

FTIR/ATR 光谱应用于人体血液血红蛋白的快速定量分析

尹浩^{a,b} 潘涛^{a,b} 田佩玲^c 韩筠^d 韦相才^c
张清健^c 方俊宇^c 钟泉^d 冯华^d

a["重大工程灾害与控制"教育部重点实验室(暨南大学) 广州市天河区黄埔大道西 601 号 510632]

b(暨南大学光电工程系 广州市天河区黄埔大道西 601 号 510632)

c(广东省计划生育科学技术研究所 广州市越秀区梅东路 17 号 510600)

d(暨南大学数学系 广州市天河区黄埔大道西 601 号 510632)

摘 要 利用傅里叶变换红外光谱(FTIR)和衰减全反射(ATR)技术,建立了人体血液血红蛋白(HGB)的快速定量分析方法。采集人体全血样品 8 个,用常规化学方法测定血红蛋白浓度作为光谱校正模型的参考化学值。每个样品用蒸馏水溶血,分别配制成 2 倍、3 倍、4 倍、5 倍、6 倍稀释的溶血液样品,和全血样品一起共得到 6 组 48 个样品用于光谱测定。基于 11 点 Savitzky-Golay 平滑的二阶导数光谱,采用 MLR 分析和交叉检验对每组样品分别建立定量模型。每组分别采用全谱(4000—600 cm^{-1})、指纹领域(1800—800 cm^{-1})建立 MLR 模型,并建立每个波数的一元线性回归模型,从中遴选效果最好的单点模型。结果表明,每组样品的最优单点模型都有良好的预测效果。直接测定的全血样品组的最优单点模型的采用波数、交叉检验均方误差(RMSECV)、相对交叉检验均方误差(RRMSECV)、预测相关系数(R_p)分别为 1759 cm^{-1} 、4.9 $\mu\text{g/L}$ 、3.6%、0.825。

关键词 傅里叶变换红外光谱, 衰减全反射, 血液, 血红蛋白。

中图分类号: O657.33 文献标识码: A 文章编号: 1004-8138(2009)02-0431-06

1 前言

中红外光谱分析是一种既可以分析物质结构又可以测定物质含量的方法。它具有测定时间短,精度高,并且可以进行多指标同时测定、非破坏性测定、连续测定等优点,是在线(On Line)、实时(Real Time)、原位(In Situ)定量分析测定的方法之一。中红外光谱用于生物体系和生物过程分析的主要难点是水分子或有机溶剂的强烈的红外吸收干扰了待测组分的红外光谱。傅里叶变换红外光谱(FTIR)和衰减全反射(ATR)技术的提出^[1],克服了上述困难,已经在农产品、食品的品质测定^[2],细胞代谢计测^[3,4],发酵过程监测^[5],酶活性分析^[6]等方面有广泛应用。

人体血液血红蛋白(HGB)含量的测定是评价人群贫血患病率和铁营养状态的基本方法之一。另外,地中海贫血是我国南方地区发病率很高的一种遗传性血红蛋白病,目前尚无有效的治疗手段,建立地中海贫血人群筛查和婚前孕产前诊断是最根本的预防地中海贫血患儿出生的方法。血

国家自然科学基金(10771087);广东省科技计划粤港关键领域重点突破项目(2007A020905001);

广东省科技计划项目(2007B030501008;2007B020714001);广州市科技攻关项目(2007Z3-E0281);

教育部留学归国人员科技启动基金(2005-383)1)资助

联系人,电话:(020)85220484 转 413;手机:(0)13660337400;E-mail: tpan@jnu.edu.cn

作者简介:尹浩(1973—),男,广东省江门市人,讲师,博士研究生,主要从事生物医学信息技术、光谱技术方面研究。

收稿日期:2008-12-31;接受日期:2009-01-19

血红蛋白含量是地中海贫血症血液筛查的一个基本指标。但是现有的测定方法基于生化分析,需要预处理和化学试剂,耗费检测成本且不利于环保,还需要专业人员的操作,不适用于大人群的筛查,也不便于在基层医院、社区推广。因此,建立无需化学试剂、直接、快速定量分析人体血红蛋白浓度的方法是很有研究价值和实际意义的。目前国内外还未见有利用 FTIR/ATR 光谱技术直接快速定量测定人体血液血红蛋白浓度的工作报告。

本文利用傅里叶变换红外光谱和衰减全反射技术,探讨建立人体血液血红蛋白的直接快速定量分析方法。常规方法测定血红蛋白浓度时,首先要用蒸馏水对血浆进行溶血,然后用生化试剂进行化学反应再来测定。蒸馏水溶血是一种简单方便的样品预处理,其作用是缓解血液中各种成分的相互作用,减少光谱中其他成分对于血红蛋白的干扰,从而有可能得到更好的预测效果。为此,本文除了直接测定全血样品,也考虑通过测定不同稀释配比的溶血液样品来换算原全血样品的血红蛋白浓度,探讨建立适当的测定方法。采用多元线性回归(MLR)和交叉检验(Cross-validation)对每一样品组分别建立校正模型。每组分别采用全谱、指纹领域建立 MLR 模型,并建立每个波数的一元线性回归模型,从中遴选效果最好的单点模型。探讨利用 FTIR/ATR 光谱技术直接快速定量测定人体血液血红蛋白浓度的可行性和测定误差。

2 实验与方法

2.1 实验材料、仪器和测量方法

广州市某研究所提供的人体全血样品 8 个。用 BC-3000Plus 全自动血液细胞分析仪测定样品的血红蛋白浓度,作为光谱定标的参考化学值。用蒸馏水溶血,每个样品都分别配制成 2 倍、3 倍、4 倍、5 倍、6 倍稀释的溶血液样品,和全血样品一起共得到 6 组 48 个样品用于光谱测定。样品血红蛋白化学值的数据分析参见表 1。

表 1 各组血液样品血红蛋白化学值的数据分析

血液样品	HGB (g/L)			
	最小值	最大值	平均值	标准差
全血	122.0	148.0	135.0	8.2
2 倍溶血	61.0	74.0	67.5	4.1
3 倍溶血	40.7	49.3	45.0	2.7
4 倍溶血	30.5	37.0	33.8	2.1
5 倍溶血	24.4	29.6	26.9	1.6
6 倍溶血	20.3	24.7	22.5	1.4

光谱实验仪器采用德国 Bruker 公司的 VERTEX 70 型傅里叶变换红外光谱仪(FTIR)和英国 Specac 公司的水平衰减全反射(ATR)样品测量附件。水平 ATR 的样品槽为 ZnSe 晶体,45° 入射角,3 次反射。光谱采集范围为 4000—600 cm^{-1} ,探测器为 KBr,扫描次数为 32,分辨率为 4 cm^{-1} 。实验室温度为 25±1,湿度为 46% RH。

采集光谱时,取 0.075mL 的血液样品置于 ATR 样品槽中,用衰减全反射方式获得样品的 FTIR/ATR 吸收光谱。每个样品重复采集 3 次光谱,计算平均光谱。

2.2 定量方法与模型评价指标

首先进行光谱预处理。合理的光谱平滑化和光谱求导(一阶、二阶导数光谱等)处理可以消除基线飘移和平缓背景干扰形成的高频随机噪声的影响,从而提供比原光谱更高的分辨率和更清晰的光谱轮廓变化,提高光谱信噪比^[7-9]。本文采用 11 点平滑的 Savitzky-Golay 方法计算二阶导数谱,其公式如下:

$$d^2A(\nu) = \frac{1}{429d\nu^2} \{15[A(\nu+5\Delta\nu) + A(\nu-5\Delta\nu)] + 6[A(\nu+4\Delta\nu) + A(\nu-4\Delta\nu)] - [A(\nu+3\Delta\nu) + A(\nu-3\Delta\nu)] - 6[A(\nu+2\Delta\nu) + A(\nu-2\Delta\nu)] - 9[A(\nu+\Delta\nu) + A(\nu-\Delta\nu)] - 10A(\nu)\} \quad (1)$$

其中: $\nu(\text{cm}^{-1})$ ——波数; $A(\nu)$ ——吸光度; $\Delta\nu$ ——光谱数据的波数间隔, 在本实验中 $\Delta\nu = 1.9285(\text{cm}^{-1})$ 。

采用多元线性回归和交叉检验对全血、2 倍溶血、3 倍溶血、4 倍溶血、5 倍溶血、6 倍溶血的 6 个样品组分别建立定量模型。其数学关系如下:

$$d^2A\nu = C_{\text{HGB}}d^2A_{\text{HGB}}(\nu) + \epsilon \quad (2)$$

其中: $d^2A\nu$ ——样品在波数 ν 的吸光度的二阶导数值; $d^2A_{\text{HGB}}(\nu)$ ——血红蛋白成分在波数的单位浓度吸光度的二阶导数值; C_{HGB} ——样品中血红蛋白的浓度; ϵ ——血红蛋白以外的其他成分对光谱的吸收干扰。

基于每个波数的各已知样品的吸光度和血红蛋白浓度的化学值数据, 根据关系式(2), 用线性回归方法确定血红蛋白在每个波数的单位吸光度的二阶导数值, 得到血红蛋白的单位吸光度的二阶导数光谱。这样校正后, 再基于血红蛋白的单位吸光度的二阶导数光谱和未知样品的二阶导数光谱, 根据关系式(2), 用线性回归方法确定未知样品中血红蛋白浓度的预测值。然后通过比较预测值与化学值来评价预测效果。在上述计算过程中, 使用了 Leave-One-Out 的交叉检验来计算每个样品的预测值。

在计算预测值的环节, 分别采用三种模式: (1) 采用全谱段($4000\text{—}600\text{cm}^{-1}$)的数据, 用多元回归计算预测值; (2) 采用指纹领域($1800\text{—}800\text{cm}^{-1}$)的数据, 用多元回归计算预测值; (3) 分别采用全谱中每一个波数的数据, 用单元回归计算预测值, 再从中遴选效果最好的波数点, 得到相应的预测值。

模型评价指标包括预测相关系数(R_p)、交叉检验均方误差(RMSECV)、相对交叉检验均方误差(RRMSECV%), 其中:

$$R_p = \frac{\sum_{i=1}^n (C_{\text{HGB},i} - \bar{C}_{\text{HGB}})(C_{\text{HGB},i} - \bar{C}_{\text{HGB}})}{\sqrt{\sum_{i=1}^n (C_{\text{HGB},i} - \bar{C}_{\text{HGB}})^2 (C_{\text{HGB},i} - \bar{C}_{\text{HGB}})^2}} \quad (3)$$

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (C_{\text{HGB},i} - C_{\text{HGB},i})^2}{n}}; \text{RRMSECV} = \frac{\text{RMSECV}}{C_{\text{HGB}}} \times 100(\%) \quad (4)$$

其中: $C_{\text{HGB},i}$, $C_{\text{HGB},i}$ ——分别为第 i 样品的血红蛋白浓度的预测值和化学值; $\bar{C}_{\text{HGB},i}$ ——各组样品的血红蛋白浓度预测值的平均值; \bar{C}_{HGB} ——各组样品的血红蛋白浓度化学值的平均值; n 是各组中样品的总数。

3 结果与讨论

图 1 为全血、2 倍溶血、3 倍溶血、4 倍溶血、5 倍溶血、6 倍溶血的 6 个样品组 48 个样品的 FTIR/ATR 光谱。在 3290cm^{-1} 、 1637cm^{-1} 附近对应水分子的强烈吸收。在指纹区域($1800\text{—}800\text{cm}^{-1}$), 不同样品有明显不同吸收, 这部分区域包含了丰富的血液中(水分以外的)血红蛋白等主要成分的吸收信息。

基于样品的二阶导数光谱和参考化学值, 采用多元线性回归和交叉检验对全血、2 倍溶血、3 倍溶血、4 倍溶血、5 倍溶血、6 倍溶血的 6 个样品组分别建立定量模型。按照本文 2.2 节的方法, 计算

全谱段MLR 模型、指纹领域MLR 模型、每个波数点的单点模型的预测结果。在每个单点模型中,综合考虑按照交叉检验均方误差最小和预测相关系数最大两方面的优选目标,并侧重于交叉检验均方误差最小来综合评价,遴选最好的单点模型,找到最优的波数点。以全血样品组的单点模型为例,如图 2 所示,给出所有单点模型对应的交叉检验均方误差和预测相关系数,经过综合考虑,确定波数 1759cm^{-1} 的单点模型对应的预测效果最好。其余类似,也可以找到的 2 倍溶血、3 倍溶血、4 倍溶血、5 倍溶血、6 倍溶血等 5 个样品组的最优单点模型。全血、2 倍溶血、3 倍溶血、4 倍溶血、5 倍溶血、6 倍溶血的 6 个样品组的全谱段MLR 模型、指纹领域MLR 模型和最优单点模型的预测效果汇总在表 2 中。

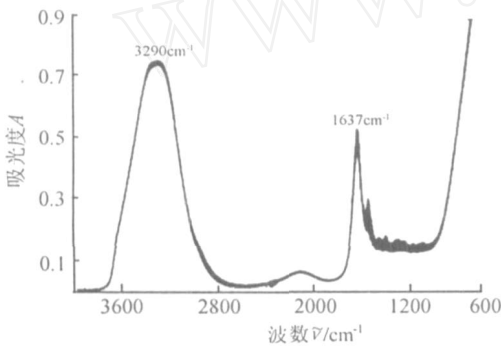


图 1 48 个样品的 FTIR/ATR 光谱

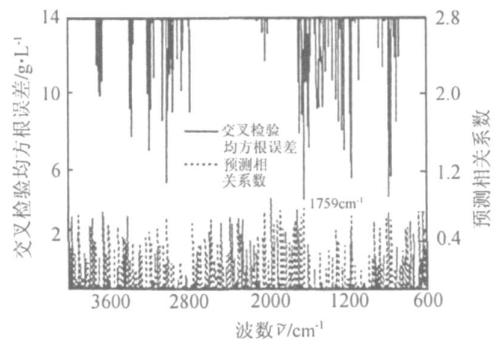


图 2 每个单点模型对应的预测相关系数和交叉检验均方误差

从表 2 可以看出,每一组样品的最优单点模型都有良好的预测效果。最优单点模型的预测效果明显优于全谱、指纹谱两类模型,相对交叉检验均方误差都在 5% 以内。全谱区 ($4000\text{—}600\text{cm}^{-1}$) 模型与指纹领域 ($1800\text{—}800\text{cm}^{-1}$) 模型的预测效果没有显著的差别。但指纹谱区能否代替全谱区得到同样的效果还有待进一步证实。在最优单点模型中,采用的波数 1759 、 1718 、 943cm^{-1} 位于指纹谱区内, 3188 、 3003 、 3375cm^{-1} 在指纹谱区之外,它们都没有落在明显的吸收峰顶处,这说明血液这种具有复杂背景和多成分互相干扰的样品,即使是中红外光谱也不能单纯用吸收峰的信息来简单计算。

图 3 给出全血、2 倍溶血、3 倍溶血、4 倍溶血、5 倍溶血、6 倍溶血的 6 个样品组全部 48 个样品的最优预测值与化学值的比较。全部样品的相关性很好,整体相关系数达到 0.998。并且在最优单点模型中,直接测定的全血样品模型与其他溶血后测定的模型相比预测效果没有显著差别,这说明采用 FTIR/ATR 光谱直接定量分析人体血液的血红蛋白浓度是可行的。

收集更多的样品建模,遴选高信噪比和相对稳定的特征波数,采用多个特征波数联合建模是进一步提高模型预测精度的关键。对于全谱、指纹谱或者其他部分谱模型要采用偏最小二乘法(PLS)方法消除干扰等,这些是本文下一步需要开展和完善的研究方向。

4 结论

分别对全血、2 倍、3 倍、4 倍、5 倍、6 倍稀释的溶血液样品组进行 FTIR/ATR 光谱测定。采用

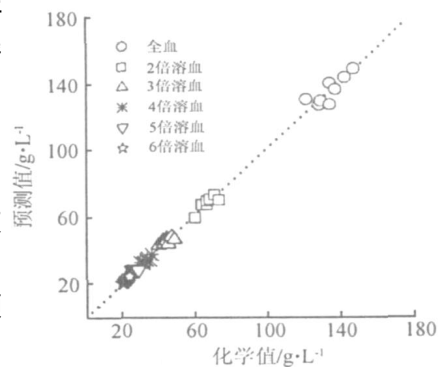


图 3 最优单点模型对应的预测值和化学值的比较

MLR 分析和交叉检验对每组样品分别建立定量模型。每组都分别采用全谱、指纹领域建立 MLR 模型, 并建立每个波数的一元线性回归模型, 从中遴选效果最好的单点模型。结果表明, 每组样品的最优单点模型都有良好的预测效果。直接测定的全血样品组的最优单点模型的采用波数、交叉检验均方误差、相对交叉检验均方误差、预测相关系数分别为 1759cm^{-1} 、 4.9g/L 、 3.6% 、 0.825 。从而表明, 利用 FTIR/ATR 光谱技术直接快速定量测定人体血液血红蛋白浓度是可行的, 并具有较高的分析精度。

表 2 各样品组的全谱模型、指纹领域模型和最优单点模型的预测效果对比

血液样品	采用 $4000\text{--}600\text{cm}^{-1}$		采用 $1800\text{--}800\text{cm}^{-1}$		最优单点模型			
	RM SECV (g/L)	RRM SECV (%)	RM SECV (g/L)	RRM SECV (%)	采用波数 (cm^{-1})	RM SECV (g/L)	RRM SECV (%)	R_p
全血	14.2	10.5	14.7	10.9	1759cm^{-1}	4.9	3.6	0.825
2倍溶血	3.8	5.7	3.8	5.7	3188cm^{-1}	2.4	3.5	0.830
3倍溶血	3.3	7.4	3.1	6.8	1718cm^{-1}	2.1	4.6	0.655
4倍溶血	2.0	5.8	1.9	5.6	943cm^{-1}	1.1	3.3	0.844
5倍溶血	1.5	5.4	1.9	7.0	3003cm^{-1}	1.3	4.7	0.729
6倍溶血	1.3	5.7	1.3	5.9	3375cm^{-1}	0.5	2.2	0.941

参考文献

- [1] Mirabella FM, Harrick NJ. *Internal Reflection Spectroscopy: Review and Supplement* [M]. New York: Harrick Scientific Corporation, 1985.
- [2] Wilson RH. Fourier Transform Mid-Infrared Spectroscopy for Food Analysis[J]. *Trends in Analytical Chemistry*, 1990, 9(2): 127—131.
- [3] Nakanishi K, Hashimoto A, Pan T *et al* Mid-Infrared Spectroscopic Measurement of Ionic Dissociative Materials in Metabolic Pathway[J]. *Applied Spectroscopy*, 2003, 57(12): 1510—1516.
- [4] Pan T, Hashimoto A, Kanou M *et al* Development of a Quantification System of Ionic Dissociative Metabolites using an FT-IR/ATR Method[J]. *Bioprocess and Biosystems Engineering*, 2003, 26(2): 133—139.
- [5] Axlens PH, Braddock WD, Brockman HL *et al* Use of Internal Reflectance Infrared Spectroscopy for in Situ Study of Supported Lipid Monolayers[J]. *Applied Spectroscopy*, 1995, 49(5): 526—531.
- [6] Fujii T, Miyahara Y. Infrared ATR Spectroscopy of Substrates in Aqueous Solution using Cryoenrichment and Its Application in Enzyme Activity Assays[J]. *Applied Spectroscopy*, 1998, 52(2): 128—133.
- [7] Savitzky A, Golay MJE. Smoothing and Differentiation of Data by Simplified Least Squares Procedures[J]. *Analytical Chemistry*, 1964, 36(8): 1627—1637.
- [8] 陈洁梅, 潘涛, 陈星旦. 二阶导数光谱预处理在用 FTIR/ATR 方法定量测定葡萄糖-6-磷酸和果糖-6-磷酸中的应用[J]. *光学精密工程*, 2006, 14(1): 1—7.
- [9] 潘涛, 陈洁梅. 解离性代谢物和 pH 值同时定量分析的 FT-IR/ATR 光谱方法[J]. *光谱学与光谱分析*, 2008, 28(10): 31—32.

Rapid Quantitative Analysis for the Human Blood Hemoglobin Applied through FTIR/ATR Spectrum

YIN Hao^{a,b} PAN Tao^{a,b} TIAN PeiLing^c HAN Yun^d WEI Xiang-Cai^c
ZHANG Qing-Jian^c FANG Jun-Yu^c ZHONG Quan^d FENG Hua^d

a (Key Laboratory of Disaster Forecast and Control in Engineering, Ministry of Education of the People's Republic of China (Jinan University), Guangzhou 510632, P. R. China)

b (Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, P. R. China)

c (Family Planning Research Institute of Guangdong, Guangzhou 510600, P. R. China)

d (Department of Mathematics, Jinan University, Guangzhou 510632, P. R. China)

Abstract The rapid quantitative analysis method of the human blood hemoglobin (HGB) was

developed by a Fourier transform infrared (FTIR) spectrometer and attenuated total reflection (ATR) techniques. Eight samples of the human blood were collected, hemoglobin concentration was measured by the conventional chemical method, and it was as the reference chemical value of the calibration model for the spectrum. Each sample with distilled water hemolysis, were configured to 2 times, 3 times, 4 times, 5 times, 6 times dilute hemolytic solution sample respectively, and the whole blood samples had been together a total of 6 groups of 48 samples for spectrometry. Based on the second derivatives of the spectra were calculated by using 11 points Savitzky-Goray smoothing, the MLR analysis and the cross-validation, the quantitative models of each sample group were established respectively. To each group the MLR model was established by using the whole region (4000—600 cm^{-1}) and the fingerprint region (1800—800 cm^{-1}) respectively. The linear regression model of the each wavenumber was established, and the best single-point model was selected from above models. The results showed that the best single-point models of each sample group had a good prediction effect. To the best single-point model for the whole blood sample group which by direct determination, the adopting wavenumber, the root mean square error cross validation (RMSECV), the relative root mean square error cross validation (RRMSECV), the prediction correlation coefficient (Rp) were 1759 cm^{-1} , 4.9 g/L , 3.6%, 0.825, respectively.

Key words FTIR, ATR, Blood, Hemoglobin.

欢迎您投稿“高效、保质”的中文核心期刊 《光谱实验室》

这是您的发明、发现获得“优先权”的可靠保障!

发表周期多数(50% - 75%)为 4—8 个月,

少数(20% - 45%)为 1—4 个月,个别(0- 7%)为 15—30 天

及时发表科技论文,是尽早实现其社会效益的前提,也是作者创造性劳动得到尊重、为在世界上取得“优先权”的可靠保障,因为发明、发现的“优先权”通常是以出版时间为准的。因此,本刊把尽快发表作者的论文,视为自己的神圣职责。

确保论文质量是论文早日发表的条件。作者发表论文总是要反映自己在工作有所发明、有所发现和有所创造的成绩,而不是去暴露自身的“缺欠”和“毛病”;换言之,作者发表论文总是要为自己争“光”,而不是自己要出“丑”。因此,作者投稿之前,除了自己要反复检查外,一定要多请您周围的同事、专家挑“毛病”,把“毛病”消灭在投稿之前,再投本刊才能发表得快。如果本刊挑出毛病,再请作者修改,反复“折腾”,不仅消耗双方精力,而且必然延长发表时间。保证质量的基本要求就是论文要做到“齐、清、定”。“齐”即全稿包括表、图和照片等齐全;“清”即文字图片打印(书写)清楚,不得有模糊不清的文字、图片和数字,段落要分明,便于排版和校对;“定”即做到稿件内容完整,在编辑过程中无须增删修改。

来稿请用 Word 或北大方正排版,用电子邮件发到本部电子信箱[E-mail: (1) gpsys@263.net; (2) gpsysh@public.sti.ac.cn; (3) gps@chinajournal.net.cn; (4) gpsys@periodicals.net.cn]。为避免某一电子信箱的服务器发生故障而延误收稿,建议作者向本刊两个或两个以上信箱同时发送电子邮件。

本刊收到作者来稿后,都会在 16 小时(遇公休日顺延)内发出“关于收到稿件的通知”。因此,作者发送稿件后 5 日以上都没有消息,一定要及时来电查询。

一篇论文出版,常常需要反复沟通“作者—编辑部—审者—编辑部—作者”之间的联系,其中与作者的联系是最重要的一环,一旦脱节,必然中断编辑过程。因此作者来稿时,务必将联系人的详细地址、办公室电话、手机号码、传真号码和电子信箱等(通讯方式要尽可能全)告诉编辑部,以便能及时与您联系。否则,由此而耽误出版由作者自己负责。

《光谱实验室》编辑部