

# 基于分子对接的反向虚拟筛选方法

王存新<sup>1</sup>, 许先进<sup>2</sup>

(1. 北京工业大学生命科学与生物工程学院, 北京 100124;  
2. 密苏里大学-哥伦比亚分校道尔顿心血管研究中心, 密苏里 哥伦比亚 65211)

**摘要:** 基于分子对接的反向虚拟筛选方法在药物靶点确定、老药新用以及药物副作用/毒理研究领域具有重要的应用前景,吸引了药物发现领域研究人员的广泛关注. 首先对分子对接方法和蛋白质数据库进行细致的介绍,然后列举目前可以用于反向虚拟筛选的网络服务器,并列举该方法在药物设计领域的一些具体应用,最后对该方法目前所存在的问题进行讨论.

**关键词:** 反向虚拟筛选; 分子对接; 药物发现; 药物靶点确定; 老药新用; 毒副作用

**中图分类号:** Q 71; Q 74

**文献标志码:** A

**文章编号:** 0254 - 0037(2019)11 - 1164 - 09

**doi:** 10. 11936/bjutxb2019060001

## Docking-based Inverse Virtual Screening Approach

WANG Cunxin<sup>1</sup>, XU Xianjin<sup>2</sup>

(1. College of Life Science and Bioengineering, Beijing University of Technology, Beijing 100124, China;  
2. Dalton Cardiovascular Research Center, University of Missouri-Columbia, Columbia, MO 65211, USA)

**Abstract:** Docking-based inverse virtual screening (IVS) has attracted much attention in the drug discovery field, due to its applications in target identification, drug repositioning, and side effects/toxicity prediction. First, molecular docking methods and protein target database were introduced in details. Then, currently available web servers were briefly reviewed. A number of applications in the field of drug design were also presented. Finally, challenges of the methods were discussed.

**Key words:** inverse virtual screening; molecular docking; drug discovery; target identification; drug repositioning; side effects/toxicity

虚拟筛选(virtual screening)方法现已成为新药研发过程中不可或缺的重要技术. 传统的虚拟筛选方法是指对于给定的靶点蛋白质,运用基于物理化学的计算机算法程序,对数据库中的小分子化合物进行打分排序,从中挑选出靶点蛋白质潜在的配体分子,然后通过实验加以验证. 近些年,另一虚拟筛选方法逐渐获得研究人员的青睐. 该方法与传统的虚拟筛选方法正好相反,对于给定的配体分子,需要

从数据库中挑选出能与之结合的靶点蛋白质,故称之为反向虚拟筛选(inverse virtual screening, IVS)方法<sup>[1-5]</sup>.

IVS方法之所以获得新药研发人员的青睐,主要是因为它在以下几个方面有重要的应用前景:药物靶点确定(target identification)、老药新用(drug repositioning)以及药物副作用/毒理(side effects/toxicity)研究. 对药物靶点的确定有助于理解药物

收稿日期: 2019-06-03

基金项目: 科技部国际科技合作资助项目(2010DFA31710)

作者简介: 王存新(1943—),男,教授,主要从事生物信息学及分子设计方法学方面的研究, E-mail: cxwangbjut@126.com

通信作者: 许先进(1987—),男,博士后,主要从事蛋白质与配体相互作用及复合物结构预测方面的研究, E-mail: xianjin37@gmail.com

发挥作用的机理,进而有效地对药物进行改进以增强其药效.老药新用,即为现有的药物寻找新的靶点蛋白质,以达到治疗新的疾病的目的.由于已经成药的化合物都已通过毒副作用及临床实验的筛查,该策略可以极大缩短药物研发时间和费用.在药物副作用/毒理研究方面,如果在新药研发初期能够有效预测药物的副作用,即确定除靶标蛋白质之外其他能与药物结合的蛋白质,必将极大地提高药物在后期临床实验中的通过率.IVS 技术正是为这些应用而量身定制的.

目前已经发展出众多不同种类的 IVS 方法,根据方法的不同大致可以分为以下几类:基于配体的方法(ligand-based methods)、结合位点比较(binding site comparisons)、蛋白质-配体相互作用指纹分析(protein-ligand interaction fingerprints)以及基于分子对接的方法(docking-based methods)<sup>[1-4]</sup>.基于配体的方法的理论基础是分子相似性准则(molecular similarity principle),即结构相似的分子倾向于具有相似的生物活性<sup>[6-7]</sup>.这类方法依赖于数据库中与给定配体分子结构相似的化合物是否有已知的活性或毒性信息,如果没有,这类方法将无能为力.结合位点比较和蛋白质-配体相互作用指纹分析的方法同样存在类似的限制,即对于给定的配体分子,必须至少有一个已知的能与该配体结合的蛋白质靶点或者有一个已知结构的蛋白质-配体复合物结构.故这 3 类方法又可以被统称为“基于知识的”方法(“knowledge-based” methods).相比较而言,基于分子对接的方法是目前唯一一种无须“已知”靶点的 IVS 方法,因此该方法正受到越来越多研究人员的关注.

简而言之,基于分子对接的 IVS 方法是指对于给定的配体分子,运用分子对接程序,将该小分子与数据库中的蛋白质受体进行逐个对接,并输出对应的打分分值,然后根据该分值对数据库中的蛋白质受体进行排序,排在前面的就是该配体分子的潜在靶点.因此基于分子对接的 IVS 方法一般由以下两大部分构成:用作搜索引擎的分子对接方法和蛋白质靶点数据库.本文将对这两大内容分别进行介绍.为方便使用,现已有不少研究组将基于分子对接的 IVS 方法整合到网络服务器供研究人员免费使用.另外,作者列举了 IVS 在药物靶点确定、老药新用以及药物副作用/毒理研究领域的一些具体应用实例.最后,讨论了该类方法目前所存在的问题以

及发展前景.

## 1 分子对接方法

对于已知结构的一对受体和配体分子,可以通过物理化学原理和科学计算算法,预测受体和配体分子间的结合模式及结合自由能,这类方法被称为分子对接方法.分子对接方法的思想可以追溯到 100 多年前 Fischer<sup>[8]</sup>提出的“锁钥模型”,即受体与配体的结合是一个几何匹配的过程.随着研究的不断深入,简单的“锁钥模型”的局限性日渐显现.“锁钥模型”中,受体和配体在结合前后自身的构象没有改变,即复合物的形成是一个刚性对接过程,这显然与实际不符.另外,除了几何匹配外,能量匹配(静电、氢键、疏水相互作用)也在复合物的形成过程中扮演着至关重要的角色.随后, Koshland<sup>[9]</sup>于 1958 年提出“诱导契合”的匹配概念.在该模型中,复合物的形成是受体与配体相互适应的过程,受体和配体的构象在结合过程中都会发生变化,最终形成形状和能量最优匹配的结合模式.

近年来,分子对接方法已经从最初的蛋白质-小分子对接拓展到任意生物大分子之间(蛋白质-蛋白质、蛋白质-多肽、蛋白质-DNA/RNA 等)的对接.经过数 10 年的发展应用,世界各国的研究人员已经开发出相当数量的分子对接程序,有兴趣的读者可以参看最近的综述文献<sup>[10-15]</sup>.本文将主要集中在适用于 IVS 的蛋白质-小分子对接程序上.

### 1.1 主要的分子对接程序

据不完全统计,针对蛋白质-小分子的分子对接程序现已超过 50 种之多<sup>[16]</sup>.不同的程序都有各自特点,从方法上很难比较孰优孰劣,对结合模式和结合自由能的预测能力也因体系而异,理论上任意一种分子对接程序都可以用作 IVS 的搜索引擎.表 1 列出了常用的蛋白质-小分子对接程序.更多的分子对接程序可以见参考文献[16].各个程序具体的使用方法可以参看对应的网站或程序手册.需要指出的是,现有的分子对接程序都是针对单个受体和单个配体的对接,即对于给定的一个蛋白质受体和一个配体分子,预测这一对蛋白质-配体的结合模式及结合自由能.而在 IVS 研究中,研究人员需要将配体分子逐个对接到数据库中的每一个蛋白质受体上,然后根据所计算的结合自由能将蛋白质受体进行排序,挑选出潜在的靶点蛋白质.

表1 常用的蛋白质-小分子对接程序

Table 1 Prevalent docking programs for protein-ligand complexes

程序名称	下载链接	算法及特点
AutoDock <sup>[17]</sup>	<a href="http://autodock.scripps.edu/">http://autodock.scripps.edu/</a>	遗传搜索算法与基于经验的打分函数;对接过程中可以考虑受体的部分残基侧链的柔性,配体分子可以完全柔性对接
AutoDock Vina <sup>[18]</sup>	<a href="http://vina.scripps.edu/">http://vina.scripps.edu/</a>	搜索算法和打分函数都不同于 AutoDock,其计算速度和成功率都有很大的提高,且能在多核机器上并行运算
DOCK <sup>[19]</sup>	<a href="http://dock.compbio.ucsf.edu/">http://dock.compbio.ucsf.edu/</a>	锚定片段生长(anchor-and-grow)搜索算法;用 AMBER 力场经验势能函数打分
Glide <sup>[20-21]</sup>	<a href="https://www.schrodinger.com/glide">https://www.schrodinger.com/glide</a>	商业化软件,由 Schrödinger 公司提供;蒙特卡罗搜索算法,分级筛选搜索可能结合位点;使用不同精度的打分函数,可加速虚拟筛选过程
GOLD <sup>[22]</sup>	<a href="https://www.ccdc.cam.ac.uk/solutions/csd-iscovery/components/gold/">https://www.ccdc.cam.ac.uk/solutions/csd-iscovery/components/gold/</a>	商业化软件,遗传搜索算法,配体柔性,部分考虑蛋白柔性
MDock <sup>[23-24]</sup>	<a href="http://zoulab.dalton.missouri.edu/resources_mdock.html">http://zoulab.dalton.missouri.edu/resources_mdock.html</a>	通过使用多个蛋白质受体或配体小分子构象来考虑蛋白质或小分子的柔性;打分函数为基于知识的原子对统计势

## 1.2 分子对接程序的算法简介

分子对接算法是基于复合物形成的能量最低原理假设,即天然的蛋白质-配体复合物结构在复合物构象空间中能量最低。目前的分子对接程序一般包含两大部分:构象搜索和打分函数(能量评估)。构象搜索是指通过计算机搜索算法,对蛋白质-配体复合物的构象空间进行搜索,产生各种可能的结合模式。打分函数则是对这些可能的结合模式进行能量评估,寻找最低能量对应的结合模式,即近天然复合物构象。

值得注意的是,最低能量的分值在 IVS 的研究中通常是直接或间接地对数据库中蛋白质受体进行排序挑选。作者将集中介绍打分函数项在 IVS 研究中的应用。

## 1.3 打分函数

现有的对接程序中的打分函数主要分为以下3类:基于物理的打分函数、基于经验的打分函数、基于知识的打分函数。基于物理的打分函数又被称为基于力场(force field-based)的结合自由能计算方法,该方法考虑分子的内能、溶剂效应以及熵效应,因此计算十分耗时,分子对接程序 DOCK<sup>[19]</sup>和 GOLD<sup>[22]</sup>中包含该类打分函数。相比较而言,基于经验的打分函数计算速度要快很多,该类打分函数将考虑残基成对偏好性、几何互补性及静电、氢键、疏水相互作用能等因素对蛋白质-配体结合的贡献,各贡献项的权重则可以通过对已知的复合物结构及结合自由能值拟合得到。使用该类打分函数的分子对接程序包括 AutoDock/Vina<sup>[17-18]</sup>等。基于知

识的打分函数则依赖于已知结构的蛋白质-配体复合物,通过对这些数据结构进行统计分析,现已发展残基-残基接触势<sup>[25]</sup>、残基成对偏好性和原子-原子接触势<sup>[26]</sup>等打分函数。例如分子对接程序 MDock<sup>[23-24]</sup>中的 ITScore 就是基于原子对接触势的打分函数。

理论上,运用上述打分函数计算出来的分值可以直接用于对受体数据库中的蛋白质受体进行排序,然后根据排序的顺序挑选出潜在的靶点蛋白质。实际上,不少基于分子对接的 IVS 方法正是这么做的。例如网络服务器 TarFisDock<sup>[27]</sup>就是直接使用分子对接程序 DOCK4.0 中的打分函数计算得到的分值对靶点蛋白质进行排序,输出排在前面的 2%、5% 或者 10% 的蛋白质靶点用于进一步的实验测试。然而目前分子对接中的打分函数都不能严格地计算蛋白质-配体的结合自由能,在 IVS 中直接使用该分值对不同的蛋白质进行排序将会产生众多的假阳性蛋白质靶点。目前已有不少研究尝试减少预测结果中的假阳性靶点的数目,下面列举了其中的一些重要方法。

在首个基于分子对接的 IVS 方法(INVDOCK)中<sup>[28]</sup>,为了对蛋白质的排序列表进行筛选,作者针对配体引入了一个能量阈值  $\Delta E_T = -4.184\alpha N \text{ kJ/mol}$ ,其中  $N$  是配体分子的原子数目, $\alpha$  是常数(约 1.0),该值可以通过对蛋白质数据库中已知复合物结构进行拟合得到。在 IVS 研究中,如果某个蛋白质的对接分值优于该阈值  $\Delta E_T$ ,该蛋白质将被选为潜在的蛋白质靶点。更进一步,考虑到在体内

环境中配体会与蛋白质的天然配体进行竞争结合, 作者引入了另一个能量阈值  $\Delta E_c$ , 该值是蛋白质与其天然配体的打分分值, 通过对已知的复合物结构计算得到. 在 IVS 研究中, 如果某个蛋白质的对接分值低于  $\beta\Delta E_c$ , 该蛋白质将被选为潜在的蛋白质靶点, 其中,  $\beta \leq 1$  (作者推荐值 0.8).

在最近的 Santiago 等<sup>[29]</sup>的工作中, 针对单个蛋白质的阈值概念同样被引入到 IVS 研究中; 不同的是, 他们挑选了一个含有 1 990 个类似药物分子的配体数据库, 并将这些小分子对接到受体数据库中的每个蛋白质上, 对于特定蛋白质受体, 取前 200 或者前 20 分值的平均值, 或者是所有分值的玻尔兹曼加权平均值, 视作该蛋白的阈值. 在 IVS 中, 如果对接得到的分值优于对应的阈值, 则该蛋白质受体被选为配体分子潜在的蛋白质靶点. 根据文中的结果分析, 使用前 20 分值的平均值作为阈值要优于另外 2 个.

除了对蛋白质受体设置阈值外, 另一种可以减少假阳性靶蛋白的方法是使用多种不同的打分函数对数据库中的蛋白质受体分别进行排序, 然后挑选出在不同种打分函数的排序中排列都比较靠前的蛋白质受体, 用于进一步的实验测试. 例如在 Li 等<sup>[30]</sup>的工作中, 使用了 2 种不同种类的打分函数, 一种是基于经验的 ICM 打分函数, 另一种是基于知识的 PMF 打分函数, 结合了这 2 种不同类型的打分函数 (consensus scoring), 从而有效地提高了 IVS 的成功率.

除此之外, 针对具体的研究对象, 研究人员发展了各种不同的策略以提高 IVS 的成功率. 例如在 Yang 等<sup>[31]</sup>构建的预测严重药物不良反应 (serious adverse drug reactions) 的 IVS 网络服务器中引入的 2DIZ (2-directional Z-transformation) 算法. 感兴趣的读者可进一步参看最近的英文综述文章<sup>[5]</sup>.

## 2 靶点蛋白质数据库

IVS 中除了作为搜索引擎的分子对接程序外, 另一重要组成部分就是靶点蛋白质数据库. 该数据库由已知结构的蛋白质受体组成, 并且是能够结合小分子配体的蛋白质. 随着 X 射线 (X-ray)、核磁共振 (nuclear magnetic resonance, NMR) 以及冷冻电镜 (Cryo-EM) 等实验技术的快速发展, 越来越多的蛋白质结构被解析出来. 到目前 (2019 年 5 月 11 日) 为止, 蛋白质结构数据库 (protein data bank, PDB)<sup>[32]</sup>中已经存有 140 774 个蛋白质结构, 为 IVS

的靶点蛋白质数据库的构建提供了丰富的资源. 目前已经有不少可从 PDB 演变而来的子数据库, 可直接用于基于分子对接的 IVS.

在现有的子数据库中, Kellenberger 等<sup>[33]</sup>、Desaphy 等<sup>[34]</sup>构建的 sc-PDB (screening-PDB, <http://bioinfo-pharma.u-strasbg.fr/scPDB/>) 数据库, 从 PDB 中搜集了所有结合了配体分子的蛋白质结构信息, 配体分子包括小分子、多肽 (少于 9 个残基) 以及核苷酸 (小于 4 个碱基). 在最近更新的 2017 年版本中, 包含了 16 034 个 PDB 结构, 其中含有 4 782 种蛋白质受体及 6 326 个配体分子. 该数据的优点是受体数量大, 包含的信息多, 缺点是没有对众多信息进行有效的分类, 例如数据库中的蛋白质种类繁多, 并且来自各种不同的物种, 使得 IVS 的后期挑选变得更加复杂. 因此, 根据不同的研究目的, 研究人员进一步构建了更加精细的蛋白质受体数据库, 例如治疗靶点数据库 (therapeutic target database, TTD)<sup>[35]</sup>、潜在药物靶点数据库 (potential drug-target database, PDTD)<sup>[36]</sup>、药物不良反应数据库 (drug adverse reaction database, DART)<sup>[37]</sup>、小分子-转录因子数据库 (small molecule transcription factor database, SM-TF)<sup>[38]</sup> 以及蛋白质-配体复合物结合亲和力实验数据库 (a comprehensive collection of binding affinities for the protein-ligand complexes in the protein data bank, PDBbind)<sup>[39-40]</sup>. 下面我们将对这几个数据库进行简要介绍.

治疗靶点数据库 TTD (<https://db.idrblab.org/ttd/>) 从现有的文献中收集已知的潜在治疗靶点, 包括蛋白质靶点和核酸靶点. 最近更新的版本 (2017 年 9 月), 总共含有 3 101 个靶点, 其中已经成药的靶点 445 个, 正在进行临床试验的 1 121 个, 其他正在研究阶段的共有 1 535 个. 需要提醒的是其中很多靶点的三维结构还未被解析出来, 使用该数据库进行 IVS 研究需要对数据库进行进一步的处理.

与治疗靶点数据库 TTD 类似, 潜在药物靶点数据库 PDTD (<http://www.dddc.ac.cn/pdtd/>) 也是针对治疗靶点的数据库, 不同的是, PDTD 只包含有蛋白质受体, 并且是已知三维结构的蛋白质受体. 最近的一次更新是 2008 年, 该版本包含 1 207 个条目, 覆盖 841 个已知的或潜在的药物靶蛋白. 这些蛋白质又进一步根据治疗领域或者是生物化学特性进行了分类.

药物不良反应数据库 DART (<http://bidd.nus.edu.sg/group/drt>) 则是针对有药物不良反应的蛋白

质受体,该数据库收集了147个已被确定的具有药物不良反应的蛋白质受体,另外还含有89个潜在的受体.该数据库最后一次更新是在2003年.

小分子-转录因子数据库 SM-TF (<http://zoulab.dalton.missouri.edu/SM-TF/>) 则是从PDB中收集了所有能够结合小分子的转录因子结构,转录因子的功能失常与众多疾病直接相关,是很多药物的靶点蛋白质.最近更新的版本含有934个条目,覆盖了176种不同的转录因子.

蛋白质-配体复合物结合亲和力实验数据库 PDBbind ([www.pdbbind.org.cn](http://www.pdbbind.org.cn)) 收集了PDB数据库中所有具有实验测得的结合亲和力数据的蛋白质-配体复合物结构.对于每个蛋白质-配体复合物,该数据库除了提供结合亲和力的实验数据( $K_d$ ,  $K_i$ , 或  $IC_{50}$ )外,还提供清理过后的蛋白质三维结构(PDB文件)和对应的配体结构(MOL2文件和SDF文件).最近更新的版本含有16151个蛋白质-配体复合物,另外还包含2416个蛋白质-蛋白质复合物、896个蛋白质-核酸复合物以及125个核酸-配体复合物.

根据研究目的的不同,研究人员可以从PDB或sc-PDB数据库出发,构建个性化的数据库,这些数据库还可以运用同源建模(homology modeling)的方法进行扩充.

### 3 网络服务器

为方便以实验为主的研究人员,使用基于分子对接的IVS技术,一些研究组已经将该方法自动化,并构建成便于使用的网络服务器.通常,使用者只需通过网络服务器上传自己感兴趣的配体分子,服务器会完全自动地运行IVS,最后给出该配体分子潜在的受体蛋白质,以供研究人员进一步实验测试.现有的网络服务器包括 TarFisDock、SerPreSA、DRAR-CPI以及 idTarget.

TarFisDock (<http://www.dddc.ac.cn/tarfisdock>)<sup>[27]</sup> 使用的搜索引擎是 DOCK4.0,蛋白质数据库是潜在药物靶点数据库 PDTD,使用者可以选择整个 PDTD 库或者是其中的子库进行 IVS 研究.

SerPreSA (<http://cpi.bio-x.cn/sepresa/>)<sup>[31]</sup> 使用的搜索引擎是 DOCK4.0,所有蛋白质数据库由91种与严重药物不良反应(severe adverse drug reactions, SADR)相关的蛋白质,因此该服务器主要用于预测配体分子潜在的严重药物不良反应.

DRAR-CPI (<http://cpi.bio-x.cn/drar/>)<sup>[41]</sup> 使

用的搜索引擎是 DOCK6.0,蛋白质数据库中含有353个药物靶点蛋白质,可用于预测给定配体分子的可能的靶点蛋白质,寻找该配体分子潜在的药用价值或者毒副作用.

idTarget<sup>[42]</sup> (<http://idtarget.rcas.sinica.edu.tw>) 使用 MEDock<sup>[43]</sup> 作为搜索引擎,蛋白质数据库使用 sc-PDB,可用于对配体分子潜在蛋白质受体进行广泛搜索.

除了上述网络服务器外,据作者所知,Bullock等<sup>[44]</sup>推出的 DockoMatic2.0是目前唯一一个可在本地机器运行基于分子对接IVS的开源程序.该程序使用 AutoDock 或者 AutoDock Vina 作为搜索引擎,虽然该方法没有以上网络服务器使用方便,但它允许使用者构建自己的蛋白质数据库,可适用于不同的研究目的.

### 4 应用实例

基于分子对接的IVS技术在药物研发领域已有广泛的应用,例如对药物靶点的确定、老药新用、药物副作用/毒理研究等,下面将列举一些这些方面的应用实例,更多的应用实例可参考最近的英文文献<sup>[5]</sup>.

天然产物在药物的发展史上扮演着至关重要的作用,对基于天然产物的传统药物作用靶点的确定不仅能够揭开其作用机理,还能够为现代药物设计提供重要的治疗靶点<sup>[45]</sup>.最近,IVS方法被广泛应用于确定传统中药(traditional Chinese medicine, TCM)的蛋白质靶点研究当中<sup>[46-49]</sup>.例如 Zhao等<sup>[48]</sup>则运用 INVDOCK 程序筛选了膜荚黄芪(*Astragalus membranaceus* Bunge)中治疗心血管疾病的有效成分黄芪甲苷(astragaloside-IV)的蛋白质靶点,所用数据库是从 DrugBank<sup>[50]</sup>中收集的188个与心血管疾病相关的蛋白质.作者运用实验方法在分子水平上验证了39个预测的蛋白质中的3个靶点蛋白质,并进一步通过药物-靶点网络的方法对筛选结果加以分析验证.Chen等<sup>[49]</sup>运用网络服务器 idTarget<sup>[42]</sup>,并结合基于配体的反向虚拟筛选网络服务器 PharmMapper<sup>[51]</sup>,预测传统中药组分丹参素的抗癌靶点,筛选结果显示 GTPase HRas 是丹参素的一个潜在蛋白质靶点.

在老药新用方面,Li等<sup>[52]</sup>针对 DrugBank2.5 数据库<sup>[50]</sup>中的药物分子及蛋白质靶点进行了大批量的分子对接实验.具体地,作者使用分子对接程序 ICM<sup>[53]</sup>将4621个已上市或还在实验阶段的药物分

子分别对 252 个人类蛋白质药物靶点上, 对于给定的配体分子, 结合使用了 2 种不同类型的打分函数对蛋白质进行排序, 根据已知的蛋白质靶点数据, 预测成功率达到 48.8%, 这也意味着另外 51.2% 的预测靶点中同样存在着真实的靶点蛋白质。更进一步, 作者通过实验的方法验证了用于治疗慢性骨髓性白血病的尼洛替尼 (Nilotinib) 是 MAPK14 的强效抑制剂 ( $IC_{50} = 40 \text{ nmol/L}$ )。MAPK14 是治疗炎症的靶蛋白, 尼洛替尼则有可能会成为治疗类风湿性关节炎的药物。

除了上面的 2 类应用之外, IVS 在药物副作用/毒理研究领域也有非常广泛的应用。例如 Chen 等<sup>[28]</sup>测试了 INVDOCK 程序针对 8 种常见药物的副作用和毒性作用, 这 8 种药物分子分别为阿司匹林 (aspirin)、庆大霉素 (gentamicin)、布洛芬 (ibuprofen)、因地那韦 (indinavir)、新霉素 (neomycin)、青霉素 G (penicillin G)、4H-他莫昔芬 (4H-tamoxifen) 和维生素 C (vitamin C)。结果显示 83% 的实验已知的药物副作用和毒性作用成功被预测。随后, 该方法同样被应用到 11 种抗艾滋病药物 (anti-HIV drugs) 的副作用和毒性作用研究中<sup>[54]</sup>, 并达到了类似的预测成功率 (86%)。Ma 等<sup>[55]</sup>则运用 IVS 方法研究了三聚氰胺 (melamine) 的毒性机制。三聚氰胺曾经在中国的婴儿奶粉中被发现, 在被喂食该奶粉的婴幼儿体内检测出肾结石症状。预测结果给出了 4 个可能的致病靶点蛋白质 (glutathione peroxidase 1, beta-hexosaminidase subunit beta, l-lactate dehydrogenase 和 lysozyme C), 但这里靶点还需进一步的实验加以验证。

## 5 困难与挑战

经过 10 多年的发展, 基于分子对接的 IVS 方法已经有了很大的进步, 该方法已被应用于药物研发领域, 在非药物研发领域也有所应用<sup>[56-57]</sup>。但该方法仍然面临不少困难和挑战, 下面将从分子对接方法和蛋白质数据库 2 个方面分别进行讨论。

在分子对接方法上, 目前所用到的打分函数还很不精确, 不能严格地计算蛋白质-配体的结合自由能, 可以说都是非常粗略地对结合自由能进行估算。这也可以从各种分子对接程序在最近的 D3R (drug design data resource, <https://drugdesigndata.org/>) 竞赛中的表现看出<sup>[58-63]</sup>。D3R 每年提供 1~2 个蛋白质-配体数据库, 全世界近 30 个研究组参加预测数据库中蛋白质-小分子复合物的结合模式和

结合自由能。令人遗憾的是, 现有的方法在结合自由能的预测上还有很长的路要走。另外, 目前的 IVS 研究中都未考虑受体蛋白质的柔性问题。现有的分子对接程序在处理蛋白质柔性问题尚有很大的困难, 对接过程中要处理上百或上千种不同蛋白质的柔性, 势必要极大地增加计算量, 而且预测结果也不一定有改善。因此, 目前的基于分子对接的 IVS 方法还远不能用于精准预测, 类似于传统的虚拟筛选方法, 只能将可能的蛋白质靶点从一个大的数据库中缩小到一个较小的范围, 然后通过实验的方法加以确定<sup>[64-70]</sup>。

关于蛋白质结构数据库, 最大的问题是数据的不完整性。虽然目前已被解析出结构的蛋白质的数目相当可观, 但存在较大冗余, 且只是整个蛋白质组的很小的一部分。好在随着时间的推移及新的实验技术的出现, 蛋白质结构数据库会越来越完善。另外, 还可以用同源建模的方法对数据库进行扩充。

虽然目前基于分子对接的 IVS 方法存在着上述问题, 但相信随着分子对接方法的逐步优化、计算机能力的迅速提升以及更多蛋白质结构的解析, 这些问题将会逐步得到解决。因此, IVS 方法必将有更大的发展和应用前景。

## 参考文献:

- [1] ROGNAN D. Structure-based approaches to target fishing and ligand profiling [J]. *Molecular Informatics*, 2010, 29 (3): 176-187.
- [2] KOUTSOUKAS A, SIMMS B, KIRCHMAIR J, et al. From in silico target prediction to multi-target drug design: current databases, methods and applications [J]. *Journal of Proteomics*, 2011, 74(12): 2554-2574.
- [3] MA D L, CHAN D S, LEUNG C H. Drug repositioning by structure-based virtual screening [J]. *Chemical Society Reviews*, 2013, 42(5): 2130-2141.
- [4] XIE L, XIE L, BOURNE P E. Structure-based systems biology for analyzing off-target binding [J]. *Current Opinion in Structural Biology*, 2011, 21(2): 189-199.
- [5] XU X, HUANG M, ZOU X. Docking-based inverse virtual screening: methods, applications, and challenges [J]. *Biophysics Reports*, 2018, 4(1): 1-16.
- [6] WILLETT P, BARNARD JM, DOWNS GM. Chemical similarity searching [J]. *Journal of Chemical Information and Computer Sciences*, 1998, 38(6): 983-996.
- [7] BENDER A, GLEN RC. Molecular similarity: a key technique in molecular informatics [J]. *Organic & Biomolecular Chemistry*, 2004, 2(22): 3204-3218.

- [8] FISCHER E. Einfluss der configuration auf die wirkung der enzyme [J]. Berichte der Deutschen Chemischen Gesellschaft, 1894, 27: 3189-3232.
- [9] KOSHLAND D E. Application of a theory of enzyme specificity to protein synthesis [J]. Proceedings of the National Academy of Sciences of the United States of America, 1958, 44: 98-104.
- [10] BROOIJMANS N, KUNTZ I D. Molecular recognition and docking algorithms [J]. Annual Review of Biophysics and Biomolecular Structure, 2003, 32(1): 335-373.
- [11] SOUSA S F, FERNANDES P A, RAMOS M J. Protein-ligand docking: current status and future challenges [J]. Proteins: Structure, Function, and Bioinformatics, 2006, 65(1): 15-26.
- [12] GRINTER S Z, ZOU X. Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design [J]. Molecules, 2014, 19(7): 10150-10176.
- [13] 许先进, 王存新. 分子对接方法在药物发现之外领域的应用 [J]. 北京工业大学学报, 2017, 43(12): 1872-1880.  
XU X J, WANG C X. Applications of molecular docking: beyond the drug discovery [J]. Journal of Beijing University of Technology, 2017, 43(12): 1872-1880. (in Chinese)
- [14] CIEMNY M, KURCINSKI M, KAMEL K, et al. Protein-peptide docking: opportunities and challenges [J]. Drug Discovery Today, 2018, 23(8): 1530-1537.
- [15] PORTER KA, DESTA I, KOZAKOV D, et al. What method to use for protein-protein docking? [J]. Current Opinion in Structural Biology, 2019, 55: 1-7.
- [16] SOUSA SF, RIBEIRO AJ, COIMBRA JT, et al. Protein-ligand docking in the new millennium—a retrospective of 10 years in the field [J]. Current Medicinal Chemistry, 2013, 20(18): 2296-2314.
- [17] MORRIS G M, GOODSELL D S, HALLIDAY R S, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function [J]. Journal of Computational Chemistry, 1998, 19(14): 1639-1662.
- [18] TROTT O, OLSON A J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading [J]. Journal of Computational Chemistry, 2010, 31(2): 455-461.
- [19] KUNTZ I D, BLANEY J M, OATLEY S J, et al. A geometric approach to macromolecule-ligand interactions [J]. Journal of Molecular Biology, 1982, 161(2): 269-288.
- [20] FRIESNER R A, BANKS J L, MURPHY R B, et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy [J]. Journal of Medicinal Chemistry, 2004, 47(7): 1739-1749.
- [21] HALGREN T A, MURPHY R B, FRIESNER R A, et al. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening [J]. Journal of Medicinal Chemistry, 2004, 47(7): 1750-1759.
- [22] JONES G, WILLETT P, GLEN R C, et al. Development and validation of a genetic algorithm for flexible docking [J]. Journal of Molecular Biology, 1997, 267(3): 727-748.
- [23] HUANG S Y, ZOU X. An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials [J]. Journal of Computational Chemistry, 2006, 27(15): 1866-1875.
- [24] HUANG S Y, ZOU X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking [J]. Proteins: Structure, Function, and Bioinformatics, 2007, 66(2): 399-421.
- [25] ZHANG C, LIU S, ZHOU Y Q. Accurate and efficient loop selections by the DFIRE-based all-atom statistical potential [J]. Protein Science, 2004, 13(2): 391-399.
- [26] ZHANG C, VASMATZIS G, CORNETTE J L, et al. Determination of atomic desolvation energies from the structures of crystallized proteins [J]. Journal of Molecular Biology, 1997, 267(3): 707-726.
- [27] LI H, GAO Z, KANG L, et al. TarFisDock: a web server for identifying drug targets with docking approach [J]. Nucleic Acids Research, 2006, 34(Suppl 2): W219-W224.
- [28] CHEN Y Z, ZHI D G. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule [J]. Proteins: Structure, Function, and Bioinformatics, 2001, 43(2): 217-226.
- [29] SANTIAGO D N, PEVZNER Y, DURAND A A, et al. Virtual target screening: validation using kinase inhibitors [J]. Journal of Chemical Information and Modeling, 2012, 52(8): 2192-2203.
- [30] LI YY, AN J, JONES SJ. A computational approach to finding novel targets for existing drugs [J]. PLoS Computational Biology, 2011, 7(9): e1002139.
- [31] YANG L, LUO H, CHEN J, et al. A server for the prediction of populations susceptible to serious adverse

- drug reactions implementing the methodology of a chemical-protein interactome [J]. *Nucleic Acids Research*, 2009, 37(Suppl 2): W406-W412.
- [32] BERMAN HM, WESTBROOK J, FENG Z, et al. The protein data bank [J]. *Nucleic Acids Research*, 2000, 28(1): 235-242.
- [33] KELLENBERGER E, MULLER P, SCHALON C, et al. sc-PDB: an annotated database of druggable binding sites from the Protein Data Bank [J]. *Journal of Chemical Information and Modeling*, 2006, 46(2): 717-727.
- [34] DESAPHY J, BRET G, ROGNAN D, et al. sc-PDB: a 3D-database of ligandable binding sites-10 years on [J]. *Nucleic Acids Research*, 2014, 43(D1): D399- D 404.
- [35] LI Y H, YU C Y, LI X X, et al. Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics [J]. *Nucleic Acids Research*, 2017, 46 ( D1 ): D1121-D1127.
- [36] GAO Z, LI H, ZHANG H, et al. PDTD: a web-accessible protein database for drug target identification [J]. *BMC Bioinformatics*, 2008, 9(1): 104.
- [37] JI Z L, HAN L Y, YAP C W, et al. Drug adverse reaction target database (DART) [J]. *Drug Safety*, 2003, 26(10): 685-690.
- [38] XU X, MA Z, SUN H, et al. SM-TF: A structural database of small molecule-transcription factor complexes [J]. *Journal of Computational Chemistry*, 2016, 37 (17): 1559-1564.
- [39] WANG R, FANG X, LU Y, et al. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures [J]. *Journal of Medicinal Chemistry*, 2004, 47 (12): 2977-2980.
- [40] LI Y, SU M, LIU Z, et al. Assessing protein-ligand interaction scoring functions with the CASF-2013 benchmark [J]. *Nature Protocols*, 2018, 3(4): 666-680.
- [41] LUO H, CHEN J, SHI L, et al. DRAR-CPI: a server for identifying drug repositioning potential and adverse drug reactions via the chemical-protein interactome [J]. *Nucleic Acids Research*, 2011, 39 (suppl 2): W492-W498.
- [42] WANG J C, CHU P Y, CHEN C M, et al. idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach [J]. *Nucleic Acids Research*, 2012, 40(W1): W393- W399.
- [43] CHANG D T, OYANG Y J, LIN J H. MEdock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm [J]. *Nucleic Acids Research*, 2005, 33(Suppl 2): W233- W238.
- [44] BULLOCK C, CORNIA N, JACOB R, et al. DockoMatic 2.0: high throughput inverse virtual screening and homology modeling [J]. *Journal of Chemical Information and Modeling*, 2013, 53(8): 2161-2170.
- [45] JI H F, LI X J, ZHANG H Y. Natural products and drug discovery: can thousands of years of ancient medical knowledge lead us to new and powerful drug combinations in the fight against cancer and dementia? [J]. *EMBO Reports*, 2009, 10: 194-200.
- [46] YUE Q X, CAO Z W, GUAN S H, et al. Proteomics characterization of the cytotoxicity mechanism of ganoderic acid D and computer-automated estimation of the possible drug target network [J]. *Molecular & Cellular Proteomics*, 2008, 7: 949-961.
- [47] FENG L X, JING C J, TANG K L, et al. Clarifying the signal network of salvianolic acid B using proteomic assay and bioinformatic analysis [J]. *Proteomics*, 2011, 11: 1473-1485.
- [48] ZHAO J, YANG P, LI F, et al. Therapeutic effects of astragaloside IV on myocardial injuries: multi-target identification and network analysis [J]. *PLOS ONE*, 2012, 7: e44938.
- [49] CHEN S J, REN J L. Identification of a potential anticancer target of danshensu by inverse docking [J]. *Asian Pacific Journal of Cancer Prevention*, 2014, 15: 111-116.
- [50] WISHART D S, KNOX C, CUO A C, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration [J]. *Nucleic Acids Research*, 2006, 34: D668-D672.
- [51] LIU X, OUYANG S, YU B, et al. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach [J]. *Nucleic Acids Research*, 2010, 38: W609-W614.
- [52] LI Y Y, AN J, JONES S J. A computational approach to finding novel targets for existing drugs [J]. *PLOS Computational Biology*, 2011, 7: e1002139.
- [53] ABAGYAN R, TOTROV M, KUZNETSOV D. ICM-A new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation [J]. *Journal of Computational Chemistry*, 1994, 15: 488-506.
- [54] JI Z L, WANG Y, YU L, et al. In silico search of putative adverse drug reaction related proteins as a potential tool for facilitating drug adverse effect prediction

- [J]. *Toxicology Letters*, 2006, 164: 104-112.
- [55] MA C, KANG H, LIU Q, et al. Insight into potential toxicity mechanisms of melamine: an in silico study [J]. *Toxicology*, 2011, 283: 96-100.
- [56] 许先进, 苏计国, 刘斌, 等. 持久性有机污染物 4, 4'-DDE 和 CB-153 的反向虚拟筛选 [J]. *物理化学学报*, 2013, 29(10): 2276-2285.
- XU X J, SU J G, LIU B, et al. Reverse virtual screening on persistent organic pollutants 4, 4'-DDE and CB-153 [J]. *Acta Physico-Chimica Sinica*, 2013, 29(10): 2276-2285. (in Chinese)
- [57] CALVARESI M, ZERBETTO F. Baiting proteins with C60 [J]. *ACS Nano*, 2010, 4(4): 2283-2299.
- [58] GAIEB Z, PARKS CD, CHIU M, et al. D3R grand challenge 3: blind prediction of protein-ligand poses and affinity rankings [J]. *Journal of Computer-Aided Molecular Design*, 2019, 33(1): 1-8.
- [59] GAIEB Z, LIU S, GATHIAKA S, et al. D3R grand challenge 2: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies [J]. *Journal of Computer-Aided Molecular Design*, 2018, 32(1): 1-20.
- [60] GATHIAKA S, LIU S, CHIU M, et al. D3R grand challenge 2015: evaluation of protein-ligand pose and affinity predictions [J]. *Journal of Computer-Aided Molecular Design*, 2016, 30(9): 651-668.
- [61] XU X, MA Z, DUAN R, et al. Predicting protein-ligand binding modes for CELPP and GC3: workflows and insight [J]. *Journal of Computer-Aided Molecular Design*, 2019, 33(3): 367-374.
- [62] DUAN R, XU X, ZOU X. Lessons learned from participating in D3R 2016 grand challenge 2: compounds targeting the farnesoid X receptor [J]. *Journal of Computer-Aided Molecular Design*, 2018, 32(1): 103-111.
- [63] XU X, YAN C, ZOU X. Improving binding mode and binding affinity predictions of docking by ligand-based search of protein conformations: evaluation in D3R grand challenge 2015 [J]. *Journal of Computer-Aided Molecular Design*, 2017, 31(8): 689-699.
- [64] KITCHEN D B, DECORNEZ H, FURR J R, et al. Docking and scoring in virtual screening for drug discovery: methods and applications [J]. *Nature Reviews Drug Discovery*, 2004, 3(11): 935-949.
- [65] BENDER A, GLEN R C. A discussion of measures of enrichment in virtual screening: comparing the information content of descriptors with increasing levels of sophistication [J]. *Journal of Chemical Information and Modeling*, 2005, 45(5): 1369-1375.
- [66] REDDY A S, PATI S P, KUMAR P P, et al. Virtual screening in drug discovery—a computational perspective [J]. *Current Protein and Peptide Science*, 2007, 8(4): 329-351.
- [67] HUANG S Y, GRINTER S Z, ZOU X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions [J]. *Physical Chemistry Chemical Physics*, 2010, 12(40): 12899-12908.
- [68] SCIOR T, BENDER A, TRESADERN G, et al. Recognizing pitfalls in virtual screening: a critical review [J]. *Journal of Chemical Information and Modeling*, 2012, 52(4): 867-881.
- [69] PYZER-KNAPP E O, SUH C, GÓMEZ-BOMBARELLI R, et al. What is high-throughput virtual screening? a perspective from organic materials discovery [J]. *Annual Review of Materials Research*, 2015, 45: 195-216.
- [70] FORLI S, HUEY R, PIQUE M E, et al. Computational protein-ligand docking and virtual drug screening with the AutoDock suite [J]. *Nature Protocols*, 2016, 11(5): 905-919.

(责任编辑 郑筱梅)