

探索应用反向对接技术研究蛋白激酶抑制剂选择性的可行性

刘廷林^{1,2}, 谢焕章², 魏于全^{1,2}, 杨胜勇^{2*}

(1. 四川大学生命科学学院, 四川 成都 610065; 2. 四川大学华西医院生物治疗国家重点实验室, 四川 成都 610041)

摘要: 本文探索了应用反向对接技术研究激酶抑制剂选择性的可行性。首先通过收集晶体结构数据或使用同源建模方法建立了包含 422 个激酶的靶点三维结构数据库, 在此基础上建立了反向对接激酶靶点筛选方法, 并对相关参数和打分函数进行了优选。最后, 选取了 7 个典型的选择性激酶抑制剂作为例子, 用于测试该反向对接激酶靶点筛选方法。结果表明, 这些激酶抑制剂的选择性作用靶点均具有较高的打分值 (打分值均排在所测试激酶的 35% 以内)。本文的研究提示, 反向对接技术可以应用于激酶靶点虚拟筛选, 并进而应用于激酶抑制剂的选择性研究。

关键词: 激酶抑制剂; 选择性; 分子对接; 反向对接

中图分类号: R916

文献标识码: A

文章编号: 0513-4870 (2009) 07-0758-06

The feasibility of application of reverse docking method to the selectivity studies of protein kinase inhibitors

LIU Ting-lin^{1,2}, XIE Huan-zhang², WEI Yu-quan^{1,2}, YANG Sheng-yong^{2*}

(1. College of Life Science, Sichuan University, Chengdu 610065, China; 2. State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu 610041, China)

Abstract: This investigation is to explore the feasibility of applying reverse docking method to the selectivity studies of protein kinase inhibitors. Firstly, a database that consists of 422 protein kinase structures was established through collecting the reported crystal structures or homology modeling. Then a reverse docking based method of protein kinase target screening was established, followed by the optimization of related parameters and scoring functions. Finally, seven typical selective kinase inhibitors were used to test the established method. The results show that the selective targets of these inhibitors have relatively high scoring function values (ranking in the first 35% of the tested kinase targets according to the scoring function values). This implies that the reverse docking method can be applied to the virtual screening of kinase targets and further to the selectivity studies of protein kinase inhibitors.

Key words: kinase inhibitor; selectivity; molecular docking; reverse docking

激酶在细胞的生命过程中起着至关重要的作用, 其功能异常会导致肿瘤、糖尿病、炎症等许多重大疾病的发生^[1,2]。鉴于此, 激酶已作为最重要的疾病治疗靶点之一成为当前研究的热点。据统计, 目前全世界在研或开发项目中的药物约三分之一与激酶相关^[3],

4]。

根据人类基因谱图, Manning 等^[5]确定了 518 种人类激酶, 其中 478 种激酶都有一个序列和结构高度保守的催化域 (又称激酶域), 而该催化域正是 ATP 竞争性抑制剂的作用位点, 这就使得高选择性的激酶 ATP 竞争性抑制剂的获取非常困难^[6]。而激酶抑制剂的选择性直接关系到药物的毒副作用, 这对于后期药物开发具有重大影响。另外, 高选择性的激酶抑制剂也可作为探针分子用于研究激酶介导的细胞信

收稿日期: 2009-02-02.

基金项目: 国家自然科学基金资助项目 (30772651, 20872100).

*通讯作者 Tel: 86-28-85164063, Fax: 86-28-85164060,

E-mail: yangsy@scu.edu.cn

号转导通路。因此, 选择性激酶抑制剂的研发近年来变得炙手可热^[7-10]。

选择性激酶抑制剂研发的核心问题是如何确定抑制剂的选择性, 当前普遍采用的方法是对研究对象(抑制剂)在酶水平上直接进行多个激酶的活性抑制实验, 从而获得激酶活性抑制谱^[10]。为了获得完整的激酶活性抑制谱, 需要对所有激酶进行活性抑制实验, 但其成本昂贵。因此, 为了节约成本, 常规的做法是只选择部分激酶进行实验, 而这样得到的选择性是不全面的, 可信度不高。为了同时兼顾成本和可信度, 作者提出, 在进行实验之前, 首先应用分子反向对接技术对激酶进行初步筛选(反向对接技术^[11, 12], 即使用同一个小分子分别对接到多个激酶的活性位点), 选取具有较高打分函数值的激酶, 再对这些选取的激酶开展酶水平试验, 这样可以减少所需实验测试的激酶数量, 在保持一定可信度的基础上降低实验研究成本。

然而, 如前所述, 激酶序列和结构具有高度保守性, 这使得人们怀疑反向对接技术是否可用于激酶的虚拟筛选。虽然序列和结构具有高度保守性, 但每个激酶在序列和结构方面仍然存在着不同程度的差异, 这些差异最近已被作为设计选择性激酶抑制剂的依据^[13], 并且已有成功的例子, 例如被 FDA 批准上市用于治疗肿瘤的激酶抑制剂 Imatinib (Gleevec)、Gefitinib (Iressa)、Erlotinib (Tarceva)、Sorafenib (Nexavar)、Dasatinib (Sprycel) 等, 正是由于其作用靶激酶与其他激酶在序列和结构上的差异而显示出选择性^[14-16]。这些提示我们, 通过反向对接技术对激酶进行虚拟筛选是有可能实现的。本文研究的主要目的就是初步探索这种方法的可行性, 为后续进一步的深入研究奠定基础。

材料与方法

激酶三维结构的收集与建模 所有已报道的激酶结构数据均从 Protein Data Bank (PDB) 数据库中获得, 对其中不完整的三维结构坐标用 DeepView^[17]或 Chimera^[18]程序补全。对于没有晶体结构报道的激酶, 则使用 SWISS-MODEL 程序^[17]通过同源建模方法建立, 其同源建模模版为与其同源性最高的已知晶体结构的激酶, 所有经过建模的蛋白结构均使用 Accelrys 公司的 InsightII 软件包进行了优化。

反向对接 本研究使用一个自编的反向对接驱动和接口程序 TS (Target Screening), 而对接操作则

调用最新版的 DOCK (版本 6.2)^[19, 20]核心模块实现(该程序简称为 TS-DOCK)。蛋白质的分子表面信息数据由 DMS 程序生成, 距小分子配体 0.65 nm 的距离范围定义为活性位点区域, 活性位点通过 SPHGEN 程序产生, 使用格点生成程序 GRID 生成受体格点文件。将所有激酶蛋白采用上述方法处理后储存为靶点库。

使用 Chimera 程序对小分子化合物加氢, 使用 Gasteiger 电荷。小分子化合物以 MOL2 文件的形式提交给 TS-DOCK, 由 TS-DOCK 实现小分子与靶点库中所有激酶的自动对接及打分排序。

结果与讨论

1 激酶靶点数据库的构建

从晶体结构数据库中收集了 49 个激酶的三维结构数据, 通过同源建模方法建立了 373 个激酶的三维结构(为使通过同源建模获得的激酶结构的可信度更高, 只对与已知结构的激酶具有 30%以上同源性的未知结构的激酶进行了同源建模), 所有这些三维结构(共 422 个)构成了激酶靶点数据库, 用于后续的虚拟筛选研究。

2 对接参数与打分函数的优选

在使用反向对接进行激酶靶点虚拟筛选之前, 必须先对分子对接中的相关参数和打分函数进行优化。

首先从已知晶体结构的激酶中选取了 6 个具有较高分辨率的激酶-配体复合物, 即 Aurora A-2J4Z (前者为激酶名称, 后者为 PDB 号)、ABL1-2F4J、TGF β R1-1PY5、p38 α -1YQJ、CDK5-1UNL 和 GSK3 β -1J1B。先将这些复合物中的小分子配体取出, 然后使用 TS-DOCK 程序将该配体小分子对接回去, 并调整受体活性位点的大小、配体的最大取向数、bumps 过滤参数、Simplex 优化参数等相关参数, 以获得最佳的对接效果, 即对接后的配体与晶体结构中的配体的 RMSD (Root Mean Square Deviation) 值最小。图 1 给出了最终对接后的小分子与激酶复合物的叠合结果。由图 1 可以看出, 对接后的小分子与激酶复合物的叠合情况良好(所有 RMSD 值均小于 0.18 nm)。

为优选打分函数, 作者选取了 10 个激酶, 每个激酶分别选取 3 个配体分子, 配体分子选取的原则是该配体分子已有实验活性数据 (IC₅₀) 报道, 另外, 三者的活性数据之间至少有一个数量级的差异。30

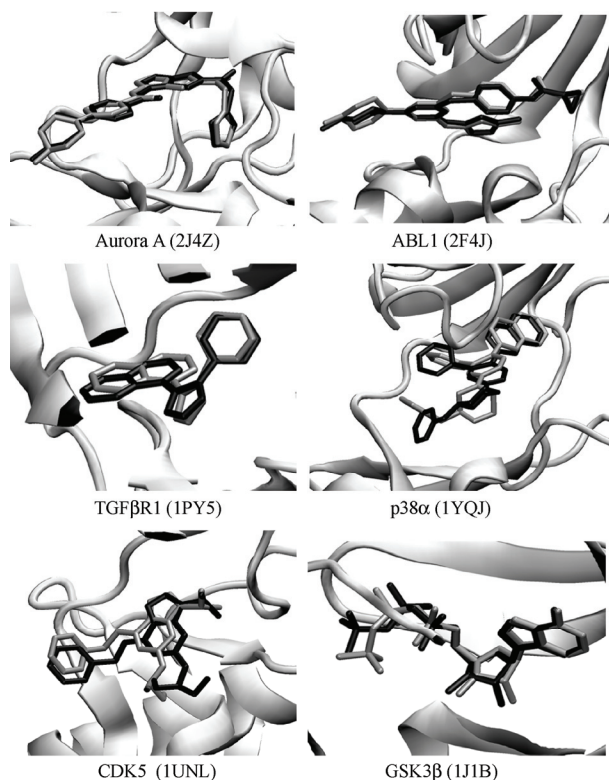


Figure 1 Superposition of docked ligands with their corresponding ones in co-crystallized complexes. Dark: docked ligands, Light: co-crystallized ligands

个配体分子的结构如图 2 所示。作者首先将这些小分子对接到对应的靶点, 然后使用 DOCK 程序中提供的 Contact、Grid-based、Continuous、GB/SA 和 Amber 5 种打分函数分别进行打分。其中, Contact Score 是根据受体配体的形状匹配来进行相互作用评价; Grid-based Score 是在分子力场下基于受体配体之间的结合自由能来评价相互作用, 而结合能又包括静电相互作用和范德华力的贡献; Continuous Score 是一种不用事先构建格点的简易打分方式, 适合单次对接评价; GB/SA Score 是一种考虑了溶剂可及表面影响的打分函数; Amber Score 是加入分子动力学和共轭梯度优化的打分函数。本实验统计打分函数表现的规则是: 对于每一个激酶靶点, 如果使用各打分函数产生的排序与实验活性值排序一致, 则得分计为 1, 否则计为 0。最后由 10 个得分值加和得总分 (total score)。总分值越高, 表明该打分函数对所有激酶靶点的总体表现越好。表 1 给出了 5 种打分函数的表现情况, 从表 1 可见, Grid-based 打分函数有最好的表现, 即 10 个靶点中有 9 个靶点的小分子对接得分值排序与其活性一致, 仅有一个靶点例外, 而且, 在各打分函数中, Grid-based 是速度最快的打分

函数之一, 仅次于 Contact Score, 而打分效果却远优于 Contact Score。因此, 本研究采用 Grid-based 打分函数。

Table 1 Performance of the selected five scoring functions in the scoring function test

No.	Kinase	Contact	Grid-based	Continuous	GB/SA	Amber
1	ABL1	1	1	1	1	1
2	CDK2	0	1	1	0	1
3	CHK2	1	1	1	1	1
4	ErbB2	1	1	0	0	1
5	Erk1	0	0	1	1	0
6	FGFR1	1	1	0	1	0
7	c-Kit	1	1	1	1	1
8	p38-alpha	0	1	1	1	1
9	PKCa	0	1	0	1	1
10	PLK1	1	1	1	1	1
Total score		6	9	7	8	8

1: If the predicted rank is consistent with that of the experimental data; 0: If the predicted rank is not consistent with that of the experimental data; Total score: The sum of ten scores

3 应用反向对接技术确定激酶抑制剂的选择性

Karaman 等^[10]使用同位素标记法测试了目前已上市或正在进行临床试验的 38 个小分子激酶抑制剂对 317 个激酶的抑制活性谱, 发现若干个具有较高选择性的抑制剂分子。本实验选取了其中选择性最高的 7 个抑制剂分子, 即 Lapatinib、GW-2580、PTK-787、CP-690550、CP-724714、SB-431542 和 VX-745 (图 3)。由于 Karaman 等使用的 317 个激酶中不仅包括了 287 个蛋白激酶 (非突变型), 也包括 27 个突变型蛋白激酶和 3 个酯酶, 在本次测试中, 为使问题简化, 作者仅选择了 287 个非突变蛋白激酶构成靶点库。分别将这 7 个小分子通过 TS-DOCK 反向对接程序进行靶点筛选, 并使用 Grid-based 打分函数进行打分排序。

表 2 给出了 7 个选择性的小分子抑制剂及其选择性作用靶点、实验 IC₅₀ 值、分子对接打分和打分值排序, 其中, 打分值排序 (表 2 中的 rank by TS Dock) 是该激酶的对接得分在所有参加测试的激酶靶点中所处的位次。由表可见, GW-2580 和 Lapatinib 仅有一个选择作用靶点, 其对应分子对接得分值的排序分别为 31 和 36。VX-745 和 CP-724714 的选择性作用靶点均有两个, 活性较高的对接打分排序也相对靠前。SB-431542、CP-690550 和 PTK-787 的选择性稍差, 这里仅给出了抑制活性较高 (小于

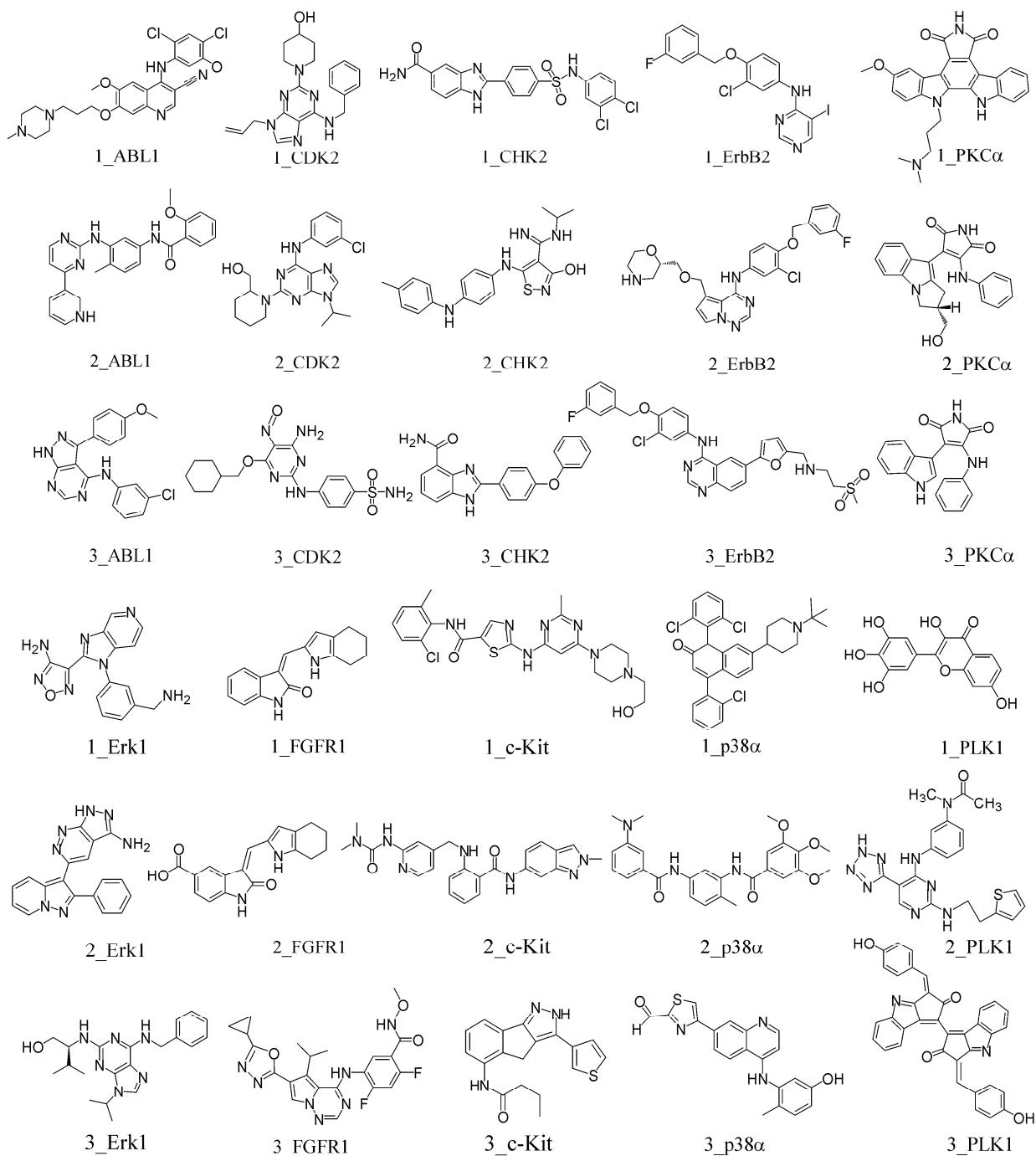


Figure 2 Thirty small molecular kinase inhibitors used in the scoring function test

200 nmol·L⁻¹) 的作用靶点, 结果显示这些靶点均排在了前 100 名, 即在所测试激酶数目的 35% 以内。这提示, 在使用实验方法开展激酶抑制剂的选择性测试之前, 可以首先通过反向对接技术对激酶进行虚拟筛选, 只选取打分值较高者进行实验测试, 从而在一定程度上降低实验成本并节约测试时间。

通过反向对接会产生许多“假阳性”的靶点, 其

原因主要来自于两个方面: ① 目前分子对接方法和打分函数还存在许多缺陷; ② 蛋白激酶结构的保守性较高, 分子对接方法对这种相似性太高的激酶活性区域的区分困难。在实际应用中, 当决定选择哪些激酶用于选择性测试实验时, 可以首先排除掉那些打分值太低的激酶 (如打分值排序在后 50% 的激酶), 在剩下的激酶中, 再根据打分值的区间分布, 并结合其他实验数据或经验选择。

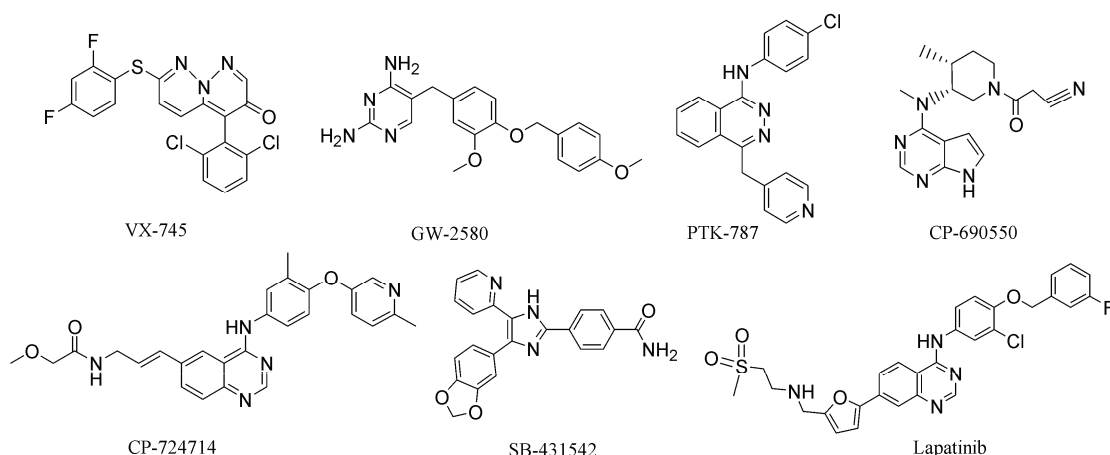


Figure 3 Seven kinase inhibitors with a relatively high selectivity

Table 2 Selected seven small molecular kinase inhibitors together with their acting targets, experimental IC_{50} values, docking scoring function values and ranks

Inhibitor	Selective Targets	IC_{50} /nmol·L ⁻¹	Dock Score	Rank by TS-Dock
PTK-787	Kit	5.1	-46.640789	43
	FLT1	9.6	-50.156010	9
	CSF1R	18	-44.726871	78
	PDGFRB	25	-44.636539	80
	VEGFR2	62	-44.195251	92
	PDGFRA	96	-44.178055	93
VX-745	p38- α ph	2.8	-49.556278	36
	p38- β ph	74	-46.608341	97
CP-724714	EGFR	42	-61.442013	31
	ErbB2	43	-57.848377	69
CP-690550	JAK3	2.2	-45.676849	34
	DCAMKL3	4.5	-43.147728	82
	JAK2	5	-43.494331	73
SB-431542	TGFBR1	170	-52.975182	11
	ACVR1B	190	-51.167480	19
	CSNK1D	170	-47.083771	82
Lapatinib	EGFR	2.4	-65.667709	36
GW-2580	CSF1R	1.6	-52.520477	31

结论

本文探索了使用反向对接技术进行激酶抑制剂选择性研究的可行性。首先建立了包含 422 个激酶三维结构的数据库，并对反向对接的参数和打分函数进行了优选。然后选取了 7 个典型的选择性激酶抑制剂作为例子，用于测试该反向对接激酶靶点筛选方法。结果表明，这些激酶抑制剂的选择性作用靶点均具有较高的打分值（打分值排在所测试激酶的 35% 以内）。这些研究结果提示，在使用实验方法进行激酶抑制剂的选择性测试之前，可以首先通过反向对接技术对激酶进行初筛，选择具有较高打分值的激

酶进行酶水平的抑制活性测试。

总之，虽然激酶的同源性较高，但每个激酶在序列和三维结构上仍存在一定的差异，利用这些差异，通过反向对接技术进行激酶抑制剂选择性研究具有一定可行性。在下一步的研究中，作者将进一步优化并完善激酶靶点数据库，评测并优化对接策略，优选打分函数等，以提高激酶虚拟筛选的准确率，最终实现实用的基于反向对接的激酶靶点虚拟筛选，并将之应用于激酶抑制剂选择性研究。

References

- [1] Giamas G, Stebbing J, Vorgias CE, et al. Protein kinases as targets for cancer treatment [J]. *Pharmacogenomics*, 2007, 8: 1005–1016.
- [2] Noble ME, Endicott JA, Johnson LN. Protein kinase inhibitors: insights into drug design from structure [J]. *Science*, 2004, 303: 1800–1805.
- [3] Liao JJ. Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors [J]. *J Med Chem*, 2007, 50: 409–424.
- [4] Doerig C, Meijer L. Antimalarial drug discovery: targeting protein kinases [J]. *Expert Opin Ther Targets*, 2007, 11: 279–290.
- [5] Manning G, Whyte DB, Martinez R, et al. The protein kinase complement of the human genome [J]. *Science*, 2002, 298: 1912–1934.
- [6] Scapin G. Protein kinase inhibition: different approaches to selective inhibitor design [J]. *Curr Drug Targets*, 2006, 7: 1443–1454.
- [7] Dancey J, Sausville EA. Issues and progress with protein kinase inhibitors for cancer treatment [J]. *Nat Rev Drug Discov*, 2003, 2: 296–313.

- [8] Peng T, Pei JF, Zhou JJ. Three-dimensional quantitative structure-activity relationship study of tyrosine kinase inhibitors [J]. *Acta Phys Chim Sin* (物理化学学报), 2003, 19: 163–166.
- [9] Kéri G, Orfi L, Eros D, et al. Signal transduction therapy with rationally designed kinase inhibitors [J]. *Curr Signal Transduct Ther*, 2006, 1: 67–95.
- [10] Karaman MW, Herrgard S, Treiber DK, et al. A quantitative analysis of kinase inhibitor selectivity [J]. *Nat Biotechnol*, 2008, 26: 127–132.
- [11] Chen YZ, Zhi DG. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule [J]. *Proteins: Struct Funct Genet*, 2001, 43: 217–226.
- [12] Li HL, Gao ZT, Kang L, et al. TarFisDock: a web server for identifying drug targets with docking approach [J]. *Nucleic Acids Res*, 2006, 34: 219–224.
- [13] Deng XQ, Xiang ML, Jia R, et al. Progress in the design of selective ATP-competitive kinase inhibitors [J]. *Acta Pharm Sin* (药学报), 2007, 42: 1232–1236.
- [14] Thaimattam R, Banerjee R, Miglani R, et al. Protein kinase inhibitors: structural insights into selectivity [J]. *Curr Pharm Design*, 2007, 13: 2751–2765.
- [15] Vulpetti A, Bosotti R. Sequence and structural analysis of kinase ATP pocket residues [J]. *Farmaco*, 2004, 59: 759–765.
- [16] Wu W, Lu C, Chen SY, et al. The signal transduction pathway of multi-target kinase inhibitors as anticancer agents in clinical use or in phase III [J]. *Acta Pharm Sin* (药学报), 2009, 44: 242–257.
- [17] Arnold K, Bordoli L, Kopp J, et al. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling [J]. *Bioinformatics*, 2006, 22: 195–201.
- [18] Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—a visualization system for exploratory research and analysis [J]. *J Comput Chem*, 2004, 25: 1605–1612.
- [19] Kuntz ID, Blaney JM, Oatley SJ, et al. A geometric approach to macromolecule-ligand interactions [J]. *J Mol Biol*, 1982, 161: 269–288.
- [20] Moustakas DT, Lang PT, Pegg S, et al. Development and validation of a modular, extensible docking program: DOCK 5 [J]. *J Comput Aided Mol Des*, 2006, 20: 601–609.

《Goodman & Gilman 药理学和治疗学手册》已出版

科学出版社于 2009 年 4 月出版的中文版《Goodman & Gilman 药理学和治疗学手册》(Goodman & Gilman's Manual of Pharmacology and Therapeutics) 是《Goodman & Gilman 治疗学的药理学基础》(第 11 版) 的缩编本, 涵盖了临床药理学基础知识, 并概括了药物代谢动力学、药物效应动力学及药理学的重点内容; 吸纳了药理学、治疗学及相关学科的最新成果, 及时介绍了各权威学术机构针对相关疾病制订或修订的最新诊疗指南和用药原则, 并讨论了重大的患者特异性问题; 选择性推荐了美国 FDA 批准的部分新药或药物的新用途。定价 298.00 元。

联系人: 温晓萍 电话: 010-64034601, 传真: 010-64019761