

# 近红外光谱方法预测生物柴油主要成分

孔翠萍\* 褚小立 杜泽学 陆婉珍

(石油化工科学研究院, 北京 100083)

**摘要** 采用近红外光谱快速测定法对生物柴油的成分(脂肪酸甲酯、单甘酯、二甘酯、三甘酯和甘油)进行了研究。采用气相色谱方法获得其成分的基础数据,通过偏最小二乘方法与近红外光谱数据进行回归运算,分别建立以文冠果油生物柴油为例的单原料油校正模型及多种原料油生物柴油的混合校正模型,并以花椒油生物柴油为例考察了校正模型的适用性。结果表明:通过偏最小二乘方法可以建立适合多种原料油生物柴油的通用校正模型。对于新型生物柴油,向校正集中添加 10 个以上样本,扩充校正模型后,便可较为准确地测定这类新生物柴油样本的成分含量。此方法分析速度快、成本低、操作便捷、重复性好,适合于生物柴油生产过程的中间控制分析。

**关键词** 近红外光谱; 生物柴油; 化学计量学; 甲酯

## 1 引言

动植物油脂通过酯交换反应得到生物柴油,它是一种长链脂肪酸的单烷基酯,工业上应用的主要是脂肪酸甲酯。生物柴油不含硫和芳烃,十六烷值高,润滑性能好<sup>[1]</sup>,是一种优质清洁燃料。目前,很多国家强制使用生物柴油,促进了生物柴油产业的发展,其产量逐年迅速增长<sup>[2]</sup>。

动植物油脂主要是各种脂肪酸甘油三酯的混合物,在与甲醇的反应过程中会生成脂肪酸单甘酯、二甘酯等中间产物和副产物甘油,影响生物柴油的产率及品质。因此,在生产过程中,需要及时准确地测定脂肪酸甲酯、单甘酯、二甘酯、三甘酯和甘油的含量,以便精确控制反应过程及产品质量。目前,对这些组分的测定均采用气相色谱方法<sup>[3,4]</sup>,分析前需对样品进行衍生化(如硅烷化)处理,定性和定量过程较为复杂,操作难度大,分析时间长,不适用于生物柴油生产过程的中间控制分析。

近红外光谱法(NIR)不需要对样品进行预处理,操作简单,分析快速,非常适合油品的定量和定性分析<sup>[5]</sup>,石油化工业中得到了较为广泛的应用<sup>[6~10]</sup>。已有将 NIR 光谱用于测定生物柴油中甲酯、甲醇和甘油等成分的报道<sup>[11~14]</sup>,均得到了满意结果。但这些研究所用样本的原料相对单一,建模样本较少,所测定的成分也不完全,且大都是实验室合成的样品,不具有实际的工业应用价值。

本研究从我国多个中试和工业装置中收集了 200 余个生物柴油样本,涉及 6 种原料油及多种加工工艺,其中包括来源复杂的地沟油和酸化油,样本的覆盖范围广,所得的结果更具参考性,建立的校正模型具有实际应用价值。

## 2 实验部分

### 2.1 生物柴油样品和仪器

收集了 224 个生物柴油样品,从反应器取样后取其甲酯相,并蒸馏除去(120 °C)其中的甲醇及少量的水,置于 20 mL 密封小瓶中保存。其原料油来源分别为花椒油(47 个)、苦山杏油(16 个)、文冠果油(44 个)、地沟油(40 个)和酸化油(石炼厂 43 个、中粮集团 34 个)。

6890 气相色谱仪(美国安捷伦公司),氢火焰离子化检测器(FID),程序升温冷柱头柱上进样系统,UltraA lby-HT1 不锈钢毛细柱。Antaris II FT-NIR 光谱仪(美国 Thermo Scientific 公司)。

2009-11-26 收稿; 2010-01-27 接受

本文系中石化科研资助项目(N0 S108071)

\* E-mail: lingdrean@gmail.com

## 2.2 基础数据测定

采用气相色谱方法<sup>[4]</sup>测定生物柴油成分的基础数据, 色谱条件: 分别采用油酸甲酯、单油酸甘油酯、1,3-二油酸甘油酯、三油酸甘油酯作为定性和定量标准。采用双内标法进行定量分析, 以正十五烷为内标测定甲酯含量, 以三癸酸甘油酯为内标测定甘油、单甘酯、二甘酯和三甘酯的含量。

以空白样品池为参比, 采集生物柴油样品的 N R 谱图, 光谱范围 4000~10000  $\text{cm}^{-1}$ , 分辨率 8  $\text{cm}^{-1}$ , 扫描次数 128。

## 2.3 数据分析

采用石油化工科学研究院编制的“化学计量学软件 3.0 版”在 PC 计算机上处理光谱数据。用 K-S 方法将生物柴油样品分为校正集和验证集, 其中校正集用于建立分析模型, 验证集用于检测模型的准确性。采用马氏距离、光谱残差和最邻近距离 3 个指标判断模型对未知样本的适用性。

校正集的光谱经一阶微分处理后分别与气相色谱测定的成分数据(甘油、甲酯、单甘酯、二甘酯以及三甘酯的质量分数)通过偏最小二乘方法(PLS1)进行回归运算, 建立校正模型, PLS 所用的最佳主因子数由留一法交互验证所得的预测残差平方和(PRESS 值)确定。为考察校正模型的适用性和准确性, 本研究对不同种类的生物柴油样品分别进行建模与验证研究, 包括: 以文冠果油生物柴油为例的单原料油的校正模型与验证; 多种原料油的通用校正模型的建立与验证; 以花椒油生物柴油为例, 研究了 N R 光谱校正模型的适用性。

## 3 结果与讨论

### 3.1 基础数据获取

采用 2.2 节的方法测定生物柴油成分的基础数据。典型的生物柴油样品色谱图见图 1; 224 个生物柴油的成分分布见表 1。可见样本的覆盖范围较宽, 具有较强的代表性。与标准方法<sup>[15]</sup>相比, 此方法的准确性和重复性较好。

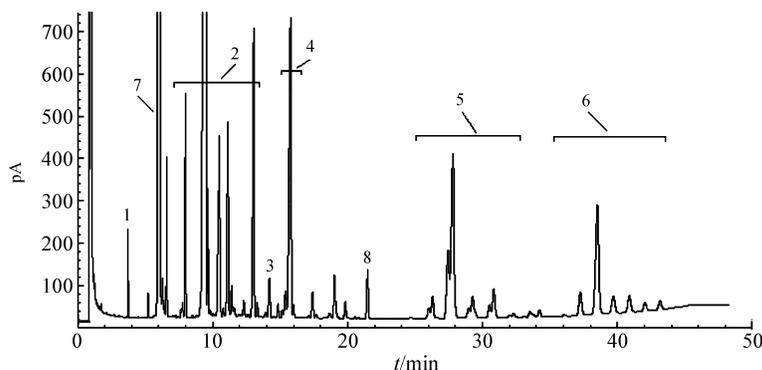


图 1 典型的生物柴油色谱图

Fig 1 Chromatogram of a typical biodiesel sample

1. 甘油 (Glycerol); 2. 甲酯 (Methyl esters); 3. C16 单甘酯 (C16 Monoglyceride); 4. C<sub>18</sub> 单甘酯 (C<sub>18</sub> Monoglyceride); 5. 二甘酯 (Diglycerides); 6. 三甘酯 (Triglycerides); 7. 正十五烷 (内标 1, Pentadecane internal standard 1); 8. 三癸酸甘油酯 (内标 2, Triolein internal standard 2)。

表 1 样品集性质统计

Table 1 Statistical compositions of sample set

性质统计 Compositions	甘油 Glycerol	甲酯 Methyl esters	单甘酯 Monoglycerides	二甘酯 Diglycerides	三甘酯 Triglycerides
最大值 Maximum (%)	0.536	101.64	9.19	9.34	9.19
最小值 Minimum (%)	0.000	66.47	0.00	0.00	0.00
平均值 Average (%)	0.119	91.87	2.56	1.51	0.43
标准偏差 Standard deviations (%)	0.104	8.39	2.90	2.12	1.41

### 3.2 近红外光谱解析

图 2 是所收集的 6 类原料油生物柴油的 NIR 光谱图。由图 2 可见, 光谱中主要包含 C—H 倍频和合频的吸收信息, 如 8609 和 8316  $\text{cm}^{-1}$  处的吸收是甲基 C—H 的二级倍频峰, 7216 和 7108  $\text{cm}^{-1}$  处的吸收是亚甲基 C—H 的第一组合频峰, 5828 和 5766  $\text{cm}^{-1}$  处的吸收分别是甲基和亚甲基 C—H 的一级倍频峰, 6005  $\text{cm}^{-1}$  处为甲酯中的 C—H 吸收峰。

尽管 6 类原料油生物柴油的 NIR 光谱非常相近, 但其成分上却存在一定差异。例如, 花椒油生物柴油主要是亚麻酸、亚油酸和油酸甲酯; 苦山杏油生物柴油主要是油酸和亚油酸甲酯; 酸化油和地沟油生物柴油则主要是多种脂肪酸甲酯的混合, 具体视原料油种类而定。这种差异可以在其 NIR 光谱的主成分分析 (PCA) 得分图上表现出来。图 3 为 6 类 52 个有代表性生物柴油样品的前 3 个主成分得分分布图。从图 3 可见, 不同原料生产的生物柴油有聚类分布的倾向。由于酸化油原料来源复杂, 所以其分布较宽。其余 4 种生物柴油的分布则相对集中。但从整体来看, 尽管不同种类生物柴油的 NIR 光谱之间存在差异, 但这些差异并非十分显著, 说明脂肪酸基团对光谱会产生一定的影响, 但 NIR 光谱主要反映的是甲酯基团的信息。因此, 有可能建立适合多种原料油生物柴油的通用校正模型。

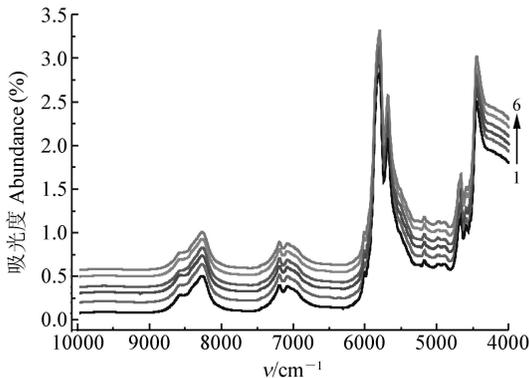


图 2 6 类原料油生物柴油的 NIR 光谱图

Fig 2 NIR spectra of 6 kinds of biodiesel

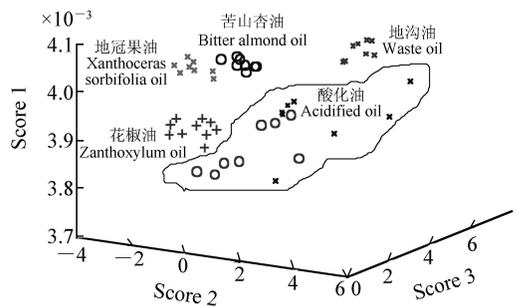


图 3 6 类 52 个有代表性生物柴油样品的前 3 个主成分得分分布图

Fig 3 Principal component scores of 52 biodiesel samples of 6 kinds

### 3.3 校正模型的建立

**3.3.1 文冠果油单原料油模型与验证** 将 44 个文冠果油生物柴油样品分为校正集 (36 个样品) 和验证集 (8 个样品), 考察了不同光谱区间对模型参数的影响, 以及单种原料油所建模型的预测准确性。

生物柴油在 4000~ 10000  $\text{cm}^{-1}$  范围的光谱包含了 O—H 键的一级倍频、C—H 键的一级倍频、二级倍频及其合频信息。从图 2 可以看出, C—H 一级倍频区 (6000~ 6550  $\text{cm}^{-1}$ ) 和部分 C—H 第一合频区 (4550~ 4000  $\text{cm}^{-1}$ ) 的吸收度值超过了 1.5 AU, 这些区间的光谱吸收强度可能与成分浓度之间存在较强的非线性关系。另外, 10000~ 9000  $\text{cm}^{-1}$  区间的光谱存在较大的噪声, 且没有显著的特征吸收。若这些光谱区间参与模型建立, 将会使校正模型预测能力下降。表 2 给出了利用全谱和去除这些光谱区间 (即选取 4550~ 5500  $\text{cm}^{-1}$  和 6550~ 9000  $\text{cm}^{-1}$  区间) 所建模型及其预测的对比结果。可以看出, 选用 4550~ 5500  $\text{cm}^{-1}$  和 6550~ 9000  $\text{cm}^{-1}$  光谱区间建立模型的预测能力明显优于全谱的结果。

表 2 文冠果油生物柴油不同光谱区间的校正和预测结果比较

Table 2 Calibration and prediction results of xanthoceras sorbifolia oil biodiesel in different spectral ranges

成分 Composition	波数范围 Wavelength region	主因子数 Number of principal components	$R^2$	SECV (%)	SEP (%)
甘油 Glycerol		4	0.9579	0.044	0.063
甲酯 Methyl esters		6	0.9976	0.54	0.49
单甘酯 Monoglycerides	4000~ 10000 $\text{cm}^{-1}$	5	0.9906	0.42	0.46
二甘酯 Diglycerides		5	0.9874	0.40	0.37
三甘酯 Triglycerides		4	0.9240	0.22	0.18

续表 2(Continued to Table 2)

成分 Composition	波数范围 Wavelength region	主因子数 Number of principal components	$R^2$	SECV (%)	SEP (%)
甘油 Glycerol		3	0.9737	0.036	0.057
甲酯 Methyl esters		6	0.9992	0.32	0.65
单甘酯 Monoglycerides	4550~ 5500 $\text{cm}^{-1}$	5	0.9978	0.21	0.23
二甘酯 Diglycerides	6550~ 9000 $\text{cm}^{-1}$	5	0.9976	0.17	0.20
三甘酯 Triglycerides		4	0.9419	0.19	0.18

注 (Note):  $R^2$  为校正集交互验证得到的相关系数 (Correlation coefficient); SECV 为校正集交互验证得到的预测标准偏差 (Standard error of cross validation); SEP 为验证集预测标准偏差 (Standard error of prediction)。

从单种原料油的建模和预测结果可以看出,对测定生物柴油中的甘油、甲酯、单甘酯、二甘酯和三甘酯的含量, NIR 光谱方法可以给出比较准确的结果。

**3.3.2 多种原料油模型与验证** 若能建立适合所有原料油来源的生物柴油全局校正模型 (Global models), 则对该方法的推广和使用都将带来很大的便利。将 6 种原料油来源的 224 个生物柴油样品分为校正集 (167 个样品) 和验证集 (57 个样品), 以考察建立全局模型的可行性。

采用与文冠果油生物柴油建模相同的条件建立了多种原料油生物柴油的校正模型。表 3 给出了模型建立过程的交互验证结果和验证集的预测结果; 图 4 中 A、B 和 C 分别为甲酯、单甘酯和二甘酯的 NIR 光谱交互验证预测值与色谱方法的相关图。与单原料的文冠果油结果相比, 由于成分变得相对复杂, 光谱之间的干扰更为显著, 建立模型所用的主因子数明显增加, 同时校正和预测结果都有不同程度的降低。尽管如此, 仍得到了满意的结果, 完全满足工业生产对过程分析误差的要求。

上述结果表明, 建立多种原料油来源的生物柴油全局校正模型是完全可行的, 但随着模型中原料油种类的增多, 成分的复杂性将引起光谱与浓度间的非线性变动关系, PLS 方法所建模型的预测准确性将下降。为得到准确性更高的预测结果, 需要分别建立单种原料油的 PLS 校正模型, 或采用更为复杂的算法, 如局部权重回归、神经网络或支持向量机等建立非线性校正模型。

表 3 多种原料油生物柴油的校正和预测结果

Table 3 Results of calibration and prediction of multiple kinds of biodiesel samples

成分 Composition	主因子数 Number of principal components	$R^2$	SECV (%)	SEP (%)
甘油 Glycerol	10	0.9635	0.028	0.047
甲酯 Methyl esters	11	0.9898	1.29	1.43
单甘酯 Monoglycerides	13	0.9945	0.31	0.47
二甘酯 Diglycerides	11	0.9800	0.46	0.51
三甘酯 Triglycerides	12	0.9465	0.53	0.87

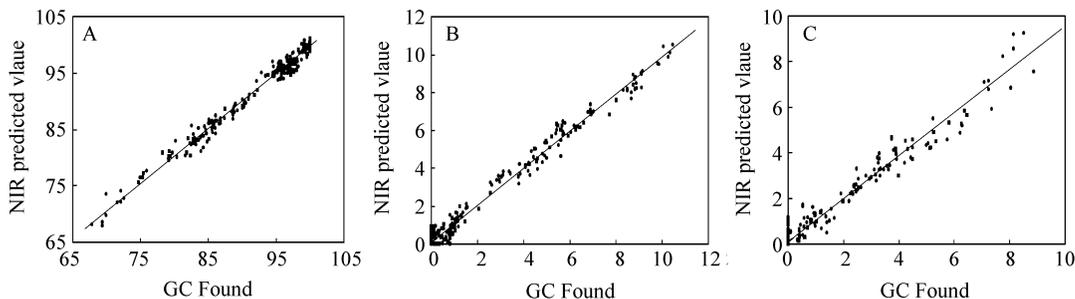


图 4 近红外光谱法与气相色谱法测定生物柴油中甲酯含量 (A) 单甘酯含量 (B) 和二甘酯含量 (C) 的相关图

Fig. 4 Correlation of NIR results with GC of the content of methyl esters (A), monoglycerides (B) and diglycerides (C) in biodiesel

●: 校正集样本 (Calibration set); ■: 验证集样本 (Prediction set)。

**3.3.3 模型适用性的研究** 以花椒油生物柴油为例, 研究了多种原料油生物柴油的 NIR 校正模型对新原料油生物柴油的适用性。若校正集中不含待测类型的样本, 其预测结果将会存在一定的系统误差, 且模型的适应性判据也会给出提示。为考察向校正集添加少量样本对扩充模型适应性的效果, 从 47 个

花椒油生物柴油中任意选取 10 个样品, 用于模型的扩充, 剩余 37 个作为验证集样本。

将花椒油生物柴油样本从多种原料油校正集中剔除, 建立剩余 5 种原料油生物柴油的校正模型。然后, 再将任意选取的 10 个花椒油生物柴油样本添加到校正集中, 建立含有 10 个花椒油生物柴油的多原料油生物柴油的校正模型。用以上建立的两类模型分别对 37 个验证集样本进行预测分析 (表 4)。由表 4 可见, 未含花椒油生物柴油的模型仍具有较好的预测趋势, 但存在一定的系统误差, 预测准确性也有明显下降。当模型中添加了 10 个花椒油生物柴油样本后, 消除了系统误差, 且预测准确性有了显著改善, 与 3.3.2 节建立的模型基本相当。

上述结果说明, 对于模型中未含的新类型原料油生物柴油样本, 只需向多种原料油校正集中添加 10 个以上样本便可较为准确地测定新类型生物柴油的成分。

表 4 不同校正模型对 37 个花椒油生物柴油的预测结果

Table 4 Prediction results of 37 zanthoxylum oil biodiesel samples with different calibration models

成分 Composition	不含花椒油生物柴油校正模型的预测结果 Calibration model without zanthoxylum oil biodiesel samples		含 10 个花椒油生物柴油校正模型的预测结果 Calibration model including 10 zanthoxylum oil biodiesel samples	
	R <sup>2</sup>	SEP (%)	R <sup>2</sup>	SEP (%)
甘油 Glycerol	0.8377	0.190	0.8785	0.059
甲酯 Methyl esters	0.9810	2.22	0.9863	1.13
单甘酯 Monoglycerides	0.9828	1.01	0.9855	0.61
二甘酯 Diglycerides	0.9718	1.10	0.9753	0.83
三甘酯 Triglycerides	0.7354	1.26	0.7476	0.98

### 3.4 近红外分析方法的重复性

由于光谱测量的稳定性, NIR 分析方法具有良好的重复性。随机选取一个样本重复扫描 7 次的 NIR 光谱, 并分别调用建立的多种原料油的生物柴油模型, 其质量分数的相对标准偏差分别为: 甘油 0.74%, 甲酯 0.26%, 单甘酯 2.86%, 二甘酯 0.70%, 三甘酯 1.41%。说明采用 NIR 分析方法测定混合油生物柴油的成分具有较高的精密性。

### 3.5 小结

通过以上研究, 可以得到以下结论: (1) 通过偏最小二乘方法可以建立适合多种原料油生物柴油的通用近红外校正模型, 其结果完全满足过程控制分析的要求。但若想得到更准确的预测结果, 需要建立单种原料油的 PLS 校正模型, 或者采用更为复杂的校正算法; (2) 对于新型生物柴油, 只需向校正集添加少量样本, 扩充模型覆盖范围后, 便可较为准确地对该类样本进行预测分析。这样, 基于本研究建立的这些基础校正模型, 针对新类型原料油或加工工艺, 只需添加少量样本, 便可直接应用于生物柴油的实际工业生产中; (3) 近红外光谱方法有望较好解决生物柴油传统分析方法的诸多弊端, 可以快速准确地测定生物柴油的主要成分含量, 适合于生物柴油生产过程的中间控制分析。

## References

- 1 M N En-Ze(闵恩泽). *Journal of Chemical Industry and Engineering* (化工学报), **2006** 57(8): 1739~1745
- 2 DU Ze-Xue(杜泽学), YANG Guo-Jun(阳国军). *Progress in Chemistry* (化学进展), **2009** 21(11): 2341~2347
- 3 LI Chang-Xiu(李长秀), YANG Hai-Ying(杨海鹰), WANG Li-Qin(王丽琴), TIAN Song-Bai(田松柏). *Chinese Journal of Chromatography* (色谱), **2006** 24(5): 524~528
- 4 LI Chang-Xiu(李长秀), TANG Zhong(唐忠), YANG Hai-Ying(杨海鹰). *Journal of Instrumental Analysis* (分析测试学报), **2005** 24(5): 66~68
- 5 CHU Xiao-Li(褚小立), XU Yu-Peng(许育鹏), LU Wan-Zhen(陆婉珍). *Chinese J. Anal. Chem.* (分析化学), **2008** 36(5): 702~709
- 6 XU Guang-Tong(徐广通), YUAN Hong-Fu(袁洪福), LU Wan-Zhen(陆婉珍). *Chinese J. Anal. Chem.* (分析化学), **1999** 27(1): 29~33
- 7 WU De-Hui(吴德会). *Spectroscopy and Spectral Analysis* (光谱学与光谱分析), **2008** 28(7): 1530~1534

- 8 LU Wan-Zhen(陆婉珍). *Modern Near Infrared Spectroscopy Analytical Technology* (现代近红外光谱分析技术). Beijing (北京): China Petrochemical Press(中国石化出版社), 2007: 35~ 82
- 9 QIAN Ping(钱平), SUN Guo-Qin(孙国琴), ZHANG Cun-Zhou(张存洲). *Spectroscopy and Spectral Analysis*(光谱学与光谱分析), 2008, 28(12): 2851~ 2854
- 10 CHU Xiao-Li(褚小立), YUAN Hong-Fu(袁洪福), LU Wan-Zhen(陆婉珍). *Petroleum Refinery Engineering*(炼油技术与工程), 2005, 35(4): 26~ 29
- 11 Knothe G. *JAACS*, 1999, 76(7): 795~ 800
- 12 Felizardo P, Baptista P, Menezes J C. *Anal. Chim. Acta*, 2007, 595(1-2): 107~ 113
- 13 Kawai S, Kohda J, Nakano Y. *Near Infrared Spectroscopy*, 2009, 17(1): 51~ 58
- 14 Baptista P, Felizardo P, Menezes J C. *Anal. Chim. Acta*, 2008, 607(2): 153~ 159
- 15 ASTM D6584-07. Standard Test Method of Determination of Free and Total Glycerin in B-100 Biodiesel Methyl Esters by Gas Chromatography. American National Standard

## Prediction of Main Chemical Compositions in Biodiesel by Near Infrared Spectroscopy

KONG Cuiping\*, CHU Xiao-Li, DU Ze-Xue, LU Wan-Zhen  
(Research Institute of Petroleum Processing, Beijing 100083)

**Abstract** A rapid method was established for the determination of the main chemical compositions in biodiesel (including fatty acid methyl esters, monoglycerides, diglyceride, triglyceride and glycerol) by the combination of near infrared spectra (NIR) and chemometrics. The reference values were obtained by the gas chromatographic method, which was correlated with the NIR spectra data by partial least squares regression method. The calibration models for multi-feedstock biodiesel and a single feedstock biodiesel were established respectively. The applicability of the calibration models was examined with the zanthoxylum oil biodiesel as an example. The results showed that a universal calibration model suitable for various biodiesel from different sources can be established by partial least-squares method. For this biodiesel manufactured from other fatty oils, the same calibration model can be extended by adding 10 more new samples. The calibration model so obtained will provide more accurate results. The NIR method is suggested as a rapid method for the quality control in the process of biodiesel production.

**Keywords** Near infrared spectroscopy, Biodiesel, Chemometrics, Methyl ester

(Received 26 November 2009; accepted 27 January 2010)

## 《色谱在生命科学中的应用》

全书共分五章,介绍了目前在生命科学研究中常用的色谱技术,详细讨论了液相色谱、气相色谱和毛细管电泳等色谱技术在脱氧核糖核酸分析、生物技术药物的纯化、蛋白质和核酸性质的鉴定、生物标志物检测,以及在蛋白质组学研究中的应用与发展。

该书为化学工业出版社出版的色谱技术丛书之一,可供从事色谱分析的人员及大专院校分析化学相关专业师生学习参考。该书由廖杰、钱小红、董方霆、张养军编著,定价 29.00 元。