

人工智能与天然药物 动态监测快报

总第 005 期 2024-03-20



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

政策法规	12
政府工作报告首提“创新药”	2
美国参议院委员会交锋中通过“生物安全草案”	2
研究前沿	3
通过 AI/ML 和 CADD 工具在安全、基于 Web 的 SaaS 平台中优化药物发现	3
《自然-生物技术》: 从靶点发现直至 2 期临床, AI 如何加速潜在”first-in-class “疗法开发	5
D3Rings: 一种快速准确的环系统识别和类药物环状化合物深度生成方法	7
将深度学习 QSPR 模型适用于特定的药物发现工程	8
基于片段的综合药物设计和虚拟筛选技术探索 thiazolidine -2,4-diones 的抗糖尿病潜力: 设计、合成和体内研究	9
发现和优化天然产物启发的 SARS-CoV-2 2'-O-甲基转移酶抑制剂的信息学和计算方法	10
将人工智能应用于药物发现: 工具、技术和应用	10
人工智能驱动的针对 SARS-CoV-2 主要蛋白酶的共价药物设计策略: 结构见解和分子机制	11
人工智能用于靶向多药治疗: 药物发现的下一个前沿	11
靶向泛素特异性蛋白酶 7 的天然产物的虚拟筛选	12
SiteMine: 蛋白质结构数据库中的大规模结合位点相似性搜索	12
近期会议与活动	12

政策法规

政府工作报告首提“创新药”

2023-03-08

2024年3月5日，十四届全国人大二次会议5日上午在人民大会堂开幕，国务院总理李强代表国务院，向十四届全国人大二次会议作政府工作报告。

在“积极培育新兴产业和未来产业”部分，李强总理表示：实施产业创新工程，完善产业生态，拓展应用场景，促进战略性新兴产业融合集群发展。巩固扩大智能网联新能源汽车等产业领先优势，加快前沿新兴氢能、新材料、创新药等产业发展，积极打造生物制造、商业航天、低空经济等新增长引擎。制定未来产业发展规划，开辟量子技术、生命科学等新赛道，创建一批未来产业先导区。

值得注意的是，这是“创新药”一词，首次进入政府工作报告。在此前的政府工作报告中，更多是提及到推进“生物医药”等新兴产业的发展，促进产业升级。

信息来源：<https://news.yaozh.com/archive/42148.html>

美国参议院委员会交锋中通过“生物安全草案”

2023-03-08

2024年3月6日早上9点30分（美东时间），美国参议院国土安全委员会听证会举行。此次听证会中，编号为“S3558”的法案引起关注，即《生物安全法案》的修正案，法案目的是防止外国对手窃取美国基因数据和个人健康信息，提到了华大基因和药明康德。

尽管有反对意见，这个修正案还是通过了听证会决议。即便，这项法案要落实，还要经过参议院全体会议通过；众议院委员会及全体会议通过后，到达国会；最终总统签署后才落地，但此次推进仍引起关系人士的担忧。

信息来源：<https://news.yaozh.com/archive/42180.html>

研究前沿

通过 AI/ML 和 CADD 工具在安全、基于 Web 的 SaaS 平台中优化药物发现

2024-01-25

AIDDISON 是一款由人工智能驱动的药物发现平台。该平台通过充分整合计算辅助药物设计 (CADD) 工具和 AI 技术, 实现了对新颖分子的虚拟筛选和生成。通过各种来源, 如公共数据库、虚拟化学集合和生成模型, AIDDISON 可以识别或生成数千个可行的分子起点。通过基于性质的过滤, 选择最有成功概率的分子进行进一步研究, 通过对接和形状对比来评估其生物活性。最终, 有潜在活性的分子设计可送入 SYNTHIA 逆合成软件评估其可合成性和所需试剂。

AIDDISON 的目标是使科学家能够更迅速、更高效地进行药物发现。其平台的开发仍在积极进行, 未来版本将引入更多功能。同时, AIDDISON 注重信息安全, 遵循 ISO 27001 标准, 为数字产品提供最高级别的信息安全认证。

AIDDISON 的搜索功能涵盖了多个方面, 包括 2D 相似性搜索、2D 药效团搜索、De Novo 设计 (生成模型) 和基于形状搜索。

1. 2D 相似性搜索: 利用 FPSim2, 通过对结构特征 (ECFP) 编码成分子指纹进行比较, 快速在大量化合物中找到与目标分子相似的结构。结果从包括 Sigma-Aldrich 目录、PubChem、ChEMBL 和 ZINC 等多个数据库中返回, 保证了结果的标准化。

2. 2D 药效团搜索: 2D 药效团搜索通过比较 BioSolveIT 的 Feature Trees (FTrees) 计算与目标分子的相似性。这是一种非常高效且有效的工具, 用于进行支架跳跃和基于配体的筛选, 适用于庞大的虚拟化学空间。

3. De Novo 设计 (生成模型): 基于生成式 AI 模型, 使用 REINVENT 2.0 从目标分子生成具有期望化学性质的虚拟分子。强化学习奖励函数包括药物相似性 (QED) 和与目标结构的相似性, 同时还评估可合成性。用户可通过自定义模型扩展评分组件, 最佳分数的分子返回供进一步分析。

4. 基于形状搜索: 基于 Cresset's Flare, 通过 3D 叠合评估新颖分子的结合亲

和性。使用形状和基于性质的集团对比模式，与已知活性的参考分子进行叠合和评分（例如图 2）。可通过 Gallery View 可视化 3D 结构对齐结果。

5.分子对接：AIDDISON 采用基于 Flare Docking 工作流程的分子对接方法（图 3）。用户可以通过 PDB 的 API 或上传 PDB 格式的蛋白质结构，选择口袋的方式包括算法、手动指定或基于结晶结合的配体周围的口袋。利用 Flare Docking 工作流程，系统自动添加氢原子和电荷，为待对接分子生成多个低能构象。通过 LF RankScore 评估，选取最优位构，该能量函数专为实现与已报告的晶体配体相似的构象而调整。对接结果在 Gallery View 中可视化，展示新分子在口袋内的位置，并显示蛋白质的溶剂可及表面和参考配体。计算评估通过 Table View 进行，用户可按 VSScore 和 Flare 软件的“结合自由能”等多种计算量排序，辅助选择具有最佳性能的候选物。

信息来源：<https://baijiahao.baidu.com/s?id=1789119988615096355>

推荐评论：AIDDISON 是一款结合了计算辅助药物设计工具和 AI 技术的药物发现平台。该平台整合了公共数据库、虚拟化学集合和生成模型，能够识别或生成数千个可行的分子起点，提供例如包括 2D 相似性搜索、2D 药效团搜索、De Novo 设计、基于形状搜索等功能，为用户提供多种药物设计和评估工具。

《自然-生物技术》：从靶点发现直至 2 期临床，AI 如何加速潜在” first-in-class “疗法开发

2024-03-11

2024 年 3 月 8 日，临床阶段生成式 AI 驱动的生物科技公司英矽智能在《自然-生物技术》(Nature Biotechnology) 杂志发表研究论文，阐述由其生成式 AI 发现和设计的潜在全球首创 TNIK 抑制剂 INS018_055 从算法开发走到 2 期临床试验的历程，并首次披露该药物的临床前和 1 期临床试验数据。

英矽智能的研究团队首先利用组织纤维化相关的组学和临床数据集，训练其靶点发现平台 PandaOmics。通过深度特征合成、因果关系推断等过程，前述平台提名潜在靶点列表。随后，PandaOmics 中的自然语言处理 (NLP) 模型通过分析涵盖专利、出版物、研发基金、临床试验等文本数据的数百万个文本文件，进一步评估潜在靶点的新颖性以及与其疾病的关联性，TNIK 被确定为最有潜力的抗纤维化靶点。

既往研究曾揭示 TNIK 与多种纤维化驱动生物通路存在间接关联，但从未提出将其作为特发性肺纤维化 (IPF) 治疗靶点。TNIK 靶点还与衰老的多种标志物有关，可能具有更广泛的治疗潜力。

确定 TNIK 靶点后，英矽智能研发团队利用生成化学引擎 Chemistry42，根据基于结构的药物设计 (SBDD) 策略生成具有所需特性的创新分子结构。经过多次迭代筛选，最终获得候选分子 INS018_055，整个过程共合成并测试了不到 80 个分子。

据悉，Chemistry42 结合 40 多种生成化学算法和超过 500 个预训练的奖励模型，支持新颖化合物从头生成，能根据专家反馈进行虚拟筛选并优化生成结果。

为评估 INS018_055 的抗纤维化活性，研究团队使用肺纤维化小鼠模型进行测试。用 INS018_055 治疗两周后，效果与尼达尼布一致。研究团队还在肺纤维化小鼠模型中测试了 INS018_055 和吡非尼酮联合治疗的效果，两周给药后，发现小鼠模型肺功能显著改善。他们还观察到，INS018_055 减少了肺纤维化小鼠模型的肺部炎症。此外，INS018_055 还在另外两种动物模型中减轻了皮肤和肾脏纤维化。

发现 TNIK 靶点仅 18 个月后，2021 年 2 月，INS018_055 被提名为临床前候选化合物。2021 年 11 月，INS018_055 在澳大利亚的首次人体微剂量试验中完成首批健康受试者给药。英矽智能称，该项人体微剂量试验结果超出预期，展现候选药物良好的药代动力学和安全性特征，不仅完成了 AI 制药临床概念验证，还为后续临床试验奠定了基础。

随后，英矽智能在新西兰和中国启动 INS018_055 的 I 期临床试验，分别纳入 78 名和 48 名健康受试者，完成单次剂量递增（SAD）和多次剂量递增（MAD）队列研究，均表明 INS018_055 具有良好的安全性、耐受性和药代动力学特征，支持后续 2 期临床试验的开展。

信息来源：https://www.sohu.com/a/763371916_120546887

推荐评论：英矽智能的成功例子进一步表明 AI 辅助药物研发是大势所趋，我们必须及时调整现有的研究模式。目前，药物研发底层逻辑正从经验驱动向数据驱动转变。高质量数据将成为药物研发新的制高点（英矽智能的成功很大程度上源于 in-house 的数据）。高质量数据如何而来，在我看来算法并不直接产生数据（至少现在不行），实验数据，或者说独有的数据，才是关键。对于天物药物研发而言，我们要进一步挖掘新颖的活性分子，阐明它们的发作用靶点，解析它们与靶蛋白相互作用的结构基础，发展衍生物，为训练 AI 提供独一无二的高质量数据，推动新颖先导分子的发现及创新药物研发进程。

D3Rings:一种快速准确的环系统识别和类药物环状化合物深度生成方法

2024-01-11

不断探索分子的化学空间，寻找对特定靶点具有高亲和力和特异性的配体是药物发现的一个重要课题。环状化合物，特别是具有不同支架的天然化合物的关注，为药物设计提供了对新型分子结构的重要见解。然而，它们的环结构的复杂性阻碍了广泛接受的用于环状化合物的系统鉴定和分类的方法和软件的适用性。Ma, Minfei 等开发了一种新的方法 D3Rings，用于鉴定无环、单环、螺环、稠环和桥环、笼环化合物以及大环化合物。通过使用 D3Rings，我们在 ChEMBL、DrugBank 和 COCONUT 三个不同的数据库中完成了环状化合物的统计。结果表明，天然产物中的环结构丰富，特别是螺环、大环、稠环和桥环。在此基础上，训练了三个深度生成模型，即 VAE、AAE 和 CharRNN，并用于构建两个类似于 DrugBank 和 COCONUT 但比它们大 10 倍的数据集。然后，放大的数据集被用于探索分子化学空间，重点关注复杂的环结构，用于新药的发现和开发。用新生成的针对三种严重急性呼吸系统综合征冠状病毒 2 型靶蛋白的类 COCONUT 数据集进行对接实验表明，扩展的化合物数据库改善了分子对接结果。在排名靠前的对接分子中，环状结构表现出最好的对接得分。这些结果表明，探索结构新颖的环状化合物的化学空间和不断扩大类药物化合物库的重要性，以促进发现对特定靶标具有高结合亲和力的强效配体。D3Rings 现已可在 <http://www.d3pharma.com/D3Rings/> 免费使用。

信息来源: Ma

MF, et al. D3Rings: A Fast and Accurate Method for Ring System Identification and Deep Generation of Drug-Like Cyclic Compounds. JOURNAL OF CHEMICAL INFORMATION AND MODELING, 2024, 64(3): 724-736

推荐评论: 基于深度学习驱动的天然产物虚拟结构、光谱数据生成有助于弥补天然产物实验数据规模小、成本高等缺点，对于指导以天然产物为基础的结构和活性成分研究具有重要意义。

将深度学习 QSPR 模型适用于特定的药物发现工程

2024-02

将分子结构与化合物性质联系起来的机器学习 (ML) 模型可以帮助药物化学和药物设计工作。在制药行业, 这些 ML 全球模型可以作为一种“开箱即用”的解决方案在发现项目中使用, 以帮助药物设计、合成优先级和实验选择。然而, 药物发现项目通常侧重于化学空间的有限部分 (如化学系列), 在这些部分, 全球模型可能不适用。有时会生成局部 ML 模型来关注特定的项目或系列。Fluetsch, Andrin 等对基于 ML 的全局模型、局部模型和混合全局局部策略进行了基准测试。对诺华公司 300 多个药物发现项目和 10 个吸收、分布、代谢和排泄 (ADME) 试验进行了分析。在这项工作中, 提出了基于迁移学习方法的混合全局-局部策略, 以利用历史 ADME 数据 (全局) 和项目特定数据 (局部) 来适应模型预测。微调预训练的全局 ML 模型 (用于权重初始化, WI) 是性能最好的方法。与全局和局部模型相比, 所有测定的平均绝对误差的平均改善分别为 16% 和 27%。有趣的是, 当分析训练集大小的影响时, 发现即使在低数据场景下 (例如, 类似于每个项目 10 个分子), WI 微调也是成功的。总之, 这项工作强调了结构域自适应在分子性质预测领域的潜力, 以完善新化合物数据分布的现有预训练模型。

信息来源:

Fluetsch A, et al. Adapting Deep Learning QSPR Models to Specific Drug Discovery Projects, MOLECULAR PHARMACEUTICS, 2024, DOI: 10.1021/acs.molpharmaceut.3c01124

推荐评论: 通过迁移学习方法, 研究提出了混合全局-局部策略, 利用历史数据和项目特定信息来预测分子药代动力学性质。

复杂阳离子重排结果的计算预测

2024-01-18

近年来，人们对计算机辅助有机合成的兴趣再次高涨。Klucznik, T 等展示了配备了由物理有机化学规则以及量子力学和动力学计算增强的机械步骤综合知识库的计算机，可以使用反应网络方法来分析一些最复杂的有机转化的机制：即阳离子重排。这种重排导致分子的碳骨架发生显著变。研究者通过三组实验验证了这一算法（1）预测尾部至头部萜烯（THT）环化的结果，其中在微小结构细节不同的模块化前体中编码了显著不同的结果；（2）比较溶液中或超分子胶囊中 THT 环化的结果；（3）分析复杂的反应混合物。研究结果支持了一个愿景：计算机不再仅仅操纵已知的反应类型，而是将有助于合理化和发现新的、机械上复杂的转换。

信息来源：Klucznik T, et al. Computational prediction of complex cationic rearrangement outcomes. Nature, 2024, 625(7995) DOI10.1038/s41586-023-06854-3

基于片段的综合药物设计和虚拟筛选技术探索 thiazolidine -2,4-diones 的抗糖尿病潜力：设计、合成和体内研究

2023-12-05

Gupta 等研究了利用和整合基于片段的药物设计和虚拟筛选技术，探索新合成的噻唑烷-2,4-二酮衍生物的抗糖尿病潜力。该研究涉及通过基于片段的药物设计来设计噻唑烷-2, 4-二酮化合物的新变体。筛选过程包括通过对接算法进行基于药效团的虚拟筛选，以及鉴定新的 12 种得分最高的化合物。验证显示化合物 SP4e、SP4f 有突出的抗糖尿病药物的潜力，值得进一步探索其治疗应用。

信息来源：Gupta S, et al. Integrated fragment-based drug design and virtual screening techniques for exploring the antidiabetic potential of thiazolidine-2,4-diones: Design, synthesis and in vivo studies. EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY, 2023, 261: 115826

发现和优化天然产物启发的 SARS-CoV-2 2'-O-甲基转移酶抑制剂的 信息学和计算方法

2024-01

Hanna, George S.等研究了一种使用计算分析的方法，进行合理的药物设计，以识别和开发受天然产品启发的抗病毒药物。这样做的目的是通过靶向高度保守的 RNA 2'-O-甲基转移酶 (2'-O-MTase)，鉴定天然产物原型，以帮助开发安全、有效和负担得起的 Betacoronavirus 复制广谱抑制剂。Machaeriols RS-1 (7) 和 RS-2 (8) 是使用先前概述的信息学方法鉴定的，首先筛选天然产物原型，然后进行计算机引导合成。这两种分子都基于一种罕见的天然产物基团。从亚马逊特有的 *Machaerium* 属中分离出的 machaeriols (3-6) 比阳性对照 Sinefengin (2) 更有效地抑制严重急性呼吸系统综合征冠状病毒 2'-O-MT 酶，并且在计算机建模中表明了不同的分子相互作用。文中强调了计算驱动筛选的潜力，以利用自然产物库并提高隔离或合成模拟物开发的效率。

信息来源: Hanna GS, et al. Informatics and Computational Approaches for the Discovery and Optimization of Natural Product-Inspired Inhibitors of the SARS-CoV-2 2'-O Methyltransferase. JOURNAL OF NATURAL PRODUCTS, 2024, 87(2): 217-227

将人工智能应用于药物发现：工具、技术和应用

2024-3-10

Agu, Peter Chinedu 等对于人工智能 (AI) 方法对药物开发领域的影响进行了综述，主要聚焦于工智能模型和数据资源如何通过为传统方法提供更实惠、更方便的选择来重塑药物开发过程。

文中综述了药物开发的知名信息来源，分子表示技术，及基于人工智能创建药物发现模型的算法。详细介绍了人工智能算法如何用于预测药物毒性、药物生物活性和药物物理化学性质，以及从头药物设计、结合亲和力预测和其他基于人工智能的药物-靶标相互作用模型等。

信息来源: Agu PC, et al. Piquing artificial intelligence towards drug discovery: Tools, techniques, and applications, DRUG DEVELOPMENT RESEARCH, 2024, 85 (2): e22159

人工智能驱动的针对 SARS-CoV-2 主要蛋白酶的共价药物设计策略：结构见解和分子机制

2024-01

Abadi, MHHE 等综述探讨了 Mpro 抑制剂的研究进展以及人工智能在药物设计中的应用。人工智能利用庞大的数据集和先进的算法来简化有前景的 Mpro 抑制剂的设计和鉴定。人工智能驱动的药物发现方法，包括分子对接、预测建模和基于结构的药物再利用，处于识别有效抗病毒治疗潜在候选药物的前沿。在新冠肺炎可能威胁全球健康的时代，寻求针对 Mpro 的有效抗病毒解决方案可能对抑制病毒至关重要。

信息来源：Abadi MHHE, et al. AI-driven covalent drug design strategies targeting main protease (m^{pro}) against SARS-CoV-2: structural insights and molecular mechanisms, JOURNAL OF BIOMOLECULAR STRUCTURE & DYNAMICS 2024, 23087691

人工智能用于靶向多药治疗：药物发现的下一个前沿

2024-01

Cichonska, Ann 等认为在药物发现方面，靶向多药学正在重新定义治疗设计。预先选择的药理学特征，如激酶药物，有望增强疗效和降低毒性。人工智能（AI）提供了一种变革性的解决方案。机器学习和深度学习技术能够建模蛋白质结构，生成新化合物，并解码其多药物效应，为更系统、更具预测性的多靶点药物设计开辟了途径。他们综述探讨了人工智能在识别协同共靶点并将其与导致不良反应的反靶点区分开来方面的用途，然后讨论了在多药生态学背景下，人工智能实现的对接、生成化学和蛋白质组化合物相互作用的蛋白化学计量建模方面的进展，并分析了未来挑战。

信息来源：Cichonska A, et al. AI for targeted polypharmacology: The next frontier in drug discovery, CURRENT OPINION IN STRUCTURAL BIOLOGY, 2024, 84: 102771

靶向泛素特异性蛋白酶 7 的天然产物的虚拟筛选

2024-02-25

Li, Hongju 等基于 watvina 分子对接的虚拟筛选方法和生物学评价,发现了新的 USP7 抑制剂靶向催化活性位点。从 3760 种天然产物中筛选出三种命中物,并通过酶促和动力学分析验证其为 USP7 抑制剂。三种天然产物适合作为开发新型 USP7 抑制剂的先导化合物,尤其是抗克隆癌症药物。

信息来源: Li, HJ, et al. Virtual screening of natural products targeting ubiquitin-specific protease 7. JOURNAL OF BIOMOLECULAR STRUCTURE & DYNAMICS, 2024.
DOI10.1080/07391102.2024.2316779

SiteMine: 蛋白质结构数据库中的大规模结合位点相似性搜索

2024-02-25

蛋白质结合位点相似性分析可以解决药物发现和设计方面的难题,如药物再利用、分析蛋白质-配体和蛋白质-蛋白质复合物、配体杂合性研究或功能预测。虽然已有许多工具,但它们在运行时间、提供结构叠加以及对不同应用领域的适用性等方面各有优缺点。SiteMine 是一种一体化的数据库驱动、提供比对的结合位点相似性搜索工具。作者在 ProSPECCTs 基准上对 SiteMine 的性能进行了评估,结果表明它在大多数数据集上都表现出了良好的性能。该方法为基于结合位点对比的结构化分子设计提供了最先进的新方法。作者应用 SiteMine 通过筛选类似的配体结合口袋,帮助预测新型结合位点的潜在结合体。预测的结合位点和配体定义的结合位点都具有可靠的性能。适用于与已知结构相似度较低的蛋白质。

SiteMine 可应用于选择性分析和发现已知药物的新靶点。

信息来源:

https://mbd.baidu.com/newspage/data/landingsuper?context=%7B%22nid%22%3A%22news_10112402070699183659%22%7D&n_type=1&p_from=3;

Reim T, Ehrh C, Graef J, et al. SiteMine: Large - scale binding site similarity searching in protein structure databases. Archiv der Pharmazie, 2024: e2300661

近期会议与活动

时间	题目	地点	相关链接
25-26 Apr 2024	20th International Congress on Advances In Natural Medicines, Nutraceuticals & Neurocognition	London, United Kingdom	https://www.clocate.com/international-congress-on-advances-in-natural-medicines-nutraceuticals-and-neurocognition/55207/
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/
26-27 Apr 2024	2nd Global Conference on Pharmaceutical Sciences	Japan(Hybrid)	https://www.clocate.com/global-conference-on-pharmaceutical-sciences/90644/
22-24 May 2024	Drug Design & Medicinal Chemistry Congress	Basel, Switzerland	https://www.clocate.com/drug-design-and-medicinal-chemistry-congress/36332/
23-24 May 2024	11th International Conference on Natural, Traditional & Alternative Medicine	Zurich, Switzerland	https://www.clocate.com/international-conference-and-expo-on-natural-traditional-and-alternative-medicine/72721/
02-03 May 2024	23rd Drug Discovery Innovation Programme	Boston, MA, United States	https://www.clocate.com/drug-discovery-innovation-programme/84002/
02-03 May 2024	ICETM 2024: International Conference on Ethnomedicine and Traditional Medicine	Amsterdam, Netherlands (Hybrid)	https://www.clocate.com/icetm-international-conference-on-ethnomedicine-and-traditional-medicine/56851/
2024年4月 12-13日	2024第二届合成生物学产业嘉年华暨展览会	南京	https://www.instrument.com.cn/news/20231114/691866.shtml

特别鸣谢（快报信息评论专家）

姓名	单位（按姓氏拼音排序）
年 寅	中国科学院昆明植物研究所
胡贵林	中国科学院昆明植物研究所

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