

专题: 生物分子模拟中的机器学习

## 靶向 PD-L1 蛋白的计算机辅助药物筛选\*

林开东<sup>1)</sup> 林晓倩<sup>1)2)</sup> 林绪波<sup>1)†</sup>

1) (北京航空航天大学, 医学科学与工程学院/生物与医学工程学院, 北京市生物医学工程高精尖创新中心, 北京 100191)

2) (北京航空航天大学沈元学院, 北京 100191)

(2023 年 6 月 29 日收到; 2023 年 8 月 10 日收到修改稿)

针对 PD-1/PD-L1 免疫检查点的单克隆抗体抑制剂逐渐进入市场并在多种类型的肿瘤治疗中取得一定的积极效果. 然而, 随着应用范围的不断扩展, 抗体药物的局限性以及过多同质化研究等问题逐渐显现出来, 小分子化合物抑制剂成为了研究者们关注的新焦点. 本文旨在利用基于配体和基于结构的结合活性预测方法实现针对 PD-L1 靶点的小分子化合物虚拟筛选, 从而帮助加速小分子药物的开发. 通过从相关研究文献及专利收集 PD-L1 小分子抑制活性数据集, 根据不同分子表征方法和算法构建机器学习活性判定分类模型和活性强度预测回归模型, 两类模型从大型类药小分子库 (ZINC15) 中筛选获得 68 种高 PD-L1 抑制活性候选化合物. 其中 10 种化合物不仅具备良好的药物相似性和药代动力学, 还在分子对接中与已报道的热点化合物表现出同等水平的结合强度和相似的作用机制, 这一现象在后续分子动力学模拟和结合自由能估计中得到进一步验证. 本文提出了一个融合基于配体方法和基于结构方法的计算机辅助药物研发工作流程, 其在大型化合物数据库中有效筛选出有潜力的 PD-L1 小分子抑制剂, 有望助力加速肿瘤免疫治疗的应用.

**关键词:** PD-1/PD-L1, 虚拟筛选, 机器学习, 分子动力学模拟**PACS:** 05.10.-a, 02.70.-c**DOI:** 10.7498/aps.72.20231068

## 1 引言

阻断免疫检查点蛋白与其配体的结合作为肿瘤治疗的方法之一, 近年来在临床应用中迅速发展. 正常生理状态下, 部分负调节因子作为免疫检查点来抑制 T 细胞的过度激活, 确保免疫反应保持自我耐受<sup>[1]</sup>. 然而不幸的是, 肿瘤细胞可以利用这种机制诱导 T 细胞衰竭, 形成促进肿瘤生长和侵袭的微环境, 从而逃避免疫系统的攻击<sup>[2-4]</sup>. 为了重新激活并增强 T 细胞介导的抗肿瘤功能, 前人已经设计了一系列疗法来阻断免疫检查点蛋白与其配体的结合<sup>[5,6]</sup>. 其中, 细胞程序性死亡蛋白 1 (programmed cell death protein 1, PD-1) 是最受关注的免疫检查点之一, 其通常表达于活化的

T 细胞、自然杀伤性细胞、B 淋巴细胞和其他免疫细胞的表面. PD-1 与其在肿瘤细胞上高度表达的配体 PD-L1 相互作用后, 发生一定的构象变化, 介导胞内信号通路从而抑制 T 细胞的增殖、活化和细胞杀伤性功能<sup>[7-12]</sup>. 前人的研究已表明, PD-1 或 PD-L1 的基因敲除或抗体抑制可以增强小鼠免疫系统的抗肿瘤功能, 这表明阻断 PD-1 和 PD-L1 之间的相互作用可能为肿瘤免疫治疗提供一种有效的策略<sup>[13,14]</sup>.

PD-1/PD-L1 抗体抑制剂药物的研发目前取得非常显著的进展. 2014 年, 美国食品药品监督管理局批准了第一款 PD-1 抗体 Pembrolizumab 用于治疗晚期黑色素瘤之后, 一系列 PD-1/PD-L1 单克隆抗体被应用在了非小细胞肺癌、肝癌和食管胃交界癌等多种肿瘤疾病的临床治疗中<sup>[15-17]</sup>. 然

\* 国家自然科学基金 (批准号: 21903002) 和北京航空航天大学沈元学院卓越研究基金 (批准号: 230121202) 资助的课题.

† 通信作者. E-mail: [linxbseu@buaa.edu.cn](mailto:linxbseu@buaa.edu.cn)

而,随着应用的不断深入,抗体半衰期长、慢性免疫毒性、组织渗透有限、存储和运输成本高等难以避免的缺点逐渐暴露出来<sup>[18-20]</sup>.另一方面,抗体的同质化研究过多,造成了较大的资源浪费.小分子抑制剂具备较好的肿瘤组织渗透性、相对稳定的生物安全性和良好的口服利用性等优势,已成为 PD-1/PD-L1 抑制剂药物的下一个研发热点<sup>[18,21-23]</sup>.

百时美施贵宝 (Bristol-Myers Squibb, BMS) 于 2015 年公开了一系列非肽基联苯类小分子抑制剂,对 PD-1/PD-L1 结合具有强大的阻断抑制活性<sup>[24]</sup>. Holak 团队<sup>[25]</sup>曾报道, BMS 化合物通过与 PD-L1 结合诱导其发生二聚化,以间接的方式抑制 PD-1/PD-L1 相互作用.基于这一机制,其他的公司和学术团队开发了一系列不同骨架的衍生物,如 Incyte 公司的 INCB086550 (NCT04629339/NCT03762447)<sup>[26]</sup>、红日药业的 IMM-010 (NCT04343859)<sup>[27,28]</sup>、再极药业的 MAX-10181 (NCT04122339)、贝达药业的 BPI-371153 (NCT05341557)、歌礼药业的 ACS61 (NCT05287399) 以及和誉生物医药的 ABSK043 (NCT04964375) 等化合物目前已进入到了临床试验阶段.由于目前 PD-L1 小分子抑制剂的市场空白,寻找更多样骨架且具备良好用药性质的化合物仍具备重要意义.

数据驱动的计算方法已成为药物开发的重要工具, PD-1/PD-L1 小分子抑制剂亦不例外<sup>[29]</sup>.基于结构的方法,如药效团分析、分子对接和分子动力学 (molecular dynamics, MD) 模拟,在先前的研究中被广泛应用于靶向 PD-L1 二聚体的小分子抑制剂的虚拟筛选<sup>[29-33]</sup>.本文基于各种机器学习算法和分子描述符或指纹构建了一系列基于配体方法的分类和回归模型,以预测 ZINC15<sup>[34]</sup> 中类药物化合物对 PD-1/PD-L1 相互作用的抑制活性.具有高预测活性的化合物将继续通过药物相似性、药代动力学筛选并以分子对接和 MD 模拟进行基于结构方法上的验证,以最终获得具备 PD-L1 抑制潜力的小分子化合物.

## 2 研究方法

### 2.1 数据获取及整理

本文用于训练及测试的数据集来源于与 PD-L1 小分子抑制剂相关的 37 篇研究性论文及 16 项专利 (见补充材料表 S1 ([online](#))), 为避免因实验技

术手段不同而导致化合物对 PD-L1 抑制活性测定数据水平不一致的情况,本文仅收录了以半抑制浓度 (half-maximal inhibitor concentration,  $IC_{50}$ ) 为指标的均相时间分辨荧光 (homogeneous time-resolved fluorescence, HTRF) 的实验数据.

参考研究性论文及专利中的结构示意图,本文利用 ChemDraw 获取并检查化合物的简化分子线性输入规范 (simplified molecular input line entry system, SMILES) 字符串.  $IC_{50}$  不高于 1  $\mu\text{mol/L}$  的化合物定义为阳性样本 (即对 PD-L1 具备抑制活性), 而  $IC_{50}$  高于 10  $\mu\text{mol/L}$  的化合物定义为阴性样本 (即对 PD-L1 不具备抑制活性), 为减小用于分类模型构建的两类样本数量不平衡问题, 一项细胞水平的高通量筛选 (high throughput screening, HTS) 实验记录 (PubChem BioAssay AID: 2316) 部分数据被引入补充阴性样本. 最后, 仅具备明确  $IC_{50}$  测定值而非测定范围的化合物被收录于回归模型的数据集中,  $IC_{50}$  的对数转化值  $pIC_{50}$  (即  $-\lg IC_{50}$ ) 作为样本标签.

### 2.2 数据集聚类

为了确保机器学习模型尽可能均匀地获得数据集中不同结构的信息, 本文对数据集中的化合物先进行了聚类处理. 首先, 2130 个分类模型阳性样本和 1099 个回归模型样本分别转化为 600 和 350 种仅保留环形结构和连接环形结构的最短路径的 Mureko 骨架<sup>[35]</sup>. 分子骨架以 2 为最大半径, 2048 为向量长度转化为圆形扩展指纹 (extended-connectivity finger print, ECFP)<sup>[36]</sup> 后, 以平均为链接算法、欧氏距离为计算方式对两类骨架的 ECFP 进行分层聚类, 簇的数量由 5—20 之间对应的最佳轮廓系数决定, 后续模型训练将从簇内分层抽样进行数据集划分 (图 1).

### 2.3 分子表征

由 ChemDraw 得到的化合物 SMILES 字符串分别转化为三类分子描述符 (RDKit, PaDEL 1D&2D, Mordred) 和三类分子指纹 (ECFP, MACCS, PubChem) 以作为化合物的特征向量. 开源 Java 软件 PaDEL-Descriptor v2.0 用于计算 PaDEL 1D&2D 描述符<sup>[37]</sup>、MACCS 分子指纹<sup>[38]</sup> 和 PubChem 分子指纹<sup>[37]</sup>. RDKit 描述符和 ECFP 分子指纹<sup>[36]</sup> 均由化学信息处理程序包 RDKit 转化,

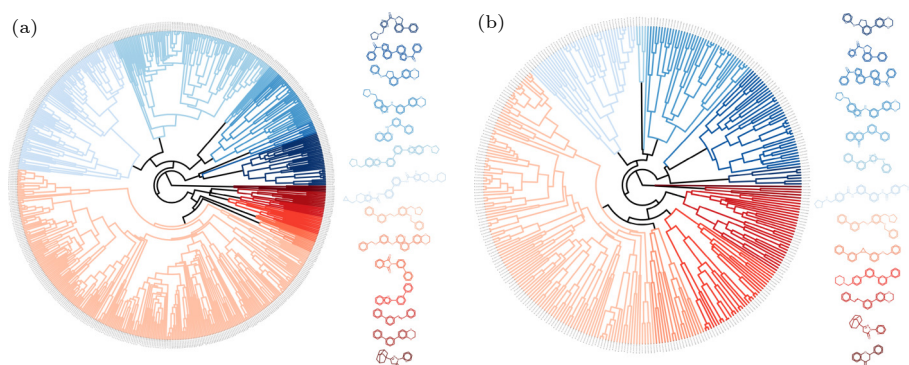


图1 数据集聚类 (a) 分类模型阳性样本骨架的 14 个簇; (b) 回归模型样本骨架的 13 个簇. 环形树状图的颜色代表不同的分子结构骨架簇, 颜色越接近, 骨架越相似, 一旁的化学结构图是每个簇中最具代表性的骨架

Fig. 1. Dataset clustering: (a) 14 clusters of scaffolds of active compounds in the classification models; (b) 13 clusters of scaffolds of compounds in the regression models. The colors of the circular dendrograms represent different clusters of scaffolds, and the closer the colors are, the more similar the scaffolds are. The chemical structure diagrams on the side are the most representative scaffolds of each cluster.

其中 ECFP 分子指纹为 2048 维向量, 指示最大以两个原子为半径的结构碎片存在与否. 此外, 描述符计算软件 Mordred<sup>[39]</sup> 被用于最后一类特征向量的转化.

## 2.4 机器学习

分别使用五种机器学习算法构建 PD-L1 小分子抑制剂分类模型和抑制活性  $IC_{50}$  预测回归模型, 其中逻辑回归 (logistic regression, LR)、邻近算法 (K-nearest neighbor, KNN)、支持向量机 (support vector machine, SVM)、随机森林 (random forest, RF) 和多层感知机 (multilayer perceptron, MLP) 用于分类任务模型构建, SVM、岭回归 (ridge regression)、高斯过程回归 (Gaussian process regression, GPR)、RF 和 MLP 用于回归任务模型构建. 两类模型数据集的 80% 为训练数据集, 另外 20% 为测试数据集. 所有分子描述符及分子指纹的特征值删除空缺之后, 基于训练集进行 min-max

标准化处理. 各类算法的最佳超参数由基于 Scikit-learn 网格搜索方法的五重交叉验证 (5-fold cross validation) 确定, 此过程中, 分类模型以马修斯相关系数 (matthews correlation coefficient, MCC), 回归模型以预测值的平均绝对误差 (mean absolute error, MAE) 为超参数选择标准.

## 2.5 模型评价标准

分类模型性能由灵敏度 (sensitivity, SE)、特异度 (specificity, SP)、准确度 (accuracy, ACC)、马修斯相关系数 (Matthews correlation coefficient, MCC) 进行泛化评估, 其计算公式如下:

$$SE = \frac{TP}{TP + FN}, \quad (1)$$

$$SP = \frac{TN}{TN + FP}, \quad (2)$$

$$ACC = \frac{TP + TN}{TP + FN + TN + FP}, \quad (3)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}, \quad (4)$$

其中, TP(true positive) 为分类模型判定为具备抑制活性且实验结果亦为具备抑制活性的化合物样本数, FP(false positive) 为分类模型判定为具备抑制活性但实验结果为不具备抑制活性的化合物样本数, TN(true negative) 为分类模型判定为不具备抑制活性且实验结果亦为不具备抑制活性的化

合物样本数, FN(false negative) 为分类模型判定为不具备抑制活性但实验结果为具备抑制活性的化合物样本数.

回归模型性能由平均绝对误差 (mean absolute error, MAE)、均方根误差 (root mean-square error, RMSE) 和决定系数 (correlation coefficient,  $R^2$ )



进行泛化评估, 其计算公式如下:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_{\text{true}} - Y_{\text{pred}}|, \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{\text{true}} - Y_{\text{pred}})^2}, \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_{\text{true}} - Y_{\text{pred}})^2}{\sum_{i=1}^n (Y_{\text{true}} - \bar{Y}_{\text{true}})^2}, \quad (7)$$

其中,  $n$  为回归模型待评估数据集样本数,  $Y_{\text{true}}$  为化合物  $\text{pIC}_{50}$  实验测定值,  $Y_{\text{pred}}$  为回归模型对化合物  $\text{pIC}_{50}$  的预测估计值.

## 2.6 药物相似性及 ADMET 检验

本文选取来自 ZINC15 数据库<sup>[34]</sup> 的 7400926 个可商业购买 (截至 2023 年 2 月 24 日)、电中性且收录三维结构信息的类药小分子作为候选化合物筛选池, 尽管这些分子已通过 ZINC15 的药物相似性检验, 但由于本文所筛选的化合物针对蛋白质相互作用, 传统的药物相似性指标 (quantitative estimate of drug-likeness, QED) 已经不再适用<sup>[40]</sup>. Kosugi 和 Ohue<sup>[41]</sup> 提出了一种更加适合于针对蛋白质相互作用抑制剂筛选的药物相似性指标 (quantitative estimate index for compounds targeting protein-protein interactions, QEPPi), 化合物的 QEPPi 分数为一个介于 0—1 之间的值, 数值越大代表化合物的蛋白质相互作用抑制剂药物相似性越高, 本文以 0.7 为阈值筛选高类药性化合物.

ADMET 代表化合物的吸收 (absorption)、分配 (distribution)、代谢 (metabolism)、排泄 (excretion) 和毒性 (toxicity) 等重要用药性质, 在早期药物筛选中十分重要, 本文采用 ADMETlab 2.0 对化合物进行 ADMET 筛选<sup>[42]</sup>.

## 2.7 分子对接

基于结构的活性预测方法需要蛋白质靶点的晶体结构信息, 本文选取的 PD-L1 二聚体结构来源于蛋白质数据库 PDB (ID: 7DY7). 本文选择 PD-L1 二聚体中的结合界面区域为对接结合口袋, 基于如下理由: 1) 前期的研究表明 PD-L1 二聚体

中的结合界面区域是很有潜力的药物结合位点<sup>[43,44]</sup>; 2) 分子对接结果显示小分子在该区域的结合亲和力最强.

分子对接利用 AutoDock Vina 进行<sup>[45]</sup>, AutoDockTools 将 PD-L1 二聚体转化为 pdbqt 格式, 对接网格盒子尺寸为  $40 \text{ \AA} \times 30 \text{ \AA} \times 40 \text{ \AA}$ , 中心坐标为  $(144.799 \text{ \AA}, -9.364 \text{ \AA}, 16.163 \text{ \AA})$ . 每个化合物各随机生成 50 种对接姿态和位置, 根据对接得分进行排名, 最佳得分前十位小分子对应的结合构象将用于分子动力学模拟的初始构象. 此外, 为验证通过上述流程得到的候选化合物作用机制是否与前人的研究结果相近, 本文利用 LigPlot+ 研究小分子与 PD-L1 二聚体的相互作用<sup>[46]</sup>.

## 2.8 分子动力学模拟

本文所有的分子动力学模拟采用 CHARMM 36 全原子力场<sup>[47,48]</sup>, 化合物配体的力场参数通过 CGenFF 程序获取<sup>[47-50]</sup>. 十个高对接分数的候选化合物和两个对照化合物 (BMS202 以及 INCB086550) 与 PD-L1 二聚体复合系统搭建于  $10 \text{ nm} \times 10 \text{ nm} \times 10 \text{ nm}$  盒子内, 每个盒子填充水分子并以 NaCl 中和体系电荷数, 整个搭建过程由 GROMACS<sup>[51]</sup> 工具 gmx solvate 和 gmx genion 实现. 所有的分子动力学模拟工作均使用 GROMACS 2019.6 程序包运行, 模拟步长为 2.0 fs, 采用等温等压 (NPT) 系综, 模拟时长为 100 ns. 模拟盒在  $x$  轴、 $y$  轴和  $z$  轴 3 个方向上均设定了周期性边界条件. 采用半各向同性的 Parrinello-Rahman<sup>[52]</sup> 方法将系统压强维持在 1 bar ( $1 \text{ bar} = 10^5 \text{ Pa}$ ), 压缩系数设定为  $4.5 \times 10^{-5}$ , 弛豫时间为 5 ps. 此外, 系统采用 Nose-Hoover<sup>[53,54]</sup> 控温方法将配体-蛋白质复合物和溶剂进行耦合, 温度维持在 310 K, 弛豫时间为 1 ps. 长距离静电相互作用使用 Particle-Mesh Ewald (PME)<sup>[55]</sup> 方法计算, 短距离静电相互作用的截止距离设置为 1.2 nm, LINCS (LINear constraint solver)<sup>[56]</sup> 算法用于约束含 H 原子的键长.

## 2.9 MM/PBSA 结合自由能计算

本文以 gmx\_MMPBSA<sup>[57]</sup> 为工具利用分子力学泊松-玻尔兹曼表面积法 (molecular mechanics Poisson-Boltzmann surface area, MM/PBSA) 来计算各个体系最后 10 ns 轨迹的平均结合自由能

( $\Delta G$ ), 其计算公式如下:

$$\Delta G = G_{\text{complex}} - (G_{\text{protein}} + G_{\text{ligand}}), \quad (8)$$

其中  $G_{\text{complex}}$  为蛋白质-配体复合物的自由能,  $G_{\text{protein}}$  和  $G_{\text{ligand}}$  分别为蛋白质和配体各自的自由能. 各组的自由能计算公式如下:

$$G = E_{\text{MM}} + G_{\text{sol}} - T\Delta S, \quad (9)$$

$$E_{\text{MM}} = E_{\text{vdw}} + E_{\text{ele}}, \quad (10)$$

$$G_{\text{sol}} = E_{\text{PB}} + E_{\text{SA}}, \quad (11)$$

其中  $E_{\text{MM}}$  为气相结合能量, 由范德瓦耳斯项  $E_{\text{vdw}}$  和静电项  $E_{\text{ele}}$  构成;  $G_{\text{sol}}$  为溶剂化自由能, 由极性  $E_{\text{PB}}$  和非极性  $E_{\text{SA}}$  贡献构成;  $T\Delta S$  为熵变, 因其计算成本较高且未必能够提高自由能的精度, 本文暂不对其进行计算.

### 3 研究结果

#### 3.1 分类模型性能

30 种分类模型分别基于 LR, KNN, SVM, RF 和 MLP 五种算法以及 RDKit, PaDEL 1D&2D, Mordred 分子描述符和 ECFP, PubChem, MACCS

分子指纹 6 种分子表征方式所建立. 图 2 为各个模型经五折验证并设置最佳超参数后在测试集上的性能表现, 其中以 ECFP 为输入的 KNN 模型表现出最佳性能, SE(阳性样本预测正确率) = 0.9937, SP(阴性样本预测正确率) = 0.9781, ACC(全部样本预测正确率) = 0.9859, MCC(综合评价两类样本预测性能指标) = 0.9720. 考虑到分子描述符或分子指纹的计算较为耗时, 同时获取 740 万余个类药小分子的全部六种输入向量的计算量更大, 因此, 仅采用以 ECFP 为输入用于后续的分类筛选. 另一方面, 相对于 LR 和 MLP 两种模型, KNN, SVM 和 RF 模型对于可能具备噪声数据的任务具有较好的鲁棒性, 因此, 选择 KNN, SVM 和 RF 模型作为分类筛选的模型; 当有两种或以上的模型支持小分子为活性时, 则认定该化合物对 PD-L1 具备抑制活性.

#### 3.2 分类模型解释

为了进一步探索输入向量的特征与分类之间的相关性, 本文用活性和非活性化合物之间的 RDKit, PaDEL 和 Mordred 三种描述符来分析基于特征重要性而构建的 RF 模型中权重最高的

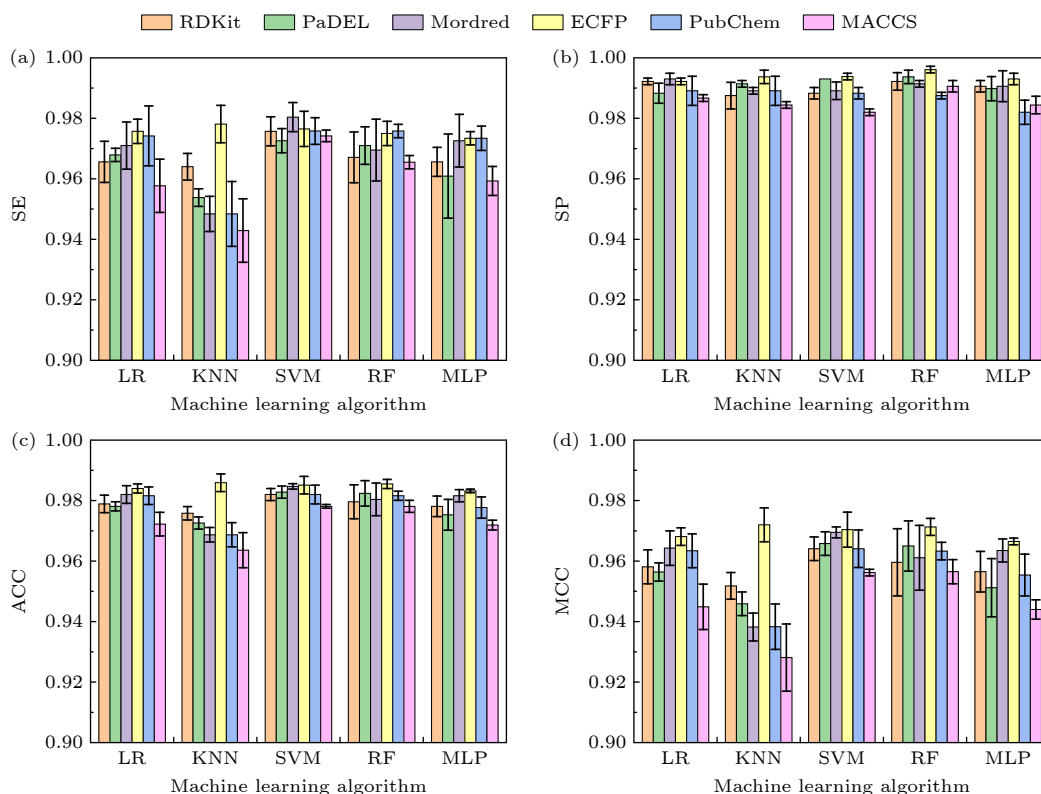


图 2 分类模型性能表现 (a) 灵敏度; (b) 特异度; (c) 准确度; (d) 马修斯相关系数

Fig. 2. Performance of binary models for classification tasks: (a) SE; (b) SP; (c) ACC; (d) MCC.

5个特征的分布差异(见补充材料图 S1 (online)). 尽管部分特征在两类样本间的分布差异是十分显著的,但由于分子描述符的特征信息难以解释,特征值根据理化性质及矩阵运算得出,其大小本身不具备具象含义,我们的认知只能停留在较浅薄的层面.

与描述符不同的是,分子指纹 ECFP 的位点指示局部亚结构的存在或不存,这可能揭示分子结构与对 PD-1/PD-L1 的抑制活性之间的关系. 分别统计了前十位活性化合物中比例显著高于非活性化合物的子结构和前十位活性化合物中比例显著低于非活性化合物的子结构(图 3). 带有黄色芳香性原子的环状结构在活性化合物中的计数远多于其在非活性化合物中的计数,这意味着芳香性结构片段可能对化合物的 PD-L1 抑制活性有正向贡献,相反,在非活性化合物中常出现的带灰色原

子的脂肪链片段则对活性没有正向贡献.

值得注意的是, ECFP985 和 ECFP1161 片段清楚地表示了化合物的联苯特性,这也是 BMS 化合物的核心结构特征之一 [24,58]. 此外,苯甲基 (ECFP253)、吡啶 (ECFP1453) 和与苯环相连的醚键 (ECFP1971) 通常存在于多数表现出 PD-L1 抑制活性的小分子结构中. 前人的分子对接分析表明,苯甲基能够与 PD-L1 的 Ala121 和 Met115 产生强烈的疏水相互作用 [59]. Lu 等 [60] 发现 PD-L1-BMS202 复合物中,吡啶环的氮原子周围有很大的空间,因此他们在该位点附加了一系列的取代基,以提高配体的结合亲和力. 由于该片段的重要性,许多其他研究团队也将工作重点放在吡啶的修饰上 [59,61,62]. 除了环状结构,连接环状结构的路径可能也是值得关注的组成部分,其中,以醚键连接六元环或五元环的活性化合物约占 50% [43,63-65].

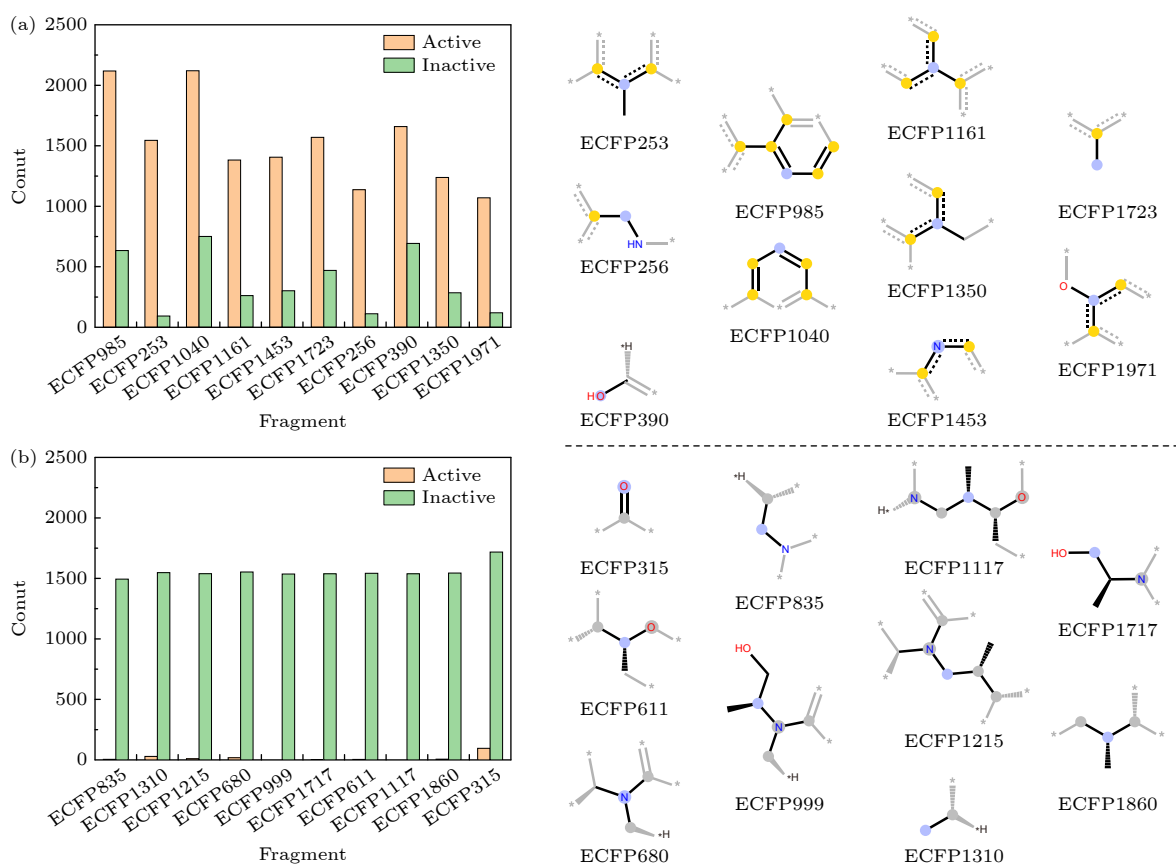


图 3 分子结构片段在两类样本间的计数差异 (a) 活性化合物计数占优的结构片段; (b) 非活性化合物计数占优的结构片段. 结构片段的中心原子以蓝色突出显示; 芳香性原子被着色为黄色, 而脂肪烃链原子则被着色为灰色

Fig. 3. Count difference of fragments of structures between active and inactive compounds: (a) Fragments in active compounds with a proportion higher than inactive compounds; (b) fragments in active compounds with a proportion lower than inactive compounds. The center atoms of substructure fragments are highlighted in blue; aromaticity atoms are colored in yellow and aliphatic hydrocarbon atoms are colored in gray.

### 3.3 回归模型性能

30种回归模型分别基于SVM, Ridge Regression, GPR, RF和MLP五种算法以及RDKit, PaDEL 1D&2D, Mordred分子描述符和ECFP, PubChem, MACCS分子指纹6种分子表征方式所建立. 图4为各个模型经五折验证后、最佳超参数设置下在测试集上的性能表现, 其中以ECFP为输入的SVM模型以MAE = 0.4503, RMSE =

0.6375, GPR模型以MAE = 0.4557, RMSE = 0.6375的性能表现显著优于其他模型. 图5展示了两种模型在训练集及测试集所有样本点的预测结果及偏差, 绝大部分的样本点预测偏差在 $\pm 1$ 范围内, 以ECFP为输入的SVM模型在测试上的决定系数 $R^2 = 0.782$ , 以ECFP为输入的GPR模型在测试上的决定系数 $R^2 = 0.787$ , 两类模型预测值的平均值将作为候选化合物的PD-L1抑制活性预测值.

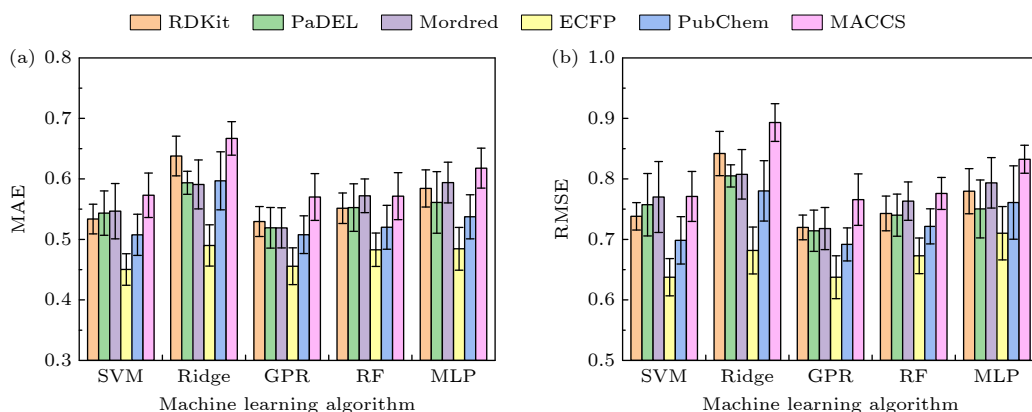


图4 回归模型性能表现 (a) 平均绝对误差; (b) 均方根误差

Fig. 4. Performance of continuous models for regression tasks: (a) MAE; (b) RMSE.

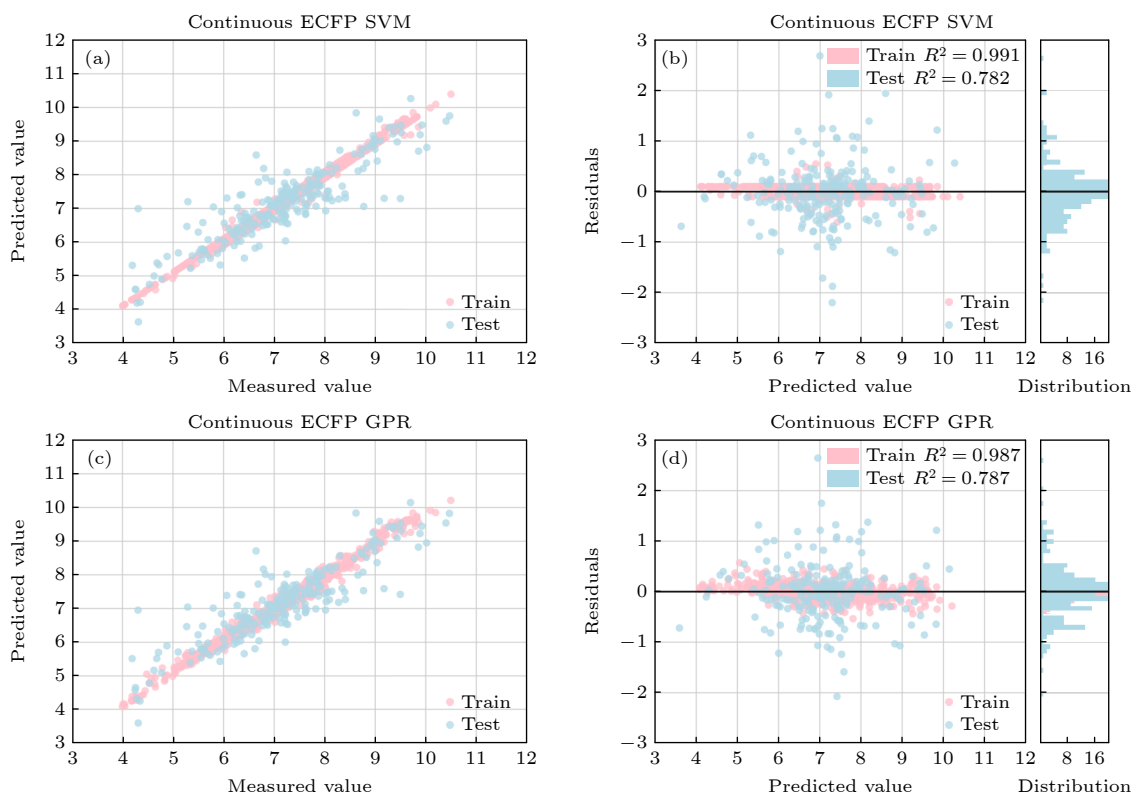


图5 两种最佳性能回归模型预测结果 (a), (c) 样本预测值与标签值分布; (b), (d) 样本预测偏差

Fig. 5. Prediction results of two best-performing continuous models: (a), (c) Distribution of predicted values and label values; (b), (d) prediction residuals.

### 3.4 虚拟筛选

ZINC15 数据库中电中性且收录三维结构信息的类药物小分子构成本文的化合物筛选池 (7400926 个小分子). 整个虚拟筛选过程包含 4 个环节: 1) 同时使用以 ECFP 为输入的 KNN, SVM 和 RF 等 3 种分类模型, 其中至少有 2 种判定结果为对 PD-L1 蛋白具有抑制活性, 则认定该分子具有抑制活性. 2) 同时使用以 ECFP 为输入的 SVM 和 GPR 回归模型预测 pIC<sub>50</sub> 值, 两模型 pIC<sub>50</sub> 的平均值小于 7 (IC<sub>50</sub> > 100 nmol/L) 的化合物将被剔除. 3) 计算小分子化合物的 QEPPi 分数 (见补充材料表 S2 (online)), 并保留 QEPPi 得分超过 0.7 的候选化合物. 尽管初始筛选池的化合物在 ZINC15 中根据 Lipinski 五原则<sup>[66]</sup> 被定义为类药化合物, 但其中部分化合物因分子量过小可能难以充分结合到 PD-L1 二聚体的相互作用界面, 因此, 重新评估他们的药物相似性仍具有重要意义.

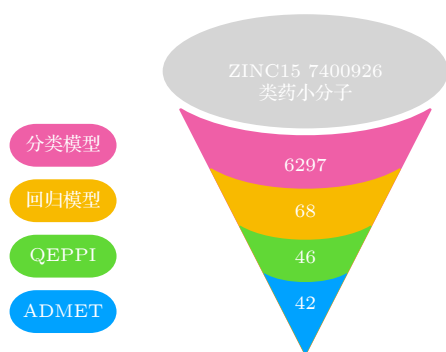


图 6 虚拟筛选流程

Fig. 6. Workflow of virtual screening.

义. 4) 使用 ADMETlab 2.0<sup>[42]</sup> 计算小分子的 ADMET 性质, 以进一步筛选具有应用潜能的小分子 (见补充材料表 S3 (online)). 最终, 通过整个虚拟筛选流程, 42 个候选化合物被认为对 PD-L1 具有较强的抑制活性并具备良好的药用性能 (图 6).

### 3.5 分子对接

为了验证筛选结果并通过基于结构的方法进一步识别有抑制潜力的化合物, 利用 AutoDock Vina 将 42 个候选化合物以及 BMS202 和 INCB-086550 对接到 PD-L1 二聚体的 IgV 结构域 (PDB ID:7DY7). 具有最低对接评分的 10 种化合物被认为有潜力的 PD-L1 抑制剂并继续用于之后的分析 (表 1), 其中值得注意的是, 10 种候选化合物对接分数均低于临床 II 期试验化合物 INCB086550 的对接分数, 因此可以认为先前构建的分类和回归模型是筛选 PD-L1 小分子抑制剂的有效工具. 分析 12 种化合物与 PD-L1 二聚体的相互作用可知 (见补充材料图 S2 (online)), TYR56(A), TYR56(B), MET115(A), MET115(B), ALA121(A) 和 TYR123(A) (括弧中 A 或 B 分别表示 PD-L1 二聚体中的单体 A 或 B) 等氨基酸残基与所有配体均发生了较强的疏水相互作用, 这意味着这些小分子的结合位置和结合模式可能是近似的. 此外, 除了 ZINC000021874692 和 ZINC000021874694 这一对化合物为旋光异构体, 其余化合物的结构差异较大, 相似度较低, 意味着其可为未来的湿实验筛选提供较高的容错空间 (见补充材料图 S3 (online)).

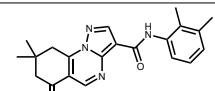
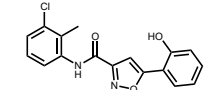
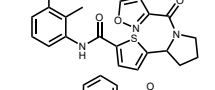
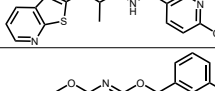
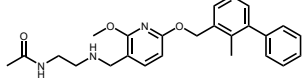
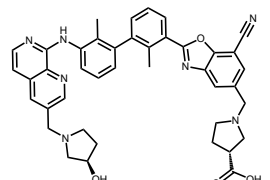
表 1 最低对接评分的 10 种候选化合物及 BMS202 和 INCB086550 的对接结果

Table 1. Docking results for the top 10 hits with BMS202 and INCB086550.

序号	化合物	化学结构式	对接分数/(kcal·mol <sup>-1</sup> )
Hit 1	ZINC000021723762		-11.8
Hit 2	ZINC000021874692		-10.7
Hit 3	ZINC000175468610		-10.7
Hit 4	ZINC000019770413		-10.6
Hit 5	ZINC000021874694		-10.6
Hit 6	ZINC000952973550		-10.5



表 1 (续) 最低对接评分的 10 种候选化合物及 BMS202 和 INCB086550 的对接结果  
Table 1 (continued). Docking results for the top 10 hits with BMS202 and INCB086550.

序号	化合物	化学结构式	对接分数/(kcal·mol <sup>-1</sup> )
Hit 7	ZINC000064987401		-10.5
Hit 8	ZINC000003908573		-10.4
Hit 9	ZINC000020538424		-10.4
Hit 10	ZINC000004063088		-10.3
Ctr 1	BMS202		-11.1
Ctr 2	INCB086550		-10.0

### 3.6 分子动力学模拟

为验证配体与 PD-L1 二聚体结合的稳定性, 计算了 12 种配体-蛋白质复合物 100 ns 轨迹中蛋白质骨架和配体小分子构象基于蛋白质骨架叠合的均方根偏差 (root mean square deviation, RMSD) (图 7). 所有体系的 PD-L1 二聚体骨架和配体的 RMSD 在 50 ns 后均稳定在 0.8 nm 以下, 模拟轨迹达到平衡, 蛋白质与配体结合稳定. 值得注意的是, 以 INCB086550 为例的分子量较大的小分子与以 ZINC000019770413 为例的分子量较小的小分子

相比, 配体构象 RMSD 波动较大. 由图 8 的蛋白质-配体结合模式可知, 起到关键作用的结构片段集中在配体的一端, 而另一端暴露于结合口袋之外. 若配体的分子量较大, 分子骨架较长, 其在结合区域外的部分更多, 这部分运动更为自由, 这可能是部分小分子构象 RMSD 波动相对较大的原因.

### 3.7 MM/PBSA 结合自由能计算

分子动力学模拟的最后 10 ns 轨迹被用于 MM/PBSA 计算以评估配体小分子与 PD-L1 二聚

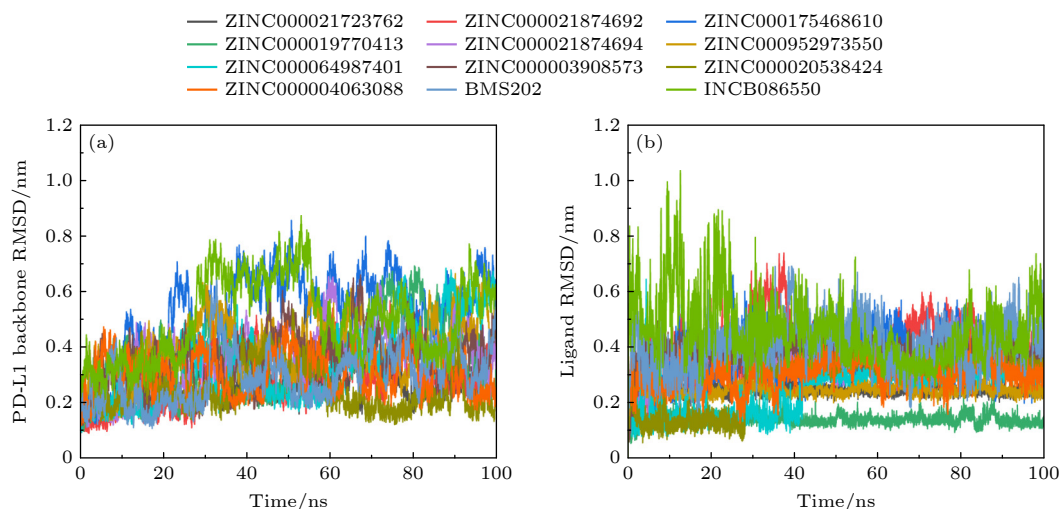


图 7 RMSD (a) PD-L1 二聚体骨架; (b) 配体  
Fig. 7. RMSD: (a) Backbone of PD-L1 dimer; (b) ligand.

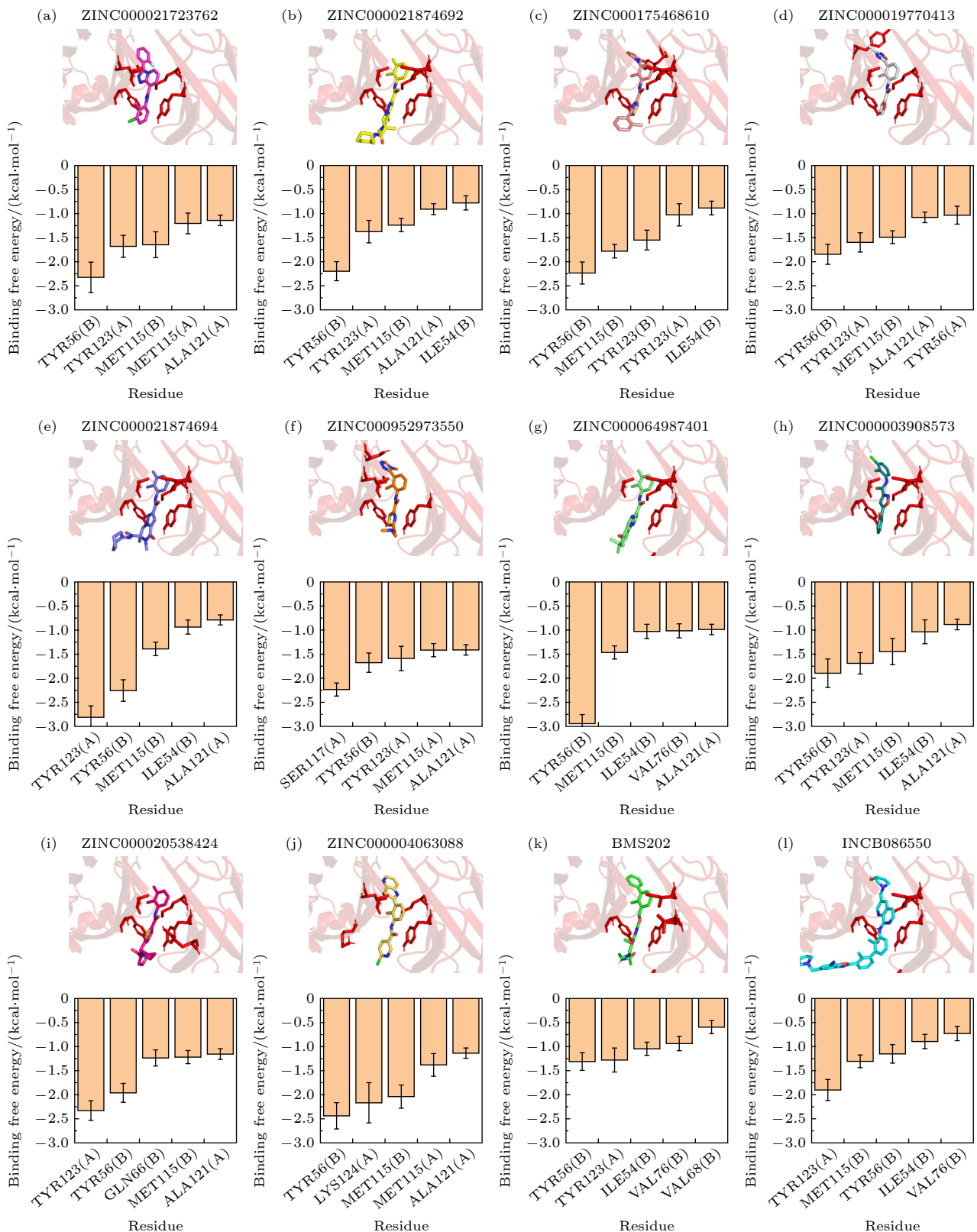


图 8 配体与 PD-L1 结合模式和关键残基 (a) ZINC000021723762; (b) ZINC000021874692; (c) ZINC000175468610; (d) ZINC000019770413; (e) ZINC000021874694; (f) ZINC000952973550; (g) ZINC000064987401; (h) ZINC000003908573; (i) ZINC000020538424; (j) ZINC000004063088; (k) BMS202; (l) INCB086550

Fig. 8. Ligand binding mode with PD-L1 and key residues: (a) ZINC000021723762; (b) ZINC000021874692; (c) ZINC000175468610; (d) ZINC000019770413; (e) ZINC000021874694; (f) ZINC000952973550; (g) ZINC000064987401; (h) ZINC000003908573; (i) ZINC000020538424; (j) ZINC000004063088; (k) BMS202; (l) NCB086550.

表 2 结合自由能计算结果  
Table 2. Results of MMPBSA.

化合物	$E_{vdw}/(\text{kcal}\cdot\text{mol}^{-1})$	$E_{ele}/(\text{kcal}\cdot\text{mol}^{-1})$	$E_{PB}/(\text{kcal}\cdot\text{mol}^{-1})$	$E_{SA}/(\text{kcal}\cdot\text{mol}^{-1})$	$\Delta G/(\text{kcal}\cdot\text{mol}^{-1})$
ZINC000021723762	-58.86 ± 0.12	-8.76 ± 0.10	35.53 ± 0.09	-4.03 ± 0.00	-36.12 ± 0.12
ZINC000021874692	-50.42 ± 0.11	-6.50 ± 0.14	35.65 ± 0.16	-4.50 ± 0.01	-25.78 ± 0.12
ZINC000175468610	-42.46 ± 0.09	-14.34 ± 0.08	36.20 ± 0.13	-4.02 ± 0.00	-24.61 ± 0.10
ZINC000019770413	-47.31 ± 0.08	-3.45 ± 0.08	26.32 ± 0.07	-3.48 ± 0.00	-27.91 ± 0.08
ZINC000021874694	-55.17 ± 0.09	-17.87 ± 0.13	45.67 ± 0.14	-4.59 ± 0.00	-31.97 ± 0.12
ZINC000952973550	-55.82 ± 0.08	-13.47 ± 0.11	45.83 ± 0.11	-4.19 ± 0.00	-27.65 ± 0.11
ZINC000064987401	-48.50 ± 0.08	-14.71 ± 0.12	38.59 ± 0.11	-4.06 ± 0.00	-28.68 ± 0.12
ZINC000003908573	-44.75 ± 0.07	4.59 ± 0.11	20.20 ± 0.13	-3.74 ± 0.00	-23.70 ± 0.12
ZINC000020538424	-49.50 ± 0.09	-8.50 ± 0.13	32.61 ± 0.16	-4.17 ± 0.01	-29.57 ± 0.10
ZINC000004063088	-64.14 ± 0.09	-5.38 ± 0.11	36.75 ± 0.10	-4.21 ± 0.00	-36.98 ± 0.10
BMS202	-40.23 ± 0.08	-3.57 ± 0.07	22.35 ± 0.10	-4.31 ± 0.01	-25.75 ± 0.09
INCB086550	-50.65 ± 0.14	-9.87 ± 0.17	44.44 ± 0.27	-5.13 ± 0.02	-21.22 ± 0.15

体间的结合作用强度. 由表 2 可知, 大部分候选化合物的结合自由能均与对照组的结合自由能处于同一水平, 而 ZINC000021723762, ZINC000021874694 和 ZINC000004063088 甚至显著优于对照组水平, 范德瓦耳斯相互作用为驱动这三类化合物区别于其他化合物的主要因素.

将结合自由能分解至每个氨基酸残基与配体分子的相互作用上, 提取了每个体系中对结合强度贡献度最大的 5 个关键残基. 如图 8 所示, TYR56(B), TYR123(A), MET115(B), ALA121(A) 在绝大部分体系中都起到关键作用, 这意味着 12 种配体结合于 PD-L1 二聚体相互作用交界面的模式是近似的. 前人研究揭示出, MET115(B) 和 ALA121(A) 与化合物中的芳香环能够发生强烈疏水相互作用, 而 TYR56(B) 能够与化合物的苯环形成  $\pi$ - $\pi$  堆积以增强结合稳定性<sup>[32,33,59,67]</sup>, 这些作用在结合自由能贡献中均得到了体现.

## 4 结论

通过从相关研究文献及专利收集 PD-L1 小分子抑制活性数据集, 并对数据集内化合物以结构相似性进行分层抽样, 本文构建了各 30 种基于不同分子表征方法和算法的机器学习活性判定分类模型和活性强度回归预测模型, 其中性能最佳分类模型分类正确率可达 98% 以上, 回归模型预测平均绝对误差在 0.5 以下, 具备良好的应用价值. 将以上模型并结合药用性质筛选工具构成的虚拟筛选方法应用于 ZINC15 大型小分子数据库, 获得了 42

种有潜力的 PD-L1 小分子抑制剂, 其中的 10 种通过分子对接、分子动力学模拟和结合自由能估计验证发现, 候选化合物的作用机制与对照化合物相近, 部分化合物的结合强度甚至优于对照化合物, 这意味着本文前期建立的机器学习模型是可以帮助加速 PD-L1 小分子抑制剂虚拟筛选的有效工具.

然而, 完整的药物研发并不是一个简单的工程, 本文仅使用计算方法帮助加速 PD-L1 小分子抑制剂的研发, 在完整的研发产业链中处于较上游的位置. 为尽可能使得本文的筛选结果得到更进一步的有效认证, 后续分子水平、细胞水平以及动物模型水平的湿实验验证仍是必不可少的, 计算机技术目前仅能扮演锦上添花的角色. 相信随着生物计算领域与生物技术领域的合作加深, 药物研发将能够向一个尽可能高效、低成本的方向发展.

## 参考文献

- [1] Waldman A D, Fritz J M, Lenardo M J 2020 *Nat. Rev. Immunol.* **20** 651
- [2] Wherry E J, Kurachi M 2015 *Nat. Rev. Immunol.* **15** 486
- [3] Zou W, Wolchok J D, Chen L 2016 *Sci. Transl. Med.* **8** 328rv4
- [4] Jiang X, Wang J, Deng X, Xiong F, Ge J, Xiang B, Wu X, Ma J, Zhou M, Li X, Li Y, Li G, Xiong W, Guo C, Zeng Z 2019 *Mol. Cancer* **18** 10
- [5] Topalian S L, Drake C G, Pardoll D M 2015 *Cancer Cell* **27** 450
- [6] Postow M A, Callahan M K, Wolchok J D 2015 *J. Clin. Oncol.* **33** 1974
- [7] Tang Q, Chen Y, Li X, Long S, Shi Y, Yu Y, Wu W, Han L, Wang S 2022 *Front. Immunol.* **13** 964442
- [8] Ai L, Xu A, Xu J 2020 *Adv. Exp. Med. Biol.* **1248** 33
- [9] Dermani F K, Samadi P, Rahmani G, Kohlan A K, Najafi R

- 2019 *J. Cell. Physiol.* **234** 1313
- [10] Parry R V, Chemnitz J M, Frauwirth K A, Lanfranco A R, Braunstein I, Kobayashi S V, Linsley P S, Thompson C B, Riley J L 2005 *Mol. Cell. Biol.* **25** 9543
- [11] Patsoukis N, Brown J, Petkova V, Liu F, Li L, Boussiotis V A 2012 *Sci. Signal.* **5** ra46
- [12] Hui E, Cheung J, Zhu J, Su X, Taylor M J, Wallweber H A, Sasmal D K, Huang J, Kim J M, Mellman I, Vale R D 2017 *Science* **355** 1428
- [13] Iwai Y, Ishida M, Tanaka Y, Okazaki T, Honjo T, Minato N 2002 *Proc. Natl. Acad. Sci. U.S.A.* **99** 12293
- [14] Iwai Y, Terawaki S, Honjo T 2005 *Int. Immunol.* **17** 133
- [15] Gong J, Chehrizi-Raffle A, Reddi S, Salgia R 2018 *J. Immunother. Cancer* **6** 8
- [16] Akinleye A, Rasool Z 2019 *J. Hematol. Oncol.* **12** 92
- [17] Shiravand Y, Khodadadi F, Kashani S M A, Hosseini-Fard S R, Hosseini S, Sadeghirad H, Ladwa R, O'Byrne K, Kulasingham A 2022 *Curr. Oncol.* **29** 3044
- [18] Zhang J, Zhang Y, Qu B, Yang H, Hu S, Dong X 2021 *Eur. J. Med. Chem.* **218** 113356
- [19] Johnson D B, Nebhan C A, Moslehi J J, Balko J M 2022 *Nat. Rev. Clin. Oncol.* **19** 254
- [20] Mould D R, Meibohm B 2016 *BioDrugs* **30** 275
- [21] Wu Q, Jiang L, Li S C, He Q J, Yang B, Cao J 2021 *Acta Pharmacol. Sin.* **42** 1
- [22] Zhan M M, Hu X Q, Liu X X, Ruan B F, Xu J, Liao C 2016 *Drug Discov. Today* **21** 1027
- [23] Liu C, Seeram N P, Ma H 2021 *Cancer Cell Int.* **21** 239
- [24] Chupak L S, Zheng X 2015 WO Patent 2015034820 A1
- [25] Zak K M, Grudnik P, Guzik K, Zieba B J, Musielak B, Dömling A, Dubin G, Holak T A 2016 *Oncotarget* **7** 30323
- [26] Koblisch H K, Wu L, Wang L S, Liu P C C, Wynn R, Rios-Doria J, Spitz S, Liu H, Volgina A, Zolotarjova N, Kapilashrami K, Behshad E, Covington M, Yang Y O, Li J, Diamond S, Soloviev M, O'Hayer K, Rubin S, Kanellopoulou C, Yang G, Rupar M, DiMatteo D, Lin L, Stevens C, Zhang Y, Thekkat P, Geschwindt R, Marando C, Yeleswaram S, Jackson J, Scherle P, Huber R, Yao W, Hollis G 2022 *Cancer Discov.* **12** 1482
- [27] Jiang J, Zou X, Liu Y, Liu X, Dong K, Yao X, Feng Z, Chen X, Sheng L, Li Y 2021 *Front. Pharmacol.* **12** 677120
- [28] Wang Y, Liu X, Zou X, Wang S, Luo L, Liu Y, Dong K, Yao X, Li Y, Chen X, Sheng L 2021 *Pharmaceutics* **13** 598
- [29] Sobral P S, Luz V C C, Almeida J, Videira P A, Pereira F 2023 *Int. J. Mol. Sci.* **24** 5908
- [30] Luo L, Zhong A, Wang Q, Zheng T 2021 *Mar. Drugs* **20** 29
- [31] Acúrcio R C, Pozzi S, Carreira B, Pojo M, Gómez-Cebrián N, Casimiro S, Fernandes A, Barateiro A, Farricha V, Brito J, Leandro A P, Salvador J A R, Graça L, Puchades-Carrasco L, Costa L, Satchi-Fainaro R, Guedes R C, Florindo H F 2022 *J. Immunother. Cancer* **10** e004695
- [32] Lung J, Hung M S, Lin Y C, Hung C H, Chen C C, Lee K D, Tsai Y H 2020 *Molecules* **25** 5293
- [33] Guo Y, Liang J, Liu B, Jin Y 2021 *Int. J. Mol. Sci.* **22** 10924
- [34] Sterling T, Irwin J J 2015 *J. Chem. Inf. Model.* **55** 2324
- [35] Bemis G W, Murcko M A 1996 *J. Med. Chem.* **39** 2887
- [36] Rogers D, Hahn M 2010 *J. Chem. Inf. Model.* **50** 742
- [37] Yap C W 2011 *J. Comput. Chem.* **32** 1466
- [38] Durant J L, Leland B A, Henry D R, Nourse J G 2002 *J. Chem. Inf. Comput. Sci.* **42** 1273
- [39] Moriwaki H, Tian Y S, Kawashita N, Takagi T 2018 *J. Cheminform.* **10** 4
- [40] Bickerton G R, Paolini G V, Besnard J, Muresan S, Hopkins A L 2012 *Nat. Chem.* **4** 90
- [41] Kosugi T, Ohue M 2021 *Int. J. Mol. Sci.* **22** 10925
- [42] Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, Yin M, Zeng X, Wu C, Lu A, Chen X, Hou T, Cao D 2021 *Nucleic Acids Res.* **49** W5
- [43] Jing T, Zhang Z, Kang Z, Mo J, Yue X, Lin Z, Fu X, Liu C, Ma H, Zhang X, Hu W 2023 *J. Med. Chem.* **66** 6811
- [44] Qin M, Meng Y, Yang H, Liu L, Zhang H, Wang S, Liu C, Wu X, Wu D, Tian Y, Hou Y, Zhao Y, Liu Y, Xu C, Wang L 2021 *J. Med. Chem.* **64** 5519
- [45] Trott O, Olson A J 2010 *J. Comput. Chem.* **31** 455
- [46] Laskowski R A, Swindells M B 2011 *J. Chem. Inf. Model.* **51** 2778
- [47] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, Mackerell Jr. A D 2010 *J. Comput. Chem.* **31** 671
- [48] Yu W, He X, Vanommeslaeghe K, MacKerell Jr. A D 2012 *J. Comput. Chem.* **33** 2451
- [49] Vanommeslaeghe K, MacKerell Jr. A D 2012 *J. Chem. Inf. Model.* **52** 3144
- [50] Vanommeslaeghe K, Raman E P, MacKerell Jr. A D 2012 *J. Chem. Inf. Model.* **52** 3155
- [51] Abraham M J, Murtola T, Schulz R, Páll S, Smith J C, Hess B, Lindahl E 2015 *SoftwareX* **1** 19
- [52] Parrinello M, Rahman A 1981 *J. Appl. Phys.* **52** 7182
- [53] Nosé S 1984 *Mol. Phys.* **52** 255
- [54] Hoover W G 1985 *Phys. Rev. A* **31** 1695
- [55] Essmann U, Perera L, Berkowitz M L, Darden T, Lee H, Pedersen L G 1995 *J. Chem. Phys.* **103** 8577
- [56] Hess B, Bekker H, Berendsen H J C, Fraaije J G E M 1997 *J. Comput. Chem.* **18** 1463
- [57] Valdés-Tresanco M S, Valdés-Tresanco M E, Valiente P A, Moreno E 2021 *J. Chem. Theory Comput.* **17** 6281
- [58] Chupak L S, Ding M, Martin S W, Zheng X, Hewawasam P, Connolly T P, Xu N, Yeung K S, Zhu J, Langley D R, Tenney D J, Scola P M 2015 WO Patent 2015160641 A3
- [59] Wang T, Cai S, Wang M, Zhang W, Zhang K, Chen D, Li Z, Jiang S 2021 *J. Med. Chem.* **64** 7390
- [60] Lu L, Qi Z, Wang T, Zhang X, Zhang K, Wang K, Cheng Y, Xiao Y, Li Z, Jiang S 2022 *ACS Med. Chem. Lett.* **13** 586
- [61] Dai X, Wang K, Chen H, Huang X, Feng Z 2021 *Bioorg. Chem.* **114** 105034
- [62] Le Biannic R, Magnez R, Klupsch F, Leleu-Chavain N, Thiroux B, Tardy M, El Bouazzati H, Dezitter X, Renault N, Vergoten G, Bailly C, Quesnel B, Thuru X, Millet R 2022 *Eur. J. Med. Chem.* **236** 114343
- [63] Russomanno P, Assoni G, Amato J, D'Amore V M, Scaglia R, Brancaccio D, Pedrini M, Polcaro G, La Pietra V, Orlando P, Falzoni M, Cerofolini L, Giuntini S, Fragai M, Pagano B, Donati G, Novellino E, Quintavalle C, Condorelli G, Sabbatino F, Seneci P, Arosio D, Pepe S, Marinelli L 2021 *J. Med. Chem.* **64** 16020
- [64] Liu L, Yao Z, Wang S, Xie T, Wu G, Zhang H, Zhang P, Wu Y, Yuan H, Sun H 2021 *J. Med. Chem.* **64** 8391
- [65] Cheng B, Ren Y, Cao H, Chen J 2020 *Eur. J. Med. Chem.* **199** 112377
- [66] Lipinski C A 2004 *Drug Discov. Today Technol.* **1** 337
- [67] Liang J, Wang B, Yang Y, Liu B, Jin Y 2023 *Int. J. Mol. Sci.* **24** 1280



## SPECIAL TOPIC—Machine learning in biomolecular simulations

## Virtual screening of drugs targeting PD-L1 protein\*

Lin Kai-Dong<sup>1)</sup> Lin Xiao-Qian<sup>1)2)</sup> Lin Xu-Bo<sup>1)†</sup>

1) (*Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, School of Engineering Medicine, Beihang University, Beijing 100191, China*)

2) (*Shen Yuan Honors College, Beihang University, Beijing 100191, China*)

( Received 29 June 2023; revised manuscript received 10 August 2023 )

## Abstract

Monoclonal antibody inhibitors targeting PD-1/PD-L1 immune checkpoints are gradually entering the market and have achieved certain positive effects in the treatments of various types of tumors. However, with the expansion of application, the limitations of antibody drugs and problems such as excessive homogenization of research gradually appear, making small-molecule inhibitors the new focus of researchers. This study aims to use ligand-based and structure-based binding activity prediction methods to achieve virtual screening of small-molecule inhibitors targeting PD-L1, thereby helping to accelerate the development of small molecule drugs. A dataset of PD-L1 small-molecule inhibitory activity from relevant research literature and patents is collected and activity judgment classification models with intensity prediction regression models are constructed based on different molecular featurization methods and machine learning algorithms. The two types of models filter 68 candidate compounds with high PD-L1 inhibitory activity from a large drug-like small molecule screening pool (ZINC15). Ten of these compounds not only have good drug similarities and pharmacokinetics, but also exhibit comparable binding affinities and similar mechanisms of action with previous reported hotspot compounds in molecular docking. This phenomenon is further verified in subsequent molecular dynamics simulation and the estimation of binding free energy. In this study, a virtual screening workflow integrating ligand-based method and structure-based method is developed, and potential PD-L1 small-molecule inhibitors are effectively screened from large compound databases, which is expected to help accelerate the application and expansion of tumor immunotherapy.

**Keywords:** PD-1/PD-L1, virtual screening, machine learning, molecular dynamics simulation

**PACS:** 05.10.-a, 02.70.-c

**DOI:** 10.7498/aps.72.20231068

\* Project supported by the National Nature Science Foundation of China (Grant No. 21903002) and the Excellence Research Fund of Shen Yuan Honors College, Beihang University, China (Grant No. 230121202).

† Corresponding author. E-mail: [linxbseu@buaa.edu.cn](mailto:linxbseu@buaa.edu.cn)

## 靶向PD-L1蛋白的计算机辅助药物筛选

林开东 林晓倩 林绪波

### Virtual screening of drugs targeting PD-L1 protein

Lin Kai-Dong Lin Xiao-Qian Lin Xu-Bo

引用信息 Citation: *Acta Physica Sinica*, 72, 240501 (2023) DOI: 10.7498/aps.72.20231068

在线阅读 View online: <https://doi.org/10.7498/aps.72.20231068>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

基于机器学习 $J_1$ - $J_2$ 反铁磁海森伯自旋链相变点的识别方法

Identifying phase transition point of  $J_1$ - $J_2$  antiferromagnetic Heisenberg spin chain by machine learning

物理学报. 2021, 70(23): 230701 <https://doi.org/10.7498/aps.70.20210711>

$U_{1x}Th_xO_2$ 混合燃料力学性能的分子动力学模拟

Effects of Th doping on mechanical properties of  $U_{1x}Th_xO_2$ : An atomistic simulation

物理学报. 2021, 70(12): 122801 <https://doi.org/10.7498/aps.70.20202239>

基于机器学习的无机磁性材料磁性基态分类与磁矩预测

Classification of magnetic ground states and prediction of magnetic moments of inorganic magnetic materials based on machine learning

物理学报. 2022, 71(6): 060202 <https://doi.org/10.7498/aps.71.20211625>

基于波动与扩散物理系统的机器学习

Machine learning based on wave and diffusion physical systems

物理学报. 2021, 70(14): 144204 <https://doi.org/10.7498/aps.70.20210879>

结合机器学习的大气压介质阻挡放电数值模拟研究

Numerical study of discharge characteristics of atmospheric dielectric barrier discharges by integrating machine learning

物理学报. 2022, 71(24): 245201 <https://doi.org/10.7498/aps.71.20221555>

机器学习辅助绝热量子算法设计

Machine learning assisted quantum adiabatic algorithm design

物理学报. 2021, 70(14): 140306 <https://doi.org/10.7498/aps.70.20210831>