

核磁共振光谱结合选择性约束的多元曲线分辨解析 绿豆种子萌发中的代谢变化

李响, 李中峰, 高俊, 王霞, 张欣*, 张卓勇*

首都师范大学化学系, 北京 100048

摘要 绿豆作为一种大宗粮食作物, 营养丰富, 可以直接食用, 也可培育为豆芽或制成淀粉及其制品等。研究种子的萌发过程能够为粮食研究提供重要的科学依据和数据基础, 为了研究植物在代谢过程中代谢物的变化, 分别对绿豆种子在 24, 36 和 48 h 的萌发样品进行代谢物的提取, 通过¹H-NMR 测试, 分析绿豆在不同萌发时间内代谢物的变化。¹H-NMR 除无损、稳定等特点外, 尚存在一些不足, 例如该方法灵敏度较低, 在复杂体系中谱峰会存在严重重叠现象, 不易分辨代谢物特征峰, 因此需要化学计量学方法对其中的信息进行提取。采用了多元曲线分辨-交替最小二乘法(MCR-ALS)分析 NMR 的代谢组学数据。MCR 为双线性模型, 能够将测定的复杂数据矩阵分解并提供各主成分 NMR 纯光谱和其对应的相对浓度, 在代谢组学中能够获得具有实际理化意义的结果, 便于检索光谱中对应谱库的代谢物及其浓度变化。仅使用非负约束往往导致双线性模型无法获得唯一解, 在本文中, MCR 迭代过程除了施加非负约束还使用了选择性约束, 对于判断不包含相应代谢物的主成分, 可将其¹H-NMR 谱图强度约束为零。结果表明, 乙酸、胆碱在绿豆萌发前期存在, 萌发后期含量呈现升高趋势; 蔗糖、缬氨酸、 α -酮戊二酸色氨酸、苯丙氨酸、肌苷在萌发之前未见存在, 随绿豆萌发含量逐渐增加; O-甲基鲨肌酸、磷酸胆碱、 α -葡萄糖在萌发前存在, 在萌发初期不断增加, 直至萌发后期, 因被消耗而含量逐渐降低, 表现出先增加后降低的趋势。乳酸、2-氨基-4-氧戊酸、苹果酸、柠檬酸、胆碱、棉籽糖家族寡糖、 α -半乳糖、N1-甲基烟酸在绿豆萌发期间一直存在, 含量未见明显变化。MCR 能够方便地利用先验性质提供 NMR 中包含的代谢物信息, 此方法可推广到不同植物的代谢研究中。

关键词 多元曲线分辨-交替最小二乘; 种子萌发; 核磁共振氢谱

中图分类号: O651 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)10-3010-05

引言

种子萌发是粮食作物生长过程中极为重要的一个环节, 通常分为三个阶段, 第一阶段, 种子吸胀, 细胞质水合化; 第二阶段, 种子吸水速度缓慢, 萌发过程的初始阶段处于相对稳定状态, 代谢变化较少; 第三阶段, 胚根突破种皮幼苗成长。此期间种子中的各种成分急剧变化^[1], 例如磷脂质、甘油三酯、烃类含量不断下降, 而单甘油酯、甾醇类以及游离脂肪酸含量升高^[2]。研究种子代谢过程中成分的变化能够揭示种子萌发生长中的关键代谢通路。通过核磁共振(NMR)、液相色谱-质谱联用(LC-MS)等高通量技术期望同

时鉴定数十甚至数百种的代谢物来推测生物体内发生的代谢变化^[3], 然而谱学技术获得的代谢数据极其复杂, 需要利用化学计量学方法提取其中有效的信息, 了解生物功能, 识别代谢机理和生物标志物^[4]。主成分分析(PCA), 偏最小二乘(PLS), 正交偏最小二乘(OPLS)等以潜变量之间相互正交为核心约束的方法在 NMR 代谢组学中应用极为广泛, 然而数学上的正交并非仪器测量中自然物质存在的信号规律, 经这些方法分析所获得的结果往往很难与光谱库进行直接检索。多元曲线分辨(MCR)方法能够灵活地利用已知条件对数据实现约束, 并获得符合理化意义的波谱结果^[5]。

本研究通过不同约束下的多元曲线分辨-交替最小二乘法(MCR-ALS)解析¹H-NMR 谱, 以实现非靶向代谢组学研

收稿日期: 2019-08-29, 修订日期: 2020-01-08

基金项目: 国家自然科学基金青年科学基金项目(21705112), 北京市教育委员会科学研究项目(KM201710028009, KM201910028014), 北京市优秀人才培养资助青年骨干个人项目(18540090004)和首都师范大学青年创新团队(009175301300)资助

作者简介: 李响, 女, 1996年生, 首都师范大学化学系硕士研究生 e-mail: 2180702014@cnu.edu.cn

* 通讯联系人 e-mail: xinzhang@cnu.edu.cn; zhangzhuoyong@cnu.edu.cn

究分析植物特定时期的差异性代谢,解释绿豆的不同萌发时间阶段代谢物变化规律。

1 实验部分

1.1 样品处理

将绿豆用去离子水浸泡 2 h 后,用 10% NaClO 溶液消毒 15 min,用去离子水清洗 7 次^[6]。将清洗过的绿豆放入烧杯后,置于恒温培养箱中进行萌发,选取萌发时间分别为 24, 36 和 48 h 的样本各 30 例,实验总样本数为 90 例。将萌发后的绿豆样品经液氮处理后研磨,置于 -80 °C 冷冻,冻干。冻干后的粉末样品每份称取 100 mg,分别加入 3 mL 50% 甲醇超声,提取代谢物,取上清液,氮吹后,冻干。测试前加入 600 μ L 重水超声溶解。采用 VARIAN 600 MHz 核磁共振氢谱仪(美国瓦力安公司)进行 ¹H-NMR 测试,谱宽设定为 20 ppm(12 kHz),扫描次数 128,温度为 27 °C,采样数据点为 16 k,以乳酸化学位移 1.336 ppm 的质子信号为标准进行定标。

1.2 数据处理

将测得的核磁数据,利用 MestReNova 软件进行相位校正、基线校正和化学位移漂移校正。相位校正首先通过一阶相位校正方法对数据进行处理,见式(1)和式(2)

$$\text{Re}'_k = \text{Re}_k \cos\left(\alpha + \frac{k}{N}\beta\right) - \text{Im}_k \sin\left(\alpha + \frac{k}{N}\beta\right) \quad (1)$$

$$\text{Im}'_k = \text{Re}_k \sin\left(\alpha + \frac{k}{N}\beta\right) - \text{Im}_k \cos\left(\alpha + \frac{k}{N}\beta\right) \quad (2)$$

其中 $k=0, \dots, N-1$; Re_k 和 Im_k 分别为数据(k)的真实和虚拟主成分数, Re'_k 和 Im'_k 分别为相位校正后数据(k)的真实和虚拟主成分数, N 为数据总数。

基线校正采用了多项式拟和方法,平均积分范围内的所有数据点对数据进行分段积分,积分范围包含 0.80~9.20 ppm,去除重水峰 4.50~5.23 ppm,去除甲醇峰 3.35~3.39 ppm,积分区间为 0.01。最后,通过对总面积进行归一化处理。

将积分后得到的数据导入 Matlab 中,得到 90×841 个变量的数据矩阵。

多元曲线分辨率(MCR)是基于双线性模型的一种化学计量学方法,在数学上将混合物的响应,分解为所研究系统中涉及的每个主成分的纯信号贡献^[7-8]。

多元曲线分辨-交替最小二乘法(MCR-ALS)以交替方法对数据进行迭代,得到数据的浓度分布和光谱矩阵,原理如式(3)

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (3)$$

式(3)中, $\mathbf{D}(m \times n)$ 为核磁共振数据矩阵,有 m 个样本, n 个波谱变量, $\mathbf{C}(m \times p)$ 和 $\mathbf{S}^T(p \times n)$ 为提取后主成分的浓度矩阵和 NMR 波谱矩阵, \mathbf{C} 对应代表绿豆代谢物的浓度矩阵, \mathbf{S}^T 为绿豆代谢物主成分的核磁共振氢谱图。 \mathbf{E} 为误差矩阵。 p 是该体系的主成分个数。

在双线性模型中,不充分的约束会导致最终模型受到旋转模糊的影响,存在非唯一解。在此工作中, NMR 提供的代

谢物信号不可能出现负值,代谢物的信号贡献也不可能为负值,所以可以同时光谱和浓度矩阵施加非负约束。此外,在不同时段存在特定代谢物的生成和消失,可通过选择性约束对不同时段中的代谢物的浓度矩阵进行约束。

2 结果与讨论

首先, MCR 模型需要确定主成分的个数,这将直接影响到 MCR 对结果的解析。依据 PCA 结果确定主成分个数为 3,这时解释原变量方差的百分比为 99.46%,选择更多的主成分可能会过度拟合噪声。

由于在仅使用非负约束的模型中存在不唯一解,将浓度下降趋于零的部分作为约束对象,推测此主成分对应的代谢物是在代谢过程中降解的成分,于是使用选择性约束以实现在迭代过程中寻找更接近真实的解。

在以往报道中,核磁定量分析可以通过 T_1 回溯转换法得以实现但方法较为繁琐,需要内标和待观察物的 T_1 值、以及非全弛豫一维谱图^[6]。通过对 NMR 数据得到的矩阵运用 MCR-ALS 方法进行处理,从而获得三种主成分以及各主成分所包含的代谢物。将每种主成分的响应数据绘制出箱型图,可以根据箱型图直观地看出每种主成分中各个代谢物的相对含量变化趋势,此方法较为直观和简便。

图 1 为经过预处理后的核磁共振谱,每条谱线对应一个待测样本的氢谱。通过图 1 可以看出在各豆种子萌发的 NMR 图谱中存在一定差异,然而由于代谢成分较为复杂,各化合物的特征峰高度重叠,难以获得有用的代谢规律信息。

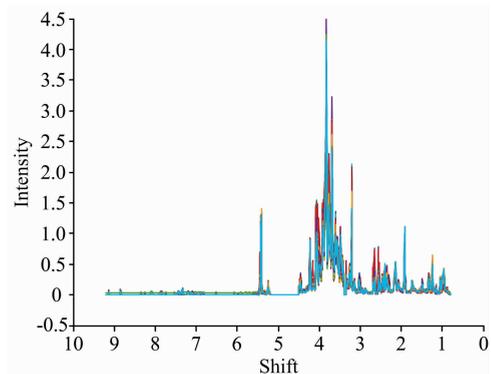


图 1 绿豆萌发代谢物的原始 ¹H-NMR 图谱
Fig. 1 The origin ¹H-NMR spectra of the extracted content during seed germination

图 2 为 MCR 仅在对浓度和光谱方向使用非负约束下的结果,图中分别显示了三个时间段样本浓度变化的箱型图及其对应主成分的 NMR 纯光谱。NMR 纯光谱从图(a2), (b2) 和(c2)中可以看出三种主成分的光谱信息,具体特征峰对应代谢物如表 1。箱型图由最小值(min), 下四分位数(Q1), 中位数(median), 上四分位数(Q3), 最大值(max)绘制而成。从浓度变化图 2(a1)中可以看出第一主成分在第一阶段含量明显降低。如图 2(b1)所示,第二主成分在整个萌发期间含

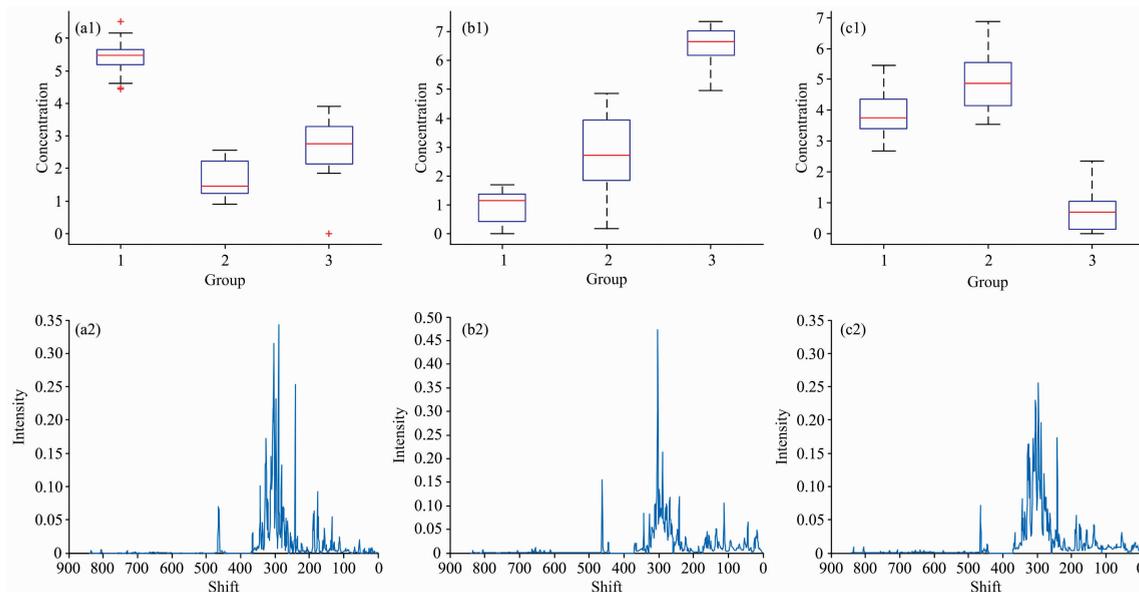


图 2 仅使用非负约束条件下, MCR 获得的三种主成分对应物质的含量变化及其¹H-NMR 图

(a): 第一主成分; (b): 第二主成分; (c): 第三主成分

Fig. 2 The boxplot of concentrations (a1—c1) and the ¹H-NMR spectra (a2—c2) of the three principal components obtained by MCR with non-negative constraints on both concentration and spectra directions

(a): The first principal component; (b): The second principal component; (c): The third principal component

量呈上升趋势。如图 2(c1)所示, 第三主成分在萌发初期含量升高, 在萌发后期含量大幅下降。

其中第二主成分 24 h 组浓度趋向为 0, 推测此主成分对应的代谢物在实际代谢变化中浓度在下降, 可以使用选择性约束对这一位置进行约束。

如图 3 所示, 依照仅非负约束 MCR 结果的提示对第二

主成分, 在 24 h 组的样本浓度方向施加选择性约束。图 3 (a1), (b1)和(c1)三幅图分别为在此约束条件下, 第一主成分、第二主成分以及第三主成分在 24, 36 和 48 h 含量而得出的误差图。图 3(a2), (b2)和(c2)分别为第一主成分、第二主成分以及第三主成分的¹H-NMR 图。如图 3(a1)所示, 第一主成分在萌发初期含量下降, 在萌发后期含量小幅度升

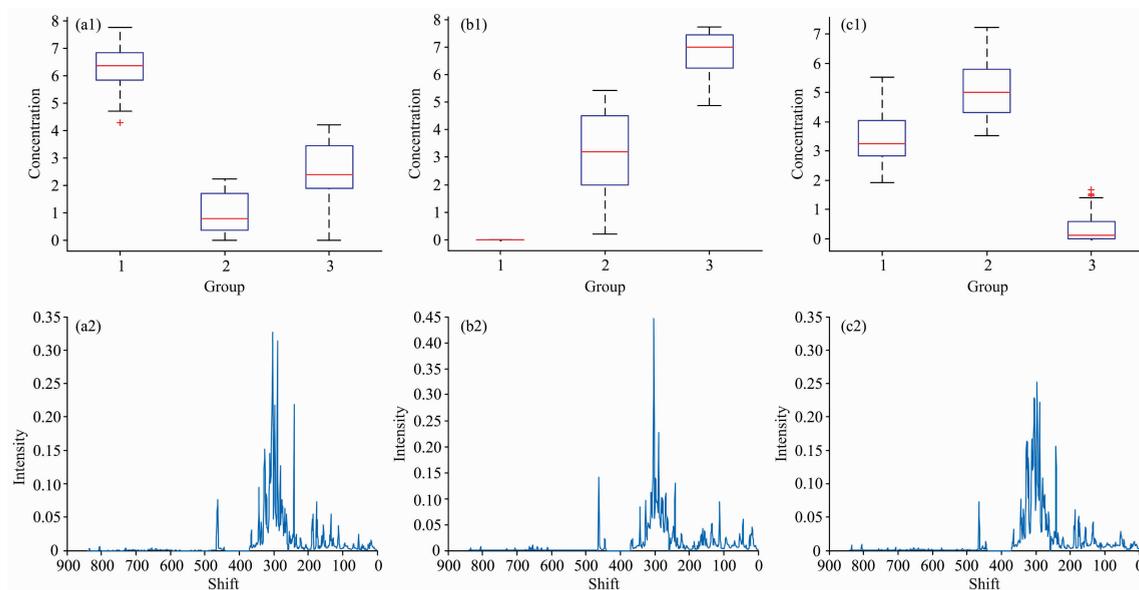


图 3 MCR 结合选择性约束(第二个主成分, 24 h 组)得到的各主成分的¹H-NMR 和浓度箱型图

(a): 第一主成分; (b): 第二主成分; (c): 第三主成分

Fig. 3 The boxplot of concentrations (a1—c1) and the ¹H-NMR spectra (a2—c2) obtained by MCR with selectivity constraint

(a): The first principal component; (b): The second principal component; (c): The third principal component

高。如图 3(b1)所示,第二主成分在整个萌发期间含量保持上升趋势。如图 3(c1)所示,第三主成分在萌发初期含量升高,在萌发后期含量大幅度下降。

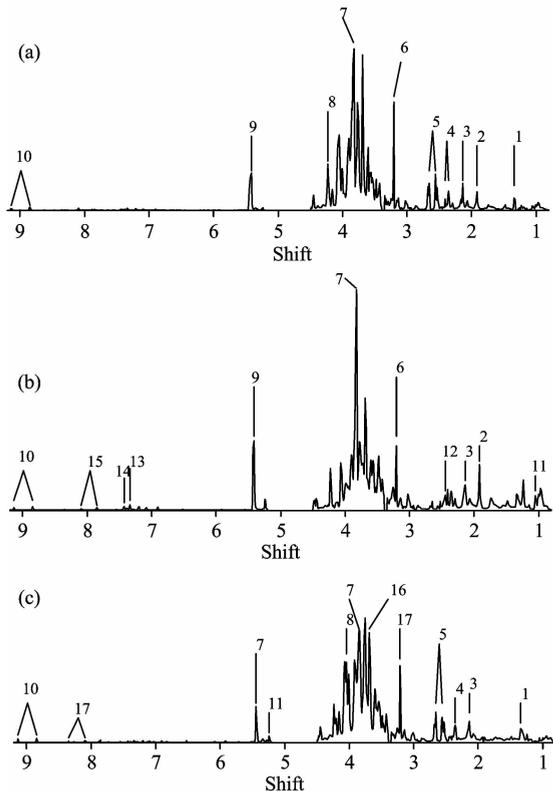


图 4 MCR 结合选择性约束(第二个主成分, 24 h 组)得到的各主成分的¹H-NMR 图

(a): 第一主成分; (b): 第二主成分; (c): 第三主成分

Fig. 4 The ¹H-NMR spectra (a—c) obtained by MCR with selectivity constraint

(a): The first principal component;
(b): The second principal component;
(c): The third principal component

根据文献[6]对绿豆种子提取物中的代谢物特征峰进行识别,可从 MCR 结果的图谱中找到如表 1 所示代谢物。

其中,第一主成分中包含的代谢物有:乳酸、乙酸、2-氨基-4-氧戊酸、胆碱、棉籽糖家族寡糖、 α -半乳糖、蔗糖、N1-甲基烟酸;第二主成分中包含的代谢物有:缬氨酸、乙酸、 α -酮戊二酸、2-氨基-4-氧戊酸、胆碱、棉籽糖家族寡糖、蔗糖、色氨酸、苯丙氨酸、肌苷、N1-甲基烟酸;第三主成分中包含

的代谢物有:乳酸、2-氨基-4-氧戊酸、苹果酸、柠檬酸、O-甲基鲨肌酸、磷酸胆碱、棉籽糖家族寡糖、 α -半乳糖、N1-甲基烟酸。

表 1 绿豆提取物的谱峰归属

Table 1 Peak assignment of mung bean extract

编号	代谢物	化学位移
1	乳酸(lactate)	1.33(d)
2	乙酸(acetate)	1.92(s)
3	2-氨基-4-氧戊酸(AOPA)	2.14(s)2.86(dd)3.01(dd)
4	苹果酸(malate)	2.38(dd)2.69(dd)
5	柠檬酸(citrate)	2.55(dd)2.68(dd)
6	胆碱(choline)	3.20(s)3.52(m)4.06(m)
7	棉籽糖家族寡糖(stachyose)	3.54(t)3.58(dd)3.68(s)3.68(m)3.73(dd)
8	α -半乳糖(α -Gal)	5.27(d)4.07
9	蔗糖(sucrose)	5.42(d)4.22(d)3.89(m)3.84(m)
10	N1-甲基烟酸(NMNA)	4.44(s)8.08(dd)8.83(d)
11	缬氨酸(valine)	0.99(d)1.04(d)2.26(m)
12	α -酮戊二酸(α -KG)	2.45(t)3.02(t)
13	色氨酸(Trp)	3.30(dd)3.47(dd)4.05(dd)7.19(t)7.28(t)
14	苯丙氨酸(Phe)	7.32(m)7.36(m)7.42(m)
15	肌苷(inosine)	8.26(s)8.35(s)
16	O-甲基鲨肌酸(OMSI)	3.14(t)3.34(t)3.36(dd)3.61(s)
17	磷酸胆碱(PC)	3.22(s)3.60(m)4.17(m)
18	α -葡萄糖(α -Glc)	5.25(d)

注: s: 单峰; d: 双重峰; dd 双重峰; t 三重峰; m 多重峰

Note: s: Singlet; d: Doublet; dd: Doublet of doublets; t: Triplet; m: Multiplet

3 结论

通过 MCR-ALS 结合¹H-NMR 谱,实现了对绿豆种子萌发的非靶向代谢组学研究,模型充分利用了测定体系的先验信息进行约束,得到了具有理化意义的光谱信息和相对浓度变化代谢信息。结果表明乙酸、胆碱在绿豆萌发前期存在,萌发后期含量呈现升高趋势;蔗糖、缬氨酸、 α -酮戊二酸、色氨酸、苯丙氨酸、肌苷在萌发之前未见存在,随绿豆萌发时间不断延长,不断合成,并且含量逐渐增加;O-甲基鲨肌酸、磷酸胆碱、 α -葡萄糖在萌发前存在,在萌发初期不断增加,直至萌发后期,因被消耗而含量逐渐降低,表现出先增加后降低的趋势。乳酸、2-氨基-4-氧戊酸、苹果酸、柠檬酸、胆碱、棉籽糖家族寡糖、 α -半乳糖、N1-甲基烟酸在绿豆萌发期间一直存在,含量未见明显变化。

References

- [1] Nonogaki H, Bassel G W, Bewley J D. *Plant Science*, 2010, 179(6): 574.
- [2] AliAbdel-Rahman E S A, El-Fishawy F A, El-Geclawy M A, et al. *International Journal of Food Engineering*, 2007, 3(5): 6.
- [3] Puig-Castellvi F, Alfonso I, Tauler R. *Anal. Chim. Acta*, 2017, 964: 55.
- [4] Motegi H, Tsuboi Y, Saga A, et al. *Sci. Rep.* 2015, 5: 15710.
- [5] Puig-Castellvi F A, Roma Tauler I. *Sci. Rep.* 2016, 6: 30982.
- [6] WU Xiang-yu, LI Ning, TANG Hui-ru(吴香玉, 李 宁, 唐惠儒). *Chinese Journal of Magnetic Resonance(波谱学杂志)*, 2014, 31(4).

[7] Garrido M, Rius F X, Larrechi M S. *Anal. Bioanal. Chem.*, 2008, 390(8): 2059.

[8] Mostafa A, Shaaban H. *Acta Pharm.*, 2019, 69(2): 217.

Analysis of Metabolism During Mung Bean Germination by $^1\text{H-NMR}$ Combining With MCR-ALS With Selectivity Constraint

LI Xiang, LI Zhong-feng, GAO Jun, WANG Xia, ZHANG Xin*, ZHANG Zhuo-yong*

Department of Chemistry, Capital Normal University, Beijing 100048, China

Abstract The research of the Seed germination progress could provide important scientific information and corresponding database for food crops. As one of the primary agricultural products, mung bean is nutrients-rich and can be eaten directly, manufactured as products or cultured to bean sprouts. In order to study the changes of metabolites in plants during the process of metabolism, we extracted the metabolites from the germination samples of mung bean seeds at 24 h, 36 h and 48 h, and analyzed the changes of metabolites in mung bean seeds by $^1\text{H-NMR}$. Except for its non-destructive and stable characteristics, $^1\text{H-NMR}$ still has some shortcomings of low sensitivity and low resolution in the spectrum. In complex systems, there is overlap in spectrum bands for different compounds, which makes it difficult to give enough information on metabolites. Therefore, chemometric methods are needed to interpret the information in NMR spectroscopy. In this paper, multivariate curve resolution-alternating least squares (MCR-ALS) was used for data processing and analysis of the metabolite data based on NMR. MCR is a bilinear model, which can decompose the given data matrix to the pure spectra and their corresponding relative concentration. MCR can provide physical and chemical meaningful results, which facilitate the analysis of metabolites based on the corresponding spectral library and provide their concentration changes. In this paper, besides the non-negative constraints, selectivity constraint was also used in MCR iteration to reduce the rotation ambiguity in MCR modeling. The changes of metabolites and differential metabolites in mung bean at different germination stages were analyzed. In the determination of samples without certain metabolites, the $^1\text{H-NMR}$ spectra are constrained to zero. From the MCR results, we can find that acetate and choline exist in the pre-germination stage of mung bean, and increasing in the late stage of germination. Sucrose, valine, α -ketoglutarate tryptophan, and phenylalanine inosine do not exist in the pre-germination stage, while they gradually increased during the seed germination of the samples. OMSI, phosphorylcholine and α -glucose exist before the germination and increases in the pre-germination stage. Until the late stage of germination, they were consumed and the content decreased. Lactate, 2-amino-4-oxopentanoic acid, Malic acid, citric acid, choline, stachyose, α -galactose, N1-methylnicotinic acid always exist during the germination of mung bean, and there is no significant change.

Keywords Multivariate curve resolution-alternating least squares; Seed germination; Nuclear magnetic resonance spectroscopy

(Received Aug. 29, 2019; accepted Jan. 8, 2020)

* Corresponding authors