

汤菲菲,姚燕,邵玉涛,等. 基于理论吸收光谱的棉纤维波长筛选方法[J]. 江苏农业科学,2019,47(7):207-209,219.
doi:10.15889/j.issn.1002-1302.2019.07.049

基于理论吸收光谱的棉纤维波长筛选方法

汤菲菲¹,姚燕¹,邵玉涛²,蔡晋辉¹,沈晓敏¹,廖文汉¹

(1. 中国计量大学计量测试工程学院热工所,浙江杭州 310018; 2. 阿拉尔质量技术监督局,新疆阿拉尔 843300)

摘要:马克隆值是评价棉纤维品质的一个重要指标,提出一种基于 Gaussian 软件的波长筛选方法,以期提高红外光谱法测定棉纤维马克隆值的准确性。采用基于哈特里-福克方程(简称 HF)的从头算计算法、半经验分子轨道 AM1 方法和密度泛函理论 DFT 对棉纤维分子进行量子化学计算,并将计算结果与棉纤维红外光谱图进行比较分析,从而得到最优理论计算方法及其基组,进而通过分析该最优理论红外光谱图得到特征波长。该方法筛选出的特征波长为 3 727~3 567、3 005~2 700、1 250~1 000 cm^{-1} 。分别利用此特征波长与连续投影法(SPA)选出的特征波长建立偏最小二乘(PLS)模型,对比分析的结果表明,使用 Gaussian 软件 DFT 密度泛函理论的 B3LYP/6-311+G** 方法可以较为准确地得到最优光谱图,该理论光谱波长筛选方法可以优选出高信噪比的波长。基于 Gaussian 软件的波长筛选方法能够简化红外建模步骤并提高红外光谱定量模型的性能。

关键词:棉花;纤维品质;纤维分子结构;马克隆值;Gaussian 软件;量子化学;波长筛选;模型优化;效果验证

中图分类号: O657.33 **文献标志码:** A **文章编号:** 1002-1302(2019)07-0207-03

棉花是我国重要的战略物资,是仅次于粮食的第二大类农作物,在农业经济格局中具有重要地位。棉花质量的好坏对于其价格、生产工艺成本以及后续产品的质量起着至关重要的作用。马克隆值是反映棉花纤维细度与成熟度的综合指标,且会对成纱工艺、质量及织物质量产生很大的影响,是棉花贸易、棉纺工程的重要指标。

传统棉纤维马克隆值的测量需要借助 Y145C 型气流仪、便携式气流仪等仪器^[1],这种检验方法不仅费时费力,而且由于个人经验的不同以及把握的标准不同,容易出现人为误差。红外光谱是现代结构化学和分析化学最常用和不可缺少的工具,具有测试迅速、操作方便、重复性好、灵敏度高、不损坏样本结构等优点,因此被广泛应用于科学研究的各个领域^[2]。但红外光谱对于复杂对象待测组分的特征吸收频率会受到其他组分的干扰,得到的冗余信息较多,从而影响吸收峰位置、强度,甚至峰的形状,进而影响模型的预测精度。因此,进行光谱定量分析时,必须选择与待测组分相关的高信息量波长,以降低其他组分带来的噪声干扰^[3]。Gaussian 软件是一个成熟、界面友好的量子化学软件,可执行各类不同精度和理论档次的量化计算,并可进行分子和化学反应性质的理论预测。因此,可以利用 Gaussian 软件模拟红外光谱并在进行红外光谱试验之前筛选出特征波长,缩小可用的光谱区范围,节约试验时间,提高模型预测精度。

本研究提出在进行红外光谱试验之前利用 Gaussian 软件对理论模拟光谱进行波长筛选,在最优波段运用红外光谱测

量马克隆值的方法,以期缩短红外光谱试验的时间,减少冗余光谱信息,提高预测模型的精度。采用 3 种方法对棉纤维分子进行量子化学计算,将计算结果与实际试验光谱图进行对比分析,从而得到最优的理论计算方法及其基组,并得到与实际试验较为吻合的理论红外光谱图,通过分析该理论红外光谱图进一步得到棉纤维分子的频带分配信息,进而由频带分配信息得出特征波长;分别将该特征波长与连续投影法(SPA)选择的特征波长建立偏最小二乘(PLS)模型,并对模型优化效果进行验证。

1 红外吸收光谱原理解析

红外吸收光谱是由分子振动、转动能级的跃迁所形成的^[4]。在有机分子中,不同基团选择性地吸收不同频率的红外辐射,引起分子的能量变化,发生不同振动能级的跃迁,形成独特的红外吸收光谱。绝大多数有机化合物和无机离子的基频吸收带出现在中红外光谱区,由于基频振动是红外光谱中吸收最强的振动,因此该区被广泛应用于化合物分析、结构分析和其他化学过程的定性定量分析中^[5]。本研究对棉纤维结构以及内部分子振动的分析也在中红外区域进行。

2 Gaussian 软件的使用

Gaussian 软件作为强大的工具,可以研究许多化学领域的课题,也正是因为该软件在计算分子振动频率方面具有较高的精度,本研究采用该软件对棉纤维分子进行结构优化和红外振动光谱的模拟计算。整个计算过程为利用 GaussView 5.0、Chemdraw/Chem 3D、Hyperchem 等绘图软件绘制输入文件^[6],选择合适的基组对该文件进行几何优化,几何优化的目的是寻找分子稳定的几何形态,即将建立的模型优化到一个能量极小的点上^[7];然后使用相同的基组选择合适的方法进行频率计算^[8],并通过 GaussView 5.0 显示计算结果。

收稿日期:2017-11-03

基金项目:国家自然科学基金(编号:51606181)。

作者简介:汤菲菲(1993—),女,浙江丽水人,硕士研究生,主要从事近红外光谱的分析研究。E-mail: tangfeifei1993@126.com。

通信作者:姚燕,博士,副教授,主要从事近红外光谱的分析研究。E-mail: yaoyan@cjl. edu. cn。

整个理论计算过程如图 1 所示,通过该过程能够在理论上得到较为准确的分子理论振动谱。

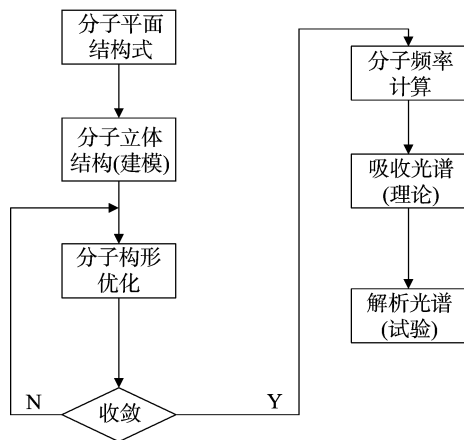


图1 理论计算过程

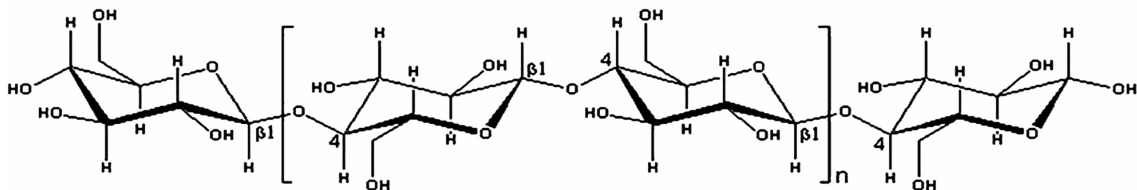


图2 棉纤维分子结构

式中: $n + ik$ 表示复折射率,它为一复数,等于介电常数 $\hat{\epsilon}$ 的平方; v_j, S_j, Γ_j 分别为各阻尼振荡器的共振频率、振荡强度、脉宽; ϵ_∞ 为高频部分对吸收系数和折射率的贡献; i 为复数虚部。

吸收系数 $\alpha(v)$ 与复折射率的虚部 k 、光速 c 和频率 v 之间存在以下关系:

$$\alpha(v) = 4\pi vk/c. \quad (2)$$

吸光度分别采用 HF、AM1、DFT 方法算出的结果,EXP 代表的是试验光谱,根据式(1)、式(2)可得到理论计算光谱。图 3 代表的是在红外光谱指纹区(1 330 ~ 400 cm^{-1})使用 3 种不同方法计算出的结果以及实际试验得到的棉纤维吸光度信息,图 4 代表的是在红外光谱特征区(4 000 ~ 1 330 cm^{-1})使用 3 种不同方法计算出的结果以及实际试验得到的棉纤维吸光度信息。

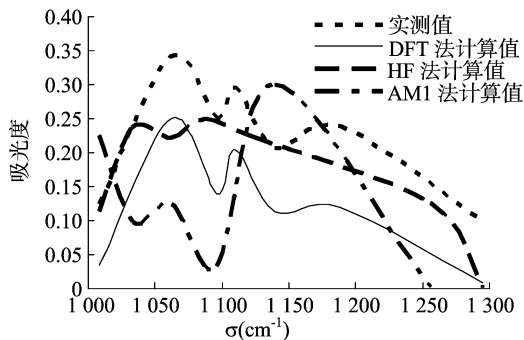


图3 红外光谱指纹区棉纤维吸光度信息

由图 3、图 4 可知,利用 DFT 密度泛函理论的 B3LYP/6-311 + G** 方法所得到的计算结果在吸收峰数量、波形形状以及频率位置、吸光度上与试验吸收峰最为接近,而 HF、AM1 方法得到的吸收谱与试验结果相差较大,这可能与理论本身

3 棉纤维的红外光谱理论解析

3.1 基组的选择

采用 GaussView 5.0 软件绘制棉纤维分子结构图,并分别采用哈特里-福克方程(HF)从头算理论、AM1 半经验理论和 DFT 密度泛函理论对棉纤维分子进行量子化学计算^[7]。每种理论计算方法的几何优化和频率计算均在相同的理论水平和基组上完成,HF 方法选择的基组是 HF/6-31G(d),DFT 方法选取的基组为 B3LYP/6-311 + G(d,p),一般表示为 B3LYP/6-311 + G**^[9]。计算结果没有虚频出现,说明得到的分子结构是一个稳定的构象。棉纤维分子结构如图 2 所示。

假设理论光谱的每个特征共振峰都是一个阻尼振荡器,其特征吸收系数和折射率可根据如下公式得到^[10]:

$$\hat{\epsilon}(v) = (n + ik)^2 = \epsilon_\infty + \sum_j \frac{S_j v_j^2}{v_j^2 - v^2 - i v \Gamma_j}. \quad (1)$$

精度不够或选取的基组偏小有关。相对于 HF 和 AM1 方法,DFT 采用了 B3LYP/6-311 + G** 较大基组,并有效地考虑了电子交换和电子相关能量,因而有较高的计算正确率。

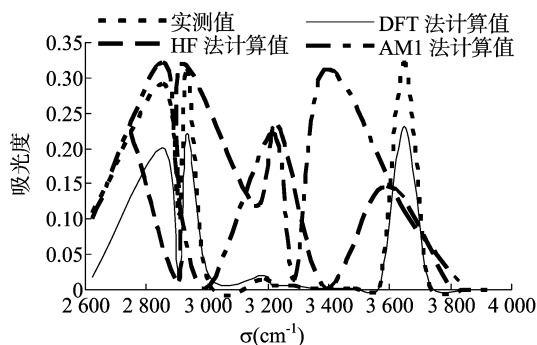


图4 红外光谱特征区棉纤维吸光度信息

3.2 基于最优光谱图的特征波长筛选

从图 5、图 6 可以看出,Gaussian 软件鉴定基团的结果与理论结果是很符合的,且理论光谱(DFT)与试验光谱的变化趋势大体相同,只是存在一定程度的偏移,主要原因为(1)试验是在常温下进行的,而理论计算在绝对零度的条件下进行,随着温度的升高,棉纤维的吸收峰强度也会增强^[11];(2)理论计算时采用的是单分子模型,只能得到分子内的振动模式,而忽略了分子之间的范德华力、氢键弱作用力以及晶格振动等。因此,理论计算结果与试验结果不可能完全相符,但可从另一角度验证吸收峰来源判断的正确性。

纤维素大分子的官能团主要是羟基(—OH)、亚甲基(—CH₂)和醚键[R—O—R'(R')₂]。从图 5、图 6 可以看出,在 4 000 ~ 400 cm^{-1} 范围内,棉纤维最强的红外吸收带在 1 060 cm^{-1} 附近,在其两侧有很多弱的肩峰,在 2 855、2 932、

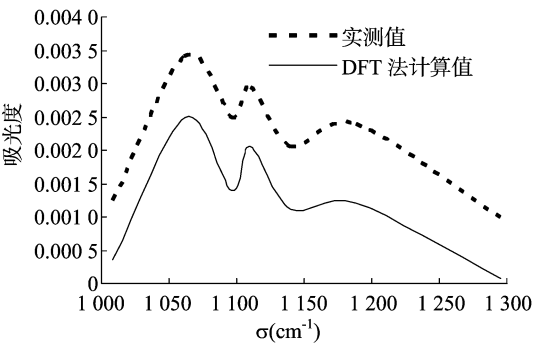


图5 红外光谱指纹区棉纤维吸光度信息

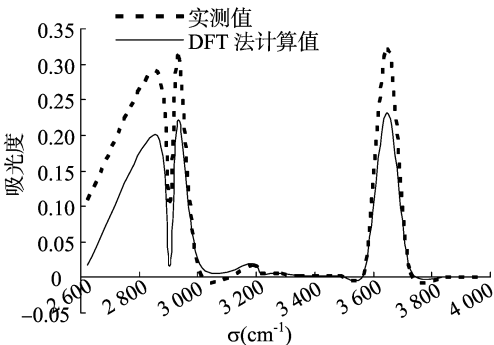


图6 红外光谱特征区棉纤维吸光度信息

3 647 cm^{-1} 处也有较强的红外吸收带。根据红外光谱基团吸收带的可知,一般情况下,羟基、亚甲基、醚键的实际吸收带分别为 3 630、2 900、1 100 cm^{-1} 左右。结合基团吸收频率与分子振动之间的关系,可以得出棉纤维分子的频带分配结果(表 1)。根据吸光度波长筛选方法^[12]可得到,选取的特征波段为 3 727 ~ 3 567、3 005 ~ 2 700、1 250 ~ 1 000 cm^{-1} ,波长点个数为 715 个。

表 1 棉纤维分子频带分配结果

频率 (cm^{-1})	化学键振动分配	备注
3 647	伸缩振动	—OH 基团,纤维素
2 932	反对称伸缩振动	—CH ₂ 基团,纤维素
2 855	对称伸缩振动	—CH ₂ 基团,纤维素
1 380 ~ 1 650	弯曲振动	—OH 基团,H ₂ O
1 060	伸缩振动	R—O—R(R'),纤维素

3.3 特征波长筛选方法有效性检验

连续投影算法(SPA)是一种使矢量空间共线性最小化的前向变量选择算法,其优势在于提取全波段的几个特征波长,能够消除原始光谱矩阵中冗余的信息,可用于光谱特征波长的筛选。近年来,国内外学者在利用光谱分析技术检测作物和食品中某些重要成分的含量时,采用连续投影算法进行有效波长的选取^[13]。作为对比,本研究采用连续投影算法提取特征波长,分别建立偏最小二乘模型,以验证基于 Gaussian 软件波长筛选方法的有效性。试验收集了 30 个棉纤维样本,随机选择 25 个样本为训练集,5 个为测试集,采用大容量棉花纤维测试仪(HVI)^[14]测得棉纤维样本的马克隆值,并使用美国 Thermo Fisher Scientific 公司生产的 Nicolet 系列 NEXUS670 型号的傅里叶变换红外光谱仪测量 30 个样本的红外吸收光谱。红外试验在室温(20 ~ 23 $^{\circ}\text{C}$)环境下进行,光谱仪的扫描方式为漫反

射方式,分辨率为 8 cm^{-1} ,自动扫描 32 次,记录其平均光谱。

对试验数据分别进行偏最小二乘回归,由结果(图 7)可知,SPA 算法优选的波长为 13 个,此时的预测平均标准偏差(RMSE)最小,为 0.16,2 种不同波长筛选方法得到的预测结果对比如表 2 所示。SPA 方法提取出了 13 个特征波段,分别为 1 060、1 087、1 248、2 757、2 790、2 855、2 932、2 853、3 002、3 621、3 647、3 696、3 705 cm^{-1} ,它们均分布在本研究方法筛选出的特征波段中,由表 2 可知,采用 SPA 方法预测相关系数达到 0.825,表明该算法确实提取了建模所需要的有用信息,简化了模型的计算,但是与之相比,本研究提出的波长筛选方法建立的模型预测精度更高。

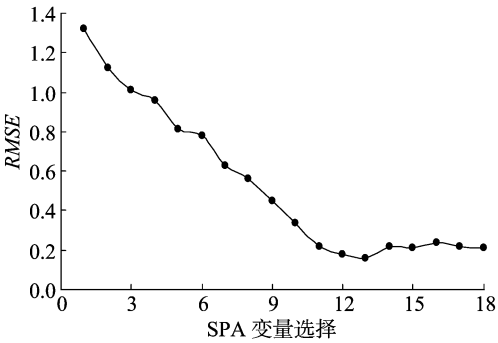


图7 SPA 算法选择变量数

表 2 2 种不同算法结果比较

模型	R^2	RMSEP	RMSEC
SPA - PLS	0.825	0.16	0.09
Gaussian - PLS	0.946	0.07	0.04

注:RMSEP、RMSEC 分别表示预测集均方根误差、校正集均方根误差。

4 结论

本研究提出一种基于理论吸收光谱筛选出最优波长并测定马克隆值的新方法,该方法利用 Gaussian 软件采用 DFT 密度泛函理论的 B3LYP/6 - 311 + G** 基组得到最优光谱图,分别将该方法筛选出的特征波长与 SPA 法得到的特征波长建立 PLS 模型,通过比较相关系数以及均方根误差可知,该方法能够有效地筛选出有用的特征波长信息,且选择效果优于经典的连续投影法。通过理论模拟红外吸收光谱进行波长筛选,并在筛选出的波长范围内进行后续的红外光谱试验,可简化建模计算,提高棉纤维马克隆值的预测精度。

参考文献:

[1] 孟莉莉. 棉花检验中的马克隆值测试方法[J]. 中国纤检,2014 (7):78 - 79.
[2] 杨新河,王丽丽,黄建安,等. 傅里叶变换红外光谱鉴别黑茶样品[J]. 食品科学,2012,33(14):203 - 205.
[3] 陈艺芳. 可见 - 近红外 - 中红外光谱用于人血清白蛋白分析的方法研究[D]. 广州:暨南大学,2016.
[4] 黄莉. 红外光谱结合模式识别技术在微量车漆物证分析中的应用[J]. 法制与社会,2016(35):273,277.
[5] 岳飞亨. 基于太赫兹波的持久性有机物检测技术研究[D]. 杭州:浙江大学,2012.

燥 1、2 min 对样品预测精确度的提升要优于传统烘箱 40 ℃ 干燥 2 h 的结果。微波法不仅在干燥效率上优于传统烘箱,同时对样品预测精确度的提升方面也优于传统的烘箱干燥法。

3 结论

本研究利用近红外技术建立了卷烟不同物理状态下的定量分析模型,并将微波和烘箱应用于卷烟近红外的分析过程中。结果表明:水分含量对卷烟烟丝、烟末样品的近红外光谱吸收强度和主成分分布有较大影响。随着水分含量的升高,在谱图区间的 6 944、5 155 cm^{-1} 附近的吸收强度也越大,通过一阶求导等预处理方法不能有效地消除水分含量带来的光谱差异;通过对比传统的烘箱干燥法和微波干燥法的卷烟近红外预测结果发现,经微波干燥后的样品预测准确度整体要优于传统烘箱干燥的结果,采用微波干燥 1、2 min 后样品的预测准确度要优于烘箱 40 ℃ 干燥 2 h 的结果,但对于烟丝样品不适宜进行干燥处理。微波干燥法不仅在干燥效率上优于烘箱法,同时对提高样品预测准确度起到了积极的作用,是一种快速、精确的卷烟近红外干燥方法。此外,微波加热不均匀性一直是行业研究的热点和难点,严重影响了微波干燥技术在工业化应用中的推广。本研究仅通过固定样品位置来降低微波场的不均匀性影响,如何从根本上克服微波加热的不均匀性问题,需要在接下来的工作中进行深化研究。

参考文献:

- [1] Roggo Y, Chalou P, Maurer L, et al. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies[J]. Journal of Pharmaceutical and Biomedical Analysis, 2007, 44 (3): 683 - 700.
- [2] Teye E, Huang X, Afoakwa N. Review on the potential use of near infrared spectroscopy (NIRS) for the measurement of chemical residues in food [J]. American Journal of Food Science and Technology, 2013, 1(1): 1 - 8.
- [3] 吴习宇, 赵国华, 祝诗平. 近红外光谱分析技术在肉类产品检测中的应用研究进展[J]. 食品工业科技, 2014, 35 (1): 371 - 374, 380.
- [4] 段焰青, 周红, 李青青, 等. 烟样水分质量分数对其常规化学成分近红外测定的影响[J]. 云南大学学报(自然科学版), 2005, 27 (5): 424 - 428.
- [5] Liu H L, Jiaqiang E, Deng Y W, et al. Experimental study on pyrolysis characteristics of the tobacco stem based on microwave heating method[J]. Applied Thermal Engineering, 2016, 106 (5): 473 - 479.
- [6] Hesas R H, Daud W M, Sahu J N, et al. The effects of a microwave heating method on the production of activated carbon from agricultural waste: A review [J]. Journal of Analytical and Applied Pyrolysis, 2013, 100: 1 - 11.
- [7] 秦西云, 杨宇虹. 近红外分析的样品前处理新技术研究[J]. 中国农学通报, 2008, 24 (10): 150 - 152.
- [8] 全国烟草标准化技术委员会. 烟草及烟草制品试样的制备和水分测定 烘箱法: YC/T 31—1996 [S]. 北京: 中国标准出版社, 2001.
- [9] 全国仪器分析测试标准化技术委员会. 分子光谱多元校正定量分析通则: GB/T 29858—2013 [S]. 北京: 中国标准出版社, 2014.
- [10] 黄锋, 陈清, 王乐, 等. 片烟增湿与干燥的薄层动力学模型[J]. 中国烟草学报, 2014, 20 (6): 34 - 40.
- [11] 全国烟草标准化技术委员会. 烟草及烟草制品 总植物碱的测定 连续流动(硫氰酸钾)法: YC/T 468—2013 [S]. 北京: 中国标准出版社, 2013.
- [12] 全国烟草标准化技术委员会. 烟草及烟草制品 水溶性糖的测定 连续流动法: YC/T 159—2002 [S]. 北京: 中国标准出版社, 2004.
- [13] 全国烟草标准化技术委员会. 烟草及烟草制品 总氮的测定 连续流动法: YC/T 161—2002 [S]. 北京: 中国标准出版社, 2004.
- [14] 全国烟草标准化技术委员会. 烟草及烟草制品 钾的测定 连续流动法: YC/T 217—2007 [S]. 北京: 中国标准出版社, 2007.
- [15] 全国烟草标准化技术委员会. 烟草及烟草制品 氯的测定 连续流动法: YC/T 162—2011 [S]. 北京: 中国标准出版社, 2011.
- [16] Vadiwambal R, Jayas D S. Non - uniform temperature distribution during microwave heating of food materials - a review [J]. Food and Bioprocess Technology, 2010, 3 (2): 161 - 171.
- [17] 全国家用电器标准化技术委员会. 家用微波炉性能试验方法: GB/T 18800—2008 [S]. 北京: 中国标准出版社, 2008.
- [18] Qin Y H, Gong H L. NIR models for predicting total sugar in tobacco for samples with different physical states [J]. Infrared Physics and Technology, 2016, 77: 239 - 243.
- [19] Jolliffe I T, Cadima J. Principal component analysis: a review and recent developments [J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2016, 374 (2065): 20150202.
- [20] Liu Z C, Cai W S, Shao X E. Outlier detection in near - infrared spectroscopic analysis by using monte carlo cross - validation [J]. Science in China Series B - chemistry, 2008, 51 (8): 751 - 759.
- [10] Yu B, Zeng F, Yang Y, et al. Torsional vibrational modes of tryptophan studied by terahertz time - domain spectroscopy [J]. Biophysical Journal, 2004, 86 (3): 1649 - 1654.
- [11] 赵永会, 顾丹丹, 冯俊霞, 等. 二维红外光谱法分析纤维素[J]. 理化检验(化学分册), 2015, 51 (11): 1497 - 1500.
- [12] 梁瑜. 基于吸光度的波长筛选方法用于近红外光谱定量模型的优化[J]. 电子测试, 2016 (21): 62 - 64.
- [13] 刘明博, 唐延林, 李晓利, 等. 水稻叶片氮含量光谱监测中使用连续投影算法的可行性[J]. 红外与激光工程, 2014, 43 (4): 1265 - 1271.
- [14] 马浩彩, 郝秀美. 使用美棉纺细号纱棉结的控制[J]. 棉纺织技术, 2011, 39 (3): 142 - 144.

(上接第 209 页)

- [6] 颜志刚. 基于太赫兹时域光谱技术的生物分子和农药分子的检测技术研究[D]. 杭州: 浙江大学, 2008.
- [7] Yu W, Yang L, Zhang T L, et al. Crystal structure and geometry - optimization study of 4 - methyl - 3', 5' - dinitro - 4' - methyl benzylidene aniline [J]. Journal of Molecular Structure, 2006, 794 (1/2/3): 255 - 260.
- [8] 胡皆汉, 叶金星, 程国宝, 等. β - 榄香烯振动光谱的量子化学从头计算[J]. 光谱学与光谱分析, 2001, 21 (2): 163 - 168.
- [9] Fischer B M, Walther M, Uhd Jepsen P. Far - infrared vibrational modes of DNA components studied by terahertz time - domain spectroscopy [J]. Physics in Medicine and Biology, 2002, 47 (21): 3807 - 3814.