

全血胆固醇、甘油三酯近红外光谱分析与模型优化

黄富荣¹ 罗云瀚¹ 郑仕富² 陈哲¹

(¹暨南大学光电信息与传感技术广东普通高校重点实验室, 广东 广州 510632)

²暨南大学第一附属医院临床检验中心, 广东 广州 510632

摘要 为验证近红外光谱测量全血中胆固醇和甘油三酯的可行性, 采用近红外透射光谱结合化学计量学方法建立胆固醇和甘油三酯定量分析模型。获取全血在近红外全波段 800~2500 nm 范围的透射光谱后, 通过采用间隔偏最小二乘法(iPLS)进行特征波段的选择实现对分析模型的优化。经优化后的胆固醇、甘油三酯分析模型的特征吸收波段分别为 1650~1730 nm 和 2260~2340 nm, 预测相关系数分别为 0.79 和 0.865, 预测均方根误差分别为 0.5 mol/kL 和 0.28 mol/kL。研究结果表明近红外光谱技术可用于测量全血胆固醇、甘油三酯的含量, 运用间隔偏最小二乘法可确定特征吸收波段和优化分析模型。

关键词 近红外光谱术; 全血; 胆固醇; 甘油三酯; 间隔偏最小二乘法

中图分类号 O657.33 **文献标识码** A **doi**: 10.3788/AOS201131.1030001

Near-Infrared Spectroscopic Analysis and Model Optimization on Cholesterol and Triglyceride in Whole Blood

Huang Furong¹ Luo Yunhan¹ Zheng Shifu² Chen Zhe¹

¹Key Laboratory of Optoelectronic Information and Sensing Technologies of Guangdong Higher Education Institutes, Jinan University, Guangzhou, Guangdong 510632, China

²Clinical Laboratory Center, First Affiliated Hospital, Jinan University, Guangzhou, Guangdong 510632, China

Abstract The quantitative models of cholesterol and triglyceride analysis in whole blood are established by using near-infrared transmission spectroscopy combined with chemo-metrics methods. First the transmission spectra of whole blood in wavelength range of near infrared full spectrum (800~2500 nm) is acquired, and then the calibration models for cholesterol and triglyceride analysis are optimized by using interval partial least squares (iPLS). The results show that the optimal wave band of calibration models for cholesterol and triglyceride lies 1650~1730 nm and 2260~2340 nm respectively, and the prediction coefficient of optimal analysis model is 0.79 and 0.865 respectively, the root mean square errors of prediction of related prediction model are 0.5 mol/kL and 0.28 mol/kL respectively. The study demonstrates that compared with whole spectra data model, the iPLS model can not only improve precision, but also simplify the model. Near-infrared spectroscopy combined with iPLS can be used successfully to determine the content of cholesterol and triglyceride in whole blood.

Key words near infrared spectroscopy; whole blood; cholesterol; triglyceride; interval partial least square method

OCIS codes 300.6340; 300.1030; 300.6170

1 引言

近红外光谱(NIRS)分析技术是国内外发展较快的一种新型定性、定量分析技术, 具有分析速度快、谱图信息量大、分析效率高、样品无需预处理、分

析成本低、无试剂和无污染等优点, 在农业、食品、医药、石油、木材、烟草和饲料等行业得到了广泛的应用^[1~6]。

近年来, 采用近红外光谱技术结合化学计量学

收稿日期: 2011-04-18; **收到修改稿日期**: 2011-05-15

基金项目: 国家 863 计划(2009AA04Z315)、广东省自然科学基金(S2011040001850, 8151063201000017)和中央高校基本科研业务费专项资金(21611342)资助课题。

作者简介: 黄富荣(1979—), 男, 博士, 助理研究员, 主要从事近红外光谱技术在生物医学方面的研究。

E-mail: furong_huang@163.com

方法分析血液重要生化指标(如葡萄糖、胆固醇和甘油三酯等)的研究得到了国内外科工作者的关注,已在实验室用于分析人体血液模拟样品、血清以及血浆中的葡萄糖、胆固醇和甘油三酯等成分。Chen等^[7]用Nexus 670傅里叶变换近红外光谱仪分析实验室配制的含葡萄糖、甘油三酯、乳酸盐、尿素、抗坏血酸盐和丙盐等的血液模拟样品,结果表明在全谱区内有和频与第一倍频两个波段适于人体生化指标的分析。Hazen等^[8]利用近红外光谱分析技术结合偏最小二乘(PLS)建模对242个血清样品的葡萄糖、甘油三酯、胆固醇等生化指标进行预测。预测结果表明:主成分数为10时,在波段2061~2353 nm内建立胆固醇最优预测模型的均方根误差(RMSEP)为0.33 mol/kL;主成分数为13时,在波段2061~2353 nm内建立的甘油三酯最优预测模型的RMSEP为0.12 mol/kL。Heise等^[9]根据胆固醇、甘油三酯的C-H, O-H等官能团在倍频和和频两个区域有吸收的物理现象,在倍频区1666~1815 nm、合频区2212~2374 nm内建立124个血浆样品胆固醇、甘油三酯等含量近红外分析模型,在这两个吸收区间建模胆固醇的RMSEP分别为0.22 mol/kL和0.2 mol/kL;甘油三酯的RMSEP分别为0.155 mol/kL和0.145 mol/kL,本实验是根据含氢的物质在和频和倍频区有吸收的物理意义进行简单的建模特征波段选择,没有运用数学方法更精确地确定吸收的特征区间。国内陈华才等^[10]也探讨了人体血清生化指标的近红外分析,研究结果表明利用近红外光谱分析技术可以实现血清中葡萄糖、胆固醇和甘油三酯等生化指标的快速检测。

从国内外的研究现状可以得到近红外光谱用于血液模拟样品、血清和血浆等这些结构成分较全血简单的样品的分析是可行的。然而从全血到血清、血浆需要经过离心、分离等多个步骤,如果能直接测量全血,则不但能简化步骤,还能节约检测时间和检测费用。而目前关于近红外光谱用于全血的成分分析却鲜有报道。本文针对全血中的胆固醇、甘油三酯两种成分,采用近红外透射光谱结合间隔偏最小二乘法(iPLS),分析其测量的可行性,并通过光程长和特征波段的选取探讨其优化分析模型。

2 实验材料和方法

2.1 实验材料、仪器和测量方法

实验中114个全血样品由暨南大学第一附属医院临床检验中心提供。

理化分析仪器采用Hitachi 7600型全自动生化分析仪(日本日立公司)。光谱实验仪器采用XDS Rapid Content光栅型近红外光谱分析仪(丹麦Foss公司)。光谱采集范围为800~2500 nm,探测器为硅(800~1100 nm)、硫化铅(1100~2500 nm),光谱分辨率8 nm。实验室温度(23±1) °C,相对湿度(RH)为46%。

光程长短会影响到光能吸收,从而影响到光谱形成,而获取一个理想光谱是近红外光谱分析的保证。因而在做实验前,首先采集不同光程长(0.5, 1, 2和4 mm)全血样品,分析比较不同光程下吸光度大小来选取合适的光程长。图1是0.5, 1, 2和4 mm四种光程长的全血近红外光谱图,由图可知,当光程长为1, 2和4 mm时,在全谱区范围内吸光度都较大,在长波区域还出现吸收过饱和。而在光程长为0.5 mm时,全血样品的光谱在800~1400 nm吸光度范围是在0.6~0.9 AU;在1600~1800 nm光谱范围内的吸光度范围为0.55~0.75 AU;在2200~2400 nm波段,吸光度在0.7~1.1 AU范围。整个全谱范围内吸光度都在1 AU附近,因而,选择0.5 mm样品池作为全血光谱扫描的样品池。然后取适量的全血样品置于合适光程(0.5 mm)的石英比色皿中,用透射方式扫描获得114个全血的近红外吸收光谱,每个样品重复采集3次光谱,然后计算平均光谱。胆固醇、甘油三酯的理化分析是采用目前临床生化指标检测所使用的全自动生化分析仪方法,所得化学值作为参考值。

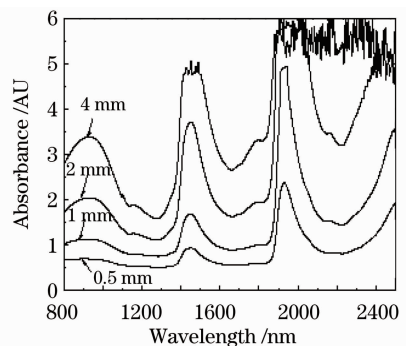


图1 不同光程全血样品近红外光谱图
Fig. 1 NIR spectra of blood samples at different thicknesses

2.2 定标集与验证集划分

异常样品包括由测定方法的变化和实验操作人员的失误等因素引起化学值异常样品以及由光谱仪本身的误差和测量性能参数的变化等因素引起的光谱异常样品。采用稳健主成分分析算法及杠杆值与

学生残差 T 检验值对异常样品剔除。首先采用稳健主成分分析算法剔除 2 个光谱异常的全血胆固醇样品,然后采用杠杆值与学生残差 T 检验值剔除 5 个化学值异常的全血胆固醇样品,共剔除了 7 个异常样品。采用稳健主成分分析算法及杠杆值与学生残差 T 检验值共剔除 9 个甘油三酯光谱异常样品及化学值异常样品。根据浓度范围和分布均匀的原

表 1 定标集与预测集的划分

Table 1 Reference measurements in calibration set and prediction set

	Cholesterol / (mol/kL)					Triglyceride / (mol/kL)			
	Min	Max	Mean	S. D.		Min	Max	Mean	S. D.
Calibration set	2.76	7.42	4.93	0.97	Calibration set	0.41	2.97	1.19	0.56
Prediction set	3.42	6.6	5.11	0.68	Prediction set	0.5	2.92	1.27	0.56

2.3 iPLS 算法及步骤

为了优选待测成分的最佳建模谱区, iPLS 是近年来发展的一种方法^[11~13]。其原理是首先将整个光谱分割成若干个等宽子区间,再在每个子区间进行偏最小二乘法回归,建立待测成分的局部回归模型。采用留一交互验证(LOO-CV)和计算交互验证均方差(RMSECV),分别比较各局部模型的精度,当 RMSECV 值最小,表明该子区间为最佳建模区间。它的优点是简便,可视化,运算量小,能快速找到最优建模谱区等。iPLS 算法已在蜂蜜的水分和酸度^[14]、苹果的糖度^[15]以及蛋白质二级结构^[16]等的测定中得到应用。

iPLS 算法步骤如下:1)首先采用中心化、自动化等预处理方法对原始光谱进行预处理;2)确定全谱区范围内建模的最佳主成分数;3)然后通过 iPLS 算法将整个光谱区域划分为 n 个等宽的子区间;4)在每个子区间上进行 PLS 回归,建立全血样品胆固醇或甘油三酯的“局部回归模型”,即可得到 n 个局部回归模型;5)以交互验证均方差(RMSECV)值为各模型的精度衡量标准,分别比较全光谱模型和各局部模型的精度,取精度最高的局部模型所在的子区间为入选区间;6)为了得到更准确的波段,将光谱区域重新分区,重复以上 3)~5)的步骤,比较不同分区的入选区间建模的效果,从而确定最佳特征波段;7)在最佳特征波段内进行内 PLS 回归,建立最优预测模型。

3 结果与讨论

3.1 全血光谱图

图 2 是 114 个全血样品的近红外光谱图,全血光谱在 1450 nm 和 1940 nm 等波段附近有明显的

吸收峰,这是全血的水分吸收。全血样品除了含有大量的水分外,还含胆固醇、甘油三酯、葡萄糖和蛋白质等富含 C-H, N-H 基团的有机物质,但光谱中除了水的吸收峰未能看到其他显著的吸收峰,这是因为水分强烈吸收将其他物质的吸收信息覆盖。另外,从全血光谱图还可以看到谱图的轮廓不够清晰,基线漂移严重。这可能是由于全血样品中含有大量的红细胞存在的缘故,这些红细胞的存在会引起严重的光散射,从而使全血光谱的发生基线漂移。由于全血不同成分吸收及红细胞散射使得某些波段存在光谱重叠和相互干扰,从而影响分析模型的质量和精度。因此选择特征波段对近红外光谱分析结果具有非常重要的影响。

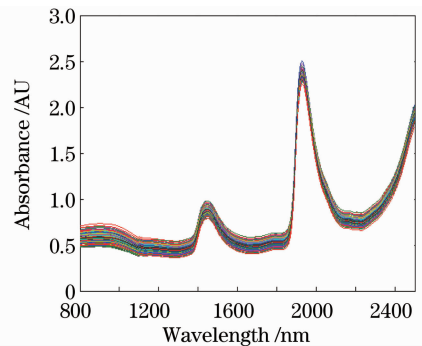


图 2 114 个全血样品近红外光谱图

Fig. 2 NIR spectra of the 114 whole-blood samples

3.2 特征波段的选择

对原始光谱采用中心化预处理后,确定全谱区范围内的最佳主成分数,再采用 iPLS 法将全谱区划分区间。建模的波长范围过窄,会把待测成分的有用信息丢失;建模的波长范围过宽,会引进更多对待测成分分析的无用信息。因而为了保证在尽可能窄的波长范围内,得到尽可能多的待测成分的信息,实验中首先在全谱 800~2500 nm 范围内划分 20 个区

间,即每个区间波长范围是 85 nm。各子区间优化结果如图 3 所示,由图 3(a)胆固醇模型可知,第 11, 18 区间的 RMSECV 值在虚线附近(其中 11 区间最小),其他各区间 RMSECV 值都很大,在虚线以上,并且大部分区间都出现了吸收饱和,因而把 11 区间作为建模入选区间。由图 3(b)甘油三酯模型可知,只有第 18 区间 RMSECV 值在虚线以下,其他区间

都在虚线以上,并且达到吸收饱和,因而把 18 区间作为建模入选区间。从以上结果可以看出,由于受全血样品的水强烈吸收和血红细胞的散射吸收影响,对于胆固醇或甘油三酯的信息只能在其两个特征谱区第一倍频区(11 区间)、合频区(18 区间)可以被提取出来,其他区间信息都被覆盖。

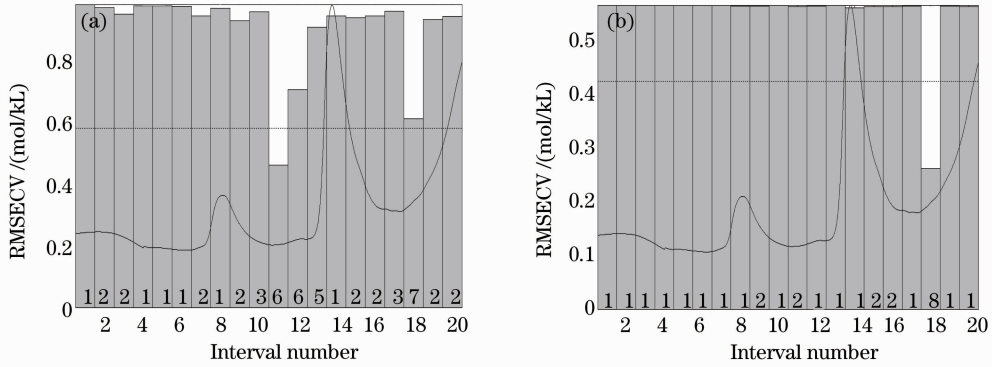


图 3 (a)胆固醇和(b)甘油三酯在全谱区内运行 iPLS 结果示意图(分割数 20)

Fig. 3 Location of the optimum intervals corresponding to full spectrum of (a) cholesterol and (b) triglyceride obtained by iPLS(interval number 20)

3.3 特征吸收波段的确定

为了得到特征吸收波段,将全谱区间分成不同分割区,观察各个不同分割区间下最优区间,再从入选区间中选一个最优因子数较低、交互验证均方差(RMSECV)值较小,相关系数 R 较大的区间,该区间即是特征吸收波段,也是最优建模波段。分别将全谱区(800~2500 nm,共 850 个数据点)划分为 n 个子区间(n : 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25),各分区优化结果如表 2 所示。对胆固醇模型,在分割区间数为 20,21,23 时,交叉检验均方差和相

关系数都非常接近,但由于它们各自所对应的最优因子数分别为 6, 8, 8, 根据在检验均方差和相关系数都非常接近的情况下选择因子数较小的模型会更加稳定的原则,选择分割区间为 20 时,第 11 区间(1650~1730 nm)作为最优特征波段,该区间交叉验证均方差 RMSECV 为 0.41 mol/kL;相关系数 R 为 0.894。甘油三酯模型在分割区间数为 21 时,第 19 区间交叉验证均方差 RMSECV 最小为 0.255 mol/kL;相关系数 R 最大为 0.887,从而确定甘油三酯的最优建模波段是 2260~2340 nm。

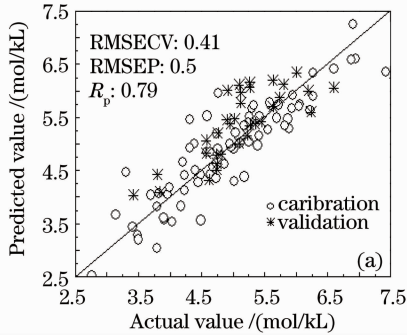
表 2 不同分割区间的 iPLS 优化结果

Table 2 Results of iPLS calibration model with different interval numbers

(a) Cholesterol					(b) Triglyceride				
Interval	Range /nm	Factors	RMSECV / (mol/kL)	R	Interval	Range /nm	Factors	RMSECV / (mol/kL)	R
15	1710~1820	5	0.72	0.662	15	2270~2390	6	0.304	0.83
16	1650~1760	8	0.43	0.897	16	2290~2390	5	0.292	0.85
17	1700~1800	10	0.53	0.841	17	2300~2400	5	0.305	0.836
18	1650~1740	9	0.448	0.889	18	2220~2310	7	0.308	0.837
19	1700~1780	10	0.44	0.891	19	2230~2320	7	0.288	0.85
20	1650~1730	6	0.41	0.894	20	2245~2330	5	0.258	0.885
21	1690~1770	8	0.41	0.906	21	2260~2340	6	0.255	0.887
22	1650~1730	9	0.46	0.882	22	2270~2345	6	0.296	0.857
23	1690~1760	8	0.42	0.901	23	2280~2350	5	0.307	0.83
24	2290~2360	6	0.73	0.674	24	2290~2360	4	0.294	0.85
25	1680~1750	6	0.44	0.891	25	2300~2360	5	0.297	0.84

3.4 最优预测模型的建立

确定最优建模波段后,在此波段内建立预测模型,该预测模型就是最优预测模型。在 1650~1730 nm 波段内,对未参与定标的 33 个全血样品胆固醇的含量预测,预测结果如图 4(a)所示,预测相关系数 R_p 为 0.79,



预测均方根差 (RMSEP) 为 0.5 mol/kL。在 2260~2340 nm 波段内,对未参与定标的 32 个全血样品甘油三酯的含量预测,预测结果如图 4(b)所示,预测相关系数 R_p 为 0.865, RMSEP 为 0.28 mol/kL。

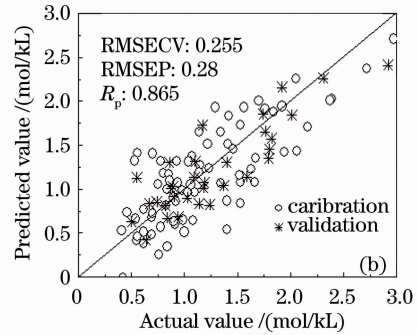


图 4 (a)胆固醇、(b)甘油三酯最优预测模型图

Fig. 4 Predicted value versus actual value of (a) cholesterol and (b) triglyceride

4 结 论

采用近红外透射光谱结合间隔偏最小二乘法确定全血样品胆固醇、甘油三酯近红外分析特征波段,并建立相应预测模型。结果表明,对于全血中胆固醇的分析,最优建模波段是在全谱划分为 20 个分区时第 11 分区内(1650~1730 nm),所建立最优预测模型的预测相关系数 R_p 为 0.79,预测均方根差为 0.5 mol/kL。对于全血中甘油三酯的分析,最优建模波段是在全谱划分为 21 个分区时第 19 分区内(2260~2340 nm),所建立最优预测模型的预测相关系数 R_p 为 0.865,预测均方根差为 0.28 mol/kL。以上结果表明,近红外光谱技术可以成功地检测全血胆固醇、甘油三酯的含量,运用间隔偏最小二乘法可以找到最优建模波段、优化分析模型。但由于受到全血其他非检测成分(如水、血红蛋白等)强吸收及血红细胞散射等因素的影响,预测精度不高,特别是对甘油三酯的检测,相对预测均方根差达 22.1%,这与临床全自动生化分析仪对甘油三酯的检测精度小于等于 6%的要求还有一定差距。因此,如何更好地克服全血其他成分(如水、血红蛋白等)强吸收及血红细胞散射等因素的影响,需要进一步重点研究。此外,只使用了 114 个全血样品,样品数量较少,其中胆固醇和甘油三酯病人的血液样品只有 21 个,将来增加样品数量,并针对高血脂和高胆固醇病人建立专门模型,有望进一步完善校正模型,提高预测精度。

参 考 文 献

- 1 P. Williams, K. Norris. Near-Infrared Technology in the Agricultural and Food Industries (2nd Ed.) [M]. Minnesota (USA): the American Association of Cereal Chemists, Inc. St. Paul, 2001
- 2 Lu Wangzhen. Modern Near Infrared Spectroscopy Analytical Technology (2nd Ed.) [M]. Beijing: Chinese Petrochemistry Press, 2007
陆婉珍. 现代近红外光谱分析技术(第二版)[M]. 北京:中国石化出版社, 2007
- 3 Sun Guangming, Liu Fei, Zhang Fan *et al.*. Fat determination of proline in herbicide-stressed oilseed rape leaves based on near infrared spectroscopy [J]. *Acta Optica Sinica*, 2010, **30**(4): 1192~1196
孙光明, 刘飞, 张帆等. 基于近红外光谱技术检测除草剂胁迫下油菜叶片中脯氨酸含量的方法[J]. *光学学报*, 2010, **30**(4): 1192~1196
- 4 Zhao Jianhua, Zhao Chongwen, Wei Zhoujun *et al.*. Study on monitoring multi-component toxic gases based on near-infrared spectroscopic method [J]. *Acta Optica Sinica*, 2010, **30**(2): 567~573
赵建华, 赵崇文, 魏周君等. 基于近红外光谱技术的多组分毒性气体检测研究[J]. *光学学报*, 2010, **30**(2): 567~573
- 5 Liu Yande, Chen Xingmiao, Ouyang Aiguo. Non-destructive measurement of soluble solid content in gannan navel oranges by visible/near-infrared spectroscopy [J]. *Acta Optica Sinica*, 2008, **28**(3): 478~481
刘燕德, 陈兴苗, 欧阳爱国. 可见/近红外光谱法无损检测赣南脐橙可溶性固形物[J]. *光学学报*, 2008, **28**(3): 478~481
- 6 Zhao Jiewen, Guo Zhiming, Chen Quansheng *et al.*. Feasibility study on use of near infrared spectroscopy in quantitative analysis of catechins in green tea [J]. *Acta Optica Sinica*, 2008, **28**(12): 2302~2306
赵杰文, 郭志明, 陈全胜等. 近红外光谱法快速检测绿茶中儿茶素的含量[J]. *光学学报*, 2008, **28**(12): 2302~2306
- 7 J. Chen, M. A. Arnold, G. W. Small. Comparison of combination and first overtone spectral regions for near-infrared calibration models for glucose and other biomolecules in aqueous solutions [J]. *Analytical Chemistry*, 2004, **76**(18): 5405~5413
- 8 K. H. Hazen, M. A. Arnold, G. W. Small. Measurement of

- glucose and other analytes in undiluted human serum with near-infrared transmission spectroscopy[J]. *Analytica Chimica Acta*, 1998, **371**: 255~267
- 9 H. M. Heise, R. Marbach, A. Bittner *et al.*. Clinical chemistry and near infrared spectroscopy: multicomponent assay for human plasma and its evaluation for the determination of blood substrates [J]. *J. Near Infrared Spectrosc.*, 1998, **6**(1): 361~374
- 10 Chen Huacai, Yang Zhongguo, Chen Xingdan *et al.*. Rapid determination of biochemical compound in human serum using Fourier transform near infrared spectroscopy [J]. *Chinese J. Analysis Laboratory*, 2005, **24**(7): 17~20
陈华才, 杨仲国, 陈星旦 等. 傅里叶变换近红外光谱法快速检测人血清生化成分[J]. *分析实验室*, 2005, **24**(7): 17~20
- 11 Liu Bing, Bi Kaishun, Sun Lixin *et al.*. Determining danshensu and hesperidin in rukuaxiaopian using near-infrared spectroscopy [J]. *World Science and Technology (Modernization of Traditional Chinese Medicine and Materia Medica)*, 2009, **11**(3): 388~396
刘 冰, 毕开顺, 孙立新 等. 近红外光谱结合不同偏最小二乘法测定乳块消片醇沉液中丹参素和橙皮苷含量[J]. *世界科学技术——中医药现代化*, 2009, **11**(3): 388~396
- 12 Wang Jiahua, Li Pengfei, Cao Nanning *et al.*. Study on the combination weight PLS model for determining SSC of peach based on the optimal information regions obtained from iPLS methods [J]. *J. Infrared Millim. Waves*, 2009, **28**(5): 386~391
王加华, 李鹏飞, 曹楠宁 等. 基于 iPLS 原理最优化信息区间的桃糖度组合权重 PLS 模型研究[J]. *红外与毫米波学报*, 2009, **28**(5): 386~391
- 13 L. Norgaard, A. Saudland, J. Wanger *et al.*. Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy [J]. *Appl. Spectrosc.*, 2000, **54**(13): 413~419
- 14 Luiz C. M. Pataca, W Mdomiro Borges Neto, Ronei J. Poppi *et al.*. Determination of apparent reducing sugars, moisture and acidity in honey by attenuated total reflectance-Fourier transform infrared spectroscopy [J]. *Talanta*, 2007, **71**(5): 1926~1931
- 15 Zou Xiaobo, Zhao Jiewen, Xia Rong *et al.*. Near infrared determination of sugar content in apples based on multiresolution decomposition and interval partial least square (iPLS) method [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2006, **37**(6): 79~82
邹小波, 赵杰文, 夏 蓉 等. 苹果糖度近红外光谱小波去噪和 iPLS 建模[J]. *农业机械学报*, 2006, **37**(6): 79~82
- 16 Susana Navea, Roma Tauler, Anna de Juan. Application of the local regression method interval partial least-squares to the elucidation of protein secondary structure [J]. *Anal. Biochem.*, 2005, **336**(2): 231~242