

文章编号: 1674-9057(2018)03-0529-08

doi:10.3969/j.issn.1674-9057.2018.03.024

新型噻唑-2-乙胺类 HAT 抑制剂的 QSAR 研究

段家喜, 赵 赛, 易忠胜, 聂瑾芳

(桂林理工大学 a. 化学与生物工程学院; b. 广西高校食品安全与检测重点实验室;
c. 岩溶地区水污染控制与用水安全保障协同创新中心, 广西 桂林 541004)

摘要: 采用 VSMVI 变量筛选方法从大量描述符中筛选最优子集, 再由 MLR 回归方法建立了 83 个噻唑-2-乙胺类化合物与非洲人类锥虫病(HAT)抑制活性之间的二维定量结构-活性相关(2D-QSAR)模型, 最优模型的拟合相关系数($r^2 = 0.8892$)和交叉验证相关系数($q^2 = 0.8574$)表明模型具有良好的稳健性、拟合能力和预测能力。模型描述符在一定程度上反映了分子的二维结构和疏水性对抑制活性具有重要影响。同时采用基于 CoMFA 和 CoMSIA 方法的三维定量结构-活性相关(3D-QSAR)建立了相关性显著、预测能力强的定量模型(CoMFA: $r^2 = 0.924$, $q^2 = 0.516$; CoMSIA: $r^2 = 0.944$, $q^2 = 0.531$), 其中 CoMSIA 的疏水场贡献率最高, 说明了分子的疏水作用对抑制活性的重要影响。

关键词: 噻唑-2-乙胺类化合物; 非洲人类锥虫病(HAT); 二维定量结构-活性相关(2D-QSAR); 基于变量相互作用的变量筛选方法(VSMVI); 比较分子场分析方法(CoMFA); 比较分子相似性指数分析法(CoMSIA)

中图分类号: R917

文献标志码: A

0 引言

非洲人类锥虫病(human african trypanosomiasis, HAT)^[1]是由一种非洲虫类——舌蝇又名采采蝇(tsetse fly, 拉丁名 *Glossina*)传播锥虫而引起的人体寄生虫疾病。虽然有调查指出非洲地区的感染人数每年都在减少, 但感染总人数仍然较大^[2]。HAT 发病分为阶段 1(早期, 血淋巴)^[3]和阶段 2(晚期, 脑膜脑炎)^[4]。目前, 已有药物苏拉明^[5]和戊烷脒^[6]用于 HAT 的一阶段治疗, 美拉肿醇^[7]和依氟鸟氨酸^[8]用于 HAT 的二阶段治疗。然而, 实际用于临床的 HAT 药物数量少, 伴有毒性和锥虫的高度抗原变异所带来的耐药性等问题迫切需要开发新型、安全和易于管理的 HAT 药物。

HAT 抑制剂的设计对发掘潜在的新型 HAT 药物提供了帮助。近些年设计针对的靶标不同导致抑制剂的种类繁多, 有针对锥虫体内磷酸二酯酶

(cAMP)靶标而设计的抑制剂, 其中 TbrPDEB1、2 两种 cAMP 已被证实可作为 HAT 的治疗靶点^[9]; 有针对人体一些靶标蛋白酶^[10]、亮氨酸 tRNA 合成酶^[11]和拓扑异构酶^[12]而设计的抑制剂; 也有直接针对分子药效基团表现出来的体外活性来设计优化分子的结构。其中不乏抑制活性表现较好且具有潜在药物价值的 HAT 抑制剂, 比如联脒类、硝唑类和苯并硼唑类等^[13-14]多种化合物已被列为候选药物, 其中大部分候选药物由于伴随着一定的毒性和较低的临床药效等缺点, 实际上很难通过治疗 HAT 的临床研究阶段。HAT 抑制剂由于其结构多样性, 药效基团的复杂多变, 阻碍了进一步开发新型高效的潜在药物。

对已经通过实验测定抑制活性的 HAT 抑制剂进行定量结构-活性相关(QSAR)^[15]研究来指导开发新型潜在药物价值的抑制剂成为了一个比较简单的方法。本文结合 2D-QSAR 和基于比较分子

收稿日期: 2017-04-16

基金项目: 国家自然科学基金项目(21267008; 21467006); 广西自然科学基金项目(2014GXNSFAA118034); 广西高等学校高水平创新团队及卓越学者计划项目(桂教人[2014]49)

作者简介: 段家喜(1991—), 男, 硕士, 研究方向: 环境毒理学, 1316539887@qq.com。

通讯作者: 易忠胜, 博士, 教授, yzs@glut.edu.cn。

引文格式: 段家喜, 赵赛, 易忠胜, 等. 新型噻唑-2-乙胺类 HAT 抑制剂的 QSAR 研究 [J]. 桂林理工大学学报, 2018, 38(3): 529-536.

场分析法 (CoMFA)、比较分子相似性指数分析法 (CoMFA) 的 3D-QSAR^[16] 研究噻唑-2-乙胺类化合物与 HAT 抑制活性之间的关系, 建立具有显著预测能力的噻唑-2-乙胺类化合物的活性预测模型, 并考察它们的结构对 HAT 抑制活性的影响。

1 数据与方法

1.1 数据来源

选用的 83 个新型噻唑-2-乙胺类衍生物抑制活性数据来源于文献 [17]。其生物抑制活性采用化合物对锥虫的半数抑制浓度表示 (IC_{50} , nmol/L), 其负对数值 pIC_{50} 作为建立数学模型的因变量, 以消除量纲的影响。通常对于大样本集的模型建立, 一般将样本集分为训练集和测试集, 建立的模型由测试集作为外部样本集检验其预测能力, 故将 83 个样本按生物活性值大小排列后均匀选取训练集 (63 个化合物, 76%) 和测试集 (20 个化合物, 24%), 确保在模型建立和外部测试中选用的化合物分布均匀广泛。分子结构如表 1 和图 1 所示。

1.2 2D-QSAR 建模

化合物的结构式均由 Chemoffice 构建, 再用 SYBYL 软件的 Tripos 标准力场进行分子能量最小优化。相应的优化参数设置: 能量收敛标准为 0.005 kJ/mol, 最大迭代次数为 1 000, 对待优化的分

子附加 Gasteiger-Huckel 电荷。将优化后的分子结构导入 DRAGON 软件计算相应的结构描述符。计算得到各类描述符共 4 885 个, 这些描述符全面反映了分子的各种物化特性和结构特征。除去化合物分子结构中对建模影响不显著的描述符 (包括方差为零和 90% 的样本为零的描述符), 以及只选相关系数大于 0.95 的多个描述符中的 1 个, 最终初选得到 579 个描述符。建模采用通用变量筛选与建模程序 (GVSM), 首先使用基于变量相互作用的变量筛选方法 (variable selection method based on variables interaction, VSMVI)^[18-19] 进行变量寻优, 然后利用多元线性回归分析 (multiple linear regression, MLR) 方法建立 2D-QSAR 模型, 建立模型后进行 y -randomization 检验, 并计算相应的应用域。

1.3 CoMFA 和 CoMSIA 建模

选取对 HAT 具有最优抑制活性的 70 号化合物作为模板分子, 以分子结构中的酰胺基团为公共骨架, 以化合物最优构象, 对 83 个化合物采用 SYBYL 软件的 Align-Database 模块进行分子骨架的叠合, 分别建立 CoMFA 和 CoMSIA 模型。CoMFA 建模参数: 选用 Tripos 标准场, 静电场能和立体场能的阈值设定为 30 kcal · mol⁻¹; CoMSIA 建模参数: 衰减系数设为 0.3, 其余均为默认值。

2 结果与讨论

2.1 2D-QSAR 模型

2.1.1 2D-QSAR 模型建立 以抑制剂的半数抑制活性负对数值 pIC_{50} 为因变量, 采用 VSMVI 方法从 579 个描述符中筛选最优子集。筛选过程中利用留一法交叉验证 (LOOCV) 进行验证, 根据交叉验证系数 (q^2) 和交叉验证的均方根误差 ($RMSEV$) 随变量数变化的曲线 (图 2), 确定了最优模型的变量数为 6 (当模型增加一个变量, 对应的 q^2 增加量低于 5% 时不再增加模型变量, 实际当 q^2 值在变量数为 6 增加到 7 时, 相应 q^2 值的增量只接近 2%)。

$$pIC_{50} = (-15.9270 \pm 4.4180) + (13.8073 \pm 2.1053)BIC4 + (-8.7436 \pm 0.6880)GATS4i + (20.1235 \pm 2.9012)SpMin2_Bh(s) + (-0.0242 \pm 0.0048)P_VSA_e_3 + (-0.0439 \pm 0.0045)RDF100s + (-0.9088 \pm 0.0845)CATS2D_06_DL,$$

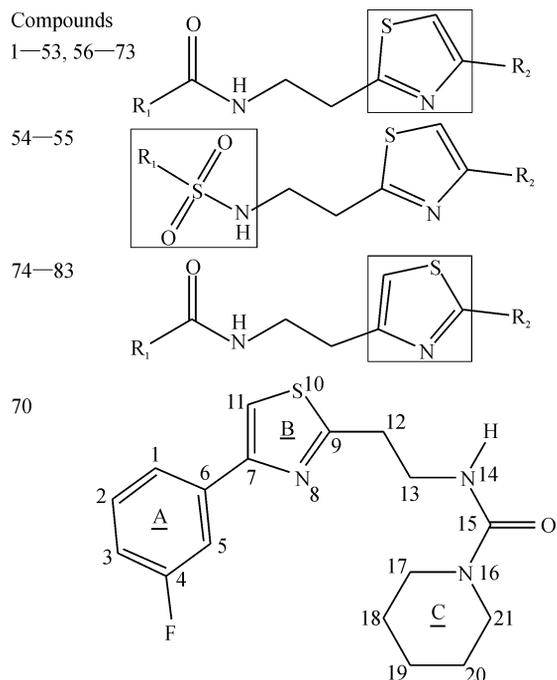


图 1 分子结构图

Fig. 1 Molecular structure

表1 分子结构与对应 pIC_{50} 实验值和3个模型的预测值
Table 1 Molecular structures, experimental and predicted pIC_{50} values

Compounds	R_1	R_2	$pIC_{50}/$ ($mmol \cdot L^{-1}$)	2D-QSAR	CoMFA	CoMSIA
				$Pred_1$	$Pred_2$	$Pred_3$
1	Phenyl	Phenyl	6.199	5.480	5.560	5.603
2	Phenyl	3-Cyanophenyl	5.721	6.297	6.087	5.777
3	Phenyl	4-Cyanophenyl	4.370	4.508	4.461	4.469
4	Phenyl	4-(Trifluoromethyl)phenyl	3.839	4.499	3.374	3.948
5	Phenyl	3-Nitrophenyl	5.565	5.401	5.534	5.301
6*	Phenyl	4-Nitrophenyl	4.533	5.049	4.785	4.938
7	Phenyl	3-Methoxyphenyl	5.214	5.618	5.188	5.479
8	Phenyl	4-Methoxyphenyl	4.738	4.847	4.870	4.722
9	Phenyl	2-Fluorophenyl	6.807	6.190	6.541	6.492
10	Phenyl	3-Fluorophenyl	6.633	6.355	6.424	6.456
11*	Phenyl	4-Fluorophenyl	6.662	5.861	6.076	5.774
12	Phenyl	3-Chlorophenyl	6.345	6.368	6.084	6.136
13	Phenyl	4-Chlorophenyl	4.876	4.739	4.833	5.058
14*	Phenyl	3-Bromophenyl	6.257	5.765	5.991	5.982
15	Phenyl	4-Bromophenyl	4.460	4.127	4.674	4.695
16	Phenyl	2,4-Difluorophenyl	6.790	6.598	6.387	6.519
17	Phenyl	2,5-Difluorophenyl	5.788	6.650	6.274	6.341
18*	Phenyl	2,6-Difluorophenyl	5.914	4.713	6.200	6.211
19	Phenyl	3,4-Difluorophenyl	6.839	6.096	6.331	6.398
20	Phenyl	3,5-Difluorophenyl	5.790	6.352	6.227	6.373
21*	Phenyl	2,4,5-Trifluorophenyl	5.672	5.709	5.850	6.490
22	Phenyl	2-Furanyl	5.597	5.806	5.577	5.373
23	Phenyl	2-Thiophenyl	5.790	6.463	5.741	5.621
24	Phenyl	2-Pyridyl	5.440	5.352	5.622	5.867
25*	Phenyl	3-Pyridyl	5.333	5.642	5.878	5.732
26	Phenyl	4-Pyridyl	4.987	4.732	5.689	5.417
27	3-Cyanophenyl	Phenyl	6.767	6.289	7.079	6.678
28*	4-Cyanophenyl	Phenyl	4.391	4.581	4.069	4.683
29	3-Nitrophenyl	Phenyl	6.824	6.678	6.773	6.913
30	2-Fluorophenyl	Phenyl	6.337	6.823	6.272	6.628
31*	3-Fluorophenyl	Phenyl	6.417	6.490	6.445	6.204
32	4-Fluorophenyl	Phenyl	5.848	6.054	5.969	5.467
33	2-Chlorophenyl	Phenyl	6.316	5.550	6.284	6.631
34	3-Chlorophenyl	Phenyl	5.824	6.270	6.299	5.898
35	4-Chlorophenyl	Phenyl	4.462	4.391	4.692	4.474
36	2-Pyrrolyl	Phenyl	5.697	6.164	5.912	5.697
37*	2-Furanyl	Phenyl	6.317	5.887	6.134	6.338
38	2-Thiophenyl	Phenyl	6.593	6.245	6.003	6.594
39	3-Thiophenyl	Phenyl	6.721	6.480	6.550	6.544
40*	3-Pyridyl	Phenyl	5.799	6.636	5.327	5.841
41	4-Pyridyl	Phenyl	4.783	4.905	5.522	6.313
42*	Cyclopentyl	Phenyl	6.572	8.977	7.310	6.941
43*	Cyclohexyl	Phenyl	6.785	7.133	6.367	6.384
44	N-Pyrrolidinyl	Phenyl	6.903	6.954	6.866	6.711
45*	N-Pyrrolyl	Phenyl	5.759	6.611	6.248	5.770
46	N-Piperidinyl	Phenyl	7.690	7.485	7.586	7.604
47	N-Morpholinyl	Phenyl	6.292	6.133	6.367	6.26
48	N-Azepanyl	Phenyl	7.287	7.194	7.658	7.337
49	tert-Butylamino	Phenyl	5.738	5.674	5.856	5.816

续表 1

Compounds	R ₁	R ₂	pIC ₅₀ / (mmol · L ⁻¹)	2D-QSAR	CoMFA	CoMSIA
				Pred ₁	Pred ₂	Pred ₃
50	Benzylamino	Phenyl	4.045	3.896	4.107	4.187
51	Cyclohexylamino	Phenyl	4.996	5.022	5.035	4.984
52	Phenylamino	Phenyl	3.510	3.795	3.264	3.322
53	Phenoxy	Phenyl	4.209	4.296	4.331	4.136
54*	Phenyl	Phenyl	4.120	5.883	6.201	5.235
55	4-Methylphenyl	Phenyl	4.277	4.818	4.568	4.405
56	3-Cyanophenyl	2-Fluorophenyl	6.406	6.655	6.648	6.778
57	3-Cyanophenyl	3-Fluorophenyl	6.983	6.976	6.708	6.686
58*	3-Cyanophenyl	4-Fluorophenyl	6.886	5.981	6.683	6.745
59	3-Nitrophenyl	2-Fluorophenyl	6.876	6.981	6.775	6.611
60	3-Nitrophenyl	3-Fluorophenyl	6.889	7.051	6.637	7.229
61	3-Nitrophenyl	4-Fluorophenyl	6.833	6.557	6.681	6.940
62	3-Fluorophenyl	3-Fluorophenyl	6.488	6.670	6.423	6.618
63	3-Fluorophenyl	4-Fluorophenyl	6.469	6.322	6.369	6.109
64	2-Thiophenyl	2-Fluorophenyl	6.752	6.371	6.876	6.706
65*	2-Thiophenyl	3-Fluorophenyl	6.827	6.908	6.313	6.982
66*	2-Thiophenyl	4-Fluorophenyl	6.752	6.276	5.862	6.550
67	<i>n</i> -Pyrrolidinyl	2-Fluorophenyl	7.305	7.033	7.125	7.123
68*	<i>n</i> -Pyrrolidinyl	3-Fluorophenyl	7.287	7.531	7.117	7.279
69	<i>n</i> -Pyrrolidinyl	4-Fluorophenyl	6.500	6.609	6.588	6.649
70	<i>n</i> -Piperidinyl	3-Fluorophenyl	8.046	8.066	7.954	7.909
71*	<i>n</i> -piperidinyl	4-Fluorophenyl	7.921	6.966	7.575	7.553
72	<i>n</i> -Piperidinyl	2,4-Difluorophenyl	8.004	7.690	7.880	8.055
73	<i>n</i> -Piperidinyl	3,4-Difluorophenyl	8.013	8.034	8.180	8.251
74	Phenyl	3-Fluorophenyl	6.697	6.586	6.222	6.488
75	Phenyl	4-Fluorophenyl	6.783	6.661	6.191	6.356
76*	Phenyl	4-Chlorophenyl	4.943	5.246	5.297	5.341
77	2-Fluorophenyl	Phenyl	6.271	6.504	6.340	6.233
78	3-Fluorophenyl	Phenyl	6.074	6.089	6.352	6.244
79	2-Chlorophenyl	Phenyl	5.818	5.492	6.264	5.998
80	3-Chlorophenyl	Phenyl	6.192	6.469	5.761	5.946
81	4-Chlorophenyl	Phenyl	4.514	4.855	4.971	4.682
82	2-Thiophenyl	Phenyl	6.652	6.876	6.844	6.846
83	Cyclohexyl	Phenyl	6.717	7.056	6.671	6.621

注: *为外部测试集,其他为训练集; pIC₅₀为实验值; Pred为模型的预测值。

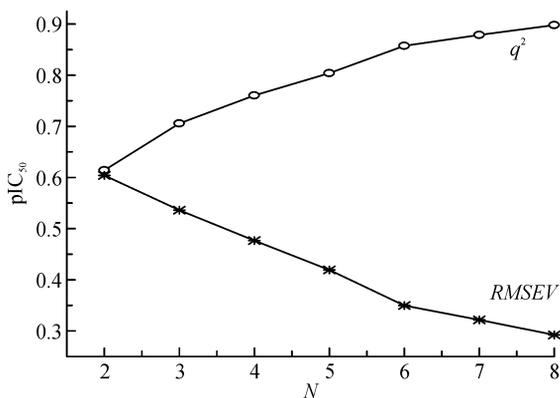


图 2 q^2 和 $RMSEV$ 随变量数的变化

Fig. 2 Changes in q^2 and $RMSEV$ with the number of variables

其中, $RMSEV = 0.3709$, $RMSE = 0.3498$ 为模型的均方根误差; q^2 和 r^2 为 LOOCV 交叉验证系数, $r^2 = 0.8892$, $q^2 = 0.8574$; $F = 74.9253$ 为模型的 Fisher 统计量, “±” 表示回归系数的标准偏差。从模型的统计参数可以看出, 模型具有良好的拟合能力和预测能力, 而且 r^2 和 q^2 值相差不大, 说明模型具有良好的稳定性。图 3 为用训练集建立的 QSAR 模型预测全部样本的 pIC₅₀ 值的相关性图, 由 2D-QSAR、CoMFA 和 CoMSIA 模型得出的实验值与预测值可知, 3 个模型都表现出较强的预测能力。

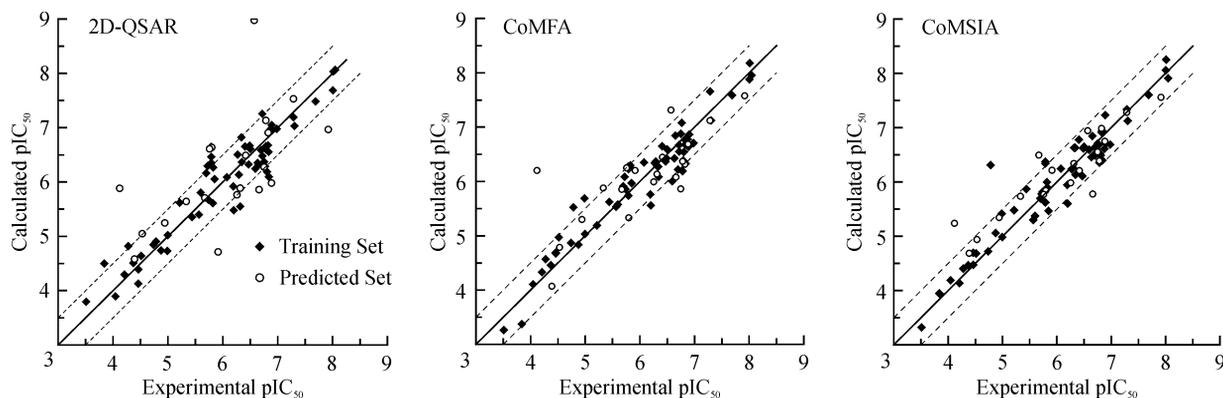


图3 模型 2D-QSAR、CoMFA、CoMSIA 实验值与预测值相关图

Fig. 3 Correlation diagrams of experimental value and predicted value for model 2D-QSAR, CoMFA, CoMSIA

2.1.2 模型验证 对模型进行 10 000 次的 y 随机化 (y -randomization) 检验, 结果见表 2。可以看出随机检验的最大拟合相关系数 (r_{\max}^2) 和最大交叉检验相关系数 (q_{\max}^2) 都远远小于真实模型, 而且均方根误差 ($RMSE$, $RMSEV$) 的平均值 (y -rand_{average}) 均大于 1, 远远超过真实模型的误差, y 随机化检验的结果说明了高质量模型的建立并非偶然因素所致。

表 2 10 000 次 y -randomization 检验结果Table 2 Results of y -randomization validation after 10 000 times

projects	q^2	r^2	$RMSE$	$RMSEV$
Real model	0.857 4	0.889 2	0.396 8	0.349 8
y -rand q_{\max}^2	0.223 5	0.370 4	0.784 7	0.871 4
y -rand r_{\max}^2	0.223 5	0.370 4	0.784 7	0.871 4
y -rand _{average}	-0.133 3	0.089 9	1.077 8	1.202 7

2.1.3 模型应用域和离群值 在 QSAR 研究中, 模型建立后有必要对模型确定一个适用的最佳预测范围, 即应用域 (applicability domain, AD)^[20]。本文采用 Hotelling T2 和 Leverage 方法, 并用 Williams 图^[21]来表示模型的应用域, 图 4 给出了最优模型的应用域 (Y 离群警戒值 $\delta^* = \pm 2$, X 离群值警戒值 $h^* = 0.286$)。可以看出, 训练集化合物 32、43、46 和测试集化合物 2、16 超过或接近 X 离群值警戒值, 从这些化合物的分子结构选定的描述符角度来看, 与其他化合物有较大的差异, 直接导致预测出现严重的偏差。其他化合物在 Williams 图上表现得都比较集中, 表明模型整体预测效果较好。对于应用域上的离群值并不能直接删除再建立新的模型, 应当谨慎处理, 否则, 会导致整个模型的扭曲而不具备真实性。

2.1.4 分子结构与 HAT 抑制活性的关系 表 3 列出了模型中反映分子结构特征或物化特性的描述符以及与之相对应的标准回归系数。可以看出, 最优模型中反映分子 2D 结构的 GATS4i 描述符和反映供体分子疏水性的 CATS2D_06_DL 描述符对模型的影响最大, 相应的标准回归系数分别为 -0.611 4 和 -0.530 6; 其次是反映分子径向分布的 RDF100s 描述符, 以及对模型的影响能力相似的反映分子特征的 Burden 特征值的 SpMin2_Bh(s) 和反映分子中化学结合信息的 BIC4 描述符。对模型影响较小的为反映原子桑德森键电负性的 P_VSA_e_3 描述符。最优模型中的描述符在一定程度上反映了分子的特征结构和分子的疏水性, 是化合物对 HAT 抑制活性的最主要影响因素。

2.2 CoMFA 和 CoMSIA 模型

2.2.1 CoMFA 和 CoMSIA 模型建立 采用 CoMFA

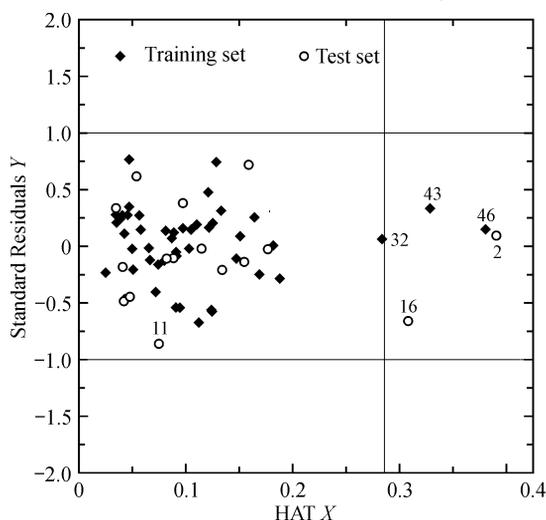


图 4 2D-QSAR 模型的 Williams 图

Fig. 4 Williams figure for 2D-QSAR model

表3 模型 2D-QSAR 中描述符的含义以及标准回归系数
Table 3 Meanings of the descriptors and the standard regression coefficients in 2D-QSAR

描述符符号	描述	归类	标准回归系数
BIC4	描述分子化学键的结合信息	分子信息指数	0.369 2
GATS4i	电离电势加权自相关	2D 结构自相关	-0.611 4
SpMin2_Bh(s)	质量加权负荷特征值	Burden 特征值描述符	0.402 1
P_VSA_e_3	原子桑德森电负性	P_VSA-like 原子特征描述符	-0.236 2
RDF100s	分子结构, 径向分布函数	RDF 描述符	-0.477 6
CATS2D_06_DL	供体分子的疏水性	CATS 2D 描述符	-0.530 6

方法计算关于立体场和静电场的描述符, 然后用偏最小二乘法 (partial least square, PLS)^[22] 建立 CoMFA 模型。模型给出了 62.8% 的立体场贡献率和 37.2% 的静电场贡献率, 其统计学参数为: 交叉验证相关系数 q^2 为 0.516 (因此可以认为模型具有较好的预测能力), 最佳主成分数 N 为 7, 模型拟合的相关系数 r^2 为 0.924, 标准偏差 SEE 为 0.310, 统计量 F 值为 93.349, 该模型对整个样本集的预测值见表 1 中的 $Pred_2$ 。

采用 CoMSIA 方法计算关于立体场、静电场、疏水场、氢键供体场和氢键受体场的 5 种力场描

述符, 使用基于不同力场的描述符建立了 7 个主成分的 CoMSIA 模型。模型的统计参数列于表 4, 可以看出基于立体场、静电场和疏水场的模型 M_7 相对其他模型的综合表现更优, 模型给出了 51.2% 的立体场贡献率、15.7% 的立体场贡献率和 33.0% 的静电场贡献率, 该模型对整个样本集的预测值见表 1 中的 $Pred_3$ 。

2.2.2 CoMSIA 模型的三维等高线图 图 5a 为立体场等势图, 在 A 环的 C_1 、 C_3 和 C 环的 C_{20} 的周围有黄色区域, 说明在此区域引入小体积的基团有利于提高化合物的 HAT 抑制活性。如 9 号化合

表 4 5 种力场不同组合的 CoMSIA 模型统计参数
Table 4 Statistic results of CoMSIA with different combinations of 5 fields

Model	Field	q^2	r^2	SEE	F	N
M_1	0.057A/0.046D/0.480H/0.132S/0.284E	0.534	0.942	0.274	106.848	8
M_2	0.043D/0.490H/0.151S/0.316E	0.528	0.944	0.269	110.912	8
M_3	0.059A/0.505H/0.139S/0.297E	0.523	0.942	0.272	108.355	8
M_4	0.098D/0.077A/0.319S/0.506E	0.517	0.845	0.437	50.116	6
M_5	0.053D/0.066A/0.559H/0.322E	0.511	0.938	0.281	100.667	8
M_6	0.057D/0.111A/0.647H/0.185S	0.416	0.922	0.316	78.692	8
M_7	0.512H/0.157S/0.330E	0.531	0.944	0.267	112.425	8

注: A—氢键受体场; D—氢键供体场; H—疏水场; S—立体场; E—静电场。

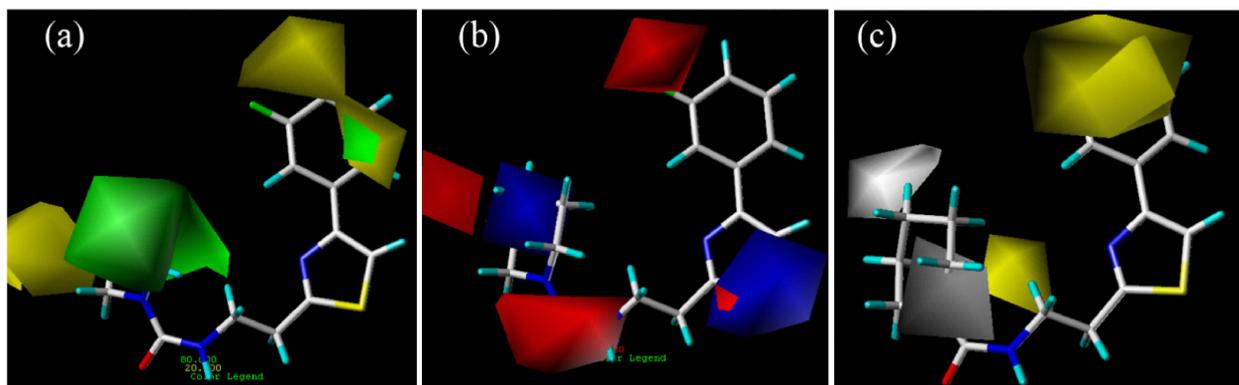


图 5 立体场 (a)、静电场 (b) 和疏水场 (c) 的等势图

Fig. 5 3D contour plot from CoMSIA analysis in steric (a), electrostatic (b) and hydrophobic (c) field

物($pIC_{50} = 6.807$)在 A 环 C_1 位置没有引入取代基, 比同一位置引入氟基团的 18 号化合物($pIC_{50} = 5.914$)的抑制活性强。75 号化合物($pIC_{50} = 6.783$)在 A 环的 C_3 位置引入氟基团, 抑制活性远强于在同一位置引入三氟甲基的 4 号化合物($pIC_{50} = 3.839$), 也强于在同一位置下引入甲醚基团的 7 号化合物。在 C 环的 C_{18} 周围有一个绿色区域, 说明在此区域引入大体积基团对化合物的 HAT 抑制活性有利。如 80 号化合物($pIC_{50} = 6.162$)在 C 环的 C_{18} 位置引入氯基团较在 C_{19} 的位置上引入氯基团的 35 和 81 号化合物($pIC_{50} = 4.462, 4.514$)抑制活性明显加强。

图 5b 为静电场等势图, A 环的 C_4 和 C 环连接的 C_{15} 周围有红色区域, 说明在此区域引入负电性基团可以提高化合物的 HAT 抑制活性。如, 在 A 环的 C_4 位置取代基分别为氟基的 74 号化合物、氯基的 12 号化合物和氰基的 2 号化合物, 其抑制活性依次下降($pIC_{50} = 6.697 > 6.345 > 5.721$), 这是由于共轭效应下 C_4 位置上连接的强吸电子基团会吸引整个苯环的电子云往取代基的位置偏移, 从而增加负电性。在 B 环的 S_{10} 和 C 环的 C_{19} 周围有蓝色区域, 说明在这两个区域正电性基团或者减少此区域的负电性将有利于提高化合物的 HAT 抑制活性。如, B 环内硫的电负性要强于氮, 所以整个环上的电子云密度 S_{10} 位置是要低于 N_8 位置(S_{10} 位置的负电性减弱), 导致 43 和 30 号化合物的 HAT 抑制活性分别强于 83 和 77 号化合物($pIC_{50} = 6.785 > 6.717, 6.337 > 6.271$)。在 C 环的对位上 32 号化合物为氟基取代, 28 号化合物为氰基取代, 由于在苯环上取代考虑到共轭效应和氰基吸电能力强于氟基, 在 C_{19} 的位置上 28 号化合物的电子云密度更为密集, 正电性反而减弱, 导致 28 号化合物的抑制活性减弱($pIC_{50} = 5.848 > 4.391$)。

图 5c 为疏水场等势图, 在 A 环的 C_3 和 C_4 周围出现大面积的黄色区域, 说明此区域取代的疏水性基团能增加化合物的 HAT 抑制活性。如, 11 号化合物在 A 环的对位 C_3 上取代的是氟基为疏水基团, 而 3 号化合物在同一位置取代的是极性氨基为亲水基团, 所以 11 号化合物的 HAT 抑制活性明显较强($pIC_{50} = 6.662 > 4.370$)。10 号化合物和 2 号化合物分别在 A 环间位 C_4 上用氟基和氰基取

代导致了相同的结果。在 C 环的 N_{16} 和对位的周围出现灰色区域, 表明在此区域引入亲水基团有利于增加化合物的 HAT 抑制活性。如, 57 和 58 号化合物在 C 环的 C_{20} 位置上的氰基为亲水性基团, 其表现的抑制活性要强于在同一位置用氟基取代的 62 和 63 号化合物($pIC_{50} = 6.983 > 6.488, 6.886 > 6.469$), 也强于同一位置用硝基取代的 60 和 61 号化合物。

3 结 论

采用 VSMVI 方法进行变量筛选, 以 MLR 回归方法建立了噻唑-2-乙胺类化合物的 HAT 抑制活性的 6 变量最优 2D-QSAR 模型, 同时采用 CoMFA 和 CoMSIA 方法从立体场效应、静电作用和疏水作用等方面分析了结构对其 HAT 抑制活性之间的关系。QSAR 模型的描述符和力场贡献率均表明疏水作用对 HAT 抑制活性的重要影响。力场等势图显示在 A 环内侧(C_4, C_5)引入强吸电子基团与苯环的共轭效应能增加电子云密度有利于提高化合物的抑制活性, A 环外侧和对位的疏水性和大体积基团会大幅度降低化合物的抑制活性; C 环对位不适合引入强吸电子基团, C 环内侧间位(C_{18})适合引入吸电子基团, 外侧间位(C_{20})适合引入极性亲水性基团。QSAR 模型的相关性显著、预测能力强, 可为 HAT 抑制剂的抑制活性研究提供有力的支持。

参考文献:

- [1] Stich A, Abel P M, Krishna S. Human african trypanosomiasis [J]. British Medical Journal, 2002, 325 (7357): 203 - 206.
- [2] Simarro P P, Franco J R, Diarra A, et al. Diversity of human African trypanosomiasis epidemiological settings requires fine-tuning control strategies to facilitate disease elimination [J]. Research & Reports in Tropical Medicine, 2013, 4: 1 - 6.
- [3] Matthews K R. The developmental cell biology of *Trypanosoma brucei* [J]. Journal of Cell Science, 2005, 118 (2): 283 - 290.
- [4] Legros D, Ollivier G, Gastelluetcheorry M, et al. Treatment of human African trypanosomiasis-present situation and needs for research and development [J]. Lancet Infectious Diseases, 2002, 2 (7): 437 - 440.
- [5] Walther M M, Figg W D, Linehan W M. Intravesical suramin: a novel agent for the treatment of superficial transitional-cell carcinoma of the bladder [J]. World Journal of Urology, 1996, 14: S8 - S11.

- [6] Wang C C. Molecular mechanisms and therapeutic approaches to the treatment of African trypanosomiasis [J]. Annual Review of Pharmacology, 1995, 35: 93 – 127.
- [7] Apted F I C. Present status of chemotherapy and chemoprophylaxis of human trypanosomiasis in the Eastern Hemisphere [J]. Pharmacology & Therapeutics, 1980, 11: 391 – 413.
- [8] Zuccotto F, Martin A C R, Laskowski R A, et al. Dihydrofolate reductase; a potential drug target in trypanosomes and leishmania [J]. Journal of Computer-Aided Molecular Design, 1998, 12 (3): 241 – 257
- [9] Oberholzer M, Marti G, Baresic M, et al. The *Trypanosoma brucei* cAMP phosphodiesterases TbrPDEB1 and TbrPDEB2: flagellar enzymes that are essential for parasite virulence [J]. FASEB Journal, 2007, 21 (3): 720 – 731.
- [10] Ehmke V, Heindl C, Rottmann M, et al. Potent and selective inhibition of cysteine proteases from *Plasmodium falciparum* and *Trypanosoma brucei* [J]. Chemmedchem, 2011, 6 (2): 273 – 278.
- [11] Ding D Z, Meng Q Q, Gao G W, et al. Design, synthesis, and structure-activity relationship of *Trypanosoma brucei* Leucyl-tRNA Synthetase Inhibitors as Antitrypanosomal Agents [J]. Journal of Medicinal Chemistry, 2015, 54: 1276 – 1287.
- [12] Deterding A, Dungey F A, Thompson K A, et al. Antitrypanosomal activities of DNA topoisomerase inhibitors [J]. Acta Tropica, 2005, 93: 311 – 316.
- [13] Wenzler T, Boykin D W, Ismail M A, et al. New treatment option for second-stage african sleeping sickness; in vitro and in vivo efficacy of aza analogs of DB289 [J]. Antimicrobial Agents & Chemotherapy, 2009, 53 (10): 4185 – 4192
- [14] Jacobs R T, Plattner J J, Nare B, et al. Benzoxaboroles; a new class of potential drugs for human African trypanosomiasis [J]. Future Medicinal Chemistry, 2011, 3 (10): 1259 – 1278.
- [15] Verma R P, Hansch C. Camptothecins: a SAR/QSAR study [J]. Chemical Reviews, 2009, 109 (1): 213 – 235
- [16] Kubinyi H. 3D QSAR in drug design: theory, methods and applications [J]. 1993, 3: 413 – 418
- [17] Patrick D A, Wenzler T, Yang S, et al. Synthesis of novel amide and urea derivatives of thiazol-2-ethylamines and their activity against *Trypanosoma brucei* rhodesiense [J]. Bioorganic & medicinal chemistry, 2016, 24: 2451 – 2465.
- [18] 易忠胜. 变量筛选新方法及其在环境内分泌干扰效应 QSAR 中的应用 [D]. 南京: 南京大学, 2010.
- [19] Yi Z S, Zhang A Q. A QSAR study of environmental estrogens based on a novel variable selection method [J]. Molecules, 2012, 17 (5): 6126 – 6145.
- [20] Baskin I I, Kireeva N, Varnek A. The one-class classification approach to data description and to models applicability domain [J]. Molecular Informatics, 2010, 29: 581 – 587.
- [21] Netzeva T I, Worth A, Aldenberg T, et al. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52 [J]. Alternatives to Laboratory Animals Atla, 2005, 33 (2): 155 – 173.
- [22] Wold S, Sjöström M, Eriksson L. PLS-regression; a basic tool of chemometrics [J]. Chemometrics & Intelligent Laboratory Systems, 2008, 58 (2): 109 – 130.

QSAR study on novel HAT inhibitors of amide and urea derivatives of thiazol-2-ethylamines

DUAN Jia-xi, ZHAO Sai, YI Zhong-sheng, NIE Jin-fang

(a. College of Chemistry and Bioengineering; b. Guangxi College and University Key Laboratory of Food Safety and Detection; c. Collaborative Innovation Center for Water Pollution Control and Water Safety in Karst Area, Guilin University of Technology, Guilin 541004, China)

Abstract: 2D-QSAR models are developed to predict the activity against *trypanosoma brucei* rhodesiense of amide and urea derivatives of thiazol-2-ethylamines. The VSMVI (variable selection method based on variable interaction) technique is used to select optimal subset from large-size molecular descriptors. Multiple linear regression (MLR) method is employed to construct the QSAR models. The correlation coefficient of leave-one-out cross validation ($q^2 = 0.8574$) and fitting correlation coefficient of models ($r^2 = 0.8892$) show that the model is robust and has good fitting and prediction for selecting bioactivities. The descriptor of the 2D-QSAR model reflects, to some extent, that the two-dimensional structure and hydrophobicity of the molecules play an important role on the inhibitory activity. Simultaneously, the models of three-dimensional quantitative structure-activity relationship (3D-QSAR) based on CoMFA and CoMSIA show a significant correlation and strong predictive ability (CoMFA: $r^2 = 0.924$, $q^2 = 0.516$. CoMSIA: $r^2 = 0.944$, $q^2 = 0.531$). The highest contribution rate of CoMSIA hydrophobic field indicates that the hydrophobic interaction of molecules has an important effect on the inhibitory activity.

Key words: amide and urea derivatives of thiazol-2-ethylamines; VSMVI; HAT; 2D-QSAR; CoMFA; CoMSIA