

AIDD 与 CADD 提升药物成功率的思考

陈柏宇, 吕泸楠, 徐小迪, 张 荣, 李 炜*, 付 伟**

(复旦大学药学院, 上海 201203)

摘 要 随着人工智能 (AI) 和计算科学的迅速发展, 特别是人工智能药物设计 (AIDD) 与计算机辅助药物设计 (CADD) 技术的引入, 自然语言处理、图像识别、深度学习和机器学习等多种技术为新药开发提供了革命性的新途径, 大幅提升了研发流程的效率和成功率。在药物发现过程中, AI 技术加速了药物靶点的识别、候选药物的筛选、药理评估及质量检验, 有效降低了研发风险和成本。本文深入探讨 AIDD 和 CADD 技术在药物研发中的应用, 分析它们在提升药物设计成功率和药物研发效率方面的思考与探索, 并探讨这些技术的未来发展趋势及可能面临的挑战。

关键词 人工智能药物设计; 计算机辅助药物设计; 药效团; 分子生成; 基于片段的药物设计

中图分类号 TP181; R914 **文献标志码** A **文章编号** 1000-5048(2024)03-0284-11

doi: [10.11665/j.issn.1000-5048.2024011302](https://doi.org/10.11665/j.issn.1000-5048.2024011302)

引用本文 陈柏宇, 吕泸楠, 徐小迪, 等. AIDD 与 CADD 提升药物成功率的思考 [J]. 中国药科大学学报, 2024, 55(3): 284–294.

Cite this article as: CHEN Baiyu, LYU Lunan, XU Xiaodi, *et al.* Reflections on improving drug success rates with AIDD and CADD[J]. *J China Pharm Univ*, 2024, 55(3): 284–294.

Reflections on improving drug success rates with AIDD and CADD

CHEN Baiyu, LYU Lunan, XU Xiaodi, ZHANG Ying, LI Wei*, FU Wei**

School of Pharmacy, Fudan University, Shanghai 201203, China

Abstract The rapid advancements in artificial intelligence (AI) and computational sciences, particularly through the introduction of artificial intelligence drug design (AIDD) and computer-aided drug design (CADD) technologies, have revolutionized pathways in drug development. These include techniques such as natural language processing, image recognition, deep learning, and machine learning. By employing advanced algorithms and data processing techniques, these technologies have significantly enhanced the efficiency and success rate of R&D processes. In drug discovery, AI technologies have accelerated the identification of drug targets, screening of candidate drugs, pharmacological assessments, and quality control, effectively reducing R&D risks and costs. This article delves into the application of AIDD and CADD in drug development, analyzing their roles in enhancing the success rates and efficiencies of drug design, exploring their future trends, and addressing the potential challenges.

Key words artificial intelligence drug design; computer-aided drug design; pharmacophore; molecular generation; fragment-based drug design

This study was supported by the National Natural Science Foundation of China (No. 82273853, No. 82073765)

药物研发是一个多步骤、高投入、高风险的过程, 包括靶点确证、先导化合物发现及优化、临床试验等阶段。统计显示, 一种新药从研发开始到上市平均需要投入约 5 亿美元及耗时 12 年^[1-2]。显然,

这一传统的药物研发模式, 因其漫长的周期、低成功率和高昂的费用, 已不再适应当前的经济和医疗健康需求。

20 世纪 80 年代, 由 Martin Karplus 等科学家领

收稿日期 2024-01-13 通信作者 *Tel: 13917505901 E-mail: wei-li@fudan.edu.cn

**Tel: 021-50772526 E-mail: wfu@fudan.edu.cn

基金项目 国家自然科学基金项目 (No. 82273853, No. 82073765)

衍的研究团队将计算机辅助药物设计(computer-aided drug design, CADD)技术引入到新药研发领域^[3]。在当时的科技背景下, CADD 经历了高速发展, 显著加速了多款针对 HIV 和青光眼等疾病的药物的研发进程^[4]。1979 年, 世界上首家 CADD 公司 Tripos 成立, 开启了该领域的商业化进程。1990 年, Schrödinger 公司成立, 迅速成为 CADD 及人工智能药物设计(artificial intelligence drug design, AIDD)领域的领头羊。到了 2006 年, Berg 和 Kiromic BioPharm 等公司成立, 利用人工智能技术寻找新的药物靶点和生物标志物。2010 年之后, 以 Insilico、Deep Genomics 等为代表的公司专注于开发 AIDD 技术, 并逐渐兴起。2015 年, 中国的晶泰科技成立, 标志着中国药物研发进入了新的阶段。2020 年, AlphaFold2 的蛋白结构预测技术超越了以往所有方法。2021 年, Insilico 开发出 ISM001-005, 这是全球首个完全由 AI 技术驱动的针对特定间质性肺纤维化(IPF)靶点和药物的发现, 包括靶点发现和验证。这说明在如今, 无论是 CADD 还是 AIDD, 都是药物设计至关重要的一环。

1 CADD 和 AIDD 的基本原理

1.1 CADD

CADD 是一种多学科交叉的综合技术, 基于计算化学和计算生物学的理论与方法, 运用特定的计算机程序来分析、模拟和预测药物分子的结构与性质的关系、药物分子与受体生物大分子之间的相互作用机制, 以及生物大分子之间的识别和结合过程^[5-7]。

通常, CADD 方法分为两大类: 基于配体的药物设计(ligand-based drug design, LBDD)和基于结构的药物设计(structure-based drug design, SBDD)^[8]。LBDD, 亦称间接药物设计, 主要包括定量构效关系(QSAR)、药效团建模和相似性搜索等。而 SBDD, 亦称直接药物设计, 侧重于分子对接、从头药物设计和分子动力学模拟等方法。SBDD 基于受体结构的特征, 以及受体与配体的结合方式, 采用互补配对和诱导契合的原理, 来优化先导化合物的搜索。

1.1.1 基于结构的药物设计 SBDD 广义上是指基于配体结构和受体蛋白结构的药物设计^[8]。狭义上, SBDD 专指基于受体结构的药物设计, 即依据与药物作用的靶点(如酶、受体、离子通道、抗原、核酸、多糖等)的三维结构, 采用分子识别原理(互补性)来设计能调控受体的先导物, 或根据已知药物

的作用力与构效关系推测新化合物的药效, 以此发现活性分子。本文讨论的 SBDD 是指这一狭义上的基于受体结构的药物设计方法。

1894 年, Fisher 提出了“锁钥原理”来解释配体与受体的结合机制, 即配体与受体关系类似于锁与钥匙的配合, 配体能够精确插入受体的空腔中, 从而引发生理作用。然而, 这一理论假设配体与受体均为完全刚性, 忽略了它们的构象变化。实际上, 配体小分子与靶点蛋白均具有一定的柔性, 其构象在生理空间上不断变化。在配体与靶点结合过程中, 除了空间形状的匹配, 还需满足能量匹配, 即结合自由能的变化决定它们是否能成功结合。此外, 配体和蛋白之间会形成氢键、疏水作用、范德华力、静电作用等多种微观作用力。为了更准确地描述这一现象, 1959 年 D.E.Koshland 提出了“诱导契合”学说, 认为蛋白质结合口袋的结构和配体小分子的结构会因相互作用而发生变化。在 SBDD 中, 研究的重点即是配体和受体蛋白的作用模式。

1.1.2 基于配体的药物设计 LBDD 是在生物靶点结构未知的情况下, 通过分析具有特异性结合能力的配体的结构信息, 来发现先导化合物的方法^[8]。相比基于受体结构的药物设计, LBDD 以“从小分子到小分子”的原则进行, 其核心是分子相似性原理, 即结构相似的分子通常具有相似的生物活性。分子相似性搜索主要分为子结构搜索、基于分子指纹的相似性搜索和基于分子形状的相似性搜索。

在 LBDD 中, 描述符是分析分子相似性的基本要素^[9]。这些描述符主要分为 3 类: 一维(1D)描述符, 源于化合物的基本属性; 二维(2D)描述符, 基于分子的图形或结构片段计算得出; 三维(3D)描述符, 涉及分子的形状等特征。此外, 分子指纹技术作为一种分子的抽象表示方法, 将分子编码成一系列向量, 以便进行分子间的比较。分子描述符和分子指纹技术作为成熟的分子表征方法, 在 AIDD 中也得到了广泛应用。

1.1.3 基于片段的药物设计与基于药效团的药物设计 基于药效团的药物设计实际上发展于基于片段的药物设计(fragment-based drug design, FBDD)。FBDD 是一种利用小型化学片段进行药物开发的策略^[10]。这些小片段被用于探测靶点蛋白的潜在结合位点。相比于传统的高通量筛选(high throughput screening, HTS), FBDD 的片段具有更高的结合效率, 允许它们与大分子靶点的特定区域形成更紧密、更具特异性的相互作用^[11]。在 FBDD 中, 初步

识别出能够与靶点结合的小分子片段后,通过多种技术(如表面等离子体共振、质谱、核磁共振)对片段与靶点的结合进行详细分析^[12]。这一信息用于指导片段的化学改造,通过结构优化提高其亲和力、药代动力学特性和生物活性,最终发展成为高效的药物候选物。FBDD的关键优势在于其对化学空间的有效探索,能够发现传统方法可能忽略的创新化合物。此外,它在处理难以靶向的蛋白质(如蛋白-蛋白相互作用)时显示出特别的潜力。然而,FBDD也面临着挑战,比如需对低亲和力片段进行精确的检测和结构改造,以及将小分子片段转化为具有临床潜力的药物^[13–15]。

与传统的基于结构和基于配体的药物设计方法相比,基于药效团的药物设计采取了不同的方法。此方法不仅考虑配体与靶点蛋白的结合和相关片段的基本特性,而且还涉及根据不同的药效特征元素对配体小分子进行分类^[16–17]。药效团中的药效特征元素见表 1。这些药效特征元素在小分子与靶点蛋白的相互作用中具有不同的贡献。

1.2 AIDD

在 AIDD 中,首要步骤是学习并理解输入化合物的结构和特性,以便生成具有预期活性和成药性的新分子^[19–21]。传统的 CADD 方法,包括分子对接、高通量筛选、蛋白质晶体解析和药物临床分析等,提供了大量的丰富数据,这极大地促进了近年来 AIDD 的快速发展^[22]。AIDD 的基本流程首先包括数据获取和处理,涉及对已有药学数据(包括蛋白信息、小分子结构及药效信息等)的收集、清洗和标准化处理,随后是特征工程和表征学习,例如蛋白质和分子的表征^[22]。在 AIDD 过程中,不论是性质预测、分子生成还是反应预测,表征方法的准确性和简洁性都对模型性能有直接影响。确定合适的表征方法后,使用科学合理的算法构建模型是 AIDD 中最关键的环节,也是研究人员的主要研究焦点。目前的 AIDD 研究仍主要集中在计算层面,跨学科合作和实验验证方面还有待完善。

表 1 药效团中药效特征元素^[18]

特征元素	药效团内容
氢键受体	sp ² 或sp ³ 杂化的氧原子/与碳原子以双键形式相连的S原子/与碳原子以双键或者三键相连的氮原子
氢键供体	氢原子以及与之相连的氧原子和氮原子
疏水中心	只要和不带电原子或电负性中心相连的一组连续的碳原子都可以形成疏水中心
电荷中心	与受体形成盐桥或较强的静电相互作用,带有电荷的原子/在生理pH下会发生电离的中性基团
芳环中心	形成 π - π 相互作用,五元或六元芳环

1.2.1 表征学习 适当的表征方法不仅能为模型提供丰富的化学、生物和空间信息,而且能有效地减少计算时间和降低计算负担,从而提高结果的准确性并提升模型效能^[23]。在较早期的研究中,分子表征常采用卷积神经网络(convolutional neural network, CNN)或递归神经网络(recurrent neural network, RNN)^[24]。CNN 主要用于图像信息的处理,核心在于通过内核在图像上进行卷积以提取特征,其关键在于图像结构上的平移不变性:内核无论移至图像的哪一部分,其内部结构保持不变。另一方面,RNN 主要用于处理自然语言等序列数据。无论是图像还是语言数据,它们均属于欧氏空间数据,其特点是结构规则。然而,在现实世界中,许多数据结构是不规则的,最典型的例子是图结构。传统的 CNN 和 RNN 方法在处理此类非欧氏空间数据时往往表现不佳。

目前,在分子表征方法领域,主要分为传统的基于字符串的方法、基于图的方法以及基于图像的方法^[25–26]。其中,基于字符串的表征方法在 CADD 领域已得到发展和成熟,包括简化分子输入线性表达系统(simplified molecular input line entry system, SMILES)、国际化学标识符(international chemical identifier, InChI)和分子指纹等方式。这类字符串表示方法通常由美国信息交换标准代码(ASCII)字符组成,相比其他表征方法具有更紧凑的格式,更易于人类读写。此外,基于字符串的数据可以简单地存储和处理。

1.2.2 分子生成模型 随着“大模型”概念的提出,人工智能领域迅速发展,2023 年 3 月发布的 ChatGPT-4 模型,拥有超过 1 万亿的参数数量,标志着人工智能领域的一个重要里程碑^[25]。

在分子生成模型的应用中,科学家们致力于通过精确的生成模型来设计具有高活性和高成药性的分子。从化学空间上而言,潜在的药物化合物数量将达到 10^{23} ~ 10^{60} ,这为新药研发提供了巨大的潜力和挑战^[8]。药物产生生理作用的关键在于与靶点

蛋白的相互作用,其中配体小分子与靶点蛋白结合后引起特定的构象变化,从而调节生理活动。引入人工智能的新药物设计方法为药学家们开辟了多种新途径。与 CADD 早期相似, AIDD 最初也是在几乎不涉及药学概念的情况下进行模型训练。但自 2019 年引入 AttentiveFP(注意力机制)^[27]后,药学家们认识到仅依赖大量数据的训练无法取得理想效果,逐渐将更多药学概念引入模型中。从这个角度看, CADD 与 AIDD 不仅最终目标相同,在实施方法上也有诸多共同点。现有分子生成模型分类见表 2。

表 2 按照药物设计理念分类的分子生成方法

种类	分子生成方法
经典方法	基于规则的方法
	递归神经网络
	生成对抗网络
	变分自编码器
	强化学习
融入药学思想的生成模型	基于受体结构的生成模型
	基于配体结构的生成模型
	基于药效团的生成模型
	基于片段的生成模型
其他	基于谱学的生成模型
	基于化学反应的生成模型

2 CADD 药物设计方法介绍及案例

2.1 基于蛋白结构的药物设计

2.1.1 靶蛋白结构预测 CADD 方法的流程如下:首先,从已有的数据库出发,利用计算工具和方法,如蛋白结构预测和分子对接,来研究小分子、靶蛋白以及受体-配体复合物。接下来,采用量子化学和分子动力学计算,结合 X 射线晶体学和核磁共振技术,对分子与蛋白的三维结构进行深入分析。最后,通过合理的分子设计过程,发现活性先导化合物^[28]。这一流程不仅涵盖了从理论计算到实验验证的各个方面,还强调了结构信息在药物设计中的重要性。

2.1.2 受体-配体复合物结构预测 在 CADD 中,一个关键任务是确定受体-配体复合物的结构,这通常基于蛋白结构。对于那些已有晶体结构的受体-配体复合物,通过 X 射线晶体学、核磁共振等技术手段直接获得其精确结构是最理想情况^[3]。然而,正如前文所述,许多蛋白结构尚未明确,且在药物

设计过程中,设计的小分子与蛋白的结合情况往往无法通过实验直接确定。因此,科学家们常常采用各种计算方法,如分子对接、分子动力学模拟等,来预测受体-配体复合物的结构。

(1) 分子对接

目前最常用的分子对接工具包括 Autodock^[29]、AutodockVina^[30]、CDOCKER^[31]、GLIDE^[32]、DOCK6^[33]、GOLD^[34]、FLEX^[35] 和 SwissDock^[36]。从方法上而言,分子对接方法包括^[37]:(1)刚性对接,即配体和靶点结构都是刚性的,在对接过程中保持结构不变;(2)半柔性对接,即配体结构是柔性的,靶点是刚性的,这也是目前最常用的方法;(3)柔性对接,即配体和靶点结构都是柔性可变的。对于柔性配体小分子,常用的搜索算法有如系统搜索算法、随机算法或者模拟算法。而对于柔性的靶点蛋白,通常使用的方法有分子动力学方法和蒙特卡洛方法。此外,目前也有许多基于深度学习的方法用于预测配体和特定靶点蛋白质的结合模式,如 DSDP^[38]、Diffdock^[39] 等。这些基于深度学习的方法能够大幅提高虚拟筛选的效率,可以对受体结合位置和配体结合构象与方向进行直接预测。

(2) 分子动力学模拟

分子力学(molecular mechanics, MM)是一种使用经典力学近似处理分子以节省量子力学所需的计算资源的方法^[40]。20 世纪,MM 方法在探索配体-蛋白质结构、结合模式和优化先导化合物等各方面发挥着重要作用。分子力学中一个重要概念即“力场(force field, FF)”^[41],它由分子力学势能函数计算,代表整个过程中不同能量项的综合。MM 势能函数被用于各种采样方法,如分子动力学模拟^[42]以及蒙特卡洛(Monte Carlo)方法^[43]。分子动力学模拟是最经典的计算机辅助药物设计方法之一。作为最流行的采样算法之一,分子动力学模拟利用各种积分算法,如 Verlet 算法、Leap-frog 算法和 Beeman 算法等,来解释经典的牛顿运动方程。分子动力学模拟分析在给定体系中分子与靶点蛋白的轨迹、构象变化与相互作用模式以及体系随时间的变化过程。分子动力学模拟通常会构建一个蛋白体系,最初的蛋白结构由实验所得,之后可以用不同方法对其结构进行建模^[42]。X 射线晶体学作为获取三维蛋白质结构的重要实验方法,它要求蛋白质形成稳定的晶体,同时晶体质量决定了结构的分辨率。这样的特性限制了通过 X 射线晶体学获得高质量蛋白质结构,特别是膜蛋白。为了解

决这个问题,科学家们使用冷冻电镜,通过绕过形成晶体的步骤以直接确定不稳定或难以确证的蛋白结构。然而,样品质量依然是冷冻电镜中决定是否得到高分辨率结构的最关键因素。但在实验模拟过程中,因为各种因素而导致实验结构缺失,则需要对蛋白质结构进行建模或预测。作为传统的计算机辅助方法的同源建模与最近几年兴起的 AlphaFold 都能构建较好的未知蛋白结构^[44]。在分子动力学模拟过程中,系统的原子和分子在固定时间内相互作用,提供系统的动态特征。原子轨迹一般由牛顿运动定律决定,并采用具有不同力场的分子力学方法来计算系统能量。

在分子对接之后,使用分子动力学模拟能够提高蛋白柔性,并获得具有明确结合口袋及高柔性配体的靶点蛋白构象。此外分子动力学模拟可以应用于对接评分及先导化合物优化。通过结合自由能计算,分子动力学模拟能够准确地评估亲和力,提高化合物筛选成功率。在先导化合物优化中,可以对小的化合物集(不超过几百个)进行分子动力学模拟,并以此确定配体与靶点的相互作用,并以此为基础进一步优化先导化合物。

2.2 基于配体结构的药物设计

2.2.1 基于药效团的分子设计 通常药效团的构建包括以下步骤:(1)选择一组针对特定靶点的生物活性配体作为训练集;(2)为训练集中的每个配体创建构象空间以描述配体的构象柔性,提取药效特征元素;(3)联配训练集中的配体,确定化学特征并构建药效团模型;(4)对模型进行分析和验证。对 Log 文件分析(Cost 值分析、训练集匹配)并对测试集进行验证。目前已有的药效团生成方法包括: Catalyst^[45]、LigandScout^[46]、MOE^[47]、PharmMapper^[48]和 PharmaGist^[49]。具有不同骨架但有类似相互作用的配体可以通过基于虚拟筛选方法来进行选择。

尽管经典的药效团模型在药物设计中发挥了重要作用,但它存在一个显著缺陷:它使用静态模型来表示本质上动态的生物系统,且模型中的相互作用被简化为基本的几何特征。为克服这一限制,动态药效团模型的发展结合了传统药效团模型的原理和分子动力学模拟的技术。这种模型能够计算并揭示传统球形几何模型中配体结合的微妙细节,并提供分子轨迹中不同结合模式和特征频率的统计数据。这种方法的优势在于它能够捕捉和描

述配体与受体间的动态相互作用,从而提供更准确和全面的药物设计视角。

2.2.2 定量构效关系 定量构效关系(QSAR)旨在解决配体的生物活性和结构特性^[50]。其中生物活性指的是药代动力学特征,包括 ADMET 和其他特性。结构特性指的是配体的物理化学性质。它被广泛应用于药物发现过程中的先导化合物发现和优化,并被广泛应用于预测新配体的生物活性。一个合理的 QSAR 模型应当满足以下要求^[50]:(1)从传统的试验方案中获得具有生物活性的足量配体(≥ 20 个化合物)的数据集;(2)选择适当的化合物来构建训练集和测试集;(3)配体的描述符之间没有自相关诱发过度预测或过拟合;(4)通过内部/外部验证集验证最终的 QSAR 模型,以保证模型的可靠性。

根据研究方法的不同, QSAR 建模方法^[50]可以分为 6 类:(1)1D-QSAR,将生物活性与全局分子物理化学特征如 LogP 及 pK_a 联系起来,如基于推导描述符的方法,基于维度的 QSAR 方法;(2)2D-QSAR,将生物活性与配体的结构特征联系起来而不考虑三维表征;(3)3D-QSAR,将生物活性与配体周围的非共价相互作用联系起来;(4)4D-QSAR,另外还包含在 3D-QSAR 基础上的配体构象的几何;(5)5D-QSAR,描述了 4D-QSAR 的不同诱导拟合构象;(6)6D-QSAR 进一步结合 5D-QSAR 的不同溶解模型。此外,基于构建配体的生物活性与结构特征之间关系的技术, QSAR 方法被分为线性和非线性方法。线性方法包括线性回归、多元线性回归、偏最小二乘法和主成分分析/回归。非线性方法包括人工神经网络、贝叶斯神经网络等。为了满足药物设计中的 QSAR 要求,人们开发了各种与 QSAR 相关的工具,如 Cloud 3D-QSAR^[51]和 Web-4D-QSAR^[52]。虽然 QSAR 具有一定优势,但也存在一些挑战。例如,高质量数据集的需求使得构建一个可靠的 QSAR 模型变得困难。

2.3 5-HT_{1A} 受体激动剂开发

本课题组采用动态药效团方法构建了 5-HT_{1A} 受体的药效团模型。并通过 4 次虚拟筛选以及人工挑选,从 228401 个分子中筛选出 10 个分子,如表 3 所示,并进行活性测试。最终得到 EC₅₀ 为 7 nmol/L 的高活性小分子 FW01^[53]。后续对该分子进行进一步设计优化,以及定量构效分析,延展得到一系列全新的药物分子,如图 1 所示。

表 3 经筛选得到的 5-HT_{1A} 受体小分子活性数据

Hit	相对分子质量	氢键受体	氢键供体	ALogP	[³ H]-8-OH-DPAT K_i /(nmol/L)	[³⁵ S]GTPγS EC ₅₀ /(nmol/L)
FW01	476.6	5	1	5.84	51.9±16.4	7
FW02	347.5	4	2	3.35	70.5±14.8	77
FW03	350.4	4	1	4.04	96.5±13.1	404
FW04	394.5	5	2	3.77	103.5±23.4	128
FW05	429.0	4	1	6.10	123.6±32.0	534
FW06	294.3	4	3	3.27	133.4±9.7	434
FW07	397.5	4	1	4.60	225.2±20.2	340
FW08	344.8	4	2	4.20	294.9±5.0	572
FW09	446.0	5	1	4.74	320.1±76.9	597
FW10	396.9	4	1	4.27	351.7±19.0	ND
5-羟色胺					1.8±0.1	3

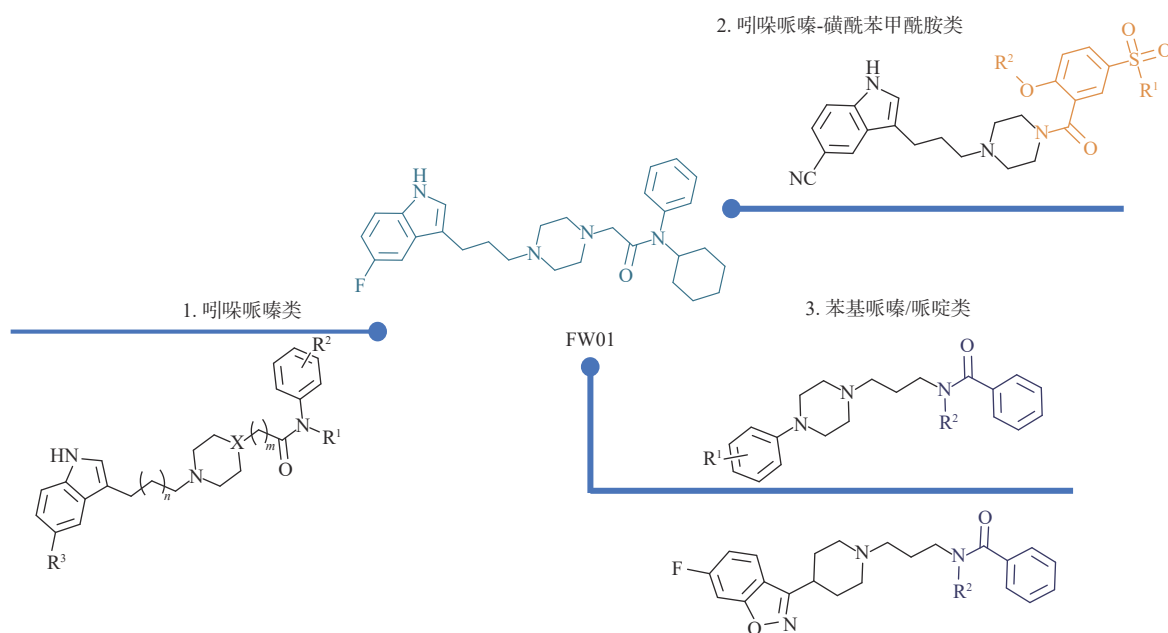


图 1 使用 CADD 方法, 从化合物 FW01 开始发现的创新药物

2.4 CADD 方法小结

CADD 作为传统药物设计方法的代表, 至今已发展成为一套成熟的研究方法和体系。然而, 当前 CADD 方法面临着若干挑战。首先, 计算资源的限制。CADD 常常需要大量计算资源, 尤其是在进行复杂的分子模拟和高通量筛选时, 这限制了某些研究的规模和深入程度。其次, 模型准确性的不足。对于大分子系统, 如蛋白质复合物, 准确的预测依然是一个挑战。再次, 处理复杂生物系统的困难。生物系统的复杂性和多变性增加了使用 CADD 方法进行精确模拟的难度。最后, QSAR 模型的局限性。虽然许多 CADD 方法依赖于 QSAR 模型, 但这些模型在处理非线性和高维数据时存在较大

误差。

3 AIDD 药物生成方法介绍及案例分析

3.1 经典分子生成方法

当人工智能技术最初被引入药物设计领域时, 药学家们主要采用神经网络直接应用于药物分子的生成, 即基于网络的分子生成模型, 依赖于神经网络学习分子的各种属性, 进而生成新的化合物。这种初步尝试标志着人工智能进入药物设计领域, 尽管相对简单, 但为后续更复杂和精细化的方法铺平了道路。使用的模型包括: (1) 递归神经网络 (RNN), 通过内部循环结构来处理序列数据, 常常用于处理如 SMILES 字符串等分子表征。在分子

生成中,其可以从前一层输入的原子和键的信息中预测下一个原子和键。(2)自动编码器(VAE),由编码器和解码器组成。编码器将输入数据(如分子表示)编码为潜在空间,而解码器从潜在空间中解码以生成新的数据。该方法可以学习数据的连续潜在空间。(3)生成对抗网络(GAN),由生成器和判别器组成。生成器生成新的数据,而判别器尝试区分真实数据和生成的数据。该方法可以生成高质量的分子,然而训练过程可能会不稳定。(4)混合模型(hybrid models)通过强化学习对预训练好的模型进行微调,每一步生成的分子后都对参数进行更新。

这些早期的基于人工智能的分子生成模型面临着几个关键问题:首先,它们需要大量的数据来进行有效的训练;其次,生成的分子有时可能缺乏化学合理性;最后,这些模型生成的分子的可合成性和成药性往往较低。尽管存在这些局限性,这些方法代表了药学家们在人工智能辅助药物设计领域的初步探索,为后续开发性能更优的模型奠定了基础。这些初步尝试不仅展示了人工智能在药物设计中的潜力,也为未来的研究提供了重要的经验和启示。

3.2 整合药理学思想的分子生成方法

为克服前述的挑战,药学家们开始专注于提高模型的可解释性。正如之前所提到的,传统药物设计方法通过将配体小分子的活性问题分解为片段或局部问题来提供解决方案。将深度学习应用于这些方法的一个优势是,将问题切分为片段的同时也有助于扩展数据集的规模。由于针对特定靶点的活性药物数据集通常规模有限,除了采用迁移学习等计算方法外,引入片段和药效团等概念本身也能在一定程度上扩大有限的配体数据规模。此外,注意力机制和 Transformer 模型的成功表明,深度学习本身也受益于先验经验的约束和微调。因此,结合药理学理念的分子生成方法应运而生,旨在更加精确和高效地解决药物设计中的挑战。

3.2.1 基于受体结构的分子生成模型 基于受体结构的模型主要以蛋白质的活性位点(口袋)数据为输入,通过在这些口袋中进行分子生长来生成全新的分子结构。这种方法在探索新的化学空间和生长出独特的分子骨架方面具有优势,但同时也面临着挑战:生成的分子有时可能缺乏化学合理性和成药性。这要求对模型进行更加精细的调整和优

化,以确保其生成的分子既新颖又具有实际应用价值。Lingo3DMol 一种采用基于片段的 SMILES 编码的分子生成模型,使用局部及全局坐标预测的新方法,即基于受体活性口袋的 atom based 分子生长方法。它的优点在于能够防止生成不合理结构的分子,但分子生成的活性、可合成性差,分子生成起点选择困难,对非共价相互作用力识别力差。AlphaDrug 采用了 Lmsr Transformer,即通过增加从蛋白质编码器到配体解码器的跳过连接来更有效地学习蛋白质和分子之间的关系。这使得模型能够更好地捕获蛋白质目标的信息。该模型结合了分子生成的生长和连接过程,但也因此导致模型计算效率低下等问题^[54–55]。

3.2.2 基于配体结构的分子生成模型 基于配体结构的分子生成模型主要以配体的结构或序列信息作为输入,依据结构-活性关系进行分子设计。这种方法在很大程度上规避了因缺乏靶点信息或靶点蛋白结构不明确而导致的问题,因此在药物设计领域得到了广泛应用。首先,化合物的小分子结构相对简单且直观,对药学家而言易于理解和解读。这与蛋白质的复杂结构形成了鲜明对比,使得药学家更容易从小分子结构中提取关键信息。其次,小分子的种类繁多,为构建数据集提供了丰富的资源。这不仅有利于进行数据分类和回归分析,而且还能有效提高深度学习模型的精度^[56–59]。QEPPi 应用了针对蛋白质-蛋白质相互作用(PPI)的药物样本子集进行模型训练,在 PPI 模型中表现优异,但在特征不足的蛋白表明效果差,同时本身该方法可拓展性有待研究。DeLA-Drug 采用自动化设计以及采样与替代策略,能够自动设计与已知生物活性化合物相似的药物分子,并能从单一的初始化合物开始生成分子;但该方法要求数据量较大,生成分子的可合成性也有待优化。RT 采用 Regression Transformer 模型,能够同时处理预测和生成任务,并在化学或蛋白质空间进行基于属性的局部探索,但该方法的可控性有待优化,同时计算成本较高。Tora3D 使用自回归方法预测一组扭转角,并从中重建 3D 构象,在生成过程中保持结构的有效性,使用能量来指导构象的生成。但该方法仍未解决分子活性的问题。

3.2.3 基于片段的分子生成模型 基于片段的分子生成方法,如前所述,涉及将配体小分子根据特定规则划分为若干较小的片段。接下来,通过片段连接、片段生长、片段优化和片段自组装等过程,合

成全新的分子结构。采用这种基于片段的方法,不仅可以显著提高生成分子的化学合理性,还能保证分子骨架的合理性。这一策略有效地解决了传统方法中可能出现的分子结构不合理问题,同时还增加了分子设计的多样性和创新性。基于片段的药物发现通常从不同片段的筛选开始。通过筛选形成一个库,以便进一步扩展到更大的优化分子。片段生长是最常用的策略,生长从口袋的一个核心开始,随后添加片段并扩展到其他部分。Frag-G/M 通过 conditional transformer 进行模型构建,确保分子生成具有特定属性,这使得其在相同数据量下,能生成更多合理分子,同时生成分子具有更高的骨架多样性。但该方法计算资源消耗较大,实际应用收到数据和计算能力的限制。DiffHopp 是一个专为 Scaffold-Hopping 设计的 3D 扩散生成模型,使用几何图神经网络作为去噪器。其生成的分子具有相对高的化学多样性。然而,该方法中对 Scaffold 与其他部分的定义可能没有捕捉到原始信息中包含的关键药效团属性。DrugEX v3 使用 Graph Transformer 在强化学习框架进行了训练,允许输入预设信息(如骨架信息)以提高通用性,其生成的分子具有较好的有效性与预测亲和力。Sc2Mol 采用两步法,即首先生成只包含碳原子和单键的骨架,然后修饰得到具有所需特性的分子。该方法不需要任何预定义,允许从头开始随机生成,但该方法对化合物侧链训练结果较差,同时需要更大量的数据量^[60-63]。

3.2.4 基于药效团的分子生成模型 将 LBDD 细化为基于片段的药物设计,从某种角度来看可以视为对化学空间的一种优化。而基于药效团的药物设计方法则在药学空间上对小分子进行更深入的编辑,考虑配体小分子与靶点蛋白的相互作用,从而理论上能够有效提高模型生成分子的活性和成药性。作为基于片段分子生成模型的进一步延伸,基于药效团的分子生成方法尚未被充分研究,仍存在大量的可探索空间。这一方法的探索和应用可能为药物设计领域带来更为精确和高效的策略。STRIFE 使用片段扩展策略,并引入片段热点图,能够考虑局部蛋白质结构,从而提高了其在片段扩展中的应用性;除了从蛋白质目标自动提取药效团信息外,STRIFE 还允许自行输入药效信息。但该方法结合力只考虑了氢键,有进一步扩展的空间。DEVELOP 方法融合了 3D 药效团信息,并显著提

高了生成化合物的质量。适用于药物发现的 hit-to-lead 和 lead 优化阶段的各种设计任务。但该方法对 3D 信息利用仍然有限。PGMG 通过输入药效团信息,并引入变量来模拟药效团和分子之间存在多对多的关系。生成的分子具有较好的活性与多样性。然而现有的模型对具有生物活性的分子生成的数据利用仍然有限^[64]。

3.3 分子生成方法的不足

目前,分子生成领域主要集中在 3 种方法:基于结构、基于配体和基于片段的分子生成。近年来,基于片段的分子生成方法的应用逐渐增多。这种方法的优势在于高效地探索化学空间,生成的分子具有更广泛的结构多样性和更高的命中率,同时为后续的优化提供了良好的起点。然而,多数当前的深度学习模型尚缺乏充分的实验证据支持,即便是少数进行了下游实验验证的模型,其普遍性也有限。此外,模型训练通常需要大量的活性数据集。基于药效团的生成模型虽然能有效提高生成效率,但其依赖于对药效团信息的精确输入和生物活性数据的训练,目前在药效团信息利用率方面仍存在局限。由此可见,在 AIDD 中,分子生成方法与药物设计之间仍存在一定程度的割裂。

4 通过 AIDD 与 CADD 提高药物设计成功率的方法

目前,CADD 已在药物研发的多个领域(如靶点发现、机制研究、虚拟筛选和先导化合物优化)得到广泛应用,并已建立起丰富的理论和实践基础。在分子机制研究和动力学特性研究中,分子动力学模拟至今仍是一个核心技术,目前尚无其他方法能够同等精确地计算结合能、配体与靶点的自由能变化以及捕捉目标结构和动态特征。然而,分子动力学模拟所需的大量计算资源是其发展的一个限制因素。因此,采用更快速的人工智能方法来加速药物设计和研发流程显得尤为必要。

AIDD 结合了算法和程序与传统药物设计方法及 CADD。无论研究人员在药物设计的哪个阶段应用人工智能,其终极目标均为设计并发现具有临床价值的药物小分子。因此,将 AIDD 的重点放在药物设计过程上是研究人员必须适应的重要转变。

在探讨如何通过 CADD 的理念和方法来增强 AIDD 的成功率时,需要着重考虑以下几个关键点。

第一,数据的整合与分析。在 AIDD 中,数据的整合和分析至关重要。CADD 的优势在于其能力处理和分析大量化学和生物学数据。AIDD 应借鉴这一优势,通过利用深度学习技术优化数据处理流程,以提高预测模型的准确性。具体而言,目前阶段应将药效团信息更完整更充分地应用在 AI 之中,将基于药效团的药物设计理念充分地应用在 AIDD 之中。例如,在基于片段的分子生成模型中,将片段的分子图进行扩展,使之包含药效团信息,能更易满足实际药物需求。

第二,算法优化。CADD 中应用的算法,如分子对接和动力学模拟,为 AIDD 提供了坚实的基础。通过对这些算法进行进一步开发和优化,可以显著提升 AIDD 在药物设计中的效率和效果。同时,随着大模型的发展,AIDD 也能利用大模型的海量数据、自我学习以及处理复杂任务的能力以完成更高精度的分子生成任务。目前随着 ChatGPT 的推出,Transformer 模型日趋成熟,但同时生成模型中,扩散模型也逐渐发展,相较于 Transformer,扩散模型具有更好的生成质量与细节处理,同时具有更好的鲁棒性及多样性。在未来可能是药学家们研究的热点。

第三,CADD 中的虚拟筛选技术,能够快速评估大量化合物,对 AIDD 同样至关重要。CADD 的筛选技术存在大量假阳性的情况,如果将直接输入到人工智能模型中,会直接导致数据质量下降。因此,优化这些筛选技术,提高筛选准确率,将使 AIDD 更有效地识别潜在的药物候选物。

第四,提高 AIDD 工具的用户界面友好性和系统的可访问性,可以使非计算专家有效利用这些工具,从而促进 AIDD 应用的推广。同时,鼓励科学界和工业界之间的合作与知识共享,有助于 AIDD 工具的快速迭代和改进。

综合来看,CADD 与 AIDD 的融合为药物发现和开发领域带来了根本性的变化。通过将 CADD 的强大数据处理与分析能力与 AIDD 的先进算法和人工智能技术相结合,通过不断的技术革新和跨学科合作,特别是生成式大模型技术的运用,AIDD 的未来展望非常广阔。关键在于,AIDD 需要深入理解 CADD 和传统药物设计与药物化学中的概念,比如针对某一特定体系的药效团和基于片段的药物设计思想的内涵,骨架跃迁中生物电子等排体的概念和成功案例所运用的关键技术,在此基础上建

立多模态生成式大模型。对于某一靶点的药物开发,针对该体系的特点进行微调,构建多参数的生成式大模型,能够真正将药物设计人员和经典药物化学家多年建构的药物设计能力赋能给 AIDD,而 AIDD 在不断优化算法的基础上,则会大大提升分子生成的准确性。此外,科学界和工业界应继续密切合作,共享知识,可以促进该领域的迅速进展和应用。总之,CADD 和 AIDD 的交叉融合预示着药物研发领域将继续经历创新潮流,最终实现其根本目的,能够实现短期内发现并开发出具有临床价值的新药物。

References

- [1] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs[J]. *J Health Econ*, 2016, 47: 20-33.
- [2] DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs[J]. *J Health Econ*, 2016, 47: 20-33.
- [3] Vemula D, Jayasurya P, Sushmitha V, et al. CADD, AI and ML in drug discovery: a comprehensive review[J]. *Eur J Pharm Sci*, 2023, 181: 106324.
- [4] National Medical Products Administration. Novel Drug Screening R&D Platform-CADD[EB/OL].(2020-08-13)[2023-11-23]. <http://www.cnpharm.com/c/2020-08-13/748891.shtml>.
- [5] Liu XW, Shi DF, Zhou SY, et al. Molecular dynamics simulations and novel drug discovery[J]. *Expert Opin Drug Discov*, 2018, 13(1): 23-37.
- [6] Yan T, Yu L, Zhang N, et al. The advanced development of molecular targeted therapy for hepatocellular carcinoma[J]. *Cancer Biol Med*, 2022, 19(6): 802-817.
- [7] Boomsma W, Nielsen SV, Lindorff-Larsen K, et al. Bioinformatics analysis identifies several intrinsically disordered human E3 ubiquitin-protein ligases[J]. *PeerJ*, 2016, 4: e1725.
- [8] Huang HJ, Yu HW, Chen CY, et al. Current developments of computer-aided drug design[J]. *J Taiwan Inst Chem Eng*, 2010, 41(6): 623-635.
- [9] Matsuzaka Y, Uesawa Y. Ensemble learning, deep learning-based and molecular descriptor-based quantitative structure-activity relationships[J]. *Molecules*, 2023, 28(5): 2410.
- [10] Bissaro M, Sturlese M, Moro S. The rise of molecular simulations in fragment-based drug design (FBDD): an overview[J]. *Drug Discov Today*, 2020, 25(9): 1693-1701.
- [11] Kashyap A, Singh PK, Silakari O. Counting on fragment based drug design approach for drug discovery[J]. *Curr Top Med Chem*, 2019, 18(27): 2284-2293.
- [12] Sheng CQ, Zhang WN. Fragment informatics and computational fragment-based drug design: an overview and update[J]. *Med*

- Res Rev*, 2013, **33**(3): 554-598.
- [13] Konteatis ZD. In silico fragment-based drug design[J]. *Expert Opin Drug Discov*, 2010, **5**(11): 1047-1065.
- [14] Liu CF, Hu LD, Dong GY, *et al.* Emerging drug design strategies in anti-influenza drug discovery[J]. *Acta Pharm Sin B*, 2023, **13**(12): 4715-4732.
- [15] Li QX, Kang CB. Perspectives on fragment-based drug discovery: a strategy applicable to diverse targets[J]. *Curr Top Med Chem*, 2021, **21**(13): 1099-1112.
- [16] Giordano D, Biancaniello C, Argenio MA, *et al.* Drug design by pharmacophore and virtual screening approach[J]. *Pharmaceuticals*, 2022, **15**(5): 646.
- [17] Dror O, Schneidman-Duhovny D, Inbar Y, *et al.* Novel approach for efficient pharmacophore-based virtual screening: method and applications[J]. *J Chem Inf Model*, 2009, **49**(10): 2333-2343.
- [18] Yang SY. Pharmacophore modeling and applications in drug discovery: challenges and recent advances[J]. *Drug Discov Today*, 2010, **15**(11/12): 444-450.
- [19] Zhong FS, Xing J, Li XT, *et al.* Artificial intelligence in drug design[J]. *Sci China Life Sci*, 2018, **61**(10): 1191-1204.
- [20] Yang X, Wang YF, Byrne R, *et al.* Concepts of artificial intelligence for computer-assisted drug discovery[J]. *Chem Rev*, 2019, **119**(18): 10520-10594.
- [21] Sarkar C, Das B, Rawat VS, *et al.* Artificial intelligence and machine learning technology driven modern drug discovery and development[J]. *Int J Mol Sci*, 2023, **24**(3): 2026.
- [22] Jiménez-Luna J, Grisoni F, Weskamp N, *et al.* Artificial intelligence in drug discovery: recent advances and future perspectives[J]. *Expert Opin Drug Discov*, 2021, **16**(9): 949-959.
- [23] Li Z, Jiang MJ, Wang S, *et al.* Deep learning methods for molecular representation and property prediction[J]. *Drug Discov Today*, 2022, **27**(12): 103373.
- [24] Lin CX, Guan YF, Li HD. Artificial intelligence approaches for molecular representation in drug response prediction[J]. *Curr Opin Struct Biol*, 2024, **84**: 102747.
- [25] Chuang KV, Gunsalus LM, Keiser MJ. Learning molecular representations for medicinal chemistry[J]. *J Med Chem*, 2020, **63**(16): 8705-8722.
- [26] Chen BY, Pan ZQ, Mou MJ, *et al.* Is fragment-based graph a better graph-based molecular representation for drug design? A comparison study of graph-based models[J]. *Comput Biol Med*, 2024, **169**: 107811.
- [27] Xiong ZP, Wang DY, Liu XH, *et al.* Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism[J]. *J Med Chem*, 2020, **63**(16): 8749-8760.
- [28] Marín M, López M, Gallego-Yerga L, *et al.* Experimental structure based drug design (SBDD) applications for anti-leishmanial drugs: a paradigm shift[J]. *Med Res Rev*, 2024, **44**(3): 1055-1120.
- [29] Goodsell DS, Morris GM, Olson AJ. Automated docking of flexible ligands: applications of AutoDock[J]. *J Mol Recognit*, 1996, **9**(1): 1-5.
- [30] Eberhardt J, Santos-Martins D, Tillack AF, *et al.* AutoDock vina 1.2. 0: new docking methods, expanded force field, and python bindings[J]. *J Chem Inf Model*, 2021, **61**(8): 3891-3898.
- [31] Wu YJ, Brooks III CL. Covalent docking in CDOCKER[J]. *J Comput Aided Mol Des*, 2022, **36**(8): 563-574.
- [32] Friesner RA, Banks JL, Murphy RB, *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy[J]. *J Med Chem*, 2004, **47**(7): 1739-1749.
- [33] Miyamoto Y, Yamauchi J, Sanbe A, *et al.* Dock6, a Dock-C subfamily guanine nucleotide exchanger, has the dual specificity for Rac1 and Cdc42 and regulates neurite outgrowth[J]. *Exp Cell Res*, 2007, **313**(4): 791-804.
- [34] David L, Mdahoma A, Singh N, *et al.* A toolkit for covalent docking with GOLD: from automated ligand preparation with KNIME to bound protein-ligand complexes[J]. *Bioinform Adv*, 2022, **2**(1): vbac090.
- [35] Cross SSJ. Improved FlexX docking using FlexS-determined base fragment placement[J]. *J Chem Inf Model*, 2005, **45**(4): 993-1001.
- [36] Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS[J]. *Nucleic Acids Res*, 2011, 39(Web Server issue): W270-W277.
- [37] Kaur T, Madgulkar A, Bhalekar M, *et al.* Molecular docking in formulation and development[J]. *Curr Drug Discov Technol*, 2019, **16**(1): 30-39.
- [38] Huang YP, Zhang H, Jiang SY, *et al.* DSDP: a blind docking strategy accelerated by GPUs[J]. *J Chem Inf Model*, 2023, **63**(14): 4355-4363.
- [39] B Fortela DL, Mikolajczyk AP, Carnes MR, *et al.* Predicting molecular docking of per- and polyfluoroalkyl substances to blood protein using generative artificial intelligence algorithm DiffDock[J]. *Biotechniques*, 2024, **76**(1): 14-26.
- [40] Ryazantsev MN, Nikolaev DM, Struts AV, *et al.* Quantum mechanical and molecular mechanics modeling of membrane-embedded rhodopsins[J]. *J Membr Biol*, 2019, **252**(4/5): 425-449.
- [41] Ding Y, Yu K, Huang J. Data science techniques in biomolecular force field development[J]. *Curr Opin Struct Biol*, 2023, **78**: 102502.
- [42] Hollingsworth SA, Dror RO. Molecular dynamics simulation for all[J]. *Neuron*, 2018, **99**(6): 1129-1143.
- [43] Dheyab MA, Aziz AA, Rahman AA, *et al.* Monte Carlo simulation of gold nanoparticles for X-ray enhancement application[J]. *Biochim Biophys Acta Gen Subj*, 2023, **1867**(4): 130318.
- [44] Ruff KM, Pappu RV. AlphaFold and implications for intrinsically disordered proteins[J]. *J Mol Biol*, 2021, **433**(20): 167208.
- [45] Loman TE, Ma YB, Ilin V, *et al.* Catalyst: fast and flexible modeling of reaction networks[J]. *PLoS Comput Biol*, 2023, **19**(10): e1011530.

- [46] Wolber G, Langer T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters[J]. *J Chem Inf Model*, 2005, **45**(1): 160-169.
- [47] Benford D, Leblanc JC, Setzer RW. Application of the margin of exposure (MoE) approach to substances in food that are genotoxic and carcinogenic: example: aflatoxin B1 (AFB1)[J]. *Food Chem Toxicol*, 2010, **48**(Suppl 1): S34-S41.
- [48] Liu XF, Ouyang SS, Yu B, *et al.* PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach[J]. *Nucleic Acids Res*, 2010, **38**(Web Server issue): W609-W614.
- [49] Schneidman-Duhovny D, Dror O, Inbar Y, *et al.* PharmaGist: a webserver for ligand-based pharmacophore detection[J]. *Nucleic Acids Res*, 2008, **36**(Web Server issue): W223-W228.
- [50] Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design: a review[J]. *Curr Top Med Chem*, 2010, **10**(1): 95-115.
- [51] Wang YL, Wang F, Shi XX, *et al.* Cloud 3D-QSAR: a web tool for the development of quantitative structure-activity relationship models in drug discovery[J]. *Brief Bioinform*, 2021, **22**(4): bbaa276.
- [52] Ataíde Martins JP, Rougeth de Oliveira MA, Oliveira de Queiroz MS. Web-4D-QSAR: a web-based application to generate 4D-QSAR descriptors[J]. *J Comput Chem*, 2018, **39**(15): 917-924.
- [53] Zhu C, Li XW, Zhao BY, *et al.* Discovery of aryl-piperidine derivatives as potential antipsychotic agents using molecular hybridization strategy[J]. *Eur J Med Chem*, 2020, **193**: 112214.
- [54] Feng W, Wang LW, Lin ZY, *et al.* Generation of 3D molecules in pockets via a language model[J]. *Nat Mach Intell*, 2024, **6**: 62-73.
- [55] Qian H, Lin C, Zhao DW, *et al.* AlphaDrug: protein target specific *de novo* molecular generation[J]. *PNAS Nexus*, 2022, **1**(4): pgac227.
- [56] Kosugi T, Ohue M. Quantitative estimate index for early-stage screening of compounds targeting protein-protein interactions[J]. *Int J Mol Sci*, 2021, **22**(20): 10925.
- [57] Creanza TM, Lamanna G, Delre P, *et al.* DeLA-drug: a deep learning algorithm for automated design of druglike analogues[J]. *J Chem Inf Model*, 2022, **62**(6): 1411-1424.
- [58] Born J, Manica M. Regression Transformer enables concurrent sequence regression and generation for molecular language modelling[J]. *Nat Mach Intell*, 2023, **5**(4): 432-444.
- [59] Wang JK, Zeng YD, Sun HY, *et al.* Molecular generation with reduced labeling through constraint architecture[J]. *J Chem Inf Model*, 2023, **63**(11): 3319-3327.
- [60] Du HY, Jiang DJ, Zhang O, *et al.* A flexible data-free framework for structure-based *de novo* drug design with reinforcement learning[J]. *Chem Sci*, 2023, **14**(43): 12166-12181.
- [61] Liao ZR, Xie L, Mamitsuka H, *et al.* Sc₂Mol: a scaffold-based two-step molecule generator with variational autoencoder and transformer[J]. *Bioinformatics*, 2023, **39**(1): btac814.
- [62] Liu XH, Ye K, van Vlijmen HWT, *et al.* DrugEx v3: scaffold-constrained drug design with graph transformer-based reinforcement learning[J]. *J Cheminf*, 2023, **15**(1): 24.
- [63] Zhu HM, Zhou RY, Cao DS, *et al.* A pharmacophore-guided deep learning approach for bioactive molecular generation[J]. *Nat Commun*, 2023, **14**(1): 6234.
- [64] Imrie F, Hadfield TE, Bradley AR, *et al.* Deep generative design with 3D pharmacophoric constraints[J]. *Chem Sci*, 2021, **12**(43): 14577-14589.



[专家介绍] 李炜, 副教授, 博士, 毕业于复旦大学, 自 2007 年起在复旦大学药学院药物化学教研室任教, 曾于 2012-2013 年在哈佛医学院附属的 McLean Hospital 的 Medicinal Chemistry Laboratory 担任访问学者, 研究方向涉及精神疾病相关的药物化学, 专注于阿片受体配基及其在疾病诊断治疗中的应用, 同时致力于多样性化合物库的构建, 并探索其在中枢神经系统疾病治疗药物的发现中的应用。



[专家介绍] 付伟, 教授, 博士, 毕业于吉林大学, 相继于中国科学院上海药物所、美国休斯顿大学及巴塞罗那超级计算中心进行博士后研究。自 2006 年起至今在复旦大学药学院任职教授。专注于中枢神经系统和免疫炎症代谢类疾病的药物研发, 提出“动态结合模式”的新药物设计概念, 成功发现多个具有研究价值的药物先导化合物。主持多项国家级科研项目。担任多个专业组织的专家和评审, 如教育部留学归国人员基金评审专家和上海市生物医药智库专家。负责课程《药物设计学》获国家线上线下一流课程, 配套教材获国家级教材, 并担任美国化学会会员及 Elsevier 旗下 *Chemical Biology Drug Design* 杂志编辑。