

羟氨酸类组蛋白去乙酰化酶抑制剂的 分子对接及 CoMFA 研究

张 亮, 相玉红, 张卓勇
(首都师范大学化学系, 北京 100048)

摘要 采用 Surflex-dock 对接方法将组蛋白去乙酰化酶抑制剂与组蛋白酶进行对接, 通过配体与受体的相互作用揭示出此种抑制剂的结合模式. 研究表明, 对接自由能与抑制剂活性有较好的相关性. 利用比较分子立场分析方法(Comparative molecular field analysis, CoMFA)对此类抑制剂进行了三维定量构效关系研究, 并考察了不同步长、探针原子以及电荷对构建模型的影响. 所建最佳模型交叉验证相关系数 $q^2 = 0.668$, 非交叉验证相关系数 $r^2 = 0.984$, 说明该模型预测能力较好, 可用于指导新型药物合成.

关键词 组蛋白去乙酰化酶; Surflex-dock; 比较分子力场分析法; 抑制剂

中图分类号 O641

文献标识码 A

文章编号 0251-0790(2009)S0-0052-06

肿瘤是仅次于心血管疾病的第二大“杀手”.^[1] 细胞生长和分化需要一系列基因有序激活, 此过程受染色质结构的调控, 染色质重构异常破坏了这种有序的过程, 导致细胞恶变、异常增生^[2]. 与肿瘤的发生和转移相关的蛋白种类繁多, 而对肿瘤细胞生长调控具有普遍生物学意义的蛋白才最有可能成为广谱低毒抗肿瘤药物的作用靶点^[3]. 组蛋白酶是调节染色体结构和基因表达的关键激酶, 组蛋白的乙酰化状态在基因转录过程中起主要调控作用. 组蛋白的乙酰化状态主要由组蛋白去乙酰化酶(Histone Deacetylase, HDAC)和组蛋白乙酰基转移酶(Histone acetyl transferases, HAT)决定, 当 HDAC 表达明显增强时, 就会造成组蛋白乙酰化状态失衡, 从而导致肿瘤和癌症的发生^[4~6]. 初步的临床研究表明, 人类可以安全地通过抑制 HDAC 的活性来获得组蛋白去乙酰化水平^[7].

目前, 人的 HDAC 的晶体结构尚未得到, 但有文献报道^[8]已经获得了一种古细菌的组蛋白去乙酰化酶类似蛋白(Histone deacetylase-like protein, HDLP)的 X 射线晶体结构, 且 HDLP 与人的 HDAC 氨基酸序列相似, 属于同源蛋白, 从而根据 HDLP 的结构确定了 HDAC 的活性位点, 并揭示出组蛋白去乙酰化酶抑制剂的作用机制.

组蛋白去乙酰化酶抑制剂(Histone Deacetylase Inhibitor, HDACI)可诱导细胞死亡、分化和(或)细胞周期停滞, 调控宿主免疫应答和肿瘤脉管系统, 影响肿瘤生长, 其对细胞的多种效应可相互协同介导抗肿瘤活性^[5]. 目前, HDAC 抑制剂大部分处于临床前研究阶段, 部分已进入临床研究. 根据分子中活性基团的结构特点, 可以将 HDAC 抑制剂分为 4 类^[6~8]: (1) 短链脂肪酸类(Short chain fatty acid, SCFA), 如苯丁酸(Phenylbutyrate)和丙戊酸(Valproic acid); (2) 羟氨酸类(Hydroxamic acid), 如 Trichostatin A(TSA), Suberoylanilide hydroxamic acid(SAHA), Scriptaid 和 Oxamflatin 等; (3) 环肽类, 如 Trapoxin 和 Depsopeptide 等; (4) 苯酰胺类化合物, 如 MS-275.

羟氨酸类抑制剂是目前研究最多也是最深入的一类抑制剂. 此类抑制剂结构主要包括金属螯合区、疏水区 and 连接区域 3 部分. 金属螯合区一般为氮氧原子, 能够与蛋白中的金属离子发生螯合作用, 疏水区一般为烃基, 深入到蛋白口袋的底部与氨基酸残基相互作用, 中间通过碳链将两部分连接起来. 本文针对此类抑制剂进行了研究, 通过其与蛋白的相互作用建立模型, 揭示出此类抑制剂的主要结合模式, 并预测新型抑制剂的结构.

收稿日期: 2009-07-28.

基金项目: 北京市优秀人才资助项目(批准号: 20081D0501600194)资助.

联系人简介: 张卓勇, 男, 博士, 教授, 主要从事分子光谱和化学计量学方面研究. E-mail: gusto2008@vip.sina.com

1 计算部分

Surflex-Dock 分子对接方法是基于生物已知受体三维结构的药物设计方法之一, 采用经验打分函数和搜索引擎将小分子对接到受体的活性位点处, 并寻找合理的取向和构象使配体和受体的几何形状和相互作用匹配达到最佳, 通过打分直接揭示出药物结构与活性的关系, 计算结果较为准确. 打分函数主要包括疏水、极性、排斥力、熵及溶剂化作用.

CoMFA 方法是研究系列化合物三维定量构效关系最为标准和通用的方法. 其基本原理是: 系列化合物由于结构相似, 往往以相同方式作用于受体的结合部位, 配体和受体之间存在着非键相互作用力, 其作用强弱的变化来自于每个化合物周围分子静电场和立体场的变化. CoMFA 能够充分考虑分子的三维结构信息, 通过计算分子周围一定范围内三维空间中假想网状格点上的立体场和静电场参数, 再用偏最小二乘法(PLS) 寻找这些三维特征信息与化合物活性的联系. 这是一种基于小分子结构的间接药物设计方法^[9].

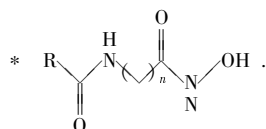
本文所有工作均在美国 Tripos 公司的分子模拟建模软件包 Sybyl 7.3 上完成. 化合物的结构输入和优化采用 Sybyl 7.3 中的 Build/edit 下的 Sketch 程序, 计算中除特别指出外, 所用参数均为缺省值.

1.1 Surflex-dock 分子对接方法

26 个抑制剂分子均选自文献[10,11], 结构见表 1. 在 Sybyl 7.3 中构建出 26 个分子的结构, 加氢并负载电荷, 采用 Tripos 力场进行能量优化 1000 次, 能量收敛标准为 0.05 kcal/mol. 所用对接的 HDAC 蛋白分子从 PDB 数据库上查得, 序列号为 1C3R. 1C3R 为二聚体, 包括 A, B 两条链, 删除其中的水和杂质离子, 对其加氢加电荷(MMFF94), 通过原复合物中的 TSN504 配体形成原型分子, 进行对接. 每个分子将产生 10 个构象, 选出结合方式和结合能都较好的分子进行研究.

Table 1 Structural and biological activity indices of 26 compounds*

Comp.	R	n	pIC50(Exp.)	Comp.	R	n	pIC50(Exp.)
1	Diphenyl	5	7.42	14	6-Quinolinylnyl	5	7.44
2	Naphthalen-2-yl	5	7.68	15	2,2-Diphenylvinyl	5	6.70
3	4-(N,N-Dimethylamino)phenyl	5	6.96	16	[Anthracen-10(9H)-ylidene]methyl	5	6.47
4	Phenyl	5	6.62	17	(9H-Fluoren-9-ylidene)methyl	5	8.30
5	4-Phenylbuta-1,3-dienyl	5	7.54	18	4-Phenylbut-3-enyl	5	6.20
6	2-Phenyl-vinyl	5	7.49	19	2-Phenylvinyl	5	7.21
7	(2E,4E)-2-Methyl-4-phenylbuta-1,3-dienyl	5	7.46	20	2-(2,4-Dimethylbenzyl)vinyl	5	8.10
8	2-(4-Pyridinyl)vinyl	5	7.00	21	(1E,3E)-3-Methyl-4-phenylbut-1,3-dienyl	5	4.64
9	2-(2-Pyridinyl)vinyl	5	7.34	22	4-Methoxyphenyl	5	6.85
10	2-(3-Pyridinyl)vinyl	5	6.85	23	2-Naphthalenyl	4	7.06
11	2-(2-Naphthalenyl)vinyl	5	8.15	24	2-Fluoro-4-chlorophenyl	5	6.82
12	2-(7-Quinolinylnyl)vinyl	5	8.05	25	4-Nitrophenyl	5	6.74
13	2-(8-Quinolinylnyl)vinyl	5	7.30	26	Diphenyl	6	7.00



1.2 比较分子立场(CoMFA)分析

药效构象的确定是 3D-QSAR 研究中至关重要的一个环节. 选取对接结果中活性最高的 9 号分子赋予 Gasteiger-Hückel, MMMFF94, Gasteiger-Marsili 3 种电荷, 并采用 Tripos 力场进行优化 1000 次, 能量收敛标准为 0.21 kJ/mol. 以优化好的分子结构为活性构象, 并在此基础上变换其取代基构建其它化合物, 进行优化, 用于 CoMFA 叠合.

叠合方式的确定是 CoMFA 方法中的关键, 叠合规则直接关系到所建模型的好坏以及预测能力的高低. 确定活性构象后, 根据一定的规则, 如药效团或公共骨架等进行分子叠合, 使每个分子采取具有最大力场相似性的空间取向进行计算. 此处采用两种骨架(图 1)进行叠合: (1) 按照最大公共骨架

进行叠合; (2) 按照与 Zn 发生螯合的原子为骨架进行叠合, 叠合图如图 2 所示.

CoMFA 计算在 Sybyl7.3 中 QSAR 模块上完成. 将叠合的模型放置到三维的网格当中并按一定的步长将网格划分, 选择不同探针原子, 用于计算探针

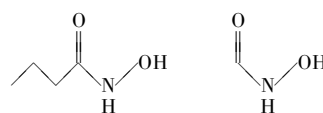


Fig. 1 Common structure used in align database (left) and the atoms chelated with Zn (right)

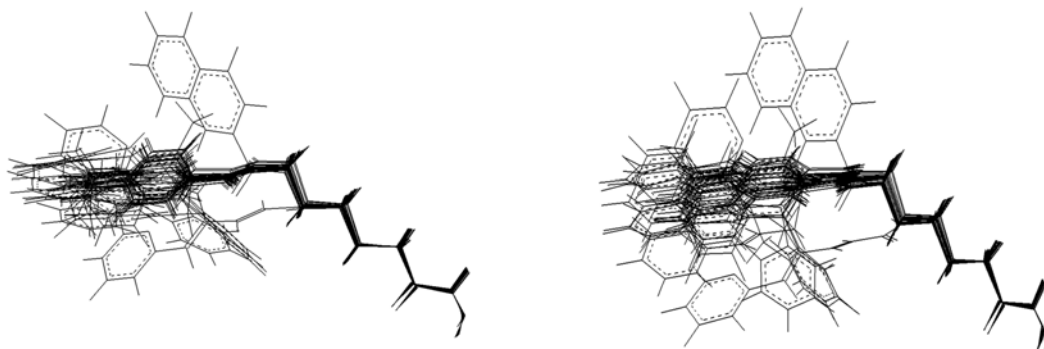


Fig. 2 Final alignment based on two different structures

与化合物的相互作用能. 首先做偏最小二乘 (PLS) 分析, 采用 Leave-one-out (LOO) 方法进行交叉验证, 然后在获得的最佳主分数下做非交叉验证进行回归计算, 建立相应的 CoMFA 模型, 用所建的 CoMFA 模型对预测集的 4 个化合物进行活性预测, 以评价模型的实际预测能力. 一般认为, 交叉验证相关系数 q^2 大于 0.5, 说明模型具有较好的预测能力.

2 结果与讨论

2.1 Surflex-dock 对接结果

由表 2 可以看到, 得出的 10 个最优构象中, 其中 8 个打分在 7.00 以上, 与原蛋白中的构象的 RMSD 为 0.081 ~ 0.104 nm (一般认为打分在 7.00 以上, 均方根偏差 RMSD 在 0.2 nm 以下的结果比较可信). 从图 3 中也可以看出, 对接后 TSN504 与蛋白中的构象和位置基本一致, 说明此方法能很好地还原蛋白中的配体构象.

Table 2 Result of docking TSN504 and protein

Conf.	Score	RMSD
1	9.36	0.84
2	9.24	0.83
3	9.20	0.81
4	9.19	0.82
5	9.74	1.04
6	7.32	1.77
7	7.15	0.97
8	7.10	0.97
9	6.88	0.95
10	6.51	2.03

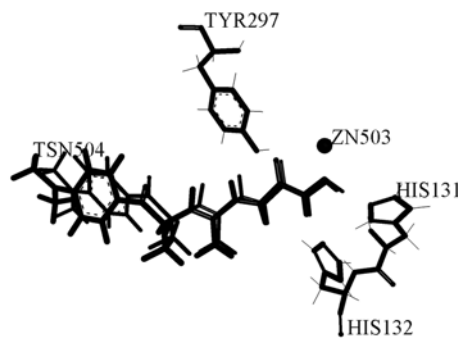


Fig. 3 TSN504's location and conformation before and after docking

图 4 为抑制剂与组蛋白酶的主要相互作用. 从图 4 中可以看出, 小分子抑制剂与蛋白酶结合, 起主要相互作用的是抑制剂周围一定范围内的氨基酸残基, 对接时选取小分子为活性位点中心进行计算. 以化合物 2 为例分析化合物与抑制剂周围 0.5 nm 氨基酸残基的对接作用模式, 小分子中—OH 上的氢和—NH—上的氢原子分别与残基 HIS131 和 HIS132 上的 N 原子发生了静电和氢键相互作用, 其键长分别为 0.2072 和 0.1929 nm, C=O 上的氧与残基 TYR297 上的氢也存在静电及氢键作用, 键长为 0.2093 nm; 且两个氧原子与锌离子有螯合作用, 距离为 0.1906, 0.2341 nm. 分子的疏水部分落入由残基 TYR92, TYR196, ASN20 等形成的疏水口袋中. 由于上述残基的作用, 使化合物 2 表现出对 HDAC 的抑制活性^[11]. 与文献[12,13]报道的羟氨酸类抑制剂的主要官能团作用完全一致.

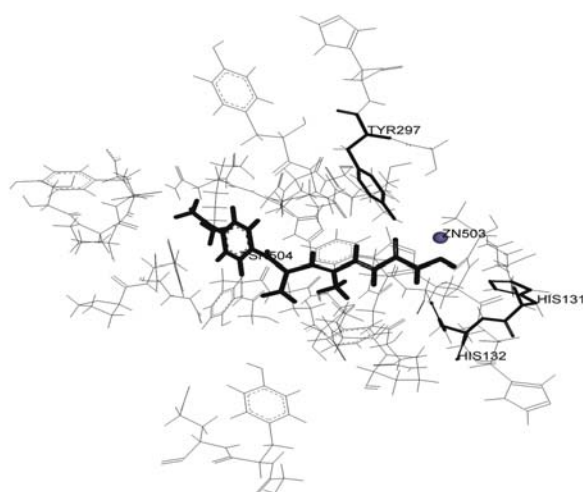


Fig. 4 Docking interaction pattern of 1C3R active residues with compound 2

16 号分子, —R 基为不在同一平面的 3 个环状结构, 体积较大, 引起了母体的偏移, 活性降低(半抑制浓度为 338 nm), 18 号分子, 虽然—R 基体积不大, 但相对较长, 同样会与蛋白发生碰撞(半抑制浓度为 625 nm); 4 号分子, 体积太小, —R 基只有一个苯环, 距离酶较远, 疏水作用较弱, 活性也较低(半抑制浓度为 238 nm). 所以在设计新型分子时, 空间位阻是一个不可忽视的因素. 与文献[14]报道的取代基不能太大也不能太小完全吻合.

2.2 CoMFA 模型分析

影响 CoMFA 模型好坏的因素有电荷、阈值、柱滤值、步长、网格大小和探针等, 不同的电荷反应不同的分子间相互作用, Gasteiger-Hückel 电荷同时考虑 σ 电子和 π 电子的作用, 而 Gasteiger-Masili 电荷只考虑 σ 电子. 不同的步长也会对结果造成一定的影响: 步长太大, 探针太疏, 则会漏掉一些重要信息, 步长太小, 探针太密, 则会引起噪音. 不同的探针用于指示原子之间的不同相互作用.

通过对影响模型的各种因素进行讨论得出的结果见表 3, 当按照最大公共骨架叠合时, 负载 Gasteiger-Hückel 电荷, 静电场和立体场阈值分别为 63 和 42 kJ/mol, 柱滤值为 16.8 kJ/mol, 采用 sp^3 杂化的 C 原子作为探针, 步长为 2.0 时所建模型最好. 最佳主成分数为 6, 交叉验证系数 q^2 为 0.668, 由最佳主成分数建立的 CoMFA 模型的非交叉验证回归系数 $R^2 = 0.984$, $F = 199.261$, 标准偏差 $SE = 0.111$. 在 CoMFA 模型中, 立体场贡献百分比为 65%, 静电场贡献百分比为 35%. 按照主要相互作用的原子叠合, 负载 Gasteiger-Hückel 电荷, 静电场和立体场阈值分别为 63 和 42 kJ/mol, 柱滤值为 16.8 kJ/mol, 采用 sp^2 杂化的 C 原子作为探针, 步长为 2.0 时所建模型最好. 最佳主成分数为 6, 交叉验证系数 q^2 为 0.654, 非交叉验证回归系数 $R^2 = 0.994$, $F = 436.406$, 标准偏差 $SE = 0.069$, 立体场贡献百分比为 65.4%, 静电场贡献百分比为 34.6%. 两个不同模型虽然所得交叉验证相关系数相近, 但以整合部分为骨架叠合所得模型. 立体场贡献为静电场的 1.88 倍, 说明立体场贡献明显高于静电场, 改变取代基体积是提高活性的主要途径.

用预测集的 4 个分子检验模型的预测能力, 相关散点图见图 5. 图 5 中每个分子都位于直线附近,

Table 3 Influences of different value of charge calculation method

Charge	N	q^2	R^2	SE	F
Gasteiger-Hückel	6	0.668	0.984	0.111	199.261
Gasteiger-Masili	6	0.386	0.989	0.094	233.590
MMFF94	6	0.577	0.997	0.049	850.972

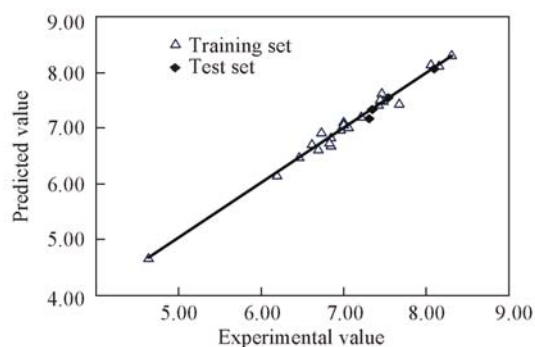


Fig. 5 Actual and predicted activities of 26 inhibitors in training and test set

且线性关系($r^2 = 0.9842$)接近 1, 表明模型比较可靠, 具有较好的预测能力.

2.4 CoMFA 等势图

图 6(A) 为 CoMFA 模型立体场等势图, 图中为 17 号分子. 其中绿色区和黄色区代表立体场对化合物活性的影响, 绿色区表示引入体积大的取代基团, 可以提高化合物活性, 黄色区表示引入体积大的基团可能降低化合物的活性. 从化合物 17 上看, 在取代基左端有一大块绿色区, 表明如果有相对较大的基团在该区取代会使化合物活性升高, 环状的上下方都有黄色区, 在此区域加入体积小的基团可能会增大化合物的活性. 5 号、7 号分子的表面识别区左端为苯环, 中间为细长的芳香链, 能够深入到蛋白空隙中且没有空间位阻, 活性较高. 且 7, 2, 12, 17 随着肽键左端所加基团体积增大, 活性明显升高. 但是, 16 号分子的左端为不共平面的 3 个环, 虽然体积较大, 但环与肽键的距离太短未伸到口袋底部, 在黄色区域与蛋白发生碰撞, 导致活性降低, pIC_{50} 为 6.47.

图 6(B) 所示为 CoMFA 模型静电场等势图. 静电场贡献明显比立体场小, 其中蓝色区和红色区代表静电场对化合物活性的影响, 蓝色区表示引入电正性大的取代基团, 可以提高化合物活性, 红色区表示引入电负性大的基团可能会提高化合物的活性. 如: 24 号分子肽键左端苯环 4 号位为电负性大的 F 取代基, 使其活性大大降低 ($pIC_{50} = 6.82$), -F 刚好落在蓝色区域; 3 号分子肽键左端苯环 4 号位为 -N,N-2 甲基, 由于 N 的电负性较大, 明显降低了化合物的活性. 17, 19 号分子左端为环状基团, 电正性较大, 活性较高.

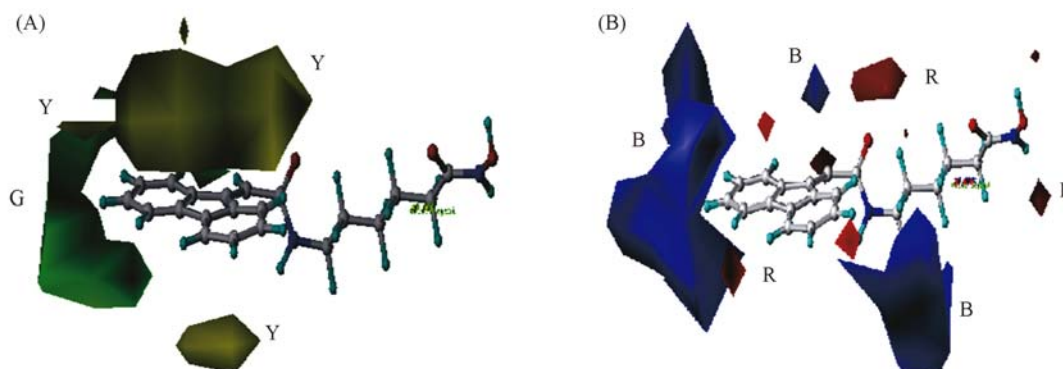


Fig. 6 Contour plot of the CoMFA steric fields(A) and electrostatic field map(B)

综上所述可以得出, 在羟胺酸类 HDACi 中, 肽键左端基团为脂肪链并连接体积较大的基团会提高化合物的活性. 脂肪链通常包括 3~5 个 C, 如果太短, 左端的表面识别区就会与蛋白产生空间位阻, 若太长, 环状到达空腔底部, 同样会与蛋白发生碰撞. 环状结构一般为体积较大的苯环, 萘或多个环的结构, 能够充分与蛋白相互作用, 活性增强. 此结论与 Surflex-dock 对接结果基团体积要适中相符合. 根据以上结论猜测, 取代基如图 7 结构的化合物可能有活性.

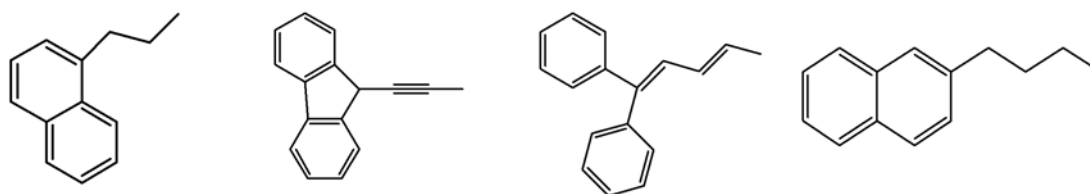


Fig. 7 Groups which maybe have activity

3 结 论

本文采用 Surflex-dock 和 CoMFA 方法对羟胺酸类 HDACi 进行了分析, 实验结果表明, Surflex-dock 方法能很好地还原蛋白质结构, 能够用于虚拟筛选和新型分子的设计, CoMFA 所建模型预测能力较好, 能很好地指导羟胺酸类化合物的结构修饰, 并能预测新化合物的活性, 两种方法相结合为新化合物的合成提供了理论依据.

参 考 文 献

- [1] XIE Ai-Hua(谢爱华), LIAO Chen-Zhong(廖晨钟), LI Bo-Yu(李伯玉), *et al.*. Chinese J. New Drugs(中国新药杂志)[J], 2005, **14**(1): 10—14
- [2] ZOU Chen(邹琛), ZHOU Jun(周俊), LU Guo-Ping(陆国平). Chinese Journal of Pathophysiology(中国病理生理杂志)[J], 2007, **23**(12): 2487—2490
- [3] LIU Ai-Lin(刘艾林), LONG Lun(龙娟), WANG Nan(王楠), *et al.*. Acta Pharmaceutica Sinica(药学学报)[J], 2005, **40**(7): 585—590
- [4] Ragno R., Simeoni S., Valente S., *et al.*. 2006, **46**: 1420—1430
- [5] Wade P. A.. Hum. Mol. Genet. [J], 2001, **10**: 693—698
- [6] Cress W. D., Seto E.. J. Cell Physiol. [J], 2000, **184**: 1—16
- [7] Johnston R. W.. Nature Reviews Drug Discovery[J], 2002, **1**: 287—299
- [8] De Ruijter AJM, Albert H. van Gennip AH, Caron HN, *et al.*. Biochemical journal[J], 2003, **370**: 737—749
- [9] Juvalle D. C., Kulkarni V. V., Deokar H S, *et al.*. Organic & Biomolecular Chemistry[J], 2006, **4**: 2858—2868
- [10] Finnin M. S., Donigian J. R., Cohen A., *et al.* Nature[J], 1999, **401**: 188—193
- [11] LUAN Feng(栾锋), YAO Xia-jun(姚小军), ZHANG Hai-xia(张海霞), *et al.*. Journal of Lanzhou University(Natural Sciences)[J], 2006, **42**: 71—78
- [12] WANG Bao-Lei(王宝雷), WANG Jian-Guo(王建国), MA Yi(马翼), *et al.*. Acta Chimica Sinica[J], 2006, **64**(13): 1373—1378
- [13] Andrianov V, Gailite V, Lola D, *et al.*. European Journal of Medicinal Chemistry[J], 2008, **44**: 1—19
- [14] LI Shun-Lai(李顺来), ZHENG Yan(郑艳), WANG Feng-Ying(汪凤颖). Journal of Beijing University of Technology(北京工业大学学报)[J], 2006, **44**(4): 75—79

Studies on Hydroxamic Acid Histone Deacetylase Inhibitors(HDACI) by Molecular Docking and CoMFA

ZHANG Liang, XIANG Yu-Hong, Zhang Zhuo-Yong*

(Department of Chemistry, Capital Normal University, Beijing 100048, China)

Abstract Surflex-dock is used in docking histone deacetylase inhibitor and histone molecules. The interaction between ligand and receptor can revealed the binding mode of the inhibitors of HDAC. The result showed that the of docking has good correlation with the inhibitor activity. A further study on the three-dimensional quantitative structure-activity relationship(3D-QSAR) was carried out using the method of comparative molecular field analysis(CoMFA). The influences of different steps, probe atomic charge on model building were investigated. A 3D-QSAR model has been established, in which cross-validation coefficient q^2 is 0.668, and non-cross-validation correlation coefficient r^2 is 0.984. The established model has a good predictive ability, and it can be used to guide the drug design and synthesis.

Keywords HDAC; Surflex-dock; CoMFA; Inhibitor