

• 数据挖掘与循证医学 •

基于机器学习与分子动力学模拟挖掘天南星治疗重度抑郁症的潜在抗炎靶点和关键活性成分

郑远腾^{1,3}, 夏漓傑^{1,3}, 张秀军^{1,2*}, 靳 梦^{2,3*}

1. 华北理工大学心理与精神卫生学院, 河北 唐山 063200

2. 华北理工大学心理与精神卫生学院、公共卫生学院, 河北, 唐山 063210

3. 齐鲁工业大学(山东省科学院)生物研究所, 山东省科学院药物筛选技术重点实验室, 山东 济南 250103

摘要:目的 利用生物信息学的方法结合多种机器学习模型, 寻找天南星治疗重度抑郁的潜在抗炎靶点, 并从免疫炎症角度探讨其发病机制。采用分子对接、药物相似性评估和分子动力学模拟, 筛选天南星治疗重度抑郁的关键活性成分, 以期对相关治疗药物与治疗策略的开发提供参考。方法 利用 CIBERSORT 与基因集变异分析 (gene set variation analysis, GSVA) 验证免疫炎症假说, 并寻找免疫标志物。基于 TCMSP、GEO 和 Swiss Target Prediction 数据库, 获得天南星潜在作用靶点与重度抑郁相关靶点, 并利用费舍尔精确检验探究相关靶点的关联性。采用加权基因共表达网络 (weighted gene co-expression network analysis, WGCNA) 分析获得与免疫炎症相关的基因共表达模块。通过将炎症相关基因、天南星潜在作用靶点、重度抑郁相关靶点、与免疫炎症相关的绿松石模块 (Turquoise) 取交集获得天南星治疗重度抑郁的候选抗炎靶点。采用基因本体 (gene ontology, GO) 和京都基因与基因组百科全书 (Kyoto encyclopedia of genes and genomes, KEGG) 富集分析探究候选抗炎靶点涉及的生物学过程和相关通路。构建 8 个机器学习模型, 以筛选潜在抗炎靶点。利用受试者工作特征 (receiver operator characteristic, ROC) 曲线与相关分析筛选出潜在抗炎靶点并进行外部验证。使用分子对接与药物相似性评估筛选天南星的关键活性成分, 并采用分子动力学模拟验证分子对接与药物相似性评估的结果。结果 正常组与重度抑郁患者的炎症得分存在显著性差异 ($P < 0.05$), CD8 T 细胞为重度抑郁患者的免疫标志物。天南星潜在作用靶点和重度抑郁相关靶点具有关联性, 与免疫炎症最相关的基因共表达模块为绿松石模块。候选抗炎靶点具有 8 个, 涉及的生物学过程为炎症反应调节, 相关通路为 Fc-RI 信号通路与 T 细胞受体信号通路。随机森林为最优模型, 且纳入的靶点都具有较高的辨别力。ROC 曲线与相关分析指出, 花生四烯酸 5-脂氧合酶 (arachidonate 5-lipoxygenase, ALOX5) 为治疗重度抑郁的潜在抗炎靶点, 且外部验证具有组间差异 ($P < 0.05$)。分子对接结果显示, 有 6 种天南星主要活性成分对接成功。药物相似性评估显示, 8,11,14-二十二碳三烯酸甲酯、24-表柏甾醇具有较好的成药性。分子动力学模拟验证了分子对接与药物相似性评估的结果。结论 ALOX5 和 8,11,14-二十二碳三烯酸甲酯为天南星治疗重度抑郁的潜在抗炎靶点及关键活性成分。

关键词: 天南星; 重度抑郁; 机器学习; 分子对接; 分子动力学模拟; 8,11,14-二十二碳三烯酸甲酯; 花生四烯酸 5-脂氧合酶
中图分类号: Q811.4; R285; TP18 **文献标志码:** A **文章编号:** 0253-2670(2024)15-5174-15

DOI: 10.7501/j.issn.0253-2670.2024.15.017

Exploration of potential anti-inflammatory targets and key ingredients of *Arisaematis Rhizoma* in treatment of major depressive disorder based on machine learning and molecular dynamics simulation

ZHENG Yuanpeng^{1,3}, XIA Lijie^{1,3}, ZHANG Xiujuan^{1,2}, JIN Meng^{2,3}

1. School of Psychology and Mental Health, North China University of Science and Technology, Tangshan 063200, China

2. School of Public Health, School of Psychology and Mental Health, North China University of Science and Technology, Tangshan 063210, China

收稿日期: 2024-02-23

基金项目: 山东省重点研发计划 (2022CXGC020515); 泰山学者青年专家项目 (tsqn202312233); 山东省自然科学基金项目 (ZR2021ZD29)

作者简介: 郑远腾, 硕士研究生, 研究方向为认知与行为的神经机制。E-mail: zhengyuanteng22@163.com

*通信作者: 张秀军, 教授, 博士生导师, 研究方向为药理学、毒理学研究。E-mail: zhxj@ncst.edu.cn

靳 梦, 女, 硕士生导师, 研究员, 研究方向为神经药理学研究。E-mail: mjin1985@hotmail.com

3. Key Laboratory of Drug Screening Technology, Shandong Academy of Sciences, Biology Institute, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250103, China

Abstract: Objective To identify potential anti-inflammatory targets in treatment of major depressive disorder and explore its pathogenesis from the perspective of immune inflammation, bioinformatics methods combined with multiple machine learning models were utilized. Molecular docking, drug similarity assessment, and molecular dynamics simulation were employed to screen the key active ingredients of Tiannanxing (*Arisaematis Rhizoma*), in order to provide reference for the development of related therapeutic drugs and strategies. **Methods** CIBERSORT and gene set variation analysis (GSVA) were used to validate the immune inflammation hypothesis and identify immune biomarkers. Based on the TCMSP, GEO, and Swiss Target Prediction databases, potential targets of *Arisaematis Rhizoma* and targets related to major depressive disorder were obtained. Furthermore, the association between potential targets of *Arisaematis Rhizoma* and targets related to major depressive disorder was explored using Fisher's exact test. Weighted gene co-expression network analysis (WGCNA) was employed to obtain gene co-expression modules most associated with immune inflammation. Anti-inflammatory candidate targets were obtained by intersecting inflammation-related genes, *Arisaematis Rhizoma* potential targets, major depressive disorder-related targets, and Turquoise module most associated with immune inflammation. Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis were conducted to explore the biological processes and pathways involved in the anti-inflammatory candidate targets. Eight machine learning models were constructed to screen potential anti-inflammatory targets. Potential anti-inflammatory targets were selected using receiver operator characteristic (ROC) curves and correlation analysis. Additionally, external validation was conducted for the potential anti-inflammatory targets. Finally, molecular docking and drug similarity assessment were used to screen key active ingredients of *Arisaematis Rhizoma*. The molecular dynamics simulation was employed to verify the results of molecular docking and similarity assessment. **Results** There was a significant difference in the inflammation scores between the normal group and major depressive disorder group ($P < 0.05$), and T cells CD8 was identified as an immune biomarker for patients with major depressive disorder. The potential targets of *Arisaematis Rhizoma* were associated with major depressive disorder-related targets. The gene co-expression module most associated with immune inflammation is Turquoise. There were eight candidate anti-inflammatory targets, involving the biological process of inflammation regulation, Fc epsilon RI signaling pathway and the pathways of T cell receptor signaling pathway. Random forest was the optimal model, and the included targets all had high discrimination ability. The ROC curve and correlation analysis indicate that arachidonate 5-lipoxygenase (ALOX5) is a potential anti-inflammatory target for treatment of major depressive disorder, and external validation demonstrates significant inter-group differences ($P < 0.05$). Molecular docking results indicated that only six active ingredients successfully docked among the main active ingredients of *Arisaematis Rhizoma*. Drug similarity assessment revealed that 8,11,14-docosatrienoic acid methyl ester and 24-epicampestrol had good pharmacological properties, and molecular dynamics simulation verified the results of molecular docking and drug similarity assessment. **Conclusion** Our results implied that ALOX5 and 8,11,14-docosatrienoic acid methyl ester are potential anti-inflammatory target and key active ingredient of *Arisaematis Rhizoma* in treatment of major depressive disorder, respectively.

Key words: *Arisaematis Rhizoma*; major depressive disorder; machine learning; molecular docking; molecular dynamics simulation; 8,11,14-docosatrienoic acid methyl ester; arachidonate 5-lipoxygenase

重度抑郁症 (major depressive disorder) 是一种严重的心理障碍, 其主要特征之一是持续的情绪低落。患者经历长时间的情感沮丧, 不仅影响日常生活, 还对自我评价产生负面影响。典型症状包括兴趣丧失、自责感、食欲减退、注意力不集中和自杀倾向。该病患病率较高, 易复发, 可导致较高的致残、致死率^[1]。根据最新的临床诊治统计数据, 全球范围内大约有 2.8 亿人患有重度抑郁症, 且每年患病人数仍在增多^[2]。值得关注的是, 在以残疾和死亡总和排名的统计中, 重度抑郁症仅居心脏病、卒中和艾滋病毒等疾病之后。而且该病会伴随患者多

年^[3]。根据世界卫生组织预测, 到 2030 年, 重度抑郁症可能成为对人类健康构成重大威胁的主要疾病之一^[4]。目前, 重度抑郁症的发病机制尚未完全明确, 科学界存在多种假说, 如单胺假说、受体假说和免疫炎症假说等^[5]。近年来, 越来越多的研究表明, 免疫调节异常与重度抑郁症的发病机制密切相关。尽管免疫炎症假说并非当前抗重度抑郁药物研究的主流方向, 但其在重度抑郁症机制领域仍备受关注^[6]。该假说指出, 过度活化的免疫系统会产生炎性细胞因子, 这些炎性因子对重度抑郁症的发生发展具有重要作用^[7]。

中药治疗情志类疾病具有靶点多、不良反应小等优势,在重度抑郁症及伴发症的治疗方面具有广阔的应用前景^[8]。天南星 *Arisaematis Rhizoma* 是天南星科的多年生药用植物,含有黄酮类、木脂素、萜类和生物碱等多种成分^[9]。研究发现,从天南星科植物的干燥根茎中提取得到的一种天然小分子多酚类化合物具有抗炎、抗氧化和抗肿瘤等多种药理活性,且毒性低、副作用少^[10]。本研究拟探讨天南星的抗炎相关通路和潜在靶点。网络药理学是一门新兴学科,基于系统生物学和网络科学理论,其目标是通过生物系统中分子、基因、蛋白质、信号通路等多层次的网络结构进行综合分析,揭示药物与生物体之间的复杂相互作用^[11],为新药研发提供更系统、全面的理论基础。随着高通量测序技术的发展,基于生物信息学的大数据挖掘分析被广泛用于探索各种疾病相关的生物标志物^[12]。生物信息学有利于探究药物的潜在治疗靶点和相关通路。在监督学习的机器学习范畴中,特征和结局事件同时输入分类模型,旨在拟合它们之间的关系。这种智能预测和决策的方法对于筛选潜在的治疗靶点发挥着重要作用。分子对接是一种计算机辅助药物研究的关键方法,用于模拟和预测分子之间的相互作用。这一技术主要通过分析受体和药物分子的结构,以理论模拟的方式预测它们之间的结合模式和亲和力^[13]。相较于分子对接,分子动力学模拟是一种更加复杂和全面的计算机模拟方法,用于研究分子在时间和空间尺度上的运动和相互作用。分子动力学模拟能够模拟分子的运动轨迹,并提供有关分子结构、动力学行为和热力学性质的详细信息。在药学领域,分子动力学常被用于药物动力学研究,以进行新药开发的筛选和预测^[14]。

本研究采用生物信息学方法结合多种机器学习模型寻找治疗重度抑郁的潜在抗炎靶点,并通过分子对接、药物相似性评估和分子动力学模拟筛选天南星的关键活性成分。旨在为重度抑郁的临床治疗提供新的研究思路,为重度抑郁的治疗方案与相关治疗药物的开发提供科学依据。

1 方法

1.1 数据来源

从高通量基因表达数据库 (gene expression omnibus, GEO) 获取重度抑郁数据集 (GSE76826、GSE98793)^[15]。重度抑郁的数据集皆是来源人的血液样本,其中 GSE76826 数据集用来获得重度抑郁

相关靶点, GSE98793 数据集用来验证潜在抗炎靶点。GSE76826 数据集是储存于 GEO 数据库的微阵列芯片数据集,患者均被《精神疾病诊断与统计手册 (第 4 版)》(Diagnostic and Statistical Manual of Mental Disorders, 4 ed, DSM-IV) 诊断为重度抑郁症,抑郁状态是使用汉密尔顿抑郁症结构化访谈指南 (SIGH-D) 评定量表测量,去除缓解组后一共 10 例患者,还包括 12 例正常人 (对照组)。验证数据集 GSE98793 排除了 64 例同时患有广泛性焦虑症和重度抑郁症的患者后,剩余的样本数为 128,其中 64 例为重度抑郁症患者,64 例为正常人 (对照组)。矩阵信息和平台信息被下载以进行 ID 转换、缺失值处理、标准化处理等。将探针 ID 与基因符号进行替换以获得基因名称,删除存在缺失值的基因所属行,并进行标准化处理。

1.2 基因集变异分析与免疫浸润

在 Gene Card (<https://www.genecards.org/>) 疾病数据库中以 “inflammation” “inflammatory” 为关键词检索获得炎症相关基因^[16]。炎症相关基因的筛选条件为相关性评分 (relevance score) > 8, 将筛选后的炎症相关基因作为基因集,采用基因集变异分析 (gene set variation analysis, GSVA) 来计算重度抑郁患者的炎症得分^[17]。在 GSE76826 数据集中,对所有基因的表达值进行核密度估计,将离散的表达值转换为连续的概率分布。然后,根据估计的表达值进行排序,并基于炎症相关基因在排序中的位置进行积分计算,从而评估这些基因在不同样本中的富集程度。最后,采用 Wilcoxon 秩和检验分析炎症得分在对照组和重度抑郁症组之间的差异。利用 CIBERSORT (<https://cibersort.stanford.edu/>) 免疫浸润分析推测重度抑郁患者的免疫细胞分布情况^[18],在对照组和重度抑郁症组中,检验不同类型的免疫细胞评分在正常组和重度抑郁组之间的差异。

1.3 天南星活性成分的筛选及潜在作用靶点的预测

从中药系统药理学数据库与分析平台^[19] (TCMSP) 检索天南星的活性成分并进行筛选,筛选条件:口服生物利用度 (oral bioavailability, OB) ≥ 30%, 且类药性 (drug-likeness, DL) ≥ 0.18。从 PubChem 数据库 (<https://pubchem.ncbi.nlm.nih.gov/>) 获得天南星活性成分的 SMILES 号^[20],利用 Swiss Target Prediction 数据库^[21] (<http://www.swisstargetprediction.ch/index.php>) 对天南星的主要活性成分进行潜在作用靶点的预测,删除重复靶点

后, 得到天南星潜在作用靶点。

1.4 重度抑郁相关靶点的挖掘与预测

R 软件的 Limma 包被用于筛选重度抑郁症组与对照组的差异基因^[22]。在本研究中, 重度抑郁差异基因的阈值被设定为 $\log_2FC$ [FC 表示差异倍数 (fold change)] 的绝对值 ≥ 0.5 , 并且校正后的 P 值 (P_{adj}) < 0.05 。R 软件的 ggplot2 包被用于制作火山图以展示差异基因的分布和表达量。

1.5 天南星潜在作用靶点与重度抑郁相关靶点的关联性

将筛选得出的天南星潜在作用靶点与重度抑郁相关的靶点取交集, 并以 2×2 列联表的形式展示, 使用费舍尔精确检验来分析它们之间的关联性。

1.6 加权基因共表达网络 (weighted gene co-expression network analysis, WGCNA) 分析

为了挖掘与免疫炎症最相关的基因共表达模块, 通过 WGCNA R 包对 GSE76826 数据集进行 WGCNA 分析^[23]。首先, 使用 WGCNA 的 goodSamplesGenes 方法清除离群基因和样本。然后利用幂函数 $am_n = |C_{mn}|^\beta$ (其中 C_{mn} 为 Gene m 和 Gene n 之间的 Pearson 相关系数) 构建加权邻接矩阵, 其中 β 是软阈值参数, 用于强调基因之间的强相关性。选择幂为 4 后, 将邻接矩阵转换为拓扑重叠矩阵 (topologically overlapping matrix, TOM), 该矩阵度量基因的网络连通性定义为与其他基因的邻接之和。使用基于 TOM 的不相似性度量进行平均连接分级聚类, 确保生成的基因树图中每个模块至少包含 30 个基因, 并且合并那些模块间的相似度小于 0.25 的部分。最终得到 32 个共表达模块, 其中灰色模块被认为是无法分配给任何模块的基因集合。

1.7 基因本体论 (gene ontology, GO)、京都基因和基因组百科全书 (Kyoto encyclopedia of genes and genomes, KEGG) 富集分析

将 8 个抗炎候选靶点转化为相应的 Entrez ID (Entrez gene identifie), 使用 R 语言及相关 R 包 (包括 clusterProfiler、org.Hs.eg.db、enrichplot、ggplot2、pathview、ggnewscale、stringr) 进行 GO 和 KEGG 的富集分析^[24], 以揭示抗炎候选靶点的功能和通路信息。显著性差异的富集结果以 $P < 0.05$ 为界定标准。

1.8 机器学习筛选潜在抗炎靶点

构建 8 种机器学习模型进一步筛选潜在抗炎靶点。利用 glmnet、randomForestSRC、kernlab、caret、

e1071R、gbm、xgboost 等 R 包来建立 8 种模型。此外, 利用 caret、pROC 等 R 包分别计算准确率 (accuracy)、精确率 (precision)、敏感性 (sensitivity)、特异性 (specificity)、曲线下面积 (area under the curve, AUC) 评估模型性能。具体公式如下。

$$\text{准确率} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{精确率} = TP / (TP + FP)$$

$$\text{敏感性} = TP / (TP + FN)$$

$$\text{特异性} = TN / (TN + FP)$$

TP 表示真阳性, TN 表示真阴性, FP 表示假阳性, FN 表示假阴性

1.8.1 最小绝对收缩和选择算子 (logistic least absolute shrinkage and selection operator, Lasso)

Lasso 是一种利用 L1 正则化 (L1-norm regularization) 的方法, 旨在通过变量选择提高预测精度和模型的可解释性^[25]。十折交叉验证证明最优 λ 值为 0.072 5, 其他超参为 $\alpha = 1$ 、family = binomial、path = TRUE、relax = TRUE。

1.8.2 Ridge 回归 Ridge 回归使用 L2 正则化 (L2-norm regularization), 它是模型系数的平方和的倍数, 添加到损失函数中有助于防止过拟合^[26]。十折交叉验证证明最优 λ 值为 0.248 7, 其他超参为 $\alpha = 0$ 、family = binomial、path = TRUE、relax = TRUE。

1.8.3 Elastic Net 回归 Elastic Net 回归结合了 Ridge 回归和 Lasso 的正则化项, 同时使用 L1 和 L2 正则化, 使得 Elastic Net 同时具有稀疏性 (L1) 和抗多重共线性的特性 (L2)^[27]。十折交叉验证证明最优 λ 值为 0.088 1, 其他超参为 $\alpha = 0.75$ 、family = binomial、path = TRUE、relax = TRUE。

1.8.4 支持向量机递归特征消除 (support vector machine recursive feature elimination, SVM-RFE) SVM-RFE 是一种强大的方法, 其目标是在两者之间建立阈值允许基于单个或多个标签预测的类特征向量。递归特征消除是一种通过递归地训练模型并删除对模型性能贡献较小的特征来选择特征的方法^[28]。十折交叉验证证明最佳特征子集为花生四烯酸 5-脂氧合酶 (arachidonate 5-lipoxygenase, ALOX5)、3-羟基-3-甲基戊二酸单酰辅酶 A 还原酶 (3-hydroxy-3-methylglutaryl-coenzyme A reductase, HMGCR)、补体 C5a 受体 1 (complement C5a receptor 1, C5AR1)、黏膜相关淋巴样组织淋巴瘤易位蛋白 1 (mucosa-associated lymphoid tissue lymphoma translocation protein 1, MALT1), 其他超参为

methods=svmRadial (径向基核函数)。

1.8.5 随机森林 (random forest, RF) RF 是一种集成学习方法, 由多个决策树组成。每个决策树都在不同的数据子集上进行训练, 而且在每个节点上, 随机选择一部分特征用于划分。最终的预测结果是所有决策树的回归问题或分类问题^[29]。最佳超参为 mtry=2、nodesize=5、replace=TRUE、sample.fraction=0.5, 其他超参为 spltrule=auc、importance=TRUE、proximity=T, forest=T。

1.8.6 梯度提升机 (gradient boosting machine, GBM) GBM 也是一种集成学习方法, 其通过迭代训练弱学习器, 每一次训练都试图修正上一轮训练的残差。最终的模型是所有弱学习器的加权和^[30]。十折交叉验证后, 最佳超参为 n.trees=100、mtry=2、replace=TRUE、n.minobsinnode=1、shrinkage=0.1、interaction.depth=3, 其他超参为 distribution=bernoulli。

1.8.7 极端梯度提升 (extreme gradient boosting, XGBoost) XGBoost 是 GBM 的一种优化版本, 引入了正则化项和树的剪枝策略, 以提高模型的泛化性能。它还使用了分桶 (bucketing) 和稀疏感知算法来处理高维稀疏数据^[31]。最佳超参为 eta=0.01、max_depth=3、min_child_weight=3、subsample=0.5、colsample_bytree=0.5, 其他超参为 objective=binary:logistic、verbose=0。

1.8.8 偏最小二乘 Logistic 回归 (partial least squares Logistic regression, PLSLR) PLSLR 是一种用于解决分类问题的统计建模方法, 结合了 PLSR 和 Logistic 回归的思想, 适用于处理多变量数据集中的分类任务^[32]。十折交叉验证后, 超参为 ncomp=2、method=pls、family=binomial、tuneLength=16。

1.9 分子对接与药物相似性评估

分子对接是一种用于药物发现的常见方法, 可准确预测蛋白质结合点和小分子配体构象, 并评估它们的结合亲和力。本研究采用 AutoDock Tools 1.5.6 和 AutoDock Vina 2 种工具进行分子对接^[33]。首先, 从 UniProt 数据库 (<https://www.uniprot.org/>) 检索蛋白质的三维结构^[34], 并将其转化为 PDBQT 蛋白质受体文件。然后从 PubChem 数据库 (<http://pubchem.ncbi.nlm.nih.gov/>) 下载小分子配体文件^[35], 对其进行修饰并保存为 PDBQT 配体文件, 将活性成分 (配体) 与靶点 (受体) 进行对接^[36]。设置的参数为 Grid center: X=-14.97, Y=-43.08, Z=-

-28.46; Grid size: X=44, Y=52, Z=96; Grid space: 0.375; Exhaustiveness: 10; Poses: 10。结果根据亲和力 (affinity) 与均方根偏差 (root mean square deviation, RMSD) 选择最佳构象进一步分析。为了明确天南星活性成分的成药性, 通过 SwissADME (<http://www.swissadme.ch/>) 评估天南星活性成分的成药性^[37]。

1.10 分子动力学模拟

利用 GROMACS (2023) 在 Ubuntu (20.04) 操作系统中进行 50 ns 的分子动力学模拟^[38], 从而验证分子对接和药物相似性评估的结果。首先, 处理蛋白的 PDB 文件, 生成相关的拓扑结构。然后, 使用 ACPYPE 工具 (<https://www.bio2byte.be/acpype/submit/>) 处理了小分子 (配体) 的拓扑结构^[39], 并将其信息添加到蛋白质拓扑文件中, 形成复合物信息文件。采用 AMBER99SB-ILDN 力场和 TIP3P (transferable intermolecular potential 3 point) 水模型进行计算, 将系统限制在边长 1.0 nm 的正立方体盒子中。添加 SPC216 水溶剂, 并通过添加 Na⁺/Cl⁻ 离子以平衡电荷, 避免蛋白与小分子碰撞。采用周期性边界的方式进行模拟。在 300 K 的温度下, 对蛋白和小分子配体复合物进行 1 000 步的能量最小化。随后, 在 300 K 温度下进行正则平衡, 使用 Parrinello-Rahman 压力耦合在 1 bar 条件下进行等温等压平衡 (100 ps, 动力学步长为 2 fs), 系统的坐标和能量每 10 ps 保存 1 次。最后, 每个体系进行 50 ns 的模拟, 对分子轨迹进行校正后用以后续评估和分析。使用 LINCS (LINear constraint solver) 方法对氢键进行约束, 计算了长程静电相互作用, 使用 PME (particle mesh ewald) 方法计算短程电力和范德华力相互作用。这些步骤有助于确保模拟的准确性和稳定性。gmx_MMPBSA 是一种新的工具用来计算 GROMACS 中的结合自由能^[40]。因此, 采用 gmx_MMPBSA 计算体系稳定后的结合自由能^[41], 具体公式如下。

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - (\Delta G_{\text{receptor}} + \Delta G_{\text{ligand}})$$

ΔG_{bind} 为结合自由能, $\Delta G_{\text{complex}}$ 为复合物的结合自由能, $\Delta G_{\text{receptor}}$ 为受体的结合自由能, ΔG_{ligand} 为配体的结合自由能

1.11 统计学分析

研究数据采用 R (4.3.2)、Python (3.7.12) 软件分析, 利用 Wilcoxon 秩和检验进行统计分析, 所有统计分析均采用双侧检验, $P < 0.05$ 作为差异显著的判断标准, $P < 0.01$ 或 $P < 0.001$ 作为差异极显著

的判断标准。

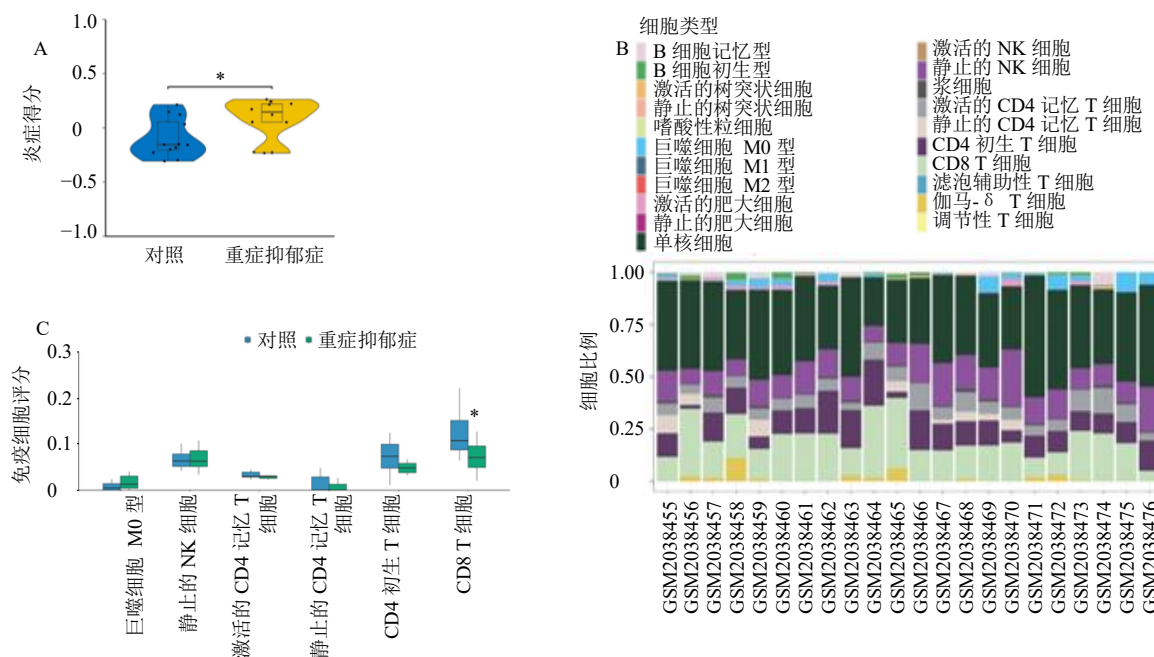
2 结果

2.1 验证免疫炎症假说与探究免疫标志物

以相关性评分 (relevance score) >8 为标准, 获得 889 个炎症相关基因。根据炎症相关基因, 利用 GSVA 在 GSE76826 数据集中计算炎症得分。运用 Wilcoxon 秩和检验比较对照组和重度抑郁症组在炎症得分上的差异。结果显示, 对照组和重度抑郁症组在炎症得分上存在显著性差异 ($P < 0.05$), 如图 1-A 所示。进而探究免疫细胞的分布, 并寻找差异免疫细胞, 如图 1-B 所示。结果显示, CD8 T 细胞免疫细胞评分在对照组和重度抑郁症组之间存在显著性差异 ($P < 0.05$), 如图 1-C 所示。

2.2 天南星治疗重度抑郁的靶点筛选

2.2.1 天南星活性成分的筛选和潜在作用靶点预测



A-炎症得分的小提琴图; B-22 种免疫细胞的堆积条形图; C-免疫细胞箱形图; * $P < 0.05$ 。

A-violin graph of inflammation score; B-stacked bar chart of 22 types of immune cells; C-boxplot of immune cell; * $P < 0.05$.

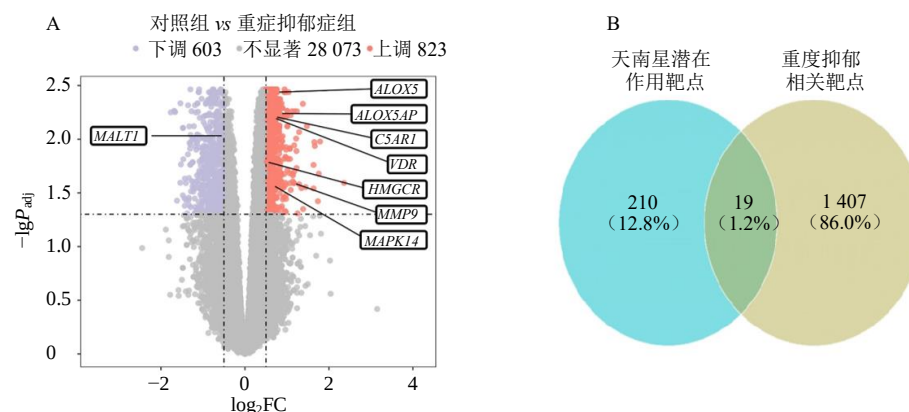
图 1 GSVA 炎症得分与差异免疫标志物

Fig. 1 GSVA inflammation scores and differential immune markers

表 1 天南星的活性成分

Table 1 Active ingredients in *Arisaematis Rhizoma*

Mol ID	活性成分	OB/%	DL
MOL013146	8,11,14-二十二碳三烯酸甲酯 (8,11,14-docosatrienoic acid, methyl ester)	43.23	0.30
MOL013156	3-苯基丙酸乙酯 (3-phenylpropyl acetate)	38.88	0.56
MOL001510	24-表柏甾醇 (24-epicampesterol)	37.58	0.71
MOL000358	β-谷甾醇 (β-sitosterol)	36.91	0.75
MOL000359	谷甾醇 (sitosterol)	36.91	0.75
MOL000449	豆甾醇 (stigmasterol)	43.83	0.76
MOL000953	胆固醇 (cholesterol)	37.87	0.68



A-重度抑郁症差异基因火山图; B-天南星潜在作用靶点和重度抑郁相关靶点交集韦恩图。

A-volcano plot of differential genes in major depressive disorder (MDD); B-Venn diagram of intersection of potential targets of *Arisaematis Rhizoma* and targets associated with MDD.

图 2 天南星潜在作用靶点与重度抑郁相关靶点的关联性

Fig. 2 Correlation between potential targets of *Arisaematis Rhizoma* and targets associated with MDD

个,如图 2-B 所示。此外,将天南星潜在作用靶点与重度抑郁相关靶点的交集以 2×2 列联表的形式展示,并通过费舍尔精确检验探究靶点之间的关联性。结果表明,天南星潜在作用靶点和重度抑郁相关靶点具有关联性,如表 2 所示。

2.3 WGCNA 分析

为了挖掘与免疫炎症最相关的基因共表达模

块,利用 WGCNA 寻找与重度抑郁症组及 T 细胞 CD8 免疫细胞评分最相关的基因共表达模块。首先开展层次聚类,在此之后去除极端值,软阈值选择为 4,如图 3-A 所示。与免疫炎症最相关的基因共表达模块为绿松石模块(Turquoise)。此外,绿松石模块中包含了 7 005 个基因,如图 3-B~C 所示。

表 2 天南星潜在作用靶点与重度抑郁相关靶点的费舍尔精确检验

Table 2 Fisher's exact test for potential targets of *Arisaematis Rhizoma* and targets associated with MDD

类别	重度抑郁相关靶点		合计	P
	有交集	无交集		
天南星潜在作用靶点	有交集	210	229	0.00***
	无交集	0	1 407	
合计	1 426	210	1 636	

*** $P < 0.001$

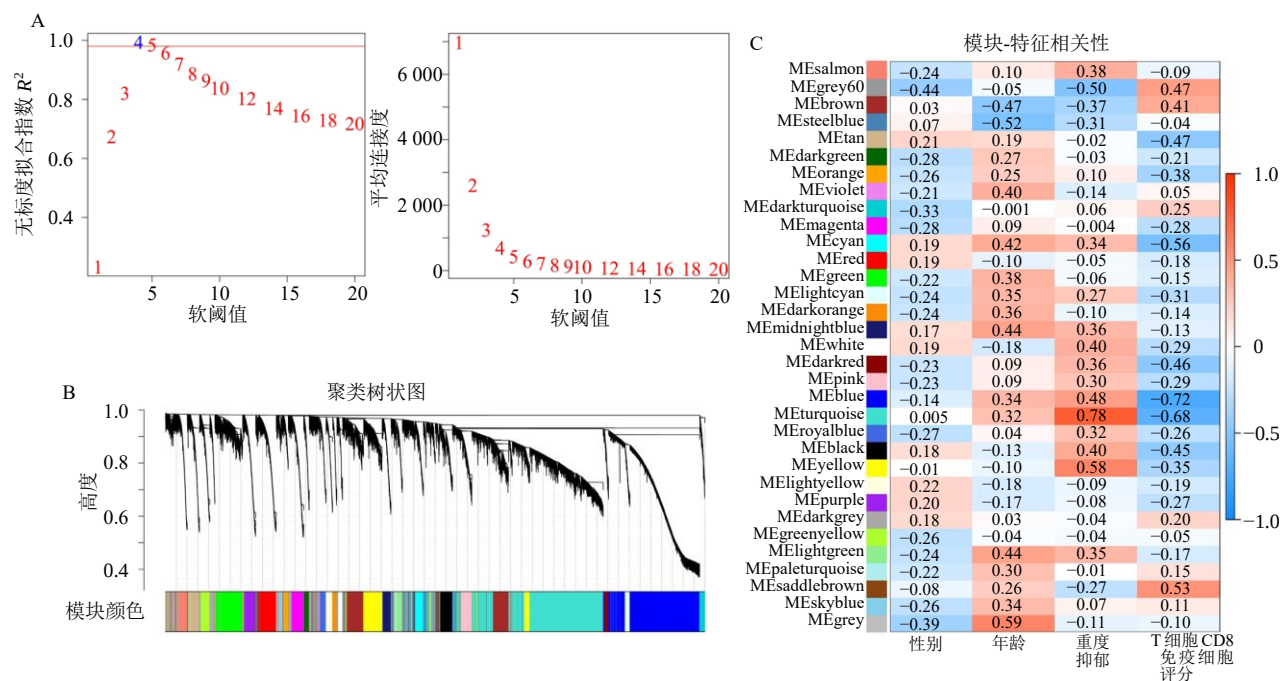
2.4 候选抗炎靶点的 GO 和 KEGG 富集分析

考虑到基因间的共表达关系并减少假阳性,本研究将炎症相关基因、天南星潜在作用靶点、重度抑郁相关靶点、绿松石模块取交集,得到 8 个抗炎候选靶点 ALOX5、ALOX5AP、C5AR1、维生素 D 受体(vitamin D receptor, VDR)、HMGCR、MMP9、丝裂原活化蛋白激酶 14 (mitogen-activated protein kinase 14, MAPK14)、MALT1 如图 4-A 所示。为了深入探究天南星治疗重度抑郁的作用机制,进行 GO 功能富集分析与 KEGG 通路富集分析。在 GO 富集的结果中,排名第 1 位的生物过程是炎症的反应调节(regulation of inflammatory response),如图 4-B 所示。KEGG 排名前 10 的通路中包括 Fc-RI 信

号通路(Fc epsilon RI signaling pathway)和 T 细胞受体信号通路(T cell receptor signaling pathway),如图 4-C 所示。上述结果表明,天南星可能通过炎症反应调节、Fc-RI 信号通路和 T 细胞受体信号通路来治疗重度抑郁。

2.5 机器学习筛选潜在抗炎靶点

2.5.1 筛选最佳机器学习模型 为了筛选天南星治疗重度抑郁的潜在抗炎靶点,采用 8 种机器学习方法进行筛选,如表 3 所示。RF 的 AUC 值在 GSE76826 数据集中最高,其他指标也表现出较高水平。因此,本研究通过受试者工作特征(receiver operator characteristic, ROC)曲线展示纳入 RF 的抗炎候选靶点,如图 5-A 所示。ALOX5 具有最高的 AUC 值,



A-对不同软阈值幂 (β) 的无标度拟合指数和不同软阈值幂的平均连接度分析; B-聚类树状图; C-基因共表达模块与重度抑郁分组及 CD8T 细胞免疫细胞评分的相关分析。

A-analysis of scale-free fit index for various soft threshold powers (β) and the mean connectivity for various soft threshold powers; B-clustering dendrogram; C-correlation analysis of gene co-expression modules, MDD groups and T cells CD8 immune cell score.

图 3 WGCNA 分析
Fig. 3 WGCNA analysis

表明 *ALOX5* 具有较好的辨别力。

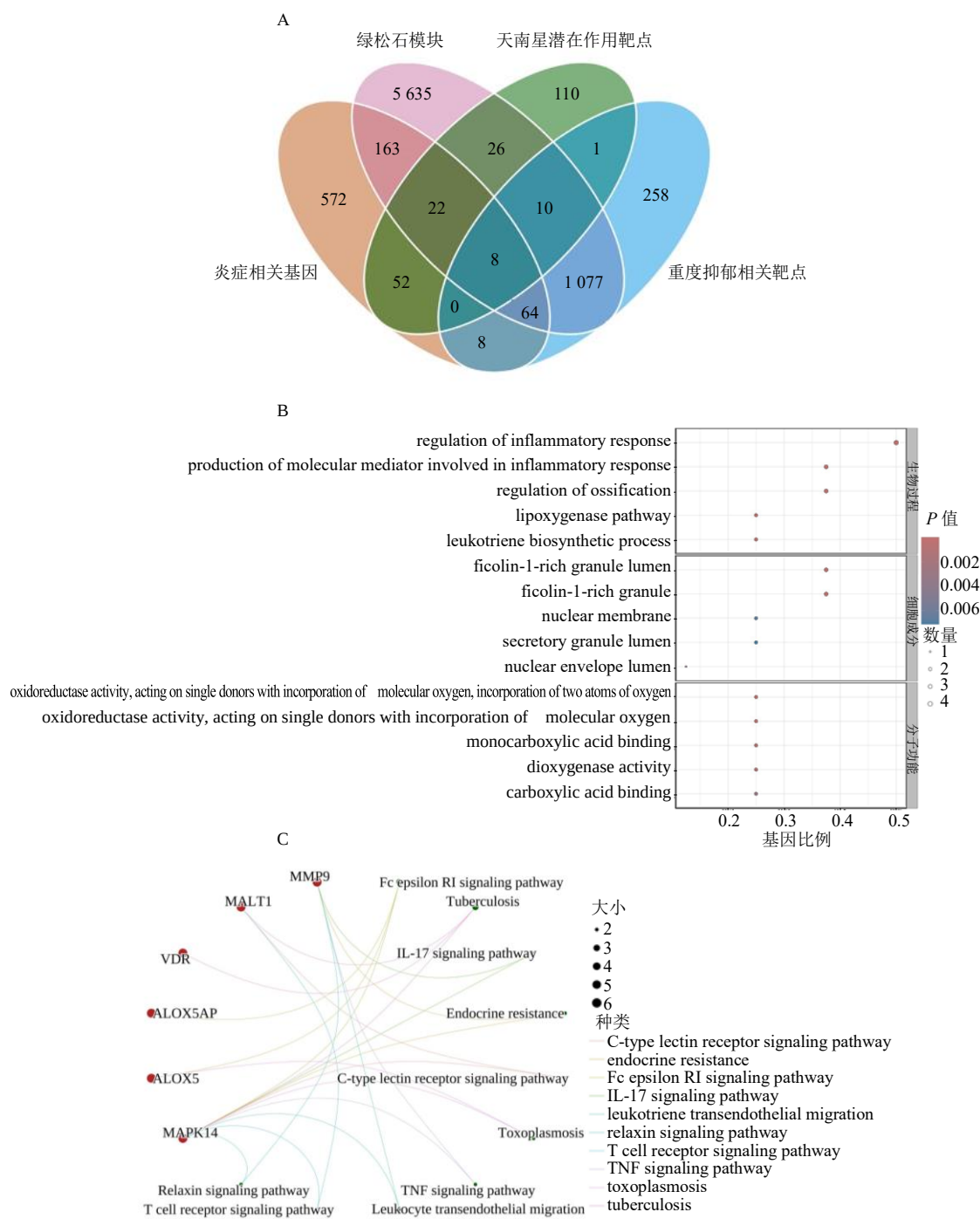
2.5.2 潜在抗炎靶点的筛选与验证 以上结果显示, RF 是最佳模型, 其中有 4 个抗炎候选靶点纳入模型中。采用 Wilcoxon 秩和检验分析 4 个抗炎候选靶点在 GSE76826 数据集中的差异表达。结果发现 4 个抗炎候选靶点在对照组和重度抑郁组之间存在显著性差异 ($P < 0.01$ 、 0.001), 如图 5-B 所示。将重度抑郁组转换为哑变量, 利用 Spearman 相关探究抗炎候选靶点与重度抑郁组、炎症得分、CD8 T 细胞免疫细胞评分之间的相关关系。4 个抗炎候选靶点与炎症得分、重度抑郁组呈正相关。同时, 4 个抗炎候选靶点与 CD8 T 细胞免疫细胞评分呈负相关, 如图 5-C 所示。在 GSE76826 数据集中, *ALOX5* 表现出较好的辨别力, 且在对照组和重度抑郁组之间存在显著性差异 ($P < 0.001$)。此外, *ALOX5* 与免疫炎症指标的相关性最强, 结果提示 *ALOX5* 为治疗重度抑郁的潜在抗炎靶点。为了验证 *ALOX5* 的可靠性, 在 GSE98793 数据集中, 采用 Wilcoxon 秩和检验分析 *ALOX5* 在对照组和重度抑郁组的差异表达。如图 5-D 所示, *ALOX5* 在对照组和重度抑郁

症组中存在显著性差异 ($P < 0.05$)。

2.6 分子对接与药物相似性评估

基于以上结果, 选择 *ALOX5* 蛋白作为天南星治疗重度抑郁的潜在抗炎靶点。为了进一步验证天南星的主要成分和 *ALOX5* 蛋白的亲合力及稳定性, 将天南星的 7 种主要活性成分与 *ALOX5* 蛋白进行分子对接。结合能反映了受体与配体结合的可能性, 结合能越低, 受体与配体的亲合力越高, 构象越稳定。如表 4 所示, 6 种天南星的主要活性成分与 *ALOX5* 蛋白对接成功, 且对接结合能均小于 -5 kcal/mol ($1 \text{ kcal} = 4.2 \text{ kJ}$), 表明上述 6 种活性成分均可以与 *ALOX5* 自由结合, 如图 6 所示。此外, 8,11,14-二十二碳三烯酸甲酯与 *ALOX5* 蛋白的结合能最低。因此, 推测该活性成分和 *ALOX5* 蛋白的亲合力最高且最稳定。

分子对接的局限性在于其依赖于打分函数来评估配体与受体的结合自由能^[42], 但实际应用中所使用的打分函数往往考虑的因素有限, 例如温度、压力和溶剂等因素通常未被充分考虑^[43]。打分函数的主要目的在于准确判断结合位置并对结合方式进行



A-炎症相关基因、天南星潜在作用靶点、重度抑郁相关靶点、绿松石模块交集的韦恩图；B-交集基因的 GO 富集分析；C-交集基因的 KEGG 富集分析。

A-Venn diagram of intersection of inflammation-related genes, potential targets of *Arisaematis Rhizoma*, targets associated with MDD, and Turquoise module; B-GO enrichment analysis of intersection genes; C-KEGG enrichment analysis of intersection genes.

图 4 候选抗炎靶点的 GO 和 KEGG 富集分析

Fig. 4 GO and KEGG enrichment analysis of candidate anti-inflammatory targets

排序，但其与实际的结合自由能相关性较低^[44]。此外，分子对接也通常假设蛋白质的大部分结构是固定的，只有少数氨基酸残基在运动^[45]。然而，根据“锁-钥”理论模型，蛋白质与配体的结合可能会引

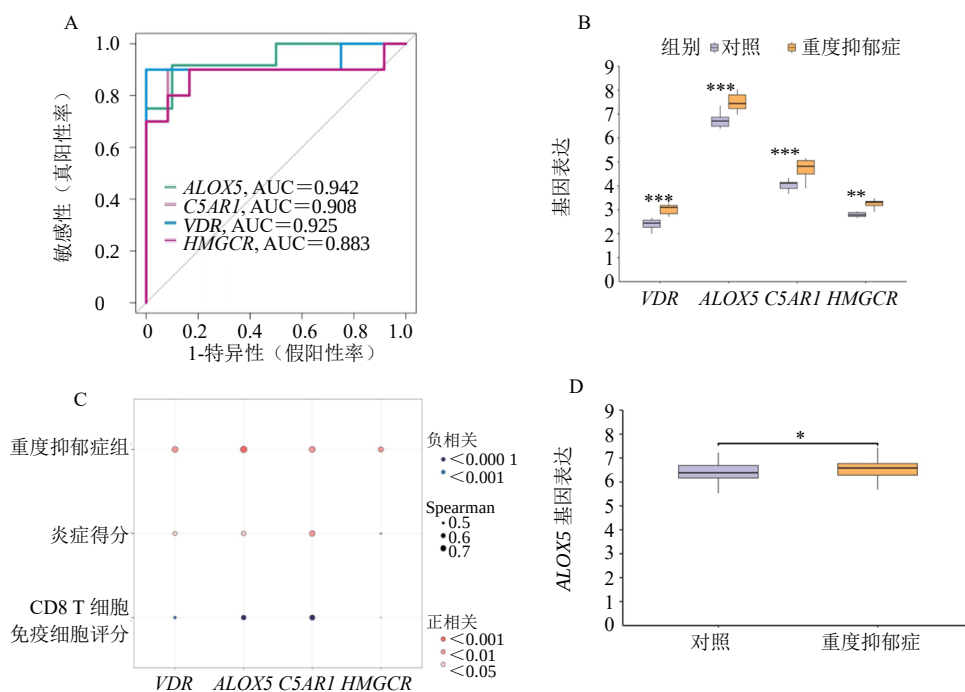
起蛋白质整体构象的改变^[46]，分子对接未能充分考虑蛋白质整体结构的变化。因此，使用分子动力学模拟与药物相似性评估来弥补分子对接的局限性。

将对接成功的天南星活性成分进行药物相似性

表 3 8 种机器学习模型在 GSE76826 数据集集中的性能表现

Table 3 Performance of eight machine learning models in GSE76826 dataset

模型	精确率	敏感性	特异性	准确率	AUC
Lasso	0.90	0.82	0.91	0.86	0.96
Ridge	0.87	0.81	0.87	0.83	0.93
Elastic net	0.90	0.88	0.89	0.88	0.97
SVM-RFE	0.91	0.83	0.90	0.86	0.96
RF	0.90	0.90	0.92	0.91	0.98
GBM	0.89	0.80	0.90	0.84	0.95
XGBoost	0.86	0.83	0.86	0.82	0.93
PLSLR	0.90	0.90	0.91	0.90	0.97



A-ROC 曲线展示 4 个抗炎候选靶点在 GSE76826 数据集集中的 AUC 值；B-抗炎候选靶点在 GSE76826 数据集集中的箱型图；C-抗炎候选靶点与重度抑郁症组、炎症得分、CD8 T 细胞免疫细胞评分的相关分析；D-*ALOX5* 在 GSE98793 数据集集中表达的箱型图；* $P < 0.05$ ** $P < 0.01$ *** $P < 0.001$ 。

A-ROC curves showing the AUC of four anti-inflammatory candidate targets in GSE76826 dataset; B-boxplots of anti-inflammatory candidate targets in GSE76826 dataset; C-correlation analysis of anti-inflammatory candidate targets, MDD group, inflammation scores, and immune cell scores of T cells CD8; D-boxplot of expression of *ALOX5* in the GSE98793 dataset; * $P < 0.05$ ** $P < 0.01$ *** $P < 0.001$.

图 5 潜在抗炎靶点的筛选与验证

Fig. 5 Screening and validation of potential anti-inflammatory targets

评估，如表 5 所示，8,11,14-二十二碳三烯酸甲酯的相对分子质量 (molecular weight, M_w) 最低。此外，所有活性成分的氢键供体 (H-bond donors) 和氢键受体 (H-bond acceptors) 数量均小于 5 和 10。在亲脂性 (Log P) 方面，仅 3-苯基丙酸乙酯的值小于 5，而 8,11,14-二十二碳三烯酸甲酯的值接近 5。对于摩尔折射率 (molecular refractivity, MR)，24-表柏甾醇

和 8,11,14-二十二碳三烯酸甲酯的数值都在 40~130。基于以上结果，8,11,14-二十二碳三烯酸甲酯表现出最佳的成药性。同时，24-表柏甾醇也表现出良好的成药性。

2.7 分子动力学模拟结果

分子对接结果表明，天南星的 6 种活性成分与 ALOX5 蛋白对接成功；药物相似性评估结果表明，

表 4 6 种天南星的主要活性成分与 ALOX5 蛋白分子对接结合能

Table 4 Molecular docking binding energy of six main active ingredients in *Arisaematis Rhizoma* and ALOX5 protein

活性成分	结合能/(kcal·mol ⁻¹)
8,11,14-二十二碳三烯酸甲酯	-11.30
3-苯基丙酸乙酯	-7.75
24-表柏甾醇	-8.62
β-谷甾醇	-8.81
谷甾醇	-8.73
豆甾醇	-8.89

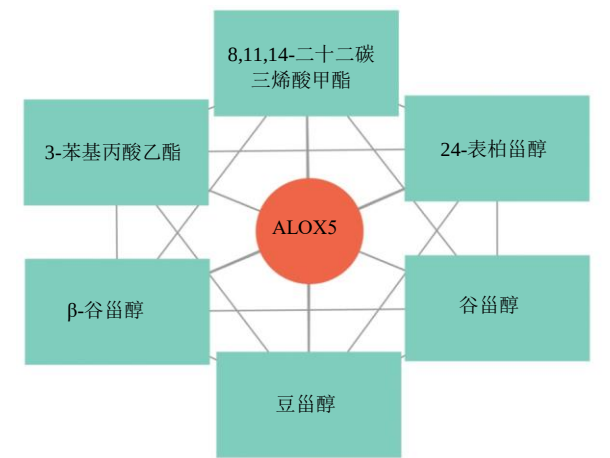


图 6 天南星 6 种活性成分与 ALOX5 蛋白的互作关系
Fig. 6 Interactions of six active ingredients of *Arisaematis Rhizoma* with ALOX5 protein

表 5 天南星 6 种活性成分的药物相似性评估

Table 5 Druglike assessment of six active ingredients of *Arisaematis Rhizoma*

活性成分	相对分子质量	氢键受体数	氢键供体数	亲脂性	摩尔折射率
8,11,14-二十二碳三烯酸甲酯	348.56	2.00	0.00	5.50	112.54
24-表柏甾醇	400.68	1.00	1.00	6.54	128.42
3-苯基丙酸乙酯	414.71	4.00	2.00	3.61	131.15
β-谷甾醇	458.55	1.00	1.00	6.73	133.23
谷甾醇	414.71	1.00	1.00	6.73	133.23
豆甾醇	412.69	1.00	1.00	6.62	132.75

用。静电势能(ΔEEL)通常指的是由于电荷之间的相互作用而导致的势能。由此可知,氢键和其他极性相互作用较弱。极性溶剂化作用(ΔEGB)表示溶剂极性对分子间相互作用的影响,2 个体系的极性溶剂化作用为正值表示极性作用较弱。极性溶剂化作用很大程度上抵消了静电势能。表面溶剂化作用(ΔESURF)表示溶剂分子与分子表面相互作用,如水分子在分子表面上形成氢键或其他相互作用。结果显示,体系中表面溶剂化作用较弱,可能 2 个体

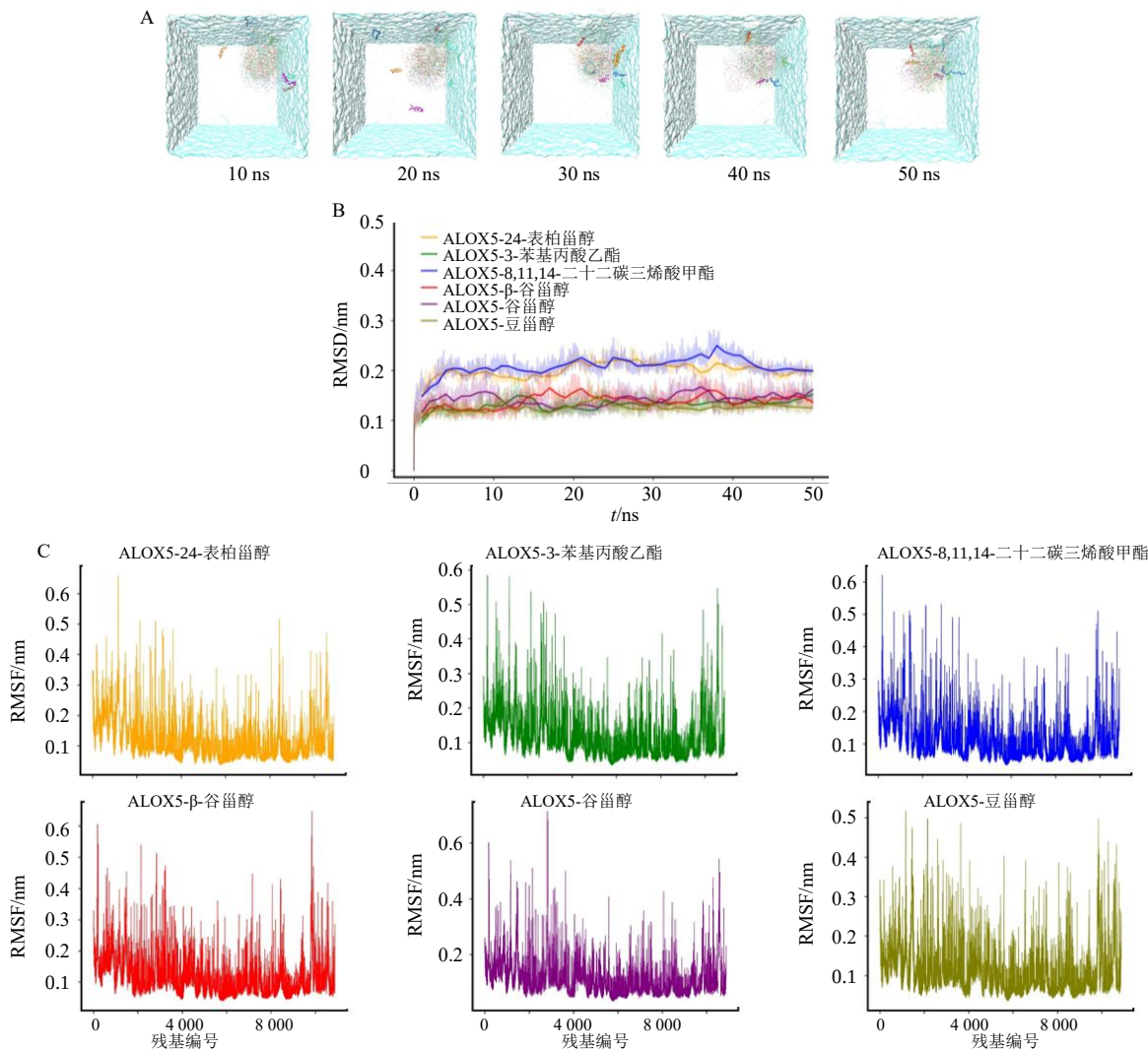
系,8,11,14-二十二碳三烯酸甲酯与 24-表柏甾醇具有较好的成药性。因此,为了验证分子对接和药物相似性评估的结果,选择天南星的 6 种活性成分与 ALOX5 蛋白进行分子动力学模拟,见图 7-A。

2.7.1 RMSD RMSD 可以反映受体-配体复合物在运动过程中的稳定性。6 种复合物体系均在 20 ns 后 RMSD 值达到稳定状态,体系的 RMSD 值波动程度均较小(<0.1 nm)。由此可见,天南星的 6 种活性成分均可以和 ALOX5 蛋白稳定结合,见图 7-B。

2.7.2 均方根波动(root mean square fluctuation, RMSF) 均方根波动可以反映受体-配体复合物氨基酸骨架原子在动力学模拟过程中的波动性。在 0~50 ns,6 种复合物的 RMSF 具有相似的波动趋势。6 种复合物的 RMSF 峰值与谷值位置基本相同,推测 6 种活性成分的起效部位近似,可能都是通过结合 ALOX5 蛋白,激活 Fc-RI 信号通路,进而调控炎症反应,见图 7-C。

2.7.3 结合自由能 结合自由能反映了受体-配体之间亲和力的差异。为了进一步验证 8,11,14-二十二碳三烯酸甲酯和 24-表柏甾醇结合 ALOX5 蛋白后的亲和力差异。采用 gmx_MMPBSA 计算 8,11,14-二十二碳三烯酸甲酯和 24-表柏甾醇结合 ALOX5 蛋白后的结合自由能。范德华力(ΔVDWAALS)越低表明非极性相互作用越强。根据结果可以推测疏水作用等非极性相互作用在结合中发挥了积极的作

系的结合以疏水作用等非极性的方式来结合。因此,2 个体系的非极性相互作用对于配体和受体的结合更为有利,其中范德华力发挥了主导作用。结合自由能(ΔTOTAL)越小,越有利于蛋白质和配体的结合,8,11,14-二十二碳三烯酸甲酯和 24-表柏甾醇的结合能分别为(-7.85±7.75)、(-2.47±5.48) kcal/mol,表明 8,11,14-二十二碳三烯酸甲酯的结合自由能比 24-表柏甾醇更低,8,11,14-二十二碳三烯酸甲酯更容易和 ALOX5 蛋白结合,如表 6 所示。



A-在水溶液系统中，6 种活性成分和 ALOX5 蛋白的模拟轨迹（0~50 ns）；B-受体-配体复合物 0~50 ns 的 RMSD；C-受体-配体复合物 0~50 ns 的 RMSF。
A-simulation trajectories of six active ingredients and ALOX5 protein in the aqueous solution system were obtained from 0 to 50 ns; B-root mean square deviation of receptor-ligand complexes from 0 to 50 ns; C-root mean square fluctuation of receptor-ligand complex from 0 to 50 ns.

图 7 分子动力学模拟结果
Fig. 7 Results of molecular dynamics simulation

表 6 8,11,14-二十二碳三烯酸甲酯和 24-表柏甾醇与 ALOX5 蛋白的结合自由能及能量值
Table 6 Binding free energy and energy values of 8,11,14-docosatrienoic acid, methyl ester and 24-epicampestrol with ALOX5 protein

ALOX5-配体	$\Delta VDWAAALS/(kcal \cdot mol^{-1})$	$\Delta EEL/(kcal \cdot mol^{-1})$	$\Delta EGB/(kcal \cdot mol^{-1})$	$\Delta ESURF/(kcal \cdot mol^{-1})$	$\Delta TOTAL/(kcal \cdot mol^{-1})$
ALOX5-8,11,14-二十二碳三烯酸甲酯	-12.23 ± 11.24	-1.12 ± 4.03	7.30 ± 7.37	-1.80 ± 1.68	-7.85 ± 7.75
ALOX5-24-表柏甾醇	-3.52 ± 7.15	-0.28 ± 1.52	1.75 ± 3.40	-0.43 ± 0.89	-2.47 ± 5.48

3 讨论

本研究基于机器学习与分子动力学模拟挖掘天南星治疗重度抑郁的潜在抗炎靶点和关键活性成分。研究结果指出，重度抑郁的发病可能跟炎症反

应及 CD8 T 细胞有关，这和以往的研究结果一致。Krishnadas 等^[47]研究发现，约有 1/3 的重度抑郁症患者具有比正常人更高水平的炎症标志物。值得注意的是，与自身免疫性疾病和传染病患者相比，重

度抑郁患者的炎症标志物显著升高^[48]。因此,炎症反应可能是导致重度抑郁的重要发病机制之一,并在免疫细胞的类型上具有一定的特异性。Evans 等^[49]研究表明,焦虑和重度抑郁症状都与 CD8 T 细胞的数量显著相关,这和本研究的结果一致,本研究发现 CD8 T 细胞的免疫细胞评分在对照组和重度抑郁组具有显著性差异。因此,推测焦虑和重度抑郁会导致持续的炎症反应,引起 T 细胞的过度激活,进而导致 CD8 T 细胞受损,最终使 CD8 T 细胞数量减少。

为了探究天南星的潜在作用靶点与重度抑郁相关靶点的关联性,本研究通过费舍尔精确检验,发现天南星的潜在作用靶点与重度抑郁的相关靶点存在 19 个交集靶点,具有关联性。以往的研究和本研究均发现,CD8 T 细胞可能是重度抑郁的重要炎症标志物。由于基因之间存在相似的表达模式,从而更好地进行协同作用。因此,根据重度抑郁分组与 CD8 T 细胞免疫细胞评分,利用 WGCNA 分析寻找与之相关性最高的基因共表达模块。考虑到基因间的共表达关系并减少假阳性,将炎症相关基因、天南星潜在作用靶点、重度抑郁相关靶点、基因共表达模块取交集,获得 8 个抗炎候选靶点。KEGG 和 GO 富集分析显示,8 个抗炎候选靶点可能涉及炎症反应的生物学过程和 T 细胞受体信号通路。构建 8 种机器学习筛选潜在抗炎靶点,其中 RF 的 AUC 值最高,表明 RF 纳入的靶点具有最好的辨别力。上述结果和以往的研究结果一致,RF 具有“开箱即用”的特性,在众多机器学习算法中 RF 只需要最小的超参调整便可以达到最高的准确性^[50]。RF 纳入的靶点为 *VDR*、*ALOX5*、*C5AR1*、*HMGCR*。在 GSE76826 数据集中,4 个靶点在对照组和重度抑郁组之间存在显著性差异 ($P < 0.01$ 、 0.001),且都和炎症得分、CD8 T 细胞免疫细胞评分、重度抑郁分组显著相关。*ALOX5* 在 GSE76826 数据集中的 AUC 值最高,且和炎症得分、CD8 T 细胞免疫细胞评分、重度抑郁分组的相关系数(绝对值)最大。基于以上结果,*ALOX5* 可能为治疗重度抑郁的潜在抗炎靶点。GO 和 KEGG 富集分析显示,*ALOX5* 涉及炎症反应的调控过程和 Fc-RI 信号通路。Wong 等^[51]研究发现 Fc-RI 信号通路还调节重要炎症介质的释放,如细胞因子、白细胞介素、白三烯和前列腺素。基于本研究结果,推测天南星通过作用于 *ALOX5* 蛋白,激活 Fc-RI 信号通路,进而调控炎症反应与 CD8 T 细胞,在治疗重度抑郁方面

发挥作用。

为了进一步筛选天南星的关键活性成分,通过分子对接、药物相似性评估和分子动力学模拟的方法筛选天南星的关键活性成分。分子对接结果表明,天南星的 6 种活性成分均可以和 *ALOX5* 蛋白稳定结合,其中 8,11,14-二十二碳三烯酸甲酯结合能最低。药物相似性评估显示,8,11,14-二十二碳三烯酸甲酯和 24-表柏甾醇具有较好的成药性。分子动力学模拟进一步佐证了分子对接和药物相似性评估的结果。综上推测,天南星治疗重度抑郁的关键活性成分是 8,11,14-二十二碳三烯酸甲酯。8,11,14-二十二碳三烯酸甲酯在重度抑郁治疗中的报道相对较少。陈思软等^[52]研究发现,8,11,14-二十二碳三烯酸甲酯作为胆南星的活性成分之一对癫痫具有治疗效果。8,11,14-二十二碳三烯酸甲酯可能具有透过血脑屏障的能力,有望用于治疗精神类或神经系统疾病。但这一推测仍需进一步的实验验证。

本研究仍有一些局限性:基于公共数据库来探究相关靶点,数据库数据的准确性和及时性有待科学验证;研究发现的主要活性成分并不能充分代表天南星,因此需要考虑相关的药效学和分子生物学实验,以对研究结果进行更深入的验证。由于这些潜在活性成分、作用靶点以及相关机制尚未得到验证和阐释,因此本研究具有进一步的发展空间和研究意义。

综上所述,*ALOX5* 和 8,11,14-二十二碳三烯酸甲酯为天南星治疗重度抑郁的潜在抗炎靶点及关键活性成分。后续可对 *ALOX5* 和 8,11,14-二十二碳三烯酸甲酯治疗重度抑郁的作用开展进一步的实验探究,为相关药物开发提供有力支撑。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] First M B, Reed G M, Hyman S E, *et al.* The development of the ICD-11 Clinical Descriptions and Diagnostic Guidelines for Mental and Behavioural Disorders [J]. *World Psychiatry*, 2015, 14(1): 82-90.
- [2] Suseelan S, Pinna G. Heterogeneity in major depressive disorder: the need for biomarker-based personalized treatments [J]. *Adv Clin Chem*, 2023, 112: 1-67.
- [3] Ferrari A J, Charlson F J, Norman R E, *et al.* Burden of depressive disorders by country, sex, age, and year: Findings from the global burden of disease study 2010 [J]. *PLoS Med*, 2013, 10(11): e1001547.
- [4] Athira K V, Bandopadhyay S, Samudrala P K, *et al.* An

- overview of the heterogeneity of major depressive disorder: Current knowledge and future prospective [J]. *Curr Neuropsychopharmacol*, 2020, 18(3): 168-187.
- [5] Jesulola E, Micalos P, Baguley I J. Understanding the pathophysiology of depression: From monoamines to the neurogenesis hypothesis model—Are we there yet? [J]. *Behav Brain Res*, 2018, 341: 79-90.
- [6] Beurel E, Toups M, Nemeroff C B. The bidirectional relationship of depression and inflammation: Double trouble [J]. *Neuron*, 2020, 107(2): 234-256.
- [7] Slavich G M, Irwin M R. From stress to inflammation and major depressive disorder: A social signal transduction theory of depression [J]. *Psychol Bull*, 2014, 140(3): 774-815.
- [8] Zheng X, Zhang X L, Wang G J, *et al.* Treat the brain and treat the periphery: Toward a holistic approach to major depressive disorder [J]. *Drug Discov Today*, 2015, 20(5): 562-568.
- [9] Huang C F, Yang R S, Liu S H, *et al.* Evidence for improved neuropharmacological efficacy and decreased neurotoxicity in mice with traditional processing of *Rhizoma Arisaematis* [J]. *Am J Chin Med*, 2011, 39(5): 981-998.
- [10] Liu S S, Sheng W L, Li Y, *et al.* Chemical constituents from *Alismatis Rhizoma* and their anti-inflammatory activities *in vitro* and *in vivo* [J]. *Bioorg Chem*, 2019, 92: 103226.
- [11] Prathipati P, Mizuguchi K. Systems biology approaches to a rational drug discovery paradigm [J]. *Curr Top Med Chem*, 2016, 16(9): 1009-1025.
- [12] 高晟玮, 薛亚然, 张垚, 等. 基于生物信息学探讨 SARS-CoV-2 介导免疫炎症反应诱发动脉粥样硬化进展的潜在机制与防治中药筛选 [J]. *中草药*, 2023, 54(8): 2523-2535.
- [13] Alonso H, Bliznyuk A A, Gready J E. Combining docking and molecular dynamic simulations in drug design [J]. *Med Res Rev*, 2006, 26(5): 531-568.
- [14] Marshall G. Computer-aided drug design [J]. *Annu Rev Pharmacol Toxicol*, 1987, 27: 193-213.
- [15] Chicco D. geneExpressionFromGEO: An R package to facilitate data reading from gene expression omnibus (GEO) [J]. *Methods Mol Biol*, 2022, 2401: 187-194.
- [16] Rebhan M, Chalifa-Caspi V, Prilusky J, *et al.* GeneCards: A novel functional genomics compendium with automated data mining and query reformulation support [J]. *Bioinformatics*, 1998, 14(8): 656-664.
- [17] Hänzelmann S, Castelo R, Guinney J. GSEA: Gene set variation analysis for microarray and RNA-seq data [J]. *BMC Bioinformatics*, 2013, 14: 7.
- [18] Newman A M, Liu C L, Green M R, *et al.* Robust enumeration of cell subsets from tissue expression profiles [J]. *Nat Methods*, 2015, 12(5): 453-457.
- [19] Ru J L, Li P, Wang J N, *et al.* TCMSP: A database of systems pharmacology for drug discovery from herbal medicines [J]. *J Cheminform*, 2014, 6: 13.
- [20] Kim S, Chen J, Cheng T J, *et al.* PubChem 2019 update: Improved access to chemical data [J]. *Nucleic Acids Res*, 2019, 47(D1): D1102-D1109.
- [21] Gfeller D, Grosdidier A, Wirth M, *et al.* Swiss TargetPrediction: A web server for target prediction of bioactive small molecules [J]. *Nucleic Acids Res*, 2014, 42: W32-W38.
- [22] Ritchie M E, Phipson B, Wu D, *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies [J]. *Nucleic Acids Res*, 2015, 43(7): e47.
- [23] Pei G, Chen L, Zhang W. WGCNA application to proteomic and metabolomic data analysis [J]. *Methods Enzymol*, 2017, 585: 135-158.
- [24] Zhang X, Yu X P, Yu Z C, *et al.* Network pharmacology and bioinformatics to identify molecular mechanisms and therapeutic targets of Ruyi Jinhuang Powder in the treatment of monkeypox [J]. *Medicine*, 2023, 102(17): e33576.
- [25] Muthukrishnan R, Rohini R. LASSO: A feature selection technique in predictive modeling for machine learning [A] // 2016 IEEE International Conference on Advances in Computer Applications (ICACA) [C]. Coimbatore: IEEE, 2016: 18-20.
- [26] Liman M D, Osanga S I, Alu E S, *et al.* Regularization effects in deep learning architecture [J]. *J Nigerian Soc Physical Sci*, 2024, 6(2): 1911.
- [27] Al Dabal M A A. A comparative study of ridge, LASSO and elastic net estimators [D]. Ottawa: Carleton University, 2021.
- [28] Sanz H, Valim C, Vegas E, *et al.* SVM-RFE: selection and visualization of the most relevant features through non-linear kernels [J]. *BMC Bioinform*, 2018, 19: 1-18.
- [29] Denisko D, Hoffman M M. Classification and interaction in random forests [J]. *Proceed Nat Acad Sci*, 2018, 115(8): 1690-1700.
- [30] Gong M J, Bai Y, Qin J, *et al.* Gradient boosting machine for predicting return temperature of district heating system: A case study for residential buildings in Tianjin [J]. *J Build Eng*, 2020, 27: 100950.
- [31] Babajide Mustapha I, Saeed F. Bioactive molecule prediction using extreme gradient boosting [J]. *Molecules*, 2016, 21(8): 983.
- [32] Abdi H. Partial least square regression (PLS regression)

- [J]. *Encycl Res Methods Social Sci*, 2003, 6(4): 792-795.
- [33] Eberhardt J, Santos-Martins D, Tillack A F, *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.
- [34] Berman H, Henrick K, Nakamura H, *et al*. The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data [J]. *Nucleic Acids Res*, 2007, 35: D301-D303.
- [35] Kim S, Thiessen P A, Bolton E E, *et al*. PubChem substance and compound databases [J]. *Nucleic Acids Res*, 2016, 44(D1): D1202-D1213.
- [36] Peri S, Navarro J D, Kristiansen T Z, *et al*. Human protein reference database as a discovery resource for proteomics [J]. *Nucleic Acids Res*, 2004, 32: D497-D501.
- [37] Daina A, Michielin O, Zoete V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules [J]. *Sci Rep*, 2017, 7: 42717.
- [38] Berendsen H J C, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation [J]. *Comput Phys Commun*, 1995, 91(1/2/3): 43-56.
- [39] Kagami L, Wilter A, Diaz A, *et al*. The ACPYPE web server for small-molecule MD topology generation [J]. *Bioinformatics*, 2023, 39(6): btad350.
- [40] Valdés-Tresanco M S, Valdés-Tresanco M E, Valiente P A, *et al*. gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS [J]. *J Chem Theory Comput*, 2021, 17(10): 6281-6291.
- [41] Hou T J, Wang J M, Li Y Y, *et al*. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations [J]. *J Chem Inf Model*, 2011, 51(1): 69-82.
- [42] Guedes I A, de Magalhães C S, Dardenne L E. Receptor-ligand molecular docking [J]. *Biophys Rev*, 2014, 6(1): 75-87.
- [43] Xie H D, Li Y P, Yu F, *et al*. An investigation of molecular docking and molecular dynamic simulation on imidazopyridines as B-raf kinase inhibitors [J]. *Int J Mol Sci*, 2015, 16(11): 27350-27361.
- [44] Raha K, Merz K M. Large-scale validation of a quantum mechanics based scoring function: Predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes [J]. *J Med Chem*, 2005, 48(14): 4558-4575.
- [45] Meng X Y, Zhang H X, Mezei M, *et al*. Molecular docking: A powerful approach for structure-based drug discovery [J]. *Curr Comput Aided Drug Des*, 2011, 7(2): 146-157.
- [46] Frimurer T M, Peters G H, Iversen L F, *et al*. Ligand-induced conformational changes: Improved predictions of ligand binding conformations and affinities [J]. *Biophys J*, 2003, 84(4): 2273-2281.
- [47] Krishnadas R, Cavanagh J. Depression: An inflammatory illness? [J]. *J Neurol Neurosurg Psychiatry*, 2012, 83(5): 495-502.
- [48] Osimo E F, Pillinger T, Rodriguez I M, *et al*. Inflammatory markers in depression: A Meta-analysis of mean differences and variability in 5 166 patients and 5 083 controls [J]. *Brain Behav Immun*, 2020, 87: 901-909.
- [49] Evans D L, Ten Have T R, Douglas S D, *et al*. Association of depression with viral load, CD8 T lymphocytes, and natural killer cells in women with HIV infection [J]. *Am J Psychiatry*, 2002, 159(10): 1752-1759.
- [50] Jordan M I, Mitchell T M. Machine learning: Trends, perspectives, and prospects [J]. *Science*, 2015, 349(6245): 255-260.
- [51] Wong B R, Grossbard E B, Payan D G, *et al*. Targeting Syk as a treatment for allergic and autoimmune disorders [J]. *Expert Opin Investig Drugs*, 2004, 13(7): 743-762.
- [52] 陈思软, 赵姣娇, 孔德志, 等. 基于网络药理学和实验验证探究胆南星对癫痫的作用机制 [J]. *药学报*, 2022, 57(4): 1031-1043.

[责任编辑 潘明佳]