RESEARCH PAPER

Application of untargeted volatile profiling in inflammatory bowel disease research

Natalia Arroyo-Manzanares[1] · María García-Nicolás[1] · Fuensanta Abellán-Alfocea[1] · Laura Prieto-Baeza[1] ·
Natalia Campillo[1] · Blanca del Val Oliver[2] · José Zarauz-García[3] · Luis Sáenz[3] · Pilar Viñas[1]

Abstract
Inflammatory bowel disease (IBD) diagnosis depends on criteria based on histological, endoscopic, radiological, and clinical results. These studies show drawbacks as being expensive, invasive, and time-consuming. In this work, an untargeted metabolomic strategy based on the monitoring of volatile compounds in serum by headspace gas chromatography–mass spectrometry is proposed as a complementary, fast, and efficient test for IBD patient diagnosis. To develop the method and build a chemometric model that allows the IBD diagnosis, serum samples including IBD patients and healthy volunteers were collected. Analyses were performed by incubating 400 µL of serum for 10 min at 90 °C. For data processing, an untargeted metabolomic strategy was used. A total of 96 features were detected, of which a total of 10 volatile compounds could be identified and confirmed by means of the analysis of real standards. The chemometric treatment consisted of a discriminant analysis of orthogonal partial least squares (OPLS-DA) obtaining a 100% of classification rate, since all the analyzed samples were correctly classified.

Keywords Inflammatory bowel disease · Headspace · Gas chromatography–mass spectrometry · Untargeted metabolomic strategy · Chemometrics

Introduction

Inflammatory bowel disease (IBD) is a chronic, multifactorial immune disorder characterized mainly, but not exclusively, by inflammation of the intestine. In general, IBD encompasses two different pathologies: ulcerative colitis (UC) and Crohn's disease (CD). Clinically, similar symptoms are shared for both diseases, such as bloody diarrhea, abdominal pain, weakness, and weight loss, while they tend to differ in the complications and prevalence, as well as the location and depth of inflammation [1, 2]. In cases where it is impossible to differentiate between UC and CD, since patients show overlapping pathological features of UC and CD, it often results in an interim diagnosis of unclassified IBD or indeterminate colitis (IC). There is a fourth entity, named microscopic colitis (MC), often misdiagnosed and confused with irritable bowel syndrome by not producing obvious macroscopic lesions.

There is no specific test for IBD diagnosis. In addition, the disease occurs in a fluctuating way, with periods of inflammatory flare-ups alternating with other periods of inflammatory remission. First, clinical symptoms are examined, including anemia, abdominal pain, weakness, weight loss, and diarrhea with blood and/or mucus. In general, the clinical suspicion of IBD prompts the practice of blood and stool tests, colonoscopy, and some imaging studies that confirm the diagnosis excluding other causes.

Analysis of blood can show an elevated white blood cell count and elevated C-reactive protein (CRP) levels. CRP is a plasma protein synthesized in the liver that can be used as a marker for indirect determination of the degree of inflammation in IBD. Under normal circumstances, CRP

✉ Pilar Viñas
pilarvir@um.es

1 Department of Analytical Chemistry, Faculty of Chemistry, University of Murcia, Regional Campus of International Excellence "Campus Mare Nostrum", E-30100 Murcia, Spain

2 Internal Medicine Service - Gastroenterology and Hepatology Section, Hospital General Universitario Rafael Méndez, Lorca, Spain

3 Laboratory Medicine Department, Hospital General Universitario Rafael Méndez, Lorca, Spain

is synthesized in small amounts (< 1 mg L^{-1}); however, because of some stimuli, usually inflammation, hepatocytes rapidly increase their synthesis [2, 3].

Regarding stool analysis, calprotectin is an abundant and widely distributed protein, mainly found in monocytes, reactive macrophages, and polymorphonuclear leukocytes. Its biological function is not exactly known, but its protective activity in inflammatory, proliferative, and infectious processes stands out, in which its plasma levels are increased from 8 to 40 times. Therefore, calprotectin determination in feces has been recently proposed as a new marker for the diagnosis of IBD [4, 5], which can be used to predict the risk of relapse or monitor response to treatment. In addition, it offers greater sensitivity and specificity than CRP.

An elevated level of calprotectin (> 50 μg g^{-1}) identifies patients who are more likely to have IBD and who should undergo colonoscopy. In contrast, a fecal calprotectin level < 50 μg g^{-1} makes the diagnosis of IBD highly unlikely. Colonoscopy allows direct visualization of the intestinal mucosa and obtaining biopsies, making it the preferred test for IBD diagnosis. However, it has some risks and limitations, since it is an invasive procedure with relatively high cost and requires bowel preparation and patient sedation. Furthermore, it is estimated that more than 60% of colonoscopies performed in young patients are potentially preventable since their results are normal [6].

In addition, disease status often needs to be reassessed with complementary explorations that can safely, quickly, and reliably detect the presence of inflammation. In the case of radiological tests, in patients with CD, magnetic resonance imaging with bowel contrast is preferable to computerized tomography scan since its diagnostic precision is greater and avoids frequent exposure to radiation.

Thus, the diagnosis of IBD depends on criteria based on the results of clinical, endoscopic, histological, and radiological examinations, which are often costly, invasive, and time-consuming. Therefore, low-cost, rapid, and non-invasive diagnostic tools are needed. In recent years, metabolomics approaches have been showing promising results in the diagnosis of IBD. The metabolomic analysis consists of the exhaustive and quantitative study of the metabolome, understanding this as the complete set of small molecules called metabolites that are synthesized by a biological system [7]. Sample selection in metabolomics studies is key, as each type of sample provides different biochemical information: blood can provide useful information on systemic metabolism, while fecal profiles are indicative of digestive metabolism, and urine provides a profile of endogenous metabolism. Due to the chemical diversity of metabolites, the choice of analytical technique will also be decisive in the metabolomic study.

To date, proton nuclear magnetic resonance spectroscopy (^{1}H-NMR) is the most widely used technique in metabolomics studies for IBD diagnosis. Studies have been conducted on biofluids such as plasma, serum, feces, and urine from patients with this disease and have been compared with IBD-healthy individuals. These studies have focused primarily on non-complex and small molecules such as amino acids or derivative metabolites. It has been reported that there are differences in metabolic profiles between IBD patients and healthy controls [8–13], as well as between IBD subtypes [8, 9, 13].

Liquid chromatography with tandem mass spectrometry (LC-MS/MS) has also been applied for monitoring the inflammation lipid mediator leukotriene E4 in plasma [14], and the lipoid profile in urine [15] as biomarkers of IBD activity. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) has also been applied for the analysis of fecal samples in patients with CD [16]. Recently, the potential of LC coupled to high-resolution mass spectrometry (HRMS) has also been investigated, demonstrating that there is a loss of "metabolic diversity" among IBD patients since gut metabolites were frequently depleted, thus leading to alterations in immune response and cell proliferation, among other biological processes [17].

Gas chromatography coupled to mass spectrometry (GC-MS) has allowed the analysis of volatile metabolites in serum [18] or feces [19] as an assessment approach to IBD or its subtypes. Specifically, Kohashi et al. [18] mainly targeted water-soluble metabolites for UC diagnostic, reporting that the components of the urea cycle (citrulline, ornithine, and urea) and the tricarboxylic acid cycle (fumaric acid, succinic acid, and malic acid) showed the most significant alterations, and all of them except urea exhibited a significant decrease of serum levels in the UC patients. In addition, significant decreases in serum levels of several amino acids such as isoleucine, lysine, histidine, leucine, and methionine have been detected in UC patients. Ahmed et al. [19] monitored between 234 and 290 metabolites to discriminate inactive and active CD or UC. The statistically significant metabolites in separating the groups were broadly classified into aldehydes, ketones, secondary alcohols, esters, and short and branched-chain fatty acids. More recent studies focus on the analysis of fecal [20] and exhaled volatile organic compounds (VOCs) [21] using GC coupled to ion mobility spectrometry (IMS).

The goal of this work is to investigate a totally untargeted metabolomic approach based on the total VOC profile obtained by GC-MS for the diagnosis of IBD. The novelty includes that a headspace (HS) analysis is proposed as a sample treatment of serum and introduction system, avoiding more complex procedures proposed to date, which are based on solid-phase microextraction (SPME) [18] or the combination of liquid-liquid extraction and lyophilization and derivatization steps [17]. On the other hand, this is the first study carried out in a totally untargeted mode.

Material and methods

Instrumentation and software

HS-GC-MS analyses were carried out on an 8890 gas chromatograph (Agilent Technologies, Santa Clara, CA, USA) equipped with an HS system (Gerstel, Mülheim, Germany) and coupled to a mass spectrometer (5977B) quadrupole mass selective detector (Agilent Technologies) with inert ion source based on electron impact (EI). The chromatographic separation was carried out on an HP-5MS capillary column (30 m × 0.25 mm inner diameter × 0.25 μm film thickness), also from Agilent Technologies.

Analysis and data acquisition were carried out using the MSD Chemstation Data Analysis application (Version G1701EA), and data were processed using MS-DIAL (Version 4.80, RIKEN) and SIMCA-P (Umetrics, Malmö, Sweden).

Standards and reagents

A total of 10 analytical standards were used for the identification of volatile compounds in serum samples: butanal, pentanal, toluene, hexanal, ethylbenzene, p-xylene, heptanal, 1-ethyl-3-methylbenzene, 1,3,5-trimethylbenzene, and 2-ethyl-1-hexanol. All of them were supplied by Sigma-Aldrich (St. Louis, MO, USA) and individual stock solutions were prepared at 1000 μg mL^{-1} in methanol (MeOH). All standard solutions were stored at −20 °C.

High-quality MeOH was provided by ChemLab (Zedelgem, Belgium). The carrier gas used was helium with 99.99% purity and was provided by Air Liquide (Madrid, Spain).

A mixture of alkanes (from C8 to C40) at 500 μg mL^{-1} in dichloromethane was also supplied by Sigma-Aldrich. The alkane mixture was prepared at 1 μg mL^{-1} in hexane and used as quality control throughout the analyses.

HS-GC–MS method

A volume of 400 μL of sample was incubated at 90 °C for 10 min, while being stirred at 750 rpm. Then, 1.5 mL of the headspace was injected into the GC system using a 2.5 mL syringe (90 °C) in splitless mode. The carrier gas was helium with a flow rate of 1 mL min^{-1}. The program oven was as follows: initial temperature of 40 °C held for 5 min, which was increased to 130 °C at 5 °C min^{-1} and subsequently to 200 °C at 35 °C min^{-1} (total run 25 min). The MS was operated in EI mode at 70 eV. Temperatures of the ion source, transfer line, and quadrupole were set at 230, 280, and 150 °C, respectively. Data acquisition was carried out in the range of 20–400 m/z.

Samples

The ethics committees of Murcia University (Favourable Report ID: 2908/2020) and Rafael Méndez University Hospital (Lorca, Spain) approved this study. Informed consent was obtained from all patients and volunteers, and samples were used in accordance with the hospital guidelines.

A total of 56 patients with IBD and 48 volunteers with no suspicion of IBD disease were recruited from the Rafael Méndez University Hospital. IBD diagnosis was established by endoscopic, histological, and radiological criteria. Since IBD is a very broad disease, the patients who participated in this study were selected with different clinical pictures to cover a wide spectrum of the disease.

Table 1 shows the clinical and demographic characteristics of the study population. The study cohort comprised 55% females in the IBD-healthy volunteer's group and 64% females in patients with IBD. The mean age of cohort patients with IBD disease and IBD-healthy volunteer was 48.9 and 43.0, respectively. Table 1 also shows CRP, calprotectin, and other parameters such as cholesterol, Fe, and hemoglobin (HGB), since they had been previously described as abnormalities exhibited by patients with IBD [22, 23], although no significant differences were found between the two groups. It should be noted that some IBD-healthy volunteers showed high CRP, suggesting another inflammatory process, not related to IBD.

IBD patients who participated in the study received a wide variety of treatments, including those specific to the inflammatory disease such as aminosalicylates (mesalazine), thiopurines (azathioprine), corticosteroids (budesonide, beclomethasone dipropionate, prednisone, and methylprednisolone), and biological (anti-TNF, anti-interleukins IL12/23, anti-integrins, JAK inhibitors…) as well as mineral (iron or calcium) and vitamin (folic acid, cyanocobalamin, calciferol) supplements. In the case of the IBD-healthy volunteers, some of them also had some type of medication including antihypertensives, antibiotics, analgesics, antipsychotics, or antidepressants.

Each blood sample was collected in a special tube for blood collection. This tube contains a separating gel that allows the cells to be separated from the blood, obtaining serum as the final result after being centrifuged at 3000 rpm for 10 min at room temperature. Then, the serum was transferred to a clean tube and stored at −20 °C until use.

For their analysis, serum samples were tempered at room temperature for approximately half an hour and vigorously vortexed for 2 min for sample homogenization. Then, 400 μL of serum was placed into a 20 mL glass vial and hermetically

Table 1 Clinical and demographic characteristics of the study population

	IBD-healthy volunteers	Patients with IBD	*P*-value*
Number of participants	45	56	
Gender			
Male, *n* (%)	20 (45)	20 (36)	
Female, *n* (%)	25 (55)	36 (64)	
Age (years)			
Median (IQR)	43.0 (23.0–63.4)	48.9 (38.2–57.1)	0.250
Cholesterol (mg/dL)			
Median (IQR)	186.0 (161.5–208.0)	192.5 (170.5–215.0)	0.810
Missing (%)	11	5	
Fe (µg/dL)			
Median (IQR)	86.3 (83.0–100.5)	84.8 (56.0–100.0)	0.740
Missing (%)	51	64	
HGB (g/dL)			
Median (IQR)	13.6 (13.0–14.5)	14.1 (13.3–15.1)	0.584
Missing (%)	7	14	
LEU (mg/dL)			
Median (IQR)	6.2 (6.1–7.6)	7.5 (5.7–8.6)	0.571
Missing (%)	9	18	
CRP (mg/dL)			
Median (IQR)	0.5 (0.1–0.8)	0.7 (0.1–0.6)	0.280
Missing (%)	16	2	
CAL (µg/g)			
Median (IQR)	88.6 (68.0–127.0)	297.4 (29.5–287.4)	0.350
Missing (%)	10	3	

HGB hemoglobin, *LEU* leucocytes, *CRP* C-reactive protein, *CAL* calprotectin, *IQR* interquartile range

*Both groups were compared using a *t* test

sealed with 18 mm aluminum screw caps provided with silicone septum and injected in the HS-GC-MS system.

Data processing

Data from HS-GC-MS were converted to analysis Base Framework (ABF) formats and processed using MS-DIAL, which includes peak picking, deconvolution, compound identification, and peak alignment. The peak identification was carried out using an alkane mix–based retention index by Kovats's method and GC-MS metabolomics MSP spectral library (RIKEN).

The features (including both identified and unidentified compounds) were used to carry out the multivariate study. The chemometric analysis consisted of a discriminant analysis of orthogonal partial least squares (OPLS-DA) using a unit variance (UV) scale and was carried out using SIMCA-P software. The OPLS-DA model was built using 80% of the samples and was externally validated with the remaining 20%. The ellipses for each category using a confidence probability level of 95% were obtained with Excel software using available online algorithms. R2X(cum), R2Y(cum), Q2(cum), and classification rate (CR) were evaluated in order to study the model success. The cross-validation procedure was a sevenfold cross-validation without shuffling. R2X(cum) and R2Y(cum) represent the cumulative fraction of the variance explained by a specific component. Q2(cum) indicates the predictive ability of the chemometric model, which should have a value greater than 0.5 [24]. In addition, validation of the model was carried out using a permutation test using 50 random permutations.

Quality control samples were included at the start and end of an assay run, and at regular intervals throughout the assay (one QC every 5 samples analyzed). QC was a mixture of serum samples with the aim of obtaining a representative sample of the qualitative and quantitative composition of the samples. Principal component analysis (PCA) was carried out to check possible signal deviations or sensitivity loss.

Results and discussion

Optimization of HS-GC–MS method

The optimization of the HS-GC-MS method was carried out with the aim of achieving the best results in terms of

intensity and separation between peaks. Specifically, oven program, volume of serum sample, and temperature and time of incubation were investigated.

Considering that the objective of an untargeted metabolomic method is to obtain as much information as possible, it is essential to achieve a good separation between peaks in the shortest possible time, so the first parameter optimized was the oven program. Among the different temperature programs assayed, the best results were obtained with the following conditions: initial temperature of 40 °C held for 5 min, which was increased to 130 °C at 5 °C min^{-1} and subsequently to 200 °C at 35 °C min^{-1}, thus providing a total run time of 25 min.

The amount of the sample available limited the range of sample volumes investigated. Experiments were carried out using 100, 250, and 400 μL of serum. As expected, using the greater amount of sample, the number of signals and their intensity increased (Fig. 1a). Thus, 400 μL was selected for further experiments.

Next, the effect of the incubation temperature was investigated. Experiments at temperatures between 80 and 100 °C were performed. When the temperature increased from 80 to 90 °C, the intensity of the signals also increased. This is because high temperatures facilitate the release of volatile organic compounds with high boiling points. However, when the temperature increased from 90 to 100 °C, practically no significant differences were observed, for which reason 90 °C was finally selected as the optimum value.

Finally, the incubation time was studied between 5 to 20 min. As can be seen in Fig. 1b, the time range studied did not have a very significant effect on the intensity of the signals obtained. A small increase was observed in the signals that appear at the beginning of the chromatogram when increasing the time from 5 to 10 min. However, after 10 min, no differences were observed. An incubation time of 10 min was therefore selected as optimum.

Metabolic features derived from untargeted metabolomic analysis and peak identification

For data analysis, an untargeted metabolomic strategy was proposed using MS-DIAL. According to the acquisition method, the mass range was established between 20 and 400 m/z. For peak detection, data points were smoothed with a linearly weighted smoothing average using a level of 3 scans and an average peak width of 20 scans. Noise was defined by ion amplitude less than 1000. The deconvolution was carried out to the detected m/z–retention time features, thus features with identical peak widths and retention times were combined into single arrays. In this case, a resolution of 0.5 was set, since higher values could decrease the number of resolved chromatographic peaks and lower values could recognize noise as chromatographic peaks.

For peak identification, an alkane mix–based retention index by Kovats's method was used. The alkane mix (C_8 to C_{40}) was analyzed using the proposed methodology and a dictionary with retention time was created and used for identification purposes. Deconvoluted spectra were matched against a GC-MS metabolomic MSP spectral library from RIKEN. The retention index, retention time, and m/z tolerances were set at 20, 0.5 min, and 0.5 Da, respectively. A match criterion of more than 85% was considered. Finally, alignment was carried out using a retention time tolerance of 0.075 min and spectral similarity tolerance of 70%.

A total of 96 features were detected, of which a total of 10 compounds could be identified and confirmed by means of the analysis of real standards (Fig. 2). Table 2 summarizes the identified compounds.

The presence of butanal, pentanal, toluene, hexanal, ethylbenzene, p-xylene, and heptanal has been previously described in serum human samples [25, 26]. 1-Ethyl-3-methylbenzene has been associated with smokers and has been detected in exhaled breath [27], while 1,3,5-trimethylbenzene and 2-ethyl-1-hexanol have been detected in biological
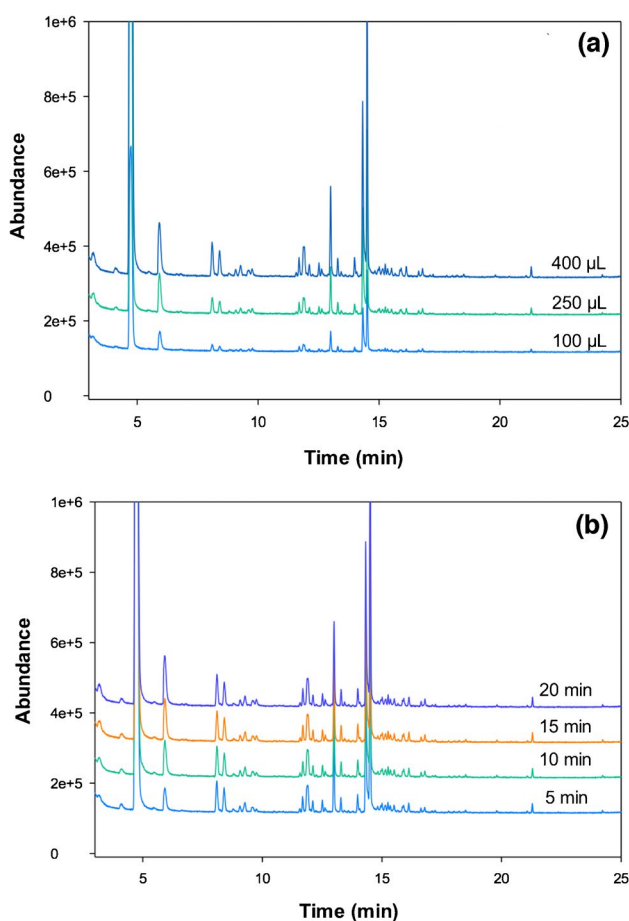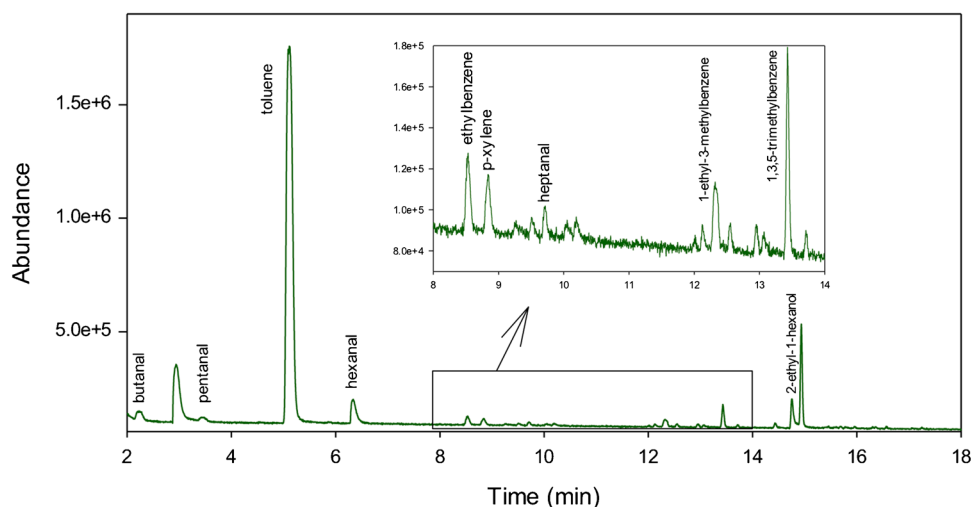

Fig. 1 Optimization of **a** volume of sample, **b** time of incubation

Fig. 2 HS-GC-MS chromatogram of a serum human sample showing the identified compounds

human and rat samples after exposure to these compounds [28, 29].

Subsequently, a statistical study of the identified compounds in the samples was carried out. Therefore, a *t* test was performed using the peak area values of the samples of each category since data fit a normal distribution. Only pentanal (*p*-value = 0.01) and hexanal (*p*-value = 0.01) presented statistically significant differences between the mean from IBD-healthy volunteers and IBD patients at the 95.0% confidence level. However, no significant differences were obtained for butanal, toluene, p-xylene, heptanal, 1-ethyl-3-methylbenzene, 1,3,5-trimethylbenzene, and 2-ethyl-1-hexanol (*p*-value > 0.05).

Chemometric models for the identification of IBD patients

The 96 features (including known and unknown compounds) were used to create the dataset for multivariate analysis, so it had a dimension of 101 (samples) × 96 (features). The 101 samples included both serum samples of IBD-healthy

Table 2 Detected compounds in serum samples

RT (min)	Compound	Ions (*m/z*)
2.3	Butanal	44, 43, 72, 41
3.5	Pentanal	44, 58, 29, 41
5.1	Toluene	91, 92, 77, 65
6.3	Hexanal	56, 44, 41, 43
8.5	Ethylbenzene	106, 91, 65, 77
8.8	p-Xylene	91, 106, 105, 77
9.8	Heptanal	70, 41, 44, 43
12.3	1-Ethyl-3-methylbenzene	120, 105, 91, 77
13.4	1,3,5-Trimethylbenzene	105, 120, 91, 77
14.8	2-Ethyl-1-hexanol	57, 41, 43, 70, 83

volunteers (45 samples) and serum samples of patients with IBD (56 samples).

Initially, the residual normal probability plot was obtained. The residuals were random and normally distributed, since the normal probability plot has all the points lying on a straight line. The normal distribution of the data makes it possible to create the models using the raw data without transformations. In addition, all experimental runs ranged between − 2.5 and + 2.5; therefore, no outliers were detected (Supplemental Figure S1).
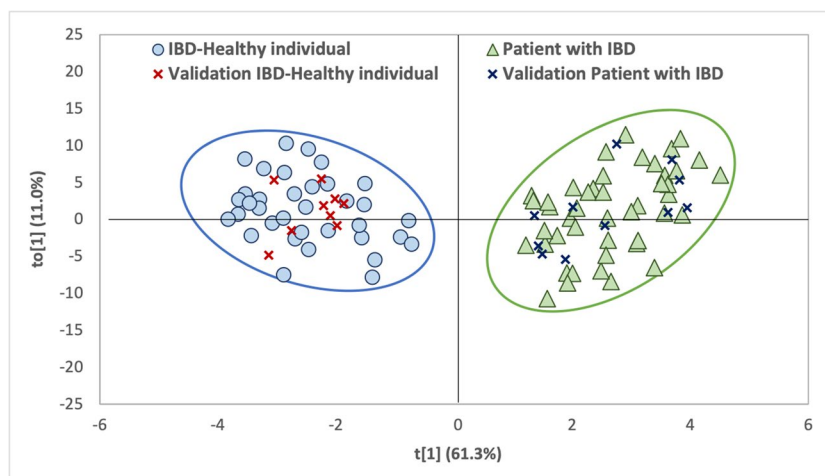
PCA was firstly applied over the training set to visualize any possible grouping of samples, but it was not enough to discriminate the categories (Supplemental Figure S2). Therefore, an OPLS-DA using UV scale or autoscaling, the most applied in metabolomics and that uses the standard deviation as the scaling factor, was proposed [30].

For the construction and validation of the models, the dataset was randomly divided into two subsets. Eighty percent of samples were used for the model optimization (45 serum samples from patients with IBD and 36 from IBD-healthy volunteers) and the remaining 20% for its validation (11 serum samples from patients with IBD and 9 from IBD-healthy volunteers). The obtained chemometric model (1 + 1 component, R2X = 0.828, R2Y = 0.723, and Q2 = 0.543) perfectly separated the samples from patients with IBD and IBD-healthy volunteers as can be seen in Fig. 3. Q2 > 0.5 is usually admitted for good predictability, although it is difficult to give a general limit that corresponds to good predictability, since this strongly depends on the properties of the dataset. For this reason, in order to demonstrate the model predictability, CV-ANOVA was carried out (Supplemental Table S1) and the model has also been validated by a permutation test that consists of comparing the Q2 obtained for the original dataset with the distribution of Q2 values calculated when original Y values are randomly assigned to the individuals [31]. Supplemental Figure S3 shows the permutation

Fig. 3 OPLS-DA score plot for the classification of serum samples of IBD-healthy and IBD patients including validation set (predicted samples). Ellipses for each category using a confidence probability level of 95% are shown

plot for the OPLS-DA model using 50 random permutations. This plot confirms the validity of the model since all Q2 and R2 values to the left are lower than the original points to the right and the regression line of the Q2-points intersects the vertical axis below zero. Moreover, after applying the model to the validation set, a CR of 100% was obtained, since all samples were correctly classified. The OPLS-DA models met all quality criteria; it was proposed due to its ability to model data with noisy and multicollinear variables, such as spectral metabolic data. However, its results were also compared with those obtained using a PLS-DA model (Supplemental Information), demonstrating that the OPLS-DA allows obtaining a better separation between the samples of patients with IBD and IBD-healthy volunteers.

The variable influence on the projections (VIP) plot (Supplemental Figure S4) revealed that hexanal and pentanal were two of the compounds that most influenced the classification of the samples, in agreement with the data obtained from the ANOVA analysis. Further modeling of men and women separately allowed us to evaluate the influence of the confounding factors by means of shared and unique structures (SUS) plot. The SUS plot compared the contribution of volatile compounds to differentiate IBD-healthy females and females with IBD versus IBD-healthy males and males with IBD (Supplemental Figure S5). The majority of features are closely clustered around the diagonal of the SUS plot; therefore the volatile compounds responsible for the differentiation of patients with IB are the same in women and men, which is an added value to the proposed methodology.

Conclusions

This work proposes an analytical strategy based on the use of HS-GC-MS and its combination with chemometric models as a possible complementary tool for the diagnosis of patients with IBD.

The untargeted metabolomic strategy allowed us to obtain 96 features that were used to build an OPLS-DA model that would allow us to separate patients with IBD and IBD-healthy individuals. The external validation did not give rise to false positives or negatives (100% of the classification rate). The results showed a higher concentration of hexanal and pentanal in patients with IBD. In addition, a similar volatile profile was found for female and male IBD patients.

The main advantages of the proposed method are its speed in obtaining reliable results, since it takes less than an hour to apply the HS-GC-MS method and the chemometric model, and the no need for sample treatment, since the volatiles generated in the headspace when heating the sample are monitored.

Although the results obtained demonstrate the potential of this methodology, it should continue to be investigated increasing the number of samples, and in order to demonstrate its applicability for the differentiation of patients with CD or UC and evaluate the influence in the classification of the other inflammatory diseases.

Code availability Not applicable.

Declarations

Ethics approval The ethics committees of Murcia University (Favourable Report ID: 2908/2020) and Rafael Méndez University Hospital (Lorca, Spain) approved this study.

Consent to participate Informed consent was obtained from all patients and volunteers, and samples were used in accordance with the Rafael Méndez University Hospital (Lorca, Spain) guidelines.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

1. Kim DH, Cheon JH. Pathogenesis of inflammatory bowel disease and recent advances in biologic therapies. Immune Network. 2017;17:25–40. https://doi.org/10.4110/in.2017.17.1.25.
2. Wright EK, Ding NS, Niewiadomski O. Management of inflammatory bowel disease. Med J Australia. 2018;209:318–23. https://doi.org/10.5694/mja17.01001.
3. Gisbert JP, González-Lama Y, Maté J. Role of biological markers in inflammatory bowel disease. Gastroenterol Hepatol. 2007;30:117–29. https://doi.org/10.1157/13100073.
4. Sipponen T, Kolho KL. Fecal calprotectin in diagnosis and clinical assessment of inflammatory bowel disease. Scand J Gastroenterol. 2015;50:74–80. https://doi.org/10.3109/00365521.2014.987809.
5. Khaki-Khatibi F, Qujeq D, Kashifard M, Moein S, Maniati M, Vaghari-Tabari M. Calprotectin in inflammatory bowel disease. Clin Chim Acta. 2020;510:556–65. https://doi.org/10.1016/j.cca.2020.08.025.
6. Tavabie OD, Hughes SA, Loganayagam A. The role of faecal calprotectin in the differentiation of organic from functional bowel disorders. Br J Gen Pract. 2014;64:595–6. https://doi.org/10.3399/bjgp14X682525.
7. Castillo-Peinado LS, Luque de Castro MD. Present and foreseeable future of metabolomics in forensic analysis. Anal Chim Acta. 2016;925:1–15. https://doi.org/10.1016/j.aca.2016.04.040.
8. Schicho R, Shaykhutdinov R, Ngo J, Nazyrova A, Schneider C, Panaccione R, et al. Quantitative metabolomic profiling of serum, plasma, and urine by 1H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals. J Proteome Res. 2012;11:3344–57. https://doi.org/10.1021/pr300139q.
9. Williams HRT, Willsmore JD, Cox IJ, Walker DG, Cobbold JFL, Taylor-Robinson SD, et al. Serum metabolic profiling in inflammatory bowel disease. Dig Dis Sci. 2012;57:2157–65. https://doi.org/10.1007/s10620-012-2127-2.
10. Stephens NS, Siffledeen J, Su X, Murdoch TB, Fedorak RN, Slupsky CM. Urinary NMR metabolomic profiles discriminate inflammatory bowel disease from healthy. J Crohn's Colitis. 2013;7:42–8. https://doi.org/10.1016/j.crohns.2012.04.019.
11. Gall GL, Noor SO, Ridgway K, Scovell L, Jamieson C, Johnson IT, et al. Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome. J Proteome Res. 2011;10:4208–18. https://doi.org/10.1021/pr2003598.
12. Bjerrum JT, Nielsen OH, Hao F, Tang H, Nicholson JK, Wang Y, et al. Metabonomics in ulcerative colitis: diagnostics, biomarker identification, and insight into the pathophysiology. J Proteome Res. 2010;9:954–62. https://doi.org/10.1021/pr9008223.
13. Williams HRT, Cox IJ, Walker DG, North BV, Patel VM, Marshall SE, et al. Characterization of inflammatory bowel disease with urinary metabolic profiling. Am J Gastroenterol. 2009;104:1435–44. https://doi.org/10.1038/ajg.2009.175.
14. Stanke-Labesque F, Pofelski J, Moreau-Gaudry A, Bessard G, Bonaz B. Urinary leukotriene E4 excretion: a biomarker of inflammatory bowel disease activity. Inflamm Bowel Dis. 2008;14:769–74. https://doi.org/10.1002/ibd.20403.
15. Fan F, Mundra PA, Fang L, Galvin A, Moore XL, Weir JM, et al. Lipidomic profiling in inflammatory bowel disease: comparison between ulcerative colitis and Crohn's disease. Inflamm Bowel Dis. 2015;21:1511–8. https://doi.org/10.1097/MIB.0000000000000394.
16. Jansson J, Willing B, Lucio M, Fekete A, Dicksved J, Halfvarson J, et al. Metabolomics reveals metabolic biomarkers of Crohn's disease. PLoS ONE. 2009;4:e6386. https://doi.org/10.1371/journal.pone.0006386.
17. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Microbiol. 2019;4:293–305. https://doi.org/10.1038/s41564-018-0306-4.
18. Kohashi M, Nishiumi S, Ooi M, Yoshie T, Matsubara A, Suzuki M, et al. A novel gas chromatography mass spectrometry-based serum diagnostic and assessment approach to ulcerative colitis. J Crohn's Colitis. 2014;8:1010–21. https://doi.org/10.1016/j.crohns.2014.01.024.
19. Ahmed I, Greenwood R, Costello B, Ratcliffe N, Probert CS. Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease. Aliment Pharmacol Ther. 2016;43:596–611. https://doi.org/10.1111/apt.13522.
20. Bosch S, Wintjens DSJ, Wicaksono A, Kuijvenhoven J, der Hulst R, Stokkers P, et al. The faecal scent of inflammatory bowel disease: detection and monitoring based on volatile organic compound analysis. Dig Liver Dis. 2020;52:745–52. https://doi.org/10.1016/j.dld.2020.03.007.
21. Tiele A, Wicaksono A, Kansara J, Arasaradnam RP, Covington JA. Breath analysis using enose and ion mobility technology to diagnose inflammatory bowel disease - a pilot study. Biosensors. 2019;9:55. https://doi.org/10.3390/bios9020055.
22. Agouridis AP, Elisaf M, Milionis HJ. An overview of lipid abnormalities in patients with inflammatory bowel disease. Ann Gastroenterol. 2011;24(3):181–7.
23. Kaitha S, Bashir M, Ali T. Iron deficiency anemia in inflammatory bowel disease. World J Gastrointest Pathophysiol. 2015;6(3):62–72. https://doi.org/10.4291/wjgp.v6.i3.62.
24. Shahfiza N, Osman H, Hock TT, Abdel-Hamid AHZ. Metabolomics approach for multibiomarkers determination to investigate dengue virus infection in human patients. Acta Biochim Pol. 2017;64:215–9. https://doi.org/10.18388/abp.2015_1224.

25. Silva LK, Hile GA, Capella KM, Espenship MF, Smith MM, De Jesús VR, et al. Quantification of 19 aldehydes in human serum by headspace SPME/GC/high-resolution mass spectrometry. Environ Sci Technol. 2018;52:10571–9. https://doi.org/10.1021/acs.est.8b02745.

26. Mochalski P, King J, Klieber M, Unterkofler K, Hinterhuber H, Baumann M, et al. Blood and breath levels of selected volatile organic compounds in healthy volunteers. Analyst. 2013;138:2134–45. https://doi.org/10.1039/c3an36756h.

27. Marco E, Grimalt JO. A rapid method for the chromatographic analysis of volatile organic compounds in exhaled breath of tobacco cigarette and electronic cigarette smokers. J Chromatogr A. 2015;1410:51–9. https://doi.org/10.1016/j.chroma.2015.07.094.

28. Ernstgård L, Norbäck D, Nordquist T, Wieslander G, Wålinder R, Johanson G. Acute effects of exposure to 1 mg/m3 of vaporized 2-ethyl-1-hexanol in humans. Indoor Air. 2010;20:168–75. https://doi.org/10.1111/j.1600-0668.2009.00638.x.

29. Świercz R, Wąsowicz W, Majcherek W. Mesitylene (1,3,5-trimethylbenzene) in the liver, lung, kidney, and blood and 3,5-dimethylbenzoic acid in the liver, lung, kidney and urine of rats after single and repeated inhalation exposure to mesitylene. Pol J Environ Stud. 2006;15:485–92.

30. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics. 2006;7:142. https://doi.org/10.1186/1471-2164-7-142.

31. Triba MN, Le Moyec L, Amathieu R, Goossens C, Bouchemal N, Nahon P, Rutledge DN, Savarin P. PLS/OPLS models in metabolomics: the impact of permutation of dataset rows on the K-fold cross-validation quality parameters. Mol BioSyst. 2015;11:13–9. https://doi.org/10.1039/c4mb00414k.