

人工智能与天然药物 动态监测快报

总第 007 期 2024-07-15



中国科学院昆明植物研究所
云南省植物学会
中国科学院成都文献情报中心

目 录

政策规划	3
国常会审议通过《全链条支持创新药发展实施方案》	3
北京发布《北京市加快医药健康协同创新行动计划（2024-2026 年）》	4
研究前沿	5
对接得分机器学习：针对 155 个靶标的特定目标机器学习模型，提高基于对接的虚拟筛选效果.....	5
首次发现海洋多氧樟脑内酯作为治疗溃疡性结肠炎的潜在药物.....	6
AI 驱动的一站式创新药物发现平台 DrugFlow	7
基于强化学习模型的 De Novo Molecular 可合成性导向设计：AI 辅助发现抗 IBD 导向靶点 CXCR4.....	8
NOD2 的 FOCUS：利用用户知情的机器学习框架推进 IBD 药物发现	9
基于中药微球捕鱼策略和生物层干涉测量技术的药物靶点发现.....	10
HI-Mol 模型，仅使用训练集 2% 的数据即可实现分子生成	11
Nat Comput Sci MISATO: 蛋白质-配体复合物的机器学习数据集用于基于结构的药物发现.....	12
Nat Mach Intell ChemCrow: 结合大型语言模型与化学工具的应用	13
Cell 复旦团队利用 AI 技术，从全球微生物中一性发现近百万种候选抗生素	14
Science David Baker 团队：用计算方法发现化学结构多样的大环分子	15
近期会议与活动	16

政策法规

国常会审议通过《全链条支持创新药发展实施方案》

2024-07-08

国务院总理李强 7 月 5 日主持召开国务院常务会议，研究部署推进数字经济高质量发展有关工作，审议通过《全链条支持创新药发展实施方案》。会议指出，发展创新药关系医药产业发展，关系人民健康福祉。要全链条强化政策保障，统筹用好价格管理、医保支付、商业保险、药品配备使用、投融资等政策，优化审评审批和医疗机构考核机制，合力助推创新药突破发展。要调动各方面科技创新资源，强化新药创制基础研究，夯实我国创新药发展根基。

目前，《方案》全文尚未公布。2024 年 3 月公布的《方案》的征求意见稿核心要点主要有：

创新药重点研发目录遴选原则：重大创新+解决为需求的临床需求+明确临床价值
研发端：重点支持的技术研发方向：工程细胞基因编辑、新药靶点智能挖掘、细胞靶向干预、人工智能药物发现、新型药物筛选和评价模型等前沿技术研究。

重点支持的药物研发方向：新型成药技术平台建设，突破放射性药物、口服生物药、新型疫苗和佐剂、高质量药用辅料等规模化制备技术，开发儿童多单元定时定位递送、智能响应型递送等新型递送载体和释药技术。发展多分子偶联、靶向蛋白降解、免疫调节、基因干预、干细胞再生、微生物组治疗、mRNA 疫苗与核酸药物等新型治疗药物以及具备中医药特色诊疗优势的创新药。

信息来源：

https://mp.weixin.qq.com/s?__biz=MzIzNjY5MTA2Mw==&mid=2247526202&idx=1&sn=81fdad67ba55dc2b8f2f58fafd7d6538&chksm=e9127c8f223097a7420146080aaf05c933d886f255116cc749c1766bffe7576cec2c29afc225&scene=27;

https://weibo.com/ttarticle/p/show?id=2309405011754903273520#_loginLayer_1720681558682

北京发布《北京市加快医药健康协同创新行动计划（2024-2026年）》

2024-05-23

5月23日，北京市人民政府办公厅印发《北京市加快医药健康协同创新行动计划（2024-2026年）》。《计划》提出，在细胞基因治疗、脑机接口、合成生物学等前沿技术领域部署“核爆点”专项，催生具有颠覆性、引领性的科学发现和技术突破。加强基因编辑等产业亟需的关键核心技术研究。支持细胞疗法、基因药物等市场潜力大的品种快速形成产业增量。在炎症、感染等领域，推动全新靶点、全新机制的重磅药物研发。

《计划》提出发展目标：到2026年，医药健康产业总规模达到1.25万亿元（其中，医药工业营业收入达到2400亿元），固定资产投资每年100亿元以上。实现引领全球的科学发现和技术突破5-8项，新增获批上市创新药10个、创新医疗器械30个；“三医”联动发展能级进一步提升，新建1-2家研究型医院；引进培育多层次专业人才1万人以上；市医药健康产业投资基金撬动社会风险投资每年200亿元；高品质特色园区10家。产业承载力进一步提升，具有全球影响力的现代化医药健康产业集群加速形成。

信息来源：

https://mp.weixin.qq.com/s?__biz=MzIzNjY5MTA2Mw==&mid=2247526202&idx=1&sn=81fdad67ba55dc2b8f2f58fafd7d6538&chksm=e9127c8f223097a7420146080aaf05c933d886f255116cc749c176

[6bffe7576cec2c29afc225&scene=27;](https://weibo.com/p/show?id=2309405011754903273520#_loginLayer_1720681558682)

https://weibo.com/p/show?id=2309405011754903273520#_loginLayer_1720681558682

研究前沿

对接评分机器学习: 针对 155 个靶标的特定目标机器学习模型, 提高基于对接的虚拟筛选效果

2024-07-03

在药物发现中, 分子对接方法在准确预测能量方面面临挑战。用于分子对接的评分函数往往不能完全准确地模拟复杂的蛋白质-配体相互作用, 导致虚拟筛选和靶标预测的偏差和不准确性。沈阳药科大学的 LIU HH 等研究报道了“对接评分 ML”, 该方法通过对 155 个已知癌症治疗靶点的 20 多万个对接复合物的分析进行开发。所使用的评分功能基于 ChEMBL 的生物活性数据, 并使用监督机器学习和深度学习技术进行了微调。通过使用多种数据集对该方法进行了广泛验证, 如选择性机制验证, DUDE, DUD-AD 和 LIT-PCBA 数据集, 并对舒尼替尼等药物进行了多靶点分析。为了提高预测精度, 研究了特征融合技术。通过将图形卷积网络(GCN)的功能与多个对接函数相结合, 结果表明该方法明显优于传统方法, 说明对接评分 ML 是一种高效、准确的虚拟筛选和反向对接工具。

信息来源: LIU HH, et al. Docking Score ML: Target-Specific Machine Learning Models Improving Docking-Based Virtual Screening in 155 Targets. *J. Chem. Inf. Model.* 2024
<https://doi.org/10.1021/acs.jcim.4c00072>

首次发现海洋多氧樟脑内酯作为治疗溃疡性结肠炎的潜在药物

2024-07-03

杉烷内酯是海洋软珊瑚的特征性代谢物，结构复杂，生物活性广泛。然而，很少有一个深入的药理学研究。本研究利用基于 $^1\text{H}/^{13}\text{C}$ HSQC 的小分子精确识别技术 (SMART) 对长柄马苏蒿 (*Sinularia pediculata*) 进行了系统的化学研究，分离鉴定了 31 个膜型二萜类化合物，其中包括 6 个新化合物。在生物实验中，几种化合物显示出明显的抗炎活性，抑制 NO 的产生。综合分析构效关系 (SAR)，发现两种生物活性最强、毒性较小的化合物 8 和 9 可以通过抑制 NF- κ B 和 MAPK 信号通路抑制炎症，减少炎症因子的分泌。在右旋糖酐硫酸钠 (DSS) 诱导的急性结肠炎小鼠模型中，8 和 9 表现出良好的抗炎作用和修复结肠上皮的能力，这为 cembranolides 作为潜在溃疡性结肠炎 (UC) 药物的应用提供了新的见解。

信息来源：CUI YY, et al. The First Discovery of Marine Polyoxygenated Cembranolides as Potential Agents for the Treatment of Ulcerative Colitis. *J. Med. Chem.*, 2024
<https://doi.org/10.1021/acs.jmedchem.4c00950>

AI 驱动的一站式创新药物发现平台 DrugFlow

2024-06-26

浙江大学侯廷军等在 *Journal of Chemical Information and Modeling* 上发表论文《DrugFlow: An AI-Driven One-Stop Platform for Innovative Drug Discovery》，报导了一站式创新药物发现平台 DrugFlow。这是一个人工智能驱动的一站式平台，提供了一个干净、方便和基于云的界面，以简化早期药物发现工作流程。通过无缝集成一系列创新的人工智能算法，包括分子对接、定量结构-活性关系建模、分子生成、ADMET(吸收、分布、代谢、排泄和毒性)预测和虚拟筛选，DrugFlow 可以为早期药物发现的几乎所有关键阶段提供有效的人工智能解决方案，包括命中识别和命中/先导物优化。作者希望该平台能够提供足够有价值的指导，以帮助现实世界的药物设计和发现。该平台可在 <https://drugflow.com> 上获得。

信息来源：CHAO, et al. *Chem. Inf. Model.* 2024. <https://doi.org/10.1021/acs.jcim.4c00621>

基于强化学习模型的 De Novo Molecular 可合成性导向设计： AI 辅助发现抗 IBD 导向靶点 CXCR4

2024-06-12

人工智能(AI)从头分子生成为药物发现提供了新结构的线索。然而，所生成分子的靶亲和力和可合成性对人工智能技术的成功应用提出了关键挑战。杭州师范大学的 Jiang Xiaoying 等开发了一种先进的强化学习模型，以弥合从头分子生成理论与药物发现实际方面之间的差距。该模型以化学反应模板和商用构建模块为起点，采用正向反应预测生成分子，同时进行实时对接和药物相似度预测，确保可合成性和药物相似度。应用该模型设计了靶向炎症相关受体 CXCR4 的活性分子，并根据 AI 提出的合成路线成功制备了活性分子。在随后的体外和体内实验中，一些分子表现出有效的抗 CXCR4 和抗炎活性。表现最好的化合物 XVI 缓解了炎症性肠病的相关症状，并表现出合理的药代动力学特性。

信息来源：

JIANG Xiaoying, et al. Synthetically Feasible De Novo Molecular Design of Leads Based on a Reinforcement Learning Model: AI-Assisted Discovery of an Anti-IBD Lead Targeting CXCR4. *J. Med. Chem.* 2024, 67, 12, 10057–10075. <https://doi.org/10.1021/acs.jmedchem.4c00184>

NOD2 的 FOCUS: 利用用户知情的机器学习框架推进 IBD 药物发现

2024-06-06

加拿大 Toronto 大学的 Choudhary Ruhi 等研究了一种可定制用户知情合成框架 (FOCUS), 这是一种为药物发现量身定制的生成机器学习模型。FOCUS 集成了领域专业知识, 并使用近端策略优化(PPO)来指导蒙特卡罗树搜索(MCTS), 以有效地探索化学空间。它生成潜在候选药物的 SMILES 表示, 优化其对 NOD2、PEP 和 MCT1 受体的可药物性和结合功效。该模型具有高度的解释性, 允许基于详细周期报告的用户反馈和专家驱动的调整。FOCUS 利用 SHAP 和 LIME 等工具, 提供了一个透明的决策过程分析, 强调对接分数和交互指纹等功能。与 Muramyl 二肽(MDP)的比较研究表明, 其相互作用谱有所改善。FOCUS 将先进的机器学习与专家洞察力相结合, 加速了药物发现流程。

信息来源: Choudhary Ruhi and Mahadevan Radhakrishnan. FOCUS on NOD2: Advancing IBD Drug Discovery with a User-Informed Machine Learning Framework. ACS Med. Chem. Lett. 2024. <https://doi.org/10.1021/acsmchemlett.4c00148>

基于中药微球捕鱼策略和生物层干涉测量技术的药物靶点发现

2024-05-29

北京大学 Zhang, H 等提出了一种结合传统中药微球 (TCM-MPs) 和生物层干涉测量 (BLI) 技术的新方法, 用于发现复杂基质中的关键药物靶点。微球由内向外依次由 Fe_3O_4 磁性纳米层、油酸改性层、光亲和基团 (TAD) 层和活性小分子层组成。TAD 在紫外光下产生高活性的卡宾, 可以实现药物活性小分子的自组装和固定, 具有非选择性。研究以参芪降糖颗粒 (SJG) 为例, 利用构建的中药 MP 对人肾小球系膜细胞 (HMCs) 裂解物的相关蛋白进行了筛选。筛选出 28 个差异蛋白。根据生物信息学的靶点分析, 选择 GNAS 作为关键靶点, 参与胰岛素分泌和 cAMP 信号通路。为了进一步验证 GNAS 与小分子的相互作用效应, 建立了一种基于生物层干涉测量法 (BLI) 与 UHPLC-Q/TOF-MS/MS 相结合的反向捕鱼技术。结果显示, 26 个小分子可能与 GNAS 相互作用, 其中 7 个小分子具有强结合活性。HMCs 的体外实验表明, 7 种活性化合物可以通过与 GNAS 结合显著激活 cAMP 途径。开发的中药 MP 靶向捕捞策略与 BLI 反向捕捞技术相结合, 从复杂的靶蛋白系统中筛选出与活性成分直接相互作用的关键蛋白, 对发现中药复杂系统的药物靶点具有重要意义。

信息来源: Zhang, H, et al. Discovery of drug targets based on traditional Chinese medicine microspheres (TCM-MPs) fishing strategy combined with bio-layer interferometry (BLI) technology. ANALYTICA CHIMICA ACTA, 2024, 1305: 342542

HI-Mol 模型，仅使用训练集 2% 的数据即可实现分子生成

2024-07-11

寻找新分子一直是化学中一个基本而关键的问题，因为它与实现重要应用(如药物发现和材料设计)有着密切的关系。基于生成模型最近在其他领域(例如图像和视频)在学习高维数据分布方面突破的启发，一些研究已经考虑使用大型分子数据集训练深度生成模型来学习分子分布。尽管分子生成方法在基于大规模数据集来寻找化学上有效的新的分子的任务上表现出优秀的性能，然而，现有的生成模型在小样本数据集的分子生成任务上仍面临挑战。

2024 年 5 月 2 日，韩国科学技术院 Seojin Kim 等人在人工智能顶级会议 ICML2024 上发表文章 Data-Efficient Molecular Generation with Hierarchical Textual Inversion。作者提出了一种新的分子生成方法，分子生成层次文本反转(Molecular generation with Hierarchical textual Inversion, HI-Mol)。具体而言，作者提出了一种分子专用的文本反转方案来捕获分子的层次信息。与在给定训练数据上优化单个共享标记的视觉域文本反转不同，作者为分子文本反转设计了多级标记。该方法广泛地利用了给定分子的信息，有效地缓解了分子数量有限的问题。HI-Mol 仅使用训练集 2% 的数据即可实现高效的分子生成，超越了现有的基线方法，有望在高效生成分子方面开启尚未被探索但至关重要的研究方向。这项工作将促进分子生成的研究，并可以加快许多重要生成任务的发展，例如新药物的发现与设计。

信息来源：Kim et al. Data-Efficient Molecular Generation with Hierarchical Textual Inversion. ICML. 2024;
AIDD Pro. <https://mp.weixin.qq.com/s/WQX8XmEcddZfwqYjPyezKQ>

Nat Comput Sci | MISATO: 蛋白质-配体复合物的机器学习数据集用于基于结构的药物发现

2024-06-12

2024年5月10日, Grzegorz M. Popowicz 团队在 Nature Computational Science 发表文章 MISATO: machine learning dataset of protein - ligand complexes for structure-based drug discovery。作者提出了一个实验验证过的蛋白质-配体结构的数据库 MISATO (molecular interactions are structurally optimized, 分子相互作用结构优化), 展示了该数据库有助于更好地训练与药物发现及其他领域相关的模型, 包括量子化学、通用结构生物学和生物信息学。作者提供基于量子化学的结构整理和优化, 包括配体几何结构的正则化。作者为数据库补充了缺失的动态和化学信息, 包括在时间尺度上允许检测某些系统的瞬态和隐蔽状态的分子动力学模拟。这些对于成功的药物设计非常重要。因此, 作者用尽可能多的物理参数补充实验数据, 减轻 AI 模型隐式学习所有这些信息的负担, 从而集中于主要的学习任务。MISATO 数据库提供用户友好的格式, 可以直接导入机器学习代码。作者还提供了各种预处理脚本来筛选和可视化数据集, 提供了 AI 基线模型示例, 这些示例包含计算量子化学性质 (化学硬度和电子亲和力)、结合亲和力计算、蛋白质柔性或诱导契合特征预测。这些量子力学 (QM)、分子动力学 (MD) 和 AI 基线模型在实验数据上进行了广泛验证。作者希望将 MISATO 转变为一个雄心勃勃的社区项目, 对整个药物发现领域产生深远影响。。

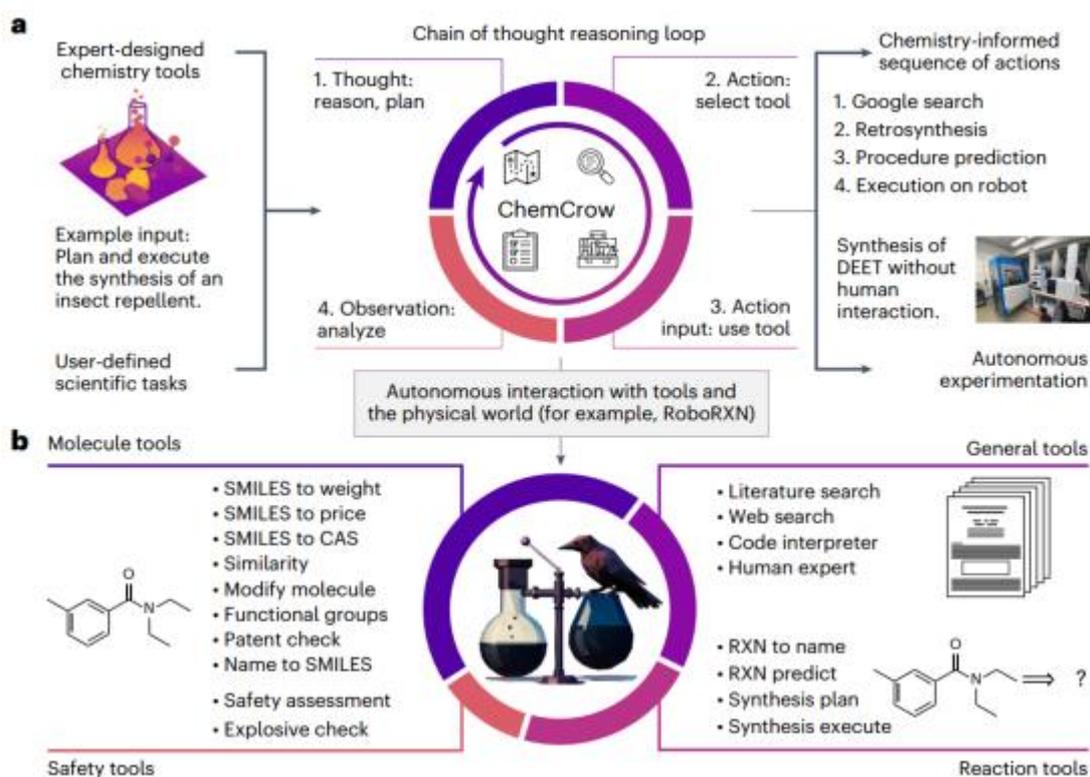
信息来源: 智药邦. Nat Comput Sci | MISATO: 蛋白质-配体复合物的机器学习数据集用于基于结构的药物发现. https://mp.weixin.qq.com/s/m4a_N_mtxLMf0DjXaSKtVA

Nat Mach Intell | ChemCrow: 结合大型语言模型与化学工具的应用

2024-06-22

2024年5月8日, Philippe Schwaller 和 Andrew D. White 研究团队在《Nature Machine Intelligence》期刊上发表了一篇题为“Augmenting large language models with chemistry tools”的研究文章。该文章展示了研究团队开发的一种专为化学任务设计的增强型大语言模型(LLM)“ChemCrow”, 能够执行有机合成、药物发现及材料设计等多种任务。

ChemCrow 利用 18 个专门设计的化学工具, 如图 1b 所示, 这些工具不仅赋予 ChemCrow 关于分子和反应属性的知识, 还具有在物理实验室直接执行任务的能力。这种工作流程有效地结合了链式思考推理和与任务相关的工具, 使 LLM 从一个高度自信但通常错误的信息源转变为一个能够反思任务、使用合适工具进行行动并观察响应的推理引擎。



信息来源: 智药邦. Nat Mach Intell | ChemCrow: 结合大型语言模型与化学工具的应用.
https://mp.weixin.qq.com/s/hJsD_3sD54I2tW6rQIKTsA

Cell | 复旦团队利用 AI 技术, 从全球微生物中发现近百万种候选抗生素

2024-06-15

2024 年 6 月 5 日, 复旦大学类脑智能科学与技术研究院与美国宾夕法尼亚大学的研究人员合作, 在国际顶尖学术期刊 Cell 上发表了题为: *Discovery of antimicrobial peptides in the global microbiome with machine learning* 的研究论文。

该研究提出了一种基于机器学习 (Machine Learning) 的方法, 用于预测全球微生物群中的抗菌肽 (AMP), 利用来自环境和宿主相关栖息地的 63410 个宏基因组和 87920 个原核基因组的大量数据集, 识别并创建了一个包含近 100 万种候选抗菌肽序列库——AMPSphere, 研究团队还免费公开了这个抗菌肽序列库。

为了验证预测结果, 研究团队从中合成了 100 中抗菌肽, 并在体外和体内实验中验证了其对抗临床相关耐药病原体和人体肠道细菌的活性, 结果显示, 其中有 79 中抗菌肽有活性, 有 63 种能够特异性地完全抑制一种病原体的生长, 这些活性抗菌肽通过破坏细菌膜发挥抗细菌活性。这项研究证明了人工智能方法从全球微生物组中鉴定功能性抗菌肽的能力。

信息来源: 智药邦. Cell | 复旦团队利用 AI 技术, 从全球微生物中一性发现近百万种候选抗生素 <https://mp.weixin.qq.com/s/qw9MGiBISgt7-eRTjO-HyQ>

Science | David Baker 团队：用计算方法发现化学结构多样的大环分子

2024-05-22

2024 年 4 月 25 日，David Baker 教授团队在 Science 上发表最新研究 *Expansive discovery of chemically diverse structured macrocyclic oligoamides*。介绍了一种用于识别由 α 、 β 、 γ 和其他 17 种氨基酸骨架组成的有序大环分子的计算方法，并利用这种方法预测了由大于 42,000 种单体组合组成的 1,490 万个大环分子。该方法分三步：第一步，使用基于网格的自适应搜索 AIMNet 潜在能量图来识别每个单体的低能构象；第二步，计算与这些单体构象相关的刚体转换，以及由这些单体成对组合构建的所有二肽构象，并将其存储在哈希表中；第三步，通过识别哈希表中刚体变换组合接近于零的条目对，快速、系统地生成大环分子。

本研究成果开辟了易于合成的类药大环分子的广阔空间，将大大促进基于结构的药物设计。

信息来源：智药邦. Science | David Baker 团队：用计算方法发现化学结构多样的大环分子.
<https://mp.weixin.qq.com/s/nqlsvOMQw1mgeHdpyoH8tw>

近期会议与活动

时间	题目	地点	相关链接
18-20 Dec, 2024	Middle East Organic & Natural Product Expo 2024	Dubai, United Arab Emirates	https://www.clocate.com/middle-east-organic-and-natural-product-expo/18595/
24-26 Oct, 2024	The Global Summit on Agriculture and Plant Science (GSAPS 2024)	Bern, Switzerland	https://www.clocate.com/the-global-summit-on-agriculture-and-plant-science-gsaps/101308/
04-05 November, 2024	7th International Conference on Natural Products and Medicinal Plants Research	Toronto, Canada	https://naturalproducts.conferences.com/
10-11 Oct, 2024	10th World Congress on Medicinal Plants and Marine Drugs	Paris, France	https://www.clocate.com/world-congress-on-medicinal-plants-and-marine-drugs/88219/
8月1-4日, 2024	复旦大学药学院团队第三期药物设计培训班	上海	https://mp.weixin.qq.com/s/j3PTCCRNbGzxixxYtwSUlg
8月18-21日, 2024	2024《中国天然药物》青年科学家论坛会议	西宁	https://gz.qhu.edu.cn/xwzx/xsjl/4e3cd061e2634c4398555cc66f501c02.htm

人工智能与天然药物动态监测快报

编写委员会

主 编：庄会富 陆颖

执行主编：田雅娟

编 委：张于 颜秉超 胡坤 卿立燕 杨帅

编 辑：杜宁 田雅娟 黄蓉 史继强

工作联系人：杜 宁 电话：0871-65223176 邮箱：duning@mail.kib.ac.cn

田雅娟 电话：028-81258707 邮箱：tyj@clas.ac.cn