基于QSPR研究绿原酸类化合物与血清白蛋白的结合*

□高 静 张燕玲 乔延江**

(北京中医药大学中药信息工程研究中心/北京市"中药基础与新药研究重点实验室" 北京 100102)

摘 要:本文选取32个小分子化学成分和9个天然产物化学成分为训练集,利用Dragon软件计算分子描述符,应用WEKA软件中CfsSubsetEval评价方法和BestFirst-D1-N5搜索方法进一步筛选描述符,利用Linear Regression方法建立了分子结构与其血清蛋白结合常数值logK的定量结构-性质相关性模型(相关系数为0.9557)。应用该模型对10个绿原酸类化合物的logK进行预测,结果显示绿原酸的logK预测值为4.51,与文献报道的logK值在4~7范围内基本相符。研究结果为进一步探讨绿原酸与血清蛋白的结合情况提供指导,为研究药物与血清蛋白的结合提供思路。

关键词:分子描述符 线性回归模型 绿原酸 血清白蛋白

doi: 10.3969/j.issn.1674-3849.2012.04.005

绿原酸(Chlorogenic acid, CA)是由咖啡酸和奎尼酸生成的缩酚酸,是植物细胞在有氧呼吸过程中经莽草酸途径的中间产物合成的一种苯丙素类化合物,是一种具有重要生物活性的小分子物质,主要存在于金银花、栀子等药材中,具有广泛的药理作用,如抗菌、抗病毒、保肝利胆、抗肿瘤、降血压、降血脂、清除自由基和兴奋中枢神经系统等,被认为是抗菌解毒、消炎类中成药的主要有效成分,且通常被作为定性定量的指标。但从20世纪60年代开始国内外的专家学者就对绿原酸是否是致敏原存有争议,至今依然没有得出确切的研究结论[1~3]。有研究表明绿原酸以半抗原形式存在,本身不具有致敏性,但它与血清蛋白结合形成复合物能产生免疫原性,从而引起过敏反应[4]。所以研究绿原酸与血清蛋白的结合情况对于其过敏反应的研究具有一定的指导意义。

血清蛋白是人体内含量最丰富的蛋白质,负责身体内各种内源与外源性物质的转运,对它们的分布与沉淀有重要作用。当药物被吸收后,进入循环系统,许多药物会与血清蛋白发生较强的、可逆的结合,生成游离型和结合型的药物,通常游离型药物具有药理活性,而结合型药物不能进行被动转运而失去其药理活性[5]。

定量结构-性质关系(Quantitive Structure Property Relationship, QSPR)研究在于建立化合物定量

收稿日期:2012-07-09

修回日期:2012-08-13

* 北京中医药大学自主课题(2011-CXTD-11):中药信息工程,负责人:乔延江;北京中医药大学自主课题(2011-JDJS-12):活血化瘀中药治疗冠心病的作用机制研究,负责人:张燕玲;教育部博士点基金(20092213120006):基于花生四烯酸代谢途径的清热解毒中药有效成分信息辨识研究,负责人:张燕玲。

** 通讯作者:乔延江,本刊编委,教授,博士生导师,北京中医药大学副校长,主要研究方向:中药信息工程,中药质量控制,Tel:010-84738620,E-mail:yjqiao@263.net。

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

的结构与性质的关系数学模型，并将之用于对模型外化合物的性质进行预测。QSPR 方法用于药物代谢、毒性、生物利用度、肠吸收等方面的研究已有成功案例[6]。本文建立药物与血清蛋白结合常数的 QSPR，用来预测绿原酸类化合物与血清蛋白的结合。

一、材料和方法

1. 材 料

分子描述符利用 Dragon professional version 2.1 计算。化合物结构式利用 ISIS_Draw 绘制并生成 mol 格式文件；药物-血清蛋白结合常数 logK 值及分子结构来源于文献[7~10]（如图 1）。中药化学成分结构较为复杂，分子量较大，因此本文为了保障所构建模型适用于中药化学成分的预测，从文献中选取了 9 个天然产物化学成分作为训练集的一部分，同时参与模型计算。

2. 化合物结构和描述符的计算方法

本实验利用 Dragon2.1[11]软件计算 42 个分子的组成、拓扑、二维（2D）自相关等 1481 个描述符[12]。由于不同描述符的量纲不同、数值相差较大，影响建模的准确率。因此本文采用标准化变换对所有描述符数据进行预处理，即对每个变量先减去样本平均值，再除以样本方差。

为了使建模数据更加准确，本实验选用了 Weka3.6.0 机器学习平台中的 CfsSubsetEval[13]评价方法和 BestFirst-D1-N5[14]搜索方法进一步筛选描述符，希望找到能区分所有样本的最佳描述符组合。

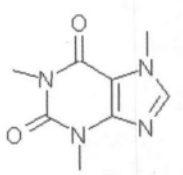
二、结果与讨论

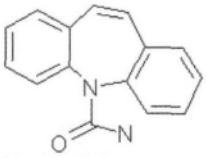
1. 描述符筛选结果

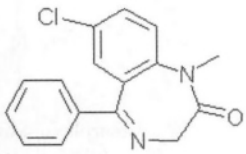
应用上述提到的筛选方法，最后筛选出 20 个描述符用于建立模型（如表 1），包括拓扑描述符、2D 自相关、原子中心碎片、BUCT 等。

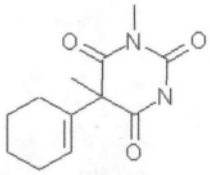
2. 模型的建立

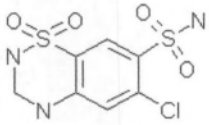
应用 Weka3.6.0 中的 Linear Regression（最小平方线性回归）方法对上述筛选得到的 20 个描述符建立回归模型，建模时先用全部属性建立一个完整模式，将这些属性按它们的标准化系数进行降序排列，然后逐一去掉每个项直至模型进行预测的估计误差不在降低[14]。因此，在属性筛选的基础上进一步

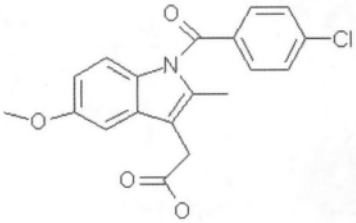

Caffeine
logK=2.20

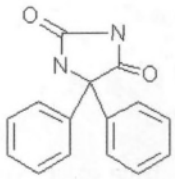

Carbamazeine
logK=4.89

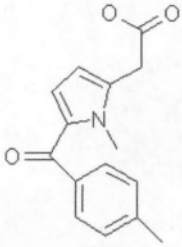

Diazepam
logK= 4.51

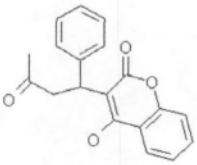

Hexobarbital
logK=3.08

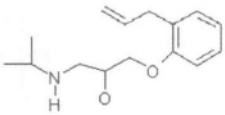

Hydrochlorothiazide
logK=3.00

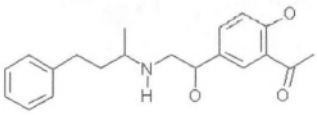

Indomethacine
logK=4.81

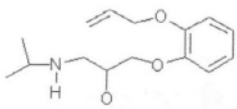

Phenytoin
logK= 3.43


Tolmetin
logK=4.96

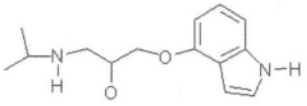

Warfarin
logK=2.57


Alprenolol
logK=2.72

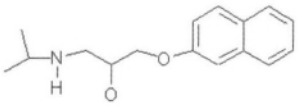

Labetalol
logK=2.66


Oxprenolol
logK=2.66

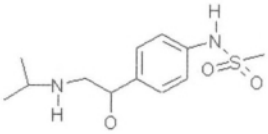
?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net



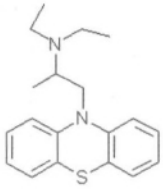
Pindolol
logK=2.11



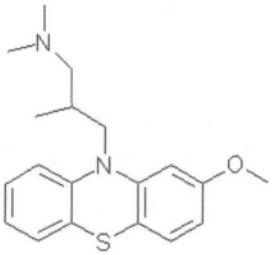
Propranolol
logK= 3.04



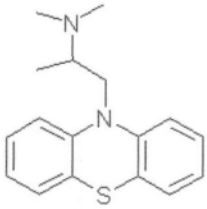
Sotalol
logK=2.18



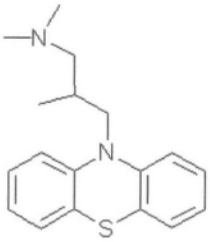
Etopropazine
logK=3.36



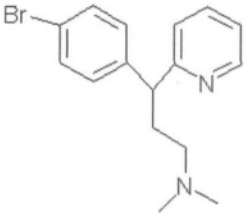
Metotrimeprazine
logK= 4.41



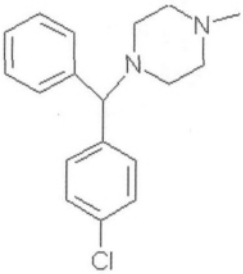
Prometazine
logK=4.38



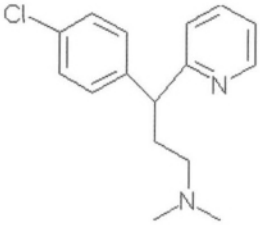
Trimeprazine
logK=3.68



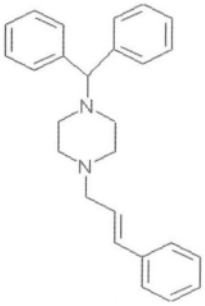
Brompheniramine
logK=3.26



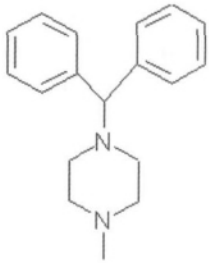
Chlorcyclizine
logK=3.92



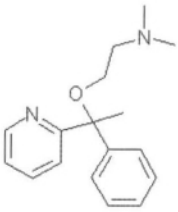
Chlorpheniramine
logK=3.11



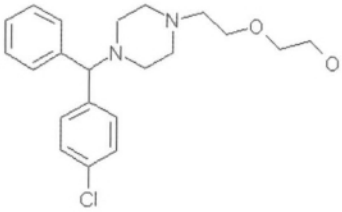
Cinarizine
logK=3.57



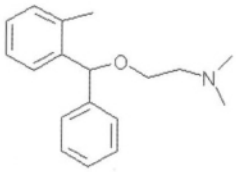
Cyclizine
logK=3.72



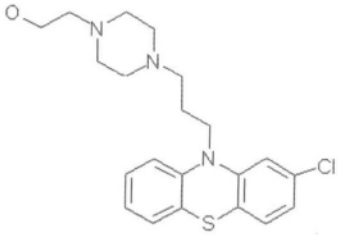
Doxylamine
logK=2.85



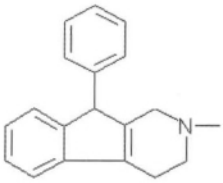
Hydroxyzine
logK= 3.61



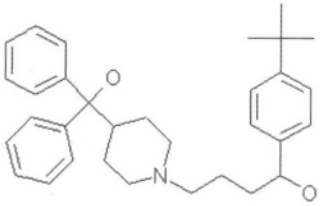
Orphenadrine
logK=3.98



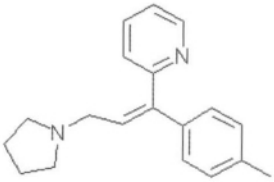
Perphenazine
logK=3.01



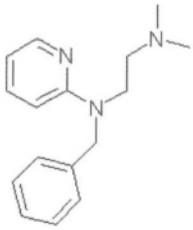
Phenindamine
logK= 4.61



Terfenadine
logK= 3.62



Tripeleuamine
logK=3.23



Tripolidine
logK=3.58

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

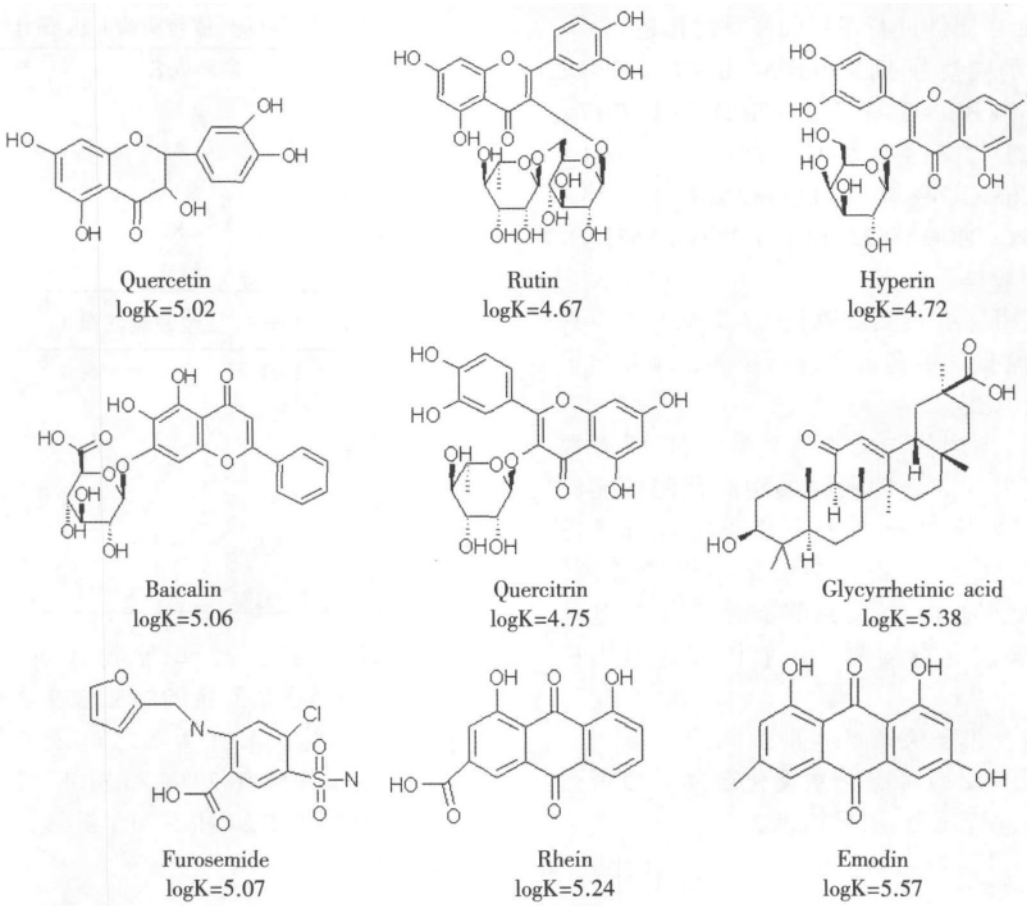


图 1　建模分子的结构式及 logK 值

表 1　筛选得到的描述符的名称

No.	Abbreviation	Full name
1	nCIR	Number of circuits
2	nN	Number of Nitrogen atoms
3	piPC09	Molecular multiple path count of order 09
4	BELe1	Lowest eigenvalue n.1 of Burden matrix/weighted by atomic Sanderson electronegativities
5	nR07	Number of 7-membered rings
6	T(N…N)	Sum of topological distances between N…N
7	BELm2	Lowest eigenvalue n.2 of Burden matrix/weighted by atomic masses
8	BEHv2	Highest eigenvalue n.2of Burden matrix/weighted by atomic van der Walls volumes
9	BEHp2	Highest eigenvalue n.2of Burden matrix/weighted by atomic polarizabilities
10	JGI4	Mean topological charge index of irder4
11	RDF075p	Radial Distribution Function-7.5/ weighted by atomic polarizabilities
12	MATS7p	Moran autocorrelation -lag7/ weighted by atomic polarizabilities
13	MATS7v	Moran autocorrelation -lag7/ weighted by atomic van der Walls volumes
14	R5e	R autocorrelation of lag 5/ weighted by atomic Sanderson electronegativities
15	R7e+	R maximal autocorrelation of lag7/weighted by atomic Sanderson electronegativities
16	nCOORPH	Number of esters(aromatic)
17	ATS7e	Broto-Moreau autocorrelation of a topological structure-lag7/weighted by atomic Sanderson electronegativities
18	C-040	R-C(=X)-X/R-C#X/X-=C=X
19	H-050	R-C(=X)-X/R-C#X/X-=C=X
20	nNHR	The number of secondary amine(aliphatic)

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

优选了其中 12 个属性进行了回归模型的构建。

最终得到的模型为：logK=0.6754×nR07−0.9247×piPC04+2.1165×BELm2−1.5029×BEHv2+1.7817×BELe1+1.6021×JGI4+1.6031×ATS7e+1.4694×MATS7p+1.2768×R7e++1.0248×nCOOR−1.2118×nNHR−0.7098×C−040+0.3622，相关系数 r=0.9557。

3. 模型的验证

为了验证模型的预测准确性，从文献[15]中找到 5 种已知结合常数的中药分子进行预测，预测结果与实验值比较（见表 2）。

结果显示，预测值与实验值基本相符，没有数量级上的较大差异。表明所建模型的预测性能较好，也体现了本文选择训练集分子时着重考虑了天然产物化合物，以提高模型适用性的意义。应用 QSPR 模型对 10 个绿原酸类化合物与血清蛋白的结合常数进行预测，10 个化合物的结构见图 2。

4. 绿原酸类化合物预测结果

用上述得到模型对绿原酸类化合物与血清蛋白的结合常数进行预测，结果见表 3。

表 2 预测 logK 值与实验 logK 值比较

中药分子名称	实验 logK 值	预测 logK 值
桑色素	5.17	4.93
川陈皮素	4.68	4.61
秋水仙碱	4.73	5.05
和厚朴酚	4.82	4.61
大黄素	5.69	5.15

表 3 模型预测结果

分子编号	LogK 值	分子编号	LogK 值
1	3.84	6	3.24
2	3.38	7	4.71
3	5.00	8	4.51
4	5.00	9	2.92
5	3.96	10	4.27

结果显示，8 号分子绿原酸的 logK 值为 4.51。毕淑云等[5]用光谱法测定的绿原酸在 290K−320K 与人血清白蛋白的 logK 在 4~7 之间；孙艳辉等[16]用同样方法测定绿原酸在 298K 和 310K 与牛血清白蛋白的 logK 分别为 5.27 和 5.20。虽然预测值与文献

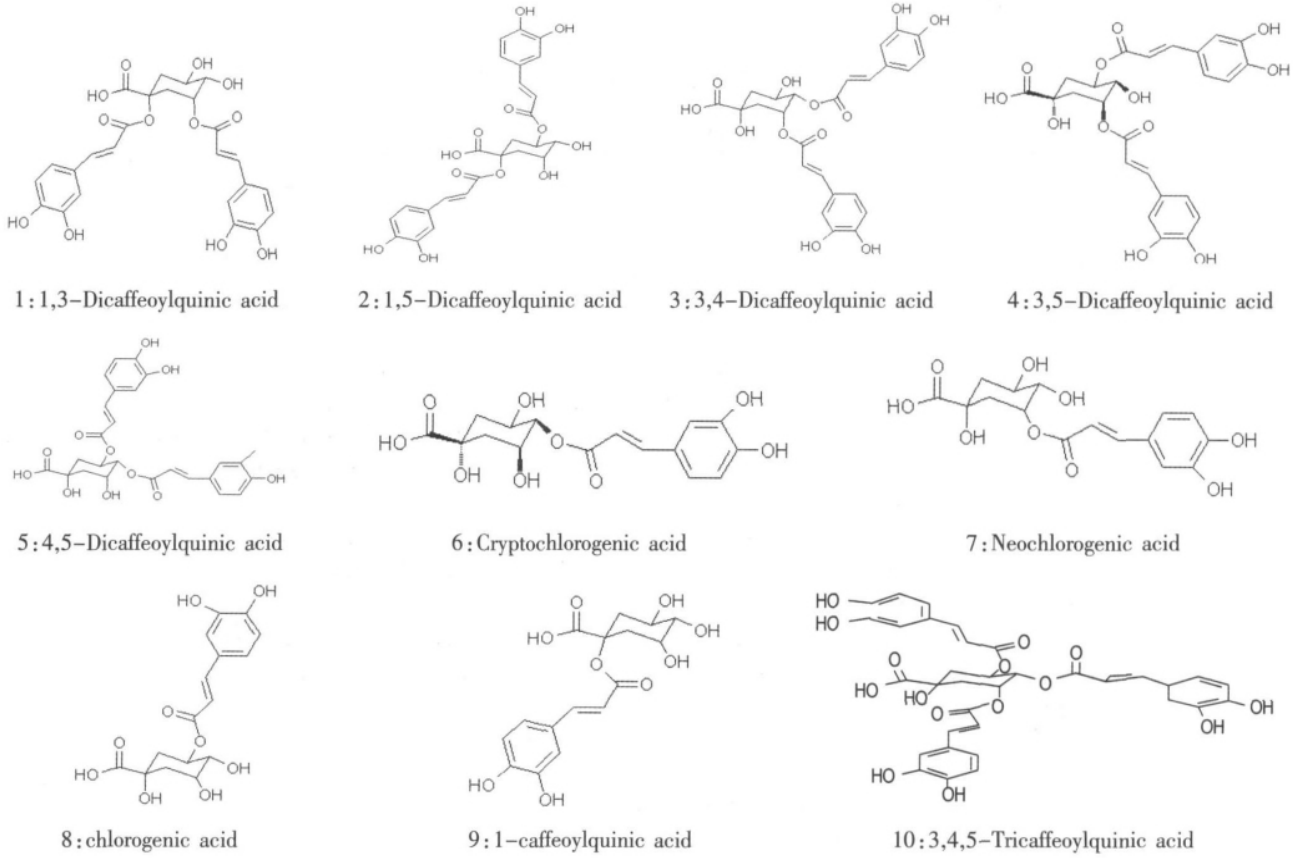


图 2 10 个绿原酸类化合物名称与结构

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

的报道值之间存在一定偏差，但是基本是一致的，表明所建立的模型有较好的预测能力。

在预测结果中，互为同分异构体的 1~5 号分子与 6~9 号分子，分别呈现出不同的 logK 值，表明该模型能够识别出结构相近的化合物，这对于研究绿原酸类化合物与血清白蛋白的结合有一定的指导意义。结果显示，3、4、7、8、10 号分子均与血清蛋白有较强的结合。

三、结 论

本文利用 QSPR 模型研究绿原酸类化合物与血清白蛋白的结合情况，建立了分子结构与血清蛋白结合常数 logK 之间的 QSPR 模型，应用该模型预测绿原酸类化合物的 logK 值。结果显示，本模型预测的绿原酸的 logK 值与文献报道实验值基本相符，表明所建模型有较好的预测能力。

本实验模型对药物分子的预测结果仍有一定的局限性，但对了解药物与血清白蛋白的结合情况，预测结合常数具有一定的指导意义。研究结果可为“绿原酸类化合物可能是半抗原、其引起过敏反应是因为与血清蛋白结合形成复合物而产生免疫原性”的实验研究提供参考。

参考文献

1 Layton L L. Nondiffusible allergenic contaminant isolated from samples of chlorogenic acid causing allergic reaction: pure chloro-genic acid not an allergen. *Allergy*, 1966, 5(5):268.

2 张瑞霞, 汤纳平, 林海霞, 等. 绿原酸与双黄连粉针剂致敏性比较评价.世界科学技术-中医药现代化, 2010, 12(6):1005~1008.

3 Freedman S O. Chlorogenic acid: an allergen in green coffen bean. *Nature*,1961:192~241.

4 彭博, 贺蓉, 徐启华, 等. 绿原酸对 RBL-2 H3 细胞脱颗粒作用的研究.中国中药杂志, 2011, 36(7):921~916.

5 毕淑云, 闫莉莉, 孙艳涛. 光谱法研究绿原酸与人血清白蛋白的相互作用.化学分析计量, 2010, 19(2):31~34.

6 Colmenarejo G, Alvarez-Pedraglio A, Lavandera JL. Cheminformatic Models To Predict Bindjing Affinities to Human Serum Albumin. *J Med Chem*, 2001, 44:4370~4378.

7 Martínez-Gómez M A, Sagrado S, Villanueva-Cama s RM, *et al*. Characterization of basic drug-hunman serum protein interactions by capillary electrophoresis. *Electrophoresis*, 2006, 2:3410~3419.

8 Martínez-Gómez M A, Carril-Avilés MM, Sagrado S, *et al*. Characterization of basic drug-hunman serum protein interactions by capillary electrophoresis. *Joumal of Chromatography A*, 2007, 1147:261~269.

9 Martínez-Pla J J, Martínez-Gómez M A, Martín-Biosca Y. High-throughput capillary electrophoresis frontal analysis method for the study of drug interactions with hunman serum albumin at near-physiological conditions. *Electriphoresis*, 2004, 25:3176~3185

10 周能. 药活性成分与血清蛋白的相互作用及其对药物结合影响的研究.中南大学博士学位论文, 2007:1~179.

11 Todeschini R, Consonni V, Pavan M. Dragon software, version 2.1. Milano Chemometrics and QSAR Research Group: Milano, 2002.

12 Todeschini R, Consonni V. Handbook of Molecular Descriptors. Wily-VCH: Weinheim, 2000:1~510.

13 Remco R. Bouckaert, Eibe Frank, Mark Hall. Weka Manual for Version 3.6.0. The University of Waikato: New Zealand, 2008:10~210.

14 Witten LH, Frank E(董琳,邱泉,于晓峰, 等译). 数据挖掘 :实用机器学期技术.北京 :机械工业出版社, 2006:282~283.

15 朱瑞瑞, 倪永年. 血清白蛋白与中药活性成分相互作用研究进展.化学研究与应用, 2011, 23(10),1281~1289.

16 孙艳辉, 魏亚玲, 吴霖生, 等. 光谱法研究绿原酸与牛血清白蛋白的相互作用.食品工业科技, 2010, 30(4):129~131.

Study on Combination of Chlorogenic Acid Compounds and Serum Albumin Based on QSPR

Gao Jing, Zhang Yanling, Qiao Yanjiang

(Research Center of Traditional Chinese Medicine Information Engineering / Beijing Key Laboratory of Chinese Pharmaceuticals and Drug Innovation, Beijing University of Chinese Medicine, Beijing 100102, China)

Abstract: Molecular descriptors of the training set, which include 32 small molecules and 9 natural compounds, were calculated by the Dragon software. CfsSubsetEval evaluation methodology and BestFirst-D1-N5 searching method were further used to screen molecular descriptors. Quantitative structure-properties relationship (QSPR) model was established by linear regression method to analyze the relationship between molecular structure and the binding constant logK of serum albumin. The model, with the correlation coefficient of 0.9557, was used to predict logK value of 10 chlorogenic acids. The predictive logK of chlorogenic acid is 4.51, which is consistent with the value 4~7 that reported in literatures. The results will be useful for further study on the combination of chlorogenic acid and serum albumin.

Keywords: Molecular descriptors, linear regression model, chlorogenic acid, serum albumin

（责任编辑:申萌萌　张志华,责任译审:王　晶）

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