

• 专家论坛 •

CADD和AIDD的药物化学刍议

郭宗儒*

(中国医学科学院、北京协和医学院药物研究所, 北京 100050)

摘要: 人工智能辅助药物发现 (AIDD) 是计算机辅助药物发现 (CADD) 的新版本, 现代AI的学习能力极大地提升了CADD的能力。20世纪中叶, QSAR开拓了计算机辅助药物研究。在药物靶标结构信息未知的情况下, 用回归的方法揭示具有相似骨架结构的化合物中子结构与活性之间的定量关系, 以指导药物分子的设计。经过几十年的发展, CADD技术在基于靶标结构的药物发现 (SBDD)、基于片段的药物发现 (FBDD)、基于靶标结构的虚拟筛选 (SBVS) 等方面获得了广泛的应用。而AIDD的重大变革, 将显著提升计算机辅助药物设计、虚拟筛选和药物靶标发现的能力。本文从药物化学实践的角度, 以实例解读CADD和AIDD的关系。

关键词: 计算机辅助药物发现; 人工智能辅助药物发现; 诺氟沙星; 维奈托克; halicin

中图分类号: R914 文献标识码: A 文章编号: 0513-4870(2023)10-2931-11

Perspective of CADD and AIDD in medicinal chemistry

GUO Zong-ru*

(Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China)

Abstract: Artificial intelligence-aided drug discovery (AIDD) is a new version of computer-aided drug discovery (CADD). AIDD is featured in significantly promoting the performance of conventional CADD. AI markedly enhances the learning ability of CADD. In the 1960s, CADD was established from conventional QSAR approaches, which mainly used regression approaches to derive substructure-activity relationship for compounds with a common scaffold, and guide drug molecular design, figure out the binding features of drugs, and identify potential drug targets. Since the 1990s, structural biology has provided three-dimensional structures of drug targets, enabling drug discovery based on target structure (SBDD), fragment-based drug discovery (FBDD), and structure-based virtual screening (SBVS) with CADD approaches. In the past 30 years, many first in class (FIC) and best in class (BIC) drugs were discovered with CADD. Now, AIDD will further revolutionize CADD by reducing human interventions and mining big chemical and biological data. It is expected that AIDD will significantly enhance the abilities of CADD, virtual screening and drug target identification. This article tries to provide perspectives of CADD and AIDD in medicinal chemistry with case studies.

Key words: computer-aided drug discovery; artificial intelligence-aided drug discovery; norfloxacin; venetoclax; halicin

计算机的应用和互联网普及深刻地影响着人类社会, 促进了计算机化学的发展, 也促进了新药的创制。从20世纪60年代计算机的应用, 人们经历了数据输入

输出的穿孔纸带和穿孔卡片。个人计算机兴起, 从8位、16位、32位计算机, 发展到今天的64位多核个人计算机, 数据存储能力从最初的以千字节 (KB) 到今天常见的万亿字节 (TB)。日新月异的科学技术极大地改变了药物发现和分子设计的范式。如今, 人工智能辅助药物发现 (AIDD) 为计算机辅助药物发现 (CADD) 注

收稿日期: 2023-06-05; 修回日期: 2023-07-22.

*通讯作者 E-mail: zrguo@imm.ac.cn

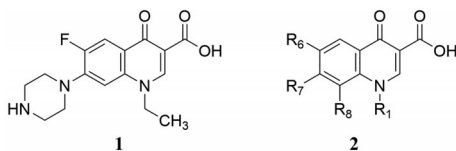
DOI: 10.16438/j.0513-4870.2023-0702

入了新的活力。本文通过解读若干用 CADD 方法发现的药物 (或进入临床研究的化合物), 从药物化学的角度分析 CADD 和 AIDD 的关系。

1 二维和三维定量构效关系

20 世纪 60 年代计算机用于药物研究, Hansch-藤田 稔夫开创了药物 (农药) 定量构效关系 (QSAR) 的研究, 他们用原子或基团的热力学参数作为表征化合物结构描述符 (descriptors), 以线性回归方法建立分子结构参数与活性 (或其他性质) 的回归方程, 以表征结构与活性的定量关系^[1]。虽然 Hansch 分析是性质 (电性、疏水性和立体性) 与活性 (药效或药代) 的相关性, 但这无疑是对既往 (定性的) 构效关系的突破, 因为事件以数学形式表达是科学进步的标志。蓬勃发展的 QSAR 大都属于回顾性的描述和结构的优化, 完全依赖 QSAR 创制的新药很少 (后面叙述)。与此同时, Free 和 Wilson 提出的数学模型是对系列分子的周边基团与活性作回归处理, 可视作以结构作为描述符, 但缺乏物化/生物学意义^[2]。QSAR 的本质是提取分子的特征向量 (即选出合适的描述符), 建立特征向量与生物活性的函数关系。

据笔者所知, 日本杏林公司研制的抗菌药, 拓扑异构酶抑制剂诺氟沙星 (**1**, norfloxacin), 是完全基于 Hansch-藤田模型的成功案例。在不断演化的 QSAR 方程指导下, 活性化合物与“未知靶标”的结合能力得到优化, 催生了诺氟沙星的诞生。



研制者古贺等^[3]对喹诺酮酸骨架 (**2**) 的不同位置作逐步变换, QSAR 结果指导下一轮设计。例如, 首轮 8 个化合物的构效方程, 揭示了 R_1 基团的最适 STERIMOL (描述抑制或基团的三维尺寸) 轴向长度参数为 4.2, 与乙基和环丙基相近。

$$\lg 1/\text{MIC} = -0.492(\pm 0.18)[L_{(1)}]^2 + 4.102(\pm 1.59)[L_{(1)}] - 1.999 \quad (1)$$

$$n = 8, r = 0.955, s = 0.126, F_{2,5} = 25.78, L_{(1)\text{opt}} = 4.2$$

式中 n 是样本数, r 是相关系数, s 是标准偏差, $F_{2,5}$ 是显著性检验, $L_{(1)\text{opt}}$ 是 STERIMOL 立体参数的 1 位取代基轴向最适距离。

经过对 6、7 和 8 位基团的 3 轮优化, 最终得到基于 71 个化合物数据的 QSAR 方程:

$$\begin{aligned} \lg 1/\text{MIC} = & -0.362(\pm 0.25)[L_{(1)}]^2 + \\ & 3.036(\pm 2.21)[L_{(1)}] - 2.499(\pm 0.55)[E_{s(6)}]^2 - \\ & 3.345(\pm 0.73)[E_{s(6)}] + 0.986(\pm 0.24)I_{(7)} - \\ & 0.734(\pm 0.27)I_{(7N-CO)} - 1.023(\pm 0.23)[B_{4(8)}]^2 - \\ & 3.724(\pm 0.92)I_{B_{4(8)}} - 0.205(\pm 0.05)[\Sigma\pi_{(6,7,8)}]^2 - \\ & 0.485(\pm 0.10)\Sigma\pi_{(6,7,8)} - \\ & 0.681(\pm 0.39)\Sigma F_{(6,7,8)} - 5.571 \quad (2) \end{aligned}$$

$$n = 71, r = 0.964, s = 0.274, F_{11,59} = 70.22$$

方程中的回归系数代表了不同位置基团的电性、疏水性或立体性对抗菌活性的影响权重, 研制出首创药物诺氟沙星, 1986 年美国 FDA 批准上市。后来研制活性更强的沙星类药物也都依据了 QSAR 方程的设计原则。

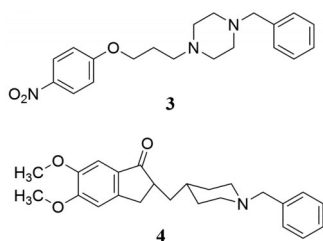
Hansch-藤田模型和 Free-Wilson 模型基于分子的拓扑结构, 又称 2D-QSAR。该模型未考虑分子的三维结构, 而且处理的样本为相同骨架的同系物分子, 骨架对活性的贡献隐含在回归方程的常数项中, 因而不能预测新的骨架, 只限于先导物的优化。后来出现的三维结构与活性的定量关系 (3D-QSAR) 方法, 其中 1988 年 Cramer III 等^[4]提出的“比较分子场方法” (CoMFA) 得到比较广泛的应用。

CoMFA 方法假定, 具有相同骨架的系列化合物以相同的方式与靶标结合, 它们的活性是药物与受体相互作用的力场函数, 因而各化合物的生物活性取决于分子场的差异。分子场包括静电场、立体场和疏水场等, 分别反映药物分子和靶标之间的相互作用的特性。CoMFA 分 3 个步骤: ① 确定药物分子的活性构象, 将同系列分子按照共同骨架叠合; ② 在叠合好的分子簇周围定义分子场的三维空间, 该空间按照一定的步长均匀划分成格点; 分别用表征分子场性质的探针 (分别为带电荷的、体积的、疏水的原子、离子或基团) 在格点上移行, 计算探针与分子中诸原子的相互作用能; ③ 通过偏最小二乘 (PLS) 方法建立化合物活性和分子场特征之间的关系, 一般用三维彩图表征这类化合物的电性、立体性和疏水性的三维构效关系^[4]。由 CoMFA 又演化出比较分子相似性方法 (CoMSIA)。CoMFA 和 CoMSIA 在 20 世纪 80/90 年代受到学术界的重视, 主要原因是关联三维结构与活性的描述符是近于连续、而且是自生成的, 能够清晰地勾勒出高活性化合物的结构特征。然而应用 CoMFA 和 CoMSIA 方法主要是描述性的, 也不能产生新的骨架, 难以准确预测高活性分子, 因此, 尚未发现单纯由 CoMFA 设计的新药报道。

2D/3D-QSAR 是在靶标结构未知情况下, 通过研究系列化合物的结构或物化性质与活性的定量依存关系,

揭示药物与靶标的作用特征, 借以研制高活性化合物, 所以可归纳为基于配体结构的药物设计。新世纪交替, 结构生物学的蓬勃发展, 解析了众多蛋白质和其配体复合物的三维结构。以计算机辅助基于靶标受体结构的药物发现 (SBDD) 逐渐替换了 2D/3D-QSAR 模式。

日本卫材公司研制治疗轻度或中度阿尔茨海默病症状的胆碱酯酶 (AChE) 抑制剂, 从普筛得到的苗头化合物 (**3**) 出发, 经传统药物化学方法的结构变换与 3D-QSAR (量子化学参数和分子形状比较) 的分析, 以及根据 AChE 与乙酰胆碱的复合物晶体结构, 将抑制剂作分子对接 (docking), 研制出多奈哌齐 (**4**, donepezil)。于 1996 年经 FDA 批准在美国上市^[5]。多奈哌齐被认为是 3D-QSAR 辅助药物设计的另一个成功案例。



多奈哌齐是 AChE 可逆性抑制剂, 与底物乙酰胆碱的结合模式有相当大的区别。图 1a 是多奈哌齐-AChE 复合物晶体结构图, 图 1b 是多奈哌齐与 AChE 的分子对接投影图, 二者对应非常相似。多奈哌齐与 AChE 由 5 个关键结合位点构成: ① 茛菪酮的苯环与 Trp279 形成 π - π 堆积作用 (KBI1); ② 甲氧基与 Arg289 (经结构水介导) 形成氢键结合 (KBI2); ③ 茛菪酮的羰基氧原子经水分子介导与 Phe288 氢键结合 (分子模拟的羰基与 Tyr70 或 121 有氢键结合) (KBI3); ④ 哌啶经质子化形成的铵离子没有像乙酰胆碱的季铵离子沉入底部与

Trp84 结合, 而是在“半途”与 Phe330 形成 N^+ - π 相互作用、与 Asp72 形成盐桥以及与 Tyr121 氢键结合 (KBI4); ⑤ 苯基沉入到峡道底部, 与 Trp84 发生 π - π 堆积作用 (KBI5)^[6]。

2 基于受体结构的药物发现

以互补性原理为基础的 SBDD 是继 QSAR 之后 CADD 的重要进展。用 X-射线衍射测定受体蛋白或复合物的晶体三维结构, 或用二维核磁共振解析蛋白质在溶液中的构象, 以及冷冻电子显微镜技术使结构生物学不断成熟, 成为药物发现的常规技术, 它们与基于片段的药物发现 (FBDD)、分子对接 (docking) 和分子动力学模拟 (MD simulation) 等技术结合, 增加 CADD 的能力, 促进了许多新药物分子实体 (NME) 的发现。表 1^[7-29] 列出一些代表性的上市药物、先导物 (或苗头分子) 和上市药物的化学结构 (限于篇幅略去分子结构的演化过程)。

下面简要叙述维奈托克 (**10**, venetoclax) 的结构演化过程。B 细胞淋巴瘤 (Bcl) 蛋白家族中包含有抗凋亡蛋白如 Bcl-2 和 Bcl-xL, 与促凋亡蛋白如 Bak、Bax 和 Bad 相互作用和制约, 精确地调控细胞行为。一些肿瘤为了避免和逃逸凋亡, 高表达 Bcl-2 或 Bcl-xL, 因而成为研制抗肿瘤药物的靶标。Bcl-xL 和 Bcl-2 的疏水性沟槽是结合促凋亡蛋白 Bak、Bax 和 Bad 的结合部位。该蛋白-蛋白相互作用的面积广泛 (750~1 500 Å²) 且表浅, 没有特征性结合位点。基于蛋白的三维结构, 采用基于片段设计方法 (SAR by NMR), 以确定片段性质及其定位。筛选了近万个化合物得到了 49 个 K_d 值 < 5 mmol·L⁻¹ 苗头化合物为第一结合位点, 其中 4'-氟-联苯-4-甲酸的二维核磁显示可引起 Bcl-xL 疏水沟槽内的 Gly94 和 Gly138 的 ¹⁵N 化学位移变化, 计算得出 K_d 为 300 μ mol·L⁻¹。第二结合位点筛选了 3 000 多个分子,

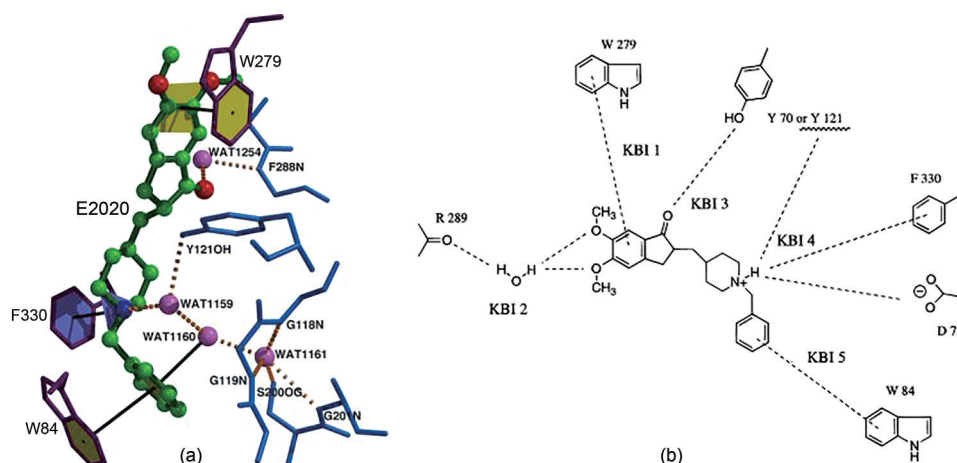
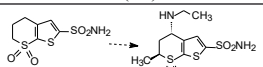
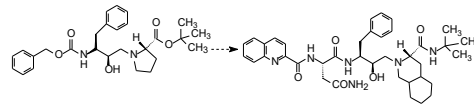
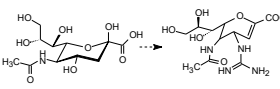
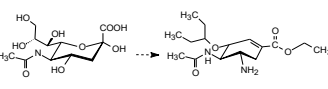
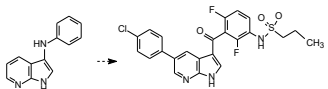
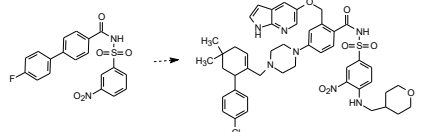
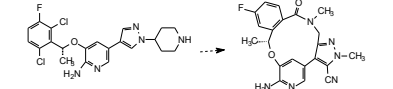
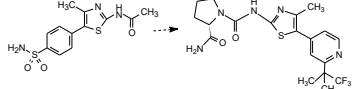
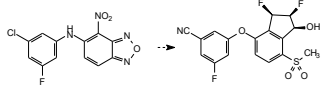
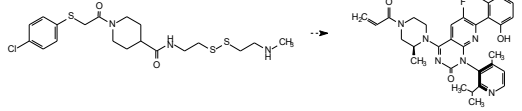
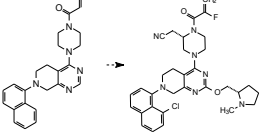


Figure 1 a: Co-crystal structure of donepezil and acetylcholinase; b: Schematic diagram of binding mode of donepezil to acetylcholinase through docking

Table 1 Some typical new molecular entities based on CADD

Drug	Name	Target	From lead (hit) to NME	Indication	Year	Ref.
5	Dorzolamide	Carbonic anhydrase		Glaucoma	1995	[7,8]
6	Saquinavir	HIV protease		AIDS	1997	[9]
7	Zanamivir	Sialidase		Influenza	1999	[10,11]
8	Oseltamivir	Sialidase		Influenza	1999	[12-14]
9	Vemurafenib	B-raf kinase		Melanoma	2012	[15,16]
10	Venetoclax	Bcl-xL		Chronic lymphocytic leukemia	2016	[17-19]
11	Lorlatinib	Anaplastic lymphoma kinase		NSCLC	2018	[20,21]
12	Alpelisib	PI3K		Advanced breast cancer	2019	[22]
13	Belzutifan	HIF-2α		VHL syndrome CCRCC	2021	[23,24]
14	Sotorasib	KRAS		NSCLC	2021	[25-27]
15	Adagrasib	KRAS		NSCLC	2022	[28,29]

得到 24 个 ($K_d < 5 \text{ mmol} \cdot \text{L}^{-1}$) 化合物, 例如萘酚和联苯等。二者经不同连接基得到 **10a**, $K_i = 1.4 \mu\text{mol} \cdot \text{L}^{-1}$, 但 NMR 显示与 Bcl-xL 结合并非最佳状态, 还发现有第三结合位点, 经设计合成得到 **10b**, $K_i = 36 \text{ nmol} \cdot \text{L}^{-1}$, 图 2 为 **10b** 与 Bcl-xL 结合图。

化合物 **10b** 是基于 Bcl-xL 结构设计的, 没有考虑对 Bcl-2 蛋白的抑制, 由于对人体多种高表达 Bcl-2 肿瘤抑制作用很弱, 以致不能阻止 Bcl-2 蛋白的抗凋亡作用, 所以抑癌谱较窄。研究者意识到当初对靶标可药性 (druggability) 的认识有局限性。

10b 高活性抑制 Bcl-xL, 但介质中若含有 1% 人血

清, 活性下降 69 倍, 含 10% 血清则完全失活。后来证明血清中白蛋白 III (HAS-III) 可与 **10b** 的酸性基团 $-\text{CO}-\text{NH}-\text{SO}_2-$ (拉电子的羰基与磺酰基使 NH 具有酸性) 结合, 降低了抑制 Bcl-xL 的活性。NMR 研究发现, **10b** 的第一片段氟代联苯基与 Bcl-xL 的结合和与同 HAS-III 作用环境不同, Bcl-xL 在氟端尚留有空间, 而且发生部分溶剂化, 而 HAS-III 结合的氟苯基被非极性残基完全包围, 没有空隙。提示该片段也可加入或变换为极性基团, 以阻止与 HAS-III 的结合^[30]。

为此, 变换氟代联苯基部分, 引入含有极性基团的碳环或碳链, 发现 **10c** 对 Bcl-xL 具有高抑制活性而

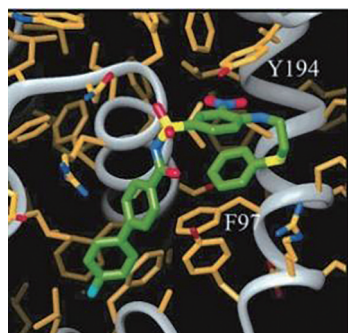
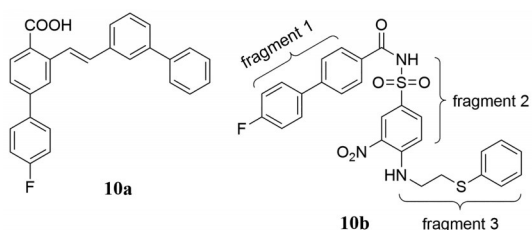


Figure 2 Binding mode of **10b** to Bcl-xL by NMR

HAS-III 不影响其作用。**10c** 可促进放射治疗或紫杉醇对非小细胞肺癌 (高表达抗凋亡蛋白 Bcl-xL) 的抑制, 表明有促凋亡作用。小鼠移植对多种细胞毒药物无效的人肿瘤 A549 细胞, **10c** 与紫杉醇合用, 抑制率达 60%~70%, 且未见增加毒性。

然而 **10c** 只是基于 Bcl-xL 结构设计的, 没有考虑对 Bcl-2 蛋白的抑制, 对人体多种高表达 Bcl-2 的肿瘤抑制作用很弱, 不能阻断 Bcl-2 蛋白的抗凋亡作用, 所以抑瘤谱窄。Bcl-xL 与 Bcl-2 蛋白序列的同源性虽然只有 49%, 但三维结构却很相似^[31], 例如两个蛋白都有疏水型沟槽, 只是 Bcl-2 的沟槽较宽和深, 这个区别为继续修饰化合物结构提供了着力点。为了提高抗肿瘤活性, 设定的新目标是对 Bcl-xL/Bcl-2 双靶标作用。

化合物 **10d** 是变换片段 1 时所合成的化合物, NMR 研究表明, **10d** 的苯乙基呈伸展型构象, 结合于 Bcl-xL 疏水沟槽; 而苯乙基与 Bcl-2 的结合则深入到疏水沟槽的深部, 埋入其中。这为设计双靶标抑制剂提供了修饰位置。图 3 是 **10d** 经 NMR 研究确定的与 Bcl-xL 和 Bcl-2 的结合模式。然而 **10d** 对 Bcl-xL/Bcl-2 的活性都不够强。

以 **10c** 的 4',4'-二甲基哌啶的二甲基基作为“把手”进

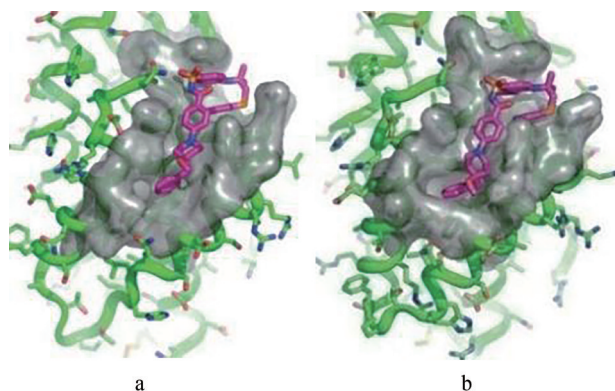
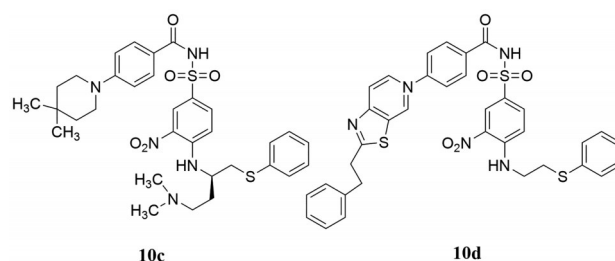


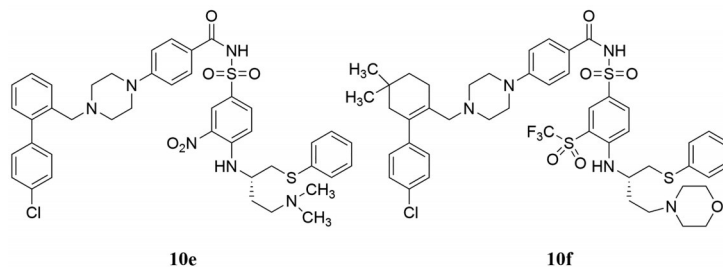
Figure 3 Binding of **10d** to Bcl-xL (a) and to Bcl-2 (b)

行基团变换或引长, 经多轮 SAR 优化, 得到化合物 **10e**, 对高表达 Bcl-xL 和 Bcl-2 的细胞活性 EC_{50} 分别为 0.018 和 $0.016 \mu\text{mol}\cdot\text{L}^{-1}$, 而结合 HAS-III 的作用很弱。然而 **10e** 的溶解性和药代性质有缺陷, 仍需优化。

下一步是优化物化和药代性质, 几经变换不同位置的基团, 得到了 **10f**, 对 Bcl-xL 细胞 $EC_{50} = 4.2 \text{ nmol}\cdot\text{L}^{-1}$, Bcl-2 细胞 $EC_{50} = 5.9 \text{ nmol}\cdot\text{L}^{-1}$, 对 HAS-III 的作用很弱, $AUC = 6.26 \mu\text{mol}\cdot\text{L}^{-1}\cdot\text{h}$ 。**10f** 确定为候选化合物, 代号为 ABT-263, 定名为 navitoclax, 进入临床研究^[32]。

Navitoclax (**10f**) 的 II 期临床显示对患者有抗肿瘤作用, 但出现血液毒性, 与临床前实验发现与剂量依赖性降低血小板相吻合, 是由于抑制了 Bcl-xL 蛋白的缘故。这个结果质疑了 Bcl-xL 可药性。下一步的结构变换是去除对 Bcl-xL 抑制作用, 只保留和提高抑制 Bcl-2 的活性。

10f 与 Bcl-xL 和 Bcl-2 沟槽中主要结合位点都是 P2 和 P4 疏水腔, 二者之差异难以区分。因而从药物化学的构效关系作优化探索。通过系统地除去或变换重要的结合基团, 得到去除苯硫基的化合物 **10g**, 对 Bcl-2



虽然失去了部分活性 ($K_i = 59 \text{ nmol}\cdot\text{L}^{-1}$), 但明显降低了抑制 Bcl-xL 作用 ($K_i = 5\,540 \text{ nmol}\cdot\text{L}^{-1}$), 提示可区分两个靶标蛋白。

化合物 **10g** 与 Bcl-2 的晶体结构显示, 片段 3 占据的 P4 空间变小, 降低了与 Bcl-xL 作用。另一个特征是由于 **10g** 与 Bcl-2 二聚体结合, 引起第 2 个 Bcl-2 蛋白的色氨酸残基 (Trp30) 嵌入到 **10g** 结合的 P4 腔内, 导致 Trp30 的吲哚环与硝基苯形成 π - π 叠合作用, 氮原子与 Bcl-2 的 Asp103 发生氢键结合。图 4 是 **10g** 与 Bcl-2 二聚体的晶体图, 绿色的吲哚环与硝基苯发生 π - π 叠合, 氮原子与 Asp103 发生氢键结合。为了模拟这个结合特征将吲哚环经醚键连接在母核苯环上, 化合物 **10h** 结合 Bcl-2 有高度选择性 (图 5), $K_i < 0.1 \text{ nmol}\cdot\text{L}^{-1}$, 与 Bcl-xL 结合的 $K_i > 660 \text{ nmol}\cdot\text{L}^{-1}$, 活性相差千倍。

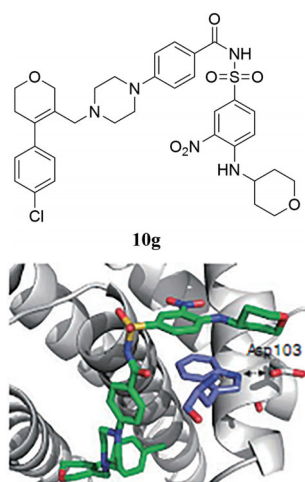


Figure 4 Crystallography of complex of **10g** with Bcl-2

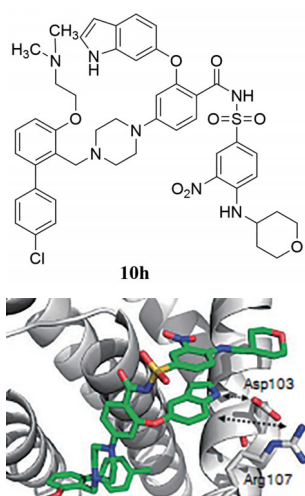
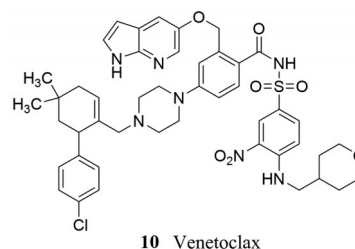


Figure 5 Crystallography of complex of **10h** with Bcl-2

进而将前述药代优化的片段加以整合 (干预蛋白-蛋白相互作用的分子结构作局部变换往往不影响整体的结合), 最终优化出化合物 **10**, 对 Bcl-2 高表达的

急性淋巴白血细胞 $\text{EC}_{50} = 8 \text{ nmol}\cdot\text{L}^{-1}$, 而对 Bcl-xL 高表达的 H146 细胞 $\text{EC}_{50} > 4\,000 \text{ nmol}\cdot\text{L}^{-1}$ 。 **10** 消除了抑制血小板的不良反, 小鼠灌胃 $100 \text{ mg}\cdot\text{kg}^{-1}$, $\text{AUC} = 2\,261 \mu\text{g}\cdot\text{h}\cdot\text{mL}^{-1}$, 血小板计数未见变化。 **10** 定为候选化合物, 命名为维奈托克, 经临床试验, 证明对 17 号短臂染色体缺失的慢性淋巴白血病有效, 于 2016 年 4 月 FDA 批准上市^[33]。



3 基于靶标结构的虚拟筛选和药物发现

基于靶标结构的虚拟筛选 (SBVS) 虽然也属分子对接范畴, 但与前述先导物优化的分子对接的目标和所处的阶段不同, SBVS 旨在发现先导物, 是研究的初始阶段, 因而有更多不确定性因素。虚拟筛选基于受体与配体互补性结合, 其成功的标志是实验证实的有预期活性的结果, 筛选出的结构新颖, 不同于已有的分子骨架, 而且是值得进一步研究的类药性分子等。

虚拟筛选常用的分子对接软件有商用的 GLIDE、DOCK 系列、AutoDock 系列和 GOLD 等。筛选的化合物库的规模为数百万到上亿个分子。分子对接找到化合物经打分函数表征和排序, 还要剔除假阳性化合物。Zhu 等^[34]最近总结分析了近 15 年来基于靶标结构的虚拟筛选的文献, 分析了不同对接软件的应用状况、靶标的分布, 以及在创制新药方面的作用。下面列举成功的或重要进展的实例简要解析虚拟筛选在新药创制特征。

3.1 抗 COVID-19 病毒药恩西曲韦

2019 年全球暴发冠状病毒的流行性感染 (COVID-19), 日本盐野义公司着手研制病毒增殖的重要蛋白酶 3CL^{pro} 抑制剂, 以阻断多聚蛋白的成熟而终止病毒的复制和繁殖。设定的研制目标是可口服非肽类非共价结合的小分子药物。研制策略是基于冠状病毒蛋白酶 3CL^{pro} 的结构, 经虚拟筛选以获得先导化合物。为进行分子对接的 3CL^{pro} 模板是使用 3CL^{pro} 与抑制剂 ML-188 (**11**) 复合物的晶体结构 (图 6), 图中的蓝圈表示 3CL^{pro} 发生氢键的接受体位置, 红圈为疏水基团, 作为药效团过滤 (pharmacophore filter) 的特征。对接的软件是 GLIDE, 用公司的数十万个化合物作分子对接, 并用药效团过滤器甄别, 将对接高打分的前 300 个苗头分子作抑酶实验, 同时经质谱分析以排除假阳性。最后确定

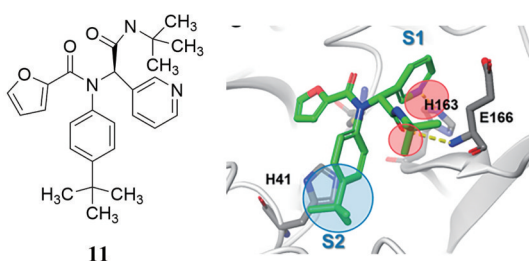


Figure 6 Crystallography of complex of **11** with 3CL^{pro}. The common H-bond acceptors are circled in red; the common hydrophobic features are circled in blue

了化合物 **12** 为先导物。

根据 **11** 与 3CL^{pro} 复合物的晶体结构, 将二氟甲基的位置与苯环环化, 变换成苯并吡唑, 与 Tyr26 发生 π - π 相互作用, 酰胺片段用三唑置换以与 His163 发生 π - π 相互作用 (三唑或咪唑可视作酰胺的等排结构), 再对个别基团微调, 优化出抗 COVID 药物恩西曲韦 (**13**, ensitrelvir), 于 2022 年 2 月在日本上市^[35]。图 7 是恩西曲韦与 3CL^{pro} 蛋白酶复合物的晶体图。

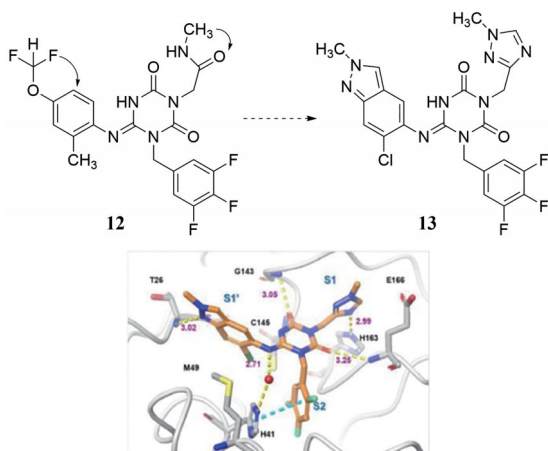


Figure 7 X-ray costructure of **13** and 3CL^{pro} protein

3.2 乙酰辅酶 A 羧化酶抑制剂福尔索司他

乙酰辅酶 A 羧化酶 (ACC) 是脂肪酸从头合成的限速酶, 已知有 ACC1 和 ACC2 两种亚型都与脂肪酸代谢密切相关。抑制 ACC 及其二聚化可阻止脂肪酸的体内合成, 促进脂肪酸氧化代谢, ACC 是研制治疗肥胖症、糖尿病和非酒精性脂肪肝 (NASH) 的靶标。

Harriman 等^[36]为研制 ACC1 和 ACC2 抑制剂, 采用基于结构的虚拟筛选寻找先导化合物。以索拉芬 A (**14**, soraphen A) 与人 ACC2 的 BC 域 (hACC2 BC) 复合物晶体结构为筛选模板。索拉芬 A 是从黏细菌 *Sorangium cellulosum* 中分离出的一种 I 型聚酮类抗生素, 不仅有抗真菌作用, 而且有强效抑制 ACC 酶活性。用分子对接软件 GLIDE 筛选了 130 万个化合物, 基于

对参与结合的重要氨基酸能够发生氢键结合、静电和疏水相互作用以及可置换高能结构水分子的综合打分, 从中选出 250 个多样性结构的化合物。经体外实验测定对 hACC1 和 hACC2 的抑制活性, 发现数个结构不同的苗头分子, 经底物动力学分析, 证明不是结合于酶的催化中心。确定了化合物 **15** (ND-022) 为先导物, 其活性为 hACC1 IC₅₀ = 3.9 $\mu\text{mol}\cdot\text{L}^{-1}$; hACC2 IC₅₀ = 6.6 $\mu\text{mol}\cdot\text{L}^{-1}$ 。图 8 是 hACC2 BC 分别与 ND-022 (**15**) 和与索拉芬 A 复合物晶体结构的叠合图, 显示有相同的结合位点和部分的结构重叠。

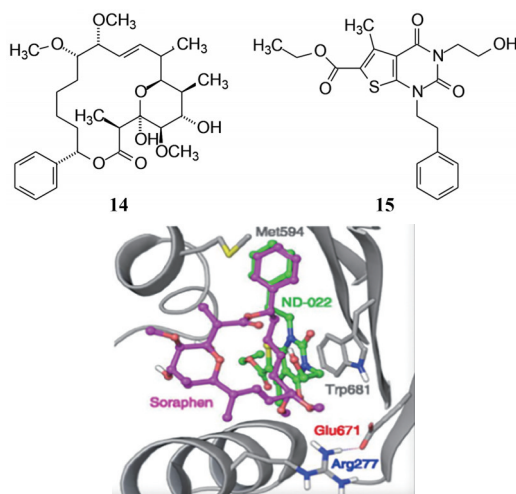


Figure 8 Cocystal structure of **15** (green) complexed with hACC2 BC overlaid with the structure of hACC2 BC complexed with **14** (purple)

化合物 **15** 经结构优化 (活性、选择性和成药性), 研制出化合物 **16** (代号 ND-630), 对 hACC1 和 hACC2 的 IC₅₀ 分别为 2.1 和 6.1 $\text{nmol}\cdot\text{L}^{-1}$, 图 9 是 **16** 与 hACC2 BC 复合物晶体衍射图。 **16** 定名为福尔索司他 (firsocostat), 现处于临床 II 期, 治疗 NASH 引起的肝纤维化。

3.3 FTO 抑制剂恩他卡朋类减肥药的研制

始自 SBVS 另一个进入临床研究的实例是研制以脂肪堆积和肥胖相关蛋白 (FTO) 为靶标的药物, 治疗肥

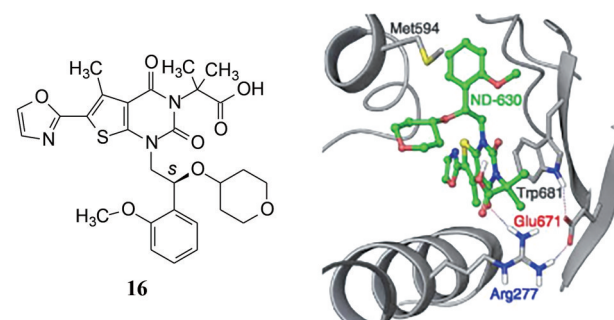
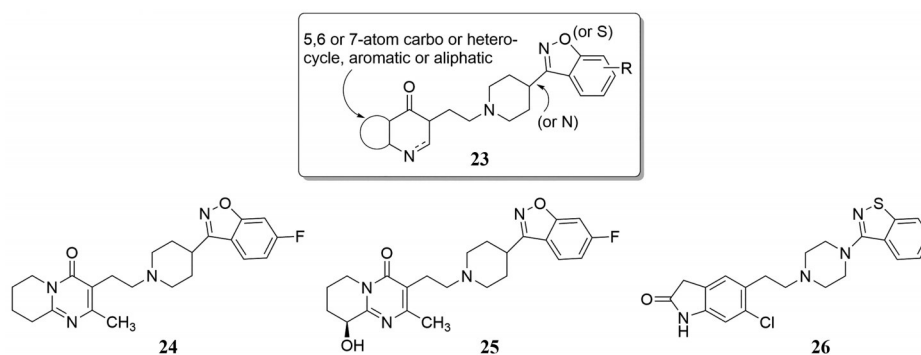


Figure 9 Structure of **16** modeled into the cocystal structure of the primary amide of **16** complexed with hACC2 BC



ziprasidone) 上市于 2006 年^[44], 是中枢多巴胺 2 (D2) 受体和 5-羟色胺 2A (5HT2A) 受体拮抗剂, 它们都是非典型抗精神病治疗药。AI 设计的药物没有超越药物化学家的思维框架, 或许是 AI 学习的训练集结构的多样性不够, 以及表征模型描述符过于明确, 从而生成的分子显示出训练集中过多的结构痕迹。

4.3 抗菌药的研究—Halicin 的发现

MIT 的科学家用 AI 研究新的抗菌药, 实现了分子结构的突破。抗菌药长时间的应用或滥用, 使细菌产生耐药性, 药物失去效力。研究者训练了一个预测抗菌分子的深度神经网络, 从“药物再利用中心库” (drug repurposing hub) 中发现了与现有抗菌药的结构迥异的广谱抗菌化合物。简要的研制要点如下。

4.3.1 训练集 从 FDA 批准的药库选出 1 760 个结构多样和活性不同的药物及 800 个天然化合物, 去掉重复分子共 2 335 个化合物。以一定浓度测定对大肠杆菌 *E. coli* BW25113 的抑制率, 以 80% 作为活性和无活性 (hit/non-hit) 的截断值, 其中有 120 个对大肠杆菌有抑制作用。这 120 个分子的多样性结构应该有不同的作用机制和抑菌靶标。

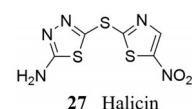
4.3.2 建立二元分类模型 分子的化学结构用图论 (graph theory) 表示, 原子为图中的节点 (node), 节点的初始特征有明确的物理化学定义, 如原子类型、表面电荷、杂化态、手性、芳香性、原子连接氢原子等, 连接键为图的边 (edge)。例如单/双/叁/共轭键、环状、顺反异构等。将这些连续向量初始特征值经消息传播的迭代, 即模型沿着编码了近邻的节点 (原子) 和边 (键型) 的连接信息作多次的消息传播, 从而找到节点最终稳定的特征向量表达方式, 再加和成表达分子最终数字向量。

按初筛结果将 2 335 个分子经二进制编码分成 hit 和 non-hit, 用这些数据训练二元分类模型, 该模型以后预测新化合物是否可能抑制大肠杆菌的生长。AI 运行的切入点, 是选择模型的特征, 为化学分子找到一个可以量化的表现形式。既往使用描述符 (descriptor) 或向量指纹 (vector fingerprint) 的有无作为表征结构

的量化特征, 因而生成的分子往往带有训练集中分子的结构印迹。MIT 的数据转化采用了消息传递神经网络 (message passing neural networks model, MPNN) 避免这种情况发生。

4.3.3 虚拟筛选生成潜在的活性化合物 将 MPNN 模型提取的分子特征向量, 经训练产生深度学习模型, 用此模型虚拟筛选药物再定位中心库 (drug repurposing hub)^[45] 中 6 111 个已有临床试验的候选药, 得到 99 个对 *E. coli* 的抑制预测有高得分值的化合物。

4.3.4 实验测定与 halicin 发现 实验测定抑菌活性, 按照设定的截断值选出 51 个, 经对结构与训练集化合物的相似性、(现实或曾经的) 临床状态和安全性 (用 AI 模型预测) 进行甄别, 得到了曾作为激酶抑制剂而研制的化合物 **27**^[46], 命名为 halicin。27 的 MIC 值为 $2 \mu\text{g}\cdot\text{mL}^{-1}$, 与含硝基的抗菌化合物 (如甲硝唑) 谷本相似度系数 (Tanimoto similarity coefficient) 为 0.21, 提示相似度较低。27 具有广谱杀菌作用, 对结核分枝杆菌等 36 种碳青霉烯耐药菌株显示杀菌作用, MIC 在 $\mu\text{g}\cdot\text{mL}^{-1}$ 水平。深入研究杀菌机制是干扰细菌形成跨膜的电化学梯度。电化学梯度能协助细菌产生能量, 没有这种梯度, 细菌就会死亡。重塑电化学梯度不是简单的几个突变就能完成的, 因此 27 也最大程度上减少了耐药的产生^[47]。



4.3.5 发现 halicin 的启示 广谱和抗耐药菌的 halicin 发现和因此发现新的作用机制, 证明 AIDD 确实可超越人们的思维范畴。首先, 用于训练模型的不拘一格, 样本是从宏观的抑制 *E. coli* 的表型作 hit 和 non-hit 的二元区分, hit 的作用靶标可能包含了 RNA、酶、离子通道或细胞膜等, 因而从更广泛空间发现新的作用环节和化合物, 避免了局限于给定靶标作分子生成。其次, 将化学结构 (组成分子的原子性质和键合方式) 依据图

论原理转换成AI可学习处理的数字向量,通过深度神经网络的处理和自学,形成的特定模型较少有研制者的人为干预,避免了AI提取hit中的描述符或指纹向量,构建出相似于hit结构的me-too分子,也难以发现干扰细菌电化学(pH)梯度的新抗菌环节。人们不知道AI模型在学习和运作中怎样产生的知识,这正是研制者让AI替他们想和做出想象不到的事情。Halicin的发现标志着AI超越了人脑,同时找到了崭新的作用靶标和强效分子。

5 小结

AIDD是深化的CADD,是用更多的数据和更复杂的算法在更强大的计算机进行药物发现。AIDD的学习能力是传统CADD所没有的。20世纪60年代QSAR开拓了计算机辅助药物研究,是在不知靶标结构的情况下,用构效关系方程揭示同类药物的结构变换引起活性变化的量变关系,映射出药物与靶标的结合特征,以指导新化合物的设计,所以属于结构优化的范畴。20世纪90年代以来,结构生物学提供靶标三维结构,得以进行基于靶标结构的药物发现、基于片段的药物发现、基于靶标结构的虚拟筛选等,在靶标-配体结合的原子和基团的微观水平进行分子设计,基本过程是结合人类智慧与计算机协助以发现和优化先导化合物。近30年来诞生了许多同类第一(FIC)和同类最佳(BIC)的新药。AIDD则是新的变革和飞跃,例如发现halicin较少有人为的诱导因素,自学训练的多样结构和模型的稳健性,使得AI生成了不同于现有的抗菌药结构,同时也揭示了新的作用环节,这些都是研究者所未及的。

致谢: 本文完稿后承蒙中山大学徐峻教授的宝贵意见,谨致谢忱。

作者贡献: 郭宗儒负责文献整理、论文撰写与修改。

利益冲突: 作者声明不存在利益冲突。

References

- [1] Hansch C, Fujita T. p - π - σ Analysis. A method for the correlation of biological activity and chemical structure [J]. J Am Chem Soc, 1964, 86: 1616-1626.
- [2] Free SM Jr, Wilson JW. A mathematical contribution to structure-activity studies [J]. J Med Chem, 1964, 7: 395-399.
- [3] Koga H, Itoh A, Murayama S, et al. Structure-activity relationships of antibacterial 6,6-disubstituted and 7,8-disubstituted 1-alkyl 1,4-dihydro-4-oxiquinoline-3-carboxylic acids [J]. J Med Chem, 1980, 23: 1358-1363.
- [4] Cramer RD III, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins [J]. J Am Chem Soc, 1988, 110: 5959-5967.
- [5] Inoue A, Kawai T, Wakita M, et al. The simulated binding of (\pm)-2,3-dihydro-5,6-dimethoxy-2-[[1-(phenylmethyl)-4-piperidinyl]methyl]-1*H*-inden-1-one hydrochloride (E2020) and related inhibitors to free and acylated acetylcholinesterases and corresponding structure-activity analyses [J]. J Med Chem, 1996, 39: 4460-4470.
- [6] Sussman JL, Harel M, Frolow F, et al. Atomic structure of acetylcholinesterase from Torpedo Californica: a prototypic acetylcholine-binding protein [J]. Science, 1991, 253: 872-879.
- [7] Ponticello GS, Freedman MB, Habecker CN, et al. Thienothio-pyran-2-sulfonamides: a novel class of water-soluble carbonic anhydrase inhibitors [J]. J Med Chem, 1987, 30: 591-597.
- [8] Greer J, Erickson JW, Baldwin JJ, et al. Application of the three-dimensional structures of protein target molecules in structure-based drug design [J]. J Med Chem, 1994, 37: 1035-1054.
- [9] Krohn A, Redshaw S, Ritchie JC, et al. Novel binding mode of highly potent HIV proteinase inhibitors incorporating the (*R*)-hydroxyethylamine isostere [J]. J Med Chem, 1991, 34: 3340-3342.
- [10] von Itzstein M, Wu WY, Kok GB, et al. Rational design of potent sialidase-based inhibitors of influenza virus replication [J]. Nature, 1993, 363: 418-423.
- [11] von Itzstein M, Dyason JC, Oliver SW, et al. A study of the active site of influenza virus sialidase: an approach to the rational design of novel anti-influenza drugs [J]. J Med Chem, 1996, 39: 388-391.
- [12] Williams MA, Lew W, Mendel WD, et al. Structure-activity relationships of carbocyclic influenza neuroaminidase inhibitors [J]. Bioorg Med Chem Lett, 1997, 17: 1837-1842.
- [13] Kim CU, Lew W, Williams MA, et al. Structure-activity relationship studies of novel carbocyclic influenza neuraminidase inhibitors [J]. J Med Chem, 1998, 41: 2451-2460.
- [14] Kim CU, Lew W, Williams MA, et al. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis and structural analysis of carbocyclic sialic acid analogues with potent antiinfluenza activity [J]. J Am Chem Soc, 1997, 119: 681-690.
- [15] Shuker SB, Hajduk PJ, Meadows R, et al. Discovering high-affinity ligands for proteins [J]. Science, 1996, 274: 1531-1534.
- [16] Tsai J, Lee J, Wang W, et al. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity [J]. Proc Natl Acad Sci U S A, 2008, 105: 3041-3046.
- [17] Petros AM, Dinges J, Augeri DJ, et al. Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis [J]. J Med Chem, 2006, 49: 656-663.
- [18] Bruncko M, Oost TK, Belli BA, et al. Studies leading to potent, dual inhibitors of Bcl-2 and Bcl-xL [J]. J Med Chem, 2007, 50: 641-662.
- [19] Park CM, Bruncko M, Adickes J, et al. Discovery of an orally bioavailable small molecule inhibitor of prosurvival B-cell

- lymphoma 2 proteins [J]. *J Med Chem*, 2008, 51: 6902-6915.
- [20] Huang QH, Johnson TW, Bailey S, et al. Design of potent and selective inhibitors to overcome clinical anaplastic lymphoma kinase mutations resistant to crizotinib [J]. *J Med Chem*, 2014, 57: 1170-1187.
- [21] Johnson TW, Richardson PF, Bailey S, et al. Discovery of (10*R*)-7-amino-12-fluoro-2, 10, 16-trimethyl-15-oxo-10, 15, 16, 17-tetrahydro-2*H*-8, 4- (metheno)pyrazolo[4, 3-*h*] [2, 5, 11] -benzoxadiazacyclotetradecine-3-carbonitrile (PF-06463922), a macrocyclic inhibitor of anaplastic lymphoma kinase (ALK) and c-ros oncogene 1 (ROS1) with preclinical brain exposure and broad spectrum potency against ALK-resistant mutations [J]. *J Med Chem*, 2014, 57: 4720-4744.
- [22] Bruce I, Akhlaq M, Bloomfield GC, et al. Development of isoform selective PI3-kinase inhibitors as pharmacological tools for elucidating the PI3K pathway [J]. *Bioorg Med Chem Lett*, 2012, 22: 5445-5450.
- [23] Scheuermann TH, Stroud D, Sleet CE, et al. Isoform-selective and stereoselective inhibition of hypoxia inducible factor-2 [J]. *J Med Chem*, 2015, 58: 5930-5941.
- [24] When PM, Rizzi JP, Dixon DD, et al. Design and activity of specific hypoxia-inducible factor-2 α (HIF-2 α) inhibitors for the treatment of clear cell renal cell carcinoma: discovery of clinical candidate (*S*)-3-((2,2-difluoro-1-hydroxy-7-(methylsulfonyl)-2,3-dihydro-1*H*-inden-4-yl)oxy)-5-fluorobenzonitrile (PT2385) [J]. *J Med Chem*, 2018, 61: 9691-9721.
- [25] Ostrem JM, Peters U, Sos ML, et al. K-Ras (G12C) inhibitors allosterically control GTP affinity and effector interactions [J]. *Nature*, 2013, 503: 548-551.
- [26] Shin Y, Jeong JW, Wurz RP, et al. Discovery of *N*-(1-acryloylazetid-3-yl)-2-(1*H*-indol-1-yl)acetamides as a covalent inhibitor of KRAS^{G12C} [J]. *ACS Med Chem Lett*, 2019, 10: 1302-1308.
- [27] Lanman BA, Allen JR, Allen JG, et al. Discovery of a covalent inhibitor of KRASG12C (AMG 510) for the treatment of solid tumors [J]. *J Med Chem*, 2020, 63: 52-65.
- [28] Fell JB, Fischer JP, Baer BR, et al. Discovery of tetrahydropyridopyrimidines as irreversible covalent inhibitors of KRAS-G12C with *in vivo* activity [J]. *ACS Med Chem Lett*, 2018, 9: 1230-1234.
- [29] Fell JB, Fischer JP, Brian R. et al. Identification of the clinical development candidate MRTX849, a covalent KRASG12C inhibitor for the treatment of cancer [J]. *J Med Chem*, 2020, 63: 6679-6693.
- [30] Wendt MD, Shen W, Kunzer A, et al. Discovery and structure-activity relationship of antagonists of B-cell lymphoma 2 family proteins with chemopotential activity *in vitro* and *in vivo* [J]. *J Med Chem*, 2006, 49: 1165-1181.
- [31] Petros AM, Olejniczak ET, Fesik SW. Structural biology of the Bcl-2 family of proteins [J]. *Biochim Biophys Acta*, 2004, 1644: 83-94.
- [32] Park CM, Bruncko M, Adickes J, et al. Discovery of an orally bioavailable small molecule inhibitor of prosurvival B-cell lymphoma 2 proteins [J]. *J Med Chem*, 2008, 51: 6902-6915.
- [33] Souers AJ, Levenson JD, Boghaert ER, et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets [J]. *Nat Med*, 2013, 19: 202-208.
- [34] Zhu H, Zhang Y, Li W, et al. A comprehensive survey of prospective structure-based virtual screening for early drug discovery in the past fifteen years [J]. *Int J Mol Sci*, 2022, 23: 15961-15979.
- [35] Unoh Y, Uehara S, Nakahara K, et al. Discovery of S-217622, a noncovalent oral SARS-CoV-2 3CL protease inhibitor clinical candidate for treating COVID-19 [J]. *J Med Chem*, 2022, 65: 6499-6512.
- [36] Harriman G, Greenwood J, Bhat S, et al. Acetyl CoA carboxylase inhibition by ND-630 reduces hepatic steatosis, improves insulin sensitivity, and modulates dyslipidemia in rats [J]. *Proc Natl Acad Sci U S A*, 2016, 113: 1796-1805.
- [37] Yang J, Loos RJ, Powell JE, et al. FTO genotype is associated with phenotypic variability of body mass index [J]. *Nature*, 2012, 490: 267-272.
- [38] Wu K. Effect of entacapone on bodyweight loss in obese population: a randomized, double-blind, placebo-controlled, pilot trial, clinical trial registration. NCT02349243 [EB/OL]. 2015 [2022-09-15]. <https://clinicaltrials.gov/ct2/show/NCT02349243>.
- [39] Peng S, Xiao W, Ju D, et al. Identification of entacapone as a chemical inhibitor of FTO mediating metabolic regulation through FOXO1 [J]. *Sci Transl Med*, 2019, 11: eaau7116.
- [40] Yoshimaga H, Uemachi N, Ohno T, et al. 2,6-Disubstituted pyridine derivatives [P]. US 2020/0239432 A1. 2020-09-23
- [41] Yoshinaga H, Ikuma Y, Ikeda J, et al. Condensed lactam derivative [P]. US 2020/0745401 A1. 2020-07-09
- [42] Fenton C, Scott LJ. Risperidone: a review of its use in the treatment of bipolar mania [J]. *CNS Drugs*, 2005, 19: 429-444.
- [43] Jones MP, Nicholl D, Trakas K. Efficacy and tolerability of paliperidone ER and other oral atypical antipsychotics in schizophrenia [J]. *Int J Clin Pharmacol Ther*, 2010, 48: 383-399.
- [44] Schmidt AW, Lebel LA, Howard HR Jr, et al. Ziprasidone: a novel antipsychotic agent with a unique human receptor binding profile [J]. *Eur J Pharmacol*, 2001, 425: 197-201.
- [45] Corsello S, Bittker J, Liu Z, et al. The drug repurposing hub: a next-generation drug library and information resource [J]. *Nat Med*, 2017, 23: 405-408.
- [46] De SK, Stebbins JL, Chen LH, et al. Design, synthesis, and structure-activity relationship of substrate competitive, selective, and *in vivo* active triazole and thiazazole inhibitors of the c-Jun N-terminal kinase [J]. *J Med Chem*, 2009, 52: 1943-1952.
- [47] Stokes JM, Yang K, Swanson K. et al. A deep learning approach to antibiotic discovery [J]. *Cell*, 2020, 180: 688-702.