

· 论 著 ·

拉曼光谱结合偏最小二乘法分析大鼠血液中紫杉醇含量

滕美玉^{1,2}, 宋佳^{2,3}, 赵毅⁴, 逯城宇², 邢高杨², 李兰洲², 闫国栋², 王迪²

1. 吉林省吉测检测技术有限公司(长春 130012)
2. 吉林大学 生命科学学院(长春 130012)
3. 天津科技大学 生物工程学院(天津 300457)
4. 吉林大学第一医院 创伤骨科(长春 130012)

【摘要】 利用偏最小二乘法(PLS)结合拉曼光谱技术,建立了血液中紫杉醇含量的预测模型。本实验利用拉曼光谱对312个样本进行了扫描,采用高效液相色谱技术(HPLC)对血液中紫杉醇含量进行了常规分析,利用蒙特卡罗偏最小二乘法(MCPLS)剔除异常样本,确定了校准集和预测集,采用可移动窗口偏最小二乘法(MWPLS)以逼近度(D_a)为指标优化了最佳预处理方法、波长变量和隐变量数等参数,并最终建立了紫杉醇的预测模型。其校准集和预测集的预测值与真实值之间的相关系数(R_c^2 和 R_p^2)分别为0.933 1和0.926 4。最后对预测模型进行了独立验证实验,结果表明20个验证样本的相关误差为 $9.36\% \pm 2.03\%$,表明模型具有很好的拟合度和预测能力。

【关键词】 拉曼光谱; 偏最小二乘法; 紫杉醇

Analysis of paclitaxel concentration in rat plasma by Raman spectrums combined with partial least square

TENG Meiyu^{1,2}, SONG Jia^{2,3}, ZHAO Yi⁴, LU Chengyu², XING Gaoyang², LI Lanzhou², YAN Guodong², WANG Di²

1. Jilin JiCe Detective Technical Co.LTD, Changchun 130012, P.R.China
2. College of Life Science, Jilin University, Changchun 130012, P.R.China
3. College of Biotechnology, Tianjin University of Science and Technology, Tianjin 300457, P.R.China
4. Trauma Department of Orthopedics, The First Hospital of Jilin University, Changchun 130012, P.R.China

Corresponding author: WANG Di, Email: jluwangdi@jlu.edu.cn

【Abstract】 Partial least square (PLS) combining with Raman spectroscopy was applied to develop predictive models for plasma paclitaxel concentration detection. In this experiment, 312 samples were scanned by Raman spectroscopy. High performance liquid chromatography (HPLC) was applied to determine the paclitaxel concentration in 312 rat plasma samples. Monte Carlo partial least square (MCPLS) method was successfully performed to identify the outliers and the numbers of calibration set. Based on the values of degree of approach (D_a), moving window partial least square (MWPLS) was used to choose the suitable preprocessing method, optimum wavelength variables and the number of latent variables. The correlation coefficients between reference values and predictive values in both calibration set (R_c^2) and validation set (R_p^2) of optimum PLS model were 0.933 1 and 0.926 4, respectively. Furthermore, an independent verification test was performed on the prediction model. The results showed that the correlation error of the 20 validation samples was $9.36\% \pm 2.03\%$, which confirmed the well predictive ability of established PLS quantitative analysis model.

【Key words】 Raman spectroscopy; partial least square; paclitaxel

引言

DOI: 10.7507/1001-5515.201607051

基金项目: 吉林省教育厅“十二五”科学技术研究项目(440020031107)

通信作者: 王迪, Email: jluwangdi@jlu.edu.cn

紫杉醇为红豆杉的次级代谢产物,是从红豆杉中获得的复合双萜。紫杉醇化学名为 5β , 20-环氧-1,

2 α , 4, 7 β , 10 β , 13 α -六羟基紫杉烷-11-烯-9-酮-4, 10-二乙酸酯-2-苯甲酸酯-13[(2'R, 3'S)-N-苯甲酰-3-苯基异丝氨酸酯]。紫杉醇的分子量为 853.92 Da, 分子式为 C₄₇H₅₁NO₁₄^[1]。它可以与促进微管聚合的 β 微管蛋白相互作用产生细胞毒性进而达到抗癌活性^[2]。即使在不存在鸟苷三磷酸的条件下,紫杉醇也可以结合到微管蛋白 β 亚单位,促进微管蛋白 α 和 β 亚单位聚合,稳定微管,从而发挥抗癌作用^[3]。紫杉醇是临床上广泛应用的抗癌药物,已用于治疗卵巢转移癌、膀胱癌、肺癌、食管癌等^[4]。准确而快速地测定紫杉醇含量,对于紫杉醇的临床应用具有重要意义。

拉曼光谱是印度物理学家拉曼(Sir C. V. Raman)于1928年首次发现的。拉曼光谱是一种非弹性散射的电磁辐射,是分子振动和辐射之间能量交换的结果^[5]。在入射光和样本分子中发生能量交换,这种非弹性的散射称为拉曼反应。拉曼光谱具有无损伤和高特异性的优点^[6],是快速检测及鉴定领域的新兴技术。在目前的研究中,拉曼光谱已在各个领域当中得到应用。在食品检测中,拉曼光谱已被用于牛奶中三聚氰胺的检测^[7];在食品加工产业中,拉曼光谱已被用于对猪肉质量的检测^[8]以及对肉制品烹制温度的预测^[9]。

偏最小二乘法(partial least square, PLS)具有很好的选择性和预测准确性,适用于复杂的多组分光谱^[10],是使用最广泛的多变量校准方法之一。PLS能够消除数据中共线性的影响,有效降低光谱数据维度。在目前的研究中,PLS被用于建立物质含量预测模型,例如对总花色苷含量的预测^[11]。此外,波长的选择、异常值的剔除、校准集和预测集的选择都决定了模型的预测精准度,而蒙特卡罗偏最小二乘法(Monte Carlo partial least square, MCPLS)常用于剔除光谱样本中的异常值。

本文尝试采用PLS结合拉曼光谱建立紫杉醇含量的定量分析技术,通过对最优光谱预处理的筛选,筛选波长变量和隐含层数,获得最优的逼近度(degree of approach, D_a)值,进而筛选出最优的预测模型,以达到准确和快速地测定紫杉醇含量。

1 实验部分

1.1 材料

紫杉醇(江苏红豆杉药业有限公司);聚氧乙烯蓖麻油(上海共振生物科技有限公司);乙醇(北京化工厂);乙酸乙酯(国药集团化学试剂有限公司);乙腈(北京化工厂)。

1.2 动物饲养

实验方案经吉林大学实验动物中心许可(许可证号 SCXK-(JI)2013-0003)。SD大鼠(6周龄,180~220 g,雄性)容纳在透明塑料笼中,并保持在12 h明/暗循环光照中(开灯时间7:00-19:00),温度(23 \pm 1) $^{\circ}$ C,自由饮水进食。实验8 h前,动物禁食、自由饮水。所有的实验均在安静的房间内进行,并且单鼠单笼。共纳入26只SD大鼠,每只取12份血样,共计312份血样。

1.3 血浆样品制备

将100~600 ng的紫杉醇溶解在混合溶液[聚氧乙烯蓖麻油:乙醇(体积/体积)=1:1]中,然后将一定量的紫杉醇溶液加入到大鼠血浆中。振荡数次,在37 $^{\circ}$ C下孵育30 min,血浆样品的处理参照之前发表过的方法^[12]稍作修改。取800 μ L乙酸乙酯,加入到100 μ L的血浆样品中,经1 min涡旋混合后,将溶液以10 000 r \cdot min⁻¹离心10 min,共离心两次。将有机层分离并在37 $^{\circ}$ C氮气下蒸干,残余物溶解于100 μ L乙腈,涡旋混合1 min。溶液以10 000 r \cdot min⁻¹离心10 min后,将离心液转移到注射瓶中进行高效液相色谱(high performance liquid chromatography, HPLC)分析。

1.4 紫杉醇含量的测定

本实验所用的色谱柱为Agilent ZORBAX Eclipse XDB-C18柱(4.6 mm \times 250 mm, 5 μ m),配制流动相为甲醇:水:乙腈的体积比为23:41:36,液相流速为1.0 mL \cdot min⁻¹,检测波长为227 nm,柱温为40 $^{\circ}$ C,样品的进样量为20 μ L。高效液相色谱图如图1所示。

1.5 拉曼光谱数据

本实验利用InVia Raman Microscope光谱对312个紫杉醇血浆样本进行随机扫描,扫描波长范围为0~4 500 nm,扫描间隔为1 nm,入射夹缝宽为12 nm。每个样本扫描三次,取平均值(如图2所示)。

1.6 建立PLS定量分析模型

程序脚本的编写采用Matlab2010Ra(美国Math Works)软件。

1.6.1 MCPLS对异常样本的剔除 样本进行留一交叉验证用来筛选初始隐变量数(hidden variables, nLV),如式(1)所示,当为校准集时,结果为校准集均方根误差(root mean square errors of calibration, RMSEC);当为预测集时,结果为预测集均方根误差(root mean square errors of prediction, RMSEP)。

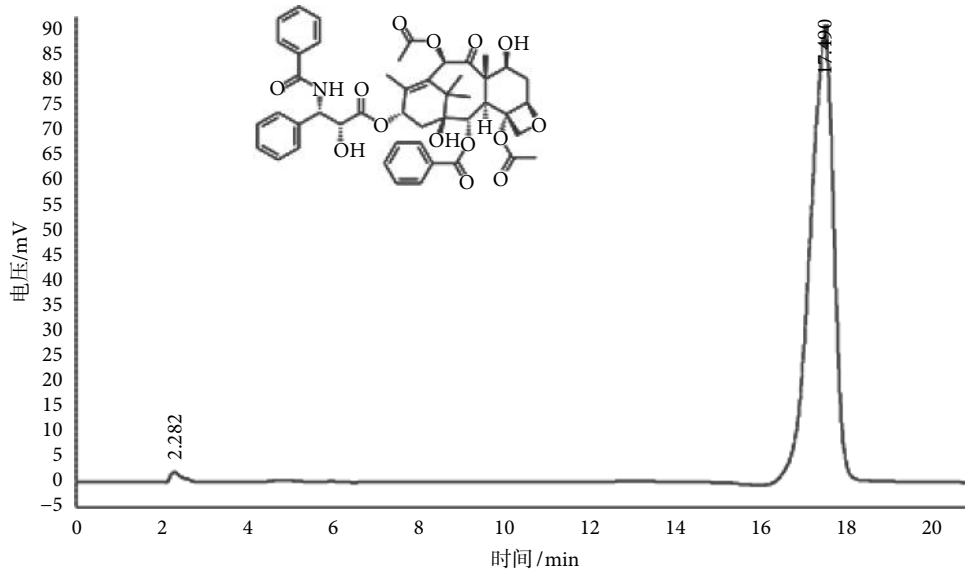


图1 紫杉醇化学结构及高效液相色谱图

Fig.1 The chromatogram and structure of paclitaxel

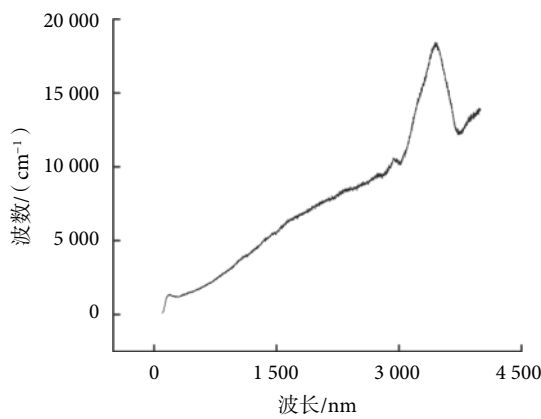


图2 紫杉醇血浆样本拉曼光谱平均值

Fig.2 The average of Raman spectroscopy of plasma paclitaxel samples

$$RMSE = \sqrt{\frac{\sum (y_p - y_r)^2}{n}} \quad (1)$$

式中 n 为样品数量, y_p 为实验值, y_r 为预测值。

利用 MCPLS 进行样品异常值的剔除, 随机选择 50% 的样品作为校准集建立 PLS 模型, 其余的样品作为预测集, 反复进行 10 000 次, 保证每个样品都充当过预测集。计算预测集样品中的预测残差 (predictive residual error, PRE)、平均预测残差 (the mean of the PRE, MPRE) 和标准偏差 (standard deviation of PRE, SDPRE)。根据散点图, 将同时具有较高 MPRE 和较高 SDPRE 的数值作为异常值剔除。

1.6.2 校准集的建立 为了防止过拟合现象, 随机选取样品数的 10%、20%、30%、40%、50%、60%、70%、80%、90% 作为校准集进行 PLS 模型的建立, 其余作为预测集, 反复进行 50 000 次计算, 计算拟合度 (degree of fitting, Df) 如式 (2) 所示, 拟合度的相对标准偏差 (relative standard deviation, RSSDa) 与对应的校准集数作图 (其中 c 值为 10), 选择 RSSDa 最小的校准集数作为校准集的样品数。

$$Df = \frac{c}{\frac{RMSEC}{n_c} + \frac{RMSEP}{n_p} + |RMSEC - RMSEP|} \quad (2)$$

式中, n_c 为校准集样品数, n_p 为预测集样品数, c 为常数。

1.6.3 PLS 定量分析模型的建立 采用快速傅里叶变换 (fast Fourier transform, FFT)、卷积平滑 (savitzky-golay smoothing)、一阶导数 (first derivative, 1st) 和二阶导数 (second derivative, 2nd) 对原始光谱进行除噪处理, 选择窗口为 5、7、9、11、13 和 15 时的除噪效果, 利用可移动偏最小二乘法 (moving window partial least square, MWPLS) 优选 PLS 模型建立中的窗口 (W) 和隐变量数 (nLV) 等, 获得 PLS 定量分析模型。模型筛选中以逼近度 D_a 为评价指标 [如式 (3) 所示], 最高值则为模型参数最优。其中 nLV 值范围为 1 ~ 20, 间隔为 1; W 值分别为 25、49、74、98、123、148、172、197、221、246、271、295、320、345 和 369 (所有波长点数 1% ~ 15%); nW 范围值为 5 ~ 75, 间隔为 5。

$$D_a = \frac{c}{\frac{n_c}{n_t} \times \text{RMSEC} + \frac{n_p}{n_t} \text{RMSEP} + |\text{RMSEC} - \text{RMSEP}|} \quad (3)$$

1.6.4 独立验证实验 20 只大鼠分别给予 1 ~ 5 mg·kg⁻¹ 紫杉醇随机尾静脉注射。大鼠尾静脉取约 1.0 mL 血。将样品 4 000 r·min⁻¹ 离心 10 min。取上清转移到新的 1.5 mL EP 管并储存于 -80℃ 用于分析。20 个血浆样品作为独立验证集来测试性能最佳的定量分析模型。对实验值和预测值进行线性拟合分析,以可决系数(亦称确定系数) R^2 为模型预测效果的评价指标。 R^2 的值越接近 1, 说明回归直线对观测值的拟合程度越好,模型的预测效果越好。

2 结果与讨论

2.1 异常值的剔除及校准集样本的选择

在实验的过程中,人为因素、仪器误差均有可能造成对样本信息的采集错误,这些错误数据对模型的建立产生了不利的影响,需要予以剔除^[13]。利用 MCPLS 对样本进行筛选,MPRE 和 SDPRE 值较高的样本数据被视为异常值并剔除,本实验中共剔除了 11 个异常样本(见图 3)。

为了防止模型的过拟合^[14],实验中按照 1.6.2 所述,以 RSDDa 为评价指标,采用 MCPLS 建立血浆中紫杉醇拉曼光谱定量分析模型的校准集样本数,通过计算,当校准集为总样本的 50% 时 RSDDa 最低(见图 4)。校准集和预测集样本的紫杉醇含量统计结果如表 1 所示。

2.2 血浆中的紫杉醇浓度 PLS 预测模型

为了获得最优的定量 PLS 模型,根据 D_a 值,进一步筛选隐变量数 (nLV , 1 ~ 20)、窗口大小 (W , 25 ~ 369) 和波长变量 (nW , 5 ~ 75)。为了消除人为

和仪器对样本信息采集的干扰^[15],实验中分别运用卷积平滑、导数变换和快速傅里叶变换进行光谱除噪和消除基线漂移。通过 MWPLS 确定波长和 D_a 值之间的关系,在 1% ~ 15% 波长变量范围内,共选择 125 个波长变量用来建立紫杉醇含量的质量分析模型(见图 5)。

如表 2 所示,卷积平滑处理可有效去除光谱噪音,使模型的拟合度及预测能力达到良好平衡。当 W 为 49、 nLV 为 5、 nW 为 5 时, D_a 最大,模型效果最佳。

在得到紫杉醇含量的最优 PLS 模型基础上,进

表 1 血浆中紫杉醇含量统计值

Tab.1 The statistics of plasma paclitaxel

组别	数量	最大值/ (ng·mL ⁻¹)	最小值/ (ng·mL ⁻¹)	均值 ± 标准差/ (ng·mL ⁻¹)
校准集	135	70.20	0.66	22.65 ± 16.43
预测集	135	70.65	0.33	24.16 ± 16.81

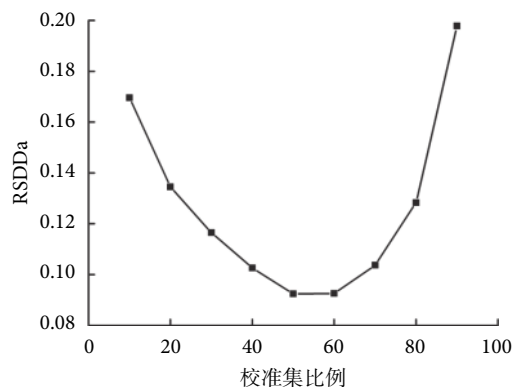


图 4 校准集比例对 RSDDa 的影响

Fig.4 The influence of calibration set ratio on RSDDa

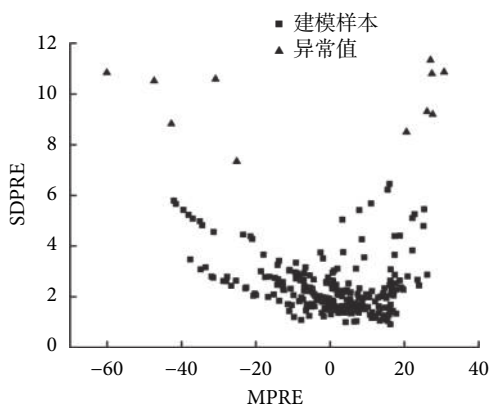


图 3 利用平均预测残差与标准偏差选择建模样本

Fig.3 Sample selection for modeling via MPRE versus SDPRE

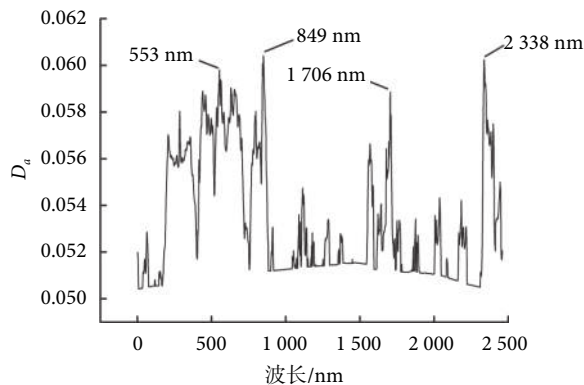


图 5 波长和 D_a 值的关系

Fig.5 The relationship between wavelengths and D_a value

表 2 光谱预处理方法及相关参数对 PLS 法建立的紫杉醇定量分析模型相关指标的影响

Tab.2 The result of suitable spectra preprocessing methods and related indexes in PLS models for paclitaxel quantitative analysis

预处理方法	初始数值	一阶导数	二阶导数	快速傅里叶变换	卷积平滑
Windows	—	11	5	7	13
W	25	98	25	25	49
nW	5	5	5	5	5
nLV	4	2	2	2	5
RMSEC/(ng·mL ⁻¹)	15.988 0	13.528 6	15.665 5	15.183 3	14.464 0
RMSEP/(ng·mL ⁻¹)	17.656 5	15.911 3	18.067 0	16.555 5	16.364 3
D _a	1.106	1.220	1.118	1.232	1.247

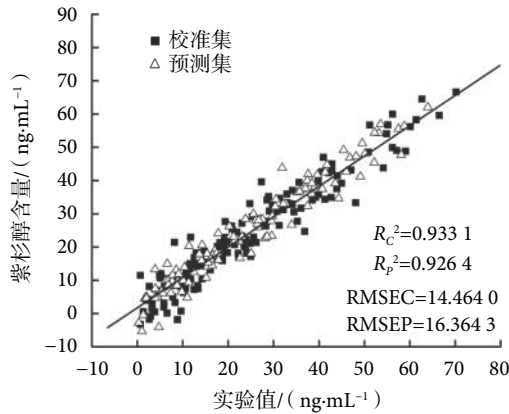


图 6 校准集及预测集中实验值和预测值的相关性

Fig.6 The correlation between experimental values and predictive values of paclitaxel in both calibration set and validation set

行模型的预测值与真实值之间的相关性考察。结果显示,模型的预测能力很好,并且未出现过拟合现象,模型的 R_c^2 和 R_p^2 分别为 0.933 1 和 0.926 4 (见图 6)。

2.3 独立验证

采用 20 个独立的紫杉醇血浆样本验证模型的预测性能,结果表明,20 个样本的实验值为 $(24.42 \pm 16.18) \text{ ng}\cdot\text{mL}^{-1}$, 预测值为 $(24.80 \pm 15.57) \text{ ng}\cdot\text{mL}^{-1}$ 。20 个验证样本的相关误差为 $9.36\% \pm 2.03\%$ 。实验表明,拉曼光谱结合 PLS 可以很成功地分析血液中紫杉醇的含量(见图 7)。

3 结论

已有研究表明,拉曼光谱在物质含量的预测中已有很广泛的应用,例如在食品生产和检测行业^[7-9],而在紫杉醇含量的预测上,还没有相关报道。本实验第一次利用拉曼光谱对大鼠血液中紫杉醇代谢进行预测,采用 PLS 结合拉曼光谱建立血液中紫杉醇含量的定量分析模型,经过优化筛选,最优分析模型的预测能力能达到 92.46%。

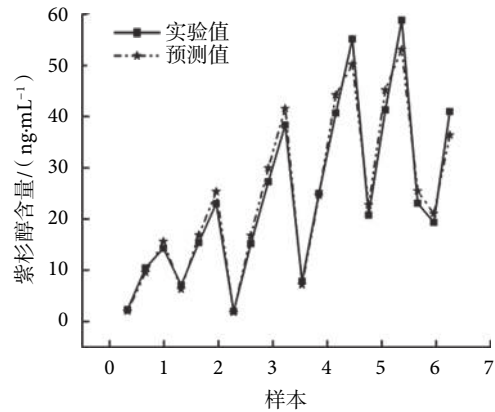


图 7 利用独立验证集检测最佳偏最小二乘法模型的预测能力

Fig.7 The performance of optimum PLS model was tested by independent validation set

本实验证实了利用拉曼光谱结合化学计量学的方法,能够有效地筛选信息,去除噪音以及无效波长变量,大大减少输入矩阵,从而减少计算量,缩短计算时间,进而达到对血液中的紫杉醇含量进行快速、无损的分析;并且,模型的泛化能力强、预测精准度高,能够满足实际检测的应用。

参考文献

- 1 Panchagnula R. Pharmaceutical aspects of paclitaxel. *Int J Pharm*, 1998, 172(1/2): 1-15.
- 2 He L, Jagtap P G, Kingston D G, *et al*. A common pharmacophore for Taxol and the epothilones based on the biological activity of a taxane molecule lacking a C-13 side chain. *Biochemistry*, 2000, 39(14): 3972-3978.
- 3 Guenard D, Gueritte-Voegelein F, Potier P. Taxol and taxotere: discovery, chemistry, and structure-activity relationships. *Acc Chem Res*, 1993, 26(4): 160-167.
- 4 Gangadevi V, Muthumary J. Preliminary studies on cytotoxic effect of fungal taxol on cancer cell lines. *African Journal of Biotechnology*, 2007, 6(12): 1382-1386.
- 5 De Beer T, Burggraeve A, Fonteyne M, *et al*. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. *Int J Pharm*, 2011, 417(1/2): 32-47.

(下转第591页; Continued on Page 591)