

基于太赫兹时域光谱技术的黄酮类化合物研究

殷明¹, 王建林¹, 黄浩亮¹, 黄秋萍², 杨萌萌¹, 傅正平³, 陆亚林^{2*}

1. 中国科学技术大学国家同步辐射实验室, 安徽 合肥 230026

2. 中国科学技术大学合肥微尺度物质科学国家研究中心, 安徽 合肥 230026

3. 中国科学技术大学化学与材料科学学院, 安徽 合肥 230026

摘要 黄酮类化合物常以游离或糖苷形式广泛存在于植物体内的一大类多酚物质, 具有抗氧化、抗菌、抗病毒、抑制肿瘤生长等药理作用, 作为一种重要的中药活性成分, 其具有较高的药用价值和开发前景。利用太赫兹时域光谱技术 (THz-TDS) 研究了 8 种常见的黄酮类化合物 (黄芩素、槲皮素、柚皮素、大豆素、黄芩苷、葛根素、染料木素和天麻素) 在 0.2~2.5 THz 波段的生物分子特性, 结果表明这些黄酮类物质在太赫兹波段具有明显不同的特征吸收峰, 并研究了它们在 78~320 K 范围内随温度变化的太赫兹吸收特性, 结果显示随着温度的降低, 特征吸收峰逐渐增强, 并且吸收峰频率位置发生蓝移。另外, 通过化学计量学方法结合太赫兹吸收谱对黄酮类物质进行定性鉴别和定量分析研究, 首先利用主成分分析 (PCA) 提取光谱特征变量, 然后将前五个主成分分量作为支持向量机 (SVM) 的输入变量建立分类模型, 通过优化模型选择最优参数, 最终得到 100% 的分类准确度。另外, 采用偏最小二乘回归 (PLSR) 模型和人工神经网络 (ANN) 模型对淀粉中含有不同浓度含量的黄酮类物质进行定量分析, 经过对比这两种方法, ANN 模型得到了最高的预测精度, 其中, 柚皮素和大豆素在预测集中的浓度预测相关系数分别为 $R^2=0.9944$, $R^2=0.9964$, 均方根误差分别为 $RMSE=1.9325$, $RMSE=1.5441$ 。综上所述, 利用 THz-TDS 技术研究了黄酮类物质在太赫兹波段的生物分子特性, 并结合化学计量学方法为黄酮类物质提供了一种快速、有效、无损的定性鉴别和定量分析方法, 在中草药的检测方面具有潜在的应用价值, 对其他生物分子的研究也具有较强的借鉴意义。

关键词 黄酮类化合物; 太赫兹时域光谱; 化学计量学; 定性鉴别; 定量分析

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)12-3919-06

引言

黄酮类化合物泛指两个苯环通过中央三碳链连接而成的一类化合物, 常以游离或糖苷形式广泛存在于植物体内的一大类多酚物质, 具有抗氧化、抗菌、抗病毒、抑制肿瘤生长等药理作用。目前中草药市场上以次充好、以假乱真的现象比较常见, 严重扰乱市场秩序和消费者身体健康, 为了确保药品质量和保护消费者权益, 采用一种可靠、有效的中草药检测方法是必不可少的。传统的中草药检测方法包括光谱法^[1]、色谱法^[2-3]、质谱法^[4-5]等, 但是每种方法各有优劣, 没有一种方法可以解决所有问题。

太赫兹 (Terahertz, THz) 波是指频率在 0.1~10 THz 范

围内的电磁波, 该频段内包含了大量有关物质的物理、化学和结构信息以及多种分子振动模式, 很多生物分子在太赫兹波段都有明显的特征吸收峰, 太赫兹谱学在食品、药品、生物医学领域有重要的应用价值^[6]。本文利用太赫兹时域光谱技术 (THz-TDS) 研究常见的黄酮类化合物在太赫兹波段的光谱特性, 同时利用几种化学计量学方法对不同种类黄酮类物质进行种类鉴别和定量分析, 为中草药的检测提供一种无损、快速、有效的分析方法, 在中药材市场检测中具有重要的应用价值。

1 实验部分

1.1 设备

收稿日期: 2019-09-12, 修订日期: 2020-01-18

基金项目: 国家自然科学基金项目 (51627901, 11704373), 国家重点研发计划项目 (2016YFA0401004, 2017YFA0402904), 中科大重要方向培育基金项目 (WK2340000071, WK2310000076) 资助

作者简介: 殷明, 1987年生, 中国科学技术大学国家同步辐射实验室博士后研究员 e-mail: yinming@ustc.edu.cn

* 通讯联系人 e-mail: yllu@ustc.edu.cn

本文实验装置使用自搭的透射式 THz-TDS 系统,其原理图如图 1 所示。通过测量得到太赫兹脉冲透过样品和参考信号的振幅和相位信息,利用样品信号 $[E_s(\nu)]$ 和参考信号 $[E_r(\nu)]$ 计算得到透射系数 $[T(\nu)]$,透射系数定义如式(1)

$$T(\nu) = \frac{E_s(\nu)}{E_r(\nu)} = \left[\frac{4n}{(n+1)^2} \right] \exp \left[-\frac{\alpha d}{2} + \frac{i2\pi\nu(n-1)d}{c} \right] \quad (1)$$

式(1)中, ν 是频率, d 为样品厚度, c 为真空中的光速, α 表示吸收系数, n 表示折射率,根据式(1)中的实部和虚部,可以计算出样品的折射率(n)和吸收系数(α)如式(2)和式(3)所示

$$n(\omega) = \varphi(\omega) \frac{c}{\omega d} + 1 \quad (2)$$

$$\alpha(\omega) = \frac{2k(\omega)\omega}{c} = \frac{2}{d} \ln \left[\frac{4n(\omega)}{\rho(\omega)[n(\omega)+1]^2} \right] \quad (3)$$

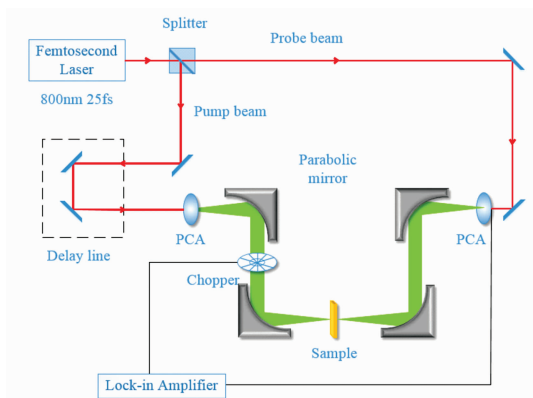
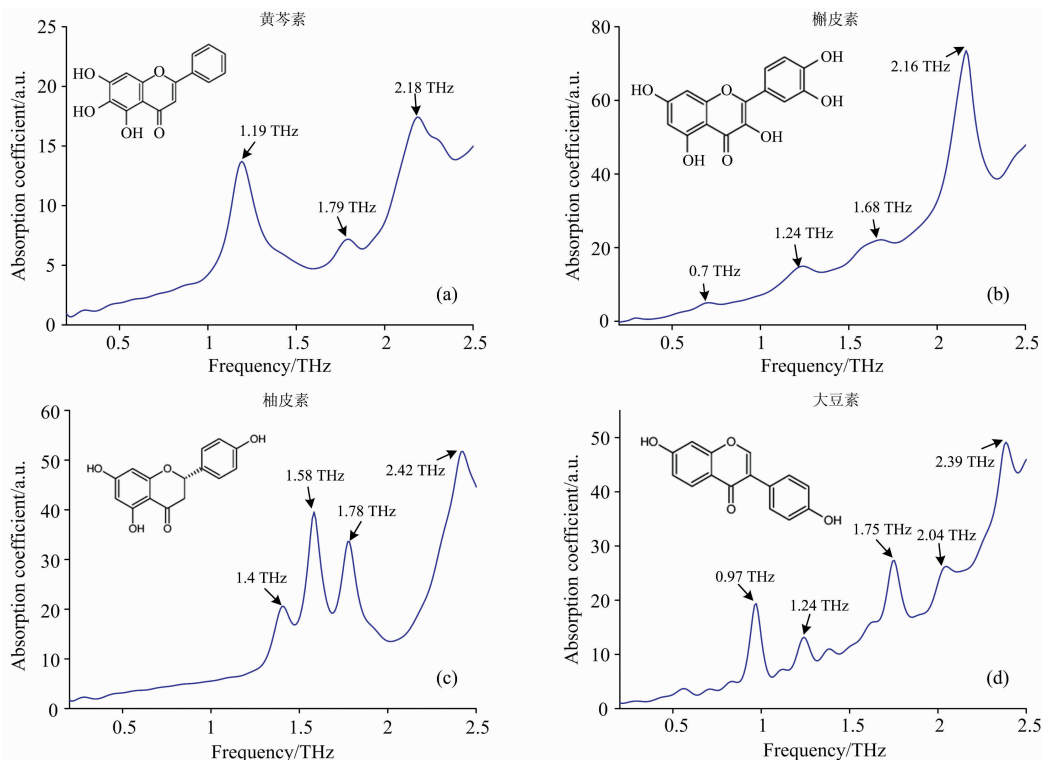


图 1 透射式太赫兹时域光谱系统原理图

Fig. 1 Schematic diagram of transmission terahertz time-domain spectroscopy system



其中, ω 是角频率, $\varphi(\omega)$ 为相位差, $\rho(\omega)$ 为幅值比, $k(\omega)$ 表示消光系数。

1.2 样品制备

研究了 8 种常见的黄酮类化合物,它们包括黄芩素、槲皮素、柚皮素、大豆素、黄芩苷、葛根素、染料木素和天麻素。所有样品纯度均大于 99%,并购买于同一批次,实验前均放在恒温恒湿箱中保存,未经过其他预处理。实验样品制备采用粉末压片法,首先将样品在玛瑙研钵中充分研磨,然后与高密度聚乙烯粉末(HDPE)按照 1:2 比例进行充分混合,最后在压片机下制成厚度为 1 mm 的薄片。在做分类鉴别研究时,每种黄酮类物质制备 10 个样品,一共 80 个样品作为待测样品;在做定量预测分析时,将黄酮类物质与淀粉混合配制不同浓度的待测样品,浓度范围从 10%~90% (ω/ω) (间隔 10%),每种浓度制备 10 个样品,共 90 个样品作为待测样品。在建立定性和定量模型时将每种黄酮类物质按照比例随机分成校正集(70%)和预测集(30%)。

2 结果与讨论

2.1 光谱分析

实验在室温下利用 THz-TDS 测量了所有黄酮类化合物样品,为了保证实验数据准确性,每个样品测量 3 次取平均值得到样品的太赫兹时域信号,原始时域信号经过 Savitsky-Golay 滤波预处理后,再通过计算得到样品在 0.2~2.5 THz 范围内的吸收系数,8 种黄酮类物质的 THz 吸收谱如图 2 所示。从图中可以看出,虽然这些黄酮类物质具有相似的分子结构,但每种物质在太赫兹波段都有明显不同的特征吸收峰,体现了太赫兹对生物分子的指纹谱特性。因此,可以通过太赫兹吸收谱对黄酮类物质进行种类鉴别。

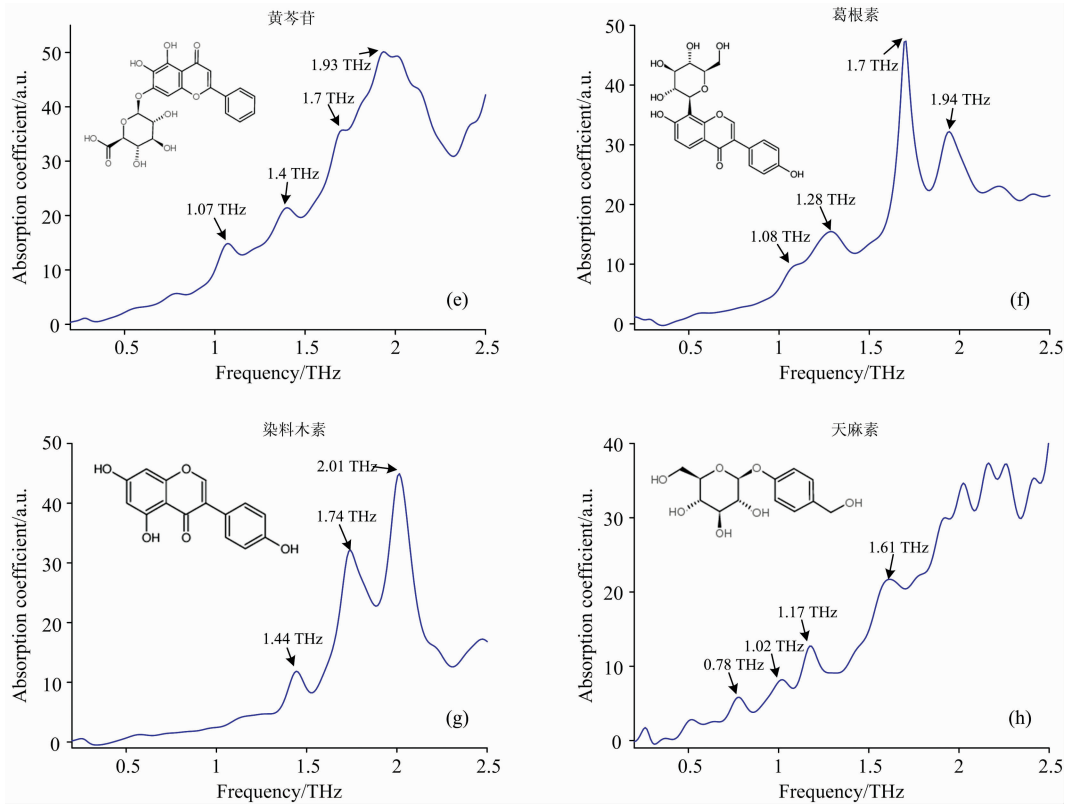


图 2 8 种黄酮类物质的太赫兹吸收谱

Fig. 2 Terahertz absorption spectra of eight flavonoids

2.2 低温分析

实验测试了柚皮素和大豆素在 78~320 K 温度范围内的太赫兹吸收光谱, 结果如图 3 所示, 从图中可以看出, 随着温度的降低, 柚皮素和大豆素的三个吸收峰逐渐增强, 并且吸收峰频率位置发生轻微的蓝移, 柚皮素的三个吸收峰 1.4, 1.58 和 1.78 THz 分别蓝移到 1.42, 1.6 和 1.80 THz, 平移量为 0.02 THz。大豆素的三个吸收峰 0.97, 1.24 和

1.75 THz 分别蓝移到 0.98, 1.25 和 1.76 THz, 平移量为 0.01 THz。通常, 随着温度的降低, 吸收峰变得更尖锐主要是能量振动态分布随温度变化的结果, 而由温度引起的吸收峰位置的频移被认为是多种机制作用的结果, 其中, 蓝移主要是由于热膨胀引起的键长增加以及振动势的非谐性导致的, 红移主要是由于分子间弱相互作用力导致的^[7-10]。

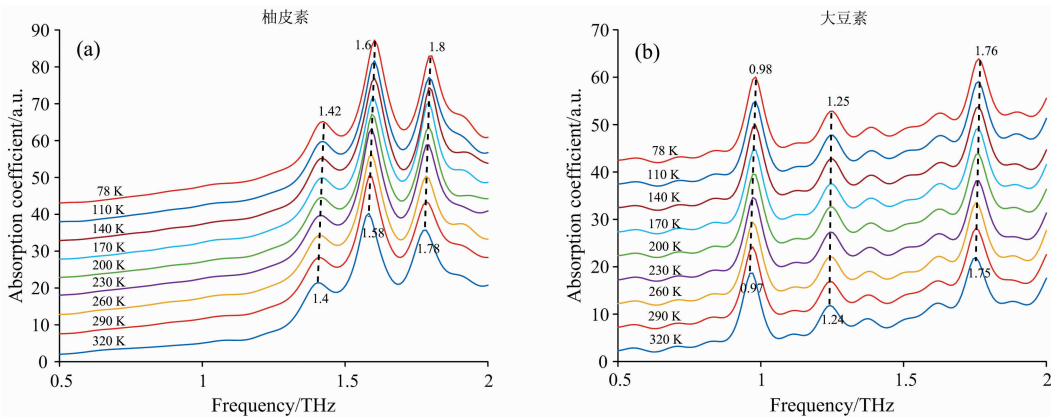


图 3 柚皮素和大豆素随温度变化的吸收谱

Fig. 3 Temperature dependent of absorption spectra of naringenin and daidzein

2.3 定性鉴别

主成分分析(principal component analysis, PCA)是一种基于统计学的无监督模式识别方法, 可以用来提取特征变

量、降低数据维度以及可视化数据结构。支持向量机(support vector machines, SVM)是一种机器学习算法, 具有很强的学习能力和泛化性能, 可以有效地解决非线性的分类问

题。本文首先通过 PCA 方法对原始光谱数据提取特征变量, 图 4 所示为所有样品吸收谱做主成分分析得到的三维得分图, 其中, 前五个主成分的累计方差贡献率超过了 98%, 代表了原始数据的主要信息。因此, 可以提取前五个主成分作为 SVM 分类模型的输入变量。利用 SVM 在校正集上建立分类模型, 选择 RBF 径向基核函数优化模型, 利用网格搜索寻找最优参数, 得到正则化参数 $c=0.7356$ 和核参数 $g=$

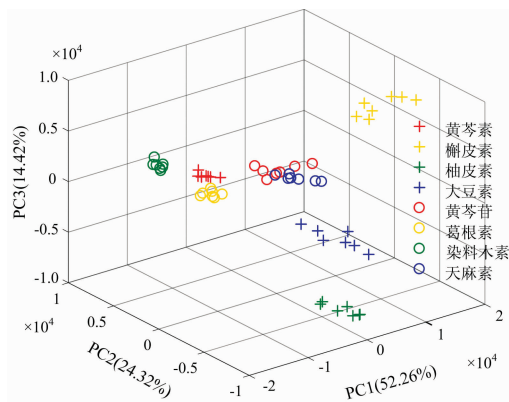


图 4 8 种黄酮类物质太赫兹吸收谱的 PCA 三维得分图

Fig. 4 PCA three-dimensional score map for terahertz absorption spectra of eight flavonoids

1.4353, 交叉验证 CV 达到 100% 最高值, 最后利用训练好的分类模型在预测集上对 8 种黄酮类物质做分类鉴别, 最终得到分类准确率达到 100%。对比原始光谱数据作为支持向量机的输入变量, 本文利用 PCA 提取前五个主成分分量作为 SVM 输入变量来建立校正模型, 不仅提取了特征变量, 而且减少了运算量, 提高了运行效率和模型精确度。

2.4 定量分析

利用偏最小二乘回归 (partial least squares regression, PLSR) 和人工神经网络 (artificial neural network, ANN) 回归模型对淀粉中不同含量的黄酮类物质进行定量分析, 其中 ANN 模型采用 RBF 神经网络。在做定量预测分析过程中, 通常用相关系数 R^2 和均方根误差 RMSE 两个指标来衡量模型的性能, 其中, R^2 和 RMSE 的表达式如式 (4) 和式 (5) 所示

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5)$$

其中, y_i 和 \hat{y}_i 分别为第 i 个样品的真实值和预测值, \bar{y} 是所有样品真实值的平均值, n 是样品个数。

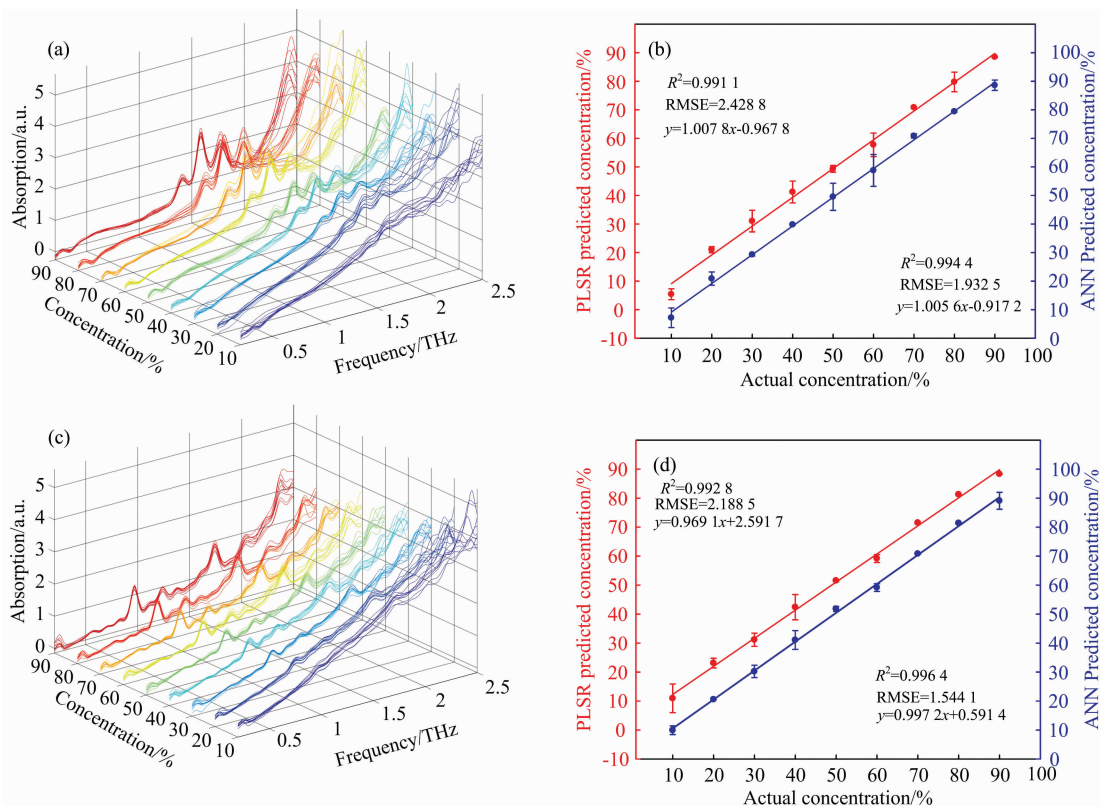


图 5 淀粉中不同浓度含量的柚皮素和大豆素的定量分析

(a) 和 (c) 分别为不同浓度的柚皮素和大豆素的太赫兹时域光谱; (b) 和 (d) 分别为柚皮素和大豆素的预测结果

Fig. 5 Quantitative analysis of different concentrations of naringenin and daidzein in starch

(a) and (c) are terahertz time-domain spectra of naringenin and daidzein with different concentrations;

(b) and (d) are the predicted results of naringenin and daidzein

实验测量得到了淀粉中不同浓度柚皮素和大豆素的太赫兹时域光谱如图 5(a) 和 (c) 所示。从图中可以看出, 随着淀粉中黄酮类物质浓度含量的增加, 吸收峰强度也随着增强。利用两种回归方法对淀粉中黄酮类物质的浓度建立了定量预测模型, 其预测结果如图 5(b) 和 (d) 所示, 图中显示了淀粉中黄酮类物质的真实浓度和预测浓度之间的关系。对淀粉中不同浓度含量的柚皮素的定量分析中, 利用 PLSR 回归模型得到的预测结果为: 有效成分含量预测值与其真实值之间的相关系数 $R^2=0.9911$, 均方根误差 $RMSE=2.4288$ 。利用 ANN 回归模型得到的预测结果为: $R^2=0.9944$, $RMSE=1.9325$ 。对淀粉中不同浓度含量的大豆素的定量分析中, 利用 PLSR 回归模型得到的预测结果为: $R^2=0.9928$, $RMSE=2.1885$ 。利用 ANN 回归模型得到的预测结果为: $R^2=0.9964$, $RMSE=1.5441$ 。对比两种模型预测结果可以看出, ANN 模型具有更高的相关系数以及更低的均方根误差, 因此, 结果证明了利用 THz-TDS 对淀粉中不同浓度含量的黄酮类化合物做定量分析时, ANN 模型具有较高的预测精度。

3 结 论

利用 THz-TDS 研究了 8 种常见的黄酮类化合物在

0.2~2.5 THz 波段的生物分子特性, 结果显示均具有明显不同的特征吸收峰, 并研究了它们在 78~320 K 温度范围内随温度变化的吸收特性, 结果表明, 随着温度的降低, 特征吸收峰逐渐增强, 主要是由能量振动态分布的影响, 并且吸收峰位置发生蓝移, 主要是由于热效应以及振动势的非谐性导致的。另外, 通过化学计量学方法对这些黄酮类化合物进行种类鉴别和定量分析, 首先利用 PCA 提取光谱特征变量, 然后将前五个主成分分量作为 SVM 的输入变量进行分类鉴别, 通过优化模型得到最优参数, 得到分类准确度为 100%。另外, 采用 PLSR 和 ANN 回归模型对淀粉中含有不同浓度的黄酮类物质做定量检测, 结果 ANN 方法得到了最高的预测精度。综上所述, 本文利用 THz-TDS 技术研究了黄酮类物质在太赫兹波段的生物分子特性, 并结合化学计量学方法为黄酮类物质提供了一种快速、有效、无损的分类鉴别和定量分析方法, 对中草药的检测和质量监控领域具有实际的应用价值。

References

- [1] SUN Heng, JIN Hang, HU Qiang, et al(孙恒, 金航, 胡强, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(6): 1702.
- [2] CAI Shuang, RUAN Cheng-jiang, DU Wei, et al(蔡爽, 阮成江, 杜维, 等). Journal of Analytical Science(分析科学学报), 2019, 35(3): 311.
- [3] YANG Li-xin, LIU Dai, FENG Xue-feng, et al(杨立新, 刘岱, 冯学锋, 等). China Journal of Chinese Materia Medica(中国中药杂志), 2002, (3): 166.
- [4] ZHANG Wei-bing, WANG Zhi-cong, ZHANG Ling-yi(张维冰, 王智聪, 张凌怡). Chinese Journal of Analytical Chemistry(分析化学), 2013, 41(12): 1851.
- [5] ZHANG Wei-bing, WANG Zhi-cong, ZHANG Ling-yi(张维冰, 王智聪, 张凌怡). Chinese Journal of Analytical Chemistry(分析化学), 2014, 42(3): 415.
- [6] LI Tian-ying, JIANG Ling, ZHANG Long, et al(李天莹, 蒋玲, 章龙, 等). Science and Technology of Food Industry(食品工业科技), 2019, (12): 359.
- [7] Zhou Lu, Chen Ligang, Ren Guanhua. Physical Chemistry Chemical Physics, 2018, 20, 27205.
- [8] Shen Y C, Upadhy P C, Linfield E H, et al. Appl. Phys. Lett., 2003, 82: 2350.
- [9] Takahashi M, Okamura N, Fan X, et al. J. Phys. Chem. A, 2017, 121: 2558.
- [10] Walther M, Fischer B W, Jepsen P U. Chem. Phys., 2003, 288: 261.

Research on Flavonoids Based on Terahertz Time-Domain Spectroscopy

YIN Ming¹, WANG Jian-lin¹, HUANG Hao-liang¹, HUANG Qiu-ping², YANG Meng-meng¹, FU Zheng-ping³,
LU Ya-lin^{2*}

1. National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei 230026, China

2. Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China, Hefei 230026, China

3. School of Chemistry and Materials Science, University of Science and Technology of China, Hefei 230026, China

Abstract Flavonoids are a large class of polyphenols widely distributed in plants in the form of free or glycoside, which has anti-oxidation, anti-bacterial, anti-viral, anti-tumor growth and other pharmacological effects. As an important active component of traditional Chinese medicine, they have high medicinal value and development prospect. In this paper, the biomolecular properties of eight common flavonoids, including baicalein, quercetin, naringenin, daidzein, baicalin, puerarin, genistein and gastrodin, were studied by terahertz time-domain spectroscopy (THz-TDS) in the 0.2~2.5 THz band. The results showed that these flavonoids have different characteristic absorption peaks in the terahertz band. The terahertz absorption characteristics with temperature variation in the range of 78~320 K were studied. The results showed that the characteristic absorption peaks gradually increased with the decrease of temperature, and the frequency position of absorption peak was blue-shifted. In addition, Qualitative identification and quantitative analysis of flavonoids were carried out by chemometrics combined with terahertz absorption spectra. First, the spectral characteristic variables were extracted by principal component analysis (PCA), then the first five principal components were used as input variables of support vector machine (SVM) to establish a classification model, and the optimal parameters were selected through the optimization model, and finally, the classification accuracy of 100% was obtained. In addition, the partial least squares regression (PLSR) model and the artificial neural network (ANN) model were used to analyze the flavonoids with different concentrations in starch quantitatively. By comparing the two methods, the ANN model obtained the highest prediction accuracy. The correlation coefficients of naringenin and daidzein in the prediction set were $R^2=0.9944$, $R^2=0.9964$, and the root means square error was $RMSE=1.9325$ and $RMSE=1.5441$, respectively. In summary, the biomolecular properties of flavonoids were studied by THz-TDS technology, and a rapid, effective and non-destructive qualitative identification and quantitative analysis of flavonoids were provided. This method has potential application value in the detection of Chinese herbal medicine, and has better reference significance for the study of other biomolecules.

Keywords Flavonoids; Terahertz time-domain spectroscopy; Chemometrics; Qualitative identification; Quantitative analysis

(Received Sep. 12, 2019; accepted Jan. 18, 2020)

* Corresponding author