

近红外光谱法在蛹虫草诱变筛选及其发酵条件优化中应用

郭伟良¹, 张卓勇², 逯家辉¹, 宋佳¹, 王迪¹, 滕利荣^{1*}

1. 吉林大学生命科学学院, 吉林 长春 130012

2. 首都师范大学化学系, 北京 100048

摘要 应用近红外(NIR)光谱技术结合化学计量学法建立测定蛹虫草菌丝体中腺苷、蛋白、多糖和虫草酸4种主要有效成分含量的定量分析模型,可应用于蛹虫草高产突变株的筛选和发酵条件的优化。468株蛹虫草突变株在不同的摇瓶发酵条件下进行发酵,收集菌丝体粉末样品并采集NIR光谱,同时采用常规方法分别测定样品的4种有效成分含量。偏最小二乘法(PLS)用于建立NIR光谱与有效成分含量间相关模型。采用蒙特卡罗偏最小二乘法(MCPLS)识别异常样本和选择合适的校正集样本数;采用移动窗口偏最小二乘法(MWPLS)筛选波长变量;以逼近度(Da)为考察指标,选择最有效的光谱预处理方法和最适的PLS隐变量数,最终得到测定蛹虫草菌丝体中腺苷、蛋白、多糖和虫草酸含量的最优PLS模型,其校正集样本参考值与预测值间相关系数(R_c)分别为0.929 43, 0.984 79, 0.907 85和0.851 31;预测集误差(RMSEP)为0.667 14, 0.020 65, 0.011 31和0.011 59,表明模型具有很好的拟合度和预测性能,应用该方法进行蛹虫草诱变筛选和发酵条件的优化是可行的。

关键词 近红外(NIR)光谱;蒙特卡罗偏最小二乘法(MCPLS);移动窗口偏最小二乘法(MWPLS);蛹虫草
中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2010)08-2077-06

引言

蛹虫草(*Cordyceps militaris* (L.) Link)又名北虫草、北冬虫夏草,其子座在春至秋季在半埋于林地上或腐枝落叶层下的鳞翅目昆虫蛹上长出。在我国各省均有分布,是中国传统珍贵中药材,许多学者对蛹虫草的药理作用进行大量研究,确认蛹虫草的药理作用基本上与冬虫夏草相同,可以作为冬虫夏草的代用品^[1-3]。蛹虫草含有30多种元素,主要活性物质包括腺苷、蛋白、多糖和虫草酸等。目前蛹虫草生产主要以固体栽培为主,液体深层发酵研究的时间不长,蛹虫草发酵普遍存在生产成本低,周期长,产量低,产品质量差等问题。而发酵水平的提高主要从两方面入手:一是通过菌种改良方法,二是发酵工艺的优化。然而在大量的菌种筛选工作和发酵工艺优化过程中,采用常规的化学分析方法逐个有效成分含量的测定非常费时,费力,费资金^[4-6]。

近红外(near infrared, NIR)光谱技术,具有分析速度快、效率高、成本低、可多组分同时测定和直接对样品进行无损检测等优点^[7-10],非常适合蛹虫草高产突变株筛选和发酵条件优化。但是由于NIR光谱中有效信息率低,对复杂样

品进行NIR光谱分析需从复杂、重叠、变动的光谱中提取微弱信息,非常困难^[11]。偏最小二乘法(partial least square, PLS)由于具有较强的提供信息的能力而成为化学计量学中最受推崇的多变量校正方法,目前是最有效地分析方法之一,它从自变量矩阵和因变量矩阵中提取偏最小二成分,有效地降维,并消除自变量间可能存在的复共线关系,明显改善数据结果的可靠性和准确度^[12]。然而在建立间接测量数学模型时,模型的稳健性和泛化能力一直是个瓶颈问题,选择具有代表性样本作为建立模型的样本可改善所建立模型的稳健性,合理地选择校正集样本数量和预测集样本数量可有效提升模型的泛化能力,防止模型出现过拟合现象。蒙特卡罗偏最小二乘法(MCPLS)通过大量的随机计算和评估,能有效的识别模型中异常的样本和确定合理的校正集样本数和预测集样本数^[13, 14],能较好地改善模型的稳健性和提升模型的泛化能力。近红外光谱的本质是物质的诸如振动和转动能级的反映,相邻的波长通道一般具有高的相关性,如果一个波长通道能够有资格用来建立校正模型,那么以其为中心的一个区域也应该可以用来建立校正模型,如果一个波长通道受到与目标物无关的性质的干扰,那么以其为中心的区域也应该受到与目标物无关的性质的干扰,基于这一特性,可

收稿日期: 2009-09-26, 修订日期: 2009-12-28

基金项目: 中国医学基金会新药发展基金项目(20061108)资助

作者简介: 郭伟良, 1983年生, 吉林大学生命科学学院博士生 * 通讯联系人 e-mail: tenglr@jlu.edu.cn

©1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

采用移动窗口偏最小二乘法(MWPLS)有效筛选测定组分的特征波长变量^[15, 16]。

本文应用近红外光谱法结合偏最小二乘法(NIR-PLS)建立测定蛹虫草菌丝体中腺苷、蛋白、多糖和虫草酸的含量定量分析模型,采用MCPLS法识别异常样本和确定校正集样本数和预测集样本数,采用MWPLS法筛选波长变量等方法的优化,得到稳健性和预测性能都比较好的定量分析模型。该方法可应用于蛹虫草高产突变株的筛选和发酵条件的优化。

1 实验

1.1 实验材料

Table 1 Statistical adenosine, protein, polysaccharides and *Cordyceps* *militaris* acid contents of the *Cordyceps militaris*

Components	Samples numbers	Average/(g · g ⁻¹)	Ranges/(g · g ⁻¹)
Adenosine	531	2 873 81(× 10 ⁻³)	0 268 57~ 7 254 91(× 10 ⁻³)
Protein	507	0 523 70	0 142 51~ 0 768 62
Polysaccharides	531	0 068 51	0 030 54~ 0 212 46
<i>Cordyceps militaris</i> acid	518	0 083 25	0 010 80~ 0 155 65

1.3 仪器和样品 NIR 光谱的采集

使用 UV-3150 紫外可见近红外分光光度计(岛津公司,日本)和 ISR-3100 积分球附件(岛津公司,日本)扫描样品的 NIR 光谱,扫描波长范围为 800~ 2498 nm,光谱通带密度为 12 nm,以 BaSO₄ 标准品为空白,光谱采样间隔为 3 nm,每个样品扫描 3 次光谱取其平均光谱为样品光谱。

1.4 分析方法

(1) 根据 2005 版《中国药典》中介绍的方法,利用 HPLC 法测定上清中腺苷的含量^[16]。

(2) 采用凯氏定氮法测定样品中蛋白含量^[17]。

(3) 采用蒽酮硫酸法测定胞内多糖含量^[18]。

(4) 利用比色法测定上清液中虫草酸含量^[19]。

1.5 PLS 模型的建立

以下程序编写均采用 Matlab(Mathwork 公司,美国)。

1.5.1 异常样本的剔除

采用 MCPLS 法识别模型中异常样本,其基本程序如下:

(1) 以交互验证均方根误差(RMSECV)为评价指标,采用留一交互验证法初步确定模型的隐变量数。

(2) 随机选出 50% 的样本作为建立模型的校正集样本,剩余样本作为预测集样本,计算每个预测集样本的预测残差(PRE)。

(3) 重复(2) n_0 次,在此过程中, n_0 需要足够大,以保证每个样品都有机会多次被选中作为预测集样本。 n_0 越大,结果就越稳定可靠。本文考察了 n_0 为 10 000 和 100 000 时结果基本一致,为了保证结果的可靠性,本文选择 n_0 为 100 000。经过此随机过程,每个样本将获得一组 PRE,计算它们的平均值(MPRE)和标准偏差(SDPRE)。

甘露醇标准品、腺苷标准品(美国 Sigma 公司);甲醇(北京化工色谱);3,5-二硝基水杨酸、蒽酮、葡萄糖、亚硝基胍等均为分析纯;蛋白胍、酵母粉等均为生化试剂。

1.2 样品的制备

以蛹虫草菌株 CGMCC5.699(中国科学院微生物研究所)为出发菌株,采用亚硝基胍对其进行诱变,获得 468 株突变株。采用 250 mL 的摇瓶进行液体深层发酵,接种量为 3% (v/v),装液量为 80~ 150 mL,摇床转速为 100~ 150 r · min⁻¹,发酵温度为 23~ 26 °C,发酵时间为 2~ 5 d,发酵结束后,5 000 r · min⁻¹ 离心 5 min 收集菌丝体,冷冻干燥,粉碎过 60 目筛制备供试样品。样品的腺苷、蛋白、多糖和虫草酸含量统计结果如表 1 所示。

(4) 绘制 MPRE 对 SDPRE 的散点图,MPRE 和 SDPRE 高的样本为异常样本。

1.5.2 校正集样本数的选择

为了防止模型出现过拟合现象和具有很好的泛化能力,校正集样本数与预测集样本数的选择很关键。本文采用 MC-PLS 法选择合适的校正集样本数和校正集样本数配比。其基本程序如下:

(1) 分别随机选择 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% 和 90% 的样本作为校正集样本,剩余样本作为预测集样本,建立 PLS 模型并计算其拟合度(D_f),其计算公式如下:

$$D_f = \frac{c}{\frac{RMSEC}{n_c} + \frac{RMSEP}{n_p} + |RMSEC - RMSEP|} \quad (1)$$

式中 n_c 和 n_p 分别是校正集样本数和预测集样本数, RMSEC 和 RMSEP 分别为校正均方根误差和预测均方根误差, c 是根据作图需要而设定的常数,本文测定腺苷、蛋白、多糖和虫草酸含量 PLS 模型的 c 值分别为 0.5, 0.02, 0.01 和 0.01。

(2) 重复(1) n_r 次, n_r 需要足够大以保证结果的稳定可靠,本文 n_r 采用 50 000,经过 n_r 次随机过程,可得到不同校正集样本数所建立 PLS 模型的一组 D_f 值,计算它们平均值(MD_f), MD_f 值最高时,校正集样本数最合适。

1.5.3 光谱预处理方法,波长变量和隐变量数的选择

由于采集的近红外原始光谱含有杂噪音和基线漂移的干扰,分别考察了采用卷积平滑(savitzky-golay smoothing)、快速傅里叶变换(FFT)、一阶导数、二阶导数、标准正态转换(SNV)和小波变换方法对光谱进行预处理时除噪效果,其中在进行卷积平滑、FFT、一阶导数和二阶导数预处理时,

考察了处理窗口为 10、20 和 30 nm 时对除噪效果的影响,在进行小波变换时,分别采用 db1、db2 和 db3 的小波基分解,分解水平为 2、3 和 4 时对除噪效果的影响。在采用原始光谱和预处理后的光谱进行 PLS 建模时,均采用 MWPLS 法选择各组分的特征波长变量,其基本程序如下:

(1) 将相邻波长的 W 个光谱波长变量划入一个窗口(W 为窗口大小),对处于该窗口的波长变量采用不同的 PLS 的隐变量数建立 PLS 模型,选取最佳 PLS 隐变量数,得到该窗口的最优模型。

(2) 去掉窗口内最前端的一个波长变量,并在紧邻窗口后的位置上增加一个波长变量,计算新窗口的光谱区域的 PLS 最优模型。

(3) 重复步骤(2),直到最后一个数据点进入窗口中。

(4) 比较各个窗口的 PLS 模型的逼近度(Da)值,选择 Da 值最高的前 n_w 个窗口波长变量组合进行建模, Da 计算公式如式(2)。

$$Da = \frac{c}{\frac{n_v}{n_1} \times RMSEC + \frac{n_p}{n_1} RMSEP + |RMSEC - RMSEP|}$$

(2)

式中 n_1 为总样品数, c 同式(1)。

在此程序中, W 的大小和 n_w 对波长变量的筛选最为关键,本文分别考察 W 为 20~100, 每间隔 10 取一个值时对 Da 值的影响,同时考察了 n_w 为 20~200, 每间隔 20 取一个值时对 Da 值的影响, Da 值最高时, W 和 n_w 值最合适。

1. 5. 4 最优 PLS 模型的建立

模型经过异常样本的剔除、选择合适的校正集样本数、选择光谱预处理方法、选择有效波长变量和选择合适隐变量数等建模参数的优化,建立测定蛹虫草菌丝体中腺苷、蛋

白、多糖和甘露醇含量最优 PLS 模型。

2 结果与讨论

2 1 样品的近红外光谱

图 1 是所有蛹虫草菌丝体粉末样品的 NIR 光谱,由图可以看出谱线复杂,各组分的特征信息难以分辨和解析,简单的常规光谱分析方法难以进行定量分析,本文采用 PLS 法建立 NIR 光谱与有效成分含量间的相关模型。

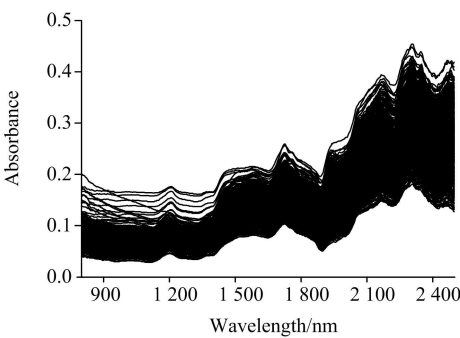


Fig 1 NIR spectra of the samples

2 2 异常样本的剔除

按照 1 5 1 的方法,采用 MCPLS 方法识别异常样本,结果如图 2 所示,由图可以看出,用于建立测定腺苷含量 PLS 模型的 531 个样本中,存在 8 个样本偏离主体样本,具有较高的 SDPRE 和 MPRE,因此这些样品被识别为异常样本。而用于建立测定蛋白、多糖和虫草酸含量 PLS 模型的样本中,分别存在 15、25 和 23 个异常样本,这些样本在之后的建模过程中将被剔除。

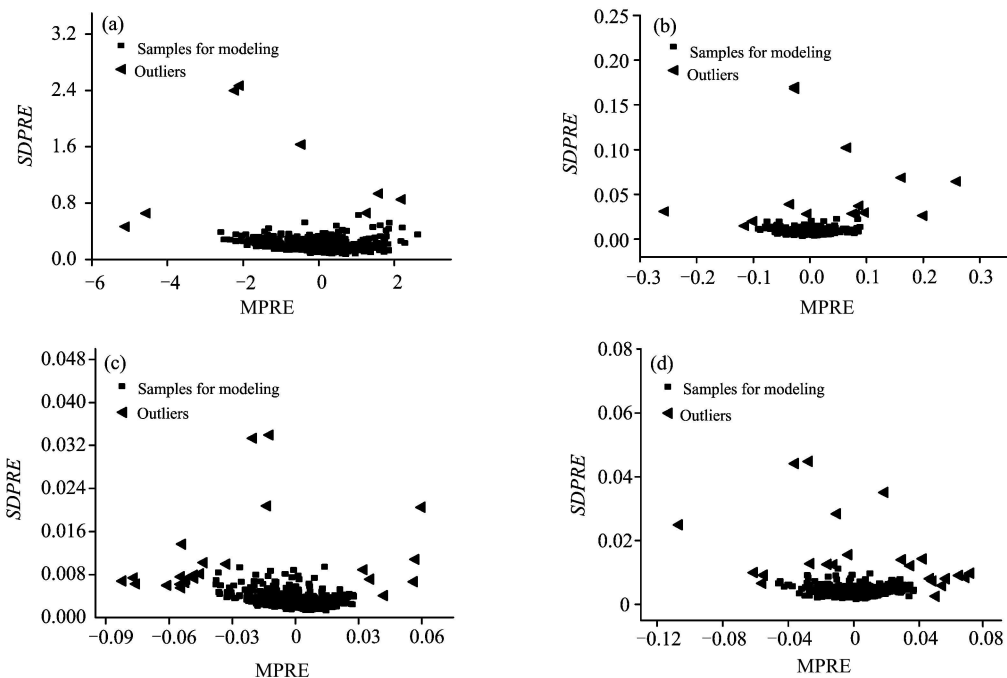


Fig 2 Scatter plot of MPRE vs SDPRE

(a): Adenosine model; (b): Protein model; (c): Polysaccharide model; (d): *Cordyceps militaris* acid model

2.3 校正集样本数的选择

按照 1.5.2 方法考察选择不同的校正集样本与总样本比例对 MD_f 的影响, 结果如图 3 所示, 由图可以看出, 在建立测定腺苷、蛋白、多糖和虫草酸含量的模型时, 校正集样本数分别为总样本数的 60%, 60%, 60% 和 70% 时, MD_f 最

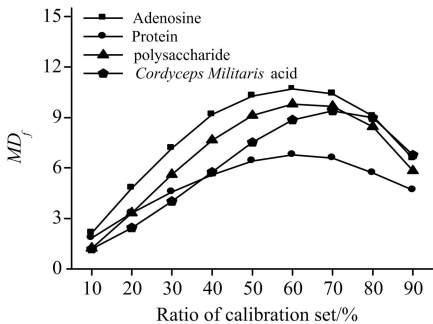


Fig 3 Effect of the ratio of calibration samples on MD_f

高, 即所选择校正集样本数最为合理。
2.4 光谱预处理方法和波长变量的筛选

按照 1.5.3 方法考察预处理方法对 D_a 的影响, 以筛选有效的光谱预处理方法, 同时采用 MWPLS 方法筛选有效成分的特征光谱波长变量, 结果如表 2 所示。在采用原始光谱建立 PLS 模型时, 除了建立测定蛋白质含量 PLS 模型时, 采用原始光谱全光谱建立模型的 D_a 值较 MWPLS 法选择的波长变量建立模型的 D_a 值高外, 其他 PLS 模型均是采用 MWPLS 法选择的波长变量建立模型的 D_a 较高, 表明 MWPLS 法选择波长变量很有效, 同时大大减小输入数据矩阵, 从而大大减少计算量和计算时间。建立测定腺苷、蛋白、多糖和虫草酸含量 PLS 模型时, 最有效的光谱预处理方法分别为 SNV、二阶导数、FFT 和卷积平滑光谱; 在进行 MWPLS 波长变量筛选时, 最合适的 W 分别为 60, 70, 20 和 60; 最优的 n_w 值分别为 180, 140, 200 和 180; 最合适的 PLS 隐变量数分别为 11, 11, 10 和 11。

Table 2 Results of using various preprocessing methods and wavelength variables

Components	Preprocessing methods	Windows	W	n_w	RMSEC	RMSEP	D_a	n'_{LV}
Adenosine	Original spectra				0.570 80	0.690 32	0.675 43	11
	Original spectra		100	200	0.601 07	0.677 28	0.705 08	11
	Savitzky-Golay smoothing	20	80	180	0.661 95	0.696 32	0.703 55	13
	FFT	10	100	180	0.632 90	0.708 02	0.676 21	15
	First order derivative	20	100	120	0.613 59	0.697 03	0.683 15	14
	Second order derivative	30	100	60	0.697 68	0.762 02	0.633 79	14
	Standard normal variate		60	180	0.589 43	0.667 14	0.714 68	11
	Wavelet db2 level	2	80	180	0.660 24	0.689 79	0.712 11	13
Protein	Original spectra				0.018 80	0.022 06	0.878 82	10
	Original spectra		40	200	0.016 92	0.022 05	0.837 63	11
	Savitzky-Golay smoothing	10	90	120	0.018 11	0.021 29	0.886 28	13
	FFT	10	90	60	0.019 12	0.021 87	0.870 49	15
	First order derivative	10	70	160	0.020 38	0.020 65	0.963 30	7
	Second order derivative	30	70	140	0.020 58	0.020 65	0.967 46	11
	Standard normal variate		70	160	0.017 34	0.021 37	0.870 12	11
	Wavelet db1 level	2	90	100	0.018 47	0.021 46	0.882 64	13
Polysaccharide	Original spectra				0.010 94	0.011 72	0.831 05	8
	Original spectra		30	140	0.010 71	0.011 30	0.867 32	9
	Savitzky-Golay smoothing	20	20	100	0.011 52	0.011 52	0.867 62	9
	FFT	10	20	200	0.011 02	0.011 31	0.874 74	10
	First order derivative	30	70	40	0.011 47	0.011 75	0.842 66	6
	Second order derivative	30	40	160	0.011 55	0.011 69	0.856 85	5
	Standard normal variate		30	120	0.010 93	0.011 37	0.866 57	8
	Wavelet db3 level	2	20	140	0.011 30	0.011 43	0.871 94	8
Cordyceps militaris acid	Original spectra				0.012 21	0.012 16	0.816 55	8
	Original spectra		30	180	0.010 85	0.011 72	0.835 05	11
	Savitzky-Golay smoothing	10	60	180	0.011 52	0.011 45	0.860 87	11
	FFT	10	70	140	0.011 79	0.011 83	0.844 31	11
	First order derivative	10	100	80	0.011 80	0.011 84	0.843 33	8
	Second order derivative	20	100	40	0.012 39	0.012 49	0.798 94	10
	Standard normal variate		20	40	0.016 28	0.018 65	0.516 66	5
	Wavelet db2 level	2	30	200	0.011 61	0.011 66	0.857 01	11

n'_{LV} : The number of latent variables

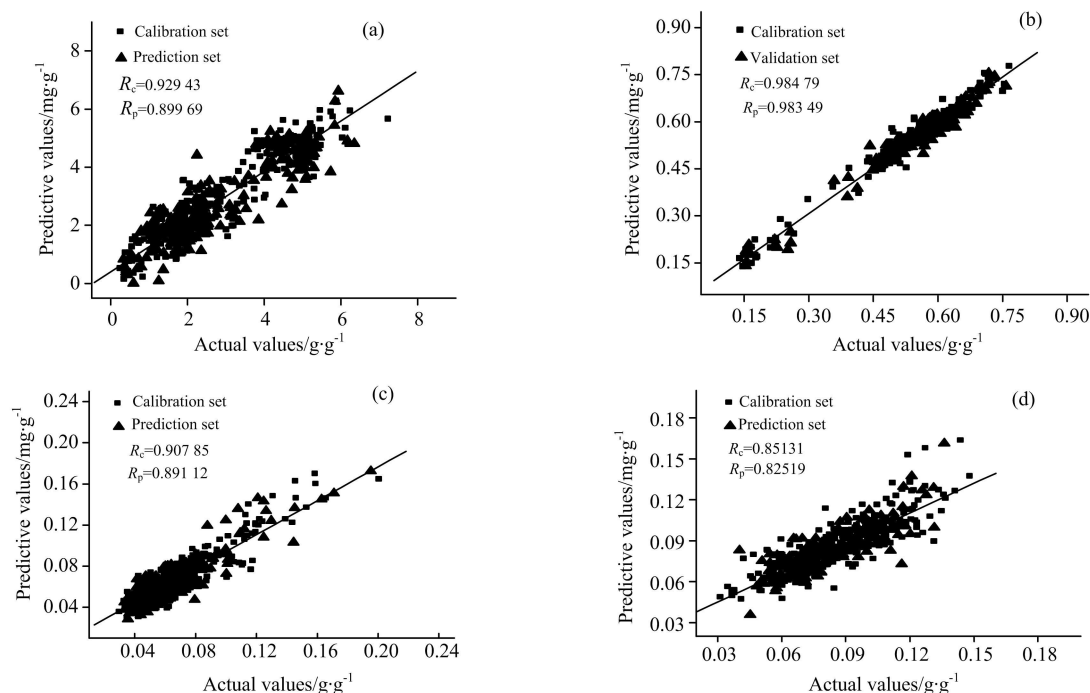


Fig 4 Correlation between actual values and NIR predicted values obtained by the optimum models

(a): Adenosine model; (b): Protein model; (c): Polysaccharide model; (d): *Cordyceps militaris* acid model

2.5 最优模型的建立

经过上述优化,采用最有效的光谱预处理方法对原始近红外光谱进行预处理,然后采用 MWPLS 选择各组分的特征波长变量,选择最合适的 PLS 隐变量数,建立测定蛹虫草菌丝体中腺苷、蛋白、多糖和虫草酸含量的最优 PLS 模型,采用所建立的最优 PLS 模型对所有样品中的腺苷、蛋白、多糖和虫草酸含量进行预测,预测值与化学测量值间的相关性如图 4 所示。这些模型的拟合度和预测精度均比较好,尤其是测定蛋白含量的 PLS 模型, R_c 达到 0.984 79。这些模型的预测精度完全可以满足其在蛹虫草高产突变株的筛选和发酵条件优化中的应用。

3 结 论

本文采用不同的发酵条件下对不同的蛹虫草突变株进行

发酵,收集菌丝体固体粉末样品,这些样品来源广泛,保证模型的泛化能力,将此模型应用于蛹虫草高产突变株的筛选及其发酵条件的优化中是可行的。采用 MCPLS 识别异常样本和选择合适的校正集样本在总样本中所占比例,使模型具有更好的泛化能力和预测外延性能。采用 MWPLS 方法筛选了各有效成分的特征光谱变量,很好地提高了 PLS 模型的稳健性和预测性能。采用 NIR 光谱方法结合数学建模方法建立可同时快速无损测定蛹虫草菌丝体中腺苷、蛋白、多糖和虫草酸的含量。该方法具有较好的预测精度,具有快速、无损、无污染、同时测定多组分、成本低等众多优点,很适用于蛹虫草突变株的大量筛选和发酵条件优化,同时该方法也可推广于真菌突变株的筛选、发酵条件优化和发酵产品质量鉴定等领域的应用。

参 考 文 献

- [1] LIN Qun ying, SONG Bin, LI Tai hui (林群英, 宋 斌, 李泰辉). Chinese Bulletin of Microbiology (微生物学通报), 2006, 33(4): 154.
- [2] Yu K W, Suh H J, Bae S H, et al. J. Microbiol. Biotechnol., 2001, 11: 266.
- [3] Song C H, Jeon Y J, Yang B K, et al. J. Microbiol. Biotechnol., 1998, 8: 536.
- [4] Russell R, Paterson M. Phytochemistry, 2008, 69: 1469.
- [5] Mao Xian Bing, Eksriwong T, Chauvatcharin S, et al. Process Biochemistry, 2005, 40: 1667.
- [6] Lim Jong min, Yun Jong won, et al. Process Biochemistry, 2006, 41: 1620.
- [7] Argekar A P, Kunjir S S, Purandare K S. Journal of Pharmaceutical and Biomedical Analysis, 1996, 14: 1645.
- [8] Espinosa Mansilla A, Acedo Valenzuela M I, Munoz A., et al. Analytica Chimica Acta, 2001, 427: 129.
- [9] McClure W F. Journal of Near Infrared Spectroscopy, 2003, 11(6): 487.
- [10] GUO Weiliang, MENG Qingfan, LU Jiahui, et al. ICCSA 2006, LNCS 3980, 2006, 938.
- [11] Blanco M, Coello J, Iturriaga H. Chemometrics and Intelligent Laboratory Systems, 2000, 50: 75.

- [12] Swierenga H, Wulfert F, Noord O E, et al. *Analytica Chimica Acta*, 2000, 411: 121.
- [13] Swierenga H, Groot P J, Weijer A P, et al. *Chemometrics and Intelligent Laboratory Systems*. 1998, 41: 237.
- [14] Du Y P, Liang Y Z, Jiang J H, et al. *Analytica Chimica Acta*, 2004, 501: 183.
- [15] Kasemsumran S, Du Y P, Murayama K, et al. *Analytica Chimica Acta*, 2004, 512: 223.
- [16] Chinese Pharmacopoeia Commission. *Pharmacopoeia of the People's Republic of China*, People's Medical Publishing House, 2005. 75.
- [17] YU Jiar-liang, LI Kai-xiong, LU Shi-ling, et al(于见亮, 李开雄, 卢士玲, 等). *Food Research and Development*(食品研究与开发), 2008, 29(2): 141.
- [18] ZHANG Wei-jie(张惟杰). *Biochemical Research Technology of Sugar Compound*(糖复合物生化研究技术). Hangzhou: Zhejiang University Press(杭州: 浙江大学出版社), 2003. 12.
- [19] CAI You-hua, FAN Wen-xia, LIU Xue-ming, et al(蔡友华, 范文霞, 刘学铭, 等). *Modern Food Science and Technology* (现代食品科技), 2008, 24(1): 76.

Application of Near Infrared Spectroscopy in Screening *Cordyceps Militaris* Mutation Strains and Optimizing Their Fermentation Process

GUO Wei-liang¹, ZHANG Zhuo-yong², LU Jiar-hui¹, SONG Jia¹, WANG Di¹, TENG Li-rong^{1*}

1. College of Life Science, Jilin University, Changchun 130012, China

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

Abstract To fast screen high yield *Cordyceps militaris* mutations strains and optimize their fermentation process, near infrared (NIR) spectroscopy technology combined with chemometrics has been applied to establishing models for simultaneous determination of adenosine, protein, polysaccharide and *Cordyceps militaris* acid contents in *Cordyceps militaris* powder samples. Fermentations were implemented in Erlenmeyer flask with 468 *Cordyceps militaris* mutations strains under various fermentation conditions and *Cordyceps militaris* powder samples were collected. Then their NIR spectra were obtained using UV-VIS-NIR spectrometer and their adenosine, protein, polysaccharide and *Cordyceps militaris* acid contents were determined using reference methods. Partial least squares (PLS) method was employed to model the relationships between NIR spectra and the above mentioned components' contents in *Cordyceps militaris* powder samples. Monte Carlo partial least square (MCPLS) was applied to identify the outliers and select suitable number of calibration samples. Moving window partial least square (MWPLS) was applied to select the characteristic wavelength of the components. The degree of the approaching (Da) was employed as criterion for selecting effective pretreatment methods investigated. The optimum models for determination of adenosine, protein, polysaccharide and *Cordyceps militaris* acid contents in *Cordyceps militaris* powder samples were obtained with the above mentioned optimization. Their correlation between actual and predictive values of calibration samples (R_c) was 0.929 43, 0.984 79, 0.907 85, and 0.851 31, respectively. Their root mean square error of prediction set (RMSEP) was 0.667 14, 0.020 65, 0.011 31, and 0.011 59, respectively. The obtained results demonstrated that the fitting and the predictive accuracy were satisfactory. It is feasible to apply this method to screen the high yield *Cordyceps militaris* mutation strains and optimize their fermentation process.

Keywords Near infrared spectroscopy; Monte Carlo partial least squares; Moving windows partial least square; *Cordyceps militaris*

(Received Sep. 26, 2009; accepted Dec. 28, 2009)

* Corresponding author