• 专题:双清论坛“组合靶标与原创药物” •
DOI: 10.3724/BNSFC-2025-0009

人工智能在药物靶标发现中的应用*

李诗良 李洪林**

华东师范大学药学院 人工智能新药创智中心,上海 200062

[摘 要] 靶标发现作为创新药物研发的核心起始环节正面临双重瓶颈:一方面,现有研究高度集中于少数已验证靶标,药物研发同质化严重;另一方面,传统实验和计算方法在通量、精度以及对复杂难治疾病的多靶标识别方面能力有限。为应对“靶标荒”并加速源头创新,人工智能(Artificial Intelligence,AI)技术正以前所未有的深度与广度渗透至靶标发现的全流程。通过自然语言处理模型对海量文献的知识进行挖掘,深度整合多组学等生物医药多模态数据,并利用卷积神经网络提取生物影像特征、图神经网络建模分子间的相互作用网络,AI模型已极大地加速了新靶标的识别与验证过程,并向靶标发现AI多智能体协同优化系统演进,将为发现“多靶标—多通路”精准治疗策略提供新范式。本文系统性地阐述了AI在靶标发现中的创新应用、现存挑战及未来发展方向,旨在为AI驱动的新一代药物研发提供战略指导。

[关键词] 靶标发现;人工智能;知识图谱;多智能体

药物研发是一个高风险、高投入且周期漫长的过程。据统计,单个新药从发现到上市平均需10~15年,研发成本高达26亿美元,而最终成功率不足10%[1,2]。对药物作用机制的认识不足是导致临床II/III期试验因疗效不佳而失败的关键因素[3]。截至2022年,已确定的成功药物靶标不到500个,而人类蛋白质组中也仅1.8%拥有已验证的化学探针[4]。面对靶标枯竭危机,靶标挖掘的创新突破迫在眉睫,其中包括由结构基因组学联盟(Structural Genomics Consortium,SGC)发起的“Target 2035”计划,旨在通过开放科学合作,为人类蛋白质组中的绝大部分蛋白开发高质量的化学与生物探针,从而从根本上系统性地破解靶标发现的瓶颈[5]。

作为创新药物研发的源头,靶标发现指识别能够被药物调控并产生治疗效果的特定生物分子或细胞通路的过程。传统靶标发现方法主要分为实验、组学和计算三类。这些方法已经取得了诸多突破性成果,如CRISPR筛选技术发现BRD2(Transcriptional Regulator Bromodomain-containing Protein 2)是调控SARS-CoV-2宿主感染的潜在关键因子[6],大规模全基因组关联研究(Genome-Wide Association Study,GWAS)鉴定出炎症性肠病潜在治疗靶标白介素-23(IL-23)[7]。然而,各类方法均存在一定的局限性。实验方法能直接观测生物学效应,但通量有限且成本高昂;组学方法通过整合多组学数据预测靶标,但受样本质量限制;传统计算方法虽效率高,但依赖结构数据完整性和算法精度,在新型靶标预测方面表现欠佳。特别是在研究癌症、神经退行性疾病等复杂疾病的多靶标机制时,这些传统方法往往难以全面解析。

随着生物医学数据的爆炸式增长,AI技术为药物靶标发现开辟了全新路径。AI算法能整合多模态数据,构建动态生物分子网络模型,并深度挖掘海量文献与临床数据,建立起可推理的生物医学知识图谱。这种技术范式不仅能解析传统的线性分子相互作用,更能揭示深层次的网络调控机制,从而显著提升潜在药物靶标的筛选效率和预测准确性。具体而言,卷积神经网络(Convolutional Neural Network,CNN)通过局部卷积与全局池化的协同策略,实现对生物医学影像的多尺度特

收稿日期:2025-06-30; 修回日期:2025-08-27
* 本文根据国家自然科学基金委员会第399期“双清论坛”讨论的内容整理。
** 通信作者,Email:hlli@hsc.ecnu.edu.cn
本文受到国家自然科学基金项目(82425104)和国家重点研发计划项目(2022YFC3400501)的资助。

引用格式: 李诗良,李洪林. 人工智能在药物靶标发现中的应用. 中国科学基金,2026,40(2):217–227.
Li SL,Li HL. Applications of artificial intelligence in drug target discovery. Bulletin of National Natural Science Foundation of China,2026,40(2):217–227.(in Chinese)

征提取;图神经网络(Graph Neural Network,GNN)融合拓扑分析与深度学习算法,精准捕捉生物分子网络的复杂关联特征;自然语言处理(Natural Language Processing,NLP)技术突破文献挖掘瓶颈,通过语义空间映射显著提升靶标功能预测的灵敏度;而多智能体通过智能体间的协同博弈与知识共享,有效解决数据异构性问题,实现靶标发现流程的自主优化。这些技术的协同创新与集成应用,正在推动靶标发现领域取得突破性进展,为创新药物研发注入新的动力。

本综述旨在系统梳理AI靶标发现的技术演进、应用案例与核心挑战,并展望其在靶标轴/群和瞬时构象靶标识别等前沿领域中的机遇,为创新药物研发提供新思路。

1 传统靶标发现方法概述与局限性

药物靶标发现是现代药物开发过程中至关重要的起始步骤,旨在识别与疾病发生发展密切相关的生物分子,如特定的蛋白质或基因,从而为创新药物的研发奠定基础[8]。在过去的几十年中,传统靶标发现方法在推动新药研发方面发挥了核心作用(图1)。然而,随着生物医学科学的飞速发展和对更高效、更精准药物需求的增长,这些方法的固有局限性也日益凸显,不断促使着新技术的探索与应用[9]。

1.1 基因微扰:通过揭示基因功能来识别靶标

基因微扰策略通过直接操控基因组推断基因功能,其发展历程体现了对精确性和通量性的不懈追求。最先兴起的是基因敲除技术,其通过在胚胎干细胞中进行同源重组来永久性破坏特定基因[10]。随后,RNA干扰(RNAi)技术利用小干扰RNA(siRNA)在转录后降解目标mRNA,实现了快速、高通量的基因沉默[11]。2012年出现的革命性CRISPR-Cas9基因编辑技术,利用向导RNA(sgRNA)引导Cas9核酸酶精确切割DNA,实现永久性的基因敲除[12]。

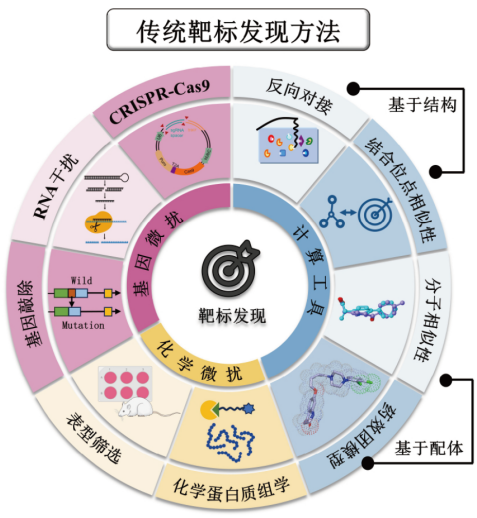


图1 传统靶标发现方法

Fig.1 Traditional Drug Target Discovery Methods

1.2 化学微扰:从表型到靶标的逆向探索

化学微扰策略采用与基因微扰相反的逻辑,即先通过小分子诱导期望的细胞表型,再反向追溯其分子靶标[13]。表型筛选是这一策略的经典范式,其历史可追溯至药物发现的早期,如青霉素的发现[14]。然而,表型筛选最大的瓶颈在于关键靶标解析的困难,这极大地阻碍了后续的药物优化[14,15]。为了解决这一难题,化学蛋白质组学应运而生[16]。该技术利用化学探针在整个蛋白质组范围内系统性地鉴定小分子相互作用的靶标。其中,基于活性的蛋白质分析尤为关键,它利用能与酶活性位点共价结合的探针,不仅能识别靶标,还能确认其功能状态,从而为表型筛选提供了强大的机理解析工具。

1.3 传统计算驱动的靶标发现方法

1.3.1 基于结构的靶标发现

基于结构的靶标发现方法以生物大分子的三维结构为基础,通过计算模拟来预测小分子与潜在靶标的相互作用[17]。其中,反向对接(reverse docking)是核心技术之一,它将单个小分子化合物与一个包含大量蛋白质结构的数据库进行对接计算,通过评价结合亲和力来预测最有可能的靶标[18]。该方法的优势在于能够直观地探索分子与靶标的结合模式,从而识别全新的靶标。此外,基于靶标结合位点的相似性分析也是一种重要策略。该策略通过比较不同蛋白质结合位点的几何构象、残基信息及理化属性,可以发现全局结构不相关但局部功能可能相似的靶标,这为理解药物的脱靶效应和多靶标作用机制提供了精准的识别工具[19]。

1.3.2 基于配体的靶标发现

当靶标的三维结构未知时,基于配体的靶标发现方法便显示出其独特价值。这类方法的核心思想是相似的分子具有相似的生物活性,通过分析已知活性配体的结构特征来反向推断其作用靶标[17]。药效团匹配是其中的代表性技术,它通过识别已知活性分子共有的关键药效特征(如氢键供体/受体、疏水中心等)及其空间排布,构建药效团模型,然后在靶标数据库中搜索具有相似药效团特征的蛋白质[20]。这种方法不依赖于靶标的结构信息,尤其适用于天然产物或结构新颖化合物的作用机制研究,并且可以通过引入统计学显著性评分(Z'-score)等标准进一步提升预测的准确性[21]。基于分子相似性的方法则通过计算查询分子与大型数据库中具有已知靶标注释的化合物之间的结构相似度,来预测其潜在靶标。这些方法可以利用二维分子指纹,或更复杂

的三维形状及化学特征进行比较[22]。

2　人工智能赋能的药物靶标发现核心技术与方法

2.1　靶标发现中的多模态数据

随着高通量生物技术的进步，研究者得以从多个层面采集异构、生物学意义各异的数据类型，构建多维度的疾病状态图谱。这些来源多样、粒度不一的数据被统称为“多模态数据”，其融合分析正逐渐成为提高靶标发现准确性、系统性与可解释性的关键路径。

分子组学数据（如基因组、转录组、表观组、蛋白质组、代谢组）提供了疾病起源与分子调控的核心信息。其中，基因组变异有助于定位致病基因（TCGA、1 000 Genomes Project 等）；转录组数据（GEO、Array Express等）揭示基因表达变化，探索信号通路，特别是单细胞和空间转录组（Human Cell Atlas、HuBMAP等）提升了数据分辨率；表观组学（ENCODE、Roadmap Epigenomics等）用于理解非编码区域功能；而蛋白质组（PRIDE、ProteomicsDB等）和代谢组（HMDB等）则反映翻译后水平与代谢通路响应。多组学协同能构建从遗传变异到功能输出的因果链，全面解析靶标机制。

蛋白质结构信息对理解靶标功能、识别作用位点及评估成药性至关重要。实验手段（X 射线晶体学、冷冻电镜等）及结构预测工具（如AlphaFold）丰富了蛋白质三维结构数据（Protein Data Bank、AlphaFold Protein Structure Database），有效揭示了小分子结合位点、蛋白互作界面、变构调控口袋等关键区域信息。同时，PDBbind、BindingDB等数据库提供靶标与配体结合构象及亲和力信息，支持结构驱动的靶标评估。

靶标功能与所处组织微环境密切相关。传统病理图像（如HE染色、免疫组化）记录蛋白组织分布（Human Protein Atla，HPA），空间组学技术（空间转录组、空间蛋白组）实现分子表达与组织结构精准配准（SpatialDB、STNet等），可在微环境尺度识别靶标分布及细胞互作，这些数据为靶标筛选提供生理相关参考。

靶标生物学意义体现在对细胞功能或状态的影响。高通量成像、流式细胞术等提供细胞表型数据（如Cell Painting项目），可以捕捉细胞动态变化。功能扰动数据（如全基因组 CRISPR/Cas9或RNAi筛选）则通过系统识别功能性靶标（Cancer Dependency Map-DepMap、LINCS等），帮助理解靶标在细胞生存、疾病进展中的作用，两者结合可以深入挖掘靶标作用通路。

靶标通常处于复杂的分子互作网络中。生物网络（蛋白互作网络、信号通路，如STRING、KEGG、Reactome）能够识别关键调控节点，通过整合分子互作信息、组学、表型与文献证据构建的生物医学知识图谱（Heionet、PharmKG、OpenBioLink等）连接靶标、疾病、药物等实体，形成语义化信息，有助于信息关联与基于先验知识的靶标推理。

临床数据是连接分子机制与疾病表型的关键资源，为靶标发现提供外部验证。电子病历、真实世界数据库（Real-World Data，RWD）、临床试验结果等构成丰富的临床信息（如MIMIC-III/IV、SEER）。将分子数据与患者生存、药物响应等临床结局关联（cBioPortal），可识别有预后或预测价值的候选靶标。临床数据可用于推断脱靶风险，是靶标发现“闭环化”的重要环节。

多模态数据为靶标发现提供了全面、立体的生物学信息支撑。不同模态数据在分子调控、结构表征、功能响应、组织分布及临床表型等方面展现出各自的独特优势，其协同使用能够显著提升靶标识别的系统性与准确性（图2）。面对不同数据间存在的结构差异与语义异质性，亟需发展高效的数据标准化与集成策略，进一步提升数据互操作性与生物学可解释性，从而为后续的数据融合建模奠定坚实基础。

2.2　生物网络与链路预测

生物网络是将生物系统以图的方式进行表示，将生物学实体整合为系统，进行靶标预测或信息挖掘。主要流程是：（1）收集数据。组成生物网络的数据来源通常是组学、实验、文献挖掘或者临床等。（2）构建可视化互作网络。网络可以表示为$G=(V,E)$，其中V和E分别是节点和边的集合。生物网络中的节点可以代表蛋白质、基因、疾病、代谢通路和药物等，网络中的边代表的是各种生化物理功能的相互作用，这种相互作用的关系通常是关联、因果、相似性等。（3）网络分析。按照节点类型是否相同可分为同构网络和异构网络。组成同构网络的节点是相同类型的元素，如蛋白质相互作用网络，节点是蛋白质，边代表蛋白质相互作用，Amanatidou等[23]使用蛋白质网络鉴定了新的与非酒精性脂肪肝相关的潜在蛋白。异构网络是将生命科学和临床医学中不同类型的元素作为节点，并将它们多样化的关系作为边，组成一个复杂的网络。在靶标预测时一般使用GNN进行表征，将节点变成一个储存网络拓扑信息的低维特征向量，特征向量相近的药物更有可能有相同的靶标，反之亦然。在药物靶标识别的应用中，还有一种重要的网络是相似性网络，其核心思想是相似的药物可以作用于相似的靶标，从而治疗相似的疾病。图网络分析过程的主要算法有中心性分析、模块社区发现、网络扩散算法等。中心性是判断网络中节点重要性的指标，比如在癌症靶

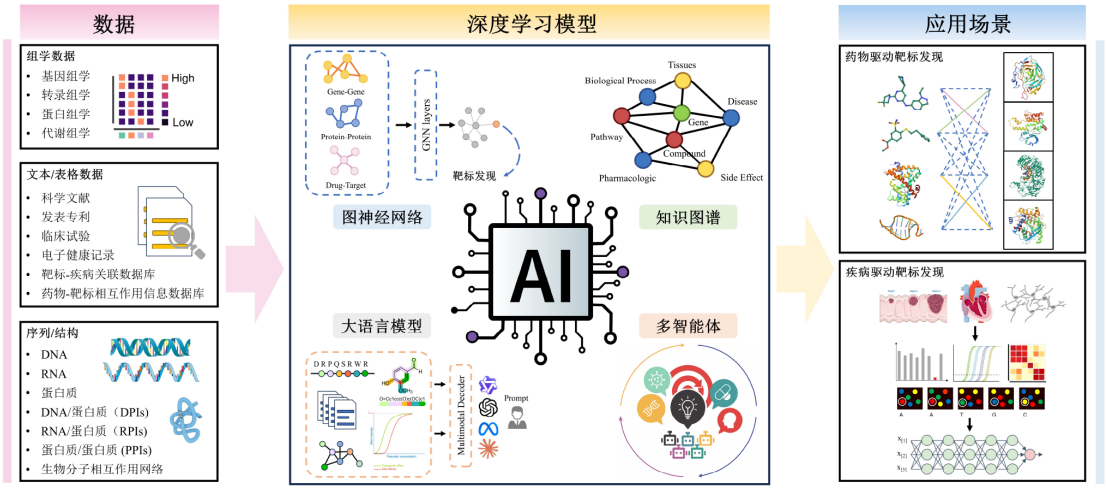


图2 人工智能赋能的药物靶标发现核心技术与方法

Fig.2 Technologies and Methods for AI-driven Drug Target Discovery

标蛋白发现中，可以度量通路中每一个蛋白节点的中心性来判断其在癌症发生中的重要性。生物网络是呈高度模块化的，社区发现算法Louvain[24]能够识别生物系统内具有紧密连接的区块，常被用于药物的重定位分析[25]。常用的网络传播算法如随机游走，得出的随机游走概率可以评价节点到锚点的接近程度，衡量药物与预测靶标存在联系的可能。

链路预测是药靶识别的重要途径，其目标是在生物网络中发现潜在的药物—靶标或疾病—靶标关联。该方法通常通过整合多源生物信息构建异构网络，并利用GNN的节点信息传递机制实现关系预测。例如，Luo等[26]开发的DTInet可用于从异构网络中预测新的药物—靶标相互作用(Drug-Target Interaction，DTI)，从而识别老药的新适应症。尽管基于生物网络的靶标预测取得了显著进展，但仍面临三大关键挑战：其一，数据质量与整合的挑战，核心在于如何对充满噪声与偏差的多源异构数据进行高效清洗、对齐与标准化；其二，模型的可解释性难题，即如何打开算法“黑箱”，揭示支撑预测结果的生物学逻辑与因果关系；其三，网络的静态局限性，由于生命系统本质上是动态演化的，因此，构建能反映特定时间、空间和病理状态下变化的动态网络模型，是该领域亟待突破的前沿方向。

2.3 知识图谱与智能推理

知识图谱(Knowledge Graph，KG)是AI系统理解复杂世界、实现深层语义理解的基石[27]。本质上，KG是一个由实体、关系和属性构成的多关系图，凭借其强大的知识表达与推理能力，已在搜索引擎优化、智能问答、推荐系统等众多领域展现出巨大价值[28]。研究人员通过构建生物医药知识图谱(Biomedical Knowledge Graphs，BKGs)以支持靶标发现，BKGs的构建过程遵循一套标准化流程，包括从权威生物医学数据库和文献中收集并预处理药物、基因、蛋白质、疾病、信号通路等异构数据，定义图谱的模式层以确定实体和关系类型及其属性，将多源数据通过本体映射或标识符匹配进行集成与映射，最终将构建好的图谱导入图数据库(如Neo4j)进行高效存储、查询和管理，并通过索引优化确保其能支持复杂的生物医学分析任务。

BKGs通过结构化整合多源实体及其复杂关系，为系统性发现潜在药物靶标构建了计算框架(图2)。释放BKGs潜力的核心在于知识图谱推理(Knowledge Graph Reasoning，KGR)，其目标是从已知知识网络中预测并补全缺失的“疾病—靶标”关联，为实验验证提供高质量候选集，提升药物研发效率。为此，学术界已开发出多种先进推理模型，主要可分为基于路径规则、基于嵌入表示和基于多模态融合三类范式。例如，PoLo模型[29]是基于路径与逻辑规则的推理算法的典型代表，其采用强化学习与逻辑规则混合策略在预定义元路径上推理，不仅可预测靶标，还能揭示其潜在生物学通路。在靶标发现中，审查PoLo等模型提供的推理路径，可深入洞察靶标与疾病关联的生物学逻辑，指导后续实验设计。基于嵌入表示与GNN的推理算法克服了对预定义路径的依赖，能更充分利用图拓扑结构，将实体和关系映射到低维向量空间进行链接预测，如KG-Predict[30]采用两阶段策略，利用CompGCN等图卷积网络学习异构图谱的拓扑与语义嵌入，再通过InteractE解码层打分排序，可有效识别与糖尿病、萎缩性胃炎等复杂疾病相关的候选基因[31]。基于多模态融合的推理算法旨在整合分子结构、文本描述、基因序列等信息，获得对生物实体更全面的理解。代表模型CamE[32]通过三重协同注意力机制，高效融合多模态特征，在DRKG-MM等基准数据集上表现优异。此外，还有

像KGE-UNIT[33]构建了多任务学习框架,其结合知识图谱嵌入(Knowledge Graph Embedding,KGE)与CNN分别提取图谱结构与生物实体模态特征,在DTI预测中成果显著。这类模型的应用价值在于,多模态推理能从更综合视角审视疾病复杂性,发现传统单模态方法难以揭示的隐藏关联,加速药物再利用与创新药开发。

2.4 靶标发现深度学习模型

2.4.1 卷积神经网络

CNN作为深度学习的核心架构之一,凭借其对网格状数据进行分层特征提取的内在能力,已成为靶标发现与预测领域的关键赋能技术。通过级联的卷积层与池化层,CNN能够自动学习从局部到全局的复杂模式,适用于处理一维生物序列、基因表达热图以及蛋白质三维结构体素等多种数据类型。

靶标CNN在靶标发现中的应用始于对蛋白质三维结构的分析。AtomNet[34]是最早将3D CNN应用于结构化药物发现的模型之一,它通过对蛋白质结合位点环境进行体素化编码,实现了对分子生物活性的预测。随后,DeepSite[35]进一步确立了3D CNN在这一领域的效用,该模型直接从蛋白质三维坐标中识别潜在的药物结合位点并对其进行评分。随后,CNN的应用范围从结构数据扩展至非图像的组学数据。DeepInsight[36]提出了一种将基因表达谱转换为二维图像表征的方法,从而使标准的CNN架构能够用于识别与疾病相关的生物标志物,拓宽了其在多组学分析中的应用场景。

与此同时,针对蛋白质静态结构无法完全反映其生物功能的局限性,研究者们致力于将蛋白质动态学信息整合到CNN模型中。TRAPP-CNN[37]通过分析分子动力学轨迹衍生的四维网格数据,评估在蛋白质动态过程中出现的瞬时结合口袋的可成药性,将蛋白质的构象柔性纳入了考量。

近年来,CNN的应用已进一步拓展至高通量表型筛选领域。DeepProfiler[38]等模型可从海量的细胞显微图像中学习形态学特征,用于量化细胞对基因敲除或药物处理等扰动的反应。这种基于图像的表型分析方法,为推断药物靶标或作用通路提供了新的高通量解决方案。

2.4.2 图神经网络

GNN已在靶标发现领域展现出巨大潜力,其根本优势在于对图结构数据或网络数据的原生处理能力(图2)。在生命科学中,众多关键实体均可自然地表示为图:例如,将原子视为节点、化学键视为边的化学分子结构,将氨基酸残基视为节点、其空间邻近关系视为边的蛋白质三维结构,以及将生物分子(如基因、蛋白质)或概念(如疾病、通路)视为节点、它们之间的相互作用或关联视为边的蛋白质相互作用网络与生物医学知识图谱。GNN通过在这些图上进行消息传递,能够学习并表征节点关键的拓扑结构与化学环境信息,从而为预测靶标及其功能提供了强大的新范式。

GNN在该领域的应用可追溯至宏观系统生物学层面的探索。例如,大规模生物医学知识图谱的构建,如药物重定位知识图谱(Drug Repurposing Knowledge Graph,DRKG)[39],将药物、基因、疾病等数百万个实体节点连接起来。通过在此类知识图谱上应用GNN模型,研究人员可以系统性地发掘连接特定疾病与潜在新靶标的隐藏关系,从而生成新的治疗假设并对其进行排序。随后,GNN被迅速应用于核心的DTI预测任务。GraphDTA[40]是该领域的代表性模型之一,它采用GNN学习药物分子的图表示,并结合蛋白质的一维序列信息,实现了对药物与靶标结合亲和力的有效预测。

近年来,研究趋势转向利用GNN处理更复杂的异构网络以解决系统层面的问题。例如,ModulePred[41]通过构建包含多种生物实体的异构模块网络,并利用GNN学习节点表示,以预测疾病与基因功能模块之间的潜在关联,为发现疾病相关靶标提供了更系统化的视角。当前,GNN模型的设计正朝着整合更丰富信息、更精细化捕捉相互作用特征的方向发展。例如,GNNBlockDTI[42]通过构建模块化的图结构,有效地整合了药物分子图与蛋白质的接触图,以更精准地捕捉相互作用的关键特征。同样地,H^2GnnDTI[43]采用层次化异构图神经网络,将药物的原子结构、蛋白质的序列及其三维结构信息统一在一个框架下进行学习,旨在实现对相互作用机制更全面的表征。

2.4.3 大语言模型

大语言模型(Large Language Models,LLMs)凭借其卓越的NLP和序列数据建模能力,正为靶标发现开辟新的前沿(图2)。与专注于图像或图结构数据的模型不同,LLMs的核心优势在于能够处理两大类关键数据:海量的非结构化文本数据(如科学文献、专利、临床试验报告)和被视为一种“语言”的生物序列数据(如蛋白质的氨基酸序列)。

LLMs在生物医学领域的应用突破主要体现在蛋白质语言模型(Protein Language Models, PLMs)上。例如,ProtT5[44]通过在庞大的蛋白质序列数据库上进行预训练,学习蛋白质序列的内在“语法”和功能模式,使其能够仅从一维序列信息预测蛋白质的功能特性。在此基础上,Meta AI开发的ESM-2[45]模型通过在数亿级别的进化多样性序列上进行训练,将PLMs的能力提升到了新的高度。这一规模化训练不仅使其能够(通过配套的ESM-Fold工具)精准预测蛋白质的三维结构,还能评

估序列突变对蛋白质功能的影响，这对于评估一个潜在靶标在疾病中的作用及其可成药性至关重要。

与此同时，LLMs在处理海量非结构化文本方面的原生优势也被应用于靶标发现。通过在专业生物医学语料库上进行预训练或微调，可以构建强大的知识提取工具。例如，BioMedLM[46]这类模型能够高效地从文献中解析并整合基因、疾病及化合物之间的复杂关系。更进一步，最新的生成式模型，如Tx-LLM[47]，超越了传统的知识提取范畴。这类模型不仅能识别已知的关联，还能主动生成并评估全新的靶标—疾病配对假设，极大地加速了早期靶标的筛选与验证过程。

2.5 靶标发现多智能体系统

近年来，以LLMs为核心的AI技术正为科学研究带来范式革命[48]。从技术原理上看，LLMs在多智能体系统中扮演“大脑”或“协调器”的角色：它将复杂的科学问题拆解为子任务，并分配给多个能调用数据库、执行分析软件等专用工具的智能体。最后，核心LLMs整合并推理各智能体的返回结果，以协同方式完成单一模型无法应对的复杂工作流。在此框架下，LLMs正从被动工具发展为自主决策的智能体，有望加速靶标识别、机制阐明和个性化治疗方案的制定[49]。

通过协同整合领域知识与外部工具，LLMs多智能体系统能够自主地解构并执行药物发现中的复杂工作流。大型科技公司和学术机构积极探索“LLMs即科学家”范式[50]，谷歌的AI辅助科学家系统[51]已成功应用于加速药物重定位和提出新的治疗靶标，如成功识别了肝纤维化的表观遗传靶标。DrugAgent系统[52]通过协调多个专职智能体（查询数据库、运行模型、搜索文献和推理），生成附带详细推理过程的DTI预测，提高了准确性和可信度。CLADD框架[53]也利用多个协作智能体，通过检索增强生成技术动态调用外部知识，在无需微调的情况下处理多种药物发现任务，包括靶标预测。“元生”（OriGene）系统[54]则通过协调多个专业智能体，对多组学、临床数据和文献专利等多模态数据进行推理，系统化地识别原创性治疗靶标。此外，多智能体系统还通过自动化复杂生物医学工作流和支持高级推理来加速药物研发。TxAgen[55]专注于治疗药物推理，其通过整合生物医学知识分析药物相互作用进而生成个性化治疗策略。BioMaster、BRAD、SRAgent以及BioAgents等[56-59]致力于自动化和简化生物信息学工作流，其通过角色扮演、数据库交互、执行分析流程和本地化部署等方式，解放科学家，加速靶标验证到临床研究的各个环节。这些应用共同展示了多智能体系统在整合信息、提供可解释决策和自动化流程方面的巨大潜力。

2.6 总结与展望

药物靶标发现的核心挑战在于全面解析疾病从分子机制到临床表现的复杂病理过程，因此，高效的AI模型构建需基于对不同数据模态的深入理解及其与特定模型架构的合理适配（表1）。在这一背景下，科学文献和专利等非结构化文本数据因蕴含海量知识而天然适合由LLMs进行解析，凭借其卓越的语言理解与推理能力，能够高效挖掘潜在的靶标—疾病关联。与此同时，以图结构形式化表示的生物网络和知识图谱，其拓扑属性与GNN的优势高度契合，使得捕捉深层次关系和预测药物—靶标相互作用成为可能。进一步地，蛋白质序列和三维结构作为靶标功能解析的关键信息来源，分别适合由PLMs和CNN进行建模：前者通过学习蛋白质的“生物语言”理解其内在属性，从而实现序列表征、突变效应预测、新序列设计等多元应用；而后者则精于从二维或三维结构中提取空间特征，广泛应用于结合位点鉴定、配体—靶标结合亲和力等关键任务。当前，AI靶标发现研究正加速迈向多模态数据融合的范式，通过系统整合组学、文本、表格、网络、序列及结构等多层信息，有望构建更具可解释性与更高成功率的协同框架（如多智能体系统），从而显著提升新药研发的效率与成功率。

3 AI驱动的药物靶标发现应用与成功案例

药物研发的源头核心难题——药物靶标的发现与确证，长期以来依赖疾病生物学家通过繁重的人工劳动整合海量、多样且异构的生物医学数据，以“直觉驱动”的方式提出假说，再经漫长的实验验证。靶标发现任务本身（包含“由药寻靶”和“由病寻靶”两种模式）所需处理的数据量巨大、模式复杂，碎片化的人工工作流已难以有效实现跨领域深度集成推理。人工智能凭借其处理高维、多源异构大数据的天然优势，能自动提取特征、识别跨模态数据间非直观的复杂关联，生成更具潜力的靶标假说。针对不同数据类型和任务目标，多种适配的神经网络架构正被有效应用，显著加速靶标发现过程，降低成本与盲目性，深刻变革传统靶标发现范式，在提升药物研发源头效率与临床成功率方面展现出极高的实用价值与广阔前景。

长期以来，复杂疾病（如癌症、阿尔茨海默病、心血管疾病等）常无法得到预期的治疗效果，这是细胞内分子网络和跨组织信号通路的系统性、复杂性失调导致的。例如，在癌症治疗领域，新型靶向抗癌药物的研发始终是重大科学难题。针对这一挑战，威尔康奈尔医学院的研究团队提出的CaRNetS创新计算框架[60]，成功识别出驱动不同癌症谱系的关键调控因子及其相互作用蛋白。

表1 AI驱动药物靶标发现的代表性模型框架及工作

Table 1 Representative Models and Studies in AI-driven Drug Target Discovery

框架	常见数据类型	代表方法	数据集	下游任务	项目链接
链路预测	图结构数据	DTI-net[26]	DisGeNET、IntAct、GEO	药物—靶标相互作用预测、药物重定位	https://github.com/luoyunan/DTINet
知识图谱	图结构数据	KGE-UNIT[33]	Luo's图结构数据、BioKG	药物—靶标、药物—药物相互作用预测	https://github.com/zcc1203/KGE-UNIT
卷积神经网络	图像数据、结构数据	DeepSite[35]	PDBBind refined set (2015)、CSAR Hi-Q (2013)	蛋白质结合位点预测	https://open.playmolecule.org/tools/deepsite
		DeepProfiler[38]	Cell Painting、L1000	药物作用机制分析、药物重定位、基因功能预测、活性分子筛选	https://github.com/cytomining/DeepProfiler
图神经网络	图结构数据	GraphDTA[40]	Davis、KIBA、BindingDB、LSP	药物—靶标结合亲和力预测、虚拟筛选、药物重定位	https://github.com/thinng/GraphDTA
		H^2GnnDTI[43]	DrugBank、BindingDB、STITCH	药物—靶标相互作用预测、药物重定位、药物毒性预测	https://github.com/LiminLi-xjtu/H2GnnDTI
大语言模型	文本数据、序列数据、表格数据	ESM-2[45]	Metagenomic Atlas	蛋白质结构预测、蛋白质序列表征、突变效应预测等	https://github.com/facebookresearch/esm
		BioMedLM[46]	PubMed生物医学文献数据库	生物医学问答、命名实体识别、关系抽取、文献综述生成	https://github.com/stanford-crfm/BioMedLM
多智能体	多模态数据	OriGene[54]	TRQA-lit、TRQA-db、CPTAC、TCGA、GEO、Human Protein Atlas、PubMed、OpenTarget、GEPIA、UCSC Genome Browser	疾病相关靶标自动识别、药物重定位	https://github.com/GENTE-Llab/OriGene
		DrugAgent[52]	PAMPA、HIV、DAVIS	ADMET性质预测、药物靶标预测、药物重定位	https://anonymous.4open.science/r/DrugAgent-B2EA

在神经内分泌癌中,发现了NAMPT和PPAT作为潜在靶标,并验证了 Teglarinad Chloride和Mercaptopurine Hydrate的抑制效果;在肝癌中鉴定出BRD9依赖性与TP-472的疗效;在肾癌中则证实了MDM2抑制剂RG7112的特异性作用。阿尔茨海默病(Alzheimer's Disease,AD)是一种以多种病理生理过程为特征的神经退行性疾病,其随机对照临床试验的高失败率在很大程度上是由于对阿尔茨海默病的病理生理学及其治疗调节缺乏了解[61]。为此,研究人员开发了基于网络拓扑的深度学习框架NETTAG[62],通过评估其推断的AD风险基因与已知药物靶标之间的网络接近度,针对AD风险基因的蛋白质产物识别可重新利用的药物,发现吉非贝齐是潜在预防和治疗AD的强效候选药物。Standigm ASK™[63]结合KG和神经网络专注于挖掘特发性肺纤维化(Idiopathic Pulmonary Fibrosis,IPF)相关的上皮—间充质转化(Epithelial-Mesenchymal Transition,EMT)靶标,成功识别了三个与IPF治疗有关的新基因AMFR、MDFIC和NR5A2,实验结果揭示了这些基因对TGFβ(一种已知的EMT激活剂和纤维化诱导因子)的反应性以及它们在EMT过程中的关键作用。上述疾病靶标发现的成功案例展示了AI和多组学数据整合在复杂疾病中的应用价值。

药物发现通常有基于表型和基于靶标两种途径,对候选药物靶标的先验认识是提高药物有效性和临床成功率的关键(图2)。AI在DTI预测、蛋白—蛋白相互作用、DNA/RNA—蛋白相互作用与复合物结构预测任务以及揭示新的药物候选物或药物重定位等靶标发现领域已得到广泛应用。深度学习模型DeepDTnet[64]采用深度神经网络图表征算法融合多源生物医学数据,通过实验验证了DeepDTnet预测结果,即拓扑替康(一种已获批准的拓扑异构酶抑制剂)是人类视黄酸受体相关孤儿受体γt(Retinoic acid–related Orphan Receptor γt,RORγt)的直接抑制剂,在小鼠多发性硬化症模型中显示出潜在的治疗效果。香港中文大学团队开发SCOPE-DTI深度学习框架[65]以中药明星成分人参皂苷Rh1为例,利用SCOPE-DTI预测其潜在靶标,通过表达筛选出AKR1B10等3个抗癌靶标,经后续实验验证Rh1与AKR1B10的高结合亲和力,为天然产物靶标发现提供了有利工具。BenevolentAI则利用其专有的人工智能平台,将最初获批用于治疗类风湿性关节炎的巴瑞替尼(baricitinib)鉴定为治疗新冠肺炎的潜在药物。除了直接的DTI的预测,AI方法已经成为靶标发现研究流程中的常用工具,例如,以AlphaFold为代表的深度学习模型对靶标蛋白结构预测领域产生了

革命性影响。近期加州大学的研究团队借助AlphaFold3预测PHGDH蛋白三维结构,发现该蛋白存在空间构型与已知转录因子家族中的DNA结合结构域高度相似的亚结构域,首次建立了PHGDH与阿尔茨海默病因果关系的直接证据[66];复旦大学团队同样利用AlphaFold-Multimer预测帕金森疾病全新治疗靶标FAM171A2的三维结构,从而能够利用虚拟筛选发现有效抑制FAM171A2蛋白和病理性突触核蛋白结合并抑制多巴胺能神经元对该致病蛋白纤维的摄取的分子[67]。

LLMs的最新进展催生了基于多智能体的框架,这些框架能够实现自动化科学推理、协调多种工具,并将文献知识与结构化数据相融合,从而针对治疗性靶标发现与验证所特有的复杂性进行专门设计。为此,上海交通大学、复旦大学、上海人工智能实验室等机构的研究者开发了名为“元生”(OriGene)的多智能体AI系统[54],旨在实现治疗靶标的自动化发现与验证。在实证研究中,系统通过多组学挖掘锁定G蛋白偶联受体GPR160为肝癌新靶标,体外实验验证了小分子抑制剂对Huh-7细胞系的抑制作用,揭示其阻断可激活T细胞免疫应答(“直接杀伤+免疫调节”双效机制),最终在3例患者肿瘤片段模型和12例患者来源类器官(Patient-Derived Organoids,PDO)中均显示显著的抑瘤效果。对于结直肠癌,系统优选精氨酸酶ARG2(特异性高表达于癌组织),在HCT116细胞系验证抑制剂剂量效应后,自主升级实验方案并在4例转移性患者PDO模型中实现强效肿瘤抑制。通过深度集成多智能体协同决策、自我进化机制与实验验证闭环,多智能体实现了从靶标假设生成到实验确证的流程自动化,为突破药物研发的靶标发现瓶颈提供了强大的AI驱动范式。

AI推动靶标发现的价值,不仅体现在临床前研究阶段,更在于显著加速了从新靶标发现到新药开发的全流程进程。PandaOmics是英矽智能开发的AI驱动靶标发现平台[68],AI技术通过整合多模态数据与先进算法体系,实现了从靶标识别、验证到候选药物设计的全流程智能化,已经成功应用于癌症、肌萎缩侧索硬化症、IPF以及衰老等靶标发现[69-71]。研究人员借助PandaOmics,成功识别出一个全新的IPF治疗靶标TNIK。从靶标发现到确立临床候选化合物 Rentosertib(ISM001-055)[72],这一过程仅耗时18个月。近期公布的Rentosertib IIa期临床试验结果显示,该药物达到了主要研究终点,并表现出可控的安全性和耐受性,为IPF及其他器官纤维化患者带来了全新希望。这一成功案例不仅有力验证了AI在首创靶标挖掘中的颠覆性能力,更标志着药物研发正式迈入AI驱动的新范式。

尽管精准疾病靶标的发现推动了原创新药的研发,疾病的异质性使得个性化医疗的需求日益增加。肿瘤新抗原(neoantigens)是由肿瘤细胞基因突变(如点突变、插入/缺失、融合基因等)产生的异常蛋白质片段。因其源于个体特异性突变,新抗原不易诱导免疫耐受或引发自身免疫反应,成为精准免疫治疗的理想靶标,而传统的新抗原筛选仍面临着数据复杂度高、个体差异大和验证成本高昂的问题。为此,Feng等[73]开发的PISTE算法实现了对T细胞受体(T Cell Receptor,TCR)—抗原肽—人类白细胞抗原(Human Leukocyte Antigen,HLA)相互作用复杂景观的智能导航。结果显示,该模型对TCR—抗原肽—HLA结合的预测准确率超过90%。在一项前列腺癌前瞻性研究中,基于PISTE预测的新抗原成功在75%的患者体内诱导了免疫反应,显示了AI模型在肿瘤新抗原筛选助力个性化医疗方面的临床应用潜力。

总体而言,人工智能已深刻融入药物靶标发现的各个环节,从克服海量异构数据整合的挑战,到识别复杂生物网络中的关键节点,再到精准预测分子间相互作用并指导实验验证,展现出强大的变革力量,一系列标志性成功案例有力印证了AI在该领域的颠覆性能力,为癌症、神经退行性疾病、纤维化疾病等重大难治性疾病开辟了全新的治疗路径。AI驱动的靶标发现正从根本上升级药物研发的源头创新范式,并以前所未有的效率将科学洞见转化为具有明确临床潜力的治疗方案,标志着精准医疗与个性化治疗新时代的加速到来。

4 结 语

药物靶标发现作为药物研发的起点,长期面临高失败率、高成本和研发周期冗长等挑战。传统实验筛选方法虽在靶标挖掘中发挥了重要作用,但在大规模数据整合、多靶标机制解析以及复杂疾病的异质性应对方面表现出明显的局限性。

近年来,AI技术的快速发展正推动靶标发现进入以“精准化、高效化和系统化”为特征的新阶段。在技术层面,GNN、KG和LLMs等先进方法,能够从庞杂的组学、表型、结构及临床数据中自动提取潜在靶标特征,并构建复杂疾病调控网络的多尺度表征。这类方法不仅具备更强的泛化与解释能力,也提升了靶标预测的可靠性。此外,多智能体系统已逐步应用于生物医学知识整合、分子生成和研究设计等环节,进一步推动靶标发现的系统性与自动化。在应用层面,AI驱动的靶标发现已在多个场景取得初步成果,包括蛋白质结构预测、候选靶标的筛选与优先级评估以及个体化靶标识别。这些成果不仅验证了AI在复杂生物体系中的推理潜力,也表明其具备从研究

工具向转化应用过渡的能力。尤其在难成药靶标、新靶向区域(如非编码RNA、蛋白瞬态构象)识别等领域,AI正拓展传统方法难以触及的成药空间。

展望未来,AI驱动的靶标发现有望在以下几个方向实现突破。(1)从相关推断到因果建模:结合因果推理与网络动力学,识别疾病核心调控因子,推动病因层面的靶标发现。(2)从单靶标到系统调控:聚焦“靶标群”或“靶标轴”的协同干预策略,多靶标协同干预提升疗效并克服耐药。(3)从静态结构到动态构象:利用生成模型捕捉蛋白瞬态构象,挖掘新型可药结合口袋。(4)从预测到虚拟验证:借助虚拟细胞与数字类器官平台,模拟验证靶标功能与机制。(5)从工具辅助到策略共创:推动AI深度嵌入靶标发现全流程,构建筛选—验证—干预的闭环优化体系。

尽管上述技术路径展现出巨大的潜力,当前仍面临诸多挑战。例如,缺乏统一评估标准阻碍了构象生成模型对瞬时可药口袋识别能力的系统性比较;数据共享机制与伦理规范尚未完善,限制了高质量跨尺度生物数据的整合与应用。因此,推动药物靶标发现走向智能化与系统化,需要算法突破、数据开放、多学科融合与政策协同共同驱动。

总体而言,随着AI与结构生物学、系统药理学等学科的深度融合,靶标发现正经历从“经验驱动”向“机制驱动”、从“单点预测”向“系统干预”的根本转变。这一趋势有望最终推动药物研发模式向智能化、精准化和以患者为中心的新格局演进。

参 考 文 献

[1] DiMasi JA,Grabowski HG,Hansen RW. Innovation in the pharmaceutical industry:New estimates of R&D costs. Journal of Health Economics,2016,47:20—33.

[2] Paul SM,Mytelka DS,Dunwiddie CT,et al. How to improve R&D productivity:The pharmaceutical industry's grand challenge. Nature Reviews Drug Discovery,2010,9(3):203—214.

[3] Sun DX,Gao W,Hu HX,et al. Why 90% of clinical drug development fails and how to improve it? Acta Pharmaceutica Sinica B,2022,12(7):3049—3062.

[4] Zhou Y,Zhang YT,Lian XC,et al. Therapeutic target database update 2022:Facilitating drug discovery with enriched comparative data of targeted agents. Nucleic Acids Research,2022,50(D1):D1398—D1407.

[5] Ackloo S,Antolin AA,Bartolome JM,et al. Target 2035–an update on private sector contributions. RSC Medicinal Chemistry,2023,14(6):1002—1011.

[6] Samelson AJ,Tran QD,Robinot R,et al. BRD2 inhibition blocks SARS-CoV-2 infection by reducing transcription of the host cell receptor ACE2. Nature Cell Biology,2022,24(1):24—34.

[7] Sewell GW,Kaser A. Interleukin-23 in the pathogenesis of inflammatory bowel disease and implications for therapeutic intervention. Journal of Crohn's & Colitis,2022,16(S2):ii3—ii19.

[8] Jia ZC,Yang X,Wu YK,et al. The art of finding the right drug target:Emerging methods and strategies. Pharmacological Reviews,2024,76(5):896—914.

[9] Thafar M,Raies AB,Albaradei S,et al. Comparison study of computational prediction tools for drug-target binding affinities. Frontiers in Chemistry,2019,7:782.

[10] Wyatt PG,Gilbert IH,Read KD,et al. Target validation:Linking target and chemical properties to desired product profile. Current Topics in Medicinal Chemistry,2011,11(10):1275—1283.

[11] Künne T,Swarts DC,Brouns SJJ. Planting the seed:Target recognition of short guide RNAs. Trends in Microbiology,2014,22(2):74—83.

[12] Jiang FG,Doudna JA. CRISPR-Cas9 structures and mechanisms. Annual Review of Biophysics,2017,46:505—529.

[13] Schenone M,Dančík V,Wagner BK,et al. Target identification and mechanism of action in chemical biology and drug discovery. Nature Chemical Biology,2013,9(4):232—240.

[14] Gaynes R. The discovery of penicillin—new insights after more than 75 years of clinical use. Emerging Infectious Diseases,2017,23(5):849—853.

[15] Jacobsen JOB,Kelly C,Cipriani V,et al. Phenotype-driven approaches to enhance variant prioritization and diagnosis of rare disease. Human Mutation,2022,43(8):1071—1081.

[16] Rix U,Superti-Furga G. Target profiling of small molecules by chemical proteomics. Nature Chemical Biology,2008,5(9):616—624.

[17] Gong DH,Tang JX,Wang BL,et al. Method development for potential drug target identification and drug discovery. Scientia Sinica Chimica,2025,55(8):2223—2242.

[18] Li HL,Gao ZT,Kang L,et al. TarFisDock:A web server for identifying drug targets with docking approach. Nucleic Acids Research,2006,34(Web Server Issue):W219—W224.

[19] Li SL,Cai CQ,Gong JY,et al. A fast protein binding site comparison algorithm for proteome-wide protein function prediction and drug repurposing. Proteins,2021,89(11):1541—1556.

[20] Liu XF,Ouyang SS,Yu B,et al. PharmMapper server:A web server for potential drug target identification using pharmacophore mapping approach. Nucleic Acids Research,2010,38(Web Server Issue):W609—W614.

[21] Wang X,Shen YH,Wang SW,et al. PharmMapper 2017 update:A web server for potential drug target identification with a comprehensive target pharmacophore database. Nucleic Acids Research,2017,45(W1):W356—W360.

[22] Gong JY,Cai CQ,Liu XF,et al. ChemMapper:A versatile web server for exploring pharmacology and chemical structure association based on molecular 3D similarity method. Bioinformatics,2013,29(14):1827—1829.

[23] Amanatidou AI,Dedoussis GV. Construction and analysis of protein-protein interaction network of non-alcoholic fatty liver disease. Computers in Biology and Medicine,2021,131:104243.

[24] Blondel VD,Guillaume JL,Lambiotte R,et al. Fast unfolding of communities in large networks. Journal of Statistical Mechanics:Theory and Experiment,2008,2008(10):10008.

[25] Chu YY,Shan XQ,Chen TH,et al. DTI-MLCD:Predicting drug-target interactions using multi-label learning with community detection method. Briefings in Bioinformatics,2021,22(3):bbaa205.

[26] Luo YN,Zhao XB,Zhou JT,et al. A network integration approach for

drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nature Communications,2017,8:573.

[27] Ji SX,Pan SR,Cambria E,et al. A survey on knowledge graphs:Representation,acquisition,and applications. IEEE Transactions on Neural Networks and Learning Systems,2022,33(2):494—514.

[28] Peng CY,Xia F,Naseriparsa M,et al. Knowledge graphs:Opportunities and challenges. Artificial Intelligence Review,2023,56(11):13071—13102.

[29] Liu YS,Hildebrandt M,Joblin M,et al. Neural multi-hop reasoning with logical rules on biomedical knowledge graphs. The Semantic Web. Cham:Springer International Publishing,2021:375—391.

[30] Gao ZX,Ding PJ,Xu R. KG-Predict:A knowledge graph computational framework for drug repurposing. Journal of Biomedical Informatics,2022,132:104133.

[31] Wang XY,Yang K,Jia T,et al. KDGene:Knowledge graph completion for disease gene prediction using interactional tensor decomposition. Briefings in Bioinformatics,2024,25(3):bbae161.

[32] Xu DR,Zhou JB,Xu T,et al. Multimodal biological knowledge graph completion *via* triple co-attention mechanism. 2023 IEEE 39th International Conference on Data Engineering (ICDE). Anaheim,CA,USA. IEEE,2023:3928—3941.

[33] Zhang CC,Zang TY,Zhao TY. KGE-UNIT:Toward the unification of molecular interactions prediction based on knowledge graph and multi-task learning on drug discovery. Briefings in Bioinformatics,2024,25(2):bbae043.

[34] Wallach I,Dzamba M,Heifets A. Atomnet:A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. (2015-10-10)/[2025-06-30]. https://doi.org/10.48550/arXiv.41510.02855.

[35] Jiménez J,Doerr S,Martínez-Rosell G,et al. DeepSite:Protein-binding site predictor using 3D-convolutional neural networks. Bioinformatics,2017,33(19):3036—3042.

[36] Sharma A,Vans E,Shigemizu D,et al. DeepInsight:A methodology to transform a non-image data to an image for convolution neural network architecture. Scientific Reports,2019,9:11399.

[37] Yuan JH,Han SB,Richter S,et al. Druggability assessment in TRAPP using machine learning approaches. Journal of Chemical Information and Modeling,2020,60(3):1685—1699.

[38] Moshkov N,Bornholdt M,Benoit S,et al. Learning representations for image-based profiling of perturbations. Nature Communications,2024,15:1594.

[39] Ioannidis VN,Song X,Manchanda S,et al. Drkg-drug repurposing knowledge graph for covid-19. Journal of Biomedical Informatics,2020,115:103696.

[40] Nguyen T,Le H,Quinn TP,et al. GraphDTA:Predicting drug–target binding affinity with graph neural networksFree. Bioinformatics,2021,37(8):1140—1147.

[41] Jia XH,Luo WW,Li JQ,et al. A deep learning framework for predicting disease-gene associations with functional modules and graph augmentation. BMC Bioinformatics,2024,25(1):214.

[42] Deng GJ,Shi CS,Ge RQ,et al. Efficient substructure feature encoding based on graph neural network blocks for drug-target interaction prediction. Frontiers in Pharmacology,2025,16:1553743.

[43] Jing YY,Zhang DX,Li LM. H2GnnDTI:Hierarchical heterogeneous graph neural networks for drug–target interaction prediction. Bioinformatics,2025,41(4):btaf117.

[44] Pokharel S,Pratyush P,Heinzinger M,et al. Improving protein succinylation sites prediction using embeddings from protein language model. Scientific Reports,2022,12:16933.

[45] Lin ZM,Akin H,Rao R,et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science,2023,379(6637):1123—1130.

[46] Bolton E,Venigalla A,Yasunaga M,et al. Biomedlm:A 2.7 b parameter language model trained on biomedical text. (2024-03-27)/[2025-06-30]. https://doi.org/10.48550/arXiv.42403.18421.

[47] Chaves JMZ,Wang E,Tu T,et al. Tx-llm:A large language model for therapeutics. (2024-06-10)/[2025-06-30]. https://doi.org/10.48550/arXiv.42406.06316.

[48] Lu J,Choi K,Eremeev M,et al. Large language models and their applications in drug discovery and development:A primer. Clinical and Translational Science,2025,18(4):e70205.

[49] Yang Y,Peng Q,Wang J,et al. Multi-llm-agent systems:Techniques and business perspectives. (2024-11-21)/[2025-06-30]. https://doi.org/10.48550/arXiv.42411.14033.

[50] Zheng T,Deng Z,Tsang HT,et al. From automation to autonomy:A survey on large language models in scientific discovery. (2025-05-19)/[2025-06-30]. https://doi.org/10.48550/arXiv.42505.13259.

[51] Gottweis J,Weng WH,Daryin A,et al. Towards an ai co-scientist. (2025-02-26)/[2025-06-30]. https://doi.org/10.48550/arXiv.42502.18864.

[52] Liu S,Lu Y,Chen S,et al. Drugagent:Automating ai-aided drug discovery programming through llm multi-agent collaboration. (2024-11-24)/[2025-06-30]. https://doi.org/10.48550/arXiv.42411.15692.

[53] Lee N,De Brouwer E,Hajiramezanali E,et al. Rag-enhanced collaborative llm agents for drug discovery. (2025-02-22)/[2025-06-30]. https://doi.org/10.48550/arXiv.42502.17506.

[54] Zhang Z,Qiu Z,Wu Y,et al. Origene:A self-evolving virtual disease biologist automating therapeutic target discovery. (2025-06-06)/[2025-06-30]. https://doi.org/10.1101/2025.1106.1103.657658.

[55] Gao S,Zhu R,Kong Z,et al. Txagent:An ai agent for therapeutic reasoning across a universe of tools. (2025-03-14)/[2025-06-30]. https://doi.org/10.48550/arXiv.42503.10970.

[56] Mehandru N,Hall AK,Melnichenko O,et al. Bioagents:Democratizing bioinformatics analysis with multi-agent systems. (2025-01-10)/[2025-06-30]. https://doi.org/10.48550/arXiv.42501.06314.

[57] Pickard J,Prakash R,Choi MA,et al. Automatic biomarker discovery and enrichment with BRAD. Bioinformatics,2025,41(5):btaf159.

[58] Su H,Long W,Zhang Y. Biomaster:Multi-agent system for automated bioinformatics analysis workflow. (2025-01-26)/[2025-06-30]. https://doi.org/10.1101/2025.1101.1123.634608.

[59] Youngblut ND,Carpenter C,Prashar J,et al. Scbasecount:An AI agent-curated,uniformly processed,and continually expanding single cell data repository. (2025-05-06)/[2025-06-30]. https://doi.org/10.1101/2025.1102.1127.640494.

[60] Forbes AN,Xu D,Cohen S,et al. Discovery of therapeutic targets in cancer using chromatin accessibility and transcriptomic data. Cell Systems,2024,15(9):824—837.

[61] Cheng FX,Wang F,Tang J,et al. Artificial intelligence and open science in discovery of disease-modifying medicines for Alzheimer's disease. Cell Reports Medicine,2024,5(2):101379.

[62] Xu JL,Mao CS,Hou Y,et al. Interpretable deep learning translation of GWAS and multi-omics findings to identify pathobiology and drug repurposing in Alzheimer's disease. Cell Reports,2022,41(9):111717.

[63] Han S,Lee JE,Kang S,et al. Standigm ASK™:Knowledge graph and

artificial intelligence platform applied to target discovery in idiopathic pulmonary fibrosis. Briefings in Bioinformatics,2024,25(2):bbae035.
[64] Zeng XX,Zhu SY,Lu WQ,et al. Target identification among known drugs by deep learning from heterogeneous networks. Chemical Science,2020,11(7):1775—1797.
[65] Chen Y,Ji X,Zhang Z,et al. Scope-dti:Semi-inductive dataset construction and framework optimization for practical usability enhancement in deep learning-based drug target interaction prediction. (2025-03-12)/[2025-06-30]. https://doi.org/10.48550/arXiv.42503.09251.
[66] Chen JC,Hadi F,Wen XZ,et al. Transcriptional regulation by PHGDH drives amyloid pathology in Alzheimer's disease. Cell,2025,188(13):3513—3529.e26.
[67] Wu KM,Xu QH,Liu YQ,et al. Neuronal FAM171A2 mediates α-synuclein fibril uptake and drives Parkinson's disease. Science,2025,387(6736):892—900.
[68] Kamya P,Ozerov IV,Pun FW,et al. PandaOmics:An AI-driven platform for therapeutic target and biomarker discovery. Journal of Chemical Information and Modeling,2024,64(10):3961—3969.
[69] Pun FW,Liu BHM,Long X,et al. Identification of therapeutic targets for amyotrophic lateral sclerosis using PandaOmics - an AI-enabled biological target discovery platform. Frontiers in Aging Neuroscience,2022,14:914017.
[70] Pun FW,Leung GHD,Leung HW,et al. Hallmarks of aging-based dual-purpose disease and age-associated targets predicted using Panda Omics AI-powered discovery engine. Aging,2022,14(6):2475—2506.
[71] Pun FW,Leung GHD,Leung HW,et al. A comprehensive AI-driven analysis of large-scale omic datasets reveals novel dual-purpose targets for the treatment of cancer and aging. Aging Cell,2023,22(12):e14017.
[72] Xu ZJ,Ren F,Wang P,et al. A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis:A randomized phase 2a trial. Nature Medicine,2025,31(8):2602—2610.
[73] Feng ZY,Chen JY,Hai YL,et al. Sliding-attention transformer neural architecture for predicting T cell receptor–antigen–human leucocyte antigen binding. Nature Machine Intelligence,2024,6(10):1216—1230.

Applications of Artificial Intelligence in Drug Target Discovery

Shiliang Li Honglin Li*

Innovation Center for AI and Drug Discovery, School of Pharmacy, East China Normal University, Shanghai 200062, China

Abstract Target discovery,a core step in drug development,is currently facing a dual bottleneck. Current research efforts remain highly concentrated on a small number of validated targets,resulting in serious homogeneity across drug pipelines. At the same time,traditional experimental and computational approaches are limited in throughput,precision,and their capacity to identify multiple targets in complex and refractory diseases. To overcome these challenges,artificial intelligence (AI) is now transforming this landscape. Natural language processing enables large-scale knowledge mining from biomedical literature,while the deep integration of multi-omics and multimodal datasets broadens biological insights. Convolutional neural networks extract features from bioimaging,and graph neural networks capture the complexity of molecular interaction networks. Together,these advances greatly accelerate the identification and validation of novel targets. Looking forward,the field is evolving toward AI-driven,multi-agent collaborative-optimization systems,offering a new paradigm for discovering "multi-target—multi-pathway" precision therapies. This paper systematically examines the innovative applications,challenges,and future directions of AI in target discovery,providing strategic guidance for the next generation of AI-enabled drug development.

Keywords target discovery;artificial intelligence;knowledge graph;multi-agent system

李洪林 博士生导师,教授,华东师范大学药学院院长、人工智能新药创智中心主任,原临港国家实验室副主任。主要研究方向为药物靶标识别和药物发现的计算方法及其应用,主持国家自然科学基金青年科学基金项目(A类)及延续资助等项目10余项。获上海市自然科学奖一等奖、教育部自然科学奖一等奖、第十三届中国青年科技奖等奖项。

李诗良 博士生导师,研究员,华东师范大学药学院药物探索联合实验室主任,紫江青年学者。主要研究方向为药物靶标及新药发现,主持国家自然科学基金面上项目、上海市"科技创新行动计划"启明星项目等项目5项。获第二十二届工博会高校展区特等奖等奖项。

(责任编辑 贾祖冰 张 强)

* Corresponding Author,Email:hlli@hsc.ecnu.edu.cn