

血清中胆固醇的近红外光谱测量及模型适用性研究

杨皓旻^{1, 2}, 卢启鹏^{1*}

1. 中国科学院长春光学精密机械与物理研究所, 应用光学国家重点实验室, 吉林 长春 130033
2. 中国科学院研究生院, 北京 100049

摘 要 使用近红外光谱分析方法进行了血清中胆固醇含量的快速、无试剂测量研究, 并考察、分析了模型的适用性。利用傅里叶光谱仪测量 1, 2 和 6.5 mm 厚血清的近红外光谱, 在合频与第一、第二倍频波段内, 结合偏最小二乘法(PLS)对胆固醇建立了模型。1, 2, 6.5 mm 厚血清模型的预测标准差(RMSEP)分别是 0.15, 0.16, 0.29 mmol · L⁻¹, 相对预测误差平均值(MPEP)是 2.9%, 3.1%, 4.8%。为测试模型的适用性, 模型建立后一个月内分别测量了其他两批血清的光谱, 使用已建立的模型对样品进行预测, 1 mm 模型的预测效果最好。实验结果表明, 对于血清中胆固醇的含量, 可以建立较高精度的分析模型。此外, 建模采用的样品厚度较薄时, 分析模型适用性更好。

关键词 近红外光谱; 人血清分析; 偏最小二乘法
中图分类号: O651.3 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2011)02-0375-04

引 言

常规医学检查往往需要化验血液, 血清化验是其中的重要内容。通常包括化验血清中多种成分, 如蛋白类、脂类、葡萄糖等, 以判断人体的健康状况。常规化验需要试剂与血清反应, 样品不可重复使用, 因此分析多种成分需使用不同的试剂和多份血清样品, 增加了被检人员的身体和经济负担, 也容易造成环境污染。

近红外光谱分析技术是一种非破坏性的分析技术, 已广泛应用于食品、农产品、化工、制药、土壤检测、环境监测等多个行业的质量监控和在线分析^[1-3], 并不断进行扩展应用的研究。近红外光谱技术最突出的优点是可对样品实施无损、多成分同时分析, 非常适合于血清常规检查这类日常、大批量样品的分析工作。

近红外光谱技术应用于人体血液成分分析的研究已有多年, 20 世纪 80 和 90 年代, 有文献报道利用近红外光谱技术分析血液及相应模拟溶液, 对于其中的蛋白类、脂类、葡萄糖等都达到了一定的精度^[4-12], 而无创检测离实际应用还有距离^[13, 14], 因而研究人员把工作重点转入研究更深层次的基础问题^[15]。近红外光谱技术如能应用于血液、血清的临床

化验, 可以实现快速分析, 减少试剂与血液的用量, 并为无创生化检测提供帮助。

本文利用傅里叶变换近红外光谱仪测量了 1, 2, 6.5 mm 三种厚度的人血清透射光谱, 结合偏最小二乘法(PLS)对血清中胆固醇含量建立三种预测模型。在测量建模样品之后一个月内, 测量了另外两批样品的光谱, 用建立的模型预测, 考察模型的适用性。

1 实验过程

1.1 样品获取

由杭州某医院取得 40 份日常化验血清样品, 其中胆固醇的含量经 Hitachi-7170 全自动生化分析仪测定, 作为样品成分的真实值。表 1 是 40 份血清样品中胆固醇浓度的分布情况。

Table 1 Cholesterol concentration distribution in human serum samples

成分	最小值	最大值	平均值	标准差
胆固醇/(mmol · L ⁻¹)	2.87	7.25	4.42	0.99

收稿日期: 2010-05-24, 修订日期: 2010-08-26
基金项目: 国家自然科学基金项目(60878052), 国家自然科学基金重点项目(60938002)和应用光学国家重点实验室开放基金(O9Q13FQ090)资助
作者简介: 杨皓旻, 1985 年生, 中国科学院长春光学精密机械与物理研究所博士生 e-mail: yanghaomin_3@163.com
* 通讯联系人 e-mail: luqipeng@126.com

1.2 样品光谱采集

自医院取得冷冻的血清样品,经室温解冻后,使用 MPA 型傅里叶变换近红外光谱仪(布鲁克公司,德国)采集样品透射光谱。该光谱仪配备铟镓砷探测器,扫描光谱范围 $12\,500 \sim 4\,000\text{ cm}^{-1}$ 。分别使用 1, 2, 6.5 mm 的样品池。以空样品池光谱作为背景,光谱分辨率为 32 cm^{-1} ,扫描 20 次平均,仪器自测信噪比优于 1.1×10^5 。样品温度为室温。测量过程中,先使用 6.5 mm 的样品池依次测量所有样品,然后更换 2 mm 样品池,最后使用 1 mm 样品池。

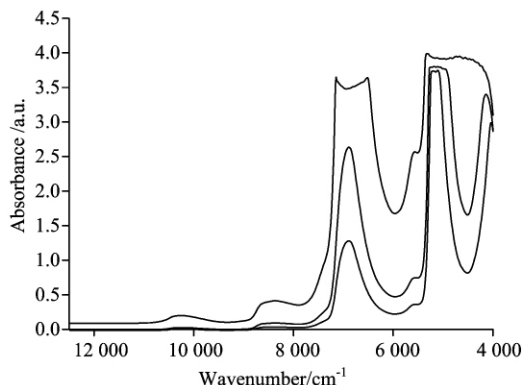


Fig 1 Spectra of sera with three kinds of thickness

2 结果与分析

2.1 血清吸收光谱

图 1 是三种厚度同一份血清样品的光谱,在 $6\,900$ 与 $5\,150\text{ cm}^{-1}$ 附近,水的吸收非常强烈,掩盖了血清中其他成分的光谱信息,严重影响对其他成分的分析,建立模型时需要避开这两个波段。

2.2 建立定标模型

用于建立模型的样品组一共 40 份,按照真实值将其均匀分为定标集和预测集两部分,其中 28 份定标,12 份预测。

使用 OPUS 软件(布鲁克公司,德国)分析血清光谱。利用软件中 QUANT-2 组件建立偏最小二乘法(PLS)模型。该软件提供优化功能,可设置多种优化条件并分别建模,列出验证结果,便于使用者从中选择最优的建模条件。对于这三种厚度,分别建立了一系列的 PLS 模型,每个模型使用 $1 \sim 20$ 个因子,尝试了五个波段,分别使用多元散射校正、矢量归一化、一阶导数等 11 种预处理方法处理光谱。

经过软件优化后,以定标模型的预测标准差(RMSEP)和决定系数(R_p^2)为评价标准,得到各成分的最优定标条件与结果,如表 2。

Table 2 Models for three kinds of thickness

成分	波段 / cm^{-1}	预处理方法	R_c^2	RMSEC / $(\text{mmol} \cdot \text{L}^{-1})$	R_p^2	RMSEP / $(\text{mmol} \cdot \text{L}^{-1})$	Factors	MPEP /%
1 mm	6 100~5 450 4 450~4 200	无	0.997 7	0.08	0.981 0	0.15	18	2.9
2 mm	6 100~5 730	矢量归一化	0.992 7	0.14	0.976 8	0.16	18	3.1
6.5 mm	12 500~7 500 6 100~5 450	MIN-MAX 归一化	0.827 1	0.46	0.926 7	0.29	6	4.8

Factors 表示 PLS 模型所使用的因子数, MPEP 表示平均相对预测误差,等于成分相对误差绝对值的平均

胆固醇含有 C—H, N—H, O—H 等分子基团,这些基团在合频与一、二倍频的特征波段分别分布在 $5\,000 \sim 4\,100$, $6\,250 \sim 5\,500$ 和 $9\,000 \sim 8\,300\text{ cm}^{-1}$ 左右,因而建立的模型大致利用这三个波段范围内的光谱。

由表 2 可见,样品的厚度不相同,模型使用的波段也不相同,样品越薄,使用的波段越靠近长波方向。对比三种厚度的模型可以看出 1 mm 模型的效果最好,这可能是由于模型包括 $4\,450 \sim 4\,200\text{ cm}^{-1}$ 波段,该波段是合频波段,谱带的重叠程度相对轻微,特征相对明显。

2.3 模型适用性

模型用于实际成分分析时,所用样品光谱必然来自不同批次的测量。对于不同批次采得的样品,样品自身性质的差异、测量时的仪器条件与周围环境都可能影响样品光谱,最终影响模型的预测结果。预测结果对这些条件的敏感程度,反映了模型对于不同样品和不同外部条件的适用性。

为了验证模型的适用性,在本次模型建立后收集了两组新样品。两组新样品是同一所医院提供的日常血清化验样品,分别有 8 份与 22 份,胆固醇的含量通过生化分析仪测得。新样品的光谱测量条件与建模样品的条件相同。

将新样品光谱代入模型计算,得到胆固醇含量的预测值,对预测值与真实值进行一元线性回归,回归方程与直线如图 2。

经过三种厚度模型计算,得到新样品的真实值与预测值的回归直线。三条回归直线都不同于理想直线,而且存在如下趋势:样品厚度越薄,得到的回归直线的斜率和决定系数越接近于 1,详细的预测结果如表 3。

由图 2 和表 3 可见,三种厚度模型中,1 mm 模型的回归直线与理想直线最接近,决定系数与直线斜率都接近于 1,预测标准差最小,说明 1 mm 模型对于这两组样品的适用性最好。2 与 6.5 mm 模型的回归效果不相上下,从回归直线上看,6.5 mm 模型的回归直线更接近于理想直线,从预测结果上看,2 mm 模型的预测标准差较小。

综合建模与适用性测试两部分结果可见,对于同一批样品,利用 1 或 2 mm 厚度样品光谱建立的模型可用于预测该批次的其他样品,且具有较高的精度,而建立的 6.5 mm 厚度模型精度较差,模型的决定系数较低。这三种模型预测其他批次样品时,1 mm 模型的适用性较好,预测结果最接近真实值并且精度没有显著的下降,而 2 与 6.5 mm 模型效果

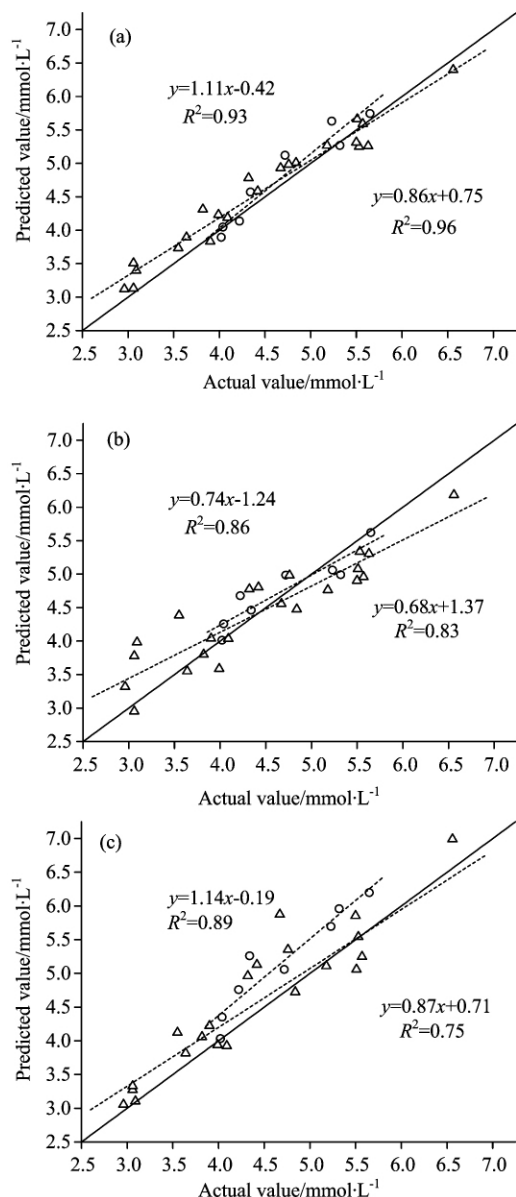


Fig 2 Results of applicability test for three thickness models

○: Pred 1; △: Pred 2
(a): 1 mm; (b): 2 mm; (c): 6.5 mm

下降不少, 预测标准差 (RMSEP) 和平均相对预测偏差 (MPEP) 都扩大了一倍左右。

总体说来, 1 mm 模型的预测精度与适用性都要好于其他两种模型。

Table 3 Results of applicability test				
样品组	样品厚度/mm	R^2	RMSEP/(mmol·L ⁻¹)	MPEP/%
第一组	1	0.93	0.23	3.7
	2	0.75	0.24	4.4
	6.5	0.89	0.53	10
第二组	1	0.96	0.26	5.4
	2	0.69	0.44	8.9
	6.5	0.75	0.53	8.5

3 结 论

利用傅里叶近红外光谱仪分别测量 1, 2, 6.5 mm 血清的光谱, 对其中胆固醇的浓度建立了预测模型。结果表明, 采用适宜的波段和方法可以对胆固醇实现较高精度的分析。三种模型的 RMSEP 分别达到 0.15, 0.16, 0.29 mmol·L⁻¹, MPEP 则是 2.9%, 3.1%, 4.8%。

为了验证模型对不同的样品适用性, 在模型建立后测量了两组新样品, 不同样品厚度下建立的模型测试效果各不相同。1 mm 模型的效果最佳, RMSEP 不大于 0.26 mmol·L⁻¹, MPEP 不大于 5.4%, 与模型定标时的效果最为接近。

本次实验用于建模的血清样品主要覆盖胆固醇的正常浓度范围, 后续实验中, 将大量搜集样品, 对测量过程和建模方法加以完善, 有望进一步提升模型的分析能力和稳定性, 建立更具实用性的模型。

References

[1] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄, 赵龙莲, 韩东海, 等). The Foundation and Application of Near Infrared Spectroscopy Analysis(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005.

[2] LU Wan-zhen(陆婉珍). Modern Near-Infrared Spectroscopy Analytical Technology, 2nd ed.(现代近红外光谱分析技术, 第2版). Beijing: China Petrochemical Press(北京: 中国石化出版社), 2006.

[3] DING Hai-quan, LU Qi-peng, PIAO Ren-guan, et al(丁海泉, 卢启鹏, 朴仁官, 等). Optics and Precision Engineering(光学精密工程), 2007, 15(12): 1946.

[4] Peuchant E, Salles C, Jensen R. Analytical Chemistry, 1987, 59(14): 1816.

[5] Heise H M, Marbach R. Proceeding of SPIE, 1991, 1575: 507.

[6] Haaland D M, Robinson M R, Koepp G W, et al. Applied Spectroscopy, 1992, 46(10): 1575.

[7] Hall J W, Pollard A. Clinical Biochemistry, 1993, 26(6): 483.

[8] Hazen K H, Arnold M A, Small G W. Analytica Chimica Acta, 1998, 371(2-3): 255.

[9] da Costa P A, Poppi R J. Analytica Chimica Acta, 2001, 446(1-2): 39.

- [10] CHEN Hua-cai, YANG Zhong-guo, LI Hui-ying, et al(陈华才, 杨仲国, 李惠英, 等). Acta Laser Biology Sinica(激光生物学报), 2004, 13(6): 429.
- [11] Chen J, Arnold M A, Small G W. Analytical Chemistry, 2004, 76(18): 5405.
- [12] Maruo K, Oota T, Tsurugi M, et al. Applied Spectroscopy, 2006, 60(12): 1423.
- [13] Arnold M A, Small G W. Analytical Chemistry, 2005, 77(17): 5429.
- [14] CHEN Xing-dan(陈星旦). Optics and Precision Engineering(光学精密工程), 2008, 16(5): 759.
- [15] Amerov A K, Chen J, Small G W, et al. Analytical Chemistry, 2005, 77(14): 4587.

Study on Measurement of Cholesterol in Serum by Near-Infrared Spectroscopy and Applicability of Models

YANG Hao-min^{1, 2}, LU Qi-peng^{1*}

1. State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
2. Graduate University of Chinese Academy of Sciences, Beijing 100049, China

Abstract To study rapid analysis of cholesterol concentration in human serum with no reagent, near-infrared spectroscopy was used. Applicability of analytical models was studied. Spectra of serum were measured by a FT-NIR spectrometer with 1, 2 and 6.5 mm optical path length respectively. Partial least-square (PLS) models were calibrated for cholesterol in combination, first overtone and second overtone spectral regions. Root mean square error of prediction (RMSEP) of these models is 0.15, 0.16 and 0.29 mmol · L⁻¹, and mean percent error of prediction (MPEP) is 2.9%, 3.1% and 4.8%, respectively. To validate applicability of these models, other two groups of serum spectra were measured in one month after the models were calibrated. These two sample groups were calculated using calibrated models. Prediction result of 1mm model is the best. The result of this experiment indicates that it's possible to calibrate precise models for cholesterol. If model is based on thinner samples, applicability is better.

Keywords NIR spectroscopy; Human serum analysis; PLS

(Received May 24, 2010; accepted Aug. 26, 2010)

* Corresponding author