



基于分子对接和步进序列群的蓖麻毒素 核酸适配体序列优化*

刘 佳^{1,2)} 杨志芳²⁾ 王 闯^{2,3)} 肖 兰^{2,3)} 郭 磊^{2)**} 吴海霞^{1)**} 谢剑炜²⁾

⁽¹⁾ 河北科技大学化学与制药工程学院, 石家庄 050091;

⁽²⁾ 军事医学研究院毒物药物研究所, 抗毒药物与毒理学国家重点实验室, 北京 100850;

⁽³⁾ 中央民族大学药学院, 教育部民族医药重点实验室, 北京 100081)

摘要 **目的** 有效结合分子对接预测和表面等离子体共振实验评价技术, 获得亲和力更强、序列最短的最优适配体。**方法** 针对前期筛选出的靶向蓖麻毒素的3条80 nt单链DNA适配体(L14、P3、L7), 在明确各自二维随机区茎环序列与靶蛋白结合能力的基础上, 以H-DOCK分子对接为指导, 分别确定蓖麻毒素适配体随机区的最短结合单元, 从而构建两端延长步进序列群, 以表面等离子体共振技术测定序列群序列的亲和力和动力学参数, 明确适配体的结合关键结构, 从而筛选得到最优适配体。**结果** 3条全长适配体的随机区适配体L14r、P3r、L7r均可形成一定的茎环结构, 其中L14r较L14的亲和力增强9倍、L7r增强2倍、P3r基本不变。对随机区适配体和蓖麻毒素进行分子对接, 结果显示, L14r、P3r、L7r的对接分数值皆优于阴性序列40T, 结合关键氨基酸个数分别为11、8、9个, 存在距离小于5 Å的预测结合位点分别为20、12、15个, 具有良好的与蓖麻毒素的结合能力。进一步明确了蓖麻毒素活性口袋所容纳的适配体最短结合单元L14rm、P3rm、L7rm的序列构成, 在此基础上构建出两端延长步进序列群。针对该步进群, 基于结合关键氨基酸个数、结合位点个数、对接得分等参数的变化和表面等离子体共振测定结果筛选出最优适配体。所获得的最优适配体L14rm、L7rm-2亲和力继续增强了1~2倍。**结论** 随机区适配体能有效地与蓖麻毒素结合, 较之全长适配体亲和力更强, 分子对接结合步进序列群设计, 仅使用17条序列, 便有效获得了3条最优适配体并明确其结合作用。3条结合蓖麻毒素的最优适配体——L14rm、P3r、L7rm-2的 K_D 值分别为(64±30)、(167±19)、(120±1) nmol/L, 亲和力提高到全长适配体的14、1、4倍。

关键词 蓖麻毒素, 适配体, 分子对接, 步进序列群, 表面等离子体共振, 结构优化

中图分类号 Q52

DOI: 10.16476/j.pibb.2023.0280

蓖麻毒素(ricin communis agglutinin 60, RCA60)是最早发现的植物毒素蛋白, 也是一种凝集素, 在油料作物——蓖麻的种子(蓖麻子)中高丰度存在, 一般可占到植物种子重量的1%~5%。RCA60的分子质量约66 ku, 结构由A、B两条链组成, 其中B链是一种对半乳糖(胺)具有特异性的凝集素, 可与真核细胞膜上的糖蛋白或糖脂的半乳糖位点结合, 介导A链进入胞内, 而A链具有N-糖苷酶活性, 能够识别真核细胞28S核糖体RNA(ribosomal RNA, rRNA)中含有5'-GAGA-3'序列的特定核苷酸(nucleotide, nt)区域即 α 帚曲霉蛋白-蓖麻毒素环(α -sarcin-ricin loop, SRL)结构域, 脱去其特定的腺嘌呤(如水解小鼠细胞28S

rRNA的A₄₃₂₄位点腺苷酸上的N-糖苷键), 使核糖体失活并不可逆地抑制蛋白质合成, 继而导致细胞死亡^[1]。

RCA60是同时被列入国际《禁止化学武器公约》和国际《禁止生物和毒素武器公约》的唯一一种蛋白质毒素^[2], 其小鼠静脉染毒的半数致死剂量为5~10 μ g/kg^[3], 具有剧烈毒性, 且来源广泛、易于制备、无特效解毒药。近年来国内外发生了包括误食蓖麻子在内的多起中毒事件, 以及多起

* 国家自然科学基金(21974152)资助项目。

** 通讯联系人。

郭磊 Tel: 010-68225893, E-mail: guolei@bmi.ac.cn

吴海霞 Tel: 0311-81668373, E-mail: wuhaixia@hebust.edu.cn

收稿日期: 2023-07-17, 接受日期: 2023-08-06

RCA60 白色粉末信件等恐怖事件,说明该类型毒素对公共安全和民众健康存在高度危害。一直以来,发展灵敏快速的分析检测方法、筛选有效的抑制剂等都是RCA60防治研究的重要内容。

通过体外文库技术——指数富集的配体系统进化 (systematic evolution of ligands by exponential enrichment, SELEX) 所筛选到的核酸适配体 (aptamer),具有高亲和力、高特异性、可功能化修饰等特点,且易于实现在体外的大量、快速合成,被誉为“化学抗体”^[4],是一类高效的亲和识别元件。对于RCA60,目前已经报道了几种RNA或单链DNA (ssDNA) 适配体,例如Fan等^[5]筛选出特异性结合RCA60 A链的RNA适配体RA80, Lamont等^[6]筛选出的特异性结合B链的DNA适配体SSRA1、本课题组针对RCA60全蛋白进行筛选所获得的ssDNA适配体A3、C1、C5^[7]、L7、L14、P3^[8]等。

一般而言,通过SELEX技术所获得的适配体由两端固定引物和中间的一段非引物序列组成,典型全长为70~130 nt,但绝大多数情况下,并不是全长序列均参与靶分子的识别,核心识别区域的长度为20~50 nt不等,其他则为辅助支撑结构或无效序列^[9],因此,在后续的应用或机理研究中需要首先针对全长适配体进行一定的结构裁剪^[10]。

自1990年SELEX出现至今,国内外已经出现了多种针对适配体的裁剪方法,绝大多数为经验试错法。针对模拟产生的全长适配体的二级结构,人为界定其局部高级结构(茎环、假结等)后,再采用去除未配对结构、去除或缩短茎、缩小环等手法获得截短结构^[11],还可对于截短结构再次进行随机裁剪^[12]、定点突变^[13],或通过遗传算法等构建小型突变文库等获取更优结构^[14]。前期本课题组^[5]通过构建步进式文库,构建向左缩进、向右缩进、两端收缩、两端延伸序列群的途径,以合理“穷尽”适配体序列中与靶蛋白产生亲和力的大多数可能结果,针对重组促红细胞生成素 α ,成功裁剪优化出一条27 nt的ssDNA适配体。对于RCA60的全长适配体,早期仅采用二维结构模拟及简单的局部剪裁等思路,例如SSRA1是将60 nt全长的适配体保留随机区结构,经酶联免疫吸附法获得其与RCA60在不同溶液中的解离常数(dissociation constant, K_D)值为5~22 nmol/L。

多种适配体的结构剪裁方法极大依赖于经验和直觉,具有较大的工作量以及不确定性。本工作以

分子对接模拟为指导、灵活引入步进式序列群策略,驱动实验结果进行评估验证。首先将本课题组前期筛选出的3条全长ssDNA适配体L14、P3、L7,在去掉引物区保留其随机区核苷酸序列的基础上,通过三维(3D)分子对接预测了随机区适配体L7r、P3r、L14r与RCA60的结合情况,并通过所预测的蓖麻毒素活性口袋具体结合位点,分别确定了3条适配体的最短结合单元。继而构建基于最短结合单元的两端延长步进序列群,使用表面等离子体共振技术(surface plasmon resonance, SPR)测定该序列群与RCA60的亲和力和动力学参数,成功快速筛选得到亲和力更强、序列最短的最优适配体。

1 材料与方法

1.1 材料与试剂

微量RCA60、蓖麻凝集素(ricin agglutinin, RCA120)、相思子毒素(abrin)均由本实验室微量制备,电泳纯度大于95%。L14 (CAGCTCAGAAGCTTGATCCTGTGACAGCAGGGGGAGTGTGCGTAATAAGCGAGGAGATGACTCGAAGTCGTGCATAATCA)、P3 (CAGCTCAGAAGCTTGATCCTGTGAGTCGACCTGAGCCCGAGCAGGTTCGGAATCCGTGGACTCGAAGTCGTGCATCTGCA)、L7 (CAGCTCAGAAGCTTGATCCTGTGGACAGGAGCTGGTTAAATAGGCAGCACCGAGCAGACTCGAAGTCGTGCATCTGCA)、L14rm (CGTAATAAGCG)、P3r (GAGTCGACCTGAGCCCGAGCAGGTTCGGAATC)、L7rm-2 (TGGACAGGAGCTGGTTAA)等ssDNA寡核苷酸序列由生工生物工程公司合成(HPLC纯化)。N-羟基琥珀酰亚胺(N-hydroxysuccinimide, NHS)、1-乙基-3-(3-二甲基氨基丙基)碳酰二亚胺盐酸盐(1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride, EDC)、乙醇胺、氢氧化钠、乙酸钠-乙酸缓冲液(pH 4.0)均购自Cytiva公司(美国),其中EDC、NHS、乙醇胺来源于氨基偶联试剂盒(Cytiva, BR100050); 4-(2-羟乙基)哌嗪-1-乙磺酸(4-(2-hydroxyethyl) piperazine-1-ethanesulfonic acid, HEPES)、Tween 20购自于Sigma公司(Missouri, 美国)。Zeba Spin脱盐柱(截止分子量7 ku)购自于Thermo(美国)。超纯水为18.2 M Ω ·cm,由Milli-Q A10水净化系统(Millipore, MA, 美国)制备。所有其他试剂均为

分析纯及以上纯度, 购自于国药化学试剂有限公司(北京, 中国)。

RCA60是高毒性生物毒素, 所有相关实验均应根据生物安全操作指南进行。所有实验都在通风良好的通风橱中进行, 并佩戴防护手套和护目镜。实验结束后, 通过沸水蒸煮 30 min 或高压灭菌(121°C, 0.2 MPa) 20 min, 对含有相关毒素的样品进行完全洗消, 然后用碱液(0.1 mol/L NaOH, pH>10)浸泡 1 h 以上^[15]。

1.2 SPR

所有亲和力及动力学参数测定均在 Biacore T200 分子互作仪上进行(Cytiva, 美国), 缓冲液为 HBS-T (10 mmol/L HEPES, 140 mmol/L NaCl, 0.5% Tween 20, 用 NaOH 调 pH 为 7.4), 包括缓冲液配制用水及进样针清洗、仪器管路清洗用水在内的仪器用水均经过灭菌处理并用 0.22 μm 滤膜(津腾, 中国)过滤后使用。

蛋白质在共价偶联前经脱盐处理, 通过 Nanodrop 仪(Eppendorf, 德国)测定其 260 nm、280 nm 处的紫外吸收, 并使用 Lowry-Kalcker 经验公式“蛋白质浓度(g/L) = $1.45A_{280} - 0.74A_{260}$ ”计算获得蛋白质浓度^[16], 继而加入 10 mmol/L 乙酸钠-乙酸缓冲液(pH 4.0)稀释至 50 mg/L。

1.2.1 芯片活化与蛋白质的共价偶联

使用 Series S Sensor Chip CM5 (Cytiva, 美国) 芯片^[17], 使用 100 mmol/L EDC、391.2 mmol/L NHS 混合溶液流经 CM5 芯片中的 Fc 2 通道表面, 流速为 10 $\mu\text{L}/\text{min}$, 持续时间为 900 s, 以活化 CM5 芯片上葡聚糖的羧基。向 Fc 2 通道注入 50 mg/L 的 RCA60 (pH 4.0) 溶液, 流速为 10 $\mu\text{L}/\text{min}$, 持续时间 120 s, 荷正电的 RCA60 (等电点为 7.0^[18]) 通过静电作用吸引到荷负电的葡聚糖表面后, 其氨基与葡聚糖上被活化的羧基实现共价偶联。每次蛋白质偶联水平平均固定为 5 000 响应单位(response unit, RU), 最后用 HBS-T 缓冲液冲洗 2 min 至基线平稳^[5]。Fc 1 为空白通道。

1.2.2 样品处理及测定

所有寡核苷酸序列均以超纯水配制为 100 $\mu\text{mol}/\text{L}$ 的储备液, 按使用量分装并置于 -20°C 条件下储存, 避免反复冻融。使用时, 所有寡核苷酸序列均以 HBS-T 缓冲液稀释至 2 $\mu\text{mol}/\text{L}$, 95°C, 5 min 变性后并缓慢恢复至室温, 然后稀释为合适的浓度梯度, 提交至 SPR 进行测定, 测定温度为 25°C。在 Biacore T200 中, 装载入已偶联好蛋白质

的 CM5 芯片, 缓冲液为 HBS-T, 流速为 30 $\mu\text{L}/\text{min}$, 结合时间 120 s, 解离时间 180 s, 再生条件为 0.1 mmol/L NaOH, 30 $\mu\text{L}/\text{min}$, 30 s, 再生后基线变化维持在 ± 5 RU 范围内。

对于全长及剪裁的各种适配体, 首先在 1 $\mu\text{mol}/\text{L}$ 的浓度下进行单次结合实验, 判断序列变化对结合相互作用的影响; 再对出现阳性结果的样品进行多循环动力学(multiple cycle kinetics, MCK)测定, 进行亲和力和动力学参数评价。所有操作由 Biacore T200 Control Software (2.0.0) 软件控制。

1.2.3 数据分析

使用 Biacore T200 Evaluation Software (2.0.0) 软件, 采用 Kinetics, 1 : 1 结合模式进行拟合, 获得结合速率常数(association velocity constant, k_a)、解离速率常数(dissociation velocity constant, k_d)、 K_D 等数据。以 $U < 15$, χ^2 接近 1 筛选出合理的拟合结果。

1.3 分子对接

所有寡核苷酸序列通过 Mfold (<http://www.mfold.org/mfold/applications/dna-folding-form.php>) 获取其二维(2D)结构, 条件为 5 mmol/L Mg^{2+} , 140 mmol/L Na^+ , 从中选取 Gibbs 自由能(ΔG)最低的模拟结构。基于上述 2D 结构, 通过 RNAcomposer (<https://rnacomposer.cs.put.poznan.pl/>) 实现 3D 结构预测, 并采用 OpenBabel 2.4.1 软件获取 pH 7.4 下的 RCA60 (PDB ID: 3RTJ) 和多条候选寡核苷酸序列的 3D 结构, 并使用 SYBYL-X2.0 基于 Tripos 力场进行能量最小化以减少构建模拟系统中的非正常构象。针对 RCA60, 设置其 A 链的 rRNA N-糖苷酶活性口袋中的 14 个关键残基(R48、D75、N78、Y80、V81、D96、D100、G121、Y123、R134、E177、R180、E208、W211) 为适配体对接位点, 将 RCA60 与多条序列通过 H-DOCK (<http://hdock.phys.hust.edu.cn/>) 进行分子对接^[19]。计算前 100 个姿势的 DS 值和配体 RMSD 值, 进一步选择 10 种较低 DS 的姿势。

2 结果与讨论

针对经 SELEX 筛选获得的核酸适配体时, 一般均需通过结构剪裁, 获取较短的适配体序列, 既保留了其针对靶分子的核心序列, 也节省了经济成本, 这也成为了适配体应用于生物传感、即时诊断、DNA 纳米技术、药物靶向递送之前的必备步骤。

另外,在筛选和优化适配体时,合适的适配体与靶分子间结合作用的预测及评价方法是必要的。预测方法方面,已涌现出使用高通量测序技术、FASTAptamer、APtaCluster等工具对实验产生的大量序列进行测序、聚类、模型搜索以及多维度评分,之后对筛选的结果进行结构模拟并进一步优化策略,其中通过Mfold、Ufold、RNAcomposer、ARES等方法进行2D、3D结构预测,并使用H-DOCK、GRAMM、Z-DOCK等对适配体和靶分子进行分子对接预测,这种策略允许更精确地评估适配体-靶分子复合物的结合能力,可为进一步的序列截短或突变提供有价值的信息^[20],但目前尚未见基于分子对接指导适配体结构剪裁的系统报道。实验方法方面,普遍使用凝胶电泳迁移率变动分析^[21]、SPR^[22]、生物膜干涉技术^[23]、等温滴定量热法^[24]、微量热泳动法^[25]、分析超速离心技术^[26],根据结合作用前后的特征值变化判定亲和力及得到其他相关特征值如动力学参数、热力学参数等^[27]。其中SPR是一种实时无标记的检测技术,在分子相互作用评价方面应用甚广,可以高通量地

提供亲和力及动力学特征参数。本工作即以分子对接和SPR为核心技术,以RCA60为靶蛋白,提出了并评估以3D分子对接模拟及步进序列群指导适配体优化的策略,可仅提交少量序列至SPR进行结合评价实验。

2.1 适配体全长与随机序列的分子对接与SPR测定

本团队前期通过毛细管电泳-SELEX技术,采用非涂层毛细管或者中性涂层毛细管,进行了4轮筛选,分别筛选到了L14、P3、L7等3条可与RCA60结合的适配体序列,全长80 nt,均存在多个茎环结构,具有较低的 ΔG 和 K_D 值。经SPR测定,与已报道的ssDNA适配体SSRA1相比,P3略低于SSRA1的 K_D 值,但L7、L14的 K_D 值要高2~4倍(图1)。该SPR评价方法所得的结果与筛选者Lamont等^[6]获取的 K_D 值有所差异,应该是由于使用的评价方法不同引起的。本工作所有的评价实验都基于SPR进行,这就保持了较好的评价一致性和可信度。

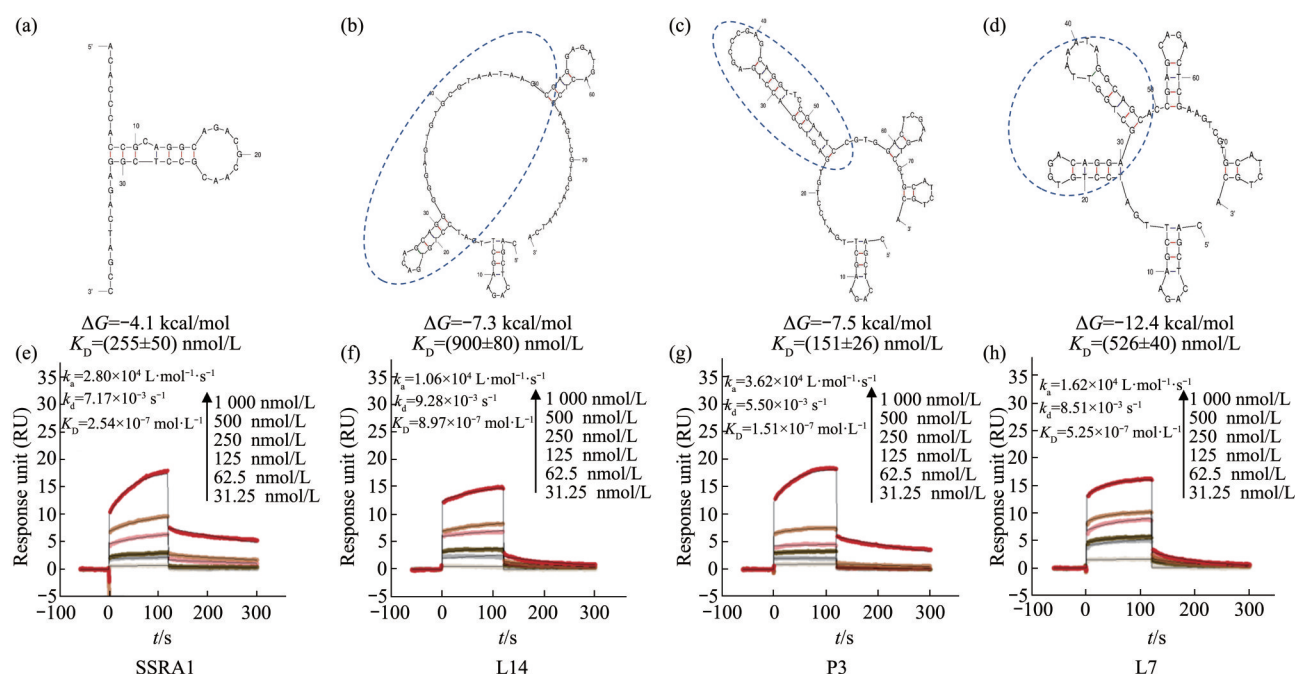


Fig. 1 The 2D structures, Gibbs' energy, and the SPR evaluation results of four aptamers, SSRA1, L14, P3 and L7

The Mfold results of SSRA1 (a), L14 (b), P3 (c) and L7 (d); and the multiple cycle kinetics results of SSRA1 (e), L14 (f), P3 (g) and L7 (h).

2.2 随机区适配体与分子对接

Goto等^[28-29]根据RCA60和抑制剂的复合物的X射线晶体衍射结果揭示出,RCA60的活性口袋

中,G121、Y123、R134、E177、R180、E208、W211组成了初级口袋,R48、D75、N78、V81、D96、D100组成了次级口袋,两者被Y80残基划

分, Y80类似于“门”残基, 其运动可导致抑制剂从初级口袋滑向次级口袋, 发生相互作用, 这也可能是抑制剂作用不佳的根本原因^[30-31]。通过使用H-DOCK, 在pH 7.4的条件下, 针对RCA60的活性口袋14个氨基酸残基进行分子对接, 预先模拟判断裁剪的适配体与蛋白质的活性口袋是否存在结合, 并使用DS值概算比较各适配体间结合力大小, 选取合适的适配体序列进行下一步评价与验证。

先对全长适配体进行直接裁剪, 保留随机区域的发卡结构(具体剪裁位置如图1的2D结构图中的虚线位置所示), 分子对接结果显示, L14r (34 nt, DS -262.8, RMSD 92.5 Å)、P3r (32 nt, DS -250.1, RMSD 82.0 Å)、L7r (32 nt, DS -292.4, RMSD 99.4 Å)的DS值皆优于阴性序列40 T (40 nt,

DS -239.3, RMSD 91.5 Å)。其中L14r共与RCA60的11个关键氨基酸残基结合(R48、D75、N78、Y80、D96、D100、G121、R134、R180、E208、W211), 距离小于5 Å的预测结合位点有20个, P3r中的4个核苷酸与RCA60活性口袋中8个残基(R48、N78、Y80、D96、D100、R134、R180、W211)紧密连接, 包括4个次级口袋残基与3个初级口袋残基, 共存在12个距离小于5 Å的预测结合位点, L7r的2D结构为通过1个核苷酸相连的2个茎环形式, 其分子对接预测的结合位点则产生在2个茎部分, 与9个关键氨基酸相连(R48、N78、Y80、D96、D100、Y123、R134、E208、W211), 存在15个距离小于5 Å的预测结合位点(图2)。根据分子对接结果所示, 推测三者具有良好的与RCA60结合能力。

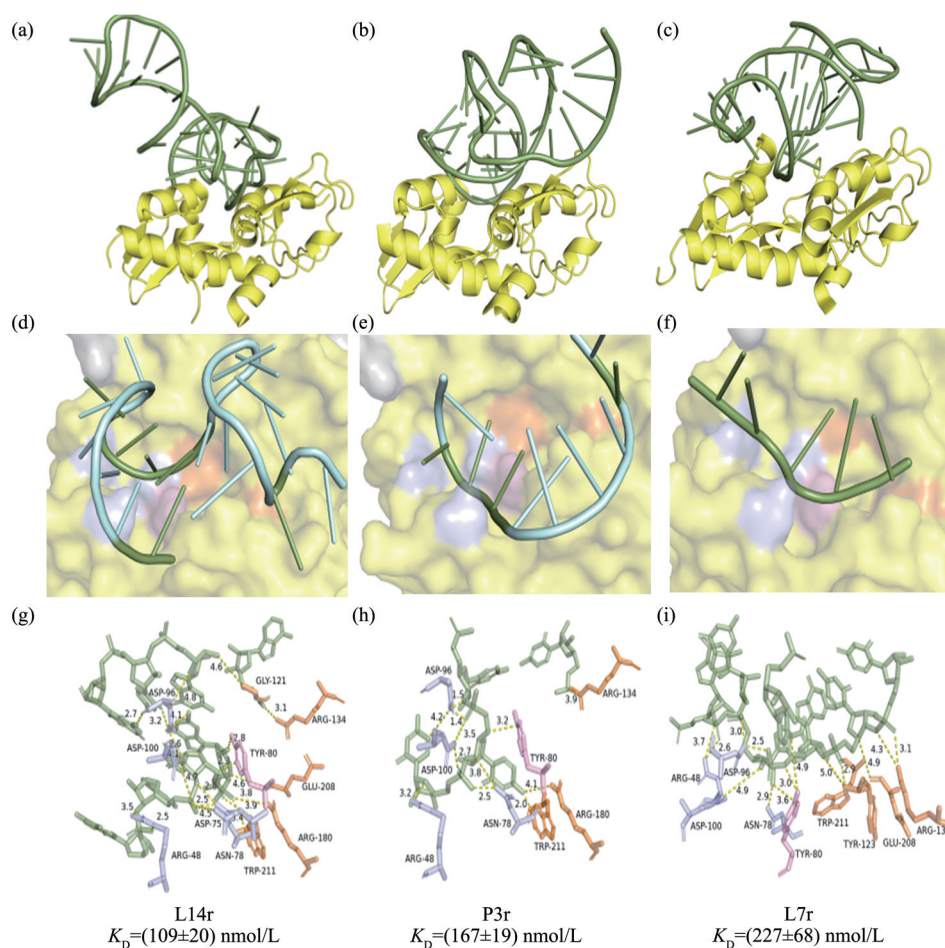


Fig. 2 The H-DOCK docking results of L14r, P3r and L7r

(a-c) The partial diagrams of RCA60 and L14r, P3r and L7r; (d-f) the binding of L14r, P3r and L7r to the primary and secondary active pockets of RCA60; (g-i) the docking bag details. Color annotation, yellow: RTA; green: aptamer; orange: primary pocket of RCA60; purple: secondary pocket of RCA60.

对3条随机区适配体进行SPR评价的结果表明, L14r、P3r、L7r与RCA60的 K_D 值分别为 (109 ± 20) 、 (167 ± 19) 、 (228 ± 68) nmol/L, 与L14、P3、L7适配体的 K_D 值 (900 ± 80) 、 (151 ± 26) 、 (526 ± 40) nmol/L相比, L14r、L7r的 K_D 值下降至全长适配体的1/9和1/2, 提示这两条全长适配体中两侧过长的序列结构在与靶蛋白作用的过程中阻碍了结合。对于L14r, 还可以从分子对接结果直观性的发现, 与活性口袋紧密结合仅有一段核苷酸, 提示可能还有进一步优化空间。P3全长序列与截短序列 K_D 值基本一致, 两端长链形成的空间结构对结合影响较不明显, 说明固定序列为无效序列。

2.3 确定最短结合单元及构建步进延长序列群

将L14r、P3r、L7r进行分子对接后, 可顺利获

取3条适配体与RCA60活性口袋的结合位点。针对上述3条随机区适配体, 仅选取其与RCA60活性口袋所结合的序列部分, 将之往外扩展至该部分能自折叠为一定的3D结构, 将上述序列作为最短适配体结合单元(minimum aptamer binding motif, MABM)。将L14r、P3r、L7r的MABM分别命名为L14rm、P3rm和L7rm, 分别比原随机区适配体截短了23、14、22 nt。进一步, 对上述MABM构建两端延长型的序列群, 以期延长的游离核苷酸链能通过产生更多的相互作用位点而起到一定的辅助结合效果, 除了有利于溶液条件下的亲和力评价, 亦可明确MABM与RCA60结合的具体作用规律。序列群组成如图3所示, 共17条序列。

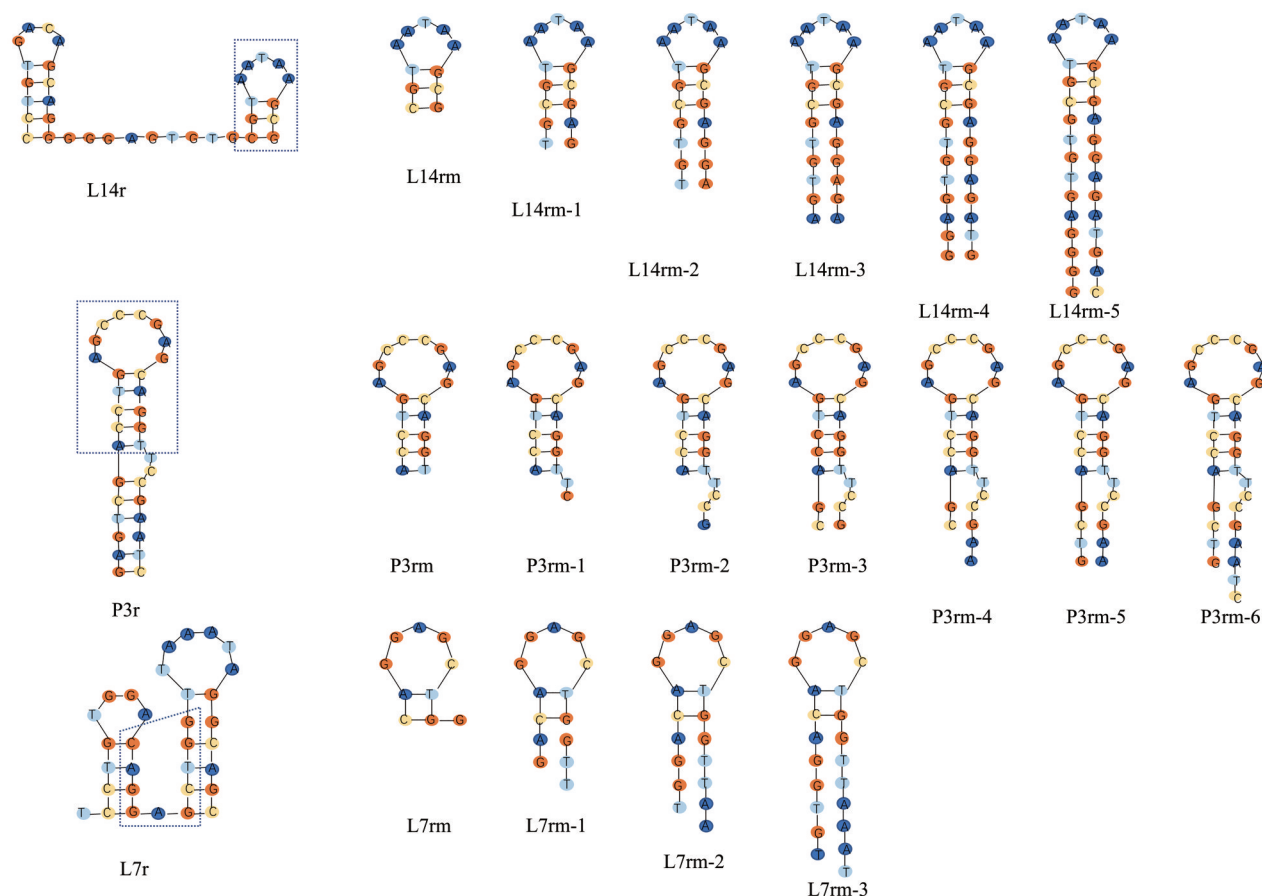


Fig. 3 The Mfold results of elongated sequences of L14r, P3r, and L7r

The dashed box represents the MABM. Color annotation, orange: G; yellow: C; blue: A, light blue: T.

图4给出了3条MABM及其延长序列群的分子对接结果, 总结归纳了序列群的每条寡核苷酸与

RCA60活性口袋中14个关键氨基酸的最短距离, 以此来推断两者结合的紧密程度。L14rm仅保留了

L14r 中 3' 端的茎部, 分子对接预测 L14rm (DS -232.1, RMSD 94.5 Å) 与 10 个关键氨基酸相连 (R48、N78、Y80、D96、D100、Y123、E177、R180、E208、W211), 存在 18 个距离小于 5 Å 的预测结合位点。以步进 $n=2$ 构建其延长序列群, 共含 5 条序列。P3rm 是含有 5 个碱基对的独立发卡结构, DS 值为 -235.0, RMSD 值为 109.9 Å, 与 10 个关键氨基酸相连 (R48、D75、N78、Y80、D100、Y123、R134、E177、R180、E208)。其延长序列群则构建了 6 条序列。

L7r 预测结合位点位于 2 个茎末端之间 (GGAGCT), 将两侧延长两个碱基, 形成了含有 2 个碱基对的发卡结构, 即将其作为 L7r 的 MABM (L7rm)。L7rm (DS -272.5, RMSD 91.1 Å) 分子

对接预测其与 13 个关键氨基酸相连 (R48、D75、N78、Y80、D96、D100、G121、Y123、R134、E177、R180、E208、W211), 存在 20 个距离小于 5 Å 的预测结合位点。其延长序列群则构建了 3 条序列。

经 SPR 测定, L14rm、L7rm 的 K_D 值为 (64 ± 30)、174 nmol/L, 说明两者与 RCA60 具有很好的结合, 验证了分子对接中预测的结合位点的正确性, 但 P3rm 的结构过于柔性, 未测得有效的 K_D 值, 此时将其延长序列群重新逼近原 P3r 序列, 应可有助于恢复一定的结合力。因此, 在确保 MABM 能嵌合进入 RCA60 活性口袋的前提下, 构建两端延长序列群的策略是较为有效、简捷的。

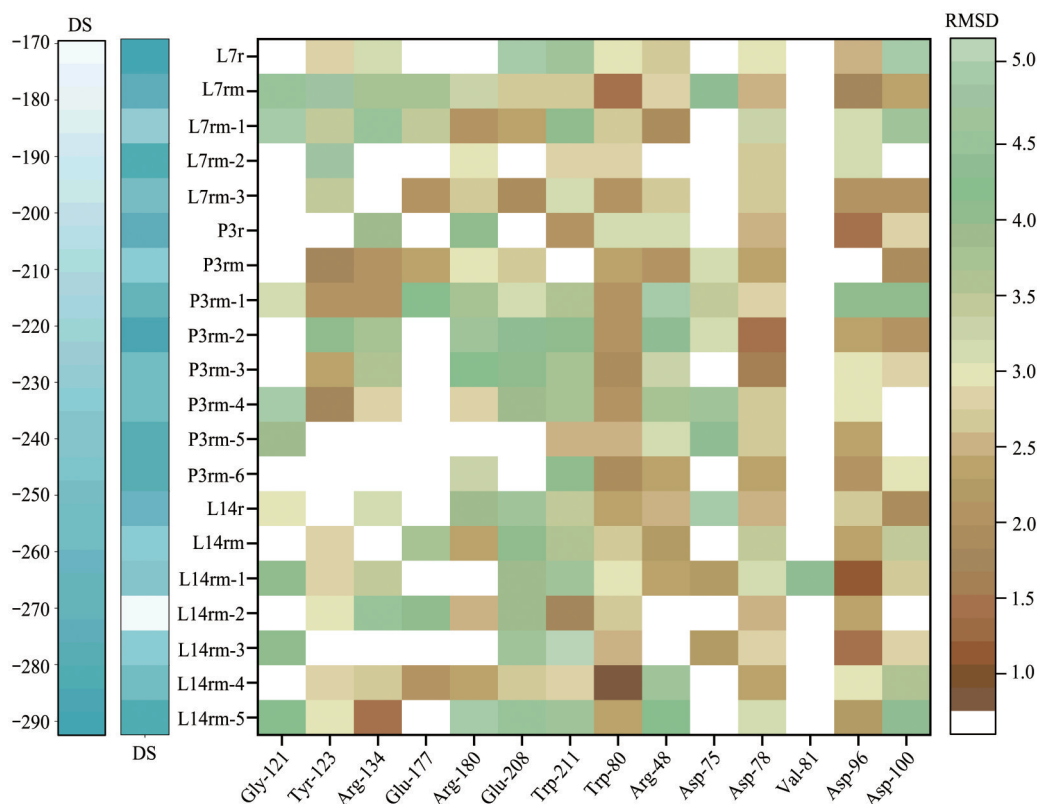


Fig. 4 The H-DOCK results of elongated sequences of L14rm, P3rm and L7rm

Two kinds of color annotation are shown. Color annotation for DS is from light blue to sky blue according to the decreased value; for minimum distances of different aptamer from 14 key amino acid residues in the primary and secondary pockets of RCA60 is from light green to dark brown, where white represents no interaction.

两端延长序列群中, L14rm-1~L14rm-5 两端分别延长了 2 对碱基, L14rm-1~L14rm-3 DS 值逐渐升高, 连接关键氨基酸数目整体减少, 提示此时未形成稳定的茎部结构; L14rm-4 和 L14rm-5, 分子

对接结果为 L14rm-4 (DS -258.1, RMSD 94.4 Å)、L14rm-5 (DS -281.4, RMSD 133.8 Å), 关键性氨基酸连接数目逐步增多 (图 4), 对 DS 值较低的 L14rm-5 (DS -281.4, RMSD 133.8 Å) 进行 SPR

测定, K_D 值为 (158 ± 46) nmol/L (表1, 图5)。在 H-DOCK 结果中, L14rm-4 的 5'端 A3、G4、T5、G6、T7、G8、C9 与 RCA60 的 R125、E127、Q128、L133、I205、T206、N209、S210、R213、S228、P229、I230、Q231 形成相互作用、L14rm-5 的 5'端的 G1、G2、G3、G4、A5、G6、T7、G8 与 RCA60 的 N88、N97、A101、D110、V111、Q112、N113、R114、Y115、T116、F117 形成相互作用, 提示此时延长的核苷酸链起到一定的固定支撑的作用, 辅助两者结合, 但最优适配体仍然为 L14rm。

P3rm 不与 RCA60 进行结合, 将之逐步延长两端直到构成 P3r 结构, 其中 P3rm-2、P3rm-6 存在较低的 DS 值, K_D 值分别为 (375 ± 175) 、 (629 ± 126) nmol/L (表1, 图5), 进一步延长, 两侧核苷酸链形成第二个茎后恢复至 P3r 的结构, 产生更

优的结合能力, 提示 P3r 中第二个茎为辅助结合的关键性结构, 最终选择 P3r 为 P3 的最优适配体。

L7rm-1 (12 个氨基酸, 20 个结合位点)、L7rm-2 (6 个氨基酸, 7 个结合位点)、L7rm-3 (10 个氨基酸, 14 结合位点) 连接氨基酸与距离小于 5 Å 的预测结合位点皆有所减少, 提示结构的延长阻碍了茎环结构进入蛋白质的活性口袋。而对接结果 L7rm-1 (DS -227.1, RMSD 95.0 Å)、L7rm-2 (DS -283.8, RMSD 95.2 Å)、L7rm-3 (DS -249.4, RMSD 112.4 Å) 显示 DS 值先下降再升高, L7rm-2 与 RCA60 结合效果更佳。SPR 测定结果表明, L7rm-1、L7rm-3 不具有结合能力, 而 L7rm-2 的 K_D 值为 (120 ± 1) nmol/L (表1, 图5), 为 L7 的最优适配体。

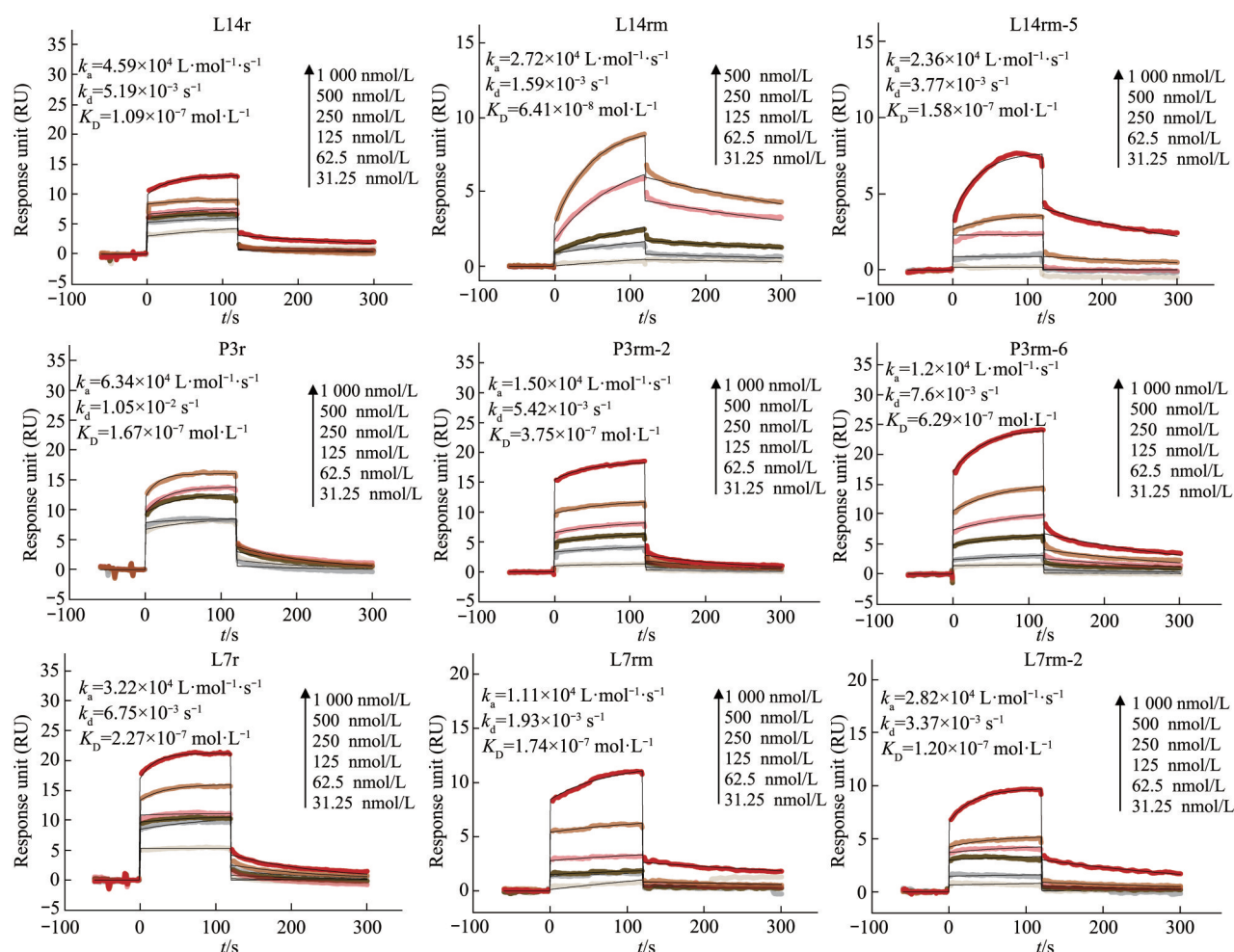


Fig. 5 SPR multiple cycle kinetic measurement results of L14r/P3r/L7r, L14rm/P3rm/L7rm and thereafter elongated sequences with RCA60

Table 1 The affinity and docking results of all truncated aptamers with RCA60						
	$k_a/(\text{L}\cdot\text{mol}^{-1}\cdot\text{s}^{-1})$	$k_d/(\text{s}^{-1})$	$K_D/(\text{nmol}\cdot\text{L}^{-1})$	DS	RMSD/Å	$\Delta G/(\text{J}\cdot\text{mol}^{-1})$
L14	$(1.1\pm0.4)\times10^4$	$(9.3\pm2.8)\times10^{-3}$	900±80	—	—	−7.3
L14r	$(4.6\pm1.8)\times10^4$	$(5.1\pm2.9)\times10^{-3}$	109±20	−262.8	92.5	−2.5
L14rm	$(2.7\pm1.2)\times10^4$	$(1.6\pm0.5)\times10^{-3}$	64±30	−232.1	94.5	0.3
L14rm-5	$(2.4\pm0.3)\times10^4$	$(3.8\pm1.4)\times10^{-3}$	158±46	−281.4	133.8	−1.3
P3	$(3.6\pm1.0)\times10^4$	$(5.5\pm1.9)\times10^{-3}$	151±26	—	—	−7.5
P3r	$(6.3\pm1.4)\times10^4$	$(1.0\pm0.0)\times10^{-2}$	167±19	−250.1	82.0	−7.5
P3rm-2	$(1.5\pm0.6)\times10^4$	$(5.4\pm0.9)\times10^{-3}$	375±175	−279.1	98.8	−2.8
P3rm-6	$(1.2\pm0.1)\times10^4$	$(7.6\pm1.1)\times10^{-3}$	630±126	−285.8	95.2	−2.3
L7	$(1.6\pm0.2)\times10^4$	$(8.5\pm1.2)\times10^{-3}$	526±40	—	—	−12.4
L7r	$(3.2\pm1.7)\times10^4$	$(6.7\pm1.7)\times10^{-3}$	227±68	−292.4	99.4	−5.3
L7rm	1.1×10^4	1.9×10^{-3}	174	−272.5	91.1	0.6
L7rm-2	$(2.8\pm0.1)\times10^4$	$(3.4\pm0.1)\times10^{-3}$	120±1	−283.8	95.2	−0.32

3 结 论

针对目前适配体剪裁仍较依赖于经验驱动的现状, 本文针对蓖麻毒素的适配体, 以分子对接模拟为指导、灵活引入步进式序列群策略, 驱动实验结果评估验证, 避免了人为判断因素, 方便、快捷地获得了最优适配体序列。所采用的分子对接的结果参数(DS值、RMSD值)以及对接模型的模拟预测具有较高的可信度, 有效预测了各序列的结合可能性。所选取的最短适配体结合单元, 在两端延长步进序列群的辅助下, 仅使用了17条适配体, 就得到了亲和力更强、序列最短的最优适配体。L14rm、P3r、L7rm-2的 K_D 值分别为(64±30)、(167±19)、(120±1) nmol/L。最优适配体与RCA60的亲和力提高到全长适配体的14、1、4倍, 也说明了本套基于分子对接和步进序列群的蓖麻毒素适配体序列优化策略的独特价值。预期将在适配体的结构剪裁方面有所裨益, 也为蓖麻毒素的最优适配体在传感和抑制剂等方面的应用奠定基础。

参 考 文 献

[1] Szewczak A A, Moore P B, Chang Y L, *et al.* The conformation of the sarcin/ricin loop from 28S ribosomal RNA. *Proc Natl Acad Sci USA*, 1993, **90**(20): 9581-9585

[2] 梁龙辉, 夏俊美, 刘昌财, 等. 剧毒性II型核糖体失活蛋白蓖麻毒素和相思子毒素的检测鉴定方法研究进展. *色谱*, 2021, **39**(3): 260-270

Liang L H, Xia J M, Liu C C, *et al.* *Chines Journal of Chromatography*, 2021, **39**(3): 260-270

[3] Audi J, Belson M, Patel M, *et al.* Ricin poisoning: a comprehensive review. *JAMA*, 2005, **294**(18): 2342-2351

[4] 王巍, 贾凌云. 适配体筛选方法研究进展. *分析化学*, 2009,

37(3): 454-460

Wang W, Jia L Y. *Chin J Anal Chem*, 2009, **37**(3): 454-460

[5] Fan S, Wu F, Martiniuk F, *et al.* Protective effects of anti-ricin A-chain RNA aptamer against ricin toxicity. *World J Gastroenterol*, 2008, **14**(41): 6360-6365

[6] Lamont E A, He L, Warriner K, *et al.* A single DNA aptamer functions as a biosensor for ricin. *Analyst*, 2011, **136**(19): 3884-3895

[7] Tang J J, Xie J W, Shao N S, *et al.* The DNA aptamers that specifically recognize ricin toxin are selected by two *in vitro* selection methods. *Electrophoresis*, 2006, **27**(7): 1303-1311

[8] 沈晓书. CE-SELEX技术筛选特异性识别蓖麻毒素的寡核苷酸适配子[D]. 北京: 军事医学科学院, 2006

Shen X S. *In vitro* Selection of High-specificity DNA Aptamers for Ricin by CE-SELEX[D]. Beijing: Academy of Military Medical Sciences, 2006

[9] He X Q, Guo L, He J L, *et al.* Stepping library-based post-SELEX strategy approaching to the minimized aptamer in SPR. *Anal Chem*, 2017, **89**(12): 6559-6566

[10] Gao S X, Zheng X, Jiao B H, *et al.* Post-SELEX optimization of aptamers. *Anal Bioanal Chem*, 2016, **408**(17): 4567-4573

[11] Huang P J, Liu J W. A DNA aptamer for theophylline with ultrahigh selectivity reminiscent of the classic RNA aptamer. *ACS Chem Biol*, 2022, **17**(8): 2121-2129

[12] Shangguan D H, Tang Z W, Mallikaratchy P, *et al.* Optimization and modifications of aptamers selected from live cancer cell lines. *Chembiochem*, 2007, **8**(6): 603-606

[13] 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

[14] Lapa S A, Chudinov A V, Timofeev E N. The toolbox for modified aptamers. *Mol Biotechnol*, 2016, **58**(2): 79-92

[15] Söderström M, Bossée A, Dorner B G, *et al.* Recommended Operating Procedures for Analysis in The Verification of Chemical Disarmament. Finland: University of Helsinki, 2017: 547-548

[16] 范华杰, 李闻捷. 蛋白质定量方法的合理应用. *中华国际医药*

- 杂志, 2003, **2**(2): 3
- Fan H J, Li W J. Chin Int J Med, 2003, **2**(2): 3
- [17] Majka J, Speck C. Analysis of protein-DNA interactions using surface plasmon resonance. Adv Biochem Eng Biotechnol, 2007, **104**: 13-36
- [18] Dong H N, Park E J, Kim M S. Characterization of two ricin isoforms by sodium dodecyl sulfate-capillary gel electrophoresis and capillary isoelectric focusing. Bull Korean Chem Soc, 2011, **32**(12): 4253-4257
- [19] Yang Z F, Wang C, Xiao L, *et al.* SCX-tip-aided LC-MS detection of active ricin *via* oligonucleotide substrates for depurination kinetics. Analyst, 2023, **148**(12): 2782-2792
- [20] Sun D, Sun M, Zhang J, *et al.* Computational tools for aptamer identification and optimization. Trends Analyt Chem, 2022, **157**: 116767
- [21] Shrivastava G, Hyodo M, Ara M N, *et al.* The screening of RNA aptamers specific for carbonic anhydrase I using the systematic evolution of ligands by an exponential enrichment method (SELEX). Nucleosides Nucleotides Nucleic Acids, 2014, **33**(11): 697-708
- [22] Luo L, Yang J W, Li Z, *et al.* Label-free differentiation and quantification of ricin, abrin from their agglutinin biotoxins by surface plasmon resonance. Talanta, 2022, **238**(Pt 1): 122860
- [23] Gao S X, Zheng X, Hu B, *et al.* Enzyme-linked, aptamer-based, competitive biolayer interferometry biosensor for palytoxin. Biosens Bioelectron, 2017, **89**(Pt 2): 952-958
- [24] Su H X, Xu Y C. Application of ITC-Based characterization of thermodynamic and kinetic association of ligands with proteins in drug design. Front Pharmacol, 2018, **9**: 1133
- [25] Magnez R, Bailly C, Thuru X. Microscale thermophoresis as a tool to study protein interactions and their implication in human diseases. Int J Mol Sci, 2022, **23**(14): 7672
- [26] Bogutzki A, Curth U. Analytical ultracentrifugation for analysis of protein-nucleic acid interactions. Methods Mol Biol, 2021, **2263**: 397-421
- [27] Yan J H, Xiong H J, Cai S D, *et al.* Advances in aptamer screening technologies. Talanta, 2019, **200**: 124-144
- [28] Goto M, Higashi S, Ohba T, *et al.* Conformational change in ricin toxin A-chain: a critical factor for inhibitor binding to the secondary pocket. Biochem Biophys Res Commun, 2022, **627**: 1-4
- [29] van den Berg R M, Joosen M J A, Savransky V, *et al.* Inactivation of ricin by constituents present in a skin decontamination lotion. Chem Biol Interact, 2022, **365**: 110055
- [30] Ho M C, Sturm M B, Almo S C, *et al.* Transition state analogues in structures of ricin and saporin ribosome-inactivating proteins. Proc Natl Acad Sci USA, 2009, **106**(48): 20276-20281
- [31] Saito R, Goto M, Katakura S, *et al.* Pterin-based small molecule inhibitor capable of binding to the secondary pocket in the active site of ricin-toxin A chain. PLoS One, 2022, **17**(12): e0277770

Molecular Docking and Stepping Sequence Cluster Design Prompted Sequence Optimization of Ricin Nucleic Acid Aptamers*

LIU Jia^{1,2)}, YANG Zhi-Fang²⁾, WANG Chuang^{2,3)}, XIAO Lan^{2,3)},
GUO Lei²⁾**, WU Hai-Xia¹⁾**, XIE Jian-Wei²⁾

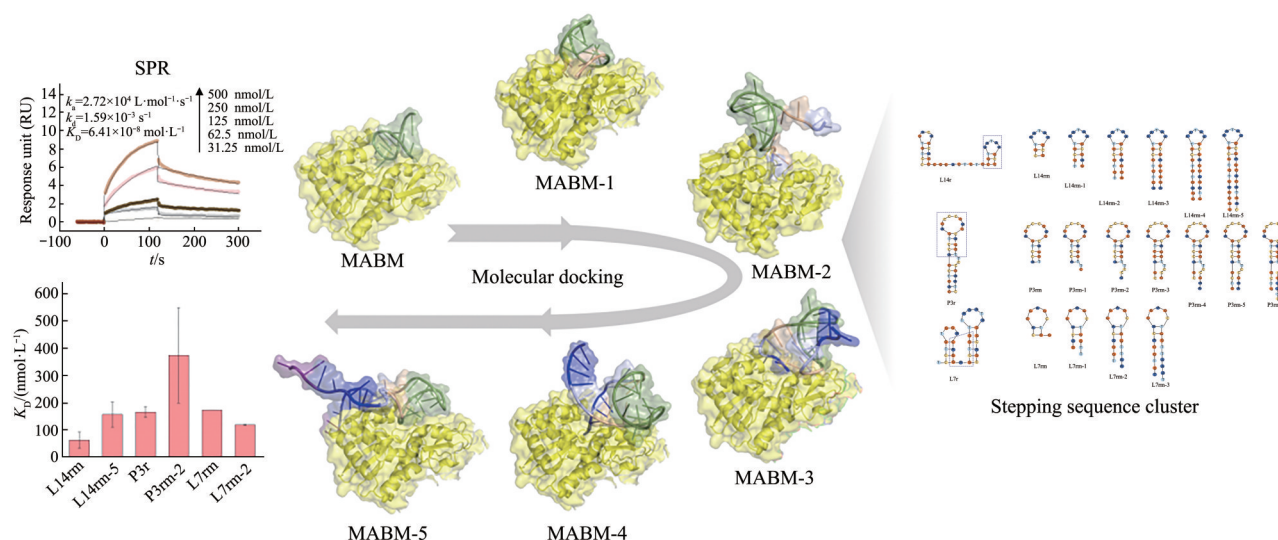
⁽¹⁾School of Chemical and Pharmaceutical Engineering, Hebei University of Science and Technology, Shijiazhuang 050091, China;

⁽²⁾State Key Laboratory of Toxicology and Countermeasures, Institute of Toxicology and Pharmacology,

Academy of Military Medical Sciences, Beijing 100850, China;

⁽³⁾School of Pharmacy, Minzu University of China, Key Laboratory of Ethnic Medicine, Ministry of Education, Beijing 100081, China)

Graphical abstract



Abstract Objective To obtain the finest aptamers with stronger affinity and shortest sequences from the combination of molecular docking simulation and surface plasmon resonance (SPR) evaluation experimentation. **Methods** Towards three previously screened single-stranded DNA aptamers against ricin with 80 nt, L14, P3, L7, on the confirmation basis of binding ability between their 2D stem-loop sequences of random region with target proteins, the H-DOCK molecular docking was performed to guide the determination of their minimum aptamer binding motifs (MABM), and an elongated stepping sequence cluster was sequentially constructed. The affinity and kinetic parameters of the designed clusters were measured by SPR, in which the binding key

* This work was supported by a grant from The National Natural Science Foundation of China (21974152).

** Corresponding author.

GUO Lei. Tel: 86-10-68225893, E-mail: guolei@bmi.ac.cn

WU Hai-Xia. Tel: 86-311-81668373, E-mail: wuhaixia@hebust.edu.cn

Received: July 17, 2023 Accepted: August 6, 2023

structures of the selected aptamers were depicted, and finally the finest aptamers could be selected. **Results** All three random region aptamers, L14r, P3r and L7r, can form some certain hairpin structures, and the affinity of L14r is increased by nine times than L14, L7r increased by 2 times, and P3r kept the same. The results of molecular docking between the random region aptamers and ricin showed that, the docking scores of L14r, P3r and L7r were all lower than negative sequence of 40T, the number of key binding amino acids were 11, 8 and 9, and the predicted binding sites with distances less than 5 Å were 20, 12 and 15, respectively, indicating good binding ability with ricin. Further, the sequence composition of MABM, L14rm, P3rm and L7rm, were deduced from the binding structures confined in the ricin active pockets, and the elongated stepping sequence cluster was built. On the parameters including the number of key binding amino acids, binding sites, the docking scores, as well as the results of SPR evaluation, the finest aptamers were evolved, in which the affinity of L14rm and L7rm-2 continued to be increased by 1–2 times. **Conclusion** The random region aptamer can effectively bind ricin with stronger affinity than the full-length aptamer. Molecular docking and stepping sequence clusters design can aid the fast evolution of three finest aptamers from only 17 sequences as well as the investigation of binding interaction. The K_D values of the three finest aptamers against ricin, L14rm, P3r and L7rm-2, were (64 ± 30) , (167 ± 19) and (120 ± 1) nmol/L, respectively, and the affinity was increased to 14, 1 and 4 times of the full-length aptamers.

Key words ricin, aptamer, molecular docking, stepping sequence cluster, surface plasmon resonance, structural optimization

DOI: 10.16476/j.pibb.2023.0280