

氮杂茈萜衍生物靶向虚拟筛选研究

李悦青¹, 黄雅俊¹, 赵伟杰¹, 王希诚^{*2}

(1. 大连理工大学 精细化工国家重点实验室, 辽宁 大连 116012;

2. 大连理工大学 工业装备结构分析国家重点实验室, 辽宁 大连 116024)

摘要: 为寻求新的具有强生理活性的茈萜类化合物, 设计了 237 个全新结构的氮杂茈小分子. 基于靶标虚拟筛选平台 TarFisDock, 针对病毒类靶标作广泛筛选, 所设计化合物均集中结合于 HIV 蛋白酶的活性位点, 这一结果已经为 HIV 蛋白酶正向对接所证实. 通过统计所设计分子的体积、脂水分分配系数、小分子-蛋白间氢键, 归纳出分子的理化性质与对接结果中结合能量的关系, 并考察了分子结构对能量得分的影响. 筛选结果提示 HIV 蛋白酶可能是氮杂茈衍生物抑制 HIV 的高活性靶标.

关键词: 氮杂茈衍生物; 靶向虚拟筛选; 抗病毒

中图分类号: O625 **文献标志码:** A

0 引言

茈萜类化合物是具有二苯乙烯母核结构的化合物及其聚合物的总称. 天然单体茈萜类化合物的取代基主要是羟基和甲氧基, 取代位置主要在 3、5 位, 也有 4 位异戊烯基和 2 位羧基取代, 结构以反式为主, 目前至少已在 24 个科 53 个属的植物中发现此类物质^[1]. 研究发现, 茈萜类化合物具有广泛的抗病毒活性, 如抑制人类免疫缺陷病毒(HIV)、单纯疱疹病毒、巨细胞病毒、T 淋巴细胞病毒、流感病毒和乙肝病毒^[2~4]等, 其中以抑制 HIV 的研究占多数. 近年来雷小平等^[5,6]先后对侧链氨基取代及氮杂环取代的茈萜衍生物进行过合成与活性研究, 结果表明其均具有抑制癌细胞增殖的作用. 李悦青^[7]曾设计用吡啶、咪唑、苯并咪唑环替代母核结构中的一个苯环, 合成了一系列氮杂环茈萜衍生物, 实验证实其中一些化合物具有一定的抗病毒作用. 为进一步考查氮杂茈衍生物的抗病毒活性, 本文设计 237 个全新的小分子, 采用计算

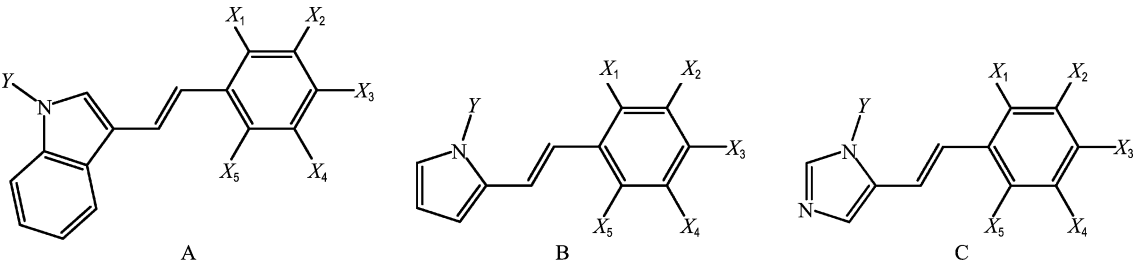
机辅助药物设计中对接方法, 基于 TarFisDock^[8]药物靶标虚拟筛选平台, 在病毒类靶标中进行靶向虚拟筛选; 在计算机模拟的层面, 研究小分子的一些理化性质对小分子-蛋白相互作用的影响.

1 小分子设计及筛选方法

1.1 小分子设计及准备

本文设计的小分子结构见图 1, 母核结构中依次用吡啶、吡咯、咪唑环代替一个苯环(分别对应图 1 中的 A、B、C 三种结构), 同时为了调节分子的脂水分分配系数和体积, 在环上引入各类仲胺取代基、甲氧基、苄氧基、羟基、卤素和烯基等.

利用 ISIS Draw 绘制设计分子的 2D 结构, 并用 SPL(Sybyl programming language)脚本程序转为三维结构, 采用 Tripos 公司的 SYBYL6.9 分子设计软件包给小分子加全氢和 Gasteiger-Huckel 电荷, 选择 TRIPOS 力场优化, 并保存为 mol2 格式文件.



$X_1 \sim X_5 = \text{H}, \text{OH}, \text{OCH}_3, \text{Cl}, \text{OCH}_2\text{Ph}, \text{OCH}_2\text{CH}=\text{C}(\text{CH}_3)_2$
 $Y = \text{CH}_2\text{N}(\text{CH}_3)_2, \text{CH}_2\text{N}(\text{CH}_2\text{CH}_3)_2, \text{CH}_2\text{N}(\text{CH}_2\text{CH}_2\text{CH}_3)_2, \text{CH}_2\text{CH}=\text{C}(\text{CH}_3)_2, 1\text{-甲基-四氢吡咯}, 4\text{-甲基-吗啉}$

图 1 设计分子的结构

Fig. 1 Structures of designed molecules

1.2 筛选平台及范围

应用药物靶标筛选工具 TarFisDock 在 PDTD^[9] 数据库的病毒类靶标(viral infections)中进行反对接虚拟筛选,其中涉及 18 类药物靶点的 69 个靶标蛋白:HIV Protease 28 个, HIV Reverse Transcriptase 6 个, HIV Integrase 2 个, Neuraminidase 9 个, Thymidylate Synthetase 4 个, Thymidine Kinase 4 个, Pokeweed Antiviral Protein 2 个, Human Class I Histocompatibility Antigen 2 个, Rhinovirus 2 个, Nonstructural Protein 2 个, Poliovirus 1 个, Thymidine Monophosphate Kinase 1 个, Holliday Junction DNA Helicase Ruvb 1 个, Influenza Virus Hemagglutinin 1 个, Intercellular Adhesion Molecule 1 个, Inosine-5'-monophosphate Dehydrogenase 1 个, SARS Coronavirus Main Proteinase 1 个, Undefined 1 个.将 237 个小分子的 mol2 文件逐一提交,任务返回的计算结果按能量得分 *E* 排序,每个小分子输出结合前十位的靶标蛋白,统计结果见表 1.

2 结果与讨论

从表 1 可以看出,靶向虚拟筛选结果中仅出现了 6 类药物靶点中的 39 个靶标蛋白.其中, HIV 蛋白酶(HIV Protease)出现的次数最多,共计 2 307 次,占总次数的 97.34%;排第二位的是唾液酸苷酶(Neuraminidase),出现了 52 次,为 2.19%,显示出高度集中于一类靶点(HIV 蛋白酶)的趋势.虽然这与 HIV 蛋白酶靶标蛋白所占比例较大有一定关系(共 28 个,40.58%),但揭示出靶向虚拟筛选层面上的靶标单一性.

表 1 靶向虚拟筛选中各靶标蛋白出现的次数

Tab. 1 Frequency of the proteins in the targeted virtual screening

蛋白编号	蛋白类型	出现次数
1HVR	HIV Protease	173
1AAQ	HIV Protease	139
1HTG	HIV Protease	137
7HVP	HIV Protease	131
1HVI	HIV Protease	122
1AJV	HIV Protease	114
2UPJ	HIV Protease	113
1HPS	HIV Protease	112
1HSG	HIV Protease	106
1IDA	HIV-2 Protease	103
1HPX	HIV Protease	96
1HVJ	HIV Protease	96
1HIH	HIV Protease	91
4PHV	HIV Protease	91
1HVL	HIV Protease	86
1HTF	HIV Protease	79
1HVK	HIV Protease	75
1DIF	HIV Protease	74
1AJX	HIV Protease	70
1HBV	HIV Protease	61
9HVP	HIV Protease	61
1TCX	HIV Protease	60
1HOS	HIV Protease	50
1HPV	HIV Protease	48
1IVE	Neuraminidase	25
7UPJ	HIV Protease	17
1TSD	Thymidylate Synthetase	7
1IVD	Influenza A Subtype N2 Neuraminidase	7
1NNC	Neuraminidase Subtype N9 (TERN)(NA)	7
1ING	Influenza A Subtype N2 Neuraminidase	4
1HTE	HIV Protease	2
1NNB	Neuraminidase	2
1NSC	Neuraminidase	2
2SIM	Neuraminidase	2
3HVT	Reverse Transcriptase	2
1INF	Influenza Virus B/Lee/40 Neuraminidase	2
1NSD	Neuraminidase	1
2PLV	Poliovirus (Type 1, Mahoney Strain)	1
1UK4	SARS Coronavirus Main Proteinase (SARS-CoV 3CLpro)	1

在出现的 39 个靶标蛋白中,复合物为非肽的 HIV 蛋白酶 1HVR 排在首位,出现的频率是 173 次,能量得分介于 $(-54.11\sim-31.7)\times 4.186\ 8$ $\text{kJ}\cdot\text{mol}^{-1}$,表明约 73%的虚拟筛选分子与 HIV 蛋白酶的非肽活性抑制位点可以结合,提示该位点可能是氮杂茛苈生物抑制 HIV 的高活性位点.

为了检验靶向筛选结果的可靠度,根据反对接计算结果,针对出现次数最多的靶标 1HVR,采用 Kuntz 研究组开发的 DOCK6.1 程序与设计分子进行正向对接计算:从 PDTD 获得蛋白 PDB 和 mol2 文件,用 dms 程序计算活性位点区的分子表面,sphgen 生成描述活性位点信息的球集,sphselector 选取距离原 1HVR 晶体复合物中配体原子均方根偏差 0.3 nm 范围内的球集,showbox 产生球集坐标外延 0.5 nm 的网格范围文件,grid 计算格点的能量、接触及碰撞文件.所有计算在 SGI 工作站上完成,统计结果见表 2.

表 2 小分子与 1HVR 的结合能量得分分布
Tab. 2 Energy score distribution of complex combined by small molecule and 1HVR

得分范围/ $(4.186\ 8\ \text{kJ}\cdot\text{mol}^{-1})$	小分子数
$-55\leq E<-50$	6
$-50\leq E<-45$	81
$-45\leq E<-40$	95
$-40\leq E<-35$	35
$-35\leq E<-30$	13
$-30\leq E<-25$	1
$-25\leq E<-20$	0
$-20\leq E<0$	2
$E\geq 0$	4

从表 2 可以看出,小分子与 1HVR 的结合能量得分集中在 $(-55\sim-30)\times 4.186\ 8$ $\text{kJ}\cdot\text{mol}^{-1}$,共 230 个,约占全体小分子的 97%,这与反对接计算的能量得分范围基本相符.随后考查了小分子的理化性质及结构对结合的影响.

2.1 理化性质对结合的影响

2.1.1 分子体积对结合的影响 用 DS Viewer Pro5.0 计算小分子的体积 V ,介于 $(134.7\sim 449.6)\times 10^{-3}\ \text{nm}^3$,统计分子体积与得分之间的

变化关系,见图 2.

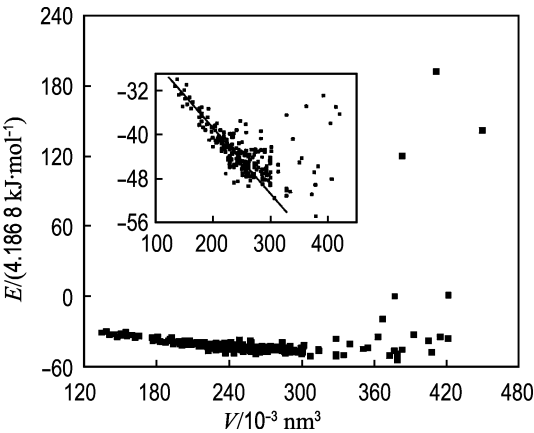


图 2 分子体积对对接结果的影响

Fig. 2 Effect of molecular volume on docking result

可以看出,在一定范围内,随着分子体积的增大,能量得分呈下降的趋势.体积太大或太小,都不利于与活性位点充分结合,并导致能量增加,表 2 中得分较差的几个分子都属于这种情况.图 2 表明,当分子体积在 $(200\sim 300)\times 10^{-3}\ \text{nm}^3$ 时,会取得比较好的对接结果.

2.1.2 脂水分配系数对结合的影响 脂水分配系数(P)是反映小分子疏水性的参数,也是影响药物吸收的一个重要性质,利用 Chem3D Ultra 8.0 计算小分子的 $\text{Clog } P$,考查它与对接结果之间的关系,结果见图 3.两者之间并没有表现出明显的数值变化关系, $\text{Clog } P$ 分布在 1.247 2~10.852 2.对超过 3 000 个上市药物的研究发现,药物分子的 $\text{Clog } P$ 主要集中在 0~5.图 3 统计结果显示,有 137 个(57.8%)分子的 $\text{Clog } P$ 位于 0~5,此区间能量得分基本分布在 $(-50\sim-30)\times 4.186\ 8$ $\text{kJ}\cdot\text{mol}^{-1}$,说明分子设计比较合理.

2.1.3 氢键对结合的影响 用 DS Viewer Pro5.0 显示小分子与活性位点周围的残基形成的氢键,在结合构象中与 HIV 蛋白酶残基形成氢键的分子有 23 个,得分前 50 位的分子仅占 5 个.可以看出,氢键在小分子与蛋白结合过程中所起的作用较小.

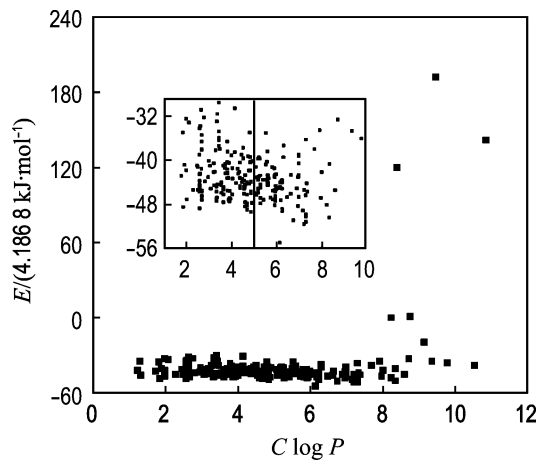


图 3 脂水分配系数与对接结果之间的关系
Fig. 3 Relationship between Clog *P* and docking result

2.2 结构对结合的影响

2.2.1 氮杂环对结合的影响 根据母核结构中氮杂环的不同将设计分子分为 A、B、C 三类,把除杂环外其他结构相同的 3 个分子列为一组(共 78 组,不能列为一组的 3 个小分子不计),每组内按对接得分优劣顺序排名,统计各组中 3 种结构的得分排名情况,结果见表 3.

表 3 三种结构得分排名所占比率
Tab. 3 Ranking ratio of three kinds of structures

排名	A 结构	B 结构	C 结构
1	46	5	27
2	12	26	40
3	20	47	11

从表 3 可以看出,具有 A 结构的分子总体对接结果最好,而含 B 结构的最差.即取代基相同时,3 种结构对接结果优劣排序依次为吡啶环>咪唑环>吡咯环.

2.2.2 取代基对结合的影响

(1)取代基相同,取代位置不同

选取具有相同组成、苯环上取代基为两个甲氧基的 84 个小分子,根据取代基位置不同分为 4 组(每组 21 个),每组内按对接得分优劣顺序排名,考查取代位置对得分的影响,统计结果见表 4.

从表 4 可以看出,3,5-取代时结合得分整体情况最好,2,6-取代时则最差.反映出在立体异构体中,小分子与蛋白结合时的低能结构和单独存在时的稳定结构的一致性.

表 4 不同位置取代时得分排名所占比率
Tab. 4 Ranking ratio of substituent at different positions

排名	2,4-取代	2,5-取代	2,6-取代	3,5-取代
1	5	5	1	10
2	6	9	3	3
3	7	4	4	6
4	3	3	13	2

(2)取代基不同

采取固定一环(苯环或氮杂环)上的取代基、变化另一环上取代基的方法,比较了对接得分受影响情况.发现苯环取代基不变时,氮杂环上取代基为—CH₂N(C₃H₇)₂的得分总体情况最优,其次为含氮、氧的六元杂环(4-甲基-吗啉)取代.杂环为吡啶环时,苯环上取代基为两个甲氧基的结果最好;杂环为吡咯环时,苯环上取代基为两个甲氧基或两个苄氧基得分较好;杂环为咪唑环时,苯环上取代基为两个苄氧基得分最优.

3 结 论

氮杂茛苈生物物的靶向虚拟筛选结果显示:病毒靶标库中的 HIV 蛋白酶可能是所设计分子的高活性靶标.正向分子对接计算初步证实靶标筛选的精确性.基于正向对接计算结果,本文归纳出了分子结构与 HIV 蛋白酶结合的一些规律.当分子体积在(200~300)×10⁻³ nm³,小分子与蛋白的结合能较低;当分子的 Clog *P* 处于 0~5 时,对接计算的结合能基本分布在(−50~−30)×4.186 8 kJ·mol⁻¹;在对接计算中,观察到的小分子-蛋白间形成氢键较少,在分子设计中分子的氢键给体和受体数目可以不作为主要考查因素;不同氮杂芳香环取代对结合能的影响,提示吡啶类衍生物更具有研究的潜力.

参考文献:

- [1] 斯建勇. 天然芪类化合物的研究概况[J]. 天然产物研究与开发, 1994, **6**(4):71-79
- [2] 夏丽萍, 谢金洲. 白藜芦醇的药理作用进展[J]. 中国误诊学杂志, 2006, **6**(2):220-222
- [3] 丁 宇, 范 红. 天然植物活性成分白藜芦醇的抗炎作用研究进展[J]. 华西医学, 2008, **23**(1):199-200
- [4] 周 侠, 吕秋军, 王 林, 等. 白藜芦醇及其衍生物抗乙型肝炎病毒体外实验研究[J]. 中国药理学杂志, 2005, **40**(24):1904-1906
- [5] 雷小平, 张亮仁, 刘 波. 6-取代氨基-1, 2-双(对氯苯)-己烯-1 类化合物的设计、合成及钙调素拮抗活性的研究[J]. 药学报, 1997, **32**(6):426-430
- [6] 雷小平, 刘 波, 朱 娜, 等. 杂环取代的二苯乙烯类化合物抗癌活性研究[J]. 中国新药杂志, 2000, **9**(8):540-542
- [7] 李悦青. 芪类化合物的设计、合成及其活性研究[D]. 大连: 大连理工大学, 2006
- [8] LI Hong-lin, GAO Zhen-ting, KANG Ling, *et al.* TarFisDock: a web server for identifying drug targets with docking approach [J]. **Nucleic Acids Research**, 2006, **34**:219-224
- [9] GAO Zhen-ting, LI Hong-lin, ZHANG Hai-lei, *et al.* PDTD: a web-accessible protein database for drug target identification [J]. **BMC Bioinformatics**, 2008, **9**(1):104-110

Research on targeted virtual screening of aza stilbene derivatives

LI Yue-qing¹, HUANG Ya-jun¹, ZHAO Wei-jie¹, WANG Xi-cheng^{*2}

(1. State Key Laboratory of Fine Chemicals, Dalian University of Technology, Dalian 116012, China;

2. State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116024, China)

Abstract: To seek new stilbene analogs with high activity, 237 aza stilbene derivatives with whole new structures are designed. During screening targets of viral infection supported by the target-based virtual screening service of TarFisDock, most of the designed molecules can act with the active site of HIV protease. The reliability of the result is also confirmed by docking HIV protease and 237 molecules. The effects of some physicochemical properties on the docking energy score, such as molecular volume, octanol/water partition coefficient, hydrogen bond formed between small molecules and protein are explored. And molecular structure factors affecting docking results are also discussed. Screening results suggest that HIV protease may be the highly-active target of aza stilbene derivatives to inhibit HIV.

Key words: aza stilbene derivative; targeted virtual screening; anti-virus