

指数富集配基的系统进化原理及应用进展

刘 晓 李长友 陈元鼎*

中国医学科学院 医学生物研究所, 昆明 600118
(中国协和医科大学)

摘要 过去十几年, 指数富集配基的系统进化 (SELEX) 技术得到了充分的发展, 目前已经成为在数量众多的靶分子中, 高亲和性和特异性地确定适配子的一种崭新的方法. SELEX 的流程即是人工合成随机寡核苷酸序列、适配子筛选、扩增、再循环的过程. 在这一过程中, 适配子修饰基团文库的应用使得 SELEX 流程更加经济、快捷、高通量. 回顾了 SELEX 的研究和应用, 期待着这一方法能够在理论和实践上得到进一步的发展.

关键词 指数富集配基的系统进化, 适配子, 原理, 应用

学科分类号 Q81

指数富集配基的系统进化 (systematic evolution of ligands by exponential enrichment, SELEX) 技术^[1], 是指寡核苷酸依靠自身对相应靶物质的高亲和力和高特异性, 在一含有极多随机序列 (包括靶序列) 的寡核苷酸池中, 通过多次的循环选择富集, 来获取所需要的特定结构和/或功能适配子的过程. 其中的适配子 (“aptamers” 由拉丁词根 *aptus* 意为匹配和希腊后缀 *-mer* 相连而成), 通过柱层析等选择性技术, 从绝大部分非功能性 RNA 或 DNA 分子文库中被选择出来. 这一技术适用于富集任何特殊需要的生命小分子物质.

SELEX 技术的发展基于上世纪 70 年代和 80 年代的单克隆、核酶、PCR 技术, 这是核酸定向进化的必然方向和阶段性成果. 1990 年, 美国的 Joyce, Szostak 和 Gold 三家实验室各自独立地建立起核酸文库, 并逐步发展成一项能从超过 10^{14} ~ 10^{15} 核酸分子文库中, 同时扫描独立的有着不同功能的核苷酸分子, 筛选出具有特异性结合能力配基的新技术, 它被分别称作: 体外选择 (*in vitro* selection), 体外进化 (*in vitro* evolution), 指数富集配基的系统进化 (SELEX). 它们的基本原理和操作步骤都类似, SELEX 的称法最为通用. 但直到 Ellington 等^[2] 和 Tuerk 等^[1] 分别利用随机 RNA 文库筛选出染料分子及 T4 DNA 聚合酶的特异寡核苷酸配基, 这项技术才逐渐引起人们越来越多的重视. 至今已衍生出多项子技术.

1 SELEX 技术的原理

SELEX 技术, 即为适配子筛选、扩增、再循环的过程, 主要分以下 4 个步骤: a. 合成随机寡

核苷酸序列文库 (含适配子); b. 选择性分离出同靶物质高亲和性和特异性结合的适配子; c. 用基于转录的技术来扩增适配子; d. 用亲和层析或硝酸纤维素膜过滤等方法, 通过多次再循环来得到适配子.

1.1 SELEX 技术的基本原理

首先是利用现有的分子生物学技术, 人工合成一个含有 10^{14} ~ 10^{15} 个单链寡核苷酸序列的随机文库, 序列长度往往在 25~35 之间. 而单链的随机寡核苷酸序列, 尤其是 RNA, 容易形成可与蛋白质、核酸等靶物质特异性共价结合的二级结构, 如发卡 (hairpin)、茎环 (stem-loop)、G-四聚体 (G-tetramer)、假节 (pseudoknot)、鼓包 (bulge) 等等. 在这一高亲和力特异性结合的基础之上, 靶物质如氨基酸、多肽、甚至金属离子都可以同随机文库 (池) 相互作用, 选择性分离出适配子 (aptamers), 然后通过 RT-PCR 等建立在转录基础上的技术, 生成次一级文库, 再与靶物质结合, 反复多次循环, 即可获得与靶物质特异性高亲和力结合的适配子.

在这一过程中, 随机序列寡核苷酸文库在一个给定温度的选择性缓冲液池中, 可以同靶物质亲和作用的特异性序列非常少, 只有通过物理分离技术加以纯化. 大分子如蛋白质可以通过硝酸纤维素膜过滤; 小分子如核酸、多肽等, 可以通过简单的修饰, 添加上可以与固相支持物产生亲和作用的基

* 通讯联系人.

Tel: 0871-8334689, E-mail: chenyd@km169.net

收稿日期: 2002-04-28, 接受日期: 2002-05-28

团, 经简单的冲洗即可以达到纯化目的. 高亲和性适配子的富集效率, 取决于多重循环中每一次循环的严谨性. 经过多次的选择性分离扩增后, 达到了饱和状态, 富集的文库被克隆, 相应的序列信息也被获得, 并在以后的工作中进一步探讨它们的性质.

适配子纯化循环的次数取决于每一次循环的严谨程度. 一般而言, 对于大多数靶物质在 8~15 次循环中即可得到亲和富集, 2 天内可完成一次 SELEX 循环, 包括克隆和测序. 一个典型的 SELEX 过程可能持续 2~3 个月. 一旦序列确定, 适配子可由化学合成来生产, 少量适配子的纯化和分析不超过 3 天, 可以产生足够数量 (几个摩尔) 的适配子用于后继的开发和研究. 最近 SELEX 过程实现了自动化, 将人为因素的干扰减少到最低. 在微型自动化分析仪上可同时执行多个独立的重复 SELEX 过程, 使适配子的发现更加快捷、经济和高通量.

1.2 适配子的序列结构

SELEX 过程中的适配子是包括修饰序列的单链寡核苷酸序列, 通常由三个部分组成: 3' 和 5' 端的固定序列以及中间的功能随机序列.

固定序列存在的目的在于增加文库的稳定性和为扩增作准备, 同时固定序列中往往含有较多的 G/C 碱基对, 有利于在以后的 PCR 过程中提高退火温度. 引物中往往含有特定的限制性酶切位点和 T7 启动子, 以利于体外转录、克隆、测序. 这一部分往往被截短来降低核苷酸的延展性, 并且对直接与靶物质相互作用, 或形成一个同靶物质相互作用的结构并不重要. 目前的技术已经可以在 SELEX 过程中的随机序列中删减功能适配子. 中间序列即为适配子特定功能区, 重点为修饰基团的选择和应用, 这一点我们将在下文中进一步讨论.

随机核苷酸片段不能太短, 太短可能导致没有足够的二级结构同靶分子相结合, 同时库容量太小; 随机寡核苷酸片段也不能太长, 太长可能由于全套文库单链寡核苷酸序列数量一定, 导致降低了选择分离出来的配基概率, 同时对整个反应体系的 RNA 量要求太高. 因此, 随机文库长度多为几十到几百个碱基, 一般为 70~80 个核苷酸, 但也有少数随机 RNA 文库中有二个以上的随机序列.

1.3 SELEX 的修饰基团文库

目前的修饰基团大致可以分为以下几个部分: 抗核苷酸酶修饰基团、有利于适配子复制、扩增的

修饰基团、结构类似物修饰基团、可通过光交联或不通过光直接交联的修饰文库.

1.3.1 抗核苷酸酶修饰基团: 这方面最为成功的修饰为: 在嘧啶 C-5 环 2' 位上导入氨基 (—NH₂) 或氟基 (—F) 等功能基团, 在生物液中成功地提高了适配子抗核苷酸的能力. 在嘧啶 C-5 位上的结构类似物的替代, 可能对适配子的结构造成一定的影响, 如苯甲酰基团的替代, 提高了适配子对 SELEX 过程中所用酶的耐受性, 这一点可以极大地提高核苷酸文库的各种组合几率. 但硫代磷酸的修饰是个例外, 因为硫代磷酸修饰的寡核苷酸可非特异性地与蛋白质相互作用, 因此不能用它来确定特异性配基^[3].

1.3.2 有利于适配子复制、扩增的修饰基团: 这一类配基的存在, 可以极大地提高核苷酸文库的分子差异性, 以利于适配子的特异性和亲和性相互作用. 这方面有许多精妙的例子: Piccirilli 等^[4] 和 Bain 等将异构 G/异构 C 和 2,6-二氨基嘧啶/黄嘌呤导入 DNA 和 RNA, 目的在于扩充基因编码; Tor 等^[5] 利用体外翻译将甲基 C/ (6-氨基) 异构 G 来介导一个 RNA 序列的单一位点上的氨基修饰, 也可以通过生物素或 EDTA 等, 将一个单一烷基胺基团导入 RNA 转录后的位点来达到人为控制反应的目的.

1.3.3 结构类似物修饰基团: Schweitzer 等^[6] 和 Guckian 等发现了一个新的碱基对类似物, 在识别中不需要氢键, 非极性地形形成针对 AT 碱基对的类似物 (6-甲基嘌呤和 2,4-二氟甲苯), 并由酶介导与 DNA 相互作用, 进一步发挥促进核苷酸复制的功能. 用经过修饰的嘧啶核苷类似物 5- (1-戊基) -2-脱氧尿嘧啶的适配子可以选择性地识别人凝血酶.

1.3.4 可通过光交联或不通过光直接交联的修饰文库: 含有某些修饰的核苷酸可以由紫外光处理来获得一些活性基团, 这些基团可能同邻近的其他分子有共价联系, 这一种特殊类型的适配子被称之为光交联的适配子. 含有 5-溴尿嘧啶和 5-碘尿嘧啶的修饰基团可以被用于产生交联适配子. 这一反应发生在核苷酸的碱基 C-5 环和相邻且结构适合的电子密集的氨基酸 (如 Try、Tyr、His、Phe、Cys 等) 之间. 5-溴尿嘧啶和 5-碘尿嘧啶分别在 308 nm 和 325 nm 波长的单色光源下活性最大化^[7]. 在无光条件下, 偶然发现可与靶物质发生共价交联的适配子. 虽然这些非依赖性交联的化学过程不需要交

联的光源, 却一样需要 5-碘尿嘧啶, 在诊断应用中具有光交联适配子的所有优势^[8].

2 SELEX 的应用进展

SELEX 技术诞生于上世纪 90 年代初. 当时正处在以人类基因组计划 (human genome plan, HGP) 为代表的分子生物学理论和技术大发展时期. 十余年来 SELEX 技术的发展也自然地同后基因组计划, 尤其是蛋白质组学的发展紧密联系在一起.

2.1 基因调控与信号传导研究

基因调控与信号传导的探索和研究一直是生命科学的前沿和热点, SELEX 技术的发展自然与此不能分割. Turnage 等^[9]用 SELEX 技术发现了黑腹果蝇 (*Drosophila melanogaster*) 基因组中的两个 Arg 富集区, 介导了翻译过程中的 RNA 连接, 并且通过黑色基因 SU (s) 的抑制来调控编码蛋白的活性. Emmerich 等^[3]用 SELEX 技术发现日本山茶 (*japonicum*) 慢根瘤菌中, 一个不完全的插入重复序列 (GCGGG-N₍₅₎-GTTCGC), 对 DNA 连接的反应调节子 RegR 的作用具有关键性意义, 这对固氮技术的研究与发展有着重要的意义. Xu 等^[10]利用 SELEX 技术发现, Hypusine 在序列特异性的真核启动子 5A (eIF-5A) 功能作用中是必不可少的. 这一研究表明 eIF-5A-RNA 作用的生理信号依赖 Hypusine 和靶 RNA 的核心基序. Man 等^[11]利用 SELEX 的变种荧光相关亲和分析研究了伤寒噬菌体抑制子 Mnt 与 DNA 操纵子之间的作用. Amarasinghe 等^[12]作用在剪接抑制子 IS3 上的蛋白 PSI (P element somatin inhibitor) 中, 有一个 RNA 体外连接的选择性位点, 存在于它的 KH 结构域上. Libri 等^[13]利用 SELEX 研究了酵母内含子 5' 剪接位点 (5SS) 识别中, 通过 U1 的 snRNP 错配现象, 解释了主要通过抑制结合来稳定 suRNA/5SS 错配的相互作用.

2.2 蛋白质功能的研究

Shtatland 等^[14]利用基因组 SELEX 研究了噬菌体 MS2 衣壳蛋白, 同 *E. coli* 外膜形成相关基因 rffg, 和乳糖利用抑制子 bgR 之间的相互作用, 得到了高亲和力的适配子. Wen 等^[15]研究了细菌噬菌体 Ff 基因 5 蛋白 (g5p) 高亲和力寡核苷酸的 SELEX 选择, 发现 G 富集基序对于形成 g5p 连接启动子结构, 提供了实际的 g5p 连接位点具有重要的意义. White 等^[16]用“套索 (toggle)” SELEX

得到了与猪凝血酶特异性的适配子, 这对基因诊断治疗具有重要的意义. White 等^[17]利用 SELEX 技术研究了一种 PUM-HD (短小杆菌素同源结构域) 及其同有的 RNA 连接位点, 发现在 PUM2 氨基端有 Ser 和 Glu/Ala 富集区, 并且 PEB (PUM2 连接元件) 与果蝇反应元件的 3' 端具有亲缘关系, 但又互不相同. Ulrich 等^[18]利用体外筛选与锥虫细胞粘附受体相结合的 RNA 适配子, 抑制了锥虫对细胞的侵害, 对 Chagas 病的诊断治疗有着重要的意义.

2.3 核酸研究

Wang 等^[19]利用 SELEX 发现了抗 DNA 抗体对双链 DNA (dsDNA) 和单链 DNA (ssDNA) 连接基序上的不同. 单链 DNA 含有 cacc、caccc、accc 或 cccc 区组, 双链 DNA 含有 5' gcg3' / 3' cgc5' 序列, 提示了 SELEX 技术对于研究 DNA 抗原方面具有的重要价值. Bouvent 等^[20]利用 NRE (核仁蛋白识别元件) 发现了 RNA 茎环上的 RBD1 和 RBD2 (折叠元件结构域), 这对了解模式蛋白识别 RNA 的结构过程具有重要意义. Roychowdhury-Saha 等^[21]报道了利用 RNA 适配子的靶向识别性获得了对黄素腺嘌呤核苷酸 (FAD) 特异性的黄素. 这项工作对于酶学, 尤其是氧化还原酶的研究具有重要意义.

2.4 肿瘤学研究

Blank 等^[22]研究了大鼠脑神经胶质瘤微血管内皮调节蛋白相关的 DNA 适配子的系统进化, 这对研究脑肿瘤的发生具有重要意义. Hicke 等^[23]利用肌腱蛋白-C (TN-C) 的纯化蛋白, 和表达 TN-C 的多形性胶质母细胞瘤细胞得到了 TN-C 的适配子, 对于肿瘤组织的诊断和治疗具有重要意义. Lato 等^[24]发现了 ATP 特异性的含硼适配子, 这对应用硼中子捕获治疗法 (BNCT) 进行针对肿瘤的治疗具有重要意义.

2.5 病毒学研究

近 2 年 SELEX 技术在病毒学方面的应用, 从 20 世纪 90 年代的 HIV 发展到应用在 HCV 和 HIV 两个方面. Fukuda 等^[25]提纯了对 HCV 的 NS3 (非结构蛋白 3) 蛋白水解酶具特异性的 RNA 适配子, 并检定了性质, 发现含有保守序列 GA (A/U) UGGGAU, 适配子中的茎和环对蛋白酶的抑制是必需的. Ali 等^[26]发现在 HCV 介导 IRES (核糖体进入位点) 的翻译起始中, RNA 同蛋白质相连需要 PTB (多聚嘧啶结合蛋白). Andreola

等^[27]报道在体外抗病毒活性实验中, 发现了抗 HIV-1 的 RNaseH 活性的 DNA 适配子, 这对于发展特异性抑制逆转录酶活性药物有着重要的意义. Biroccio 等^[28] 选择出了 HCV 对 RNA 依赖性的 RNA 多聚酶具有高亲和力和特异性的适配子, 这对 HCV 的诊治研究具有重要意义.

SELEX 技术是系统论和分子生物学相结合的产物. 理论上, 它可以高亲和性和特异性地识别几乎一切小分子靶物质. 正因为如此, 近年来它在生命科学的各个领域都得到了快速的发展, 并且建立起了自己的数据库系统 (www.mgs.bionet.nsc.ru/mgs/systems/selex/). SELEX 技术同经典的分子识别技术——克隆抗体技术相比, 实现了从单一靶物质筛选到多种靶物质筛选的飞跃, 具有与特异性靶物质结合力更强、特异性更高、异质性更好、筛选周期更短、条件可以操控和性质可以变化等诸多优势, 因此, 虽然 SELEX 的研究应用依然在初期阶段, 但它的进展惊人, 无论在科研、生产还是商品市场上, SELEX 技术都有着潜在的巨大价值.

参 考 文 献

- 1 Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 1990, **249** (4968): 505~ 510
- 2 Ellington A D, Szostak J W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, 1990, **346** (30): 818~ 822
- 3 Emmerich R, Strehler P, Henneke H, *et al.* An imperfect inverted repeat is critical for DNA binding of the response regulator RegR of *Bradyrhizobium japonicum*. *Nucleic Acids Res*, 2000, **28** (21): 4166~ 4171
- 4 Piccirilli J A, Krauch T, Moroney S E, *et al.* Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature*, 1990, **343**: 33~ 37
- 5 Tor Y, Dervan P B. Site-specific enzymatic incorporation of an unnatural base, N⁶-(6-aminohexyl) isoguanosine, into RNA. *J Am Chem Soc*, 1993, **115**: 4461~ 4467
- 6 Schweitzer B A, Kool E T. Hydrophobic, non-hydrogen-bonding bases and base pairs in DNA. *J Am Chem Soc*, 1995, **117**: 1863 ~ 1872
- 7 Willis M C, Hicke B J, Uhlenbeck O C, *et al.* Photocrosslinking of 5-iodouracil-substituted RNA and DNA to proteins. *Science*, 1993; **262**: 1255~ 1257
- 8 Jensen K B, Atkinson B L, Willis M C, *et al.* Using *in vitro* selection to direct the covalent attachment of human immunodeficiency virus type 1 Rer protein to high-affinity RNA ligands. *Proc Natl Acad Sci USA*, 1995, **92**: 12220~ 12224
- 9 Turnage M A, Brewer-Jensen P, Bai W L, *et al.* Arginine rich regions mediate the RNA binding and regulatory activities of the protein encoded by the *Drosophila melanogaster* suppressor of sable gene. *Mol Cell Biol*, 2000, **20** (21): 8198~ 8208
- 10 Xu A, Chen K Y. Hypusine is required for a sequence-specific interaction of eukaryotic initiation factor 5A with postsystematic evolution of ligands by exponential enrichment RNA. *J Biol Chem*, 2001, **276** (4): 2555~ 2561
- 11 Man T K, Stormo G D. Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay. *Nucleic Acids Res*, 2001, **29** (12): 2471~ 2478
- 12 Amarasinghe A K, MacDiarmid R, Adams M D, *et al.* An *in vitro*-selected RNA-binding site for the KH domain protein PSI acts as a splicing inhibitor element. *RNA*, 2001, **7** (9): 1239~ 1253
- 13 Libri D, Duconge F, Levi L, *et al.* A role for the PSF-U mismatch in the recognition of the 5' splice site of yeast introns by U1 snRNP. *J Biol Chem*, 2002, **277** (20): 18173~ 18181
- 14 Shtatland T, Gill S C, Javornik B E, *et al.* Interactions of *Escherichia coli* RNA with bacteriophage MS2 coat protein: genomic SELEX. *Nucleic Acids Res*, 2000, **28** (21): E93
- 15 Wen J D, Gray C W, Gray D M. SELEX selection of high-affinity oligonucleotides for bacteriophage Ff gene 5 protein. *Biochemistry*, 2001, **40** (31): 9300~ 9310
- 16 White R, Rusconi C, Scardino E, *et al.* Generation of species cross reactive aptamers using "toggle" SELEX. *Mol Ther*, 2001, **4** (6): 567~ 573
- 17 White E K, Moore-Jarrett T, Ruley H E. PUM2, a novel murine puf protein, and its consensus RNA-binding site. *RNA*, 2001, **7** (12): 1855~ 1866
- 18 Ulrich H, Magdesian M H, Alves M J, *et al.* *In vitro* selection of RNA aptamers that bind to cell adhesion receptors of *Trypanosoma cruzi* and inhibit cell invasion. *J Biol Chem*, 2002, **277** (23): 20756~ 20762
- 19 Wang Y, Mi J, Cao X. Anti-DNA antibodies exhibit different binding motif preferences for single stranded or double stranded DNA. *Immunol Lett*, 2000, **73** (1): 29~ 34
- 20 Bouvent P, Allain F H, Finger L D, *et al.* Recognition of pre-formed and flexible elements of an RNA stem-loop by nucleolin. *J Mol Biol*, 2001, **309** (3): 763~ 775
- 21 Roychowdhury-Saha M, Lato S M, Shank E D, *et al.* Flavin recognition by an RNA aptamer targeted toward FAD. *Biochemistry*, 2002, **41** (8): 2492~ 2499
- 22 Blank M, Weinschenk T, Priemer M, *et al.* Systematic evolution of a DNA aptamer binding to rat brain tumor microvessels. selective targeting of endothelial regulatory protein pigpen. *J Biol Chem*, 2001, **276** (19): 16464~ 16468
- 23 Hicke B J, Marion C, Chang Y F, *et al.* Tenascin-C aptamers are generated using tumor cells and purified protein. *J Biol Chem*, 2001, **276** (52): 48644~ 48654
- 24 Lato S M, Ozerova N D, He K, *et al.* Boron-containing aptamers to ATP. *Nucleic Acids Res*, 2002, **30** (6): 1401~ 1407
- 25 Fukuda K, Vishnuvardhan D, Sekiya S, *et al.* Isolation and characterization of RNA aptamers specific for the hepatitis C virus nonstructural protein 3 protease. *Eur J Biochem*, 2000, **267** (12): 3685~ 3694
- 26 Ali N, Pruijn G J, Kenan D J, *et al.* Human La antigen is required for the hepatitis C virus internal ribosome entry site-mediated translation. *J Biol Chem*, 2000, **275** (36): 27531~ 27540
- 27 Andreola M L, Pileur F, Calmels C, *et al.* DNA aptamers selected against the HIV-1 RNase H display *in vitro* antiviral activity. *Biochemistry*, 2001, **40** (34): 10087~ 10094
- 28 Biroccio A, Hamm J, Incitti I, *et al.* Selection of RNA aptamers that are specific and high-affinity ligands of the hepatitis C virus RNA-dependent RNA polymerase. *J Virol*, 2002, **76** (8): 3688 ~ 3696

Aptamers of Identification and Application in Systematic Evolution of Ligands by Exponential Enrichment Process

LIU Xiao, LI Chang-You, CHEN Yuan-Ding*

(Institute of Medical Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Kunming 600018, China)

Abstract In the past decade, systematic evolution of ligands by exponential enrichment (SELEX) technology has been under development, and is a general approach for identification of aptamers or oligonucleotide ligands that bind with high-affinity and specificity to a wide range of selected molecules on their versatility. The SELEX protocol is a process of synthesis random oligonucleotide sequents library pool, select special aptamers, amplify and iterative *in vitro*. Combinatorial modify groups libraries offer the convenience in the SELEX process, making the screening process fast, easy and high-throughput. Here the current designs and application of modified nucleotides for in SELEX process are reviewed, and expected to further the utility of this method in both practical and theoretical terms.

Key words systematic evolution of ligands by exponential enrichment, aptamers, identification, application

* Corresponding author. Tel: 86-871-8334689, E-mail: chenyd@km169.com

Received: April 28, 2002 Accepted: May 28, 2002