基于偏最小二乘回归的红外光谱波段筛选用于甘露醇−氯化钙共晶的定量分析

李祖頔[1], 张 珂[2], 张泽飞[2], 钱 帅[1], 魏元锋[1], 张建军[2], 高 缘[1*]

(1. 中国药科大学中药学院, 江苏 南京 211198; 2. 中国药科大学药学院, 江苏 南京 211198)

摘要: 甘露醇−氯化钙金属有机骨架 (MOF) 共晶显著改善了*β*-甘露醇的可压片性, 可开发作为新型片剂填充剂, 但研究者在该辅料放大生产过程中发现产物中含有甘露醇单体, 显著影响辅料的功能特性。本研究拟对甘露醇−氯化钙共晶体系的多组分进行定量研究。实验以MOF共晶辅料甘露醇−氯化钙共晶为模型化合物, 采集红外光谱, 以偏最小二乘回归 (PLSR) 为基础, 采用去除异常波段、归一化进行光谱预处理, 采用遗传算法 (GA) 和竞争性自适应重加权算法 (CARS) 2种不同变量筛选方法筛选关键变量, 建立并比较了共晶辅料制备产物中甘露醇−氯化钙MOF共晶含量的定量校正模型。采用GA法和CARS法2种不同变量筛选方法分别筛选出160、14个变量, CARS-PLSR法建立的甘露醇−氯化钙共晶模型具有最佳性能, 模型的平均相对误差 (MRE) 和校正均方根误差 (RMSEC) 分别为0.008 8和0.892 5, 与全光谱模型相比, 模型的决定系数 (R^2) 由0.978 3提升为0.994 4。本研究建立的共晶体系定量方法预测精度高、检测速度快、稳定性好, 对优化此类共晶辅料制备工艺条件及质量控制方法的研究具有重要意义。

关键词: 傅里叶红外光谱; 偏最小二乘回归; 甘露醇−氯化钙共晶; 波段筛选; 定量分析

中图分类号: R917 **文献标识码**: A **文章编号**: 0513-4870(2023)04-1041-08

Infrared spectral band screening based on partial least squares is used for the quantitative analysis of mannitol-calcium chloride cocrystal

LI Zu-di[1], ZHANG Ke[2], ZHANG Ze-fei[2], QIAN Shuai[1], WEI Yuan-feng[1], ZHANG Jian-jun[2], GAO Yuan[1*]

(*1. School of Traditional Chinese Pharmacy, China Pharmaceutical University, Nanjing 211198, China; 2. School of Pharmacy, China Pharmaceutical University, Nanjing 211198, China*)

Abstract: Mannitol-calcium chloride metal organic framework (MOF) cocrystal significantly improved the tabletability of *β*-mannitol and could be developed as a new tablet filler. However, mannitol monomer was found in the product during the scale-up production of the excipient, which significantly affected the functional properties of the excipient. In this study, we intend to quantify the multi-component eutectic system of mannitol-calcium chloride. In this experiment, the MOF cocrystal excipient mannitol-calcium chloride cocrystal was used as the model compound, and infrared spectrum was collected. Based on partial least squares regression (PLSR) method, the abnormal bands were removed and the spectrum was preprocessed by normalization. The quantitative correction model of mannitol-calcium chloride MOF cocrystal content in cocrystal excipients was established and compared by two different variable screening methods, genetic algorithm (GA) and competitive adaptive

收稿日期: 2022-09-26; 修回日期: 2022-11-04.

基金项目: 国家自然科学基金资助项目 (81873012, 82074029, 82104401); 高校“双一流”项目 (CPU2018GY11, CPU2018GY27); 江苏省自然科学基金资助项目 (SBK2020042291); 中国博士后科学基金资助项目 (2020M671665).

*通讯作者 Tel: 86-25-83379418, E-mail: newgaoyuan@163.com

DOI: 10.16438/j.0513-4870.2022-1079

reweighting algorithm (CARS). Two different variable screening methods, GA method and CARS method, were used to screen out 160 and 14 variables, respectively. The mannitol-calcium chloride cocrystal model established by CARS-PLSR method had the best performance, and the average relative error (MRE) and corrected root mean square error (RMSEC) of the model were 0.008 8 and 0.892 5, respectively, the determination coefficient (R^2) of the model was increased from 0.978 3 to 0.994 4. The quantitative method of eutectic system established in this study has high prediction accuracy, fast detection speed and good stability, which is of great significance for optimizing the preparation process conditions and quality control methods of such eutectic excipients.

Key words: Fourier transform infrared spectroscopy; partial least squares regression; mannitol-calcium chloride cocrystal; variable selection; quantitative analysis

共晶是指两种或两种以上的分子通过氢键、π-π堆积作用、范德华力等非共价键以固定的化学计量比结合形成的晶体[1]。与原晶体相比, 共晶的分子相互作用力、晶体结构、晶格能和表面性质等均可发生较大改变, 故其理化性质也会发生不同程度的改变[2]。甘露醇−氯化钙金属有机骨架 (metal-organic framework, MOF) 共晶是β-甘露醇与氯化钙以配位键形成的MOF共晶, 与市售片剂填充剂β-甘露醇相比, 这种具有三维无限配位骨架的共晶有更高的结合强度和更低的弹性回复率, 显著改善了β-甘露醇的成片性能, 且甘露醇−氯化钙MOF共晶能参与湿法造粒过程, 拓宽了其在片剂开发中的应用, 具有广阔的应用前景[3]。在探索适宜工艺放大的甘露醇−氯化钙MOF共晶晶体工程学制备方法过程中, 研究者发现产物中含有少量甘露醇单体, 且甘露醇单体可压性较差, 混入会影响共晶辅料的可压性。为了对甘露醇−氯化钙MOF共晶辅料的制备方法进一步优化, 保证产品的功能特性, 亟需对该共晶体系进行定量研究。

共晶体系多组分的定量是晶体学研究的难点。共晶大多是在氢键的作用下形成的, 遇水或其他溶剂时会发生解离, 因此常规化学分析法 (滴定法、HPLC) 只能测得单体的化学计量比, 不能直接检测共晶的含量。而粉末X射线衍射 (powder X-ray diffraction, PXRD)、差示扫描量热 (differential scanning calorimetry, DSC) 和热重分析 (thermogravimetric analysis, TGA) 等技术可能受到样品粒径、择优取向和样品厚度等因素的影响, 影响定量计算的准确性[4-7]。光谱表征技术 (红外、近红外和拉曼光谱等) 可研究样品分子水平上的差异, 并且具有快速、无损、样品准备时间短等优点[8]。然而, 单纯的光谱表征技术由于制样的影响, 检测噪音较高, 灵敏度和准确度较低, 难以实现共晶的精确定量[9]。

偏最小二乘回归等多元回归方法可从光谱中提取定性和定量信息, 克服信号重叠和光谱响应信号与样品浓度之间非线性的问题[10-12], 提高光谱检测技术的灵敏度和准确度[13-17]。由于样品光谱谱带过宽或无信息盲区的混入等, 会影响全光谱模拟模型的准确率。应用变量筛选算法可有效去除光谱中的无关信息, 并筛选出与成分信息相关的关键变量, 与全光谱模型相比可显著降低模型复杂度, 提高拟合效率, 优化模型的精密度和重复性[18-21]。

本研究采用傅里叶红外光谱 (Fourier transform infrared spectroscopy, FTIR) 定量分析甘露醇−氯化钙MOF共晶产物中的共晶, 以偏最小二乘回归 (partial least squares regression, PLSR) 为基础, 结合不同变量筛选算法用于共晶含量的快速测定, 采用遗传算法 (genetic algorithm, GA)、竞争性自适应重加权算法 (competitive adaptive reweighted sampling, CARS) 2种变量筛选方法, 建立并比较了晶体工程学制备共晶辅料产物含量的定量校正模型, 对优化此类共晶辅料制备工艺条件及质量控制方法的研究具有重要的参考价值。

材料与方法

试剂与仪器 甘露醇160C [罗盖特 (中国) 精细化工有限公司]; 无水氯化钙 (西陇化工股份有限公司); 无水乙醇 (国药集团化学试剂有限公司); 溴化钾 (德国CNW公司)。电子天平 (德国Sartorius公司); 水浴锅 (上海恩谊科技有限公司); 悬臂式电动搅拌器 (深圳力振科技有限公司); 真空干燥箱 (上海博迅实业有限公司); Eclipse Ci-POL透反射偏光显微镜 (南京斯普利仪器设备有限公司); 熔点仪 (天津天大天发公司, YRT-3); 粉末X-射线衍射仪 (Bruker公司, D8 Advance衍射仪); IRAffinity-1S傅里叶变换红外光谱仪 (日本Shimadzu公司)。

甘露醇−氯化钙共晶的制备 根据文献[3]方法, 通过将甘露醇 (0.1 mol) 和氯化钙 (0.1 mol) 加入100 mL乙醇/水 (8∶2, v∶v) 中, 在65 ℃下溶解, 得到澄清溶液, 置于60 ℃烘箱中密封静置结晶, 数天后溶液底部出现数颗无色透明的晶粒。采用PXRD和红外光谱对晶粒

进行鉴定。为了最大限度地减少粒径效应, 将样品分别在玛瑙研钵中研磨并通过100目筛。

二元混合物的制备 取甘露醇-氯化钙共晶样品适量, 研钵研磨均匀。将其与甘露醇称量至离心管混合, 配制共晶质量分数分别为100%、99.5%、99%、98%、96%、94%、92%、90%、85%、80%、70%、60%的混合粉末得到红外测定标准集样品, 配制共晶质量分数分别为95%、85%、75%的混合粉末得到红外测定验证集样品。

PXRD 测定时以2°·min^{-1}的速度扫描, 步长为0.02°, 扫描范围为5°～40°, 管压和安培数分别设为40 kV和40 mA。使用Mercury 4.1.0软件利用β-甘露醇和共晶的单晶结构计算各自的PXRD图谱, 随后将实验得到的PXRD图与利用计算得到的PXRD图进行对比。其中β-甘露醇的晶体结构文件从剑桥晶体数据库 (Cambridge Structural Database, CSD) 下载, 其CSD编号为224659; 共晶的晶体结构文件CSD编号为1866028。

FTIR 按质量比1∶100称取二元混合物样品和溴化钾, 于玛瑙研钵中研磨混合, 混合粉末压制成透明薄片。FTIR仪扫描范围4 000～400 cm^{-1}, 扫描次数为32。数据用Nicolet Omnic红外光谱处理软件 (Version 8.0) 处理。每个含量水平的样品平行测量3次, 采集3张光谱, 将平均值用作样品的原始光谱数据。

数据分析 对于定量分析, 将二元混合物分为包括36个光谱的校准集和包括9个光谱的预测集。使用PLSR建立二元混合物的定量模型。在定量分析中, 使用单一预处理方法及其组合对原始光谱进行预处理。所有数据处理均采用Matlab R2018b实现。图形绘制采用Origin 2021软件、Matlab R2018b软件。使用决定系数 (R^2)、校正集均方根误差 (RMSEC) 及验证集均方根误差 (RMSEP) 作为优化建模参数的指标, 并检查模型预测性能[22]。同时, 以检测限 [LOD, 公式 (1)] 和定量限 [LOQ, 公式 (2)] 评价模型的预测误差[23]。

$$\mathrm{LOD} = 3.3 \times \frac{\mathrm{SD}}{m} \tag{1}$$

$$\mathrm{LOQ} = 10 \times \frac{\mathrm{SD}}{m} \tag{2}$$

其中, SD为3次检测的标准差, m为标准集线性的斜率。

结果与讨论

1 甘露醇-氯化钙共晶的表征

采用偏光显微镜、熔点仪、PXRD和红外光谱对所制备的甘露醇-氯化钙共晶及甘露醇进行表征, 共晶和甘露醇的形状、熔点、PXRD图谱和红外光谱显示出明显的差异, 且与文献[24]报道一致。

1.1 偏光显微镜 (PLM)

使用偏光显微镜观察发现甘露醇-氯化钙共晶和β-甘露醇的外观和形状不同 (图1A)。两者均具有明暗相间的双折射现象, 表现出晶体特征。其中共晶样品为较大的双棱锥状, β-甘露醇为细条状晶体。

1.2 熔点仪

分别取β-甘露醇和甘露醇-氯化钙共晶样品粉末适量, 装至末端封口毛细管中, 于熔点仪上测定熔点, 升温速度为3 ℃·min^{-1}。β-甘露醇的熔点为168.1～169.2 ℃, 与文献[24]报道一致。共晶的熔程为149.5～150.3 ℃, 熔程短且单一, 提示共晶具有很好的相纯度。

1.3 PXRD

各样品的PXRD图谱如图1B所示。从图中可看出, 甘露醇-氯化钙共晶和β-甘露醇PXRD图谱中的衍射峰与根据单晶结构计算得到的PXRD图谱一一对应, 但不同衍射峰的相对强度有所差异, 这可能是晶体粉末的择优取向导致的。PXRD结果表明成功制备出了甘露醇-氯化钙共晶。

1.4 FTIR

各样品FTIR图如图1C所示。β-甘露醇在3 100～3 600 cm^{-1}存在3 290和3 400 cm^{-1}两个尖锐的吸收峰, 与文献[24]报道一致, 对应β-甘露醇中羟基 (-OH) 的伸缩振动。共晶在3 000 cm^{-1}以上范围内则存在4个明显的吸收峰, 分别位于3 484、3 379、3 314和3 184 cm^{-1}处, 其中3 484 cm^{-1}处的吸收峰对应结晶水上的羟基 (-OH), 而甘露醇中的羟基 (-OH) 形成共晶后由于配位键的形成, 羟基上电子云密度下降, 故羟基伸缩振动峰发生了明显的红移, 其中3 290 cm^{-1}处的吸收峰红移至3 184 cm^{-1}, 3 400 cm^{-1}处的吸收峰发生红移并裂为3 379和3 314 cm^{-1}处的两个峰。

2 共晶定量模型

2.1 模型的建立

以PLSR法测定标准集的累计贡献率如图2A所示, 第一主成分占据光谱矩阵贡献率为64.65%, 浓度矩阵为74.34%。前10个主成分占据光谱矩阵贡献率为99.99%, 浓度矩阵为99.97%, 基本携带样品的所有信息。PLSR法不同主成分数与模型评价指标的关系如图2B所示, 随着主成分数的增加, R^2不断增加, MRE、RMSEC不断下降, 说明主成分数越多, 共晶标准集的拟合效果越好。主成分数增加至5时, RMSEP显著改善, 此时验证集的预测值和实际值如图2C, 结果较好, 说明前5个主成分在拟合中起关键作用; 主成分数由5增加至10时, RMSEP不降反增, 此时整个模

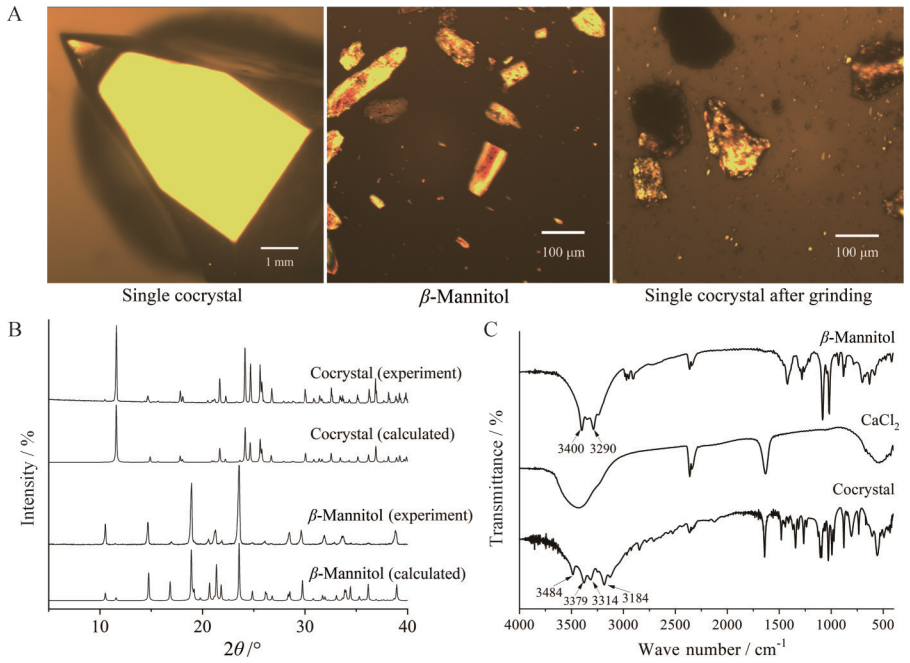

Figure 1 A: Polarized light microscope (PLM) images of mannitol- calcium chloride single cocrystal was obtained by cooling crystallization of mannitol calcium chloride solution; PLM images of β-mannitol; PLM images of mannitol-calcium chloride cocrystal after grinding. B: Powder X-ray diffraction (PXRD) patterns of mannitol- calcium chloride cocrystal and β-mannitol. C: Fourier transform infrared spectrometer (FTIR) spectra of cocrystal and β-mannitol

型存在过拟合情况, 即把标准集中与浓度无关的信息也引入至模型的拟合中。这表明模型不宜添加后续主成分, 确定主成分个数为5之后, 使用5个主成分对模型进行拟合, PLSR对共晶单晶预测的含量为99.67%, 与实际值误差0.33% (图2D)。

2.2 光谱预处理优筛

红外光谱中的异常值、无关变量和噪音过多会降低光谱的灵敏度和准确度, 进而影响整个模型的定性

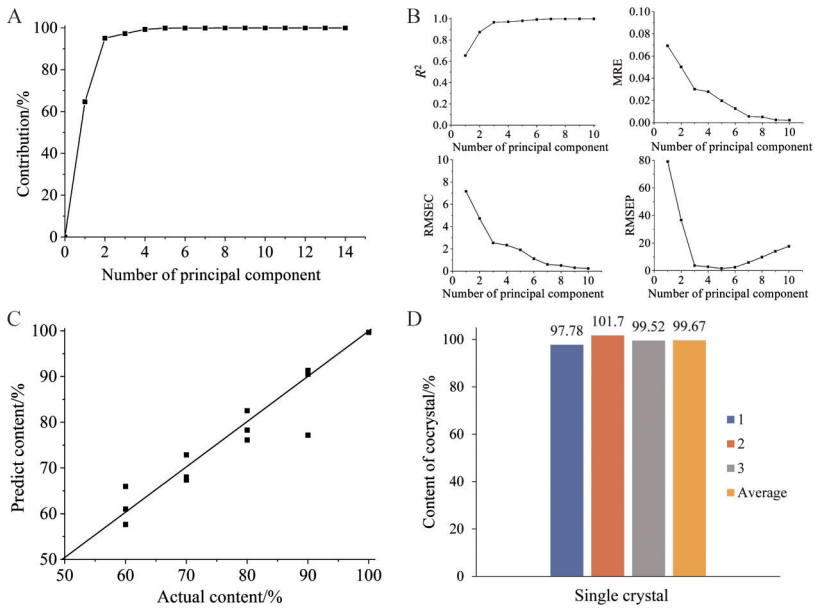

Figure 2 A: Diagram of the relationship between the number of principal components and the cumulative contribution rate. B: Relationship between the number of principal components and model evaluation indexes of partial least squares regression (PLSR) method. C: The predicted and actual values of the validation set when the number of principal components of PLSR method was 5. D: Cocrystal content of detection set when PLSR principal component is 5. R^2: Coefficient of determination; MRE: Magnitude of relative error; RMSEC: Root mean square error of calibration; RMSEP: Root mean square error of prediction

和定量效果[25]。预处理的方式会直接影响模型的仿真效果。

本研究采用直接观察法去除了异常波段或异常样本, 降低环境中的无关变量对光谱的影响, 如图3A、B所示, 在2 000～2 500 cm^{-1}波段内, 存在空气中游离的二氧化碳和水蒸气峰[26]。将红外测定标准集光谱舍去该波段前后的结果如表1所示, RMSECV和RMSEP分别降低了2.02%和8.88%, 表明舍去该异常波段能明显减少模型中的无关信息, 改善模型的预测效果。

此外, 比较了原始光谱 (已去除异常波段) 和经过多种不同预处理方法后的模型效果, 考察了归一化 (normalization)[25,27]、Savitsky-Golay (SG) 滤波平滑[28]、多元散射校正 (multiplicative scatter correction, MSC)[25]、标准正交变换 (standard normal variate correction, SNV)[25]、一阶导数法 (first derivative, 1stD) 和二阶导数法 (second derivative, 2ndD) 及叠加预处理方法对模型预测效果的影响。结果如表1所示, 归一化预处理后, R^2值改善1.03%, RMSEC改善25.37%, RMSECV改善23.83%, 模型精密度得到明显改善; RMSEP降低79.09%, 说明模型在预测未知含量的甘露醇-氯化钙样品时准确度提高。因此, 本研究选择归一化作为最佳的预处理方法。同一共晶标准品平行测定3次收集的样品原始光谱和预处理的光谱显示在图3C、D中。

2.3 波段压缩与优选

2.3.1 遗传算法 (GA) GA法[29]主要思路参考了达尔文的适者生存思想, 运算过程中, 信息交换和变异的基本单元称为染色体, 每个个体携带的染色体总数是固定的。用GA对红外光谱波长进行压缩和优选时, 先随机选择一定数量的测定波长为初始波长集合, 用集合内的波长进行模型拟合, 以RMSECV作为整个模型的评价指标, RMSECV越小, 则该个体的存活概率越高。将红外测定标准集样品光谱数据划分为100个连

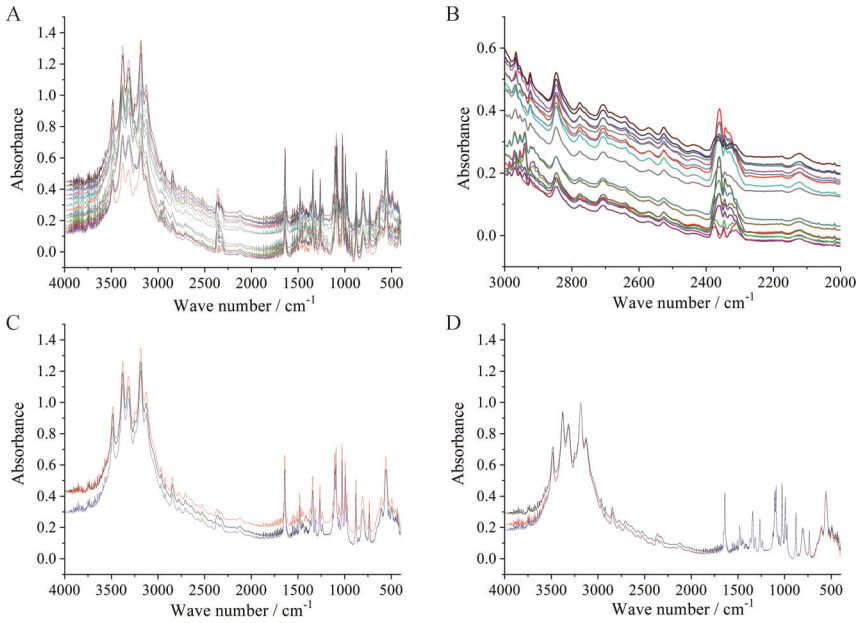

Figure 3 A: Sample absorbance at 4 000 to 400 cm^{-1} wavelength. B: Sample absorbance at 3 000 to 2 000 cm^{-1} wavelength. C, D: Raw FTIR spectra of parallel determination and spectra pretreated with normalization (C) and of cocrystal standard (D)

Table 1 PLSR modeling and evaluation table after eliminating abnormal bands and influence of different pretreatment methods on simulation performance of the model. RMSECV: Root mean square error of cross validation; SG: Savitsky-Golay; MSC: Multiplicative scatter correction; SNV: Standard normal variate correction; 1stD: First derivative; 2ndD: Second derivative

Pretreatment method	Principal component number	R^2	RMSEC	RMSECV	RMSEP
4 000−400 cm^{-1}	5	0.978 3	1.731 7	2.313 7	7.502 0
4 000−2 500 cm^{-1} and 2 000−400 cm^{-1} (raw)	5	0.977 9	1.745 7	2.266 9	6.835 8
Normalization	5	0.988 0	1.302 8	1.726 6	1.429 6
SG	\	0.987 8	1.311 2	1.736 7	1.361 2
MSC	5	0.990 4	1.161 9	1.622 0	2.411 4
Normalization + MSC	5	0.990 5	1.159 0	1.617 9	2.437 2
SNV	5	0.990 0	1.184 9	1.774 0	2.092 8
MSC + SNV	5	0.990 5	1.158 7	1.617 3	2.434 4
SG + 1stD	\	0.986 8	1.366 8	2.013 3	1.573 5
SG + 2ndD	\	0.991 0	1.124 3	1.945 7	3.208 4

续的微型光谱, 作为遗传算法运算的染色体, 以GA筛选后得到的RMSECV值同筛选的染色体间的关系如图4A所示, 其中, 红色标出的染色体数10为最小染色体最优解, 表示选择波长数目最少的情况下, 携带的样品浓度信息达到极大值; 绿色标出的染色体数61表示全局最优解, 为GA剔除掉无效光谱波长后留下的光谱波段; 蓝色标出的染色体数53为局部最优解。这3个信息点都是RMSECV的极小值。

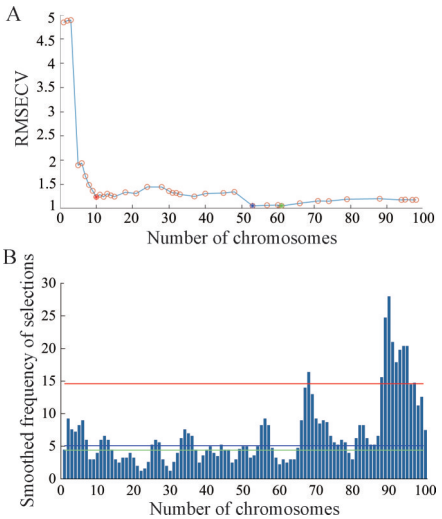

Figure 4 A: The relationship between RMSECV and chromosomes obtained by genetic algorithm (GA). B: The number of times a chromosome has been selected. Red indicator line: The 10 chromosomes marked in red in Figure 4A; the same goes for blue and green lines

每个染色体被选择的次数如图4B所示, 被选择的次数最多, 表明该染色体对应的光谱波段可使RMSECV值更小, 即携带更多的浓度信息的概率更高[29]。与横轴平行的红色指示线的波段, 即为图4A中红色标出的10个染色体。蓝色、绿色同理。

2.3.2 CARS法 竞争自适应重加权算法(CARS)[18]用于消除光谱的无关信息变量, 首先使用蒙特卡洛算法随机采样并使用PLSR进行相关性运算, 以权重大小对所有波长排序, 之后通过指数递降函数方式和自适应重加权采样两阶段对波长进行快速取舍和选择优化。使用CARS对甘露醇−氯化钙共晶标准集进行光谱压缩和筛选, 选定的波长数量、RMSECV值和回归系数的路径如图5所示。图中的趋势表明, 在前36个迭代周期, 光谱波长数的筛选与RMSECV值的同步优化, 回归系数路径分布收缩在一定范围内; 在36个迭代周期之后, 波长数的减低引起RMSECV值的升高(图5A、B), 表明筛选后的所有波段都携带特有的信息, 后续若再对波段进行压缩, 会造成原有信息的大幅丢失, 这与36个周期后RMSECV值迅速增大相一致。因此选择第36个迭代周期后的波长子集为最优, 图5C中显示了变量筛选过程中光谱回归系数的变化。

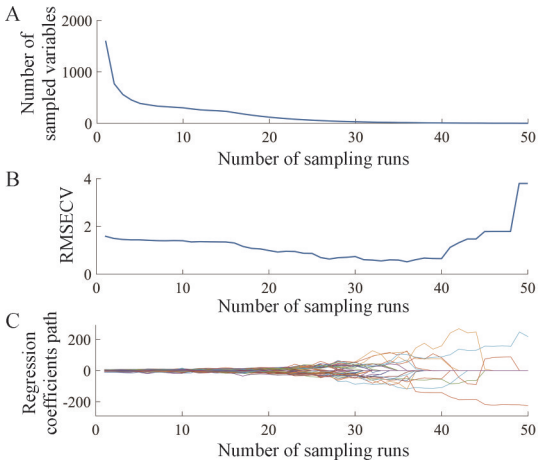

Figure 5 Competitive adaptive weighted resampling method (CARS) variable screening results. Iteration period and number of screening wavelengths (A), RMSECV (B) and regression coefficient path graph (C)

2.4 甘露醇−氯化钙共晶体系定量模型优化结果

基于全光谱和经过预处理、波段压缩与筛选后的结果建立PLSR模型, 结果见表2、3。经过GA法波段筛选后, 筛选得到的波长数从原来的1 607个大幅减少, 可减少无关的变量信息, 与全光谱模型的预测性能相比, 相关的评价指标R^2、MRE、RMSEC、RMSECV结果更优, 证明了GA波段筛选有对甘露醇−氯化钙共晶红外光谱信息有一定的提纯作用, 但GA法的编程实现比较复杂, 其算法搜索速度较慢, 要得到精确解需较多的训练时间。而经CARS波段筛选后, 保留的波长数仅14个, R^2、MRE、RMSEC、RMSECV值分别由0.988 0、0.012 3、1.302 8、1.726 6变为0.994 4、0.008 8、0.892 5、1.242 1, 使用提纯后的光谱波长数进行建模, 模型仍具有良好的准确性和稳定性, 建模结果可说明其压缩变量的可行性。将偏最小二乘回归原始模型和经过去除异常波段、归一化处理和波段筛选优化后的模型进行对比, 75%组别的回收率明显提升, 检测限降低至0.66%、2.01%, 定量限降低至0.65%、1.97%, 模型的准确度和灵敏度大幅改善, 优化后的PLSR模型检测性能更佳。

结论

本研究证明了FTIR结合化学计量学方法快速检测甘露醇−氯化钙MOF共晶与甘露醇二元混合物中共晶的可行性。为了提高检测的精度和效率, 将光谱数

Table 2 Comparison of modeling results of different band screening methods

Method	Number of wavelengths	R^2	MRE	RMSEC	RMSECV	RMSEP
Normalization	1 607	0.988 0	0.012 3	1.302 8	1.726 6	1.429 6
GA	160	0.984 1	0.013 5	1.498 4	1.832 4	0.737 4
	848	0.988 9	0.011 3	1.252 0	1.748 7	1.332 5
	976	0.986 5	0.013 2	1.379 7	1.875 3	1.352 6
CARS	14	0.994 4	0.008 8	0.892 5	1.242 1	1.772 4

Table 3 Evaluation table of PLSR model before and after optimization. LOD: Limit of detection; LOQ: Limit of quantitation

Processing mode	Accuracy (calculate by cocrystal recovery rate)			LOD/%	LOQ/%
	75%	85%	95%		
The original model	66.61 ± 10.29	83.35 ± 0.78	97.55 ± 0.51	2.30	6.96
GA	73.25 ± 1.26	84.75 ± 0.67	95.63 ± 0.07	0.66	2.01
CARS	74.12 ± 1.22	84.67 ± 0.73	95.39 ± 0.09	0.65	1.97

据降维与特征光谱区域选择相结合, 应用特征波长筛选在有效降低检测工作量的同时, 保证了模型的预测效果。其中, CARS法有更佳的变量筛选能力, 其筛选的变量更少且具代表性, 所建立的模型预测效果更佳。基于PLSR结合CARS法筛选关键变量, 采用压缩率99%的波段建立的含量定量校正模型优于基于全波长建立的偏最小二乘回归模型, 可快速、灵敏、可靠、直接地测定产物中的共晶, 对优化此类共晶辅料制备工艺条件及质量控制方法的研究具有重要的参考价值。

作者贡献: 李祖峁、张珂、张泽飞负责实验操作、结果分析及起草、撰写文章; 钱帅、魏元锋、张建军负责实验操作、结果分析; 高缘负责修改文章。

利益冲突: 作者排名顺序无争议, 稿件不涉及涉密和知识产权争议。

References

[1] Grothe E, Meekes H, Vlieg E, et al. Solvates, salts, and cocrystals: a proposal for a feasible classification system [J]. Cryst Growth Des, 2016, 16: 3237-3243.

[2] Singaraju AB, Nguyen K, Gawedzki P, et al. Combining crystal structure and interaction topology for interpreting functional molecular solids: a study of theophylline cocrystals [J]. Cryst Growth Des, 2017, 17: 6741-6751.

[3] Cheng H, Wei Y, Wang S, et al. Improving tabletability of excipients by metal-organic framework-based cocrystallization: a study of mannitol and $CaCl_2$ [J]. Pharm Res, 2020, 37: 130.

[4] Sadul SK, Perumal A, Somanath T. Development and validation of method for the estimation of tacrolimus crystalline API in tacrolimus extended release capsules 0.5 mg, 1 mg and 5 mg by power X-ray diffractometer (PXRD) [J]. Res J Pharm Technol, 2020, 13: 5375-5381.

[5] Talik P, Czerniecka E, Hubicka U, et al. Quantification of active pharmaceutical ingredients in commercially available poly pharmaceutical tablets by means of DSC [J]. Acta Pol Pharm, 2017, 74: 1057-1062.

[6] Bruni G, Maggi L, Mustarelli P, et al. Enhancing the pharmaceutical behavior of nateglinide by cocrystallization: physicochemical assessment of cocrystal formation and informed use of differential scanning calorimetry for its quantitative characterization [J]. J Pharm Sci, 2019, 108: 1529-1539.

[7] Yazdani A, Höhne GW, Misture ST, et al. A method to quantify crystallinity in amorphous metal alloys: a differential scanning calorimetry study [J]. PLoS One, 2020, 15: e0234774.

[8] Braun DE, Maas SG, Zencirci N, et al. Simultaneous quantitative analysis of ternary mixtures of *D*-mannitol polymorphs by FT-Raman spectroscopy and multivariate calibration models [J]. Int J Pharm, 2010, 385: 29-36.

[9] Bell SEJ, Charron G, Cortés E, et al. Towards reliable and quantitative surface-enhanced Raman scattering (SERS): from key parameters to good analytical practice [J]. Angew Chem Int Ed, 2020, 59: 5454-5462.

[10] Meacham-Hensold K, Montes CM, Wu J, et al. High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity [J]. Remote Sens Environ, 2019, 231: 111176.

[11] Qiao J, Wang G, Li W, et al. A deep belief network with PLSR for nonlinear system modeling [J]. Neural Netw, 2018, 104: 68-79.

[12] Rajbanshi J, Bhattacharya S. Assessment of soil erosion, sediment yield and basin specific controlling factors using RUSLE-SDR and PLSR approach in Konar river basin, India [J]. J Hydrol (Amst), 2020, 587: 124935.

[13] de Oliveira Penido CAF, Pacheco MTT, Novotny EH, et al. Quantification of cocaine in ternary mixtures using partial least squares regression applied to Raman and Fourier transform infrared spectroscopy [J]. J Raman Spectrosc, 2017, 48: 1732-1743.

[14] Chen H, Tan C, Lin Z, et al. Detection of melamine adulteration in milk by near-infrared spectroscopy and one-class partial least squares [J]. Spectrochim Acta A Mol Biomol Spectrosc, 2017,

173: 832-836.

[15] Hu Y, Erxleben A, Ryder AG, et al. Quantitative analysis of sulfathiazole polymorphs in ternary mixtures by attenuated total reflectance infrared, near-infrared and Raman spectroscopy [J]. J Pharm Biomed Anal, 2010, 53: 412-420.

[16] Ziemons E, Bourichi H, Mantanus J, et al. Determination of binary polymorphic mixtures of fluconazole using near infrared spectroscopy and X-ray powder diffraction: a comparative study based on the pre-validation stage results [J]. J Pharm Biomed Anal, 2011, 55: 1208-1212.

[17] Feng Y, Li X, Xu K, et al. Qualitative and simultaneous quantitative analysis of cimetidine polymorphs by ultraviolet-visible and shortwave near-infrared diffuse reflectance spectroscopy and multivariate calibration models [J]. J Pharm Biomed Anal, 2015, 104: 112-121.

[18] Li H, Liang Y, Xu Q, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration [J]. Anal Chim Acta, 2009, 648: 77-84.

[19] Salari M, Dehghani MH, Azari A, et al. High performance removal of phenol from aqueous solution by magnetic chitosan based on response surface methodology and genetic algorithm [J]. J Mol Liq, 2019, 285: 146-157.

[20] Watanabe E, Yamashita Y, Maki A, et al. Cerebral blood flow measurement during epilepsy using multi-channel near infra-red spectroscopic topography [J]. Neurosci Res, 1997, 21: 302.

[21] Zhu Z, Chen C, Chen C, et al. Prediction of tumor size in patients with invasive ductal carcinoma using FT-IR spectroscopy combined with chemometrics: a preliminary study [J]. Anal Bioanal Chem, 2021, 413: 3209-3222.

[22] Trampuž M, Teslić D, Likozar B. Process analytical technology-based (PAT) model simulations of a combined cooling, seeded and antisolvent crystallization of an active pharmaceutical ingredient (API) [J]. Powder Technol, 2020, 366: 873-890.

[23] Bhavana V, Chavan RB, Mannava MKC, et al. Quantification of niclosamide polymorphic forms - a comparative study by Raman, NIR and MIR using chemometric techniques [J]. Talanta, 2019, 199: 679-688.

[24] Yoshinari T, Forbes RT, York P, et al. Moisture induced polymorphic transition of mannitol and its morphological transformation [J]. Int J Pharm, 2002, 247: 69-77.

[25] Liu T. A New Wavelengths Selection Method for Molecular Spectra Based on Antcolony Optimization and Fundamental Applications (基于蚁群算法的分子光谱波长选择新方法与应用基础研究) [D]. Hangzhou: Zhejiang University, 2017.

[26] Dousseau F, Therrien M, Pézolet M. On the spectral subtraction of water from the FT-IR spectra of aqueous solutions of proteins [J]. Appl Spectrosc, 1989, 43: 538-542.

[27] Jiang WW. Study on Quantitative Regression Method for Organic Infrared Spectroscopy Based on Integrated Learning (基于集成学习的有机物红外光谱定量回归方法研究) [D]. Hefei: Hefei University of Technology, 2020.

[28] Savitzky A, Golay MJ. Smoothing and differentiation of data by simplified least squares procedures [J]. Anal Chem, 1964, 36: 1627-1639.

[29] Mirjalili S, Jin SD, Sadiq AS, et al. Genetic algorithm: theory, literature review, and application in image reconstruction: methods and applications [J]. Stud Comput Intell, 2020, 811: 69-85.