

研究论文

3-氨基磺酰苯甲酸类 AKR1C3 抑制剂的 3D-QSAR 和分子对接研究

吴建军¹ 马玉卓^{1*} 戴雪娥² 刘鹰翔²¹广东药学院药科学院 广州 510006; ²广州中医药大学中药学院 广州 510006

摘要 醛酮还原酶 1C3 (AKR1C3)作为治疗前列腺癌的新靶点已成为研究热点, 3-氨基磺酰苯甲酸衍生物对其具有高效的选择性和抑制活性。本文采用比较分子场分析(COMFA)和比较分子相似性指数分析(COMSIA)方法, 将经分子对接后的 34 个优势构象组成训练集和 11 个优势构象组成测试集, 构建三维定量构效关系(3D-QSAR)模型。COMFA 模型的交叉验证系数(q^2), 非交叉验证系数(R^2), 标准偏差(SEE)和 F 值分别为 0.761, 0.973, 0.122, 185.963; 自举法回归系数为 $R^2_{bs}=0.98$ 。最佳组合 COMSIA 模型的 q^2 , R^2 , SEE, F 和 R^2_{bs} 分别为 0.734, 0.984, 0.097, 147.850, 0.994。COMFA 和 COMSIA 模型的系统外部测试 R^2_{pred} 分别为 0.864 和 0.756, r^2_m 分别为 0.8127 和 0.5377。这些结果表明, 所建立的 QSAR 模型具有较高的可靠性和较强预测能力。经三维等势图分析可知, 在 2、5 或 6 位适当增加取代基体积, 或在 5 位引入氢键受体, 或在 7 位引入负电性取代基则能提高化合物的生物活性。该模型为进一步设计具有更优选择性和活性的化合物提供了理论依据。

关键词 3-氨基磺酰苯甲酸衍生物 AKR1C3 抑制剂 分子对接 COMFA COMSIA

3D-QSAR and Molecular Docking Study of 3-Sulfamoylbenzoic Acid Derivatives as AKR1C3 Inhibitors

Wu Jianjun¹, Ma Yuzhuo^{1*}, Dai Xue'e², Liu Yingxiang²¹ College of Pharmacy, Guangdong Pharmaceutical University, Guangzhou 510006; ² School of Chinese Materia Medica, Guangzhou University of Chinese Medicine, Guangzhou 510006

Abstract The human aldo-keto reductase family 1 member C3 (AKR1C3) has become a significant emerging target of therapeutic interest in prostate cancers. In the present work, 3D-QSAR model was established by a training set of 34 compounds and validated by the evaluation of a test set of 11 compounds, using the comparative molecular field analysis (COMFA) and comparative molecular similarity indices analysis (COMSIA) methods. The q^2 , R^2 , SEE, F and the R^2_{bs} were 0.761, 0.973, 0.122, 185.963 and 0.98 respectively in COMFA model, and were 0.734, 0.984, 0.097, 147.850, and 0.994 respectively in optimum COMSIA model. The established 3D-QSAR model shows strong stability and good predictive ability. The systemic external validation parameters for COMFA and COMSIA models were R^2_{pred} (0.864 and 0.756) and r^2_m (0.8137 and 0.5377), which validate the quality and predictive ability of 3D-QSAR model. Based on the 3D contour maps, if the volume of the substituent are increased appropriately at the 2, 5 or 6 positions of the template molecule, or hydrogen bond acceptor substituent was added to the 5 position of the template molecule, or the electronegative substituent was added to the 7 position of the template molecule, the biological activity of compounds will be improved. The generated models can provide useful information for designing new compounds with higher selectivity and stronger activity.

Keywords 3-sulfamoylbenzoic acid derivatives, AKR1C3 inhibitors, Molecular docking, COMFA,

吴建军 男, 26 岁, 硕士生, 从事计算机辅助药物设计研究。*联系人, 马玉卓 女, 副教授, 从事基于靶酶的小分子抑制剂的设计和构效关系研究。E-mail: mayuzhuo66@163.com

广东省科学计划项目(2011B031800232)资助

2015-03-18 收稿, 2015-05-26 接受

COMSIA

前列腺癌(Prostate Cancer, PC)是一种雄性激素依赖性疾病^[1]。在男性恶性肿瘤中,其发病率位居首位(28%),致死率为 10%,仅次于肺癌^[2]。雄激素和雄激素受体通路异常促使前列腺癌的发生和发展。因此,抑制雄激素的生物合成和阻断雄激素受体信号通路是前列腺癌的传统治疗方法,即雄激素剥夺疗法(Androgen deprivation therapy, ADT)^[3,4]。最初,ADT 对晚期或转移性前列腺癌有效,但经过一段时间治疗后,大部分患者发展为致命性的去势前列腺癌(castrate resistant prostate cancer, CRPC)^[5]。目前用于治疗 CRPC 的药物有阿比特龙和恩杂鲁胺^[6-9],然而阿比特龙会产生高血压的不良反应,需与强的松联合用药,并且两者在治疗过程中均会观察到药物抵制作用。因此,对作用于新靶点治疗 CRPC 药物的研发具有重大意义。

研究发现,在 CRPC 中,AKR1C3 基因表达能力增加了 5.3 倍,AKR1C3 迅速上调,并参与雄激素的合成^[10, 11]以应对雄激素水平的下降。AKR1C3 也具有前列腺素 PGF 合成酶活性,能促进肿瘤细胞的生长^[12];与 AKR1C 的其他亚型相比,其浓度水平和催化能力均较高^[13]。Tian 等^[14]研究表明,AKR1C3 的过度表达与前列腺癌的发生、发展密切相关,是近几年研究治疗前列腺癌的新颖作用靶点^[15]。近几年研究 AKR1C3 抑制剂的报道较多,且化合物的结构多样^[16-19]。Gobec 研究小组^[20]采用基于片段的药物虚拟筛选方法发现结构新颖的 AKR1C3 抑制剂。Denny 研究小组^[21]合成了一系列 3-(3,4-二氢异喹啉-2(1H)-磺胺)苯甲酸类化合物,并进行结构修饰,发现了具有强活性和高选择性的 AKR1C3 抑制剂。

本文在此基础上,采用分子对接、比较分子场分析(COMFA)和比较分子相似性指数分析(COMSIA)方法对该系列化合物进行三维定量构效关系(3D-QSAR)研究,为进一步设计更高活性和选择性的新化合物提供理论依据,避免盲目性。

1 材料与方法

1.1 研究的化合物及其活性数据

综合考虑化合物结构多样性和活性范围合理分布,将 3-氨磺酰基苯甲酸类化合物分为训练集(34 个)和测试集(11 个)。所有化合物均由 ChemDraw 2010 构建,采用 Tripos 公司 sybyl-X 2.0 中 Minimize 分子力学程序进行能量最小化以期得到低能稳定构象。在此过程中,采用 Tripos 力场,对所有化合物加载 Gasteiger-Huckel 电荷;采用 Powell 能量梯度法,最大优化次数为 1000 次,梯度下降系数设置为 0.001,能量变化收敛条件限定为 0.05kJ/(mol×Å),其他均为 sybyl-X 2.0 缺省值^[22]。所有化合物的活性数据均用转化为 pIC50(-logIC50)表示并列于表 2 中。

1.2 分子对接和活性构象

本文采用 Molecular Operating Environment 2008 (MOE 2008)的 docking 模块完成分子对接,将该系列化合物分别对接到 AKR1C3 受体结合位点(PDB:4FAM)(www.rcsb.org)中以期得到较好的活性构象。对于不同种类的受体,最佳的 placement 和 scoring 函数也是有差异的。因此,本文采用 Self-docking analysis 方法^[23],以控制变量法来选择针对 AKR1C3 受体的最佳对接函数。首先选择 3R94、3UG8、4FA3、4FAM 和 4HMN 五个配体-受体共晶结构作为 Self-docking 的受体和配体来源,该五个共晶结构均是不同配体分子与 AKR1C3 受体蛋白形成的复合物,如表 1。控制变量以 Alpha Triangle, Alpha PMI, Proxy Triangle, Triangle Matcher 作为 placement 函数和 ASE, Affinity dG, Alpha HB, London dG 作为 scoring 函数。分别将五个配体从相应的复合物中提取出来,再在不同的 placement 和 scoring 函数条件下对接到对应的受体中。比较对接后最佳构象与受体的相互作用模型和共晶结构中配体-受体作用模型。以均方根偏差

(RMSD)作为对接可靠性的衡量标准,用平均 RMSD 值和 RMSD 小于 2.0Å 的构象个数作为选择最佳对接函数的评价标准。选择最佳对接函数,并将该系列化合物进行对接,挑选出最佳构象。

表 1 五个 AKR1C3 共晶结构(PDB 编号: 3R94、3UG8、4FA3、4FAM 和 4HMN)的基本信息

Tab. 1 Information of five AKR1C3 crystal structures (PDB code: 3R94, 3UG8, 4FA3, 4FAM and 4HMN)

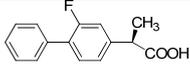
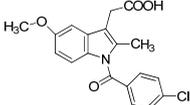
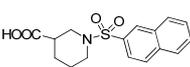
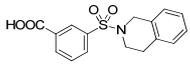
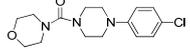
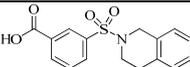
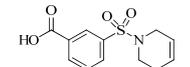
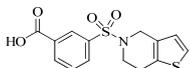
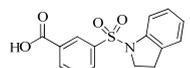
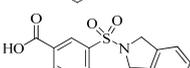
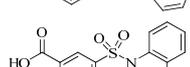
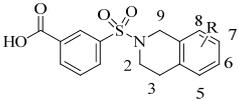
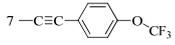
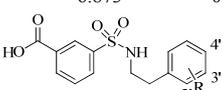
PDB code	配体分子	分辨率/Å	氨基酸数目	关键氨基酸
3R94		2.01	331	Tyr55,His117,Phe311, Met20,Phe306
3UG8		1.73	331	Tyr55,Trp227, Phe306,Met120
4FA3		2.0	331	Tyr55, Phe306, Trp227, Phe311
4FAM		2.0	662	Tyr55,His117,Phe306, Trp227,Phe311,Leu54
4HMN		2.4	331	Tyr55,Phe311,Trp227, Tyr24,Phe306,Phe311

表 2 化合物的分子结构及其生物活性

Tab. 2 The molecular structures of the compounds and their biological activities

NO.	化合物	Expt.pIC50	COMFA (pIC50)		COMSIA(pIC50)		Docking E/(cal/mol)
			Pred.	Res.	Pred.	Res.	
1		7.886	7.848	0.038	7.854	0.032	-24.8265
2		6.409	6.294	0.115	6.384	0.025	-20.6961
3		7.602	7.707	0.105	7.613	-0.011	-24.3181
4*		6.648	6.366	0.312	6.950	-0.272	-24.7875
5		7.328	7.204	0.124	7.351	-0.023	-24.1972
6*		6.699	6.892	-0.193	7.155	-0.457	-24.9750
							
	R						
7	9-Me	7.569	7.621	-0.052	7.538	0.031	-25.6099
8	2-Me	8.066	7.937	0.129	8.028	0.038	-26.1135
9	5-NH2	7.276	7.289	-0.013	7.316	0.040	-24.8839
10*	5-NO2	8.051	8.065	-0.014	8.029	0.022	-24.8826
11	5-Cl	7.959	7.630	-0.051	7.701	-0.258	-24.3858
12*	5-Br	7.959	7.761	0.198	7.634	0.325	-24.3858
13	5-I	7.854	7.773	0.081	7.728	0.126	-26.6662
14	5-OH	7.796	7.774	0.022	7.809	-0.013	-25.9062
15*	5-OMe	7.796	8.002	-0.206	8.159	-0.330	-26.7227
16	6-Me	7.770	7.697	0.073	7.769	0.001	-26.5604
17	6-NO2	7.658	7.730	-0.072	7.657	0.001	-25.0395
18	6-CN	7.538	7.480	0.058	7.537	0.001	-26.4371
19	6-Cl	8.060	7.857	0.203	7.831	0.229	-24.5537
20*	6-Br	8.215	8.152	0.063	7.882	0.333	-24.8797
21	6-I	7.409	7.671	-0.262	7.682	-0.273	-24.8486
22	6-OH	7.569	7.733	-0.064	7.521	-0.048	-25.9274

23*	6-OMe	7.420	7.393	0.027	7.442	-0.022	-26.4266
24	7-Me	7.886	7.750	0.136	7.849	0.037	-26.4768
25	7-C≡CH	7.377	7.506	-0.120	7.444	-0.067	-26.1963
26		5.883	5.951	-0.129	5.885	-0.002	-25.1097
27		7.167	7.114	0.053	7.213	-0.046	-26.6913
28*	7-NO2	7.770	7.225	0.545	7.213	0.384	-24.7625
29	7-F	7.678	7.727	-0.049	7.646	0.032	-24.3355
30	7-Cl	7.699	7.701	-0.002	7.712	-0.013	-24.1788
31*	7-Br	7.921	7.734	0.187	7.828	0.093	-24.3355
32	7-I	7.854	7.729	0.125	7.837	0.017	-23.5615
33	7-CN	7.469	7.485	-0.016	7.499	-0.030	-23.5687
34	7-OMe	7.538	7.673	-0.135	7.552	-0.014	-26.1341
35*	8-Cl	7.721	7.559	0.162	7.986	-0.265	-24.1472
36	6,7-diOMe	6.796	6.873	-0.077	6.805	-0.188	-27.0146
							
37	H	6.658	6.725	-0.067	6.846	-0.188	-25.7744
38	2'-F	6.481	6.629	-0.148	6.479	0.002	-24.7400
39	3'-Cl	6.678	6.631	0.047	6.621	0.057	-23.8609
40	4'-Cl	6.854	6.618	0.236	6.851	0.003	-25.1621
41	3-OMe	6.131	6.202	-0.071	6.089	0.042	-22.3299
42	4-OMe	5.842	5.814	0.028	5.829	0.013	-25.8915
43	3-OPh	6.658	6.513	0.145	6.565	0.093	-24.7360
44*	3-Me	6.276	6.319	-0.043	6.245	0.031	-28.5283
45	4-Me	6.194	6.309	-0.115	6.193	0.001	-26.0215

1.3 活性化合物分子叠加

分子叠加在创建一个可靠而有效的 3D QSAR 模型过程中至关重要。在 sybyl-X 2.0 的 Align database 模块中, 将选出的训练集和测试集以化合物 1 为模板分子, 以 1~10 号原子为公共骨架进行叠合。如图 1。

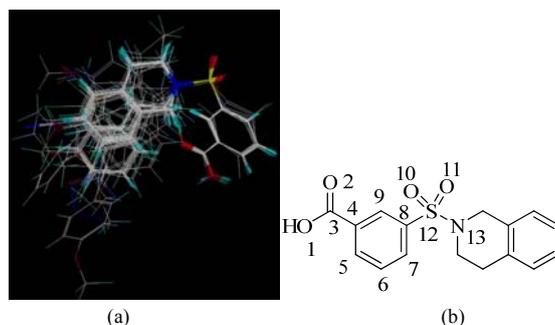


图 1 (a) 45 个化合物分子叠合图; (b) 模板分子 1 和 1~10 号骨架原子

Fig. 1 (a) The molecular superimposition of the 45 compounds; (b) The template molecule 1 and the common atoms from 1 to 10

表 3 自我对接结果(不同函数条件下的 RMSD)

Tab. 3 Self-Docking Validation(the RMSD in different functions)

Rescoring function 1: ASE; Rescoring function2: ASE;

Refinement: Forcefield; Retain: 10

PDB	Alpha triangle	Alpha PMI	Proxy triangle	Triangle Matcher
3R94	0.3085	6.5110	0.3922	0.3920
3UG8	0.4425	0.4189	1.0022	0.4522
4FA3	1.7895	1.8227	0.5911	0.5909
4FAM	0.3048	0.2633	0.2655	0.2656
4HMN	0.8940	5.0306	1.2198	1.4130
aRMSD ^a	0.9349	3.5117	0.8677	0.7784
Inf.2Å ^b	12	6	12	13

Rescoring function 1: Affinity dG; Rescoring function2:ASE;
Refinement: Forcefield; Retain: 10

PDB	Alpha triangle	Alpha PMI	Proxy triangle	Triangle Matcher
3R94	6.5951	2.2560	5.4099	5.4188
3UG8	3.1073	0.4244	0.4484	0.4525
4FA3	0.5217	0.5185	0.5668	0.5210
4FAM	0.2765	0.8709	0.8688	0.8697
4HMN	0.8917	5.1325	0.9247	0.9249
aRMSD	2.8481	2.301	2.055	2.046
Inf.2Å	10	8	11	11

Rescoring function 1: Alpha HB; Rescoring function2:ASE;
Refinement: Forcefield; Retain: 10

PDB	Alpha triangle	Alpha PMI	Proxy triangle	Triangle Matcher
3R94	0.3821	1.6188	1.1690	4.5294
3UG8	0.4535	0.4192	3.6054	4.2488
4FA3	0.8173	1.2739	0.5279	0.5280
4FAM	0.8812	0.8733	0.8738	0.8728
4HMN	0.8963	5.1300	0.9363	0.9356
aRMSD	0.8576	2.3288	1.7781	2.779
Inf.2Å	12	11	12	11

Rescoring function 1: London dG; Rescoring function2:ASE;
Refinement: Forcefield; Retain: 10

PDB	Alpha triangle	Alpha PMI	Proxy triangle	Triangle Matcher
3R94	0.3920	2.2569	2.2573	0.3936
3UG8	0.4497	0.4246	0.8653	0.8654
4FA3	1.6407	0.5368	0.5743	0.5742
4FAM	0.8725	4.9412	0.8760	0.2894
4HMN	0.8873	5.1298	0.8978	0.8982
aRMSD	1.061	3.3223	1.3677	0.7552
Inf.2Å	13	7	7	10

a 表示 RMSD 平均值, b 表示 RMSD ≤ 2.0Å 的个数。

1.4 COMFA 和 COMSIA 模型的构建

在 sybyl-X 2.0 中, 采用 Tripos 标准力场, 以 Lennard-Jones 和 Coulomb(常数介电函数)为计算函数, 将叠合好的训练集分子放入一个步长为 0.2 nm 的三维晶格中, 以 sp^3 杂化的 C^+ 作为立体探针(范德瓦尔斯半径 1.52 Å), +1 价的氢原子作为静电探针, 分别探测每个网格点的立体场和静电场的大小和分布; 立体场和静电场的贡献截止值(cut-off)设置为 30.0 kcal/mol。在 COMSIA 模型中, 除了计算立体场和静电场外, 还引入了疏水场、氢键受体场和氢键供体场, 更能充分说明不同场对分子活性的影响。COMSIA 为了避免分子表面附近格点处势能的急剧变化及异常值的出现, 采用了依赖于距离的高斯函数计算分子场能量。分子相似性指数计算函数^[24]如下:

$$A_{F,K}^q(j) = \sum_{i=1}^n w_{probe,k} w_{ik} e^{-\alpha r_{iq}^2} \quad (1)$$

在式(1)中, A 表示分子 j 中所有原子在格点 q 处的相似性指数; $w_{probe,k}$ 是半径为 1Å、电荷为+1、疏水性为+1、氢键受体为+1、氢键供体为+1 的探针原子; w_{ik} 是原子 i 的理化性质 k 的实际值; r_{iq} 表示在格点 q 处探针原子与测试分子中原子 i 之间的距离; α 表示衰减因子, 系统默认为 0.3。

采用偏最小二乘法(PLS)构建系列化合物结构与活性之间的定量构效关系。以 COMFA 和 COMSIA 值作为自变量, pIC_{50} 作为因变量, 进行模型构建。采用“留一法”(LOO)进行交叉验证(cross-validation)分析获得 q^2 和最佳主成分数(N)。再根据 N , 系统默认柱滤值为 0 kcal/mol, 进行非交叉验证(non-cross-validation)分析得到 R^2 、预测差平方和(PRESS)、标准偏差(SEE)和 F 值。此外还进行了圈数为 100 的自举法(bootstrapping)进行模型分析, 得到 R_{bs}^2 和 SD_{bs} , 进一步验证模型的统计质量和稳定性。引

入系统外部预测参数 R_{pred}^2 和 r_m^2 进一步验证模型的可靠性和预测能力^[25,26], 如:

$$R_{\text{pred}}^2 = 1 - \frac{\sum (Y_{\text{pred}(\text{test})} - Y_{\text{test}})^2}{\sum (Y_{\text{test}} - Y_{\text{mean}(\text{train})})^2} \quad (2)$$

在式(2)中, $Y_{\text{pred}(\text{test})}$ 代表测试集化合物的预测活性(pIC₅₀), Y_{test} 代表测试集化合物的真实活性(pIC₅₀), $Y_{\text{mean}(\text{train})}$ 代表训练集化合物活性的平均值。

$$r_m^2 = r^2 (1 - \sqrt{r^2 - r_0^2}) \quad (3)$$

在式(3)中, r 和 r_0 分别代表具有截距和不具有截距的测试集化合物的实验值和预测值的相关系数。

一个具有可靠预测能力的 3D-QSAR 模型需满足的一般条件有^[27]: $q^2 > 0.5$, $R^2 > 0.6$, $r_m^2 > 0.5$, $R_{\text{pred}}^2 > 0.5$, $((r^2 - r_0^2)/r^2) < 0.1$ 。

2 结果与分析

2.1 分子对接与活性构象

本文采用 Self-docking analysis 分析方法讨论 placement-scoring 函数对对接结果的影响。利用控制变量法确保 Refinement 为 Forcefield, Retain 为 10, Rescoring function2 为 ASE 因素不变, 再分别讨论不同 Rescoring function 1 条件下不同 Placement 函数对接效果的影响, 以便选择最佳对接函数来提高对接准确性。经结果分析, 如表 3, 最终选择 London dG-Triangle Matcher 为最佳的对接函数。对接参数设置为 Placement: Triangle Matcher, Rescoring 1: London dG, Rescoring 2: ASE, Retain: 10, Refinement: Forcefield。AKR1C3 受体蛋白 4FAM 在该设置参数下进行配体-受体自我对接, RMSD 值为 0.2612Å, 打分为-24.0782 kcal/mol, 如图 2。RMSD 值较小且对接后配体构象与原晶体中配体构象重合得非常好, 因此, 对接函数可靠性较强。将本研究的化合物均对接到受体的活性口袋, 并选择与配体构象最接近的构象。

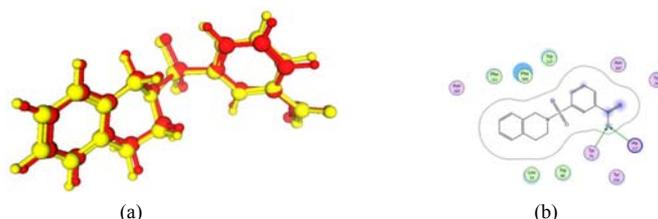


图 2 (a)对接后的优势构象与共晶结构中配体的叠合, 红的代表共晶结构中的配体, 黄色代表对接后的优势构象, 其均方根偏差(RMSD)为 0.2612Å; (b)优势构象与 AKR1C3 受体蛋白相互作用的二维结合模型

Fig.2 (a) The superimposition of the best docked conformation and the co-crystallized ligand in AKR1C3 complex (PDB: 4FAM). Red and yellow represent native co-crystallized and docked ligand respectively, the RMSD is 0.2612 Å; (b) Two-dimensional representation of interaction of the docked conformation and AKR1C3 receptor (4FAM)

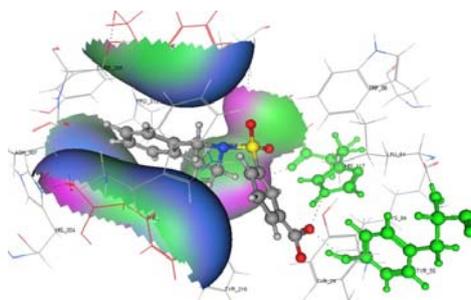


图 3 对接后的优势构象与 4FAM 结合口袋的相互作用模型。高斯面表示 SP1 口袋, 红色残基表示组成 SP1 口袋氨基酸, 绿色残基是重要氨基酸 Tyr-55 和 His-117

Fig.3 The interaction of the docked conformation and the binding pocket. The Gaussian surface represents the SP1 pocket; the red residues form the SP1 pocket; the green Tyr-55 and His-117 residues are the key residues

配体-受体相互作用模式可以从理论上解释药物作用机理, AKR1C3 蛋白具有 $(\alpha/\beta)_8$ ‘桶状’三维折叠结构, β 折叠与 α 螺旋式反向平行,具有对底物有选择性的3个Loops区。该蛋白在催化过程中,NADP⁺起着催化辅助因子的重要作用,且和氨基酸残基 His-117、Tyr-55 组成了氧离子催化域,对配体的锚定具有重要作用。从模板分子1与受体对接的相互作用模型(如图3)可知,羧基上的氧原子与氨基酸 His-117 和 Tyr-55 以氢键方式结合,因此,对于本系列化合物而言,3-苯甲酸部位中的羧基是关键作用基团。3,4-二氢异喹啉部位占据 SP1 口袋^[28],苯甲酸中苯环被 SP3 口袋包围,磺胺起着重要的桥梁作用。SP1 口袋比 SP3 口袋小,因此二氢异喹啉环上取代基的体积不宜太大。

2.2 COMFA 和 COMSIA 模型的分析

由训练集构建的 COMFA 模型,其 $q^2=0.761$ 、 $R^2=0.973$ 、 $N=6$ 、 $F=185.963$ 、 $SEE=0.122$ 、 $R^2_{bs}=0.980$ 、 $SD_{bs}=0.007$,如表4。这一结果表明,该模型的统计质量和稳定性较高,具有较好的预测能力;其立体场和静电场对模型的贡献分别是0.751和0.249,可见立体场效应比静电场效应对化合物活性的影响大。对立体场(S)、静电场(E)、疏水场(H)、氢键供体场(D)和氢键受体场(A)进行不同组合并构建 COMSIA 模型,分别对每个模型进行 PLS 分析,所得到的9个模型都具有很好的相关性和预测能力,其中 SEA 组合的模型最佳: $q^2=0.734$ 、 $R^2=0.984$ 、 $SEE=0.097$ 、 $N=10$ 、 $F=147.850$ 、 $R^2_{bs}=0.994$,其立体场、静电场和氢键受体场的贡献分别为0.444、0.491和0.066。对 COMFA 和 COMSIA(SEA)模型分别进行外部预测,得到 R^2_{pred} 和 r^2_m 值,如表5。从实验值与预测值的线性回归图(图4)可知,COMFA 和 COMSIA(SEA)模型具有较好的外部预测能力。相比于 COMFA 模型,COMSIA 模型采用了平滑的高斯函数,避免了势能的急剧变化和 cut-off 值的设定;COMSIA(SEA)模型还引入了氢键受体场来表征化合物结构与活性之间的关系。因此这两个模型用于揭示3-氨磺酰基苯甲酸衍生物的构效关系是必要的。

表4 COMFA 和 COMSIA 模型结果
Tab. 4 The results of COMFA and COMSIA models

Models	q^2	N	SEE	R^2	F	R^2_{bs}
COMFA(SE)	0.761	6	0.122	0.973	185.963	0.980
COMSIA(SED)	0.624	5	0.104	0.982	129.732	0.991
COMSIA(SEA)	0.734	10	0.097	0.984	147.850	0.994
COMSIA(SHD)	0.657	10	0.109	0.980	117.162	0.994
COMSIA(EHD)	0.528	8	0.111	0.919	126.024	0.994
COMSIA(SEHA)	0.640	6	0.134	0.965	127.220	0.998
COMSIA(SEHD)	0.575	10	0.079	0.989	223.998	0.998
COMSIA(SHDA)	0.640	6	0.175	0.940	73.077	0.994
COMSIA(SED)	0.604	9	0.109	0.979	131.345	0.987
COMSIA(SEHDA)	0.567	10	0.079	0.990	227.965	0.997

表5 COMFA 和 COMSIA 模型的外部验证
Tab. 5 The external validation results of COMFA and COMSIA models

模型	R^2_{pred}	$(r^2-r_0^2)/r^2$	r^2_m	r^2_m (LOO)	r^2_m (overall)
COMFA	0.864	0.0072	0.8137	0.9627	0.9417
COMSIA(SEA)	0.756	0.0994	0.5377	0.9268	0.9268

2.3 三维等势图的分析

COMFA 和 COMSIA 三维等势图采用 Stedev coeff 方法来显示,以模板分子1表示不同场对活性的贡献。在立体场中,绿色(G)表示在该区域引入大基团有利于提高化合物的活性,黄色(Y)区域则表示在该区域引入大基团对化合物的活性不利^[29,30]。从图5(a)中可知,G主要分布在5和6位,但在6外缘覆盖有Y区域,这表明在一定体积范围内5和6位引入较大基团有利于提高化合物的活性,但当体积超过

一定范围后则会使活性降低。如化合物 11(7.959)、12(7.959) 大于 1(7.886)，又如化合物 20(8.215)>19(8.060)>16(7.770)>21(7.409)。从对接模型图 3 可知,3,4-二氢异喹啉-2-1(H)部位占据 AKR1C3 受体蛋白(由 Ser-118、Asn-167、Phe-306、Phe-311、Tyr-319 组成)的 SP1 区,当 5 和 6 位取代基体积过大时,容易与口袋中残基发生碰撞,从而影响配体和受体之间的结合。7 和 8 位区域被 Y 覆盖,表明增大取代基的体积使化合物的生物活性降低。如 29、30、32 取代基的体积均大于 H 取代,而其活性小于化合物 1 的活性。在 2 位中间有一片 G 覆盖,这表明在 2 位取代基增大体积有利于提高生物活性,如化合物 8 在 2 位的取代基是 Me,其体积比 H 大,化合物 8(8.066)>1(7.886)。从对接图形可知,在 2 位结合处的受体蛋白空间很大,因此增大体积有利于配体与受体的相互作用。图 5(b)为 COMFA 的静电等势图,在静电场中,红色(R)表示在该区域引负电性基团有利于提高化合物的活性,蓝色(B)则表示引入正电性基团则有利于化合物生物活性的提高。由该图可知,7 位区域被 R 所占据,表明 7 位引入负电性基团能使化合物的活性得到提高。如负电性基团取代的化合物 28(7.770),29(7.678),30(7.699),31(7.921),32(7.854),33(7.469),34(7.538),35(7.721)活性均大于乙炔基取代的 25(7.377)。B 区域在氨磺酰基氮原子附近出现,R 区域出现在磺酰基附近,表明磺胺基团上的电负性 O 原子和电正性-NH⁺对化合物的活性起着至关重要的作用。苯甲酸上的羧基也被 R 区域覆盖,表现出 O 的电负性,其与重要氨基酸 His-117 和 Tyr-55 形成氢键,如图 3,是活性化合物的关键基团。

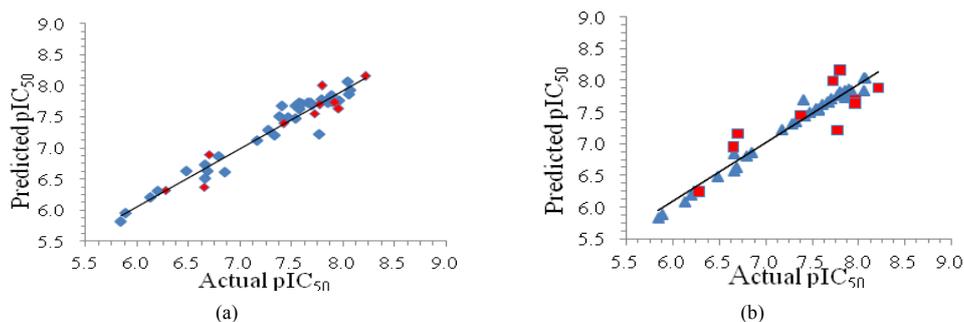


图 4 (a)COMFA 模型中化合物的预测值与实验值的线性回归图;(b)COMSIA(SEA)模型中化合物的预测值和实验值的线性回归图;红色代表 11 个测试集化合物,蓝色代表 34 个训练集化合物

Fig.4 (a) Linear regression between actual and predicted activities of 45 compounds from the whole set in COMFA model. (b) Linear regression between actual and predicted activities of 45 compounds from the whole set in the COMSIA (SEA) model; the red color presents 11 compounds from the test set and the blue color presents 34 compounds from the training set

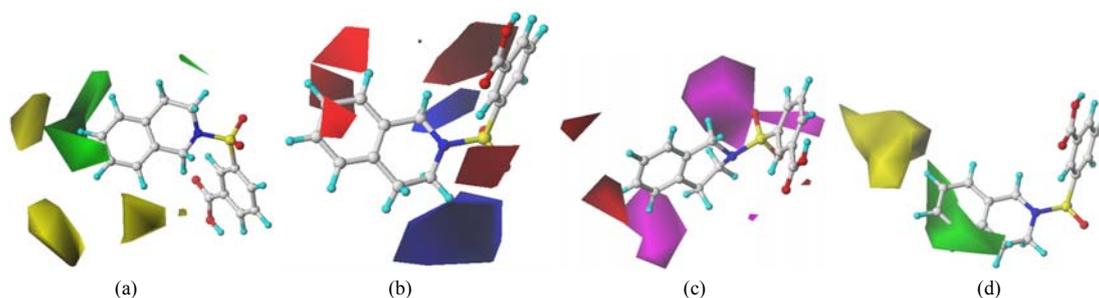


图 5 COMFA 模型立体等势图(a)和静电等势图(b); COMSIA 模型氢键受体等势图(c)和立体等势图(d)

Fig.5 Contour maps of COMFA and COMSIA models: (a) COMFA steric contour map; (b) COMFA electrostatic contour map; (c) COMSIA hydrogen-bond acceptor contour map; (d) COMSIA steric contour map

在 COMSIA 模型中,立体场和静电场的三维等势图与 COMFA 模型中立体场、静电场的等势图相似。如图 5(d),G 区域主要出现在 5 和 6 位取代位置,Y 区域主要出现在 7 和 8 取代位点附近。图 5(c)为 COMSIA 模型中的氢键受体场等势图,在氢键受体场中,紫色(P)区域代表氢键受体有助于化合物活性的提高,红色(R)表示在该区域内引入氢键受体使活性降低。在 5 位附近出现 P 区域,表明在 5 位引入氢键受体可以提高化合物的活性。如化合物 10(8.051)的 5 位为-NO₂取代,是氢键受体,其活性大于

化合物 1(7.886); 化合物 9(7.276)的 5 位是-NH₂ 取代, 虽有氢键受体作用, 但主要还是氢键供体作用, 因此其活性比化合物 1 的活性小。R 区域在 6 和 7 位取代位置出现, 表明在该位置引入氢键供体对提高活性不利。例如, 化合物 22(7.569)和 23(7.420)的 6 位分别被-OH 和-OMe 取代其活性均小于 H 取代的化合物 1(7.886)。此外, 在磺胺的 N 和 O 原子附近也出现 P 区域, 表明磺胺结构有利于化合物活性的提高, 是该类化合物与 AKR1C3 受体作用的重要部分, 羧基附近也被 P 覆盖, 表明羧基以氢键受体形式与关键氨基酸相互作用。

3 结论

AKR1C3 是治疗激素依赖型和激素非依赖性前列腺癌的新作用靶点, 其抑制剂是一类具有发展前景的抗癌药。本文以训练集化合物及其生物活性构建了 COMFA 和 COMSIA 模型, 并用测试集进行了验证。COMFA 模型结果为 $q^2=0.761$, $R^2=0.973$, $N=6$, $SEE=0.122$, $F=185.963$, $R^2_{bs}=0.980$, $SD_{bs}=0.007$; COMSIA(SEA)模型结果为 $q^2=0.734$, $R^2=0.984$, $N=10$, $SEE=0.097$, $F=147.850$, $R^2_{bs}=0.994$, $SD_{bs}=0.005$ 。其外部测试 R^2_{pred} 分别为 0.864 和 0.756; r^2_m 分别为 0.8137 和 0.5377。结果表明, 所构建的 3D-QSAR 模型具有较高的可靠性和较强的预测能力。配体与受体相互作用模型显示该系列化合物主要占据受体结合位点的 SP1 和 SP3 口袋, 羧基是相互作用的关键基团, 其与重要氨基酸 Tyr-55 和 His-117 形成氢键。QSAR 模型显示, 在 2、5 或 6 位适当增加取代基体积, 或在 5 位引入氢键受体, 或在 7 位引入电负性取代基则能提高化合物的生物活性。本研究得到的 3D-QSAR 模型为设计更强活性和更高选择性的 AKR1C3 抑制剂提供了理论依据。

参考文献

- [1] B J Feldman, D Feldman. Nat. Rev. Cancer, 2001, 1(1): 34-45.
- [2] R Siegel, D Naishadham, A Jemal. CA: A Cancer Journal for Clinicians, 2013, 63(1): 11-30.
- [3] G Attard, C S Cooper, J S de Bono. Cancer Cell, 2009, 16(6): 458-462.
- [4] J A Locke, E S Guns, A A Lubik et al. Cancer Res., 2008, 68(15): 6407-6415.
- [5] G Attard, J Richards, J S de Bono. Clinical Cancer Res., 2011, 17(7): 1649-1657.
- [6] J S de Bono, C J Logothetis, A Molina et al. New Engl. J. Med., 2011, 364(21): 1995-2005.
- [7] A H M Reid, G Attard, D C Danila et al. J. Clin. Oncol., 2010, 28(9): 1489-1495.
- [8] G Attard, A H M Reid, R J Auchus et al. J. Clin. Endocrinol Metab., 2012, 97(2): 507-516.
- [9] H I Scher, K Fizazi, F Saad et al. New Engl. J. Med., 2012, 367(13): 1187-1197.
- [10] M Stanbrough, G J Bubley, K Ross et al. Cancer Res., 2006, 66(5): 2815-2825.
- [11] N Sharifi. Mol. Endocrinol., 2013, 27: 708-714.
- [12] K H Chang, C E Ercole, N Sharifi. Brit. J. Cancer, 2014, 111: 1249-1254.
- [13] C Cai, S Chen, P Ng et al. Cancer Res., 2011, 71(20): 6503-6513.
- [14] Y Tian, L Zhao, Y Li et al. Diagn Pathol., 2014, 9: 42.
- [15] A O Adeniji, M Chen, T M Penning. J. Steroid Biochem. Mol. Biol., 2013, 137: 136-149.
- [16] P Bydal, V Luu-The, F Labrie et al. Eur. J. Med. Chem., 2009, 44(2): 632-644.
- [17] M Sinreih, I Sosič, N Beranič et al. Bioorg. Med. Chem. Lett., 2012, 22(18): 5948-5951.
- [18] S Endo, D Hu, T Matsunaga et al. Bioorg. Med. Chem., 2014, 22(19): 5220-5233.
- [19] 田原 僮, 赵丽晶, 王政 等. 肿瘤防治研究, 2013, 40(7): 720-722.
- [20] P Brožič, S Turk, A O Adeniji et al. J. Med. Chem., 2012, 55(17): 7417-7424.
- [21] S M Jamieson, D G Brooke, D Heinrich et al. J. Med. Chem., 2012, 55(17): 7746-7758.
- [22] 李鸣建, 冯惠, 冯长君 等. 化学通报, 2015, 78(2): w011.
- [23] J M Planesas, R M Claramunt, J Teixidó et al. J. Chem. Inf. Model, 2011, 51(4): 777-787.
- [24] J Zheng, G Xiao, J Guo et al. Chem. Biol. Drug Des., 2011, 78(2): 314-321.
- [25] H Abedi, H Ebrahimzadeh, B Jahan et al. Struct. Chem., 2013, 24: 433-444.
- [26] P P Roy, S Paul, I Mitra et al. Molecules, 2009, 14(5): 1660-1701.
- [27] P P Roy, K Roy. Q SAR & Combin. Sci., 2008, 27(3): 302-313.
- [28] M C Byrns, Y Jin, T M Penning. J. Steroid Biochem. Mol. Biol., 2011, 125(1): 95-104.
- [29] 王琴, 梅虎, 孙家英 等. 化学通报, 2011, 74(1): 54-60.
- [30] 王勤金, 班树荣, 李青山. 化学通报, 2014, 77(6): w014.