

计算机技术在药物设计与发现中的应用进展

於振兴

重庆理工大学 重庆

【摘要】随着组学技术与结构生物学的快速发展，计算机技术在药物设计与发现中的作用愈加突出。本文系统综述了当前主流的计算机辅助药物设计策略，重点关注生物信息学方法（如加权基因共表达网络分析 WGCNA、单细胞 RNA 测序 scRNA-seq 及机器学习）在药物靶点识别中的应用，以及基于分子结构的虚拟筛选、分子对接、分子动力学模拟、结合自由能计算和丙氨酸扫描等技术在候选小分子筛选与机制验证中的关键作用。在此基础上，进一步探讨了多种计算技术的整合应用。最后，展望了在人工智能与多模态数据驱动下，计算机辅助药物设计的未来发展趋势与面临的挑战。本研究可为交叉学科背景下的新药研发提供理论支持与技术参考。

【关键词】计算机辅助药物设计；WGCNA；scRNA-seq；虚拟筛选；分子动力学模拟；结合自由能计算；丙氨酸扫描

【收稿日期】2025 年 4 月 25 日

【出刊日期】2025 年 5 月 22 日

【DOI】10.12208/j.ijmd.20250018

Advances in the application of computational technologies in drug design and discovery

Zhenxing Yu

Chongqing University of Technology, Chongqing

【Abstract】 With the rapid advancement of omics technologies and structural biology, computational technologies have played an increasingly prominent role in drug design and discovery. This review provides a comprehensive overview of the current mainstream strategies in computer-aided drug design (CADD), with a particular emphasis on the application of bioinformatics approaches—such as Weighted Gene Co-expression Network Analysis (WGCNA), single-cell RNA sequencing (scRNA-seq), and machine learning—in the identification of drug targets. Additionally, structure-based techniques including virtual screening, molecular docking, molecular dynamics simulations, binding free energy calculations, and alanine scanning are highlighted for their critical roles in the screening of candidate small molecules and the elucidation of underlying mechanisms. On this basis, the integration and synergistic application of multiple computational technologies are further discussed. Finally, the future development trends and challenges of CADD under the influence of artificial intelligence and multimodal data are explored. This study aims to provide theoretical support and technical reference for novel drug development in interdisciplinary contexts.

【Keywords】 Computer-aided drug design; WGCNA; scRNA-seq; Virtual screening; Molecular dynamics simulations; Binding free energy calculations; Alanine scanning

随着新药研发成本的持续攀升和周期的不断延长，传统的药物发现方法已难以满足现代医学对高效率、低成本和高精确度的需求。近年来，计算机技术的快速发展为药物设计与发现提供了有力工具，极大地推动了药物研发模式的转型。计算机辅助药物设计（Computer-Aided Drug Design, CADD）作

为当前新药研发的重要手段，凭借其高通量、高效率与预测性强等优势，广泛应用于先导化合物筛选、靶点识别、活性优化等多个阶段。

与此同时，生物信息学的发展使研究者能够从海量的组学数据中挖掘潜在的药物靶点。如加权基因共表达网络分析（Weighted Gene Co-expression

Network Analysis, WGCNA) 能够有效识别疾病相关模块基因, 结合单细胞 RNA 测序 (single-cell RNA sequencing, scRNA-seq) 技术进一步提升靶点发现的精度和特异性。在此基础上, 通过虚拟筛选、分子对接、分子动力学模拟和结合自由能计算等一系列结构生物学方法, 可深入分析候选化合物与靶标之间的结合机制与稳定性, 从而提高药物设计的可靠性。

本文将围绕生物信息学和分子模拟两大类计算机技术的基本原理及其典型应用进行综述, 重点探讨二者在现代药物研发中的整合策略与协同价值, 以期为新药研发提供参考和思路。

本综述采用 PubMed 数据库进行文献检索, 检索策略包括主题词与自由词组合, 主要关键词有“computer-aided drug design”、“molecular dynamics simulations”、“virtual screening”、“machine learning”、“AlphaFold”等。纳入标准: 同行评议的综述或研究论文, 关注计算机辅助药物设计应用; 排除标准: 会议摘要、专利、学位论文及与主题无关文献。检索结果先经标题和摘要筛选, 再阅读全文评估并纳入文献。

1 生物信息学技术在药物靶点发现中的应用

药物靶点的准确识别是药物研发的核心步骤之一。传统的实验筛选方法不仅耗时费力, 且效率较低。随着高通量测序技术的发展和组学数据的不断积累, 生物信息学已成为现代药物靶点挖掘的重要手段, 可在海量数据中识别与疾病发生发展密切相关的分子靶点。典型技术包括 WGCNA、机器学习模型以及 scRNA-seq 等。

1.1 WGCNA 在共表达模块识别与靶点筛选中的应用

WGCNA 是一种基于表达谱数据构建基因共表达网络的系统生物学方法, 能够通过构建模块来识别与特定表型 (如疾病分组、临床指标) 显著相关的基因集群^[1]。其优势在于不仅考虑基因间的表达趋势相关性, 还可将多个协同表达的基因归入统一模块, 便于识别潜在的功能核心基因 (hub gene) 或相关通路。在癌症研究中, WGCNA 分析可用于识别与结直肠癌进展相关的关键基因模块, 从而为靶向治疗提供潜在的生物标志物^[2]。

在药物靶点筛选中, 通常以 GEO 或 TCGA 等

数据库中的转录组数据为基础, 利用 WGCNA 构建共表达网络, 识别与疾病密切相关的模块, 并结合基因功能富集分析 (GO、KEGG 等) 进一步筛选候选靶点。该过程为后续分子对接与药物筛选提供了明确的研究方向。

1.2 机器学习在靶点识别中的策略

近年来, 机器学习方法在生物信息学中的应用日益广泛, 尤其在疾病相关基因预测、药物-靶点关系构建、靶点优选等方面表现出较强的分析能力。常用算法包括支持向量机 (Support Vector Machine, SVM)、随机森林 (Random Forest, RF) 和梯度提升树 (XGBoost) 等。

通常将表达矩阵与样本表型作为输入, 通过训练分类模型识别对表型差异最具影响力的特征基因。此外, 还可结合蛋白质互作网络 (Protein-Protein Interaction, PPI) 数据, 对候选基因进行评分与排序, 优先筛选可能具有调控作用的关键节点。这种数据驱动的预测方法有助于从多维角度验证 WGCNA 的筛选结果, 进一步提升靶点识别的准确性和可靠性。

1.3 单细胞 RNA 测序在靶点精准化中的应用

单细胞 RNA 测序技术 (scRNA-seq) 能够在单细胞水平揭示细胞异质性和基因表达差异, 是实现精准医学的重要技术手段。在药物靶点研究中, scRNA-seq 不仅可识别特定细胞类型中的差异表达基因, 还能解析疾病微环境中的关键细胞亚群及其信号通路, 从而辅助筛选具有细胞特异性的治疗靶点。该技术已广泛应用于肿瘤学、免疫学、发育生物学等领域, 助力研究人员探索细胞命运决定、分化过程及细胞间相互作用等关键生物学过程^[3]。为处理单细胞转录组数据, 研究者通常采用多种生物信息学方法, 包括数据预处理、降噪、细胞聚类及差异表达分析等^[4]。

由于单细胞数据中常存在技术噪声, 如何精确地区分真实生物学信号与噪声是一个重要的挑战。因此, 近年来开发了诸多专门的算法和工具, 如 Seurat、Scanpy 等, 以提升数据分析的准确性与鲁棒性^[5,6]。scRNA-seq 通常与 WGCNA 联合应用, 用于从单细胞数据中提取关键共表达模块, 并识别在特定细胞亚群中特异性高表达且具有调控潜力的核心基因。进一步结合机器学习方法可构建靶点评分

体系,实现靶点的多维度优选与精准识别。

1.4 数据整合与资源平台支持

生物信息学靶点发现过程中常用的数据来源包括 GEO、TCGA 等数据库,蛋白互作网络数据来源于 STRING、BioGRID 等平台。此外,此外, DrugBank 等化合物-靶点数据库可用于已有药物与靶点间相互关系的查询与交叉验证。通过多源数据整合,可有效提升分析的系统性与靶点筛选的可靠性。

2 分子模拟技术在候选药物筛选与验证中的应用

在完成潜在药物靶点识别后,寻找并优化能够与其高亲和力结合的小分子化合物,成为新药研发的重要步骤。作为计算机辅助药物设计的核心组成部分,分子模拟技术涵盖了从虚拟筛选、分子对接到分子动力学模拟与结合自由能计算等多个关键环节,能够在结构层面预测候选分子的结合位点、稳定性和生物活性,为先导化合物的筛选与优化提供理论依据。

2.1 虚拟筛选

虚拟筛选 (Virtual Screening, VS) 是一种基于计算机的高通量筛选技术,可依据靶标蛋白的三维结构 (结构基础虚拟筛选) 或已知活性配体的特征 (配体基础虚拟筛选),从化合物库中筛选出可能具有活性的小分子化合物。常用的数据库如 ZINC、PubChem、DrugBank 等,均提供数百万级的可筛选小分子资源。

在实际应用中,研究者常依据已解析的蛋白晶体结构或通过同源建模获得靶点结构,并结合 AutoDock、DOCK、MOE 等软件进行虚拟筛选。通过打分函数对化合物的结合能力进行排序,筛选出前几十至上百个潜在高亲和力的小分子进入后续分析环节。

2.2 分子对接

分子对接 (Molecular Docking) 是一种计算模拟方法,旨在预测小分子与靶标蛋白之间的最优结合构象及其结合自由能,是药物设计中的关键技术之一。分子对接的基本理论来源于 Fisher 提出的“钥匙与锁”模型,该模型认为配体 (“钥匙”) 和受体 (“锁”) 之间必须在结构和空间上匹配。然而,受体与配体在结合过程中常伴随构象变化,使该过程

更加复杂化^[7,8]。该方法通过模拟小分子在靶点结合位点中的多种构象排列,评估其空间与能量匹配程度,最终筛选出结合稳定性较高的复合物^[9]。

在对接过程中,通常会锁定靶蛋白的活性口袋区域,并利用打分函数 (如 Vina Score、Gold Score) 对结合能进行评估。分子对接结果不仅提供小分子与关键残基的相互作用信息,如氢键、疏水作用与 π - π 堆积等,还能指导后续化合物的结构优化。

2.3 分子动力学模拟

分子动力学 (Molecular Dynamics, MD) 模拟用于模拟蛋白-配体复合物在生理条件下的动态行为,通过计算分子在一定时间尺度内的运动轨迹,评估其稳定性与构象变化。在药物设计领域,分子动力学模拟通过量化配体与受体结合模式的稳定性,弥补了传统实验研究的不足。例如,分子对接所得的配体-受体复合物可进一步通过 MD 模拟验证其结合可靠性与自由能变化。此外,MD 模拟还可揭示结合过程中受体构象可能发生的诱导契合现象,从而为深入理解蛋白质-配体的复杂相互作用机制提供详尽的信息^[10]。其在疾病相关靶点研究及药物筛选中的应用已显著提升研究效率并降低研发成本^[11]。

常用的模拟软件包括 GROMACS、AMBER、NAMD 等。在药物设计中,研究者可将分子对接得到的复合物作为初始构象,构建模拟体系 (加入水分子、离子环境),进行数十到上百纳秒的动力学模拟,从而分析复合物的均方根偏差、均方根波动、氢键变化等参数。这些指标有助于评估候选化合物在体内环境中是否稳定结合靶点,从而提升预测准确性。

2.4 结合自由能计算与丙氨酸扫描

基于分子动力学模拟结果,可进一步采用 MM/PBSA 或 MM/GBSA 方法计算复合物的结合自由能,从而快速筛选潜在候选药物。结合自由能计算是对小分子与蛋白结合稳定性的进一步定量评估,常用方法包括 MM-PBSA (分子力场+泊松-玻尔兹曼表面面积) 和 MM-GBSA (分子力场+广义玻尔兹曼表面面积)^[12]。该方法基于 MD 模拟轨迹,估算蛋白-配体复合物的总自由能差,为候选分子的活性预测提供定量依据。

此外,丙氨酸扫描 (Alanine Scanning) 是一种

突变分析技术，广泛用于识别配体结合位点上的关键氨基酸残基。该方法通过将关键残基逐一突变为丙氨酸，并重新计算结合自由能，以评估各残基在结合中的贡献程度。该方法常与 MM-GBSA 联合使用，可在原子水平识别影响结合力的关键位点，为后续药物结构优化提供重要参考。通过上述详细的评估方法，可确认热点氨基酸残基，并对蛋白质-配体的相互作用进行更加深入的理解^[13]。

表 1 常用分子模拟方法及其优点与局限性比较

方法	应用阶段	优点	局限性
虚拟筛选	初步筛选	快速高通量，适合大库	打分函数易误判
分子对接	初筛验证	可视化结合模式，低计算资源	缺乏体系动态信息
分子动力学模拟	稳定性验证	模拟生理环境下相互作用	计算资源需求大
MM-GBSA 结合能计算	定量分析	结合力定量评价准确	依赖轨迹质量
丙氨酸扫描	识别关键残基	揭示重要位点	不能反映所有作用类型

3 生物信息学与分子模拟的整合应用

随着多组学数据和蛋白质结构信息的不断丰富，生物信息学与分子模拟技术的整合已成为新药研发领域的重要发展趋势。通过有机结合上游的靶点预测与下游的药物筛选流程，可实现以疾病为驱动的药物发现，提高新药研发的靶向性与效率。

3.1 技术整合流程

典型的整合流程主要包括以下步骤：

①差异表达分析与 WGCNA 构建：基于疾病组织或细胞的转录组数据，筛选差异表达基因，构建 WGCNA 共表达网络，识别与疾病显著相关的模块及中心基因。

②靶点评分与优选：结合 PPI 网络分析、机器学习模型、实验验证及 scRNA-seq 数据，对模块内基因进行综合评分，优选表达特异性强、功能重要的候选靶点。

③结构建模与对接筛选：利用已解析的蛋白结构或同源建模方法获取靶点三维结构，并基于该结构进行虚拟筛选与分子对接，初步筛选出结合能力较强的小分子化合物。

④分子模拟验证与能量分析：对优化化合物进行分子动力学模拟，评估其结合稳定性；结合 MM-GBSA 等方法定量计算结合自由能，并通过丙氨酸扫描识别关键结合残基，为后续优化提供依据。

如图 1 所示，该整合策略不仅增强了靶点预测

2.5 常用分子模拟方法及其优点与局限性比较

分子模拟技术已成为结构基础药物设计的重要组成部分，涵盖从初步虚拟筛选到精确结合自由能分析的多个阶段。不同模拟方法在适用场景、精度、计算资源需求等方面存在显著差异，合理选择并组合使用相关技术对于提高新药筛选的效率与可靠性具有重要意义。表 1 总结了各方法的适用阶段、主要优势与局限性。

的生物学意义，还提高了候选分子筛选的结构可靠性，为高通量药物研发提供了系统性的路径支持。

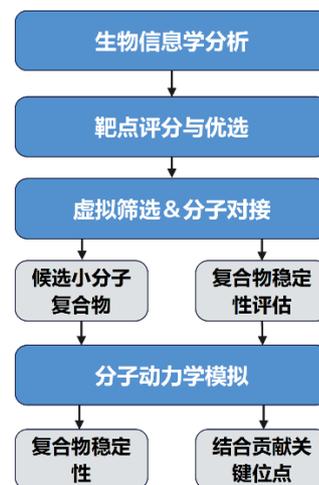


图 1 典型的整合流程

3.2 典型应用案例

近年来，已有多研究成功应用该整合策略。例如在肾间质纤维化治疗研究中，研究者通过 WGCNA 筛选模块核心靶点，结合 scRNA-seq 识别相关细胞亚群及关键基因，经过体内外实验验证后，开展分子对接与分子动力学模拟，最终获得四种与移植后肾纤维化密切相关的潜在中心生物标志物及四种潜在有效的小分子^[14]。在纤维化、代谢性疾病及中枢神经系统疾病等领域，类似策略亦被广

泛应用，显著缩短了前期药物发现的周期。

此外，AI 驱动的多模态数据融合模型也正逐步融入该流程，为整合分析提供更强的预测能力和智能化支持。

4 总结与展望

本文系统综述了当前主流计算机技术在药物设计与发现中的应用，重点探讨了生物信息学方法（如 WGCNA、机器学习和 scRNA-seq）在药物靶点识别中的作用，以及分子模拟方法（如虚拟筛选、分子对接、分子动力学模拟与结合自由能计算）在候选化合物筛选与验证中的价值。在此基础上，本文进一步论述了两类技术的整合应用实践，强调其在新药研发流程中的协同效应与前景潜力。

随着 AI 算法和蛋白质结构预测技术的快速发展，尤其是 AlphaFold2 和最新 AlphaFold3 在蛋白质折叠及复合物预测上实现了前所未有的高精度，计算机辅助药物设计正在迈向智能化和自动化新时代^[15,16]。AlphaFold 可在数小时内提供接近实验分辨率的结构模型，显著提高了同源建模和虚拟筛选的准确性；但其在无序区域或多蛋白复合物的预测上仍存在局限，且对长序列的运算成本较高、模型可解释性不足，这些问题在高通量药物筛选中需结合实验验证和其他算法加以补充。多组学数据与单细胞数据的持续积累，将进一步推动靶点发现的精准化，助力个体化医疗和新药的快速开发。

然而，当前整合技术在数据标准化、计算资源需求以及实验验证支持等方面仍面临诸多挑战。多组学与结构数据在标准化处理上存在格式差异和批次效应问题，常导致下游分析结果偏倚。应对策略包括规范化算法（如 ComBat、Harmony 等）、构建统一数据模型，以及采用深度学习方法自动校正非生物学变异。此外，结构预测和分子模拟对计算资源需求巨大，可通过加速硬件（GPU/TPU）、分布式计算和混合精度算法降低成本。最后，结合实验数据反馈以迭代优化算法参数，是保障整合结果可靠性的关键。未来应加强算法优化、标准体系建设与交叉学科协作，进一步推动生物信息学与分子模拟技术在药物研发中的深度融合与临床转化。

参考文献

[1] LANGFELDER P, HORVATH S. WGCNA: an R package

for weighted correlation network analysis[J]. BMC Bioinformatics, 2008, 9(1): 559.

[2] WANG P, ZHENG H, ZHANG J, et al. Identification of key gene modules and genes in colorectal cancer by co-expression analysis weighted gene co-expression network analysis[J]. Bioscience Reports, 2020, 40(9): BSR20202044.

[3] DENG G, ZHANG X, CHEN Y, et al. Single-cell transcriptome sequencing reveals heterogeneity of gastric cancer: progress and prospects[J]. Frontiers in Oncology, 2023, 13.

[4] JOVIC D, LIANG X, ZENG H, et al. Single-cell RNA sequencing technologies and applications: A brief overview[J]. Clinical and Translational Medicine, 2022, 12(3): e694.

[5] R S, JA F, D G, et al. Spatial reconstruction of single-cell gene expression data[J]. Nature biotechnology, 2015, 33(5).

[6] WOLF F A, ANGERER P, THEIS F J. SCANPY: large-scale single-cell gene expression data analysis[J]. Genome Biology, 2018, 19(1): 15.

[7] FAN J, FU A, ZHANG L. Progress in molecular docking[J]. Quantitative Biology, 2019, 7(2): 83-89.

[8] LEMIEUX R U, SPOHR U. How Emil Fischer was led to the lock and key concept for enzyme specificity[J]. Advances in Carbohydrate Chemistry and Biochemistry, 1994, 50: 1-20.

[9] GUEDES I A, DE MAGALHÃES C S, DARDENNE L E. Receptor-ligand molecular docking[J]. Biophysical Reviews, 2014, 6(1): 75-87.

[10] HOLLINGSWORTH S A, DROR R O. Molecular Dynamics Simulation for All[J]. Neuron, 2018, 99(6): 1129-1143.

[11] SOHRABY F, BAGHERI M, ARYAPOUR H. Performing an In Silico Repurposing of Existing Drugs by Combining Virtual Screening and Molecular Dynamics Simulation[J]. Methods in Molecular Biology (Clifton, N.J.), 2019, 1903: 23-43.

[12] POHORILLE A, JARZYNSKI C, CHIPOT C. Good Practices in Free-Energy Calculations[J]. The Journal of Physical Chemistry B, 2010, 114(32): 10235-10253.

[13] BARILLARI C, MARCOU G, ROGNAN D. Hot-spots-guided receptor-based pharmacophores (HS-Pharm): a knowledge-based approach to identify ligand-anchoring atoms

- in protein cavities and prioritize structure-based pharmacophores[J]. *Journal of Chemical Information and Modeling*, 2008, 48(7): 1396-1410.
- [14] NIU X, XU C, CHEUK Y C, et al. Characterizing hub biomarkers for post-transplant renal fibrosis and unveiling their immunological functions through RNA sequencing and advanced machine learning techniques[J]. *Journal of Translational Medicine*, 2024, 22(1).
- [15] JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold[J]. *Nature*, 2021, 596(7873): 583-589.
- [16] ABRAMSON J, ADLER J, DUNGER J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3[J]. *Nature*, 2024, 630(8016): 493-500.

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<https://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS