

利用三维荧光光谱与平行因子分析法测定 细胞培养基中多类代谢成分的含量

宋晓康^{1,2}, 赵强^{1,3*}, 张元志², 王贻坤², 邓国庆², 朱灵^{2**}

¹安徽建筑大学环境与能源工程学院, 安徽 合肥 230601;

²中国科学院合肥物质科学研究院安徽光学精密机械研究所, 安徽 合肥 230031;

³环境污染控制与废弃物资源化利用安徽省重点实验室, 安徽 合肥 230601

摘要 将三维荧光光谱技术与平行因子分析法相结合, 实现了细胞培养基样本中色氨酸(TRY)、还原烟碱腺嘌呤二核苷酸(NADH)、黄素腺嘌呤二核苷酸(FAD)和维生素 B6(VB6)含量的定性和定量分析。首先制备了 25 个细胞培养基样本, 使用六因素五水平正交试验确定每个样本中四类荧光组分的混合方式和浓度梯度; 采用荧光分光光度计扫描各样本的激发-发射矩阵, 依次进行波长差异性校准、水拉曼和瑞利散射去除、内滤效应校正、光强标准化以及高杠杆点去除等预处理, 然后采用水拉曼谱峰面积作为光强的当量单位, 将荧光光强进行标准化; 采用核一致矩阵确定模型的最佳组分数为 4, 最后使用平行因子分析法对三维荧光光谱数据集进行三线分解, 得到上述种荧光团的平均回收率分别为 100.2%±15.5%、107.4%±37.1%、100.5%±6.4%、99.5%±6.5%, 预测均方根误差(RMSEP)分别为 0.124、43.312、1.601、0.639 μg/mL, 相对平均偏差分别为 13.216%、36.937%、6.112%、6.331%, 检测限分别为 0.013、52.628、0.003、0.012 μg/mL。TRY 的发射光谱与 NADH 的激发光谱重合, 且 FAD 的激发光谱与 NADH 的发射光谱重合, 因此出现了荧光共振转移现象, 该现象导致模型对 NADH 浓度的预测能力低于对其他三种荧光团的预测能力。模型在各荧光团的定性分析方面表现良好, 均通过了 S4C6T3 的拆半分析校验。实验结果表明, 三维荧光光谱结合平行因子分析法能够快速测定细胞培养基中多种代谢类荧光组分的含量, 在细胞能量和物质代谢检测方面具有广阔的应用前景。

关键词 光谱学; 三维荧光光谱; 平行因子分析; 代谢类荧光团; 正交试验设计; 内滤校正; 拆半分析

中图分类号 O657.3

文献标志码 A

DOI: 10.3788/CJL202249.0911001

1 引言

细胞代谢相关的内源性荧光团包括烟酰胺腺嘌呤二核苷酸(NADH)、黄素腺嘌呤二核苷酸(FAD)、色氨酸(TRY)、维生素 B6(VB6)等, 其中: NADH、FAD 普遍存在于真核生物和原核生物中, 是参与有氧呼吸的重要辅酶, 也是线粒体电子传递链上的中间产物, 在细胞的物质代谢和能量代谢中发挥着重要作用^[1-4]。TRY 是一种人体必需氨基酸, 可作为底物参与蛋白质的生物合成, 同时能调控蛋白质的代谢, 从而影响蛋白质的沉积; 此外, TRY 是内分泌激素释放的调节剂, 可作为信号分子调控

三羧酸循环及其下游靶蛋白的活性, 从而促进蛋白质合成。VB6 是一种水溶性维生素, 在碱性条件下会缓慢降解, 它参与氨基酸、碳水化合物和脂质的代谢, 并参与血红蛋白的合成。因此, 定性和定量测定细胞代谢过程中的多类内源性荧光基团的含量, 对于评估细胞生长状态以及预测疾病的发生和发展过程具有重要意义^[5-6]。

体外细胞培养作为一种最直接的研究上述问题的手段, 能在细胞层级上研究疾病的发生和发展机制, 评估细胞的耐药性和趋化性等问题; 但细胞培养基是一种复杂的分析体系, 其包含的多种营养成分(维生素、无机盐、脂肪及核酸衍生物)^[7]会对上述待

收稿日期: 2021-09-01; 修回日期: 2021-09-21; 录用日期: 2021-10-08

基金项目: 国家自然科学基金(51978003)、安徽省科技重大专项计划(17030801007)

通信作者: *rommel99@163.com; **zhul@aiofm.ac.cn

测荧光团的定性和定量分析造成严重干扰,从而成为快速、准确测定代谢类荧光基团含量的难点和关键点。

目前,研究人员已采用色谱法、质谱法和核磁共振谱法实现了荧光团的分离和定量分析^[8-9],其中的三维荧光光谱法是发展快速的荧光分析技术,不仅具有灵敏度高、选择性强以及实时检测的优势,还克服了传统二维荧光光谱中谱峰重叠严重的问题^[10-11]。本文对三维荧光光谱数据集进行水拉曼单位标准化,以提高仪器的动态检测范围;同时,采用平行因子分析法建模,在未进行任何预处理的前提下,利用该算法的高阶优势对培养基中 NADH、FAD、TRY 和 VB6 的混叠荧光光谱进行“数学分离”,用以代替费时和复杂的化学分离,准确、快速地测定了上述 4 种荧光团的含量。

2 方法原理

2.1 平行因子分析法

在设定激发波长数为 I 、发射波长数为 J 的条件下,分别采集 K 个多组分混合样品的荧光光谱图,获得激发-发射光谱矩阵数据,然后将这些光谱矩阵数据按照样本次序进行叠加,得到大小为 $I \times J \times K$ 的三维响应数阵 \mathbf{X} 。 x_{ijk} 是该立方阵中的元

素,其数学表达式为

$$x_{ijk} = \sum_{n=1}^N \mathbf{a}_{in} \mathbf{b}_{jn} \mathbf{c}_{kn} + e_{ijk} \quad (i=1, \dots, I; j=1, \dots, J; k=1, \dots, K), \quad (1)$$

式中: \mathbf{a}_{in} 是大小为 $I \times K$ 的相对激发矩阵; \mathbf{b}_{jn} 是大小为 $J \times N$ 的相对发射光谱矩阵; \mathbf{c}_{kn} 是大小为 $K \times N$ 的相对浓度矩阵; e_{ijk} 是大小为 $I \times J \times K$ 的残差数据阵 \mathbf{E} 中的元素; N 表示的是模型因子数,对应于模型的最佳组分数^[12]。

平行因子分析法可以视为主成分分析(PCA)法由二维空间向三维空间的一种扩展,解决了 PCA 法存在的旋转不定性问题^[13-14]。在三维荧光光谱立方数据阵中,每一个样本的光谱数据都表示为

$$\mathbf{X}_{..k} = \mathbf{A} \text{diag}(\mathbf{c}_{(k)}) \mathbf{B}^T + \mathbf{E}_{..k} \quad (k=1, \dots, K), \quad (2)$$

式中: k 为样本编号;对角阵 $\text{diag}(\mathbf{c}_{(k)})$ 表达了不同样本之间的荧光强度以及单个样本中所含不同组分类型的信息; \mathbf{A} 和 \mathbf{B} 为轮廓矩阵,表达了全部组分的定性信息。 \mathbf{A} 和 \mathbf{B} 的求解过程是先确定建模的组分数 N ,并初始化矩阵 \mathbf{A} 和 \mathbf{B} ,表达式如式(3)、(4)所示,采用了交替最小二乘法,且要求残差平方和最小。

$$\mathbf{A} = \left[\sum_{k=1}^K \mathbf{X}_{..k} \mathbf{B} \text{diag}(\mathbf{c}_{(k)}) \right] \left[\sum_{k=1}^K \text{diag}(\mathbf{c}_{(k)}) \mathbf{B}^T \mathbf{B} \text{diag}(\mathbf{c}_{(k)}) \right]^{-1}, \quad (3)$$

$$\mathbf{B} = \left[\sum_{k=1}^K \mathbf{X}_{..k} \mathbf{A} \text{diag}(\mathbf{c}_{(k)}) \right] \left[\sum_{k=1}^K \text{diag}(\mathbf{c}_{(k)}) \mathbf{A}^T \mathbf{A} \text{diag}(\mathbf{c}_{(k)}) \right]^{-1}. \quad (4)$$

求解 \mathbf{c} 是对其每一行进行更新,循环 k 次来更新整个矩阵,得到

$$\mathbf{c}_{(k)} = \left\{ \left[(\mathbf{A}^T \mathbf{A}) \cdot (\mathbf{B}^T \mathbf{B}) \right]^{-1} \text{diagm}(\mathbf{A}^T \mathbf{X}_{..k} \mathbf{B}) \right\}^T. \quad (5)$$

重复计算式(3)~(5),直到满足收敛标准

$$S^{(m)} = \sum_{i=1}^n \left\| \mathbf{X}_{..k} - \mathbf{A}^{(m)} \text{diag}(\mathbf{c}_{(k)})^{(m)} (\mathbf{B}^T)^{(m)} \right\|_F^2, \quad \left| \frac{S^{(m)} - S^{(m-1)}}{S^{(m-1)}} \right| \leq \epsilon, \quad (6)$$

式中: S 表示残差平方和; m 是迭代次数; ϵ 为收敛阈值。

2.2 核一致诊断法

平行因子分析法建模需要先给定最佳的组分数再进行三线性成分分解,本文采用核一致诊断法来确定模型组分数。核一致函数值 C_c 的表达式为

$$C_c = \left[1 - \frac{\sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F (g_{def} - t_{def})^2}{\sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F t_{def}^2} \right] \times 100\%, \quad (7)$$

式中: g_{def} 表示 Tucker3 计算得到的核心阵元素; t_{def} 表示对角阵元素为 1 的三维数据阵中的元素; F 为建模时选定的组分数。随着选择的组分数增多,核一致函数值会从 100% 逐渐下降,并接近于零,甚至是负数。因此,阈值设定为 60%,该值对应的组分数即为最佳组分数。当核一致函数值低于 60% 时,认为模型偏离三线性,组分数选择不合理^[14-15]。

2.3 品质因子

分析方法的品质因子,如预测均方根误差(RMSEP)、检测限(LOD)、回收率、线性相关性、

95%置信区间,均可以用来评估、分析结果的可靠性和适应性,同时也可以作为选择方法和优化实验条件的目标函数。

预测均方根误差(RMSEP, E_{RMSEP})和平均相对误差(ARE, E_{ARE})的计算公式分别为

$$E_{\text{RMSEP}} = \left[\frac{1}{k_s - 1} \sum_{k'=1}^{k_s} (C_{\text{real},k'} - C_{\text{pred},k'})^2 \right]^{\frac{1}{2}}, \quad (8)$$

$$E_{\text{ARE}} = (1/k_s) \sum_{i=1}^l \left| (C_{\text{real},i} - C_{\text{pred},i}) / C_{\text{true},i} \right| \times 100\%, \quad (9)$$

式中: C_{real} 和 C_{pred} 分别表示各组分的真实浓度和模型的预测浓度; k_s 表示样本个数; l 表示样本浓度的水平数,也是浓度梯度的个数。RMSEP和ARE的值越小,表明算法获得的预测值比较接近真实值,模型比较可靠。国际纯粹应用化学联合会(IUPAC)规定,对于各种光学分析方法,检出限可以用仪器噪声确定。检测限(LOD, f_{LOD})的计算公式为^[16]

$$f_{\text{LOD}} = (X_L - X_0) / k_c = K_p S_0 / k_c, \quad (10)$$

式中: X_L 为可测量的最小分析信号; X_0 为空白多次测量信号的平均值; K_p 是由置信水平确定的系数(当置信水平设定为设定95%时, $K_p = 3.2$); S_0 为空白多次测量信号的标准偏差; k_c 为校正曲线的斜率,即灵敏度。

3 实验部分

3.1 仪器与试剂

采用日立公司F-7000荧光分光光度计扫描样本得到激发-发射光谱;激发和发射光的狭缝宽度均为5 nm,光电倍增管(PMT)检测器的电压设置为650 V,扫描速度为2400 nm/min;激发光波段为250~500 nm,步长为5 nm;发射光波段为250~700 nm,步长为5 nm。采用岛津UV-2550紫外可见分光光度计测量样本的吸光度,波段范围为200~750 nm,步长为2 nm。

试剂 NADH ($C_{21}H_{27}N_7Na_2O_{14}P_2 \cdot xH_2O$)、FAD($C_{27}H_{31}N_9Na_2O_{15}P_2 \cdot xH_2O$)、TRY($C_{11}H_{12}N_2O_2$)和VB6($C_8H_{11}NO_3 \cdot HCl$)均购于Sigma-Aldrich公司,4-羟乙基哌嗪乙磺酸(HEPES)购于阿拉丁试剂(上海)有限公司,罗丹明B购于生工生物工程(上海)股份有限公司,CHO细胞无蛋白化学培养基购于上海百赛生物科技股份有限公司。

3.2 样品配制

使用电子天平分别称取75 mg TRY、100 mg NADH、41.28 mg VB6、50 mg FAD,将其分别移至

50 mL具塞试管中,然后用10 mmol/L的HEPES缓冲液溶解并定容至50 mL,得到这4种荧光物质的储备液。然后,使用无蛋白化学培养基溶液进行稀释,将TRY储备液稀释至12 $\mu\text{g}/\text{mL}$,NADH储备液不进行稀释(直接作为标准使用液),将FAD和VB6储备液分别稀释至0.2 mg/mL和0.164 mg/mL,从而得到4种荧光物质的标准使用液。

最后采用六因子五水平正交试验设计表 $L_{25}(5^6)$ 给出的方式混合上述荧光组分的标准使用液,得到最终用于光谱检测的25个实验样本。采用正交试验能以最少的试验次数获得与全因子设计等效的试验结果。表1中的第2列数字依次代表了TRY、NADH、FAD和VB6浓度的水平值,取自 $L_{25}(5^6)$ 标准表的后4列。实验流程如下:先移取TRY、NADH、FAD、VB6标准使用液至10 mL试管中,然后用无蛋白细胞培养基进行稀释,得到TRY、NADH、FAD、VB6质量浓度分别为0.12,20,2,1.65 $\mu\text{g}/\text{mL}$ 的1号样本;如此反复,完成25个样本的制备工作。

表1 正交试验的水平数以及TRY、NADH、FAD、VB6浓度配制表

Table 1 Levels of orthogonal test and concentration of TRY, NADH, FAD, and VB6

| No. | Level | Mass concentration / ($\mu\text{g} \cdot \text{mL}^{-1}$) | | | |
|-----|-------|---|------|-----|------|
| | | TRY | NADH | FAD | VB6 |
| 1 | 1111 | 0.12 | 20 | 2 | 1.65 |
| 2 | 2222 | 0.24 | 40 | 4 | 3.3 |
| 3 | 3333 | 0.48 | 80 | 8 | 6.6 |
| 4 | 4444 | 0.96 | 160 | 16 | 13.2 |
| 5 | 5555 | 1.92 | 320 | 32 | 26.4 |
| 6 | 2345 | 0.24 | 80 | 16 | 26.4 |
| 7 | 3451 | 0.48 | 160 | 32 | 1.65 |
| 8 | 4512 | 0.96 | 320 | 2 | 3.3 |
| 9 | 5123 | 1.92 | 20 | 4 | 6.6 |
| 10 | 1234 | 0.12 | 40 | 8 | 13.2 |
| 11 | 3524 | 0.48 | 320 | 4 | 13.2 |
| 12 | 4135 | 0.96 | 20 | 8 | 26.4 |
| 13 | 5241 | 1.92 | 40 | 16 | 1.65 |
| 14 | 1352 | 0.12 | 80 | 32 | 3.3 |
| 15 | 2413 | 0.24 | 160 | 2 | 6.6 |
| 16 | 4253 | 0.96 | 40 | 32 | 6.6 |
| 17 | 5314 | 1.92 | 80 | 2 | 13.2 |
| 18 | 1425 | 0.12 | 160 | 4 | 26.4 |
| 19 | 2531 | 0.24 | 320 | 8 | 1.65 |
| 20 | 3142 | 0.48 | 20 | 16 | 3.3 |
| 21 | 5432 | 1.92 | 160 | 8 | 3.3 |
| 22 | 1543 | 0.12 | 320 | 16 | 6.6 |
| 23 | 2154 | 0.24 | 20 | 32 | 13.2 |
| 24 | 3215 | 0.48 | 40 | 2 | 26.4 |
| 25 | 4321 | 0.96 | 80 | 4 | 1.65 |

4 分析与讨论

4.1 荧光强度的标准化和归一化

荧光强度的标准化不仅提升了仪器的动态检测范围,而且能在平行因子分析法建模中提高算法对化学组分的分辨率。标准化流程如图 1 所示。第一步是对原始数据进行校正,即将空白样本的激发发射谱(EEM blanks)、350 nm 激发波长下纯净水的拉曼光谱(Rm350)与校正因子矩阵 C 相乘,以此来校正荧光分光光度计的系统误差。首先采集质量浓度为 3 g/L 的罗丹明 B 溶液的激发谱,计算得到仪器激发光单色器的校正因子(E_{cor});

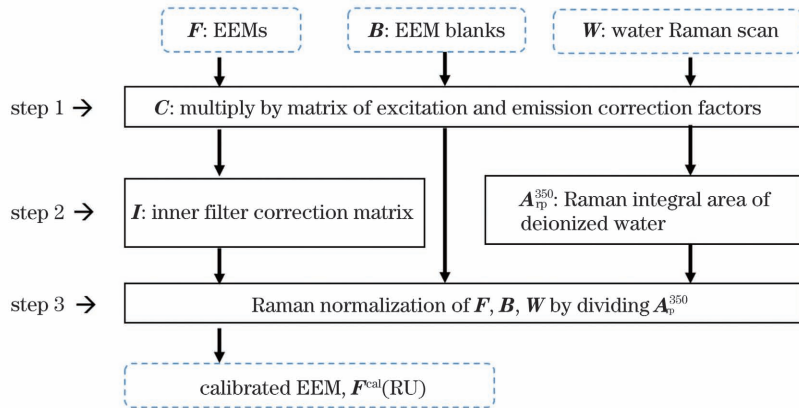


图 1 原始数据集的标准化处理流程图

Fig. 1 Standardizing process procedure of raw dataset

采集纯净水的拉曼光谱图($E_x = 350$ nm),其对应的特征峰位于 397 nm 附近,计算峰面积 A_p^{350} 。第三步是对荧光强度值进行标准化,标准化公式为

$$F^{\text{cal}}(\text{RU}) = (1/A_p^{350}) \cdot F^{\text{IFE}}, \quad (11)$$

即:使用水的拉曼峰面积,将 AU 单位转化为拉曼单位(RU350),最终得到 $F^{\text{cal}}(\text{RU})$ [19]。该方法不仅可以校正由光源强度和检测器响应波动导致的实测数据的差异性,还扩大了仪器的动态检测范围。这是因为在实验中为了保证数据集中浓度较高和较低样本的荧光光谱数据均具有较高的信噪比,通常会根据样本的实测光谱的强弱来反向调整仪器参数(如,单色器的狭缝和光电倍增管的电压)。仪器调整后,实测数据如果继续用 AU 单位表示,将会丢失样本的相对浓度信息,而拉曼单位(RU350)是一种表示光强的相对单位,仪器参数变动不会对其造成干扰,因此使用该单位要优于使用 AU 单位。

然后,去除数据集中对荧光强度归一化产生干扰的数据点(即由样本介质产生的瑞利散射光和拉曼散射光的光谱数据),并使用埃尔米特插值法进行插值处理。最后,对各样本的荧光强度进行归一化,

然后采集涂覆有 BaSO_4 的标准漫反射板的荧光同步谱($\Delta\lambda = 0$),计算得到仪器发射光单色器的校正因子(E_{mcor})^[17-18],再根据公式 $C = E_{\text{mcor}}(E_{\text{mcor}})^T$ 计算得到校正因子 C ;然后对实测荧光光谱 $F(\text{AU})$ 进行校正,得到 F^{corr} ,即 $F^{\text{corr}} = C \cdot F(\text{AU})$ 。第二步是校正内滤效应的和计算水拉曼谱的积分面积。内滤校正矩阵 I 由激发波段和发射波段的吸光度计算得到,计算公式为 $I_{\lambda_{\text{ex}}, \lambda_{\text{em}}} = 10^{0.5(A_{\lambda_{\text{ex}}} + A_{\lambda_{\text{em}}})}$,其中 $A_{\lambda_{\text{ex}}}$ 为激发光谱波段下的吸光度, $A_{\lambda_{\text{em}}}$ 为发射光谱波段下的吸光度,再按照公式 $F^{\text{IFE}} = I \cdot F^{\text{corr}}$ 计算得到 F^{IFE} 。

其主要目的是使平行因子分析算法聚焦于不同样本的三维荧光光谱的波形差异,而不是强度信息。这是因为浓度高(或者光谱强度高)的样本具有较高的杠杆值,会导致不同的荧光组分的光谱呈现共线性^[20]。对方阵数据集的样本维度数据[对应于样本激发-发射光谱(EEM)的强度]进行归一化能够降低这种共线性效应,从而提高模型对荧光组分的分辨能力;并且程序中会保存所有样本的归一化系数,在完成三线性成分分解之后,会用该系数完成逆向归一化,得到各荧光团真实的光谱强度,用于后续荧光组分的定量分析。

4.2 荧光组分的提取与模型校验

平行因子法建模的思路是先设定较宽范围的组分数,并依次运行平行因子算法,然后识别并去除载荷矩阵、得分矩阵中的高杠杆值和异常值,最后采用核一致诊断法确定模型的最佳组分数,从而得到最佳模型。本实验中首先设定的模型组分数为 3~5,采用平行因子分析法分解后得到对应的荧光发射光谱和激发光谱,如图 2 所示。分析图 2 可知:各组分的激发光谱在紫外波段重叠严重,这主要是因为这

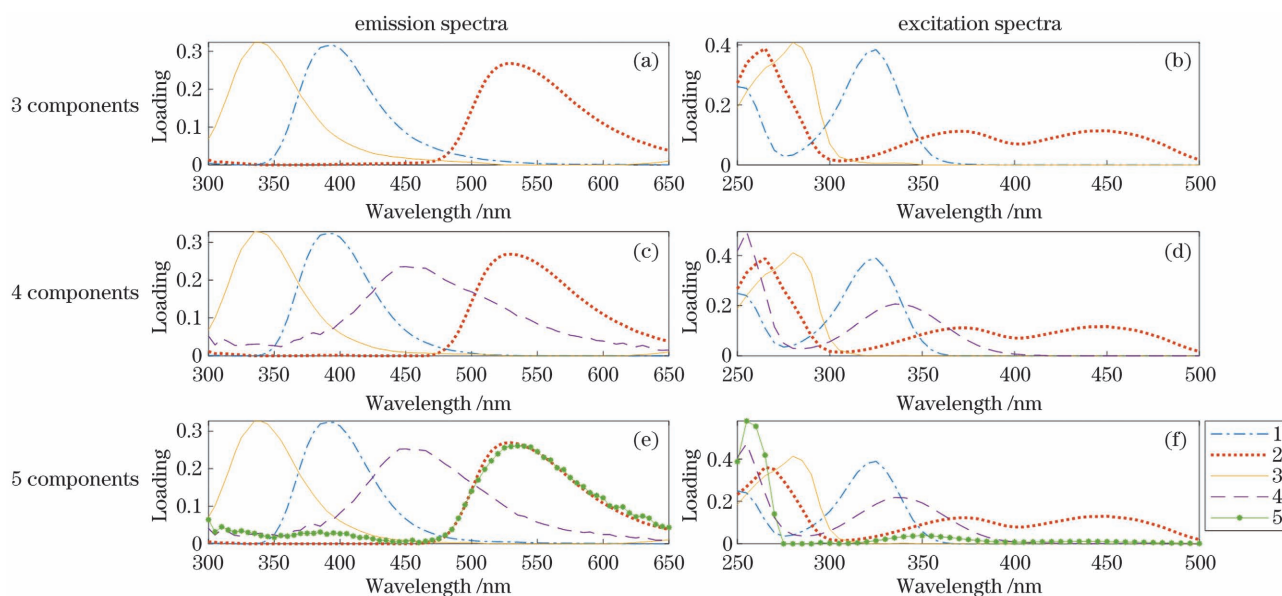


图 2 设定 3~5 组分下得到的荧光发射光谱和激发光谱

Fig. 2 Fluorescence emission and excitation matrix spectra based on 3 to 5 components configuration

几类荧光团在该波段下存在严重的吸收现象;当组分数增加到 5 时,第 5 组分与第 2 组分的激发光谱和发射光谱非常接近,即相关性较高,在后续的核心一致诊断中会再次进行判断,确定是否将其作为单独组分纳入最终的模型中。

数据集中异常值的来源包括实验过程中的粗大误差和仪器波动等因素,这些因素会导致数据集中存在高杠杆点或强影响点。高杠杆点又称为自变量因子空间中的离群点,需要使用帽子矩阵作为判据对其进行判断,即:当数据点的帽子值大于帽子均值的 2~3 倍时,就予以剔除。回归到本实验中,以四组分设定下的平行因子分析法的建模为例,作为典

型的三线性成分模型,它存在 3 个自变量因子空间,分别是样本维度的得分值[score,如图 3(a)所示]、发射光谱维度的载荷值[loading 1,如图 3(b)所示]和激发光谱维度的载荷值[loading 2,如图 3(c)所示],因此需要依次识别不同维度下的高杠杆点^[20-21]。3 个维度下高杠杆值的计算结果如 3(d)~(f)所示。在样本维度上,8 号样本的杠杆值为 0.68[如图 3(d)所示],大于高杠杆值的判定阈值,予以剔除;在两个波长维度上,激发波长 $E_x = 280$ nm 和发射波长 $E_m = 350$ nm 附近的杠杆值虽然高于其他波段,但均未超出判定阈值,因此予以保留。

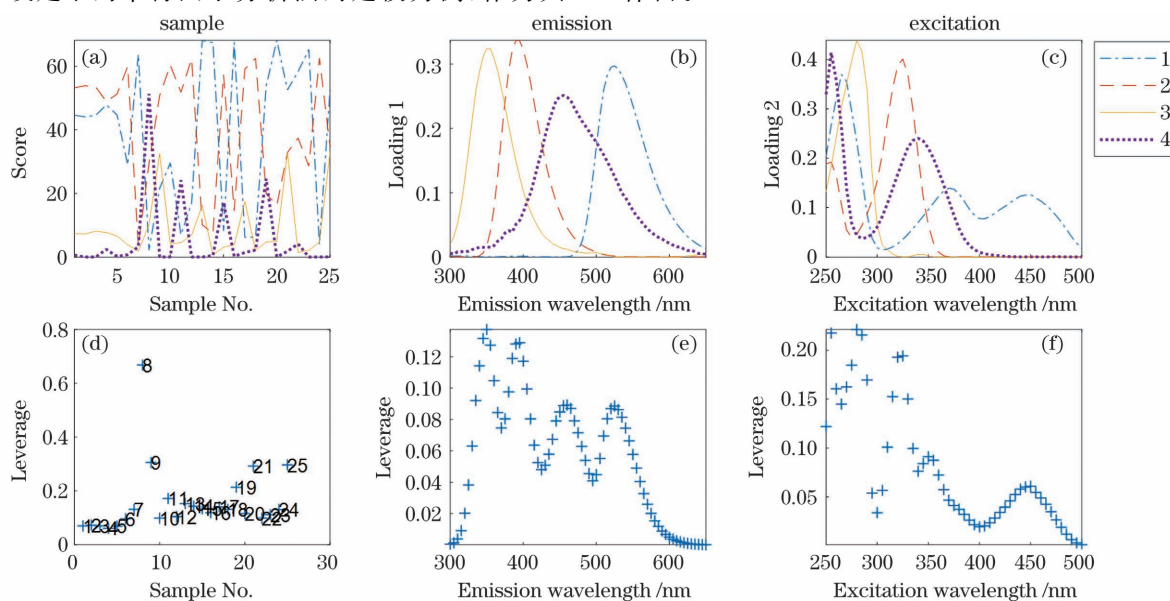


图 3 立方阵中 3 个维度上高杠杆点的识别图

Fig. 3 High leverage data identification from 3 dimensions within the cubic matrix

在建模初期设定了组分数的范围是 3~5,但是在定量分析中需要给出最佳且唯一的组分数。已经有学者归纳出了多种确定最佳组分数的方法^[22-23],如残差分析法、加一法等,但较为常用的是由 Bro 提出的核一致诊断法^[14]。如式(7)所示,在设定不同组分数的前提下,核一致函数采用

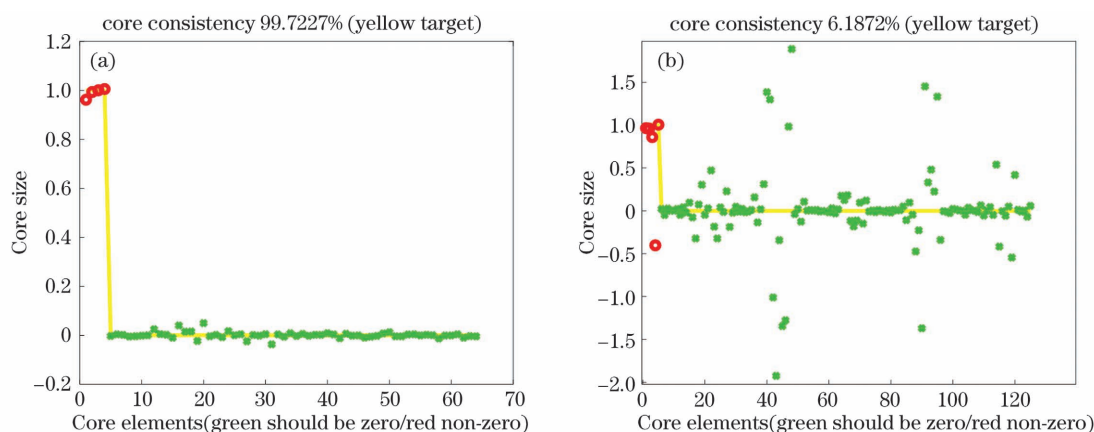


图 4 三组分和四组分建模得到的核一致函数值。(a)三组分;(b)四组分

Fig. 4 Core consistency value of three and four components models. (a) Three components; (b) four components

在确定模型的最终组分数为 4 后,需要再次进行模型精炼,即设定更加严苛的迭代收敛标准(10^{-8}),并且重复运行 10 次,将残差最小的模型定为最佳模型。经过三线分解得到了 VB6、FAD、TRY 和 NADH 的三维荧光光谱,如图 5 所示。VB6、FAD、TRY 和 NADH 的最佳激发波长/发射波长分别是 332 nm/400 nm、450 nm/550 nm、280 nm/350 nm 和 350 nm/450 nm,这与标准图谱

Tucker3 方法计算所有组分的核心阵与取得最佳成分时核心阵的拟合程度。如图 4(a)所示,四组分的核一致函数值为 99.7227%。当组分数增加到 5 时,如图 4(b)所示,函数值骤降为 6.1872%。根据组分数的判断阈值(60%),可得出最佳组分数为 4。

数据中给出的结果一致^[24]。

最后采用拆半分析法^[20,25]校验上述模型。数据集的分配方式采用 S4C6T3,即:将整个数据集拆分为 4 个子集(A、B、C、D),然后将其两两配对得到 6 个组合(AB、CD、AC、BD、AD、BC);之后,将这些组合进行两两配对,构成 3 个测试集(AB vs. CD、AC vs. BD、AD vs. BC);然后对每个测试集内的第一个和第二个组合对应的数据集均采用平行因子

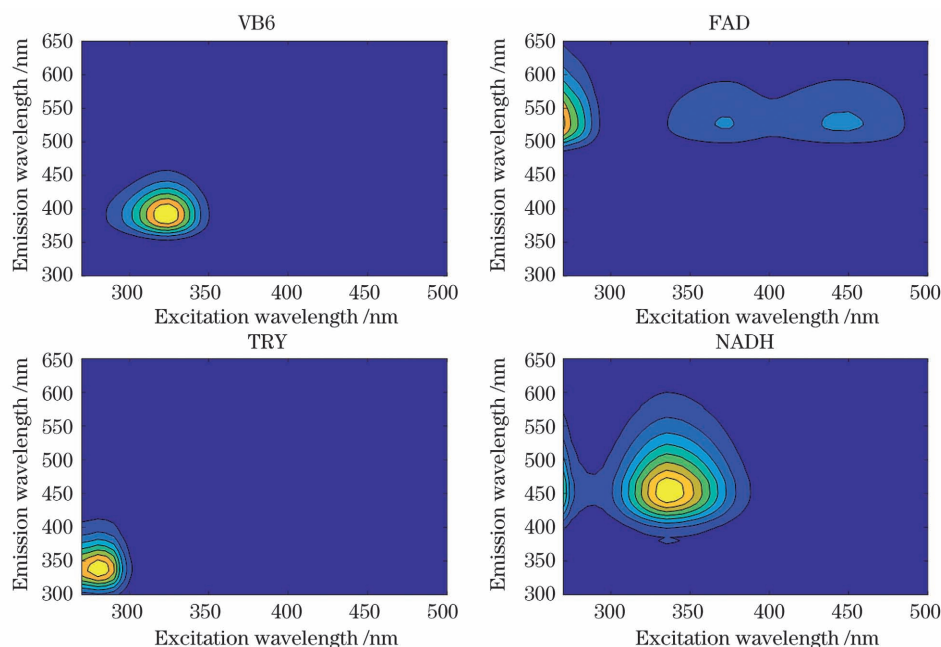


图 5 建模后四组分(VB6、FAD、TRY、NADH)的三维等高线图

Fig. 5 Three-dimensional contour maps of four components (VB6, FAD, TRY, and NADH) after modeling

分析算法进行荧光组分分离,然后对比分离效果的差异性。图 6 和图 7 分别表示拆半分析得到的激发光谱图和发射光谱图,下面以图 6(a)为例进行具体说明。实线对应于第一个组合 AB (1st composition, 简称为“1st comp”),虚线对应于第二个组合 CD(2nd comp),1st comp 1(VB6)表示在第一个组合的数据集上建模分离所得的第一个组分,也就是 VB6。以此类推,图 6(b)~(c)和图 7(a)~

(c)均采用该方式标注光谱曲线。此外,实线与虚线的重合程度越高,表明模型对组分分离的一致性越好。本文采用假设性检验方法来定量评估这种相似度,设定的标准是:当概率 $p > 0.95$ 时,认为模型通过拆半分析。对图 6 和图 7 的各组光谱曲线进行假设性检验,结果发现均满足上述判定条件,证明了本模型在组分分离和浓度预测方面的有效性。

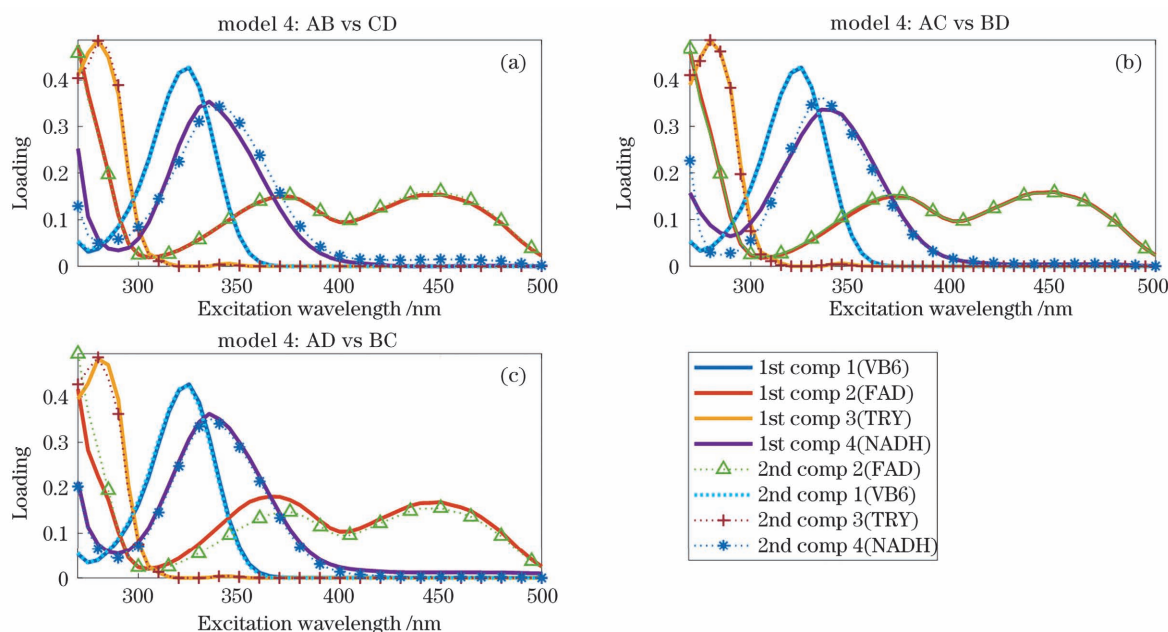


图 6 拆半分析(S4C6T3)后所得的激发光谱

Fig. 6 Excitation spectra after split-half analysis (S4C6T3)

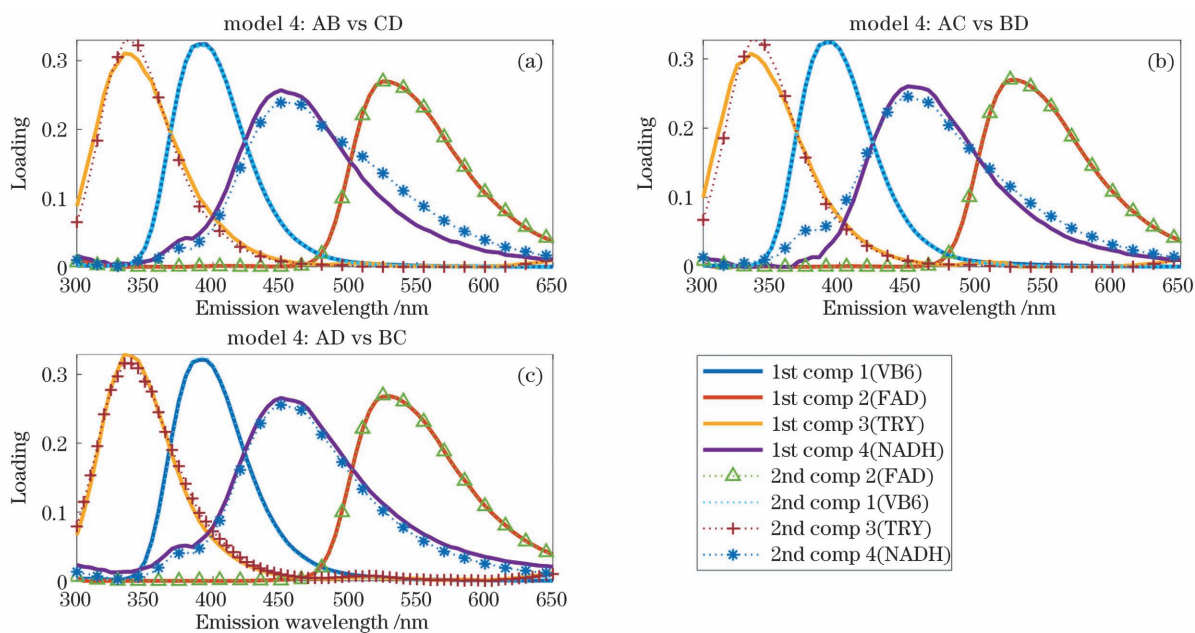


图 7 拆半分析(S4C6T3)后所得的发射光谱

Fig. 7 Emission spectra after split-half analysis (S4C6T3)

4.3 多类代谢荧光团浓度的定量分析

根据荧光分析法的基本原理,在一定频率及强度的激发光照射下,荧光强度与荧光团的含量成正比,而且在平行因子分析法建模后,已获取单荧光组分的浓度和光谱强度。因此,可以绘制两者的标准曲线。如图 8 所示,原始数据点使用○表示,拟合结果如蓝色标准曲线(1st fitted curve)所示;但是由于异常点的存在,因此采用肖维勒准则^[26]识别这些数

据点(如图中×所示),并予以剔除。再次拟合的结果如红色标准曲线(2nd fitted curve)所示。图中的 95%置信带用于评估模型对浓度预测的精度,置信带越窄,预测的精度越高。此外,检测下限(LOD)、预测均方根误差(RMSEP)、回收率以及相对平均偏差(ARE)也能验证实验方法的可靠性,它们的计算公式如式(8)~(10)所示,结果如表 2 和表 3 所示。

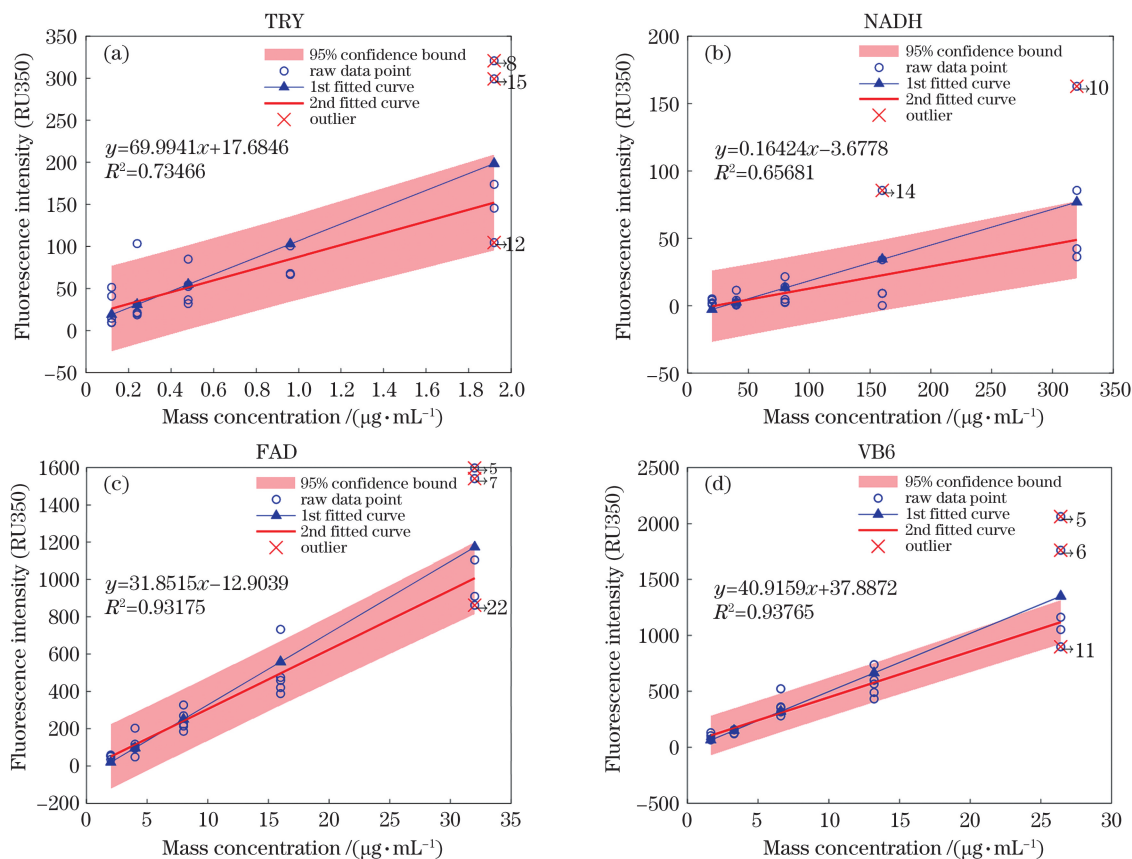


图 8 TRY、NADH、FAD 和 VB6 的荧光强度与组分浓度之间的关系图

Fig. 8 Relationship between fluorescence intensity and concentration of TRY, NADH, FAD and VB6

表 2 采用 PARAFAC 预测的 TRY、NADH、FAD、VB6 的浓度和回收率

Table 2 Predication mass concentration and recovery of TRY, NADH, FAD, and VB6 by PARAFAC

| Fluorophore | Test sets | Added mass concentration / ($\mu\text{g}\cdot\text{mL}^{-1}$) | Predicted mass concentration / ($\mu\text{g}\cdot\text{mL}^{-1}$) | Recovery / % |
|-------------|-----------|--|--|--------------|
| TRY | T1 | 1.920 | 1.827 | 95.2 |
| | T2 | 0.120 | 0.118 | 98.6 |
| | T3 | 0.240 | 0.193 | 80.6 |
| | T4 | 0.480 | 0.496 | 103.3 |
| | T5 | 0.960 | 1.185 | 123.4 |
| NADH | T1 | 160 | 230.428 | 144.0 |
| | T2 | 320 | 279.800 | 87.4 |
| | T3 | 20 | 29.673 | 148.3 |
| | T4 | 40 | 37.255 | 93.1 |
| | T5 | 80 | 51.249 | 64.1 |

续表

| Fluorophore | Test sets | Added mass concentration / ($\mu\text{g}\cdot\text{mL}^{-1}$) | Predicted mass concentration / ($\mu\text{g}\cdot\text{mL}^{-1}$) | Recovery /% |
|-------------|-----------|--|--|-------------|
| FAD | T1 | 8 | 7.450 | 93.1 |
| | T2 | 16 | 15.326 | 95.8 |
| | T3 | 32 | 35.081 | 109.6 |
| | T4 | 2 | 2.030 | 101.5 |
| | T5 | 4 | 4.090 | 102.5 |
| VB6 | T1 | 3.3 | 2.941 | 89.1 |
| | T2 | 6.6 | 6.813 | 103.2 |
| | T3 | 13.2 | 13.741 | 104.1 |
| | T4 | 26.4 | 27.479 | 104.0 |
| | T5 | 1.65 | 1.60 | 97.1 |

表 3 采用 PARAFAC 得到的算法的精度

Table 3 Precision of the algorithm obtained by using PARAFAC

| Fluorophore | Average recovery /% | RMSEP /($\mu\text{g}\cdot\text{mL}^{-1}$) | ARE /% | LOD /($\mu\text{g}\cdot\text{mL}^{-1}$) |
|-------------|---------------------|---|--------|---|
| TRY | 100.2±15.5 | 0.124 | 13.216 | 0.013 |
| NADH | 107.4±37.1 | 43.312 | 36.937 | 52.628 |
| FAD | 100.5±6.4 | 1.601 | 6.112 | 0.003 |
| VB6 | 99.5±6.5 | 0.639 | 6.331 | 0.012 |

通过对上述图表进行分析可以发现:NADH 的残差平方和较大, R^2 较小,95%置信带的区间较宽。这说明本模型对 NADH 浓度的预测能力相对较差。首要的原因是荧光共振能量转移现象的存在,即: $\lambda=350\text{ nm}$ 既是 TRY 的荧光发射光谱的特征波长又是 NADH 的激发光谱的特征波长, $\lambda=450\text{ nm}$ 既是 NADH 的发射光谱的特征波长又是 FAD 的激发光谱的特征波长。因此,在上述波段处,荧光组分之间会发生严重的自吸现象,导致多组分的荧光产生过程已经不满足三线性建模的前提条件(即多组分的荧光强度具有线性叠加特性)。

另外,由于单色器光栅的分光特性所产生的倍频峰干扰,如图 5 中 FAD 的荧光峰(E_x/E_m)位于 450 nm/550 nm 位置,对应的 1/2 倍频峰位于 275 nm/550 nm 位置,导致平行因子分析算法将该倍频峰视为单独的荧光组分,化学分辨率随之下降。但本文通过计算核一致矩阵选定了最佳的组分数,降低了该因素对模型的干扰。综合特征峰的比对、标准曲线以及表 3 中各品质因子的分析结果可以发现,三维荧光光谱法结合平行因子分析法能够在存在多种干扰的细胞培养基中,直接、快速地实现对 NADH、FAD、TRY 和 VB6 的定量分析。

5 结 论

细胞培养基中代谢类荧光团的实测荧光光谱会

受到各种营养成分以及生长因子等背景的干扰,而且各类荧光团之间也会因为化学结构类似而导致光谱重叠严重,难以直接对单组分进行定性和定量分析。为了解决该问题,本文首先使用罗丹明 B 和标准反射板校正荧光分光光度计;然后采用水拉曼单位(RU350)进行荧光强度归一化,以提高仪器的动态检测范围;接着采用三维荧光光谱结合平行因子分析算法对实测光谱数据进行建模分析,利用该算法的“二阶优势”,取代复杂的“化学分离”,对培养基中含有的 TRY、NADH、FAD 和 VB6 荧光组分进行“数学分离”,得到上述组分的荧光特征峰(E_x/E_m)分别位于 280 nm/350 nm、350 nm/450 nm、450 nm/550 nm、332 nm/400 nm。

在定量分析方面,NADH 的预测均方根误差、相对平均偏差以及检测限明显大于其他三种荧光组分。这是因为 NADH 在 350 nm 处与 TRY,在 450 nm 处与 FAD 发生了荧光共振能量转移现象。其他三种组分,即 TRY、FAD 和 VB6,受该物理现象的影响较小,它们的平均回收率在(99.5±6.5)%~(100.5±6.4)%范围内,检测限在 0.003~0.013 $\mu\text{g}/\text{mL}$ 范围内。这表明该方法已经达到了痕量分析的检测级别,满足对培养基中代谢类荧光组分定量检测的需求,在体外细胞代谢和免疫研究、体外医学诊断、细胞分选等领域具有很大的应用潜力。

参 考 文 献

- [1] Deal J, Harris B, Martin W, et al. Demystifying autofluorescence with excitation scanning hyperspectral imaging [J]. *Proceedings of SPIE*, 2018, 10497: 1049715.
- [2] Lagarto J L, Dyer B T, Talbot C B, et al. Characterization of NAD (P) H and FAD autofluorescence signatures in a Langendorff isolated-perfused rat heart model [J]. *Biomedical Optics Express*, 2018, 9(10): 4961-4978.
- [3] Liu Z, Pouli D, Alonzo C A, et al. Mapping metabolic changes by noninvasive, multiparametric, high-resolution imaging using endogenous contrast [J]. *Science Advances*, 2018, 4(3): eaap9302.
- [4] Quinn K P, Sridharan G V, Hayden R S, et al. Quantitative metabolic imaging using endogenous fluorescence to detect stem cell differentiation [J]. *Science Reports*, 2013, 3: 3432.
- [5] Dang C V. Links between metabolism and cancer [J]. *Genes & Development*, 2012, 26(9): 877-890.
- [6] 李静, 曾红娟, 庞小峰. 肝癌细胞和正常肝细胞的自体荧光光谱研究 [J]. *光学学报*, 2009, 29(8): 2261-2263.
Li J, Zeng H J, Pang X F. Study of autofluorescence spectrum between hepatic carcinoma cell and hepatic cell [J]. *Acta Optica Sinica*, 2009, 29(8): 2261-2263.
- [7] 李伟风, 樊振林, 张涓瑜, 等. 用于重组蛋白药物生产的 CHO 细胞无血清培养基的研究进展 [J]. *中国细胞生物学学报*, 2021, 43(4): 905-916.
Li W F, Fan Z L, Zhang H Y, et al. Advances of serum-free medium for CHO cells for the production of recombinant protein [J]. *Chinese Journal of Cell Biology*, 2021, 43(4): 905-916.
- [8] Maeda S, Kobori H, Tanigawa M, et al. Methemoglobin reduction by NADH-cytochrome b_5 reductase in *Prosilocerus akamusi* larvae [J]. *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology*, 2015, 185: 54-61.
- [9] Wei T, Zang J, Zheng Y D, et al. Characterization of a novel nicotine hydroxylase from pseudomonas sp. ZZ-5 that catalyzes the conversion of 6-hydroxy-3-succinylpyridine into 2, 5-dihydroxypyridine [J]. *Catalysts*, 2017, 7(9): 257.
- [10] 杨丽丽, 王玉田, 鲁信琼. 三维荧光光谱结合二阶校正法用于石油类污染物的识别和检测 [J]. *中国激光*, 2013, 40(6): 0615002.
Yang L L, Wang Y T, Lu X Q. Identification and measurement of petroleum pollutant by three-dimensional matrix fluorescence with second-order calibration methods [J]. *Chinese Journal of Lasers*, 2013, 40(6): 0615002.
- [11] 朱焯炜, 阙立志, 吴亚敏, 等. 三维荧光光谱结合 PARAFAC 和 GA 对中国白酒品牌的鉴别 [J]. *中国激光*, 2015, 42(6): 0615002.
Zhu Z W, Que L Z, Wu Y M, et al. Identification of Chinese liquors by three-dimensional fluorescence spectra combined with PARAFAC and genetic algorithm [J]. *Chinese Journal of Lasers*, 2015, 42(6): 0615002.
- [12] 梁逸曾, 吴海龙, 俞汝勤. 化学计量学 [M]//分析化学手册. 3 版. 北京: 化学工业出版社, 2016.
Liang Y Z, Wu H L, Yu R Q. *Chemometrics [M]//Handbook of analytical chemistry*. 3rd ed. Beijing: Chemical Industry Press, 2016.
- [13] Bro R. PARAFAC. Tutorial and applications [J]. *Chemometrics and Intelligent Laboratory Systems*, 1997, 38(2): 149-171.
- [14] Bro R. Multiway analysis in the food industry: models algorithms and applications [D]. Amsterdam: University of Amsterdam, 1998: 231-235.
- [15] 王玉田, 边旭, 商凤凯, 等. 三维荧光光谱结合自加权交替三线性分解算法检测农药类混合物 [J]. *光谱学与光谱分析*, 2018, 38(12): 3780-3784.
Wang Y T, Bian X, Shang F K, et al. Three-dimensional fluorescence spectra combined with a self-weighted alternating trilinear decomposition algorithm to detect pesticide mixtures [J]. *Spectroscopy and Spectral Analysis*, 2018, 38(12): 3780-3784.
- [16] 王婷, 郭晋君. 分析化学中检出限的分类及计算方法 [J]. *广东化工*, 2013, 40(17): 183, 189.
Wang T, Guo J J. Classifications and calculation methods of detection limit in analytical chemistry [J]. *Guangdong Chemical Industry*, 2013, 40(17): 183, 189.
- [17] 刘中仕, 荆西平. 荧光光谱的罗丹明 B 校正原理 [J]. *化学通报*, 2005, 68(10): 771-775.
Liu Z S, Jing X P. Principles of corrected fluorescence spectroscopy by rhodamine B [J]. *Chemistry*, 2005, 68(10): 771-775.
- [18] 李哲, 余安澜, 左都罗, 等. 大口径光谱系统狭缝弯曲的分析与矫正 [J]. *光学学报*, 2015, 35(6): 0630004.
Li Z, Yu A L, Zuo D L, et al. Analysis and correction of slit curvature in spectrograph with high NA [J]. *Acta Optica Sinica*, 2015, 35(6): 0630004.
- [19] Murphy K R, Butler K D, Spencer R G, et al. Measurement of dissolved organic matter fluorescence in aquatic environments: an interlaboratory comparison [J]. *Environmental Science & Technology*, 2010, 44(24): 9405-9412.

- [20] Murphy K R, Stedmon C A, Graeber D, et al. Fluorescence spectroscopy and multi-way techniques: PARAFAC[J]. *Analytical Methods*, 2013, 5(23): 6557.
- [21] Yu S H, Zhang Y J, Zhao N J. Outlier detection of time series three-dimensional fluorescence spectroscopy [J]. *Spectroscopy and Spectral Analysis*, 2015, 35(6): 1624-1627.
- [22] Cuss C W, Guéguen C, Andersson P, et al. Advanced residuals analysis for determining the number of PARAFAC components in dissolved organic matter[J]. *Applied Spectroscopy*, 2016, 70(2): 334-346.
- [23] Liu Y J, Postma G, Wu H L, et al. Angle distribution of loading subspace (ADLS) for estimating chemical rank in multivariate analysis: applications in spectroscopy and chromatography[J]. *Talanta*, 2019, 194: 90-97.
- [24] Murphy K R, Stedmon C A, Wenig P, et al. OpenFluor: an online spectral library of auto-fluorescence by organic compounds in the environment [J]. *Anal Methods*, 2014, 6(3): 658-661.
- [25] Bro R, Vidal M. EEMizer: automated modeling of fluorescence EEM data[J]. *Chemometrics and Intelligent Laboratory Systems*, 2011, 106(1): 86-92.
- [26] Ni W G, Qi J Z, Liu L J, et al. A pulse signal preprocessing method based on the Chauvenet criterion [J]. *Computational and Mathematical Methods in Medicine*, 2019, 2019: 1-9.

Utilizing Three-Dimensional Fluorescence Spectra and Parallel Factor Analysis Algorithm to Quantify Concentration of Multiple Metabolic Fluorophores in a Cell Culture Medium

Song Xiaokang^{1,2}, Zhao Qiang^{1,3*}, Zhang Yuanzhi², Wang Yikun²,
Deng Guoqing², Zhu Ling^{2**}

¹ School of Environment and Energy Engineering, Anhui Jianzhu University, Hefei 230601, Anhui, China;

² Anhui Institute of Optics and Fine Mechanics, Hefei Institute of Physical Science, Chinese Academy of Sciences, Hefei 230031, Anhui, China;

³ Anhui Provincial Key Laboratory of Environmental Pollution Control and Resource Reuse, Anhui Jianzhu University, Hefei 230601, Anhui, China

Abstract

Objective Nicotinamide adenine dinucleotide (NADH), flavin adenine dinucleotide (FAD), tryptophan (TRY), vitamin B6 (VB6), etc., are domain endogenous fluorophores for monitoring cell metabolism. Specifically, NADH and FAD are found in eukaryotes and prokaryotes and are involved in the aerobic respiration of many important coenzymes. They are also the intermediate product of the mitochondrial electron transport chain, which is involved in cell material and energy metabolism. TRY is an essential amino acid that functions as a substrate in protein biosynthesis and regulates protein metabolism, thereby influencing protein deposition. TRY is also a regulator of endocrine hormone release and a signal molecule that regulates the activity of the tricarboxylic acid cycle and its downstream target proteins to promote protein synthesis. VB6 is a water-soluble vitamin that degrades slowly under alkaline conditions. It participates in the metabolism of amino acids, carbohydrates, and lipids, as well as the synthesis of hemoglobin. Therefore, qualitative and quantitative analysis of the contents of multiple endogenous fluorophore groups in cell metabolism is critical for assessing cell growth state and developing disease prediction models. *In vitro* cell culture is one of the most direct methods for studying the aforementioned issues as it can study the occurrence and development mechanisms of diseases at the cell level, as well as evaluate the drug resistance and chemotaxis of cells. Cell culture medium is a complex analysis medium. It contains various nutrients (vitamins, inorganic salts, fat, and nucleic acid derivatives), which causes serious interference to the qualitative and quantitative analysis of the fluorophores to be measured, which further becomes the difficulty and key point for the rapid and accurate determination of the content of metabolic fluorophore groups.

Methods To achieve the rapid quantitative analysis of TRY, NADH, FAD, and VB6 in a cell culture medium, this study used a three-dimensional (3D) fluorescence spectrum separation method based on parallel factor analysis. In

terms of sample preparation, the deployment and fluorophore concentrations were determined using an orthogonal experiment table with 6 factors, 5 levels, and 25 tests. Each sample's three-dimensional (3D) excitation-emission contour was scanned using a 3D fluorescence spectrophotometer. Wavelength difference correction, the removal of Rayleigh scattering and Raman scattering peaks, inner filtration effect correction, light intensity standardization, and high leverage point removal were all performed sequentially. The core consistency function was used to determine that the optimal model's factor was 4. Further, the parallel factor analysis model was established to decompose the 3D fluorescence spectrum dataset into two loading matrices and one score matrix. Two loadings could treat as the emission and excitation spectra of many pure chemical components (fluorophores). Finally, the parallel factor analysis model was validated by split-half analysis. Results showed that the emission and excitation spectra of all four pure fluorophores passed S4T6C3 validation and demonstrated a strong capability in predicting the concentrations of fluorophores.

Results and Discussions The optimal excitation/emission wavelengths of TRY, NADH, FAD, and VB6 are 280/350, 350/450, 450/550, and 332/400 nm, as determined by the two loading matrices and one score matrix of the parallel factor analysis model (Fig. 5), which are consistent with the standard spectra in the database. In terms of quantitative analysis, the average recoveries of these four fluorophores are $100.2\% \pm 15.5\%$, $107.4\% \pm 37.1\%$, $100.5\% \pm 6.4\%$, and $99.5\% \pm 6.5\%$. Meanwhile, the root prediction errors (RMSEP) of each fluorophore are 0.124, 43.312, 1.601, and 0.639 $\mu\text{g}/\text{mL}$, and the relative mean deviations are 13.216%, 36.937%, 6.112%, and 6.331%. The limits of detection (LOD) are 0.013, 52.628, 0.003, and 0.012 $\mu\text{g}/\text{mL}$ (Table 3). Nonetheless, TRY's emission spectrum coincides with NADH's excitation spectrum, and FAD's excitation spectrum coincides with NADH's emission spectrum. These two cases indicate that the presence of the fluorescence resonance energy transfer phenomenon limits the model's quantitative analysis capability for NADH concentration, resulting in a large relative average deviation of NADH (Table 3 and Fig. 8).

Conclusions Because of the similar chemical structure of various fluorophore groups and cell culture medium, the spectra heavily overlap, making it difficult to conduct the qualitative and quantitative analysis of single components directly. To begin, we use Rhodamine B and a calibration standard baffle fluorescence spectrophotometer to solve the problem. Moreover, the water Raman unit (RU350) is used for fluorescence intensity normalization, which increases the dynamic detection range of the instrument, and then the parallel factor analysis algorithm, which inherits the advantage of "second order" or "mathematical separation," decomposes the entire dataset into four fluorescent components (TRY, NADH, FAD, and VB6), without the complex "chemical separation." The average recoveries of these fluorophores are in the range of $(99.5\% \pm 6.5\%) - (100.5\% \pm 6.4\%)$, and the detection limits are in the range of 0.003–0.013 $\mu\text{g}/\text{mL}$, indicating that the method reaches the detection level of trace analysis. This method can meet the demand for the quantitative detection of metabolic fluorescent components in a culture medium, indicating a significant advantage in cell metabolism and immune research, medical diagnosis *in vitro*, and cell sorting.

Key words spectroscopy; three-dimensional fluorescence spectra; parallel factor analysis; cell metabolism fluorophores; orthogonal experimental design; inner filter calibration; split-half analysis