



面向核酸适配体的人工智能方法与应用*

刘上华¹⁾ 张洪旗¹⁾ 刘入铭¹⁾ 曾红娟¹⁾ 邓科君¹⁾ 鄢丹^{2,3)} 汤丽霞¹⁾ 林昊^{1)**}⁽¹⁾ 电子科技大学生命科学与技术学院, 信息生物学研究中心, 成都 611731; ⁽²⁾ 首都医科大学附属北京友谊医院药理学部, 北京 100050;⁽³⁾ 北京市临床药学研究所, 北京 100035)

摘要 核酸适配体 (nucleic acid aptamer) 是一类能够高效、特异识别蛋白质、小分子和完整细胞等多种靶标的单链寡核苷酸分子, 凭借其分子质量小、化学与热稳定性强、易于修饰且免疫原性低等优势, 在诊断、治疗与分析检测领域展现出广阔的应用前景。本文首先对现有核酸适配体数据库进行了系统梳理, 随后概述了高通量 SELEX 数据的模拟筛选流程; 接着综述了从序列特征提取到二级和三级结构预测的各类生物信息学方法; 然后评述了分子对接、分子动力学及结合自由能计算结合机器学习与深度学习技术在核酸适配体-靶标相互作用预测中的应用; 最后总结了基于理性设计、生成式 AI 与进化算法的计算设计策略, 并列举了精准诊断、生物传感和食品安全等典型案例。通过多层次、多方法的整合, 本文为核酸适配体的发现、设计与应用提供了一套全面的生物信息学工具与思路, 并对未来智能闭环和多模态融合的研究方向进行了展望。

关键词 核酸适配体, SELEX 数据分析, 结构预测, 相互作用预测, 生成式设计, 生物信息学

中图分类号 Q-1, Q-31, Q74

DOI: 10.3724/j.pibb.2025.0355

CSTR: 32369.14.pibb.20250355

核酸适配体是一类由单链 DNA 或 RNA 组成的短链寡核苷酸, 长度一般是 25~80 个碱基, 通过其独特的三维折叠结构能够高效、特异地识别并结合蛋白质、小分子、病毒颗粒, 乃至整细胞或金属离子^[1-5]。与传统抗体相比, 核酸适配体具有分子质量小、合成及化学修饰便捷、热稳定性和化学稳定性强、免疫原性低及组织穿透性良好的显著优势, 使其在医学诊断与治疗、生物传感器开发和农业与食品安全等方面展现出广阔的应用前景^[6-12]。自 1990 年首次通过指数富集配体系统进化技术 (systematic evolution of ligand by exponential enrichment, SELEX)^[13-14] 筛选出高亲和力核酸适配体以来, 该领域经历了由经典 SELEX 向高通量 SELEX (HT-SELEX)^[15-17]、以活细胞为靶标的 SELEX (Cell-SELEX)^[18]、计算机模拟 SELEX (*In silico*-SELEX)^[19-20] 等多种技术改进的快速演变, 有效缩短了筛选周期并提高了产出效率。

在 SELEX 流程中, 多轮结合-洗脱-扩增迭代可在高达 10^{15} 水平的随机文库中富集目标结合序列, 但传统实验操作烦琐、耗时长, 并且对实验条

件十分敏感^[21-25]。伴随着高通量测序平台 (如 Illumina) 的普及以及 AptaPLEX、AptaSuite 等专用生物信息学工具的出现, 研究者得以在“实验-测序-计算”闭环中, 实时监控各轮富集动态, 利用序列/丰度矩阵和 *k*-mer、二级结构、物化属性等多维特征, 通过聚类、基序 (motif) 挖掘和统计显著性分析, 大幅提升了候选序列的识别率与可重复性^[22, 26-29]。同时, 核酸二级结构预测 (RNAfold、Mfold 等) 及三维结构预测 (Rosetta FARFAR2、3dRNA、RNAComposer) 工具的结合, 为核酸适配体的功能位点定位和理性改造提供了坚实的结构基础^[30-31]。

近年来, 随着机器学习和深度学习在生物学各个领域的快速发展, 越来越多的研究团队将支持向量机 (support vector machine, SVM)、随机森林 (random forest, RF)、图神经网络 (graph neural

* 国家自然科学基金 (82130112) 资助项目。

** 通讯联系人。

Tel: 13678168394, E-mail: hlin@uestc.edu.cn

收稿日期: 2025-07-29, 接受日期: 2025-09-11

network, GNN)、Transformer、BERT (bidirectional encoder representations from transformers) 等算法引入核酸适配体-靶标相互作用预测, 显著提高了结合亲和力与特异性的预测准确率^[32-34]。生成式模型(如变分自编码器/对抗生成网络、扩散模型、强化学习等)更进一步实现了从现有高亲和力序列库或表型数据出发的端到端核酸适配体设计与优化, 并辅以蒙特卡洛树搜索(Monte Carlo tree search, MCTS)、遗传算法和多目标优化等计算策略, 实现了“生成-预测-筛选-验证”的闭环迭代流程^[35-37]。

本综述从以下几个方面展开: 第一, 介绍核酸适配体的各种数据库; 第二, 总结高通量 SELEX 数据虚拟筛选方法; 第三, 总结序列、结构特征的挖掘与预测; 第四, 详述相互作用预测与模型构

建; 第五, 系统评述生成式与优化算法在核酸适配体设计中的应用; 最后, 结合医学诊断、生物传感和食品安全等领域的典型案例, 展望核酸适配体技术在精准医疗和环境监测中的未来发展方向(图1)。与近年综述^[1-2, 38-39]相比, 本文新增体现在4点: a. 全链条与可复现, 按“数据库、筛选、结构、相互作用、设计、应用”构建一体化框架; b. 选型准则与可靠性, 区分复合体与自由态小型单链核酸的建模路径并结合化学探针与伪结处理; c. 生成式闭环, 联动生成式设计、主动或进化优化与物理和可制造性约束; d. 评测对齐与落地, 建立计算指标与实验读数的映射并提炼工程化要点。通过本综述, 期望为研究者提供完整的生物信息学工具和策略指南, 助力适配体从实验室走向临床与产业化应用。

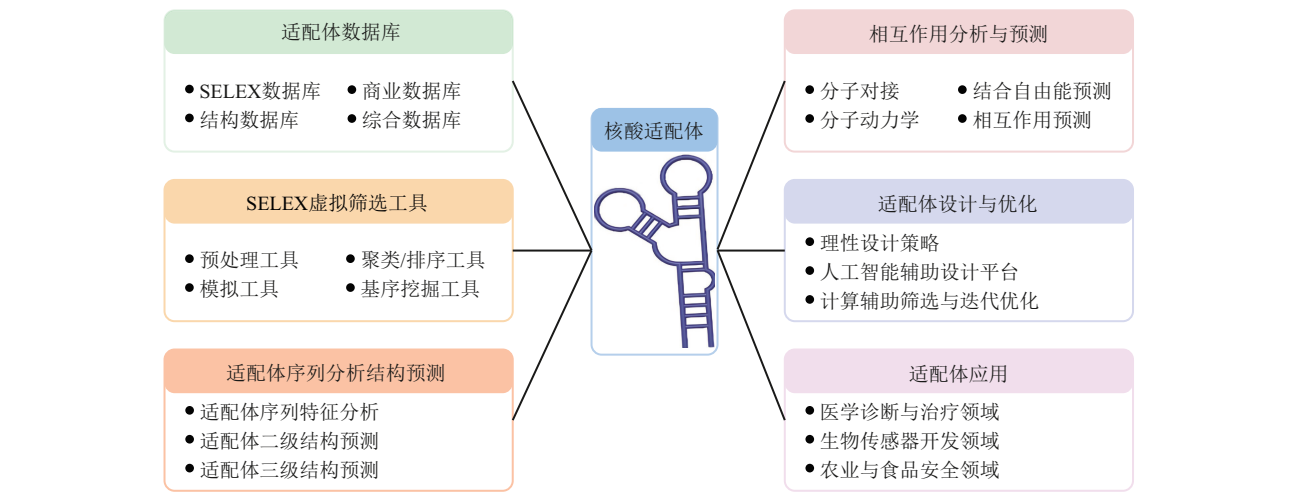


Fig. 1 Overall framework of this review
图1 本综述整体框架

1 核酸适配体的生物信息数据库与资源

近年来, 随着核酸适配体研究的迅猛发展, 多种专门的生物信息数据库应运而生, 为科研人员提供了丰富的序列、结构及靶标相关信息。根据各数据库的功能定位和应用场景, 可将其大致分为4类(表1):

第一类是基于 SELEX 技术的功能位点数据库, 如 SELEX_DB^[40] 与 HTPSELEX^[41]。SELEX_DB 主要收录传统 SELEX 实验筛选得到的功能位点序列, 但数据量很少; HTPSELEX 则侧重于高通量 SELEX 数据, 尤其针对转录因子结合位点的建模

与分析, 能够处理大规模测序数据并挖掘富集规律。

第二类是 RNA 核酸适配体的三维结构与相互作用分析数据库, 典型代表为 RNAapt3D^[42]。该数据库提供高精度的 RNA 三维结构预测, 以及 RNA-蛋白质相互作用位点分析功能, 是药物设计与结构生物学研究的重要支撑。

第三类为商业化及应用导向数据库, 如 Apta-IndexTM。此类平台汇集了核酸适配体序列的专利与市场化信息, 使产业界用户能够快速获取产品开发动态和技术进展。

第四类是综合性的核酸适配体数据库, 包括

Aptamer Database^[43]、RiboaptDB^[44]、Aptamer base^[45]、Aptabase^[46]、AptaDB^[47]、UTexas Aptamer Database^[48]与全球核酸适配体数据平台等。这些综合性数据库以数据完整性和结构化描述为特色,定期更新并严格遵循统一的数据格式。Aptamer Database收录天然与人工设计的核酸适配体序列,同时详尽记录靶标类型与实验条件,为序列空间探索与理论生物学研究提供坚实支持;RiboaptDB则专注于RNA核酸适配体与核酶序列的全面汇聚,兼顾深度计算分析与结构预测需求;Aptamer base强调实验条件的翔实记录与开放式协作机制;Aptabase以系统化的数据管理与版本控制为核心,保障数据的可追溯性与持续更新;AptaDB在数据标准化与维护流程上表现卓越,提供清晰的接口与注释;UTexas Aptamer Database是最新最全面的数据库之一,为研究者提供最新且一

致的资源;中国科学院杭州医学院开发的“全球核酸适配体数据平台”不仅整合了核酸适配体序列、结构预测、靶标评价与筛选工具,还提供实时更新的文献、专利和临床试验信息,为研究者提供了一站式的全流程服务。

尽管上述数据库在数据格式与组织方式上存在差异,将它们整合并加以利用仍是核酸适配体生物信息学研究的核心挑战。研究人员可借助批量爬取与数据清洗技术,实现跨数据库的格式统一与标准化存储;同时,利用Biopython、BioPandas等编程工具,以及RNAfold、PyMOL等序列与结构分析软件,深入挖掘数据库资源,构建高效的综合性数据整合平台^[49-51]。此类平台不仅可大幅提升数据可用性与分析效率,更将为未来核酸适配体的设计、筛选及产业化应用提供坚实的数据与技术基础。

Table 1 Summary of nucleic acid aptamer databases

表1 核酸适配体数据库汇总

数据库	特点	机构	创建时间	更新时间	数据量	网址	访问
SELEX_DB ^[40]	侧重SELEX技术获得的功能位点序列,支持基因组功能位点预测	俄罗斯科学院西伯利亚分院细胞学与遗传学研究所	2000年	不活跃	116条适配体	http://www.mgs.bionet.nsc.ru/mgs/systems/selex/	不可访问
Aptamer Database ^[43]	涵盖序列、靶标和实验条件,适于序列空间与理论研究	德克萨斯大学奥斯汀分校	2004年	不活跃	239条适配体	http://aptamer.icmb.utexas.edu/	不可访问
RiboaptDB ^[44]	丰富的RNA适配体与核酶序列,适合理论与计算生物学研究	南密西西比大学	2006年	不活跃	3 842条适配体	http://mfgn.usm.edu/ebi/riboapt/	不可访问
HTPSELEX ^[41]	高通量SELEX实验数据,支持转录因子结合位点建模	瑞士实验癌症研究所与瑞士生物信息研究所	2006年	2012	高通量SELEX数据	https://epd.expasy.org/htpselex/	可以访问
Apta-Index TM	商业化数据库,提供适配体序列与市场化应用信息	Aptagen公司	2008年	活跃	895条适配体	https://www.aptagen.com/apta-index/	可以访问
Aptamer base ^[45]	强调实验条件详细结构化描述,开放协作更新	卡尔顿大学	2012年	2012	928条适配体	http://aptamer.freebase.com	不可访问
Aptabase ^[46]	定期更新、系统性记录核酸适配体序列和相关实验数据	印度理工学院古瓦哈提分校	2021年	2021	618条适配体	https://www.iitg.ac.in/proj/aptabase/index.html	可以访问
RNAapt3D ^[42]	专注于RNA适配体三维结构预测及RNA-蛋白质相互作用分析	日本国立先进工业科学技术研究所	2022年	2022	55条适配体	https://rnaapt3d.medals.jp/	可以访问
AptaDB ^[47]	涵盖较新且丰富的核酸适配体数据,实时更新,数据量大	华东理工大学	2024年	2024	1 350条适配体	https://lmmd.ecust.edu.cn/aptadb/	可以访问
UTexas Aptamer Database ^[48]	全面且持续更新的核酸适配体数据,强调标准化及长期维护	德克萨斯大学奥斯汀分校	2024年	2024	1 495条适配体	https://sites.utexas.edu/aptamerdatabase	可以访问

续表1

数据库	特点	机构	创建时间	更新时间	数据量	网址	访问
全球核酸适配体数据库	核酸适配体数量最多, 集成了对应的文献以及智能筛选系统	中国科学院杭州医学研究所	—	活跃	4 178条适配体	https://www.aptamer.org/	可以访问

核验日期: 2025年8月31日。

2 SELEX 高通量数据虚拟筛选流程

SELEX 自 20 世纪 90 年代提出以来, 已发展为从随机寡核苷酸文库中高效筛选出特异性结合靶标分子的标准技术。随着高通量测序与合成生物学方法的快速进步, 现代 SELEX 实验通常产生包含百万至千万级别序列及其对应丰度信息的带有质量分数的序列文件格式 (fast quality, FASTQ) 原始数据。对这些海量数据进行系统化处理, 不仅能够准确追踪每一轮富集过程中候选序列的动态变化, 还能评估实验条件 (如结合强度、洗脱效率、扩增偏好等) 对结果的影响, 从而优化实验设计并提高最终核酸适配体的亲和力与特异性^[21-22, 27, 52-53]。此外, 随着应用需求的多样化, 基于传统 SELEX 方法还衍生出了多种改进版本, 常见的包括: 以活细胞为靶标的 SELEX (Cell-SELEX)^[54]、多靶标交替富集的 SELEX (Toggle-SELEX)^[55]、在活体环境中进行的 SELEX (*In vivo* SELEX)^[56]、固相靶标捕获 SELEX (Capture-SELEX)^[57]、毛细管电泳分离 SELEX (CE-SELEX)^[58]、高保真 PCR 优化 SELEX (Hi-Fi SELEX)^[59] 以及基于微流控芯片的 SELEX (microfluidic SELEX)^[60] 等。基于此, 首先需要将原始测序数据转化为标准化的“序列/丰度”表格, 随后通过模拟、聚类、基序挖掘、统计检验及可视化等一系列分析步骤, 提炼出最具价值的候选序列。下面将依次介绍 SELEX 数据分析流程中的关键工具及其特点 (表 2)。

在 SELEX 高通量数据的分析流程中, 首先要对原始测序数据进行预处理。此阶段常用的工具包括 AptapLEX^[61] 和 Aptasuite 中的 Aptamux 模块^[62-63]。它们能够对原始 FASTQ 文件完成序列去重、样本分拣、接头剪切和低质量序列过滤, 并最终输出每轮富集的“序列/丰度”矩阵。相比通用的测序质控软件, 这些工具针对 SELEX 实验特点进行了优化, 能够直接生成后续分析所需的标准化表格, 不足之处在于对标签序列 (barcode) 配置格式和参数依赖程度较高, 且大多数工具只提供命

令行接口, 缺少可视化质控报告。

在“干试验”设计和参数优化环节, Aptasim^[62] 为研究者提供了一个可定制的 SELEX 过程模拟平台。通过对初始文库的随机化、结合-洗脱-扩增步骤的数学建模, Aptasim 可生成多轮模拟数据, 用于评估不同洗脱强度、扩增效率和结合亲和力模型对富集效果的影响。该工具在优化实验方案、验证富集算法性能方面优势明显, 但其简化的噪声模型往往无法完全再现真实实验中的复杂偏差, 需要结合实际测序数据加以校正。

完成富集实验或模拟后, 便进入序列聚类与排序阶段。AptacLUSTER^[64]、FASTAptamer^[65-66]、RaptRanker^[67]、AptCompare^[68]、AptamerRunner^[69] 和 AptafastZ^[70] 等工具, 能够基于序列相似性或 *k*-mer 丰度变化, 将海量序列分为若干簇, 并结合轮次间的丰度动态, 对簇中心或个体序列进行优先级排序。AptacLUSTER 和 FASTAptamer 以轻量级的 *k*-mer 频率统计著称, 能够快速完成库中序列的初步筛选; RaptRanker 在此基础上进一步纳入二级结构信息, 从而显著提高了对高亲和力核酸适配体的识别准确性; AptCompare 则通过跨实验条件的候选序列交叉比对, 能够挖掘不同 SELEX 方案下的共性和差异。尽管上述工具各有优势, 但在面对数百万量级的高通量序列库时, 聚类运算仍十分耗时, 而且不同软件的输出格式多样, 往往需要额外脚本进行结果整合。为此, AptamerRunner 提供了一种一站式聚类与网络构建方案: 它首先使用 RNAfold 预测每条序列的二级结构, 然后结合多种相似性度量执行 all-vs-all 聚类, 并生成可直接导入 Cytoscape 的结构化网络, 大幅简化了后续可视化与分析流程。与此同时, AptafastZ 针对高通量排序需求进行了专门优化——该工具会跳过在任何富集轮次中均未出现的稀疏序列, 并通过简化的打分模型将亲和力计算速度提升 10 倍以上, 而排序精度仍可与经典算法相媲美, 为大规模亲和力排序提供了高效可靠的替代方案。

在基序挖掘环节, Aptatrace^[71]、Aptasuite 的 Aptamotif^[62]、APTANI^[72] 和 MPBind^[73] 等工具

各具特色。它们通过对每轮序列的二级结构预测及短 k -mer 片段富集趋势分析, 锁定真正参与靶标结合的功能性序列-结构片段。AptaTRACE 敏感度高, 能早期检测到潜在基序; APTANI 则利用信息论指标自动截取关键片段并绘制热图; MPBind 在传统统计学习框架中嵌入丰富的物理化学耦合特

征, 实现对序列亲和力的精细评分。但结构预测本身耗时, 且参数调优较为烦琐。此外, APV-Sankey^[74] 利用 Sankey 流程图直观展示 k -mer 在各轮次中的富集演化路径, 并集成聚类 and 丰度数据, 可交互式追踪候选基序的富集动态与突变历史, 极大提升了基序挖掘和结果解读的效率。

Table 2 SELEX data analysis tools

表2 SELEX数据分析工具

工具名称	主要模块	运行语言	GUI/Web界面	典型输出格式
AptaPLEX ^[61]	预处理	C++/Python	否	TSV, FASTQ
AptaSIM ^[62]	模拟	Python	否	FASTA
AptaCLUSTER ^[64]	聚类	Java	否	FASTA, TSV
AptamerRunner ^[69]	聚类	Python	否	GraphML/TSV
FASTAptamer ^[65-66]	聚类/排序	Python	否	TSV
RaptRanker ^[67]	聚类/排序	Python	否	TSV
AptCompare ^[68]	聚类/整合	Python	否	TSV
AptaFastZ ^[70]	排序/亲和力打分	Python	否	TSV
AptaTRACE ^[71]	基序挖掘	Python	否	TSV, JSON
AptaSuite ^[62]	全管线 (含预处理、聚类、基序)	Java/Python	部分	TSV, CSV
APTANI ^[72]	基序挖掘	Python/R	否	TSV, heatmap
MPBind ^[73]	基序挖掘/评分	Python	否	TSV
APV-Sankey ^[74]	基序可视化	Python/JavaScript	是	HTML/JSON

TSV: 制表符分隔的数值文件 (tab-separated values); FASTQ: 带有质量分数的序列文件格式; FASTA: 纯序列文件格式; GraphML: 图形结构标记语言 (graph markup language); JSON: JavaScript对象表示法 (JavaScript object notation); CSV: 逗号分隔数值文件 (comma-separated values); HTML: 超文本标记语言 (hypertext markup language); heatmap: 热图, 用颜色深浅表示数值大小的图示方法; GUI/Web界面: 图形用户界面或网页界面 (graphical user interface/web interface)。

3 核酸适配体序列分析与预测

核酸适配体序列分析与预测是核酸适配体研究的重要基础, 涉及序列特征分析、结构预测及相关生物信息学工具的综合应用。通过有效的序列分析与预测, 可深入理解核酸适配体的功能机制, 并为后续核酸适配体的理性设计提供重要指导。本节系统梳理并评述当前应用广泛且在基准中表现领先的关键工具与方法。

3.1 序列特征分析

在核酸适配体序列特征表征中, 基础组成特征的挖掘既是最直观的, 也是后续一切复杂分析的基石。首先, 通过统计单个序列中 A、T/U、G、C 4 种碱基的绝对丰度及相对频率, 可计算出 GC 含量、AT/GC 比值等常用指标, 这些参数不仅反映了序列的整体热稳定性, 还可与实验测得的熔解温度、结合动力学参数直接关联; 进一步地, 将序列

划分为固定长度的滑动窗口, 统计所有长度为 2、3, 乃至 4、5 的 k -mer 子串出现频率, 并采用归一化或词频-反文档频率 (term frequency-inverse document frequency, TF-IDF) 方法得到相对权重, 可识别出在不同 SELEX 轮次中持续富集或稀释的局部序列模式^[75-77]。此外, 越来越多的研究借助核苷酸理化属性——如堆积能 (stacking energy)、电子离子相互作用势能 (electron-ion interaction potential, EIIP)、亲水性/疏水性指数、核苷酸分子质量、等电点、转子角度等——将各碱基映射为数值序列, 构建基于物理化学属性的组合特征, 以期更全面地刻画序列的折叠倾向和结合亲和力^[78-81]。

序列复杂度与保守性分析侧重从宏观和进化角度评估核酸适配体的多样性与关键位点。信息论中的 Shannon 熵^[82]、Rényi 熵^[83]或 Tsallis 熵^[84]可用于量化碱基分布的随机性与多样性, 而压缩算法计

算的压缩比率^[85]、低复杂度检测算法^[86]简化复杂度分数等则进一步反映序列内部的重复性和低复杂区段;此外,语言学中的“重复度指数”(linguistic complexity)和混沌图(chaos game representation)所揭示的分形维数,也为序列的内部模式提供了独特视角^[87]。在保守性方面,借助Clustal Omega^[88]、MAFFT^[89]等多序列比对工具,可以识别SELEX各迭代轮中或不同物种同源序列间的高度保守区;基于比对结果生成的序列logo(WebLogo、SeqLogo)及位点互信息分析,则能直观展示每一位点上碱基的偏好性与依赖性,为功能性基序的定位和实验验证提供有力线索^[90]。

除了上述基础与宏观特征外,引入高级统计学与耦合特征能够挖掘序列中远程位点间的相互作用与全局顺序信息。自关联(autocorrelation)和交叉关联(cross-correlation)特征通过计算相隔 d 个碱基位置的理化属性之间的相关性,将碱基间的长程依赖关系量化^[91-92],常用的方法包括Moran自关联系数^[93]、Geary自交互指标^[94]、伪核苷酸组成(pseudo nucleotide composition, PseNC)及其扩展变体在传统 k -mer统计基础上融入理化耦合因子,通过一阶、二阶乃至更高阶的序列耦合数(sequence-order coupling number)刻画核苷酸在折叠与结合过程中的协同效应^[95];而基于二核苷酸自交叉协方差(dinucleotide auto-cross covariance)、三核苷酸理化性质协方差(trinucleotide physicochemical covariance)的特征,则从二核苷酸和三核苷酸层面进一步细化了局部信息与全局信息的交互模式^[96-97]。这些高级特征已在RNA二级结构预测、核酸适配体活性分类与结合亲和力评估中被广泛验证,可显著提升模型对微妙序列差异的辨识能力,为高精度的核酸适配体设计提供坚实的数据支撑。

3.2 二级结构预测

在核酸适配体二级结构预测领域,现有方法大

体可分为三大类(表3):基于热力学最小自由能的经典方法、基于统计与协同进化的模型,以及深度学习与端到端预测方法。第一类方法以RNAfold^[98]、ViennaRNA^[99-100]、Mfold^[101]和UNAFold^[102]为代表,均采用Turner参数集与动态规划或能量搜索算法,在数秒内为长序列计算出最小自由能(minimum free energy, MFE)结构及其配对概率,但对复杂伪结(RNA二级结构中一种非嵌套配对)支持有限且仅能输出静态最优或子最优构象;第二类方法以CONTRAFold^[103]和Infernal(covariance models)^[104]为典型,前者通过条件随机场(conditional random field, CRF)学习序列-结构关联,后者结合隐马尔可夫模型(hidden Markov model, HMM)与协同变异信息,对家族内高度保守区域的预测尤为精准,但对孤立序列及缺乏同源数据的场景表现受限;第三类方法则借助深度神经网络实现端到端预测,如SPOT-RNA^[105]将卷积神经网络(convolutional neural network, CNN)与长短期记忆(long short-term memory, LSTM)网络相结合直接预测碱基配对概率,E2EFold^[106]基于Transformer同时预测配对区与非配对区,MXfold2^[107]和UFold^[108]则分别将热力学特征或图像分割思路融入网络结构,显著提升了对伪结和非典型结构的识别能力,但需要大量标注数据且模型解释性尚待加强,RNA-FM^[109]基于2300万条ncRNA自监督学习,作为通用表征用于下游二级结构与可及性等任务,RNAErnie^[110]在掩码与基序感知任务上进行预训练,强化了结构/相互作用相关的表征,RiNALMo^[111]规模更大,强调对未见家族的泛化能力。今后,可以将三类方法的优势加以整合,例如在深度网络中引入物理化学约束,或在统计模型中融入神经网络表征,以期进一步提升核酸适配体二级结构预测的准确性与适用范围。

Table 3 Nucleic acid aptamer secondary structure prediction tools
表3 核酸适配体二级结构预测工具

方法类别	软件名称	核心原理	优势	局限
热力学最小自由能	RNAfold ^[98] /	动态规划+Turner能量参数	计算快速、参数成熟,支持配对概率输出	对伪结支持有限,仅输出静态最优/次优结构
	ViennaRNA ^[99-100]			
	Mfold ^[101] /UNAFold ^[102]	能量搜索+参数优化	支持伪结预测,界面与参数灵活	计算成本高,长序列耗时
统计与协同进化	CONTRAFold ^[103]	CRF	捕捉序列-结构关联,训练后速度快	依赖大量已知结构数据,孤立序列效果欠佳

续表3

方法类别	软件名称	核心原理	优势	局限
深度学习与端到端预测	Infernal (CMs) ^[104]	隐马尔可夫模型+协同变异	进化保守性强, 适用于同源家族分析	需丰富同源序列, 计算与比对耗时
	SPOT-RNA ^[105]	CNN+LSTM	可直接从序列预测配对概率, 伪结识别能力强	训练数据需求大, 模型解释性不足
	E2Efold ^[106]	Transformer	端到端预测, 兼顾配对与非配对区域	高算力需求, 对长序列内存消耗大
	MXfold2 ^[107] /UFold ^[108]	热力学特征融合或图像分割网络	对复杂结构和伪结预测更准确	模型训练与调参复杂, 解释性有限
	RNA-FM ^[109]	BERT式自监督预训练	跨任务/小样本迁移效果好、可与现有预测器组合	需针对体系与条件做任务适配, 端到端二级结构推断仍依赖下游头部或微调
	RNAErnie ^[110]	多任务预训练	在结构相关任务与RNA-RNA相互作用上表现优良, 表征更贴近结构语义	模型体量与训练资源要求较高, 公开实现更新节奏需关注
	RiNALMo ^[111]	通用RNA语言模型	对跨家族二级结构预测具优势, 适合少样本微调到适配体	不直接输出配对矩阵, 需下游预测头或微调; 模型体量较大, 推理与训练开销偏高

Turner参数: 用于核酸二级结构预测的热力学参数集合; CRF: 条件随机场 (conditional random field); HMM: 隐马尔可夫模型 (hidden Markov model, HMM); CNN: 卷积神经网络 (convolutional neural network), LSTM: 长短期记忆 (long short-term memory); Transformer: 一种基于自注意力机制的深度学习模型架构。

特别地, 针对适配体常见的伪结与非典型配对, 建议结合显式支持伪结的算法 KnotFold 与化学探针数据进行交叉验证^[112]。考虑到短链适配体的构象多样性与条件敏感性, 单一最小自由能解往往难以覆盖真实状态, 实际应用中宜同时报告配对概率矩阵/候选构象集, 并将其作为后续三维建模与相互作用预测的约束。

3.3 三级结构预测

核酸适配体三级结构预测方法主要可分为三大类。第一类为片段组装与能量最小化方法, 如 Rosetta FARFAR/FARFAR2^[113]、3dRNA^[114] 和 SimRNA^[115]。FARFAR/FARFAR2 与 3dRNA 根据已知二级结构将序列拆分为片段/结构模块, 利用片段库或模板进行初始构建, 结合蒙特卡洛/模拟退火搜索与能量函数打分, 并在全原子势能下最小化 (可辅以分子动力学精炼), 对中小尺寸 RNA 可获得近至原子级的三维模型; SimRNA 则采用粗粒度表示与知识势能进行蒙特卡洛全局采样, 可结合二级结构与距离约束, 所得模型需经全原子重建与精修以提升精度, 但其对高质量片段库的依赖较强, 且计算资源开销显著。第二类为基于模板的同源建模方法, 代表工具包括 RNAComposer^[116] 与 ModeRNA^[117], 这类工具通过将目标序列的二级结构映射到预先构建的片段或同源骨架上, 实现一

键式三维模型构建, 并可对局部环区或插入片段进行快速优化。该策略具有计算速度快、易于实施的优点, 但仅在存在合适同源模板时才能保证预测可靠性。第三类为深度学习与 AI 辅助方法。RoseTTAFold-RNA^[118] 扩展了 RoseTTAFold 的三轨注意力与图神经网络框架, 可同时学习蛋白质与核酸的几何约束与接触图, 从而端到端预测蛋白质-RNA/DNA 复合物的三维结构, 并可结合几何或能量精修提升准确度; AlphaFold-RNA^[119] 在蛋白质框架基础上针对 RNA 改造, 结合多序列比对与注意力推断距离与取向, 同时给出置信度指标; AlphaFold 3^[120] 可用于生成界面位形与置信度指标; 而对自由态、长度较短的核酸适配体, 更建议遵循“二级结构与化学探针约束→三维采样/精修”的路径。在应用层面, 此类模型可用于快速提出结合界面假设与候选构象、对关键残基/核苷酸及其突变体进行优先级排序、为对接/分子动力学与自由能计算提供合理初始构象, 并在多候选序列间完成预筛与淘汰。其性能边界主要体现在: 对自由态短链与多构象/伪结体系覆盖不足; 对小分子与金属离子配位化学的精细刻画有限; 对修饰核苷酸的参数化与识别不充分; 置信度指标并不等同于亲和力或实验可验证的稳定性。当前局限还包括: 对离子强度、pH、温度和缓冲体系等条件依赖缺乏显

式建模; RNA 专用模型向 ssDNA 适配体的迁移受限; 训练数据偏倚可能导致界面样式过拟合。总体而言, 这些方法在速度与复合体建模方面具优势, 但对伪结、构象系综与条件依赖的刻画仍不足, 建议与对接-分子动力学-自由能评估联合交叉验证, 并在报告中提供候选构象集及阈值说明 (表4)。

Table 4 Nucleic acid aptamer tertiary structure prediction tools
表4 核酸适配体三级结构预测工具

方法类别	软件名称	核心原理	官网地址	优势	局限
片段组装 与能量最小化	Rosetta	经验片段拼装+蒙特卡	https://rosic.rosettacommons.org/farfar2	原子级精度高; 可加入实验约束	计算耗时; 依赖高质量片段库
	FARFAR2 ^[113]	洛搜索+能量微调			
	3dRNA ^[114]	片段模板拼接+二级结构约束	http://biophy.hust.edu.cn/3dRNA	构建速度快; 对中短链核酸适配体友好	对非典型折链支持有限; 精度依赖二级结构质量
	SimRNA ^[115]	知识驱动粗粒度蒙特卡洛折叠采样	https://genesilico.pl/SimRNAweb	计算开销低; 适合全局折叠搜索	精度低于原子级方法; 对复杂伪结支持有限
模板同源建模	RNAComposer ^[116]	二级结构→结构片段数据库映射+局部优化	http://rnacomposer.cs.put.poznan.pl/	速度快; 易实施	需存在高同源模板; 对新颖序列无效
	ModeRNA ^[117]	同源PDB结构比对+核苷酸替换	https://iimcb.genesilico.pl/modernaserver/	简单直接; 适合家族内同源建模	仅在存在合适模板时可靠; 对远缘序列效果欠佳
	RoseTTAFoldNA ^[118]	图神经网络+多序列比对信息	https://neurosnap.ai/service/RoseTTAFold%20All-Atom	端到端预测; 速度快; 对复合物预测有潜力	对自由态短链适配体直接折叠不优; 对小分子/离子体系支持有限
深度学习与AI辅助	AlphaFold-RNA ^[119]	Transformer架构+距离矩阵预测+物理优化	https://deepmind.com/science/alphafold/alphafold-server	融合进化和结构信息; 一体化高准确度预测	对RNA应用尚在探索中; 模型和数据集仍不完善
	AlphaFold 3 ^[120]	多组分扩散式架构, 联合预测蛋白、核酸、小分子/离子等复合体与置信度	https://alphafoldserver.com/welcome	适合核酸适配体-蛋白质界面假设与候选筛选, 多构象覆盖有限; 提供pLDDT/PAE等指标	自由态短链RNA/DNA折叠与化学细节与条件依赖需交叉验证

pLDDT: 预测局部距离差评分 (predicted local distance difference test); PAE: 预测对齐误差 (predicted aligned error)。

4 核酸适配体与靶标相互作用分析与预测

上一节系统讨论了核酸适配体的序列特征挖掘及其二级、三级结构预测方法, 为后续功能分析和性能优化奠定了坚实基础。本节将延续这一思路, 回顾核酸适配体与靶标相互作用的主要计算预测策略。首先, 分子对接方法 (如 AutoDock Vina/ AutoDock4、ZDOCK、MDockPP 和 DOCK) 通过刚性或柔性搜索与经验/知识驱动的打分函数, 为核酸适配体-蛋白质复合物提供初步的结合构象和界面信息。随后, 全原子分子动力学模拟 (GROMACS、NAMD、AMBER) 结合 MM-PBSA/MM-GBSA 或热力学积分 (thermodynamic integration, TI) 和自由能微扰 (free energy perturbation, FEP) 等自由能计算, 对这些对接模型的动力学稳定性与结合亲和力进行定量评估。最后, 面对海量候选序列, 机器学习与深度学习方法

(包括 RF、SVM、MLP、Transformer、BERT、XGBoost 等) 已成为不可或缺的利器——它们不仅可从序列和结构特征中自动提取影响相互作用的关键模式, 还能与分子模拟方法结合, 构建“预测-验证-优化”闭环流程。下文将详细阐述各类方法的原理与典型应用, 并对它们的适用场景与性能特点进行综合对比。

4.1 分子对接

分子对接是研究核酸适配体与其靶标分子初步相互作用模式的重要手段, 通过在受体结构中搜索配体的最优结合位点和构象, 为后续动力学模拟与亲和力评估提供可靠的起始模型^[121-122]。本节重点综述四类常用对接软件: AutoDock (包括 Vina 与 AutoDock4)、ZDOCK、MDockPP 与 DOCK。

4.1.1 AutoDock Vina与AutoDock4

AutoDock^[123] 工具包包含两种对接引擎: AutoDock Vina 和 AutoDock4。Vina 使用改良的统计打分函数结合高效的局部搜索算法, 在保证速度

的同时能够准确预测核酸适配体-蛋白质相互作用构象,并输出经验性结合能估计。对于数千至数万候选核酸适配体的初筛,Vina通常作为首选。AutoDock4则采用经典的分项能量模型,将总结合能细分为氢键、疏水效应、离子相互作用及熵项,通过对已知配体-受体复合物的训练数据优化这些能量项的权重,以提高对疏水性或非极性结合口袋的预测精度^[124-125]。在Angiopoietin-2核酸适配体研究中,研究者常先使用Vina快速筛选候选结合构象,再以AutoDock4进行精细重排和打分,以获得更可靠的结合模型^[125-126]。

4.1.2 ZDOCK

ZDOCK^[127]为刚性蛋白-蛋白质(或蛋白质-核酸)对接方法的代表,核心基于快速傅里叶变换(fast Fourier transform, FFT)在三维网格上进行全局搜索,通过形状互补、电荷分布及统计势能三者的加权打分来快速识别潜在结合构象。在多种蛋白质-蛋白质基准测试中,ZDOCK在前1 000个预测中正确率超过70%,显示了其高效的全局扫描能力。尽管其刚性假设无法完全捕捉核酸适配体茎-环结构的柔性变化,但在初步筛选和多靶并行对接场景中,ZDOCK仍然是常用工具^[128]。

4.1.3 MDockPP

MDockPP^[129]在继承FFT全局采样策略的基础上,增加了基于知识的打分与后处理优化流程。该软件首先通过全局搜索生成大量候选结合位点,随后利用经验势能函数对构象进行细化打分与重排,从而改善了单纯统计打分的局限性^[130-131]。在针对前列腺特异性膜抗原(prostate-specific membrane antigen, PSMA)核酸适配体的对接研究中,MDockPP在11个靶点中成功预测了6个结合构象,验证了其多肽与核酸类配体的良好适应性^[132]。

4.1.4 DOCK

DOCK^[133]系列软件最早提出了基于形状匹配的对接策略,将配体分解为若干刚性碎片,通过几何重叠评估其在受体口袋中的初始摆放,然后结合AMBER(assisted model building with energy refinement)力场进行能量最小化与精炼。在细胞色素P450核酸适配体对接研究中,DOCK成功生成了多个候选结合模式,随后经分子动力学与自由能计算筛选出最稳定的结合构象,证明了几何匹配与力场精炼相结合的有效性^[134]。

以上4种对接工具各具特色:AutoDock兼顾速度与细致评分,ZDOCK擅长全局刚性搜索,

MDockPP引入知识驱动优化,而DOCK则在几何匹配与力场精炼间寻求平衡。针对核酸适配体-蛋白质相互作用的不同需求,可灵活选择或组合使用,以获得既高效又准确的结合构象预测。

4.2 分子动力学模拟

在分子对接获得初步结合构象后,分子动力学(molecular dynamics, MD)模拟可用于评估核酸适配体-靶标复合物的稳定性与结合能。MD模拟通过牛顿运动方程精确描述分子系统在溶剂环境下的热力学行为,能够捕捉键长、键角和二面角的微小变化,以及氢键和疏水相互作用的动态演变。结合MM-PBSA(molecular mechanics Poisson-Boltzmann surface area)或MM-GBSA(molecular mechanics generalized born surface area)等后处理方法,可定量计算结合自由能(ΔG),并通过能量分解分析识别关键残基或核苷酸^[135-137]。以下介绍3种在核酸适配体研究中应用最广泛的软件包。

4.2.1 GROMACS

GROMACS^[138](groningen machine for chemical simulations)是一款开源且高效的MD软件,广泛应用于生物分子模拟。它支持多种主流力场(如AMBER、CHARMM、OPLS),并提供从系统构建到模拟分析的全面工具链。在核酸适配体-蛋白质复合物研究中,GROMACS可进行能量最小化、热平衡与生产期模拟,并利用gmx mmpbsa模块计算结合自由能^[137, 139]。由于其出色的并行性能和灵活的脚本接口,GROMACS可在数小时内完成上万原子规模体系的长时程模拟。

4.2.2 AMBER

AMBER^[140](assisted model building with energy refinement)是一套商业化MD软件包,提供专门针对蛋白质和核酸体系优化的力场及模拟工具。AMBER的pmemd模块在多GPU环境下具有卓越的性能,可实现微秒级模拟。对于核酸适配体-蛋白质复合物,AMBER可结合MMPBSA.py脚本进行结合自由能计算,并通过残基能量分解分析关键结合位点。尽管AMBER为闭源软件,但其在准确性和可靠性方面具有良好口碑,被广泛应用于药物与蛋白质-核酸相互作用研究。

4.2.3 NAMD

NAMD^[141](nanoscale molecular dynamics)是一款针对大规模并行计算优化的开源软件,能够高效扩展至数千核CPU或GPU环境。NAMD支持多种力场,集成工具命令语言脚本接口,便于用户定

制复合物体系的模拟方案。在核酸适配体设计领域, NAMD 常用于长时程动力学研究, 结合 CHARMM (chemistry at Harvard macromolecular mechanics) 力场评估结合界面稳定性, 其可扩展性使得亚微米尺度模拟成为可能, 从而揭示结合/解离过程中的大尺度构象变化^[142-143]。

4.3 结合自由能预测

结合自由能 (ΔG) 是衡量核酸适配体与靶标结合亲和力的核心指标, 能够反映相互作用的热力学稳定性。目前主流的预测方法可分为两大类: 基于端点能量估算的 MM-PBSA/MM-GBSA, 以及基于路径积分的严格自由能计算 (如 TI 和 FEP)。

4.3.1 MM-PBSA/MM-GBSA

MM-PBSA 和 MM-GBSA 方法首先利用分子动力学模拟对受体-核酸适配体复合物、单独受体和单独核酸适配体体系分别进行平衡与采样, 然后基于这些模拟轨迹评估复合物在溶剂环境中的总能量变化, 包括不同能量项 (如分子内部相互作用、溶剂化自由能等)^[144-145]。由于该类方法计算效率较高、实现相对简便, 已在多项核酸适配体-蛋白质系统中得到广泛应用, 并能够较好地再现实验测得的结合趋势。然而, 需要注意的是, MM-PBSA/MM-GBSA 在评估过程中并未充分考虑结合过程中的构象重排和溶剂动力学效应, 因而更适用于对候选核酸适配体进行快速筛选和亲和力趋势分析, 而非精确预测绝对结合能水平^[146]。

4.3.2 热力学积分 (TI) 与自由能微扰 (FEP)

TI 和 FEP 方法通过对体系哈密顿量的连续变换, 严格积分得到两种状态 (结合与游离) 之间的自由能差。TI/FEP 通常需要在多个中间 λ 态下进行长时程模拟, 因而计算量大、配置复杂, 但可提供更高精度的 ΔG 预测, 适用于对少量候选进行精细打分的场景。近年来, 有研究将 TI/FEP 与增强采样 (如 Metadynamics) 相结合, 进一步提升了结合自由能计算的收敛性与可靠性^[147-148]。

4.4 核酸适配体与靶标相互作用预测

近年来, 随着 SELEX 产生的大规模序列与实验结合数据的积累, 多种机器学习方法被先后提出以预测核酸适配体与蛋白质的相互作用。如表 5 所示, 早期研究多依赖于手工构建的理化特征, 如氨基酸组成 (amino acid composition, AAC)、伪氨基酸组成 (pseudo amino acid composition, PseAAC)、伪核苷酸组成 (pseudo nucleotide composition, PseNC)、离散余弦变换 (discrete cosine transform, DCT) 和位置特异性得分矩阵 (position-specific scoring matrix, PSSM) 等, 并结合随机森林或支持向量机等经典分类算法, 对核酸适配体-蛋白质相互作用数据预测, 取得了良好的初步效果。随着深度学习的兴起, Emami 等的 AptaNet^[149] 模型将 k -mer 频率与 PseAAC 融入多层感知机, 显著提升了模型在大规模数据集上的表现; 随后, AptaTrans^[150] 利用 Transformer 架构对 k -mer 与频繁子串 (frequent common substring, FCS) 进行端到端特征学习, 进而进一步增强了泛化能力。紧接着, Morsch 等提出的 AptaBERT^[151] 借助预训练的 BERT 模型融合序列与二级结构信息, 将核酸适配体-蛋白质相互作用预测推向了新的高度。近期的 APIPred^[152] 则通过将丰富的手工特征 (包括 k -mer、反向互补 k -mer、AAC、PseAAC) 输入 XGBoost, 并结合分子对接与表面等离子共振实验对预测结果进行了验证, 展示了传统特征与现代集成学习相结合的强大潜力。最新的 PAIR^[153] 方法更进一步, 采用大规模预训练语言模型并结合对比学习框架, 在飞速发展的少样本条件下实现了端到端的核酸适配体-蛋白质配对和打分, 无需额外参数即可稳定输出高质量预测结果。总体来看, 核酸适配体-蛋白质预测领域正从传统手工特征与经典算法, 向深度预训练与自监督学习的多模态融合演进, 不断提升预测的准确度与泛化能力。

Table 5 Summary of nucleic acid aptamer-protein interaction prediction studies
表5 核酸适配体-蛋白质相互作用预测工作总结

方法来源	数据集规模	特征编码	模型架构	性能
Li等 ^[154] (2014)	训练580正/1 740负; 测试145正/435负	核苷酸组成+PseAAC	随机森林	ACC 0.774; MCC 0.372
Zhang等 ^[155] (2016)	训练580正/1 740负; 测试145正/435负	PseKNC+DCT+PSSM+无序区信息	集成随机森林	ACC 0.719; MCC 0.398
Yang等 ^[156] (2019)	训练580正/1 740负; 测试145正/435负	NAC+PseKNC+自动编码器特征	支持向量机	ACC 0.757; MCC 0.478
PPAI ^[157] (2020)	训练56正/168负; 测试143正/421负	传统特征+SMOTE算法	自适应增强和随机森林	ACC 0.806; MCC 0.555
AptaNet ^[149] (2021)	850正/2 554负	k -mer/rev- k -mer+AAC+PseAAC	MLP	ACC 0.914; MCC 0.824

续表5

方法来源	数据集规模	特征编码	模型架构	性能
AptaTrans ^[150] (2023)	训练580正/1 740负; 测试145正/435负	<i>k</i> -mer+FCS(频繁子串)	Transformer	<i>ACC</i> 0.876; <i>MCC</i> 0.679
APIPred ^[152] (2023)	850正/2 554负	<i>k</i> -mer/rev- <i>k</i> -mer+AAC+PseAAC	XGBoost	<i>ACC</i> 0.965; <i>MCC</i> 0.790
AptaBERT ^[151] (2023)	4 950样本	WordPiece Token+结构预训练	BERT Encoder+分类头	<i>ACC</i> 0.910; <i>AUC</i> 0.960
PAIR ^[153] (2024)	训练56正/168负; 测试143正/421负	ESM2+RNA-FM	对比学习+交叉注意力	<i>ACC</i> 0.931; <i>MCC</i> 0.704

ACC: 准确率 (accuracy); *MCC*: 马修斯相关系数 (Matthews correlation coefficient); *AAC*: 氨基酸组成 (amino acid composition); *PseAAC*: 伪氨基酸组成 (pseudo amino acid composition); *PseKNC*: 伪核苷酸组成 (pseudo k-tuple nucleotide composition); *DCT*: 离散余弦变换 (discrete cosine transform); *PSSM*: 位置特异性得分矩阵 (position-specific scoring matrix); *NAC*: 核苷酸组成 (nucleotide acid composition); *SMOTE*: 合成少数类过采样技术 (synthetic minority oversampling technique); *k*-mer: 长度为*k*的核苷酸序列片段; rev-*k*-mer: 反向互补的*k*-mer序列片段; *FCS*: 频繁子串 (frequent common substring); *MLP*: 多层感知机 (multilayer perceptron); *XGBoost*: 极端梯度提升树 (eXtreme gradient boosting); *BERT*: 基于Transformer的双向编码器表示模型 (bidirectional encoder representations from Transformer); *ESM2*、*RNA-FM*分别为蛋白质和RNA序列的预训练语言模型。

目前,几乎所有成熟的机器学习和深度学习框架(如AptaNet、AptaTrans、AptaBERT、APIPred等)均集中于核酸适配体-蛋白质相互作用的预测与分析,针对核酸适配体与小分子,或核酸适配体在细胞层面(如细胞表面受体)相互作用的专门机器学习模型尚未见系统报道。当前的小分子结合研究仍主要依赖分子对接、分子动力学和MM-PBSA自由能计算等,而细胞水平的靶标识别则多通过实验筛选与表型分析,尚无针对性的数据集与算法管线可以直接应用。

5 核酸适配体设计与优化的计算策略

随着生物信息学与计算技术的飞速发展,核酸适配体设计与优化已从传统的实验筛选,逐步演变为计算驱动的智能流程。本节从理性设计策略出发,梳理人工智能辅助生成平台,最后探讨高效的计算筛选与迭代优化方法,旨在为科研人员提供一套系统的计算机模拟设计思路。

5.1 理性设计策略

理性设计强调,在已有结构或序列-功能关联基础上,通过精确的序列改造与结构优化提升核酸适配体性能。首先,基于二级结构预测工具和三维建模软件,可定位结合核心的茎-环区域,然后以定点突变或碱基替换的方式加强关键相互作用。例如,在Thrombin核酸适配体研究中,通过互补碱基对替换和环区序列微调,显著提高了结合亲和力与特异性^[158]。其次,序列截短与核心最小化是降低合成成本与提高折叠效率的有效方法。该策略通过梯度截短或基于信息熵的最小化算法,去除不参与结合的冗余序列,往往能在保持结合活性的同时将序列长度缩减20%~40%^[159-160]。最后,结构引导

优化结合了二级与三级结构信息,对核心环区或结合口袋进行局部重建和能量微调,以实现界面高度互补。例如,对血管内皮生长因子(vascular endothelial growth factor, VEGF)核酸适配体进行片段重拼装后,通过分子动力学与能量最小化筛选出热力学最优构象,显著增强了结合稳定性^[161-162]。

5.2 人工智能辅助设计平台

近年来,生成式模型与深度学习技术被广泛引入核酸适配体设计领域,催生了一系列AI辅助平台。DeepSELEX^[163]使用强化学习来对碱基进行编辑从而优化突变序列;Apta-MCTS^[164]则将MCTS与核酸适配体-蛋白质分类器打分相结合,通过“选择-扩展-模拟-回溯”循环,在序列空间中迭代生成多长度、高得分的候选,展示了无需深度网络也能完成端到端aptamer生成的新思路;RaptGen^[165]基于变分自编码器(variational autoencoder, VAE)与谱系隐马尔可夫模型(profile HMM),利用SELEX轮次序列学习潜在空间,并通过HMM解码生成高多样性的候选序列;AptaDesigner(<https://github.com/JuanCRueda/AptaDesign>)则采用对抗生成网络(generative adversarial network, GAN)和条件VAE,将高亲和力和序列库与靶标信息作为条件输入,实现满足特定结合阈值或蛋白靶点需求的端到端序列生成;AptaDiff^[166]引入扩散模型(diffusion model),通过噪声添加与反向去噪的方式,在首轮SELEX数据上生成更具创新性的高亲和力序列;RhoDesign^[37]针对小分子感光RNA aptamer,结合条件VAE与图卷积网络(graph convolutional network, GCN),实现基于功能表型(如荧光响

应)的序列生成;最新的AiDTA^[167]平台更是通过深度强化学习(MCTS+策略-价值网络)在片段对接-组装框架下脱离任何SELEX或结构数据,完成真正从头生成的蛋白结合核酸适配体设计。上述平台共同构建了从“生成-预测-筛选”闭环的AI设计生态,为高效探索序列空间提供了强有力的工具,具体的输出特征和优劣势见表6。

Table 6 Overview of nucleic acid aptamer design assistance platforms
表6 核酸适配体辅助设计平台汇总

模型	核心架构	训练数据来源	生成策略	输出特性	优势	局限
DeepSELEX ^[163] (2020)	DQN/Policy Gradient	SELEX轮次 序列	强化学习碱基级编辑	优化突变序列	自适应决策; 无需潜在空间假设	状态空间大; 收敛慢; 奖励设计复杂
Apta-MCTS ^[164] (2021)	MCTS+API分 类器	数据库蛋白 质-核酸适配 体对	蒙特卡洛树搜索: 选择→ 扩展→模拟→回溯	可变长度从头 生成核酸适配 体序列	不依赖实验数据; 自 适应探索; 生成多样 性强	需高效API分类器; 搜 索空间巨大; 计算开 销大
AptaDesigner (2021)	GAN/cVAE+ 判别/条件 网络	高亲和力 序列库	对抗/条件生成+即时打分	条件核酸适配 体序列	闭环迭代; 可指定蛋 白质或亲和力阈值	训练不稳定; 需大量标 注序列
RaptGen ^[165] (2022)	VAE+profile HMM	SELEX轮次 序列	潜在空间采样+HMM解码	新核酸适配体 序列	学习序列簇结构, 扩 展SELEX覆盖	依赖大量SELEX数据; HMM解码质量有限
RhoDesign ^[37] (2024)	cVAE+GCN	功能实验验证 核酸适配体	功能条件生成	表型响应核酸 适配体	端到端功能表型指导; 结合GCN强化识别	聚焦小分子; 需功能实 验数据
AptaDiff ^[166] (2024)	Diffusion Model	多轮SELEX	噪声添加+反向去噪	高亲和力候选 序列	保持序列多样性; 去 耦合训练与采样	扩散模型训练对计算资 源需求高
AiDTA ^[167] (2025)	MCTS+ Policy-Value Network	无需SELEX/ 结构数据	片段对接→MCTS组装 +RL策略指导	从头生成蛋白 质结合序列	完全从头生成; 多靶 标验证高成功率	模型训练与模拟开销 大; 依赖高质量片段库

DQN: 深度Q网络 (deep Q-network); Policy Gradient: 策略梯度方法; MCTS: 蒙特卡洛树搜索 (Monte Carlo tree search), 一种基于随机模拟的决策算法; API分类器: 核酸适配体-蛋白质相互作用用分类器 (aptamer-protein interaction classifier); GAN: 生成对抗网络 (generative adversarial network); cVAE: 条件变分自编码器 (conditional variational autoencoder); VAE: 变分自编码器 (variational autoencoder); profile HMM: 谱集成隐马尔可夫模型 (profile hidden Markov model); GCN: 图卷积网络 (graph convolutional network); Diffusion Model: 扩散生成模型; Policy-Value Network: 策略-价值网络; RL: 强化学习 (reinforcement learning); SELEX: 指数富集配体系统进化技术 (systematic evolution of ligands by exponential enrichment)。

5.3 计算辅助筛选与迭代优化

在生成和初始设计之后, 高效筛选与迭代优化是获取最终高性能核酸适配体的关键环节。遗传算法 (genetic algorithm, GA)^[168-169] 作为最早应用于核酸适配体优化的进化策略, 通过将候选序列视作种群, 依次进行交叉、变异与适应度评估 (基于结合预测模型或结构稳定性), 在多代进化中不断提升整体亲和力水平; 模拟退火 (simulated annealing, SA)^[170-171] 则通过随机扰动序列并依据Metropolis 准则接受或拒绝突变, 能够跳出局部最优并在更广阔的序列空间中搜索; 多目标优化方法^[172-173] (如NSGA-II) 则在结合亲和力、特异性、折叠稳定性及合成成本等多个指标之间建立帕累托 (Pareto) 优化, 生成一系列最优折中候选, 为实验验证提供多样选择。此外, 将强化学习与 GA、

SA 等方法结合, 能够在迭代过程中自适应地调整搜索策略, 例如根据历史得分动态调整变异概率或交叉策略^[167]。通过与第4节中介绍的相互作用预测模型无缝衔接, 上述计算优化策略可在生成数万级候选后, 迅速筛选并优化至数十至百级的高亲和力序列, 显著缩短实验筛选周期。

6 核酸适配体的应用领域与实例

近年来, 核酸适配体在多种应用领域取得了显著进展。首先, 在医学诊断与治疗方面, 研究者利用细胞内吞型适配体-药物偶联物 (aptamer-drug conjugates, ApDCs) 实现了对肿瘤细胞的高效靶向输送, 并在体内外实验证明其具有低全身毒性及显著的治疗增效^[174]; 同时, 多种适配体-siRNA、适配体-反义寡核苷酸和适配体-CRISPR/Cas9 平台

也已在动物模型中成功递送核酸药物,大幅提升基因敲低和基因编辑的效率与特异性^[175]。此外,将核酸适配体与荧光染料、放射性同位素或纳米材料结合,可构建用于肿瘤早期诊断与影像学检测的多模态探针^[176-177],例如标记荧光的核酸适配体探针能够在手术中实现实时成像,帮助医师精确区分肿瘤与正常组织^[178];而在靶向治疗方面,核酸适配体对表皮生长因子受体(epidermal growth factor receptor, EGFR)、人表皮生长因子受体2(human epidermal growth factor receptor 2, HER2)、PSMA等肿瘤表面标志物的高亲和识别,使其成为精准递送小分子药物和核酸药物的理想载体^[179-180],其中以靶向核仁素的DNA核酸适配体AS1411在急性髓系白血病和肾细胞癌临床试验中取得了良好疗效^[181],展示了核酸适配体在精准诊疗领域的广阔应用前景。

在生物传感器开发领域,核酸适配体因其小尺寸和易功能化的特点,与纳米材料、微流控和电化学检测技术深度融合,催生了多款高灵敏、低成本、可穿戴或植入式的传感器。譬如,将金纳米粒子(AuNP)与石墨烯场效应晶体管(G-FET)相结合,通过适配体-AuNP复合物在G-FET表面形成信号增强层,实现对心肌肌钙蛋白I、甲胎蛋白(alpha-fetoprotein, AFP)等肿瘤标志物的皮摩尔(pmol/L)级快速检测^[182-185]。此外,研究团队还构建了基于体外转录放大电路的核酸适配体传感器,当目标结合后触发T7 RNA聚合酶介导的RNA聚合放大,产生可通过荧光或颜色比色法检测的RNA产物,实现了对小分子毒素(如黄曲霉毒素、四环素)和重金属离子(如 Hg^{2+} 、 Pb^{2+})的即时检测^[186-187]。最新进展包括柔性基底上的可穿戴核酸适配体传感器(aptasensor),用于汗液中乳酸、葡萄糖等代谢物的动态监测,为个性化健康管理提供了新工具^[188-190]。

在农业与食品安全领域,核酸适配体传感器同样展现出广阔应用前景。针对有机磷农药(如甲基对硫磷、马拉硫磷)和真菌毒素(黄曲霉毒素B1、赭曲霉毒素A)的检测,研究人员开发了分裂适配体-纳米金光学传感平台,通过“双臂”核酸适配体与靶标结合后重组产生可见光信号,检测限达 10^{-9} 级别,且在果蔬和乳制品等复杂基质中表现出良好的抗干扰能力^[191-192]。同时,利用激光刻蚀石墨烯(laser-induced graphene, LIG)制备的电化学适配体感受器,在环境水样或食品表面对沙门氏

菌、李斯特菌等病原菌的检测下限可低至 10^5 CFU/L,响应时间不超过15 min,为现场快速筛查与在线监控提供了可靠途径^[193-195]。更进一步的研究将核酸适配体结合微流控技术,实现了自动化、高通量的样本处理与检测,为农业生产与食品监管带来了变革性工具^[196-198]。

7 总结与展望

核酸适配体作为一类兼具高亲和力、可设计性与多靶标识别能力的寡核苷酸/核酸类分子,已在数据库建设、高通量 SELEX 数据处理、序列与结构特征挖掘、二级/三级结构预测、靶标相互作用建模以及生成式设计与优化等多个层面形成了较为系统的计算与生物信息学研究体系。本文回顾并比较了现有数据库资源与数据整合挑战,梳理了从原始 SELEX 序列到候选核酸适配体筛选的虚拟筛选流程,评述了基于传统机器学习与深度预训练的核酸适配体-靶标相互作用预测方法,分析了分子对接、分子动力学与结合自由能计算在结合构象与亲和力评估中的角色,并重点总结了以VAE、GAN、Diffusion Model、MCTS等为代表的AI驱动生成与优化平台在核酸适配体设计中的进展与潜力。通过多模态、多方法的融合与闭环迭代,当前研究已初步构建起“数据-预测-生成-验证”的智能化核酸适配体发现与设计流水线,为核酸适配体从发现走向精细化应用奠定了理论与方法基础。

尽管前文系统回顾了核酸适配体在生物信息学与AI驱动下的多维进展,但要真正推动其向临床与产业化转化,仍面临若干结构性与实践性障碍,主要可以归纳为以下5个方面:

首先,核酸适配体相关数据散落于多个分离数据库,缺乏统一的格式与注释标准,序列、二级结构、三维折叠、靶标信息、富集轮次、结合实验条件及验证结果等元数据互不兼容,极大地增加了跨平台整合、批量处理与下游模型训练的难度。现有数据库更新频率、质量控制与元数据完整性参差不齐,导致基础训练集存在偏差、噪声与信息缺失,从而影响模型的泛化能力与鲁棒性。与此同时,高质量的二级和三级结构标注极度稀缺,缺乏核酸适配体特异的结构预测基准或金标准数据,使得基于结构的表征、结合部位识别和功能关联建模难以精确开展,也制约了像Transformer、BERT这类预训练大模型在核酸适配体序列/结构领域的迁移与微调。

其次, 从计算预测到实验验证之间存在明显的断层。模型推荐的候选核酸适配体需经过多轮体外/体内实验(合成与纯化、亲和力/动力学测定如表面等离子体共振、等温滴定量热法、结合特异性、细胞递送、生物分布、免疫原性等)才能确认功能, 但当前缺乏统一的评价指标体系与标准操作规程, 不同研究在实验条件、缓冲体系和数据处理上存在较大差异, 使得验证结果难以横向比较, 阻碍构建社区公认的高质量基准集, 并削弱了模型反馈用于优化的可信度与效率。

再者, 当前计算设计与实验验证之间普遍缺乏高效联动机制。候选核酸适配体从模型输出到合成、表征再到反馈调整通常依赖分散的手工操作、多平台格式转换和人工决策, 导致迭代周期长、反馈滞后、资源利用效率低下, 难以形成“设计-验证-更新”的快速闭环, 妨碍了基于数据驱动的核酸适配体优化策略的加速推进。

此外, 当前的生成式设计预测模型(如VAE、GAN、Diffusion Model、MCTS等)多为“黑箱”结构, 在数据稀缺、少样本或新颖靶点场景下表现出收敛不稳、模式坍塌或过拟合的风险。缺乏可解释机制使得研究者难以追溯某一候选序列为何被模型生成、其关键决定因子何在, 进而限制了基于模型建议的定向突变、优化路径设计与实验优先排序的效率。模型训练与推理往往依赖大量计算资源, 在资源受限的实验室或产业环境中难以大规模部署, 同时不同工具之间在输入/输出、超参数配置、打分标准等方面缺乏一致性, 进一步增加了实际应用的工程复杂度和再现性挑战。

最后, 核酸适配体从计算发现到实际应用还面临生物学本身的不确定性与转化障碍, 包括体内稳定性差、非特异性结合、免疫应答、药代动力学行为复杂等, 这些因素往往在早期计算筛选中难以充分建模, 同时, 核酸适配体候选的安全性、批量合成一致性与监管合规性, 也需要与生物信息预测结果紧密结合, 共同构建“预测-验证-质量保证”闭环体系。

为解决上述问题, 未来应从5个相互支撑的方向同步推进, 形成系统性策略以突破这些瓶颈:

第一, 应建立一个开放、标准化、可扩展的多模态核酸适配体大数据平台。该平台不仅要整合来自高通量 SELEX、二/三级结构组学、靶标结合实验、表型响应乃至临床前验证的数据, 还应制定统一的元数据注释规范与交换格式(如统一的标准、

质量评分体系和应用程序编程接口), 实现数据的互操作、版本追踪与社区共建。为缓解数据稀缺与异构性, 可引入迁移学习、自监督/弱监督学习、联邦学习和数据增强技术, 使分散在不同实验室与机构的数据均能在保障隐私与产权的前提下为模型训练和候选生成提供价值。

第二, 应建立标准化的核酸适配体验证与评价体系, 包括统一的亲和力/特异性测定协议、数据格式与质量控制标准、可比的实验对照设计, 以及共享的多中心基准数据集。此类规范化输出不仅提升跨组结果的可比性, 还为模型训练时的监督信号提供一致性基础, 并使得后续的模型更新可以基于定量、可追踪的反馈进行。

第三, 需开发真正端到端的自动化闭环 workflow, 实现计算设计与实验验证的实时耦合。通过在微流控平台上集成快速寡核苷酸合成、高通量亲和力测定(如微阵列表面等离子体共振或单分子捕获)、实时数据采集与反馈机制, 可以实现模型推荐序列的即时合成与功能评估, 并通过主动学习或贝叶斯优化动态选取下一轮候选, 使“设计-验证-更新”成为低延迟、可迭代的闭环过程。引入数字孪生与虚实混合仿真则可在计算空间预筛实验参数, 进一步提升资源利用效率。

第四, 必须提升生成式设计模型的可控性与可靠性。除了继续推进可控扩散、因果强化学习、条件生成网络等新型架构外, 应结合多目标优化策略(例如在亲和力、结构兼容性和脱靶风险之间构建 Pareto 前沿)、软/硬约束机制以及生成路径可追踪技术, 使得设计者可以明确指定和解释生成序列在各个性能维度上的权衡。可解释生成机制将使得研究者更容易理解某一序列被提出的原因, 从而进行理性微调或二次筛选。

第五, 应构建跨学科、跨机构的协同生态, 打通从算法开发、数据标准化、实验验证到临床/产业转化的全链条。具体包括与制药公司、诊断设备厂商、合同开发与制造(contract development and manufacturing organization, CDMO)机构及监管部门合作, 共建多中心标准化评估平台与性能基准, 制定核酸适配体设计与验证的行业指南与认证流程, 并推动符合良好生产规范(good manufacturing practice, GMP)等质量体系的规模化生产和质量追踪。与此同时, 可探索针对计算-实验产出建立可追溯注册与监管科技框架, 降低落地壁垒、提升临床与市场信任度。

通过上述多维协同的系统性建设,核酸适配体研究将从现阶段的各模块优化迈向真正的“智能化系统工程”,实现更高效、可解释、可重复且易于转化的核酸适配体发现与应用路径。

参 考 文 献

- [1] Chen Z, Hu L, Zhang B T, *et al.* Artificial intelligence in aptamer-target binding prediction. *Int J Mol Sci*, 2021, **22**(7): 3605
- [2] Fallah A, Havaei S A, Sedighian H, *et al.* Prediction of aptamer affinity using an artificial intelligence approach. *J Mater Chem B*, 2024, **12**(36): 8825-8842
- [3] He J, Duan Q, Ran C, *et al.* Recent progress of aptamer–drug conjugates in cancer therapy. *Acta Pharm Sin B*, 2023, **13**(4): 1358-1370
- [4] Ji C, Wei J, Zhang L, *et al.* Aptamer-protein interactions: from regulation to biomolecular detection. *Chem Rev*, 2023, **123**(22): 12471-12506
- [5] Piasek A M, Musolf P, Sobiepanek A. Aptamer-based advances in skin cancer research. *Curr Med Chem*, 2023, **30**(8): 953-973
- [6] Zhou G, Latchoumanin O, Hebbard L, *et al.* Aptamers as targeting ligands and therapeutic molecules for overcoming drug resistance in cancers. *Adv Drug Deliv Rev*, 2018, **134**: 107-121
- [7] Uinarni H, Oghenamaro E F, Menon S V, *et al.* Breaking barriers: nucleic acid aptamers in gastrointestinal (GI) cancers therapy. *Cell Biochem Biophys*, 2024, **82**(3): 1763-1776
- [8] Bege M, Ghanem Kattoub R, Borbás A. The 20th anniversary of pegaptanib (Macugen™), the first approved aptamer medicine: history, recent advances and future prospects of aptamers in therapy. *Pharmaceutics*, 2025, **17**(3): 394
- [9] Isaei E, Sobhanipoor M H, Rahimlou M, *et al.* The application of aptamer in tuberculosis diagnosis: a systematic review. *Trop Dis Travel Med Vaccines*, 2024, **10**(1): 25
- [10] Lee M, Lee M, Song Y, *et al.* Recent advances and prospects of nucleic acid therapeutics for anti-cancer therapy. *Molecules*, 2024, **29**(19): 4737
- [11] Sansare Y, Kejamurthy P, Singh S, *et al.* Aptamers as therapeutic targets: prospects and progress in the treatment of cancers. *Nucleosides Nucleotides Nucleic Acids*, 2025: 1-39
- [12] Song K, Hwang S J, Jeon Y, *et al.* The biomedical applications of biomolecule integrated biosensors for cell monitoring. *Int J Mol Sci*, 2024, **25**(12): 6336
- [13] Saito S. SELEX-based DNA aptamer selection: a perspective from the advancement of separation techniques. *Anal Sci*, 2021, **37**(1): 17-26
- [14] Kohlberger M, Gadermaier G. SELEX: critical factors and optimization strategies for successful aptamer selection. *Biotechnol Appl Biochem*, 2022, **69**(5): 1771-1792
- [15] Mukherjee S, Murata A, Ishida R, *et al.* HT-SELEX-based identification of binding pre-miRNA hairpin-motif for small molecules. *Mol Ther Nucleic Acids*, 2022, **27**: 165-174
- [16] de Martin X, Oliva B, Santpere G. Recruitment of homodimeric proneural factors by conserved CAT-CAT E-boxes drives major epigenetic reconfiguration in cortical neurogenesis. *Nucleic Acids Res*, 2024, **52**(21): 12895-12917
- [17] Shareef S, Hariprasad P. Gold nanoparticle-assisted, label-free SELEX coupled with high-throughput NGS for generating highly sensitive and specific DNA aptamers targeting Aflatoxin B(1). *Anal Chim Acta*, 2025, **1364**: 344201
- [18] Wong K Y, Wong M S, Lee J H, *et al.* From cell-SELEX to tissue-SELEX for targeted drug delivery and aptamer nanomedicine. *Adv Drug Deliv Rev*, 2025, **224**: 115646
- [19] Rasouli Jazi H R, Zeinoddini M, Arab S S. A novel in silico SELEX method to screen and identify aptamers against vibrio cholerae. *Curr Comput Aided Drug Des*, 2023, **19**(6): 416-424
- [20] Shekar P V, Kumar A, Mulgaonkar N, *et al.* Aptamer development for SARS-CoV-2 and omicron variants using the spike protein receptor binding domain as a potential diagnostic tool and therapeutic agent. *Biomolecules*, 2025, **15**(6): 805
- [21] Chatterjee O, Kaur G A, Shukla N, *et al.* Multifaceted arsenal in SELEX nanomedicine. *Adv Colloid Interface Sci*, 2025, **342**: 103540
- [22] He M, Wang Z, Wu X, *et al.* Functional SELEX and biomedical applications of aptamers: beyond molecular recognition. *Angew Chem Int Ed*, 2025, **64**(26): e202424687
- [23] Chinchilla-Cárdenas D J, Cruz-Méndez J S, Petano-Duque J M, *et al.* Current developments of SELEX technologies and prospects in the aptamer selection with clinical applications. *J Genet Eng Biotechnol*, 2024, **22**(3): 100400
- [24] Hu Y, Jiang G, Wen Y, *et al.* Selection of aptamers targeting small molecules by capillary electrophoresis: advances, challenges, and prospects. *Biotechnol Adv*, 2025, **78**: 108491
- [25] Yang G, Liu W, Zhao Y, *et al.* Induction of binding sites for RecA aptamers by differentiated-competition capillary Electrophoresis-SELEX. *Talanta*, 2024, **267**: 125213
- [26] Gao S, Zheng X, Jiao B, *et al.* Post-SELEX optimization of aptamers. *Anal Bioanal Chem*, 2016, **408**(17): 4567-4573
- [27] Zhu C, Feng Z, Qin H, *et al.* Recent progress of SELEX methods for screening nucleic acid aptamers. *Talanta*, 2024, **266**(Pt 1): 124998
- [28] Li T, Wang J, Zhu L, *et al.* Advanced screening and tailoring strategies of pesticide aptamer for constructing biosensor. *Crit Rev Food Sci Nutr*, 2023, **63**(31): 10974-10994
- [29] Singh S, Chowdhury P, Ghosh A, *et al.* Virtual screening of truncated single stranded DNA aptamers for Staphylococcal enterotoxin type A. *J Biomol Struct Dyn*, 2023, **41**(21): 11862-11871
- [30] Ropii B, Bethasari M, Anshori I, *et al.* The assessment of molecular dynamics results of three-dimensional RNA aptamer structure prediction. *PLoS One*, 2023, **18**(7): e0288684
- [31] Xu N, Wang Z, Xu Z, *et al.* Experimentally verified flexible molecular docking and dynamic simulation of aptamer with intracellular proteins based on direct DNA 3D structure prediction. *Int J Biol Macromol*, 2025, **316**(Pt 1): 144318

- [32] Ferrè F, Colantoni A, Helmer-Citterich M. Revealing protein-lncRNA interaction. *Brief Bioinform*, 2016, **17**(1): 106-116
- [33] Buglak A A, Samokhvalov A V, Zherdev A V, *et al.* Methods and applications of in silico aptamer design and modeling. *Int J Mol Sci*, 2020, **21**(22): 8420
- [34] Uwiragiye E, Rhinehardt K L. TFIDF-random forest: prediction of aptamer-protein interacting pairs. *IEEE/ACM Trans Comput Biol Bioinform*, 2022, **19**(5): 3032-3037
- [35] Andrianova M, Kuznetsov A. Logic gates based on DNA aptamers. *Pharmaceuticals (Basel)*, 2020, **13**(11): 417
- [36] Adachi T, Nakamura S, Michishita A, *et al.* RaptGen-assisted generation of an RNA/DNA hybrid aptamer against SARS-CoV-2 spike protein. *Biochemistry*, 2024, **63**(7): 906-912
- [37] Wong F, He D, Krishnan A, *et al.* Deep generative design of RNA aptamers using structural predictions. *Nat Comput Sci*, 2024, **4**(11): 829-839
- [38] Kinghorn A, Fraser L, Liang S, *et al.* Aptamer bioinformatics. *Int J Mol Sci*, 2017, **18**(12): 2516
- [39] Lee S J, Cho J, Lee B H, *et al.* Design and prediction of aptamers assisted by in silico methods. *Biomedicines*, 2023, **11**(2): 356
- [40] Ponomarenko J V, Orlova G V, Ponomarenko M P, *et al.* SELEX_DB: an activated database on selected randomized DNA/RNA sequences addressed to genomic sequence annotation. *Nucleic Acids Res*, 2000, **28**(1): 205-208
- [41] Jagannathan V, Roulet E, Delorenzi M, *et al.* HTPSELEX—a database of high-throughput SELEX libraries for transcription factor binding sites. *Nucleic Acids Res*, 2006, **34**(Database issue): D90-D94
- [42] Sato R, Suzuki K, Yasuda Y, *et al.* RNAapt3D: RNA aptamer 3D-structural modeling database. *Biophys J*, 2022, **121**(24): 4770-4776
- [43] Lee J F, Hesselberth J R, Meyers L A, *et al.* Aptamer database. *Nucleic Acids Res*, 2004, **32**(Database issue): D95-100
- [44] Thodima V, Pirooznia M, Deng Y. RiboaptDB: a comprehensive database of ribozymes and aptamers. *BMC Bioinformatics*, 2006, **7**(Suppl 2): S6
- [45] Cruz-Toledo J, McKeague M, Zhang X, *et al.* Aptamer Base: a collaborative knowledge base to describe aptamers and SELEX experiments. *Database*, 2012, **2012**: bas006
- [46] Bachu V, Deware L, Kumar A, *et al.* Aptabase: an aptamer database [EB/OL]. Assam, India: Biosensors and Biofuel cell Lab, IIT Guwahati, 2021[2021-10-24]. <https://www.iitg.ac.in/proj/aptabase>
- [47] Chen L, Yu Z, Wu Z, *et al.* AptADB: a comprehensive database integrating aptamer-target interactions. *RNA*, 2024, **30**(3): 189-199
- [48] Askari A, Kota S, Ferrell H, *et al.* UTexas Aptamer Database: the collection and long-term preservation of aptamer sequence information. *Nucleic Acids Res*, 2024, **52**(D1): D351-D359
- [49] Komarova N, Barkova D, Kuznetsov A. Implementation of high-throughput sequencing (HTS) in aptamer selection technology. *Int J Mol Sci*, 2020, **21**(22): 8774
- [50] Rigden D J, Fernández X M. The 2024 Nucleic Acids Research database issue and the online molecular biology database collection. *Nucleic Acids Res*, 2024, **52**(D1): D1-D9
- [51] Sayers E W, Beck J, Bolton E E, *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res*, 2021, **49**(D1): D10-D17
- [52] Shamah S M, Healy J M, Cload S T. Complex target SELEX. *Acc Chem Res*, 2008, **41**(1): 130-138
- [53] Didarian R, Ozbek H K, Ozalp V C, *et al.* Enhanced SELEX platforms for aptamer selection with improved characteristics: a review. *Mol Biotechnol*, 2025, **67**(8): 2962-2977
- [54] Ohuchi S. Cell-SELEX technology. *BioResearch Open Access*, 2012, **1**(6): 265-272
- [55] White R, Rusconi C, Scardino E, *et al.* Generation of species cross-reactive aptamers using “toggle” SELEX. *Mol Ther*, 2001, **4**(6): 567-573
- [56] Chen L, He W, Jiang H, *et al.* In vivo SELEX of bone targeting aptamer in prostate cancer bone metastasis model. *Int J Nanomedicine*, 2019, **14**: 149-159
- [57] Lyu C, Khan I M, Wang Z. Capture-SELEX for aptamer selection: a short review. *Talanta*, 2021, **229**: 122274
- [58] Mosing R K, Bowser M T. Isolating aptamers using capillary electrophoresis-SELEX (CE-SELEX). *Methods Mol Biol*, 2009, **535**: 33-43
- [59] Ouellet E, Foley J H, Conway E M, *et al.* Hi-fi SELEX: a high-fidelity digital-PCR based therapeutic aptamer discovery platform. *Biotechnol Bioeng*, 2015, **112**(8): 1506-1522
- [60] Dembowski S K, Bowser M T. Microfluidic methods for aptamer selection and characterization. *Analyst*, 2017, **143**(1): 21-32
- [61] Hoinka J, Przytycka T. AptAPLEX—a dedicated, multithreaded demultiplexer for HT-SELEX data. *Methods*, 2016, **106**: 82-85
- [62] Hoinka J, Backofen R, Przytycka T M. AptASUITE: a full-featured bioinformatics framework for the comprehensive analysis of aptamers from HT-SELEX experiments. *Mol Ther Nucleic Acids*, 2018, **11**: 515-517
- [63] Hoinka J, Przytycka T M. The bioinformatics of aptamers: HT-SELEX analysis with AptASUITE. *Methods Mol Biol*, 2023, **2570**: 73-83
- [64] Hoinka J, Berezhnoy A, Sauna Z E, *et al.* AptCluster – a method to cluster HT-SELEX aptamer pools and lessons from its application//Sharan R. *Research in Computational Molecular Biology*. Cham: Springer, 2014: 115-128
- [65] Alam K K, Chang J L, Burke D H. FASTAptamer: a bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol Ther Nucleic Acids*, 2015, **4**(3): e230
- [66] Kramer S T, Gruenke P R, Alam K K, *et al.* FASTAptamer 2.0: a web tool for combinatorial sequence selections. *Mol Ther Nucleic Acids*, 2022, **29**: 862-870
- [67] 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

- [68] Shieh K R, Kratschmer C, Maier K E, *et al.* AptCompare: optimized *de novo* motif discovery of RNA aptamers via HTS-SELEX. *Bioinformatics*, 2020, **36**(9): 2905-2906
- [69] Ruiz-Ciancio D, Veeramani S, Singh R, *et al.* AptamerRunner: an accessible aptamer structure prediction and clustering algorithm for visualization of selected aptamers. *Mol Ther Nucleic Acids*, 2024, **35**(4): 102358
- [70] Wang Z, Chang D, Sargent E H, *et al.* AptFastZ: an algorithm for the rapid identification of aptamers with defined binding affinities. *Anal Chem*, 2023, **95**(48): 17438-17443
- [71] 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
- [72] Caroli J, Taccioli C, Fuente A D L, *et al.* APTANI: a computational tool to select aptamers through sequence-structure motif analysis of HT-SELEX data. *Bioinformatics*, 2016, **32**(2): 161-164
- [73] 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
- [74] Zhang Y, Wang Y, Gao Y, *et al.* APV-Sankey: a comprehensive toolbox for aptamer screening and visualization. *bioRxiv*, 2025. DOI: 10.1101/2025.02.11.637585
- [75] Avila Santos A P, de Almeida B L S, Bonidia R P, *et al.* BioDeepfuse: a hybrid deep learning approach with integrated feature extraction techniques for enhanced non-coding RNA classification. *RNA Biol*, 2024, **21**(1): 410-421
- [76] Wang H, Mennea P D, Chan Y K E, *et al.* A standardized framework for robust fragmentomic feature extraction from cell-free DNA sequencing data. *Genome Biol*, 2025, **26**(1): 141
- [77] Wu S, Xu J, Guo J T. Accurate prediction of nucleic acid binding proteins using protein language model. *Bioinform Adv*, 2025, **5**(1): vbaf008
- [78] Parvez A, Ali S D, Tayara H, *et al.* Stacking based ensemble learning framework for identification of nitrotyrosine sites. *Comput Biol Med*, 2024, **183**: 109200
- [79] Zhao R, Hettich C, Zhang J, *et al.* Excimer energies. *J Phys Chem Lett*, 2023, **14**(12): 2917-2926
- [80] Perovic V, Glisic S, Veljkovic M, *et al.* Novel entropy-based phylogenetic algorithm: a new approach for classifying SARS-CoV-2 variants. *Entropy (Basel)*, 2023, **25**(10): 1463
- [81] Sarapata K, Kania A. Revealing miRNAs patterns by employing matrix representations and energy analysis. *J Mol Graph Model*, 2024, **132**: 108835
- [82] Song M, Zhu Q, Peng J, *et al.* Improving the evaluation of cross efficiencies: a method based on Shannon entropy weight. *Comput Ind Eng*, 2017, **112**: 99-106
- [83] Baez J C. Rényi entropy and free energy. *Entropy (Basel)*, 2022, **24**(5): 706
- [84] Anastasiadis A. Tsallis entropy. *Entropy*, 2012, **14**(2): 174-176
- [85] Grumbach S, Tahi F. A new challenge for compression algorithms: genetic sequences. *InfProcess Manag*, 1994, **30**(6): 875-886
- [86] Khan M, Alteneder M, Reiter W, *et al.* Single-cell and chromatin accessibility profiling reveals regulatory programs of pathogenic Th2 cells in allergic asthma. *Nat Commun*, 2025, **16**(1): 2565
- [87] Kania A, Sarapata K. Multifarious aspects of the chaos game representation and its applications in biological sequence analysis. *Comput Biol Med*, 2022, **151**(PtA): 106243
- [88] Sievers F, Higgins D G. Clustal omega. *Curr Protoc Bioinformatics*, 2014, **48**: 3.13.1-3.13.16
- [89] Katoh K, Kuma K I, Toh H, *et al.* MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*, 2005, **33**(2): 511-518
- [90] Crooks G E, Hon G, Chandonia J M, *et al.* WebLogo: a sequence logo generator. *Genome Res*, 2004, **14**(6): 1188-1190
- [91] Antarsen J, Wellnitz B, Kramer S N, *et al.* Cross-correlation increases sampling in diffusion-based super-resolution optical fluctuation imaging. *Chem Biomed Imaging*, 2024, **2**(9): 640-650
- [92] Zhang W, Lu X, Ren J. Study on drug-mediated protein-protein interaction in single living cells by fluorescence cross-correlation spectroscopy. *Analyst*, 2025, **150**(10): 2029-2038
- [93] Chen Y. An analytical process of spatial autocorrelation functions based on Moran's index. *PLoS One*, 2021, **16**(4): e0249589
- [94] Liang Y, Liu S, Zhang S. Geary autocorrelation and DCCA coefficient: Application to predict apoptosis protein subcellular localization via PSSM. *Phys A Stat Mech Appl*, 2017, **467**: 296-306
- [95] Tahir M, Hayat M, Khan S A. iNuc-ext-PseTNC: an efficient ensemble model for identification of nucleosome positioning by extending the concept of Chou's PseAAC to pseudo-tri-nucleotide composition. *Mol Genet Genomics*, 2019, **294**(1): 199-210
- [96] Zhao S, Pan Q, Zou Q, *et al.* Identifying and classifying enhancers by dinucleotide-based auto-cross covariance and attention-based Bi-LSTM. *Comput Math Methods Med*, 2022, **2022**: 7518779
- [97] Zou H, Yang F, Yin Z. iDHS-DT: identifying DNase I hypersensitive sites by integrating DNA dinucleotide and trinucleotide information. *Biophys Chem*, 2022, **281**: 106717
- [98] Langdon W B, Petke J, Lorenz R. Evolving better RNAfold structure prediction//Castelli M, Sekanina L, Zhang M, *et al.* Genetic Programming. Cham: Springer International Publishing, 2018:220-236
- [99] Lorenz R, Bernhart S H, Höner Zu Siederdisen C, *et al.* ViennaRNA package 2.0. *Algorithms Mol Biol*, 2011, **6**: 26
- [100] Gruber A R, Bernhart S H, Lorenz R. The ViennaRNA web services. *Methods Mol Biol*, 2015, **1269**: 307-326
- [101] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, 2003, **31**(13): 3406-3415
- [102] Markham N R, Zuker M. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol*, 2008, **453**: 3-31
- [103] Do C B, Woods D A, Batzoglou S. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, 2006, **22**(14): e90-8

- [104] Nawrocki E P, Eddy S R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, 2013, **29**(22): 2933-2935
- [105] Singh J, Hanson J, Paliwal K, *et al.* RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat Commun*, 2019, **10**(1): 5407
- [106] Gotesman M. ML & RNA Folding: Extending The E2Efold Model [D]. New York: Yeshiva University, 2024
- [107] Sato K, Akiyama M, Sakakibara Y. RNA secondary structure prediction using deep learning with thermodynamic integration. *Nat Commun*, 2021, **12**(1): 941
- [108] Fu L, Cao Y, Wu J, *et al.* Ufold: fast and accurate RNA secondary structure prediction with deep learning. *Nucleic Acids Res*, 2022, **50**(3): e14
- [109] Chen J, Hu Z, Sun S, *et al.* Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. *arXiv*, 2022. <https://doi.org/10.48550/arXiv.2204.00300>
- [110] Wang N, Bian J, Li Y, *et al.* Multi-purpose RNA language modelling with motif-aware pretraining and type-guided fine-tuning. *Nat Mach Intell*, 2024, **6**(5): 548-557
- [111] Penić R J, Vlašić T, Huber R G, *et al.* RiNALMo: general-purpose RNA language models can generalize well on structure prediction tasks. *Nat Commun*, 2025, **16**(1): 5671
- [112] Gong T, Ju F, Bu D. Accurate prediction of RNA secondary structure including pseudoknots through solving minimum-cost flow with learned potentials. *Commun Biol*, 2024, **7**(1): 297
- [113] Watkins A M, Rangan R, Das R. FARFAR2: improved *de novo* Rosetta prediction of complex global RNA folds. *Structure*, 2020, **28**(8): 963-976.e6
- [114] Zhang Y, Wang J, Xiao Y. 3dRNA: 3D structure prediction from linear to circular RNAs. *J Mol Biol*, 2022, **434**(11): 167452
- [115] Boniecki M J, Lach G, Dawson W K, *et al.* SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res*, 2016, **44**(7): e63
- [116] Antczak M, Popenda M, Zok T, *et al.* New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. *Acta Biochimica Polonica*, 2016, **63**(4): 737-744
- [117] Rother M, Rother K, Puton T, *et al.* ModerNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res*, 2011, **39**(10): 4007-4022
- [118] Baek M, McHugh R, Anishchenko I, *et al.* Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA. *Nat Methods*, 2024, **21**(1): 117-121
- [119] McDonnell R T, Henderson A N, Elcock A H. Structure prediction of large RNAs with AlphaFold3 highlights its capabilities and limitations. *J Mol Biol*, 2024, **436**(22): 168816
- [120] Abramson J, Adler J, Dunger J, *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 2024, **630**(8016): 493-500
- [121] Ferreira L G, Dos Santos R N, Oliva G, *et al.* Molecular docking and structure-based drug design strategies. *Molecules*, 2015, **20**(7): 13384-13421
- [122] Santos L H S, Ferreira R S, Caffarena E R. Integrating molecular docking and molecular dynamics simulations. *Methods Mol Biol*, 2019, **2053**: 13-34
- [123] Trott O, Olson A J. 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
- [124] Mao T, Chen B, Wei W, *et al.* AutoDock and molecular dynamics-based therapeutic potential prediction of flavonoids for primary Sjögren's syndrome. *Heliyon*, 2024, **10**(13): e33860
- [125] Ni B, Wang H, Khalaf H K S, *et al.* AutoDock-SS: AutoDock for multiconformational ligand-based virtual screening. *J Chem Inf Model*, 2024, **64**(9): 3779-3789
- [126] Eberhardt J, Santos-Martins D, Tillack A F, *et al.* AutoDock vina 1.2.0: new docking methods, expanded force field, and Python bindings. *J Chem Inf Model*, 2021, **61**(8): 3891-3898
- [127] Pierce B G, Wiehe K, Hwang H, *et al.* ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics*, 2014, **30**(12): 1771-1773
- [128] Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins*, 2003, **52**(1): 80-87
- [129] Huang S Y, Zou X. MDockPP: a hierarchical approach for protein-protein docking and its application to CAPRI rounds 15-19. *Proteins*, 2010, **78**(15): 3096-3103
- [130] Duan R, Xu X, Qiu L, *et al.* Performance of hybrid strategies combining MDockPP and AlphaFold2 in CAPRI rounds 47-55. *Proteins*, 2025. DOI: 10.1002/prot.26805
- [131] Huang Y, Chen H, Xie Z, *et al.* Aptamer-mediated modulation of eEF1 enhances salt stress tolerance in rice. *BMC Plant Biol*, 2025, **25**(1): 800
- [132] Duan R, Qiu L, Xu X, *et al.* Performance of human and server prediction in CAPRI rounds 38-45. *Proteins Struct Funct Bioinform*, 2020, **88**(8): 1110-1120
- [133] Kuntz I D, Blaney J M, Oatley S J, *et al.* A geometric approach to macromolecule-ligand interactions. *J Mol Biol*, 1982, **161**(2): 269-288
- [134] Jensen S B, Thodberg S, Parween S, *et al.* Biased cytochrome P450-mediated metabolism via small-molecule ligands binding P450 oxidoreductase. *Nat Commun*, 2021, **12**(1): 2260
- [135] Guo J, Liu H. The applications of molecular dynamics simulation in studying protein structure and dynamics. *Curr Med Chem*, 2024, **31**(20): 2839-2840
- [136] Wang M, Ma A, Wang H, *et al.* Atomic molecular dynamics simulation advances of *de novo*-designed proteins. *Q Rev Biophys*, 2024, **57**: e14
- [137] Yu D, Li H, Liu Y, *et al.* Application of the molecular dynamics simulation GROMACS in food science. *Food Res Int*, 2024, **190**: 114653
- [138] Berendsen H J C, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comput Phys Commun*, 1995, **91**(1/2/3): 43-56

- [139] Chen Y, Yang J. Acceleration of the GROMACS free-energy perturbation calculations on GPUs. *ACS Omega*, 2025, **10**(22): 22858-22873
- [140] Case DA, Cheatham TE, Darden T, *et al.* The Amber biomolecular simulation programs. *J Comput Chem*, 2005, **26**(16): 1668-1688
- [141] Phillips JC, Braun R, Wang W, *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem*, 2005, **26**(16): 1781-1802
- [142] Sookhak Lari K, Davis G B, Kumar A, *et al.* The dynamics of per- and polyfluoroalkyl substances (PFAS) at interfaces in porous media: a computational roadmap from nanoscale molecular dynamics simulation to macroscale modeling. *ACS Omega*, 2024, **9**(5): 5193-5202
- [143] Steinegger K M, Allmendinger L, Sturm S, *et al.* Molecular dynamics simulations elucidate the molecular organization of poly (beta-amino ester) based polyplexes for siRNA delivery. *Nano Lett*, 2024, **24**(49): 15683-15692
- [144] Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov*, 2015, **10**(5): 449-461
- [145] Tuccinardi T. What is the current value of MM/PBSA and MM/GBSA methods in drug discovery?. *Expert Opin Drug Discov*, 2021, **16**(11): 1233-1237
- [146] Zhu Y X, Sheng Y J, Ma Y Q, *et al.* Assessing the performance of screening MM/PBSA in protein-ligand interactions. *J Phys Chem B*, 2022, **126**(8): 1700-1708
- [147] Gouda H, Kuntz I D, Case D A, *et al.* Free energy calculations for theophylline binding to an RNA aptamer: comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers*, 2003, **68**(1): 16-34
- [148] Rasouli A, Pickard F C 4th, Sur S, *et al.* Essential considerations for free energy calculations of RNA-small molecule complexes: lessons from the theophylline-binding RNA aptamer. *J Chem Inf Model*, 2025, **65**(1): 223-239
- [149] Emami N, Ferdousi R. AptaNet as a deep learning approach for aptamer-protein interaction prediction. *Sci Rep*, 2021, **11**(1): 6074
- [150] Shin I, Kang K, Kim J, *et al.* AptaTrans: a deep neural network for predicting aptamer-protein interaction using pretrained encoders. *BMC Bioinformatics*, 2023, **24**(1): 447
- [151] Morsch F, Umasankar I L, Moreta L S, *et al.* AptaBERT: predicting aptamer binding interactions. *bioRxiv*, 2023. DOI: 10.1101/2023.11.24.568626
- [152] Fang Z, Wu Z, Wu X, *et al.* APIPred: an XGBoost-based method for predicting aptamer-protein interactions. *J Chem Inf Model*, 2024, **64**(7): 2290-2301
- [153] Zhang J, Yan Z, Zeng H, *et al.* PAIR: protein-aptamer interaction prediction based on language models and contrastive learning framework//IEEE. 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Lisbon, Portugal: IEEE, 2024: 5426-5432
- [154] Li B Q, Zhang Y C, Huang G H, *et al.* Prediction of aptamer-target interacting pairs with pseudo-amino acid composition. *PLoS One*, 2014, **9**(1): e86729
- [155] Zhang L, Zhang C, Gao R, *et al.* Prediction of aptamer-protein interacting pairs using an ensemble classifier in combination with various protein sequence attributes. *BMC Bioinformatics*, 2016, **17**(1): 225
- [156] Yang Q, Jia C, Li T. Prediction of aptamer-protein interacting pairs based on sparse autoencoder feature extraction and an ensemble classifier. *Math Biosci*, 2019, **311**: 103-108
- [157] Li J, Ma X, Li X, *et al.* PPAI: a web server for predicting protein-aptamer interactions. *BMC Bioinformatics*, 2020, **21**(1): 236
- [158] Deng B, Lin Y, Wang C, *et al.* Aptamer binding assays for proteins: the thrombin example—a review. *Anal Chim Acta*, 2014, **837**: 1-15
- [159] Cox J C, Ellington A D. Automated selection of anti-protein aptamers. *Bioorg Med Chem*, 2001, **9**(10): 2525-2531
- [160] Tombelli S, Minunni M, Mascini M. Analytical applications of aptamers. *Biosens Bioelectron*, 2005, **20**(12): 2424-2434
- [161] Ng E W M, Shima D T, Calias P, *et al.* Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nat Rev Drug Discov*, 2006, **5**(2): 123-132
- [162] Zhou P, Zhang S, Li L, *et al.* Targeted degradation of VEGF with bispecific aptamer-based LYTACs ameliorates pathological retinal angiogenesis. *Theranostics*, 2024, **14**(13): 4983-5000
- [163] Asif M, Orenstein Y. DeepSELEX: inferring DNA-binding preferences from HT-SELEX data using multi-class CNNs. *Bioinformatics*, 2020, **36**(Supplement_2): i634-i642
- [164] Lee G, Jang G H, Kang H Y, *et al.* Predicting aptamer sequences that interact with target proteins using an aptamer-protein interaction classifier and a Monte Carlo tree search approach. *PLoS One*, 2021, **16**(6): e0253760
- [165] Iwano N, Adachi T, Aoki K, *et al.* Generative aptamer discovery using RaptGen. *Nat Comput Sci*, 2022, **2**(6): 378-386
- [166] Wang Z, Liu Z, Zhang W, *et al.* AptaDiff: *de novo* design and optimization of aptamers based on diffusion models. *Brief Bioinform*, 2024, **25**(6): bbae517
- [167] Guo G, Guo L, Qian J, *et al.* *De novo* design of protein-binding aptamers through deep reinforcement learning assembly of nucleic acid fragments. *bioRxiv*, 2025. DOI: 10.1101/2025.06.01.657174
- [168] Selvam R, Lim I H Y, Lewis J C, *et al.* Selecting antibacterial aptamers against the BamA protein in *Pseudomonas aeruginosa* by incorporating genetic algorithm to optimise computational screening method. *Sci Rep*, 2023, **13**(1): 7582
- [169] Torkamanian-Afshar M, Nematzadeh S, Tabarzad M, *et al.* In silico design of novel aptamers utilizing a hybrid method of machine learning and genetic algorithm. *Mol Divers*, 2021, **25**(3): 1395-1407
- [170] Zhang Y, Cao L, Yue Y, *et al.* A novel coverage optimization strategy based on grey wolf algorithm optimized by simulated annealing for wireless sensor networks. *Comput Intell Neurosci*, 2021, **2021**: 6688408
- [171] Xue Z, Sun C, Zheng W, *et al.* TargetSA: adaptive simulated annealing for target-specific drug design. *Bioinformatics*, 2024, **41**(1): btac730

- [172] Deb K, Pratap A, Agarwal S, *et al.* A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput*, 2002, **6**(2): 182-197
- [173] Taneda A. Multi-objective optimization for RNA design with multiple target secondary structures. *BMC Bioinformatics*, 2015, **16**: 280
- [174] Zahedi A M, Pirouzbakht M, Zanganeh S, *et al.* Aptamer-Drug Conjugates (ApDCs): transformative approaches in targeted cancer therapy and precision oncology. *Int J Pharm*, 2025, **681**: 125902
- [175] Driscoll J, Gondaliya P, Zinn D A, *et al.* Using aptamers for targeted delivery of RNA therapies. *Mol Ther*, 2025, **33**(4): 1344-1367
- [176] Li X, Lv M M, Liu H, *et al.* Nicked RCA-assisted self-assembling MNzyme coupled with fluorogenic aptamer for label-free detection of miRNA. *Sens Actuat B Chem*, 2025, **429**: 137278
- [177] Zhang L, Feng T, Liu Q, *et al.* Engineering thermostable fluorescent DNA aptamer for the isothermal amplification of nucleic acids. *Biosens Bioelectron*, 2025, **273**: 117183
- [178] Bouvier-Müller A, Ducongé F. Application of aptamers for *in vivo* molecular imaging and theranostics. *Adv Drug Deliv Rev*, 2018, **134**: 94-106
- [179] Lee J S, Kim M, Jin H, *et al.* DNA aptamer-conjugated lipid nanoparticle for targeted PTEN mRNA delivery to prostate cancer cells. *Int J Pharm*, 2024, **662**: 124519
- [180] Xiao Y, Pan T, Da W, *et al.* Aptamer-drug conjugates-loaded bacteria for pancreatic cancer synergistic therapy. *Signal Transduct Target Ther*, 2024, **9**(1): 272
- [181] Liu Y, Hu B, Pei X, *et al.* A non-G-quadruplex DNA aptamer targeting NCL for diagnosis and therapy in bladder cancer. *Adv Healthc Mater*, 2023, **12**(20): e2300791
- [182] Chen K, Zhao H, Wang Z, *et al.* Sandwich-type electrochemical aptasensor based on Au-modified conductive octahedral carbon architecture and snowflake-like PtCuNi for the sensitive detection of cardiac troponin I. *Biosens Bioelectron*, 2022, **212**: 114431
- [183] Zhang Y, Figueroa-Miranda G, Lyu Z, *et al.* Monitoring amyloid- β proteins aggregation based on label-free aptasensor. *Sens Actuat B Chem*, 2019, **288**: 535-542
- [184] Chen C, Chen L, Yang Y, *et al.* Aggregation-induced emission luminogen based ELISA for highly sensitive protein detection. *Sens Actuat B Chem*, 2024, **401**: 134961
- [185] Wei L, Zhu D, Cheng Q, *et al.* Aptamer-Based fluorescent DNA biosensor in antibiotics detection. *Food Res Int*, 2024, **179**: 114005
- [186] Zhang C, Dang W, Zhang J, *et al.* Development of a paper-based transcription aptasensor for convenient urinary uric acid self-testing. *Int J Biol Macromol*, 2024, **271**(Pt 1): 132241
- [187] Li Y, Su R, Li H, *et al.* Fluorescent aptasensors: design strategies and applications in analyzing chemical contamination of food. *Anal Chem*, 2022, **94**(1): 193-224
- [188] Rabiee N, Rabiee M. Wearable aptasensors. *Anal Chem*, 2024, **96**(49): 19160-19182
- [189] Abdelfattah M A, Jamali S S, Kashaninejad N, *et al.* Wearable biosensors for health monitoring: advances in graphene-based technologies. *Nanoscale Horiz*, 2025, **10**(8): 1542-1574
- [190] Radi A E, Acero Sánchez J L, Baldrich E, *et al.* Reusable impedimetric aptasensor. *Anal Chem*, 2005, **77**(19): 6320-6323
- [191] Almenhali A Z, Eissa S. Aptamer-based biosensors for the detection of neonicotinoid insecticides in environmental samples: a systematic review. *Talanta*, 2024, **275**: 126190
- [192] Tan X, Yu C, Tang J, *et al.* Progress in nanomaterials-based enzyme and aptamer biosensor for the detection of organophosphorus pesticides. *Crit Rev Anal Chem*, 2024, **54**(2): 247-268
- [193] Cavallaro N, Moreira G, Vanegas D, *et al.* A *Listeria monocytogenes* aptasensor on laser inscribed graphene for food safety monitoring in hydroponic water. *Discov Food*, 2024, **4**(1): 169
- [194] Labib M, Zamay A S, Kolovskaya O S, *et al.* Aptamer-based impedimetric sensor for bacterial typing. *Anal Chem*, 2012, **84**(19): 8114-8117
- [195] Wu W, Fang Z, Zhao S, *et al.* A simple aptamer biosensor for *Salmonellae enteritidis* based on fluorescence-switch signaling graphene oxide. *RSC Adv*, 2014, **4**(42): 22009-22012
- [196] Adampourezare M, Nikzad B, Roshangar L. Application of microfluidic device based on aptamer for the determination of food contamination; recent progress and challenges. *Biotechnol Sustain Mater*, 2025, **2**(1): 7
- [197] Hou Y, Liu Z, Huang H, *et al.* Biosensor-based microfluidic platforms for rapid clinical detection of pathogenic bacteria. *Adv Funct Mater*, 2025, **35**(1): 2411484
- [198] Mi F, Hu C, Wang Y, *et al.* Recent advancements in microfluidic chip biosensor detection of foodborne pathogenic bacteria: a review. *Anal Bioanal Chem*, 2022, **414**(9): 2883-2902

Artificial Intelligence for Nucleic Acid Aptamers: Methods and Applications^{*}

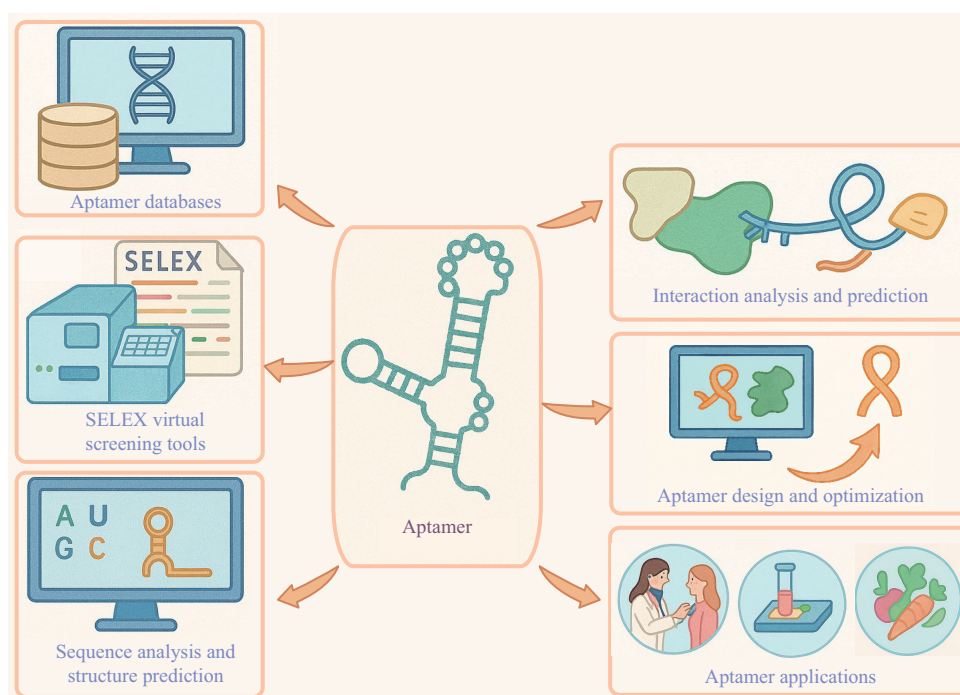
LIU Shang-Hua¹⁾, ZHANG Hong-Qi¹⁾, LIU Ru-Ming¹⁾, ZENG Hong-Juan¹⁾, DENG Ke-Jun¹⁾,
YAN Dan^{2,3)}, TANG Li-Xia¹⁾, LIN Hao¹⁾**

⁽¹⁾School of Life Science and Technology, Center for Informational Biology, University of Electronic Science and Technology of China,
Chengdu 611731, China;

⁽²⁾Department of Pharmacy, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China;

⁽³⁾Beijing Institute of Clinical Pharmacy, Beijing 100035, China)

Graphical abstract



Abstract Nucleic acid aptamers represent a class of single-stranded oligonucleotides capable of high-affinity and specific binding to diverse targets, including proteins, small molecules, cells, and metal ions. Their advantages over antibodies—such as simpler synthesis, lower immunogenicity, superior stability, and easier modification—have positioned them as powerful tools in therapeutics, diagnostics, and biosensing. This review systematically surveys the integral role of bioinformatics and artificial intelligence (AI) in modern aptamer development, spanning from *in silico* selection and structural prediction to the generative design of novel aptamer sequences. The application of high-throughput SELEX (HT-SELEX) has greatly accelerated the discovery of

* This work was supported by a grant from The National Natural Science Foundation of China (82130112).

** Corresponding author.

Tel: 86-13678168394, E-mail: hlin@uestc.edu.cn

Received: July 29, 2025 Accepted: September 11, 2025

aptamers, but also introduced computational challenges in processing large-scale sequencing data. Bioinformatics pipelines now routinely include tools like AptaPLEX and Aptasuite for preprocessing raw reads, including demultiplexing, adapter trimming, and quality filtering. Subsequent analytical steps involve clustering-based tools (e.g., FASTAptamer, AptacLUSTER) to identify enriched sequences, and motif discovery algorithms (such as Aptatrace and MPBind) that uncover conserved sequence-structure patterns associated with binding functionality. These approaches allow researchers to move beyond manual curation and extract meaningful candidates from complex selection rounds. Accurate prediction of secondary and tertiary structures is essential for understanding aptamer function and interaction mechanisms. Conventional tools, including RNAfold and Mfold, employ thermodynamics-based models to predict RNA folding, yet often struggle with pseudoknots and non-canonical pairs. Recent advances in deep learning—exemplified by SPOT-RNA, E2Efold, and Ufold—have significantly improved prediction accuracy by leveraging neural networks trained on large structural datasets. For tertiary structure, methods range from fragment assembly (Rosetta FARFAR2) and homology modeling (RNAComposer) to deep learning-aided approaches such as AlphaFold-RNA and RoseTTAFoldNA. While these tools offer new insights, predicting structures for short, flexible aptamers remains non-trivial. Predicting aptamer-target interactions draws on both physics-based and data-driven approaches. Molecular docking programs—AutoDock Vina, ZDOCK, and MDockPP—provide initial binding poses, which can be refined using molecular dynamics simulations (with GROMACS, AMBER, or NAMD) and free energy perturbation techniques to estimate binding affinity. Complementarily, machine learning models are increasingly employed to predict interactions from sequence and structural features. Early efforts used hand-engineered features with classifiers like SVM and random forest, while contemporary deep learning models (AptaNet, AptabERT, PAIR) utilize pre-trained language models to capture intricate sequence-binding relationships with superior generalization. Perhaps the most transformative development is the use of generative AI for *de novo* aptamer design. Conditional variational autoencoders (e.g., RaptGen), generative adversarial networks (e.g., Aptadesigner), and diffusion models (e.g., Aptadiff) can generate novel aptamer sequences conditioned on target properties or desired binding affinities. Reinforcement learning and evolutionary algorithms, including Monte Carlo tree search (Apta-MCTS) and NSGA-II, support multi-objective optimization toward high specificity, stability, and low immunogenicity. These approaches mark a paradigm shift from selective discovery to intentional design, greatly expanding the functional sequence space. Aptamers designed using these computational strategies are increasingly used across biomedical and environmental applications, including targeted therapeutics, diagnostic biosensors, and food-safety monitoring. Nonetheless, key challenges persist: data scarcity and heterogeneity, model interpretability, and experimental validation bottlenecks. Future progress will depend on standardized data sharing, improved explainable AI, and the integration of computational design with high-throughput experimental screening—ultimately enabling robust, clinically viable aptamer technologies.

Key words nucleic acid aptamer, SELEX data analysis, structure prediction, interaction prediction, generative design, bioinformatics

DOI: 10.3724/j.pibb.2025.0355

CSTR: 32369.14.pibb.20250355