

·综述·

DOI: 10.13865/j.cnki.cjbmb.2019.03.03

分子模拟指导核酸适配体文库设计

雷云霞, 李招发, 林俊生*

(华侨大学医学院, 福建 泉州 362021)

摘要 核酸适配体是利用配体指数富集的系统进化技术(SELEX)从随机文库中筛选获得一段有功能的单链寡核苷酸。但因筛选过程中的文库选择、洗涤次数、分离效率、缓冲液离子含量和 pH 值等多种因素的影响,迄今所报道的亲合力与特异性都很高的核酸适配体为数不多。初始文库是核酸适配体筛选的源头,作为 SELEX 技术的根本,其设计是否合理直接影响到筛选的成败和效率。分子模拟能以核酸适配体文库为主体,计算机为主要工具,发展多种结构模拟与分析工具,辅助核酸适配体文库的合理设计。本文综述了现阶段利用分子模拟进行核酸适配体初始文库设计的相关方法,希望能为从源头上提高核酸适配体筛选的成功率提供线索。

关键词 核酸适配体; 配体指数富集的系统进化技术(SELEX); 文库设计; 分子模拟; 分子对接
中图分类号 Q71

The Design of Aptamer Candidate Pool Guided by Molecular Simulation

LEI Yun-Xia, LI Zhao-Fa, LIN Jun-Sheng*

(School of Medicine, Huaqiao University, Quanzhou 362021, Fujian, China)

Abstract Aptamers are functional single-stranded oligonucleotides selected from random libraries by SELEX (systematic evolution of ligands by exponential enrichment). However, the number of reported aptamers with high affinity and specificity so far is still less than satisfactory because of library selection, washing times, separation efficiency, ion, pH value of buffer and so on. As the fundamental of SELEX process, the initial library is the primary source for screening SELEX. Properly design of initial library is directly related to the outcome and efficiency of the selection. A variety of structural simulation can be generated for aptamer library and relevant analysis tools can be developed by molecular simulation. This paper summaries the present methods of using molecular simulation to guide the design of aptamer initial library, hoping to provide clues for improving the success rate of aptamer screening from the initial source.

Key words aptamer; systematic evolution of ligands by exponential enrichment (SELEX); library design; molecular simulation; docking

核酸适配体是通过配体指数富集的系统进化技术(systematic evolution of ligands by exponential enrichment, SELEX)从随机文库中筛选获得的功能性单链寡核苷酸^[1]。通过将初始文库与靶标结合;洗涤去除非特异性结合的序列;分离与靶标特异性结合的片段;进行 PCR 扩增,得到次级文库进入下一轮筛选,这样的循环过程使能与靶标结合的核酸适配体在文库中富集。由此可见,由于起始文库的随机性和有限性,传统的核酸适配体筛选过程本身就相当于大海捞针。即使筛选全程的所有步骤都正确操作,其结果依然不确定。筛选过程中还有许多影响筛选结果的因素,例如:靶标溶液的基体与浓

收稿日期: 2018-06-15; 修回日期: 2018-08-02; 接受日期: 2018-09-01
华侨大学研究生科研创新能力培育计划(No.1611316023); 国家重点研发计划(No.2016YFE0101700)和福建省科技厅引导性项目(No.2016Y0063)资助

* 通讯作者 Tel: 0595-22690889; E-mail: junshenglin@hqu.edu.cn
Received: June 15, 2018; Revised: August 2, 2018; Accepted: September 1, 2018

Supported by Huaqiao University Graduate Research Innovation Ability Training Project (No. 1611316023); National Key Research and Development Program of China (No.2006YFE0101700); Science and Technology Department Guiding Project of Fujian Province (No. 2016Y0063)

* Corresponding author Tel: 0595-22690889; E-mail: junshenglin@hqu.edu.cn

度^[2];筛选时的洗涤次数;分离效率等。因此,初始文库作为核酸适配体筛选的源头,其设计的合理性对于核酸适配体筛选的整个环节至关重要。但实际筛选过程所用的文库往往存在多种问题。首先,核酸适配体筛选初始文库的随机序列一般有 30~60 个碱基^[3]。因此理论上初始文库容量为 $10^{18} \sim 10^{36}$,但因目前核酸合成的最大库容量只能达到 10^{16} ,实际与理论库容量相比相差很大^[4];其次,PCR 扩增具有一定的偏好性^[5,6],会使原本丰度不够大的文库结构变得更加单一;最后,正如 Carothers 等所指出的,核酸适配体结构越复杂,其与靶标结合的亲和力越高^[7]。而 Gevertz 等研究发现,核酸适配体筛选所用的初始文库大多数结构简单,其中含有的多是简单的线性茎环结构以及低分支结构,而高阶结节结构以及多分支结构含量较少^[8,9]。

初始文库是核酸适配体筛选的源头。合理的水库设计是筛选能否成功的根本因素。分子模拟能通过各种软件实现对靶标与核酸结构的模拟分析^[10],预测二者的作用位点^[11],同时可以对初始文库的结构模块进行高通量筛选^[12],是辅助文库设计的有效手段。利用分子模拟的方法辅助文库设计,可缩小初始文库的库容量,同时又不影响文库目的序列丰度和二级结构的多样性,可从文库设计方面入手,提高核酸适配体筛选的成功率和效率。核酸适配体筛选所用的初始文库由固定序列和随机序列组成。固定序列一般就是位于核酸适配体两端与引物互补用于扩增的部分,其设计原理与引物设计原理相同。现阶段,根据核酸适配体筛选目的不同,分子模拟指导的初始文库随机序列的设计主要包括两大类:一是为了优化已知结合序列的有指向性的文库设计;二是为了获得全新的能与靶标结合的序列的无指向性的文库设计。

1 有指向性的文库设计

有指向性的文库设计是指在对靶标序列有所了解的基础上,保留序列作用位点,对其他非作用位点进行随机突变或直接对靶标序列进行随机突变,构建二次筛选文库。传统的 SELEX 方法效率低的原因之一是,在对靶标可能要结合的序列一无所知的情况下,要在随机文库中找到一条或几条序列,这个过程非常困难,犹如大海捞针。利用有指向性的文库设计法设计的水库,虽然库容量远小于 10^{16} ,但文库中的无效序列比例明显减少,目标序列的丰度显著增高。已有研究表明,这种文库设计方法能够有

效提高核酸适配体的筛选效率与成功率^[13]。这种文库设计主要存在两种情况。

1.1 已知野生型靶序列情况下的有指向性的文库设计

自然界中存在着许多能与基因元件结合的靶标物质,它们能够通过与基因元件相结合,从而调控下游基因的表达。已有许多研究者利用这种已知的野生型序列,并对其进行改造,设计筛选文库。用该文库筛选获得了功能更为优良的核酸适配体。例如,人 T 淋巴细胞白血病病毒 1 型 (human T-cell leukemia virus type 1, HTLV-1) 的 REX 蛋白,通过其基因元件 (rex-binding element, XBE) 作用来调控病毒 mRNA 的胞质转运。利用该原理, Baskerville 等在基因元件的基础上,通过对其序列进行突变,构建筛选文库,从而筛选获得比基因元件亲和力高 9 倍的 REX 蛋白核酸适配体^[14]。同样, Ahirwar 等将雌激素受体 α (estrogen receptor α , ER α) 的反应元件 (estrogen response elements, ERE_s) 进行突变,通过二级结构预测软件 MC-Fold 确定序列二级结构^[15]。在二级结构水平上对文库进行初步筛选,保证文库二级结构与 EREs 类似,然后利用分子对接^[16]联用的方法 (同时使用 AutoDock Vina^[17], HADDOCK^[18] 和 PatchDock^[19] 三种软件进行对接,避免单一软件算法缺陷) 对文库序列进行评估,将分子对接评分较高的序列进行分子生物学验证,最后获得能与雌激素受体特异性结合的核酸适配体^[20]。

这种文库设计虽然具有一定的指向性,可提高筛选成功率和效率,但在设计时需要考虑的一个重要因素是在原有基础上的突变率以及随机化程度。程度太低,突变碱基占原序列比例太小则设计的水库库容量太小;突变率太高确实可以提高库容量,但文库中的无效序列也会增加^[21]。

1.2 已知核酸适配体序列的有指向性的文库设计

核酸适配体筛选的目的是获得能与靶标高特异性和高亲和力结合的核酸适配体。但因初始文库设计不合理以及 PCR 扩增的偏好性等原因,传统的 SELEX 技术在大多数情况下难以成功筛选获得高亲和力的功能性核酸适配体。在核酸适配体序列已筛选成功,但其亲和力又不理想或核酸适配体序列太长,不适于进行下游研究的情况下,在已有序列的基础上设计新的二次筛选文库,继续进行筛选,是提高核酸适配体亲和力的有效方法之一。

2016 年, Xu 等对已有的前列腺特异性膜抗原 (prostate specific membrane antigen, PSMA) 核酸适

配体进行二级结构预测,并通过分子对接技术确定其作用的关键碱基及结构模块。在此基础上,对已有核酸适配体进行碱基的截断、插入和突变,形成新的二次筛选文库。然后,利用二级结构预测软件 Vfold 2D^[22] 和三维结构预测软件 Vfold 3D^[23] 对其进行结构预测。通过分子对接技术评估文库序列与靶标结合的亲和力,最后筛选获得了序列更短、亲和力更高的核酸适配体^[24]。

凝血酶核酸适配体早在 1992 年就被筛选获得^[25],并且已被证明其独特的 G 四联体结构与其亲和力和功能性密切相关^[26]。在此基础上,Ikebukuro 等对凝血酶核酸适配体进行随机突变,结合二级结构预测软件保证其 G 四联体结构不被改变的前提下设计获得二次筛选文库。利用遗传算法对其进行模拟演化,经过几轮优化最后获得了功能性更强的凝血酶核酸适配体^[27]。

这种方法应用的前提是已经筛选获得了目的序列,但其与靶标结合效果不佳。在此基础上对该序列进行突变设计文库再次进行筛选。其目的是得到亲和力更高的核酸适配体。

有指向性的文库设计方法,从目的序列覆盖率以及筛选成功率上均比较高,但使用前提是有已知的靶标序列,要在原来的基础上进行优化,增加核酸适配体的亲和力。而对目的序列未知的新靶标筛选的文库设计不适用。

2 无指向性的文库设计

在对靶标结合序列等信息不了解的情况下,要设计出合理的初始筛选文库就要从增加文库序列丰度及结构多样性着手。研究表明,核酸适配体的亲和力以及功能与其结构有着密切关系^[28]。Bing 等对筛选出的 20 条链霉亲和素核酸适配体进行二级结构分析,在结构不同的 20 条核酸适配体中发现了一致的结构,并且该结构也存在于其他三个不同实验室筛选获得的链霉亲和素核酸适配体^[29];同样,研究发现,ATP 核酸适配体的二级结构由两侧的茎和中间的一个环构成,并且利用分子对接分析发现,二者的作用位点就在中间的环上^[30, 31]。不难看出,核酸适配体一级序列的碱基组成,与其二、三级结构的复杂程度对其与靶标的结合能力有较大的影响。因此,通过增加筛选文库的一级序列丰度与结构多样性以及复杂性(包括内环、凸环、多结节以及错配等),可以提高筛选出高亲和力核酸适配体的可能性^[7]。

2.1 嵌入式文库设计法

众多研究发现,核酸适配体的二级结构中至少有 1 个茎环结构,并且茎环结构中的环形成的口袋是其结合靶标的潜在位点^[32-34]。因此,在筛选所用的随机文库中,插入能够形成稳定茎环结构的固定序列是提高文库筛选成功几率的一种思路。2002 年,Davis 等^[35]插入一段茎环结构到随机文库中,设计出 GTP 的筛选文库(Fig.1A),为了保证茎环结构的稳定性,在茎部设计多个 CG 互补碱基,以提高其稳定性。并将其设计的一部分随机文库(Fig.1A)与序列长度相同的完全随机文库(Fig.1B)1:1混合进行筛选。最后证明,筛选获得的高亲和力核酸适配体均具有茎环结构,且全部来自于嵌入式随机文库。

核酸三叶草结构是重要的结合口袋^[36],它能够通过中间的环状穴非特异性地包裹荧光小分子^[37],同时也是靶标分子特异性结合的位点^[38]。利用这一竞争结合特性能够很好地搭建起核酸适配体传感器。基于此,Yu 等以检测亚甲基二氧吡咯戊酮为出发点,将形成三叶草结构包裹荧光素的卡因核酸适配体^[37]的中间环部分,置换成随机序列形成筛选文库(Fig.1C),成功搭建了亚甲基二氧吡咯戊酮的核酸适配体传感器^[39]。

这种方法简单快速,通过直接在随机文库中插入茎环结构来提高文库结构多样性,人为在文库中添加靶标的结合位点。但由于 PCR 扩增中的突变和错配的发生,以及随机片段碱基与固定序列碱基配对而导致的错配,都会导致预先设计好的茎环结构在最后未能保留^[40]。

2.2 定向进化法

2.2.1 一级结构序列定向进化法 由于 PCR 的扩增偏好性,在核酸适配体筛选过程中富集的核酸适配体在扩增过程中丢失,从而导致筛选的失败。Lin 等在体外核酸适配体分子进化过程中引入诱变 PCR 技术^[41],不仅解决了 PCR 扩增偏好性的问题,而且进一步强化变异逐步扩延了文库,加上渐趋苛刻的选择逐步吐弱纳强,促使核酸适配体序列趋向高亲和力^[4, 42]。其中,2'-脱氧次黄嘌呤核苷-5'-三磷酸(2'-deoxyinosine-5'-triphosphate, dITP)的引入在一级结构序列上丰富了文库容量,并且 dITP 能与核酸中的其他 5 种碱基形成配对^[43],因此,文库二级结构也会相应发生一定程度的改变。这种方法利用分子生物学较好地解决了 PCR 扩增偏好性的问题,虽然目前暂时仍未将 dITP 引入到文库的分子模

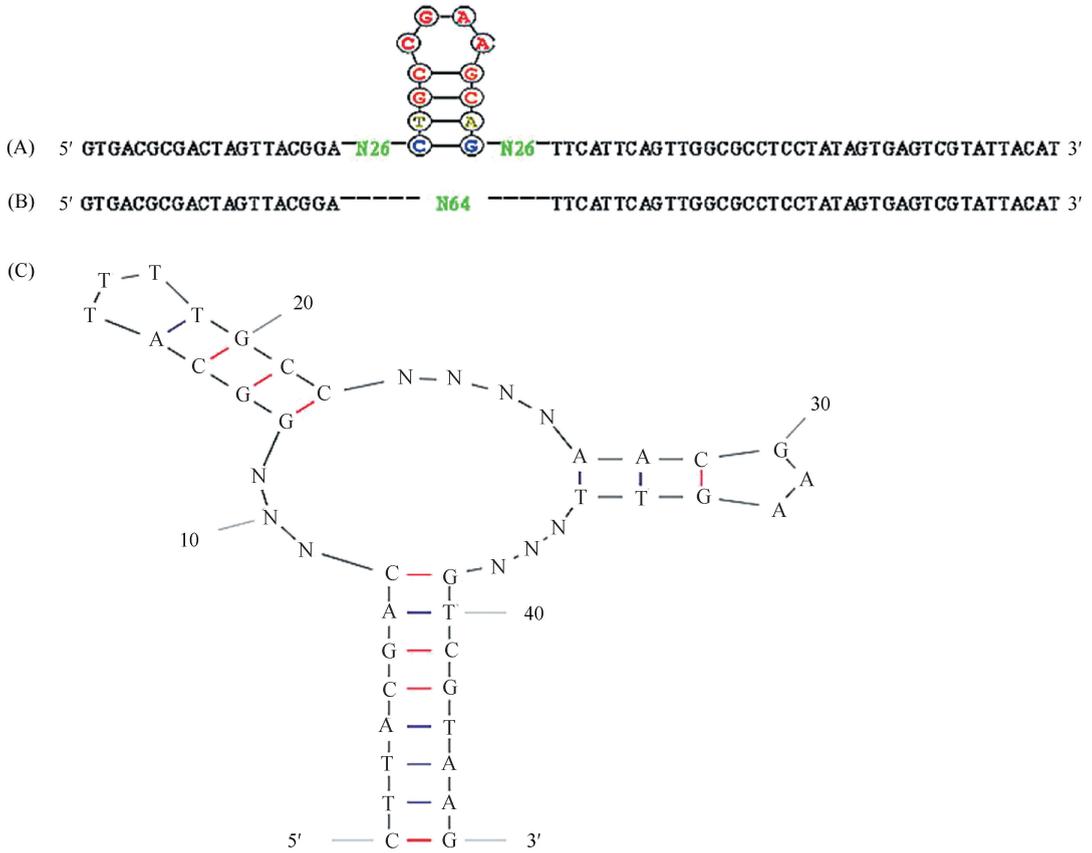


Fig.1 The library designed by embedded method (A) The library of GTP designed by embedded method. (B) Random library of GTP. (C) The sequence of the randomized three-way junctions (TWJs) pool of 3,4-methylenedioxypropylvalerone (MDPV)

拟设计中,但这仍是分子模拟指导核酸适配体文库设计的一个新思路。

2.2.2 二级结构模块定向进化法 二级结构模块定向进化法是在已知结构模块与核酸适配体亲和力和功能密切相关的前提下,在文库设计过程中定向增加结构模块在初始文库中的比例。随机文库中的结构多样性缺乏,是复杂结构(如高阶结结构)在已有核酸适配体中罕见的原因^[44]。为了进一步证明这一点,Luo 等人从核酸适配体数据库^[45]中挑选了 2 793 种已知的核酸适配体,利用 ViennaRNA^[46]对其二级结构进行预测,同时利用 CountJunctions 算法计算其结节数,结果发现,只有不到 1%的核酸适配体具有四阶或五阶结节结构,值得注意的是,天然核酸适配体往往具有高阶结节结构。因此,Luo 等人利用随机过滤法:即对随机文库中的结节结构类型进行统计,并将其中的五阶结节结构的非引物部分进行随机突变以增加文库多样性(多阶结节结构),通过这种方法设计获得的 RFPoolA 文库(Fig.2A),经统计发现,RFPoolA 文库与随机文库相比高阶结

节结构含量明显增加;利用基因过滤法:即设置一个目标文库,其中各阶结构占有固定的比值,将经过随机过滤的文库进行突变重组拷贝等操作,直至参数与目标文库一致,使文库结构分布均匀得到文库 GFPool1 (Fig. 2B),并用 GFPool1 文库利用与 Huizenga^[47]相同的筛选方法及条件,筛选获得 2 条呈五阶结节结构的 ATP 核酸适配体^[48]。

这种方法能够从结构复杂性和多样性两方面定向设计,适用于任何靶标的初始文库设计,但其局限性在于这种方法需要从结构的复杂性和多样性上对文库进行考量,相对于其他方法较为麻烦。

2.3 图案化文库设计法

Knight 的研究发现碱基在文库中的位置对文库的丰度和文库序列的折叠是有影响的^[49]。图案化文库其实就是通过限定嘌呤碱基与嘧啶碱基的相对位置以形成具有一定排布特点的文库。受蛋白质中疏水性氨基酸与亲水性氨基酸交替排布^[50]启发,Ruff 将让嘌呤(R)与嘧啶(Y)交替排布以增加碱基配对的概率,设计了两套文库设计方案:(1) 让 R

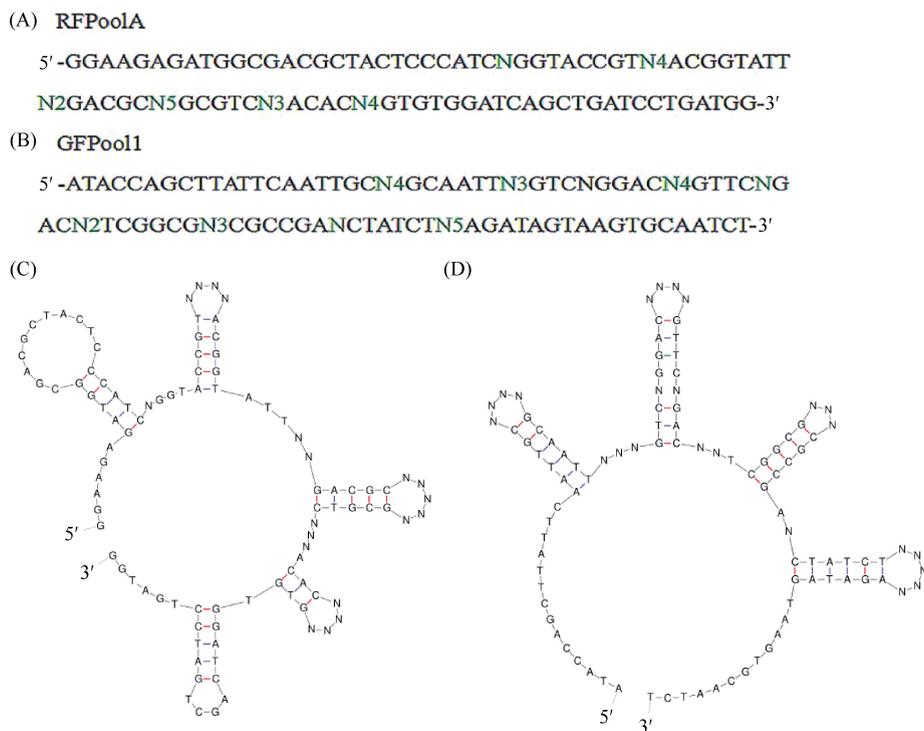


Fig.2 The sequences and secondary structure of library designed by directional evolve method (A) The sequence of RFPoolA designed by Random Filtering method. (B) The sequence of GFPool1 designed by Genetic Filtering method. (C) The secondary structure of RFPoolA predicted by Mfold. (D) The secondary structure of GFPool1 predicted by Mfold

与 Y 作为一个复合体在文库中出现以增加形成茎状结构的几率,并在它们中间插入随机序列 Nn 以形成环状、凸环等结构作为靶标的可能结合位点,其中 R 和 Y 中的两种碱基出现概率相同,形成 RY 文库 (Fig.3B); (2) 为碱基避免过度约束碱基位置导致二级结构不稳定,将原本的 R (嘌呤) 和 Y (嘧啶) 设计成含有少量嘧啶和嘌呤的碱基混合体,中间插入随机序列 Nn,形成 R * Y * 文库 (Fig.3A)。随后将相同长度的两种图案化的文库与随机文库混合,分别筛选链霉亲和素、免疫球蛋白 IgE (immunoglobulin E, IgE)、表皮生长因子受体 (epidermal growth factor receptor, VEGF) 的核酸适配体,最后发现 R * Y * 文库在数轮筛选后的文库富集情况、单克隆序列在文库中的出现频率以及核酸适配体亲和力三个方面均优于 RY 文库及随机文库^[51]。

图案化文库设计法从碱基组成及其在文库中的位置入手对文库进行设计,同时对文库的平均折叠自由能进行评估,适用于任何靶标的初始文库设计,但其设计本意是要增加文库中的茎环、凸环等复杂结构含量,虽然从理论上可以达到设计目的但并没有对文库中的结构情况进行统计。

(A) R * Y * library (R * Y *)₄N₄(R * Y *)₅N₃(R * Y *)₅N₄(R * Y *)₅N₃(R * Y *)₄

R * =45:5:45:5 A/C/G/T ; Y * =5:45:5:45 A/C/G/T

(B) RY library (RY)₄N₄(RY)₅N₃(RY)₅N₄(RY)₅N₃(RY)₄

R =50:0:50:0 A/C/G/T ; Y =0:50:0:50 A/C/G/T

Fig.3 The sequences of RY library and R * Y * library designed by patterned method (A) The sequence of R * Y * library, the R * and Y * are the compound of purines and pyrimidines, and N means random bases (N = A, C, G, U).

(B) The sequence of RY library. R: purines; Y: pyrimidines

2.4 图案化结合虚拟筛选法

在 Ruff^[51] (PT1 文库) 与 Luo^[48] (PT2 文库) 的工作基础上, Chushak 将 R 与 Y 进行不同的排列组合新设计了 2 个文库 (PT3; PT4) (Fig.4), 并利用分子模拟的方法从两个层次对 4 个文库进行浓缩优化。(1) 通过二级结构预测软件 ViennaRNA^[46] 预测文库序列的二级结构, 从最小自由能 (最小自由能越低结构越稳定) 与未配对碱基数 (未配对碱基结构更加灵活更容易与靶标结合; 未配对碱基可能作为原子的供体/受体与靶标形成氢键) 两个方面进行考量, 从二级结构上对文库进行第一轮优化; (2) 利用三维结构预测软件

Rosetta^[52]模拟文库序列三维结构,再利用分子对接软件 AutoDock Vina^[17]将文库序列与靶标进行对接打分,筛选出与靶标结合的高分序列对文库进一步浓缩和优化文库^[53]。

这种方法弥补了单纯的图案化设计法的不足,它不仅从碱基含量和位置上对文库进行了设计,还从结构稳定性及多样性上对初始文库进行浓缩,其次,它还能针对不同的靶标对文库进行优化,缩小筛选范围提高筛选效率,同时对小分子靶标的文库设计也同样适用。

PT1: **RYRY-N₂(RY)₂-N₂(RY)₂-N₂(RY)₂-N₂-RYRY**
 PT2: **RRYY-N₂-RRYY-N₂-RRYY-N₂-RRYY-N₂-RRYY-N₂-RRYY**
 PT3: **RRYY-N₂-RRYY-N₂-RRYY-N₂-RRYY-N₂-RRYY-N₂-RRYY**
 PT4: **RRYY-N₂(RY)₂-N₂(RY)₂-N₂(RY)₂-N₂-RRYY**

Fig.4 The sequences of pool PT1, PT2, PT3 and PT4

In the library, the purines (A, G) instead of R, the pyrimidines (U, C) replaced by Y and N means random bases (N=A, C, G, U)

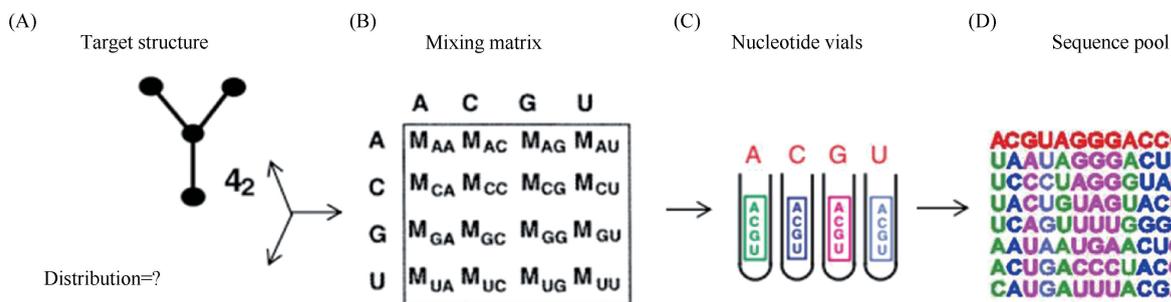


Fig.5 The process of Mixing Matrix method combined with Graph Theory Analysis (A) Propose pools with both random and biased sequence mutations around a specific sequence. (B) Introduction to the Mixing Matrix. (C) The elements specify mixing (or “contamination” or “doped”) in four vials (A, C, G, and U). (D) Apply mixing matrices to starting sequences lead to designed sequence pools

3 问题与展望

初始文库作为核酸适配体筛选的起点,其片段长度、随机程度、碱基位置及组成、文库的结构多样性等对筛选都具有较大的影响,基于计算机的分子模拟技术,能通过编程对文库的取样方式、文库中各碱基的位置以及组成比例进行设置,同时在二级结构上对文库进行分析,结合靶标特点对文库作进一步的优化和浓缩,不仅有效地缩小初始文库容量,而且能增加文库结构多样性,使文库中的目的序列在 SELEX 前就进行富集,从而从根本上提高 SELEX 技术的效率。但是目前该方法的应用也存在某些问题:(1)暂时仍无一种文库设计方法完美地适用于所有情况文库设计,研究者必须综合靶标分子的

2.5 混合矩阵结合图论分析法

2003年,Gan等设计了一种能用于文库结构分析的图论法^[54]。2005年,Gevertz等利用图论法分析发现,随机文库普遍缺乏结构多样性,并且文库的结构多样性取决于一级结构序列的取样方式^[8]。从一级结构取样方式以及结构多样性两个方面出发,Kim等^[55]设计了一种利用混合矩阵结合图论分析的文库设计的方法(Fig.5^[56],引自:http://rubin2.biomath.nyu.edu/modeling_pool_synthesis.html),这种方法首先利用4×4核苷酸转移概率矩阵法在序列水平上对文库进行取样,再结合图论法在二级结构水平上对文库结构多样性进行分析,既保证了文库序列取样的合理性,又保证了文库结构的多样性^[55],并且Gan与其团队将这种文库设计方法编程做成了文库设计软件RAGPOOLS^[56]。

这种方法适用于任何靶标,但计算时间长,需要高性能的计算设备,并且比较繁琐。

特点、核酸适配体的应用方向等多种因素,选择适合自己的文库设计方案;(2)在分子水平上对核酸适配体三维结构及核酸适配体与靶标结合的机制知之甚少,分子模拟难以百分之百模拟出靶标与核酸适配体的结合环境;(3)在计算机水平上高通量运算对计算机性能要求较高,同时对核酸适配体以及靶标三维结构的处理软件也需要进一步的完善和提高,探索出除文库结构外会影响核酸适配体与靶标结合的其他因素,以便更好的指导核酸适配体文库的设计。

参考文献 (References)

- [1] Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase [J]. Science, 1990, 249(4968): 505-510

- [2] Wang J, Rudzinski JF, Gong Q, *et al.* Influence of target concentration and background binding on in vitro selection of affinity reagents[J]. *PLoS One*, 2012, **7**(8): e43940
- [3] Kim YS, Gu MB. Advances in aptamer screening and small molecule aptasensors[J]. *Adv Biochem Eng Biotechnol*, 2014, **140**: 29-67
- [4] Lin JS, Kauff A, Diao Y, *et al.* Creation of DNA aptamers against recombinant bone morphogenetic protein 15[J]. *Reprod Fertil Dev*, 2015.doi:10.1071/RD14409
- [5] Crameri A, Stemmer WP. 10 (20)-fold aptamer library amplification without gel purification[J]. *Nucleic Acids Res*, 1993, **21**(18): 4410
- [6] Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR)[J]. *J Biosci Bioeng*, 2003, **96**(4): 317-323
- [7] Carothers JM, Oestreich SC, Davis JH, *et al.* Informational complexity and functional activity of RNA structures[J]. *J Am Chem Soc*, 2004, **126**(16): 5130-5137
- [8] Gevertz J, Gan HH, Schlick T. In vitro RNA random pools are not structurally diverse: A computational analysis[J]. *RNA*, 2005, **11**(6): 853-863
- [9] Stich M, Briones C, Manrubia SC. On the structural repertoire of pools of short, random RNA sequences[J]. *J Theor Biol*, 2008, **252**(4): 750-763
- [10] Schlick T, Pyle AM. Opportunities and challenges in RNA structural modeling and design[J]. *Biophys J*, 2017, **113**(2): 225-234
- [11] Zhang S, Zhou J, Hu H, *et al.* A deep learning framework for modeling structural features of RNA-binding protein targets[J]. *Nucleic Acids Res*, 2016, **44**(4): e32
- [12] Rabal O, Pastor F, Villanueva H, *et al.* In silico aptamer docking studies: from a retrospective validation to a prospective case study-TIM3 aptamers binding[J]. *Mol Ther Nucleic Acids*, 2016, **5**(10): e376
- [13] Knight R, Yarrus M. Analyzing partially randomized nucleic acid pools: straight dope on doping[J]. *Nucleic Acids Res*, 2003, **31**(6): e30
- [14] Baskerville S, Zapp M, Ellington AD. High-resolution mapping of the human T-cell leukemia virus type 1 Rex-binding element by in vitro selection[J]. *J Virol*, 1995, **69**(12): 7559-7569
- [15] Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data[J]. *Nature*, 2008, **452**(7183): 51-55
- [16] Chen YC. Beware of docking! [J]. *Trends Pharmacol Sci*, 2015, **36**(2): 78-95
- [17] Di Muzio E, Toti D, Politicelli F. DockingApp: a user friendly interface for facilitated docking simulations with AutoDock Vina [J]. *J Comput Aided Mol Des*, 2017, **31**(2): 213-218
- [18] Kurkuoglu Z, Koukos PI, Citro N, *et al.* Performance of HADDOCK and a simple contact-based protein-ligand binding affinity predictor in the D3R Grand Challenge 2[J]. *J Comput Aided Mol Des*, 2018, **32**(1): 175-185
- [19] Schneidman-Duhovny D, Inbar Y, Nussinov R, *et al.* PatchDock and SymmDock: servers for rigid and symmetric docking[J]. *Nucleic Acids Res*, 2005, **33**(Web Server issue): W363-W367
- [20] Ahirwar R, Nahar S, Aggarwal S, *et al.* In silico selection of an aptamer to estrogen receptor alpha using computational docking employing estrogen response elements as aptamer-alike molecules [J]. *Sci Rep*, 2016, **6**: 21285
- [21] Hall B, Micheletti JM, Satya P, *et al.* Design, synthesis, and amplification of DNA pools for in vitro selection[J]. *Curr Protoc Nucleic Acid Chem*, 2009, **Chapter 9**: Unit 9.2
- [22] Cao S, Chen SJ. A domain-based model for predicting large and complex pseudoknotted structures[J]. *RNA Biol*, 2012, **9**(2): 200-211
- [23] Cao S, Chen SJ. Physics-based de novo prediction of RNA 3D structures[J]. *J Phys Chem B*, 2011, **115**(14): 4216-4226
- [24] Xu X, Dickey DD, Chen S, *et al.* Structural computational modeling of RNA aptamers[J]. *Methods*, 2016, **103**: 175-179
- [25] Bock LC, Griffin LC, Latham JA, *et al.* Selection of single-stranded DNA molecules that bind and inhibit human thrombin [J]. *Nature*, 1992, **355**(6360): 564-566
- [26] Kuryavyi V, Majumdar A, Shallop A, *et al.* A double chain reversal loop and two diagonal loops define the architecture of a unimolecular DNA quadruplex containing a pair of stacked G (syn)-G (syn)-G (anti)-G (anti) tetrads flanked by a G-(T-T) Triad and a T-T-T triple [J]. *J Mol Biol*, 2001, **310**(1): 181-194
- [27] Ikebukuro K, Okumura Y, Sumikura K, *et al.* A novel method of screening thrombin-inhibiting DNA aptamers using an evolution-mimicking algorithm [J]. *Nucleic Acids Res*, 2005, **33**(12): e108
- [28] Szostak JW. Functional information: molecular messages [J]. *Nature*, 2003, **423**(6941): 689
- [29] Bing T, Yang XJ, Mei HC, *et al.* Conservative secondary structure motif of streptavidin-binding aptamers generated by different laboratories [J]. *Bioorg Med Chem*, 2010, **18**(5): 1798-1805
- [30] Dieckmann T, Suzuki E, Nakamura GK, *et al.* Solution structure of an ATP-binding RNA aptamer reveals a novel fold[J]. *RNA*, 1996, **2**(7): 628-640
- [31] Jiang F, Kumar RA, Jones RA, *et al.* Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex [J]. *Nature*, 1996, **382**(6587): 183-186
- [32] Jenison RD, Gill SC, Pardi A, *et al.* High-resolution molecular discrimination by RNA [J]. *Science*, 1994, **263**(5152): 1425-1429
- [33] Fan P, Suri AK, Fiala R, *et al.* Molecular recognition in the FMN-RNA aptamer complex [J]. *J Mol Biol*, 1996, **258**(3): 480-500
- [34] Geiger A, Burgstaller P, von der Eltz H, *et al.* RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity [J]. *Nucleic Acids Res*, 1996, **24**(6): 1029-1036
- [35] Davis JH, Szostak JW. Isolation of high-affinity GTP aptamers from partially structured RNA libraries [J]. *Proc Natl Acad Sci U S A*, 2002, **99**(18): 11616-11621
- [36] Li F, Lin Y, Le XC. Binding-induced formation of DNA three-way junctions and its application to protein detection and DNA strand displacement [J]. *Anal Chem*, 2013, **85**(22): 10835-10841
- [37] Neves MA, Reinstein O, Saad M, *et al.* Defining the secondary structural requirements of a cocaine-binding aptamer by a thermodynamic and mutation study [J]. *Biophys Chem*, 2010, **153**(1): 9-16
- [38] Canoura J, Wang Z, Yu H, *et al.* No Structure-switching required: a generalizable exonuclease-mediated aptamer-based assay for small-molecule detection [J]. *J Am Chem Soc*, 2018, **140**(31): 9961-9971
- [39] Yu H, Yang W, Alkhamis O, *et al.* In vitro isolation of small-molecule-binding aptamers with intrinsic dye-displacement functionality [J]. *Nucleic Acids Res*, 2018, **46**(8): e43
- [40] Yang KA, Pei R, Stefanovic D, *et al.* Optimizing cross-reactivity with evolutionary search for sensors [J]. *J Am Chem Soc*, 2012, **134**(3): 1642-1647
- [41] Kuipers OP. Random mutagenesis by using mixtures of dNTP and dTTP in PCR [J]. *Methods Mol Biol*, 1996, **57**: 351-356
- [42] Lin JS, McNatty KP. Aptamer-based regionally protected PCR for protein detection [J]. *Clin Chem*, 2009, **55**(9): 1686-1693
- [43] Oda Y, Uesugi S, Ikehara M, *et al.* NMR studies for identification of dI:dG mismatch base-pairing structure in DNA [J]. *Nucleic Acids Res*, 1991, **19**(19): 5263-5267
- [44] Khoo D, Perez C, Mohr I. Characterization of RNA determinants recognized by the arginine- and proline-rich region of Us11, a herpes simplex virus type 1-encoded double-stranded RNA binding protein that prevents PKR activation [J]. *J Virol*, 2002, **76**(23): 11971-11981
- [45] Lee JF, Hesselberth JR, Meyers LA, *et al.* Aptamer database [J]. *Nucleic Acids Res*, 2004, **32**(Database issue): D95-D100

- [46] Hofacker IL. Vienna RNA secondary structure server[J]. *Nucleic Acids Res*, 2003, **31**(13): 3429-3431
- [47] Huizenga DE, Szostak JW. A DNA aptamer that binds adenosine and ATP[J]. *Biochemistry*, 1995, **34**(2): 656-665
- [48] Luo X, McKeague M, Pitre S, *et al.* Computational approaches toward the design of pools for the in vitro selection of complex aptamers[J]. *RNA*, 2010, **16**(11): 2252-2262
- [49] Knight R, De Sterck H, Markel R, *et al.* Abundance of correctly folded RNA motifs in sequence space, calculated on computational grids [J]. *Nucleic Acids Res*, 2005, **33**(18): 5924-5935
- [50] Xiong H, Buckwalter BL, Shieh HM, *et al.* Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides [J]. *Proc Natl Acad Sci U S A*, 1995, **92**(14): 6349-6353
- [51] Ruff KM, Snyder TM, Liu DR. Enhanced functional potential of nucleic acid aptamer libraries patterned to increase secondary structure[J]. *J Am Chem Soc*, 2010, **132**(27): 9453-9464
- [52] Leaver-Fay A, Tyka M, Lewis SM, *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules[J]. *Methods Enzymol*, 2011, **487**: 545-574
- [53] Chushak YG, Martin JA, Chavez JL, *et al.* Computational design of RNA libraries for in vitro selection of aptamers [J]. *Methods Mol Biol*, 2014, **1111**: 1-15
- [54] Gan HH, Pasquali S, Schlick T. Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design [J]. *Nucleic Acids Res*, 2003, **31**(11): 2926-2943
- [55] Kim N, Gan HH, Schlick T. A computational proposal for designing structured RNA pools for in vitro selection of RNAs [J]. *RNA*, 2007, **13**(4): 478-492
- [56] Kim N, Shin JS, Elmetwaly S, *et al.* RagPools: RNA-As-Graph-Pools--a web server for assisting the design of structured RNA pools for in vitro selection [J]. *Bioinformatics*, 2007, **23**(21): 2959-2960