

## 基于化学成分群加和性分子描述符的中药浸膏粉溶化性预测研究

李文静<sup>1,2,3</sup>, 齐飞宇<sup>1</sup>, 霍晓乾<sup>1</sup>, 卢雪明<sup>4</sup>, 张志强<sup>5</sup>, 张燕玲<sup>1,2,3</sup>, 乔延江<sup>1,2,3\*</sup>, 徐冰<sup>1,2,3\*</sup>

1. 北京中医药大学 中药信息学系, 北京 102400
2. 北京市科委 中药生产过程控制与质量评价北京市重点实验室, 北京 102400
3. 中药制药与新药开发教育部工程研究中心, 北京 102400
4. 山东省食品药品审评查验中心, 山东 济南 250014
5. 北京康仁堂药业有限公司, 北京 101300

**摘要:** 目的 基于化学成分群加和性分子描述符建立中药浸膏粉溶化性预测模型, 并进行验证。方法 采用原位浊度传感器评价中药浸膏粉溶化性, 对测试方法条件进行优化并进行验证; 以 13 种代表性中药 (广藿香、当归、薄荷、瞿麦、银柴胡、乌梅、葛根、细辛、益母草、苦参、熟地黄、荆芥和川芎) 浸膏粉为研究对象, 检索每种中药所含化学成分, 根据成分的分子描述符计算出每种中药化学成分群平均性质, 进而采用偏最小二乘 (partial least squares, PLS) 法建立该平均性质与对应中药溶液浊度值的关联预测模型并验证。结果 中药浸膏粉溶化性客观评价方法的重复性和耐用性良好; 对从 13 种中药中检索得到的 1840 种成分分子描述符进行主成分分析后, 由得分图和载荷图结果可知分子描述符可以区分出溶解度有差异的化学成分类别; 随机划分校正集和验证集后, 基于化学成分群加和性分子描述符预测溶液浊度值的 PLS 预测模型的  $R^2_{X\text{ cum}}=0.873$ ,  $R^2_{Y\text{ cum}}=0.998$ ,  $Q^2_{\text{cum}}=0.869$ ; 川芎、荆芥和益母草的浊度预测误差分别为 3.4%、9.3%、23.8%。结论 基于化学成分群加和性分子描述符的中药浸膏粉溶化性预测模型具有一定的预测能力, 研究结果为从中药整体成分层面预测宏观物性提供参考。

**关键词:** 中药浸膏粉; 溶化性; 化学成分群; 分子描述符; 浊度传感器; 广藿香; 当归; 薄荷; 瞿麦; 银柴胡; 乌梅; 葛根; 细辛; 益母草; 苦参; 熟地黄; 荆芥; 川芎; 偏最小二乘法; 主成分分析

中图分类号: R283.6 文献标志码: A 文章编号: 0253-2670(2022)22-7029-10

DOI: 10.7501/j.issn.0253-2670.2022.22.006

## Prediction of solubility of traditional Chinese medicine extract powder from additive molecular descriptors of chemical ingredients group

LI Wen-jing<sup>1,2,3</sup>, QI Fei-yu<sup>1</sup>, HUO Xiao-qian<sup>1</sup>, LU Xue-ming<sup>4</sup>, ZHANG Zhi-qiang<sup>5</sup>, ZHANG Yan-ling<sup>1,2,3</sup>, QIAO Yan-jiang<sup>1,2,3</sup>, XU Bing<sup>1,2,3</sup>

1. Department of Chinese Medicine Information Science, Beijing University of Chinese Medicine, Beijing 102400, China
2. Beijing Key Laboratory for Production Process Control and Quality Evaluation of Traditional Chinese Medicine, Beijing Municipal Science & Technology Commission, Beijing 102400, China
3. Engineering Research Center of Chinese Medicine Production and New Drug Development, Ministry of Education of People's Republic of China, Beijing 102400, China
4. Center for Food and Drug Inspection of Shandong, Jinan 250014, China
5. Beijing Temages Pharmaceutical Co., Ltd., Beijing 101300, China

**Abstract: Objective** The prediction model of the solubility of traditional Chinese medicine extract powder was developed and verified from the additive molecular descriptors of chemical ingredients group. **Methods** The *in-situ* turbidity sensor was used to

收稿日期: 2022-06-01

基金项目: 国家自然科学基金项目 (82074033); 创新药物与高效节能降耗制药设备国家重点实验室开放基金项目 (GZSYS202007); 连云港市重大技术攻关“揭榜挂帅”项目 (CGJBGS2101)

作者简介: 李文静, 女, 硕士研究生, 专业方向为中药制剂新技术。Tel: 18810629117 E-mail: liwenjing23@126.com

\*通信作者: 徐冰, 副教授, 研究方向为中药质量与先进工艺控制。Tel: (010)53912118 E-mail: xubing@bucm.edu.cn

乔延江, 教授, 博士生导师, 主要从事中药信息学研究。Tel: (010)53912117 E-mail: yjqiao@263.net

evaluate the solubility of traditional Chinese medicine extract powder, and the test method conditions were optimized and verified. Taking extract powders of 13 representative traditional Chinese medicine [Guanghuoxiang (*Pogostemonis Herba*), Danggui (*Angelicae Sinensis Radix*), Bohe (*Menthae Haplocalycis Herba*), Qumai (*Dianthi Herba*), Yinchaihu (*Stellariae Radix*), Wumei (*Mume Fructus*), Gegen (*Puerariae Lobatae Radix*), Xixin (*Asari Radix et Rhizoma*), Yimucao (*Leonuri Herba*), Kushen (*Sophorae Flavescentis Radix*), Shudihuang (*Rehmanniae Radix Praeparata*), Jingjie (*Schizonepetae Herba*), and Chuanxiong (*Chuanxiong Rhizoma*)] as the research objects, the chemical components contained in each traditional Chinese medicine were retrieved, and the average properties of the chemical components of each traditional Chinese medicine were calculated according to the molecular descriptors of the components. Then, the partial least squares (PLS) method was used to establish and verify the correlation prediction model between the average property and the turbidity value of the corresponding traditional Chinese medicine solution. **Results** The objective evaluation method of the solubility of traditional Chinese medicine extract powder had good repeatability and durability. After principal component analysis of 1840 components molecular descriptors retrieved from 13 traditional Chinese medicines, it can be seen from the score plot and the load plot that the molecular descriptors can distinguish the chemical component categories with different solubility. The calibration set and the validation set were randomly divided, and the PLS prediction model of the solution turbidity value predicted based on the additive molecular descriptor of the chemical composition group had  $R^2_{X\text{cum}} = 0.873$ ,  $R^2_{Y\text{cum}} = 0.998$ ,  $Q^2_{\text{cum}} = 0.869$ . The turbidity prediction errors of *Chuanxiong Rhizoma*, *Schizonepetae Herba* and *Leonuri Herba* were 3.4%, 9.3% and 23.8%, respectively. **Conclusion** The solubility prediction model of traditional Chinese medicine extract powder from the additive molecular descriptors of chemical ingredients group had certain predictive ability, and the research results provided a reference for predicting macroscopic physical properties at the holistic composition level of traditional Chinese medicine.

**Key words:** traditional Chinese medicine extract powder; solubility; chemical constituents; molecular descriptors; turbidity sensor; *Pogostemonis Herba*; *Angelicae Sinensis Radix*; *Menthae Haplocalycis Herba*; *Dianthi Herba*, *Stellariae Radix*; *Mume Fructus*; *Puerariae Lobatae Radix*; *Asari Radix et Rhizoma*; *Leonuri Herba*; *Sophorae Flavescentis Radix*; *Rehmanniae Radix Praeparata*; *Schizonepetae Herba*; *Chuanxiong Rhizoma*; partial least squares; principal component analysis

中药浸膏粉是中药口服固体制剂成型工艺的原料之一<sup>[1]</sup>。在中药颗粒剂成型工艺中,制剂处方的辅料一般是可溶性辅料,因此颗粒产品的溶化性主要受浸膏粉溶化性的影响<sup>[2-3]</sup>。在颗粒剂工艺设计中,对浸膏粉溶化性进行评估,可指导前序工艺改进和处方辅料的筛选。中药提取物的化学成分组成及含量影响其溶化性<sup>[4]</sup>,如王洋等<sup>[5]</sup>在对荔枝核总皂苷颗粒剂进行溶化性检查时发现,颗粒溶液出现一定量的沉淀,与颗粒剂中的黄酮类和皂苷类成分的水溶性较差有关。

分子描述符(molecular descriptors)是指分子在某一方面性质的度量,如分子的物理化学性质或根据分子结构推算出来的数值指标<sup>[6]</sup>,其在药物溶解度的预测方面应用广泛。Mettou等<sup>[7]</sup>为改善化合物的水溶性,测定了39种化合物在水中[含有5%二甲基亚砜(dimethyl sulfoxide, DMSO)作为共溶剂]的溶解度,构建了化合物溶解度的定量结构-性质关系(quantitative structure-property relationship, QSPR)模型,并对验证集中16种化合物的溶解度进行测试后验证发现模型具有较高准确度。朱腾义等<sup>[8]</sup>基于分子描述符,研究了109个化合物在聚二甲基硅氧烷(polydimethylsiloxane, PDMS)中的溶

解度,利用逐步多元线性回归方法构建了预测模型 [ $R^2=0.916$ , 均方根误差(root mean squared error, RMSE)=0.264]。

随着中药物质基础研究的深入,可获得的中药化学成分信息越来越多,使得基于中药化学成分群信息研究中药的功能和性质成为可能,如中药药性和功效预测等<sup>[9-11]</sup>。但目前尚未见中药化学成分群与中药浸膏物理性质(如溶解度)关联研究的报道。本实验在课题组前期颗粒剂溶化性量化评价方法的基础上,完善并验证中药浸膏粉溶化性评价方法。随后以代表性中药(广藿香、当归、薄荷、瞿麦、银柴胡、乌梅、葛根、细辛、益母草、苦参、熟地黄、荆芥和川芎)浸膏粉为研究对象,研究基于中药化学成分群加和性分子描述符预测中药浸膏粉溶化性的模型,并基于该模型解释不同浸膏粉产生溶化性差异的物质和信息基础。

## 1 仪器与材料

### 1.1 仪器

M800 Process 型工业多参数变送器、Inpro8200/S/Epoxy/120 型浊度传感器,梅特勒-托利多仪器有限公司; RCT B S025 型控温磁力搅拌器,德国 IKA 公司。

## 1.2 材料

广藿香浸膏粉(批号 180209-524200-01)、当归浸膏粉(批号 YP170122-748502-03)、薄荷浸膏粉(批号 YP180209-277700-13)、瞿麦浸膏粉(批号 QM201709-3)、银柴胡浸膏粉(批号 YCH201709-2)、乌梅浸膏粉(批号 J180627-611300-13)、葛根浸膏粉(批号 GG201709-3)、细辛浸膏粉(批号 180209-113200-13)、益母草浸膏粉(批号 YP180209-441400-05)、苦参浸膏粉(批号 180209-046300-07)、熟地黄浸膏粉(批号 ShuD201709-2)、荆芥浸膏粉(批号 YP180209-071200-18)、川芎浸膏粉(批号 YP170122-748411-01)均由北京康仁堂药业有限公司提供,是在中试规模条件下,由中药饮片经水提、固液分离、浓缩和喷雾干燥制备而成。0、500、1000、2000、4000 FTU 福尔马胂标准浊度液,杭州齐威仪器有限公司。

## 2 方法与结果

本课题组在《中国药典》2020 年版方法的基础上,使用浊度传感器在线监测颗粒溶化过程中溶液浊度值随时间的变化<sup>[12-13]</sup>,采用 5 min 时的溶液浊度值反映颗粒溶化状态,并根据多批次市售颗粒的浊度测定结果制定了完全溶化、轻微浑浊、浑浊和重度浑浊的浊度区间,为中药颗粒剂溶化性的量化评价提供了参考。

### 2.1 中药浸膏粉溶化性评价方法的建立

**2.1.1 溶液浊度测试方法** 量取 200 mL 水置遮光布包裹的 250 mL 高型烧杯中,将烧杯置于磁力搅拌器上,将浊度传感器垂直于液面放置于烧杯中,传感器探头顶距液面约 0.5 cm,并保证传感器探头与烧杯底部的距离大于 7 cm,调整传感器距烧杯边缘的距离。传感器安装完成后,在烧杯内放入 B 型转子,设置磁力搅拌器温度和搅拌转速,水温稳定后打开 M800TCT 软件,设置开始加样后,取浸膏粉样品 10 g,精密称定,缓缓加入至烧杯中,加样后烧杯顶部用锡纸包裹遮光,记录测试开始至 300 s 时的浊度值。测试前,使用 0、500、1000、2000、4000 FTU 的福尔马胂标准浊度液对传感器进行校正。

**2.1.2 溶液浊度测试条件优化** 在溶液浊度测定过程中,采用单因素考察法对样品粉末加入方式、传感器探头位置、磁力搅拌器转速以及溶液温度 4 个因素分别进行优化,顺次考察每个因素不同水平对测试结果精密度的影响。

(1) 粉末加入方式:以广藿香中药浸膏粉为例,对浸膏粉加入方式进行优化。样品精密称定后,置称量纸(100 mm×100 mm)上,控制样品加入速度,使其分别在 0(即一次性加入)、60、120 s 内加入水中,探头位置、搅拌转速和溶液温度参考文献最佳条件进行设置<sup>[12]</sup>,即探头距烧杯边缘 1.5 cm、搅拌转速 400 r/min 和溶液温度 75 °C,并对溶化 300 s 的浊度值进行分析。0、60、120 s 3 种加入方式对应的溶液浊度值测定结果的 RSD 分别为 7.6%、2.8%、0.7%。浸膏粉在 120 s 内加入水中,可保障浸膏粉缓慢且充分地溶化,结果重复性好,若继续延长时间会降低实验效率,因此选择粉体的加入时间为 120 s。

(2) 探头位置:对探头位置进行优化,传感器探头顶距液面约 0.5 cm,调整传感器距烧杯边缘的距离。烧杯内径为 6 cm,因此探头位置选取距烧杯边缘 0、1.5、3.0 cm 分别进行测定,并对溶化 300 s 的浊度值进行分析。在探头距烧杯边缘 3.0 cm 时,传感器位于溶液旋涡中心,测得的浊度均值为 5212 FTU, RSD 为 69%,原因为粉末加入后逐渐集中至中间旋涡中,部分未及时溶解的粉末积累在探头表面或附近,表现为浊度值增大且测定结果的精密度差。当探头距烧杯边缘 0 cm,即传感器紧邻壁时,测得的浊度均值为 1125 FTU, RSD 为 3.5%,测定不稳定。而在浊度传感器的探头距烧杯边缘 1.5 cm 时,测得的浊度均值为 1173 FTU, RSD 为 1.2%,表明在该条件下,测量方法精密度好,因此选择探头距烧杯边缘的距离为 1.5 cm。

(3) 搅拌转速:对磁力搅拌器转速进行优化,选取磁力搅拌器转速分别为 300、400、500 r/min 进行试验,结果溶化 300 s 时的浊度均值分别为 1149、1192、1139 FTU,测定结果的 RSD 分别为 2.4%、0.6%、1.4%。由于在 400 r/min 时, RSD 最小,表明在该条件下,测量方法精密度好,故选取 400 r/min 作为搅拌转速。

(4) 溶液温度:对溶化过程水温进行优化,根据《中国药典》2020 年版四部颗粒剂项下的可溶颗粒溶化性检查法,样品在热水条件即 70~80 °C 下测定,因此分别选择 70、75、80 °C 进行试验,溶化 300 s 时的浊度均值分别为 1187、1120、999 FTU。在 80 °C 时,溶液浊度均值最低,可能由于在高温条件下,浸膏粉中部分成分在水中溶解度增加,从而改善了浸膏粉的溶化性。3 组浊度值经方差分析,

不同温度下测定的浊度值差异有统计学意义 ( $P < 0.05$ ), 表明测试温度对中药浸膏粉的溶化性影响较大。在 70、80 °C 条件下的 RSD 分别为 4.0%、3.4%, 而在 75 °C 下条件下的 RSD 为 0.6%, 表明 75 °C 条件下, 测量结果精密度最高, 因此选取 75 °C 的作为测试温度。

(5) 优化结果: 基于以上测试条件优化实验, 确定中药浸膏粉溶化性评价方法的最佳条件为粉末样品加入时间为 120 s, 探头置距烧杯边缘 1.5 cm 处, 搅拌转速 400 r/min, 温度 75 °C。

## 2.2 中药浸膏粉溶化性评价方法的验证

**2.2.1 重复性** 按照“2.1.1”项下方法, 根据“2.1.2 (5)”项测试条件优化的结果, 设置水温为 75 °C, 搅拌转速 400 r/min, 探头位置位于距杯壁 1.5 cm 处, 粉体加入时间为 120 s, 对广藿香浸膏粉溶化性重复测定 6 次。6 次测定获得的 300 s 浊度均值为 1106 FTU, RSD 为 1.3%, 表明方法的重复性良好。图 1 为 6 次实验中浊度随时间变化的曲线, 可见在 30~200 s 内, 浊度值均随时间变化呈逐渐上升趋势; 在 200~300 s 内, 上升趋势趋于平缓, 最后在约 350 s 达到平衡。根据夹角余弦法计算曲线间相似度, 结果 6 条曲线两两之间的夹角余弦值均在 0.997 以上, 表明溶化过程浊度变化曲线的重复性良好。

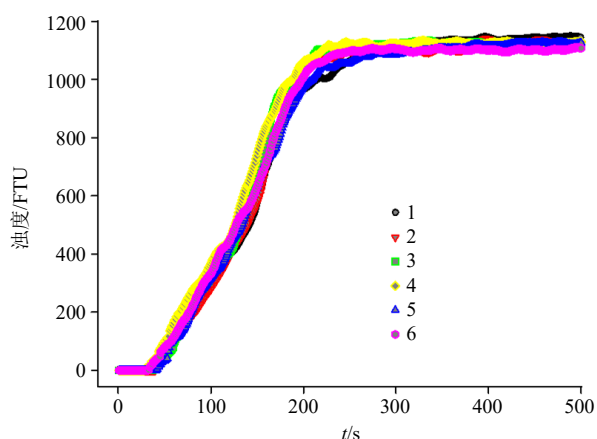


图 1 6 次实验浊度随时间变化的曲线

Fig. 1 Turbidity curves in six experiments

**2.2.2 耐用性** 为考察在溶液浊度测定过程中, 测试条件的微小波动对测定值是否产生影响, 使用 Plackett-Burman 试验设计对方法的耐用性进行研究。根据“2.1.2 (5)”项测试条件优化的结果, 实验安排如表 1 所示, 其中溶液温度 ( $x_1$ ) (75 ± 1) °C, 搅拌转速 ( $x_2$ ) (400 ± 10) r/min, 加入时间 ( $x_3$ ) (120 ± 5) s, 探头位置 ( $x_4$ ): 距杯壁距离 (1.5 ± 0.1)

cm。采用 Design Expert 软件 (V8.0.6, Stat-Ease Inc.) 设计 12 次实验, 以 300 s 时的浊度值作为响应变量, 对结果进行方差分析。12 次测定浊度最大值、最小值分别为 1258、1050 FTU, 平均值为 1126 FTU。对 Plackett-Burman 实验结果进行方差分析, 结果见表 2。F 值为 0.73, P 值为 0.601 6 > 0.05, 说明 4 个因素的变化整体上没有对响应值产生显著性影响。

表 1 Plackett-Burman 实验安排和结果

Table 1 Arrangements and results of Plackett-Burman experiment

编号	$x_1/^\circ\text{C}$	$x_4/\text{cm}$	$x_2/(\text{r}\cdot\text{min}^{-1})$	$x_3/\text{s}$	300 s 浊度值/FTU
1	74 (-1)	1.4 (-1)	390 (-1)	115 (-1)	1258
2	74	1.6 (+1)	410 (+1)	115	1080
3	74	1.4	390	125 (+1)	1078
4	74	1.4	410	115	1237
5	74	1.6	390	125	1213
6	76 (+1)	1.6	390	125	1192
7	76	1.6	410	115	1085
8	76	1.4	410	125	1071
9	76	1.4	390	115	1086
10	74	1.6	410	125	1076
11	76	1.6	390	115	1050
12	76	1.4	410	125	1084

表 2 方法耐用性考察实验结果的方差分析

Table 2 Analysis of variance results for method robustness investigation

	来源	平方和	自由度	均方根	F 值	P 值
模型		18 293.32	4	4 573.33	0.730	0.601 6
$x_1$		11 628.24	1	11 628.24	1.850	0.216 5
$x_4$		1 150.23	1	1 150.23	0.180	0.682 0
$x_2$		4 946.50	1	4 946.50	0.790	0.405 0
$x_3$		568.36	1	568.36	0.090	0.772 6
残差		44 108.00	7	6 301.14		
失拟项		44 024.96	6	7 337.49	88.360	0.081 3
纯误差		83.04	1	83.04		

## 2.3 中药化学成分数据整理

基于中药系统药理学数据库和分析平台 (TCMSP, <http://ibts.hkbu.edu.hk/LSP/tcmsp.php>) 并结合相关文献, 检索 13 种中药共得到 1840 个化合物。在 PubChem 数据库 (<https://pubchem.ncbi.nlm.nih.gov/>) 下载相应化合物, 并以 SDF 格式保存。

## 2.4 分子描述符的生成和预处理

将 SDF 格式保存的化合物导入 DRAGON 7.0 软件 (Kode Chemoinformatics) 中, 获得每个化合物的分子结构描述符。DRAGON 7.0 提供 5270 个分子描述符, 包括组成描述符、分子性质描述符、拓扑描述符和几何描述符等。对获得的分子描述符矩阵 (1840×5270) 按以下方法进行预处理: ①某些分子的一些描述符无法计算, 因此排除存在缺失值的描述符; ②删除方差为 0 的描述符; ③剔除 0 值占比超过 80% 的描述符。此外, ALOGPs 和 ALOGpS 值由 ALOGPS 2.1 软件 (<http://www.vcclab.org/lab/alogs/>) 获得。ALOGpS 表示水溶性, 其值越大表示对应的分子越易溶于水。ALOGpS 为油水分配系数, 是度量有机化合物在水中亲脂性的重要参数, 值越大越难溶于水。经筛选和整合, 最后保留了 1639 个分子描述符。

## 2.5 数据分析

采用 SAS V8 (SAS Institute Inc.) 统计软件进行实验数据分析, 组间比较采用单因素方差分析,

以  $P < 0.05$  表示差异有统计学意义。主成分分析 (principal component analysis, PCA) 和偏最小二乘 (partial least squares, PLS) 法建模和相关数据预处理采用 SIMCA 13.0 软件 (Umetrics)。PLS 模型分别采用留一交叉法验证法和外部验证集法进行性能评估。

## 2.6 不同中药浸膏粉浊度测试结果

按照“2.1.1”项下方法, 根据“2.1.2 (5)”项优化的测试条件优化的结果, 对 13 种中药浸膏粉进行溶化性评价, 测试结果如表 3 所示。13 种中药浸膏粉的浊度值均在 100~1638 FTU。乌梅、当归、银柴胡、熟地黄和川芎浸膏粉溶液属于轻微浑浊 (70~350 FTU); 薄荷、苦参、荆芥、葛根、瞿麦、细辛、广藿香和益母草属于浑浊 (350~2000 FTU)。熟地黄浊度值最低, 其主要成分包括梓醇、地黄苷 D 等环烯醚萜苷类, 环烯醚萜属于单萜类化合物, 通常易溶于水, 与葡萄糖等形成环烯醚萜苷, 相对分子质量一般较小, 大多具有极性官能团, 偏亲水性, 易溶于水。苦参浊度最高, 其主要成分为黄酮

表 3 不同中药浸膏粉浊度测试结果 ( $\bar{x} \pm s, n = 3$ )

Table 3 Turbidity measurement results of different traditional Chinese medicine extract powders ( $\bar{x} \pm s, n = 3$ )

序号	中药	浊度值/FTU	化学成分数目	主要化学成分类型
1	川芎	293.7±2.4	236	多糖类、挥发油类、有机酸类、黄酮及其苷类、生物碱类、甾体类等
2	银柴胡	207.4±4.5	58	甾醇及其苷类、环肽类、挥发油类、黄酮类、生物碱类等
3	熟地黄	100.3±1.4	76	环烯醚萜苷类、糖类、萜类、有机酸类等
4	乌梅	298.6±1.3	83	有机酸类、氨基酸类、甾醇类、黄酮类、挥发油类等
5	当归	225.9±7.1	179	多糖类、氨基酸类、挥发油类、有机酸类、黄酮类、生物碱类等
6	薄荷	779.7±11.9	217	挥发油类、黄酮及其苷类、萜类、氨基酸类等
7	苦参	1 638.0±3.4	138	黄酮类、生物碱类、苯丙素类、萜类、甾体类、有机酸类等
8	广藿香	991.7±9.3	110	黄酮及其苷类、挥发油类、木质素类、三萜类、甾体类等
9	荆芥	1 072.0±2.3	193	黄酮类、挥发油类、萜类、甾体类、脂肪酸类等
10	瞿麦	884.1±31.6	104	三萜皂苷类、黄酮类、生物碱类、有机酸类、蒽醌类等
11	细辛	1 226.0±4.6	192	挥发油类、黄酮类、苯丙素类、脂肪酸类、多糖类等
12	葛根	512.6±7.8	75	有机酸类、三萜皂苷类、黄酮及其苷类、生物碱类、香豆素类等
13	益母草	1 224.0±7.3	179	黄酮类、生物碱类、三萜类、香豆素类、挥发油类、有机酸类等

类和生物碱类等一些难溶于水的成分, 相对分子质量较大且亲水性基团占比较少, 在水中不易溶解。广藿香所含成分主要为广藿香醇、广藿香酮等<sup>[14-15]</sup>挥发性成分, 以及黄酮 (苷)、木脂素、三萜等<sup>[16-18]</sup>非挥发性成分, 这些成分相对分子质量较大且含亲水性基团较少, 大多属于难溶于水的化合物, 是导致广藿香浸膏粉溶液浊度值偏高的主要原因。

## 2.7 化学成分群的 PCA 结果

采用 PCA 对 13 种中药的 1840 个化合物的 1639 个分子描述符进行多变量分析。对原始数据矩阵 (1840×1639) 按变量方向进行中心化与标准化后, 建立 PCA 模型。前 2 个主成分的  $R^2_{cum}$  为 64.2%,  $Q^2_{cum}$  为 63.9%, 表明 2 个主成分可以解释 64.2% 的信息, 该模型的预测能力为 63.9%。绘制前 2 个主

成分对应的得分图和载荷图, 分别如图 2、3 所示。选择水溶解度差别较大的氨基酸类化合物(41 个)、有机酸类化合物(93 个)和黄酮类化合物(176 个), 绘制每类成分的置信椭圆(图 2), 氨基酸类化合物易溶于水, 不溶或微溶于醋酸乙酯、乙醚等有机溶剂; 黄酮类化合物一般难溶或不溶于水, 易溶于甲醇、乙醇等有机溶剂; 有机酸类化合物按照结构一般分为脂肪酸、芳香酸和萜类有机酸 3 大类, 其中脂肪酸大多溶于水, 而芳香酸和萜类有机酸在水中的溶解度较小。图 2 中, 右下角蓝色方形点所示氨基酸为磷酸丝氨酸, 1 L 水只能溶解 0.07 g。有机酸类化合物置信椭圆覆盖范围较大, 与黄酮类化合物接近的部分大多为萜类有机酸, 水溶性差。图 3 载荷图中标出了与溶解度相关的描述符, MLOGP、ALOGPs 和 ALOGP 是根据不同计算方法得到的  $\lg P$  值, 即油水分配系数。MLOGP 使用了基于分子结构的拓扑描述符算法; ALOGPs 使用了基于电子状态拓扑指数作为描述符的神经网络算法; ALOGP 的计算算法是 Ghose-Crippen 法的原子加和法; MLOGP2 和 ALOGP2 分别是 MLOGP 和 ALOGP 值的平方。

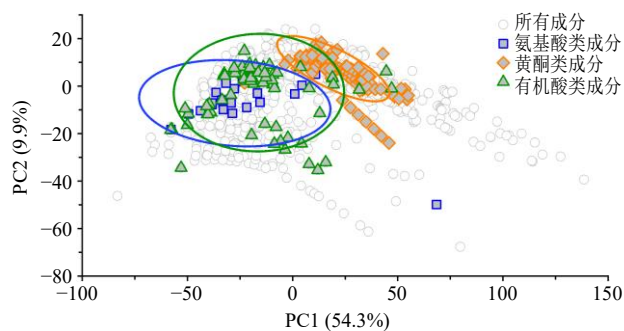


图 2 PCA 得分图

Fig. 2 Score plot of PCA model

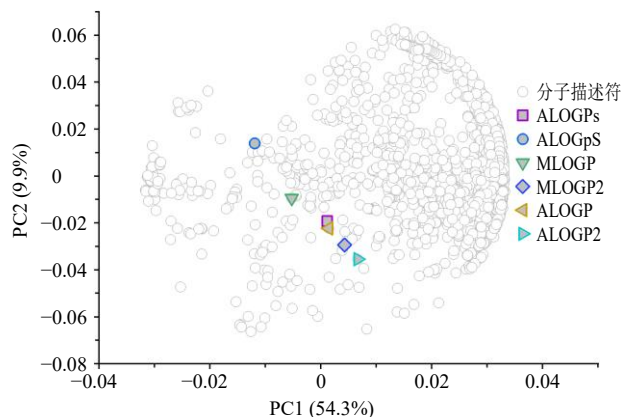


图 3 PCA 载荷图

Fig. 3 Loading plot of PCA model

## 2.8 溶化性预测模型的建立与验证

**2.8.1 加和性分子描述符的计算** 化学成分群加和性分子描述符是由单一化合物分子描述符按照其组分摩尔分数进行加权平均计算得到。对于检索到的中药化学成分群, 由于组成成分的含量信息未知, 因此本实验化学成分群加和性分子描述符由纯物质分子描述符按照等权重进行加和后, 再根据成分数目求平均计算而得。采用随机抽样法将 13 种中药划分为训练集(10 个)和验证集(3 个), 验证集的浊度范围在训练集内。将每种中药所检索到的成分分子描述符的均值作为自变量, 组成  $10 \times 1639$  矩阵, 以浊度值为因变量, 建立浊度预测模型。

**2.8.2 预测模型变量筛选与优化** 对矩阵  $10 \times 1639$  首先按列进行均值标准化, 消除变量量纲差异, 然后进行 PLS 法回归建模。当分子描述符个数为 1639 时, 在 2 个潜变量下, 模型对自变量和因变量的解释性能分别为 82.3%、53.5%, 但交叉验证性能  $Q^2_{cum}$  为 -3.3%, 表明该模型预测性能较差。

进一步根据变量重要性投影 VIP 值进行变量筛选(图 4), 首先删除  $VIP < 1$  的变量, 剩余变量个数为 497 个; 建模后, 继续删除  $VIP < 1$  的变量, 剩余 164 个变量, 此时模型  $R^2_{X cum}$  和  $Q^2_{cum}$  分别为 39.4%、41.2%。再次继续删除  $VIP < 1$  的变量, 剩余 76 个变量。

为了防止重要的变量被删去, 当变量个数到 76 时, 依次删除 5 个 VIP 值小的变量后重新建立模型。当变量剩余 56 个时, 在 5 个潜变量下, PLS 模型性能累积  $R^2_X$  为 87.3%,  $R^2_Y$  为 99.8%, 累积交叉验证性能  $Q^2_{cum}$  为 86.9%。再继续删除 5 个变量, 3 个参数均下降, 模型预测性能降低。结果表明模型最佳分子描述符个数为 56 个, 此模型作为最优模型进行下一步分析。

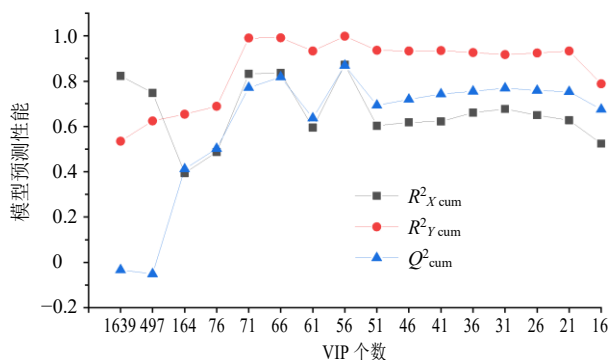


图 4 变量筛选结果

Fig. 4 Results of variable selection

预测模型得分图和载荷图见图 5、6。由得分图可看出，主成分 1 差异贡献率为 49.9%，主成分 2 差异贡献率为 11.0%。得分图可直观的反应出样品间的差异，样品间差异越大，其在得分图上距离越远。本研究中，三角形代表轻微浑浊状态的中药浸膏粉，圆形代表浑浊状态的中药浸膏粉，不同溶化状态的中药浸膏粉存在明显差异。结合载荷图可看出，VE1sign\_B(s)、VE2sign\_B(s)、VE1sign\_X、VE2sign\_Dz(v) 等 2D 矩阵描述符以及 GATS4m、GATS4e、MATS4m 等 2D 自相关描述符距离原点较远，是引起不同中药浸膏粉浊度差异的主要变量。

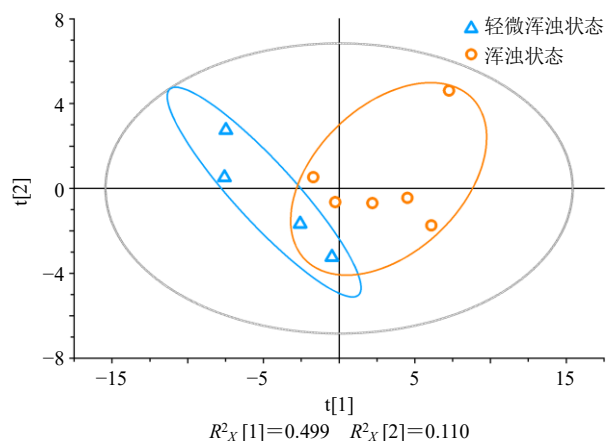


图 5 PLS 得分图

Fig. 5 Score plot of PLS model

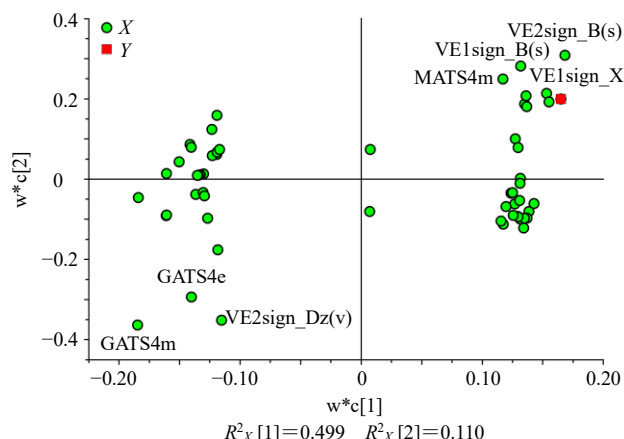


图 6 PLS 载荷图

Fig. 6 Loading plot of PLS model

通过 VIP 值可分析自变量对因变量的贡献大小。VIP 值越大，变量贡献越大。VIP>1 的分子描述符信息见表 4。其中，自变量 GATS4m、GATS4e、GATS6s、MATS4e 和 MATS4m 为基于 Geary 自相关指数和 Moran 自相关指数的描述符。自相关指数反应原子性质沿拓扑结构的分布，其中 GATS4m、GATS6s 和 GATS4e 描述符反映了化合物的静电性质对于结合的影响。MATS4e 和 MATS4m 描述符代表了化合物第 1 电离能的影响。VE1sign\_B(s)、VE2sign\_B(s)、VE2sign\_Dz(v)、VE1sign\_X、SpPosA\_X、SpMAD\_X、ChiA\_B(s)、SpMAD\_X

表 4 VIP>1 的分子描述符信息

Table 4 Molecular descriptors with VIP values larger than 1

序号	变量	VIP 值	系数影响	类别	序号	变量	VIP 值	系数影响	类别
1	GATS4m	1.604	-	2D autocorrelations	11	SpPosA_X	1.061	-	2D matrix-based descriptors
2	VE2sign_B(s)	1.426	+	2D matrix-based descriptors	12	SpMAD_X	1.059	-	2D matrix-based descriptors
3	VE2sign_Dz(v)	1.329	-	2D matrix-based descriptors	13	RBF	1.047	-	Constitutional indices
4	GATS4e	1.320	+	2D autocorrelations	14	nR10	1.035	+	Ring descriptors
5	VE1sign_B(s)	1.241	+	2D matrix-based descriptors	15	ChiA_B(s)	1.027	+	2D matrix-based descriptors
6	VE1sign_X	1.180	+	2D matrix-based descriptors	16	SpMAD_D/Dt	1.022	+	2D matrix-based descriptors
7	B02[O-O]	1.175	-	2D Atom Pairs	17	SpPosA_D/Dt	1.022	+	2D matrix-based descriptors
8	MATS4e	1.147	+	2D autocorrelations	18	S3K	1.005	+	Topological indices
9	MATS4m	1.113	+	2D autocorrelations	19	ALOGpS	1.003	+	Molecular properties
10	GATS6s	1.073	+	2D autocorrelations	20	SpPosA_B(v)	1.001	-	2D matrix-based descriptors

“-” “+” 分别表示变量对因变量的影响为负相关和正相关

“-” “+” represents the effect of the variable on the dependent variable is negatively correlated and positively correlated, respectively

D/Dt、SpPosA\_D/Dt 和 SpPosA\_B(v) 是基于 2D 矩阵描述符。B02 [O-O] 是 2D 原子对描述符，代表 O-O 在拓扑距离 2 处的存在/不存在。RBF 属于结构

描述符，表示可旋转键的数目。nR10 是环指数，代表分子中含有 10 环的个数。S3K 表示分子中原子的空间密度信息，属于拓 PathAva 化指数描述符。

ALOG<sub>PS</sub> 为分子特性描述符, 代表水溶性。

**2.8.3 预测模型验证** 模型的实际值与预测值结果如图 7 所示。除验证集的益母草外, 模型的训练集与验证集其他样本均分布在  $y=x$  线附近, 表明模型的实际值与预测值的吻合度较高, 反映了模型具有一定的预测能力。模型预测结果如表 5 所示,  $R^2=0.96$ ,  $RMSEP=177.58$ 。川芎、荆芥和益母草的预测值分别为 303.8、972.9、933.4 FTU, 相对预测误差分别为 3.4%、9.3%、23.8%, 虽然益母草的相对预测误差较大, 但验证集 3 个样本的相对预测误差值在训练集所有样本的相对预测误差范围 (即 0.1%~38.5%) 之内。从 PLS 模型训练的角度, 模型的交叉验证  $Q^2$  值并不高, 表明预测模型尚无法满足高精度预测的要求, 这体现了由化学成分群预测浊度这一物性的难度; 从模型预测应用的角度, 益母草浊度预测值未改变其所在的溶化性类别, 表明该模型的预测结果可以区分出中药浸膏粉溶液的浑浊程度。此外, 模型预测误差大的原因也可能与益母草中所含化学成分检索不充分, 以及没有对各分子描述符或化学成分进行权重设置等因素有关。

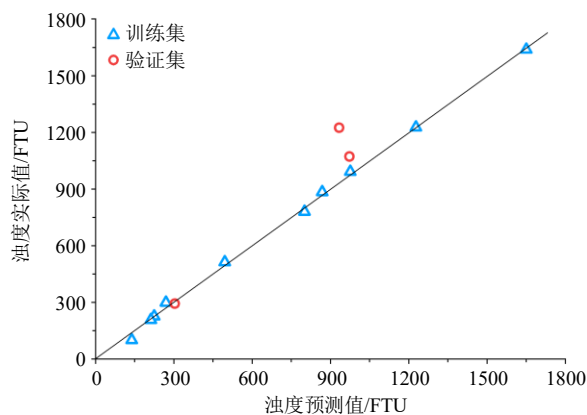


图 7 模型预测值和实际值对比

Fig. 7 Comparison of actual and predicted values

表 5 模型预测结果

Table 5 Model prediction results

中药浸膏粉	溶化性分类	预测值/FTU	实际值/FTU	相对误差/%	$R^2$	RMSEP
川芎	轻微浑浊	303.8	293.7	3.4	0.96	177.58
荆芥	浑浊	972.9	1 072.4	9.3		
益母草	浑浊	933.4	1 224.3	23.8		

### 3 讨论

#### 3.1 溶化性测试的温度

本实验采用的中药浸膏粉是中药配方颗粒的原

料。根据国家药监局在 2021 年 1 月发布的《中药配方颗粒质量控制与标准制定技术要求》<sup>[19]</sup>, 配方颗粒应具备汤剂的基本属性, 其制备工艺应与传统汤剂基本一致, 即以水为溶媒加热提取, 采用物理方法进行固液分离、浓缩和干燥。在提取过程中, 水通常加热至沸腾, 温度约 100 °C。在此温度下, 不但可以增大中药成分的溶解度外, 还能与其他成分产生“助溶”现象, 增加一些水中溶解度小的、亲脂性强的成分的溶解度<sup>[20]</sup>。在颗粒剂溶化性检查中, 采用 70~80 °C 的热水。在高温下提取出的成分, 在低温条件下能否复溶, 是影响颗粒剂溶化程度或浊度的关键。本实验研究发现, 即便在 70~80 °C, 广藿香浸膏粉的浊度也会发生显著变化, 即溶液温度的升高有助于降低浊度。因此, 对含有难溶性成分的中药浸膏粉或颗粒剂, 建议制定合理的溶化性检查温度。

#### 3.2 分子描述符的权重

目前, QSPR 主要研究单一化合物结构与性质的关系, 然而实际应用中往往会遇到处理多组分混合物的问题, 因此建立混合物的 QSPR 模型很有必要。现已存在多种计算混合物分子描述符的方法, 包括加和性描述符、非加和性整体描述符和非加和性分子片段描述符<sup>[21]</sup>。加和性描述符是各组分描述符的加权平均值; 非加和性整体描述符和非加和性分子片段描述符是将混合物作为整体, 参考组分间的相互作用, 将各组分的描述符组合起来, 建立代表混合物整体化学信息的描述符<sup>[22]</sup>。中药中所含化合物成分众多, 其物性由成分的组成、含量和结构共同决定, 在之后的研究中可运用上述方法对分子描述符进行加权, 以提高中药物性预测模型的准确度。

#### 3.3 影响中药浸膏粉溶解性的成分

植物来源的天然产物根据其生源途径, 分为初生代谢产物和次生代谢产物<sup>[23-24]</sup>。初生代谢产物是植物成长和繁殖所必需的物质, 如糖类、脂类、氨基酸、核苷酸和维生素等; 而次生代谢产物是指在特定时期和环境条件下, 由初生代谢产物派生而来的一些中、小分子化合物, 包括苯丙素类、生物碱类、萜类、有机酸类、酚类皂苷类和黄酮类等化合物, 其骨架往往比初生代谢产物更复杂。

在中药中, 一般次生代谢产物含量较低, 如益母草中黄酮类和生物碱类的质量分数分别为 0.25%~1.41%、0.11%~2.09%<sup>[25-26]</sup>; 而初生代谢产物含量较高, 如当归中多糖含量约为 15%<sup>[27]</sup>, 川芎



中多糖类化合物约为 5.71%<sup>[28]</sup>, 挥发油约 1%。次生代谢产物多为药用植物的活性成分<sup>[29]</sup>。本实验检索获得的化学成分多为次生代谢产物和相对分子质量较小的初生代谢产物, 通过这些成分可以预测中药水提物整体的溶化性, 表明次级代谢产物与初级代谢产物关系密切。此外, 植物中存在一些大分子物质(如多糖和皂苷等)可以改善难溶性小分子的溶出, 例如黄芪多糖可以增加槲皮素、芦丁、黄芩素、黄芩苷等 15 种黄酮类化合物的溶解度和稳定性<sup>[30]</sup>, 这为一些难溶性次生代谢产物出现在中药水提物中提供了可能性。

综上所述, 本研究优化了基于浊度传感器的中药浸膏粉的溶化性评价方法, 对建立的溶化性检查方法进行了重复性和耐用性的考察, 结果表明所建方法稳定可靠。在此基础上, 为了研究中药成分群与中药浊度之间的关联性, 选择了轻微浑浊和浑浊状态的 13 种代表性中药浸膏粉, 检索每种中药浸膏粉中可能的化学成分。通过 PCA 发现分子描述符可以区分出溶解度有较大差异的化学成分类别。采用分子描述符平均权重法计算中药浸膏粉的化学成分群平均物性, 并作为自变量, 采用 PLS 法建立中药浸膏粉溶化性的预测模型, 所建模型预测值与实际值符合良好, 预测结果可以指示中药浸膏粉溶化性分类。研究结果为从中药整体成分层面预测宏观物性提供了参考。

**利益冲突** 所有作者均声明不存在利益冲突

#### 参考文献

- [1] 赵晓霞, 赵巍, 张永文. 中药制剂关键质量属性确认的思考 [J]. 中草药, 2019, 50(17): 4008-4012.
- [2] 王海波, 高会芹, 李哲, 等. 玄麦甘桔含片浸膏粉的粉体学性质与制粒工艺研究 [J]. 中医研究, 2020, 33(4): 55-58.
- [3] 江雯怡, 吴文瀚, 高喜梅, 等. 中药制剂吸湿性分析与防潮技术 [J]. 中草药, 2018, 49(12): 2749-2753.
- [4] 李伟霞, 唐于平, 刘立, 等. 药对研究 (III): 药对的功效物质基础 [J]. 中国中药杂志, 2013, 38(24): 4196-4202.
- [5] 王洋, 覃仕娜, 李玉凤, 等. 荔枝核皂苷颗粒剂成型工艺研究 [J]. 中药材, 2020, 43(7): 1692-1696.
- [6] Himanen L, Jäger M O J, Morooka E V, et al. DScribe: Library of descriptors for machine learning in materials science [J]. *Comput Phys Commun*, 2020, 247: 106949.
- [7] Mettou A, Papanophytou C, Melagraki G, et al. Aqueous solubility enhancement for bioassays of insoluble inhibitors and QSPR analysis: A TNF- $\alpha$  study [J]. *SLAS Discov*, 2018, 23(1): 84-93.
- [8] 朱腾义, 姜越. 基于理论线性溶解能关系预测有机污染物在 PDMS 与水中的分配系数 [J]. 东南大学学报: 自然科学版, 2020, 50(1): 200-206.
- [9] 王小雪, 卢杉, 郑思悦, 等. 归经中药化学成分、药理作用及临床应用的实证分析 [J]. 中华中医药杂志, 2018, 33(11): 5193-5197.
- [10] 杜亚朋, 王美, 李璐遥, 等. 基于化合物稳定性探讨炮制对含环烯醚萜类成分中药药性及功效影响的研究进展 [J]. 中草药, 2021, 52(16): 5039-5051.
- [11] 李海英, 贺鹏, 贺玉婷, 等. 中药复方配伍研究的关键问题及超分子化学解决对策 [J]. 中草药, 2019, 50(12): 2757-2762.
- [12] 于佳琦, 徐冰, 李婉婷, 等. 两种在线传感器评价中药配方颗粒溶化性研究 [J]. 中华中医药杂志, 2020, 35(5): 2395-2400.
- [13] 于佳琦, 徐冰, 黄雨妍, 等. 基于在线浊度传感器的中药颗粒剂溶化性评价和分类研究 [J]. 中国中药杂志, 2020, 45(2): 259-266.
- [14] Verma R S, Padalia R C, Chauhan A, et al. Chemical composition of leaves, inflorescence, whole aerial-parts and root essential oils of patchouli {*Pogostemon cablin* (Blanco) Benth.} [J]. *J Essent Oil Res*, 2019, 31(4): 319-325.
- [15] Zhang G X, Wu Y G, Haq Muhammad Z U, et al. cDNA cloning, prokaryotic expression and functional analysis of 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCR) in *Pogostemon cablin* [J]. *Protein Expr Purif*, 2019, 163: 105454.
- [16] 李菲. 矩叶卫矛和广藿香化学成分及生物活性研究 [D]. 北京: 北京协和医学院, 2013.
- [17] Dai O, Li X H, Zhou Q M, et al. Sesquiterpenoids from the aerial parts of *Pogostemon cablin* [J]. *Phytochem Lett*, 2018, 24: 56-59.
- [18] 黄烈军. 中药广藿香化学及生物活性成分研究 [D]. 贵阳: 贵州大学, 2008.
- [19] 国家药品监督管理局. 关于发布《中药配方颗粒质量控制与标准制定技术要求》的通告 [EB/OL]. [2021-01-26]. [http://www.gov.cn/zhengce/zhengceku/2021-02/11/content\\_5586817.htm](http://www.gov.cn/zhengce/zhengceku/2021-02/11/content_5586817.htm).
- [20] 邹佳丽, 谢智勇, 姜晓飞, 等. 大黄黄连泻心汤不同配伍浸渍剂中蒽醌类成分变化研究 [J]. 世界科学技术—中医药现代化, 2008, 10(4): 61-65.
- [21] 高月, 张向倩, 张子炎, 等. QSPR 在混合物典型燃爆特性预测中的研究进展 [J]. 中国安全生产科学技术, 2020, 16(10): 20-26.
- [22] 秦玉翠. 基于人工神经网络的有机混合物定量结构性关系研究 [D]. 西安: 西安石油大学, 2018.

- [23] 廖李容. 氮添加对白羊草土壤有机碳组分及根系分泌物的影响 [D]. 北京: 中国科学院大学(中国科学院教育部水土保持与生态环境研究中心), 2021.
- [24] 孟庆峰, 张柳, 杨彩玲, 等. 贵州都匀一种聚筛蕊地衣粗多糖的提取及抑菌活性初探 [J]. 遵义医学院学报, 2019, 42(2): 158-161.
- [25] 黄锁义, 黎海妮, 余美料, 等. 益母草总黄酮的提取及鉴别 [J]. 时珍国医国药, 2005, 16(5): 398-399.
- [26] 阮金兰, 杜俊蓉, 曾庆忠, 等. 益母草的化学、药理和临床研究进展 [J]. 中草药, 2003, 34(11): U015-U019.
- [27] 赵静, 夏晓培. 当归的化学成分及药理作用研究现状 [J]. 临床合理用药杂志, 2020, 13(6): 172-174.
- [28] Yuan J F, Zhang Z Q, Fan Z C, *et al.* Antioxidant effects and cytotoxicity of three purified polysaccharides from *Ligusticum chuanxiong* Hort. [J]. *Carbohydr Polym*, 2008, 74(4): 822-827.
- [29] 冯禹壮, 徐伟, 肖春萍, 等. 氮磷钾配施对北苍术产量、品质及其相关酶基因表达的影响 [J]. 植物营养与肥料学报, 2022, 28(2): 302-312.
- [30] Liu F, Sun L L, You G J, *et al.* Effects of *Astragalus* polysaccharide on the solubility and stability of 15 flavonoids [J]. *Int J Bio Macromol*, 2020, 143(C): 873-880.

[责任编辑 郑礼胜]