·综 述·

计算机辅助筛选核酸适配体技术

邓博文，高思懿，肖博懿，吴雨龙，孙豪，王梁华，孙铭娟

海军军医大学 基础医学院 生物化学与分子生物学教研室，上海 200433

邓博文, 高思懿, 肖博懿, 吴雨龙, 孙豪, 王梁华, 孙铭娟. 计算机辅助筛选核酸适配体技术. 生物工程学报, 2022, 38(2): 678-690.
DENG BW, GAO SY, XIAO BY, WU YL, SUN H, WANG LH, SUN MJ. Computer-aided aptamers screening technologies: a review. Chin J Biotech, 2022, 38(2): 678-690.

摘 要：计算机信息技术已经渗透到我们生活的方方面面，不仅可以辅助药物的筛选，也能够模拟药物的作用。目前已有研究使用计算机辅助技术筛选适配体，对筛选效率的提升和筛选高亲和力的适配体有着重要的指导作用。文中将主要对计算机通过序列评估、结构分析、分子对接 3 个方面辅助适配体筛选的方法进行综述。

关键词：适配体；指数富集的配体系统进化技术；计算机辅助筛选；序列评估；结构分析；分子对接

Computer-aided aptamers screening technologies: a review

DENG Bowen, GAO Siyi, XIAO Boyi, WU Yulong, SUN Hao, WANG Lianghua, SUN Mingjuan

Department of Biochemistry and Molecular Biology, School of Basic Medical Science, Naval Medical University, Shanghai 200433, China

Abstract: The computer information technology that has penetrated into every aspect of our lives, can not only assist the screening of drugs, but also simulate the effect of drugs. At present, computer-aided technologies have been used to screen aptamers, which play an important role in improving the screening efficiency and screening high affinity binding aptamers. This review summarized the

Received: May 8, 2021; **Accepted:** July 28, 2021; **Published online:** August 11, 2021
Supported by: National Key Research and Development Program of China (2019YFC0312603); Innovation and Practice Ability Incubator for Undergraduate Students of Naval Medical University, China (PH2019019); Key Construction Project of Naval Curriculum, China (2019)
Corresponding author: SUN Mingjuan. Tel: +86-21-81870972; E-mail: sunmj@smmu.edu.cn
基金项目：国家重点研发计划 (2019YFC0312603)；海军军医大学本科学员创新实践能力孵化基地项目 (PH2019019)；海军课程重点建设项目 (2019)

screening methods of aptamers through computer-aided sequence evaluation, structural analysis and molecular docking.

Keywords: aptamer; exponentially enriched ligand system evolution; computer-aided screening; sequence evaluation; structural analysis; molecular docking

核酸适配体是指一种能够与特定目标物高亲和力和特异性结合的寡核苷酸片段，通过配体指数富集的配体系统进化技术 (exponentially enriched ligand system evolution, SELEX) 从随机文库中筛选获得的功能性核苷酸，大小一般约为 6–40 bp，具有一些特殊的二级结构，如发夹、茎环、假结、G-四联体等，能让适配体可以形成多种三级结构，使其能够与靶物质结合。图 1 展示了这些特殊结构的二级结构和三级结构。1990 年 Tuerk 等[1]就首次通过 SELEX 从 65 536 个序列的预测池中筛选出了两个不同的序列，它们可以与 T4 DNA 聚合酶相互作用。经过数十年的改进，SELEX 技术已经成为一种重要的研究方法和工具，其基本原理是体外化学合成一个单链寡核苷酸库，包括大量随机单链 DNA 或 RNA (10^{16}–10^{18} 个单链)，用它与靶物质 (包括金属离子、小分子、大分子、细胞、组织以及混合靶标等) 混合，同时洗脱混合液中未与靶物质结合的核酸，保留与靶物质结合的核酸分子，随后分离与靶物质结合的核酸分子，再以此为模板进行 PCR 扩增，进行下一轮筛选。通过重复筛选，一些与靶物质不

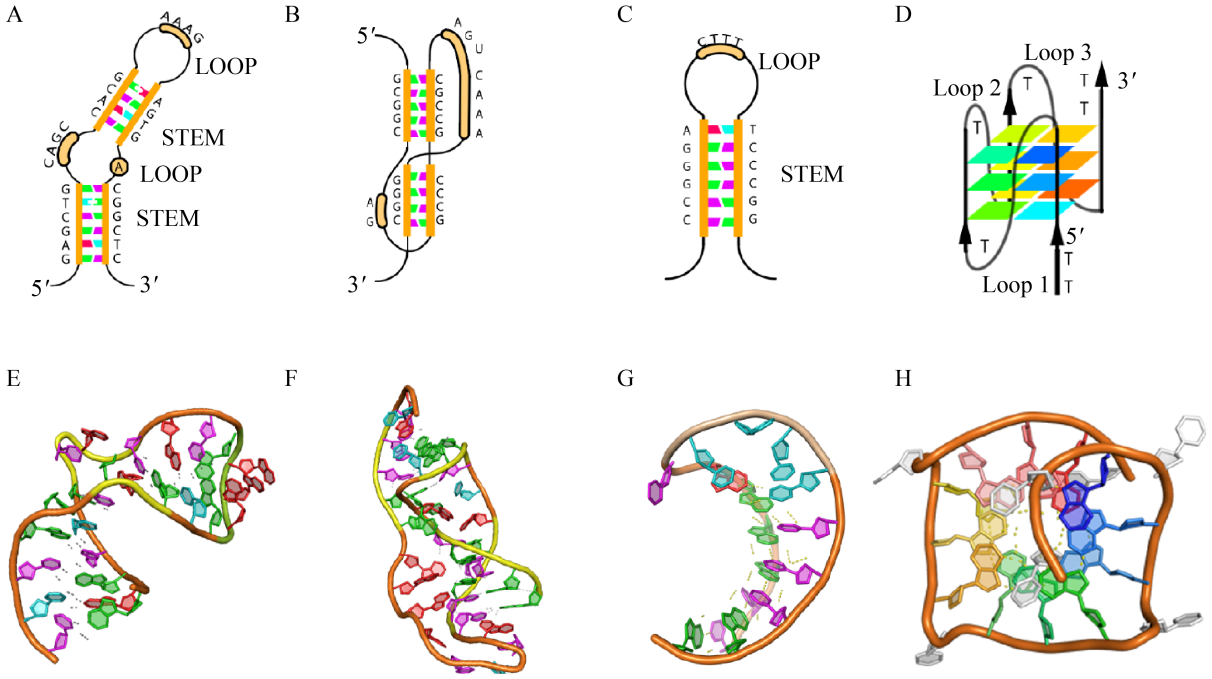

图 1　茎环、假结、发夹、G-四联体的二级结构和三级结构

Figure 1　Secondary and tertiary structures of stem-loop, hairpin, pseudoknot and G-quadruplex. (A–D) The secondary structures of stem-loop, pseudoknot, hairpin, and G-quadruplex. (E–H) The tertiary structures of stem-loop (PDB ID: 2L5Z), pseudoknot (PDB ID: 1RNK), hairpin (PDB ID: 1JWC), and G-quadruplex (PDB ID: 2M4P).

☏：010-64807509　　✉：cjb@im.ac.cn

结合的、亲和力不理想的适配体就会被洗脱，而具有高亲和力的适配体就能从 10^{16} 到 10^{18} 的随机文库中分离出来，且理论上该序列的纯度随着 SELEX 的不断筛选将会越来越高，最后库中序列大部分都是高亲和力序列[2]，图 2 为 SELEX 筛选的主要流程。这些高亲和力序列的结合具有高特异性，且可选的靶分子广泛，已经在基础研究、临床诊断和治疗中显示出了广阔的应用前景。核酸适配体还有着相对简单的化学结构，使其易于体外合成和修饰[3]，允许在寡核苷酸上的特定位置插入电化学或荧光报告分子以及表面结合剂。在适配体和靶物质结合过程中，核酸适配体的构象变化可以用来产生分析信号。许多以适配体为基础的生物传感器已经成功用于测量蛋白质的细胞分泌，但实际筛选过程却不易得到高亲和力的序列。在众多候选的序列中挑选出潜在的高亲和力序列，成为了许多适配体研究者的一大挑战。

目前通过 X 光晶体衍射[4]以及核磁共振[5]，科学家们已经分析了较多适配体的空间结构，并成功解析了许多适配体与靶分子间的作用机制，这些来之不易的研究成果为后续各种生物信息学工具奠定了基础。与此同时，随着计算机和信息技术高速发展，已有许多计算机辅助天然药物筛选的例子，同时也有越来越多的研究者发现，通过计算机辅助能提升适配体筛选的效率和提高得到高亲和力序列的概率，该技术可能对适配体未来的发展产生巨大的影响。本文将着重对计算机辅助适配体筛选相关方法以及它们的主要应用和相关结果进行分析，同时对它们的优缺点进行比较，并对近年来该领域的一些创新应用进行综述。

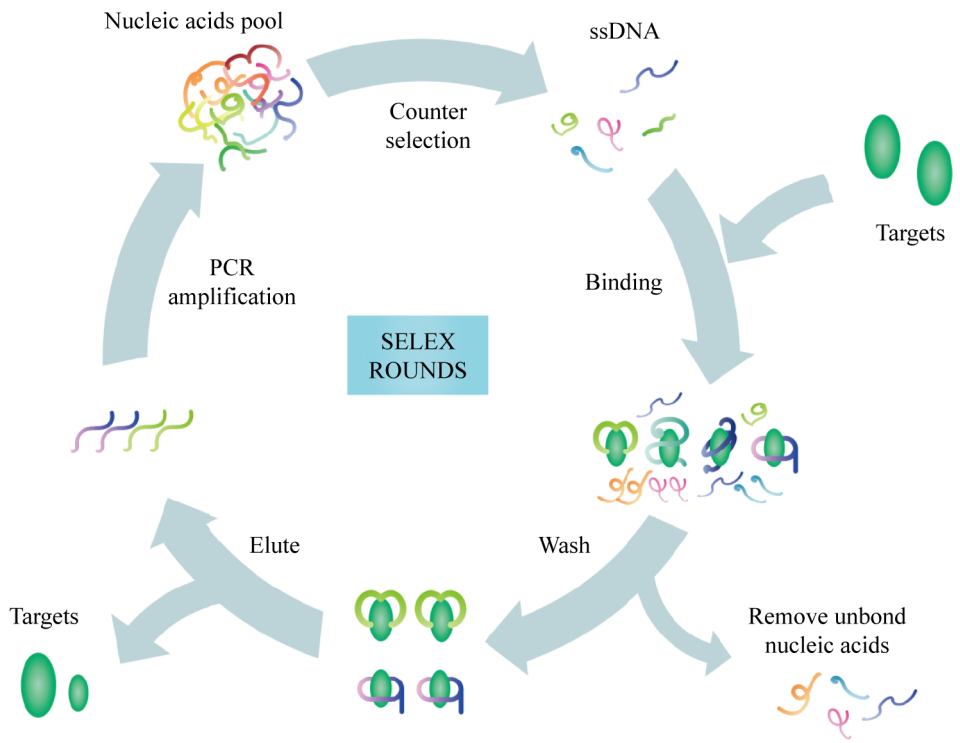

图 2　SELEX 的主要流程

Figure 2　The main process of SELEX.

1 SELEX 技术目前存在的缺陷和发展

1.1 SELEX 技术目前存在的缺陷

1.1.1 需要极大的初始文库容量

核酸适配体筛选初始文库的随机序列一般有 30–60 个碱基。因此理论上初始文库容量为 10^{18}–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[6]。

1.1.2 筛选流程较为繁琐

整个筛选流程耗时长、精力投入大，且无法确定具体筛选的次数，一般需要十几轮筛选才能够筛选出亲和力较好的序列，武汉大学基础医学院的章晓联团队总共进行了 17 轮筛选，才得到了鸟分枝杆菌的适配体[7]。

1.1.3 难以筛选出亲和力较高的序列

正如 Kanagawa 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，但由于 PCR 具有偏好性等原因[8]，导致一些富含 GC 碱基的序列更容易被扩增, 使原本丰度不够大的文库结构变得更加单一，导致某些高亲和力的序列在库中的含量越来越低，造成筛选次数增加，亲和力反而下降的现象。

1.1.4 后期处理工作十分复杂且易错过高亲和力序列

在测序、分析、筛选出合适的序列后，还需要再将其合成进行重复验证。如果要鉴定更多序列，不仅成本高，还需要投入大量时间。所以考虑到成本问题，大多数情况下还只能合成极少量的序列，因此很容易错过真正有亲和力的序列。

1.2 SELEX 技术的发展

针对以上几个问题，2004 年 Mendonsa 等提出了毛细管电泳 SELEX 这项技术, 因其无需固定相和洗脱的缘故，大大减少了筛选的次数[9]。随后也诞生了无需对靶物质进行详尽分析并且能够发现更多潜在靶点的细胞 SELEX 技术；还有无需 Sanger 测序分析，能在所有筛选轮中对库进行排序的高通量 SELEX (high throughput systematic evolution of ligands by exponential enrichment, HT-SELEX)[10]。HT-SELEX 不仅减少了筛选次数，有效地降低了 PCR 偏好性带来的误差，还能完全自动化操作，提高了筛选效率，甚至有望取代常规的 SELEX。该技术如此快速的发展，离不开计算机的辅助：通过强大的生物信息学工具，对大量数据进行全面的分析，并对结合能力以及结构进行预测，再量化丰度，模拟适配体与靶物质的相互作用，进一步综合分析得到的适配体，尽可能地保留有价值的序列。

2 计算机辅助筛选适配体的方法与工具

目前越来越多的生物信息学工具开始在生物学研究上崭露头角，它们推动了核酸适配体的研究，使得研究者们能够在进行生物活性筛选之前，利用计算机对所需的环境进行模拟，得到靶分子和配体在该环境下最有可能的结构，再利用分子对接软件模拟目标靶点与候选分子之间的相互作用，计算其结合能并评估结合程度，以此来筛选出高亲和力的适配体。且经过不断地研究和改良，研究者们已经可以通过设计算法程序对 HT-SELEX 产生的大量序列进行打分，评价其富集的程度以及结合的潜力，再折叠出其中的高分序列最有可能形成的二级结构，利用多种软件预测该高分序列的三级结构，用于之后的对接验证。下面将主要介绍如何通过计算机评估序列、分析适配体结构和分子对接的方法辅助适配体的筛选。

2.1 通过评估序列辅助筛选

核酸适配体能够识别特异的靶物质，很大

程度上是因为其独特的三维结构。序列本身的碱基排列顺序就能够对后续的三级结构的折叠产生巨大影响，可能仅仅几个碱基的不同，适配体三级结构就会有很大的差异，导致亲和力会有很大的不同。高亲和力适配体的碱基排序必然具有一定的特殊性。为了更好地了解某些特殊结构与序列的关系，科学家们也开始把核酸适配体相关的常见模体 (例如 G-四联体[11]、发夹、茎环)收录进数据库中，方便后续利用计算机分析这些特异的结构与序列之间存在的规律。例如，2000 年 Kensaku Sakamoto 团队[12]开发了针对发夹的算法，随后 Gorodkin 设计了能够分析 RNA 序列中的茎环结构的 Stem-Loop Align SearcH[13]。此后，由于发现四联体在重要的生理过程中不可或缺并且在作为治疗靶标的应用也有着非常重要的作用，使得它受到了广泛的关注。为了进一步研究四联体的序列特点，2006 年 Kikin 等[14]用 PHP 编写了基于 Web 的程序 QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/)，该 web 可用于预测核苷酸序列中形成四联体的富含鸟嘌呤的序列。后来也有不少团队利用 QGRS Mapper 来研究具有四联体结构的核酸适配体。例如，为了获得血管内皮生长因子的高亲和力适配体，Kazunori Ikebukuro[15]团队利用 QGRS Mapper 分析了它们筛选出适配体 Vap7 中可能存在的四联体结构，然后利用圆二色谱法验证了该序列中稳定的四联体结构，最后通过优化四联体的结构成功得到了亲和力更高的适配体。Zheng 等[16]也通过 QGRS Mapper 分析了石房蛤毒素的适配体，发现该适配体也存在稳定的四联体结构，筛选出了四联体结构最为稳定的适配体，之后通过保留四联体相关的序列，截短多余的序列，成功得到了 STX 的高亲和力适配体。它们虽然较好地分析出了序列中某种特殊模体的稳定性和可能性，能够在序列数量不多的情况下筛选出分数最高即最能够稳定形成上述特定结构的序列，但是它们只能针对一种特定的特殊模体进行分析，不能够分析序列中存在多种特殊模体的可能性。这就需要我们对适配体与靶物质的作用模体有一定的了解，才能够保证高效无误地筛选出具有亲和力的适配体。且这类早期的算法只能输入较少的序列，因此很难应用到大量数据的处理上。所以，为了能够分析类似于 HT-SELEX 产生的大量且无序的数据，就需要设计新的算法对这些序列进行多次的聚类和分类[17]。

为了帮助这类 SELEX 技术从海量的序列中高效地筛选出具有潜力的序列，2014 年 Ron Stewart 团队开发了一种全新的软件 MPBind[18]，并将这类新的统计框架用于预测筛选出的适配体的结合潜力。这种方法整合了多个信息量适中的数据源来进行高置信度的预测，能够计算序列的频率和富集率，同时也提出了一个新的思路即适配体的结合潜能可以分解为序列中所有聚体结合潜能的组合，再通过 MPBind 评估所有可能的聚体指标，就可以推断出适配体的结合潜能。根据这些参数，对大量的序列进行打分排名，最后就能筛选出那些得分较高的序列。但它仅仅只是考虑了一些简单的统计学参数，没有在分析中考虑到功能模体或者是高级结构，结果部分高富集率的核酸适配体亲和力并没有达到预期效果，这也侧面证明了高富集的序列不一定是高亲和力的序列。而且随着 HT-SELEX 的进一步推广，研究者发现通过大量 HT-SELEX 得到的序列池中包含大量相似的序列，对这些序列进行聚类也是进一步分析的重要方法。于是诞生了诸如 FASTAptamer[19]、AptaCluster[20]这类能够为大量适配体聚类分析的软件，它们能够将数百万个序列分类组成适

配体家族，以此来缩小筛选的范围，提升了筛选的效率。尽管它们能够对大型的适配体文库非常有效地聚类，但是它们在该过程中仅仅只是通过序列相似度筛选出了适配体家族并得到了其特征的序列，并未考虑到适配体的结构构象。而 Hiller 等[21]发现虽然 RNA 结合蛋白[22]通常以序列特异性的方式与 RNA 结合，但更偏好于结合位点的特征性结构构象；作者发现，这种结构构象显示为发夹环序列、内环序列或两个茎环之间的单链序列；作者并且一旦将双链中结合基序隔离，就会解除与蛋白质的结合。因此，RNA 的二级结构特性对于区分真实的和虚假的蛋白质结合位点非常重要。然而由于仍不清楚其作用的结构模体，通过上述方法得到的适配体家族不能用于后续的结构和功能分析。

通过对序列的分析，能够从大量的序列中筛选出更具有潜力的适配体，减少了 SELEX 筛选的轮数，同时也减少了错过高亲和力序列的可能。但仅在序列的层面上进行分析还不足以高精度地筛选出高亲和力适配体。如果要更好地辅助筛选出高亲和力适配体，还需要对它们折叠形成的高级结构有更深入的了解，进一步分析出适配体与靶物质作用的机制。

2.2 通过分析适配体的结构辅助筛选

Wang 等[23]发现它们筛选得到的凝血酶抑制剂适配体都具有一种特殊的结构，这样的结构都是由富含鸟嘌呤的核酸序列所构成的，它们与凝血酶抑制剂适配体的高亲和力密切相关。这提示在 SELEX 分析中加入结构的分析，就有可能降低单纯分析序列带来的统计学偏好性，从而提升筛选的精度。2016 年 Przytycka 团队研发了 AptaTRACE[24]，专门用于鉴定从 HT-SELEX 大量的数据中得到的不同的结构模体，AptaTRACE 还通过专注于选择的局部模体，成功减小了由于 PCR 偏好性带来的偏倚，且与 MPBind 不同，AptaTRACE 是通过序列的二级结构而不是根据丰度来衡量并选择序列的结构模体，即使它们仅占序列的一小部分，它也可以发现所选择的具有统计学意义的模体，因此即便我们完全不了解实验中的适配体与靶物质的作用模体，也能够发现相关的功能模体并排除无关的结构模体，减少 SELEX 筛选的轮数，从而减少整个过程的成本。即使 AptaTRACE 将重心从单纯的序列分析转向了结构分析并解决了部分由统计学和 PCR 相关的偏倚带来的问题，但它无法提供候选序列的排名，给后续高潜力序列的选择带来了一定的困难。为了同时兼顾到结构和提供排名，Bicciato 团队研发了 APTANI[25]，它不仅可以通过分析 HT-SELEX 数据结构模体来选择高亲和潜力的适配体，还能够输出适配体的丰度，与库中特殊模体的比对分数、结构模体和共有的序列。随后 Bicciato 使用 APTANI 分析了 HT-SELEX 实验相对应的序列文库，成功地分离出了能够特异结合鼠 IL4Ra 的适配体 C1.42。然而，即便使用的效果是可观的，由于 APTANI 是通过计算考虑适配体二级结构的分数来对候选物进行排名并确定相关的结构图案，则得出的分数更多是考虑结构的稳定性而不是结构的相似性。为了同时兼顾到打分排名、结构、聚类分析这些要点，2020 年 Michiaki 团队[26]开发了 RaptRanker，它与仅使用一轮 HT-SELEX 输入的 APTANI 不同，RaptRanker 从所有的 HT-SELEX 轮次中确定唯一序列，并基于核苷酸序列和二级结构特征通过相似性将唯一序列的所有子序列聚类，然后通过计算平均基序富集度鉴定高结合亲和力适配体。图 3 展示了 RaptRanker 的分析过程。但遗憾的是，目前该算法只能考虑到少量常规

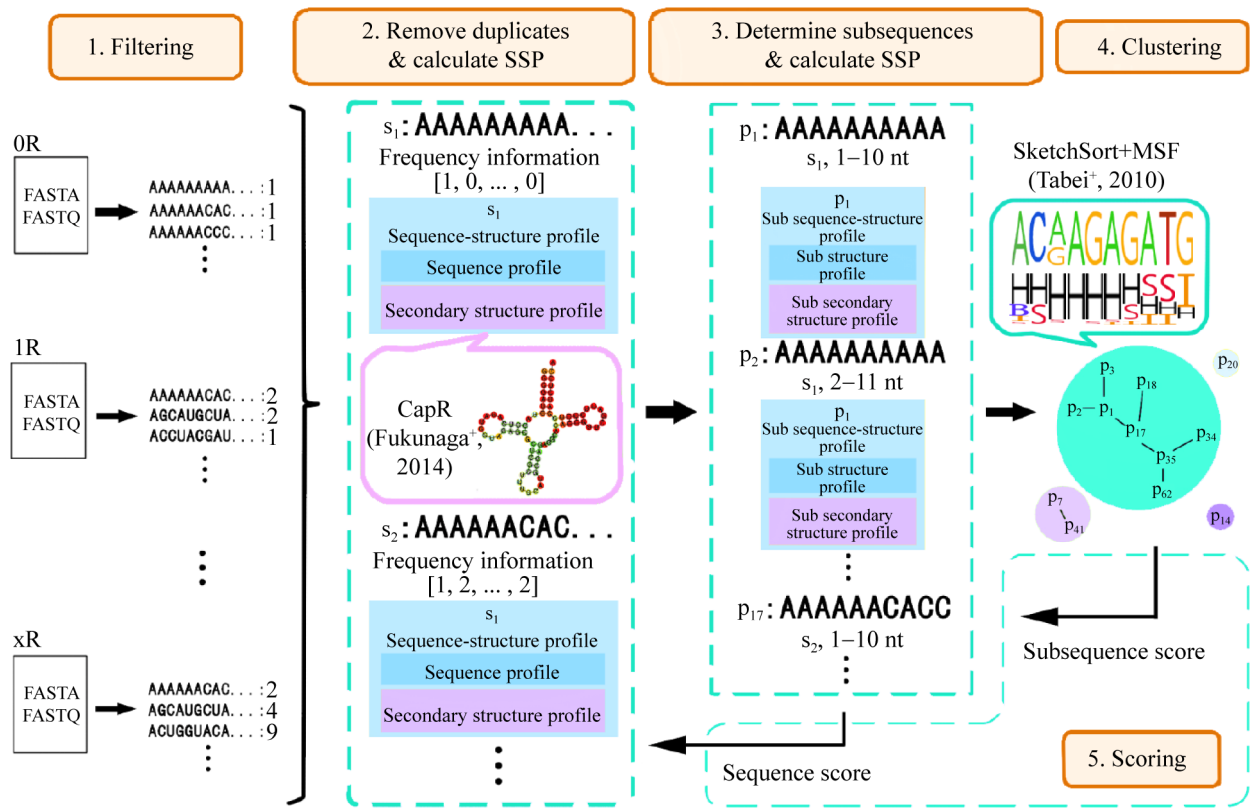

图 3　RaptRanker 的分析过程[26]

Figure 3　Analysis process of RaptRanker[26].

的二级结构，因此不能分析如四联体这类的模体，且暂时只能应用到 RNA 适配体的分析上。

通过对二级结构的进一步解析，将序列分析与结构分析相结合，有效地提高了计算机筛选出的适配体的精确度，降低了由于高富集率造成的假阳性的情况。然而，要想进一步提升该类方法筛选的精度，还需要对二级结构的预测算法有更高的要求，目前预测适配体二级结构的主流算法都是基于计算预测。Zhao 等认为尽管在过去的 40 年中提出了各种新的方法，但是在最近的 10 年中，计算预测方法的性能一直停滞不前，目前随着结构数据可用性的提高，二级结构预测正从传统的基于分数的方法成功地转向基于机器学习的方法，基于机器学习的技术将进一步提高预测的性能，而深度学习必将继续提高预测性能[27]。由于目前三级结构的预测精度不高，得到的结构与实际状况差距较大，往往需要研究者进行多步优化才能投入使用，所以暂时没有利用三级结构辅助筛选的相关报道。实际上适配体研究者更希望能够将筛选与三级结构结合起来，便于理解适配体的作用原理，为筛选出的高亲和力适配体提供更多的理论支持。Ishida 等也表示将适配体的三级结构应用在适配体筛选中有着巨大潜力，后续的工作打算将 RaptRanker 和三级结构结合起来，以望能够进一步提升其准确性[26]。

2.3 通过分子对接辅助筛选

与序列评估和结构分析不同，分子对接是一种更加直观的筛选策略，涉及使用基于计算机的模拟来对核酸-靶标相互作用进行建模，研究分子间 (如配体和受体) 相互作用并预测其结合模式和亲合力。目前大部分对接方法都致力于对适配体候选物的结合验证和适配体-靶标结合位点的确定上。通过测序经 SELEX 筛选得到的核酸库，然后将所得的各种序列类别进行计算机对接，以确定每个核酸序列和靶物质的结合强度。最后筛选出与靶物质结合得最稳定的核酸或脱氧核酸分子作为有效的适配体。虽然早期的时候这些方法大多是应用在分析蛋白质这类大分子的对接，但随着核酸适配体的逐步推广，适配体研究者也发现了如 Hex[28]、ZDOCK[29]、NPDOCK[30]、AutoDock[31] 和 AutoDock Vina[32]等几款比较适合于分析核酸适配体对接的软件。例如 Hu 等[33]就利用 Hex 研究 DNA 适配体与 PTX 的结合机制，并通过此方法证实并获得了比其他 SELEX 产生的 DNA 适配体有着更高亲和力的适配体 P-18S2。ZDOCK 是 Chen 和 Weng 于 2002 年开发的一款自动对接工具，它可以搜索 DNA 适体和靶分子之间的平移和旋转空间中所有可能的结合模式，能够对适配体目标之间的空间自由度进行三维网格搜索，它还通过考虑形状互补、静电和成对统计势来评估每个适体-靶相互作用，是一款较为全面的对接软件。Vu 等证明了 ZDOCK 服务器能够确定 DNA 适体血小板衍生生长因子受体的结合位点-β 复合物，由此发现 PDGF-B 的结合位点位于 DNA 适体的 loop1 和 loop3[34]。除此之外，Yarizadeh 等利用 ZDOCK 服务器模拟了 12 个已报道的 DNA 适体与癌胚抗原的复合物，筛选获得的 DNA 适配体 G3S1.5 被选为检测 CEA 生物标志物的最佳适体[35]。而 NPDOCK 与最初用于蛋白质对接的 Hex、ZDOCK 不同，它是一款专门用于核酸-蛋白质复合物结构分子对接的 web 服务器，使用特定的蛋白质-核酸统计潜力分析法来对形成的复合物评分，它的工作流程包括对接、对接结果评分和最佳评分模型的聚类。Kakoti 和 Goswami 使用 NPDOCK 成功得到了两个 DNA 适配体 N13 和 N53，它们是首次报道的针对人类 FABP3 的适配体[36]。适配体分子往往具有一定的柔性，它在对接的过程中往往会发生构象上的变化，因此只考虑普通的刚性对接可能会带来较大的误差，相比前面的几款软件，AutoDock 和 AutoDock Vina 加入了更多的柔性对接分析，能在确定的对接口袋中进行灵活对接[31]。它们被用于许多针对多种靶点的 DNA 适配体研究，例如 HIV-1 gp120 糖蛋白[37]、野生型表皮生长因子受体[38]、抗神经兴奋肽[39]，由于其优秀的对接性能，AutoDock 也被认为是适配体开发中最理想的算法。

分子对接很大程降低了适配体研究者分析适配体-靶标复合物的结合位点以及结合紧密程度的难度，无需将每条从 SELEX 中得到的序列都进行实验验证，通过分子对接能够从那些序列中筛选得到更具潜力的高亲和力序列，再用于后续的实验验证，大大节省了适配体筛选的成本和时间。但该方法也存在一些不足，首先分子对接的精确度很大程度上取决于适配体三级结构的精确度，但是获得精确适配体的三级结构目前还比较困难，直接测出适配体这类小分子的空间结构难度较大。由于预测软件很难考虑到体系中多种复杂的因素，通过预测软件获得的三级结构精确度不高，与实际状况下的适配体结构差距较大，往往需要通过 GROMACS 等[40]软件重新计算出体系中适配体

最合适的分子结构后，才能投入使用。其次在高度复杂的体系中进行对接，对分子对接软件的可靠性也仍是一大挑战。目前已有报道，机械学习的方法 (如贝叶斯优化和强化学习) 加速了抗体和药物筛选，这类方法也有望提升该方法的精确度，加速更多高亲和力适配体的产生。表 1 列出了上述提及的适配体筛选方法和目前可用的相关软件。

表 1　计算机辅助适配体筛选方法及其相关软件

Table 1　Computer-aided aptamer screening methods and related program

Type	Program	Access	Description	References
Sequence evaluation	QGRS Mapper	http://bioinformatics.ramapo.edu/QGRS/	Generates information on composition and distribution of putative quadruplex forming g-rich sequences (qgrs) in nucleotide sequences	[12]
	MPBind	http://www.morgridge.net/MPBind.html	A meta-motif based statistical framework and pipeline to predict selex derived binding aptamers	[18]
	FASTAptamer	https://burkelab.missouri.edu/fastaptamer.html	Counting, normalizing, ranking and sorting the abundance of each unique sequence in a population, comparing sequence distributions for two populations, clustering sequences into sequence families based on levenshtein edit distance, calculating fold-enrichment for all of the sequences present across populations, and searching degenerately for nucleotide sequence motifs	[19]
	AptaCluster	https://www.ncbi.nlm.nih.gov/CBBresearch/Przytycka/index.cgi#	Allows for an efficient clustering of whole ht-selex aptamer pools	[20]
Structural analysis	AptaTRACE	https://www.ncbi.nlm.nih.gov/CBBresearch/Przytycka/index.cgi#	A novel approach for the identification of sequence-structure binding motifs in ht-selex derived aptamers	[24]
	APTANI	http://aptani.unimore.it/	Predict specific secondary structures in each selection round and to rank aptamers by motifs embedded in their predicted structures	[25]
	RaptRanker	https://github.com/hmdlab/RaptRanker	A software for RNA aptamer selection from ht-selex experiment data based on local sequence and structure information	[26]
Molecular docking	Hex	http://hex.loria.fr/	An interactive protein docking and molecular superposition program	[28]
	ZDOCK	https://zdock.umassmed.edu/	A protein docking program searches all possible binding modes in the translational and rotational space between the two proteins and evaluates each pose using an energy-based scoring function. It combines gramm for global macromolecular docking, scoring with a statistical potential, clustering of best-scored structures, and local refinement	[29]
	NPDOCK	http://genesilico.pl/NPDock/	A web server for modeling of RNA-protein and DNA-protein complex structures	[30]
	AutoDock	http://autodock.scripps.edu/	Designed to predict how small molecules bind to a receptor of known 3D structure, to perform the docking of the ligand to a set of grids describing the target protein and pre-calculates these grids	[31]
	AutoDock Vina	http://vina.scripps.edu/	An open-source program for doing molecular docking, improving the average accuracy of the binding mode predictions compared to autodock	[32]

3 计算机在适配体上的创新应用

随着计算机在适配体上的应用逐步推广，以及计算机科学技术的飞速发展，人工智能逐步兴起，目前已有研究者将机器学习应用到了SELEX 上。在 2020 年新冠病毒暴发的这段时间里，厦门大学杨朝勇教授为了从大量序列数据中得到高性能的刺突蛋白适配体，设计开发了一款新的序列多维分析算法 SMART-Aptamer，通过无监督机器学习过程，来识别适配体家族并跟踪适配体家族大小的动态变化；以 SARS-CoV-2 的刺突蛋白作为研究对象，为了筛选出能够具有中和作用的 S 蛋白适配体，在原本的正筛选和反筛选的基础上，额外地考虑了与 ACE2 的竞争作用，更重要的是在该过程中使用了 SMART-Aptamer，该算法考虑到了 ACE2 竞争压力，利用了可变选择压力研究了其进化谱系，最后成功得到了高亲和力的适配体 CoV2-RBD-1 和 CoV2-RBD-4[41] (图 4)。该类适配体为 SARS-CoV-2 的诊断和治疗提供新的思路，同时也为深入研究冠状病毒感染的机制提供了新的工具。

计算机不仅能够应用到 SELEX 的流程中，还能够应用在初始文库的设计上。由于初始文库是核酸适配体筛选的源头，作为 SELEX 中最重要的一环，如果设计不合理，必然会影响到筛选的结果[42]。因此越来越多的研究者开始重视初始文库的设计，开始利用计算机设计初始文库。例如，在已知某二级结构与适配体的亲和力密切相关时，就可在核酸适配体数据库中[43]挑选目前已知的核酸适配体，再利用二级结构的预测软件诸如 Zuker 等[44]对挑选出的序列进行预测，由此从庞大的数据库中筛选出具有该特殊结构的适配体，构建出一个新的初始文库，不仅定向增加了含有该结构的适配体，还可保留天然适配体中的高级结构。Luo 等[45]利用相同的思路，对随机文库中的结节进行统计，并对非引物部分进行随机突变来增加文库的丰富程度，最后获得的 RFPoolA 文库，经过验证，相较于普通的随机文库，其高级结节的结构含量有明显增加。2016 年，Zhang 等[46]对已有的前列腺特异性膜抗原适配体进行二级结构的预测，再通过分子对接软件综合比较确

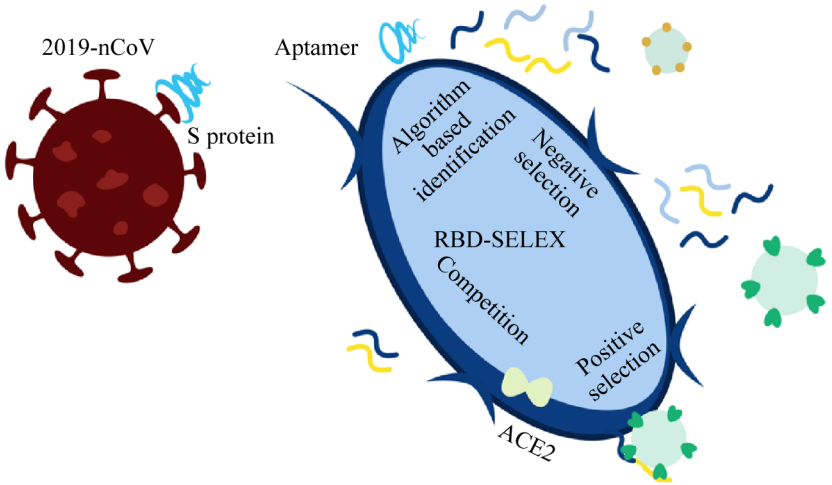

图 4　利用 SMART-Aptamer 改进的 SELEX 流程[41]

Figure 4　The SELEX procedure modified by SMART-aptamer[41].

定了其作用的关键结构，然后在此结构的基础上，对已有的核酸适配体进行单点突变、截短以及插入，通过该方法得到了 1 个新的文库，最后利用三维结构预测软件 Vfold3D[47]进行结构预测。通过分子对接进行亲和力的验证，最终成功获得了亲和力更高且较短序列的适配体。由此表明，分子模拟技术应用于初始文库的设计，确实对于提高适配体亲和力以及筛选效率有明显的效果和应用潜力。

4 总结与展望

有着巨大潜力的计算机技术为适配体的研究带来了福音，不同种类的分析软件和分子模拟软件给研究人员提供了十分有价值的信息，解决了部分 SELEX 的技术问题。例如，通过计算机设计文库，在 SELEX 开始之前就有目的地富集所需的特殊结构的序列，从根本上提高 SELEX 技术的效率，也在一定程度上解决了 SELEX 筛选出的核酸适配体结构简单不具有高级结构的问题。利用打分软件有预见性地对次级文库中的序列进行评估，可以先筛选出一些具有高级结构和稳定结构的核酸适配体；再使用适配体结构预测软件，对那些难以得到结构的适配体进行模拟，也一定程度上减轻了研究适配体结构的困难；最后可以在分子模拟平台上进行对接验证，由此在进行大量亲和力实验前，便可有方向性地得到一些“高亲和力序列”，一定程度上降低了实验室只合成低亲和力的序列用于验证最后导致筛选失败的可能，也大大减少了 SELEX 筛选的轮数，提高了筛选的效率。

结合人工智能，计算机模拟技术有望有更大的突破，但目前该方法仍存在诸多问题：(1) 没有一种标准的初始文库设计的模板，所以需要根据靶标分子的特点、适配体的具体应用，改变文库的设计。(2) 由于对核酸适配体结构的基础研究过少，难以模拟出适配体在环境中真实的三级结构，最后导致对接得到的数据误差较大。(3) 目前，大部分生信专业软件对使用者要求较高，上手难度大，且下载和配置环境十分繁琐。(4) 适配体在不同环境差别过大，目前的对接算法仍需要有更多的改进，半柔性对接仍很难模拟出核酸在复杂体系中分子构象的改变。(5) 现在还无法对修饰后的适配体二级结构进行打分评估。

目前许多研究者已将人工智能的分子模拟平台应用在高通量 SELEX 的分析上，从大量的数据中分析潜在的规律，并训练模型以得到一个可用的模型，指导接下来的实验。虽说这些通过大量数据得到的统计学规律，确实具有很大的参考价值，可实际应用并不是那么的理想。这很大程度上是因为适配体分子较大的柔性，即便是同一个体系中的相同序列得到的适配体结构也可能存在较大的差距，使得通过简单的机械学习分析得到的模型与真实的情况还有较大的差距。而且由于现有的分析技术难以解析大部分核酸适配体的三级结构，使得对适配体-靶物质之间相互作用的理解程度不如抗体-抗原之间的相互作用，这种基础研究较为匮乏的困境也是当下人工智能难以克服的，所以目前的研究更多地希望能通过某种修饰降低适配体的柔性以稳定其空间构象[48]，也有研究通过增强适配体自身的碱基相互作用提升其稳定性，如合成具有 G-四联体的适配体[49]。这也对生物信息学软件在 SELEX 上的应用提出了新的要求，需要更多算法来综合全面地评估适配体，不能只考虑宏观的热力学上的稳定性，还要考虑到适配体分子微观层面上的各种变化，降低柔性带来的影响；还应该在二级结构的预测上考虑到修饰结构，能够折叠出最稳定

的修饰后的二级结构，也需要更新三级结构预测软件的数据库，添加更多的常用于修饰的分子，如生物素等，也可添加一些已报道过的修饰基团作为基本单位，来模拟和评估修饰过的适配体。期待今后的生物信息学研究者能够攻克这些难题，辅助适配体研究者们筛选出更有价值和潜力的适配体。

REFERENCES

[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] Djordjevic M. SELEX experiments: new prospects, applications and data analysis in inferring regulatory pathways. Biomol Eng, 2007, 24(2): 179-189.

[3] 周成林, 许化溪. SELEX技术及Aptamer在小分子中的应用进展. 中国国境卫生检疫杂志, 2011, 34(4): 276-279, 288.
Zhou CL, Xu HX. The application development of SELEX and Aptamer in micromolecules. Chin J Front Heal Quar, 2011, 34(4): 276-279, 288 (in Chinese).

[4] Ruigrok VJ, Levisson M, Hekelaar J, et al. Characterization of aptamer-protein complexes by X-ray crystallography and alternative approaches. Int J Mol Sci, 2012, 13(8): 10537-10552.

[5] Sakamoto T. NMR study of aptamers. 2017.

[6] Lin JS, Kauff A, Diao Y, et al. Creation of DNA aptamers against recombinant bone morphogenetic protein 15. Reprod Fertil Dev, 2016, 28(8): 1164.

[7] 王培培, 刘焰, 罗凤玲, 等. 全细菌 SELEX 技术筛选鸟分枝杆菌的适配子. 中国病原生物学杂志, 2015, 10(2): 113-117, 179.
Wang PP, Liu Y, Luo FL, et al. Aptamers from whole-bacterium SELEX targeting *Mycobacterium avium*. J Pathog Biol, 2015, 10(2): 113-117, 179 (in Chinese).

[8] Kanagawa T. Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Biosci Bioeng, 2003, 96(4): 317-323.

[9] Mendonsa SD, Bowser MT. *In vitro* evolution of functional DNA using capillary electrophoresis. J Am Chem Soc, 2004, 126(1): 20-21.

[10] Komarova N, Barkova D, Kuznetsov A. Implementation of high-throughput sequencing (HTS) in aptamer selection technology. Int J Mol Sci, 2020, 21(22): 8774.

[11] Collie GW, Parkinson GN. The application of DNA and RNA G-quadruplexes to therapeutic medicines. Chem Soc Rev, 2011, 40(12): 5867.

[12] Sakamoto K. Molecular computation by DNA hairpin formation. Science, 2000, 288(5469): 1223-1226.

[13] Gorodkin J, Stricklin SL, Stormo GD. Discovering common stem-loop motifs in unaligned RNA sequences. Nucleic Acids Res, 2001, 29(10): 2135-2144.

[14] Kikin O, D'Antonio L, Bagga PS. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res, 2006, 34(web server issue): W676-W682.

[15] Nonaka Y, Sode K, Ikebukuro K. Screening and improvement of an anti-VEGF DNA aptamer. Molecules, 2010, 15(1): 215-225.

[16] Zheng X, Hu B, Gao SX, et al. A saxitoxin-binding aptamer with higher affinity and inhibitory activity optimized by rational site-directed mutagenesis and truncation. Toxicon, 2015, 101: 41-47.

[17] Hoinka J, Backofen R, Przytycka TM. AptaSUITE: a full-featured bioinformatics framework for the comprehensive analysis of aptamers from HT-SELEX experiments. Mol Ther Nucleic Acids, 2018, 11: 515-517.

[18] Jiang P, Meyer S, Hou Z, et al. MPBind: a meta-motif-based statistical framework and pipeline to predict binding potential of SELEX-derived aptamers. Bioinformatics, 2014, 30(18): 2665-2667.

[19] Alam KK, Chang JL, Burke DH. FASTAptamer: a bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. Mol Ther Nucleic Acids, 2015, 4: e230.

[20] Hoinka J, Berezhnoy A, Sauna ZE, et al. AptaCluster-a method to cluster HT-SELEX aptamer pools and lessons from its application. Res Comput Mol Biol, 2014, 8394: 115-128.

[21] Hiller M, Pudimat R, Busch A, et al. Using RNA secondary structures to guide sequence motif finding towards single-stranded regions. Nucleic Acids Res, 2006, 34(17): e117.

[22] Hori T, Taguchi Y, Uesugi S, et al. The RNA ligands for mouse proline-rich RNA-binding protein (mouse Prrp) contain two consensus sequences in separate loop structure. Nucleic Acids Res, 2005, 33(1): 190-200.

[23] Wang KY, McCurdy S, Shea RG, et al. A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. Biochemistry, 1993, 32(8): 1899-1904.

[24] Dao P, Hoinka J, Takahashi M, et al. AptaTRACE elucidates RNA sequence-structure motifs from selection trends in HT-SELEX experiments. Cell Syst, 2016, 3(1): 62-70.
[25] Caroli J, Taccioli C, De La Fuente A, et al. APTANI: a computational tool to select aptamers through sequence-structure motif analysis of HT-SELEX data. Bioinformatics, 2016, 32(2): 161-164.
[26] Ishida R, Adachi T, Yokota A, et al. RaptRanker: in silico RNA aptamer selection from HT-SELEX experiment based on local sequence and structure information. Nucleic Acids Res, 2020, 48(14): e82.
[27] Zhao Q, Zhao Z, Fan XY, et al. Review of machine-learning methods for RNA secondary structure prediction. 2020.
[28] Ritchie DW. Evaluation of protein docking predictions using Hex 3.1 in CAPRI rounds 1 and 2. Proteins, 2003, 52(1): 98-106.
[29] Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins, 2003, 52(1): 80-87.
[30] Tuszynska I, Magnus M, Jonak K, et al. NPDock: a web server for protein-nucleic acid docking. Nucleic Acids Res, 2015, 43(w1): W425-W430.
[31] Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: applications of autodock. J Mol Recognit, 1996, 9(1): 1-5.
[32] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem, 2010, 31(2): 455-461.
[33] Hu B, Zhou R, Li Z, et al. Study of the binding mechanism of aptamer to palytoxin by docking and molecular simulation. Sci Rep, 2019, 9(1): 15494.
[34] Vu CQ, Rotkrua P, Soontornworajit B, et al. Effect of PDGF-B aptamer on PDGFRβ/PDGF-B interaction: molecular dynamics study. J Mol Graph Model, 2018, 82: 145-156.
[35] Yarizadeh K, Behbahani M, Mohabatkar H, et al. Computational analysis and optimization of carcinoembryonic antigen aptamers and experimental evaluation. J Biotechnol, 2019, 306: 1-8.
[36] Kakoti A, Goswami P. Multifaceted analyses of the interactions between human heart type fatty acid binding protein and its specific aptamers. Biochim Biophys Acta Gen Subj, 2017, 1861(1 pt a): 3289-3299.
[37] Prokofjeva M, Tsvetkov V, Basmanov D, et al. Anti-HIV activities of intramolecular G4 and non-G4 oligonucleotides. Nucleic Acid Ther, 2017, 27(1): 56-66.
[38] Zavyalova E, Turashev A, Novoseltseva A, et al. *Pyrene*-modified DNA aptamers with high affinity to wild-type EGFR and EGFRvⅢ. Nucleic Acid Ther, 2020, 30(3): 175-187.
[39] Zhu J, Wang J, Su ZC, et al. Identification of ssDNA aptamers specific for anti-neuroexcitation peptide Ⅲ and molecular modeling studies: insights into structural interactions. Arch Pharm Res, 2008, 31(9): 1120-1128.
[40] Van Der Spoel D, Lindahl E, Hess B, et al. GROMACS: fast, flexible, and free. J Comput Chem, 2005, 26(16): 1701-1718.
[41] Song Y, Song J, Wei X, et al. Discovery of aptamers targeting the receptor-binding domain of the SARS-CoV-2 spike glycoprotein. Anal Chem, 2020, 92(14): 9895-9900.
[42] 雷云霞, 李招发, 林俊生. 分子模拟指导核酸适配体文库设计. 中国生物化学与分子生物学报, 2019, 35(3): 251-258.
Lei YX, Li ZF, Lin JS. The design of aptamer candidate pool guided by molecular simulation. Chin J Biochem Mol Biol, 2019, 35(3): 251-258 (in Chinese).
[43] Lee JF, Hesselberth JR, Meyers LA, et al. Aptamer database. Nucleic Acids Res, 2004, 32(Database issue): D95-D100.
[44] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res, 2003, 31(13): 3406-3415.
[45] Luo X, McKeague M, Pitre S, et al. Computational approaches toward the design of pools for the *in vitro* selection of complex aptamers. RNA, 2010, 16(11): 2252-2262.
[46] Zhang HL, Liu XG, Wu FB, et al. A novel prostate-specific membrane-antigen (PSMA) targeted micelle-encapsulating wogonin inhibits prostate cancer cell proliferation via inducing intrinsic apoptotic pathway. Int J Mol Sci, 2016, 17(5): E676.
[47] Cao S, Chen SJ. Physics-based *de novo* prediction of RNA 3D structures. J Phys Chem B, 2011, 115(14): 4216-4226.
[48] Wang RE, Wu H, Niu Y, et al. Improving the stability of aptamers by chemical modification. Curr Med Chem, 2011, 18(27): 4126-4138.
[49] Roxo C, Kotkowiak W, Pasternak A. G-quadruplex -forming aptamers-characteristics, applications, and perspectives. Molecules, 2019, 24(20): E3781.

(本文责编 郝丽芳)