

文章编号:1671-6833(2010)06-0104-04

## 近红外光谱法快速定量分析那格列奈

石杰<sup>1</sup>, 李长滨<sup>1</sup>, 吴拥军<sup>1</sup>, 屈凌波<sup>1</sup>, 白雁<sup>2</sup>

(1. 郑州大学 化学系, 河南 郑州 450001; 2. 河南中医学院 药学院, 河南 郑州 450008)

**摘 要:**采用近红外漫反射光谱分析技术和化学计量学的方法对那格列奈进行快速无损的定量分析. 采集41份不同浓度系列的那格列奈样品测量近红外漫反射光谱, 用多元散射校正结合一阶导数的光谱预处理方法, 结合偏最小二乘法(PLS)建立那格列奈的定量分析模型. 结果表明:交叉验证均方根误差(RMSECV)为0.448 5, 决定系数 $R^2$ 为0.999 8; 预测均方根误差(RMSEP)为0.196 0. 方法精密度RSD为1.1% ( $n=6$ ), 加标回收率为95.55%. 结果表明, 用近红外光谱分析技术对那格列奈进行定量分析结果准确可靠, 方法简便快速, 可推广用于此类样品的工业在线分析.

**关键词:**近红外漫反射光谱; 偏最小二乘法; 定量分析; 那格列奈

**中图分类号:** O561.3; R917

**文献标识码:** A

### 0 引言

近红外光谱(NIRS)是指介于可见光与中红外光之间, 波长范围为780~2 526 nm的分子光谱. 其主要谱峰反映分子中C—H, N—H, O—H等含氢基团的倍频和合频吸收. 目前近红外分析技术已广泛应用于石油化工、烟草、食品等领域的定量和定性分析以及过程控制<sup>[1-3]</sup>. 由于NIRS分析技术具有分析速度快、不需损耗化学样品和试剂、易于实现现场分析等特点, 近年来在药物分析领域应用已日趋广泛<sup>[4-5]</sup>, 它已成功应用于包括原辅料和制剂在内的生产过程等很多环节的定性和定量分析.

那格列奈为D-苯丙氨酸衍生物, 属于非磺酰脲类降血糖药, 为一种新型的餐时血糖调节剂, 能有效控制餐后血糖水平, 具有起效快、作用时间短、不易引起心血管副作用等特点. 目前其含量测定常见方法为高效液相色谱法<sup>[6-7]</sup>和流动注射化学发光法<sup>[8]</sup>, 但这两种方法都具有操作繁琐、费时等缺点. 笔者应用NIRS结合PLS建立了测定那格列奈含量的定量分析模型, 利用模型对实际样品进行预测分析. 取得了较好的实验结果, 为那格列奈质量的快速评价和在线检测提供参考和依据.

### 1 样品与实验仪器

#### 1.1 样品

那格列奈原料药(由某制药公司提供); 辅料硬脂酸镁、淀粉等为药用规格(由上述公司提供); 按厂家那格列奈片剂配方配制含主药那格列奈浓度范围为10%~30%的41个实验样品. 4个批次的那格列奈片剂(从药店购买).

#### 1.2 实验仪器

Antaris6700型近红外光谱仪(美国 Thermo Nicolet 公司), 配有 PbS 和 InGaAs 检测器、外接积分球、样品旋转器, 软件为 Nicolet 公司编制的 TQ7.0 分析软件, 50 mL 标准石英杯, 石英研钵, BP210S 型万分之一分析天平(德国赛多利斯).

### 2 样品的制备及近红外光谱的采集

按照处方量精确配制41个不同系列浓度的那格列奈样品, 浓度范围为10%~30%, 把41个实验样品随机分成2组, 一组为校正集(The Calibration set), 共31个; 另一组为预测集(The Prediction set), 共10个. 分别研细并过孔径0.15 mm筛, 以空气为参比, 按下述条件进行扫描: 测样方式为积分球漫反射, 分辨率为8 cm<sup>-1</sup>, 光谱采集范围4 000~12 000 cm<sup>-1</sup>, 扫描次数32次, 温度

收稿日期:2010-06-13; 修订日期:2010-07-26

基金项目:国家自然科学基金资助项目(J0830412); 河南省重大公益科研项目(081100912500)

作者简介:石杰(1950-), 男, 河南郑州人, 郑州大学教授, 硕士生导师, 主要从事食品与药物分析研究, E-mail: shijie@zzu.edu.cn.

(25±1)℃,相对湿度 43%。图 1 为 41 个实验室样品的原始近红外光谱叠加图。

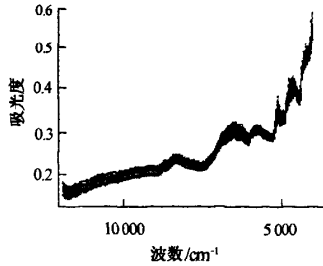


图 1 41 份那格列奈的近红外光谱图  
Fig.1 The NIR spectrum of 41 batches nateglinide

3 那格列奈定量分析模型的建立

3.1 光谱数据的预处理

样品的物理性质如颗粒大小、紧密程度、散射等都有可能引起基线的漂移,对建立定量模型产生较大影响,因此需要用多元散射校正去除近红外漫反射光谱中样品的镜面反射及不均匀性造成的噪声,消除漫反射光谱的基线及光谱的不重复性<sup>[9]</sup>。一阶导数处理过的光谱能消除因颜色差别引起的基线飘移,强化谱带特征信息,使那格列奈的特征信息更加明显的表现出来。图 2 为经过一阶导数处理后的那格列奈光谱叠加图。比较不同预处理(多元散射校正、标准归一化、常量偏移消除、标准归一化结合一阶导数、多元散射校正结合一阶导数等)后模型的 RMSECV 和 RMSEP 的大小,选择最佳预处理方法。比较结果见表 2。

选择的光谱预处理方法应使 RMSECV 和 RMSEP 两者的值尽量小并尽可能接近。综合考虑选择多元散射校正和一阶导数作为建立模型的预处理方法。

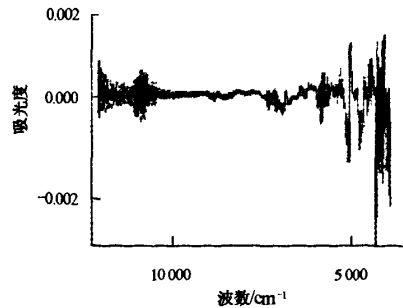


图 2 41 份那格列奈的一阶导数光谱叠加图  
Fig.2 First derivative spectra spectrum of 41 batches nateglinide

3.2 建模波段的选择

在近红外光谱分析识别模型建立过程中,波

表 1 不同预处理方法对模型性能参数的影响

Tab.1 Effect of different spectrum pretreatment methods on model

光谱预处理方法	RMSECV	RMSEP	维数
常量偏移消除	0.501 0	0.366 0	6
多元散射校正(MSC)	0.369 3	0.333 0	6
标准归一化(SNV)	0.382 9	0.325 0	4
二阶导数(SD)	0.722 9	0.355 0	2
多元散射校正+二阶导数	0.523 5	0.290 0	6
一阶导数(FD)	0.641 4	0.302 0	3
标准归一化+一阶导数	0.448 7	0.199 0	3
多元散射校正+一阶导数	0.448 5	0.196 0	3

段区间的选择要尽量覆盖样品的所有信息。波段选择可以是全波段,但是这种选择计算量较大;如果部分光谱区与样品性质间缺乏相关性,也可以另外选择适当的光谱区域<sup>[10]</sup>。根据主成分在 NIRS 吸收中的特征波段,以内部交叉验证决定系数  $R^2$  和交叉验证均方根误差 RMSECV 为指标,以突出特征信息为原则,分别对软件推荐波段及特征信息明显波段(11 493.66 ~ 4 092.21  $\text{cm}^{-1}$ ; 11 493.66 ~ 4 092.21  $\text{cm}^{-1}$ , 7 164.49 ~ 5 008.72  $\text{cm}^{-1}$ ; 8 008.45 ~ 4 008.81  $\text{cm}^{-1}$  等)进行比较,选择最佳波段。

表 2 不同光谱范围对  $R^2$  和 RMSEC 的影响

Tab.2 Effects of different spectrum range on calibration results

序号	波长范围/ $\text{cm}^{-1}$	$R^2$	RMSEC
1	11 493.66 ~ 4 092.21	0.999 8	0.127 0
2	11 493.66 ~ 4 092.21	—	—
	7 164.49 ~ 5 008.72	0.999 5	0.183 0
3	8 008.45 ~ 4 008.81	0.998 2	0.366 0
4	4 778.74 ~ 4 585.89	—	—
	4 909.88 ~ 4 782.60	0.998 5	0.334 0
	5 793.11 ~ 5 731.40	—	—
5	5 983.99 ~ 5 507.93	—	—
	4 909.88 ~ 4 500.00	0.982 6	0.355 0
	4 420.12 ~ 4 008.81	—	—

3.3 主成分数的确定

正确选择主成分数是 PLS 方法建立校正模型的必要条件。如果建立模型时使用的主成分数过少,就不能反映未知样品被测组分产生的测量数据(光谱)变化,其模型预测准确度就会降低。如果使用过多的主成分建立模型,就会将一些代表噪音的主成分加到模型中,使模型的预测能力下降<sup>[11]</sup>。本实验采用留一交叉验证法来筛选主成分数,以内部交叉验证均方差(RMSECV)和残差平方和(PRESS)作为判断依据。RMSECV 和

PRESS 值越小,表明模型的预测能力越强.由图 3 不难看出,模型的 RMSECV 和 PRESS 随着维数的增加不断减小,当维数为 3 时,所得的 RMSECV 达最小值,随后 RMSECV 和 RMSEP 变化幅度很小.因此,本实验选定的维数为 3.

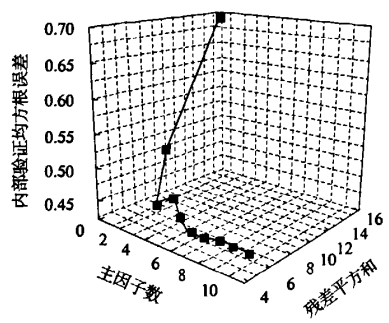


图 3 主因子数对 PLS 定量分析模型的 RMSECV 及 PRESS 值的影响  
Fig.3 Effects of rank for RMSECV and PRESS of PLS model

3.4 校正模型的建立与预测

将 41 份样品随机分为校正集和验证集,要保证验证集的含量范围在校正集的含量范围之内,并且校正集和验证集的浓度范围分布要均匀.具体情况见表 3.

表 3 校正集和验证集划分及含量分布情况  
Tab.3 Composition of the calibration and validation sample sets

样品	样品数	浓度范围/%	平均浓度/%
校正集	31	10.0~30.0	19.82
验证集	10	11.5~29.5	20.55

将收集到的 41 份样品的 NIR 光谱输入到 TQ7.0 分析软件中,采用偏最小二乘法建立定量校正模型,并以  $R^2$ 、RMSECV、RMSEP 作为模型性能的评价指标.  $R^2$  越大、RMSECV 越小,表明模型结构越合理, RMSEP 越小,表明模型的预测性能和推广能力越强.由最优的预处理模型选出波长范围区间为  $11\,493.66\sim4\,092.21\text{ cm}^{-1}$ ,采用一阶导数及多元散射校正的预处理方法,建立交叉检验定量校正模型,其  $R^2$  为 0.999 8, RMSECV 为 0.448 5, RMSEP 为 0.196 0.

一共分析了 41 份不同浓度的那格列奈样品.由图 4 可知,预测集比较均匀的分散在校正集中,而且预测值的大小跟真实值都比较接近.可知建立的模型预测能力较好.用所建立的定量模型对预测集样本进行预测,得到每一个预测集样本的预测值,再与预测集样本的真实结果进行比较,将预测

值与真实值的比值作为回收率.预测结果见表 4.

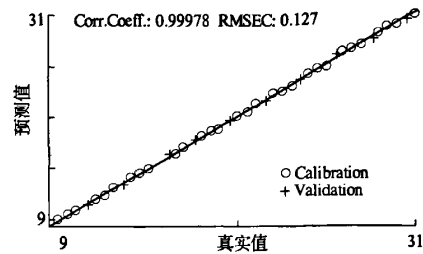


图 4 经 PLS 模型交叉检验的预测值和真实值的相关性图  
Fig.4 The scatter plot of nateglinide

表 4 预测集中那格列奈的预测结果分析

Tab.4 Analytical results of validation samples %					
序号	真实值	预测值	绝对误差	相对误差	回收率
1	11.50	11.31	-0.19	-1.65	98.30
2	13.50	13.32	-0.18	-1.33	98.70
3	16.00	16.31	0.31	1.94	102.00
4	17.50	17.68	0.18	1.03	101.00
5	19.50	19.55	0.05	0.26	100.00
6	21.50	21.39	-0.11	-0.51	99.50
7	23.50	23.62	0.12	0.51	100.50
8	25.50	25.86	0.36	1.41	101.00
9	27.50	27.40	-0.10	-0.36	99.60
10	29.50	29.37	-0.13	0.44	99.60

注:平均回收率为 100.02%, RSD 为 1.13% ( $n=10$ ).

3.5 模型对实际样品预测能力的考查

本研究建立模型的最终目的是对实际生产药物含量的测定分析或生产车间的在线分析,以找到一种简便快捷的判定药物含量成分的新方法.本实验用 4 个实际样品对所建立的模型进行考查验证.具体的验证结果如下表所示.

表 5 实际样品的预测分析结果  
Tab.5 Prediction result of practical samples %

序号	真实值	预测值	绝对误差	相对误差
1	13.16	13.29	0.13	0.99
2	19.66	19.04	-0.62	-3.15
3	26.71	27.05	0.34	1.27
4	20.00	19.46	-0.54	-2.70

4 方法学考察

4.1 精密度试验

取 30 mg 规格的那格列奈样品,相同条件下重复扫描 6 次,输入校正模型预测其百分含量.测定结果的 RSD 为 1.1% ( $n=6$ ),可见该方法的重复性良好.

4.2 加标回收率实验

取那格列奈片剂 0.500 8 g,加入那格列奈 0.030 7 g.混合均匀并过孔径 0.15 mm 筛,按测

定其他样品的方法采集样品的 NIRS, 平行测定 6 次, 利用建立的定量校正模型预测所加入原料的含量. 结果为加样回收率范围 94.39% ~ 96.14%, 平均回收率为 95.55%, RSD 为 0.72% ( $n=6$ ). 可见该方法准确可靠.

#### 4.3 计量学方法验证

训练集样品经内部交叉验证, 交叉验证均方差 (RMSECV) 为 0.448 5, 外部验证均方差 (RMSEP) 为 0.196 0. 与真实值相比, NIRS 法测定结果相对偏差都在 5% 以内, 平均相对偏差为 0.94%. 将验证集样品 NIRS 结果与实际结果在 95% 可信区间内进行配对  $t$  检验, 两种结果无显著差异.

## 5 结论

利用近红外光谱技术与化学计量学相结合的方法建立了那格列奈的定量分析模型. 系统精密度和加样回收率良好, 说明该方法具有可行性. 结果表明该方法准确可靠, 显示出良好的应用前景.

工业样品数量太大, 而实验室样品的局限性都影响定量分析模型的稳定性和适用性. 随着样品校正集个数的增多和浓度范围的扩大, 特别是工业样品和实验室样品的结合, 近红外定量分析模型的稳定性和预测性将会更好, 近红外光谱技术也将更好地应用于实际生产中.

## 参考文献

[1] 陆婉珍, 袁洪福, 徐广通, 等. 现代近红外光谱分析技术[M]. 北京: 中国石化出版社, 2000: 193-235.

- [2] 张鼎方. 近红外技术应用于成品卷烟的内在质量稳定性控制[J]. 福建分析测试, 2009, 18(2): 39-44.
- [3] 孙通, 徐惠荣, 应义斌. 近红外光谱分析技术在农产品/食品品质在线无损检测中的应用研究进展[J]. 光谱学与光谱分析, 2009, 29(1): 122-126.
- [4] 屈凌波, 陈国广, 盛龙生, 等. 近红外光谱技术及其在药物分析中的应用[J]. 药物分析杂志, 1996, 16(3): 204-206.
- [5] 周云, 臧恒昌. 近红外分析技术在中药鉴定及含量测定方面的研究进展[J]. 食品与药品, 2009, 19(01): 72-74.
- [6] MADHAVI A, REDDY G S, SURYANARAYANA M V, et al. Development of a new analytical method for determination of related components in nateglinide[J]. Chromatographia, 2008, 67: 639-645.
- [7] 何笑荣, 胡欣, 闫小燕, 等. 高效液相色谱法测定那格列奈含量及其有关物质[J]. 中国医院药学杂志, 2005, 25(1): 48-50.
- [8] 石杰, 陈桂娟, 吴拥军, 等. 流动注射化学发光法测定那格列奈[J]. 分析实验室, 2008, 27(12): 69-71.
- [9] 吴国萍, 相秉仁. 近红外光谱技术应用于摇头丸中 MDMA/MA 无损定量分析的研究[J]. 分析测试学报, 2007, 26(5): 698-701.
- [10] 梁家杰, 潘涛, 陈星旦, 等. 白砂糖色值近红外光谱分析的波段选择[J]. 红外技术, 2009, 31(2): 90-94.
- [11] 吴新生, 谢益民. 主成分-BP 算法在近红外光谱法纸浆卡伯值测量中的应用研究[J]. 分析测试学报, 2005, 24(4): 10-12.

## Rapid Quantitative Analysis of Nateglinide Based on Near Infrared Spectroscopy

SHI Jie<sup>1</sup>, LI Chang-bin<sup>1</sup>, WU Yong-jun<sup>1</sup>, QU Ling-bo<sup>1</sup>, BAI Yan<sup>2</sup>

(1. Department of Chemistry, Zhengzhou University, Zhengzhou 450001, China; 2. College of Pharmacy, Henan University of Traditional Chinese Medicine, Zhengzhou 450008, China)

**Abstract:** Near infrared spectroscopy (NIRS) and chemometrics methods were applied in fast, non-destructive, quantitative analysis of nateglinide. The quantitative analysis model was established by 41 different series near-infrared reflectance spectroscopy of nateglinide samples, combined with multiplicative scatter correction, first derivative and partial least squares (PLS). The cross-validation root mean square error (RMSECV) was 0.448 5 and determination coefficient  $R^2$  was 0.999 8; root mean square error of prediction (RMSEP) was 0.196 0. Precision of the method RSD was 1.1% ( $n=6$ ), the average recovery rate was 95.55%. It is shown that near-infrared spectroscopy technology is accurate and reliable for nateglinide quantitative analysis. The method is simple and rapid, and it may be extended to industrial on-line analysis of such samples.

**Key words:** near infrared spectroscopy; partial least squares; quantitative analysis; nateglinide