

基于拉曼光谱法的品牌仿冒药检测方法研究

王科兵¹,高群²,李皓²,陆峰²(1. 解放军第169医院药械科,湖南衡阳421001;2. 第二军医大学药学院,上海200433)

[摘要] 目的 针对目前市面上的品牌药仿冒的制假现象,以国内多个厂家生产的头孢氨苄胶囊为工具药,建立基于支持向量机和相似度法或主成分分析的两步判别方法。方法 第一步 通过模拟建立头孢氨苄胶囊品牌药及4种仿冒药的分类模型,建模交叉验证的准确率为95.63%,并对可疑样品进行预测分类,准确率为93.75%,以自制样品作为额外的预测集,假阳性率为25%。第二步 对阳性样品进行相似度计算或主成分分析,可分别将假阳性率降低至5%和0%。结果与结论 两步判别法可快速、准确地实现对品牌仿冒药的检测。

[关键词] 拉曼光谱;品牌仿冒药;支持向量机;相似度法;主成分分析

[中图分类号] R927 **[文献标志码]** A **[文章编号]** 1006-0111(2013)02-0108-05

[DOI] 10.3969/j.issn.1006-0111.2013.02.007

Study on brand counterfeit drugs detection methods by Raman spectroscopy

WANG Ke-bing¹,GAO Qun²,LI Hao²,LU Feng²(1. Department of pharmacy, the 169th Hospital of PLA, Hengyang 421002, China; 2. School of Pharmacy, Second Military Medical University, Shanghai 200433, China)

[Abstract] **Objective** To established a two-step discriminant method based on SVM and HQI or PCA, in which cephalexin capsules from several domestic manufacturers were used as tool drugs for counterfeits of brand drugs appear on the market. **Method** In the first step, a simulating classification model of cephalexin brand capsules and 4 kinds of counterfeit drugs was built, the cross-validation accuracy was 95.63%, and prediction accuracy of suspicious samples within the model was 93.75%, and the false-positive rate of additional homemade test samples was 25%. In the second step, the positive samples were applied to HQI and PCA methods, and the false positive rate could be reduced to 5% and 0% respectively. **Results and Conclusion** The two-step discrimination method established in our study could detected counterfeits of brand drugs efficiently and accurately.

[Key words] Raman spectroscopy; brand counterfeit drugs; SVM; HQI; PCA

近几年来,品牌仿冒药的假药形式已经成为世界范围内药品工业的重要问题,尤其在国内,更是频频发生^[1]。世界卫生组织(WHO)揭示了六类假药,其中之一就是品牌仿冒药^[2]。品牌药仿冒行为,严重的干扰了药品市场和生产企业的发展,更威胁着人们的身体安全。解决这类问题,需充分发挥基层药检人员的作用,从一线检验中及时发现并采取相应手段来抵制这种制假现象。因此,研究一种快速、高效的药品抽检技术,已是当务之急。

拉曼光谱技术具有简便、快速、灵敏、无需样品处理等优势,非常适合用于药品的现场快检。近年来,随着便携式拉曼光谱仪的产生与化学计量学方法的发展,拉曼光谱技术的应用范围得到了普及,将两者结合后成功应用于假药快检领域的案例已被多

次报道^[2]。本研究以头孢氨苄胶囊作为工具药,建立了基于支持向量机(SVM)和相似度(HQI)或主成分分析(PCA)的两步判别方法^[3],可实现对品牌仿冒药的准确检测,同时可为司法部门提供参考依据。

1 实验

1.1 样品准备 由于品牌药与仿冒药的外观、成分等往往具有很大的相似性,本研究所选择的不同厂家在处方、工艺、制剂等方面也同样极为相似,且仿制药之间的鉴别与品牌药仿冒药的鉴别相比,更具有难度^[4],用来模拟品牌药与仿冒药的关系非常适合。收集来自国内5个厂家生产的头孢氨苄胶囊,规格均为125 mg。选取其中1个厂家的药品作为品牌药,编号为1,其余4个厂家的药品作为该品牌药的仿冒药,编号为2~5。每个厂家各收集48个样品(来自不同批次),5个厂家共计240个样品。按照欧洲药品管理局的规定,随机选取2/3即160个样品用来建模,其余1/3即80个样品用于预

[基金项目] 国家重大科学仪器设备开发专项项目(2012YQ180132),上海市科委重点科技攻关项目(11431922500)。

[作者简介] 王科兵(1966-),男,主任药师。Tel: 13037343451, E-mail: wangkb169@163.com。

[通讯作者] 陆峰。E-mail: fenglufeng@hotmail.com。

测^[3]。此外,按照头孢氨苄胶囊的处方,实验室自制20个模拟样品作为额外的预测集样品,因此预测集样品共计100个,包含80个模型预测样品,20个额外的预测样品。

1.2 拉曼光谱采集 采用海洋光学公司的 ACCUMAN 拉曼光谱仪,Hamamatsu S7031-1006 面阵 CCD 检测器,采用 785 nm 激光激发波长,激光输出功率为 300 mw,光谱分辨率为 -6 cm^{-1} 。

采样参数为积分时间 3 s。去除胶囊壳后采集内容物的拉曼光谱,每粒胶囊均以多个位置采集到的平均光谱作为该胶囊的光谱。

1.3 光谱预处理方法 截取计算所用的特征波数范围为 $180\sim 1700\text{ cm}^{-1}$ 。采用平滑方法去除噪音。采用 Rubberband Correction 算法校正基线的漂移并扣除背景干扰,最后将所有光谱做标准化处理。

1.4 方法耐用性考察 对于拉曼光谱的采集,由于采集参数均为仪器设定,采集时不会变动,容易变动且难以控制的条件仅为采集样品时的环境光照强度。药品现场快检时,环境光照会因天气或时段的变化而变化。因此本方法根据实际情况,设定了室外采集时的中午与傍晚两个时段,室内采集时的日光灯开与关两种情况,以充分模拟现场检测时的所有情况,考察采集及预处理方法的耐用性。

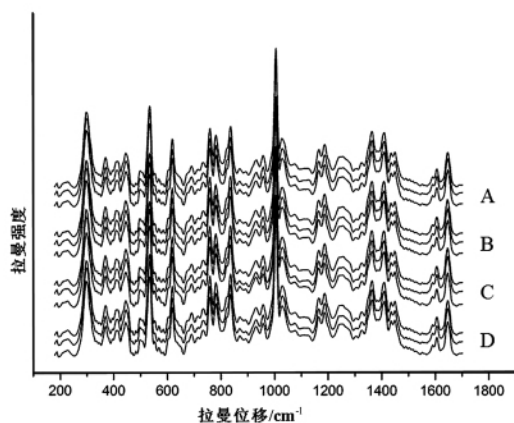


图1 厂家一的头孢氨苄胶囊不同采集条件下的拉曼光谱图
A-中午自然光; B-傍晚自然光; C-室内关灯; D-室内开灯

如图1所示,每种设定的光照环境下均采集了3张光谱图,从肉眼上看,环境内及环境间的光谱重复性很高,按照拉曼光谱耐用性考察的常规方法,需对峰位和峰形进行充分考察。

在研究中,每种光照环境下均采集了3张光谱,首先用这3张光谱的3个主峰(波数分别为 $1004.5, 835.3, 533.0\text{ cm}^{-1}$) 峰位的标准偏差(SD),以及3张光谱之间的 HQI 值作为环境内的耐用性

考察。然后求3张光谱的平均光谱作为该环境下采集的光谱,分别计算4种不同光照环境下采集光谱之间的峰位 SD 和 HQI 来考察环境间的方法耐用性,结果见表1。从表中可以看出环境内和环境间的峰位标准偏差都很小,不足1个波数。此外从光谱相似度的角度来说,环境内和环境间的 HQI 都大于0.99,达到了非常相似的程度。综合看来,本研究所用的光谱采集及预处理方法完全符合光谱耐用性考察的要求。

表1 环境内与环境间的光谱耐用性考察结果

环境状态	环境内		环境间	
	峰位 SD	HQI	峰位 SD	HQI
中午自然光	0.007-0.042	0.991-0.997		0.999
傍晚自然光	0.044-0.122	0.996-0.998	0.237-0.338	0.999
室内不开灯	0.007-0.080	0.992-0.997		0.999
室内开灯	0.016-0.107	0.990-0.996		0.999

2 结果分析与讨论

2.1 拉曼光谱分析 采集模拟体系中头孢氨苄胶囊的光谱,如图2所示:

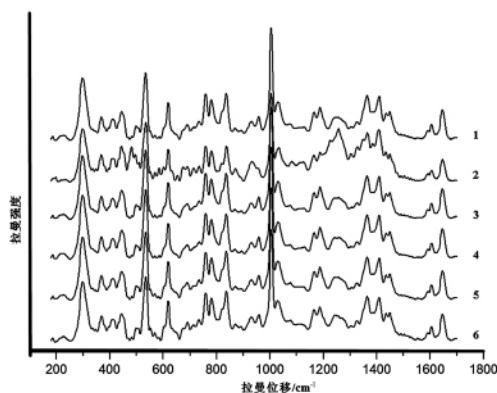


图2 头孢氨苄胶囊拉曼光谱图
1-品牌药; 2~5-仿冒药; 6-自制仿冒药

从图2中可以看出,不同厂家的头孢氨苄胶囊的光谱图,以及实验室自制的胶囊,从肉眼上看差异很小。这一方面是因为头孢氨苄胶囊的 API 含量较高,使得光谱图中反映 API 信息较多。另一方面,头孢氨苄胶囊的处方及生产工艺已十分成熟,不同厂家的处方与工艺也具有相似性。本研究体系非常适合用来模拟品牌药与仿冒药的关系,市面上所出现的品牌仿冒药首先摸索品牌药的处方,然后模仿生产工艺进行生产,最后通过仿造胶囊形状外观及包装来进行造假,其拉曼光谱图也具有很高的相似性。因此,必须借助于化学计量学的方法才能辨认出仿

冒药与品牌药的微小差异。本研究中采用3种化学计量学的方法对模型样品分别进行了建模及预测,通过比较之后结合3种方法的优势,整合为一套两步判别的方法,可实现对品牌仿冒药的准确检测。

2.2 SVM 本方法采用SVM来建立头孢氨苄胶囊品牌药与4种品牌仿冒药的分类模型。SVM非常适合用于拉曼光谱这种非线性的数据,模型准确率极高,它先增维,扩大不同厂家的头孢氨苄胶囊的拉曼光谱数据之间的边缘,然后在多维空间中用一个超平面来将类别分开^[5]。

建模时首先对160个建模样品进行内部交叉验证,75%作为校正集,25%作为验证集,采用网格状优化算法,通过内部交叉验证准确率来确定最佳参数,包括惩罚因子 c 和径向基核函数因子 g 。最终优化出交叉验证最高准确率为95.63%,模型准确,最佳参数 c 和 g 分别为32,32。采用建立的模型对80个模型预测集样品进行预测,其中类别1为品牌药,类别2~5均为该品牌药的仿冒药,建模预测集样品的分类结果见表2。

表2 模型预测集样品分类结果

真实类别	预测类别				
	1	2	3	4	5
1	16	0	0	0	0
2	0	15	0	1	0
3	0	0	16	0	0
4	3	0	0	13	0
5	0	0	0	1	15

在80个模型预测集样品中,有75个样品分类正确,模型的准确率为93.75%。其中对于品牌药类别中,16个品牌药样品均判为品牌药,无假阴性样品,阴性预报率为100%,而仿冒药中有3个样品误判为品牌药,假阳性样品数为3。

2.3 PCA 采用PCA来建立头孢氨苄胶囊品牌药与4种品牌仿冒药的分类模型。建模时同样首先对160个建模样品进行内部交叉验证,75%作为校正集,25%作为验证集。优化后所建立模型的拟合度 R^2 为0.698,模型的预测度 Q^2 为0.65,两者都远小于0.9,模型很不理想。对80个模型预测集样品进行预测,仅46个样品能正确分类,正确率为57.5%。PCA所建立的模型效果可从二维图中直观看到,这对于观察样品分布及剔除模型异常点十分有利,如图3所示。从图中也可以看出,品牌药与4类仿冒药的分类效果并不理想,大部分品牌药与仿冒药混在一起,没有明显分类界限,不能正确的对预测样品进行分类。

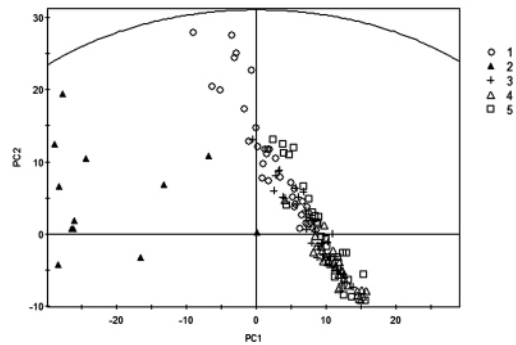


图3 PCA模型分类图

编号1-品牌药,编号2~5-品牌药的仿冒药

2.4 HQI HQI是谱库法中常用的计算待测样品光谱与标准品光谱的方法。HQI计算方法很多,本研究选择夹角余弦作为计算HQI的方法,其计算值有固定范围,从0到1,值越接近于1,表明两张光谱的相似程度越高^[6]。本研究将品牌药建模光谱随机选择10张计算平均光谱,作为品牌药的标准谱,以品牌药与仿冒药的预测光谱作为待测样品光谱,计算两者的HQI。通过比较HQI与设定某一阈值来对样品做出判别。为了评价检测方法的效能,采用建立ROC曲线的方法来实现,如图4所示。ROC曲线是根据一系列不同的二分类方式(分界值或决定阈)以灵敏度为纵坐标,1-特异度为横坐标绘制的曲线,它可以通过计算ROC曲线下的面积(AUC),用来衡量方法的检测效能^[7]。

一般来说,AUC在0.7~0.9时有一定准确性,AUC在0.9以上时有较高准确性。图4曲线的AUC为0.746,效能一般。一般来讲,最靠近左上角的ROC曲线的点是为最佳阈值,可保证假阳性和假阴性的总数最少,但此处的敏感性和特异性分别为0.67和0.75,均小于0.8,不符合快检要求。

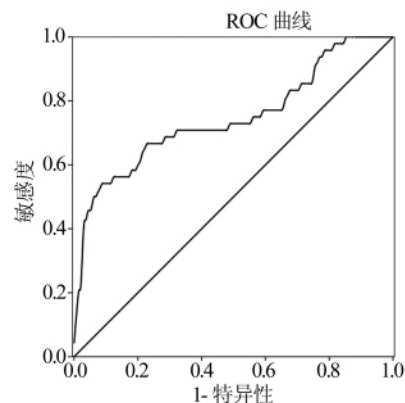


图4 HQI方法的ROC曲线图

2.5 两步判别法 上述研究中分别讨论并比较了 SVM、PCA 和 HQI 3 种方法用于仿冒药检测的优势及不足之处。3 种方法各有特色,为充分发挥 3 种方法的优势,将 3 者相互整合,建立一套完整的两步判别方法,可提高方法的准确率。

2.5.1 分析策略 对于可疑样品即预测集样品是品牌药或者仿冒药的初步判别是通过第一步来实现的。在第一步中,采用 SVM 分类法,建立了品牌药及几种仿冒药的分类模型,并对可疑样品进行预测,可疑样品在第一步中将被分到模型对应的各个类别中。在实际应用中,建立仿冒药模型的样品均为现场稽查缴获的真实仿冒药样品,因此,可疑样品被判为仿冒药(阴性样品)之后,可立即控制样品,并根据模型信息追溯样品来源,便于权威检验单位采取法定方法进一步核实,并为司法部门提供可参考的证据信息。

由于 SVM 是一种硬分类器,当预测集样品并非建模样品类别中的某一类时,预测样品依然会被归类为其中某一类。因此当模型中所收集的仿冒样品类别不完全时,很容易使得新的阴性样品误判为品牌药,产生假阳性样品。为降低假阳性结果的产生,需对阳性结果进行第二步的判别。在第二步中通过建立 HQI 或 PCA 的方法,对阳性样品进行进一步判别,降低假阳性率,提高假药样品现场初筛的准确率。

2.5.2 判别结果 本研究中,在第一步的判别中,在 80 个模型预测集样品中,有 75 个样品分类正确。其中对于品牌药类别中,16 个品牌药样品均判为品牌药,无假阴性样品,而仿冒药中有 3 个样品误判为品牌药,假阳性样品数为 3 个。此外,对自制的 20 个额外预测样品进行预测,有 5 个样品误判为品牌药,额外预测样品的假阳性率为 25%。

在第二步判别中,为了进一步对 8 个假阳性样品进行分析,将 16 个真阳性样品与 8 个假阳性样品全部纳入 HQI 或 PCA 的方法中计算进行进一步的判别。对于 HQI 方法,同样通过建立 ROC 曲线方法,评价检测方法的效能,同时优选出最佳判别阈值,如图 5 所示。ROC 曲线的 AUC 为 0.924,说明检测方法有很高的准确率,曲线的最佳判别阈值是 0.98。

如图 6 所示,以 0.98 作为判别阈值,16 个真阳性药品的 HQI 值全部大于阈值,符合要求。对 8 个假阳性样品进行判别,其中自制仿冒药样品中的 5 个假阳性样品,可进一步判别为阴性样品,降低了假阳性样品的数量,假阳性率降低为 5%,提高了判别的准确率。

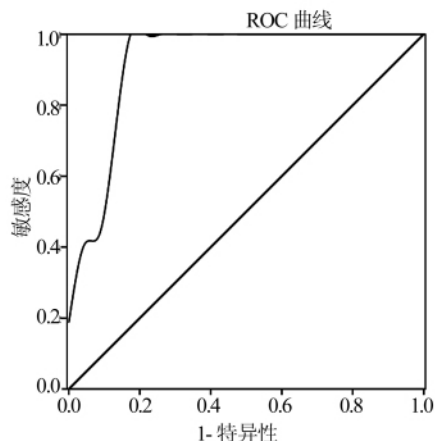


图 5 第二步判别中 HQI 方法的 ROC 曲线图

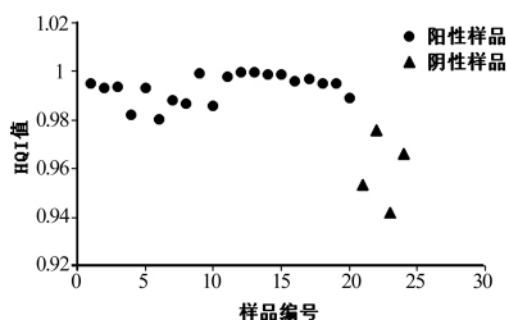


图 6 阳性样品光谱与品牌药标准谱的 HQI 结果图
编号 1~16-真阳性样品;编号 17~19-模型 3 个假阳性样品;
20~24-额外预测集中的假阳性样品。

同样,也可以通过 PCA 的方法来做进一步判别,无需建模。采用 PCA 对光谱矩阵进行降维,将多维的光谱数据,用少数几个主成分来表示,在本研究中,只选择前 2 个主成分做图,图示法可有效表示样品的分布,通过仿冒药与品牌药的分布距离,也可以对药品的真伪做出判别,如图 7 所示。

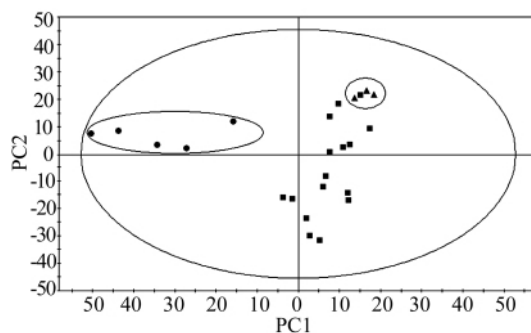


图 7 24 个阳性结果样品光谱的 PCA 图

■-模型中的真阳性样品,▲-模型中的假阳性样品,
●-自制仿冒药中的假阳性样品。

从图 7 中可以看出,假阳性样品中有 4 个样品与真阳性样品聚在一簇,无法分离,这三个样品均来

自模型。对于其余的 5 个假阳性样品,它们与真阳性样品相聚较远,且与真阳性样品间有明显界限,在第二步中可判为阴性样品,这 5 个样品均来自自制的仿冒药。因此,模型预测集的假阳性率未能降低,而额外预测集的假阳性率降低为 0,PCA 法同样有效地降低了额外预测集样品的假阳性率。

2.5.3 应用展望 本方法在第一步判别中建立了 SVM 的分类模型,可快速将样品进行分类。若可疑样品为模型中某一类样品的类别,则样品会准确归类到该类别中。在实际的现场快检中,应搜集尽可能多仿冒药的种类,模型一旦建立之后,就无需对样品进行再建模。所建立模型可直接调用来对可疑样品进行预测,并且可以为阴性判别结果进行溯源,为司法部门提供可参考的证据信息。另一方面,现场检测所遇到的新的仿冒药也应该及时纳入到模型中来,使模型不断更新完善,模型中包含的仿冒药越多,对于可疑样品的判别就越准确,阴性样品的溯源也越可信。

在实际中,新的仿冒药总是会不断出现,此时若只调用模型来判别则容易出现假阳性结果,因此需采用第二步的判别方法作为补充对阳性样品进行确证。在第二步中,我们建立了 HQI 与 PCA 两种方法,对额外的预测集样品中出现的阳性结果进行进一步分析,有效地发现了假阳性样品,提高了方法整体的正确率。

将上述两步方法结合后,一方面充分利用了 SVM 建模的稳健性,使得模型预测样品可准确归类判别,并保证了较高的阴性预报率;另一方面利用 HQI 或 PCA 直观、准确的优势,对阳性样品进一步确证,可降低假阳性率,提高现场快检的准确率及效率。

3 结论

目前,品牌药仿冒的现象层出不穷,仿冒技术也不断提高,有时采用常规方法难以检测。本研究针对市面上出现品牌药仿冒行为,以不同厂家生产的相同规格的头孢氨苄胶囊,以及实验室自制的胶囊作为工具药,恰当地模拟了市面上出现的品牌药与仿冒药关系,并建立了一套较为完整的判别方法。

本方法将拉曼光谱技术的优势与化学计量学的相关算法充分结合,准确、可靠,为品牌仿冒药的现场快速检测提供了一种简便、有效的手段。

【参考文献】

- [1] 闫冬梅. 仿冒知名品牌药的现状与对策分析[J]. 中国医学科学, 2012, 2(2): 125.
- [2] Been F, Roggo Y, Degardin K *et al.* Profiling of counterfeit medicines by vibrational spectroscopy [J]. *Forensic Sci Int*, 2011, 211: 83.
- [3] Dégardin K, Roggo Y, Been F *et al.* Detection and chemical profiling of medicine counterfeits by Raman spectroscopy and chemometrics [J]. *Anal Chim Acta*, 2011, 705: 334.
- [4] Storme-Parisa I, Rebiereb H, Matogaa M *et al.* Challenging Near InfraRed Spectroscopy discriminating ability for counterfeit pharmaceuticals detection [J]. *Anal Chim Acta* 2010 658: 163.
- [5] Dégardin K, Roggo Y, Been F *et al.* Detection and chemical profiling of medicine counterfeits by Raman spectroscopy and chemometrics [J]. *Anal Chim Acta* 2011 705: 334.
- [6] Been F, Roggo Y, Degardin K *et al.* Profiling of counterfeit medicines by vibrational spectroscopy [J]. *Forensic Sci Int* 211: 83.
- [7] Westerhuis J, Hoefsloot H, Smit S, *et al.* Assessment of PLS-DA cross validation [J]. *Metabolomics* 2008 4: 81.

[收稿日期]2013-01-08

[修回日期]2013-02-18

(上接第 97 页)

- [23] Feldman RI, Weiner H. Horse liver aldehyde dehydrogenase II Kinetics and mechanistic implications of the dehydrogenase and esterase activity [J]. *J Biol Chem*, 1972 5: 267.
- [24] Takahashi K, Weiner W. Nicotinamide adenine dinucleotide activation of the esterase reaction of horse liver aldehyde dehydrogenase dehydrogenase [J]. *Biochemistry*, 1981 20: 2720.
- [25] Mukerjee N, Pietruszko R. Human mitochondrial aldehyde dehydrogenase substrate specificity: comparison of esterase with dehydrogenase reaction [J]. *Arch Biochem Biophys*, 1992, 299: 23.
- [26] Perez MS, Hurley TD. Coenzyme isomerization is integral to catalysis in human aldehyde dehydrogenase [J]. *Biochemistry*, 2003 42: 7100.
- [27] Hammen PK, Allali HA, Hallenga K, *et al.* Multiple conformations of NAD and NADH when bound to human cytosolic and mitochondrial aldehyde dehydrogenase [J]. *Biochemistry*, 2002, 41: 7156.
- [28] Suzuki Y, Ogawa S, Sakakibara Y. Chaperone therapy for neuropathic lysosomal diseases: competitive inhibitors as chemical chaperones for enhancement of mutant enzyme activities [J]. *Perspect Medicin Chem*, 2009 3: 7.
- [29] Lowe ED, Gao GY, Johnson LN *et al.* Structure of daidzin, a naturally occurring anti-alcohol-addiction agent in complex with human mitochondrial aldehyde dehydrogenase [J]. *J Med Chem*, 2008 51: 4482.
- [30] Eun JC, Hak CK, Young CS, *et al.* Four Flavonoids from *Echinophora koreensis* and their Effects on Alcohol Metabolizing Enzymes [J]. *Arch Pharm Res*, 2009, 32: 851.
- [31] 汤子春, 邹坤, 陆海华. 开口箭与筒鞘蛇菰提取物醒酒作用机制的研究 [J]. *时珍国医国药* 2007, 18: 2958.

[收稿日期]2012-04-26

[修回日期]2012-11-12