

近红外光谱结合第一范围标定法鉴定化学降糖原料药

蔡蕊¹, 王世盛¹, 孟昱¹, 孟庆刚², 赵伟杰^{*1}

(1. 大连理工大学 制药科学与技术学院, 辽宁 大连 116024;
2. 沈阳药科大学 药学院, 辽宁 沈阳 110016)

摘要: 以建立一种快速、准确、无需样品预处理的鉴定 8 种降糖原料药(固态和液态)的近红外分析方法为目标,分别取 64 个不同批次的样品为校正集,另取 32 个样品用于定性模型的验证,使用一阶导数进行光谱预处理,结合第一范围标定法,在 12 072.3~4 000.0 cm^{-1} ,建立了 NIR 8 种降糖原料药固态定性分析模型,结果显示每种原料药均可被成功鉴定,无一错判.说明所建立的模型可对 8 种降糖原料药进行快速准确的定性分析,适用于制药企业对降糖原料药进货验收和投料前的品种鉴别.

关键词: 近红外光谱;降糖原料药;定性分析;第一范围标定法;过程分析技术
中图分类号: O657.39 **文献标志码:** A

0 引言

在制药行业,对药品生产原料的严格质量控制是确保药品质量的重要前提.传统的质量控制方法是将原料从原料仓库中取出,送到分析中心或化验室进行定性定量分析,再反馈到原料仓库,合格后才允许发放,耗时较长,无法达到及时控制质量的目的.过程分析技术(process analytical technology, PAT),可通过在生产过程中对产品进行及时检测而达到实时监控质量的目的,而无需等到最后环节的成品检验进行质控^[1~5].美国食品药品监督管理局(FDA)也鼓励制药企业将过程分析技术作为现行生产质量管理规范(cGMP),并广泛地应用到药品生产过程中^[6].由于 NIR 光谱(near infrared spectroscopy)具有穿透力强、无需样品预处理、无损、可远距离检测(光纤)等特点,在药品质量控制、真伪鉴别方面已有广泛的应用^[7,8],已成为一种重要的在生产过程中使用的过程分析技术^[9].NIR 光谱的应用,使得原料鉴别、检验工作可以直接在存储车间仓库或卸货现场进行,甚至无需打开样品内包装,检测时间只需几秒钟.但目前,该技术仅被少数国外制药企业用于原料进货入场验收时的质量鉴别、原料投放前

的质量分析,混合、干燥、压片、包衣等过程检测及终产品原位无损分析^[10,11],并且对降糖原料药的定性分析,尚未见报道.本文结合第一范围标定法对降糖原料药 NIR 定性分析的可行性进行研究,试图建立一种快速、准确、无需样品预处理的鉴定 8 种降糖原料药的方法,以期为制药企业对降糖原料药进货验收和投料前的品种鉴别提供新型的检测手段.

1 实验材料与方法

1.1 实验仪器与材料

MPA 型傅里叶变换近红外光谱仪(德国 BRUKER 公司),配有积分球漫反射采样系统、RT-PbS 检测器、样品转轮、透过单元、光纤探头测样器-铟镓砷(InGaAs)检测器、OPUS 6.5 操作软件.

降糖原料药盐酸二甲双胍、盐酸吡格列酮、格列美脲、格列喹酮、瑞格列奈、格列本脲、格列齐特、格列吡嗪,每个品种各 10 批次,均由辉瑞制药有限公司大连工厂提供.

1.2 近红外光谱采集

8 种降糖原料药固态粉末的近红外光谱图采

集(见图1):取盐酸二甲双胍、盐酸吡格列酮、格列吡嗪、格列齐特、格列美脲、格列喹酮和瑞格列奈原料粉末(无需去除聚乙烯塑料包装),使用漫反射光纤探头压住药粉,在 $12\ 500\sim 4\ 000\ \text{cm}^{-1}$ 进行光谱采集,光谱分辨率为 $8\ \text{cm}^{-1}$,样品扫描次数为16次,背景扫描次数为16次,温度 $23\sim 25\ ^\circ\text{C}$,温度变化每小时不超过 $1\ ^\circ\text{C}$,一天不超过 $2\ ^\circ\text{C}$ 。

8种降糖原料药标准品溶液的近红外光谱图采集:配制盐酸二甲双胍、盐酸吡格列酮、格列吡嗪、格列齐特、格列美脲、格列喹酮、格列本脲和瑞格列奈的最大浓度溶液(盐酸二甲双胍浓度为 $20\ \text{mg/mL}$;盐酸吡格列酮、格列齐特和格列本脲浓度为 $10\ \text{mg/mL}$;格列喹酮和瑞格列奈浓度为 $5\ \text{mg/mL}$;格列吡嗪和格列美脲浓度为 $2\ \text{mg/mL}$),装入石英液体样品池中,采用透射方式在 $12\ 500\sim 4\ 000\ \text{cm}^{-1}$ 进行光谱采集,除光谱分辨率为 $16\ \text{cm}^{-1}$ 外,其他参数同上。

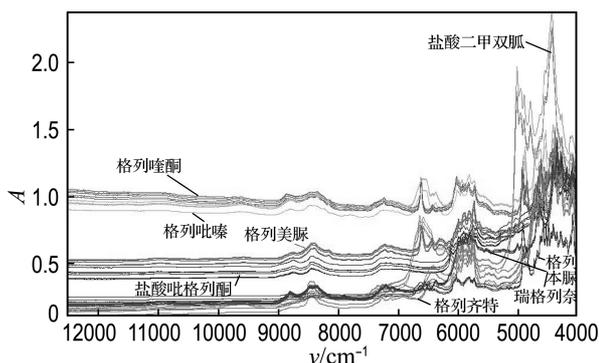


图1 固态降糖原料药的近红外光谱图

Fig. 1 NIR spectra of solid oral hypoglycemic APIs

1.3 光谱预处理方法

扫描共得到96张光谱图,其中64张光谱图作为训练样品集进行模型的建立.采集8种降糖原料药标准品固态粉末的近红外光谱图时,由于样品颗粒大小、均匀性、仪器状态等因素的影响,光谱基线会产生偏差或漂移,为消除这些不良影响,原始光谱图需要经过预处理.常用的预处理方法有消除常数偏移量、减去一条直线、矢量归一化、最小-最大归一化、多元散射校正、内部标准、一阶导数、二阶导数、一阶导数+减去一条直线、一阶导数+矢量归一化、一阶导数+MSC,共11种.而后以得到最大的选择性 S (至少大于2,越大越好)为目标,选择最佳波数范围、算法进行建模。

2 结果与讨论

2.1 处理方法

定性分析的目的是判定测试谱图与库中参考谱图的差别.必须先定义测试谱图相似性的方法,以及一个阈值.这个阈值用来判定一张谱图是否与参考谱图相似或相同。

标准算法和因子法是定性分析时常用的两种基本算法,分析过程中,两种算法将测试谱图同所有参考谱图进行比较.比较结果即是匹配值,也对应为光谱图距离 D .两张谱图越匹配,则距离越短.完全相同的谱图匹配值为零(例如:参考谱图对自己测试)。

标准算法是用欧氏距离 D 来计算光谱图距离,计算公式为

$$D = \sqrt{\sum_k [a(k) - b(k)]^2} \quad (1)$$

通过欧氏距离来比较谱图的匹配度, $a(k)$ 和 $b(k)$ 是谱图A、B的纵坐标值,即光谱强度。

因子法是先将谱图表示为因子谱的线性组合, $a = T_1 a f_1 + T_2 a f_2 + T_3 a f_3 + \dots$, f_i 为因子谱,各因子谱正交, T 为每一个因子谱的系数,其数值表征相应因子谱对所表征谱图的贡献.再用因子 T (也称得分)来计算光谱图距离,公式为

$$D = \sqrt{\sum_i (T_i a - T_i b)^2} \quad (2)$$

然后计算各品种训练集中每一样品光谱图到该样品平均光谱图的距离(H),并据此计算每一样品的阈值(D_T):

$$D_T = \max(H) + X S_{\text{dev}} \quad (3)$$

其中 $\max(H)$ 为测定的样品光谱图中到该样品平均光谱图的最大距离; S_{dev} 为 H 的标准偏差; X 为系数,一般根据经验 X 选用0.25.使用上述公式来比较谱图的匹配度。

选第一范围标定法则光谱图距离的最小和最大值由第一频率范围决定.然后计算其他频率范围的距离,其值按相同范围标定,也就是与第一频率范围同样的最小和最大值.第一范围标定法与标准算法和因子法不同,重叠谱区没有合在一起.例如:如果设定第一频率区间为 $1\ 500\sim 1\ 200\ \text{cm}^{-1}$ 、第二区间为 $1\ 500\sim 1\ 400\ \text{cm}^{-1}$,它们不会组合成一个频率范围,计算光谱图距离时 $1\ 500\sim 1\ 400\ \text{cm}^{-1}$ 的数据点被使用两次.矢量归一化预处理方法不能与第一范围标定法一起使用;因为

它是自动进行的. 与标准算法和因子法相反, 矢量归一化是对每个频率范围单独计算的, 然后从这些值中计算出平均距离.

标准算法是用欧氏距离表示光谱图距离, 而第一范围标定法使用相关因子.

计算 Pearson 相关因子 r , 即测出谱图 A 和 B 之间的相关性:

$$r = \sum a_n(k) \cdot b_n(k) \quad (4)$$

对每一个频率范围分别计算相关性. 对某一频率范围全部数据点 k 求和: a_n 和 b_n 是归一化的谱图强度, 由归一化步骤得到

$$r = \frac{\sum (a_n(k) - a_m)(b_n(k) - b_m)}{\sqrt{\sum (a_n(k) - a_m)^2} \cdot \sqrt{\sum (b_n(k) - b_m)^2}} \quad (5)$$

式中: a_m 和 b_m 是响应频率范围的平均光谱强度; 相关因子 r 的范围在 -1 (谱图相反) 到 $+1$ (谱图相同) 之间.

相关因子按照下式变换成光谱图距离:

$$D = (1 - r) \cdot 1\,000 \quad (6)$$

光谱图距离 D 值落在 0 (谱图相同) 和 $2\,000$ (谱图相反) 之间.

模型的结果一方面看所有建模光谱图是否被唯一鉴定, 另一方面还要看不同库之间的选择性, 一般用 S 表示:

$$S = D / (T_1 + T_2) \quad (7)$$

S 为选择性, D 是两个库平均光谱之间距离, T_1 与 T_2 为两个库的阈值. $S < 1$: 类交叉; $S = 1$: 类之间有接触; $S > 2$: 类之间可以完全分开, 因此 S 越大, 类之间差别越大, 模型预测的准确性越高. 根据实验效果, 选取合适的算法.

在建立 8 种降糖原料药固态模型时, 将采集到的原始光谱图进行各种不同预处理方法和分析方法的调试, 初选定波谱范围 $12\,072.3 \sim 4\,000.0 \text{ cm}^{-1}$, 选择最优参数, 确定的最优光谱预处理方法为一阶导数, 分析方法为第一范围标定法.

在建立 8 种降糖原料药溶液的近红外光谱模型时, 需将原料药溶解, 由于溶剂甲醇的吸收很强, 原始谱图所显示的信息, 除几处有细小的差别外, 用肉眼无法辨别其与甲醇的吸收谱图的区别 (如图 2、3 所示), 必须经过光谱预处理方法对谱图进行预处理, 选择较好的结果, 建立定性分析模型. 经过调试, 选定扫描波谱范围 $10\,235.7 \sim 4\,237.2 \text{ cm}^{-1}$, 确定最优预处理方法为一阶导数, 分析方法为第一范围标定法.

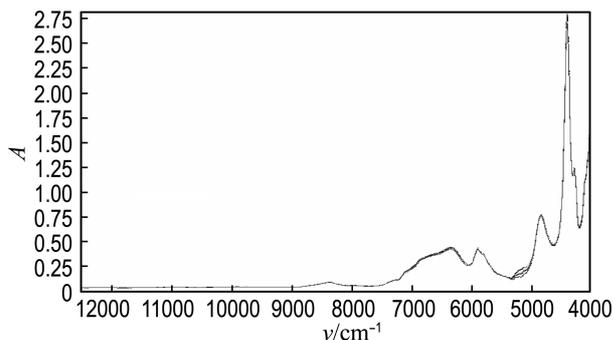


图 2 液态降糖原料药的近红外光谱图

Fig. 2 NIR spectra of liquid oral hypoglycemic APIs

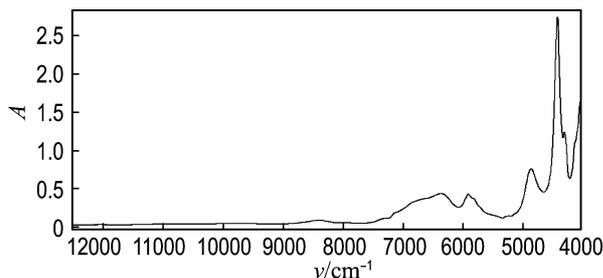


图 3 纯甲醇溶剂的近红外光谱图

Fig. 3 NIR spectra of pure methanol solvent

2.2 定性模型的建立

采用近红外漫反射光谱法对 8 种固态降糖原料药建立定性分析方法, 在扫描谱区范围 $12\,072.3 \sim 4\,000.0 \text{ cm}^{-1}$, 使用一阶导数, 结合第一范围标定法进行建模. 将 8 种降糖原料药配制成对应浓度的溶液, 采用近红外透射光谱法对其建立定性分析方法, 在扫描谱区范围 $10\,235.7 \sim 4\,237.2 \text{ cm}^{-1}$, 使用一阶导数, 结合第一范围标定法, 建立其分析模型.

对固态、液态的降糖原料药进行 NIR 定性分析, 从表 1 中可以看出, 固态模型的 S 均大于 149.161 , 远大于 2 , 说明该模型可将固态原料完全地分离; 从表 2 中可以看出, 有 6 种溶液的 S 大于 2 , 另外 2 种的 S 在 1 和 2 之间, 这说明格列吡嗪、格列喹酮的甲醇溶液应用近红外光谱法鉴别结果并未得到完全的分离, 可能是因为 2 种降糖原料药在甲醇中的溶解度有限, 无法达到一定的浓度, 甲醇溶剂在 $9\,000 \sim 4\,000 \text{ cm}^{-1}$ 都有很强的吸收, 被检测组分浓度过低则受到甲醇吸收峰的干扰更为显著, 无法体现特征信号, 致使计算结果不如固态模型的好. 因此, 相比固态降糖原料药模型, 液态降糖原料药的定性模型结果较差.

表 1 固态降糖原料药定性模型建立结果

Tab. 1 The qualitative model results of solid oral hypoglycemic APIs

No.	组一	组二	T ₁	T ₂	D	S
1	盐酸二甲双胍	格列齐特	4.518 83	1.533 26	902.734	149.161
2	盐酸吡格列酮	盐酸二甲双胍	0.413 54	4.518 83	1 001.340	203.014
3	格列吡嗪	盐酸二甲双胍	0.259 20	4.518 83	898.254	187.997
4	格列齐特	盐酸二甲双胍	1.533 26	4.518 83	902.734	149.161
5	格列美脲	盐酸二甲双胍	0.058 23	4.518 83	999.038	218.271
6	格列喹酮	盐酸二甲双胍	0.139 79	4.518 83	969.049	208.012
7	瑞格列奈	盐酸二甲双胍	0.256 99	4.518 83	935.703	195.925

表 2 液态降糖原料药定性模型建立结果

Tab. 2 The qualitative model results of liquid oral hypoglycemic APIs

No.	组一	组二	T ₁	T ₂	D	S
1	格列本脲	格列吡嗪	0.022 61	0.006 66	0.148	5.056
2	格列齐特	格列吡嗪	0.003 43	0.006 66	0.022	2.180
3	盐酸二甲双胍	格列吡嗪	0.009 08	0.006 66	0.115	7.306
4	盐酸吡格列酮	格列吡嗪	0.002 17	0.006 66	0.025	2.831
5	格列吡嗪	格列喹酮	0.006 66	0.001 85	0.017	1.998
6	瑞格列奈	格列吡嗪	0.008 10	0.006 66	0.107	7.249
7	格列美脲	格列吡嗪	0.002 00	0.006 66	0.111	12.817
8	格列喹酮	格列吡嗪	0.001 85	0.006 66	0.017	1.998

2.3 模型的验证

8 种固态降糖原料药定性分析模型的验证：取 8 种原料药样品，利用已经建立好的模型对未知样品进行近红外定性分析，当匹配值（测试光谱图与参考光谱图之间的光谱图距离）小于相应的阈值，而且所有其他物质的匹配值都大于相应的阈值时，说明被成功鉴定。以一种降糖药为例，分析结果如表 3 所示，未知样品 1 只有与盐酸二甲双胍匹配时匹配值小于阈值，因此被鉴定为盐酸二甲双胍。

表 3 未知样品 1 的定性评价结果

Tab. 3 The qualitative evaluation result of unknown Sample 1

序号	D	T	匹配样品组
1	0.195 22	0.450 23	盐酸二甲双胍
2	691.090 12	0.267 51	盐酸吡格列酮
3	844.720 96	0.368 76	格列吡嗪
4	848.343 19	0.007 78	格列齐特
5	955.186 22	1.542 99	格列美脲
6	969.975 79	0.296 77	格列喹酮
7	1 000.892 45	4.517 16	瑞格列奈
8	1 023.778 65	6.445 34	格列本脲

8 种降糖原料药溶液的近红外分析模型的验证方法同上，未知样品均被鉴定成功，说明此定性模型可以对 8 种降糖原料药进行定性分析。

3 结 论

应用近红外光谱法结合第一范围标定法，建立了 8 种降糖原料药（盐酸二甲双胍、盐酸吡格列酮、格列吡嗪、格列齐特、格列美脲、格列喹酮、瑞格列奈、格列本脲）的定性模型，使用固体光纤及液态透射方式采集谱图，分别在 $12\ 072.3 \sim 4\ 000.0\ \text{cm}^{-1}$ 、 $10\ 235.7 \sim 4\ 237.2\ \text{cm}^{-1}$ ，建立 NIR 8 种降糖原料药固态及液态的定性分析模型，两个模型对未知原料药的鉴别结果准确率均为 100%，固态分析模型得到的鉴别结果要好于液态模型。结果表明所建立的模型可以实现对 8 种降糖原料药快速准确的鉴别，比传统检测方法更加快速、简便，适用于制药企业在原料进厂、车间投料前逐桶定性鉴别，扩大了 NIR 在生产过程前端的应用，为降糖原料药质量控制提供简便、可靠的过程分析技术。

参 考 文 献：

- [1] DIEHL B, GROUT B. NIR spectroscopy— just one of many analytical tools for PAT [J]. *American Pharmaceutical Review*, 2011, 14(2):70-74
- [2] RAJAN M G, BHATTACHARJEE H, BEDI S, et al. Primer on NIR spectroscopy as a PAT tool in the tablet manufacture process [J]. *American Pharmaceutical Review*, 2010, 13(6):128-135

- [3] MAERKJ, ANDRE M, KARNER M, *et al.* Prospects for multivariate classification of a pharmaceutical intermediate with near-infrared spectroscopy as a process analytical technology (PAT) production control supplement [J]. **European Journal of Pharmaceutics and Biopharmaceutics**, 2010, **76**(2): 320-327
- [4] WATARI M. Applications of near-infrared spectroscopy to process analysis using Fourier transform spectrometer [J]. **Optical Review**, 2010, **17**(3):317-322
- [5] WU Hui-quan, KHAN M A. Quality-by-design (QbD): An integrated process analytical technology (PAT) approach for real-time monitoring and mapping the state of a pharmaceutical coprecipitation process [J]. **Journal of Pharmaceutical Sciences**, 2010, **99**(3):1516-1534
- [6] BREARLEY A M, FOULK S J. Near-infrared spectroscopy (NIR) as a PAT tool in the chemical industry: added value and implementation challenges. Chapter 15 [M] // 2nd ed. **Process Analytical Technology**. West Sussex: Wiley, 2010:493-520
- [7] 孟 昱, 汪 晴, 蔡 蕊, 等. 保健品中枸橼酸西地那非成分近红外快速定性定量分析 [J]. 大连理工大学学报, 2010, **50**(6):867-870
(MENG Yu, WANG Qing, CAI Rui, *et al.* Quantitative and qualitative analyses of Sildenafil Citrate in healthy products using near infrared spectroscopy [J]. **Journal of Dalian University of Technology**, 2010, **50**(6):867-870)
- [8] 孟 昱, 王世盛, 蔡 蕊, 等. NIR 对真伪枸橼酸西地那非片快速定性定量分析 [J]. 大连理工大学学报, 2010, **50**(1):15-19
(MENG Yu, WANG Shi-sheng, CAI Rui, *et al.* Quantitative and qualitative analyses of true or counterfeit Sildenafil Citrate tablets using near infrared spectroscopy [J]. **Journal of Dalian University of Technology**, 2010, **50**(1):15-19)
- [9] FRANCH-LAGE F, AMIGO J M, SKIBSTED E, *et al.* Fast assessment of the surface distribution of API and excipients in tablets using NIR - hyperspectral imaging [J]. **International Journal of Pharmaceutics**, 2011, **411**(1-2):27-35
- [10] BARNES S E, THURSTON T, COLEMAN J A, *et al.* NIR diffuse reflectance for on-scale monitoring of the polymorphic form transformation of pazopanib hydrochloride (GW786034): Model development and method transfer [J]. **Analytical Methods**, 2010, **2**(12):1890-1899
- [11] 陆婉珍, 褚小立. 近红外光谱(NIR)和过程分析技术(PAT)[J]. 现代科学仪器, 2007(4):13-17

Identification of oral hypoglycemic APIs by NIR spectroscopy with scaling to the first range method

CAI Rui¹, WANG Shi-sheng¹, MENG Yu¹, MENG Qing-gang², ZHAO Wei-jie^{*1}

(1. School of Pharmaceutical Science and Technology, Dalian University of Technology, Dalian 116024, China;
2. School of Pharmacy, Shenyang Pharmaceutical University, Shenyang 110016, China)

Abstract: The aim of this research is to develop a rapid, accurate and no sample preparation required method to identify 8 kinds of oral hypoglycemic active pharmaceutical ingredient (API) (solid and liquid) by using near infrared (NIR) spectroscopy. 64 NIR spectra are utilized to establish the calibration model and rest 32 spectra are utilized to validate the performance of the model. On the original data, the first derivative combined with scaling to the first range method is conducted as the optimum pre-processing method at 12 072.3-4 000.0 cm^{-1} , and 8 oral hypoglycemic APIs can be well identified by this model. With this approach, no "error determinations" are assigned to a wrong class, which indicates that the NIR method is appropriate for fast identification of oral hypoglycemic APIs in the inspection and acceptance before manufacture and warehousing of drug products in the pharmaceutical industry.

Key words: near infrared spectroscopy; oral hypoglycemic APIs; qualitative analysis; scaling to the first range method; process analytical technology