

文章编号: 1005-5630(2020)05-0027-06

DOI: 10.3969/j.issn.1005-5630.2020.05.005

# 基于太赫兹光谱的人参和西洋参鉴别

寇天一

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

**摘要:** 为了快速区分人参和西洋参, 依据两者所含的人参皂苷在太赫兹波段具有不同指纹光谱, 提出了一种基于人参皂苷的太赫兹光谱鉴别人参和西洋参的方法。采用 MATLAB 软件识别人参和西洋参的太赫兹光谱, 同时根据太赫兹光谱特征并通过标准的主成分分析 (PCA) 区分出人参、西洋参。实验结果表明, 基于太赫兹光谱的主成分分析可精确区分出人参、西洋参, 并且还可以应用于其他类似物质的区分中。

**关键词:** 人参; 西洋参; 皂苷; 太赫兹; 主成分分析 (PCA)

中图分类号: O 433.4 文献标志码: A

## Terahertz spectroscopy for accurate identification of ginseng and panax quinquefolium

KOU Tianyi

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**Abstract:** Since ginsenosides contained in the two herbs have different fingerprint spectra in the terahertz band, a method based on ginsenoside terahertz spectrum was proposed for the identification of ginseng and panax quinquefolium. The terahertz spectra of ginseng and panax quinquefolium were identified by MATLAB software, and the ginseng and panax quinquefolium were distinguished by the standard principal component analysis (PCA) according to their terahertz spectral characteristics. The results show that the principal component analysis based on terahertz spectrum can distinguish ginseng and panax quinquefolium accurately, and it can also be applied to other similar substances.

**Keywords:** ginseng; panax quinquefolium; saponins; terahertz; principal component analysis(PCA)

收稿日期: 2020-05-11

作者简介: 寇天一 (1995—), 男, 硕士研究生, 研究方向为太赫兹光谱学。E-mail:173740584@st.usst.edu.cn

## 引 言

人参和西洋参都是五加科人参属的草本植物,其干燥的根和根茎部位可入药,它们的主要化学成分为皂苷类,化学结构相当复杂,药用价值极高<sup>[1-2]</sup>。迄今为止,在人参和西洋参中已经鉴定出超过 60 多种人参皂苷,《中国药典》记载人参中最重要的成分是人参皂苷 Rg<sub>1</sub>、Rb<sub>1</sub> 和 Re,西洋参和人参十分相似,唯一的区别就是西洋参含有独有的拟人参皂苷 F<sub>11</sub>。虽然人参和西洋参可以通过传统方法(例如:基源鉴定、性状鉴定、显微鉴定)进行鉴别,但是这些鉴别手段存在主观性、片面性等问题,而记载于《中国药典》的高效液相色谱法(HPLC)是目前最权威、最全面的方法。例如,Wang 等<sup>[3]</sup>在 1999 年使用高效液相色谱-串联质谱(HPLC-MS/MS),Chan 等<sup>[4]</sup>在 2000 年使用高效液相色谱-电喷雾质谱(HPLC-ESI-MS),根据人参皂苷的分布来区分西洋参和人参。Ma 等<sup>[5]</sup>在 2006 年利用高效液相色谱-大气压化学电离质谱(HPLC-APCI-MS)对西洋参中 F<sub>11</sub> 的含量进行了研究,发现可以用一种基于 majonoside 异构体和人参皂苷 Rd 作为标记物来鉴定来自中国 and 北美的西洋参。然而,高效液相色谱法是基于物质的分离(依赖于吸附特性、表面电荷、配体特异性和蛋白质分子大小的差异),需要复杂的预处理过程,且成本高、耗时长<sup>[6-7]</sup>。因此,迫切需要寻找一种准确、快速的分析方法来鉴别人参和西洋参。

太赫兹(THz)光谱具有安全、无损、快速、易穿透等特点<sup>[8]</sup>,因此已被应用于物理学<sup>[9-10]</sup>、生物学<sup>[11-12]</sup>、化学<sup>[13-14]</sup>、药理学<sup>[15]</sup>等多个研究领域。太赫兹波段涵盖了大多数有机分子的振动和旋转运动的特征吸收频率。因此,太赫兹指纹图谱可用于物质的识别和定量分析,并且该技术已被用于检测肿瘤代谢物同分异构体 2HG<sup>[16]</sup>、甲基乙二醛<sup>[17]</sup>、同型半胱氨酸等<sup>[18]</sup>。如果 THz 光谱可以识别人参和西洋参中的特异性皂苷,则有助于人参和西洋参的鉴别。

本文基于人参和西洋参的 THz 光谱特征,通过主成分分析(PCA)研究了人参和西洋参的鉴别方法。

## 1 实验材料与方法

### 1.1 实验材料及样本制备

材料:人参皂苷 Rg<sub>1</sub>(>98%, CAS: 22427-39-0, 购买自 Sigma-Aldrich, 中国上海),拟人参皂苷 F<sub>11</sub>(>98%, CAS: 69884-00-0, 购买自 PureChem Standard, 中国成都),环烯烃共聚物(COC)粉(颗粒大小 50~100 $\mu\text{m}$  产自上海核研究所)。由于 COC 在太赫兹区的折射率弥散可以忽略,吸收可以忽略,因此非常适合于太赫兹光谱的分析<sup>[19]</sup>。实验所用的样本均是人参和西洋参块状根部,且所有样本均未进一步纯化。

在进行太赫兹光谱分析前,被检的人参、西洋参样本需要经过研磨、筛分和压片过程。人参和西洋参在 MM400 球磨机(Retsch, 德国)中研磨成颗粒大小为 40~50  $\mu\text{m}$  的粉末状,粉末样品过筛后与 COC 均匀混合,然后用压片机(4 t 压力)将样品压缩成厚 1 mm、直径 13 mm 的片剂。

### 1.2 实验仪器

实验装置为德国 Bruker 公司生产的傅里叶变换红外光谱仪,型号为 Vectex80v。其中远红外模块光源为水冷汞灯,检测器为 DLATGS/聚乙烯检测器,光谱有效覆盖区为 30~680  $\text{cm}^{-1}$ ,信噪比(SNR)优于 10000:1。仪器的光谱范围为 0.9~14 THz、分辨率为 2  $\text{cm}^{-1}$ 、扫描频率为 5 kHz,环境室温约为 22  $^{\circ}\text{C}$ ,样品仓保持真空以减少水蒸气对实验的影响。

### 1.3 主成分分析

主成分分析法(principal component analysis, PCA)是一种简化数据集的统计抽取方法。PCA 基本思路为借助正交变换,将相关联的原变量转为不相关的新变量,即将原变量系统变换成新的正交系统,使之指向样本点散布最开的正交方向,进而对多维变量系统进行降维处理。

设原变量为  $x_1, x_2, \dots, x_J$ , 主成分分析后得到的新变量为  $z_1, z_2, \dots, z_m$ , 均是  $x_1, x_2, \dots, x_J$  的线性组合( $m < J$ ,  $J$  是原变量的维数,  $m$  是新变量保留原变量最大的维数)。变量  $z_1, z_2, \dots, z_m$  构成

的坐标系是在原来坐标系经平移和正交旋转后得到的, 称  $z_1, z_2, \dots, z_m$  形成的空间为  $m$  维主平面。在主平面上, 第一主分量  $z_1$  对应于数据变异最大的方向, 对于  $z_1, z_2, \dots, z_m$ , 依次有  $V(z_1) \geq V(z_2) \geq \dots \geq V(z_m)$ ,  $V$  是对应维度变量的主成分得分。因此,  $z_1$  是携带原数据信息最多的一维变量<sup>[20]</sup>, 称为第一主成分(PC<sub>1</sub>), 第二坐标的第二大方差为第二主成分(PC<sub>2</sub>), 等等。基本算法和步骤如下:

(1) 数据标准化处理

$$y_{ij} = \frac{x_{ij} - x_j}{S_j}, (i = 1, 2, \dots, I; j = 1, 2, \dots, J) \quad (1)$$

式中:  $x_{ij}$  为第  $i$  个分区第  $j$  个指标的值;  $x_j$ 、 $S_j$  为第  $j$  个指标样本的样本均值和样本标准差, 其中  $I$  为数据表的分区数。

(2) 计算数据表  $(y_{ij})_{I \times J}$  的相关矩阵  $R$ 。

(3) 求  $R$  的  $J$  个特征值  $\lambda_1, \lambda_2, \dots, \lambda_J$ , 其中  $\lambda_1 \geq \lambda_2, \dots, \geq \lambda_J$ , 并求特征向量  $u_1, u_2, \dots, u_J$ , 而  $u_1, u_2, \dots, u_J$  又称为主轴。

(4) 计算主成分

$$z_k = \sum_{j=1}^J u_j x_j, (j = 1, 2, \dots, J; k = 1, 2, \dots, J) \quad (2)$$

本文提取  $z_1$  和  $z_2$  作为 PC<sub>1</sub> 和 PC<sub>2</sub>, 用于样品的定性分析。

## 2 实验结果和分析

人参皂苷 Rg<sub>1</sub> 属于达玛烷家族四环三萜皂苷

成员, 它的结构中含有 30 个碳原子, 这与甾醇十分相似。人参皂苷 Rg<sub>1</sub> (属于原三醇人参型皂苷, 分子量为 801.01) 具有兴奋中枢神经, 防止性功能减退, 增强记忆, 恢复疲劳, 促进 DNA、RNA 合成, 抗血小板凝集等作用。拟人参皂苷 F<sub>11</sub> (一种 ocotillol 型皂苷, 分子量为 801.02) 是达玛烷家族的三萜皂苷成员, 含有四环类固醇刚性骨架, 具体很好的治疗心肌缺血、脑缺血和中风的作用<sup>[21-22]</sup>。

将 Rg<sub>1</sub> 和 F<sub>11</sub> 这两种标志物皂苷按 25% 质量分数制备成药片状, 在同一天置于相同环境检测 8 次, 以获得这两种纯品皂苷的太赫兹吸收谱及其误差棒。

人参皂苷 Rg<sub>1</sub> 和拟人参皂苷 F<sub>11</sub> 的太赫兹光谱如图 1 所示。图 1(a) 中, 人参皂苷 Rg<sub>1</sub> 的吸收峰位为 8.36 THz 和 9.54 THz, 在较低频段并未发现太赫兹吸收峰。而拟人参皂苷 F<sub>11</sub> 则与人参皂苷 Rg<sub>1</sub> 大不相同, 如图 1(b) 所示, 拟人参皂苷 F<sub>11</sub> 在较低频段 1.76 THz、2.31 THz、3.11 THz 和 3.61 THz 处均有明显的吸收峰, 这些吸收峰相较于人参皂苷有巨大的差异。出现这种现象的原因是, 相较于其他皂苷, 拟人参皂苷 F<sub>11</sub> 中存在单独的不饱和双键(C=C 键), 主要发生的是非共轭的  $\pi$  轨道的  $\pi$ - $\pi^*$  跃迁, 其分子结构中不存在共轭体系, 是一种非共轭物质, 而人参皂苷 Rg<sub>1</sub> 等达玛烷家族四环三萜皂苷中均含有共轭体系, 这是两者本质的区别。此外, 图 1 中曲线上非常小的误差棒反映了 THz 光谱分析的再现性和准确性。

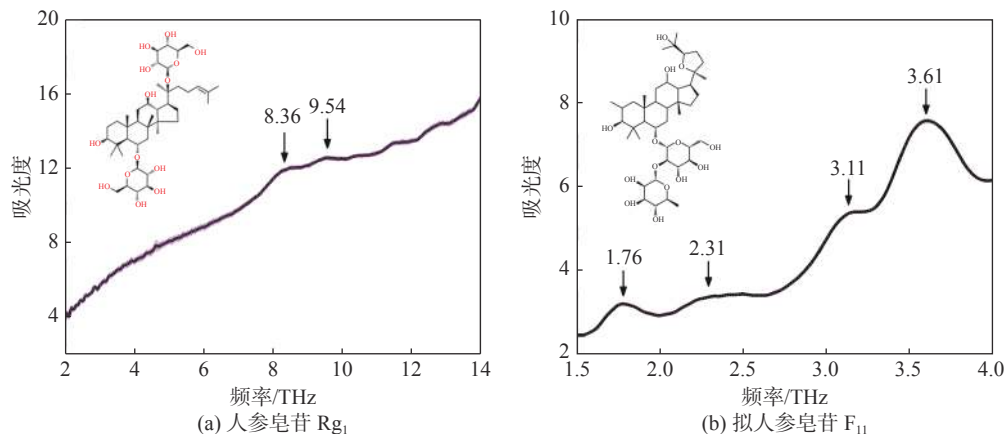


图 1 人参皂苷的太赫兹光谱  
Fig. 1 Terahertz spectra of ginsenoside

在证明了太赫兹光谱指纹可用于特异性皂苷的识别后，我们进一步做了人参和西洋参样本的太赫兹光谱测试。图 2 为人参和西洋参的太赫兹光谱，其中(a)~(c)为 3 种不同批次的人参样本，而(d)~(f)为 3 种不同批次的西洋参样本。

将这 6 个样本按上述制备过程制备成同等比例的样本压片，于同一天在相同环境下对每个样本检测 4 次，以获得人参和西洋参的太赫兹吸收谱和误差棒。

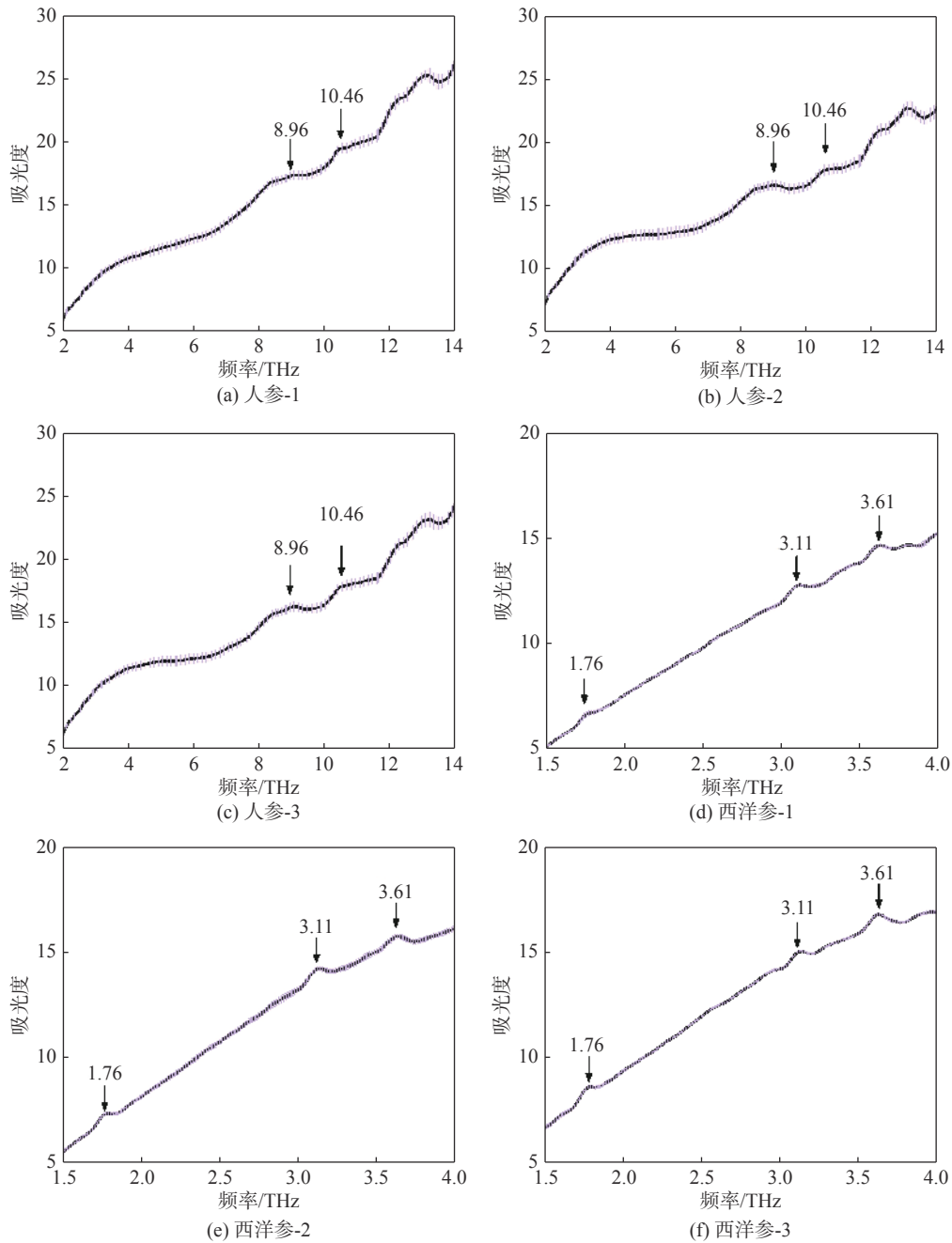


图 2 人参和西洋参的太赫兹光谱

Fig. 2 Terahertz spectra of ginseng and *Panax quinquefolium*

由图 2(a)、(b)、(c)可知，人参的光谱吸收峰与其特异性皂苷  $Rg_1$  峰位相近，均在 8~11 THz

范围内，具体峰位为 8.96 THz 和 10.46 THz。由图 2(d)、(e)、(f)可知，除了 2.31 THz 的吸

收峰太小无法检测外,所有西洋参样品均出现了拟人参皂苷  $F_{11}$  的吸收峰,相较于人参,西洋参与其特异性皂苷的峰位对应更精准。同时,在图2的6张光谱曲线上,非常小的误差棒同样反映了太赫兹光谱在人参和西洋参检测分析上的再现性和准确性。实验证明,通过人参和西洋参的太赫兹光谱吸收峰的位置,可以精准地分辨这2种五加科草本植物。

为了确定人参和西洋参的太赫兹光谱是否还可以作为与其他草本植物和非草本物质的定性鉴别依据,我们基于 MATLAB 软件,采用主成分分析(PCA)算法进一步进行分类鉴定。

用于定性分析的样本除了上述的三批次的人参和西洋参外,还对其他的草本植物(白芍和赤芍)和一些氨基酸物质(谷氨酸、赖氨酸、同型半胱氨酸、色氨酸)进行了定性分析。

图3为12个样本太赫兹光谱数据的主成分得分图( $PC_1$  值为 97.62%,  $PC_1+PC_2$  值为 99.31%)。当前  $p$  个主成分的贡献率足够大( $PC_1+PC_2+\dots+PC_p > 85.00\%$ )时,可以用前  $p$  个主成分近似代替原始数据集<sup>[23]</sup>。因此,这些光谱数据可以用  $PC_1$  和  $PC_2$  进行近似分析。在主成分分析中,任意2个样本的  $PC_1$  得分越相似,说明它们之间的相似性越强,反之,样本的  $PC_1$  得分偏差越大,差异越大<sup>[24]</sup>。可以从图3看出,人参和西洋参有十分相似的  $PC_1$  得分, $PC_1$  得分集中于 80.00% 左右。究其原因,是人参和西洋参同为五加科人参属草本植物,具有十分类似的主要成

分,例如人参皂苷  $Rb_1$ 、 $Rg_1$  和  $Re$  以及一些多糖等。同理,4种氨基酸物质也有非常相似的  $PC_1$  得分,这说明它们的主成分相差不大。 $PC_2$  得分的差异是源于各样本存在不同的物质组成成分,比如西洋参独有的拟人参皂苷  $F_{11}$ 。人参、西洋参这2种五加科人参属植物与谷氨酸、赖氨酸、同型半胱氨酸、色氨酸这4种非草本植物的物质  $PC_1$  得分相差很大,位居得分图的两侧,这说明草本植物与非草本物质的差异非常的大,因此可以很容易鉴别。另外,还可以看出,非五加科草本植物赤芍和白芍的  $PC_1$  得分介于五加科植物和氨基酸物质之间,这说明不同属的草本植物也可以被准确鉴别。对于得分图中的多种物质,完全由不同的成分决定了它们的显著差异。

### 3 结 论

本文基于太赫兹光谱,采用主成分分析(PCA)对人参和西洋参进行了鉴别。实验证明,采用太赫兹光谱能够快速、准确地区分人参皂苷  $Rg_1$  和拟人参皂苷  $F_{11}$ ,而结合 PCA 算法可更有效地区别同属五加科植物的人参和西洋参。本文鉴别方法还可以应用于其他草本植物和氨基酸物质的区分。

#### 参考文献:

- [1] 胡敏,李毓群.人参、西洋参和三七的药效比较[J].*海峡药学*,2011,23(7):112-113.
- [2] 董晓强,董文天,洪霞,等.三七、人参和西洋参化学成分与药效学之间的关系[J].*承德医学院学报*,2011,28(3):307-309.
- [3] WANG X M, SAKUMA T, ASAFU-ADJAYE E, et al. Determination of ginsenosides in plant extracts from *Panax ginseng* and *Panax quinquefolius* L. by LC/MS/MS[J]. *Analytical Chemistry*, 1999, 71(8): 1579-1584.
- [4] CHAN T W D, BUT P P H, CHENG S W, et al. Differentiation and authentication of *Panax ginseng*, *Panax quinquefolius*, and ginseng products by using HPLC/MS[J]. *Analytical Chemistry*, 2000, 72(6): 1281-1287.

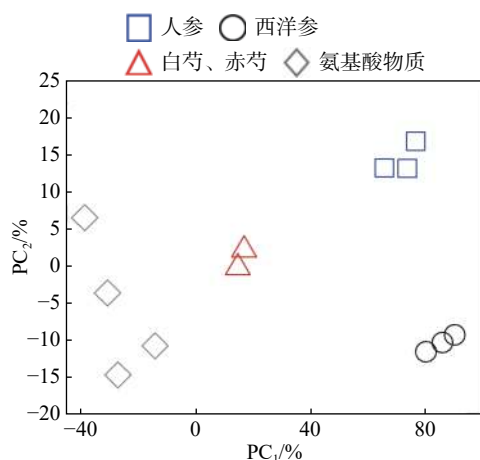


图3 12个样本的主成分得分图。

Fig. 3 Principal component score of 12 samples.

- [5] MA X Q, XIAO H B, LIANG X M. Identification of Ginsenosides in *Panax quinquefolium* by LC-MS[J]. *Chromatographia*, 2006, 64(1/2): 31 – 36.
- [6] FRANTZEN F, FAAREN A L, ALFHEIM I, et al. Enzyme conversion immunoassay for determining total homocysteine in plasma or serum[J]. *Clinical Chemistry*, 1998, 44(2): 311 – 316.
- [7] VERVOORT N, DAEMEN D, TÖRÖK G. Performance evaluation of evaporative light scattering detection and charged aerosol detection in reversed phase liquid chromatography[J]. *Journal of Chromatography A*, 2008, 1189(1/2): 92 – 100.
- [8] GE H Y, JIANG Y Y, LIAN F Y, et al. Quantitative determination of aflatoxin B1 concentration in acetonitrile by chemometric methods using terahertz spectroscopy[J]. *Food Chemistry*, 2016, 209: 286 – 292.
- [9] BILBRO L S, AGUILAR R V, LOGVENOV G, et al. Temporal correlations of superconductivity above the transition temperature in  $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$  probed by terahertz spectroscopy[J]. *Nature Physics*, 2011, 7(4): 298 – 302.
- [10] WANG Z, WU J D, YANG W, et al. Experimental observation of Bethe strings[J]. *Nature*, 2018, 554(7691): 219 – 223.
- [11] PENG Y, SHI C J, XU M Q, et al. Qualitative and quantitative identification of components in mixture by terahertz spectroscopy[J]. *IEEE Transactions on Terahertz Science and Technology*, 2018, 8(6): 696 – 701.
- [12] YANG X, ZHAO X, YANG K, et al. Biomedical applications of terahertz spectroscopy and imaging[J]. *Trends in Biotechnology*, 2016, 34(10): 810 – 824.
- [13] EL HADDAD J, DE MIOLLIS F, BOU SLEIMAN J, et al. Chemometrics applied to quantitative analysis of ternary mixtures by terahertz spectroscopy[J]. *Analytical Chemistry*, 2014, 86(10): 4927 – 4933.
- [14] ZHANG M J, HONG H C, LIN H J, et al. Mechanistic insights into alginate fouling caused by calcium ions based on terahertz time-domain spectra analyses and DFT calculations[J]. *Water Research*, 2018, 129: 337 – 346.
- [15] DU S Q, LI H, XIE L, et al. Vibrational frequencies of anti-diabetic drug studied by terahertz time-domain spectroscopy[J]. *Applied Physics Letters*, 2012, 100(14): 143702.
- [16] CHEN W Q, PENG Y, JIANG X K, et al. Isomers Identification of 2-hydroxyglutarate acid disodium salt (2HG) by Terahertz Time-domain Spectroscopy[J]. *Scientific Reports*, 2017, 7(1): 12166.
- [17] WU X, DAI Y, WANG L P, et al. Diagnosis of methylglyoxal in blood by using far-infrared spectroscopy and o-phenylenediamine derivation[J]. *Biomedical Optics Express*, 2020, 11(2): 963 – 970.
- [18] LI T, MA H Y, PENG Y, et al. Gaussian numerical analysis and terahertz spectroscopic measurement of homocysteine[J]. *Biomedical Optics Express*, 2018, 9(11): 5467 – 5476.
- [19] ZHU Z J, CHENG C, CHANG C, et al. Characteristic fingerprint spectrum of neurotransmitter norepinephrine with broadband terahertz time-domain spectroscopy[J]. *Analyst*, 2019, 144(8): 2504 – 2510.
- [20] 傅湘, 纪昌明. 区域水资源承载能力综合评价——主成分分析法的应用 [J]. *长江流域资源与环境*, 1999, 8(2): 168 – 173.
- [21] FUZZATI N. Analysis methods of ginsenosides[J]. *Journal of Chromatography B*, 2004, 812(1/2): 119 – 133.
- [22] ATTELE A S, WU J A, YUAN C S. Ginseng pharmacology: multiple constituents and multiple actions[J]. *Biochemical Pharmacology*, 2000, 58(11): 1685 – 1693.
- [23] 杨蕊菁, 赵磊, 夏鹏飞, 等. 不同产地当归药材 9 种成分的含量测定及主成分分析 [J]. *中药新药与临床药理*, 2020, 31(4): 473 – 477.
- [24] ZHAN H L, WU S X, BAO R M, et al. Qualitative identification of crude oils from different oil fields using terahertz time-domain spectroscopy[J]. *Fuel*, 2015, 143: 189 – 193.

(编辑: 刘铁英)