

基于人工智能驱动分子工厂技术的 Menin 抑制剂优化

曾 浩¹, 吴国振¹, 邹武新¹, 王 哲^{1,2}, 宋剑飞¹, 施 慧¹, 汪小涧³,
侯廷军^{1,2*}, 邓亚峰^{1,4**}

(¹ 杭州碳硅智慧科技发展有限公司, 杭州 310018; ² 浙江大学药学院, 杭州 310058; ³ 中国医学科学院、北京协和医学院药物研究所, 天然药物活性物质与功能国家重点实验室, 活性物质发现与适药化研究北京市重点实验室, 北京 100050;

⁴ 清华大学自动化系, 北京 100084)

摘要 以深度学习为代表的新一代人工智能技术已经成为推动新药研发的重要驱动力。本文创造性地提出了一种基于人工智能技术的创新药物分子设计和优化工作流程“分子工厂”, 该流程融合了自主研发的智能分子生成模型、高性能分子对接算法以及高精度亲和力预测方法, 已作为核心模块被整合进一站式药物设计软件平台 DrugFlow, 为先导化合物发现和优化提供了一整套成熟的解决方案。利用“分子工厂”模块, 针对 Menin 蛋白开展了抗耐药第 2 代抑制剂的研发。通过计算和实验的结合, 快速获得多个潜力化合物, 其中化合物 RG-10 对 Menin 野生型、M327I 突变体和 T349M 突变体的 IC_{50} 分别为 9.681 nmol/L、233.2 nmol/L 和 40.09 nmol/L; 与已进入Ⅱ期临床的阳性参照分子 SNDX-5613 相比, 其对 M327I 和 T349M 突变体的抑制活性显著提升。上述研究充分展现了“分子工厂”技术在新药研发项目中的独特优势, 能快速高效地针对特定蛋白结构产生高质量的活性分子, 对推动新药研发具有重大价值和深远意义。

关键词 分子工厂; 人工智能; 分子生成; 分子对接; Menin 抑制剂

中图分类号 TP181; R914 文献标志码 A 文章编号 1000-5048(2024)03-0326-09

doi: 10.11665/j.issn.1000-5048.2024040904

引用本文 曾浩, 吴国振, 邹武新, 等. 基于人工智能驱动分子工厂技术的 Menin 抑制剂优化 [J]. 中国药科大学学报, 2024, 55(3): 326–334.

Cite this article as: ZENG Hao, WU Guozhen, ZOU Wuxin, et al. Optimization of Menin inhibitors based on artificial intelligence-driven molecular factory technology[J]. J China Pharm Univ, 2024, 55(3): 326–334.

Optimization of Menin inhibitors based on artificial intelligence-driven molecular factory technology

ZENG Hao¹, WU Guozhen¹, ZOU Wuxin¹, WANG Zhe^{1,2}, SONG Jianfei¹, SHI Hui¹, WANG Xiaojian³, HOU Tingjun^{1,2*}, DENG Yafeng^{1,4**}

¹ Hangzhou CarbonSilicon AI Technology Co., Ltd, Hangzhou 310018; ² College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058; ³ Beijing Key Laboratory of Active Substances Discovery and Druggability Evaluation, State Key Laboratory of Bioactive Substances and Functions of Natural Medicines, Peking Union Medical College & Chinese Academy of Medical Sciences, Institute of Materia Medica, Beijing 100050; ⁴ Department of Automation, Tsinghua University, Beijing 100084, China

Abstract The new generation of artificial intelligence technology, represented by deep learning, has emerged as a crucial driving force in the advancement of new drug research and development. This article creatively proposes a workflow named “Molecular Factory” for the design and optimization of drug molecules based on artificial intelligence technology. This workflow integrates intelligent molecular generation models, high-performance molecular docking algorithms, and accurate protein-ligand binding affinity prediction methods. It has been integrated as a core module into DrugFlow, a one-stop drug design software platform, providing a comprehensive set of mature

solutions for the discovery and optimization of lead compounds. Utilizing the "Molecular Factory" module, we conducted the research of second-generation inhibitors against Menin that can combat drug resistance. Through the integration of computational and experimental approaches, we rapidly identified multiple promising compounds. Among them, compound RG-10 exhibited the IC_{50} values of 9.681 nmol/L, 233.2 nmol/L, and 40.09 nmol/L against the wild-type Menin, M327I mutant, and T349M mutant, respectively. Compared to the positive reference molecule SNDX-5613, which has entered Phase II clinical trials, RG-10 demonstrated significantly enhanced inhibitory activity against the M327I and T349M mutants. These findings fully demonstrate the unique advantages of the "Molecular Factory" technology in practical drug design and development scenarios. It can rapidly and efficiently generate high-quality active molecules targeting specific protein structures, holding significant value and profound implications for advancing new drug discovery.

Key words molecular factory; artificial intelligence; molecule generation; molecular docking; Menin inhibitor

Menin 是一种关键的染色质衔接蛋白, 对于染色质上多种蛋白复合体的形成与稳定起着至关重要的作用。这些蛋白复合体包括 MLL1 和 MLL2 甲基转移酶复合体以及 JUND 转录因子复合体等^[1]。Menin 是针对由 MLL1 和 NPM1 突变驱动的急性白血病开发药物的关键靶点, 破坏 Menin 与 MLL 的相互作用能够抑制 MLL 基因重排以及 NPM1 突变引起的白血病细胞生长。在急性淋巴细胞白血病中, 约 80% 的病例出现了 KMT2A(MLL) 突变; 而在急性髓系、淋巴样或混合表型白血病中,

约 30% 病例中 KMT2A 突变与 NPM1 突变共存。最新研究发现, 针对 Menin 的治疗对 FLT3 突变型白血病也具有显著效果。在 NPM1 突变型和 MLL 重排型白血病中, 通过抑制 Menin-MLL 相互作用并靶向 MEIS1 转录因子, 可以有效抑制 FLT3 突变。因此, Menin-MLL 和 FLT3 的联合抑制为伴有 FLT3 突变的白血病患者提供了一个协同治疗的新机会^[2]。

Menin 小分子抑制剂的研发已成为近年来的热门领域, 已进入临床的化合物结构如图 1 所示。

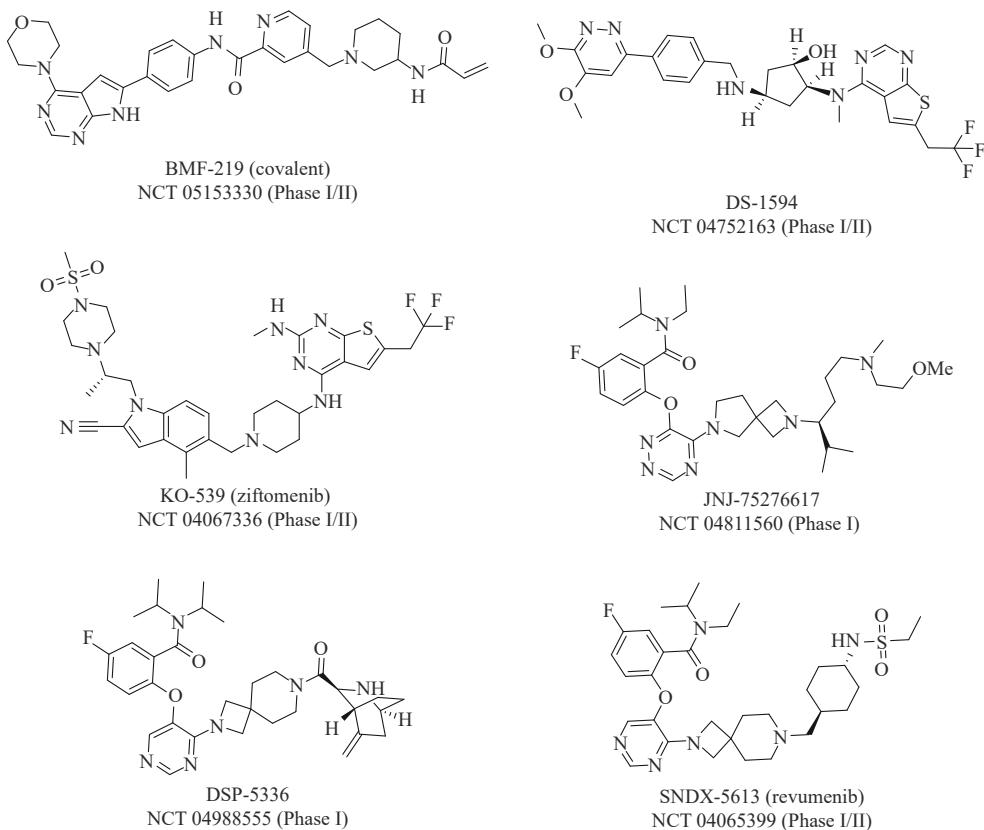


Figure 1 Chemical structure of representative Menin inhibitors

近期,由 Syndax Pharmaceuticals 公司研发的 Menin 抑制剂 revumenib(又名 SNDX-5613)在临床试验中取得积极结果,其针对携带 MLL1 或 NPM1 突变的复发性或难治性急性白血病患者。美国食品药品监督管理局(FDA)已将 revumenib 授予突破性疗法认定。在 I / II 期临床试验中,针对 MLL1 或 NPM1 突变的急性白血病患者的总体反应率达到了 53%,其中 30% 的患者实现了完全或部分的血液学缓解与恢复。然而,在 I 期临床试验中观察到,虽然患者最初对 Menin 抑制剂有反应,但随后可能会出现耐药性。研究发现,约 38.7% 的患者携带至少 1 个 MEN1 基因点突变^[3-4]。由此可见,针对 revumenib 产生耐药的 Menin 突变开发第 2 代抗耐药 Menin 抑制剂极为迫切。

新药研发是一项非常复杂系统工程,投入高、周期长、风险大。近年来,随着计算机运算能力的显著提升、智能算法的快速发展以及实验数据的不断累积,以深度学习(deep learning, DL)为代表的人

工智能(Artificial Intelligence, AI)技术被逐步应用到新药研发的多个方向,如全新药物设计、分子对接、虚拟筛选、成药性预测等,成为推动创新药物研究的新动力^[5-6]。

本文提出了一种名为“分子工厂”的药物发现工作流程,旨在通过 AI 协助新药研发人员开展高效的苗头化合物设计和优化。“分子工厂”是一个全流程的分子筛选和设计解决方案,其融合了自主研发的智能分子生成模型、高效分子过滤策略、高性能分子对接算法以及高精度亲和力预测等方法,已作为核心模块被整合进一站式药物设计软件平台 DrugFlow。如图 2 所示,用户在分析特定蛋白质靶点确定分子生成方案以及分子过滤条件后,便可利用“分子工厂”模块自动完成包括基于分子生成模型的新结构生成、基于理化性质(1D)和三维(3D)特征的分子过滤、靶标-小分子结合构象和亲和力预测在内的整个流程,最终设计出潜在的活性分子,从而提升了新药研发的效率和成功率。

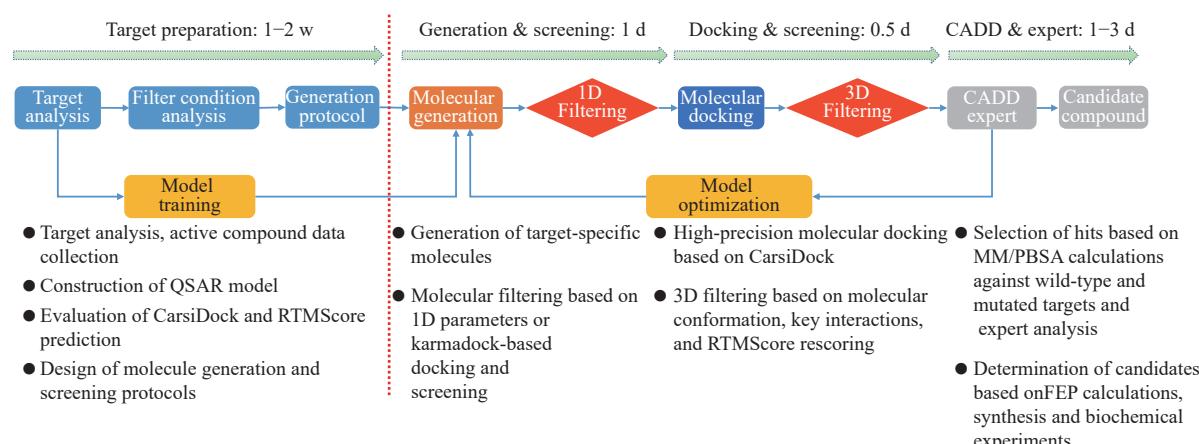


Figure 2 Process and time consumption of molecular factory assisted drug design

1 材 料

1.1 蛋白质结构

人源 Menin 蛋白和 SNDX-5613 形成的复合物共晶结构下载自蛋白质数据库,编号为 7UJ4。该结构通过 X 射线衍射获得,分辨率为 1.96 Å^[3]。复合物中的 SNDX-5613 分子即为 Syndax Pharmaceuticals 公司开发的 revumenib,该分子目前已经进入临床 II 期试验。

1.2 Menin 抑制剂活性数据

Menin 抑制剂的结构和活性数据主要来源于

ChEMBL 数据库和 BindingDB 数据库,此外本研究还收集了文献或专利报道的其他相关化合物,最终一共得到 2493 个 Menin 抑制剂。

2 方 法

本文涉及的计算部分的工作均基于“分子工厂”模块完成,“分子工厂”主要包括分子生成、分子过滤、分子对接、分子活性预测等四大功能,下面对各功能进行具体的介绍。

2.1 分子生成

为了给用户提供完善和灵活的分子生成功能,

本研究整合了多种基于配体和基于结构的分子生成模型, 并提供了友好的图形界面。分子生成模型支持多种分子设计模式, 包括 R-group 生成、Linker 生成、骨架跃迁以及从头设计等, 可满足不同的设计需求。此外, “分子工厂”允许用户上传已知活性分子的数据集, 系统可自动对生成模型进行微调, 以提高针对特定靶点的分子生成的特异性和精确度。目前“分子工厂”支持的分子生成模型包含 ResGen、Delete、smilesGPT、CarsiLinker、CarsiLinker3D 等。其中 ResGen 是一种基于蛋白口袋的三维分子生成模型, 主要用于从头设计具有高度定制化目标的药物分子^[7]。Delete 是一种基于蛋白口袋的深度结构优化模型, 以蛋白-配体相互作用为条件, 通过统一的删除(遮蔽)策略完成药物分子结构优化中的片段扩展、连接和替换等任务^[8]。SmilesGPT 是一种基于 SMILES 的分子生成模型, 使用 PubChem 中大规模的 SMILES 数据进行预训练, 该模型结合了 ADMET 性质预测模型, 可以通过微调生成具有特定性质的药物分子。CarsiLinker 是一种专注于生成连接两个分子片段的 Linker 的生成模型, 该模型结合了条件生成模型和 E(3) 等变图变分自编码器, 能够有效捕捉分子的三维结构信息, 生成具有特定构象的分子。CarsiLinker3D 在 CarsiLinker 的基础上, 增加了模型对蛋白口袋信息的使用, 并且同时支持 Linker、R-group 和骨架跃迁分子生成任务。

2.2 分子过滤

为了高效地过滤由生成模型产生的大量分子, 本研究建立了一套高效的自动化过滤流程, 包括 1D 和 3D 过滤, 可快速过滤掉成药性较差的分子。在 1D 筛选阶段, 本研究主要关注分子的理化性质, 这些性质与成药性密切相关, 包括分子量、拓扑性表面积(topological polar surface area, TPSA)、脂水分配系数(LogP)和水溶性指标(LogS)等, 此外, 本研究还排除了含有非特异性反应性子结构的化合物(PAINS)和具有潜在毒性或不利药理活性的警示结构(alert structures)的分子^[9-10]。基于这种多指标的 1D 过滤方法, 能够从生成的分子中快速剔除不符合要求的化合物, 从而为后续的 3D 过滤提供更精练的候选分子。3D 过滤方法主要是基于分子对接技术来评估化合物结合构象的合理性。具体而言, 本研究计算了对接后的化合物选定保留部分

构象的均方根偏差(RMSD), 这一指标可表征化合物的特定子结构的空间构象在生成前后的一致性。此外, 还可以分析蛋白质与配体之间的相互作用^[11], 例如考察配体能否与蛋白质的关键氨基酸残基形成特定的关键相互作用。为了评估生成分子与参考分子之间的形状相似性, 可采用最大化三维重叠算法来评估它们的形状相似性^[12-14]。鉴于 3D 形状相似性的分子更可能与靶标形成相似的几何匹配, 这一评估步骤对于判别生成的分子能否与靶标形成有效的相互作用至关重要。

2.3 分子对接

分子对接能够预测小分子配体与蛋白靶标之间的结合模式及结合强度^[15], 已成为基于靶标结构的药物设计中不可或缺的工具, 被广泛用于大型化合物库的虚拟筛选。“分子工厂”中的分子对接方法包含了基于物理原理的传统分子对接方法 AutoDock Vina 以及基于 AI 的分子对接方法 KarmaDock 和 CarsiDock。基于 AI 算法的对接方法利用深度学习算法来实现构象搜索和打分, 在结合构象和亲和力评估的可靠性有一定优势, 此外还能大幅提高计算速度。

AutoDock Vina 是一款经典的开源分子对接程序, 被广泛用于药物设计和筛选^[16-17]。KarmaDock 是一种基于 AI 的端到端 AI 对接及打分方法, 在对接速度和精度上超越了大多数传统对接/打分方法, 在 3090 英伟达 GPU 卡上对接一个分子的平均耗时仅为 0.017 s, 可实现化合物库的超高通量虚拟筛选; 该模型在 Casf2016 数据集上的富集因子(EF1%)为 23.4, 在同类方法中处于领先水平^[18]。CarsiDock 是一种基于深度学习的创新分子对接方法, 该方法基于 Transformer 并参考了 AlphaFold2 的模型架构, 采用了千万条传统分子对接方法预测得到的蛋白-配体复合物结构数据进行预训练, 然后再用 2 万条复合物晶体结构数据进行微调, 使得其在 PDBbind Core set 数据集上的对接成功率(RMSD<2Å)达到了 92.28%; 此外, 该方法通过构象合理化模块能够保证配体分子的原子坐标预测精度达到当前最佳^[19]。

2.4 分子活性预测

分子活性预测模块中还集成了基于连续介质模型的结合自由能预测方法 MM/PBSA 方法以及基于 AI 的蛋白-配体打分函数 RTMScore, 利用这两种方法可对筛选后的分子提供更高精度的活性

评估。MM/PBSA 方法是一种基于连续介质模型的终点式结合自由能预测方法, 通过分子力学和求解泊松-玻尔兹曼方程来预测配体-蛋白间的结合自由能。RTMScore 是一种基于 AI 的蛋白质-配体结合强度预测的新型打分函数, 通过 Graph Transformer 提取蛋白质的氨基酸残基以及配体的原子节点特征, 并通过混合密度网络 (mixture density network, MDN) 获取蛋白质各氨基酸残基和配体各原子间距离的概率密度分布, 最后将其转化为统计势以用于蛋白-配体间结合强度的评估^[20]。RTMScore 的筛选能力在 CASF-2016 标准数据集上显著超越了当前的其他主流方法, 其在有无天然构象的存在下分别可取得 97.3% 和 93.4% 的平均 top1 对接成功率 (DeepDock 和 PIGNet 在无天然构象时仅为 87.0%); 而在筛选能力评估中可取得 66.7% 的平均 top1 成功率和 28.00 的 1% 富集因子 (DeepDock 和 PIGNet 的两项指标分别为 55.4%、19.60 以及 43.9%、16.41)。同时本研究进一步在 DEKOIS2.0 数据集上评估了 RTMScore 筛选能力, 结果表明 RTMScore 的富集能力明显优于以相似策略构建所得的 DeepDock 和经典方法 Glide SP。

2.5 体外活性评价

本研究首先对阳性参照药物 (SNDX-5613) 进行了稀释, 以制备一系列浓度梯度的溶液, 其中稀释液作为起始点, 按照 3 倍稀释度进行连续稀释, 共设置 10 个稀释点及一个未添加药物的对照点。随后, 将 50/100 nL 的化合物分别转移至 384 孔板中, 每个孔设置两个重复样本。转移完毕后, 对 384 孔板进行 1000 r/min 的离心处理。接下来, 分别向每个孔中加入 Menin-WT、Menin-M327I 和 Menin-T349M 蛋白 5 μ L, 并在 25 $^{\circ}$ C 下孵育 10 min。之后向每个孔中加入 FITC-MLL4-43 荧光标记底物 5 μ L, 再次进行 1000 r/min 的离心, 并在 25 $^{\circ}$ C 下继续孵育 60 min。最终, 利用 BMG 荧光酶标仪读取荧光信号强度。数据分析部分, 采用非线性回归模型来拟合化合物的 IC_{50} 。

3 结果与讨论

3.1 数据集的收集与准备

本研究首先分析了 Menin 晶体结构 (PDB ID:7JU4) 中蛋白质与配体 SNDX-5613 间的相互作用。如图 3 所示, 在晶体结构中配体与 TYR-276、MET-322 及 TRP-341 形成氢键, 将其作为后续分

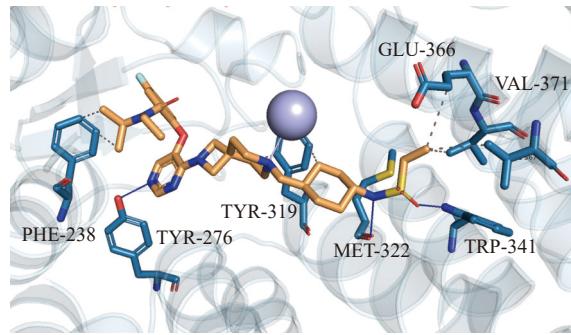


Figure 3 Binding mode of SNDX-5613 in Menin (PDB ID: 7JU4)

子 3D 过滤的参考条件。

从文献和数据库中收集 Menin 抑制剂的结构和实验活性数据, 并剔除重复数据和数据集中未定量标记活性的化合物。为了减轻实验数据差异导致的影响, 本研究根据 IC_{50} 定义了 3 种不同的活性类别, 分别为“active”(高活性)、“inactive”(低活性)和“inconclusive”(活性不确定)。其中 IC_{50} 小于 100 nmol/L 的数据标记为“active”, IC_{50} 大于 1000 nmol/L 的数据标记为“inactive”, IC_{50} 介于 100~1000 nmol/L 之间的数据标记为“inconclusive”。最终得到了 1600 条数据, 其中“active”数据为 1002 条, 其分子量、TPSA、LogP 及 LogS 的分布见图 4。由图可见, 对于 IC_{50} 小于 100 nmol/L 的分子, 其分子量主要分布在 500~800 之间, TPSA 主要分布在 70~140 之间, LogP 主要分布在 3~6 之间, LogS 主要分布在 -8~5 之间。本研究将活性分子的性质分布作为后续分子 1D 筛选的参考条件。

3.2 分子生成结果

目前 Menin 主要发生两种耐药突变: M327I 和 T349M, 这两个氨基酸位于与 SNDX-5613 磺酰胺部分附近的空间位置。这一发现提示, 针对这部分的优化可能会显著提高与 Menin 的结合选择性, 进而克服这些耐药突变。因此, 在分子生成环节中, 本研究采用了 R-group 生成策略, 把结构改造的重点集中在 SNDX-5613 中的环己烷与磺酰胺部分。

本研究利用“分子工厂”模块进行了一系列基于 Menin 蛋白结构的分子生成计算。首先利用收集到的 Menin 活性分子数据集对分子生成模型进行强化学习, 通过这一步骤, 模型能够更好地理解 Menin 抑制剂的结构特征。随后, 定义活性分子 SNDX-5613 中需要保留的部分, 通过 R-group 生成策略生成了共计 53 万个衍生物。生成分子的分子量、TPSA、LogP、LogS 的分布如图 5 所示。从分

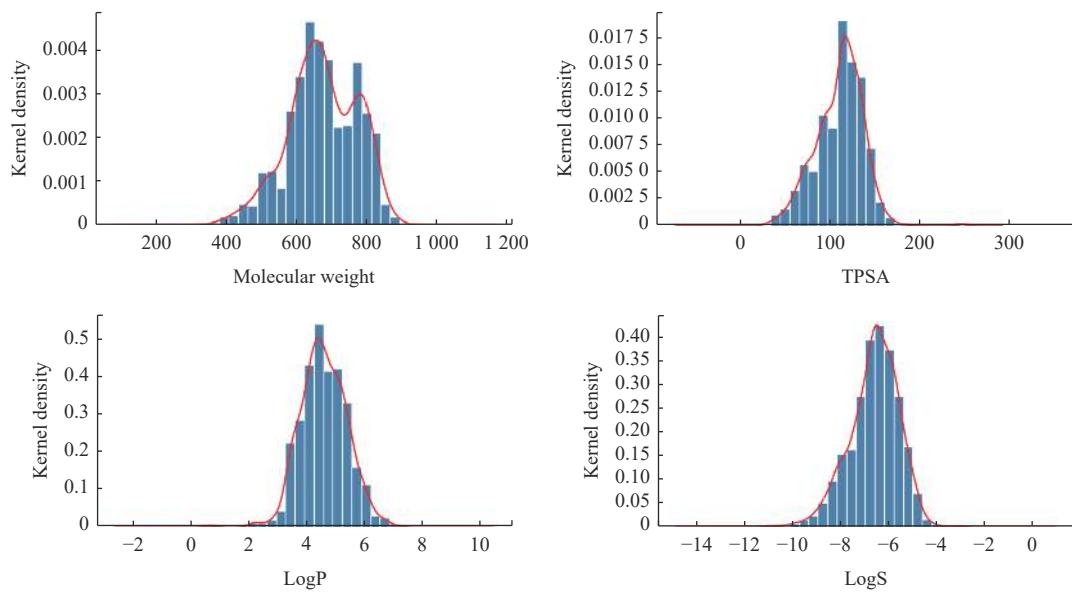


Figure 4 Kernel density distributions of molecular weight, TPSA, LogP and LogS in known actives

TPSA: Topological polar surface area; LogP: Logarithm of the *n*-octanol/water distribution coefficient; LogS: Logarithm of aqueous solubility value

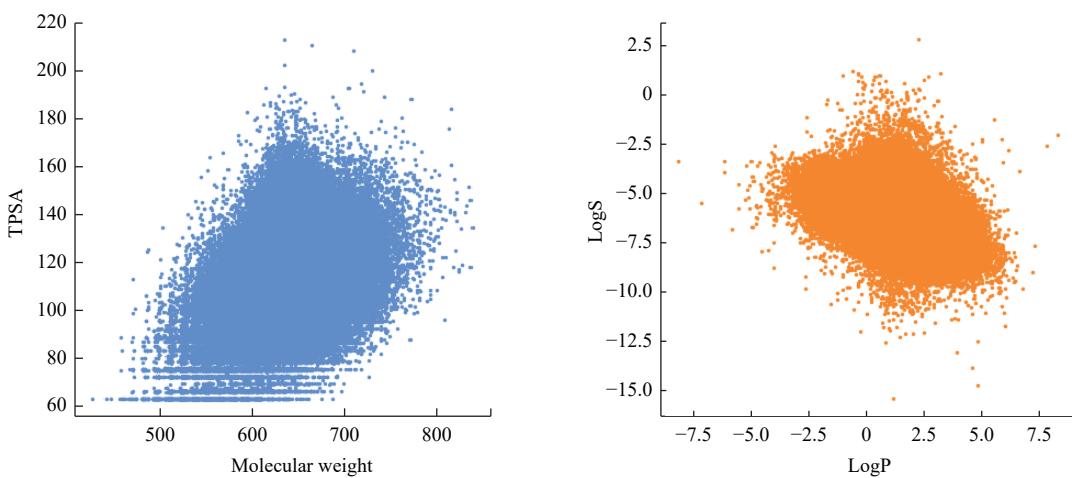


Figure 5 Distributions of molecular weight, TPSA, LogP, LogS of the generated molecules

布结果来看,绝大多数生成的分子均符合 1D 筛选的标准范围,这表明本研究的分子库具有较好的成药性特征。这些生成的分子随后采用“分子工厂”中分子过滤功能进行系统评估,以筛选出最具潜力的候选分子。

3.3 基于 1D 和 3D 的分子过滤

对于生成的分子,本研究采用了 1D 和 3D 两种策略进行了过滤。1D 过滤的参数包括分子量、TPSA、LogP、LogS、PAINS 规则以及 Alert Structures 等。共有 126223 个候选分子通过了 1D 参数过滤,这些分子的化学空间分布如图 6 所示。通过对比分析,发现 AI 模型生成的分子在化学空间中的覆盖范

围更为广泛,表明这些分子具有较高的化学多样性。此外,本研究还观察到 Menin 活性化合物的化学空间被 AI 模型生成的分子所包含,这表明 AI 模型生成的分子与已知的 Menin 活性化合物在化学性质上具有相似性。这一发现不仅证实了 AI 模型生成分子的化学多样性,也暗示了这些分子中可能存在具有潜在的活性化合物。

本研究进一步使用“分子工厂”中的 CarsiDock 方法预测了所有候选分子与 Menin 间的相互作用模型。在基于蛋白-配体相互作用的 3D 分子过滤中,重点关注蛋白质中的几个关键氨基酸残基 (TYR-276、MET-322 以及 TRP-341) 能否与化合物

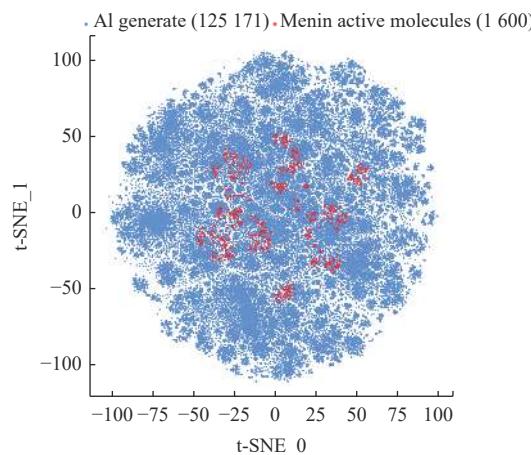


Figure 6 Comparison of the chemical spatial distribution using the t-distributed stochastic neighbor embedding (t-SNE) method

形成较为稳定的氢键。此外,本研究还评估了对接后的分子构象与参考分子中保留原子的均方根偏差(RMSD)。最终,确定了114个具有高潜力的候选分子,这些分子不仅保留了与目标蛋白的关键相互作用,同时与参考分子具有较高的3D形状相似性。

3.4 候选分子的生物活性评价

在完成分子对接分析后,采用了RTMScore方法预测了分子的亲和力;此外,还分析了上一步3D过滤中得到的蛋白-配体相互作用。通过这一综合评估过程,本研究最终筛选出了11个候选分子,这些分子在结构上与参考分子存在较为显著的差异,显示出较高的新颖性和多样性。随后,对这11个分子进行了基于MM/GBSA的结合自由能预测,以准确评估它们与目标受体间的结合亲和力。计算不仅针对野生型,也包括突变型分子,以便更全面地评估其潜在活性。在对计算结果进行深入分析,并结合专家的专业意见及合成可行性分析后,最终确定了5个分子(RG-06、RG-07、RG-10、D5-18-9、D5-18-11)作为候选化合物,进行后续的生物活性实验评价。本研究评估了这5个候选化合物对Menin蛋白野生型(Menin-WT)及其两种突变型(Menin-M327I和Menin-T349M)的抑制活性,并将这些结果与参考分子SNDX-5613进行比较。 IC_{50} 被用来量化表征化合物的活性,单位为纳摩尔(nmol/L),其中较低的 IC_{50} 表示更强的抑制效果,实验结果如表1所示。对于Menin-WT,相较于阳性参考分子SNDX-5613的 IC_{50} (11.86 nmol/L),化合物RG-06和RG-10的 IC_{50} 可分别达到32.91和

Table 1 Biological activities of five candidates against Menin-WT, Menin-M327I and Menin-T349M

Target	Compound ID	IC_{50} /(nmol/L)
Menin-WT	SNDX-5613	11.86
	RG-06	32.91
	RG-10	9.681
	D5-18-11	382.7
	D5-18-9	221.7
	RG-07	364.3
Menin-M327I	SNDX-5613	1421
	RG-06	513.8
	RG-10	233.2
	D5-18-11	625.3
	D5-18-9	363.2
	RG-07	590.7
Menin-T349M	SNDX-5613	280.2
	RG-06	53.8
	RG-10	40.09
	D5-18-11	1611
	D5-18-9	326
	RG-07	384.9

9.681 nmol/L,表明它们对野生型具有和SNDX-5613相同数量级的抑制活性。特别是RG-10,其活性甚至优于参考分子。对于Menin-M327I突变体而言,参考分子SNDX-5613的 IC_{50} 增加至1421 nmol/L,表明对该突变体的抑制活性显著下降。相比之下,RG-06和RG-10的 IC_{50} 分别为513.8和233.2 nmol/L,虽然抑制活性有所降低,但仍具有一定的抑制效果。而化合物D5-18-11和D5-18-9的 IC_{50} 较高,表明这两个化合物对M327I突变体的抑制活性较弱。对于Menin-T349M突变体而言,参考分子SNDX-5613的 IC_{50} 为280.2 nmol/L,相比野生型活性也有所降低,但幅度没有M327I大。值得注意的是,RG-06和RG-10在这一突变型中仍表现出显著的抑制活性, IC_{50} 分别为53.8和40.09 nmol/L,显著低于SNDX-5613,表明这两个化合物对Menin-T349M突变体具有明显的优势。

RG-06和RG-10对Menin-T349M突变体显著抑制活性,意味着它们在结构上能够更好地适应突变体,这为Menin抗耐药抑制剂的设计提供了有价值线索。在未来的工作将集中在进一步优化这些化合物的结构,以增强它们对不同Menin突变体的抑制活性,并在细胞和动物模型中评估它们的药效和安全性。

4 结论与展望

借助 DrugFlow 中“分子工厂”模块的智能化工作流程,综合运用分子生成、性质过滤、对接以及亲和力预测等功能,本研究从数十万个分子中最终筛选出 5 个 Menin 的潜在抑制剂,并进一步测试了它们对 Menin 蛋白野生型和突变体的抑制活性。结果表明,化合物 RG-06 和 RG-10 对 Menin 蛋白具有良好的抑制活性,尤其是 RG-10 对 Menin-WT 表现出和阳性参考分子 SNDX-5613 相近的活性,且对 Menin-M327I 和 Menin-T349M 两个突变体在抑制活性上优于参考分子近 7 倍。这样不但为 Menin 第 2 代抑制剂对开发提供了有价值的参考,同时验证了“分子工厂”在实际药物筛选和设计工作中的效果。

展望未来,本研究将进一步优化“分子工厂”流程,特别是在分子生成和亲和力预测方面,将继续探索和开发更先进的算法。此外,考虑在“分子工厂”中整合数据准备方面的功能,通过利用大语言模型的先进文本分析能力,高效地从科学文献和数据库中自动提取关键的靶点信息和化合物信息,从而帮助用户提高信息收集和分析的效率。相信随着人工智能技术的不断进步和应用,“分子工厂”将成为药物研发领域的重要工具。

References

- [1] Yokoyama A, Cleary ML. Menin critically links MLL proteins with LEDGF on cancer-associated target genes[J]. *Cancer Cell*, 2008, **14**(1): 36-46.
- [2] Issa GC, Aldoss I, DiPersio J, et al. The menin inhibitor revumenib in KMT2A-rearranged or NPM1-mutant leukaemia[J]. *Nature*, 2023, **615**(7954): 920-924.
- [3] Perner F, Stein EM, Wenge DV, et al. MEN1 mutations mediate clinical resistance to menin inhibition[J]. *Nature*, 2023, **615**(7954): 913-919.
- [4] Grisham J. Menin Inhibitors: Promising New Drug Treatment for Aggressive Leukemia | Memorial Sloan Kettering Cancer Center[EB/OL]. (2023-03-15)[2024-03-22]. <https://www.mskcc.org/news/promising-new-type-drug-aggressive-leukemia>.
- [5] Cavasotto CN, di Filippo JI. Artificial intelligence in the early stages of drug discovery[J]. *Arch Biochem Biophys*, 2021, **698**: 108730.
- [6] Gupta R, Srivastava D, Sahu M, et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery[J]. *Mol Divers*, 2021, **25**(3): 1315-1360.
- [7] Zhang O, Zhang JT, Jin JY, et al. ResGen is a pocket-aware 3D molecular generation model based on parallel multiscale modelling[J]. *Nat Mach Intell*, 2023, **5**: 1020-1030.
- [8] Zhang HT, Zhao HF, Zhang XJ, et al. Delete: deep lead optimization enveloped in protein pocket through unified deleting strategies and a structure-aware network[EB/OL]. *arXiv*, 2023: 2308.02172. <http://arxiv.org/abs/2308.02172>.
- [9] Xiong GL, Wu ZX, Yi JC, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties[J]. *Nucleic Acids Res*, 2021, **49**(W1): W5-W14.
- [10] Jiang DJ, Lei TL, Wang Z, et al. ADMET evaluation in drug discovery. 20. Prediction of breast cancer resistance protein inhibition through machine learning[J]. *J Cheminform*, 2020, **12**(1): 16.
- [11] Adasme MF, Linnemann KL, Bolz SN, et al. PLIP 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA[J]. *Nucleic Acids Res*, 2021, **49**(W1): W530-W534.
- [12] Hain E, Camacho CJ, Koes DR. Fragment oriented molecular shapes[J]. *J Mol Graph Model*, 2016, **66**: 143-154.
- [13] Nicholls A, McGaughey GB, Sheridan RP, et al. Molecular shape and medicinal chemistry: a perspective[J]. *J Med Chem*, 2010, **53**(10): 3862-3886.
- [14] Koes DR, Camacho CJ. Shape-based virtual screening with volumetric aligned molecular shapes[J]. *J Comput Chem*, 2014, **35**(25): 1824-1834.
- [15] Muhammed MT, Aki-Yalcin E. Molecular docking: principles, advances, and its applications in drug discovery[J]. *Lett Drug Design Discov*, 2024, **21**(3): 480-495.
- [16] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading[J]. *J Comput Chem*, 2010, **31**(2): 455-461.
- [17] Eberhardt J, Santos-Martins D, Tillack AF, et al. AutoDock vina 1.2. 0: new docking methods, expanded force field, and python bindings[J]. *J Chem Inf Model*, 2021, **61**(8): 3891-3898.
- [18] Zhang XJ, Zhang O, Shen C, et al. Efficient and accurate large library ligand docking with KarmaDock[J]. *Nat Comput Sci*, 2023, **3**(9): 789-804.
- [19] Cai H, Shen C, Jian TY, et al. CarsiDock: a deep learning paradigm for accurate protein-ligand docking and screening based on large-scale pre-training[J]. *Chem Sci*, 2024, **15**(4): 1449-1471.
- [20] Shen C, Zhang XJ, Deng YF, et al. Boosting protein-ligand binding pose prediction and virtual screening based on residue-atom distance likelihood potential and graph transformer[J]. *J Med Chem*, 2022, **65**(15): 10691-10706.



[专家介绍] 侯廷军, 浙江大学药学院, 求是特聘教授, 博导。长期围绕计算机辅助药物设计中的核心问题展开前沿交叉科学研究。在 *Nature Machine Intelligence*、*Nature Computational Science*、*Nature Protocols*、*Nature Communications*、*Chemical Reviews*、*Science Advances*、*PNAS*、*ACS Central Science*、*Nucleic Acids Research*、*Journal of Medicinal Chemistry* 等期刊发表 SCI 论文 500 余篇, 引用 29000 余次(Google), H 因子 85; 获授权专利和软件著作权 80 余项。任中国化学会会计(机)化学专业委员会副主任委员、中国计算机学会数字医学分会副主任委员, *Briefings in Bioinformatics*、*Journal of Cheminformatics* 等 14 种 SCI 期刊编委或顾问编委。入选科睿唯安全球高被引科学家、爱思唯尔中国高被引学者、获中国化学会会计(机)化学专委会青年计算化学家奖、药明康德生命化学研究奖、高等学校科学研究优秀成果二等奖 2 次、第四届寻找青年科学之星等学术奖项和荣誉。



[专家介绍] 邓亚峰, 毕业于清华大学, 二十年人工智能算法及产品研发经验, 现任碳硅智慧创始人兼 CEO。曾任 360 集团副总裁、人工智能研究院院长兼搜索事业部总经理, 科创板第一家人工智能上市公司格灵深瞳 CTO, 中国图形图像学会常务理事, 北京人工智能产业联盟副理事长等职务, 获得 2021 年中国人工智能年度十大风云人物称号。带领团队在 AIDD、计算机视觉、多模态大模型、机器人、智能搜索等领域做出过创新成果或先进产品, 累计申请发明专利 140 余项(已授权 98 项), 发表论文 50 篇左右。