

# 主成分分析法在花生四烯酸高产菌株 代谢组学研究中的应用

李若昕<sup>1,2</sup>, 金明杰<sup>2</sup>, 彭超<sup>2</sup>, 潘晨亮<sup>2</sup>, 李霜<sup>2</sup>, 黄和<sup>2</sup>

(1.华东理工大学生物反应器工程国家重点实验室, 上海 200237;

2.南京工业大学制药与生命科学学院, 江苏南京 210009)

**摘要:** 利用主成分分析法对于三株花生四烯酸高产高山被孢霉菌株的 GC-MS 谱图信息进行了分析, 以菌株间含有的 18 种代谢组分为原始指标, 提取出两个主成分变量, 在二元主成分平面图上实现了对三个具有不同性状菌株的良好区分, 同时发现主成分 1 可能为指向花生四烯酸积累的主成分指标, 主成分 2 可能为指向菌体生长的主成分指标。

**关键词:** 主成分分析; 花生四烯酸; 高山被孢霉; 代谢组学; GC-MS 检测

## Application of Principal Component Analysis Method in Metabolomics Analysis of High Arachidonic Acid-producing Strains

LI Ruo-xin<sup>1,2</sup>, JIN Ming-jie<sup>2</sup>, PENG Chao<sup>2</sup>, PAN Chen-liang<sup>2</sup>, LI Shuang<sup>2</sup>, HUANG He<sup>2</sup>

(1. State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China;

2. College of Pharmacy and Life Science, Nanjing University of Technology, Nanjing 210009, China)

**Abstract:** The information on GC-MS spectra of three strains of high arachidonic acid-producing *Mortierella alpina* was analysed by principal component analysis, using 18 common metabolites as the original variables. Two principal components are got and the three strains with different characters are distinguished. At the same time, principal component 1 is found to be corresponded with arachidonic acid accumulation and principal component 2 is corresponded with biomass yield.

**Key words:** principal component analysis; arachidonic acid; *Mortierella alpina*; metabolomics; GC-MS analysis

中图分类号: O657

文献标识码: A

文章编号: 1002-6630(2008)07-0247-04

代谢组学分析<sup>[1]</sup>是对一个生物系统中所有低分子量代谢物质进行全面的定性定量分析。在微生物代谢工程学中应用<sup>[2]</sup>目的是从代谢物水平上分析基因变化, 结合转录组学及蛋白质组学以优化细胞体系, 提升微生物发酵生产能力。花生四烯酸<sup>[3]</sup>(arachidonic acid, 简称 AA, 即 5,8,11,14- 二十碳四烯酸)是人体前列腺素合成的重要前体物质, 在抗血栓、降血脂、抗癌、促进婴幼儿脑组织发育等方面作用显著。虽然微生物发酵法生产花生四烯酸解决了传统化学合成法与生物提取法成本昂贵等问题<sup>[4-5]</sup>, 但是一直面临着生产投入大, 原料利用率低等技术问题。本实验运用 SAS 统计学软件, 对由 GC-MS 检测得到的三株 AA 高产菌株在合成花生四烯酸过程中所获得的各种通含胞内代谢中间产物进行主成

分分析<sup>[6-7]</sup>, 以探索高产菌株间代谢异同, 从而为从代谢组学层次进一步提升花生四烯酸发酵生产能力提供理论和方法依据<sup>[8-9]</sup>。

### 1 材料与方法

#### 1.1 试剂与仪器

四氢呋喃 上海试剂四赫维化工有限公司; 二甲基甲酰胺 国药集团化学试剂有限公司; 所有试剂均为分析纯。

全温摇瓶柜 HYG-A 太仓实验仪器厂; 超低温冰箱 Pieer 公司; GC2000 DSQ 气质联用色谱仪 美国 Thermo finnigan trace 公司。

#### 1.2 样品来源及制备

收稿日期: 2007-09-13

基金项目: 国家自然科学基金项目(20576054)

作者简介: 李若昕(1985-), 男, 硕士研究生, 研究方向为代谢工程。E-mail: ruoxin\_lea@yahoo.com

高山被孢霉菌(*Mortierella alpina*)Y1、Y2、Y3。

胞内代谢组提取方法:快速取样后,分别用0.9% NaCl与60%甲醇(-40℃)清洗,于60%甲醇-80℃冷冻10min,拣出菌球转入装有75%乙醇的冷冻管中,于80℃水浴保温30min,冷却后超声破碎,破碎后的样本真空干燥,用二次蒸馏水溶解、离心分离,取上层清液真空干燥,加入四氢呋喃与衍生化试剂,85℃反应1h,进行GC-MS检测。

### 1.3 分析方法

#### 1.3.1 胞内代谢组 GC-MS 检测条件

GC-MS 条件:DB-5MS 毛细管柱(30m × 0.32mm, 0.25 μm);进样室温度250℃,检测器温度250℃;载气为氦气,载气流速1ml/min;进样量1 μl,未分流进样;程序升温:初始温度为150℃,保持2min,升温3~280℃/min;传输线温度为250℃,离子源温度为200℃;EI源70eV;扫描范围50~600amu。

#### 1.3.2 指标测定

生物量测定采用菌丝体干重法。将摇瓶发酵的培养物抽滤,并用蒸馏水洗涤3次。抽干后,60℃烘干至恒重(含水量在4%以下),称重。

菌体总油脂采用索氏提取法提取油脂中花生四烯酸含量的测定,首先将花生四烯酸甲酯化后用GC-MS检测。

## 2 结果与分析

### 2.1 指标选取

利用GC-MS对三株花生四烯酸高产菌高山被孢霉菌株代谢组定性定量分析,得到了菌株间通含的30种代谢中间产物及其峰面积。以菌株为研究对象,选取通含物质作为指标变量,以峰面积作为定量数据,结果见表1。

### 2.2 主成分选择与建模

调用Statistics模块对原始变量进行预处理,得到指标与对象间基本情况,再运用Principal Components指令,预设主成分为3,运行得到主成分特征根(eigenvalue)及解释方差贡献率(proportion)(表2)。可看出前两个主成分的累积方差贡献率达到100%,即设定两个主成份能够代表整体数据的信息特征。

以图1为主成分设定量,再次运行Principal Components指令进行主成分分析,得到的各原始变量对于各个主成分的因子载荷量的贡献率如表3所示。

表1 变量赋值  
Table 1 Evaluate of variables

指标	菌株		
	Y1	Y2	Y3
V1 前列腺素	90216	166325	40930
V2 胱氨酸	143310	271735	96476
V3 赖氨酸	222709	67023	80512
V4 辛酸	82746	66645	78482
V5 阿糖腺苷	70622	95980	79458
V6 没食子酸	76442	71082	206029
V7 天冬酰胺	77548	92077	11949
V8 蝶啶	123435	58814	78251
V9 葡萄糖	67249	62550	204982
V10 缬氨酸	98577	99294	113009
V11 丁酸	68002	63859	101607
V12 丁醇	182595	76097	106476
V13 甘氨酸	350522	77062	114198
V14 半乳糖	135414	44376	65777
V15 乙酸	79676	210260	187218
V16 磷酸	65342	100248	150495
V17 棕榈酸	103198	400208	271805
V18 硬脂酸	67553	103573	141572

表2 主成分评价  
Table 2 Estimate of principal components

	特征值	方差	相关系数矩阵特征根	
			方差百分比(%)	累计方差百分比(%)
1	11.0168289	4.0336578	0.6120	0.6120
2	6.9831711	6.9831711	0.3880	1.0000
3	0.0000000		0.0000	1.0000

通过贡献率大小可以得到各个主成分因子的指代含义,以及原始变量的指向分类。以贡献率绝对值大于0.25作为判断指向依据。

指向Print1的原始变量有:V1(前列腺素)、V3(赖氨酸)、V8(蝶啶)、V12(丁醇)、V13(甘氨酸)、V14(半乳糖)、V15(乙酸)、V17(棕榈酸)、V18(硬脂酸);指向Print2的原始变量有:V2(胱氨酸)、V6(没食子酸)、V7(天冬酰胺)、V9(葡萄糖)、V10(缬氨酸)、V11(丁酸);指向不明显的原始变量有:V4(辛酸)、V5(阿糖腺苷)、V16(磷酸)。尽管V4对Print1贡献率为0.256830超过0.25但由于与其对Print2贡献率0.221268差别过小,故仍作为指向性不明显变量。

运用SAS软件进行主成分分析所得到的统计量描述矩阵、全部解释方差矩阵、主成分荷载矩阵,经过计

表3 变量对主成分贡献  
Table 3 Contribution of variables to principal components

特征向量	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18
Print1	0.288666	-0.090575	0.301091	0.221268	-0.239103	-0.129894	0.091857	0.291560	-0.131529	-0.150244	-0.111915	0.292806	0.300251	0.296198	-0.299005	-0.236147	-0.277166	-0.251710
Print2	0.108356	-0.360914	0.013428	0.256830	-0.230235	0.341443	-0.360403	0.095351	0.340454	0.328008	0.351342	0.089126	0.031271	0.069219	-0.046427	0.234999	-0.148348	0.207957

表4 主成分得分与主要发酵得率指标汇总  
Table 4 Data of principal component score and fermentation yield

	Print1	Print2	菌体干重(g/L)	油脂含量(%)	AA在油脂中含量(%)	AA产量(g/L)
Y1	246348.870718691229	45897.966636	34.19	55.72	60.01	11.42
Y2	80450.603728	-23383.537515	31.77	54.39	58.97	10.19
Y3	20354.546622	211003.463765	37.68	52.11	50.36	9.88

算,得到原始数据带入系数及标准化后的带入系数。标准化后的综合评价得分方程如下:

$$\text{Print1}=0.288666V_1+0.301091V_3+0.291560V_8+0.292806V_{12}+0.300251V_{13}+0.296198V_{14}-0.299005V_{15}-0.236147V_{16}-0.277166V_{17}-0.251710V_{18}+0.221268V_4-0.239103V_5-0.236147V_{16}$$

$$\text{Print2}=0.360914V_2+0.341443V_6-0.360403V_7+0.340454V_9+0.328008V_{10}+0.351342V_{11}+0.256830V_4-0.230235V_5+0.234999V_{16}$$

### 3 结 论

#### 3.1 数据汇总

将原始变量数值带入主成份数学模型后可得三菌株各自主成份得分,汇总各项发酵得率指标如表4所示。结果表明,三菌株中AA在油脂中的含量在50%~60%之间,AA产量为9~12g/L。证明了三个菌株均为花生四烯高产菌株。

#### 3.2 二元主成分散点图

以Print1为横坐标、Print2为纵坐标绘制三个诱变菌株散点图,结果如图1所示。

由图1可知,两个主成分之间没有明显的相关变化,说明主成分间相关度很底,主成分选取成功。该图展现了不同分析对象的分布情况,通过主成分分析三个诱变菌株得到良好区分。另外由于三个菌株均为花生四烯高产菌株,则三点所在区域即为花生四烯高产区域。

#### 3.3 主成分1与AA积累关系

主成分1得分与AA在油脂中的含量之间的关系见图2~3。图2~3结果表明,AA在油脂中的含量和AA产量都随主成分1得分的升高而升高,可推测不同菌株或不同培养条件下主成分1得分的高低能够在一定程度上反映AA在油脂中含量和AA产量,即指向主成分1的各胞内代谢产物指标(前列腺素、赖氨酸、蝶啶、丁醇、甘氨酸、半乳糖、乙酸、棕榈酸、硬脂酸)的联动变化影响了AA的积累。因此主成分1可能为指向AA积累的主成分指标。

#### 3.4 主成分2与菌体生长关系

主成分2与菌体干重之间的关系见图4。由图4可知,菌体干重随主成分2得分的升高而升高,可推测不

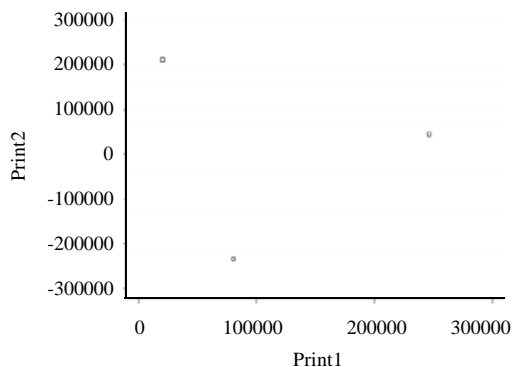


图1 主成分散点图

Fig.1 Scatter diagram of principal components

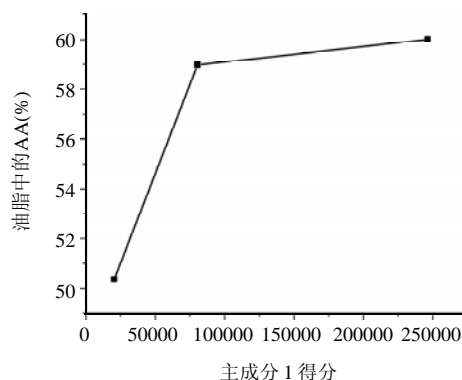


图2 主成分1得分与AA在油脂中含量作图

Fig.2 Plot of principal 1 score to AA content in oil

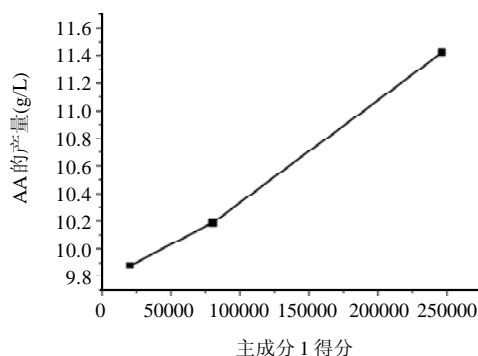


图3 主成分1得分与AA产量作图

Fig.3 Plot of principal 1 score to AA yield

同菌株或不同培养条件下主成分得分的高低能够在一定程度上反映菌体生长的水平,即指向主成分2的各胞内

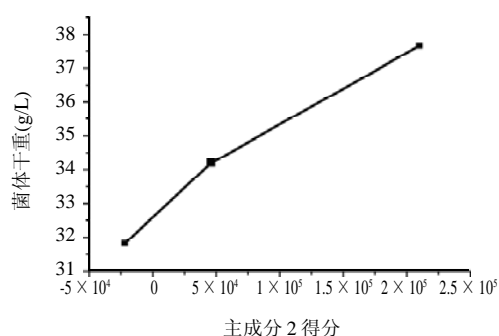


图4 主成分2与菌体干重作图

Fig.4 Plot of principal 2 score to dry biomass yield

代谢产物指标(胱氨酸、没食子酸、天冬酰胺、葡萄糖、缬氨酸和丁酸)的联动变化影响了菌体生长。因此主成分2可能为指向菌体生长的主成分指标。

#### 参考文献:

- [1] TWEEDDAL E L, NOTLEY M, ROBB L, et al. Effect of slow growth on metabolism of *Escherichia coli* as revealed by global metabolite pool analysis[J]. Bacteriol, 1998, 180 (19): 5109-5116.
- [2] URBANCZYK E, LUEDEMANN A, KOPKA J. Parallel analysis of transcript and metabolic profiles: a new approach in systems biology[J]. EMBO Rep, 2003(4): 989-993.
- [3] BENTHOLD K, TAMAS D, HANS D. Arachidonic acid supply and metabolism in human infants born at full term[J]. Lipids, 1996, 31(1): 79-83.
- [4] SINGH A, WARD O P. Production of high yields of arachidonic acid in a fed-batch system by *Mortierella alpina* ATCC 32222[J]. Appl Microbiol Biotechnol, 1997, 48: 1-5.
- [5] HAMANAKA T, HIGASHIYAMA K, FUJIKAWA, et al. Mycelial pellet intrastucture and visualization of mycelia and intracellular lipid in a culture of *Mortierella alpina*[J]. Applied Microbiology and Biotechnology, 2001, 7(5): 321-323.
- [6] 钟科军, 魏万之, 郭方道, 等. GC-MS 及主成分分析法用于咖啡香精的指纹图谱分析和微差样品的识别[J]. 分析实验室, 2006, 25(8): 119-122.
- [7] BUCHHOLZ A, HURLEBAUS J, WANDREY C, et al. Metabolomics quantification of intracellular metabolite dynamics[J]. Biomolecular Engineering, 2002, 19(1): 5-15.
- [8] 刘秀杰, 李柏林, 欧杰, 等. 亚麻酸代谢途径模型中关键节点的分析[J]. 食品科学, 2006, 27(12): 188-190.
- [9] STITT M, FERNIE A R. From measurements of metabolites to metabolomics: an on the fly perspective illustrated by recent studies of carbon-nitrogen interactions[J]. Current Opinion in Biotechnology, 2003, 14 (2): 136-144.