

紫杉醇类似物抗癌活性与分子结构的定量构效关系

陈 艳, 堵锡华

徐州工程学院化学化工学院, 江苏 徐州 221008

摘要: 目的 为了寻找新一代紫杉醇类抗癌药物, 研究紫杉醇类似物分子结构和抗癌活性 (PID_{act}) 之间的定量构效关系。方法 基于电性状态拓扑指数和电性距离矢量, 应用最佳子集回归的方法建立了 PID_{act} 值和紫杉醇类似物分子结构的定量结构-活性相关 (QSAR) 模型, 并对模型进行了交互检验和外部检验, 用 43 个紫杉醇类似物训练集样本构建的 QSAR 模型预测了外部 10 个检验集的 PID_{act} 值。结果 所建立模型具有较高的估计相关系数及 LOO (leave-one-out) 检验相关系数。结论 模型具有良好估计能力与稳定性, 训练集模型具有良好预测能力。

关键词: 紫杉醇类似物; 构效关系; 抗癌活性; 电性状态拓扑指数; 电性距离矢量

中图分类号: R286.91 文献标志码: A 文章编号: 0253-2670(2011)02-0318-06

Relationship between anticancer activity and molecular structure of paclitaxel analogues

CHEN Yan, DU Xi-hua

School of Chemistry and Chemical Engineering, Xuzhou Institute of Technology, Xuzhou 221008, China

Abstract: Objective To develop a new generation of anticancer drugs of paclitaxel, the quantitative structure-anticancer activity (PID_{act}) relationship (QSAR) of paclitaxel analogues was studied. **Methods** Based on the electrotopological state index and electronegativity distance vector, the QSAR model was developed by Leaps-Bounds regression (best subset, LBR), and was validated using cross and external-validation. PID_{act} values of ten external validation were predicted by the QSAR model built from the training set with 43 paclitaxel analogues. **Results** The model had higher calibrated correlation coefficient and LOO (leave-one-out) validated correlation coefficient. **Conclusion** The model has estimated ability and good stability, and the training set model has good predictive ability.

Key words: paclitaxel analogues; structure-effect relationship; anticancer activity; electrotopological state index; electronegativity distance vector

紫杉醇是从红豆杉属植物的树皮中提取出来的抗癌药物, 其新颖的分子结构、独特的抗癌作用机制和良好的抗癌活性引起了化学和药学工作者的关注, 成为当今抗癌药物研究的热点^[1-2]。但紫杉醇在临床应用中因其来源有限^[3], 溶解度差而受到限制, 从而激发了科学家们开展紫杉醇化合物合成、半合成和结构修饰的广泛兴趣。通过合成、半合成的方法^[4-7], 解决了资源有限的问题; 通过结构修饰即对紫杉醇各部位官能团或骨架进行改造修饰, 将修饰后的各类似物进行构效关系研究, 以期获得溶解度好, 抗癌活性高的紫杉醇类似物^[8]。

所谓构效关系研究就是建立药物分子结构与实验可测的药效学数据相关联的定量构效关系 (QSAR) 模型, 并利用 QSAR 预测所设计化合物的药效学数据, 从而为药物分子结构修饰与改造提供理论依据, 是当前药物设计的重要方法之一^[9-11]。石丙兴等^[12]利用比较分子力场分析 (CoMFA) 方法对 98 个紫杉醇衍生物进行了三维定量构效关系 (3D-QSAR) 分析, 许旋等^[13]应用分子力学、量子化学、BP 神经网络模式识别等方法研究 10 位包含 -OH 和 -OAc 的 43 个紫杉醇类似物电子结构与药效的定量关系, 均取得了较好的结果。本研究在前

收稿日期: 2010-04-17

基金项目: 江苏省自然科学基金资助项目 (09KJD150012); 徐州市科技计划研究项目 (XM08C015, XX10A060)

作者简介: 陈 艳 (1968—), 女, 江苏太仓人, 硕士, 教授, 从事药物和有机污染物的构效关系研究。

Tel: (0516)85608307 13852106096 E-mail: chenyan681110@126.com

期研究的基础上^[14-15], 将片断拓扑指数与紫杉醇类似物抗癌活性进行相关分析, 建立了较为满意的预测模型。

1 数据来源与计算方法

1.1 数据来源

从文献[11-12]中选取 53 个 10 位包含-OH 和-OAc 的紫杉醇类似物(结构母核见图 1 和 2), 紫杉醇及其衍生物的活性以 PID_{act} ($PID_{act} = -\lg [ID_{50}(A)/ID_{50}(T)]_{act}$) 表示, 其中, $ID_{50}(T)$ 和 $ID_{50}(A)$ 分别为使微管蛋白解聚速度下降一半时所需紫杉醇和其类似物的浓度。 PID_{act} 数值越大, 抗癌活性越高。

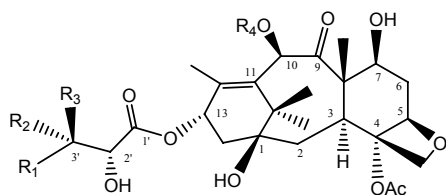


图 1 紫杉醇类似物 1~51 的结构母核

Fig. 1 Main structure of paclitaxel analogues 1—51

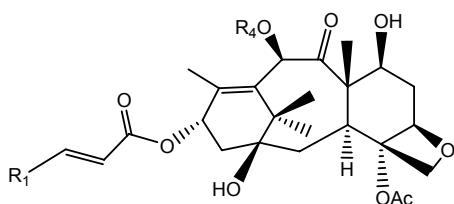


图 2 紫杉醇类似物 52 和 53 的结构母核

Fig. 2 Main structure of paclitaxel analogues 52 and 53

1.2 拓扑指数的选取

拓扑参数直接由拓扑图衍生, 此类参数可细分为两类, 即结构片段特征和拓扑指数^[16]。紫杉醇分子是个结构复杂的大分子, 有些拓扑指数的计算由于其结构的庞大而受到限制, 本研究根据所要研究的基团的位置, 截取特征结构片段, 计算其拓扑指数。对于紫杉醇衍生物可以提取的特征结构片段见图 3。

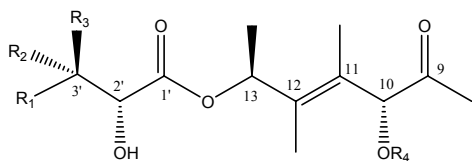


图 3 紫杉醇类似物的特征结构片段

Fig. 3 Characteristic fragment of paclitaxel analogues

采用两类拓扑指数来表征紫杉醇分子的结构, 即电性拓扑状态指数和电性距离矢量。首先应用

ChemDraw Ultra 9.0 软件分别构建所选 53 个紫杉醇类似物的特征结构片段, 然后在 Matlab 环境下, 调用上述分子结构, 应用文献[17-18]的方法编制程序, 计算得到 19 种电性拓扑状态指数(53 种紫杉醇片段分子涉及的原子结构类型, 以 e_j 表示)及 54 种电性距离矢量(以 m_j 表示), 共有 73 个描述子。根据统计学原理, 变量值的个数少于 5% 的自变量, 其对因变量的贡献可以忽略不计。因此首先对自变量集中自变量值少于 3 ($53 \times 5\%$) 的自变量予以剔除, 剩余 44 个描述子来表征分子的结构。

1.3 定量结构- PID_{act} 相关分析方法

将每种化合物的 44 种描述子作为自变量 (X), 相应的 PID_{act} 值为因变量 (Y) 构建数据集, 应用最佳子集回归 (Leaps-Bounds regression) 选择最佳变量组合, 建立相应 QSAR 模型。以逐一剔除法 (leave-one-out, LOO) 交互检验判断其稳定性与预测能力, 同时引入 Kubinyi 函数 (Kubinyi function, FIT)^[19-20], FIT 值越大, 所建的模型越稳定, 预测能力越高。其计算公式如下:

$$FIT = \frac{R^2(y-b-1)}{(y+b^2)(1-R^2)}$$

式中 y 为化合物数, b 为变量数, R^2 为判定系数

2 结果与讨论

2.1 拓扑指数与紫杉醇类似物活性的 QSAR 模型

将所选的 53 个紫杉醇类似物的 PID_{act} 实验值及其两种拓扑指数 44 个描述子引入 SPSS 统计软件, 在 95% 置信区间内, 就上述化合物数据集进行多元统计分析, 经最佳变量子集回归构建 QSAR 模型, 见表 1。对所得的不同参数组合模型进行 LOO 交叉验证, 即每次抽取 53 个化合物作为样本, 剔除 1 个化合物作为被预测对象, 再由预测值与实验值进行线性回归, 即得交叉验证系数 (Q 及 Q^2 , 见表 1)。可见, 随着变量数的增多, R 和 R^2 在逐渐增大, 从模型 1 到模型 8, Q 和 Q^2 都在不断增加, 但从模型 6 开始, 涨幅开始变缓, 而且模型 6 的 FIT 值最大, 所以本研究选用该六元数学模型为最佳 QSAR 模型:

$$PID_{act} = -2.071 + 0.033e_{13} + 0.147e_{14} - 0.023e_{16} + 0.471e_{18} + 0.041m_{18} - 2.740m_{82}$$

$$n = 53, R = 0.902, R^2 = 0.813, Q = 0.869, Q^2 = 0.755, S = 0.272, F = 33.339$$

式中 n 、 R^2 、 R 、 Q^2 、 Q 、 S 、 F 依次为样本数、判定系数、相关系数、交叉验证系数、估计标准误差及 Fisher 检验值。根据 Q^2 和 FIT 值, 可以认为该模

型具有总体稳健性与良好的预测能力。按此方程给出的计算值 (PID_{cal1}) 列于表 2, 与实验值基本吻合。

为了直观地反映方程的线性显著性, 以 PID_{act} 为横坐标, PID_{cal} 为纵坐标, 得图 4。

表 1 PID_{act} 与 e_j 及 m_j 的最佳变量子集回归结果

Table 1 Results of PID_{act} and e_j, m_j with Leaps-Bounds regression

序号	R	R^2	Q	Q^2	S	F	FIT	变量
1	0.627	0.393	0.578	0.335	0.466	33.003	0.547	e_{18}
2	0.725	0.526	0.673	0.453	0.416	27.692	0.973	e_{18}, e_{16}
3	0.759	0.576	0.710	0.504	0.398	22.171	1.074	e_{18}, e_{16}, m_{18}
4	0.790	0.624	0.743	0.552	0.378	19.928	1.154	$e_{18}, e_{16}, m_{18}, e_{13}$
5	0.827	0.683	0.785	0.616	0.351	20.300	1.298	$e_{18}, e_{16}, m_{18}, e_{13}, e_{14}$
6	0.902	0.813	0.869	0.755	0.272	33.339	2.247	$e_{18}, e_{16}, m_{18}, e_{13}, e_{14}, m_{82}$
7	0.912	0.832	0.880	0.774	0.261	31.914	2.185	$e_{18}, e_{16}, m_{18}, e_{13}, e_{14}, m_{82}, m_{36}$
8	0.918	0.843	0.881	0.776	0.255	29.674	2.019	$e_{18}, e_{16}, m_{18}, e_{13}, e_{14}, m_{82}, m_{36}, m_{77}$
9	0.922	0.850	0.881	0.776	0.253	27.010	1.818	$e_{18}, e_{16}, m_{18}, e_{13}, e_{14}, m_{82}, m_{36}, m_{77}, m_{15}$

表 2 紫杉醇类似物的结构和活性

Table 2 Structures and activities of paclitaxel analogues

类似物	R ₁	R ₂	R ₃	R ₄	PID _{act}	PID _{cal1}	PID _{cal2}
1	Ph	NHBz	H	Ac	0.00	-0.33	
2	Ph	H	H	H	-0.65	-1.17	
3	Ph	NHBz	H	H	-0.11	-0.07	
4	Ph	NHCO ₂ - <i>t</i> -Bu	H	H	0.30	0.25	
5	Ph	NHCO ₂ - <i>t</i> -Bu	H	Ac	0.30	-0.09	
6*	4-OHPh	NHBz	H	Ac	0.10	-0.04	0.00
7	3,4-Cl ₂ -phenyl	NHBz	H	Ac	-0.85	-0.37	
8	Ph	NHBz	H	Ac	-1.48	-1.48	
9	Ph	NHBz	H	H	-1.48	-1.08	
10*	Ph	NHCO ₂ - <i>t</i> -Bu	H	H	-1.00	-1.18	-1.34
11	Ph	NHCO ₂ - <i>n</i> -Bu	H	Ac	0.10	-0.10	
12	Ph	NHCOCOCH ₃	H	Ac	-0.54	-0.81	
13	Ph	NHCO- <i>t</i> -Bu	H	Ac	-0.42	-0.36	
14*	Ph	NHCOCH ₂ - <i>t</i> -Bu	H	Ac	0.16	-0.34	-0.37
15	<i>t</i> -Bu	NHBu	H	H	-0.25	-0.10	
16	<i>t</i> -Bu	NHCO ₂ - <i>t</i> -Bu	H	H	0.42	0.20	
17	<i>i</i> -PrCH ₂	NHCO ₂ - <i>t</i> -Bu	H	H	0.11	0.20	
18	<i>t</i> -BuCH ₂	NHCO ₂ - <i>t</i> -Bu	H	H	-0.16	0.22	
19	PhCH=CH	NHCO ₂ - <i>t</i> -Bu	H	H	-0.16	0.24	
20*	(CH ₃) ₂ C=CH	NHCO ₂ - <i>t</i> -Bu	H	H	0.19	0.20	0.17
21	<i>t</i> -Bu	NHCO- <i>c</i> -C ₃ H ₆	H	H	0.02	-0.10	
22	<i>t</i> -Bu	NHCO- <i>c</i> -C ₄ H ₈	H	H	0.17	-0.08	
23	<i>t</i> -Bu	NHCO- <i>c</i> -C ₅ H ₁₀	H	H	-0.18	-0.06	
24	Ph	TsNH	H	H	-0.70	-0.52	
25*	Ph	2-sulfoBzNH	H	H	-0.74	-0.45	-0.24

续表 2

类似物	R ₁	R ₂	R ₃	R ₄	PID _{act}	PID _{cal1}	PID _{cal2}
26	4-ClPh	NHBz	H	Ac	-0.38	-0.34	
27	3-ClPh	NHBz	H	Ac	-0.64	-0.35	
28	4-FPh	NHBz	H	Ac	-0.04	-0.39	
29*	4-MePh	NHBz	H	Ac	-0.38	-0.32	-0.34
30	4-MeOPh	NHBz	H	Ac	0.10	0.23	
31	2-Furyl	NHBz	H	Ac	0.05	-0.25	
32*	3-Furyl	NHBz	H	Ac	0.05	-0.05	-0.08
33	2-Pyridyl	NHBz	H	Ac	0.16	0.15	
34	3-Pyridyl	NHBz	H	Ac	0.30	0.31	
35	4-Pyridyl	NHBz	H	Ac	0.40	0.30	
36	Ph	4-ClBzNH	H	Ac	-0.38	-0.34	
37	Ph	2,4-Cl ₂ BzNH	H	Ac	-0.32	-0.38	
38*	Ph	3-ClBzNH	H	Ac	-0.30	-0.35	-0.37
39	Ph	4-MeBzNH	H	Ac	-0.20	-0.32	
40	Ph	2-MeBzNH	H	Ac	-0.28	-0.32	
41	Ph	4-FBzNH	H	Ac	-0.08	-0.39	
42	Ph	4-NO ₂ -BzNH	H	Ac	-0.30	-0.34	
43	Ph	2-FuroyNH	H	Ac	0.10	0.48	
44*	Ph	NH-CO-4-MeOPh	H	Ac	0.30	0.29	0.28
45	Ph	GluNH	H	H	0.00	-0.14	
46	Ph	NH ₂	H	Ac	-1.64	-1.44	
47	4-NMe ₂ Ph	NHBz	H	Ac	-0.66	-0.31	
48	Ph	4-CF ₃ BzNH	H	Ac	-0.78	-0.45	
49	Ph	OH	H	H	-0.48	-0.85	
50	CH ₃	OH	H	Ac	-1.78	-1.18	
51	Ph	H	H	H	-1.23	-1.17	
52*	Ph	—	—	H	-1.36	-1.49	-1.59
53	CH ₃	—	—	Ac	-2.00	-1.78	

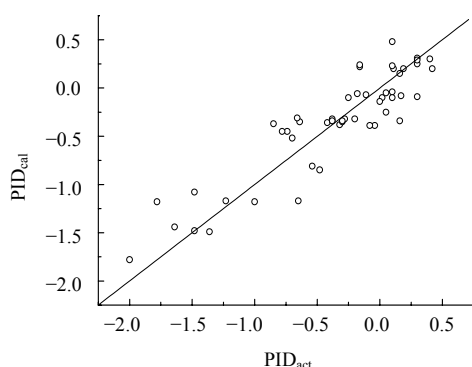


图 4 紫杉醇类似物的 PID_{act} 实验值和计算值的相关性
 Fig. 4 Correlation between PID_{act} observed values and calculated values of paclitaxel analogues

2.2 模型稳健性的再检验

2.2.1 模型中变量自相关性的检验 评价模型的稳定性及是否存在自相关性，可用变异膨胀因子 (variance inflation factors, *VIF*) [21] 予以判定。如 *VIF* = 1，表明各自变量间完全不相关；当 *VIF* < 5 时，说明变量间没有明显的自相关性，所建模型是稳定的；当 *VIF* > 5 时，说明变量间存在明显的共线性，所建模型不能用于估算与预测。

$$VIF = 1 / (1 - R^2)$$

式中，*R*² 为自变量 *X* 中某一变量与余下变量的判定系数。模型 6 中各自变量的自相关矩阵 (*VIF* 值和自相关系数 *R*²) 见表 3。可知，各自变量的 *VIF* 值均大于 1 且小于 5，说明所得模型是稳健的。

表 3 模型 6 中各描述符的自相关矩阵与 VIF 值
Table 3 Self-correlation matrix and VIF values described in model 6

变量	e_{13}	e_{14}	e_{16}	e_{18}	m_{18}	m_{82}	R^2	VIF
e_{13}	1						0.491	1.965
e_{14}	0.632	1					0.736	3.788
e_{16}	0.358	0.464	1				0.545	2.198
e_{18}	0.091	0.349	0.658	1			0.500	2.000
m_{18}	0.209	0.117	0.272	0.249	1		0.129	1.148
m_{82}	0.607	0.827	0.448	0.266	0.064	1	0.709	3.436

2.2.2 模型的外部检验 将 53 个紫杉醇类似物的活性数据分为两部分：即训练集和测试集（表 2 中用*号标记）。从全部数据中随机选取 43 个有机物的活性数据作为训练集进行 QSAR 分析，建模条件与全部数据集的最佳 QSAR 模型产生条件相同，多元线性回归分析结果： $PID_{act} = -2.411 + 0.040e_{13} + 0.153e_{14} - 0.017e_{16} + 0.439e_{18} + 0.043m_{18} - 3.023m_{82}$ ， $n=43$ ， $R=0.901$ ， $R^2=0.813$ ， $F=26.016$ ， $S=0.282$ 。

可见，训练集所建立的 QSAR 模型的拟合质量与全部数据的 QSAR 模型很接近，说明模型很稳定。用基于训练集数据所建立的 QSAR 模型来预测测试集中紫杉醇类似物的活性，预测结果（ PID_{cal2} ）列于表 2，表明基于训练集所建立的 QSAR 模型很好地预测了测试集中紫杉醇类似物的活性。

紫杉醇的抗癌作用是通过与微管受体之间的相互作用，进而促进微管蛋白装配成稳定的微管，继而抑制癌细胞内的微管解聚^[22]。这种作用与紫杉醇的化学结构相关，包括紫杉醇的空间结构和电子结构。电性拓扑状态指数是表征分子空间拓扑信息与电性性质的描述子，其不仅可编码原子间电性相互作用，还可编码原子间的拓扑环境。同时考虑到分子的生物活性应是该分子中各连接原子之间相互作用的函数，同时引入电性距离矢量，它是基于分子中的各原子固有属性、原子间的连接关系、原子间的距离等因素对分子结构进行数值化表征的一组参数，所以两类拓扑指数联合共同揭示了影响紫杉醇抗癌活性的因素。进入模型的 e_{13} 、 e_{14} 、 e_{16} 、 e_{18} 、 m_{18} 、 m_{82} ，他们对应的子结构片段为 $-OH$ 、 $-O-$ 、 $C=O$ 、 $-NH-$ ，说明在所选的结构片段中，影响紫杉醇类似物抗癌活性的因素是这 4 种子结构片段。

3 结论

基于电性状态指数和电性距离矢量描述子对 53 种紫杉醇类似物的分子结构与抗癌活性进行研

究，通过最佳子集回归选取的 6 个描述子与 PID_{act} 具有良好的相关性，所得的 QSAR 模型不仅对内部样本具有良好的估计能力，同时对外部样本具有良好的预测能力，可用于紫杉醇衍生物抗癌活性的估算和预测。

参考文献

- [1] 赵 锐, 赵玮玮. 抗癌植物药紫杉醇研究进展与动态 [J]. 中草药, 2009, 40(7): 1172-附 2.
- [2] 黄 歆, 杨尚金. 与微管作用的抗肿瘤天然产物的研究进展 [J]. 现代药物与临床, 2010, 25(1): 1-5.
- [3] Gragg G M, Schepartz S A, Suffness M, et al. The taxol supply crisis [J]. *J Nat Prod*, 1993, 56(10): 1657-1668.
- [4] Yin D L, Liu R W, Wang D H, et al. Synthesis of analogues of paclitaxel with 14-side chain from sinenxan A [J]. *J Chin Pharm Sci*, 1999, 8: 191-200.
- [5] Chen S H, Huang S, Kant J, et al. Synthesis of 7-deoxy- and 7,10-dideoxytaxol via radical intermediates [J]. *J Org Chem*, 1993, 58: 5028-5029.
- [6] 孔建强, 王 伟, 朱 平, 等. 紫杉醇生物合成的研究进展 [J]. 药学学报, 2007, 42(4): 358-365.
- [7] 刘万宏, 姚 波, 祝顺琴, 等. 紫杉醇前体生物合成途径及生物技术研究进展 [J]. 中草药, 2009, 40(8): 1327-1331.
- [8] 邓成华, 梅兴国, 余广鳌. 紫杉醇的结构修饰及构效关系 [J]. 天然产物研究与开发, 1998, 11(2): 72-81.
- [9] 袁传能, 许 旋, 徐志光. C(4)取代紫杉醇类似物的定量构效关系研究 [J]. 分子科学学报, 2008, 24(6): 417-421.
- [10] 李 燕, 王永华, 房 华, 等. 紫杉醇类衍生物的定量结构-抗癌活性关系研究进展 [J]. 中国药理学通报, 2008, 24(3): 288-293.
- [11] 田丹碧, 章 靖, 祝艳琳, 等. 紫杉醇构效关系及类似物生物活性的研究进展 [J]. 南京师范大学学报: 工程技术版, 2007, 7(4): 48-54.
- [12] 石炳兴, 梁世乐, 元英进, 等. 紫杉醇衍生物的三维定量构效关系研究 [J]. 高等学校化学学报, 2000, 21(3): 401-406.

- [13] 许旋, 徐志广, 罗一帆, 等. 紫杉醇类似物抗癌活性构效关系的神经网络模式识别研究 [J]. 华南师范大学学报: 自然科学版, 2005, 4: 73-80.
- [14] 陈艳, 堵锡华. 基于分子电性距离矢量预测 PBTAs 的热力学性质 [J]. 石油化工高等学校学报, 2009, 22(4): 29-33.
- [15] 堵锡华, 陈艳, 高淑云. 10-OAc 抗癌活性与拓扑参数的相关性 [J]. 华中科技大学学报: 自然科学版, 2009, 37(9): 130-132.
- [16] 许禄, 邵学广. 化学计量学方法 [M]. 北京: 科学出版社, 2004.
- [17] 胡黔楠, 梁逸曾, 王亚丽, 等. 直观队列命名法的基本原理及其在矩阵与拓扑指数计算中的应用 [J]. 计算机与应用化学, 2003, 20(4): 386-390.
- [18] 张婷, 梁逸曾, 赵晨曦, 等. 基于分子结构预测气相色谱程序升温保留指数 [J]. 分析化学, 2006, 34(11): 1607-1610.
- [19] Saiz-Urra L, Gonzalez M P, Teijeira M. 2D-auto-correlation descriptors for predicting cytotoxicity of naphthoquinone ester derivatives against oral human epidermoid carcinoma [J]. *Bioorg Med Chem*, 2007, 15: 3565-3571.
- [20] Saiz-Urra L, Gonzalez M P, Teijeira M. QSAR studies about cytotoxicity of benzophenazines with dual inhibition toward both topoisomerases I and II: 3D-MoRSE descriptors and statistical considerations about variable selection [J]. *Bioorg Med Chem*, 2006, 14: 7347-7358.
- [21] 李吉来, 杭焯超, 耿彩云, 等. 苯砜基羧酸酯类急性毒性的 QSAR 研究 [J]. 高等学校化学学报, 2007, 28: 117-120.
- [22] 崔萍, 霍长虹, 李力更, 等. 作用于微管的天然产物 [J]. 中草药, 2010, 41(1): 139-147.