中 779 个功能未知。这为 *E.coli* 系统的生物学研究提供了新视点和新策略。其他生化方法(如免疫共沉淀、化学交联法)以及遗传学方法，虽然也用于蛋白质相互作用的研究但目前还不适合于大规模筛选，只是作为小范围内的验证实验。

7 结 语

目前研究蛋白质相互作用尤其是大规模的蛋白质互作组仍处于初级阶段，从表 1 可以看出，不同的方法各有其优缺点，其中一种方法将成为其他研

Table 1 Comparison of high-throughput protein interaction analysis methods
表 1 高通量蛋白质相互作用方法的比较

方法	优点	缺点
酵母双杂交	真核生物蛋白可能正确折叠和进行翻译后修饰 仅需基因操作，无需蛋白质分离、纯化等步骤 可检测蛋白质之间微弱或暂时的相互作用 表型和基因型有机结合，可直接获得基因	蛋白质相互作用定位于核内，常产生假阳性和假阴性 不能研究具有自激活特性的蛋白质以及能与酵母细胞内进化保守蛋白相互作用的蛋白质
细菌双杂交	研究周期短，操作简便、快速 仅需基因操作，无需蛋白质分离、纯化等步骤 细菌转化效率高，利于筛选大文库，发现稀有蛋白质相互作用对 表型和基因型有机结合，可直接获得基因 可避免由酵母内源蛋白引起的假阳性和假阴性	表达的蛋白质缺乏翻译后修饰，结果存在假阳性和假阴性 不能研究具有自激活特性的蛋白质
酵母 / 细菌相结合的双杂交 噬菌体展示	增加了双杂交筛选的覆盖率和准确率 适合于研究膜蛋白、转录因子和调节因子 筛选过程中通过适当改变条件可以直接评价相互结合的特异性 表型和基因型有机结合，可直接获得基因	需要改造双杂交诱饵表达质粒和报告菌株 蛋白质在宿主菌中有时无法正确折叠或修饰，且构建的融合蛋白可能改变天然蛋白质的结构和功能，影响蛋白质的活性

续表

方法	优点	缺点
TAP	适合于研究多蛋白复合体 适合于研究蛋白质在生理条件下的相互作用	难以检测低丰度蛋白质 纯化步骤可能导致一些弱相互作用丢失 对实验设备要求较高
HMS-PCI	适合于研究多蛋白复合体 适合于研究蛋白质在生理条件下的相互作用 适合于检测低丰度蛋白质	不适于检测表达量高的蛋白质 纯化步骤可能导致一些弱相互作用丢失 对实验设备要求较高
蛋白质芯片	快速、简便、高灵敏度、集成化、微型化、自动化以及大规模平行检测	对靶蛋白的结合特性不强，可能与其他蛋白质发生交叉反应 定位于固体表面的蛋白质易改变空间构象，失去生物活性
生物信息学	快速、有效、节省，能够预测蛋白质相互作用	预测结果往往有偏差需要评估和校验
FRET	在活体细胞生理条件下对蛋白质相互作用进行实时、动态和定量研究	受到供体和受体距离的限制
Pull-down	适合于探测蛋白在溶液中的相互作用	通常用于对筛选结果进行验证

研究方法的有益补充，因此，利用单一方法获得的研究结果，应该采用不同方法予以交叉验证。但即便如此，往往也很难对研究结果作出正确恰当的判定。因为已发现用不同方法得到的实验结果之间相重叠匹配的部分很少。这一方面说明蛋白质间相互作用的复杂性，另一方面也说明现有的研究方法本身还存在局限性(或是尚未达到其饱和状态，或是产生了假阳性结果，抑或是不适宜用于某些特定类型的相互作用)。这些都对蛋白质相互作用的研究提出了挑战，因此有必要也必须发展和完善现有方法，并建立新方法。只有基于更准确的蛋白质相互作用的研究结果，才能构建出正确的蛋白质间连锁图，从而为最终阐明生命活动的机制提供有益的指导。

参 考 文 献

1 Fields S, Song O. A novel genetic system to detect protein-protein interactions. *Nature*, 1989, **340**(6230): 245~246

2 Bartel P L, Roecklein J A, SenGupta D, *et al.* A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat Genet*, 1996, **12** (1): 72~77

3 Uetz P, Giot L, Cagney G, *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, 2000, **403**(6770): 623~627

4 Ito T, Chiba T, Ozawa R, *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA*, 2001, **98**(8): 4569~4574

5 Giot L, Bader J S, Brouwer C, *et al.* A protein interaction map of *Drosophila melanogaster*. *Science*, 2003, **302**(5651): 1727~1736

6 Li S, Armstrong C M, Bertin N, *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science*, 2004, **303** (5657): 540~543

7 Flajolet M, Rotondo G, Daviet L, *et al.* A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene*, 2000,

242(1~2): 369~379

8 McCraith S, Holtzman T, Moss B, *et al.* Genome-wide analysis of vaccinia virus protein-protein interactions. *Proc Natl Acad Sci USA*, 2000, **97**(9): 4879~4884

9 Rain J C, Selig L, De Reuse H, *et al.* The protein-protein interaction map of *Helicobacter pylori*. *Nature*, 2001, **409**(6817): 211~215

10 Noirot P, Noirot-Gros M F. Protein interaction networks in bacteria. *Curr Opin Microbiol*, 2004, **7**(5): 505~512

11 Saito Y, Doi K, Yamagishi N, *et al.* Screening of Hsp105alpha-binding proteins using yeast and bacterial two-hybrid systems. *Biochem Biophys Res Commun*, 2004, **14**(2): 396~402

12 Millson S H, Truman A W, Wolfram F, *et al.* Investigating the protein-protein interactions of the yeast Hsp90 chaperone system by two-hybrid analysis: potential uses and limitations of this approach. *Cell Stress Chaperones*, 2004, **9**(4): 359~368

13 Stelzl U, Worm U, Lalowski M, *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 2005, **122**(6): 957~968

14 Rual J F, Venkatesan K, Hao T, *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 2005, **437**(7062): 1173~1178

15 Dove S L, Joung J K, Hochschild A. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature*, 1997, **386**(6625): 627~630

16 Adyshev D M, Kolosova I A, Verin A D. Potential protein partners for the human TIMAP revealed by bacterial two-hybrid screening. *Mol Biol Rep*, 2006, **33**(2): 83~89

17 Serebriiskii I G, Fang R, Latypova E, *et al.* A combined yeast/bacteria two-hybrid system: development and evaluation. *Mol Cell Proteomics*, 2005, **4**(6): 819~826

18 李志杰, 刘靖华, 龚小卫, 等. 用 T7 噬菌体展示筛选系统筛选与 p38 相结合的蛋白. *第一军医大学学报*, 2003, **23**(11): 1131~1133

Li Z J, Liu J H, Gong X W, *et al.* *J First Mil Med Univ*, 2003, **23** (11): 1131~1133

19 黄英, 蔡雪飞, 何茂锐, 等. T7 噬菌体展示技术筛选丙型肝炎病毒非结构蛋白 3 的相互作用蛋白. *中华肝病杂志*, 2006, **14**(8): 561~564

- Huang Y, Cai X F, He M R, *et al.* Chin J Hepatol, 2006, **14**(8): 561~564
- 20 Roman I, Figys J, Steurs G, *et al.* *In vitro* interactions between the two mitochondrial membrane proteins VDAC and cytochrome c oxidase. Biochemistry, 2005, **44**(39): 13192~13201
- 21 Lang J, Zhan J, Xu L, *et al.* Identification of peptide mimetics of xenoreactive alpha-Gal antigenic epitope by phage display. Biochem Biophys Res Commun, 2006, **344**(1): 214~220.
- 22 Sondermann H, Zhao C, Bar-Sagi D. Analysis of Ras: RasGEF interactions by phage display and static multi-angle light scattering. Methods, 2005, **37**(2): 197~202
- 23 Carter D M, Miousse I R, Gagnon J N, *et al.* Interactions between TonB from *Escherichia coli* and the periplasmic protein FhuD. J Biol Chem, 2006, **281**(46): 35413~35424
- 24 Gavin A C, Bosche M, Krause R, *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature, 2002, **415**(6868): 141~147
- 25 Forler D, Kocher T, Rode M, *et al.* An efficient protein complex purification method for functional proteomics in higher eukaryotes. Nat Biotechnol, 2003, **21**(1): 89~92
- 26 Ewing R M, Chu P, Elisma F, *et al.* Large-scale mapping of human protein-protein interactions by mass spectrometry. Mol Syst Biol, 2007, **3**: 89
- 27 Ho Y, Gruhler A, Heilbut A, *et al.* Systematical identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature, 2002, **415**(6868): 180~183
- 28 Zhu H, Bilgin M, Bangham R, *et al.* Global analysis of protein activities using proteome chips. Science, 2001, **293**(5537): 2101~2115
- 29 Zhu H, Klemic J F, Chang S, *et al.* Analysis of yeast protein kinases using protein chips. Nat Genet, 2000, **26**(3): 283~289
- 30 Ptacek J, Devgan G, Michaud G, *et al.* Global analysis of protein phosphorylation in yeast. Nature, 2005, **438**(7068): 679~684
- 31 Grelle G, Kostka S, Otto A, *et al.* Identification of VCP/p97, carboxyl terminus of Hsp70-interacting protein (CHIP), and amphiphysin II interaction partners using membrane-based human proteome arrays. Mol Cell Proteomics, 2006, **5**(2): 234~244
- 32 Lehner B, Fraser A G. A first-draft human protein-interaction map. Genome Biol, 2004, **5**(9): R63.1~R63.9
- 33 von Mering C, Huynen M, Jaeggi D, *et al.* STRING: a database of predicted functional associations between proteins. Nucleic Acids Res, 2003, **31**(1): 258~261
- 34 Ng S K, Zhang Z, Tan S H, *et al.* InterDom: a database of putative interacting protein domains for validating predicted protein interactions and complexes. Nucleic Acids Res, 2003, **31**(1): 251~254
- 35 Futschik M E, Chaurasia G, Herzel H. Comparison of human protein-protein interaction maps. Bioinformatics, 2007, **23** (5): 605~611
- 36 Chen J, Hsu W, Lee M L, *et al.* Increasing confidence of protein interactomes using network topological metrics. Bioinformatics, 2006, **22**(16): 1998~2004
- 37 Prieto C, De Las Rivas J. APID: Agile Protein Interaction DataAnalyzer. Nucleic Acids Res, 2006, **34**: W298~W302
- 38 Day R N. Imaging protein behavior inside the living cell. Mol Cell Endocrinol, 2005, **230**(1-2): 1~6
- 39 You X, Nguyen A W, Jabaiah A, *et al.* Intracellular protein interaction mapping with FRET hybrids. Proc Natl Acad Sci USA, 2006, **103**(49): 18458~18463
- 40 Arifuzzaman M, Maeda M, Itoh A, *et al.* Large-scale identification of protein-protein interaction of *Escherichia coli* K-12. Genome Res, 2006, **16**(5): 686~691

The Applications of Large-scale High-throughput Methods for Studying Protein-protein Interaction*

FU Qing-Ying, GAO Yu-Qi**

(Department of Pathophysiology and High Altitude Physiology, Department of High Altitude Military Medicine,
The Third Military Medical University, Chongqing 400038, China)

Abstract With the advent of post-genomic era, identification of protein-protein interaction has become another hot spot of protein research and promoted the invention, development and complement of relative techniques. Major large-scale high-throughput methods, such as two-hybrid system, bacteriophage display, mass spectrometry, protein microchips and bioinformatics have provided a global view of protein-protein interaction, and are hopeful to play an important role in proteomics. Each method may have its specific strengths and weaknesses as well as differences in coverage, to some extent, the data based on these methods can complete each other. Here, recent progress of these large-scale high-throughput methods and their applications for studying protein-protein interaction are reviewed.

Key words protein-protein interaction, large-scale, high-throughput screening

*This work was supported by a grant from The Major Program of The National Natural Science Foundation of China (30393131).

**Corresponding author. Tel: 86-23-68752331, E-mail: gyq@mail.tmmu.com.cn

Received: June 20, 2007 Accepted: August 8, 2007