

基于 BiRNN 模型的 DPP-4 和 SGLT-2 双靶点抑制剂设计

孙定康^{1,2}, 张鑫磊², 郭凯蕾¹, 李琦³, 梁佳龙⁴, 马丽莎², 林佳艳², 何金穗¹, 刁春妍², 王育伟^{1*}, 刘雪英^{2*}

1. 陕西中医药大学 药学院, 陕西 咸阳 712046

2. 空军军医大学 药学院 药物化学与药物分析教研室, 陕西 西安 710032

3. 河南大学 生命科学学院, 河南 开封 475004

4. 中国人民解放军陆军第九四六医院 医疗保障中心, 新疆 伊宁 835000

摘要: **目的** 设计 DPP-4 和 SGLT-2 双靶点抑制剂。 **方法** 根据现有的 DPP-4 和 SGLT-2 活性化合物训练 BiRNN 模型, 使其生成具有潜在活性的双靶点新分子, 并使用已经报道的活性化合物构建 HipHop 药效团模型, 通过药效团模型对新分子进行初筛, 然后使用分子对接、分子动力学模拟以及结合自由能计算等方法对初筛小分子进行进一步筛选最终得到候选化合物。 **结果** BiRNN 模型生成 7 494 个新分子, 经过虚拟筛选发现化合物 NM186、NM21、NM249、NM107 是相对理想的 DPP-4 和 SGLT-2 双靶点抑制剂。 **结论** NM186、NM21、NM249、NM107 可能是一种具有新型结构的 DPP-4/SGLT-2 双靶点抑制剂, 这一研究丰富了 DPP-4 和 SGLT-2 双靶点抑制剂的多样性, 为其开发提供新的设计思路。

关键词: 糖尿病; SGLT-2/DPP-4 抑制剂; 分子对接; 药效团模型; 分子动力学模拟

中图分类号: R914.2 **文献标志码:** A **文章编号:** 1674-5515(2024)07-1689-15

DOI: 10.7501/j.issn.1674-5515.2024.07.005

Design of dual-target inhibitors of DPP-4 and SGLT-2 based on BiRNN model

SUN Dingkan^{1,2}, ZHANG Xinlei², GUO Kailei¹, LI Qi³, LIANG Jialong⁴, MA Lisha², LIN Jiayan², HE Jinsui¹, DIAO Chunyan², WANG Yuwei¹, LIU Xueying²

1. Department of Pharmacy, Shaanxi University of Traditional Chinese Medicine, Xianyang 712046, China

2. Department of Medicinal Chemistry, School of Pharmacy, Air Force Medical University, Xi'an 710032, China

3. College of Life Sciences, Henan University, Kaifeng 475004, China

4. No. 946 Hospital of the People's Liberation Army Ground Force, Yining 835000, China

Abstract: Objective To design a dual-target inhibitor of DPP-4 and SGLT-2. **Methods** The BiRNN model was trained according to the existing DPP-4 and SGLT-2 active compounds to generate new dual-target molecules with potential activity, and the HipHop pharmacophore model was constructed using the reported active compounds, and the new molecules were pre-screened through the pharmacophore model, and then the small molecules were further screened by molecular docking, molecular dynamics simulation and combined free energy calculation, and the candidate compounds were finally obtained. **Results** A total of 7 494 new molecules were generated by the BiRNN model, and the compounds NM186, NM21, NM249, and NM107 were found to be relatively ideal dual-target inhibitors of DPP-4 and SGLT-2. **Conclusion** NM186, NM21, NM249 and NM107 may be dual-target inhibitors of DPP-4/SGLT-2 with novel structures, which enriches the diversity of DPP-4 and SGLT-2 dual-target inhibitors and provides new design ideas for their development.

Key words: diabetes; SGLT-2/DPP-4 inhibitors; molecular docking; pharmacophore model; molecular dynamics simulations

糖尿病是严重危害人类健康的代谢性疾病, 且发病率呈现逐年升高的趋势。2021 年我国糖尿病患

者已达 1.41 亿人, 位居全球第 1 位^[1]。其中 20%~40% 的糖尿病患者罹患糖尿病肾病并发症^[2]。现有

收稿日期: 2024-03-13

基金项目: 陕西省秦创原“科学家+工程师”队伍建设(2023KXJ-080 号)

作者简介: 孙定康, 硕士, 研究方向为药物化学。E-mail: kk327117@163.com

*通信作者: 王育伟 E-mail: wangyw09@gmail.com

刘雪英 E-mail: xylu0427@163.com

的降糖药物中, 钠-葡萄糖协同转运蛋白 2 (SGLT-2) 抑制剂已获临床验证具有明显的肾脏保护作用, 但其固有的生殖、泌尿系统感染风险限制了其运用^[3]。而研究表明二肽基肽酶-4 (DPP-4) 抑制剂可降低与 SGLT-2 抑制剂相关的泌尿生殖道感染风险^[4]。并且与 DPP-4 和 SGLT-2 单独抑制相比, 双靶点抑制能显著降低血糖、糖化血红蛋白 (HbA1c) 水平, 增加血浆胰岛素水平, 并改善肾功能^[5]。

DPP-4 抑制剂和 SGLT-2 抑制剂在降糖机制上也有着明显的协同作用, DPP-4 抑制剂可升高内源性胰高血糖素样肽-1 (GLP-1) 水平, 而 GLP-1 可与位于胰岛 β 细胞表面的 GLP-1 受体 (GLP-1R) 结合并介导其释放胰岛素, 从而降低血糖水平^[6]。SGLT-2 抑制剂主要作用于肾脏近曲小管的 SGLT-2 受体 (SGLT-2R), 抑制其对葡萄糖的重吸收, 增加尿糖的排泄, 从而降低血糖水平^[7]。基于 SGLT-2 抑制剂, 结合 DPP-4 抑制剂的特点设计 DPP-4 和 SGLT-2 双靶点抑制剂可以在降低葡萄糖在肾脏中重吸收的基础上, 同时减缓 DPP-4 对 GLP-1 的灭活作用, 起到双重作用机制发挥叠加效果的目的, 在达到相当降糖效果的同时, 减轻泌尿系统感染的不良反应。不仅具有肾脏获益, 也避免了联合用药的药物相互作用, 增强了患者的依从性降低药物的耐药性及不良反应。目前已有 DPP-4 和 SGLT-2 双靶点抑制剂进入临床研究阶段。如中国盛世泰科生物医药技术 (苏州) 有限公司的创新药 CGT-2201 目前已获得国家药品监督管理局临床试验默示许可。因此开发 SGLT-2/DPP-4 双靶点抑制剂具有重要临床意义。本研究通过机器学习、虚拟筛选、分子对接和分子动力学模拟方法寻找新型 DPP-4/SGLT-2 双靶点抑制剂。

1 材料与方法

1.1 数据准备与使用软件

从 DrugBank 数据库 (<https://go.drugbank.com>) 及 PubMed 数据库 (<https://pubmed.ncbi.nlm.nih.gov>) 中下载已经批准上市的 DPP-4 和 SGLT-2 抑制剂, 再下载数据库中活性排名靠前的小分子作为训练集构建药效团模型。

从 BindingDB 数据库 (<https://www.bindingdb.org>) 中以 “DPP4” “SGLT2” 为关键词检索到 6 896 个 DPP-4 相关的化合物和 1 271 个 SGLT-2 相关的化合物, 其中半数抑制浓度 (IC_{50}) 值小于 $100 \mu\text{mol/L}$ 且具有 DPP-4 和 SGLT-2 选择性抑制的小分子分别

有 4 036 个和 70 个。

诱饵集数据库共 100 个小分子, 由 10 个活性分子和 90 个非活性分子组成, 非活性分子来自 BindingDB 数据库中 $IC_{50} > 200 \mu\text{mol/L}$ 的小分子。活性分子来自 Binding DB 数据库中 $IC_{50} < 200 \text{ nmol/L}$ 的小分子。10 个活性 DPP-4 和 SGLT-2 抑制剂与训练集中的化合物不同, 用 YA01~YA10 编号 10 个活性分子, 用 YI11~YI100 编号 90 个非活性分子。

DPP-4 晶体结构 (PDBID:3G0B)^[8]、SGLT-2 晶体结构 (PDBID:7VSI)^[9]来源于 PDB 蛋白数据库 (<https://www.rcsb.org>)。用于虚拟筛选的 7 494 个化合物来源于 BiRNN 模型 (<https://github.com/ETHmodlab/BIMODAL>)。药效团构建、分子对接、虚拟筛选及 ADMET 活性预测等使用 Discovery Studio 3.5 软件和 Schrödinger 2020 软件完成, 并使用 Pymol 软件进行可视化。使用 Rdkit 插件 (<https://github.com/rdkit/rdkit>) 计算化合物的可合成性 (SA)、类药性 (QED) 和相对分子质量 (Mr) 等理化性质。分子动力学模拟以及结合自由能的计算使用 GROMACS 2022 版本进行, 使用 QtGrace 软件对轨迹进行分析, 计算配置见表 1。

表 1 分子模拟工作站配置

Table 1 Configuration of molecular simulation workstation

类别	规格
CPU	英特尔至强 GOLD 6248R CPU
显卡	英伟达 V100
内存	1.5T DDR4 ECC Register
硬盘	2T*24(RAID1)
操作系统	RedHat Linux/CentOS7.0/Suse Linux
软件版本	Discovery Studio 3.5/Schrödinger 2020/ GROMACS 2022

1.2 实验方法

1.2.1 基于 BiRNN 的 DPP-4 与 SGLT-2 双靶点抑制剂的生成 本研究使用基于 Tensorflow 的双向循环神经网络模型生成了具有潜在双靶点抑制活性的新型小分子。具体方法如下: 分别下载 1.1 中 BindingDB 数据库检索的 IC_{50} 值小于 $100 \mu\text{mol/L}$ 且具有 DPP-4 和 SGLT-2 选择性抑制的小分子二维结构, 并使用 Schrodinger 软件包中的 Maestro 软件将抑制剂转化为 smiles 格式, 用于特征学习, 采集的化合物共 4 106 个。由于 DPP-4 与 SGLT-2 抑制剂数量相差较大, 本研究对 DPP-4 抑制剂使用 Pipeline Pilot 中的分子相似性模块进行了聚类分析, 将所有

分子分为 70 类, 并从每类中选择 1 个代表性分子与 70 个 SGLT-2 抑制剂融合。

基于片段药物设计的方法, 使用匹配分子对算法(MMP)对 SGLT-2 和 DPP-4 抑制剂进行了拆分, 共形成了 71 组不同的分子对; 随后, 将分子对转化为 smile 结构描述符, 并使用 Random Smile 方法对其进行了扩充, 最终形成了包含 23 500 个分子对的骨架-装饰片段库。将片段库中的分子对以 8:1:1 的比例, 分为训练集(80%)、测试集(10%)和验证集(10%), 不同组之间的化合物没有重叠。在

使用训练集和验证集对模型进行训练和参数优化调整的同时, 使用测试集对模型的预测准确性(valid accuracy)进行分析。为了充分认识 SGLT-2 以及 DPP4 抑制剂中不同结构的连接特征, 使用 BiRNN 构建了含有 4 个隐藏层以及 1 272 967 个可调节参数的分子生成模型。模型训练结束后, 使用优化后的模型进行了模拟采样, 共进行 10 000 次采样。采样得到的分子使用 RDkit 软件包中的 rdinchi 模块计算分子的 inchi 描述符, 并去掉重复分子。得到的分子储存为 SDF 格式用于下一步分析, 流程见图 1。

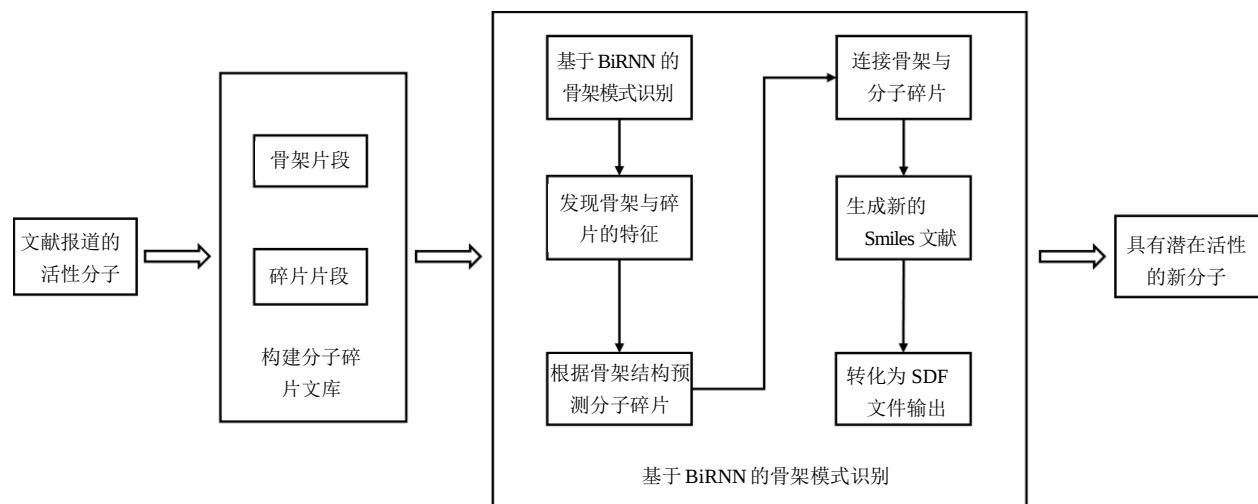


图 1 基于 BiRNN 的组合化学分子设计流程

Fig. 1 Combinatorial chemical molecule design process based on BiRNN

1.2.2 基于分子共同特征的药效团模型 (HipHop) 构建 利用 DS 软件中的 Prepare Ligand 模块分别处理 1.1 项下中的 20 个药效团训练集小分子, 参数 MaxOmitFeat 设定为 0, Principal 设定为 2, 表明构建 HipHop 药效团时, 20 个训练集分子均被定义成有活性的参考分子, 且药效团模型中所有特征元素需与训练集化合物相匹配, 使用 Feature Mapping 选取药效团的特征元素。利用 Common Feature Pharmacophore Generation 模块构建 HipHop 药效团, 模式选择 BEST, 能量阈值 Energy Threshold 设置为 10, Maximum Conformation 设置为 200, 其他参数默认。

1.2.3 药效团模型的验证 通过测试集分子(测试集包括活性分子、非活性分子)对药效团模型可靠性进行评价, 利用 Ligand Profiler 模块将 2 个靶点的 100 个测试集分子分别叠合到生成的 10 个药效团模型中, 确保运行参数与药效团构建设定参数相

同。计算 10 个药效团模型的活性命中率(YA), 并引入富集因子(EF)与拟合优度分数(GH)评价所生成的药效团模型, 其中 YA 反映药效团筛选结果的准确性, EF 反映可靠性, GH 反映稳定性, 表明筛选数据具有统计学意义, 具体计算公式如下^[10]。

$$YH = Ha/Ht$$

$$EF = Ha/Ht \times N/n$$

$$GH = Ha / (4n \times Ht) \times (3n + Ht) \times [1 - (Ht - Ha) / (N - n)]$$

其中 N 代表测试集中总分子数, 数值为 100; n 代表测试集中活性分子数, 数值为 10; Ht 代表命中分子数; Ha 代表命中分子中的活性分子数。为了评价药效团模型对测试集中活性分子和非活性分子的区分能力, 进行了测试集热图验证。热图验证中热图的横坐标代表药效团编号, 纵坐标代表药效团模型与小分子的匹配程度(Fit value), 冷色调表明 Fit value 值高, 匹配度好, 暖色调则表示 Fit value 值低, 匹配度差, Fit value 值低于 1 视为很难匹配

或不匹配。

1.2.4 基于药效团的虚拟筛选 使用 Discovery Studio 3.5 软件中的 Search 3D Database 模块将 BiRNN 模型生成的小分子分别叠合到选中的药效团模型中,筛选出药效团 Fit value>1,可与药效团特征相匹配的小分子,再对筛选后的小分子取交集得到初筛小分子。

1.2.5 初筛小分子对接研究 使用 Schrödinger 软件中的 Glide 模块将初筛小分子与 DPP-4 和 SGLT-2 分别进行高精度 (XP) 对接。蛋白质格点文件以 3G0B 和 7VSI 蛋白结构中的结晶分子为网格中心,网格框定义为 20 Å×20 Å×20 Å (1 Å=0.1 nm),使用 Receptor Grid Generation 生成格点文件用于对接,其他参数保持默认值。设置好合适的参数后,将晶体结构中的配体分子提取出来并重新对接至预先定义好的对接盒子内。将对接后配体与原复合物晶体结构中的配体构象进行对比,通过计算均方根偏差 (RMSD) 验证所设定的对接参数的可靠性。取各打分值前 2% 的小分子再取交集得到备选小分子。

1.2.6 备选小分子 ADMET 预测 药动学研究 (ADMET) 是对药物的吸收、分布、代谢、排泄和毒性的全面研究,是研究药物体内过程的重要方法,也是药物筛选中需要考虑的重要标准。DS 软件自带 ADMET 预测模块,可以对输入的小分子化合物进行 ADMET 性质预测,主要包括 6 项参数:25 °C 水溶解度 (solubility)、血脑屏障通透性 (BBB)、细胞色素 P4502D6 抑制性 (cytochrome P4502D6 inhibition)、肝毒性 (hepatotoxicity)、肠道吸收 (absorption) 和血浆蛋白结合率 (plasma protein binding, PPB)。其中 absorption、solubility 与 BBB 3 项以相关等级表示。利用 DS 软件中的 Calculate Molecular Properties 模块,点击 ADMET Descriptors 模块,对备选小分子进行 ADMET 活性预测。

1.2.7 备选小分子可合成性、类药性分析 为了评价生成分子的质量,使用 Rdkit 插件 (<https://github.com/rdkit/rdkit>) 计算化合物的 SA、QED 和 Mr 等理化性质^[11]。

1.2.8 备选小分子动力学模拟研究 分子动力学模拟 (MD) 采用 Gromacs 2022 程序,将备选小分子对接的结果作为分子动力学模拟的初始构象。选择 CHARMM36 力场。使用 Avogadro 程序对所有备选分子添加氢原子并进行去质子化。使用 CgenFF 网站对小分子施加力场。然后,将复合物溶剂化成 1 nm

厚距离蛋白质 10 Å 的 SPC 水分子立方体盒 (cubic box),盒内填充 TIP3P 水模型。向受体-配体-溶剂体系中添加 Na⁺和 Cl⁻以保持电中性,并在所有方向上施加周期性边界条件 (PBC) 以避免在模拟过程中的边缘效应。用 Particle-Mesh-Ewald (PME) 算法计算长程静电相互作用。对系统进行能量最小化。体系能量优化后,在 100 ps 的等温 (300 K) 和等压 (1atm) 条件下进行 NVT 和 NPT 系综以平衡系统。以 2 fs 时间步长进行 100 ns 的分子动力学模拟^[12]。模拟结束后,使用 QtGrace 软件分别分析每个复合物的均方根误差 (RMSD) 和蛋白质氨基酸残基的均方根波动 (RMSF)。采用分子力学/泊松-玻尔兹曼表面积 (MM/PBSA) 方法计算模拟过程最后 10 ns 中每一帧的结合自由能并取均值进行对比分析^[13],验证在动态溶剂环境中的结合能力。

2 结果

2.1 训练集药效团模型的筛选

经 DrugBank、PubMed 数据库筛选得到已经批准上市的 DPP-4 抑制剂 12 个和 SGLT-2 抑制剂 9 个,另选取数据库中活性排名靠前的小分子补充至 20 个作为训练集构建药效团模型,见表 2、3。

2.2 DPP-4 和 SGLT-2 与已知抑制剂的关键相互作用模式分析

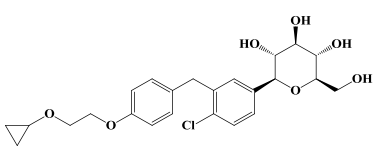
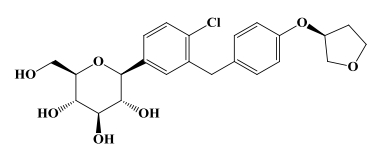
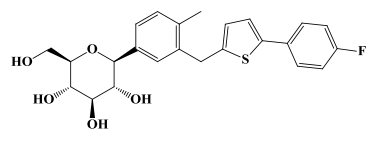
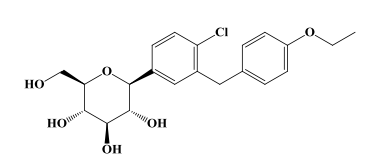
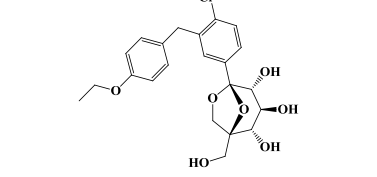
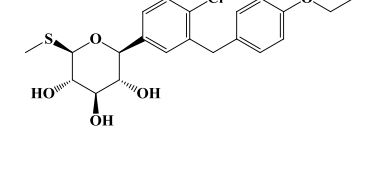
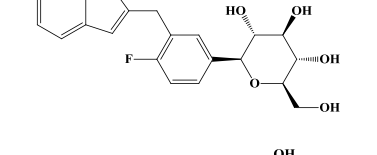
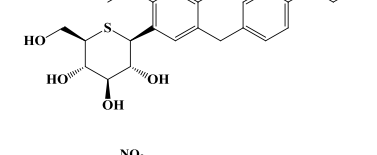
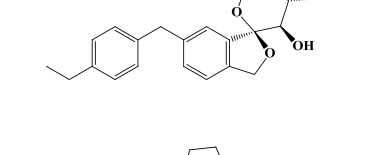
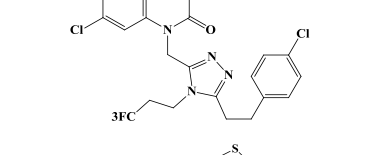
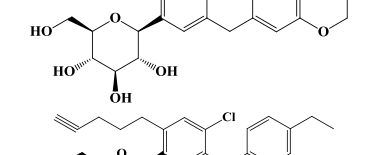
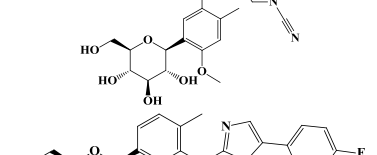
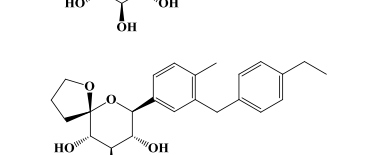
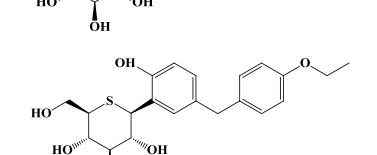
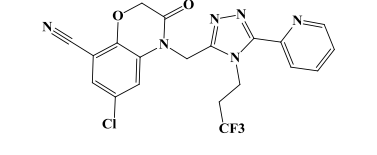
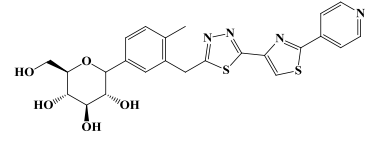
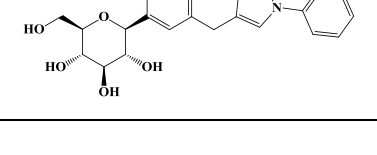
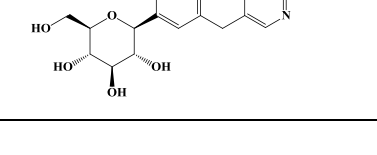
本研究对 3G0B、7VSI 复合物结构进行了构效分析。在 DPP-4 的活性口袋中,阿格列汀 (alogliptin) 哌嗪基的 NH 与氨基酸 Glu 205/206 之间存在氢键相互作用;氨基酸 Tyr 631 与杂环的羰基之间存在氢键;氰基与氨基酸 Arg 125 形成的氢键相互作用^[4],见图 2A。在 SGLT-2 的活性口袋中,恩格列净 (Empagliflozin) 的 C2-OH 与氨基酸 ASN75 和 GLU99 形成氢键;C3-OH 与氨基酸 PHE98 的主链羰基形成氢键;C4-OH 与氨基酸 SER287 形成氢键;C6-OH 与氨基酸 GLN457 形成氢键^[3],见图 2B。

2.3 双靶点抑制剂的生成

本研究构建了基于双向循环神经网络的分子生成模型,模型参数:BiRNN 输出变量 (Lstm_out=192)、学习速率 (Learning_rate=1×10⁻³)、训练步数 (Epoch=100)、样本数量 (Batch_size=140),训练 30 步后,模型的准确率趋于收敛,准确率可达到 90%,见图 3。使用 MMP 方法对部分 FDA 批准药物进行了拆分,并将骨架碎片导入分子生成模型。在完成去除重复分子后,共生成了 7 494 个不同的候选化合物分子,用于下一步的虚拟筛选。

序号	结构	IC ₅₀ /(nmol L ⁻¹)	序号	结构	IC ₅₀ /(nmol L ⁻¹)
1		3.80	11		19.00
2		1.00	12		10.30
3		26.00	13		8.47
4		1.75	14		1.60
5		3.50	15		4.00
6		19.00	16		33.00
7		26.00	17		13.00
8		22.00	18		180.00
9		0.80	19		0.92
10		7.50	20		23.00

表 3 SGLT-2 训练集化合物
Table 3 SGLT-2 training set compounds

序号	结构	IC ₅₀ /(nmol L ⁻¹)	序号	结构	IC ₅₀ /(nmol L ⁻¹)
1		2.00	11		3.10
2		2.20	12		1.10
3		0.88	13		1.80
4		7.38	14		2.26
5		2.90	15		22.00
6		3.00	16		3.00
7		0.30	17		8.10
8		53.00	18		17.40
9		110.00	19		72.60
10		32.00	20		46.90

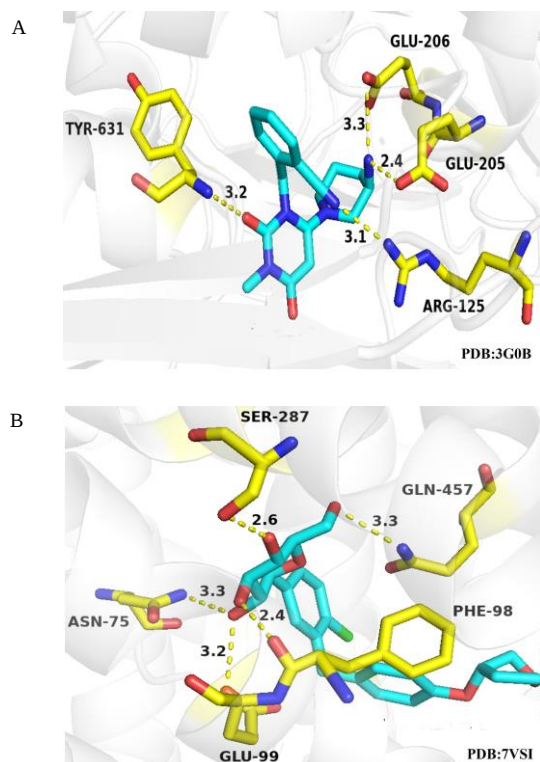


图 2 DPP-4 晶体结构与 alogliptin (A) 及 SGLT-2 晶体结构与 empagliflozin (B) 的 3D 结合作用模式

Fig. 2 3D bonding mode of DPP-4 crystal structure with alogliptin (A) and SGLT-2 crystal structure with empagliflozin (B)

2.4 HipHop 药效团的构建与验证

为了筛选出潜在的 DPP-4 和 SGLT-2 双靶点小分子抑制剂,本研究利用 Discovery Studio 3.5 中的 HipHop 定性模型方法分别构建了 DPP-4 抑制剂和 SGLT-2 抑制剂的药效团模型。构建 2 个药效团模型选用的训练集分别包含 20 个小分子化合物,它们

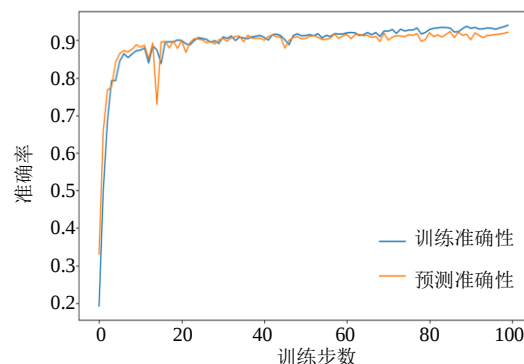


图 3 BiRNN 模型模拟结果

Fig. 3 Simulation results of the BiRNN model

具有较好的 DPP-4 和 SGLT-2 抑制活性,结构及活性见表 2、3。

生成的药效团结果参数见表 4、5,可知构建的 DPP-4 药效团模型包含 3 个特征元素,包括 1 个氢键受体,1 个电荷中心和 1 个疏水中心;构建的 SGLT-2 药效团模型包含 2 个特征元素,包括 3 个氢键受体和 1 个疏水中心。

测试集结果验证显示,DPP-4 抑制剂的 1、8、9 号药效团的 YA、GH 都为 100%且 EF 都为 10,见表 6,3 个药效团模型筛选结果的准确度均较高,且根据热图(图 4)对比发现,1 号药效团对测试集小分子区分度更强,对阴性化合物和阳性化合物的评判能力更好,因此选择 1 号药效团用于后续的虚拟筛选。SGLT-2 抑制剂的 1 号药效团是 10 个药效团中唯一的 YA 和 GH 都为 100%且 EF 为 10 的药效团,见表 7、图 5,因此选择 1 号药效团用于后续的虚拟筛选。

表 4 DPP-4 药效团模型及关键参数

Table 4 DPP-4 pharmacophore model and key parameters

模型	特征	分值	药效团与训练集分子的匹配情况	药效特征数目	药效团特征的匹配情况
1	PHA	91.208	11111111111111111111	000000000000000000	3
2	PHA	86.147	11111111111111111111	000000000000000000	3
3	PHA	82.250	11111111111111111111	000000000000000000	3
4	PHA	81.913	11111111111111111111	000000000000000000	3
5	HDA	80.143	11111111111111111111	000000000000000000	3
6	HDA	78.379	11111111111111111111	000000000000000000	3
7	PHA	77.995	11111111111111111111	000000000000000000	3
8	PHA	77.243	11111111111111111111	000000000000000000	3
9	PHA	75.558	11111111111111111111	000000000000000000	3
10	HDA	75.269	11111111111111111111	000000000000000000	3

表 5 SGLT-2 药效团模型及关键参数

Table 5 SGLT-2 pharmacophore model and key parameters

模型	特征	分值	药效团与训练集分子的匹配情况	药效特征数目	药效团特征的匹配情况
1	HDD	102.428	11111111111111111111	00000000000000000000	3
2	HDD	102.428	11111111111111111111	00000000000000000000	3
3	HDD	102.428	11111111111111111111	00000000000000000000	3
4	HDD	102.428	11111111111111111111	00000000000000000000	3
5	HDD	101.773	11111111111111111111	00000000000000000000	3
6	HDD	101.220	11111111111111111111	00000000000000000000	3
7	HDA	100.428	11111111111111111111	00000000000000000000	3
8	HDA	100.428	11111111111111111111	00000000000000000000	3
9	HDA	100.428	11111111111111111111	00000000000000000000	3
10	HDA	100.428	11111111111111111111	00000000000000000000	3

表 6 DPP-4 测试集验证药效团模型结果数据

Table 6 Data of the DPP-4 test set for the validation of pharmacophore models

模型	N	n	Ht	Ha	YA/%	GH/%	EF
1	100	10	10	10	100	100	10.0
2	100	10	11	10	91	92	9.1
3	100	10	13	10	77	80	7.7
4	100	10	15	10	67	71	6.7
5	100	10	22	9	41	46	4.1
6	100	10	35	9	26	30	2.6
7	100	10	14	9	64	67	6.4
8	100	10	10	10	100	100	10.0
9	100	10	10	10	100	100	10.0
10	100	10	53	9	17	18	1.7

N: 测试集总分子数, n: 活性分子数, Ht: 命中分子数, Ha: 命中分子中活性分子数, YA: 活性命中率, GH: 拟合优度分数, EF: 富集因子。

N: number of test sets, n: number of active molecules, Ht: number of hit molecules, Ha: number of active molecules in hit molecules, YA: activity hit ratio, GH: goodness of fit fraction, EF: enrichment factor.

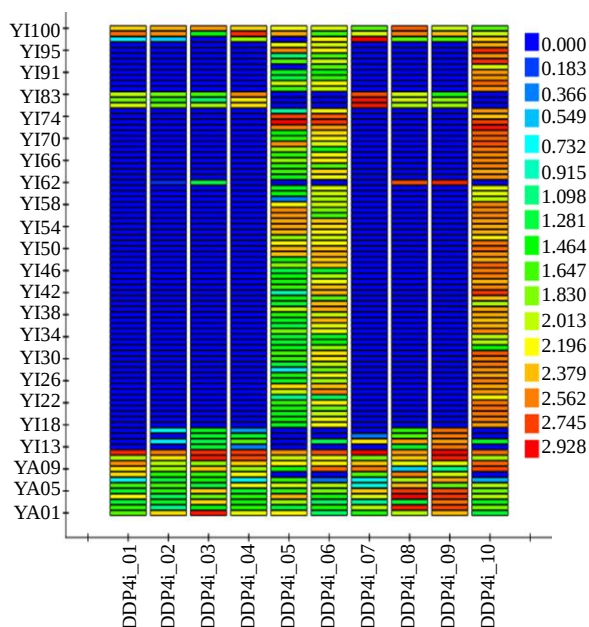


图 4 DPP-4 Ligand Profiler 热图

Fig. 4 DPP-4 Ligand Profiler heatmap

2.5 对接方法验证结果

将共结晶分子的对接构象与原复合物配体构象叠合, 并计算其 RMSD 值以考查复现率 (RMSD < 2.0 Å 表明复现率较佳)。结果如图 6 所示, DPP-4 和 SGLT-2 的对接构象与其原始复合物构象的 RMSD 值分别为 0.396 8、0.515 2 Å。

2.6 虚拟筛选结果

利用 DPP-4 的 1 号药效团和 SGLT-2 的 1 号药效团对 BiRNN 模型生成的 7 494 个小分子分别进行初筛, 设置 Fit Value > 1, 取筛选结果的交集共计 2 452 个小分子。使用 Schrödinger 软件中的 Glide 模块将 2 452 个小分子与 DPP-4 和 SGLT-2 分别进行高精度 (XP) 对接, 取打分前 2% 分别得到 49 个小分子, 对 49 个小分子取交集得到 10 个备选小分子, 具体结构及对接打分值见表 8。

2.7 备选小分子的 ADMET 预测

使用 Discovery Studio 3.5 进行备选小分子的 ADMET 性质预测, 结果见表 9。

表 7 SGLT-2 测试集验证药效团模型结果数据

Table 7 Results of the SGLT-2 test set validated pharmacophore model

模型	N	n	Ht	Ha	YA/%	GH/%	EF
1	100	10	10	10	100	100	10.0
2	100	10	10	9	90	89	9.0
3	100	10	9	8	89	86	8.9
4	100	10	9	7	78	74	7.8
5	100	10	10	4	40	37	4.0
6	100	10	10	7	70	68	7.0
7	100	10	15	7	47	48	4.7
8	100	10	12	7	58	58	5.8
9	100	10	13	7	54	54	5.4
10	100	10	16	7	44	45	4.4

N: 测试集总分子数, n: 活性分子数, Ht: 命中分子数, Ha: 命中分子中活性分子数, YA: 活性命中率, GH: 拟合优度分数, EF: 富集因子。

N: number of test sets, n: number of active molecules, Ht: number of hit molecules, Ha: number of active molecules in hit molecules, YA: activity hit ratio, GH: goodness of fit fraction, EF: enrichment factor.

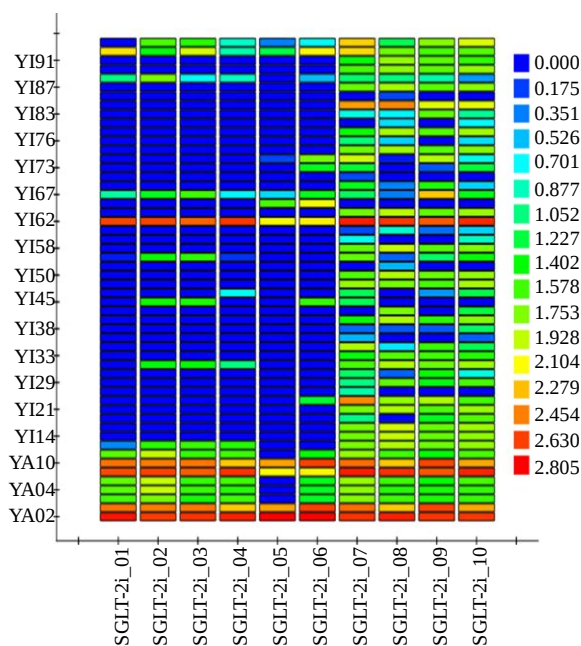


图 5 SGLT-2 Ligand Profiler 热图

Fig. 5 Heatmap of SGLT-2 Ligand Profiler

经过 ADMET 预测发现 10 个备选小分子均具有良好的水溶性和肠道吸收性,且不能通过血脑屏障,对 CYP2D6 酶无抑制性,与其他通过 CYP2D6 酶代谢的药物联用时,不会影响其他药物的代谢过程,不良反应较低。化合物 NM68、NM141、NM304 这 3 个化合物存在肝毒性。

2.8 备选小分子可合成性打分 SA_Score 以及 rdkit QED 定量评估类药性

10 个备选分子的 SA、QED、酯水分配系数 (lgP)、氢键供体数目 (HBD)、氢键受体数目 (HBA)、可旋转键 (RotB)、Mr 这些理化性质见表 10。SA 评分代表小分子合成难易程度,用 1~10 进

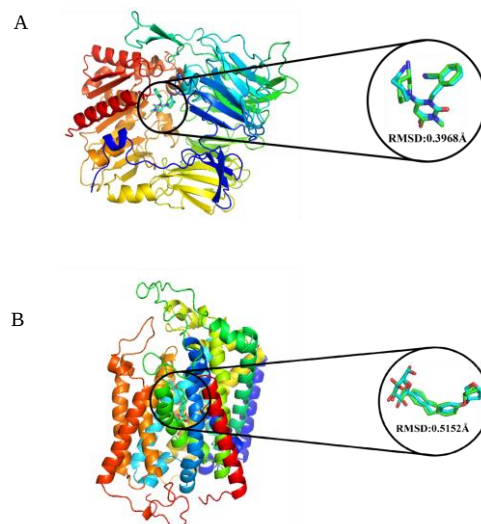


图 6 DPP-4 抑制剂 alogliptin (A) 和 SGLT-2 抑制剂 empagliflozin (B) 的构象叠合图

Fig. 6 Conformational superposition of DPP-4 inhibitor alogliptin (A) and SGLT-2 inhibitor empagliflozin (B)

行评价,越靠近 1 表明越容易合成,越靠近 10 表明合成越困难。QED 是一种将药物相似性量化为 0~1,用于定量评估药物的类药性,QED 的值越高,表示药物越可能具有良好的类药性和生物活性。HBD<5、HBA<10、RotB≤10、Mr<500 类药性较好;lgP<5 较好。10 个备选小分子均具有较好的可合成性及类药性,均具有一定的开发价值。

2.9 分子动力学模拟和结合自由能

RMSD 表示某一时刻的构象与目标构象所有原子偏差的加和,是衡量体系是否稳定的重要依据。对 10 个备选小分子对接复合物体系进行分子动力学模拟。其中 DPP-4 和 SGLT-2 与 NM186、NM21、

表 8 10 个备选小分子的结构及结合能

Table 8 Structure and docking scores of the 10 candidate small molecules

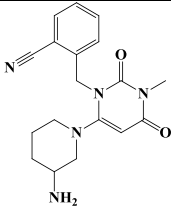
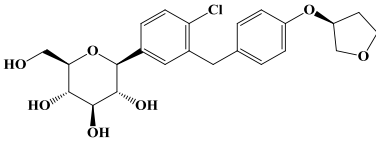
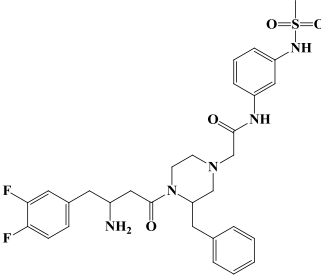
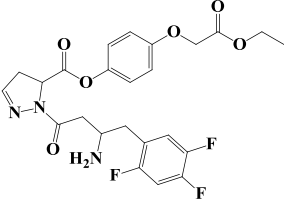
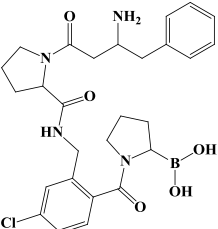
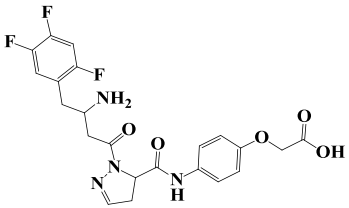
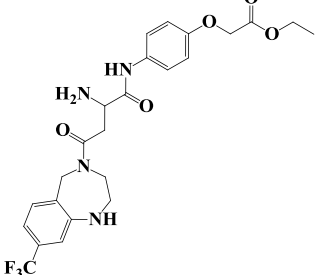
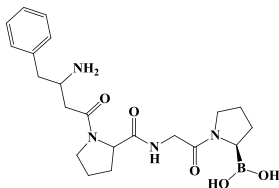
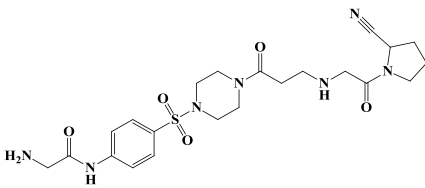
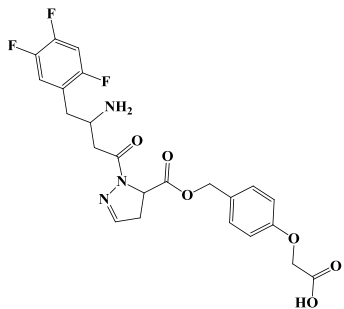
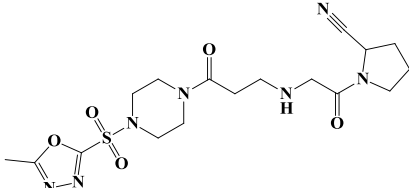
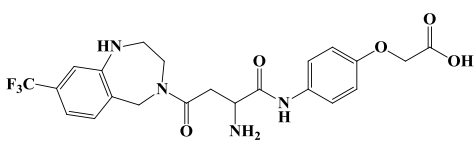
名称	结构	结合能/(kcal mol ⁻¹)	
		DPP-4 (ID:3G0B)	SGLT-2 (ID:7VSI)
alogliptin		-11.910	—
empagliflozin		—	-10.209
NM21		-14.305	-10.729
NM141		-13.725	-8.147
NM171		-10.053	-9.408
NM68		-9.510	-9.035
NM248		-9.583	-8.882

表 8 (续)

名称	结构	结合能/(kcal mol ⁻¹)	
		DPP-4 (ID:3G0B)	DPP-4 (ID:3G0B)
NM186		-13.208	-10.905
NM107		-13.701	-14.379
NM304		-12.108	-11.312
NM108		-12.851	-8.475
NM249		-12.773	-11.113

NM249、NM107 复合物体系的 RMSD 值随时间变化的结果较为稳定如图 7 所示, 在 100 ns 的模拟过程中, 除了 3G0B-NM107 这个体系在 70 ns 后趋于平稳外, 其余 7 个体系均在 30 ns 后波动逐渐趋于平稳, 复合物的 RMSD 值始终稳定在 0.1~0.2 nm, 波动较小。RMSF 表示蛋白质氨基酸在整个模拟过程中的柔性及运动剧烈程度。如图 7 所示, 除末端区域和低质量区域中的残基外, 大多数氨基酸残基的波动范围小于 0.4 nm, 表明在整个模拟过程中蛋白质的柔性变化在合理区间内。RMSD 和 RMSF 的分析结果显示, 在为期 100 ns 的分子动力学模拟过程中, 所有靶蛋白-配体复合物都达到平衡状态, 表明 4 个候选化合物 NM186、NM21、NM249、NM107

与 DPP-4 和 SGLT-2 的结合相对稳定。为充分考虑配体与蛋白结合时的溶剂效应和动态效应等因素, 采用 MM/PBSA 方法计算模拟过程最后 10 ns 中每一帧的平均结合自由能, 进一步比较动态的溶剂环境中 4 个候选化合物和 alogliptin/empagliflozin 对 DPP-4 和 SGLT-2 的抑制能力。结果如表 11 所示, 4 个候选化合物的结合自由能均低于 alogliptin/empagliflozin, 与分子对接的结果一致。进一步验证了 4 个候选化合物对 DPP-4 和 SGLT-2 的结合能力强。同时, 4 个候选化合物为结构新颖的未有相关文献报道的新化合物, 可视为虚拟筛选研究的命中化合物, 作为新型潜在的 DPP-4 和 SGLT-2 双靶点抑制剂进行下一步实验研究。

表 9 10 个备选小分子 ADMET 性质预测

Table 9 Prediction of the properties of 10 candidate small molecules for ADMET

化合物	水溶解度 ^a	血脑屏障通透性 ^b	CYP2D6 抑制性	肝毒性	血浆蛋白结合率 ^c	肠道吸收 ^d
NM21	3	3	无	无	True	1
NM68	3	3	无	有	False	1
NM107	3	4	无	无	True	1
NM108	3	3	无	无	True	0
NM141	3	3	无	有	False	0
NM171	3	4	无	无	True	2
NM186	3	3	无	无	True	1
NM248	3	3	无	无	True	0
NM249	3	4	无	无	False	1
NM304	3	3	无	有	True	1

a: 0~5 分别表示极不可能溶、微溶、少溶、溶解性较好、溶解度适中、极易溶; b: 0~4 分别表示极易透过、易透过、中等、不易透过、不能被定义; c: 化合物血浆蛋白结合率是否大≥90%的分类, False 代表低于 90%, True 代表高于 90%; d: 0~3 分别表示易吸收、中等程度吸收、不易吸收、极难吸收。

a: 0 — 5 means highly unlikely to dissolve, slightly soluble, less soluble, better solubility, moderate solubility, very soluble; b: 0 — 4 means easy to pass through, easy to pass through, medium, not easy to pass through, and can not be defined; c: Classification of whether the plasma protein binding rate of the compound is greater than or equal to 90%, False means less than 90%, True means more than 90%; d: 0 — 3 indicates easy absorption, medium absorption, difficult absorption, and extremely difficult absorption, respectively.

表 10 10 个备选小分子理化性质分析

Table 10 Physicochemical properties of 10 alternative small molecules

化合物	相对分子质量	可合成性评分	氢键供体数目	氢键受体数目	可旋转键	酯水分配系数	类药性
NM249	480.4	2.98	5	9	7	2.28	0.87
NM248	508.5	3.02	4	9	8	2.76	0.74
NM107	505.6	3.03	4	12	9	1.09	0.66
NM141	507.5	3.22	2	9	10	2.51	0.63
NM21	599.7	3.32	4	9	11	2.99	0.65
NM108	439.5	3.39	1	12	7	1.29	0.71
NM68	478.4	3.48	4	9	9	2.05	0.67
NM304	493.4	3.56	3	9	10	2.16	0.78
NM186	430.3	3.57	5	9	8	0.94	0.69
NM171	540.9	3.73	5	9	9	1.52	0.76

2.10 候选小分子与受体蛋白的结合模式分析

使用 PyMol 软件对上述化合物与受体蛋白的结合作用模式进行逐一分析, 如图 8 所示, NM21、NM107、NM186、NM249 4 个候选小分子与 DPP-4 的关键氨基酸 GLU205/206 均具有 2 组氢键相互作用, 且距离均小于 3.5 Å。NM107、NM249 与 DPP-4 的关键氨基酸 ARG125 也有一组氢键相互作用。4 个化合物除了与 DPP-4 的部分关键氨基酸有相互作用也与其它氨基酸存在着氢键作用, 相对较多的相互作用可能是其结合能力比 alogliptin 更强的有力证据。在 SGLT-2 的口袋中发现 4 个小分子与关键氨基酸 GLN457 和 GLU99 均具有两组氢键作用, 且距离均小于 3.5 Å, 其中 NM186 和 NM107 与 SGLT-2 的关键氨基酸 ASN75 具有氢键相互作用, NM186 和 NM249 与 SGLT-2 的关键氨基酸 PHE98

有氢键相互作用。这些相似的相互作用可能是 4 个化合与 DPP-4 和 SGLT-2 结合能力较强的主要原因。

3 讨论

本研究通过机器学习、虚拟筛选、分子对接和分子动力学模拟方法寻找新型 DPP-4/SGLT-2 双靶点抑制剂。在构建药效团模型时本研究选择已报道或上市的小分子, 并使用已被证实的活性化合物作为验证集去验证药效团模型, 保证构建的药效团模型可靠。选用 BiRNN 模型, 其计算量较小, 并且碎片都是来源于真实的分子, 预测的分子具有较高的可合成性, 模型通过学习已知活性化合物骨架与碎片的结合作用模式预测出具有新型结构的小分子, 准确性较高。利用 DPP-4 和 SGLT-2 的最优药效团分别去初步筛选 BiRNN 模型生成的 7 497 个小分子并取交集共计得到 2 452 个小分子。

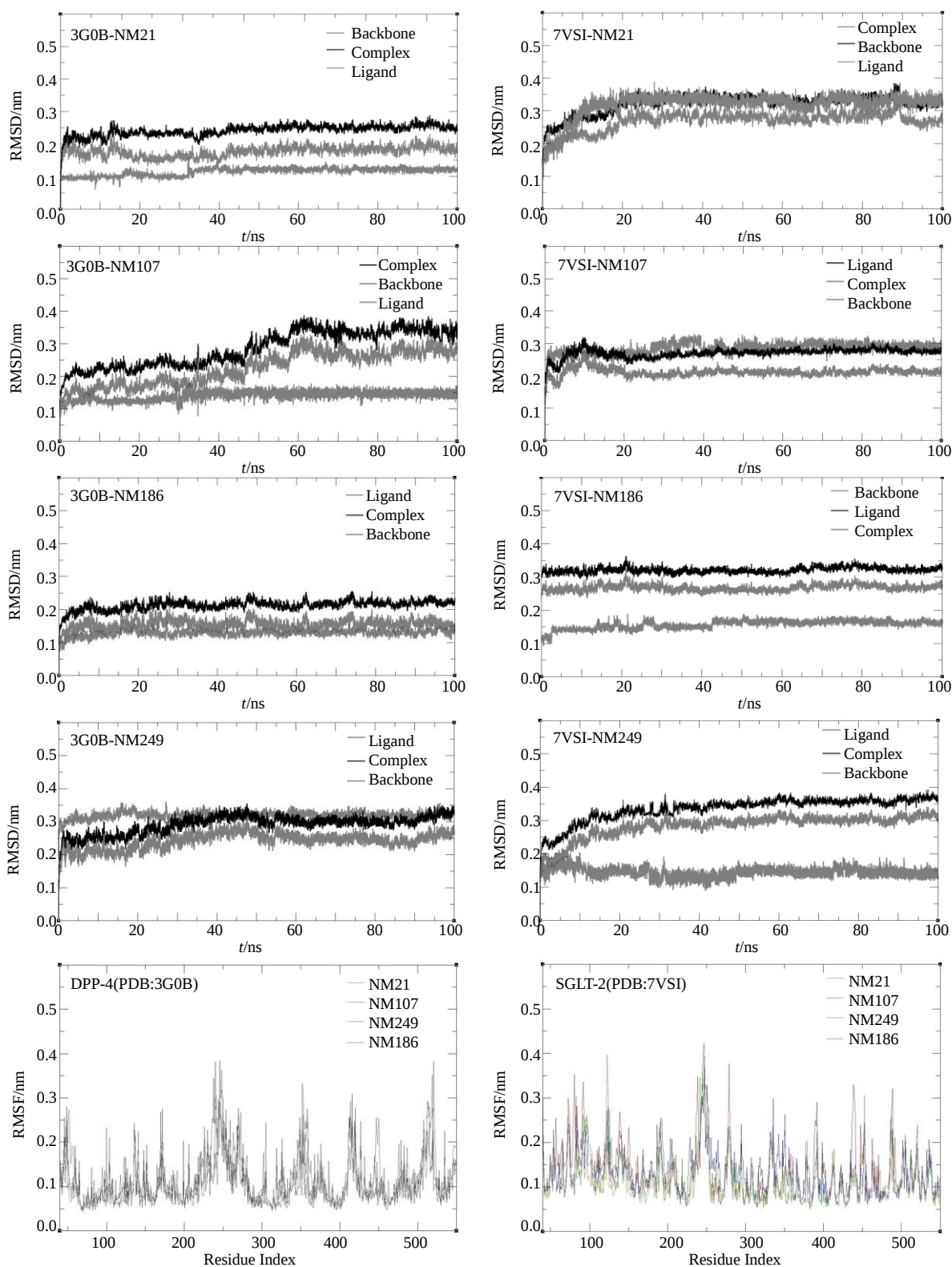


图 7 MD 模拟过程中化合物 NM21、NM107、NM186、NM249 与 DPP-4 (PDB:3G0B) 和 SGLT-2 (PDB:7VSI) 之间的 RMSD 图及氨基酸的 RMSF 图

Fig. 7 RMSD plots and amino acid RMSF plots of compounds NM21, NM107, NM186 and NM249 with DPP-4 (PDB:3G0B) and SGLT-2 (PDB:7VSI) during MD simulation

表 11 4 个候选化合物及 alogliptin/empagliflozin 的结合自由能

化合物名称	结合自由能 $\Delta E_{\text{bind}}/(\text{kcal mol}^{-1})$	
	DPP-4(ID:3G0B)	SGLT-2(ID:7VSI)
alogliptin	-57.52 ± 0.21	—
empagliflozin	—	-43.051 ± 0.91
NM21	-85.31 ± 4.76	-48.397 ± 0.36
NM186	-74.63 ± 0.62	-75.112 ± 15.26
NM249	-60.371 ± 1.54	-45.024 ± 1.07
NM107	-77.962 ± 12.51	-62.097 ± 6.04

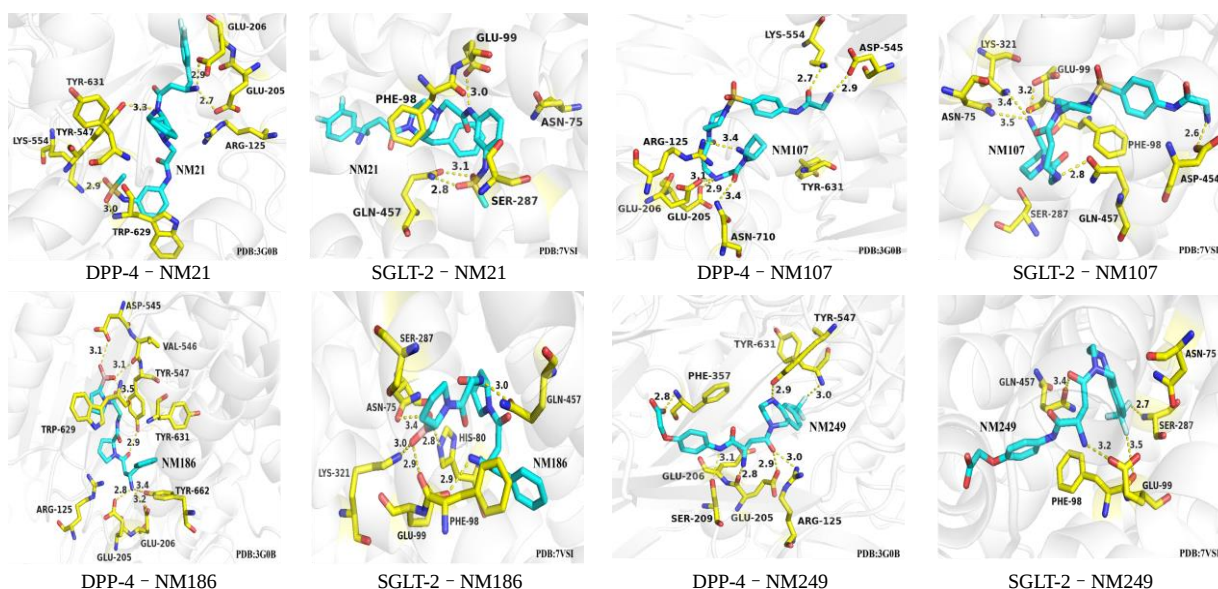


图 8 化合物 NM21、NM107、NM186、NM249 与 DPP-4、SGLT-2 之间的相互作用关系图

Fig. 8 Interaction between NM21, NM107, NM186, NM24 to DPP-4 and SGLT-2

为了验证上述 10 个备选小分子复合物体系的稳定性,采用 Gromacs 进行了 100 ns 的分子动力学模拟,以分析复合物体系的 RMSD 及 RMSF 随时间波动情况,发现化合物 NM186、NM21、NM249、NM107 复合物体系的 RMSD 值最终稳定后的幅值始终在 0.1~0.2 nm,波动较小。通过 RMSF、结合自由能、ADMET 以及理化性质分析进一步的印证了化合物 NM186、NM21、NM249、NM107 可能是潜在的 DPP-4/SGLT-2 双靶点抑制剂。4 个候选化合物作为潜在的新型 DPP-4/SGLT-2 双靶点抑制剂,为 DPP-4/SGLT-2 双靶点抑制剂的设计和研发提供了理论指导。

4 结论

本研究构建了用于 DPP-4 抑制剂和 SGLT-2 抑制剂筛选的最佳药效团模型。并且采用 BiRNN 模型构建了 DPP-4/SGLT-2 双靶点抑制剂的库共计

研究通过对复合物晶体分子的对接构象与原复合物配体构象进行叠合,并计算 RMSD 值以表明基于分子对接筛选得到结果的可靠性,结果表明配体对接构象与共结晶构象几乎完全重合,说明本研究构建的虚拟筛选模型具备可靠性。随后将初筛得到的 2 452 个小分子进行分子对接,再对对接结果分别取前 2%再取交集最终得到 10 个结合作用模式与现有 DPP-4 抑制剂 alogliptin 和 SGLT-2 抑制剂 empagliflozin 类似且对接打分较高的小分子。

7 494 个小分子。通最佳药效团对这 7 494 个小分子进行初步筛选并取出交集,再通过分子对接以及分子动力学模拟技术得到具有较高结合稳定性的小分子 NM186、NM21、NM249、NM107。

后续将会对这 4 个化合物进行化学合成以及活性测试,并进行结构优化改造以获得活性更高的 DPP-4/SGLT-2 双靶点抑制剂,本研究为 DPP-4/SGLT-2 双靶点抑制剂的开发以及基于 BiRNN 及药效团模型的小分子药物的开发提供了参考。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Williams R. IDF diabetes atlas ninth edition 2021 [C]. 2021: 31-37.
- [2] Wang J, Zhang Q B, Li S, et al. Low molecular weight fucoidan alleviates diabetic nephropathy by binding fibronectin and inhibiting ECM-receptor interaction in

- human renal mesangial cells [J]. *Int J Biol Macromol*, 2020, 150: 304-314.
- [3] Wilding J. SGLT2 inhibitors and urinary tract infections [J]. *Nat Rev Endocrinol*, 2019, 15(12): 687-688.
- [4] Fadini G P, Bonora B M, Mayur S, *et al.* Dipeptidyl peptidase-4 inhibitors moderate the risk of genitourinary tract infections associated with sodium-glucose co-transporter-2 inhibitors [J]. *Diabetes Obes Metab*, 2018, 20(3): 740-744.
- [5] Ko E J, Shin Y J, Cui S, *et al.* Effect of dual inhibition of DPP4 and SGLT2 on tacrolimus-induced diabetes mellitus and nephrotoxicity in a rat model [J]. *Am J Transplant*, 2022, 22(6): 1537-1549.
- [6] Mirza A Z, Althagafi I I, Shamshad H. Role of PPAR receptor in different diseases and their ligands: Physiological importance and clinical implications [J]. *Eur J Med Chem*, 2019, 166: 502-513.
- [7] Saisho Y. SGLT2 inhibitors: The star in the treatment of type 2 diabetes [J]. *Diseases*, 2020, 8(2): 14.
- [8] Zhang Z Y, Wallace M B, Feng J, *et al.* Design and synthesis of pyrimidinone and pyrimidinedione inhibitors of dipeptidyl peptidase IV [J]. *J Med Chem*, 2011, 54(2): 510-524.
- [9] Niu Y G, Liu R, Guan C C. *et al.* Structural basis of inhibition of the human SGLT2-MAP17 glucose transporter [J]. *Nature*, 2022, 601(7892): 280-284.
- [10] Che J X, Wang Z L, Sheng H C, *et al.* Ligand-based pharmacophore model for the discovery of novel CXCR2 antagonists as anti-cancer metastatic agents [J]. *R Soc Open Sci*, 2018, 5(7): 180176.
- [11] Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions [J]. *J Cheminform*, 2009, 1(1): 8.
- [12] Allayeh A K, El-Boghdady A H, Said M A, *et al.* Discovery of pyrano [2,3-c] pyrazole derivatives as novel potential human coronavirus inhibitors: Design, synthesis, in silico, in vitro, and ADME studies [J]. *Pharmaceuticals*, 2024, 17(2): 198.
- [13] Zhou N, Zheng C Z, Tan H T, *et al.* Identification of PLK1-PBD inhibitors from the library of marine natural products: 3D QSAR pharmacophore, ADMET, scaffold hopping, molecular docking, and molecular dynamics study [J]. *Mar Drugs*, 2024, 22(2): 83.

[责任编辑 高源]