

大规模的酵母双杂交系统在蛋白质相互作用组图谱中的应用

马海蓉 李维琪*

(中国科学院新疆理化技术研究所 乌鲁木齐 830011)

摘要: 蛋白质-蛋白质之间的相互作用是蛋白质发挥其功能的重要途径之一。通过研究蛋白质组中所有蛋白质之间的相互作用做出蛋白质相互作用对图谱是功能基因组时代许多科学家关注的问题, 而大规模的酵母双杂交系统是蛋白质相互作用对图谱的研究中应用较为广泛的策略。近两年来该策略最具代表的实例是用它进行酵母中所有蛋白之间相互作用的检查。但是巨大的蛋白质网络比我们想象要大得多, 单一的双杂交系统不能解决所有问题, 需要同其它的方法有效地结合。

关键词: 大规模的酵母双杂交系统, 蛋白质相互作用组图谱, 阵列筛选, 文库筛选

中图分类号: Q93 **文献标识码:** A **文章编号:** 0253-2654 (2003) 06-0119-05

LARGE-SCALE YEAST TWO-HYBRID SYSTEM IN THE APPLICATION ON MAP OF PROTEOME INTERACTOME

MA Hai-Rong LI Wei-Qi

(Xinjiang Technology Institute of chemistry and physics, Chinese Academy of science, Wulumuqi 830011)

Abstract: Protein fulfilling the their roles, one of important ways is through protein-protein interaction. In functional genomic era, identifying all of protein-protein interaction in proteome and mapping the protein interactions that have been attracting many scientists' attention, of which large-scale yeast two-hybrid system is one strategy of most widely used. In recent two years, ambitious projects have launched to examine all of the protein-protein interaction in *Saccharomyces cerevisiae* using large-scale yeast two-hybrid system. Nevertheless, huge protein network is larger than that we predict and single yeast two-hybrid system cannot solve all the problems, which need be complemented by other ways.

Key words: Large-scale yeast two-hybrid system, Protein interactome mapping, Array screening, Library screening

蛋白质在细胞的生命活动中扮演着重要的角色, 其功能的发挥不是孤立的, 是在纵横交错的网络中通过与其它分子相互作用完成的。其中一种重要的作用形式就是蛋

* 联系人

收稿日期: 2002-12-24, 修回日期: 2003-02-26

白质-蛋白质之间的相互作用。于是理解蛋白的功能的一个方便途径就是通过一个已知蛋白的功能去推想与之作用的另一未知蛋白的生理功能。用这种方法,通过少数已知蛋白质的功能便可以推断出更多与之联系的未知蛋白的功能。如果在蛋白质组中进行所有蛋白质之间相互作用的研究,就可能做出生物体的蛋白质相互作用组图谱,这也正是功能基因组时代许多科学家倍加关注的问题。

1 研究蛋白质相互作用组图谱的策略

目前蛋白质相互作用组图谱的研究策略主要包括两个方面:一方面是从蛋白质本身为出发点进行的,主要代表是大规模的质谱分析。该方法的优点在于通过一次性的质谱分析就能确定天然蛋白复合物的所有成分;但局限性也很明显,即获得高丰度的、天然稳定的,高纯度的蛋白质是不容易的。所以大规模的质谱分析目前主要集中在高丰度的蛋白复合物如核糖体蛋白复合物及 RNA 拼接蛋白的研究上,尚无法在全蛋白质组范围内展开^[1]。另一方面是通过基因工程的方法展开的,即酵母双杂交系统,它充分利用基因组计划已提供的生物基因组序列信息,将外源基因分别重组到“诱饵”载体和“猎物”载体上,经转化在酵母细胞中表达,由报告基因的激活来观察蛋白质之间的相互作用。该系统的优点在于:①蛋白质之间的相互作用的研究是在酵母体内进行分析的,所以无须蛋白质分离、纯化等步骤;②外源蛋白质的表达是通过重组完成的,所以蛋白表达量相对稳定;③整个过程仅需基因的操作,方法简单、方便,很容易实现自动化进行高通量蛋白质相互作用分析。因此,酵母双杂交系统是目前基因组范围内蛋白质间相互作用研究方面应用较为广泛的策略。

2 大规模的酵母双杂交系统在蛋白质相互作用组图谱中的应用

酵母双杂交系统是一种在酵母体内研究蛋白质之间的相互作用的基因系统。该系统于 1989 年由 Fields 等人^[2]首次建立。它通常由 3 个元件组成,即“诱饵”(bait)载体、“猎物”(prey)载体和报告基因。其中“诱饵”载体携带转录激活因子(如 GAL4 蛋白)的 DNA 结合结构域(DNA binding domain, DB)与蛋白 X 的融合蛋白基因;“猎物”载体携带转录激活因子的激活结构域(activate domain, AD)与蛋白 Y 的融合蛋白基因。两者在酵母细胞中共表达。如果蛋白 X 与 Y 之间存在相互作用,则使 DB 和 AD 形成一个有活性的转录激活因子,激活下游报告基因的表达。

酵母双杂交系统不仅可以用来验证两个已知蛋白之间的相互作用,还可以用来从 cDNA 文库中寻找与已知蛋白相互作用的新伙伴。最近几年,基因组计划的顺利实施正在给我们提供越来越多生物的全基因组序列和及其它们的开放阅读框(open reading frames, ORF),那么利用酵母双杂交系统和自动化操作设备,从理论上完全可以进行大规模的甚至全蛋白质组的蛋白质-蛋白质相互作用分析。事实也是如此,1996 年, Bartel 等人^[3]首次对噬菌体 T7 进行了基因组规模的蛋白质相互作用研究,结果从 55 个噬菌体蛋白中发现了 25 个相互作用对。随后大规模的双杂交项目在更大范围内展开。目前,一个雄心勃勃的项目是利用酵母双杂交系统进行酵母细胞中所有蛋白质(6,000 × 6,000)之间相互作用的研究。

大规模的酵母双杂交系统是对多个蛋白质对之间的相互作用进行高通量筛选。其策略主要包括两种:阵列筛选(Array screening)和文库筛选(Library screening)。阵列筛

选中重组到“诱饵”载体和“猎物”载体上的外源基因通常是全长的开放阅读框 (Open Reading Frame, ORF), 而文库筛选中的外源基因通常是来自基因组文库或 cDNA 文库的 DNA 片段。对于前者, 只要全长的 ORF 的靶基因在手, 很容易以简单、平行及高通量的模式进行双杂交筛选, 所以在进行全基因组范围的蛋白质相互作用分析时倍受青睐。对于后者, 筛选过程要繁琐得多, 但它能够给出蛋白质相互作用的位点, 应该说是前者的有益的补充。因此两种策略之间是一种互补而非竞争的关系。

2.1 阵列筛选 Finely 和 Brent 等人首次利用阵列筛选对果蝇 *Drosophila* 中的细胞周期异类的激酶与其相关的蛋白的相互作用进行了研究。结果作者通过 45 次测定发现了 19 对蛋白之间的相互作用。该策略最有典型意义的实例是用它来研究酿酒酵母 *S. cerevisiae* 所有蛋白之间 ($6,000 \times 6,000$) 的相互作用以描绘其蛋白质组相互作用图谱。这可能对于人类许多疾病发生的分子机理有着重要的预示作用。

Uetz 等人^[5]在 2000 年首先对此做了相关报道。其筛选步骤如下: 第一步, 将编码 6,000 个蛋白的每一个基因的全长 ORF 进行 PCR 扩增, 然后重组到“猎物”和“诱饵”载体上, 分别转化 MATa 及 MAT α 单倍体酵母细胞; 第二步, MATa 和 MAT α 单倍体细胞进行结合。接合的模式有两种: 一种是克隆阵列模式 (clone array format), 另一种是高通量的阵列模式 (high-throughput array format)。克隆阵列模式中, 携带 AD-ORF 的每一种 MAT α 细胞克隆占据 96 孔板或 384 孔板的特定位置, 让携带 DB-ORF 的 MATa 细胞克隆分别与阵列中的 MAT α 细胞克隆进行一一交配, 最后根据特定位置的阳性克隆可以推断出潜在的相互作用蛋白质对。这种方法虽然比较慢, 但却免去了测序的需要。Uetz 等人利用此模式研究了酵母中 192 个“诱饵”蛋白与 6,000 个“猎物”蛋白之间的相互作用, 结果发现了其中的 87 个“诱饵”与“猎物”之间存在 281 个潜在的相互作用对。高通量的阵列模式不是将携带 AD-ORF 的细胞克隆分别占据特定位置, 而是将所有的 AD-ORF 细胞克隆混合, 构成一个 AD 库 (即所有“猎物”细胞的集合体), 然后再将每一个携带 DB-ORF 的细胞克隆与 AD 库一一交配, 得到的二倍体阳性克隆需要通过测序确定潜在的相互作用蛋白对。在筛选速度上, 显然后者比前者快许多。Uetz 等人利用此模式完成了酵母细胞中几乎所有 6,000 个蛋白之间的相互作用, 结果从 817 个“诱饵”中发现了 652 个相互作用对, Ito 等人^[6]用一种类似的高通量筛选也完成酵母细胞中几乎所有蛋白质之间的相互作用。与 Uetz 等人做法不同的是分别将带有 DB-ORF 和 AD-ORF 的单倍体细胞克隆分别构成 62 个 DB 库和 AD 库。每一个 DB 库和 AD 库由 96 个独立的克隆组成, 然后将这些 DB 库和 AD 库进行所有可能的交配 ($62 \times 62 = 3,844$)。最后, 作者在 3,278 个蛋白中发现了 4,549 个蛋白质相互作用对, 其中 797 个蛋白中的 841 个蛋白质相互作用对具有 3 次以上的可重复性。

2.2 文库筛选 利用大规模的阵列筛选能够迅速找到潜在的相互作用对, 但是还需要进一步的实验来寻找这些相互作用对的具体作用位点即功能位点。酵母双杂交的文库筛选恰恰是利用基因组文库信息或者 cDNA 文库信息来研究某一特定生物功能相关的蛋白质之间的相互作用。应该说是阵列筛选的有益的补充。如果筛选过程在全基因组范围进行, 就有可能利用它来描绘蛋白质相互作用位点图谱和蛋白质功能图谱。

基因组规模的文库筛选中最具代表性的是 Fromont 等人^[7]的工作。他们以 10 种已知的 mRNA 前体剪接有关的蛋白为起始“诱饵”, 从酿酒酵母基因组文库中筛选到与之相互作用的蛋白质。其特点①重组到 AD 载体的片段不是基因全长的 ORF, 取而代之的是

DNA 随机片段;②筛选需要多轮进行,每一轮筛选出的阳性克隆均需要序列测定,确定“猎物”的可信度,高可信度的“猎物”再作为下一轮筛选的“诱饵”;③ 筛选过程繁琐复杂,但可以对相互作用结构域进行定位。

2.3 假阳性和假阴性 酵母双杂交系统在蛋白质组的研究中具有强大的功能。但是必须指出,该系统也存在自身局限性,不同的实验针对同样的目标得出的结果有时存在很大差异。例如,同样是高通量的酵母全蛋白质组的相互作用分析,Uetz 和 Ito 两个工作组得到的结果中只有一小部分(141 个蛋白质相互作用对)是重叠的,分别占筛出的相互作用蛋白质对总量的 21.6% 和 16.9%^[6]。再如,同样是在基因组范围内筛选与拼接蛋白 LSM2, LSM4 和 LSM8 相互作用的蛋白,分别采用阵列筛选和文库筛选两种策略,结果得到的“猎物”蛋白系列也只有一小部分是重叠的^[5]。造成这种差异的原因在于酵母双杂交系统本身存在大量的假阴性和假阳性问题。对于假阳性,可以通过多次重复实验、增加报告基因的数量或者改变报告基因的强度等方法进行改进但也难以完全消除;对于假阴性,主要是由于酵母双杂交系统本身设计的局限性而使一些蛋白之间的相互作用难以检测出:如定位在细胞膜上的蛋白、转译后需要加工修饰的蛋白以及对酵母细胞本身有毒害的蛋白等。这些蛋白功能的确定需要免疫共沉淀等生化方法以及衍生的双杂交系统如 SRS 系统、逆向双杂交系统、split-ubiquitin 系统、RNA 聚合酶Ⅲ为基础的双杂交系统等等进行补充^[8]。

3 巨大的蛋白质相互作用网络中的酵母双杂交系统的位置

大规模的双杂交系统无疑对蛋白质组相互作用图谱的绘制起到重要的推动作用。但是,蛋白质相互作用网络可能比我们的想象要大得多。如科学家预计一个细胞中平均每个蛋白同其它蛋白之间的相互作用约有 2~10 个,而我们现在发现的仅占其中的约 7%^[9],单一的酵母双杂交系统不可能解决所有问题。另外由于存在大量的假阳性及假阴性,它得出的实验结果在用别的方法进一步证明之前,只能作为一种参考,而不是一种结论。所以,它需要同别的方法有效地结合,以减少该方法自身本底的影响,得到高可信度的结果。

最近, Tong 等人^[10]在这一方面做了很好的尝试。他用一种系统的方法证实了酵母细胞中与肽识别结构域 SH3 特异结合的蛋白质相互作用网络。其做法如下:首先利用噬菌体显示技术从随机肽库中筛选出能够与 SH3 结构域特异结合的肽链,测定这些肽链的一致序列;以一致序列为基础,通过编写 PSSM (position-specific scoring matrix) 程序以及利用 BIND (Biomolecular Interaction Network Database) 计算推测出酵母细胞中包括这些一致序列在内的能够与 SH3 结构域特异结合的蛋白质,并由此描绘出可能的蛋白质-蛋白质相互作用网络;第二步,利用大规模的双杂交系统(即阵列筛选和文库筛选),用实验的方法在酵母细胞中筛选出与 SH3 结构域特异结合的蛋白质,描绘出其相互作用网络;最后,将两种方法得到的资料通过正交合并得到的重叠部分确定为 SH3 结构域蛋白质相互作用网络,并将其中关键的相互作用用免疫共沉淀法等进一步测定。这种系统的方法最大的特点是将两种独立的研究方法得到的蛋白质相互作用网络进行合并得到其中重叠的部分。首先两种方法是高度正交的,即具有各自的特点:如噬菌体显示技术是利用体外短肽结合的特点,而酵母双杂交系统则根据不完全是,有的可发现某一结构域的相互作用的体内结合;噬菌体显示技术得到的蛋白质相互作用网络

是通过计算的方法得出的,但由于使用的一致序列是推测而得的,所以带有一定的不确切性;而酵母双杂交系统得到的蛋白质相互作用网络的一些结果又可能有明显的假阳性。将这两种独立的方法进行合并,明显地减少了错误的发生率,有效地提高了可信度^[10]。

4 结束语

目前,大规模的酵母双杂交系统仍在蛋白质组相互作用图谱及功能图谱的绘制方面发挥着重要作用。但是,蛋白质相互作用网络是巨大的,如生物分子相互作用网络数据库到 2002 年 8 月为止,所记录的蛋白质-蛋白质之间的相互作用数接近 12,000 个 (<http://www.bind.ca>),并且这个数字还会在不断增加。大规模的酵母双杂交筛选不能解决蛋白质相互作用的所有问题,它需要其它的研究方法给以补充,其中包括日常的双杂交衍生系统,也包括其他任何一种蛋白质-蛋白质相互作用的研究方法(如蛋白质芯片、大规模的质谱分析、免疫共沉淀及噬菌体显示技术等等)。将这些方法有机结合,取长补短,对于巨大的蛋白质相互作用网络的构建与完善将发挥积极重要的作用。

参考文献

- [1] Gavin A C, Bosche M, Krause R, *et al.* Nature, 2002, 415: 141 ~ 147.
- [2] Fields S, Song O. Nature, 1989, 340: 245 ~ 246.
- [3] Bartel P L, Roecklein J A, SenGupta D, *et al.* Nat Genet, 1996, 12: 72 ~ 77.
- [4] Finely R L Jr, Brent R. Pro Natl Acad Sci USA, 1994, 91: 12980 ~ 12989.
- [5] Uetz P, Giot L, Cagney G, *et al.* Nature, 2000, 403: 623 ~ 627.
- [6] Ito T, Chiba T, Ozawa R, *et al.* Pro Natl Acad Sci USA, 2001, 98: 4569 ~ 4574.
- [7] Fromont-Racine M, Rain J-C, Legrain P. Nat Genet, 1997, 16: 277 ~ 282.
- [8] 马海蓉, 李维琪. 中国生物工程杂志, 2002, 23: 37 ~ 41.
- [9] Bader G D, Donaldson I, Wolting C, *et al.* Nucleic Acids Research, 2001, 29 (1): 242 ~ 245.
- [10] Tong A M Y, Drees B L, Nardelli G, *et al.* science, 2002, 295: 321 ~ 324.
- [11] Gerstein M, Lan N, Jansen R. Science, 2002, 295: 284 ~ 287.