

虚拟配体筛选:策略、前景及其限制

黄世杰编译

(军事医学科学院毒物药物研究所,北京 100850)

摘要:与高通量筛选相比,虚拟配体筛选(VS)是用计算机程序选择化合物,预测它们与靶受体的结合情况。其关键前提是知晓蛋白-配体结合所需空间和能量的标准。本文讨论了VS的靶选择,分析和制备以及如何编辑候选配体数据库。也考虑了VS运转的工具、策略、评分的精确度和结果的排序问题。

关键词:虚拟筛选;靶受体;蛋白-配体结合

中图分类号:R914.2 **文献标识码:**A **文章编号:**1001-0971(2006)06-0445-03

1 概述

20世纪80年代末和90年代初,为了克服药物开发中发现先导化合物困难所致的瓶颈,试验性高通量筛选(THS)和组合化学迅猛发展。事实上,这些努力使人失望,尤其是此途径的低命中率和高昂的花费。此种状况促成了开发替代技术的设想,90年代末虚拟筛选开始成型。本文目的是总结和评论虚拟配体筛选(VS)的概念和成功运行VS的前提。此外,诚实坦白地分析此种途径的多种限制,并寻求对这些限制的解释。下面首先讨论靶的选择、分析和前提,然后是候选配体筛选样本的编辑。接着讨论VS运行的实际需要、工具和策略,最后关注评分的精确度和筛选结果的排序问题。

2 靶的选择,验证和分析

作为认知驱动的途径,VS强烈依赖研究体系中有效信息的数量和质量。DNA和RNA作为靶受体,仅在最近才有恰当的锚定和评分工具,事实上这种靶标在VS中应用很少。相反,蛋白质在VS中应用较多,范围从酶到G蛋白偶联受体和离子通道。

3 确定靶受体的成药能力

选择的蛋白应具有能被小分子配体成功靶向的结合袋。很显然,这种结合袋的大小和几何结构,表面的粗糙和复杂度及与推测的药物样配体的形状和极性的互补性和识别性均十分重要。然而,VS开始

研究需要更多实际结合架构的可靠信息。近来,Hajduk等提出几个关键性指标帮助鉴别成药性和非成药性结合袋。对于成药性结合袋,总表面积和极性区低于 0.75 nm^2 似乎较佳,此外还应具有恰当的紧密度,表面粗糙度和复杂度。然而,Hajduk等的分析也显示单一指标的相关性不强。

4 选择靶受体最相关的几何结构

获得蛋白质空间结构的最有力方法是晶体结构分析。此法的精确度和可靠性强烈依赖衍射的分辨率。确定蛋白结构的替代技术是NMR,确定结构的精度取决于沿蛋白链的核欧沃豪斯效应(Overhauser effect)距离的局部分布图。为提高已发现结合位点建模结构的精度,新算法考虑到在同源模块化过程中配体结合的方向。

5 处理蛋白质的可变性和适应性

在VS中,重要的是获取蛋白质同分异构体和配体结合的估算值。一旦确定了这些同分异构体,VS按蛋白质的可变性进行操作,或者比较几个刚性受体(rigid receptor)同分异构体的装配。后者较令人注意,假定试验观察到的蛋白质同分异构体是处于稳定,低自由能状态。

6 晶体结构作为“金标准”

原则上,结晶的蛋白-配体复合物,可用蛋白结晶加到配体溶液中制备,然后进行晶体衍射,找到结合位点,此过程称为“浸泡”。近来用浸泡法发现,由蛋白醛糖还原酶与抑制剂唑啉司他形成的复合物有不同的同分异构体。在一种结构中,蛋白质通过结

收稿日期:2006-09-05

作者简介:黄世杰,男,副研究员,研究方向:神经药理学,Tel:010-66930670

合位点的氨基键与配体形成氢键。在另一种结构中,结合位点的氨基键发生旋转不能与同一配体形成氢键。这种差别明显影响 VS 的锚定和评分结果。在结晶条件下,配体和结合袋呈反向结合,换句话说,观察到配体引起蛋白的构象适应。在选择特异的晶体结构作为 VS 参照物时,应详细地分析配体和紧邻结合位点的 β -因子参数或者氢键网络的一致性。

7 分配质子化状态

大多数计算机筛选程序需要适当定义活性位点或残基的质子化状态。此种操作并非易事,因为局部的双电性条件调控功能基的 pK_a 值,幅度达到几个数量级。等温滴定热量计能使这种变化明显,经常使受体功能基转向供者或荷电氢键转为中性氢键。在运行 VS 各个验证步骤时,这些变化是明显的。

8 考虑与单个或多个蛋白同分异构体结合

经验显示蛋白结合袋中发生的构象适应,通常与蛋白功能需要而调整结构有关。此种功能适应必须与低能转换相应,否则造成蛋白功能的明显缺陷。因此,酶与不同底物结合或在催化构象适应时,经历多个蛋白同分异构体的结合状态。

9 配体与靶蛋白结合的“热点”地形图

已有几种工具说明结合袋的结合“热点”。这些方法或者根据全局的参数化作用力场,或者根据良好选择的经验信息。最重要的是,这些分析必须针对一个结合袋的所有构象状态,因为这将提供蛋白适应变化时结合位点分子如何识别的构成图。

10 水分子是蛋白-配体复合物的第三者

使用水基模块分析数千个配体-蛋白复合物晶体结构揭示,约 2/3 场合,配体结合中包含 1 个水分子,经常介入蛋白和配体的连接。因此,任一基于推测候选配体结合模式的预测方法,例如 VS 中,必须考虑水分子的影响。

11 候选配体的选择

HTS 和 VS 需要充分设计,充分收集小分子候选配体以供筛选。为此,大制药公司在完成化合物的国内收集外,还通过商业途径购得尽量多的化合物。

12 VS 候选配体的样本收集

以市售的化合物目录开始,目前可供购置的化合物在 1 000 万个以上。通常,分子的元素组成是 H, C, N, O, P, S, Cl 和 Br, 相对分子质量 < 500 u 都符合要求。区别化合物之间的相似性或差异性有 1 个参考点,在 VS 中,这个参考点是靶受体和两个配体结合亲和力的差异,确定两个分子应属于相同或不同的类别。

13 用于 VS 的候选配体的大小

药物优化程序的经验显示,已知有明显结合亲和力的小核心片段是进一步优化的起点。因此,VS 选择的化合物应有可优化余地,方可进入所谓的先导化合物样或片段样分子的匹配范畴。

14 选择良好平衡的候选配体用于筛选

新近, Martin 检验了常用的 Lipinski 入选标准,研究显示,它对中性和阳性电荷分子有预测能力;然而阴离子有不同的标准,其极性表面积也有预测能力。结合位点的成药能力,生物利用度标准等都是多因素决定的。用 Lipinski 标准可从筛选样本中丢弃那些推测有不良性质的化合物。

15 合成的或虚拟候选分子的数据样本

由结合位点的结构来确定适宜的设计标准,在 VS 中作为靶标,可为 VS 产生定制靶的虚拟数据库。

16 VS 运转的工具和策略

原则上,VS 运转遵循两个策略:由锚定命中点向前或向后筛选化合物。VS 中最重要步骤是候选分子锚定于靶蛋白。加上各种标准向前筛选以减少样本数,从数百万个试验化合物中减至数百或上千个有希望的候选物。应用排序向后筛选留下前位者。如今,计算机运行速率已不再是影响选择进行的因素。向前筛选开始于简单参数,例如相对分子质量、可旋转键数、亲脂性或形状,如总结构的椭圆率。其后,探讨受体结构位点的信息,结合袋热点分析,衍生出基于蛋白的药效基因。向前筛选使更精确的锚定得以进行,例如进入多个蛋白构象状态运算或反映基于蛋白的药效基因的精确锚定。最重要的是,这种分层筛选的策略可追踪监视筛选各个水平上的操作,且可人为介入,特别是易用视觉检查锚定结果。

17 选择最适宜的锚定工具

目前所有锚定工具仍然远离完美无缺的要求。近来,大众化锚定工具 DOCK 和 FlexX 开始启用,其特点是锚定已有的药效基团或特性分布。Autodock 程序预先安排的格子中计算锚定,贮存任一相互作用域的潜值。对多个配体作用的鉴别,排序能力很强。

18 VS 锚定结果的评分,排序和验证

实际应用的 VS 策略外,锚定和其后的评分是 VS 中关键的决定因素。评分的鉴别力对排序特别重要,有希望充实前位排名。过去几年已开发较好的评分函数,不限于回答单个函数,而是几个评分项目的综合结果。

19 计算结合亲和力的第一原则方法

最有力的,理论上更坚实的途径是第一原则法。使用量子机制或计算需要较少的力场,计算系统的分配函数和确定结合与非结合态的自由能的差别,随着计算机运算速度的提高,此法变得更精确,与评分关系更紧密。

20 由回归分析衍生的经验评分函数

新近基于已知亲和力结合配体晶体结构的大样本计算得到一种新回归评分函数,依靠一套训练程序,发现与分析有关的许多新项目。此种经验评分函数与训练项目有关者的符合性最佳,但通用性不大。

21 估算结合亲和力的精确度

为了估算结合常数,回归和基于认知的评分函数两者均需要试验亲和力数据作为内标。因此,它

们预测的精确度从未大于能测定的结合亲和力的精确度。根据经验,认知评分函数能较好地提取由锚定程序产生的结合位点的几何结构,最接近试验证实的结合模式。相反,回归评分函数在预测实际亲和力上较好,提供较真实的结合位点的几何结构。

22 结合亲和力: 1 个充分理解的性质

任何评分和排序的讨论必定要问如何理解 Gibbs 自由能的靶值。考虑到 1 个配体与 1 个蛋白的结合,焓(ΔH)和熵($-T\Delta S$)较 Gibbs 自由能(ΔG)分布大 3~4 倍的范围。在紧密相关的分子间可以发现焓和熵的相互补偿。此外,结合过程中许多物化现象还没有完全理解,因此也未能把它们正确地用于评分函数中。尽管理解结合过程中的物理变化存在缺陷,评分仍然进行得很满意,因为在只考虑配体和蛋白结合的相对比较下,任何分析中不确切现象,均可忽略不计。

23 结语

VS 已成为对 HTS 有力的替代和补充方法。当优化运行时,已有印象深刻的命中率报告,较 HTS 明显高 100~1 000 倍。HTS 和 VS 的比较研究表明,VS 法能捕获替代的和补充的配体。值得注意的是有经验的药物化学家经常能指出一个特别的结合袋可能是成药性的,将这些数据输入计算机运算,使这些多因素标准更明确。锚定和评分均属于多因素性质。经验和人为介入对 VS 运行中各种筛选步骤保持控制力最重要。考虑到相似性的概念,高差异分子仍然有漏筛危险。然而,可能使搜索更简单,因为许多在 VS 中遇到的参数,在相对比较时可忽略不计。此外,要把 VS 作为常规发现先导化合物的工具,仍有一段相当长的路要走。

(上接第 437 页)

3 结语

各种抗体制剂用作疾病的治疗已超过一个世纪,首先使用的是从合格供血者血浆中提取的免疫球蛋白。这第 1 代抗体药物虽然理想地反映自然免疫反应的多样性,但存在几个缺点,包括安全性差,缺乏特异性和临床应用有限。第 2 代是 mAb 药物,

具有许多优点,包括安全性较高、生产成本低、可重复性好和特异性较强。开发中有希望的 mAb 已超过 100 个。但这些药物的单特异性使其不能用于复杂抗原引起的疾病。第 3 代重组多克隆抗体回答了前 2 代抗体治疗中存在的问题,技术上精细调整后,将很快拓宽其临床使用潜力,在未来的抗体治疗药物市场中将占上风。