

研究简报

FTIR 技术结合代谢物组学分析方法在 菌株高通量筛选中的应用

陈 伟¹, 于凤鸣^{1,2}, 元英进¹

(¹ 天津大学化工学院, 天津 300072; ² 河北科技师范学院生命科学系, 河北 昌黎 066000)

关键词: 高通量筛选; 利迪链霉菌 AS 4.2501; 生物标志物

中图分类号: TQ 927

文献标识码: A

文章编号: 0438-1157 (2007) 09-2336-05

Application of FTIR and metabolomics analysis in high-throughput screening strains

CHEN Wei¹, YU Fengming^{1,2}, YUAN Yingjin¹

(¹ School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China; ² Department of Life Science, Hebei Normal University of Technology and Science, Changli 066000, Hebei, China)

Abstract: For the sake of searching new methods of high-throughput screening of mutant strains, the strains were screened by applying the technology of metabolomics. Through using Fourier Transformation Infrared spectroscope (FTIR) accompanied with Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA), the experiment successfully distinguished *Streptomyces lydicus* AS 4.2501 strains at different times of fermentation, and found the possible biomarkers when *Streptomyces lydicus* AS 4.2501 synthesized streptolydigin. The experiment, which attempted to use FTIR accompanied with Artificial Neural Network (ANN) to forecast *Streptomyces lydicus* AS 4.2501 mutation strains with different capacities of streptolydigin biosynthesis also obtained satisfying result. The results suggested that the combination of FTIR of metabolomics analysis with the data analysis of stoichiometric methods have prospective future in high-throughput screening of mutant strains.

Key words: high-throughput screening; *Streptomyces lydicus* AS 4.2501; biomarker

引 言

菌种选育在提高微生物产物产量和纯度、改变菌株性状、改善发酵过程、改变合成途径获得新产品等方面具有重要作用, 而筛选是菌种选育过程成败的一个关键步骤^[1-2]。现有菌种筛选方法如随机筛选、理性化筛选等都具有烦琐、耗时、耗资、劳

动强度大等缺点。因此, 寻找快速、方便、高通量的菌种筛选方法是菌种选育亟需解决的问题。

傅里叶变换红外光谱 (FTIR) 能够通过检测化学键中的分子振动和其他运动来测定一个样品的全部组成^[3], 具有快速、不需要试剂、高重现率、便宜等优点^[4]。早在 1974 年 Strauss 报道了 FTIR 可用于新抗生素的筛选^[5]。近年来, FTIR 技术已

经被用于抗生素含量的测定^[6]、菌株分离^[3]、生物医学研究^[7]和毒性分析^[8]等方面。红外光谱仪可以测得菌体在特定时间代谢物的红外吸收光谱图,并且红外光谱包含了样品中很多化学键或功能组分的能量变化,给出了样品组成的几乎所有信息^[9],适于不需要特殊目标代谢产物的指纹分析^[7],但 FTIR 技术很难对所有波数下的吸收峰值进行分析。而多元数据处理中的模式识别方法可以有效地实现数据的降维及主要变量的提取,并且可以通过大量样本的驯化建立专家预测系统,实现对未知样本的分类预测。应用于代谢物组学数据处理的主要模式识别方法包括无监督的模式识别方法 PCA、HCA 等和有监督的模式识别方法 ANN 等。本文旨在探讨 FTIR 技术结合代谢物组学分析方法在高通量筛选菌株中的应用前景。

实验采用的菌株是从土壤分离所得的利迪链霉菌 (*Streptomyces lydicus*) 菌株 AS 4.2501, 它所产生的利迪链菌素^[10-11]具有良好的抗变异链球菌、抗 HIVPR 以及抗 CVB6 的活性,并且它的其他代谢产物也有很强的抗真菌活性^[12],具有重要的应用价值。本文通过 PCA、HCA 方法分析,试图对出发菌株与基因阻断突变株进行区分,并在将不同发酵时间的出发菌株正确分类的基础上,找到引起分类的主要变量,即生物标志物 (biomarker); 以多个抗生素产量不同的利迪链霉菌菌体的红外吸收光谱数据为自变量,以抗生素产量为因变量,用神经网络的方法进行驯化得到专家系统,并利用该系统对大量的诱变样本的利迪链菌素合成能力进行预测,探索实现对利迪链霉菌诱变株进行高通量筛选。

1 实验材料和方法

1.1 实验菌株

利迪链霉菌 AS 4.2501; 利迪链霉菌 S-腺苷蛋氨酸合成酶基因阻断突变株 1; 3-氧酰基-ACP 合成酶 II 基因阻断突变株 2; 紫外照射和硝酸复合诱变突变株 3; 另外紫外照射后得到突变株 105 个。

1.2 生化试剂和仪器

傅里叶变换红外光谱仪 (FTIR8900, Shimadzu), 玛瑙研钵, 红外灯, 压片模具, PCR 仪 (PCR Express), 电热培养箱 (HH. BEE 600)。

1.3 实验方法

1.3.1 种子培养基的配制方法和菌体的准备 每升种子培养基中含: 可溶性淀粉 30 g, 葡萄糖 5 g, 酵母粉 2 g, 蛋白胨 4 g, 磷酸氢二钾 1.5 g, 硫酸镁 0.5 g, 氯化钠 0.5 g, 自然 pH。分装至 250 ml 的三角瓶中, 每瓶 50 ml, 121℃ 灭菌 20 min。

基因阻断突变株和出发菌株在经过种子培养后, 在发酵培养的一定时间取样, 4800 r·min⁻¹ 离心 10 min, 分离菌体和上清液。菌体用 0.5% 的 NaCl 溶液洗涤 2 次, 冷冻干燥后保存在干燥器中用于 FTIR 分析。紫外诱变菌株直接接种到发酵培养基中培养 84 h 后取样, 后处理过程同上。上清液用于抑菌活性的测定。

1.3.2 FTIR 样品制备及分析 取 3 mg 菌体样品加入到 150 mg 的 KBr 中, 在玛瑙研钵中充分研磨, 用模具制备成 KBr 晶片。将制好的 KBr 片放入傅里叶变换红外光谱仪中, 在波数 4000~400 cm⁻¹ 的范围扫描, 记录 20 次扫描结果的平均值。

1.4 数据处理

将原始数据组成 $N \times 518$ 矩阵, N 为样本数, 列为红外光谱吸收波数, 每个样本的观测值为对应波数下的吸收峰丰度。为消除样本之间上样量不同造成的影响, 对数据进行归一化处理后, 再进行化学计量学数据处理, 如下:

(1) 主成分分析 (PCA): 通过用 PCA 得分图中的聚类情况, 判断能否将出发菌株与其基因阻断突变株、同一菌株不同培养时间的菌体进行分离。在 score 图能正确聚类的基础上, 再在投影图中找到对分类贡献率比较大的波数范围, 为 biomarker 的确定提供初步信息。

(2) 等级聚类分析 (HCA): 将出发菌株不同培养时间菌体的红外吸收数据用 HCA 进行分析。输入的是 60×518 的矩阵, 其中 60 个样本分别为培养 12、24、36、48、60、72 h 的菌体各 10 株, 518 个变量为各波数下的红外吸收。根据 HCA 树状图可以判断菌体在不同时间代谢情况的相似性。

(3) 人工神经网络 (ANN): 将产利迪链菌素的利迪链霉菌 AS 4.2501 以及不产利迪链菌素的 3 个利迪链霉菌突变株的红外吸收数据 (各 10 个样本) 进行训练, 用训练后的模型对上述 4 个菌株的另外各 4 个样本进行预测; 并用紫外诱变得到的 105 个菌株中的 48 个进行训练, 用训练后的模型对另外 57 个菌株进行预测, 检验该方法应用于高

通量筛选诱变株的可行性。

2 结果与讨论

2.1 运用 FTIR 技术结合 PCA 进行菌株高通量筛选

对利迪链霉菌及其突变株菌体的 FTIR 数据进行 PCA 分析, 得到 12、24、36、48、60、72 h 点的 score 图。图 1 为 12 h 的 score 图, 从图上可以看出, 利迪链霉菌及其 3 个突变株都得到了很好的区分。其他时间点 (score 图未列出) 也有类似的结果。Bundy 等^[13]在 2005 年基于分析代谢物组的变化较好地地区分了处于稳定期的不同来源的 *Bacillus cereus*, 也有报道^[14]指出利用代谢物组学分析方法也很好地区分了不同单基因缺失的 *Saccharomyces cerevisiae* 菌株。因此基于代谢物组学分析的 FTIR 技术结合 PCA, 有助于将利迪链霉菌分成不同的功能类别。

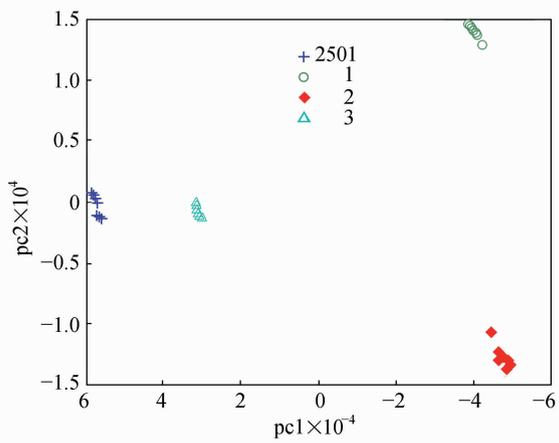


图 1 利迪链霉菌及其突变株菌体红外分析数据的 PCA 得分图

Fig. 1 PCA-derived score plot of FTIR data from *Streptomyces lydicus* AS 4.2501

and its mutant strains at 12 h

2501—AS 4.2501; 1—mutant strain of inactivating 3-oxoacyl-ACP synthase II; 2—mutant strain of inactivating S-adenosylmethionine synthetase; 3—mutant strain of co-treatment by ultraviolet and nitrous acid

在不同的生理条件下, 对同一菌株的代谢物组学分析可能得到不同的结果。基于这种推测, 利用 PCA 方法分析了同一菌株不同发酵时间的 FTIR 数据, 以区分处于不同发酵时间的菌体, 获得了满意的结果。图 2 是对 AS 4.2501 原始菌株的 score 图, 处于不同发酵时间的 AS 4.2501 得到了很好

的区分。3 个突变株处于不同发酵时间的菌体也分别得到了很好的区分 (图未列出)。随着发育进程, 生物体的代谢物积累也发生改变, 而红外光谱足以检测到这种变化。在所有的图谱中 10 个重复的样本非常紧密地聚集在一起, 表明这个方法具有良好的重现性。因此, 运用 FTIR 技术结合 PCA 不但可以区分不同基因型的菌株^[15], 而且可以区分同一个菌株的不同发育阶段。

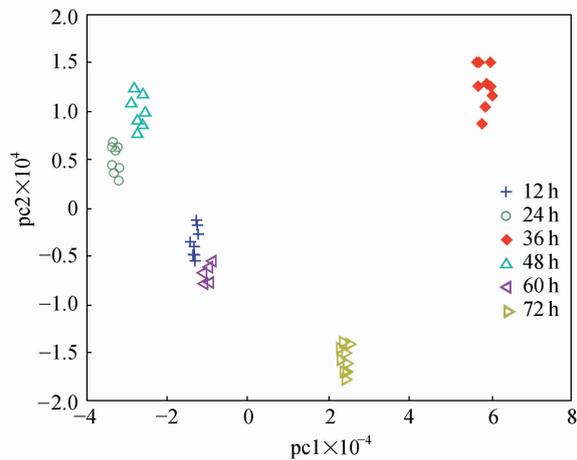


图 2 利迪链霉菌 AS 4.2501 不同发酵时间菌体红外数据的 PCA 得分图

Fig. 2 PCA-derived score plot of FTIR data from *Streptomyces lydicus* AS 4.2501

2.2 利迪链霉菌合成利迪链菌素生物标志物的筛选

本实验室对利迪链霉菌的研究结果表明, 经过 48 h 的种子培养, 再转入发酵培养基中培养 36 h 后, 在发酵液中才能检测到利迪链菌素。以后随着发酵时间的延长, 利迪链菌素的含量开始急剧上升, 发酵 72 h 后上升速度降低。等级聚类分析的结果 (图 3) 同样显示发酵 12 h 和 24 h 的菌体与发酵 36 h 的菌体确实存在很明显的差异, 但是差异随着发酵时间的延长逐渐变小。由图 2 可知, PCA 可以很好地区分不同发酵时间的利迪链霉菌 AS 4.2501 菌体。因此, 在用 PCA 得分图能正确分类的基础上, 用 PCA 投影图寻找 biomarker 具有可行性。对发酵 12 h 对 36 h 以及发酵 24 h 对 36 h 的红外数据进行 PCA 分析得到投影图 (图 4)。由图上可得, 795~810 cm^{-1} 和 1782~1797 cm^{-1} 波数范围是利迪链霉菌合成利迪链菌素的特征波长, 有可能代为利迪链霉菌合成利迪链菌素的 biomarker。795~810 cm^{-1} 属于指纹区的范围, 是 $\text{R}_1\text{R}_2\text{C}=\text{CHR}_3$ 面外弯曲振动吸收的位置,

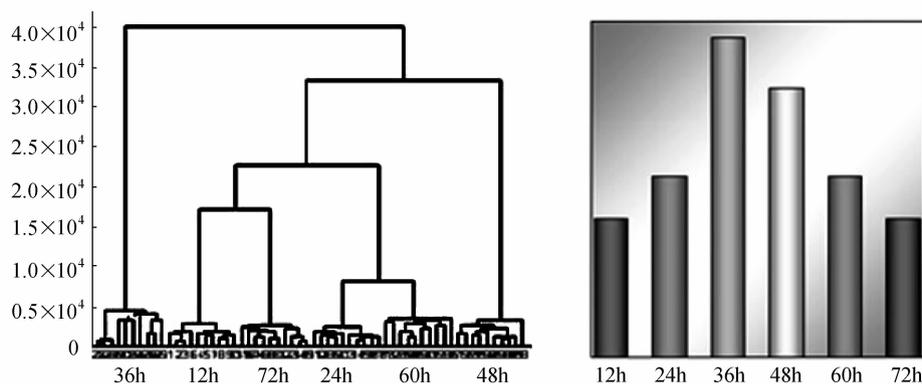
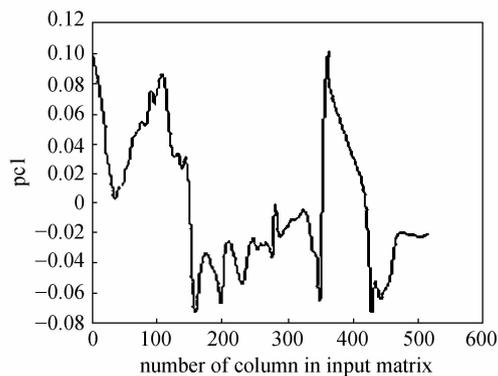


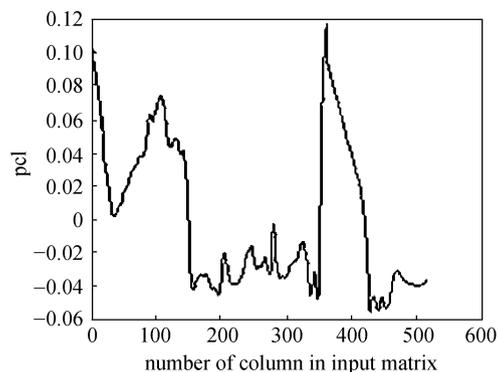
图 3 不同发酵时间利迪链霉菌 AS 4.2501 菌体红外数据的等级聚类分析结果

Fig. 3 HCA-derived plot of FTIR data from *Streptomyces lydicus* AS 4.2501 at different fermentation time

此结构在利迪链菌素的侧链部分存在; 而 $1782 \sim 1797 \text{ cm}^{-1}$ 属于羰基的特征吸收范围, 与利迪链菌素的结构中含有较多的 $\text{C}=\text{O}$ 吻合。基于 PCA 投影图寻找 biomarker 将有利于利迪链菌素高产菌株的高通量筛选。



(a)



(b)

图 4 利迪链霉菌发酵 12 h 和 24 h 对 36 h 的菌体红外数据的 PCA 投影图

Fig. 4 PCA-derived loading plot of FTIR data from *Streptomyces lydicus* AS 4.2501

(a) 12 h vs 36 h; (b) 24 h vs 36 h

2.3 应用 ANN 预测利迪链霉菌突变株的抗生素生产能力

人工神经网络能够以一定数量的生物样本作为训练样本建立数学模型, 并具有强大的泛化能力, 从而对未知样本进行预测。本实验用 3 层的 feed-forward back propagation 神经网络 (节点数分别设为 5、10 和 1, 能产利迪链菌素的设 $T=1$, 不能产利迪链菌素的设 $T=0$), 对能产利迪链菌素的利迪链霉菌 AS 4.2501 以及不能产利迪链菌素的 3 个利迪链霉菌突变株各 10 个样本的红外数据进行训练, 并用训练后的模型对上述 4 个菌株的另外各 4 个样本进行预测, 取得了令人满意的结果, 如表 1 所示。

上述结果表明, 利用红外光谱技术结合人工神经网络分析, 有可能对利迪链霉菌的诱变菌株实现高通量的筛选。因此, 随机选取了 105 个紫外诱变单菌落, 进行抑菌活性的测定, 活性的大小用抑菌圈的直径表示。构建神经网络时, 目标值的设定为: 具有最大抑菌活性的 $T=1$, 不产利迪链菌素的 $T=0$, 其他的 $T=$ 实际的抑菌圈直径/最大的抑菌圈直径。随机用 48 个样本进行人工神经网络训练, 对剩余的 57 个样本进行预测分析。结果表明, 实际测定值与模拟值正负误差在 20% 之内的达到 45 个, 占预测总数的 78.9%, 再加上 5 个不产利迪链菌素突变株的预测值均低于 0.5, 预测的基本准确率达到 87.7%。故应用 ANN 预测利迪链霉菌突变株的抗生素生产能力具有一定的可行性。

3 结 论

通过 FTIR 技术与 PCA 的结合, 有效地区分

表 1 人工神经网络预测的结果与实际结果的比较

Table 1 Comparison of simulated and actual results of tested strains

| Actual results of strains | Time of fermentation/h | Simulated results of strains | | | |
|---------------------------|------------------------|------------------------------|-------------|--------------------------|-------------------------|
| yield streptolydigin | 12 | 0.0059061 | -0.00087173 | -0.0015391 | 0.0028302 |
| | 24 | -0.0032558 | 0.01236 | -0.01224 | 0.0093315 |
| | 36 | -0.0033051 | -0.00075476 | -0.0062886 | -0.0026563 |
| | 48 | -0.0014377 | 0.0097748 | 0.0019943 | 0.0013097 |
| | 60 | 1.3329×10^{-6} | -0.00014956 | -3.0286×10^{-5} | 5.9376×10^{-6} |
| | 72 | 0.028211 | -0.2315 | 0.045933 | 0.049561 |
| not yield streptolydigin | 12 | 0.99962 | 0.99996 | 0.99941 | 0.99926 |
| | 24 | 0.88594 | 0.88603 | 0.88877 | 0.99926 |
| | 36 | 0.99998 | 0.99999 | 0.99863 | 0.99306 |
| | 48 | 0.99996 | 0.99984 | 0.99998 | 0.99999 |
| | 60 | 0.99998 | 0.99999 | 0.99999 | 0.99999 |
| | 72 | 0.99994 | 0.99995 | 0.99996 | 0.9999 |

了不同基因型利迪链霉菌菌株、处于不同发酵时间的利迪链霉菌菌株；利用 FTIR 技术与 HCA 的结合筛选利迪链霉菌合成利迪链菌素的可能生物标志物；应用 ANN 预测了利迪链霉菌突变株的抗生素生产能力。因此随着设备灵敏度的提高，以及高通量设备的应用，FTIR 技术与 PCA、HCA、ANN 技术及其他代谢物组学方法的结合使用，在菌种诱变后的高通量筛选中将具有广阔的应用前景。

References

- [1] Dai Sifa, Li Guan hong, Wu Shijin. Research advances in the breeding of modern industrial microbe. *J. of Microbiology* (微生物学报), 2000, **20** (2): 48-51
- [2] Lei Zhaozu, Qian Zhiliang, Zhang Jian. Comments on industrial strains improvement. *Industrial Microbiology* (工业微生物), 2004, **34** (1): 39-51
- [3] Kaderbhai N N, Broadhurst D I, Ellis D I, et al. Functional genomics via metabolic footprinting: monitoring metabolite secretion by *Escherichia coli* tryptophan metabolism mutants using FT-IR and direct injection electrospray mass spectrometry. *Comp. Func. Genomics*, 2003, **4**: 376-391
- [4] Fiehn O. Combining genomics, metabolome analysis, and biochemical modeling to understand metabolic network. *Comp. Func. Genomics*, 2001, **2**: 155-168
- [5] Strauss D G. A fast comparison of IR-spectra for screening new antibiotics. *J. Antibiot (Tokyo)*, 1974, **27** (10): 805-808
- [6] Meanwell R J L, Shama G. Direct FTIR assay of streptomycin in agar. *Biotechnol. Lett.*, 2005, **27** (20): 1629-1631
- [7] Naumann A M. FT-infrared and FT-Raman spectroscopy in biomedical research. *Appl. Spectrosc. Rev.*, 2001, **36**:

239-298

- [8] Harrigan G G, LaPlante R H, Cosma G N, et al. Application of high-throughput Fourier-transform infrared spectroscopy in toxicology studies: contribution to a study on the development of an animal model for idiosyncratic toxicity. *Toxicol. Lett.*, 2004, **146** (3): 197-205
- [9] Peric-Concha N, Long P F. Mining the microbial metabolome: a new frontier for natural product lead discovery. *Drug Discov. Today*, 2003, **8** (23): 1078-1084
- [10] Zhang Jianyong, Niu Jinyang, Chen Guibin, Yuan Yingjin, Hu Zongding. Mutagenicity of *Streptomyces regensis* producing a novel antibiotics AGPM. *Microbiology* (微生物学通报), 2003, **30** (2): 24-28
- [11] Shi Bingxing, Zhao Hong, Liu Xipeng, Yuan Yingjin, Hu Zongding. Improvement of the fermentation productivity of a new antibiotic AGPM by orthogonal design experiment. *Process Engineering* (过程工程学报), 2001, **4**: 442-444
- [12] Ji Shengbao, Fan Jinyong, Yuan Yingjin. Isolation and purification of the antifungal antibiotic from the fermentation broth of *Streptomyces luteogriseus* H-103. *Microbiology* (微生物学通报), 2005, **32** (3): 77-81
- [13] Bundy J G, Willey T L, Castell R S, et al. Discrimination of pathogenic clinical isolates and laboratory strains of *Bacillus cereus* by NMR-based metabolomic profiling. *FEMS Microbiol. Lett.*, 2005, **242** (1): 127-136
- [14] Li W, Tan H. Structure and function of *sanV*: a gene involved in Nikkomycin biosynthesis of *Streptomyces ansochromogenes*. *Current Microbiology*, 2003, **46** (6): 403-407
- [15] Park S J, Lee S Y, Cho J, et al. Global physiological understanding and metabolic engineering of microorganisms based on omics studies. *Appl. Microbiol. Biotechnol.*, 2005, **68** (5): 567-579