

# 傅里叶变换近红外光谱结合偏最小二乘法测定 血清甘油三酯含量

董海胜<sup>1</sup> 张丽芬<sup>2</sup> 钟悦<sup>2</sup> 黄贱英<sup>1</sup> 陈斌<sup>1\*</sup>

(<sup>1</sup> 中国航天员科研训练中心航天医学基础与应用国家重点实验室, 北京 100094)

<sup>2</sup> 北京航天城门诊部, 北京 100094

**摘要** 以日常体检用血清样品为实验材料,以临床传统甘油磷酸氧化酶-过氧化物酶-4-氨基安替比林-酚法测定得到的甘油三酯含量作为参比值,同时采集了所有血清样品近红外光谱,采集范围为 12500~4000  $\text{cm}^{-1}$ ,结合偏最小二乘法建立了血清甘油三酯定量模型。实验结果表明,变量范围选取 10796.2~8246.6  $\text{cm}^{-1}$ ,采用多元散射校正对光谱进行预处理,所建立的血清甘油三酯偏最小二乘定量校正模型的相关系数  $R^2$  为 0.9454,交叉验证校正标准差(RMSECV)为 0.146;检验集预测标准差(RMSEP)为 0.151,相关系数  $R^2$  为 0.9068。结果表明,应用近红外光谱技术结合偏最小二乘法建立血清甘油三酯定量模型,用于未知血清样品甘油三酯含量的非破坏性检测是可行的。

**关键词** 光谱学;近红外光谱;偏最小二乘;血清;甘油三酯

中图分类号 O433.4 文献标识码 A doi: 10.3788/AOS201434.s130001

## Determination of Serum Triglyceride Content by Fourier Near Infrared Spectroscopy and Partial Least Squares

Dong Haisheng<sup>1</sup> Zhang Lifen<sup>2</sup> Zhong Yue<sup>2</sup> Huang Jianying<sup>1</sup> Chen Bin<sup>1</sup>

<sup>1</sup> State Key Laboratory of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center, Beijing 100094, China  
<sup>2</sup> Beijing Aerospace City Out-Patient Department, Beijing 100094, China

**Abstract** Taking the whole blood obtained from daily clinical inspection as experimental materials, triglyceride concentration of serum is determined by traditional clinical method; glycerol phosphate oxidase-peroxidase-4-amino antipyrine-phenol method. Near infrared spectroscopy of all serum samples are collected at the same time, with the range between 12500 ~ 4000  $\text{cm}^{-1}$ . A quantitative model of serum triglyceride concentration by near infrared spectroscopy is established by combining with partial least squares (PLS) method. Experimental results show that the proper variable range is 10796.2 ~ 8246.6  $\text{cm}^{-1}$ , and with the multiplicative scatter correction as spectrum pretreatment method, the quantitative PLS calibration model for serum triglyceride is established. The correlation coefficient  $R^2 = 0.9454$ , cross-validation calibration standard deviation (RMSECV) is 0.146 for the calibration set; the predicted standard deviation (RMSEP) is 0.151, and correlation coefficient  $R^2 = 0.9068$  for the test set. The results show that, by the application of near infrared spectroscopy combined with partial least squares method, the serum triglyceride quantitative model is successfully established with high accuracy and stability, which can be used for non-destructive testing of triglyceride content in the unknown serum samples.

**Key words** spectroscopy; near infrared spectroscopy; partial least square; serum; triglyceride

**OCIS codes** 170.1610; 300.6340

## 1 引言

甘油三酯(TG)是人体内的能量物质。血清甘

油三酯含量是临床生物化学检验的常规项目,是最常用的高血脂症筛查指标<sup>[1]</sup>。临床上常规化验一般

收稿日期: 2013-09-01; 收到修改稿日期: 2013-10-01

基金项目: 国家自然科学基金(31101251,81202612)

作者简介: 董海胜(1982—),男,硕士,助理研究员,主要从事航天营养代谢与食品工程方面的研究。

E-mail: dhs303@126.com

\* 通信联系人。E-mail: chenb12@aliyun.com

采用自动生化仪,需事先购置专用试剂,将试剂与血清进行定量反应,操作步骤繁杂,包括安装试剂、冲洗试剂针、加样针、搅拌器和比色杯等步骤,试剂不可重复利用,废弃试剂排放容易造成环境污染,不同样品或指标检测存在交叉污染风险<sup>[2]</sup>。此外,当进行多指标检验时,不同检测项目均消耗样品且不可重复使用,因此,需增加采样量或采集多份血清样品,势必增加患者身体负担。由于全血样品成分复杂,极不稳定,国际通用血液甘油三酯测定一般采用血清作为测试对象。国内亦有学者尝试采用近红外光谱结合偏最小二乘测定全血甘油三酯含量,经优化后的甘油三酯分析模型预测相关系数 0.865,预测均方根误差为 0.28<sup>[3]</sup>。

近红外光谱分析是一种非侵入、非破坏性的分析技术,已广泛应用于食品、药品、化工、农业及环境监测等行业<sup>[4-9]</sup>。基于上述优势,将近红外定量分析技术应用于日常临床生物化学大批量样品的检验分析工作具有很大的潜力。近年来,近红外光谱在生命科学领域中的应用越来越多。本文探讨近红外光谱在血清甘油三酯定量分析中适用性,以实现快

速、无试剂、非侵入式分析及减少血液的用量,并为无创生化检测提供依据。

## 2 实验材料与方法

### 2.1 样品来源

实验用血清均为门诊部(北京航天城门诊部)日常体检样品,共计 96 份。

### 2.2 仪器

MPA 傅里叶变换近红外光谱仪,德国 Bruker; A25 生化分析仪,西班牙 Biosystem; 600A 医用低速离心机,河北白洋离心机厂。

### 2.3 测试步骤

将体检用新鲜血液样品离心,取上清,立即用常规生化法测得样品甘油三酯含量真实值,然后所有样品进行近红外光谱采集。随机将样品分为校正集和检验集。校正集用于建立定量模型,检验集不参与建模,用于评价所建模型的性能。样品测试、建模及验证流程详见图 1。

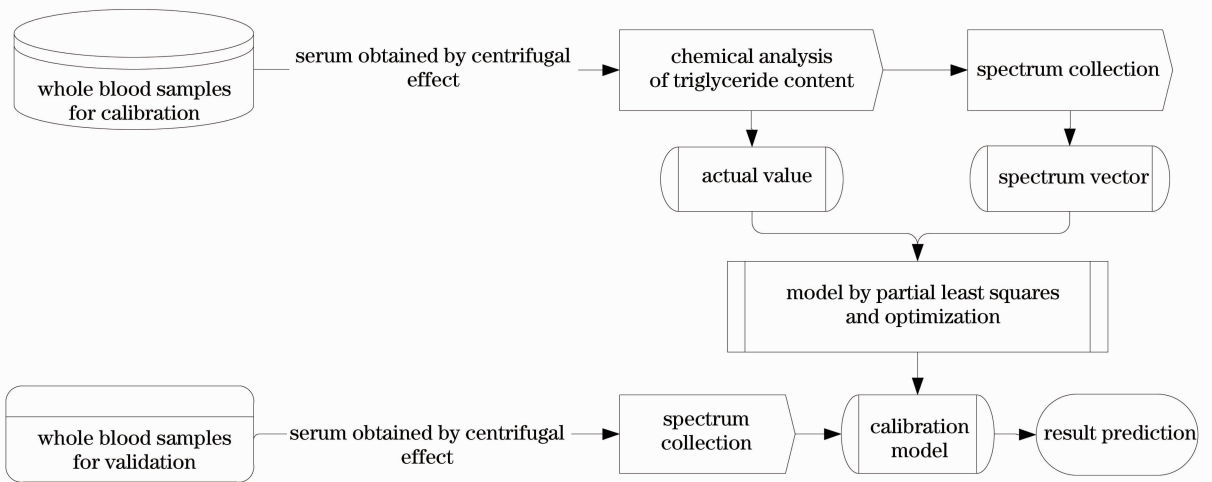


图 1 血清甘油三酯含量测试、建模及验证流程

Fig. 1 Serum triglyceride content testing, modeling and verification process

### 2.4 化学值测定

样品甘油三酯含量预先经全自动生化分析仪测定,测定值作为建模用真实值。将采集的新鲜血液样品离心,取上清按如下实验步骤测定:开机检查→安装试剂→冲洗管道→定标→质控→样品测定。测定方法为甘油磷酸氧化酶-过氧化物酶-4-氨基安替比林-酚法(GPO-PAP)。其反应原理如下:1)三酰甘油→甘油+游离脂肪酸;2)甘油+三磷酸腺苷(ATP)→甘油-3-磷酸;3)甘油-3-磷酸+O<sub>2</sub>→磷酸二羟丙酮+H<sub>2</sub>O<sub>2</sub>;4)H<sub>2</sub>O<sub>2</sub>+4-氨基比标(AAP)+

酚→醌亚胺+H<sub>2</sub>O。

### 2.5 光谱获取

血清样品约 0.5 mL,缓慢置于石英测试瓶中,避免气泡混入,将盛有样品的测试瓶置于恒温样品池中,应用 MPA 傅里叶变换近红外光谱仪采集光谱信息,采集方式为液体透射。光谱扫描范围 12500~4000 cm<sup>-1</sup>,光谱分辨率 32 cm<sup>-1</sup>,光谱取 20 次平均,样品温度为室温。

### 2.6 化学计量

应用 OPUS6.5 软件进行光谱预处理及定量模

型的建立。其中光谱预处理方法包括矢量归一化、多元散射校正、消除常数偏移量、最大最小归一化、减去一条趋势线等。定量模型的建立采用偏最小二乘法。模型性能采用交互验证标准偏差、相关系数及预测标准偏差进行评价。

### 3 结果与分析

#### 3.1 化学值测定结果

实验用血清样品随机分为校正集( $n=80$ )及检验集( $n=16$ )。样品甘油三酯浓度分布详如表 1 所示。

表 1 实验用血清甘油三酯浓度分布

Table 1 Concentration distribution of serum triglyceride of test

Groups	Concentration / (mmol·L <sup>-1</sup> )			
	Min	Max	Average	Standard error
Calibration set ( $n=80$ )	0.32	3.21	1.09	0.61
Test set ( $n=16$ )	0.44	2.09	1.08	0.51

#### 3.2 血清近红外光谱特征

血清样品近红外光谱扫描结果如图 2 所示,甘油三酯含有 C-H, O-H 等含氢基团,这些基团的组合频、一倍频及二倍频的特征谱区分别分布在 5200~4200 cm<sup>-1</sup>, 6700~5200 cm<sup>-1</sup> 和 10400~

8200 cm<sup>-1</sup>[10-11]。由于血清样品大部分组分是水分子,水分子在 7000 cm<sup>-1</sup>及 5000 cm<sup>-1</sup>附近有强烈吸收,掩盖了血清中其他成分的光谱信息,因此,建立定量模型时需要避开此谱区。

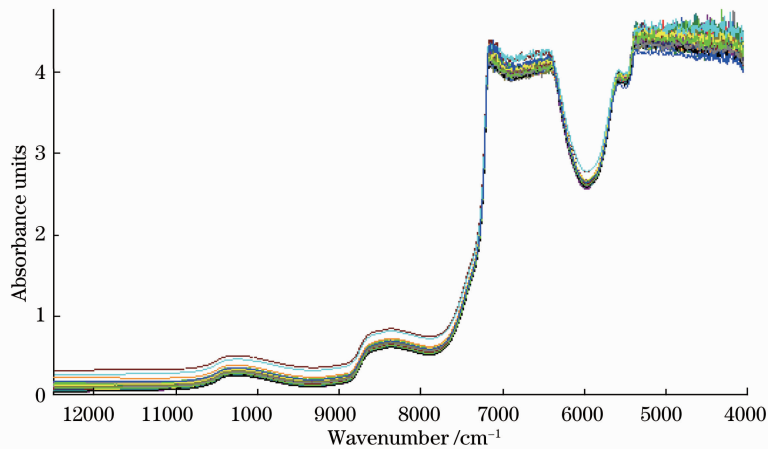


图 2 血清样品近红外光谱( $n=96$ )

Fig. 2 Near infrared spectroscopy of serum samples ( $n=96$ )

为建立血清甘油三酯预测模型,首先将光谱进行预处理,再建模,以交叉验证标准差、相关系数评价模型性能。其中光谱预处理方法综合考虑了光谱范围(变量范围)的选择,采用软件自动完成;定量模型的建立采用偏最小二乘法。血清甘油三酯定量校

正模型统计结果如表 2 所示。结果表明,光谱预处理方法以多元散射校正最优,主成份数(因子) $n=8$ ,交叉验证标准差(RMSECV)为 0.146,决定系数  $R^2$  为 0.9454,范围 10796.2~8246.6 cm<sup>-1</sup>。

表 2 不同光谱预处理方法模型性能

Table 2 Performance of model processed by different pretreatment methods

Spectrum pretreatment method	Factor	RMSECV	Variable ranges / cm <sup>-1</sup>
Multiplicative scatter correction	8	0.146	10796.2~8246.6
1 <sup>st</sup> derivative and subtract a trend line	9	0.265	11644.8~10792.3; 9099~8246.6
Constant offset elimination	9	0.346	8246.6~7394.2
1 <sup>st</sup> derivative and multiplicative scatter correction	6	0.363	9099~8246.6
1 <sup>st</sup> derivative and vector normalization	6	0.365	9099~8246.6
2 <sup>nd</sup> derivative	7	0.398	9099~8246.6
1 <sup>st</sup> derivative	7	0.428	12493.3~11644.8; 9099~7394.2

应用多元散射校正法在优化后的变量范围内建模,将校正集样品光谱代入模型计算,得到甘油三酯含量的预测值,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 3 所示。模型维数与决定系数及交叉验证标准偏差的关系分别如图 4 及图 5 所示。

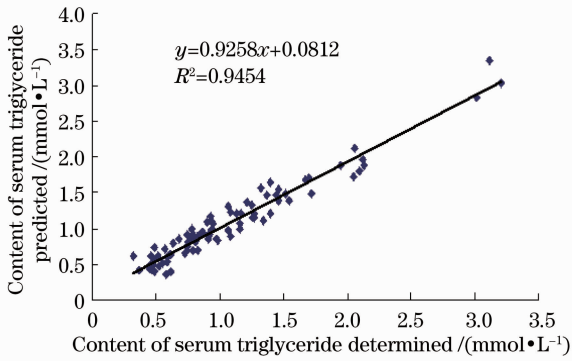


图 3 校正集化学测定值与预测值散点图  
Fig. 3 Scatter of predictive value and chemical values in calibration set

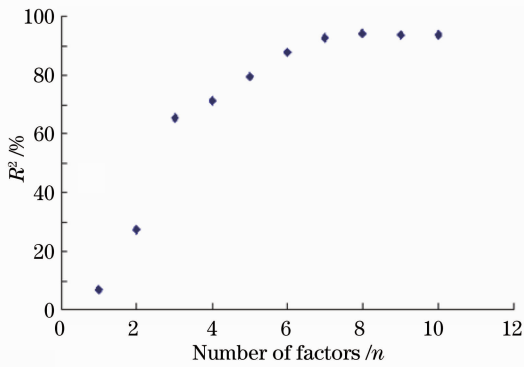


图 4 校正模型维数与决定系数的关系  
Fig. 4 Relation between  $R^2$  and the factors of calibration model

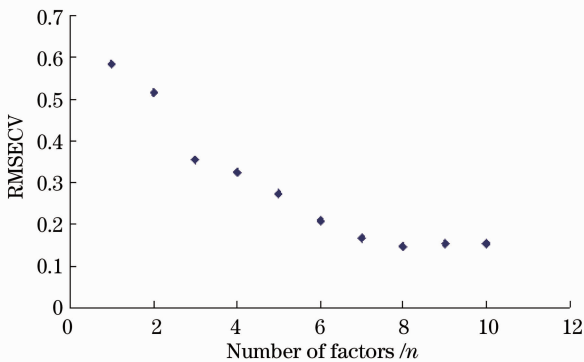


图 5 模型维数与交叉验证标准偏差的关系  
Fig. 5 Relation between RMSECV and factors of calibration model

### 3.3 预测模型评价

对所建立的甘油三酯定量模型进行初步评价,验证模型的适用性,将未参与建模的样品光谱代入模型计算(样品甘油三酯的含量真值预先通过生化分析仪测得,样品的光谱测量条件与建模样品的条件相同),得到甘油三酯含量的预测值,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 6 所示。决定系数  $R^2 = 0.9068$ , 预测标准差 (RMSEP) 为 0.151。

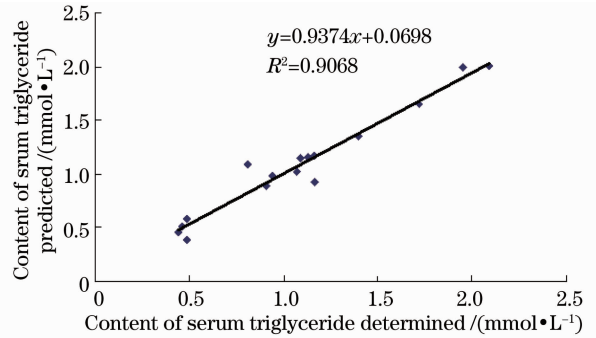


图 6 检验集化学测定值与预测值散点图  
Fig. 6 Scatter of predictive value and chemical values in test set

## 4 结 论

以常规生化检验法测得的血清甘油三酯浓度为参考真值,利用近红外光谱扫描血清的光谱,建立了甘油三酯近红外光谱定量模型。模型 RMSECV 为 0.146, 决定系数  $R^2$  为 0.9454。初步验证了模型适用性,检验集 RMSEP 为 0.151, 决定系数  $R^2$  为 0.9068。优于文献建立的全血甘油三酯含量定量模型。因此,采用适宜的波段(变量)和光谱预处理方法,应用偏最小二乘建模可以对实现血清甘油三酯浓度预测。本次实验对近红外光谱快速检测血清甘油三酯含量进行了初探,后续实验中,将增加样品量,依据医学背景,同时建立其他相关指标的定量模型,以期达到同时快速综合评定的目的。

### 参 考 文 献

- 1 Ren Jie, Zhao Dong. Present status of the epidemiology of serum lipid level in the Chinese population [J]. Chinese Journal of Contemporary Neurology and Neurosurgery, 2009, 9(1): 8-11.  
任 洁, 赵 冬. 中国人群血脂水平的流行病学现状[J]. 中国现代神经疾病杂志, 2009, 9(1): 8-11.
- 2 Wang Cong, Zhao Fengmei, Wang Huihui. Cross-contamination of serum creatinine after detection of serum phosphorus by automatic biochemical analyzer [J]. Chinese Journal of Misdiagnostics, 2011, 11(32): 7874.  
王 聪, 赵凤梅, 王慧慧. 全自动生化分析仪测定血清磷后对血清肌酐检测的交叉污染[J]. 中国误诊学杂志, 2011, 11(32):

7874.

- 3 Huang Furong, Luo Yunban, Zheng Shifu, *et al.*. Near-infrared spectroscopic analysis and model optimization on cholesterol and triglyceride in whole blood [J]. *Acta Optica Sinica*, 2011, 31(10): 1030001.  
黄富荣, 罗云瀚, 郑仕富, 等. 全血胆固醇、甘油三酯近红外光谱分析与模型优化[J]. *光学学报*, 2011, 31(10): 1030001.
- 4 Sun Guangming, Liu Fei, Zhang Fan, *et al.*. Fast determination of praline 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.
- 5 G Russell Warnick, Mary M Kimberly, Parvin P Waymack, *et al.*. Standardization of measurements for cholesterol, triglycerides, and major lipoproteins [J]. *Labmedicine*, 2008, 39(8): 482–490.
- 6 Fengxia Wang, Zhuoyong Zhang, Xiujun Cui, *et al.*. Identification of rhubarbs by using NIR spectrometry and temperature-constrained cascade correlation networks [J]. *Talanta*, 2006, 70(5): 1170–1176.
- 7 Qian Aiping, Hua Guoran, Qian Zhiyu, *et al.*. Therapy effect assessment by functional near infrared spectroscopy during laser induced interstitial thermotherapy [J]. *Chinese J Lasers*, 2011, 38(1): 0104003.  
钱爱平, 花国然, 钱志余, 等. 激光间质热疗的功能近红外疗效评估[J]. *中国激光*, 2011, 38(1): 0104003.
- 8 Li Wang, Frank S C Lee, Xiaoru Wang. Near-infrared spectroscopy for classification of licorice (*Glycyrrhiza uralensis* Fisch) and prediction of the glycyrrhizic acid (GA) content [J]. *LWT-food Science and Technology*, 2007, 40(1): 83–88.
- 9 Chi-On Chan, Ching-Ching Chu, Daniel Kam-Wah Mok, *et al.*. Analysis of berberine and total alkaloid content in Cortex *Phellodendri* by near infrared spectroscopy (NIRS) compared with high-performance liquid chromatography coupled with ultraviolet spectrometric detection [J]. *Analytica Chimica Acta*, 2007, 592(2): 121–131.
- 10 Ching-Ching Lau, Chi-On Chan, Foo-Tim Chau, *et al.*. Rapid analysis of *Radix puerariae* by near-infrared spectroscopy [J]. *J Chromatography A*, 2009, 1216(11): 2130–2135.
- 11 Zhu Xiaoli. *Molecular Spectroscopy Analytical Technology Combined with Chemometrics and Its Applications* [M]. Beijing: Chemistry Industry Press, 2011. 262–263.  
褚小立. *化学计量学方法与分子光谱分析技术* [M]. 北京: 化学工业出版社, 2011. 262–263.

栏目编辑: 韩峰