

中药活性成分中新型冠状病毒 3CL 蛋白酶抑制剂的消息传递机制神经网络虚拟筛选

张智强, 朱增文, 陈俊利, 绪连彩*

郑州轻工业大学材料与化学工程学院, 河南 郑州 450000

摘要: **目的** 在新型冠状病毒感染尚无特效药的情况下, 根据潜在的治疗靶点, 从中药活性成分中筛选出先导化合物为药物研发、中药组方等提供理论线索。**方法** 构建消息传递机制神经网络 (message passing neural networks, MPNN) 模型, 以化合物描述符 SMILES 码为输入, 以化合物对新型冠状病毒 3CL 蛋白酶抑制活性为输出, 利用开源数据对模型进行训练和优化。**结果** 用优化后的模型从 186 味清热解毒中药所含的 3 863 个活性成分中筛选出 101 个潜在的抑制剂。其中龙胆素、桑辛素 C、5-羟基-4-氧代戊酸甲酯等化合物预测活性较高, 黄芩、苍耳子、昆布、芫荽、紫苏等中药含有的潜在抑制剂数量较多。**结论** 使用 MPNN 模型虚拟筛选出的抑制剂中约 1/5 已被其他的研究报道验证有效, 证明了 MPNN 模型虚拟筛选结果的可靠性。此外, 优化的神经网络模型微调后可用于分子其他属性的预测, 在药物虚拟筛选领域有着广泛的应用前景。

关键词: 严重急性呼吸综合征冠状病毒 2 (SARS-CoV-2); 3CL 蛋白酶抑制剂; 药物虚拟筛选; 消息传递机制神经网络; 龙胆素、桑辛素 C、5-羟基-4-氧代戊酸甲酯

中图分类号: R284; R285 文献标志码: A 文章编号: 0253-2670(2024)03-0895-12

DOI: 10.7501/j.issn.0253-2670.2024.03.020

Virtual screening of SARS-CoV-2 3CL protease inhibitors from active ingredients of traditional Chinese medicines by message passing neural networks

ZHANG Zhiqiang, ZHU Zengwen, CHEN Junli, XU Liancai

School of Materials and Chemical Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000, China

Abstract: Objective In the absence of specific drugs for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections, screening of lead compounds from active ingredients of traditional Chinese medicine (TCM) based on potential therapeutic targets can provide theoretical clues for drug research and development and TCM formulations. **Methods** In this study, a message passing neural networks (MPNN) model was constructed, with the compound descriptor SMILES code as input and the inhibitory activity of compounds against SARS-CoV-2 3CL protease as output. The model was trained and optimized using open-source data. **Results** The optimized model successfully screened out 101 potential inhibitors from 3 863 active ingredients of 186 TCMs with heat-clearing and detoxifying effects. Among them, compounds such as gentianin, mycin C, methyl 5-hydroxy-4-oxopentanoate, and others exhibited the higher predicted activity, while TCMs like Huangqin (*Scutellariae Radix*), Cang'erzi (*Xanthii Fructus*), Kunbu (*Laminariae Thallus*), Yansui (*Herba Coriandri Sativi*), Zisu [*Perilla frutescens* (L.) Britt.] had a relatively higher number of potential inhibitors. **Conclusion** Approximately one-fifth of the inhibitors identified by MPNN model screening have been validated by other experimental or theoretical studies, demonstrating the reliability of the screening results. Additionally, the neural network model optimized in this study can be fine-tuned for predicting other molecular properties, indicating its extensive application prospects in drug virtual screening.

Key words: severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2); 3CL protease inhibitors; drug screening in silicon; message passing neural network; gentianin; mycin C; methyl 5-hydroxy-4-oxopentanoate

收稿日期: 2023-09-20

基金项目: 国家自然科学基金-河南联合基金项目 (U2004196); 郑州轻工业大学博士基金项目 (2011BSJJ012)

作者简介: 张智强, 硕士生导师, 讲师, 研究方向为化学信息学。E-mail: zhangzq@zzuli.edu.cn

*通信作者: 绪连彩, 硕士生导师, 副教授, 研究方向为理论与计算化学。E-mail: xulc@zzuli.edu.cn

2019 年底爆发的新型冠状病毒肺炎 (corona virus disease 2019, COVID-19) 已成为席卷全球的公共卫生事件, 截至 2023 年 7 月, 全球已报告超过 7.6 亿例确诊病例以及超过 690 万例死亡病例^[1]。虽然目前疫情已经过去, 但不排除未来再次爆发的可能。2020 年 2 月世界卫生组织将引起 COVID-19 疾病的病毒命名为严重急性呼吸综合征冠状病毒 2 (severe acute respiratory syndrome coronavirus 2, SARS-CoV-2)^[2-3]。该病毒与引起 2003 年非典型性肺炎以及 2012 年的中东呼吸综合征 (middle east respiratory syndrome, MERS) 的病原体 SARS-CoV、MERS-CoV 病毒同属于冠状病毒。冠状病毒因外形似皇冠得名, 属于 RNA 病毒, 其基因组大小为 27~32 kb, 是目前已知的最大病毒 RNA 基因组^[4]。RNA 病毒本身具有较高的变异率, 使得疫苗的有效性逐渐降低, 这给疫情的防控带来了巨大挑战, 有效治疗药物的研发显得尤为重要。

RNA 依赖性 RNA 聚合酶 (RNA-dependent RNA polymerase, RdRp) 主要负责病毒 RNA 的复制, 因此成为研究广谱抗病毒药物的关键靶点。病毒复制相关的蛋白酶如半胱氨酸蛋白酶-木瓜蛋白酶样蛋白酶 (PL protease, PL^{pro})、3C 样蛋白酶 (3C-like protease, 3CL^{pro})、弗林蛋白酶 (Furin protease) 等则在病毒蛋白的加工与多聚蛋白的裂解中具有重要作用, 可成为潜在药物靶点^[5-6], 有助于抗病毒药物的研发, 并且 3CL 蛋白酶在冠状病毒中高度保守, 同时尚未发现 3CL 蛋白酶的人类同源蛋白。因此, 3CL 蛋白酶成为抗 SARS-CoV-2 病毒药物的理想靶点^[7]。目前, 用于治疗 SARS-CoV-2 感染的匹洛那韦和利托那韦就是 3CL 蛋白酶抑制剂。

我国 SARS-CoV-2 感染患者早期的治疗过程, 中医药发挥了重要的作用^[8]。在国家卫生健康委员会颁布的《新型冠状病毒肺炎诊疗方案 (试行第 8 版)》中, 推荐了藿香正气胶囊、金花清感颗粒、连花清瘟胶囊、疏风解毒胶囊等一系列中药, 在 COVID-19 患者的治疗中发挥了重要的作用, 但其作用机制尚待明确^[9-10]。由于药物研究周期长的客观特点, 及时开发针对新发突出病毒性疾病的新型药物是无法实现的。筛选抑制 SARS-CoV-2 3CL 蛋白酶活性高的中药有效成分, 不仅能够为开发治疗 SARS-CoV-2 感染药物发现先导化合物, 也能够为未来不可预期的新型冠状病毒引发的公共卫生事件的防治积累经验^[11]。本研究通过构建消息传递机制

的神经网络 (message passing neural networks, MPNN) 模型, 采用 6 000 余个具有 SARS-CoV-2 3CL 蛋白酶抑制活性的化合物数据集对模型进行训练和评估, 使用该模型对中药活性成分数据库中收录的 186 味具有清热解毒功效的中药活性成分进行预测, 为开发治疗 SARS-CoV-2 感染的中药方剂及中药药理学研究等提供理论线索。

1 数据集与神经网络模型

1.1 数据集

本研究用于神经网络模型优化的数据集来源于 PubChem 网站上公布的 1 组对 SARS-CoV-2 3CL 蛋白酶的抑制率, 该抑制率是使用荧光共振能量转移技术在化合物浓度为 20 $\mu\text{mol/L}$ 时测得^[12]。该数据集中原始数据样本共 7 220 条, 经数据清洗、合并重复数据后, 共 6 157 个样本。其中抑制率 > 90% 的有 161 个样本, 占比为 2.6%, 抑制率 > 10% 的有 3 443 个样本, 占比 55.9%。

虚拟药物筛选的数据集来自于中药系统药理学数据库 (TCMSP)^[13]中收录的 186 味清热解毒中药的活性成分, 约 4 000 个数据样本。

1.2 消息传递机制神经网络模型 (message passing neural networks, MPNN)

MPNN 是一种图神经网络 (graph neural network, GNN) 的变体, 与其他类型的图神经网络相比, MPNN 的主要特点是在每个图卷积层中, MPNN 会计算每个节点的消息, 将其聚合到邻近节点中, 然后再次传递消息。这个消息传递过程可以发生多次, 每一次都可以聚合不同的信息。这种消息传递方式使 MPNN 能够捕捉节点的局部和全局信息。另外, MPNN 可以处理具有不同节点和边特征的图, 使得其适用于各种应用, 在分子图的表示学习中表现尤为出色^[14]。MPNN 的特点非常适合药物发现、分子属性预测等任务。

本研究使用的 MPNN 模型是基于谷歌公司发布的 MPNN 通用框架以及 DeepChem 中 MPNN 模型^[14]搭建的, 网络整体结构见图 1。该模型用于处理无向量图 $G(v, w, e_{vw})$, 其中 v 代表节点, w 代表边, e_{vw} 代表边的特征。用无向量图 G 可以描述分子特征, 节点即为分子中的原子, 节点属性即原子的属性, 本研究采用的原子属性包括元素名、价态、杂化类型以及连接的氢原子数目。无向量图 G 中的边即为原子间的键, 键的属性即为边的特征, 本研究采用键的属性有 2 个, 一是键级, 如单键、

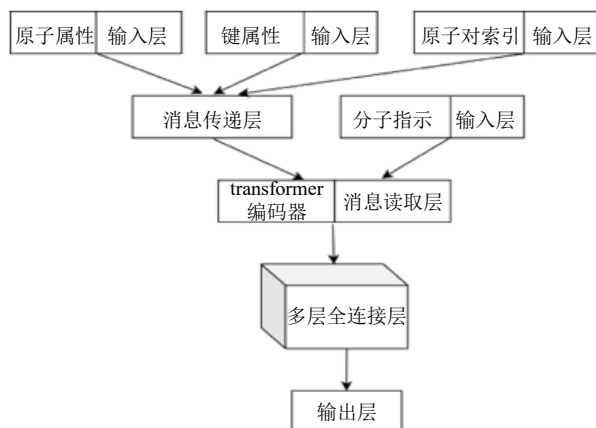


图 1 消息传递机制神经网络模型结构

Fig. 1 Structure of message passing neural network (MPNN) model

双键、三键、芳香键等；二是是否为共轭键。本研究采用的 MPNN 模型包括消息传递、消息读取、活性预测 3 个阶段。首先，将原子属性、键属性和键两端原子的索引数据送入消息传递层，通过多次迭代聚合邻居节点的信息并更新节点状态，得到最终的分子描述图；接下来将分子描述图和分子指示器数据送入消息读取层，该层的本质是 1 个变换网络（transformer）的编码器，其根据分子指示器数据将分子描述图划分成多个子图，通过应用多头注意力机制、残差连接和层归一化，从图数据中提取有用的表示，并将整个图的信息汇总为一个读出表示；最后，把读出表示送入多个全连接层做回归分析，输出分子的属性，即 20 $\mu\text{mol/L}$ 时对 SARS-CoV-2 3CL 蛋白酶的抑制率。

MPNN 模型使用 Python 3.7 搭建，在 TensorFlow 2.6 平台上进行优化。模型优化和后续的药物筛选都是在计算机上完成。

2 结果与讨论

2.1 MPNN 模型优化

MPNN 模型优化时，以化合物的简化分子线性输入规范编码（simplified molecular input line entry system, SMILES）码作为神经网络的输入数据，以化合物 20 $\mu\text{mol/L}$ 时对 SARS-CoV-2 3CL 蛋白酶抑制率为神经网络的输出数据，数据集按 9 : 1 的比例随机划分成训练集和验证集。对模型进行训练时，以均方误差为损失函数，采用 Adam 方法进行参数优化，1 个批次送入 32 个样本，训练不低于 100 轮，如果连续 10 轮训练，训练集损失函数值未下降，则降低学习率 20%，如果连续 20 轮训练，训练集损失

函数未下降，则停止训练。模型优化时会出现过度拟合，即训练集损失函数值下降，而验证集损失函数值不降反升。最终，本研究的模型训练终止于训练集损失函数值稳定在 0.01 以下，验证集损失函数值稳定在 0.03 以下停止，并将训练集损失函数最小值 0.003 的模型用于接下来的中药活性成分筛选。抑制率预测值平均误差在 ± 0.06 以内。抑制率预测值指 MPNN 模型预测的化合物浓度为 20 $\mu\text{mol/L}$ 时对 SARS-CoV-2 3CL 蛋白酶体外活性的抑制率，酶的活性被完全抑制时，取值为 1。

2.2 清热解毒中药活性成分 SARS-CoV-2 3CL 蛋白酶抑制剂筛选

SARS-CoV-2 3CL 蛋白酶抑制剂筛选是以中药为单位进行的。用优化后的 MPNN 模型预测矮地茶等 186 味清热解毒中药的活性成分在浓度为 20 $\mu\text{mol/L}$ 时对病毒 3CL 蛋白酶的抑制率。预测的数据样本有 8959 条，抑制率预测值大于 0.5 的有 347 条，占比约 4%。因为同一个活性成分可能广泛分布在不同的中药中，合并重复项后，预测样本数为 3863 条，抑制率预测值大于 0.5 的样本数为 186 条，占比约 5%。

在所有筛选的中药中，有 81 味中药至少含有 1 个预测抑制率大于 0.9 的活性成分。这 81 味中药分别为地锦草、猕猴桃根、木蝴蝶、关黄柏、半夏、半枝莲、大血藤、板蓝根、柴胡、桂枝、椿皮、灯盏辛夷、白果、白芷、白莲、白附子、矮地茶、苍耳子、葱白、薄荷、谷精草、赤芍、防风、鬼针草、了哥王根、决明子、昆布、木贼、满山红、牛蒡子、牡丹皮、猫爪草、猫眼草、积雪草、罗汉果、黄连、芥子、苦丁茶、荆芥、荷叶、菊花、虎耳草、连翘、迷蒙花、金沸草、金莲花、金银花、零陵香、马兜铃、马尾莲、马齿苋、麻黄、黄柏、黄芩、黄药子、黄连、龙胆、使君子、前胡、天葵子、山慈菇、山豆根、桑叶、桑白皮、沙棘、生姜、秦皮、青果、青蒿、余甘子、夏枯草、栀子、玄参、紫苏、紫苏子、紫草、肿节风、芫荽、辛夷、银杏叶、香薷。在所有筛选的中药活性成分中，来源于龙胆的龙胆素（TCMSP MolID: 3152）的活性最高，预测抑制率为 1.07，同时龙胆的另一个活性成分龙胆赛因（TCMSP MolID: 3170）的抑制率预测值为 1.01。需要说明的是，因为训练集数据抑制率有大于 1 和小于 0 的样本，所以抑制率预测值有可能大于 1 或小于 0。龙胆素和龙胆赛因的分子结构都含有 1 个

氧杂蒽酮母核（图 2），并且都属于多酚类化合物。其中龙胆素葡萄糖苷曾被用于治疗由 SARS 病毒感染引起的疾病^[15]。

在筛选的所有活性成分中，共 101 个化合物的抑制率预测值大于 0.9，它们绝大多数都是多酚类化合物，主要包括黄酮类、香豆素类、没食子儿茶素类、羟基肉桂酸酯类等。非酚类化合物抑制剂主要是含硫化合物、烷基酯和萘类等。

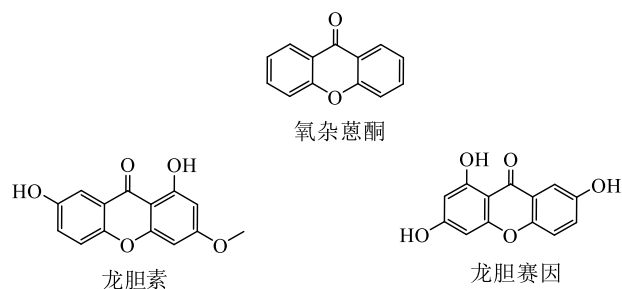


图 2 氧杂蒽酮、龙胆素 (TCMSP MolID: 3152)、龙胆赛因 (TCMSP MolID: 3170) 分子结构

Fig. 2 Molecular structures of oxanthrone, gentianin (TCMSP MolID: 3152) and gentiosin (TCMSP MolID: 3170)

2.2.1 黄酮类 SARS-CoV-2 3CL 蛋白酶抑制剂 黄酮类化合物是药用植物中广泛存在的一类活性成分，通常具有抗炎、抗病毒、抗肿瘤等活性。有很多黄酮化合物已经被研究证实为 SARS-CoV-2 3CL 蛋白酶抑制剂，如橙皮素、木犀草素、槲皮素、芹菜素、山柰酚、黄芩苷、杨梅素等^[16-17]。分子对接研究结果显示黄酮类化合物通过其吡喃环上的苯基或多羟基苯基与酶催化活性中心的 His41 之间的 π - π 堆积相互作用结合，以非共价方式占据酶的活性中心，从而抑制酶的活性^[18]。共筛选出来 19 个黄酮类化合物，分子结构如图 3 所示，抑制率预测值 0.91~1.01。黄酮类抑制剂的抑制率预测值、PubChem CID、化合物名称、实验或理论支撑数据等信息如表 1 所示。19 个黄酮化合物中，抑制率预测值最高的为芹菜素 (TCMSP MolID: 8)，该化合物在训练集中的抑制率为 0.99。而文献报道的分子对接结果显示，芹菜素与 3CL 蛋白酶的结合吉布斯自由能为 -7.3 kcal/mol (1 kcal=4.2 kJ)^[19]，在黄酮类化合物中相对较高。根据 MPNN 模型的预测结

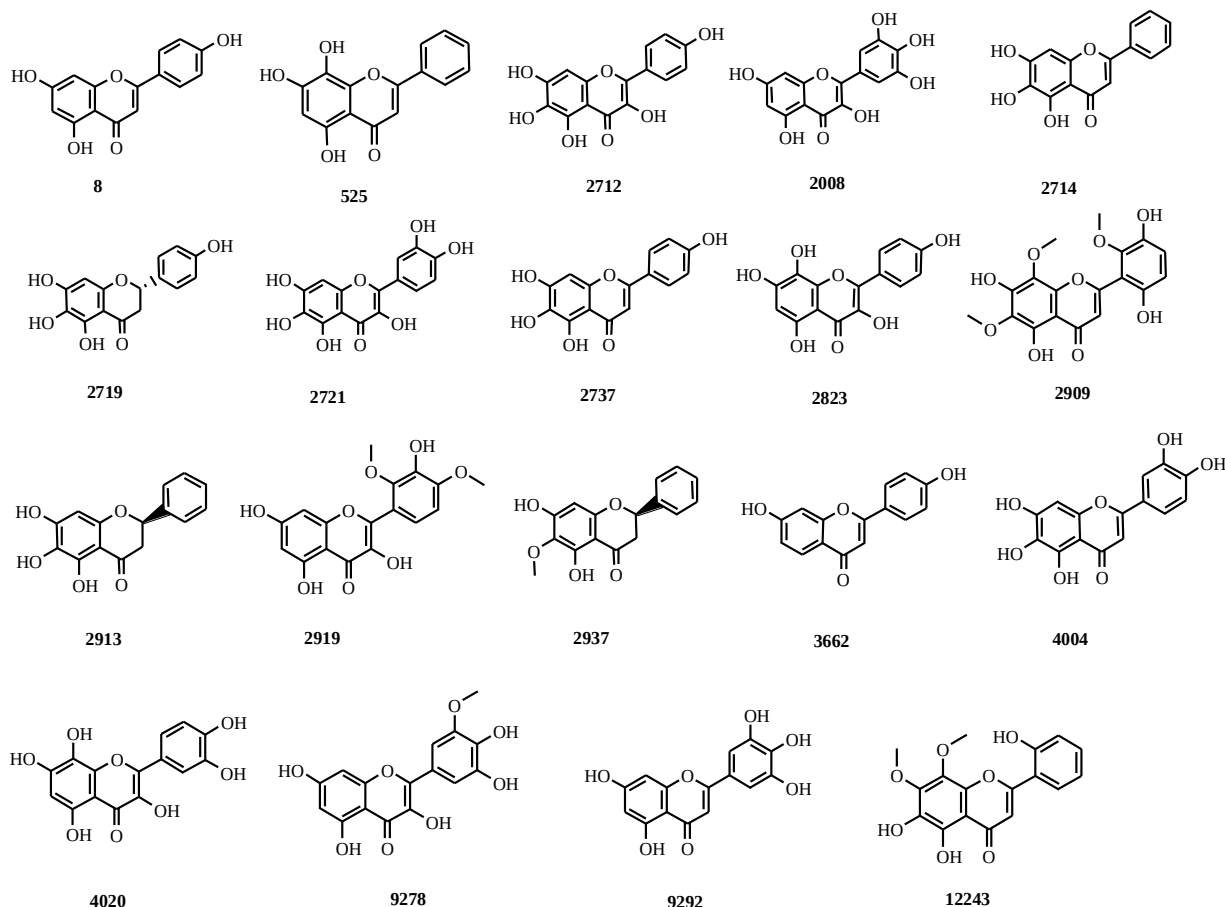


图 3 MPNN 模型筛选出的黄酮类 SARS-CoV-2 3CL 蛋白酶抑制剂分子结构和 TCMSP MolID

Fig. 3 Molecular structures and TCMSP MolID of flavonoids SARS-CoV-2 3CL protease inhibitors screened by MNPP model

表 1 MPNN 模型从中药活性成分中筛选出的黄酮类 SARS-CoV-2 3CL 蛋白酶抑制剂

Table 1 Flavonoids SARS-CoV-2 3CL protease inhibitors screened from TCMSP by MPNN model

TCMSP MolID	PubChem CID	预测抑制率	化学名称	中药来源	体外活性或分子对接结合能
8	5280443	1.01	芹菜素	木贼、积雪草	抑制率: 0.99 ^[12] -7.3 kcal·mol ⁻¹ ^[19]
525	5281674	0.91	5,7,8-三羟基黄酮	黄芩	NA
2008	5281672	0.96	杨梅素	矮地茶、满山红、苦丁茶、荷叶、连翘、 马齿苋、黄药子、沙棘	抑制率: 0.97 ^[12] IC ₅₀ : 2.86 μmol·L ⁻¹ ^[20]
2712	5281638	0.96	6-羟基山柰酚	灯盏细辛	NA
2714	5281605	0.94	黄芩黄素	木蝴蝶、关黄柏、半夏、黄芩、半枝莲、 灯盏细辛、赤芍	抑制率: 0.91 ^[12] IC ₅₀ : 0.39 μmol·L ⁻¹ ^[20]
2719	188308	0.92	6-羟基柚皮素	半枝莲	NA
2721	5281680	0.94	槲皮素	谷精草、金沸草	IC ₅₀ : 1.24 μmol·L ⁻¹ ^[20]
2737	5281697	0.96	野黄芩素	木蝴蝶、半枝莲、灯盏细辛、黄芩、紫苏	IC ₅₀ : 5.68 μmol·L ⁻¹ ^[20]
2823	5280544	0.96	草本黄素	麻黄	50 μmol·L ⁻¹ 时抑制率 0.59 ^[20]
2909	44258628	0.98	5,7,2,5-四羟基-8,6-二甲氧基 黄酮	黄芩	NA
2913	14135323	0.96	二氢黄芩素	黄芩	NA
2919	44259518	0.94	粘毛黄芩素 III	黄芩	NA
2937	25721350	0.91	二氢木蝴蝶素 A	黄芩、香薷	NA
3662	5282073	0.97	7,4'-二羟基黄酮	山豆根、梔子	NA
4004	5281642	0.92	6-羟基木犀草素	木蝴蝶	NA
4020	5280647	0.96	棉芸素	满山红	NA
9278	5282154	0.96	3'-O-甲基杨梅素	矮地茶、银杏叶	NA
9292	5281701	0.95	三叉青	银杏叶	NA
12243	15100720	0.94	5,6,2'-三羟基-7,8-二甲氧基黄酮	半枝莲	NA

NA: 无有效数据。

NA: not available.

果, 7 个黄酮类抑制剂分布于黄芩之中。黄芩是中医药最常用的清热解毒燥湿药, 在疫情期间, 黄芩是治疗效果最好的中药之一^[21]。此外, 木蝴蝶的活性成分中有 3 个黄酮化合物被筛选出来, 木蝴蝶按中医理论归肺经, 有清肺利咽的功效, 常用于治疗肺热咳嗽等症, 符合 COVID-19 的症状^[22]。

2.2.2 香豆素类 SARS-CoV-2 3CL 蛋白酶抑制剂
香豆素类化合物含有 1 个 2-*H*-1-苯并吡喃-2-酮基本母核, 如图 4 所示。天然的和半合成的香豆素类化合物通常具有较小的毒性和副作用, 具有抗氧化、抗肿瘤、抗炎、抗病毒等活性。在抑制 RNA 病毒蛋白酶方面, 香豆素类化合物也得到广泛应用, 其中就包括 SARS-CoV-2。文献报道的分子对接研究结

果表明, 香豆素类化合物与 SARS-CoV-2 3CL 蛋白酶的亲和性要强于抗 SARS-CoV-2 药物洛匹那韦和利托那韦, 香豆素类化合物主要与 3CL 蛋白酶催化活性中心的 His41 和 Cys145 结合, 从而抑制酶的活性^[23]。本研究预测出 8 个香豆素类抑制剂, 主要来自于柴胡、桑叶、防风、秦皮等中药。虽然在数量上不及黄酮类化合物多, 但是香豆素化合物的平均抑制率预测值(表 2)要高于黄酮类化合物。其中抑制率预测值最大的香豆素化合物是七叶皂苷(TCMSP MolID: 4456), 结构如图 4 所示, 其与 SARS-CoV-2 3CL 蛋白酶的结合能为 -7.1 kcal/mol^[24]。来自于地锦草等中药的伞形花内酯(skimmetin, TCMSP MolID: 2558)是训练集中出

现过的化合物, 预测的抑制率为 0.98, 训练集中的抑制率为 0.92。

2.2.3 没食子儿茶素类 SARS-CoV-2 3CL 蛋白酶抑制剂 没食子儿茶素是一种多酚儿茶素类化合物,

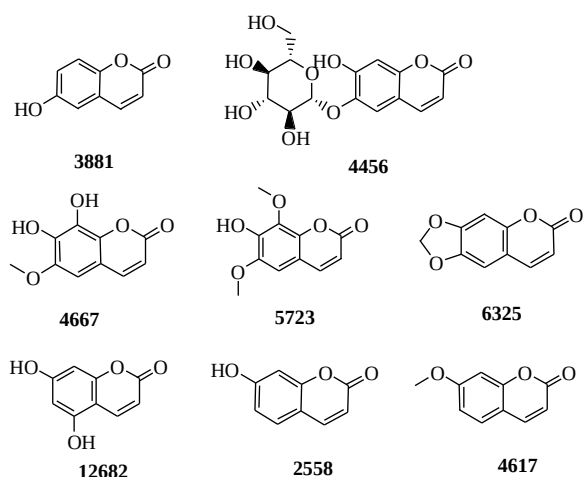


图 4 MPNN 模型筛选出的香豆素类 SARS-CoV-2 3CL 蛋白酶抑制剂分子结构和 TCMSP MolID

Fig. 4 Molecular structures and TCMSP MolID of coumarins SARS-CoV-2 3CL protease inhibitors screened by MPNN model

它们广泛存在于茶植物中, 尤其是绿茶中含量较高。这类化合物表现出广谱抗病毒活性, 包括抗寨卡病毒、疱疹病毒、肝炎病毒、流感病毒等, 同时也具有抗 SARS 和 SARS-CoV-2 的活性^[25]。体外实验数据证实, 表没食子儿茶素-3-没食子酸酯 (TCMSP MolID: 6821) 对 SARS-CoV-2 3CL 蛋白酶的抑制活性要高于木犀草素、槲皮素、山柰酚等黄酮类化合物^[26]。此外, 体外实验研究显示没食子儿茶素-3-没食子酸酯 (TCMSP MolID: 5830) 的抑制活性高于其差向异构体表没食子儿茶素-3-没食子酸酯 (TCMSP MolID: 6821)^[27]。本研究共筛选出 7 个没食子儿茶素类化合物, 预测活性等数据见表 3。但是亮可酚素 (TCMSP MolID: 7217) 的结构严格地讲与其他 6 个化合物不同 (图 5), 前者在苯并吡喃母核的杂环上多了 1 个羟基。但是, 分子对接研究表明, 亮可酚素与 SARS-CoV-2 3CL 蛋白酶有很强的亲和力, 并且可以通过改变酶的结构削弱其催化活性^[28]。

2.2.4 羟基肉桂酸酯类 SARS-CoV-2 3CL 蛋白酶抑制剂 筛选出的中药活性成分中, 羟基肉桂酸酯类

表 2 MPNN 模型从中药活性成分中筛选出的香豆素类 SARS-Cov-2 3CL 蛋白酶抑制剂

Table 2 Coumarins SARS-CoV-2 3CL protease inhibitors screened from TCMSP by MPNN model

TCMSP MolID	PubChem CID	预测抑制率	化学名称	中药来源
2558	5281426	0.98	伞形花内酯	地锦草、猕猴桃根、白芷、了哥王根、满山红、前胡、桑叶、桑白皮、夏枯草、芫荽
3881	99477	1.00	6-羟基香豆素	牡丹皮、桑叶
4456	5281417	1.01	七叶皂苷	灯盏细辛、秦皮、芫荽
4617	10748	0.90	7-甲氧基香豆素	柴胡
4667	5273569	0.98	7,8-二羟基-6-甲氧基香豆素	猕猴桃根、柴胡、防风、秦皮
5723	5318565	0.98	异嗪辛	防风、猫眼草、青蒿
6325	3083597	0.98	泽兰内酯	地锦草
12682	5324654	0.96	5,7-二羟基香豆素	桑叶

表 3 MPNN 模型从中药活性成分中筛选出的没食子儿茶素类 SARS-CoV-2 3CL 蛋白酶抑制剂

Table 3 Gallocatechins SARS-CoV-2 3CL protease inhibitors screened from TCMSP by MPNN model

TCMSP MolID	PubChem CID	预测抑制率	化学名称	中药来源
2249	65084	0.99	没食子儿茶素	余甘子、银杏叶
5830	5276890	0.96	没食子儿茶素没食子酸酯	白莲、苦丁茶
5866	9882981	0.99	(2S,3R)-2-(3,4,5-三羟基苯基)苯并吡喃-3,5,7-三醇	白莲
6821	65064	0.95	表没食子儿茶素-3-没食子酸酯	白果、余甘子
6791	72277	1.00	表没食子儿茶素	余甘子
7217	44563331	0.99	亮可酚素	荷叶
10958	199472	0.95	没食子儿茶酚没食子酸酯	矮地茶

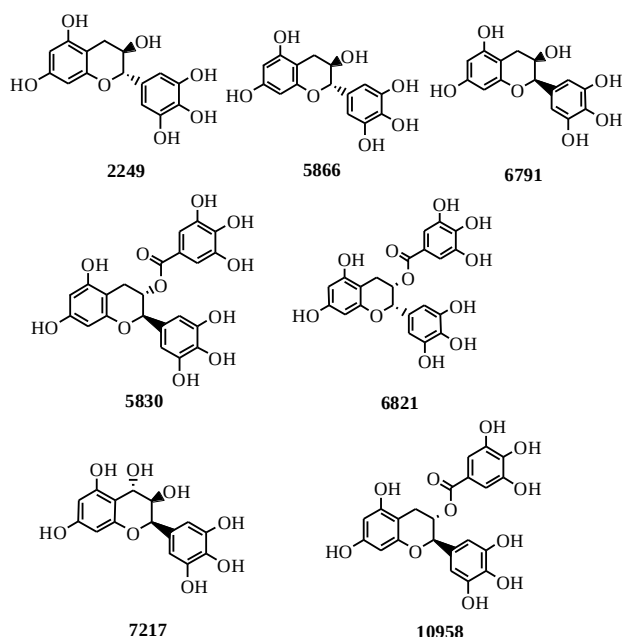


图 5 MPNN 模型筛选出的没食子儿茶素类 SARS-CoV-2 3CL 蛋白酶抑制剂分子结构和 TCMSP MolID

Fig. 5 Molecular structures and TCMSP MolID of gallic acid derivatives SARS-CoV-2 3CL protease inhibitors screened by MPNN model

化合物数量较多，主要是咖啡酸酯，如果将咖啡酸也计算在内的话，总共筛选出 19 个此类化合物（表 4）。此类抑制剂中有 2 个化合物出现在训练集中。其中一个为草芥素（TCMSP MolID: 3333，抑制率为 0.98，抑制率预测值为 0.99），另一个是迷迭香酸（TCMSP MolID: 11865，抑制率为 0.97，抑制率预测值为 0.98）。此类抑制剂中咖啡酰奎宁酸类数量最多，分子结构见图 6。奎宁酸分子结构中含有 4 个羟基可以与咖啡酸成酯。另外，天然的咖啡酰奎宁酸有单酯、二酯，因此可以得到数量众多结构相似的化合物。分子对接结果显示，咖啡酰奎宁酸单酯和二酯与 SARS-CoV-2 3CL 蛋白酶的结合吉布斯自由能相近， $-7 \sim -6$ kcal/mol，其中 5-*O*-咖啡酰奎宁酸、3,4-二氧-咖啡酰奎宁酸和 4,5-二氧-咖啡酰奎宁酸与 3CL 蛋白酶的亲和力更强，可以作为抗 SARS-CoV-2 的先导化合物^[29]。另一项分子对接的研究显示，咖啡酸本身对 SARS-CoV-2 3CL 蛋白酶也有抑制作用，只是与 3CL 蛋白酶的亲和力较咖啡酰奎宁酸化合物低一些^[30]。羟基肉桂酸酯类化合物除了通过非共价键的形式与 3CL 蛋白酶结合外，还有学者指出，它们可以通过共价键形式与 SARS-CoV-2

表 4 MPNN 模型从中草药活性成分中筛选出的羟基肉桂酸酯类 SARS-CoV-2 3CL 蛋白酶抑制剂

Table 4 Hydroxycinnamates SARS-CoV-2 3CL protease inhibitors screened from TCMSP by MPNN

TCMSP MolID	PubChem CID	预测抑制率	化学名称	中药来源
222	7067335	0.94	新绿原酸	虎耳草
223	689043	0.94	咖啡酸	芫荽
1878	6474309	0.91	4,5-二氧咖啡酰奎宁酸	苍耳子、梔子
1955	1794427	0.93	咖啡酰奎宁酸	关黄柏、灯盏细辛、牛蒡子、虎耳草、桑叶
2902	5317238	0.97	咖啡酸乙酯	灯盏细辛、黄柏、桑叶、夏枯草
3048	71720841	0.90	绿原酸甲酯	金银花
3333	5281800	0.99	草芥素	玄参
4087	6474640	0.90	1,5-二氧-咖啡酰奎宁酸	灯盏细辛、菊花
6191	5316819	0.94	[(<i>Z</i>)-2-(3,4-二羟基苯基)乙烯基]3-(3,4-二羟基苯基)丙-2-烯酸酯	紫苏
6218	689075	0.95	咖啡酸甲酯	椿皮、黄柏、紫苏
4087	6474640	0.90	1,5-二氧-咖啡酰奎宁酸	灯盏细辛、菊花
6370	5280633	0.95	新绿原酸	关黄柏
6378	12310830	0.95	绿原酸	关黄柏
6501	5281780	0.90	异绿原酸 B	苍耳子
6682	5273566	0.94	菖蒲苷 A	迷蒙花
7132	9841799	0.97	(2 <i>R</i>)-3-(3,4-二羟基苯基)-2-[(<i>Z</i>)-3-(3,4-二羟基苯基)丙烯酰基]氧 肿节风基丙酸	
7305	10324242	0.92	(1 <i>S</i> ,3 <i>S</i> ,4 <i>R</i> ,5 <i>S</i>)-3,4-双[(<i>E</i>)-3-(3,4-二羟基苯基)丙烯酰基]氧基]-1,5-二羟基环己烷-1-羧酸	苍耳子
7326	5281769	0.91	西那林(c)	苍耳子、牛蒡子、菊花
11865	5281792	0.98	迷迭香酸	薄荷、荆芥、麻黄、紫苏

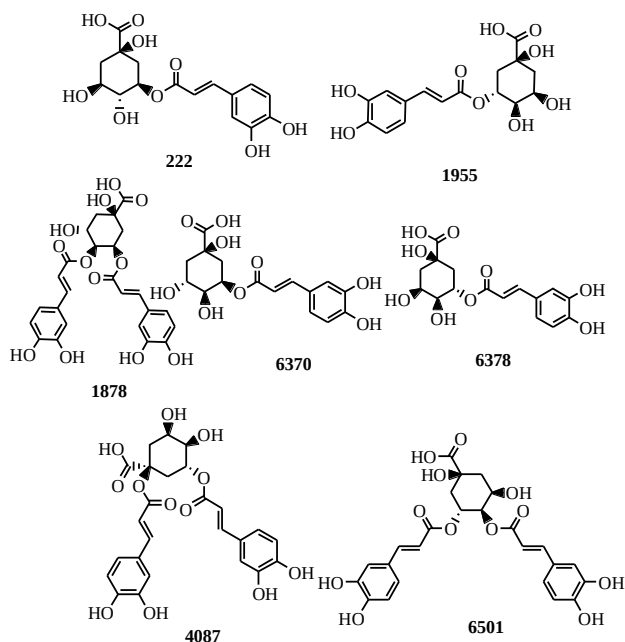


图 6 MPNN 模型筛选出的咖啡酰奎宁酸类 SARS-CoV-2 3CL 蛋白酶抑制剂分子结构和 TCMSP MolID

Fig. 6 Molecular structures and TCMSP MolID of caffeoylquinic acids SARS-CoV-2 3CL protease inhibitors screened by MPNN model

3CL 蛋白酶催化中心的 Cys145 结合, 从而非选择性地抑制酶的活性^[31]。因为羟基肉桂酸结构中含有 α -不饱和羰基, 可以与半胱氨酸的巯基发生迈克尔加成反应, 许多抗炎药物的生物活性也源于此反应^[32]。因此, 羟基肉桂酸酯类化合物的抗 SARS-CoV-2 的活性可能由多重途径实现^[17]。

2.2.5 其他酚类 SARS-CoV-2 3CL 蛋白酶抑制剂
此类抑制剂共筛选出 37 个化合物, 具体数据见表 5。其中预测抑制率值最大的是来源于桑白皮和桑叶的桑辛素 C (TCMSP MolID: 3857, 预测抑制率 1.07), 分子结构如图 7 所示。大血藤的活性成分 3,4-二羟基苯基- β -D-豆腐果苷 (TCMSP MolID: 7924, 预测抑制率 1.04) 预测活性仅次于桑辛素 C。值得注意的是, 从昆布中筛选出的鹅掌菜酚 (TCMSP MolID: 10616) 及其 4 个结构相似的化合物 (图 7), 均具有较高的预测抑制率。鹅掌菜酚和间苯磺酰呋喃酚 A (TCMSP MolID: 10620) 曾被实验证实是 SARS 3CL 蛋白酶的抑制剂, IC_{50} 分别为 8.8 和 16.7 $\mu\text{mol/L}$ ^[33]。此外, 此类抑制剂中有 2 个化合物含有羟基肉桂酰的结构片段, 但不是酯类化合物, 而是酮类化合物, 它们分别是来源于鬼针草的奥卡宁 (TCMSP MolID: 6436) 和来源于桑叶

的补骨脂乙素 (TCMSP MolID: 5009)。体外实验证实补骨脂乙素对 SARS 3CL 蛋白酶有强的抑制作用^[34]。考虑到 SARS-CoV-2 和 SARS 病毒的 3CL 蛋白酶结构相近, 并且很多抗 SARS 病毒的药物也被证实具有抗 SARS-CoV-2 的作用, 所以补骨脂乙素及鹅掌菜酚极可能对 SARS-CoV-2 3CL 蛋白酶也有抑制作用。此类抑制剂中还有 2 个来源于马尾莲的生物碱, (*R*)-番荔枝碱 (TCMSP MolID: 8797) 和 *N*-甲基网状番荔枝碱 (TCMSP MolID: 8798)。分子对接研究证实 (*R*)-番荔枝碱可以与 SARS-CoV-2 3CL 蛋白酶的催化活性中心结合, 结合吉布斯自由能为 -4.32 kcal/mol^[35]。另外, 尤克萘酮 (TCMSP MolID=3358) 与龙胆素和龙胆赛因具有相同的氧杂萘酮母核。尤克萘酮具有抗 RNA 病毒的活性^[36], 虽然目前还没有关于尤克萘酮抑制 SARS-CoV-2 的 3CL 蛋白酶的报道, 考虑到 SARS-CoV-2 是 RNA 病毒的一种, 尤克萘酮也是一种潜在的抑制剂。

本研究筛选出的 101 个中药活性成分 (预测抑制率大于 0.9) 中, 含有酚羟基的有 90 个, 占比为 90%。研究发现来源于天然产物的 SARS-CoV-2 3CL 蛋白酶抑制剂, 其中绝大多数的分子结构中都含有酚羟基^[17,19,34], 从一定程度上证明了本研究优化的 MPNN 模型是可靠的。

2.3 非酚类 SARS-CoV-2 3CL 蛋白酶抑制剂

筛选出的非酚类抑制剂见表 6, 部分抑制剂的分子结构见图 8。非酚类抑制剂中含硫化合物数量最多, 包括三硫醚、二硫醚、硫醚等。临床研究显示, 有部分 COVID-19 患者是使用精油治愈的, 其中大蒜精油中所含的主要活性成分就是硫醚^[37]。本研究筛选出的硫醚化合物共有 4 个, 分别为来源于葱白和罗汉果的甲基三硫醚 (TCMSP MolID: 7628), 来源于葱白的丙基三硫醚 (TCMSP MolID: 7630)、甲基二硫代-1-丙烯 (TCMSP MolID: 7644), 还有来源于芥子的 3-甲硫代丙基异硫氰酸 (TCMSP MolID: 13025), 它们都有较高的预测抑制率。

除硫醚外, 还有 2 个含二硫代酰基的化合物, 一个是来源于白果的甲氧基芸苔素 (TCMSP MolID: 11069), 另一个是来源于紫苏子的 *S*-丙氧硫酰基-巯基胺 (TCMSP MolID: 12897), 两者的结构见图 8, 它们对 SARS-CoV-2 3CL 蛋白酶的预测抑制率分别为 0.91 和 0.98。此外, 含硫抑制剂还包括麻黄中的 1 个巯基四氮唑化合物 1-苯基-5-巯基四

表 5 MPNN 模型从中药活性成分中筛选出的其他酚类 SARS-CoV-2 3CL 蛋白酶抑制剂
Table 5 Other phenolics SARS-CoV-2 3CL protease inhibitors screened from TCMSP by MPNN model

TCMSP MolID	PubChem CID	预测 抑制率	化学名称	中药来源
106	1057	0.97	邻苯三酚	虎耳草、青果、芫荽
141	82755	0.90	羟基酯醇	黄连
579	785	0.96	对苯二酚	胡黄连、虎耳草
1452	8768	0.98	原儿茶醛	桂枝
1812	45082883	0.93	2,5-二羟基吡啶	板蓝根
2092	7311	0.92	2,4-二叔丁基苯酚	木贼
2922	10034991	0.95	5-(2-羟乙基)-2-甲氧基苯酚	黄芩、天葵子、玄参
3374	5281862	0.94	漆酚 III	连翘
3358	5281631	0.95	尤克萘酮	连翘、生姜
3503	287064	0.98	原儿茶酸甲酯	金莲花、黄连、紫苏
3857	155248	1.07	桑辛素 C	桑叶、桑白皮
3865	194774	0.93	4-甲氧基菲-2,5-二醇	桑叶
4666	77547	0.91	原儿茶酸乙酯	柴胡
5009	5281255	0.97	异补骨脂乙素	桑叶
6179	10315050	0.94	六羟基连苯二甲酸	赤芍、使君子
6436	5281294	1.00	奥卡宁	鬼针草
6438	6123451	0.99	(2E)-2-[(3,4-二羟基苯基)亚甲基]-6,7-二羟基-1-苯并呋喃-3-酮	鬼针草
6826	71308174	0.96	螯布菜酸	余甘子
7060	6451084	0.91	石精酸 B	紫草
7134	11600642	0.99	丹参素	黄连
7272	16928	0.96	高香草醇	金莲花
7373	12820	0.96	1-(2,3-二羟基-4-甲氧基苯基)乙酮	牡丹皮
7924	6915839	1.04	3,4-二羟基苯基-β-D-豆腐果苷	大血藤
7992	14777892	0.99	4-甲氧基菲-2,7-二醇	山慈菇
8797	440586	0.94	(R)-番荔枝碱	马尾莲
8798	167718	0.90	N-甲基网状番荔枝碱	马尾莲
5982	5281343	0.93	5,7-二羟基色酮	猕猴桃根、决明子、桑白皮
10611	137388	0.93	6,6'-双鹅掌菜酚	昆布
10612	3008867	0.94	8,8'-双鹅掌菜酚	昆布
10616	145937	0.96	鹅掌菜酚	昆布
10620	130976	0.95	间苯磺酰呋喃酚 A	昆布
10626	16132364	0.93	2-O-(2,4,6-三羟基苯基)-6,6'-双鹅掌菜酚	昆布
11349	19107661	0.97	2-(2,3,4-三羟基苯基)乙腈	天葵子
12716	641363	0.92	5-(苯并呋喃-2-基)间苯二酚	桑白皮
12739	9881314	0.93	4-(2,2-二羟基苯基)乙烯基]间苯二酚	桑白皮
12789	11874	0.97	邻苯三酚羧酸	桑白皮

氮唑（TCMSP MolID: 10 750）。该抑制剂的预测抑制率是 0.96，在训练集中已被证实具有较强的抑制活性，训练集中的抑制率为 0.89。此类抑制剂中还有 1 个精油类成分，来源于香薷等中药的黄樟素

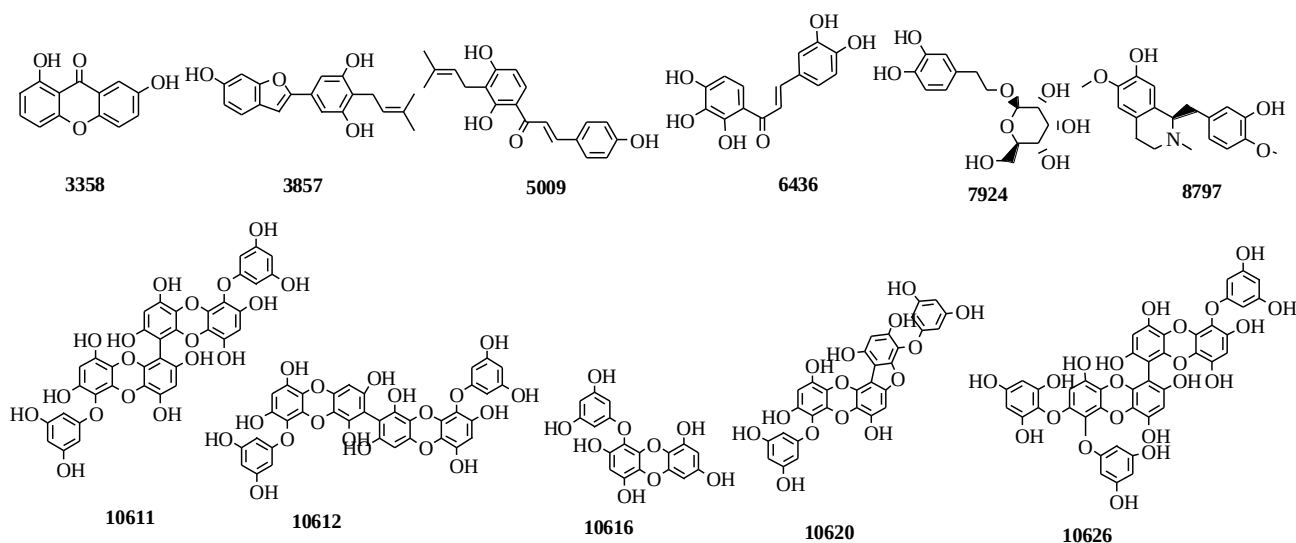


图 7 MPNN 模型筛选出的部分其他多酚类 SARS-CoV-2 3CL 蛋白酶抑制剂分子结构和 TCMSP MolID

Fig. 7 Molecular structures and TCMSP MolID of some other polyphenols SARS-CoV-2 3CL protease inhibitors screened by MPNN model

表 6 MPNN 模型从中药活性成分中筛选出的非酚类 SARS-CoV-2 3CL 蛋白酶抑制剂

Table 6 Non-phenols SARS-CoV-2 3CL protease inhibitors screened from TCMSP by MPNN model

TCMSP MolID	PubChem CID	预测活性	化学名称	中药来源
199	5144	0.94	黄樟素	马兜铃、辛夷、香薷
5483	7055	0.91	2-甲基萘	白附子、猫爪草、零陵香、芫荽
7628	19310	1.03	甲基三硫醚	葱白、罗汉果
7630	22383	0.97	丙基三硫醚	葱白
7644	5366552	1.01	甲基二硫代-1-丙烯	葱白
10750	690730	0.96	1-苯基-5-巯基四氮唑	麻黄
11069	3037463	0.91	甲氧基芸苔素	白果
11324	10464455	1.05	5-羟基-4-氧代戊酸甲酯	猫爪草
12885	79378	0.98	氯乙酸丙酯	紫苏子
12897	536887	0.98	S-丙氧硫酰基-巯基胺	紫苏子
13025	62351	0.98	3-甲硫代丙基异硫氰酸	芥子

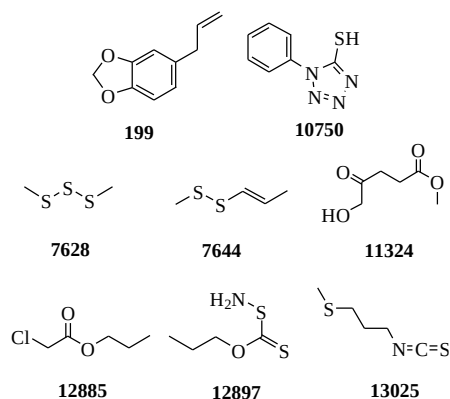


图 8 MPNN 模型筛选出的部分非酚类 SARS-CoV-2 3CL 蛋白酶抑制剂分子结构和 TCMSP MolID

Fig. 8 Molecular structures and TCMSP MolID of some non-phenols 3CL protease inhibitors screened by MPNN model

(TCMSP MolID: 199), 分子对接的结果显示黄樟素对 SARS-CoV-2 3CL 蛋白酶具有一定的亲和力, 结合吉布斯自由能为 -4.9 kcal/mol ^[37]。此外, 还有几个结构简单的化合物如 2-甲基萘 (TCMSP MolID: 5483)、氯乙酸丙酯 (TCMSP MolID: 12885)、5-羟基-4-氧代戊酸甲酯 (TCMSP MolID: 11324)。其中 5-羟基-4-氧代戊酸甲酯是非酚类抑制剂中预测抑制率最高的 1 个, 预测抑制率为 1.05, 可以作为抗 SARS-CoV-2 先导化合物做进一步的实验和理论研究。值得注意的是, 最近有学者利用氯乙酰基结构片段与 SARS-CoV-2 3CL 蛋白酶催化中心的特异性结合特点, 设计了针对病毒蛋白酶的化学探针, 用于 3CL 蛋白酶抑制剂的快速筛选^[38]。因此, 氯乙酸丙酯同样值得进行深入研究。

3 结论

本研究以 PubChem 网站上公布的生物检测数据为训练集,优化了 1 个 MPNN 神经网络用于预测化合物在 20 $\mu\text{mol/L}$ 浓度下对 SARS-CoV-2 3CL 蛋白酶的抑制率。利用优化后的模型从 186 味清热解暑中药活性成分中 (3 863 个) 筛选潜在的 SARS-CoV-2 3CL 蛋白酶抑制剂,约 5% 的活性成分预测抑制率大于 0.5,其中预测抑制率大于 0.9 的有 101 个。这 101 个化合物来自于 81 味中药。在经由 MPNN 模型虚拟筛选出的 101 个化合物中,龙胆素、桑辛素 C、5-羟基-4-氧代戊酸甲酯、3,4-二羟基苯基- β -D-豆腐果苷等化合物显示出较高的预测活性,可以作为先导化合物进一步进行理论和实验研究。在经由 MPNN 模型虚拟筛选的中药中,黄芩所含预测抑制率大于 0.9 的活性成分数量最多,同时苍耳子、昆布、芫荽、紫苏等中药也含有较多活性成分。在后续的抗 SARS-CoV-2 中药方剂的开发中,值得关注这些中药。

MPNN 模型虚拟筛选的抑制剂中,约有 1/5 的抑制剂已经有体外实验数据或分子对接数据,证明其对 SARS-CoV-2 3CL 蛋白酶具有抑制作用。此外,还有一些抑制剂虽然尚未见到相关研究报道,但已被证实具有抗 SARS 或其他 RNA 病毒的活。经由 MPNN 模型虚拟筛选出的抑制剂中有很一部分的分子结构与已证实的 SARS-CoV-2 3CL 蛋白酶抑制剂相似,它们很可能也是潜在的 SARS-CoV-2 3CL 蛋白酶抑制剂。因此,MPNN 模型虚拟筛选结果具有较高的可信度。值得注意的是,本研究使用的 MPNN 模型嵌入 transformer 的编码器层,用于提取分子特征,优化后的 MPNN 模型可以通过微调用于预测分子的其他属性。因此,MPNN 模型可作为一种范例在药物虚拟筛选领域进行推广。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Weekly epidemiological update on COVID-19 - 13 July 2023 [EB/OL]. (2023-07-13) [2023-08-10]. <https://www.who.int/publications>.
- [2] Li Q, Guan X H, Wu P, *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia [J]. *N Engl J Med*, 2020, 382(13): 1199-1207.
- [3] Wu F, Zhao S, Yu B, *et al.* A new coronavirus associated with human respiratory disease in China [J]. *Nature*, 2020, 579(7798): 265-269.
- [4] Ziebuhr J. Molecular biology of severe acute respiratory syndrome coronavirus [J]. *Curr Opin Microbiol*, 2004, 7(4): 412-419.
- [5] Anand K, Ziebuhr J, Wadhwani P, *et al.* Coronavirus main proteinase (3CL^{pro}) structure: Basis for design of anti-SARS drugs [J]. *Science*, 2003, 300(5626): 1763-1767.
- [6] Jin Z M, Du X Y, Xu Y C, *et al.* Structure of M^{pro} from SARS-CoV-2 and discovery of its inhibitors [J]. *Nature*, 2020, 582(7811): 289-293.
- [7] Zhang L L, Lin D Z, Sun X, *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors [J]. *Science*, 2020, 368(6489): 409-412.
- [8] 郭琴, 王颖, 王欢欢, 等. 基于方证对应探讨中药复方分期治疗新型冠状病毒肺炎的网络药理学机制 [J]. *中草药*, 2021, 52: 1393-1401.
- [9] 赵静, 田赛赛, 杨健, 等. 清肺排毒汤治疗新型冠状病毒肺炎机制的网络药理学探讨 [J]. *中草药*, 2020, 51(4): 829-835.
- [10] 叶飞, 黄保英, 赵莉, 等. 连花清瘟颗粒等 8 种中药体外抗新型冠状病毒活性以及细胞毒性研究 [J]. *病毒学报*, 2022, 38(2): 284-289.
- [11] 马铃, 赵建元, 郭赛赛, 等. 细胞水平新型冠状病毒 SARS-CoV-2 3CL 蛋白酶抑制剂筛选模型的建立 [J]. *药科学报*, 2020, 55(9): 2122-2126.
- [12] SARS-CoV-2 3CL-Pro protease inhibition percentage at 20 μM by FRET kind of response from peptide substrate [EB/OL]. ChEMBL, (2021-03-02) [2023-07-11]. <https://pubchem.ncbi.nlm.nih.gov/bioassay/1640020>.
- [13] 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.
- [14] Gilmer J, Schoenholz S, Riley P F, *et al.* Neural message passing for quantum chemistry [J]. *ICML*, 2017, doi: <https://doi.org/10.48550/arXiv.1704.01212>.
- [15] Khan T, Khan M A, Mashwani Z U R, *et al.* Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites [J]. *Biocatal Agric Biotechnol*, 2021, 31: 101890.
- [16] Guijarro-Real C, Plazas M, Rodríguez-Burruezo A, *et al.* Potential *in vitro* inhibition of selected plant extracts against SARS-CoV-2 chymotrypsin-like protease (3CL^{Pro}) activity [J]. *Foods*, 2021, 10(7): 1503.
- [17] Guerra Y, Celi D A, Cueva P, *et al.* Critical review of plant-derived compounds as possible inhibitors of SARS-CoV-2 proteases: A comparison with experimentally validated molecules [J]. *ACS Omega*, 2022, 7(49): 44542-44555.
- [18] Mandal A, Jha A K, Hazra B. Plant products as inhibitors of coronavirus 3CL protease [J]. *Front Pharmacol*, 2021,

- 12: 583387.
- [19] Liu H B, Ye F, Sun Q, *et al.* *Scutellaria baicalensis* extract and baicalein inhibit replication of SARS-CoV-2 and its 3C-like protease *in vitro* [J]. *J Enzyme Inhib Med Chem*, 2021, 36(1): 497-503.
- [20] 汤少梁, 贡悦, 冯雨莉, 等. 《伤寒论》治疫方用药规律分析及核心药对治疗 COVID-19 的作用机制研究 [J]. *中草药*, 2023, 54: 192-209.
- [21] 高彦宇, 刘思宇, 沈芳玲, 等. 木蝴蝶药理作用及临床应用研究进展 [J]. *中医药学报*, 2021, 49(5): 102-106.
- [22] Abdizadeh R, Hadizadeh F, Abdizadeh T. In silico analysis and identification of antiviral coumarin derivatives against 3-chymotrypsin-like main protease of the novel coronavirus SARS-CoV-2 [J]. *Mol Divers*, 2022, 26(2): 1053-1076.
- [23] Hasan A, Al Mahamud R, Jannat K, *et al.* Phytochemicals from *Solanum surattense* Burm.f. have high binding affinities for C-3 like main protease of COVID-19 (SARS-CoV-2) [J]. *J Med Plants Stud*, 2020, 8: 20-26.
- [24] Shaik F B, Swarnalatha K, Mohan M C, *et al.* Novel antiviral effects of chloroquine, hydroxychloroquine, and green tea catechins against SARS CoV-2 main protease (Mpro) and 3C-like protease for COVID-19 treatment [J]. *Clin Nutr Open Sci*, 2022, 42: 62-72.
- [25] Bahun M H, Jukić M, Oblak D, *et al.* Inhibition of the SARS-CoV-2 3CL^{pro} main protease by plant polyphenols [J]. *Food Chem*, 2022, 373(Pt B): 131594.
- [26] Nguyen T T H, Woo H J, Kang H K, *et al.* Flavonoid-mediated inhibition of SARS coronavirus 3C-like protease expressed in *Pichia pastoris* [J]. *Biotechnol Lett*, 2012, 34(5): 831-838.
- [27] Singh A, Mishra A. Leucoefdin a potential inhibitor against SARS CoV-2 M^{pro} [J]. *J Biomol Struct Dyn*, 2021, 39(12): 4427-4432.
- [28] Gonçalves C P, Marcucci M C, Rocha O V, *et al.* Computational studies of caffeoylquinic acids from Brazilian green propolis and its anti-viral potential against SARS-CoV-2 Mpro [J]. *Res Soc Dev*, 2022, 11(15): e359111536917.
- [29] Du A S, Zheng R, Disoma C, *et al.* Epigallocatechin-3-gallate, an active ingredient of Traditional Chinese Medicines, inhibits the 3CLpro activity of SARS-CoV-2 [J]. *Int J Biol Macromol*, 2021, 176: 1-12.
- [30] Zhu D F, Su H X, Ke C Q, *et al.* Efficient discovery of potential inhibitors for SARS-CoV-2 3C-like protease from herbal extracts using a native MS-based affinity-selection method [J]. *J Pharm Biomed Anal*, 2022, 209: 114538.
- [31] Zhang Z Q, Xu L C, Cheung H Y. The inhibitory effect of helenalin on telomerase activity is attributed to the alkylation of the CYS445 residue: Evidence from QM/MM simulations [J]. *J Mol Graph Model*, 2014, 51: 97-103.
- [32] Park J Y, Kim J H, Kwon J M, *et al.* Dieckol, a SARS-CoV 3CL(pro) inhibitor, isolated from the edible brown algae *Ecklonia cava* [J]. *Bioorg Med Chem*, 2013, 21(13): 3730-3737.
- [33] Solnier J, Fladerer J P. Flavonoids: A complementary approach to conventional therapy of COVID-19? [J]. *Phytochem Rev*, 2021, 20(4): 773-795.
- [34] Syahputra G, Gustini N, Bustanussalam B, *et al.* Molecular docking of secondary metabolites from Indonesian marine and terrestrial organisms targeting SARS-CoV-2 ACE-2, M pro, and PL pro receptors [J]. *Pharmacia*, 2021, 68(3): 533-560.
- [35] Sulaiman M, Zarubaev V, Samudi C, *et al.* Natural products from medicinal plants in Asia and the Pacific for RNA viruses: Hercules' fifth labour [J]. *Pharm Biol*, 2022, 60(1): 2049-2087.
- [36] Elsebai M F, Albalawi M A. Essential oils and COVID-19 [J]. *Molecules*, 2022, 27(22): 7893.
- [37] Paul A, Bondhon T A, Afrose S, *et al.* In silico binding studies with compounds present in essential oil of *Tasmannia lanceolata* leaves to 3C-like protease of SARS-CoV-2 [J]. *J Med Plants Stud*, 2020, 8(6): 108-116.
- [38] Peñalver L, Schmid P, Szamosvári D, *et al.* A Ligand Selection Strategy Identifies Chemical Probes Targeting the Proteases of SARS-CoV-2 [J]. *Angew Chem Int Ed*, 2021, 60: 6799-6806.

[责任编辑 潘明佳]