

高通量虚拟筛选CDK2/Cyclin A2 靶点抑制剂

叶成濠¹, 梁 亨¹, 李恩民², 许丽艳², 李 鹏³, 陈广慧¹

(1. 汕头大学理学院化学系, 汕头 515063;

2. 汕头大学医学院基础医学研究所, 汕头 515041;

3. 香港中文大学(深圳)生命与健康科学学院, 深圳 518172)

摘要 CDK2/Cyclin A2复合蛋白的异常表达与乳腺癌、口腔癌、食管鳞状细胞癌的发生密切相关. CDK2/Cyclin A2复合蛋白的活性位点不同于CDK2单体. 至今临床上尚无靶向此复合蛋白的药物分子. 针对CDK2/Cyclin A2复合蛋白, 以实验报道的10个抑制剂分子构建药效团模型, 通过药物体外药代动力学(ADME)、Docking、聚类分析、毒性预测, 从DrugBank, ChEMBL和TCM@Taiwan 3个数据库约90万组数据中进行高通量虚拟筛选, 进一步进行MD模拟、MM/PBSA结合自由能计算、能量分解和平均非共价作用(aNCI)分析, 筛选出3个抑制效果优于阳性实验药Roscovitine的先导分子: DrugBank-2004, DrugBank-583和ChEMBL-7122. 与CDK2蛋白相比, CDK2/Cyclin A2复合蛋白结合位点空间变大, 先导分子与Lys33, Asp86, Lys129和Asp145残基之间的排斥作用有所降低, 导致结合自由能更大.

关键词 CDK2/Cyclin A2; 分子模拟; 高通量虚拟筛选; 抑制剂; 自由能分解

中图分类号 O641

文献标志码 A

细胞周期与肿瘤的关系密切, 当细胞周期的调控机制受到破坏时, 正常细胞存在向肿瘤细胞转化的潜在风险^[1-3]. 细胞周期蛋白依赖性激酶(CDKs)隶属于丝氨酸/苏氨酸蛋白激酶, 目前已发现13个家族成员, 其中CDK2在真核细胞周期调控中发挥关键作用, 与癌症发生及发展密切相关^[4], 因此发现靶向CDK2的抑制剂成为研究热点^[5-7].

当CDK2与Cyclin A2以复合蛋白形式存在时, 其催化活性被激活, 调控细胞G1期到S期的转变^[8]. 研究发现, CDK2/Cyclin A2复合蛋白在乳腺癌、口腔癌、食管鳞状细胞癌异常高表达^[9,10]. Kontopidis等^[11]通过对比CDK2/抑制剂和CDK2/Cyclin A2复合物/抑制剂的结构, 发现CDK2/Cyclin A2的ATP结合位点中Lys33和Asp145残基发生0.1~0.2 nm的位移, 如图S1(本文支持信息)所示, 使复合蛋白ATP区域结合口袋发生改变, 导致针对CDK2的抑制剂失效, 说明CDK2和CDK2/Cyclin A2蛋白是两个不同的靶点. 因此, 精准靶向CDK2/Cyclin A2复合蛋白的抑制剂才能更好地抑制激活后的CDK2. 目前, 实验上已有针对CDK2/Cyclin A2复合蛋白的药物研究, 如NU6271^[12]和N-&-N1^[13]分子对复合蛋白表现出良好的抑制效果, 半抑制浓度(IC₅₀)分别为45和43 nmol/L; Bettayeb等^[14]融合了Meridianin和Variolins两种化合物的结构特征, 得到了针对复合蛋白具有较强抑制作用的化合物Meriolins, IC₅₀为11 nmol/L. 但由于上述抑制剂存在选择性差、溶解度低及毒性大等缺点, 至今尚无应用于临床的靶向CDK2/Cyclin A2的小分子药物.

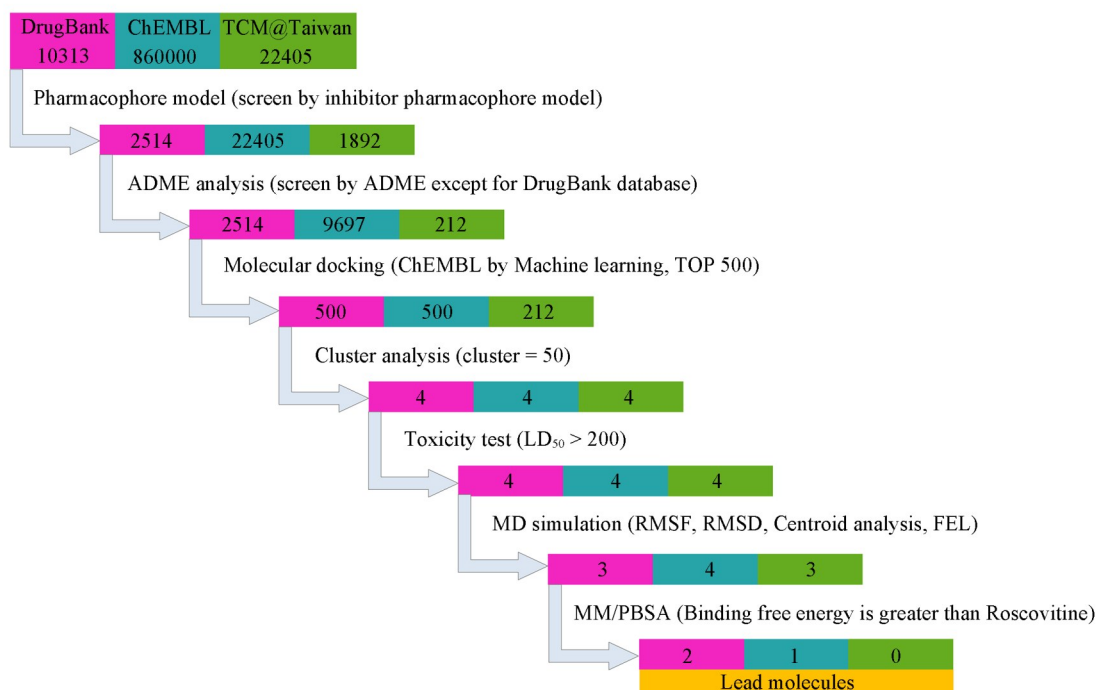
本文采用高通量虚拟筛选的方法, 旨在发现针对CDK2/Cyclin A2复合蛋白的小分子抑制剂. 基于DrugBank^[15], ChEMBL^[16]及TCM@Taiwan^[17]3个数据库近90万种小分子进行筛选, 采用药效团模型、药物体外药代动力学(ADME)计算、分子对接、机器学习、聚类、毒性预测以及分子动力学模拟方法, 发

收稿日期: 2021-06-15. 网络出版日期: 2021-09-17.

基金项目: 李嘉诚基金会交叉项目(批准号: 2020LKSFG07B)资助.

联系人简介: 陈广慧, 男, 博士, 教授, 主要从事药物分子及多孔材料高通量筛选与设计研究. E-mail: ghchen@stu.edu.cn

现对复合蛋白具有抑制作用的先导化合物,采用质心分析、自由能形貌图、MM/PBSA和平均非共价作用(aNCI)分析药物-蛋白的作用模式,筛选流程图如Scheme 1所示.



Scheme 1 Workflow of high-throughput screening of lead compounds on CDK2/Cyclin A2 target

1 计算方法

1.1 虚拟筛选

CDK2单体蛋白(PDB ID: 2A4L)^[18]和CDK2/Cyclin A2复合物蛋白(PDB ID: 3DDQ)^[19]从PDB Bank中获得,采用Swiss SPDBV 4.10^[20]对蛋白质晶体结构进行预处理;候选分子来自DrugBank(10313)^[15], ChEMBL(860000)^[16]和TCM@Taiwan(22405)^[17]数据库,如Scheme 1所示.首先,基于实验研究的10种靶向CDK2/Cyclin A2复合蛋白的抑制剂构建药效团模型,针对不同数据库进行筛选,并采用MOE 2015.2程序包^[21]完成.

其次,根据Lipinski规则将ChEMBL^[16]和TCM@Taiwan^[17]数据库中筛选得到的苗头化合物(Hit compound)进行ADME分析^[22].ADME是指药物在体内经过吸收、分布、代谢和排泄的药代动力学过程.筛选规则包括分子量 ≤ 500 ,氢键供体数目 ≤ 5 ,氮和氧原子数目总数 ≤ 10 以及分子的脂水分配系数(MolLogP) ≤ 4.15 .

再采用机器学习(Machine Learning, ML)模型,以DrugBank数据库分子与复合蛋白的对接打分数据作为训练集,预测ADME筛选后的ChEMBL数据的分子与靶点的对接打分.通过比较训练集和验证集的决定系数 R^2 来评估模型效果.将预测打分高的分子采用高精度的MMFF94力场进行优化并转为pdbqt格式的文件,并与靶点进行分子对接,采用AutoDock Vina程序^[23]完成.

然后,采用Rdkit程序^[24]建立机器学习聚类模型,处理由3个不同数据库筛选得到的分子.目的是使簇内药物相似度最大,簇与簇之间差异最小.聚类中心簇数目设置为50,以分子谱图作为特征,采用K-Means方法进行聚类.

最后,对筛选分子的毒性进行评估.通过ProTox-II服务器(http://tox.charite.de/protox_ii)^[25,26]进行,该服务器基于机器学习的方法,采用33种模型定量和定性地评估化合物的致突变性、致癌性、半数致死量(LD_{50})值和免疫毒性.

1.2 分子动力学模拟

为了验证虚拟筛选得到的苗头化合物与蛋白质的结合稳定性,对体系进行分子动力学(MD)模拟.首先,采用GAFF力场描述筛选得到的药物分子^[27,28],在B3LYP泛函和6-31G(d,p)基组的水平上,计算配体分子的约束静电势(ESP),拟合出原子电荷,采用Amber99 sb-ILDN力场描述蛋白质^[29];随后,以蛋白质-配体复合物为中心构建一个0.8 nm的TIP3P水模型盒子并设置周期性边界条件,为了保持体系的中性,使用钠离子和氯离子代替部分水分子以产生0.15 mol/L NaCl的溶剂盒子;然后,为防止在平衡模拟过程中发生剧烈重排,对体系进行能量极小化计算以及100 ps的限制性动力学模拟;最后,采用恒温恒压(NPT)系综,模拟温度为310 K,压强为1.0×10⁵ Pa,对体系进行50 ns的MD模拟.能量极小化以及MD模拟过程中均通过LINCS算法对所有氢键进行限制,并采用GROMACS 2018.4程序^[30]完成.

基于MD模拟的轨迹进行均方根波动(RMSF)、均方根位移(RMSD)和回旋半径(R_g)分析.RMSF能揭示蛋白质的各残基在模拟期间的平均位置的波动信息,采用下式计算:

$$\text{RMSF} = \sqrt{\frac{1}{T} \sum_{t_j}^T [x_i(t_j) - \bar{x}]^2} \quad (1)$$

式中: x_i 为原子的位置; T 为时间间隔; t_j 为特定的时刻; \bar{x} 为该原子的平均位置.

RMSD可以说明靶蛋白和配体之间结合的稳定性,计算公式如下:

$$\text{RMSD} = \sqrt{\frac{\sum d_i^2}{N_{\text{atoms}}}} \quad (2)$$

式中: d_i 为原子的位置; N 为系统中原子的数量.

R_g 值用于表征分子结合的紧密程度,可以采用下式计算:

$$R_g = \sqrt{\frac{\sum r^2 m}{\sum m}} \quad (3)$$

式中: r 为原子与其质心的距离; m 为一个原子的质量. R_g 越小,说明配体与靶点结合得越紧密.

对于分子的RMSD、 R_g 值和吉布斯自由能3个变量,采用GROMACS和xpm2txt.py脚本绘制三维自由能形貌图(Free energy landscape, FEL),其能量最低点代表体系的稳定构象.

1.3 MM/PBSA 结合自由能计算及能量分解

采用MM/PBSA方法计算蛋白与配体的结合自由能^[31,32].截取最后10 ns MD轨迹,每隔100 ps选1帧,共100帧进行结合自由能和能量分解计算.采用g_mmpbsa程序^[33]计算MM-PBSA结合自由能的具体公式如下:

$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \Delta G_{\text{protein}} - \Delta G_{\text{ligand}} \quad (4)$$

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S \quad (5)$$

式中: ΔG_{bind} 为结合自由能; $T\Delta S$ 为定温下复合物的熵变; ΔH 为体系的焓变,可通过下式表示:

$$\Delta H = \Delta E_{\text{MM}} + \Delta G_{\text{sol}} \quad (6)$$

式中: ΔE_{MM} 为复合物在气相下的焓变; ΔG_{sol} 为溶解自由能.将式(6)代入式(5)中得到下式:

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{sol}} - T\Delta S \quad (7)$$

由于配体与蛋白质的结合过程中熵变一般不大,在计算结合自由能的过程中可以忽略.因此式(7)可以简化为下式:

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{sol}} \quad (8)$$

式中: ΔE_{MM} 为van der Waals作用(ΔE_{vdW})和静电作用(ΔE_{ele})的总和,如下式所示:

$$\Delta E_{\text{MM}} = \Delta E_{\text{vdW}} + \Delta E_{\text{ele}} \quad (9)$$

式(8)中, ΔG_{sol} 等于极性组分溶剂化能($\Delta G_{\text{PB, sol}}$)和非极性组分溶剂化能($\Delta G_{\text{nonpolar, sol}}$)的总和.其中极性部分对自由能的贡献 $\Delta G_{\text{PB, sol}}$ 采用Poisson-Boltzmann(PB)方法计算;而非极性自由能的贡献 $\Delta G_{\text{nonpolar, sol}}$ 通过分子可及表面积(SA)计算:

$$\Delta G_{\text{sol}} = \Delta G_{\text{PB, sol}} + \Delta G_{\text{nonpolar, sol}} \quad (10)$$

最后, 将式(9)和式(10)代入式(8)中, 得到下式:

$$\Delta G_{\text{bind}} = \Delta E_{\text{vdM}} + \Delta E_{\text{ele}} + \Delta G_{\text{PB, sol}} + \Delta G_{\text{nonpolar, sol}} \quad (11)$$

可见, 采用MM/PBSA方法最终将结合自由能分解为以上4个部分.

1.4 平均非共价作用分析

aNCI分析是基于分子密度计算约化密度梯度函数(RDG), 从而描述配体与蛋白质之间的弱相互作用^[34,35]. 以MD模拟的最后一帧轨迹作为起始构象, 在相同条件下进行1 ns的MD模拟, 每隔2 ps取一帧共得到501帧. 由所得构象的分子密度得到平均密度($\overline{\rho(r)}$)和平均密度梯度($\overline{\nabla\rho(r)}$), 通过式(9)计算平均约化密度梯度(aRDG)^[35]:

$$\text{aRDG}(r) = \frac{1}{2(3\pi^2)^{1/3}} \frac{|\overline{\nabla\rho(r)}|}{\overline{\rho(r)}^{4/3}} \quad (12)$$

采用Multiwfn 3.8程序^[36]对蛋白质配体复合物进行aNCI分析, 通过等值面的形式揭示配体与蛋白的作用机制, 明确弱相互作用区域.

2 结果与讨论

2.1 CDK2/Cyclin A2抑制剂的虚拟筛选

选取已报道的10个靶向CDK2/Cyclin A2复合蛋白, 半抑制浓度在2~15600 nmol/L的抑制剂构建药效团模型^[13,14,19,37-41], 如表S1(本文支持信息)所示. 以RMSD为0.07 nm作为阈值, 对DrugBank(10313), ChEMBL(860000)和TCM@taiwan(22405)3个数据库进行虚拟筛选, 分别得到2514, 22405和1892个苗头化合物.

采用Rdkit程序^[24]通过Lipinski's规则对分子化合物的ADME性质进行评估, 再进一步筛选化合物. 发现ChEMBL和TCM@taiwan数据库中符合药效团模型的分子中分别有9697和212个满足ADME特征. 由于DrugBank数据库为FDA认证的数据库, 因此2514个分子被认为全部符合ADME要求.

将DrugBank和TCM@Taiwan数据库中2514和212个分子分别与CDK2/Cyclin A2靶点对接. 以DrugBank数据库中分子与靶点的对接数据建立机器学习模型, 预测ChEMBL数据库中9697个分子与靶点的对接打分. 以数据库中化合物的理化性质作为特征, 将DrugBank的数据以8:2分成训练集和验证集. 分别对随机森林^[42]、神经网络^[43]和XGBoost^[44]3种回归模型的预测效果进行评估, 结果如表S2(本文支持信息)所示. 发现XGBoost模型效果最佳, 在训练集和验证集中的 R^2 分别为0.97和0.93, 如图1所示. 故采用该模型预测ChEMBL top 500分子的打分并进行对接验证, 发现其对接打分绝对值均在19.1以上, 其中有95个分子在23.9以上, 如表1所示.

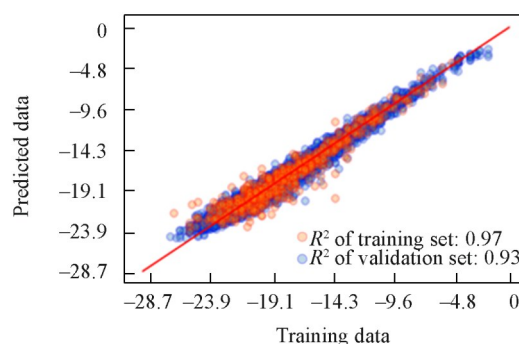


Fig. 1 Scatter plots of predicted and validated docking scores between molecules of DrugBank database and CDK2/Cyclin A2 target based on XGBoost machine learning model

The red dots represent training set data, the blue ones represent validation set data.

Table 1 Total counts of top 500 molecules predicted by XGBoost model and the number verified by docking in ChEMBL database

| Species | Docking score | |
|---|----------------|----------------|
| | (-19.1, -23.9) | (-23.9, -28.7) |
| Count of molecules by XGBoost predicted | 367 | 133 |
| Count of molecules by docking verified | 405 | 95 |

为了确保筛选得到的化合物具有化学多样性和代表性, 将每个数据库对接打分前 10% 的分子进行聚类分析. 根据 Morgan 指纹^[45]将分子划分成 50 个簇并将模型可视化, 如图 S2(本文支持信息)所示. DrugBank, ChEMBL 和 TCM@Taiwan 数据库分子的结构最大聚类距离分别为 35, 14 和 70, 差距较大. 因此, 将 3 个数据库分子分别聚类. 选取每个数据库中对接打分绝对值大于 25.1 的 top 4 苗头化合物共 12 个. 发现属于 DrugBank 数据库的 4 个分子的官能团种类较少, 结构相对简单; ChEMBL 数据库中的 4 个分子含有羟基、氰基、环丙基等取代基; TCM@Taiwan 数据库中的 4 个分子含有复杂的母环结构, 且官能团的数目较多, 表现出较大的结构多样性. 对 12 个苗头化合物进行毒性预测, 结果(表 S3, 本文支持信息)表明, 全部分子的 LD₅₀ 在 200~1500 mg/kg 区间内, 属于低毒性到中等毒性范围.

2.2 小分子与 CDK2/Cyclin A2 复合物的分子动力学模拟验证

对这 12 个分子以及阳性药 Roscovitine 与靶点形成的 13 个复合物体系分别进行 50 ns 的 MD 模拟, 通过 RMSF、RMSD、质心分析以及自由能形貌图(FEL)评估结合稳定性, 分别如图 2、图 S3 和图 S4(本文支持信息)所示. 由图 2(A)~(C)可见, 这 12 个复合物的 RMSF 与阳性药相近, 介于 0.05~0.5 nm 之间. 由图 2(D)~(F)可见, DrugBank-8482 和 TCM-37233 的 RMSD 较大, 分别为 1.0 和 0.8 nm; 而其余 10 个分子 RMSD 值与阳性药 Roscovitine 分子大致相同, 位于 0.15~0.5 nm, 且均已经达到平衡状态. 说明蛋白能与这 10 个配体形成稳定的复合物.

通过质心分析(Centroid analysis)研究不同配体与蛋白弱作用的区域. 由图 S3 可见, 经过 MD 模拟

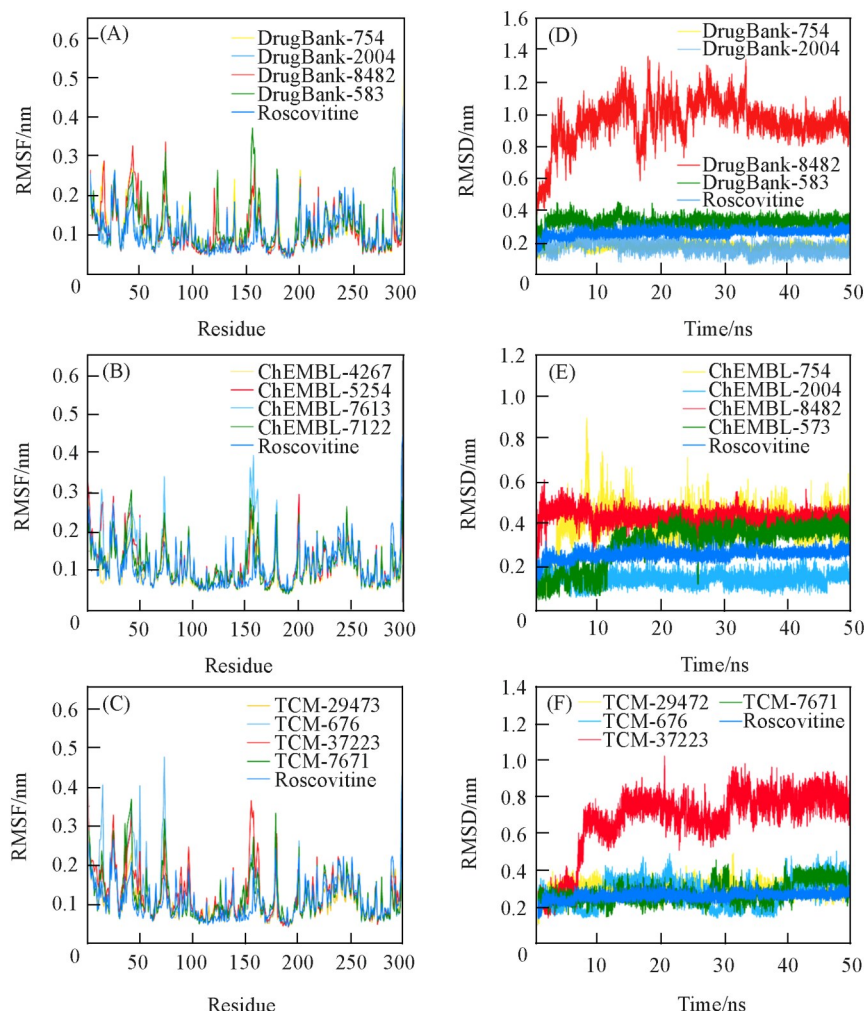


Fig. 2 RMSF(A—C) and RMSD(D—F) plots of complexes under MD simulations

(A, D) The top 4 hit molecules in the DrugBank database; (B, E) the top 4 hit molecules in the ChEMBL database; (C, F) the top 4 hit molecules in the TCM@Taiwan database.

后的配体均定位在同一结合位点,说明上述10个苗头化合物与 Roscovitine 和 CDK2/Cyclin A2 靶点的结合位点一致. 由图 S4 所示的自由能形貌图可见,这10个苗头化合物的吉布斯自由能最低点的 R_g 值均在 2.02~2.05 nm 之间,远小于阳性药物 Roscovitine 的 2.59 nm. 说明10个苗头化合物与蛋白质的结合相对紧密.

综合考虑 RMSF、RMSD、质心分析和自由能形貌图,发现潜在的10个对 CDK2/Cyclin A2 复合蛋白有抑制作用的分子,分别为 DrugBank-754, DrugBank-2004, DrugBank-583, ChEMBL-4267, ChEMBL-5254, ChEMBL-7613, ChEMBL-7122, TCM-29473, TCM-676 和 TCM-7671.

2.3 小分子与 CDK2/Cyclin A2 复合物的结合自由能的分解分析和 aNCI 分析

进一步采用 MM/PBSA 方法计算阳性药物 Roscovitine 以及上述10个候选分子与复合蛋白的结合自由能,发现 DrugBank-2004, DrugBank-583 和 ChEMBL-7122 的结合自由能分别为 -157.0, -145.3 和 -150.6 kJ/mol, 绝对值明显高于 Roscovitine 的 123.6 kJ/mol, 而其余7个分子结合自由能的绝对值均小于 Roscovitine. 如表2和图 S5(本文支持信息)所示. 因此在后面分析中只考虑上述这3个分子.

从表2可见,这3个分子与靶点的静电作用分别为 -76.7, -115.9 和 -142.0 kJ/mol, 绝对值均大于 Roscovitine 的 57.4 kJ/mol. 通过结构分析,发现 Roscovitine 上的母环是嘌呤基结构,容易与残基形成氢键作用. 形成氢键后,芳环的位置被固定,导致环上的取代基与部分残基间产生较大的排斥作用;而 DrugBank-2004, DrugBank-583 和 ChEMBL-7122 均含有多个无杂原子的芳环结构,不易与残基形成氢键,构型相对容易改变,因此与 CDK2/Cyclin A2 之间能产生较强的静电作用.

Table 2 Decomposition of binding free energy between hit molecules, the inhibitor Roscovitine and the target protein

| Compound | $E_{\text{vdW}}^a/(\text{kJ}\cdot\text{mol}^{-1})$ | $E_{\text{ele}}^b/(\text{kJ}\cdot\text{mol}^{-1})$ | $G_{\text{PB}}^c/(\text{kJ}\cdot\text{mol}^{-1})$ | $G_{\text{nonpolar}}^d/(\text{kJ}\cdot\text{mol}^{-1})$ | $G_{\text{bind}}^e/(\text{kJ}\cdot\text{mol}^{-1})$ |
|---------------|--|--|---|---|---|
| DrugBank-2004 | -212.0 | -76.7 | 151.8 | -19.8 | -157.0 |
| DrugBank-583 | -191.2 | -115.9 | 182.4 | -20.6 | -145.3 |
| ChEMBL-7122 | -215.3 | -142.0 | 226.1 | -19.1 | -150.6 |
| Roscovitine | -200.0 | -57.4 | 154.4 | -20.6 | -123.6 |

a. van der Waal interaction energy; b. electrostatic interaction energy; c. polar solvation energy; d. non-polar solvation energy; e. $G_{\text{bind}} = E_{\text{vdW}} + E_{\text{ele}} + G_{\text{PB}} + G_{\text{nonpolar}}$.

截取最后 10 ns MD 轨迹进行能量分解分析,考察每个残基对总能量的贡献. 发现3个先导分子和 Roscovitine 与复合蛋白具有相似的相互作用. 如图 3(A)~(H)所示,与 Roscovitine 类似, Ile10, Glu81, Gln85 和 Leu134 为3个先导分子与靶蛋白作用的主要残基. 值得注意的是,这3个先导化合物和 Roscovitine 与 Leu134 之间的作用自由能均很大,尤其 DrugBank-2004 分子,达到 -11.0 kJ/mol. 综上所述,3个先导分子较阳性药物 Roscovitine 具有更高的亲和力.

采用 MM/PBSA 方法进行残基能量分解,对比上述候选化合物及阳性药物 Roscovitine 与 CDK2 和 CDK2/Cyclin A2 蛋白作用的差异,结果如图3和图 S5 所示. 可见, Roscovitine 与 CDK2 的结合自由能较与 CDK2/Cyclin A2 蛋白的结合自由能大,这是由于 CDK2/Cyclin A2 结合口袋变大,降低了 Roscovitine 与残基间的亲合作用. 而与阳性药物不同,候选化合物与复合蛋白的结合自由能较与 CDK2 的结合自由能大,原因是 CDK2/Cyclin A2 结合口袋变大,削弱了配体与蛋白的 Lys33, Glu55, Asp86, Lys129 和 Asp145 残基的排斥作用.

为了直观地显示配体和受体之间的弱相互作用区域,对 DrugBank-2004, DrugBank-583, ChEMBL-7122 和 Roscovitine 与 CDK2/Cyclin A2 形成的复合物进行了 aNCI 分析. 由图 4(A)~(D)可见,上述4个分子与 CDK2/Cyclin A2 蛋白之间等值面的颜色主要是绿色和蓝色,表明配体与 CDK2/Cyclin A2 复合蛋白的 Ile10, Val18 和 Leu134 残基主要通过范德华力和氢键结合. 尽管3个候选分子与蛋白的相互作用类型和阳性药物 Roscovitine 一致,但范德华力和氢键作用明显较 Roscovitine 强.

综上所述,本文通过药效团模型、ADME、分子对接、聚类 and 毒性预测,从 DrugBank, ChEMBL 和 TCM@Taiwan 3 个数据库 90 万个分子中筛选出 12 个苗头化合物. 基于 MD 模拟和自由能形貌图结果,

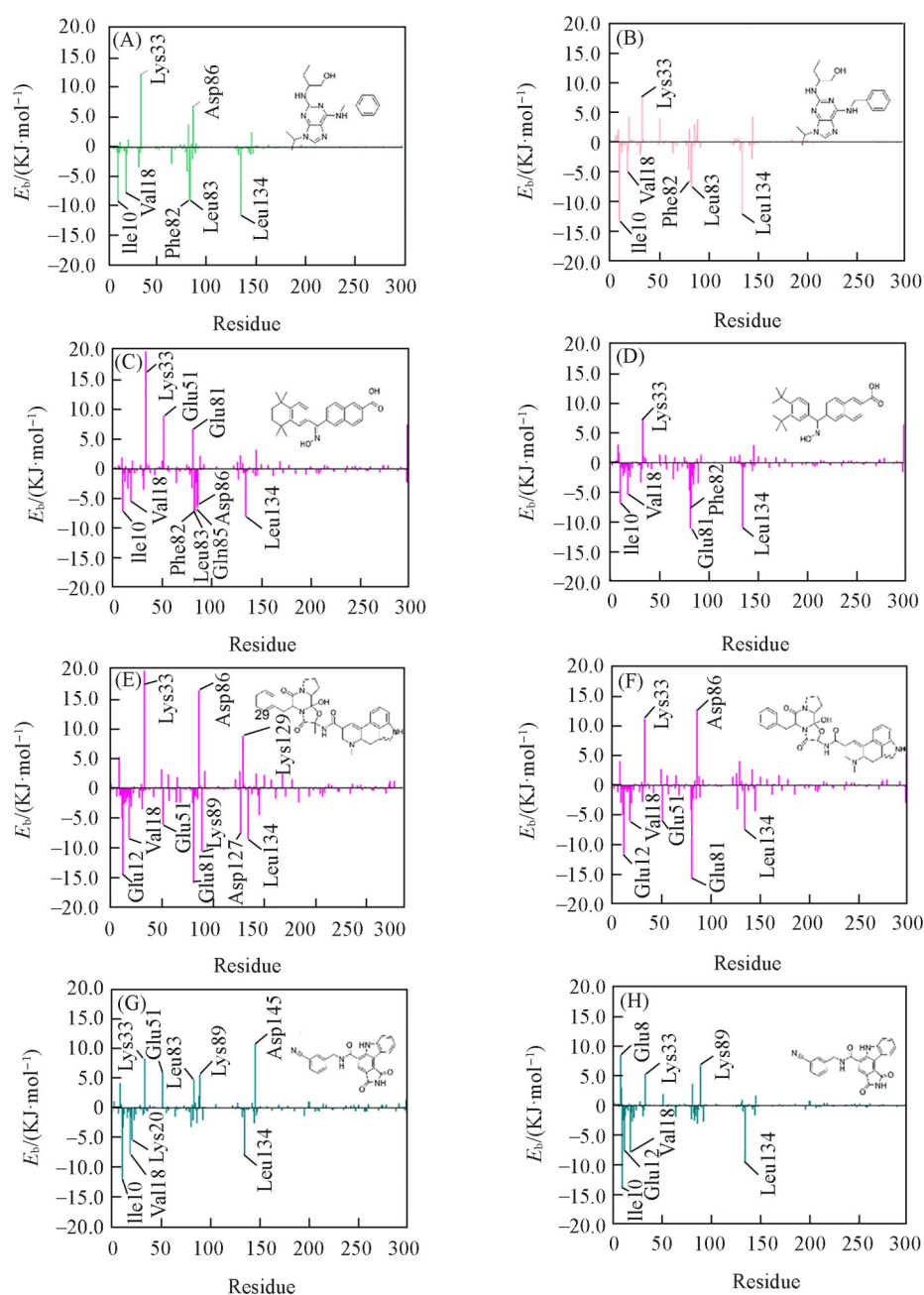


Fig.3 Decomposition of binding free energy on per-residue basis into contributions in each systems based on MM/PBSA method

(A) Known inhibitor Roscovitine-CDK2; (B) Roscovitine-CDK2/Cyclin A2; (C) DrugBank-2004-CDK2; (D) DrugBank-2004-CDK2/Cyclin A2; (E) DrugBank-583-CDK2; (F) DrugBank-583-CDK2/Cyclin A2; (G) ChEMBL-7122-CDK2; (H) ChEMBL-7122-CDK2/Cyclin A2. Insets: the structures of hit molecules.

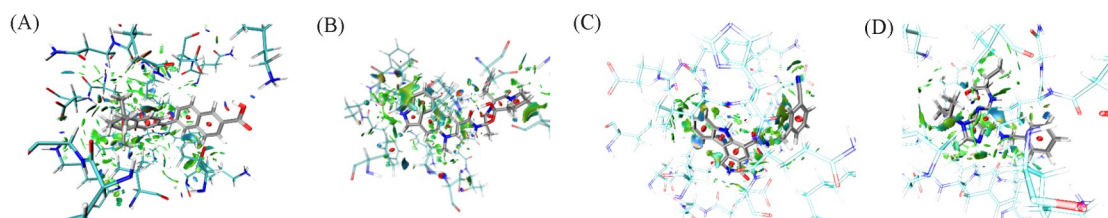


Fig.4 aNCI analysis between CDK2/Cyclin A2 and candidate drug molecules

(A) DrugBank-2004; (B) DrugBank-583; (C) ChEMBL-7122; (D) known inhibitor Roscovitine. Blue represents hydrogen bonding, green represents vDW interaction, red represents repulsion.

发现其中10个分子与CDK2/Cyclin A2复合物能稳定结合. 采用MM/PBSA方法发现其中3个先导化合物DrugBank-2004, DrugBank-583和ChEMBL-7122的结合自由能绝对值高于阳性药Roscovitine. 利用能量分解和aNCI对自由能贡献进行分析, 发现先导分子的结合自由能主要是由范德华力和静电作用构成. 由于先导化合物含有多个无杂原子的芳环结构, 不易与残基形成氢键, 有利于配体的构象转变. 因此与靶蛋白的静电作用比Roscovitine强, 提高了结合稳定性. 通过对比先导分子与CDK2/Cyclin A2和CDK2两种靶点的结合能力, 发现它们与CDK2/Cyclin A2形成更稳定的复合物. 这是由于CDK2/Cyclin A2的结合位点空间变大, 使配体分子与Lys33, Asp86, Lys129和Asp145残基之间的排斥作用减小. 本研究的高通量筛选工作能为发现针对CDK2/Cyclin A2靶点的药物研究提供理论依据.

支持信息见 <http://www.cjcu.jlu.edu.cn/CN/10.7503/cjcu20210405>.

参 考 文 献

- [1] Morgan D. O., *Nature*, **1995**, 374, 131—134
- [2] Martin A., Odajima J., Hunt S. L., Dubus P., Ortega S., Malumbres M., Barbacid M., *Cancer Cell*, **2005**, 7, 591—598
- [3] Bagella L., Sun A., Tonini T., Abbadessa G., Cottone G., Paggi M. G., de Luca A., Claudio P. P., Giordano A., *Oncogene*, **2007**, 26(13), 1829—1839
- [4] HL De Bondt J. R., Jancarik J., Jones H. D., Morgant D. O., Kim S. H., *Nature*, **1993**, 363, 595—602
- [5] Elgazwy A. S., Ismail N. S., Elzahabi H. S., *Bioorg. Med. Chem.*, **2010**, 18(21), 7639—7650
- [6] Xiao C., Sun C., Han W., Pan F., Dan Z., Li Y., Song Z. G., Jin Y. H., *Bioorg. Med. Chem.*, **2011**, 19(23), 7100—7110
- [7] Mahajan P., Chashoo G., Gupta M., Kumar A., Singh P. P., Nargotra A., *J. Chem. Inf. Model*, **2017**, 57(8), 1957—1969
- [8] Prehoda K. E., Lim W. A., *Current Opinion in Cell Biology*, **2002**, 14(2), 149—154
- [9] Kim S. J., Nakayama S., Miyoshi Y., Taguchi T., Tamaki Y. T., Matsushima Y., Torikoshi S., Tanaka T., Yoshida H., Ishihara S., Noguchi, *Annals of Oncology*, **2008**, 19(1), 68—72
- [10] Ravindranath M. H., Kelley M. C., Jones R. C., Amiri A. A., Bauer P. M., Morton D. L., *International J. Cancer*, **1998**, 75(1), 117—124
- [11] Kontopidis G. M., Pandalaneni C., McNae S. R., Gibson I., Mezna D., Thomas M., Wood M., Wang G., Walkinshaw S., Fischer M. D., *Chem. Biol.*, **2006**, 13(2), 201—211
- [12] Klon A. E., Lowrie J. F., Diller D. J., *ChemInform*, **2006**, 46(5), 1945—1956
- [13] Bettayeb K., Sallam H., Ferandin Y., Popowycz F., Fournet G., Hassan M., Echalié A., Bernard P., Endicott J., Joseph B., Meijer L., *Mol. Cancer Ther.*, **2008**, 7(9), 2713—2724
- [14] Bettayeb K., Tirado O. M., Marionneau-Lambot S., Ferandin Y., Lozach O., Morris J. C., Mateo-Lozano S., Drueckes P., Schachtele C., Kubbutat M. H., Liger F., Marquet B., Joseph B., Echalié A., Endicott J. A., Notario V., Meijer L., *Cancer Res.*, **2007**, 67(17), 8325—8334
- [15] Wishart D. S., Feunang Y. D., Guo A. C., Lo E. J., Marcu A., Grant J. R., Sajed T., Johnson D., Li C., Sayeeda Z., Assempour N., Iynkkaran I., Liu Y., Maciejewski A., Gale N., Wilson A., Chin L., Cummings R., Le D., Pon A., Knox C., Wilson M., *Nucleic Acids Res.*, **2018**, 46, 1074—1082
- [16] Mendez D., Gaulton A., Bento A. P., Chambers J., de Veij M., Felix E., Magarinos M. P., Mosquera J. F., Mutowo P., Nowotka M., Gordillo-Maranon M., Hunter F., Junco L., Mugumbate G., Rodriguez-Lopez M., Atkinson F., Bosc N., Radoux C. J., Segura-Cabrera A., Hersey A., Leach A. R., *Nucleic Acids Res.*, **2019**, 47, 930—940
- [17] Chen C. Y., *PLoS One*, **2011**, 6(1), e15939
- [18] Filgueira de Azevedo W. Jr., Leclerc S., Meijer L., Havlicek L., Strnad M., Kim S. H., *European J. Biochem.*, **1997**, 243(1/2), 518—526
- [19] Bettayeb K., Oumata N., Echalié A., Ferandin Y., Endicott J. A., Galons H., Meijer L., *Oncogene*, **2008**, 27(44), 5797—5807
- [20] Guex N., *Electrophoresis*, **1997**, 18(15), 2714—2723
- [21] Chemical Computing Group Inc., *Molecular Operating Environment(MOE)*, Version 2015, Montreal, QC, **2015**
- [22] Lipinski C. A., Lombardo F., Dominy B. W., Feeney P. J., *Advanced Drug Delivery Reviews*, **2012**, 64(1—3), 4—17
- [23] Trott O., Olson A. J., *J. Comput. Chem.*, **2010**, 31(2), 455—461
- [24] Llewellyn P. L., Bourrelly S., Serre C., Vimont A., Férey G., *Langmuir*, **2008**, 24(12), 7245—7250
- [25] Banerjee P., Dehnhostel F. O., Preissner R., *Front. Chem.*, **2018**, 6, 362
- [26] Banerjee P., Eckert A. O., Schrey A. K., Preissner R., *Nucleic Acids Res.*, **2018**, 46, 257—263
- [27] Silva A., Vranken W. F., *BMC Research Notes*, **2012**, 5, 367
- [28] Case K. B. D. A., Ben-Shalom I. Y., Brozell S. R., Cerutti D. S., Cheatham T. E. III, Cruzeiro V. W. D., Darden T. A., Duke R. E., Giambasu G., Gilson M. K., Gohlke H., Goetz A. W., Harris R., Izadi S., Liu Izmailov S. A., Kasavajhala K., Kovalenko A., Krasny R., Kurtzman T., Lee T. S., LeGrand S., Li P., Lin C. J., Luchko T., Luo R., Man V., Merz K. M., Miao Y., Mikhailovskii O.,

- Monard G., Nguyen H., Onufriev A., Pan F., Pantano S., Qi R., Roe D. R., Roitberg A., Sagui C., Schott-Verdugo S., Shen J., Simmerling C., Skrynnikov N. R., Smith J., Swails J., Walker R. C., Wang J., Wilson L., Wolf R. M., Wu X., Xiong Y., Xue Y., York D. M., Kollman P. A., *AMBER 2020*, University of California, San Francisco, 2020
- [29] Lindorff-Larsen K., Piana S., Palmo K., Maragakis P., Klepeis J. L., Dror R. O., Shaw D. E., *Proteins*, **2010**, 78(8), 1950—1958
- [30] Lindahl E., Hess B., van der Spoel D., *J. Mol. Modeling*, **2001**, 7(22), 306—317
- [31] Weis A., Katebzadeh K., Derhjelms P. S., Nilsson I., Ryde U., *J. Med. Chem.*, **2006**, 49, 6596—6606
- [32] Hou T., Wang J., Li Y., Wang W., *J. Comput. Chem.*, **2011**, 32(5), 866—877
- [33] Kumari R., Kumar R., C., *J. Chem. Inf. Model*, **2014**, 54(7), 1951—1962
- [34] Johnson E. R., Keinan S., Mori-Sanchez P., Contreras-Garcia J., Cohen A. J., Yang W., *J. Am. Chem. Soc.*, **2010**, 132(18), 6498
- [35] Contreras-Garcia J., Johnson E. R., Keinan S., Chaudret R., Piquemal J. P., Beratan D. N., Yang W., *J. Chem. Theory. Comput.*, **2011**, 7(3), 625—632
- [36] Lu T., Chen F., *J. Comput. Chem.*, **2012**, 33(5), 580—592
- [37] De Moliner E., Brown N. R., Johnson L. N., *Eur. J. Biochem.*, **2003**, 270(15), 3174—3181
- [38] Pratt D. J., Bentley J., Jewsbury P., Boyle F. T., Endicott J. A., Noble M. E. M., *J. Med. Chem.*, **2006**, 49(18), 5470—5477
- [39] Richardson C. M., Williamson D. S., Parratt M. J., Borgognoni J., Cansfield A. D., Dokurno P., Francis G. L., Howes R., Moore J. D., Murray J. B., Robertson A., Surgenor A. E., Torrance C. J., *Bioorg. Med. Chem. Lett.*, **2006**, 16(5), 1353—1357
- [40] Griffin R. J., Henderson A., Curtin N. J., Echalié A., Endicott J. A., Hardcastle I. R., Newell D. R., Noble M. E. M., Wang L., Golding B. T., *J. Am. Chem. Soc.*, **2006**, 128(18), 6012—6013
- [41] Traquandi G., Ciomei M., Ballinari D., Casale E., Colombo N., Croci V., Fiorentini F., Isacchi A., Longo A., Mercurio C., Panzeri A., Pastori W., Pevarello P., Volpi D., Roussel P., Vulpetti A., Brasca M. G., *J. Med. Chem.*, **2010**, 53(5), 2171—2187
- [42] Liaw A., Wiener M., *R News*, **2002**, 23(23), 18—22
- [43] Park D. C., El-Sharkawi M. A., Marks II R. J., Atlas Les E., Damborg M. J., *Middle East J. Entific Res.*, **1991**, 6, 442—449
- [44] Chen T., Guestrin C., *XGBoost: A Scalable Tree Boosting System*, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, **2016**, 785—794
- [45] Han X., Gross R. W., *Anal. Biochem.*, **2001**, 295(1), 88—100

High-throughput Virtual Screening of CDK2/Cyclin A2 Target Inhibitors[†]

YE Chenghao¹, LIANG Heng¹, LI Enmin², XU Liyan², LI Peng³, CHEN Guanghui^{1*}

(1. Department of Chemistry, Shantou University, Shantou 515063, China;

2. Medical Informatics Research Center, Shantou University Medical College, Shantou 515041, China;

3. School of Life and Health Sciences, The Chinese University of Hong Kong, Shenzhen 518172, China)

Abstract Abnormal expression of CDK2/Cyclin A2 protein is closely related to the occurrence of breast cancer, oral cancer, and esophageal squamous cell carcinoma. The binding pocket of CDK2/Cyclin A2 complex protein is different from that of CDK2 monomer. So far, no inhibitor that targeting CDK2/Cyclin A2 protein has been approved clinically. In order to obtain the high-efficiency inhibitor towards CDK2/Cyclin A2, the high-throughput virtual screening was performed among approximately 900000 molecules based on the three databases including DrugBank, ChEMBL and TCM@Taiwan. The 10 inhibitor molecules reported experimentally were used for pharmacophore models, to screen for hit molecules in above databases by means of docking, ADME, clustering analysis, toxicity prediction. Furthermore, 3 lead molecules including DrugBank-2004, DrugBank-583 and ChEMBL-7122 were screened out after molecular dynamics (MD) simulation, binding free energy calculation and average non-covalent interaction (aNCI) analysis, which are superior to Roscovitine because of the stronger electrostatic interaction. Moreover, as the result of the larger binding pocket than CDK2, the moderation of repulsion between lead molecules and target residues of Lys33, Asp86, Lys129 and Asp145 is achieved, and thus leading to greater binding free energies. The present work will provide a theoretical basis for the experimental research in the future.

Keywords CDK2/Cyclin A2; Molecular simulation; High-throughput virtual screening; Inhibitor; Free energy decomposition (Ed.: Y, K, S)

[†] Supported by the Li Ka Shing Foundation Cross-Disciplinary Research Program, China (No.2020LKSFG07B).