牡蛎中氨基酸的傅里叶变换近红外光谱法快速定量测定

于　颖, 宋金明, 俞立东, 孙　萱

(中国科学院 海洋研究所, 山东 青岛 266071)

摘要: 本文通过乘法散射校正(multiplicative scattering correction MSC)、一阶求导(1^{st} Derivative, Db1)和 Norris 平滑等傅里叶变换近红外光谱定量分析数学模型, 用交叉验证法确定了最优化建模参数, 构建了牡蛎中氨基酸的快速定量测定方法。结果表明, 牡蛎近红外光谱与氨基酸含量化学计量学拟合, 线性相关系数为 0.9860, 交叉验证后线性相关系数为 0.9794, 对牡蛎样品进行检测, 结果与传统的氨基酸分析仪的测试结果相比, 相对误差小于 2.5%, 方法的精密度(RSD)小于 1.0%, 本文建立的傅里叶变换近红外光谱法定量分析牡蛎中氨基酸具有快速、准确、操作简便等特点, 测定重复性优于氨基酸分析仪法。

关键词: 氨基酸; 定量测定; 牡蛎; 傅里叶变换近红外光谱法

中图分类号: Q517　　**文献标识码**: A　　**文章编号**: 1000-3096(2016)12-0001-07

doi: 10.11759/hykx20160331005

随着计算机技术和化学计量学的发展以及近红外光谱定量技术的进步, 傅里叶变换近红外光谱(FT-IR)分析技术越来越多地在许多领域被广泛使用, 已经成为一种极具竞争力的光谱定量分析方法。化合物分子中 C—H、N—H、O—H 和 C—O 等基团基频振动的合频和倍频振动的吸收峰位于近红外光谱区, 因此近红外技术比较适于分析与这些基团有直接或间接关系的成分[1]。氨基酸分子具有上述基团, 理论上采用近红外光谱法测定牡蛎中氨基酸含量是可行的, 氨基酸分子结构通式为:

$$\begin{array}{c} NH_2 \\ | \\ R\text{—}CH\text{—}COOH \end{array}$$

牡蛎又名蚝、海蛎子、蛎黄、蛎蛤、牡蛤、古贲、蚵, 属软体动物门双壳纲珍珠贝目的曲蛎科和牡蛎科[2]。牡蛎是世界上最重要的养殖贝类和我国四大养殖贝类之一。牡蛎味道鲜美, 营养丰富, 可提高人体免疫力, 具有很高的药用价值, 牡蛎肉中必需氨基酸完全程度和质量比例优于人乳和牛乳[3-4]。氨基酸是牡蛎中的重要营养成分, 是评价牡蛎质量的重要指标, 因此也是牡蛎选苗育种的重要考虑因素[5], 氨基酸在碳的生物地球化学循环中具有重要价值, 定量获取氨基酸含量是揭示其生态学功能的前提[6-7]。其常规测定方法是用盐酸将样品水解, 经氨基酸分析仪的离子交换柱分离后, 与茚三酮的溶液产生颜色反应, 再通过分光光度计比色测定氨基酸含量[8], 该方法样品前处理繁琐(水解、浓缩等)、成本高, 且非常费时(>36h), 不适用于大批量快速分析的要求。常规氨基酸测定方法测定的是不同种类氨基酸含量, 最终加和为氨基酸总量, 因一般评价牡蛎中营养价值考察的是总量, 不同种类氨基酸含量意义不明显, 因此本文仅研究牡蛎中氨基酸总量的快速检测方法。近红外光谱法定量测定大豆等样品中氨基酸已有相关报道[9], 但至目前其用于牡蛎中氨基酸快速定量的研究和探索尚属空白。

本研究采用美国 Thermo Fisher Scientific 公司的 Antaris Ⅱ傅里叶变换近红外光谱仪和相应的软件 RESULT-Integration 和 TQ Analyst 定量分析软件, 对牡蛎进行近红外光谱扫描和分析, 建立了快速测定牡蛎中氨基酸含量的分析模型。结果表明, 方法不仅快速(<5min)可行, 重复性和再现性良好, 而且由于不需要繁琐的样品分析前处理, 基本没有试剂及耗材的消费。

1　实验部分

1.1　仪器与样品

美国 Thermo Fisher Scientific 公司的 Antaris Ⅱ型号的傅里叶变换近红外光谱仪, 仪器采用 CaF_2 分

收稿日期: 2016-02-11; 修回日期: 2016-05-22

基金项目: 中国科学院仪器设备功能开发技术创新项目(yg2012055)

[Foundation: Technology Innovation Project of the Instrument and Equipment of the CAS, No. yg2015055]

作者简介: 于颖(1973-), 女, 辽宁鞍山人, 硕士, 高级工程师, 主要近红外和中红外分析研究, 电话: 0532-82898876, E-mail: yingyu@qido.ac.cn; 宋金明, 通信作者, 研究员, 博士生导师. E-mail: jmsong@qdio.ac.cn

束器，配备 InGaAs 检测器、漫反射积分球(内置背景)、透射模块、光纤模块等配件及相应的 RESULT-Integration 和 TQ Analyst 定量分析软件。

本次实验选择来自丹东、大连、昌黎、海洋岛、韩国、加拿大等不同地区的 87 组牡蛎样品作为建模与验证样品，随机抽取其中 10 个用于建模后比对和验证实验，另外 77 组作为建立数学模型的样品集。将样品水洗净后放置于组织捣碎机在 10 000~15 000 r/min 条件下研磨制成组织匀浆，将匀浆转移到 5mL 样品管中后放入–20℃冷冻 10 h，冷冻后的样品用冷冻干燥机干燥 20 h 后用于测定。

1.2 实验方法

1.2.1 牡蛎氨基酸定量结果基础数据的获取

根据国家标准方法准确称取牡蛎干粉 20 mg 左右(精确到 0.001g)于安培瓶中，加入 10 mL6 mol/L 盐酸，充氮气后密封。置于 110℃烘箱中水解 24 h 后减压蒸干，用 0.02mol/L 盐酸定容至 10 mL。量取制备好的样品 200 μL，置于 1 mL 离心管中，加入 100 μL 三乙胺溶液、异硫氰酸苯酯 100 μL，混匀，室温放置 1 h，然后加入 400 mL 正己烷振摇后放置 10 min，取下层溶液，用 0.45 μm 膜过滤后，用氨基酸分析仪检测，通过外标法获得氨基酸含量。

1.2.2 光谱采集方法

将 3~5 g 牡蛎试样直接放置在积分球采集点上，利用 RESULT-Integration 软件设计信息采集工作流，设置采集参数为在 4 000~10 000 $\mathrm{cm^{-1}}$ 范围内扫描 32 次，分辨率 8.0 $\mathrm{cm^{-1}}$，优化增益无衰减，以漫反射的方式采集光谱。每个样品分别测试 3 次，取其平均光谱。

1.2.3 数学模型的建立与优化

采集所有样品的漫反射红外透射光谱图后，在“Standards”界面添加建立模型使用的样品光谱，导入建模样品的国标法测定值，采用最小二乘法(PLS)作为建立数学模型的化学计量学方法，利用 TQ Analyst 定量分析软件进行数学预建模。根据预建模的关联结果，利用马氏距离的方法诊断光谱中异常值，采用乘法散射校正(multiplicative scattering correction MSC)、一阶求导(1st Derivative, Db1)、Norris 平滑(其中 segment lengths 选择 5, gap between segment 选择 5)等方法对光谱进行预处理，得到关联系数较高的数据模型(图 1)。

1.2.4 数据模型的验证

对模型进行交叉验证，即每次从建模样品集中取出一个样品，用剩下的(N–1)个样品建立模型，预

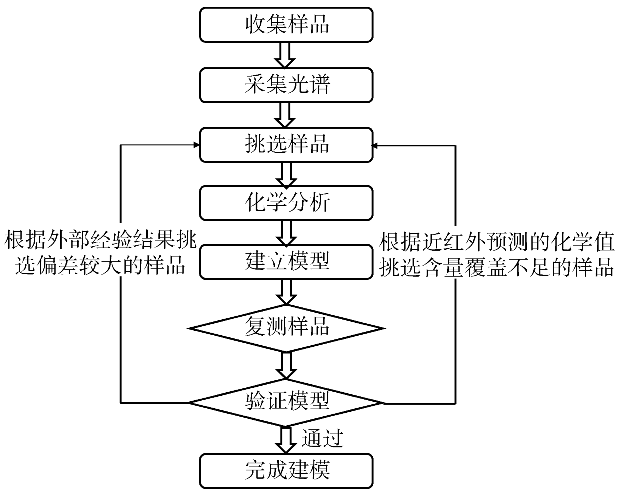

图 1 数据模型

Fig. 1 Model data

测被取出的样品，如此循环直至所有样品均被预测一次。考察预测值与国标法测定值的相关系数(R)和均方差(R_{MSECV})。

利用建立完成化学计量学数学模型，并随机抽取的俩个样品分别进行 10 次重复性实验。利用该模型检测随机抽取的十个样品，结果与氨基酸国标法比对。

表征近红外光谱定量模型建模效果的指标有校正相关系数 R_{C}、交叉验证相关系数 R_{CV}，校正集误差均方根 R_{MSRC}、交叉验证误差均方根 R_{MSSECV}，各参数的计算方程式如下:

$$R_{\mathrm{C}}=\sqrt{1-\frac{\sum(\hat{C}_{i,c}-C_i)^2}{\sum(C_i-\hat{C})^2}}$$

$$R_{\mathrm{MSEC}}=\sqrt{\frac{\sum(\hat{C}_{i,c}-C_i)^2}{n-1}}$$

$$R_{\mathrm{C,CV}}=\sqrt{1-\frac{\sum(\hat{C}_{i,cv}-C_i)^2}{\sum(C_i-\hat{C})^2}}$$

$$R_{\mathrm{MSECV}}=\sqrt{\frac{\sum(\hat{C}_{i,cv}-C_i)^2}{n-1}}$$

式中，C_i 为真实值或标准方法测定得到的参考值，$\hat{C}_{i,c}$ 是模拟拟合值，$\hat{C}_{i,cv}$ 是交叉验证时的模型拟合值，$\hat{C}$ 是参考值的平均值，n 是校正集样品数。

2 结果与分析

2.1 牡蛎的近红外光谱图和光谱范围的选择

用近红外光谱仪扫描出 87 个牡蛎试样的近红外光谱图(图 2)。利用马氏距离法鉴别光谱异常值(图

3)，其中显示分布图中的一个样品为异常，将该样品光谱值在数学模型中剔除，用主成分分析佐证其他样品光谱图无异常(图 4)。如图 2 所示，牡蛎红外吸收峰主要集中在 4 000~7 500 cm^{-1} 范围，其中在 5 257 cm^{-1}、4 900 cm^{-1} 和 4 537 cm^{-1} 处分别为氨基酸基团 O-H、N-H、C-H 等倍频吸收峰，求导后信号峰更明显见图 5 和图 6，包含氨基酸含量的重要信息。根据光谱原始图(图 2)和导数图(图 3 和图 4)中吸收峰的位置，将考查光谱范围分为 4 000~10 000 cm^{-1}、4 000~9 000 cm^{-1}、4 000~7 500 cm^{-1}、4 000~6 250 cm^{-1}，逐步剔除不提供信息的变量，根据不同波段的谱图与氨基酸含量的相关性进行光谱范围的筛选。实验结果(表 1)表明代表牡蛎中氨基酸含量变化红外吸收峰主要集中在 4 000~7 500cm^{-1}，虽然氨基酸基团- CONH 近红外吸收峰在 6 250~7 500cm^{-1} 之间，但由二阶导谱图分离可以看出该波段属于不同化学物质的合峰，且响应较弱，因此将光谱范围锁定在 4 000~6 250 cm^{-1}，通过微调，最终确定最佳光谱范围在 4 043~6 212 cm^{-1}。

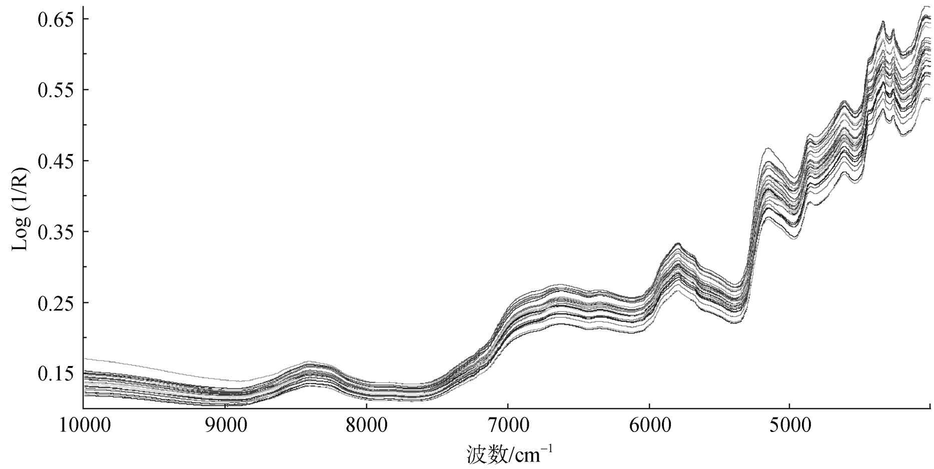

图 2　牡蛎原始近红外光谱图

Fig. 2　NIR spectrum of oyster

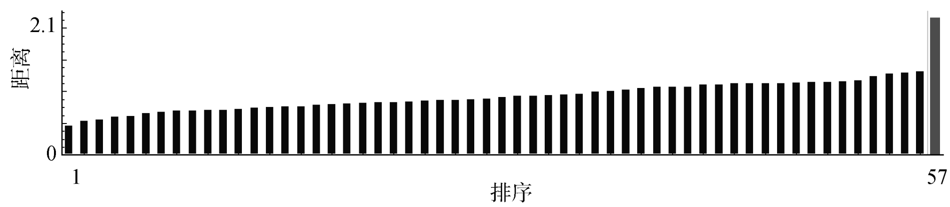

图 3　马氏距离方法鉴别异常值

Fig. 3　Identification of outliers using Markov distance method

表 1　模型的光谱范围筛选实验结果

Tab. 1　Experimental results of spectral range screening

光谱范围 (cm^{-1})	校正集		验证集	
	相关系数	误差均方根	相关系数	误差均方根
4 000~10 000	0.909 9	0.894	0.895 3	0.960
4 000~9 000	0.905 4	0.915	0.896 0	0.957
4 000~7 500	0.984 2	0.399	0.975 6	0.495
4 000~6 250	0.984 4	0.397	0.975 5	0.496

2.2　光谱预处理

近红外光谱不仅可以反映样品的化学组成及含量等信息，还会受到其密度、粒度等物理因素的干扰，甚至还包含有一部分的无关信息，如背景噪声、杂散光等。为了尽可能多地从光谱中提取与待测化学组分相关的信息，以便建立稳定可靠的定性、定量预测模型，就必须对光谱进行适当地预处理，消除其他

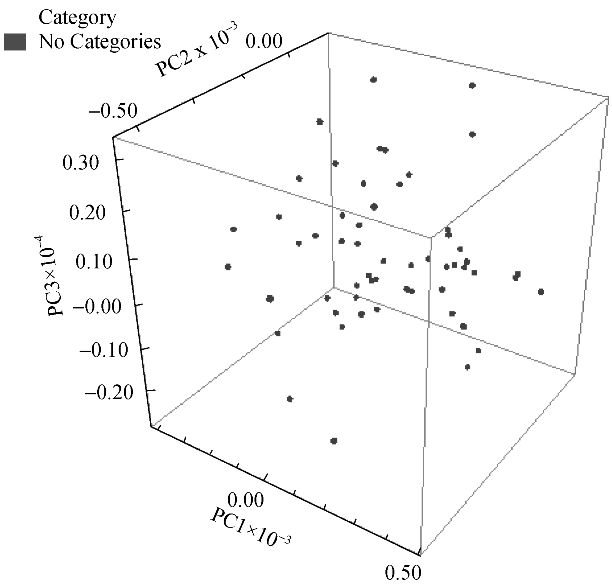

图 4　主成分分析 3D 立体图

Fig. 4　Principal component analysis 3D stereo diagram

无关信息的干扰[10-11]。一般近红外光谱建立化学计量学数据模型，都通过一定预处理方法提高光谱与数值间的拟合度，因此本实验考察了乘法散射校正(multiplicative scattering correction MSC)、一阶导数(1^{st} Derivative, Db1)、二阶导数(2^{nd} Derivative, Db2)等常用光谱预处理方法对数学模型的线性相关系数的影响(表 2)。

实验结果表明，导数可以用于消除光谱的基线漂移及背景干扰，分辨重叠峰，提高分辨率和灵敏度，但求导过程中同时也会放大噪声信号。一阶导数(Db1)已经将信号明显放大，虽然二阶导数(Db2)放大信号作用比一阶导数要好，但是同时放大了波数小于 4 500 cm^{-1} 等处的噪声，信噪比和相关系数相对反而降低。因此选择一阶导数 Db1。通过表 2 可以看出，虽然通过光谱预处理法提高了线性关系，但是提高结果并不显著，主要原因是本实验牡蛎吸收

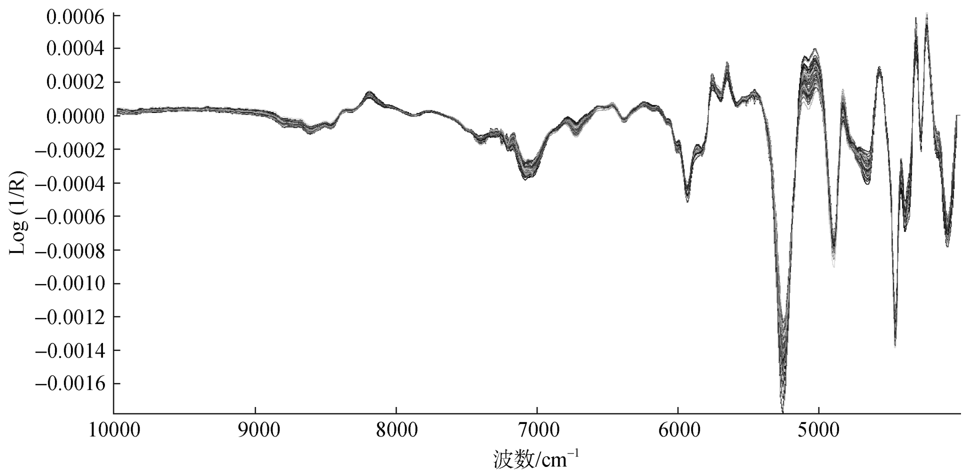

图 5　一阶导数(1^{st} Derivative, Db1)近红外光谱图

Fig. 5　First order derivative (Derivative, Db1) NIR spectroscopy

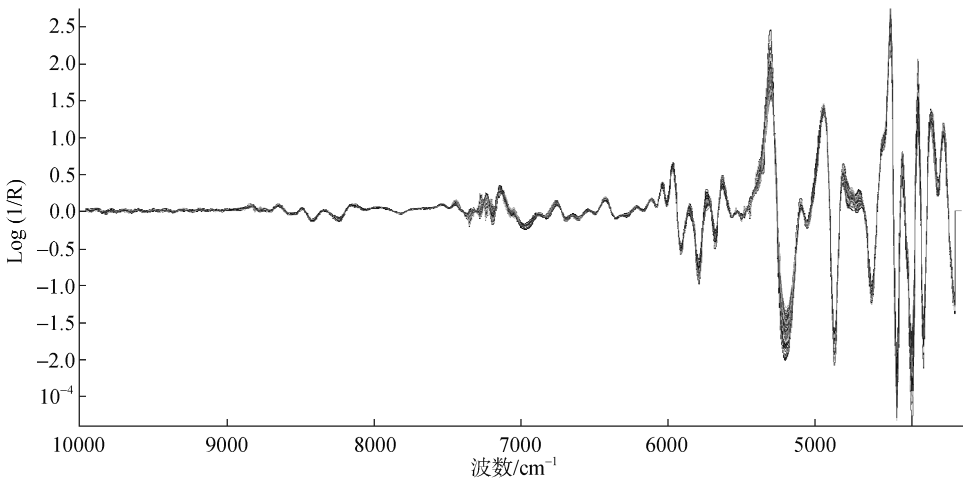

图 6　二阶导数(2^{nd} Derivative, Db2)近红外光谱图

Fig. 6　Second order derivative (Derivative, Db1) NIR spectroscopy

表 2　不同预处理方法的数学模型结果

Tab. 2　Mathematical model correlation coefficient of different pretreatment methods

预处理方法	主因子数	校正集		验证集	
		相关系数	误差均方根	相关系数	误差均方根
无	7	0.948 6	0.695	0.928 8	0.816
Db1	7	0.982 5	0.409	0.903 8	0.942
Db2	7	0.864 8	1.10	0.536 1	1.87
MSC	5	0.955 0	0.652	0.944 4	0.723
MSC+Db1	5	0.955 1	0.651	0.943 1	0.730

表 3　近红外-氨基酸含量数学模型的效果

Tab. 3　Effect of mathematical model of near infrared amino acid content

样品数量	参数	模型结果	其他参数
样品总量 87，其中建模 77 个，随机抽取 10 个进行比对实验	相关系数	0.986 0	光谱 4 043~6 212 cm^{-1} 主因子数: 5
	交叉验证相关系数	0.979 4	
	R_{MSEC}	0.379	
	R_{MSECV}	0.460	
	近红外预测值的绝对误差	±0.8%	
	交叉验证预测值的绝对误差	±0.9%	

峰 95%以上是氨基酸分子特征峰，谱图噪声少，在选定的光谱范围内几乎无无关信息，因此在没有光谱预处理的情况下也可以得到很好的线性关系，同时也说明该数据模型很好的代表了牡蛎中氨基酸含量变化，说明了近红外检测氨基酸含量方法是可行的。本实验最终采用了乘法散射校正(multiplicative scattering correction MSC)、导数(Derivative)、Norris 平滑(其中 segment lengths 选择 5, gap between segment 选择 5)等方法对光谱进行预处理。

2.3　交叉验证的结果

利用化学计量学方法最小二乘法(PLS)将牡蛎的近红外光谱与氨基酸含量进行拟合，通过交叉验证优化参数，建模样品的模型预测值与国标法测氨基酸真值相关系数为 0.9860，误差均方根(R_{MSEC})为 0.379，交叉验证相关系数为 0.9794，交叉验证误差均方根(R_{MSECV})为 0.460。预测值与真值绝对误差小于±1%。结果表明，数学模型良好，预测准确度高。

2.4　方法比对实验和重复性实验

随机抽取 10 个待测样品，分别采用新建近红外光谱法和氨基酸分析仪检测，检测结果(表 4)的最大相对偏差小于±1%，相对误差小于 2.5%。采用新建的近红外光谱法对两个样品分别进行 10 次测定，结果(表 5)表明，该方法测定牡蛎中氨基酸重复性很好，重复性测定结果的 RSD 小于 1%。因氨基酸分析仪

表 4　近红外法与氨基酸分析仪法比对实验结果

Tab. 4　Comparison of results of NIR with amino acid analyzer

样品编号	近红外法(%)	氨基酸分析仪法(%)	绝对偏差(%)	相对误差(%)
337	38.36	38.00	–0.36	–0.94
340	35.54	36.39	0.85	2.36
344	33.44	33.09	–0.35	–1.05
348	39.32	39.76	0.44	1.11
351	32.97	32.65	–0.32	–0.98
357	36.77	37.69	0.92	2.47
361	35.59	35.27	–0.32	–0.90
366	33.38	33.15	–0.23	–0.69
368	30.24	30.85	0.61	2.00
377	32.64	32.38	–0.26	–0.80

表 5　近红外光谱法重复性实验

Tab. 5　Results of repeated experiments in NIR

样品编号	近红外法测定牡蛎中氨基酸含量(%)		RSD (%)
	测定值	$\bar{x}$	
75	29.54、29.75、29.53、29.68、29.79、29.95、29.49、29.68、30.1、30.1	29.76	0.75
85	30.21、30.19、30.2、30.14、30.07、29.96、30.07、29.82、30.05、30.06	30.08	0.40

表 6 氨基酸分析仪重复性实验
Tab. 6 Results of repeated experiments in amino acid analyzer

样品编号	近红外法测定牡蛎中氨基酸含量(%) 测定值	$\bar{x}$	RSD (%)
18	38.69、39.08、39.25、39.03、36.72、36.68、39.14、39.06、38.28、38.38	38.43	2.51

法前处理复杂, 而近红外法测定氨基酸含量不需要进行任何前处理, 从而降低了误差来源, 因此近红外法重复性优于氨基酸分析仪法。

3 结论

本研究表明, 在 4 000~7 500 cm^{-1}光谱范围内, 牡蛎近红外谱图含有氨基酸分子主要基团 O-H、N-H、C-H 等的近红外倍频吸收峰, 且无其他无关信息峰, 光谱噪声小, 信息峰强度高, 因此构建近红外法氨基酸含量测定模型是可以实现的。本文利用化学计量学最小二乘法(PLS)将牡蛎近红外光谱与氨基酸含量进行拟合, 采用乘法散射校正(multiplicative scattering correction MSC)、一阶求导(1st Derivative, Db1)、Norris 平滑等光谱预处理方法优化数学模型, 建立了傅里叶变换近红外光谱定量分析牡蛎中氨基酸的数学模型, 通过对模型进行交叉实验验证, 近红外法的测定值与氨基酸真值之间的线性相关系数为 0.986 0, 与氨基酸分析仪检测氨基酸含量对比绝对误差小于±1%, 相对误差小于 2.5%, 实验结果表明建立的数学模型预测的准确度高。对于相同的样品, 氨基酸含量的近红外光谱法无需样品前处理, 降低了误差的来源, 实验重复性高于氨基酸分析仪。本文发展的傅里叶近红外光谱法为快速评价牡蛎中的氨基酸营养成分提供了可行的方法。

参考文献:

[1] 严衍禄, 吉海彦. 近代傅里叶变换近红外光谱技术及应用[M]. 北京: 科学技术文献出版社, 1994: 251.
Yan Yanlu, Ji Haiyan. Fourier transform near-Infrared spectroscopy technology and applications[M]. Beijing: Science and Technology Document Press, 1994: 251.

[2] 颜海燕, 陈新蓉, 王一农. 几种海洋动物的药用及其在临床上的应用[J]. 浙江海洋学院学报, 1999, l: 67-70.
Yan Haiyan, Chen Xinrong, Wang Yinong. Medicines of some marine animals and it's clinical practice[J]. Journal of Zhejiang Ocean University, 1999, l: 67-70.

[3] 李超柱, 陈艳华, 陈艳辉, 等. 牡蛎活性肽的分离及其免疫抑制作用的实验研究[J]. 海洋科学, 2013, 37(4): 52-55.
Li Chaozhu, Chen Yanhua, Chen Yanhui, et al. Separation and immunosuppressive activity analysis of the bioactive peptides from oyster[J]. Marine Sciences, 2013, 37(4): 52-55.

[4] 徐君卓. 世界牡蛎养殖及三倍体育苗[J]. 海洋水产科技, 1999, 57(1): 42-46.
Xu Junzhuo. Breeding and triploid breeding of world oyster[J]. Marine Fisheries Science and Technology, 1999, 57 (1): 42-46.

[5] 刘鹏超, 阙华勇, 张国范.长牡蛎早期胚胎发育中脂肪酸和氨基酸组成变化的研究[J]. 海洋科学, 2011, 35(10): 17-21.
Liu Pengchao, Que Huayong, Zhang Guofan. Changes in fatty acid and amino acid composition during early development of the Pacific oyster[J]. Marine Sciences, 2011, 35(10): 17-21.

[6] Song Jinming. Biogeochemical Processes of Biogenic Elements in China Marginal Seas[M]. Hangzhou: Springer-Verlag GmbH & Zhejiang University Press, 2010: 1-662.

[7] 宋金明, 徐永福, 胡维平, 等. 中国近海与湖泊碳的生物地球化学[M]. 北京: 科学出版社, 2008: 1-533.
Song Jinming, Xu Yongfu, Hu Weiping, et al. Biogeochemical of carbon in China's coastal waters and lakes[M]. Bin Jing: Science Press, 2008: 1-533.

[8] 国家标准化管理委员会. 食品中的氨基酸的测定[S]. GB/T5009.124-2003.
The National Standardization Management Committee. Determination of amino acids in food[S]. GB/T5009.124-2003.

[9] 李楠, 许韵华, 宋雯雯, 等. 利用近红外光谱技术快速检测大豆氨基酸含量[J]. 植物遗传资源学报, 2012, 13(6): 1037-1044.
Li Nan, Xu Yunhua, Song Wenwen, et al. A rapid method for detecting amino acid compositions in soybean by using near-infrared spectroscopy[J]. Journal of Plant Genetic Resources, 2012, 13(6): 1037-1044.

[10] Gnzález M, González P C, Alvarez G N, et al. On-line determination of fatty acid composition in intramuscular fat of Iberian pork loin by NIRs with a remote reflectance fibre optic probe[J]. Meat Science, 2005, 69(2): 243-248.

[11] 褚小立, 袁洪福, 陆婉珍.近红外分析中光谱预处理与波长选择方法进展与应用[J].化学进展, 2004, 16(4): 528-542.
Zhu Xiaoli, Yuan Hongfu, Lu Wanzhen.Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique[J]. Progress in Chemistry, 2004, 16(4): 528-542.

Quantitative analysis of amino acids in oyster by FT-NIR spectroscopy

YU Ying, SONG Jin-ming, YU Li-dong, SUN Xuan
(Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China)

Received: Feb. 11, 2016
Key words: amino acids; Quantitative determination; oyster; Fourier transform-near infrared spectroscopy (FT-NIR)

Abstract: In this study, spectra pretreatment methods of multiplicative scattering correction (MSC), Norris smoothing, and 1st Derivative (Db1) were applied to develop a FT-NIR mathematics model for quantitative analysis of amino acids in oyster. Optimum parameters of the model were studied using a cross validation method. By fitting the total amino content to the spectra, the linear correlation coefficient was found to be 0.9860 and the linear correlation coefficient of cross validation was 0.9794. The develop method was used to determine amino acids in oyster and results show no obvious difference between using the developed NIR and the traditional amino acid analyzer, with a relative error of less than 1.0% and repeatability of 1.0% RSD(n = 10). This suggests that the new NIR method is feasible for high precision determining of amino acids in oyster.

(本文编辑: 梁德海)