

文章编号: 1006-3080(2021)03-0001-06

DOI: 10.14135/j.cnki.1006-3080.20200310004

ChemCloser: 一个基于路径与片段匹配的药物设计软件

汪伟亮, 邓卫平

(华东理工大学上海新药设计重点实验室, 上海 200237)

摘要:根据先导化合物设计“me-too”药物可以显著降低新药开发的难度, 提高开发效率和成功率。如何确切把握先导化合物的空间结构, 设计结构新颖、价键合理、空间结构与先导化合物相同或相似的骨架是设计“me-too”药物的关键问题。作者设计并编写了一个基于提取结构式的长链、片段匹配和拼接的自动化全新药物骨架的设计软件: ChemCloser。该软件能为研究人员提供结构合理、与先导化合物具有相同或相似的空间结构的所有核心骨架。

关键词:“me-too”药物; 长链提取; 片段拼接; 软件; ChemCloser

中图分类号:TP319

文献标志码:A

新药研发是一个资源密集、时间周期长(10~20年), 投资高(5~26亿美元)的过程^[1-2]。如何提高开发效率、降低开发成本是大家关心的问题。在先导化合物(或已知药物)的基础上, 设计与先导化合物有类似的空间结构、带有相同药效基团的“me-too”化合物, 这样的化合物可以与先导化合物作用于同一个酶, 产生类似的药效。这种方式显然可以降低开发难度, 提高新药开发的效率和成功率。如何确切把握先导化合物的空间结构, 设计结构新颖、价键合理、空间结构与先导化合物相同或相似的骨架(scaffold-hopping)是设计“me-too”化合物的关键问题。

目前发展的药物设计方法多是针对某一个蛋白质结合位点来生成可能的药物分子结构, 包括: 蛋白质结合位点的原子或片段连接法: LUDI(Ludwigshafen)^[3]; 片段生长连接法: LEA3D(Ligand by Evolutionary Algorithm of 3 Dimensions)^[4], LigBuilder(Ligand Builder)^[5-6], eSynth(electrical Synthesis)^[7]; 按顺序构建法: LEGEND^[8], SPROUT^[9]; 随机连接法: CoG(Compound Generator)^[10], Flux(Fragment-based Ligand bUilder reaXions)^[11]; 以及分解已知活性分子后再拼接的方法: eMolFrag(electrical Molecular Fragments)^[12], molBLOCKS^[13], SPARK^[14]等。为解决设计“me-too”化合物的新型骨架的问题, 本文开发了一个基于片

段拼接、空间结构匹配的全新药物骨架设计软件: ChemCloser。该软件不需要蛋白质的结合位点信息, 只需要一个先导化合物就可以开始设计新骨架: 通过提取先导化合物的一条长链的3D坐标获得其空间信息, 并在这条长链的基础上拼接空间合理的片段, 生成与原活性化合物空间形状相似的新型结构骨架。

1 实验部分

1.1 总体设计

ChemCloser用Python 3.6语言编写, 使用RDKit^[15]作为化学支持库, 在64位Win 7系统上编译运行。ChemCloser以输入先导化合物的3D数据(pdb文件和mol文件)开始, 经过获取活性化合物的3D结构、长链、长链与片段库匹配获得符合要求的片段、拼接片段生成新结构等步骤, 最后输出新结构的3D数据(pdb文件)。总体流程如图1所示。

ChemCloser解析活性化合物、并生成新结构的流程可以分为五部分: (1)读取活性化合物的pdb文件和mol文件, 得到活性化合物的3D结构; (2)根据活性化合物的空间结构、原子间的连接关系, 获得指定的两个点之间的所有路径, 并提取其中一条路径

收稿日期: 2020-03-10

作者简介: 汪伟亮(1981年—), 男, 博士生, 研究方向: 药物化学。E-mail: 332002351@qq.com

通信联系人: 汪伟亮, E-mail: 332002351@qq.com

引用本文: 汪伟亮, 邓卫平. ChemCloser: 一个基于路径与片段匹配的药物设计软件[J]. 华东理工大学学报(自然科学版), 2021, 47(3): 1-6.

Citation: Wang Weiliang, Deng Weiping. ChemCloser: A Drug Designing Software Based on Match Route and Fragments[J]. Journal of East China University of Science and Technology, 2021, 47(3): 1-6.

(默认提取最长路径,也可以人工指定路径)作为长链。(3)将片段库中的片段与长链进行匹配。保持长链不动,平移和旋转片段,使片段与长链尽量接近。片段与长链重叠(两个点之间的距离 $< 0.1 \text{ \AA}$ 就认为是重叠)的原子数大于 3 则认为匹配成功。用户可以通过设置超参数决定使用的片段库、在匹配时是否要匹配片段和长链的杂化方式等。匹配杂化方式会导致得到的片段数变少,但准确率提高。(4)将长链和匹配成功的片段进行组合并产生新结构。ChemCloser 提供了两种组合拼接方式,默认的方式为:从片段库中取 2~3 个片段对长链进行拼接得到新结构。(5)把新结构的 3D 坐标、元素符号等信息保存到 pdb 文件中。

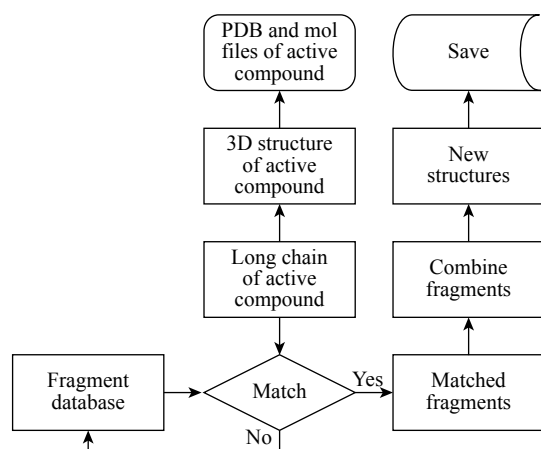


图 1 ChemCloser 的流程图
Fig. 1 Flow chart of ChemCloser

1.2 片段库

ChemCloser 采用基于片段拼接的算法生成新分子,这里的“片段”指的是 5~6 元的单环、8~10 元的并环、5~7 元的桥环、10~12 元的螺环结构等结构。ChemCloser 利用 RDKit^[15] 生成 6 个不同规模的片段库,这些片段库分别有 10 个、550 个、840 个、3.2 万个、13.6 万个、39.3 万个结构。这些基本片段最多可包含 C、H、O、N、S 等 5 种元素,其中氢元素不显式标注,其余元素用原子的 3D 坐标、元素符号、原子杂化方式等内容表示。本软件一般采用含有 550 个结构式的片段库进行匹配、拼接。

1.3 实验方法

化合物 (1)^[16] 是 DPP-IV 抑制剂的先导化合物 ($IC_{50} = 5.41 \mu\text{mol}$)。本节内容将以化合物 1 作为种子分子为例,介绍 ChemCloser 的具体工作流程:从化合物 1 中提取长链 2,并在长链 2 的基础上拼接得到香豆素并四氢吡喃结构 3、萘酚并四氢吡喃结构 4 的骨架。其中结构 4 中的萘酚并四氢吡喃是一个长效

DPP4 抑制剂的核心骨架^[16]。

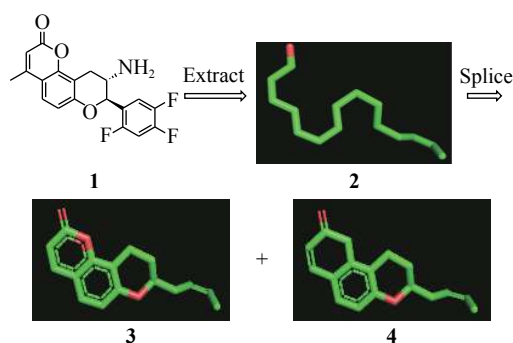


图 2 提取化合物 1 的长链并进行片段拼接

Fig. 2 Extract the long chain of compound 1 and perform fragment splicing

(1)长链 2 中所有原子间的化学键均为单键,元素种类、化学键信息不会影响每个点的空间坐标;(2)药物化学家可以根据自己的需要,在长链 2 中增加或删除结点,得到链状、树状甚至环状的结构,这些结构中的点均来自于原始的化合物 1,选择不同的点不影响其空间坐标。(3)结构式 3 中显然有 2-萘酚结构,但是 PyMol 将其显示为酮式。

2 结果与讨论

2.1 ChemCloser 的意义

化合物 1^[16] 作为 DPP-IV 抑制剂的一个先导化合物,它的稳定性、溶解性、ADME(吸收、分布、代谢和排泄)等方面可能存在问题。直接在化合物 1 的核心骨架(香豆素并四氢吡喃)上进行衍生/修饰可能会把这些问题带入新药中。

ChemCloser 的意义:在保留化合物 1 的空间结构的基础上对核心骨架进行变化,为研究人员提供所有合理的核心骨架。这些骨架中可能包含完美解决上述问题的新结构,使用这些新结构可以保持原有的药效活性,并获得良好的物理化学性质。例如化合物 4 中的萘并四氢吡喃与化合物 1 的香豆素并四氢吡喃具有几乎一样的空间形状,且化合物 4 不含酯基,不容易被酶水解而破坏其三环结构,因此具有更好的稳定性;同时将香豆素中的酯羰基改为酚羟基(PyMol 将其显示为酮式)后,为新结构提供进一步衍生和修改的可能。

2.2 软件的运行时间

ChemCloser 在读取活性化合物的 pdb 文件、mol 文件、提取活性化合物的长链等过程时运行较快,一般不超过 1 s。片段与长链进行匹配的过程的运行时间取决于片段库的选择和长链的原子数,一

般耗时 20 ~ 300 s。组合拼接过程是软件运行中耗时最长的部分,依据不同的化合物库和不同的长链,软件在片段拼接步骤消耗的时间变化较大,在 10 min ~ 10 d 甚至更久。

2.3 ChemCloser 的性能

ChemCloser 采用含有 550 个片段的片段库与长链 **2** 进行匹配,一共匹配成功 881 个状态。每一轮拼接都需要进行 881 次排列组合。第 n 轮需要 881^n 次排列组合。显然,排列组合的次数取决于片段库的大小、片段库中的片段与长链的匹配程度。如果修改超参数,使用其他较大的片段库或者不要求匹配原子的杂化方式,则匹配成功的片段状态将明显增加,ChemCloser 的排列组合性能会明显降低。基于计算机的计算性能,目前只能使用 550 个片段的片段库,且只能拼接 2 个片段。

经过两轮片段拼接、并删除价键不合理的结构,共剩余 28265 个结构。用 SHAFTS (SHApe Feature Similarity)^[18] 对 ChemCloser 生成的结构与化合物 **1** 进行相似性打分。结果见表 1。

表 1 ChemCloser 生成的新结构与化合物 **1** 的结构相似性比较
Table 1 Similarity between the new structure generated by ChemCloser and the structure of compound **1**

Index	Score range	Structure numbers	Ratio/%
1	0.90 ~ 1.00	8 550	31.1
2	0.80 ~ 0.89	11 560	42.0
3	0.70 ~ 0.79	4 190	15.2
4	0.60 ~ 0.69	2 600	9.44
5	0.50 ~ 0.59	386	1.40

从表 1 可以看出,有 73.1% 的新结构与化合物 **1** 的形状得分在 0.80 ~ 1.00 之间。说明 ChemCloser 产生的结构可以较好地保留原活性化合物的空间形状,预期它们在添加药效基团之后能与原活性化合物产生相似的药效活性。

2.4 影响性能的因素

ChemCloser 采用片段与原活性化合物的长链进行匹配的方式产生新结构,总体的空间形状被长链严格限定,所以用 ChemCloser 产生的新结构可以得到较高的形状相似性得分。如果选取的活性化合物较大,长链较长,则片段覆盖长链的部分比较少,产生的新结构的相似性会有所降低。另一个因素是活性化合物的长链的选择。选择不同的起点、终点、不同的路径都会导致 ChemCloser 得到不同的长链,进而得到不同的新结构,这些新结构也会影响形状相

似性得分。

2.5 结构式的多样性

ChemCloser 使用的片段库中包含 550 个片段(在下一个版本中可能使用含有 840 个或更多个片段的片段库),在这个规模下进行排列组合显然可以得到较好的多样性(相比与 LigBuilder^[5] 只使用 57 个片段)。但是 ChemCloser 产生的结构无法跳出长链 **2** 的限制,因此几乎无法生成其他类似形状的结构。另外,正如前面提到的,同一个活性化合物选择不同的起点、终点和路径会导致产生不同的新结构,因此只基于某一条路径生成新结构会导致多样性降低。如果在多条路径的基础上进行计算,就可以获得更多的结构式,更好的多样性。

2.6 在人鼻病毒外壳蛋白抑制剂、人类间变性淋巴瘤激酶抑制剂方面的应用

化合物 **(5)**^[19] 是一个人鼻病毒外壳蛋白抑制剂、化合物 **(6)**^[20] 是一个人类间变性淋巴瘤激酶抑制剂。ChemCloser 在它们的基础上生成一系列新结构。这些新结构与原活性化合物的结构相似性对比见表 2。表中 a) 表示化合物 **6** 中长链的起点和终点原子编号分别为 28、29; b) 表示化合物 **6** 中长链的起点和终点原子编号分别为 1、28; c) 表示化合物 **6** 中长链的起点和终点原子编号分别为 1、21。

表 2 基于化合物 **5** 和 **6** 生成的新结构的相似性值分布
Table 2 Similarity value distribution of the new structures generated based on compounds **5** and **6**

Score range	Compound 5	Compound 6 ^(a)	Compound 6 ^(b)	Compound 6 ^(c)
0.90 ~ 1.00	58.4%	36.9%	0	0
0.80 ~ 0.90	5.50%	62.5%	0	1.16%
0.70 ~ 0.80	0.20%	0	74.0%	24.3%

从表 2 可以看出,对于化合物 **5**,新结构中有 63.9% 与化合物 **5** 有较好的形状相似性(得分 0.80 ~ 1.00),说明在长链限制下生成的新结构确实能较好地保持原化合物的空间结构。

化合物 **6** 是一个环状结构,选择不同的起点和终点会得到截然不同的路径,进而生成不同的结构。例如当起点和终点分别为 28 和 29 时,生成的结构中有 99.4% 与化合物 **6** 有较好的形状相似性。而起点和终点分别为 1 和 28 时,没有得分为 0.80 ~ 1.00 之间的结构式,74.0% 的结构与化合物 **6** 的相似性得分为 0.70 ~ 0.80 之间;起点和终点分别为 1 和 21 时,仅有 1.16% 的结构式与化合物 **6** 的相似性得分在 0.80 ~ 0.90 的区间,24.3% 的结构式在 0.70 ~ 0.80 之间。这三组数据说明选择不同的起点和终点

(即不同的路径)对生成的结构有较大影响,也再一次印证了选择合适的起点和终点(即合适的路径)的重要性。

2.7 存在的问题及改进方向

2.7.1 计算性能和优化方式 ChemCloser 生成新结构的方式本质上是片段的排列组合。因此当数据量较大时,一定会发生“组合爆炸”,原因是生成的中间结构无论质量好坏都进行下一轮拼接,浪费了大量 CPU 性能。由此导致 ChemCloser 的计算性能欠佳。这个问题将在下一个版本中解决:对每一个生成的中间结构“即时打分”,舍弃得分偏低的中间结构,只用得分较高的中间结构进行下一轮拼接,这样可以同时提高软件的计算效率和准确率。

2.7.2 修改原化合物的骨架对原有生物活性的影响

活性化合物的药效基团是真正产生活性的关键基团,而活性化合物的骨架只是提供安放药效基团的基本结构。当活性化合物(基本结构+药效基团)放到酶的结合位点时,这些药效基团可以“摆放”在合理的位置,使活性化合物与酶结合。如果活性化合物的骨架改变了,但是新骨架仍然能保持所有的药效基团“摆放”在合理的位置,那么这个新化合物的药效基团还是能与酶结合,能发挥与原化合物相同或相似的活性。ChemCloser 可以保证新生成的骨架与原活性化合物有相同或相似的空间结构,且在骨架上连接药效基团后,这些药效基团仍然能“摆放”在合理的位置,并与酶结合。这样的化合物将会与原活性化合物具有相同或相似的生物活性。

2.7.3 新结构的优劣的判断 作为药物的核心骨架,应该具有较好的稳定性,适中的刚性等物理化学特性。如果核心骨架上存在酯基、酰胺、醛基、(半)缩醛、(半)缩酮等容易变化的基团,则该骨架很容易被降解(破坏);如果核心骨架是大的芳环会导致药物刚性太强,溶解性降低,导致化合物无法突破人体内的各种屏障,到达作用位点并发挥药效。我们对片段库中的基本结构进行筛选,删除化学性质不稳定、刚性过强的片段,使生成的新骨架有较好物理、化学性质。但是,即使片段库中的基本结构都是化学性质稳定、刚性适中的片段,在进行片段拼接后仍然可能生成不稳定、刚性过强的结构。目前只能人工判断新骨架的优劣,而无法让 ChemCloser 自动判断,这个问题需要在下一个版本中尝试解决。

2.7.4 提高生成结构的质量的方式 ChemCloser 生成的骨架质量差别较大,需要通过多方面对新骨架进行考察。例如用 SHAFTS^[18] 对新骨架与原活性化合物进行空间相似性打分、查找是否存在不稳定的

基团、通过引入受体的结构,计算新化合物与受体的结合数据,对比原化合物与受体的结合数据,可以进一步确定新骨架是否合理。综合应用这些方法可以在所有结构中筛选出质量最好的一个或几个骨架。

2.7.5 判断生成的化合物新颖性的方法 ChemCloser 只负责生成包括原活性化合物的骨架,以及新骨架在内的所有骨架,不负责判断新骨架是不是“新颖的”。对于 ChemCloser 生成的新骨架,必须要由药物设计者通过全面查阅文献、专利等资料后判断该骨架是不是新颖的。

2.7.6 路径的选择方式 ChemCloser 是一个半经验的药物设计软件,并不能全自动地设计活性化合物的新骨架。一般情况下,活性化合物会有多条路径,ChemCloser 默认选取活性化合物中最长的一条长链。但最长的路径并不一定是(最)合理的路径。我们认为在药物设计之前,先要研究活性化合物中哪些是药效基团,哪些是辅助基团,活性化合物的关键骨架是什么,在确定这些信息之后,再进行有取舍的路径选择。合理的路径应该保证路径上能够放置所有的药效基团,并获得活性化合物的空间信息。而这样的路径可能不再是链状结构,有可能是树状结构、甚至是图(包含环)。如果依靠计算机的路径规划算法可能很难区分哪些是必须的片段,哪些是不重要的片段,由此很难获得真正合理的路径(长链)。

3 结 论

(1) 开发了一款基于片段拼接、空间结构匹配的全新药物骨架的设计软件: ChemCloser, 它能提取活性化合物的一条长链的 3D 坐标,并在该长链的基础上拼接空间合理的片段,得到结构新颖、空间形状与活性化合物相似的分子骨架,可以为药物设计人员提供更多、更优秀的核心骨架。

(2) 以 DPP-IV 抑制剂的先导化合物 **1** 为设计模型,分析了 ChemCloser 的性能、影响性能的因素、生成的结构式的多样性。之后又考察了 ChemCloser 在人鼻病毒外壳蛋白抑制剂和人类间变性淋巴瘤激酶抑制剂方面的应用,新结构同样显示出较好的结构相似性,并且再一次印证了选择不同的长链对新结构有较大影响。

(3) 分析了 ChemCloser 存在的几个问题。① ChemCloser 采用排列组合的方式拼接生成新结构式,这种方式在数据量较大时会发生“组合爆炸”的问题,作者提出该问题的解决方案:采用“即时打分-拼接”的方式生成新结构式,该方式在解决“组合爆炸”

问题的同时,还可以提高生成的结构式的准确率。②目前 ChemCloser 不能自动判断新结构的优劣,也不能判断生成的结构是不是新颖的,这些都需要人工判断。③分析了修改原化合物的骨架对原有生物活性的影响、提高生成结构的质量的方式、路径的选择方式等问题。

参考文献:

- [1] PAUL S M, MYTELKA D S, DUNWIDDIE C T, *et al.* How to improve R&D productivity: The pharmaceutical industry's grand challenge[J]. *Nature Reviews Drug Discovery*, 2010, 9: 203-214.
- [2] AVORN J. The \$2.6 billion pill - methodologic and policy considerations[J]. *New England Journal of Medicine*, 2015, 372: 1877-1879.
- [3] BOHM H J. The computer program LUDI: A new method for the de novo design of enzyme inhibitors[J]. *Journal of Computer-Aided Molecular Design*, 1992, 6: 61-78.
- [4] DOUGUETD, MUNIER-LEHMANN H, LABESSE, G, *et al.* LEA3D: A computer-aided ligand design for structure-based drug design[J]. *Journal of Medicinal Chemistry*, 2005, 48: 2457-2468.
- [5] WANG R, GAO Y, LAI L. LigBuilder: A multi-purpose program for structure-based drug design[J]. *Journal of Molecular Modeling*. 2000, 6: 498-516.
- [6] YUAN Y, PEI J, LAI L. Ligbuilder2: A practical de novo drug design approach[J]. *Journal of Chemical Information and Modeling*, 2001, 51(5): 1083-1091.
- [7] NADERI M, ALVIN C, DING Y, *et al.* A graph-based approach to construct target-focused libraries for virtual screening[J]. *Journal of Cheminformatics*, 2016(8): 14.
- [8] NISHIBATA Y, ITAI A. Automatic creation of drug candidate structures based on receptor structure[J]. *Tetrahedron*, 1991, 47(43): 8985-8990.
- [9] GITLLET V, JOHNSON A P, MATA P, *et al.* SPROUT: A program for structure generation[J]. *Journal of Computer-Aided Molecular Design*, 1993, 7: 127-153.
- [10] BROWN N, MCKAY B, GILARDONI F, *et al.* A graph-based genetic algorithm and its application to the multiobjective evolution of median molecules[J]. *Journal of Chemical Information and Computer Sciences*, 2004, 44(30): 1079-1087.
- [11] FECHNER U, SCHNEIDER G. Flux (2): Comparison of molecular mutation and crossover operators for ligand-based de novo design[J]. *Journal of Chemical Information and Modeling*, 2007, 47(2): 656-667.
- [12] TAIRAN L, MISAGH N, Chris A, *et al.* Break down in order to build up: Decomposing small molecules for fragment-based drug design with eMolFrag[J]. *Journal of Chemical Information and Modeling*, 2017, 57(4): 627-631.
- [13] GHERSI D, SINGH M. molBLOCKS: Decomposing small molecule sets and uncovering enriched fragments[J]. *Bioinformatics*, 2014, 30(14): 2081-2083.
- [14] MARTIN S, ANDY V. Scaffold Hopping in Medicinal Chemistry [M/oL]. Wiley Online Library, NATHAN, B. 2013 [2020]. <https://onlinelibrary.wiley.com/doi/book/10.1002/9783527665143>.
- [15] Landrum G A. RDKit: Cheminformatics and Machine Learning Software. GREG, L. (www.rdkit.org) [20200522]. www.rdkit.org.
- [16] LI S, XU H, CUI S, *et al.* Discovery and rational design of natural-product-derived 2-phenyl-3, 4-dihydro-2H-benzo [f] chromen-3-amine analogs as novel and potent dipeptidyl peptidase 4 (DPP-4) inhibitors for the treatment of type 2 diabetes[J]. *Journal of Medicinal Chemistry*, 2016, 59 (14): 6772-6790. [jmedchem.6b00505](https://doi.org/10.1021/acs.jmedchem.6b00505).
- [17] LI S, QIN C, CUI S, *et al.* Discovery of a natural-product-derived preclinical candidate for once-weekly treatment of type 2 diabetes[J]. *Journal of Medicinal Chemistry*, 2019, 62 (5): 2348-2361.
- [18] LIU X, JIANG H, LI H. SHAFTS: A Hybrid approach for 3D molecular similarity calculation: 1. Method and assessment of virtual screening[J]. *Journal of Chemical Information and Modeling*, 2011, 51(9): 2372-2385.
- [19] KOTALAR P R, BELLA J, OLSON N H, *et al.* Structural studies of two rhinovirus serotypes complexed with fragments of their cellular receptor[J]. *The EMBO Journal*, 1999, 18: 6249-6259.
- [20] JOHNSON T W, Richardson P F, BAILEY S, *et al.* Discovery of (10R)-7-Amino-12-fluoro-2,10,16-trimethyl-5-oxo-10, 15, 16, 17-tetrahydro-2H-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]. *Journal of Medicinal Chemistry*, 2014, 57(11): 4720-4744.

ChemCloser: A Drug Designing Software Based on Match Route and Fragments

Wang Weiliang, Deng Weiping

(Shanghai Key Laboratory of New Drug Design & School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China)

Abstract: New drug design is a huge project, and the failure rate is very high. On the basis of the existing active compounds, the design of "me-too" drugs can increase the success rate. The "me-too" drug should have the same or similar spatial structure, pharmacophore, etc. with the original active compound, so as to ensure that the combination of the "me-too" drug and the enzyme is the same as the original active compound. The lead compound can provide the same or similar spatial structure as the original active compound, and a novel skeleton, but the skeleton of the lead compounds does not necessarily have good physicochemical properties. Designing "me-too" drugs based on lead compounds can significantly reduce the difficulty of new drug development and improve the development efficiency and success rate. How to accurately grasp the spatial structure of the lead compounds, designing a novel structure, a reasonable valence bond, and a skeleton with the same or similar spatial structure as the lead compound are the key issues for the design of "me-too" drugs. The author designs and writes an automated new drug skeleton design software based on the extraction of structural long chains, fragment matching and splicing: ChemCloser. The software can provide researchers with all core frameworks with reasonable structure and the same or similar spatial structure as the lead compounds. Researchers may find frameworks with novel structures, good physical properties and stable chemical properties in these core frameworks. After splicing pharmacodynamic groups on these skeletons, it can produce the same or similar biological activity as the original active compound.

Key words: "me-too" drugs; long chain extraction; fragment splicing; software; ChemCloser