论文

SELEX筛选核酸适配体富集水平理论分析

禹新良^{1,2*}

1. 湖南工程学院化学化工学院, 湘潭 411104

2. 湖南大学化学化工学院, 化学生物传感与计量学国家重点实验室, 长沙 410082

*通讯作者, E-mail: yxl@hnie.edu.cn

收稿日期: 2017-02-04; 接受日期: 2017-04-10; 网络版发表日期: 2017-07-04

国家自然科学基金(编号: 21190041)、湖南省自然科学基金(编号: 12JJ6011)、湖南教育厅重点资助项目(编号: 16A047)和化学生物传感与计量学国家重点实验室(湖南大学)开放课题(编号: 2016013)资助项目

摘要 指数富集的配基系统进化(SELEX)方法是一种新颖的筛选技术, 能从人工合成的DNA/RNA文库中筛选得到核酸适配体. 根据SELEX原理, 与靶标亲和性最强的核酸适配体应该具有最高的富集水平. 然而, SELEX实验很少出现这种现象. 本文采用数学方法, 论证了在满足一定实验条件时, 与靶标亲和性最强的核酸适配体并不具有最高的富集水平, 并指出了较大的靶标浓度、较低的分选效率会促使这种现象的产生.

关键词 适配体, 富集水平, 数学分析, SELEX

1 引言

核酸适配体(aptamer)是从人工合成的DNA/RNA文库中, 采用指数富集的配基系统进化(systematic evolution of ligands by EXponential enrichment, SELEX)筛选技术获得的单链寡核苷酸^[1-3]. 核酸适配体对靶标识别的特异性类似于抗体^[1,2]. 与抗体相比, 它具有许多独特的性能与优势, 如免疫原性弱、容易修饰、廉价易得等^[3-7]. 核酸适配体能识别多种类型的靶标, 如小分子(如金属离子、染料)、大分子、细胞等^[6-12]. 其识别机理复杂, 一般认为核酸适配体是通过形成特殊的三维空间结构, 如发夹(hair pin)结构、假结(pseudo knot)结构、凸环(bulge)结构、G-四分体(G-quarter)等, 并通过氨基、羟基等以分子间相互作用的方式与靶标结合, 包括碱基间的堆积、氢键、静电以及范德华力等.

核酸适配体的出现, 为生物医学、化学生物学、分析化学、环境化学等研究领域提供了独特的分子识别体系, 也为新型分子探针的获取开辟了一条全新的途径, 如用于病变分子、细胞、组织识别的分子探针^[8,13,14]. 另外, 核酸适配体还将广泛应用于蛋白质组学、基因调控等新兴领域. 因此, 有关核酸适配体的相关技术研究已成为国际化热点课题^[8-12].

SELEX技术通过设计合成单链随机寡核苷酸文库, 结合体外PCR扩增技术, 从文库中筛选出能与靶标特异性结合的核酸适配体. SELEX技术流程如下: (1) 建库. 首先, 人工合成一个单链随机寡核苷酸文库, 其含核酸序列数大约为 $10^{13} \sim 10^{15}$. 每条序列两端为固定序列区, 中间为随机序列区域. 固定序列区长度约为18~20个碱基, 在PCR扩增反应中, 该区可作为引物的结合位点. 随机序列区域长度约为20~40个碱基, 该区域可保证核酸序列三维结构的多样性, 从而增大与靶

引用格式: 禹新良. SELEX筛选核酸适配体富集水平理论分析. 中国科学: 化学, 2017, 47: 891-895

Yu X. Theoretical analysis of the aptamer enrichment in SELEX-based selection. *Sci Sin Chim*, 2017, 47: 891-895, doi: 10.1360/N032017-00019

标结合良好的核酸序列存在的概率。(2) 孵育: 在一定的实验条件下, 如缓冲体系及温度等, 将建好的文库与靶标混合温育, 结合力强的核酸序列会与靶标形成复合物。孵育一段时间(约30~60 min)后, 核酸序列、靶标、核酸-靶标的复合物三者浓度之间构成动态平衡。(3) 分离: 孵育后根据靶标性质不同选择合适的分离方法, 使与靶标结合力强的核酸序列与未结合(或结合力弱)的核酸序列进行分离。在实验中各种分离方法都存在分离效率低的问题, 即与靶标结合力强的核酸序列出现损失, 而结合力弱或不结合的序列却混入产物中。例如, 以细胞为靶标的cell-SELEX筛选中, 洗涤强度影响到分离效率; 硝酸纤维膜过滤分离对随机核酸序列有较强的背景吸附; 凝胶阻滞电泳易造成结合力强的核酸序列的损失; 微孔板分离容易导致靶标潜在位点被包被、屏蔽, 从而降低分离效率。(4) 扩增: 分离后的DNA (或RNA) 序列进行PCR (或RT-PCR) 扩增。扩增后的产品经过分离与纯化后, 用来制备寡核苷酸文库, 准备下一轮筛选。经数轮筛选, 亲和力达到或接近饱和, 扩增产品用于克隆、测序。

通常完成核酸适配体的SELEX筛选需要10~15轮循环操作, 每轮操作耗财、耗时、耗力, 而且实验条件控制要求严格, 有些实验条件细微的变化有可能显著影响实验结果^[15]。有学者对SELEX过程进行了数学分析, 从理论上讲, 亲和性最强的核酸适配体, 其富集效果也应该最好^[15-23]。但实验结果显示, 富集水平最好的序列, 其亲和性不一定最理想^[7,12,24-26]。本文基于化学平衡理论与核酸序列分离效率差异, 对SELEX过程进行数学分析, 对亲和性最强的核酸适配体、其富集水平并非最高的现象进行了理论分析。

2 数学推导

在SELEX实验中, 与靶标结合力强的核酸适配体在开始几轮的筛选中富集快, 呈指数增长, 在后面几轮则增长缓慢。因此, 核酸适配体的富集水平是判断实验进程的一个重要指标, 也是筛选过程中挑选产品用于进一步亲和性、专一性测试的重要依据。以细胞膜蛋白作为靶标、DNA核酸序列作为配体为例, 在孵育时形成热力学平衡, 其方程式可表示为^[15,22]

其中Tf是文库中未与配体序列结合的自由靶标(蛋白), Lf_i 为未与靶标结合的配体 i , C_i 是靶标-配体 i 组成的复合物, $k_{\text{on}i}$ 是靶标与配体 i 的结合速率常数, $k_{\text{off}i}$ 为靶标-配体 i 复合物的解离常数, N 是文库中核酸序列类型。靶标-配体 i 复合物解离平衡常数的计算式为

$$K_{\text{di}} = \frac{k_{\text{off}i}}{k_{\text{on}i}} = \frac{[\text{Tf}][\text{Lf}_i]}{[\text{C}_i]} \quad (2)$$

其中 $[\text{Tf}]$ 为文库中未与配体结合的自由靶标浓度, $[\text{Lf}_i]$ 为自由 i 配体的浓度, $[\text{C}_i]$ 为蛋白-配体 i 复合物的浓度。表观解离平衡常数 K_d 计算式为^[15,22]

$$K_d = \frac{[\text{Tf}][\text{Lf}]}{[\text{C}]} \quad (3)$$

其中 $[\text{C}]$ 和 $[\text{Lf}]$ 分别为靶标-配体复合物的浓度和自由配体的浓度。与靶标孵育后, 文库中配体 i 序列所占分率 F_i 为

$$F_i = \frac{[\text{L}_i]}{[\text{L}]} \quad (4)$$

其中 $[\text{L}_i] = [\text{Lf}_i] + [\text{C}_i]$ 及 $[\text{L}] = \sum_i [\text{L}_i] = [\text{C}] + [\text{Lf}]$ 。从式(2)~(4), 可推出靶标-配体 i 复合物浓度计算式为

$$[\text{C}_i] = [\text{L}] \frac{[\text{F}_i][\text{Tf}]}{K_{\text{di}} + [\text{Tf}]} \quad (5)$$

在初始文库中, 各类型核酸序列分布可表示为 $\overline{F^{(1)}} = \{F_1^{(1)}, \dots, F_N^{(1)}\}$, $\sum F_i^{(1)} = 1$ 及自由靶标浓度 $[\text{Tf}^{(1)}]$ 。类似地, 在第 r 轮筛选有 $\overline{F^{(r)}} = \{F_1^{(r)}, \dots, F_N^{(r)}\}$, $\sum F_i^{(r)} = 1$ 及靶标浓度 $[\text{Tf}^{(r)}]$ 。

适配体的SELEX筛选中, 其富集水平已有多种定义。但每一轮的适配体富集水平通常定义为配体序列 i 分率的增加, i 为解离平衡常数最小的配体序列。不妨假定解离常数大小顺序为: $0 < K_{d1} < K_{d2} < \dots < K_{dN}$ 。配体序列1的富集水平为

$$E_1^{(r+1)} = \frac{F_1^{(r+1)}}{F_1^{(r)}} \quad (6)$$

其中 $F_1^{(r)}$ 为用于第 r 轮筛选的、与靶标孵育前文库中配体序列1的分率 $[\text{Lf}_1^{(r)}]/[\text{Lf}^{(r)}]$; 而 $F_1^{(r+1)}$ 为第 r 轮筛选完成后, 产物中序列1的分率 $[\text{Lf}_1^{(r+1)}]/[\text{Lf}^{(r+1)}]$, 理论上等于孵育后复合物中序列1所占分率。对于筛选过程, 由式(3)和(5), 可推出式(7):

$$F_i^{(r+1)} = \frac{[C_i^{(r)}]}{[C^{(r)}]} = \frac{[L_i^{(r)}] \frac{F_i^{(r)} [Tf^{(r)}]}{K_{di}^{(r)} + [Tf^{(r)}]}}{[Tf^{(r)}] \frac{[Lf^{(r)}]}{K_d^{(r)}}} \quad (7)$$

从式(3)及 $[L] = [Lf] + [C]$, 可推出

$$\frac{[Lf]}{[L]K_d} = \frac{1}{K_d + [Tf]} \quad (8)$$

于是式(7)可表示为

$$F_i^{(r+1)} = \frac{\frac{F_i^{(r)}}{K_{di}^{(r)} + [Tf^{(r)}]}}{\frac{[Lf^{(r)}]}{[L^{(r)}]K_d^{(r)}}} = \frac{\frac{F_i^{(r)}}{K_{di}^{(r)} + [Tf^{(r)}]}}{\frac{1}{K_d^{(r)} + [Tf^{(r)}]}} = \frac{K_d^{(r)} + [Tf^{(r)}]}{K_{di}^{(r)} + [Tf^{(r)}]} F_i^{(r)} \quad (9)$$

式(9)可描述在第 r 轮筛选中, 新获得配体序列 i 的分率与筛选前分率之间的关系^[15,22].

在SELEX筛选中, 存在分离效率问题, 使得与靶标结合力强的配体序列出现损失, 而结合力弱或不结合的核酸序列却混入产物中, 从而降低筛选效率, 需增加筛选轮数. 本工作主要讨论靶标-配体复合物及配体分离效率. 定义输入端靶标-配体复合物浓度为 C_{in} 、配体浓度为 L_{in} ; 它们在分离后输出端分别为 C_{out} 、 L_{out} . 则靶标-配体复合物的分离效率可表示为 $K_C = C_{out}/C_{in}$, 配体序列的分离效率为 $K_L = L_{out}/L_{in}$. 整个分离系数可表示为

$$K_{CL} = K_C/K_L \quad (10)$$

分离效率处于理想状态时, 复合物中的配体序列无损失, $K_C = 1$; 自由序列不进入产物中, $K_L = 0$, 分离系数 K_{CL} 为 ∞ . 获胜核酸序列 i ($=1$)具有最高亲和力时, K_{CLi} 的最大值者为 (K_{CLi}) . 在实际的分离中, 分类得到的 i 核酸序列 $[L_i]^{part}$ 来源于两部分: 靶标-配体复合物 C_i 中的已经结合的配体序列 i , 以及自由核酸配体 L_{fi} 中的序列 i . 因此核酸序列 i 的浓度 $[L_i]^{part}$ 计算式为

$$\begin{aligned} [L_i]^{part} &= K_{Li}[Lf_i] + K_{Ci}[C_i] \\ &= K_{Li}F_i[L] + (K_{Ci} - K_{Li})[C_i] \end{aligned} \quad (11)$$

类似地, 在考虑到分离效率时, 分类得到的核酸序列 $[L_i]^{part}$ 也来源于两部分: 复合物 C 中的核酸序列与自由核酸配体 L_f 中的序列, 则得到的整体核酸序列新的浓度 $[L]^{part}$ 为

$$[L]^{part} = K_L[L] + (K_C - K_L)[C] \quad (12)$$

于是核酸序列 i 的所占分率 F_i^{part} 表示为

$$F_i^{part} = \frac{[L_i]^{part}}{[L]^{part}} = \frac{K_{Li}F_i[L] + (K_{Ci} - K_{Li})[C_i]}{K_L[L] + (K_C - K_L)[C]} \quad (13)$$

基于文献[17], 在标准实验条件下, $K_C = 100\%$, $K_L = 1\%$, 即 K_C 远大于 K_L , 因此有

$$K_C - K_{Li} \approx K_C - K_L \approx K_C \approx K_{Ci} \quad (14)$$

式(13)变为

$$F_i^{part} = \frac{[L_i]^{part}}{[L]^{part}} = \frac{K_{Li}F_i[L] + K_{Ci}[C_i]}{K_L[Lf] + K_C[C]} \quad (15)$$

将式(5)的 $[C_i]$ 项, 及式(3)的 $[C]$ 项, 代入式(15), 则有

$$F_i^{part} = \frac{K_{Li}F_i[L] + K_{Ci}[Tf]F_i/(K_{di} + [Tf])[L]}{K_L[L] + K_{CL}[Lf][Tf]/K_d} \quad (16)$$

从式(8)及式(16), 可以推得

$$F_i^{part} = F_i \frac{(K_d + [Tf])}{(K_{di} + [Tf])} \frac{\{K_{Li}K_d + (K_C + K_{Li})[Tf]\}}{\{K_LK_d + (K_C + K_L)[Tf]\}} \quad (17)$$

利用 $K_C + K_{Li} \approx K_C + K_L \approx K_C$, 式(17)可以转化为

$$F_i^{part} = F_i \frac{(K_d + [Tf])}{(K_{di} + [Tf])} \frac{(K_{Li}K_d + K_C[Tf])}{(K_LK_d + K_C[Tf])} \quad (18)$$

于是第 r 轮的配体富集水平可以表示为

$$\begin{aligned} E_i^{(r+1)} &= \frac{F_i^{(r+1)}}{F_i^{(r)}} \\ &= \frac{(K_d^{(r)} + [Tf^{(r)}])}{(K_{di}^{(r)} + [Tf^{(r)}])} \frac{\left(\frac{1}{K_{CLi}}K_{di}^{(r)} + [Tf^{(r)}]\right)}{\left(\frac{1}{K_{CL}}K_d^{(r)} + [Tf^{(r)}]\right)} \end{aligned} \quad (19)$$

当配体序列 i 的解离常数 K_{di} 为最低 K_{d1} 时, 适配体序列为“获胜”适配体, 该序列与靶标结合最好, 其富集水平用式(20)表示:

$$\begin{aligned} E_1^{(r+1)} &= \frac{F_1^{(r+1)}}{F_1^{(r)}} \\ &= \frac{(K_d^{(r)} + [Tf^{(r)}])}{(K_{d1}^{(r)} + [Tf^{(r)}])} \frac{\left(\frac{1}{K_{CL1}}K_{d1}^{(r)} + [Tf^{(r)}]\right)}{\left(\frac{1}{K_{CL}}K_d^{(r)} + [Tf^{(r)}]\right)} \end{aligned} \quad (20)$$

类似地, 可以写出亲和力稍次的配体序列 i ($=2$)的富集水平表达式. 序列1、2的富集水平之比可表示为

$$\frac{E_1^{(r+1)}}{E_2^{(r+1)}} = \frac{(K_{d2} + [\text{Tf}^{(r)}]) \cdot \left(\frac{1}{K_{\text{CL1}}} K_{\text{d1}} + [\text{Tf}^{(r)}] \right)}{(K_{\text{d1}} + [\text{Tf}^{(r)}]) \cdot \left(\frac{1}{K_{\text{CL2}}} K_{\text{d2}} + [\text{Tf}^{(r)}] \right)} \quad (21)$$

当有式(22)~(24)成立时, 式(21)右边小于1.

$$\frac{1}{K_{\text{CL2}}} < \frac{1}{K_{\text{CL1}}} \quad (22)$$

$$\frac{1}{K_{\text{CL2}}} > 1 - \left(1 - \frac{1}{K_{\text{CL1}}} \right) \cdot \frac{K_{\text{d1}}}{K_{\text{d2}}} \quad (23)$$

$$[\text{Tf}^{(r)}] > \frac{\left(\frac{1}{K_{\text{CL1}}} - \frac{1}{K_{\text{CL2}}} \right) K_{\text{d1}} K_{\text{d2}}}{\frac{1}{K_{\text{CL2}}} K_{\text{d2}} - \frac{1}{K_{\text{CL1}}} K_{\text{d1}} + K_{\text{d1}} - K_{\text{d2}}} \quad (24)$$

于是有

$$\frac{E_1^{(r+1)}}{E_2^{(r+1)}} = \frac{(K_{d2} + [\text{Tf}^{(r)}]) \cdot \left(\frac{1}{K_{\text{CL1}}} K_{\text{d1}} + [\text{Tf}^{(r)}] \right)}{(K_{\text{d1}} + [\text{Tf}^{(r)}]) \cdot \left(\frac{1}{K_{\text{CL2}}} K_{\text{d2}} + [\text{Tf}^{(r)}] \right)} < 1 \quad (25)$$

即亲和性最强的适配体其富集水平并非最高.

3 结论

对于SELEX过程核酸适配体筛选, 发现亲和性最强的适配体其富集水平并非最高. 降低靶标浓度、提高筛选过程中的分离效率, 有利于亲和性最佳的获胜核酸适配体的富集; 相反, 增大靶标浓度、降低适配体与靶标分离效率, 有利于非获胜适配体的富集.

参考文献

- 1 Famulok M, Mayer G, Blind M. *Acc Chem Res*, 2000, 33: 591–599
- 2 Gold L, Brody E, Heilig J, Singer B. *Chem Biol*, 2002, 9: 1259–1264
- 3 Tan W, Donovan MJ, Jiang J. *Chem Rev*, 2013, 113: 2842–2862
- 4 Tavittian B, Terrazzino S, Kühnast B, Marzabal S, Stettler O, Dollé F, Deverre J, Jobert A, Hinnen F, Bendriem B, Crouzel C, Giamberardino LD. *Nat Med*, 1998, 4: 467–471
- 5 Hong H, Goel S, Zhang Y, Cai W. *Curr Med Chem*, 2011, 18: 4195–4205
- 6 Meng HM, Fu T, Zhang XB, Tan W. *Nat Sci Rev*, 2015, 2: 71–84
- 7 郭秋平, 刘小丹, 谭誉宇, 王柯敏, 羊小海, 周玉, 叶玲, 赵下雨. 科学通报, 2013, 58: 2745–2750
- 8 Ye M, Hu J, Peng M, Liu J, Liu J, Liu H, Zhao X, Tan W. *Int J Mol Sci*, 2012, 13: 3341–3353
- 9 Ellington AD, Szostak JW. *Nature*, 1990, 346: 818–822
- 10 Kong RM, Chen Z, Ye M, Zhang XB, Tan WH. *Sci China Chem*, 2011, 54: 1218–1226
- 11 Fang X, Tan W. *Acc Chem Res*, 2010, 43: 48–57
- 12 Ninomiya K, Kaneda K, Kawashima S, Miyachi Y, Ogino C, Shimizu N. *Bioorg Med Chem Lett*, 2013, 23: 1797–1802
- 13 Rimmele M. *ChemBioChem*, 2003, 4: 963–971
- 14 Shangguan D, Meng L, Cao ZC, Xiao Z, Fang X, Li Y, Cardona D, Witek RP, Liu C, Tan W. *Anal Chem*, 2008, 80: 721–728
- 15 Irvine D, Tuerk C, Gold L. *J Mol Biol*, 1991, 222: 739–761
- 16 Sun F, Galas D, Waterman MS. *J Mol Biol*, 1996, 258: 650–660
- 17 Vant-Hull B, Payano-Baez A, Davis RH, Gold L. *J Mol Biol*, 1998, 278: 579–597
- 18 Djordjevic M, Sengupta AM. *Phys Biol*, 2006, 3: 13–28
- 19 Chen CK. *Comput Methods Prog Biomed*, 2007, 87: 189–200
- 20 Chen CK, Kuo TC. *Comp Chem Eng*, 2007, 31: 1007–1019
- 21 Chen CK, Kuo TL, Chan PC, Lin LY. *Comp Biol Med*, 2007, 37: 750–759
- 22 Levine HA, Nilsen-Hamilton M. *Comp Biol Chem*, 2007, 31: 11–35
- 23 Seo YJ, Chen S, Nilsen-Hamilton M, Levine HA. *Bull Math Biol*, 2010, 72: 1623–1665
- 24 Schütze T, Wilhelm B, Greiner N, Braun H, Peter F, Mörl M, Erdmann VA, Lehrach H, Konthur Z, Menger M, Arndt PF, Glökler J. *PLoS One*, 2011, 6: e29604
- 25 Yang X, Wang Y, Wang K, Wang Q, Wang P, Lin M, Chen N, Tan Y. *RSC Adv*, 2014, 4: 30934–30937
- 26 Wan J, Ye L, Yang X, Guo Q, Wang K, Huang Z, Tan Y, Yuan B, Xie Q. *Analyst*, 2015, 140: 5992–5997

Theoretical analysis of the aptamer enrichment in SELEX-based selection

Xinliang Yu^{1,2*}

¹ College of Chemistry and Chemical Engineering, Hunan Institute of Engineering, Xiangtan 411104, China

² State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China

*Corresponding author (email: yxl@hnie.edu.cn)

Abstract: SELEX (systematic evolution of ligands by EXponential enrichment) is a novel technology for aptamer selection from libraries of random DNA (or RNA) sequences. According to SELEX principle, the candidate aptamers with the strongest binding affinity to targets should have maximum enrichment. But this phenomenon is seldom found in SELEX experiments. In this paper, a mathematical analysis was carried out to explicate the reason for this phenomenon. Investigated results suggest that increasing target concentrations and decreasing efficiencies of partitioning can result in above phenomenon.

Keywords: aptamer, enrichment levels, mathematical analysis, SELEX

doi: [10.1360/N032017-00019](https://doi.org/10.1360/N032017-00019)