第47卷第1期 大连理工大学学报 Vol. 47, No. 1
2007年1月 Journal of Dalian University of Technology Jan. 2007

化学化工、动力工程

文章编号: 1000-8608(2007)01-0015-06

# 黄酮衍生物作为P糖蛋白抑制剂的构效关系研究

李 燕[1], 王永华[2], 杨 凌*[3], 张述伟[1],
蒋 达[1], 刘长厚[1], 杨胜利[3]

(1.大连理工大学 化工学院, 辽宁 大连 116012;
2.大连水产学院 生命科学与技术学院, 辽宁 大连 116023;
3.中科院大连化学物理研究所 药用资源开发研究组, 辽宁 大连 116023)

**摘要:** 黄酮类化合物对P糖蛋白外泵作用的抑制效果越来越引起人们的关注. 使用偏最小二乘法(PLS)和MolconnZ分子参数计算程序, 首次成功地构建了特异性地结合到P糖蛋白NBD2位点的黄酮类抑制剂的二维定量构效关系(2D-QSAR)模型. 所得到的模型(尤其是模型D)表现出良好的可靠性和预测性, 较好地关联了该类化合物结构特点和其对P糖蛋白抑制活性. 这些模型对进一步合成和开发黄酮类P糖蛋白抑制剂有指导作用.

**关键词:** 黄酮类; P糖蛋白; 二维定量构效关系; 偏最小二乘法
中图分类号: O621.12 文献标识码: A

## 0 引 言

癌细胞对多种结构和理化性质均不同的药物的抗药性, 即多药耐药性(multidrug resistance, 简称MDR)始终是改善癌症病人化疗效果的一大阻碍[1]. 导致细胞产生耐药性的一大原因是P糖蛋白对药物的外排作用. P糖蛋白(P-glycoprotein, 简称P-gp)是由人体mdr1基因编码的一类跨膜蛋白, 它能在ATP酶水解释放出的能量驱动下将异源性物质(包括药物分子)"泵"出细胞外[2]. 因此, 找到合适的抑制剂来阻碍P-gp对药物的泵出作用, 有助于克服癌细胞的多药耐药性并提高药效. 迄今为止尽管多种P-gp抑制剂已经被发掘出来, 但由于其在体内实验中会对人体产生较大毒性而失去进一步开发的价值[3,4].

黄酮是一类多酚类化合物, 它作为人类饮食的重要摄入成分广泛分布在植物的根、茎、叶、花及水果、茶叶中. 黄酮类化合物具有一系列对人体健康至关重要的理化作用, 如抗氧化、抗变应性、抗菌和自由基捕捉等性质[5]. 不仅如此, 许多黄酮类化合物如铁精、桑色素、根皮素、异黄酮和黄酮醇等已被证明对P糖蛋白的外泵作用有较强抑制作用, 因而其抑制癌细胞生长的效果显著[6~9]. 所以目前黄酮作为一类极有前景的P-gp抑制剂受到人们广泛关注.

由于迄今人们尚无法得到P-gp的高清晰晶体结构, 计算机技术就成为对P糖蛋白抑制剂设计和筛选的有利辅助手段. 本文目的是采用偏最小二乘法(PLS)和MolconnZ计算相结合的方式, 对一系列共46个黄酮衍生物的二维结构和其对P-gp抑制活性的关系进行研究, 建立具有良好预测性能的定量构效关系(quantitative structure activity relationship, 简称QSAR)数学模型以辅助设计和开发新型的P-gp黄酮类抑制剂.

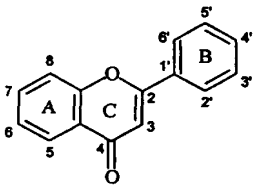
## 1 数据和方法

### 1.1 数据整理

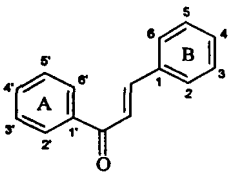
本文的数据集由46个来自文献[10]的黄酮类衍生物组成. 该类化合物结构简单, 一般由两个共轭六元环(环A和B)和连接这两个环的部分(碳链或另一个六元环C)组成, 其结构如图1所示. 这批化合物对P-gp的抑制活性采用它们与小鼠P-gp蛋白NBD2位点的解离常数$K_D$[10]的负对数, 即$-\log K_D$表示. 因此, 本文所研究的这批化合物不仅对P-gp有抑制效果, 而且是特异性地结合到其一特定位点NBD2的抑制剂.

收稿日期: 2005-06-03; 修回日期: 2006-12-11.
基金项目: 大连理工大学-中科院大连化学物理研究所合作科研探索基金资助项目; 大连理工大学青年教师培养基金资助项目; 中科院973(2003CCA03400)、863(2003AA223061)及大连化物所创新基金资助.
作者简介: 李 燕(1976-), 女, 博士, 讲师; 杨 凌*(1961-), 男, 博士, 教授, 博士生导师; 张述伟(1963-), 男, 教授, 博士生导师; 刘长厚(1940-), 男, 教授, 博士生导师; 杨胜利(1941-), 男, 教授, 中国工程院院士.

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net



结构 I



结构 II

| No | 结构 | $K_D$/(μmol·L$^{-1}$) | 3 | 4 | 5 | 6 | 7 | 8 | 2′ | 3′ | 4′ | 5′ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | I | 34.9 | | | | | —OH | | | | | |
| 2 | I | 8.9 | | | —O | | —OH | | | | | |
| 3 | I | 3.1 | | | —O | —$CH_3$ | —OH | | | | | |
| 4 | I | 6.3 | | | —O | | —$OCH_3$ | | | | | |
| 5 | I | 10.1 | | | —O | | —OH | | | | —OH | |
| 6 | I | 6.3 | | | —O | | —OH | | | —F | —F | |
| 7 | I | 2.2 | | | —O | | —OH | | | | —I | |
| 8 | I | 1.3 | | | —O | | —$OCH(CH_3)_2$ | | | | | |
| 9 | I | 0.34 | | | —O | a | —OH | | | | | |
| 10 | I | 0.99 | | | —O | | —OH | a | | | | |
| 11 | I | 0.30 | | | —O | d | —OH | | | | | |
| 12 | I | 0.20 | | | —O | | —OH | b | | | | |
| 13 | I | 0.28 | | | —O | | —OH | d | | | | |
| 14 | I | 0.045 | | | —O | c | —OH | | | | | |
| 15 | I | 0.025 | | | —O | | —OH | c | | | | |
| 16 | I | 0.7 | | | —O | | —OH | b | | | —OH | |
| 17 | I | 10.1 | —OH | | | | | | | | | |
| 18 | I | 5.3 | —OH | | —O | | —OH | | | | | |
| 19 | I | 6.7 | —OH | | —O | | —OH | | | | —OH | |
| 20 | I | 4.5 | —OH | | —O | | —OH | | | | —$OCH_3$ | |
| 21 | I | 7.0 | —OH | | —OH | | —OH | | | | —OH | —OH |
| 22 | I | 0.20 | —OH | | —O | | —OH | b | | | —$OCH_3$ | |
| 23 | I | 0.45 | —OH | | —O | | —OH | b | | | | |
| 24 | I | 0.21 | —OH | | —O | d | —OH | | | | | |
| 25 | I | 0.22 | —OH | | —O | | —OH | d | | | | |
| 26 | I | 6.8 | —OH | | —O | | —OH | | | | —F | |
| 27 | I | 4.0 | —OH | | —O | | —OH | | —F | | —F | |
| 28 | I | 1.1 | —OH | | —O | | —OH | | | | —I | |
| 29 | I | 0.06 | —OH | | —O | | —OH | | | | —$C_8H_{17}$ | |
| 30 | I | 8.9 | —$CH_3$ | | —O | | —OH | | | | | |
| 31 | I | 0.15 | —$CH_3$ | | —O | | —$CH_3$ | b | | | | |
| 32 | I | 1.3 | | | —O | —$CH_3$ | —$CH_3$ | | | | | |
| 33 | I | 36.5 | | | —O | | —OH | | | | —OH | |
| 34 | I | 37.4 | —OH | | —O | | —OH | | | —O | —OH | |
| 35 | II | 4.6 | | | | | | | | | | |
| 36 | II | 4.8 | | —OH | | | | | | | | |
| 37 | II | 2.3 | | —$OCH_3$ | | | | | | | | |
| 38 | II | 3.6 | | —F | | | | | | | | |
| 39 | II | 1.3 | | —Cl | | | | | | | | |
| 40 | II | 0.57 | | —Br | | | | | | | | |
| 41 | II | 0.25 | | —I | | | | | | | | |
| 42 | II | 0.53 | | e | | | | | | | | |
| 43 | II | 0.53 | d | —OH | | | | | | | | |
| 44 | II | 0.44 | —OH | —OH | | | | | | | | d |
| 45 | II | 2.1 | | —$C_2H_5$ | | | | | | | | |
| 46 | II | 1.0 | | —$C_4H_9$ | | | | | | | | |

注: 1)除第 33 34分子的 C2— C3键为单键外,其他分子的 C2— C3键为双键;

2) a b c d e分别代表苯甲基、二甲基烯丙基、牦牛儿基、异戊二烯基和环己基

图 1 黄酮类衍生物的结构

Fig. 1 Structure of flavonoid derivatives

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

**1.2** 计算方法

对这类黄酮化合物的定量构效关系研究是在一台戴尔工作站, LINUX RedHat8.0操作系统下,借助 SYBYL6.92软件进行的. 所有分子的二维结构来自 MDL-ISIS数据库. 首先从 700 K到 200 K用 1 000 fs的时间采用模拟退火分子动力学方法对这些结构进行了优化;其次,每个化合物构象采用 Tripos标准力场进行了能量最小化来寻找最稳定构象;最后,找到的稳定构象又采用 Gasteiger-Huckel电荷法计算了所带电荷. 此外,所有分子的疏水参数 log*P* 是基于片断计算法得来.

在以上工作基础上,采用 SYBYL软件中的 MolconnZ模块对所研究化合物的分子结构参数进行了计算. 计算的最初结果是对每一个分子,都得到了从拓扑、电拓扑状态参数、连接性指数、氢键参数到分子形状参数共 250个结构参数. 这些描述符都是分子结构的重要参数,从本质上描绘了分子的拓扑、对称性和电子状态. 但由于其数量过大,不仅单机计算时间长,而且难以从众多参数中提炼出有限个对分子生物活性最至关重要的参数,本文依一定原则对其进行了精练. 随后,采用 PLS对精炼后的参数(包括 log*P*)和其对应的抑制活性进行了关联.

PLS是一种标准统计方法[11],目前已被成功运用在许多定量构效关系研究中. 在本文中,PLS方法是将因变量矩阵 ***Y***(抑制活性)与自变量矩阵 ***X***(结构参数)关联,采用包括 Leave-one-out(LOO)和 cross-validation(CV)两步骤的"去一"交叉验证法构建模型. PLS可以从众多的参数中组合并筛选出有限个最重要的成分,并计算出该成分对因变量 ***Y***的贡献率,因此该方法尤其适用于参数数目等于或大于化合物数目的情况. 当计算获得所设定的最小均方差或最大交叉验证系数 $q^2$ 时即获得最优组分数 ($N_{op}$). $q^2$ 的计算依据式(1)进行:

$$q^2 = 1 - \frac{\sum_Y (Y_p - Y_o)^2}{\sum_Y (Y_o - Y_m)} \qquad (1)$$

其中 $Y_p$、$Y_o$和 $Y_m$ 分别是目标函数 ***Y***的预测值、实际值和平均值.

## 2　结果与讨论

定量构效关系的研究方法众多,包括各种二维定量构效法(2D-QSAR,如 Hansch法)和目前逐渐普及的三维定量构效法(3D-QSAR,如 CoMFA CoMSIA等). 二维定量构效法不仅可以将分子的生物活性与其综合性的理化性质如疏水性、分子量等相关联,还可以将活性与分子更详细的结构特点如拓扑参数、电拓扑参数、氢键受体和氢键供体等许多参数相关联. 与三维方法相比,二维定量构效法有其独特优势,因此目前在生化领域研究中仍广泛使用. 首先,二维方法没有需要事先获知分子三维结构的要求,这对那些难以获得晶体或活性结构的分子极具优势. 其次,三维方法涉及到大规模构象搜索计算和带有一定主观活性构象的选择性,而二维方法则没有这些限制,它只需计算出所有化合物的二维分子参数就可将其与分子的生物活性相关联. 因此在许多无法或不适合用三维方法计算的情况下,二维方法却可以游刃有余地发挥作用.

在本研究中,经 MolconnZ计算后每个分子得到了 250个结构参数. 而后遵循以下原则对其进行精炼: (1) 对所有分子同一参数的参数值均为零或零变异的参数值消去; (2) 数据标准化. 经上述处理后,每个分子只剩下了 81个参数. 对这些参数进行 2D-QSAR研究. 46个黄酮衍生物随机分为两个数据集: 38个分子用于构建训练集,剩下 8个分子用于构建验证集. 得到最优模型 D的交叉验证系数 $q^2 = 0.800$,非交叉验证回归系数 $r^2 = 0.912$,估计标准差 $E_{ms} = 0.266$, *F*检验值为 32.306,证明了模型的可靠性. 当用验证集验证该模型时,其预测误差值 $E_{ps}$ 只有 0.062,证明了该模型具有良好的预测性能.

表 1列出了模型 D的所有参数及其系数.

图 2对模型 D中所有分子的预测值和实际值作了比较,图中▲代表训练样本化合物,■代表预测样本化合物. 该图显示出预测值和实际值良

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

好的吻合性.

表 1 2D-QSAR模型 D的各项参数及其系数

Tab. 1 Description and coefficients of variables in model D

| 参数 | 系数 | 含义 |
|---|---|---|
| 常数 (*Const*) | - 0.589 | 截断值 |
| $X_1$(log*P*) | 0.309 | 脂水分配系数计算值 |
| $X_2$(ka1) | 0.149 | 1修正的 Kappa参数 |
| $X_3$(ka2) | - 1.682 | 1修正的 Kappa参数 |
| $X_4$(phia) | 1.963 | 分子挠性参数 |
| $X_5$(Qv) | - 0.303 | 综合极性参数 |
| $X_6$(nw HBa) | 0.163 | 弱氢键受体数 |
| $X_7$(Hmax) | - 0.345 | 最大氢键 *E*状态数 |
| $X_8$(Gmax) | - 0.199 | 最大 *E*状态数 |
| $X_9$(Gmin) | - 0.190 | 最小 *E*状态数 |

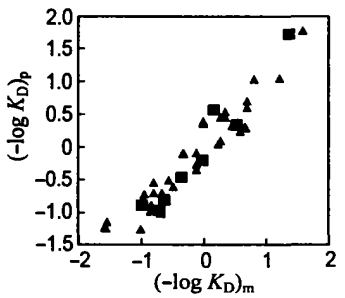



图 2 2D-QSAR模型 D实际值和预测值的比较

Fig. 2 Scatter graph of actual versus predicted - log $K_D$ values in 2D-QSAR model D

实际上,在模拟中本文采用不同结构参数组合得到了一系列较好的定量构效关系模型,这些模型如表 2所示. 一方面,一般来说模型采用的结构参数或参数组合越少越难以得到良好的模拟结果;另一方面,采用的结构参数或参数组合数越多,所建立的模型固然往往具有良好的可靠性和预测性,但对模型也越难进行直观或化学上的解释,不利于进一步指导药物分子的合成. 表 2中模型 A的各项统计结果都较好,预测误差也很小($E_{ps}$ = 0.029),但采用了 18个参数,难以对其进行全面综合的解释. 而模型 G虽只采用 6个结构参数,但不能得到满意的统计结果,其 $r^2$ 也随之而减小,小于 0.900. 综合平衡这两个因素,本文认为模型 D是最好的模拟结果,适于指导黄酮类 P-gp抑制剂的进一步筛选和合成.

目前对某一类化合物作为 P糖蛋白抑制剂研究的 QSAR文章众多,不仅常常将来自不同实验室的分子实验结果收集在一起作为数据集建模,还常忽略 P糖蛋白有多个药物结合位点的事实. 本文所用数据不仅来源于一个实验室,而且仅针对结合到 P糖蛋白一个位点—— NBD2位点的黄酮分子进行研究,在避免了不同实验室数据可能产生"噪音"问题的同时,还排除了其他位点的干扰,真正对 NBD2位点可能的结合情况进行了研究,有助于指导化学家开发专门针对该位点的黄酮类 P糖蛋白抑制剂,这也是本文创新点之一.

表 2 由不同结构参数组合得到的 2D-QSAR模型的统计结果

Tab. 2 Statistical results of 2D-QSAR models with different descriptor groups

| QSAR模型 | LOO-CV NCV | | | NCV | | $E_{ps}$ | 参数个数[1)] |
|---|---|---|---|---|---|---|---|
| | $N_{op}$ | $q^2$ | $E_{ms}$ | $r^2$ | $F$ | | |
| A | 10 | 0.800 | 0.219 | 0.943 | 44.277 | 0.029 | 18 |
| B | 2 | 0.764 | 0.249 | 0.926 | 33.683 | 0.045 | 13 |
| C | 2 | 0.780 | 0.273 | 0.911 | 27.534 | 0.069 | 11 |
| D | 1 | 0.800 | 0.266 | 0.912 | 32.306 | 0.062 | 9 |
| E | 2 | 0.783 | 0.272 | 0.904 | 34.262 | 0.048 | 8 |
| F | 7 | 0.809 | 0.274 | 0.900 | 38.606 | 0.048 | 7 |
| G | 6 | 0.812 | 0.271 | 0.899 | 45.982 | 0.049 | 6 |

注: 1) 这些参数不包括模型中的常数值 (即截断值)

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

此外,最近 Wang等[12]报告了他们采用贝叶斯神经网络(BRNN)技术建立的黄酮分子对 P-gp抑制活性的定量构效关系模型,这也是目前仅有的对黄酮类化合物作为 P糖蛋白抑制剂的 QSAR的研究报道. BRNN技术新颖,用其建立的神经网络与其他网络相比具有更好的鲁棒性,甚至不需要验证集检验模型就具有很好的预测能力. 但 BRNN方法仍具有所有神经网络的共同缺点,即它们所建立的模型"不可视",属于"黑箱"模型. 无法知道这些分子的哪些结构或理化参数与活性相关及具体相关程度. 而本文所用的 2D-QSAR方法则首次成功地将各种重要分子结构特征与分子抑制活性相关联,所建模型可以清晰地指出与活性相关的具体参数及相关系数,"可视"性强,易于接受和理解.

以模型 D为例,显然,该模型中各项参数的正确诠释和理解对进一步开发黄酮类抑制剂至关重要. $X_1$是脂 -水分配系数的计算值,其大小与分子的疏水性息息相关,一般来说其值越大分子疏水性越强. $X_1$对模型 D的较大贡献率(0.309,表 1)反映出黄酮分子疏水性对其抑制活性的重要性. 这一方面可能与黄酮分子要与 P-gp具有强疏水性的内部结合位点 NBD2结合相关;另一方面黄酮分子还必须具备一定的亲脂性或疏水性,这样才能顺利跨越构成细胞膜的磷脂双分子层,进而与跨膜蛋白 P-gp强烈结合而抑制其外泵作用. $X_2$和 $X_3$是修正的 Kappa参数,是基于分子内单键、双键和三键片断个数计算得到,反映了分子形状对抑制活性的贡献. $X_4$则是限制分子流动性的结构参数. $X_5$捕获的则是反映分子整体极性的信息. $X_6$代表的是分子中弱氢键受体和供体的数目. $X_7$~ $X_9$都属于分子电拓扑参数,编码的是分子中每个原子所处的拓扑环境和它与分子中其他所有原子间的静电作用. 从表 1可见模型 D中每个参数的贡献率都不小于 0.05,说明这些参数都对黄酮类分子的抑制活性影响重大. 尽管这些参数多是根据一定算法提炼的分子结构参数的有机组合,直观地理解所有这些参数并不容易,经本文的验证集验证,包含 9个参数的二维构效关系模型 D表现出良好的预测性能.

## 3　结　语

本文以偏最小二乘法和 MolconnZ计算相结合的方法,探讨了一系列黄酮类衍生物的二维结构与其对 P糖蛋白抑制效果间的定量关系. 所建立的多个数学模型均表现出了较好的可靠性和预测性($q^2 > 0.5, r^2 > 0.9$),其中模型 D是综合效果最好的一个. 该模型对进一步筛选合成黄酮类 P糖蛋白抑制剂具有一定指导作用.

## 参考文献:


[1] SIMON S M, SCHINDLER M. Cell biological mechanisms of multidrug resistance in tumors [J]. **Proc Natl Acad Sci**, 1994, **91**(9): 3497-3504

[2] GOTTESMAN M M, PASTAN I. Biochemistry of multidrug resistance mediated by the multidrug transporter [J]. **Annu Rev Biochem**, 1993, **62**(4): 385-427

[3] LEVEILLE-WEBSTER C R, ARIAS I M. The biology of the P-glycoprotein [J]. **J Membr Biol**, 1995, **143**(2): 89-102

[4] FORD J M. Experimental reversal of P-glycoprotein-mediated multidrug resistance by pharmacological chemosensitizers [J]. **Eur J Cancer**, 1996, **32A**(6): 991-1001

[5] MIDDLETON E JR, KANDASWAMI C, THEOHARIDES T C. The effects of plant flavonoids on mammalian cells: implications for inflammation, heart disease and cancer [J]. **Pharmacol Rev**, 2000, **52**(4): 673-751

[6] SHAPIRO A B, LING V. Effect of quercetin on Hoechst 33342 transport by purified and reconstituted P-glycoprotein [J]. **Biochem Pharmacol**, 1997, **53**(4): 587-596

[7] ZHANG S, MORRIS M E. Effects of the flavonoids biochanin A, morin, phloretin, and silymarin on P-glycoprotein-mediated transport [J]. **J Pharmacol Exp Ther**, 2003, **304**(3): 1258-1267

[8] CHIELI E, ROMITI N, CERVELLI F, *et al*. Effects of flavonols on P-glycoprotein activity in cultured rat hepatocytes [J]. **Life Sci**, 1995, **57**(19): 1741-1751


?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

[9] CASTRO A F, ALTENBERG G A. Inhibition of drug transport by genistein in multidrug-resistant cells expressing P-glycoprotein [J]. **Biochem Pharmacol**, 1997, **53**(1): 89-93

[10] DI PIETRO A, CONSEIL G, PEREZ-VICTORIA J M, *et al*. Modulation by flavonoids of cell multidrug resistance mediated by P-glycoprotein and related ABC transporters [J]. **Cell Mol Life Sci**, 2002, **59**(2): 307-322

[11] ESPINOSA-MANSILLA A, ACEDO VALENZUELA M I, MUNOZ DE LA PENA A, *et al*. Comparative study of partial least squares and a modification of hybrid linear analysis calibration in the simultaneous spectrophotometric determination of rifampicin, pyrazinamide and isoniazid [J]. **Anal Chim Acta**, 2001, **427**(1): 129-136

[12] WANG Y H, LI Y, YANG S L, *et al*. An in silico approach for screening flavonoids as P-glycoprotein inhibitors based on a Bayesian-regularized neural network [J]. **J Comput Aided Mol Des**, 2005, **19**(3): 137-147

# Research on structure-activity relationship of inhibitory effects of flavonoid derivatives on P-glycoprotein

LI Yan[1], WANG Yong-hua[2], YANG Ling*[3], ZHANG Shu-wei[1], JIANG Da[1], LIU Chang-hou[1], YANG Sheng-li[3]

(1. School of Chem. Eng., Dalian Univ. of Technol., Dalian 116012, China;
2. School of Biol. Sci. and Technol., Dalian Fisheries Univ., Dalian 116023, China;
3. Lab. of Pharmaceutical Res. Discovery, Dalian Inst. of Chem. Phys., Chin. Acad. of Sci., Dalian 116023, China)

**Abstract** Inhibitory effects of the flavonoid derivatives on P-glycoprotein (P-gp)-mediated drug efflux have attracted more and more interest of researchers. By applying a computation method combining the partial least squares method and MolconnZ module, the structure-activity relationship of a series of the flavonoid inhibitors specifically binding to the NBD2 site of P-gp was conducted. All of the resultant models (especially the model D) inhibited proper reliability and predictability, which well correlated the structural features of these compounds with their inhibitory effects on P-gp. These models will be of value for further screening and developing the flavonoid-based P-gp inhibitors.

**Key words** flavonoid; P-glycoprotein; 2D-quantitative structure-activity relationship; partial least squares method

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net