

卷烟主流烟气中巴豆醛在线光谱分析方法研究

秦云华^{1,2}, 高磊³, 李超¹, 龙雨蛟⁴, 朱明⁴, 陈达^{2*}

1. 云南中烟工业有限责任公司技术中心, 云南 昆明 650023
2. 天津大学精密仪器与光电子工程学院, 天津 300072
3. 中国昆仑工程有限公司吉林分公司, 吉林 吉林市 132000
4. 红云红河烟草(集团)有限责任公司, 云南 昆明 650231

摘要 卷烟主流烟气是卷烟燃烧时被人体吸食到体内的主要气体, 其减焦降害已成为全社会高度关注的问题。在各种卷烟主流烟气组分中, 巴豆醛以其强烈的基因毒性, 成为国家规定的卷烟中七种主要有害指标物之一。传统的巴豆醛分析方法大都采用高效液相色谱法等实验室分析方法, 需繁琐的样品前处理过程, 无法测量巴豆醛的实时浓度, 难以准确评估巴豆醛对人体健康的影响。为了快速、准确地检测卷烟主流烟气中的巴豆醛组分, 本研究搭建了一套可以直接与吸烟机耦合的傅里叶红外光谱分析系统(FTIR), 并创新性开发过采样数据驱动光谱分析方法(ODDSA), 从复杂、变动的卷烟主流烟气中准确提取巴豆醛的光谱组分信息。ODDSA方法从实验设计入手, 采用随机设计的思路尽可能模拟实际卷烟样品的分布范围, 以构建具备良好光谱数据结构的样品集。在此基础上, 创新性地将高密度小波变换引入红外光谱数据的处理过程中, 以时/频双域过采样的方式提升了光谱解析分辨率, 进而降低了其他基质组分对巴豆醛光谱信息的干扰。最后, 发展改良竞争自适应重加权采样方法, 从多倍冗余的高密度小波系数中准确提取待测物质的最佳变量组合, 由此构建高质量的巴豆醛光谱定量分析模型。为了验证ODDSA方法的有效性, 实验中采集了15种典型市售卷烟品牌, 每个品牌在线采集8支样品的主流烟气红外光谱, 随后采用随机挑选的25个验证集样本对ODDSA方法进行验证。结果表明, 检验集的线性拟合系数为0.971, 相对均方根误差为5.5%, 其预测精度能有效满足卷烟主流烟气中巴豆醛的在线分析需求, 并可拓展到环境二手烟气中其他组分的在线监测, 进而为吸烟与健康评估提供全新手段。

关键词 过采样数据驱动光谱分析方法; 卷烟主流烟气; 巴豆醛; 在线分析

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)08-2450-05

引言

在卷烟燃烧时, 卷烟主流烟气经由过滤棒直接吸入人体口腔, 其包含的化学有害物质将直接影响人体健康^[1]。在主流烟气的各类有害组分中, 巴豆醛以其强烈的基因毒性, 被公认为卷烟主流烟气中七种代表性有害成分之一, 属于三类致癌物质, 并直接与人体呼吸系统接触。大量研究表明, 人体呼吸系统对巴豆醛特别敏感, 长期接触巴豆醛易损伤口腔粘膜、气管粘膜和肺泡等器官, 引起疾病, 甚至可能导致肺癌, 其致病性与巴豆醛浓度密切相关^[2]。因此, 如何精准检测主流烟气中的巴豆醛组分含量, 成为当前分析科学、环境

科学等领域的前沿热点。

当前卷烟主流烟气中的巴豆醛浓度分析主要依赖于高效液相色谱法、气相色谱-质谱联用法等实验室分析方法。El-dridge等采用剑桥滤片中截留的巴豆醛, 通过色谱分析卷烟逐口主流烟气中的巴豆醛释放量^[3]。Zhang等采用顶空气相色谱-质谱联用法同时检测主流烟气中的巴豆醛、甲醛等四种醛类^[4]。这些分析方法均需要进行捕集、衍生化、分离等一系列繁琐的样品前处理过程, 难以有效满足在线分析的要求。由于卷烟烟气的形成是一个高度动态过程, 其含有大量的自由基等活泼成分与巴豆醛等组分进行二次反应, 导致高效液相色谱法只能对陈化烟气及其解体的部分化学成分进行分析, 无法得到与人体吸入时相仿的巴豆醛组分动态特

收稿日期: 2020-08-14, 修订日期: 2020-12-29

基金项目: 国家重点研发计划项目(2018YFF01011701), 国家自然科学基金项目(21973111, 61378048)资助

作者简介: 秦云华, 1969年生, 云南中烟工业有限责任公司技术中心研究员 e-mail: yunhuaqin@126.com

* 通讯作者 e-mail: dachen@tju.edu.cn

征,难以准确评估巴豆醛对人体和环境的潜在影响^[5]。为了进一步提升巴豆醛的实时分析精度,迫切需要发展新型、高效的主流烟气成分在线检测方法。

在各种检测技术中,红外光谱分析方法以其简单、快速、灵敏等优点成为气体在线分析的最优选技术之一,并开始得到广泛的应用。Lin 等采用开放式红外光谱技术检测农田上方的痕量温室气体^[6]。管林强等采用红外可调谐激光测量二硫化碳在 2 180.5~2 180.74 cm^{-1} 主要四条中红外吸收谱^[7]。有研究采用傅里叶红外光谱技术在线监测固定污染源中的挥发性有机组分,其灵敏度优于 1.6 ppm^[8]。大量的文献分析结果表明,红外光谱技术具备痕量气体分析的能力,有望应用于卷烟主流烟气的在线分析中。

本工作尝试将傅里叶变换红外(Fourier transform infrared, FTIR)光谱系统与直线型吸烟机系统进行耦合设计,并采用蠕动泵模拟人体的抽吸流量,在吸烟产生主流烟气的同时采集其红外光谱。在此基础上,发展过采样数据驱动光谱分析方法(oversampling data driven spectral analysis, ODDSA),从复杂、变动的主流烟气红外光谱中准确提取巴豆醛组分的特征光谱,由此构建多元定量分析模型,实现了巴豆醛组分的在线检测。相关技术为卷烟减焦除害和环境污染评估提供了一种新手段。

1 实验部分

1.1 仪器与参数

吸烟机耦合 FTIR 光谱分析系统采用模块化设计思路,其主要核心部件包括:直线型吸烟机、FTIR 光谱仪、怀特气体池、双阀聚氟乙烯采样袋、隔膜气泵和蠕动泵等部件。当卷烟在直线型吸烟机燃烧后,所产生的烟气通过双阀聚氟乙烯采样袋进行预混合,通过控制蠕动泵的吸气速度,将主流烟气匀速吸入仪器气体池内以实现红外光谱的测量。在该系统中,直线型吸烟机(SM450,英国 Cerulean 公司)采用 ISO 抽吸模式产生主流烟气,即单次抽吸容量为 35 mL,单次抽吸持续时间为 2 s,抽吸频率为 60 s。

在实验中,FTIR 光谱仪(Tensor II, Bruker, 德国)的参数设置:采集范围 800~4 000 cm^{-1} ,分辨率为 2 cm^{-1} ,扫描次数为 16 次,每个样品采集 10 次,取平均值作为该样品的光谱值。自制怀特气体池光程长为 0.8 m,主腔体尺寸为 52 mm×40 mm×52 mm,容积为 53 mL,光路在腔体内反射 8 次。双阀聚氟乙烯采样袋体积为 50 mL,隔膜气泵额定流量为 1.5 L·min⁻¹,蠕动泵的设定流量为 50 mL·min⁻¹。在实验过程中,实验环境温度控制在 23 °C,湿度控制在 60% RH,实验室内人数限制 1~3 人。

1.2 样品

实验选取 15 种不同牌号的市售中支卷烟,每个牌号采集 8 支样品,合计 120 支卷烟样品。在实验中,按照烟草行业标准 YC/T 255—2008 对样品主流烟气中的巴豆醛含量进行分析,相关样品由云南中烟责任有限公司提供。在建模过程中,随机选择 95 个样品为校正集,剩余 25 个样品为验证集,以验证建模分析效果。

1.3 光谱采集与数据分析

对每个样品连续采集 20 次光谱,取平均后作为单支卷烟主流烟气的采集光谱,共采集 15 种牌号卷烟。每种牌号各采集 8 支卷烟的主流烟气光谱,其红外光谱如图 1 所示。

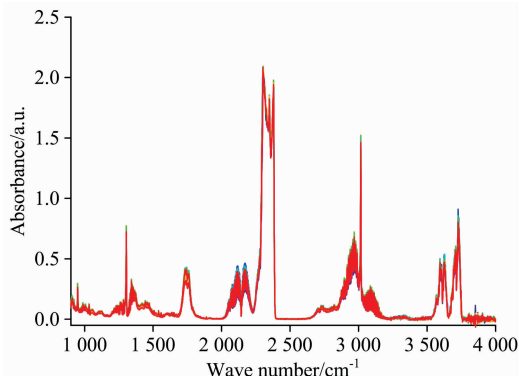


图 1 15 种卷烟主流烟气 FTIR 全谱图

Fig. 1 FTIR spectra of mainstream smoke from 15 kinds of cigarettes

1.4 过采样数据驱动光谱分析方法

卷烟主流烟气是一种极为复杂的气相混合体系,其包含着数以千计的气体组分,极有可能掩盖巴豆醛组分的红外光谱信息^[9]。为了从高度重叠的红外光谱中准确提取巴豆醛的特征信息,提出一种过采样数据驱动光谱分析方法,其主要步骤如下:(1)采用随机设计的思路尽可能采集代表性样本,以模拟实际卷烟样品的分布范围,最大限度保证各组分的浓度向量之间呈随机分布;(2)采用高密度离散小波变换法(higher-density discrete wavelet transform, HDWT)对主流烟气红外光谱数据进行时域/频域的双倍过采样,获得 HDWT 系数;(3)发展改良竞争自适应重加权采样方法(modified competitive adaptive reweighted sampling, MCARS),从 HDWT 系数中准确选择与巴豆醛信息相关的变量;(4)采用偏最小二乘法(partial least square, PLS)构建巴豆醛的多元校正模型,用于后续巴豆醛组分浓度的预测。

ODDSA 算法的核心原理:

1.4.1 高密度离散小波变换法

HDWT 算法是一种基于过采样策略的第二代小波变换,在提升待测信号的采样分辨率的同时,也能有效防止信号的畸变^[10]。与传统的离散小波变换相比,HDWT 算法在变换过程中对光谱信号的时域/频域均进行了两倍的过采样,有效拓展了复杂重叠谱的时域/频域联合特征^[11]。HDWT 具有间尺度以及近似平移不变性等特点,从而有效提高重叠峰的解析分辨率。

在 HDWT 的多尺度分解中,尺度函数和小波函数可分别由式(1)和式(2)表示

$$\phi(t) = \sqrt{2} \sum_k h_0(k) \phi(2t-k) \quad (1)$$

$$\psi_i(t) = \sqrt{2} \sum_k h_i(k) \phi(2t-k), i = 1, 2 \quad (2)$$

其中 $h_0(k)$, $h_1(k)$ 和 $h_2(k)$ 在 $k \in Z$ 时,是紧支撑的实值滤波器。利用式(1)和式(2)可推导出对偶数小波的希尔伯

特变换对, 当采样倍数为 n 时, 动态范围增加 $\log_2(n)$ 位。HDWT 通过频域和时域的同时过采样, 不仅显著提高了重叠信号的解析分辨率, 同时还保留了细节信息, 为后续主流烟气红外光谱的重叠峰解析提供了有力工具。

1.4.2 改良竞争自适应重加权采样方法

为了准确提取 HDWT 系数中的特征信息, 引入了竞争自适应重加权采样方法^[12]。采用达尔文进化论中的“适者生存”理论, 通过在每次迭代过程中将部分权重较小的变量删除, 对产生的新变量重新分配权重, 并继续迭代直至变量个数减少至设定值为止^[12]。然而, 由于红外光谱的数据量较大, 一张谱图往往包含了数千个变量, 经 HDWT 变换后, 其变量数至少拓展了 3~4 倍以上, 导致 CARS 算法的收敛性和精度均下降。为了进一步提升 CARS 算法的性能, 提出了 MCARS 方法, 其核心在于设计了一种新的权重迭代策略

$$w_i = \frac{s_i |b_i|}{\sum_{i=1}^p s_i |b_i|}, i = 1, 2, 3, \dots, p \quad (3)$$

其中, s_i 为变量 i 的方差, b_i 为变量 i 的回归系数, p 为变量个数。由式(3)可见, 该权重不仅考虑了变量在回归方程中的重要性, 同时也考虑了变量自身的变化程度, 即谱峰变化越大的地方, 变量所包含的信息越重要。因此, 将回归系数与变量方差相结合, 将极大加快 MCARS 算法的收敛性和计算精度。通过 MCARS 算法与 HDWT 算法的有效结合, 最终确定 ODDSA 模型, 实现了卷烟主流烟气中巴豆醛的在线分析。

2 结果与讨论

2.1 HDWT 计算结果

为了有效剥离各种光谱干扰对巴豆醛光谱信息的影响, 采用 HDWT 算法对光谱数据进行预处理。在光谱预处理中, 高密度离散小波的预处理结果与滤波器以及分解尺度参数密切相关。理论上, 滤波器的消失矩越高, 其导数阶数越高, 分解后的光谱信息分辨率越高, 有利于后续的信息提取。选取 3 和 4 νm 这两种小波滤波器来处理原始红外光谱数据, 其消失矩分别为 3 阶和 4 阶, 并比对不同分解尺度下的处理效果。采用留一法交叉验证的均方根误差 (root mean square error of cross validation, RMSECV) 最小化准则作为验证标准, 结果如图 2 所示。在滤波器“3 νm ”和分解尺度 4 的条件下, RMSECV 值最小。

在确定 HDWT 的最佳参数后, 主流烟气的红外光谱经 HDWT 处理后, 将原始变量从 3 038 个拓展到 11 776 个, 变换后的红外光谱如图 3 所示。HDWT 的高扩增性为原始谱图提供了额外的分辨率信息, 但同时存在许多冗余数据, 需要高效的特征选择策略来剔除不重要的系数。本文采取 MCARS 算法进行变量筛选。

2.2 MCARS 计算结果

在 MCARS 变量筛选过程中, 权重低的变量被快速剔除, 之后筛选变量的速度随迭代次数增加而减缓, 并趋于收敛。在迭代过程中, 引入 RMSECV 最低原则, 选择 36 个变

量作为最优的回归子集, 并在此基础上构建巴豆醛的多元校正模型。为了进一步验证变量筛选的合理性, 对相关变量进行了光谱重构, 结果表明, 其特征大都集中在 1 650~1 850 cm^{-1} , 与巴豆醛的特征吸收区域重叠较好。基于该多元校正模型, 即可实现未知卷烟样本的主流烟气中巴豆醛浓度的在线定量分析, 显著提升了卷烟主流烟气关键组分的分析效率和可靠性。

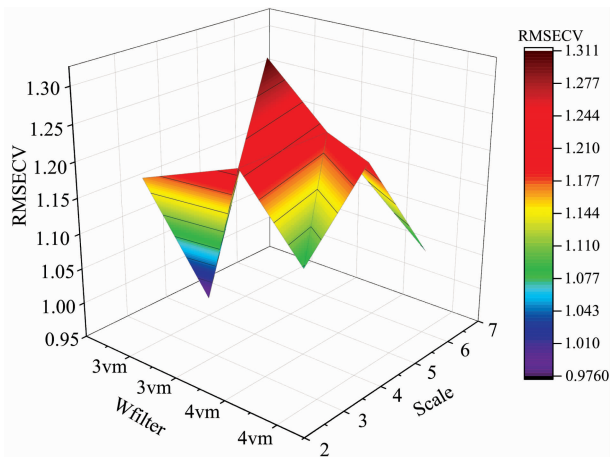


图 2 不同滤波器和分解尺度下的 RMSECV 值
Fig. 2 The RMSECV values obtained with different filters and decomposition scales

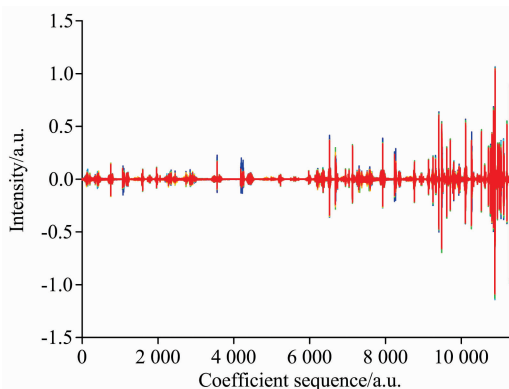


图 3 主流烟气红外光谱的 HDWT 系数图
Fig. 3 The HDWT coefficients for infrared spectra of mainstream smoke

2.3 预测结果

表 1 列出了 ODDSA 方法对主流烟气中巴豆醛浓度的预测结果, 并与其他算法的计算结果进行比对。其中, RMSEP 为检验集的均方根误差, RRMSEP 为 RMSEP 与检验集平均值的比值(代表相对误差)。由表 1 可见, 采用 HDWT 对主流烟气红外光谱进行预处理后, 变量数急剧增加, 导致冗余信息恶化了原有的模型预测精度。因此, 需要对 HDWT 系数进行变量筛选, 才能准确提取巴豆醛的特征信息。

与预期相同, CARS-PLS 和 MCARS-PLS 均在一定程度上提升了巴豆醛模型的定量预测精度。结果表明, ODDSA-PLS 获得了最佳的模型预测结果, 说明在 HDWT 与

表 1 不同建模方法预测结果对比
Table 1 Comparison of prediction results
obtained with differenu methods

建模方法	变量数	PLS 因子数	RMSEP	RRMSEP /%	R ²
PLS	3 038	12	1.81	11.4	0.950
HDWT-PLS	11 776	10	2.02	12.8	0.937
CARS-PLS	52	8	1.73	10.9	0.958
MCARS-PLS	40	7	1.68	10.6	0.960
ODDSA-PLS	36	4	0.87	5.49	0.971

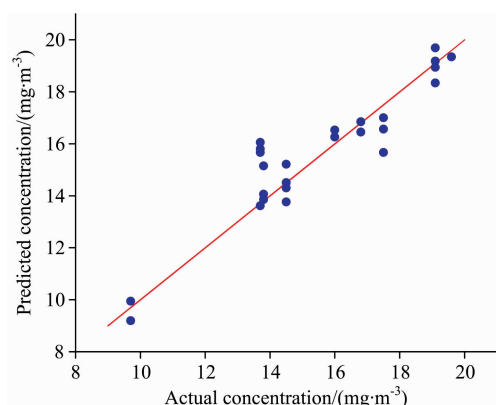


图 4 巴豆醛预测值与真实值拟合曲线

Fig. 4 The prediction results versus measurement results of crotonaldehyde

MCARS 结合后, 更有利于算法以数据驱动的方式准确提取巴豆醛的特征信息, 并有效规避其他复杂基质的干扰。为了进一步说明 ODDSA-PLS 的计算结果, 图 4 列出了 ODDSA-PLS 模型的预测值与真实值拟合曲线。由图可见, 其拟合结果具备良好的线性相关性, 能有效满足主流烟气中巴豆醛成分的定量分析要求。

3 结 论

提出了一种卷烟主流烟气中巴豆醛的在线分析方法, 并搭建了一套可直接与直线型吸烟机耦合的 FTIR 光谱分析系统, 该系统可有效模拟人体的吸烟过程, 在线采集主流烟气的红外光谱信息。在此基础上, 为了准确剥离巴豆醛的特征信息, 开发了 ODDSA 算法。该算法首先从数据结构入手, 采用随机设计的思路尽可能提升样品的代表性, 为后续的数据驱动提供信息引导。然后, 利用 HDWT 算法以过采样的方式极大拓展了红外光谱的信息量, 进而显著提升红外光谱的数据分辨率。并且开发了 MCARS 算法, 从大量冗余的 HDWT 系数矩阵中准确提取巴豆醛的特征信息, 最终构建 ODDSA 光谱定量分析模型。结果表明, ODDSA 是一种高效的红外光谱特征提取方法, 可有效消除光谱数据中其他复杂基质的干扰, 其分析精度能有效满足主流烟气中巴豆醛的在线分析需求, 进而为复杂体系的光谱在线分析提供新手段, 具备良好的普适性。

References

- [1] Liu M Z, Jiang Y, Wedow R, et al. *Nature Genetics*, 2019, 51(2): 237.
- [2] Park S L, Carmella S G, Chen M L et al. *PLOS ONE*, 2015, 10(6): e0124841.
- [3] Eldridge A, Betson T, Gama M V, et al. *Regulatory Toxicology and Pharmacology*, 2019, 107: 104402.
- [4] Zhang X, Wang R, Zhang L, et al. *International Journal of Analytical Chemistry*, 2019, 2105839, doi: 10.1155/2019/2105839.
- [5] Li C, Li E X, Zhang J, et al. *Royal Society Open Science*, 2018, 5(6): 172003.
- [6] Lin C H, Grant R H, Heber A J, et al. *Atmospheric Measurement Techniques*, 2019, 12(6): 3403.
- [7] GUAN Lin-qiang, DENG Hao, YAO Lu, et al(管林强, 邓昊, 姚路, 等). *Acta Physica Sinica(物理学报)*, 2019, 68(9): 125.
- [8] Ding Y S, Yan X Z, Wong J H, et al. *Chemical Research Toxicology*, 2016, 29(1): 125.
- [9] Han X, Tan Z, Huang Z X, et al. *Analytical Methods*, 2017, 9(24): 3720.
- [10] Chen D, Zong J, Huang Z X. *Frontiers in Chemistry*, 2018, 6: 325.
- [11] Ren G X, Wang Y J, Ning J M, et al. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2020, 230: 118079.
- [12] Li Y, Via B K, Li Y X, et al. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2020, 240: 118566.

On-Line Spectral Analysis of Crotonaldehyde Content in Cigarette Mainstream Smoke

QIN Yun-hua^{1,2}, GAO Lei³, LI Chao¹, LONG Yu-jiao¹, ZHU Ming¹, CHEN Da^{2*}

1. Technology Center of China Tobacco Yunnan Industrial Co., Ltd., Kunming 650023, China

2. School of Precision Instruments & Opto-Electronics, Tianjin University, Tianjin 300072, China

3. China Kunlun Engineering Co., Ltd., Jilin Branch, Jilin 132000, China

4. Hongyunhonghe Tobacco Group Co., Ltd., Kunming 650231, China

Abstract The cigarette mainstream smoke represents the main gas that is ingested by the human body when the cigarette is burned. The reduction of scorch and other hazardous components has become an issue of great concern to the whole society. Among various components in cigarette mainstream smoke, crotonaldehyde has become one of the seven main harmful indicators in cigarettes prescribed by the state due to its strong genotoxicity. Traditional analytical methods for crotonaldehyde usually rely on high-performance liquid chromatography and other laboratory methods, which requires complex sample pretreatment procedures. This makes it difficult to measure crotonaldehyde in real-time to evaluate its effects on health. In order to monitor the crotonaldehyde content in cigarette mainstream smoke efficiently, a Fourier Transform Infrared Spectrometer (FTIR) system was set up to a couple with a smoking machine. In this system, an innovative oversampling data driven spectral analysis (ODDSA) method was developed to accurately extract the spectral features of crotonaldehyde from the complex and fluctuating spectra of cigarette mainstream smoke. The ODDSA method started with experimental design and used the idea of random design to simulate the distribution range of actual cigarette samples, which constructed a good data structure to guide further data mining. Thereafter, the high-density wavelet transform (HDWT) was innovatively used to process the IR spectra, which enabled oversampling in time/frequency dual-domains to improve the spectral resolution. This would definitely suppress the effects of other matrix components on the analysis of crotonaldehyde. Finally, the strategy of modified competitive adaptive reweighted sampling was developed to accurately extract the interesting features from the multiple redundant HDWT coefficients, which was used to construct a qualified calibration model for the analysis of crotonaldehyde. In the experiment, 15 typical commercial cigarette brands were collected, in which 8 samples of each brand were prepared to collect their IR spectra of mainstream smoke. Thereafter, 25 samples were randomly selected to validate the performance of ODDSA. The calculation results showed that the regression coefficient of the test set was 0.971, and the relative root means square error is 5.5%. The satisfactory results indicate that the ODDSA is capable of on-line analysis of crotonaldehyde in cigarette mainstream smoke, which may well extend to on-line monitoring of other components in second-hand environmental smoke. This would provide a novel tool for the evaluation of cigarette effects on health.

Keywords Oversampling data driven spectral analysis; Cigarette mainstream smoke; Crotonaldehyde; On-line analysis

(Received Aug. 14, 2020; accepted Dec. 29, 2020)

* Corresponding author