

丝状噬菌体与噬菌体展示技术

黄仪秀 朱圣庚

(北京大学生命科学学院 北京 100871)

丝状噬菌体的利用,在分子生物学研究以及基因工程发展中起了重大作用^[1]。丝状噬菌体作为载体具有多方面的巨大应用潜力。1985年Smith^[2]最先将外源基因插入丝状噬菌体 Π 的基因III,使目的基因编码的多肽能以融合蛋白的形式展示在噬菌体表面,从而创建了噬菌体展示技术。噬菌体展示技术对生物学许多领域的研究将带来巨大影响,因此它已被誉为生命科学研究中又一次技术上的重大突破。

1 丝状噬菌体的生物学特征

丝状噬菌体的形态结构^[1]:丝状噬菌体主要包括大肠杆菌丝状噬菌体M13、fd、f1等,属于丝杆噬菌体科(Inoviridae),丝杆噬菌体属(Inovirus)。它们的亲缘关系十分密切,特性相似,都含一分子单链环状DNA,且DNA同源性高达98%以上。噬菌体颗粒形状、大小相近,都属雄性噬菌体,仅感染雄性细菌。

M13噬菌体呈长纤维状的柔性长丝,长约860nm,直径约6nm。外壳蛋白围绕基因组DNA呈规则螺旋排列。它的基因组ssDNA由6407个脱氧核苷酸组成,其DNA链内并不自我互补,只有少数碱基配对。基因组共有十个基因,分别编码10种不同的蛋白质。基因VIII(g8)所编码的gp8蛋白是噬菌体的主要外壳蛋白,约含2700个拷贝,围绕DNA呈螺旋对称的管状排列。基因III(g3)、VI(g6)、VII(g7)、IX(g9)分别编码gp3、gp6、gp7、gp9蛋白,它们在外壳蛋白中各有约5个拷贝。gp7与gp9构成噬菌体外壳顶端“塞子”,gp3与gp6蛋白构成噬菌体外壳的尾端,基因II(g2)所编码的gp2蛋白是亲代RF DNA位点特异性缺口酶,基因V(g5)所编码的gp5蛋白是ssDNA合成时螺旋去稳定蛋白。

基因I(g1)、IV(g4)所编码的gp1、gp4蛋白可能涉及噬菌体装配。基因X(g10)所编码gp10蛋白是基因III启动子的激活蛋白。

丝状噬菌体的生活周期^[1]:在侵染宿主细胞时,位于丝状噬菌体末端gp3蛋白,识别和吸附在大肠杆菌性菌毛的末端。然后其主要外壳蛋白脱落,噬菌体DNA被注入宿主细胞内,进入胞内的噬菌体DNA为正链,并以此为模板复制出负链,生成双链复制型DNA(Replicative Form DNA, RF DNA)。基因组的转录及复制都是以亲代RF DNA的负链为模板。其DNA先进进行几轮 θ 型复制后即进入类似 Φ X174的滚环复制。在进入滚环复制时,先由gp2蛋白在亲代RF DNA的正链特定位点上切开一个缺口,以环状负链DNA为模板,滚环复制出正链DNA。当复制完成一圈时,再由gp2蛋白切出单位长度的正链,并经环化形成噬菌体环状ssDNA的基因组。在复制后期,新合成的正链DNA被gp5蛋白的二聚体所包被。丝状噬菌体并不在细胞内组装成噬菌体颗粒,它们的外壳蛋白(如gp8蛋白等)合成后,预先锚在细胞质膜上,当DNA与gp5蛋白的复合物移动至细胞质膜时,gp5蛋白即被外壳蛋白所取代。子代噬菌体被释放至胞外,而宿主细胞并不被裂解。丝状噬菌体与大多数烈性噬菌体不同,既不杀死宿主细胞,也不引起细胞裂解,但却大大降低宿主细胞的生长速度(被感染细胞的生长速率仅为正常细胞生长速率的1/2~3/4)。

2 噬菌体展示技术原理

1996-08-29收稿

噬菌体展示 (phage display) 是一种基因表达产物和亲和选择相结合的技术^[1]。它是将目的基因与编码噬菌体外壳蛋白基因通过接头 (linker) 相连, 并插入到噬菌体的表达载体上, 从而使多肽或蛋白质与外壳蛋白融合并展示在噬菌体的表面。被展示的多肽或蛋白质可保持相对独立的空间结构和生物活性。通过反复的亲和选择和扩增, 可直接测定所展示的多肽或蛋白质的某些生物活性, 并分离出带有目的基因的噬菌体^[4,5]。

在丝状噬菌体所编码的 10 个蛋白质中, 与噬菌体展示有关的主要蛋白质是外壳蛋白 gp3 和 gp8。外源基因插入到外壳蛋白基因 III 或基因 VIII 时, 外源多肽与 gp3 或 gp8 的 N 端连接, 能以融合蛋白形式展示在噬菌体的表面 (图 1)。

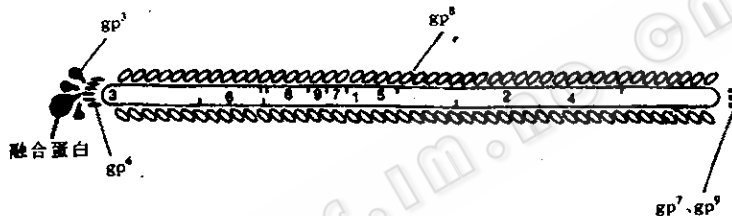


图1 噬菌体展示的示意图

贝数只有 3~5 个, 而 gp8 的拷贝数多达 2700 个左右。因此用 gp8 可筛选低亲和力的配体, 而用 gp3 则可筛选高亲和力的配体, 故大多数研究者常利用 gp3 作为外源多肽或蛋白的融合部位。

由于在 M13 噬菌体的基因组中只含有一个拷贝的基因 III 和基因 VIII, 因此, 当外源基因与基因 III 或基因 VIII 的 N 端连接时, 其重组基因产物都将形成 gp3 或 gp8 与外源蛋白的融合蛋白。其结果 gp8 的融合蛋白会妨碍外壳蛋白的正常功能; gp3 的融合蛋白则无法识别大肠杆菌的性菌毛, 使噬菌体不能感染大肠杆菌而无法增殖。目前常采用两种方法以解决上述存在问题^[6]。

(1) 在噬菌体的基因组采用双拷贝的基因 III 或基因 VIII (类型 33 或类型 88)。使一个拷贝为正常的 gp3 (或 gp8) 基因, 另一个拷贝为 gp3 (或 gp8) 与外源蛋白的重组基因, 表达产物既有少数 gp3 (或 gp8) 与外源肽段的融合蛋白, 又有正常的 gp3 (或 gp8), 以保证噬菌体对大肠杆菌的识别与正常组装。

(2) 采用噬菌质粒 (phagemid) 和辅助噬菌体

外壳蛋白 gp3 位于噬菌体的最尾端, 呈球形或线状。分子量为 42ku, 含 406 个氨基酸。其 C 端锚在外壳上, N 端游离在外。外源多肽或蛋白质与 gp3 的 N 端连接后, 可独立展示其结构域。

gp8 蛋白是噬菌体的主要外壳蛋白, 位于噬菌体的外侧, 分子量为 5.2ku, 含 50 个氨基酸, C 端与 DNA 结合, N 端伸出噬菌体外, 外源小肽可与 gp8 的 N 端融合。融合蛋白必须运送到细胞质膜上, 切除信号肽后, 才能被组装而展示。

gp3 和 gp8 的主要差别在于: 一是融合外源肽段大小的能力不同, gp3 可融合较大的外源肽段, 大至 50ku 的蛋白质已被成功地展示。而 gp8 只能融合较小的外源肽段, 如 8 肽或 9 肽, 携带肽段太大会影响噬菌体粒子的组装和侵染能力。二是拷贝数不同, gp3 的拷

(helper phage) 的双基因组系统 (类型 3+3 或类型 8+8)。在噬菌质粒 DNA 中含有外源肽段与 gp3 (或 gp8) 的重组基因, 和便于筛选的抗药性基因; 此外还含有质粒复制起点和噬菌体的基因间隔区, 可被辅助噬菌体基因 II 产物有效识别和组装。辅助噬菌体的基因组中含有正常的 gp3 和 gp8 基因, 同时还含有 DNA 复制和组装所需的全套基因。此外, 由于在其基因间隔区插入了 lacZ' 序列, 因而 gp2 蛋白对它不能有效识别。当含有重组基因的噬菌质粒进入大肠杆菌后, 再用辅助噬菌体进行感染, 这样, 所产生的子代噬菌体既展示 gp3 (或 gp8) 与外源肽段的融合蛋白, 又因同时含有正常的 gp3 和 gp8, 使噬菌体能够进行正常感染、组装和增殖。此外, 由于辅助噬菌体合成的 gp2 蛋白优先识别噬菌质粒上的正常基因间隔区, 启动滚环复制, 产生 ssDNA。通常噬菌体外壳蛋白所包装的 ssDNA 主要来自含有重组基因的噬菌质粒 DNA (一般多于包装辅助噬菌体基因组的 50 倍)。

当以一组随机多肽编码序列或基因群插入噬菌体

载体进行展示表达时,其总体称为噬菌体展示文库(phage display library)。在噬菌体展示文库中的每一个噬菌体粒子只展示一种序列的外源肽链,不同噬菌体粒子展示不同的序列的外源肽链。目前所获得噬菌体展示文库的最大容量均在 10^9 左右,这是受DNA转化效率所限制的。

噬菌体展示技术中另一个重要环节,就是如何从噬菌体展示文库中高效挑选出带有目的基因的重组噬菌体(fusion phage),这个过程称为淘选(panning)。一般采用亲和选择方法,即利用在噬菌体表面上展示的多肽可与特定的选择体(selector)进行亲和结合,从而得到分离。通常先将选择体固相化,如将选择体预先加在一块96孔板的孔中或预先结合在亲和柱上,然后使文库中重组噬菌体所展示的目的多肽与固相化的选择体亲和结合。通过洗脱,即可收集到带有目的多肽或蛋白质的重组噬菌体,并可利用这些噬菌体再感染大肠杆菌而得到进一步扩增。通过一轮亲和结合—洗脱—扩增,即所谓“淘选”的富集过程,就可富集约 $10^2 \sim 10^3$ 倍。经过几轮“淘选”,可以从噬菌体文库中筛选到带有目的基因的重组噬菌体。这是一种前所未有的高效快速选择方法。最后将选择到的重组噬菌体进行测序,读出该噬菌体所展示目的多肽或蛋白质的基因编码序列和氨基酸序列。同时可从重组噬菌体DNA中将目的基因切下,再克隆到其他宿主细胞中,经表达以获得大量的目的多肽或蛋白质。

噬菌体展示技术的意义在于:它能按照空间结构互补关系从数量巨大的多肽序列或多肽构象库中直接选择出任意所需要的对象;被选择的多肽展示在噬菌体表面,在获得多肽的同时也得到了编码该多肽的基因序列;用亲和选择(affinity selection)代替筛选(screen),被吸附的是亲和力强的融合噬菌体,其余均被淘汰,故称为淘选,由此大大简化了分离多肽及其基因的过程。这一过程与动物体内抗原对抗体克隆选择的原理相同,也与生物分子进化的自然选择相类似。从本质上说,通过噬菌体展示技术选择特异蛋白和基因是对生物获得信息过程的模拟。

3 噬菌体展示技术的应用

噬菌体展示技术是在80年代末、90年代初才兴起的重要生物技术,然而它已在广泛领域内取得重要进展,并正在不断开拓新的应用领域。就目前而言,它的

应用主要有以下几方面:

(1)制备人工抗体。噬菌体展示技术系统最先在制备特异抗体中取得成功。1990年McCafferty等^[7]从淋巴细胞中分离纯化出抗体轻链和重链可变区(V区)基因,以编码柔性链($\text{Gly}_4\text{—Ser}$)₃的序列相连,克隆到丝状噬菌体(fdCAT1)gp3的N端,使单链抗体($V_H V_L$)以融合蛋白形式在噬菌体表面展示。用固定化抗原极易从噬菌体抗体库中选择出特异抗体基因克隆^[8]。噬菌体抗体不仅能以单链形式(scFv)展示,而且也能以双链形式(Fab)展示。具体方法是将重链(V_H 和 C_H)以及轻链两者之一表达在噬菌体gp3的N端,另一条链通过质粒来表达,待两者合成后分别运送到质膜外而结合,并在噬菌体表面得到展示^[9]。噬菌体抗体较之单克隆抗体是又一次质的飞跃。有人将免疫动物获得的多克隆抗体称为第一代抗体,单克隆抗体为第二代抗体,噬菌体抗体为第三代抗体。

噬菌体抗体与杂交瘤单克隆抗体相比较有以下突出优点:①它可通过亲和吸附直接从抗体库中选择出特异抗体基因,免除杂交瘤繁杂庞大的逐个筛选过程,因而省时省力。杂交瘤筛选的细胞克隆数约为 $10^2 \sim 10^3$,而噬菌体抗体库筛选的基因克隆达 $10^8 \sim 10^9$,大大提高了获得目的抗体的机率^[10]。而且得到的抗体基因可直接用来构建各种基因工程抗体。②无需动物免疫。为从未经免疫的淋巴细胞中获得特异的抗体及其基因,常采取两项措施:一是通过多轮淘选,反复多次即能从容量很大的库中选出高亲和力的特异性抗体。二是模拟体内亲和性成熟(affinity maturation)过程,在抗体基因中引入随机或准随机突变,增加抗体基因库的多样性。③可获得人源抗体^[11]。由于人体杂交瘤系统低效和人体不能任意免疫,获得人单克隆抗体非常困难。噬菌体抗体库可包含人全套抗体基因,并可直接用于亲和选择,这就解决了抗体人源化问题。

(2)确定抗原决定簇^[12]。由于蛋白质抗原中一般只有3~8个氨基酸残基直接参与抗体的结合,因此可用随机多肽库来确定单克隆抗体在其相应蛋白质抗原上的结合位点(表位)。首先合成一定长度(9~24nt)随机序列的混合寡核苷酸,融合表达在噬菌体外壳蛋白上,然后亲和选择出与抗体结合的噬菌体,测定噬菌体DNA中编码展示多肽的序列,推算出能与抗体结合的各噬菌体肽段的氨基酸序列。如果与抗体反应的肽段

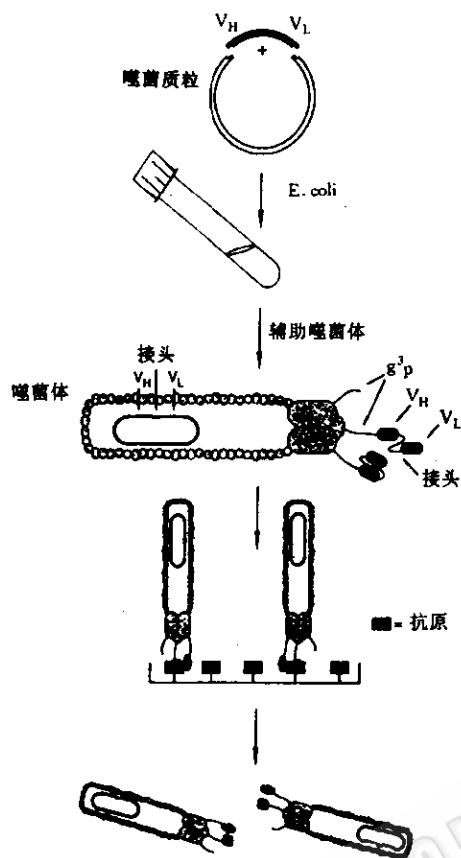


图2 噬菌体抗体库的构建与特异抗体的选择

有一个共同序列 (consensus sequence), 并且此序列与抗原的某一段氨基酸序列相同, 此序列即为序列特异性抗体的抗原决定簇 (线性表位)。然而, 对于构象特异性抗体来说, 与抗体结合的肽段可能有一个或多个共同序列, 这些共同序列既可能与抗原的某一段序列相同, 也可能在抗原上找不到相同的肽段。前一种情况是噬菌体展示多肽分段模拟构象型决定簇的结果; 后一种情况表明, 抗原决定簇间由一些在一级结构上相距很远, 而在空间结构上由于折叠而互相靠近的氨基酸残基所组成 (不连续表位)。利用噬菌体展示技术亦已确定了多种单克隆抗体的抗原决定簇。

(3) 用以研究蛋白质的相互识别和相互作用^[13]。正如抗原与抗体之间的相互识别和作用一样, 生物大分子之间和大分子与小分子之间, 如酶与底物、酶与抑制剂或激活剂、受体和配体、调节蛋白与靶结构、寡聚蛋白的单体之间识别和作用, 都可用类似的方法进行研究, 找出相互作用的互补空间结构, 已由此获取对人生

长激素、粘合素和生物素有高亲和力的多肽。某些抑制剂和激活剂可作为新药的来源。该技术还可用来研究蛋白质折叠过程各肽段和结构域之间的相互作用。

(4) 分离特定结构与功能的蛋白质^[14, 15]。由 cDNA 展示库中可直接选择出特异的蛋白质及其基因。例如, 用配体亲和选择出特异的受体; 用靶结构 (DNA 的调控序列, 蛋白质的调节亚基等) 亲和选择出特异的调节蛋白; 用癌基因产物 Jun 及 Fos 亮氨酸拉链结构找出与其相结合的蛋白。

(5) 由多肽构象库或蛋白质突变库中获取特定功能的蛋白质。这种蛋白质或是根本不存在于自然界; 或是野生型蛋白质改进性能的突变体。多肽的结构域至少需要有 30~40 个以上氨基酸残基组成才能有足够的凝聚力维持其结构稳定性。当不同的序列进行展示时即可得到多肽构象库。基因的序列内引入大量的各种突变可构成基因突变库。抗体和酶有许多性质相同, 如果由底物反应的过渡态获得其抗体, 该抗体可能具有催化底物反应的功能, 亦即具有酶促功能的抗体, 称为抗体酶 (abzyme)^[16]。这种抗体酶可以从噬菌体抗体库中选择得到。一些特殊性能的药物也可以由多肽构象库或突变库中选择出来^[13]。而当一个噬菌体展示库重复进行选择、突变和扩增, 由此可实现生物分子的快速进化, 称为实验定向进化或生物进化工程^[17]。一些著名的实验室正在用此技术获取新的抗凝溶栓药物和其他种类新的药物。这是噬菌体展示系统的又一个重要应用领域。

4 结束语

噬菌体展示的关键技术首先是构建展示库。从理论上讲, 只要构建的库足够大, 就可以从中筛选到任意所需要的多肽序列和构象。实际上并非任意序列都有生物学意义, 因此对于展示库不仅要有较大容量, 还要考虑其异质性结构。此外, 还应根据不同需要以构建不同的展示库, 如随机库、突变库、cDNA 库等。表面展示技术也可展示于其他系统, 其中 T4 和 λ 噬菌体展示系统尤引人注目, 它们不仅可不经分泌在细胞内组装后再释放, 还能在体外进行组装, 因此更适合于特殊产物的工程应用。有关选择方法也不断有新的发展, 例如将多肽结合于细胞表面受体, 可直接以细胞来进行选择。随着噬菌体展示技术的发展, 人们可以在更大程度上根据需要来改造和创建各种生物分子。

参 考 文 献

- [1] Brock T D, Madigan M T, Martinko J M *et al.* Biology of Microorganisms, 7th edition. Prentice-Hall International, Inc., New Jersey, 1994.
- [2] Smith G P. Science, 1985, **228**:1315~1317.
- [3] Smith G P, Scott J K. Methods in Enzymology, 1993, **217**:228~257.
- [4] Scott J K. Trends in Biochem Sci, 1992, **17**:241~245.
- [5] Schatz P T. Cur Opin Biotechnol, 1994, **5**:487~494.
- [6] Smith G P. Gene, 1993, **128**:1.
- [7] McCafferty J, Griffiths A D, Winter G *et al.* Nature, 1990, **348**:552~554.
- [8] Winter G, Milstein C. Nature, 1991, **349**:293~299.
- [9] Hoogenboom H R, Griffiths A D, Johnson K S *et al.* Nucleic Acids Res, 1991, **19**:4133~4137.
- [10] Winter G, Griffiths A D, Hawking R E *et al.* Annu Rev Immunol, 1994, **12**:433~455.
- [11] Marks J D, Hoogenboom H R, Bonnert T P *et al.* J Mol Biol, 1991, **222**:581~597.
- [12] Scott J K, Smith G P. Science, 1990, **249**:386~390.
- [13] Jespers L S, Messens J H, De Keyser A *et al.* Bio / Technology, 1995, **13**:378~382.
- [14] Cramer R, Suter M. Gene, 1993, **137**:69~75.
- [15] Rebar E J, Pabo C O, Science, 1994, **263**: 671~673.
- [16] Janda K D, Lo C-H L, Li T *et al.* Proc Natl Acad Sci USA, 1994, **91**:2532~2536.
- [17] Wells J A, Lowman H B. Curr Opin Biotech, 1992, **3**:355~362.