



政务微信

### 我国学者在人工智能药物设计上取得新进展

日期 2023-12-01 来源: 交叉科学部 作者: 甘智 姜杰 杜鑫 【大中小】 【打印】 【关闭】

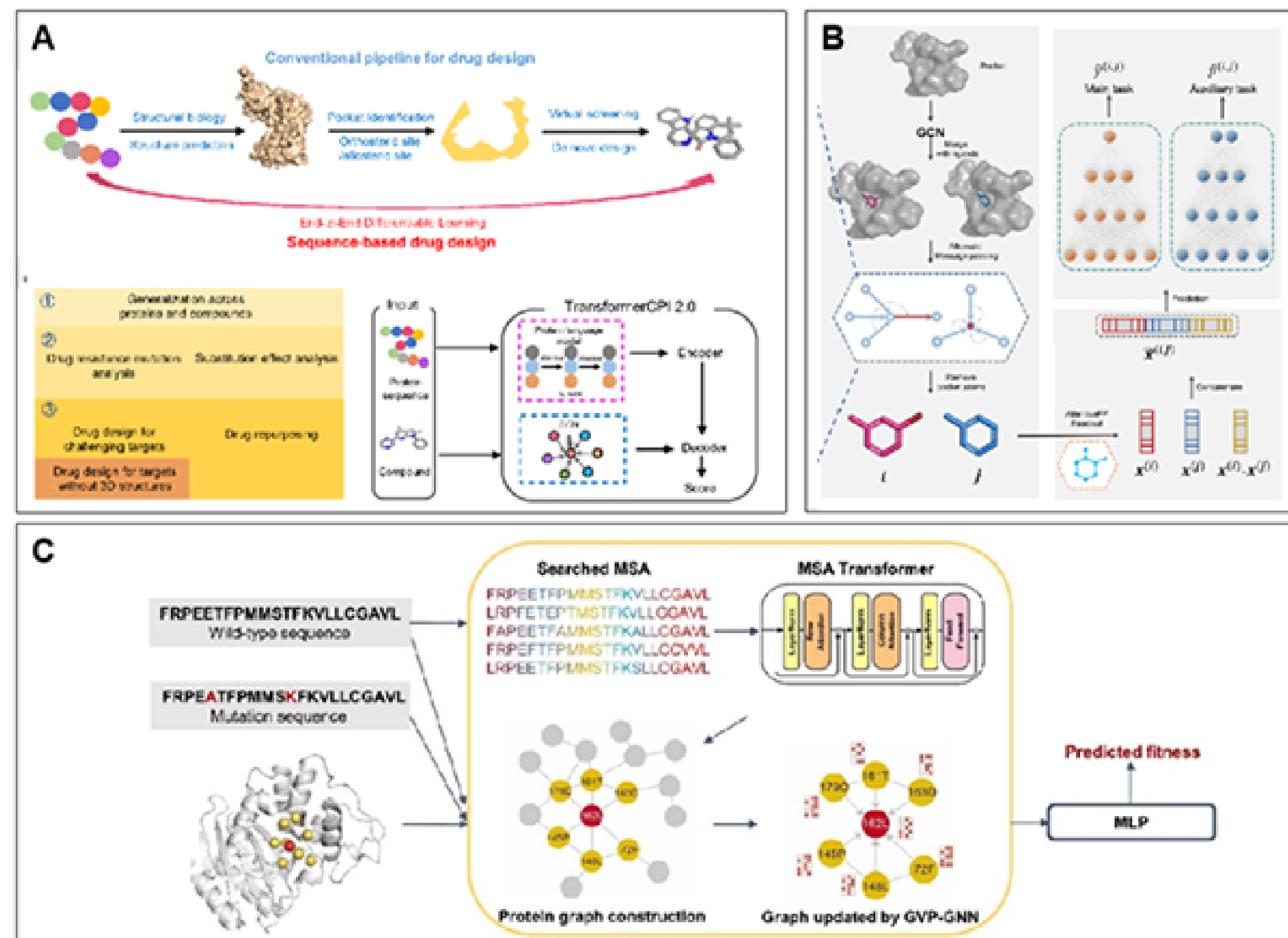


图 人工智能药物设计新方法用于 (A) 先导化合物发现、(B) 先导化合物优化、(C) 多肽药物设计

在国家自然科学基金项目(批准号: T2225002)等资助下,中国科学院上海药物研究所郑明月研究团队近期在药物设计领域取得了研究进展。通过深度学习和大数据分析,人工智能能够高效处理和解读大规模的生物信息学数据,挖掘隐藏在庞大数据集的模式和关联,提高对潜在药物靶点的识别准确性,加速药物筛选和设计的过程。针对过去计算机药物设计中存在的问题和局限,郑明月研究团队利用先进的人工智能方法,围绕先导化合物发现、先导化合物优化以及多肽药物设计,为我国在人工智能辅助药物设计领域的发展贡献了新的理念和方法,取得了以下进展。

#### 1. 先导化合物发现: 基于序列的药物设计新范式

基于蛋白质结构的药物设计(SBDD)一般涉及多个步骤的复杂流程,有其局限性,例如许多蛋白质没有高分辨率的结构,难以获得精确的配体结合位点等。虚拟筛选的复杂流程中的误差会不断累积,将导致严重的假阳性问题。相比之下,基于序列的方法可以作为基于结构药物设计的替代方法,在无法获得高质量蛋白质3D结构的场景中发挥作用,也可与其他虚拟筛选和高通量体外筛选技术结合,加速药物发现进程。

郑明月研究团队提出了一种基于序列的药物设计新概念,采用Transformer神经网络架构和预训练蛋白质语言模型的模型TransformerCPI2.0。利用Transformer神经网络架构和预训练蛋白质语言模型,设计了直接从蛋白质序列出发进行药物设计发现的模型TransformerCPI2.0(图A)。基于模型在多种具有挑战性的药物靶标预测结果,研究团队发现了E3泛素连接酶接头蛋白SPOP的新胥架抑制剂和靶向E3泛素连接酶RNF130的结合分子。逆向应用TransformerCPI2.0对已上市药物进行蛋白质组范围的靶标筛选,成功发现雷贝拉唑抗肿瘤作用的潜在靶标ADP-核糖基化因子ARF1。“从序列到药物”的端到端学习策略可以有效地简化建模和筛选流程,更加高效地探索未知化学和生物学空间,为原创药物的发现提供新的出发点和切入点。

#### 2. 先导化合物优化: PBCNet模型指导先导化合物的结构优化

先导化合物的结构优化(Lead optimization)是药物设计的核心环节,需要通过DMTA(设计-合成-测试-分析)的反复循环来提高化合物活性、特异性、成药性等性质。长期以来,该过程高度依赖药物化学家的经验以及大量的人力和资源投入。

郑明月研究团队提出了全新人工智能方法PBCNet(pairwise binding comparison network)用于指导先导化合物的结构优化(图B)。该方法采用孪生图卷积神经网络架构,通过比较一组相似配体的结合模式差异来预测二者之间的相对结合亲和力,可以较好地兼顾计算速度和精度。研究团队设计了多种测试场景,发现PBCNet模型的排序能力和预测精度均显著优于Schrödinger Glide、MM-GB/SA和多个近期报道的深度学习模型,在部分测试体系上的误差已接近于化学精度(1kcal mol<sup>-1</sup>)。在仅使用少量的结构活性数据对模型进行微调后PBCNet的排序性能已经可以接近或超过Schrödinger FEP+,并且在计算速度方面相对FEP+有多个数量级的优势。对比结果显示,使用PBCNet可以使先导化合物优化项目平均加速约4.7倍,平均节省约30%的资源投入。

#### 3. 多肽药物设计: GVP-MSA模型辅助蛋白质定向进化

定向进化是生产具有改良或全新性质蛋白质的一种有效方法。传统的定向进化实验流程包括反复的诱变和高通量筛选,实验费时费力。机器学习辅助定向进化的方法的难点在于使用有限的蛋白质突变效应实验数据建立准确的蛋白质适应度图景,即目标蛋白的序列突变体-功能的函数映射关系。

郑明月研究员和廖苍松研究员团队提出了一种迁移学习深度学习神经网络模型GVP-MSA(图C)。模型利用预训练的蛋白质语言模型处理目标蛋白的同源序列(MSA, multiple sequence alignment)信息,利用E-(3)等变的图神经网络提取蛋白质三维结构信息,使用多任务学习的方式有效地学习整合不同维度、不同功能的蛋白质数据,从而泛化到新的目标蛋白体系。模型适用于多种测试场景:单点突变效应的随机和按位置外推,对新蛋白质突变效应的零样本预测,以及由单点突变效应预测多点突变效应,均有良好的表现,验证了适应度图景迁移学习的有效性。这一工作为机器学习辅助定向进化提供了新的思路,有助于更加高效地探索蛋白质序列突变空间,快速设计具有改良或全新特性的蛋白质序列。

综合而言,这一系列研究成果标志着我国在人工智能辅助药物设计领域的不断创新和进步。这些新方法的引入为药物设计提供了更高效、更精准的工具,为未来的药物研发和创新奠定了坚实基础。研究成果发表在多个高水平国际期刊,包括《自然·通讯》(Nature Communications)、《自然·计算科学》(Nature Computational Science)、《细胞·系统》(Cell Systems)杂志等。

机构概况: 概况 职能 领导介绍 机构设置 规章制度 专家咨询 评审程序 资助格局 监督工作

政策法规: 国家科学技术相关法律 国家自然科学基金条例 国家自然科学基金规章制度 国家自然科学基金发展规划

项目指南: 项目指南

申请资助: 申请受理 项目检索与查询 下载中心 代码查询 常见问题解答 科学基金资助体系

共享传播: 年度报告 中国科学基金 大数据知识管理服务 平台 优秀成果选编

国际合作: 通知公告 管理办法 协议介绍 进程简表

信息公开: 信息公开制度 信息公开管理办法 信息公开指南 信息公开工作年度报告 信息公开目录 依申请公开

相关链接 政府 新闻 科普



版权所有: 国家自然科学基金委员会 京ICP备05002826号 京公网安备 11040202500066号  
地址: 北京市海淀区双清路9号 邮编: 100085



政府网站  
找错