Cocoa smoky off-flavor: Chemical characterization and objective evaluation for quality control

This is the author's manuscript

Original Citation:

Availability:
This version is available http://hdl.handle.net/2318/1714885 since 2019-10-31T16:29:06Z

Published version:
DOI:10.1016/j.foodchem.2019.125561

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)
This is the author's final version of the contribution published as:


The publisher's version is available at:


When citing, please refer to the published version.

Link to this full text:

[http://hdl.handle.net/]
Cocoa smoky off-flavor: chemical characterization and objective evaluation for quality control

Pamela Perotti¹, Chiara Cordero¹, Cristian Bortolini², Patrizia Rubiolo¹, Carlo Bicchi¹, Erica Liberto*

¹ Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Turin, Italy
² Soremartec Italia S.r.l., P.le P. Ferrero 1, 12051 Alba (CN), Italy
ABSTRACT

Cocoa smoky off-flavor is due to inappropriate post-harvest processing and cannot be removed in the subsequent chocolate-manufacturing steps. To date, no reliable analytical method to detect key-analytes responsible for smoky off-flavor in incoming raw material is available. This study aims to develop an analytical method, suitable for quality control, to detect smoky markers. The cocoa volatilome was first profiled by headspace solid phase microextration combined with comprehensive two-dimensional gas chromatography-mass spectrometry from a set of representative smoky and non-smoky samples; advanced fingerprinting revealed the chemicals responsible for the off-flavor. The results served to develop a 1D-GC method suitable for routine application. Ten identified smoky markers were subjected to accurate quantification, thereby defining operative ranges to accept/reject incoming bean samples. On average, these markers are present in smoky samples at 7 to 125 fold concentrations vs. those in non-smoky beans, ranging from 8 ng/g for \( p \)-ethylguaiacol to 482 ng/g for phenol.

Keywords: Cocoa volatilome, smoky off-flavor, HS-SPME, GCxGC-TOF MS, GC-MS, chemometrics

1. INTRODUCTION

Food taints and off-flavors are particularly important in food manufacturing, because they may impact consumer confidence and quality perception, while influencing the brand image (Ridgway, Lalljie, & Smith, 2010). A food taint derives from external sources of contamination, e.g. from the environment, processing or storage, whereas an off-flavor may be due to compounds formed through chemical or enzymatic reactions undergone by food components: lipid oxidation, hydrolytic processes or microbiological spoilage (Jelen, 2006; Mottram, 1998; Ridgway et al., 2010).

‘Flavor’ is a multisensory phenomenon involving olfaction, taste, texture and chemestesisis, and provides a distinctive hedonic definition of each food (Auvray & Spence, 2008). In this context, compounds causing off-flavors are ligands that, even if present at trace levels, may trigger olfactory and taste perception resulting in unpleasant and/or unexpected flavor notes.

Chocolate is a typical comfort foods having a rapid and positive impact on a person’s mood (Macht & Mueller, 2007); its main ingredient is cocoa, produced from cocoa beans (Theobroma cacao L. Malvaceae), a tree crop native of the South American continent. Top world producers of cocoa in 2017 were African countries (Ivory Coast, Ghana, Cameroon, Nigeria) followed by Indonesia and South America, with Brazil and Ecuador dominating the market (Eghbal, 2018).

The principal climatic factor influencing cocoa yield and quality is rainfall, although temperature and light exposure are important, affecting pod and bean characteristics (Budiansky, 2018; The International Cocoa Organization (ICCO), 2018). Fermentation and drying are the two fundamental post-harvest treatments that impact the final flavor quality of cocoa products. Traditional drying, where the beans are exposed to the sun for 6-10 days, is to be preferred, not least because it is the simplest and most common method used, resulting in good-quality beans.

To satisfy the ever increasing demand for cocoa, drying is sometimes speeded up by artificial processes. When carried out using heat generated from burning wood or other fuels, artificial drying
requires the smoke originated not to come into contact with the beans, to avoid any transfer of volatiles and semi-volatiles. However, in small farming communities, correct practices are sometimes neglected and the sensory quality of beans may be altered. When improperly conducted, artificial drying can develop a typical smoky off-flavor in cocoa beans; the characteristic note depends on the drying plant, the fuel (wood, diesel, etc.), the type of wood and, after drying, also the storage conditions of the beans (CABISCO/ECA/FCC, 2015; Serra Bonvehí & Ventura Coll, 1998). To date, the occurrence of smoky off-flavor has been found to be limited to African countries, where cocoa is mainly produced by small family farms, and increasing market demand, together with climate change, has increased pressure on the producers (Wessel & Quist-Wessel, 2015).

The smoky off-flavor is sometimes also described as “hammy” because it is reminiscent of smoke-cured bacon. Hammy off-flavors can also arise from over-fermentation, although in smoke-contaminated beans the hammy note is dominant, while in over-fermented cocoa it takes second place to the predominant putrid, ammoniacal or occasionally soapy/phenolic background (Aprotosoai, Vlad Luca, Miron, 2016; CABISCO, 2015; Serra Bonvehí, 1998). The smoky note has chiefly been related to the presence of phenolic compounds, which predominantly derive from lignin degradation by pyrolysis (Janairo & Amalin, 2018; Serra Bonvehí, 1998; Wang, Chambers, & Kan, 2018). Temperature, one of the principal variables impacting the formation of smoke-reminiscent odorants, conditions the chemical structure and substitution of the resulting phenol derivatives. 4-Substituted guaiacols and 4-substituted syringols prevail at lower temperatures, while at higher temperatures the reaction environment becomes richer in H-donors (H-radicals), triggering the formation of catechols/pyrogallols and o-cresols/xylenols (Janairo, 2018; Kawamoto, 2016). Very interestingly, guaiacols and methylphenols are also cocoa key-aroma compounds, their presence in high concentrations can affect cocoa’s sensory properties, influencing the native smoky note (Frauendorfer & Schieberle, 2006). This smoky perception should therefore be considered as
a taint, because it mainly derives from exposure to process smoke, or as an off-flavor, when it is due
to the neo-formation of potent odorants in beans exposed to high temperatures. In native cocoa,
the smoky note may differ in intensity; it may persist and/or may be emphasized in finished products
(chocolate or confectionary), partly because of improper manufacturing practices. Moreover, it has
been shown that odorless compounds, when combined with potent odorants, can also contribute
to the sensory profile of a perceived flavor, and increase the perceived intensity of the smoky note
(Chambers & Koppel, 2013; Jaffe, Wang, & Chambers, 2017). This synergistic effect is likewise
possible with the association between two non-smoky phenolic compounds, such as for instance
2,6-dimethylphenol at 100 ppm and eugenol at 1 ppm in propylene glycol and delivered from a
fragrance strip (Wang et al., 2018). Moreover, several flavor compounds responsible for positive
sensory attributes in foods can act as off-flavors when their concentration exceeds a certain
threshold. Known examples are sulfur compounds, such as dimethyl sulphide in beer, which has a
cabbage-like aroma at high concentrations, or 4-vinylphenol and 4-vinylguaiacol in wines (Jelen,
2006; H. Wang et al., 2018). Very few studies are available concerning cocoa smoky off-flavors
(Aprotosoaie et al., 2016; Lehrian, Keeney, & Lopez, 1978; Serra Bonvehí, 1998).

In this context comprehensive two-dimensional gas chromatography (GC×GC) coupled with time-
of-flight mass spectrometry (TOF MS) is an effective approach for detailed characterization of
complex mixtures of volatiles in food (Cordero, Kiefl, Schieberle, Reichenbach, & Bicchi, 2015).
GC×GC exploits the separation power and detection potential of the two dimensions, providing
representative 2D chromatographic patterns, and increasing sensitivity versus trace components.

Despite its potential, GC×GC-TOF MS is not yet routinely used in chemical characterization of foods;
in general, it is considered too complex for quality control laboratories, although new less
sophisticated commercial solutions avoiding thermal modulation and cryogenics make it promising
for routine analysis (Magagna et al., 2017). Conversely, methods based on 1D-GC-MS are well
accepted, cost-effective and, when integrated with automatic sample preparation and injection systems, enables fully-automatic analytical procedures for high-throughput screening to be developed.

Headspace solid phase microextraction sampling (HS-SPME) perfectly meets the above requirements, and has been widely used to characterize cocoa aroma (Ducki, Miralles-Garcia, Zumbé, Tornero, & Storey, 2008; Magagna et al., 2017; Phuong et al., 2015). It has also been adopted to screen off-flavors in several other foods, in particular to identify components responsible for unpleasant odor(s), such as haloanisoles in wine and cork, or geosmine and methylisoborneol in water (Jelen, 2006; Ridgway et al., 2010). Food consumption is mainly driven by the pleasure perceived during its intake; therefore, food sensory features became an integral part of the quality control (QC) and quality assurance. To date, the approach to detect the smoky aroma is based on a sensory evaluation by trained panelists resulting in a rather expensive and time consuming process. In addition, the lack of a reference and objective analytical method to detect and quantify chemical markers of the smoky off-flavor on incoming raw materials inspired the current research. This study aimed at developing an analytical method, suitable for routine quality control, for volatiles profiling and accurate quantitation of smoky off-flavor key-markers in cocoa beans and liquors. The final method should afford fast, accurate, objective discrimination between smoky and non-smoky cocoa products. This goal was pursued with a top-down strategy where HS-SPME-GC×GC-TOF-MS served as screening platform to identify informative odorants within a subset of samples characterized as smoky and non-smoky by an internal panel. HS-SPME-GC-MS in a fully automated set-up was then used to monitor targeted discriminating compounds and to accurately quantify them by multiple headspace extraction (MHE). Quantitative ranges for targeted compounds were then fixed as decisive markers to accept or reject incoming cocoa samples.
2. MATERIALS AND METHODS

2.1 Cocoa samples and Reference compounds

The sample set included beans (n= 54) and liquors (n= 31) of cocoa samples (*Theobroma cacao* L. main crop) of commercial grade (beans size “standard” based on counting test- federation of cocoa commerce) from different origins and harvested in different years (Table 1). Cocoa bean and liquor samples were provided by Soremartec Italia srl (Alba, Italy).

Pure reference standards for identity confirmation (key-aroma compounds and informative volatiles) were from Millipore (Milan, Italy) (Table 2), in particular acetic acid, 3-methyl butanoic acid, 3-methyl butanal, 2-phenyl ethanol, 2-heptanol, butanoic acid, 2-methyl butanal, linalool, phenylacetaldehyde, 2-ethyl-3,5-dimethyl pyrazine, 4-hydroxy-2,5-dimethyl-3(2H)-furanone, 2-ethyl-3,6-dimethyl pyrazine, (E,E)-2,4-nonadienal, dimethyl trisulfide, 2-methyl propanoic acid, ethyl-2-methyl butanoate, ethylbenzoate, 1,2,4-trimethoxybenzene, 2,6-dimethoxyphenol, carvacrol, 2-phenoxyethanol, *p*-cresol, *p*-ethylguaiacol, 1-h-pyrrole-2-carboxaldeide, phenol, 2-ethoxy-4-methylphenol, guaiacol, isoamylbenzoate naphthalene, 1,2-dimethoxybenzene, and 1,4-dimethoxybenzene.

Normal alkanes (n-alkanes n-C9 to n-C25) for Linear Retention Index (IT$_S$) determination and Internal standardization (n-heptadecane n-C17 - ISTD) were from Millipore (Milan, Italy) (Table 2).

A standard stock solution of ISTD at 1000 mg/L was prepared in degassed sunflower seed oil and stored in a sealed vial at -18°C.

2.2 Automated Head Space Solid Phase Micro Extraction: sampling devices and conditions

Automated Headspace Solid Phase Microextraction (auto-HS-SPME) was performed using a Combi-PAL AOC 5000 (Shimadzu, Milan, Italy) on-line integrated with a Shimadzu QP2010 GC–MS system
provided with Shimadzu GC–MS Solution 2.51 software (Shimadzu, Milan, Italy). SPME fiber: Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) df 50/30 μm - 2 cm length from Millipore (Bellefonte, PA, USA). Fibers were conditioned before use as recommended by the manufacturer. The standard-in-fiber procedure was adopted to pre-load the ISTD (n-C17) onto the fiber before sampling (Y. Wang, O’Reilly, Chen, & Pawliszyn, 2005). 5.0 μL of ISTD solution were placed in a 20 mL glass vial and submitted to HS-SPME at 80°C for 20 min, stirring speed 350 rpm. Cocoa samples were ground in liquid nitrogen and then stored at -80°C until analyzed. Samples were ground before headspace analysis to obtain a homogeneous powder. Cocoa powder (1.00 g) was weighed in the headspace glass vials (20 mL) and submitted to automated HS-SPME sampling. After ISTD loading, the SPME device was exposed to the headspace of cocoa for 40 min at 80° at a shaking speed of 350 rpm. Extracted analytes were recovered by thermal desorption from the fiber into the split/splitless (S/SL) injection port of the GC system at 250°C for 5 min. Each sample was analyzed in duplicate. Sampling and ISTD standardization for the preliminary screening by GC×GC-TOF MS analysis was done at 50°C under the analytical conditions reported by Magagna et al. (2017).

2.3 Quantitation

An amount of ground material appropriate to achieve headspace linearity for target analytes was processed. MHE quantification was by the External Standard approach; an aliquot of 0.100 g of ground beans was sealed in a 20 mL headspace vial and submitted to multiple consecutive extractions, exposing the fiber to the headspace for 40 minutes at 80°C before analysis. A series of calibrating solutions of reference compounds in cyclohexane, ranging from 0.1 to 50 mg/L, were used in full evaporation for MHE external calibration. Suitable volumes of standard solutions at different concentrations were submitted to multiple consecutive extractions (as for the cocoa
samples). All calibration solutions and samples were analyzed in duplicate, by full evaporation MHS-
SPME.

2.4 GC-MS and GC×GC-TOF MS instrument set-up and analytical conditions

**GC-MS analysis - Chromatographic conditions:** analyses were run on a Shimadzu QP2010 GC–MS system, controlled by Shimadzu GC–MS Solution 2.5SU1 software (Shimadzu, Milan, Italy) Injector temperature: 240°C, injection mode: splitless; carrier gas: helium, flow rate: 1 mL/min; fiber desorption time and reconditioning: 5 min; column: SolGelwax (100% polyethylene glycol) 30 m x 0.25 mm $d_c$ x 0.25 μm $d_f$ Trajan Analytical Science (Ringwood, Australia). Temperature program, from 40°C (2 min) to 200°C at 3.5°C/min, then to 240°C (5 min) at 10°C/min. MSD conditions: ionization mode: EI (70 eV); temperatures: ion source: 200°C; quadrupole: 150°C; transfer line: 260°C; scan range: 35-350 amu.

**GC×GC-TOF MS analysis - Chromatographic conditions:** GC×GC analyses were run on an Agilent 7890 GC unit coupled with a Markes BenchTOF-Select and Select-eV® option (Markes International Ltd, Llantrisant UK) operating in the EI mode at 70eV. The transfer line was set at 250°C. TOF acquisition was set at m/z 35-350 with 100 Hz sampling frequency. The GC system was equipped with a two-stage KT 2004 loop thermal modulator (Zoex Corporation, Houston, TX) cooled with liquid nitrogen and controlled by Optimode™ V.2 (SRA Instruments, Cernusco sul Naviglio, MI, Italy). Hot jet pulse time was set at 250 ms, modulation time was 3.5 s and cold-jet total flow progressively reduced with a linear function from 40% of Mass Flow Controller (MFC) at initial conditions to 8% at the end of the run.

SPME thermal desorption into the GC injector port was in split mode, split ratio 1:20. Carrier gas was helium at a constant flow of 1.3 mL/min. The oven temperature program was: from 40°C (2 min) to 200°C at 3.5°C/min and to 240°C at 10°C/min (10 min). The column set was configured as
follows: 1D SolGel-Wax column (100% polyethylene glycol) (30 m × 0.25 mm dc, 0.25 μm df) from SGE Analytical Science (Ringwood, Australia) coupled with a 2D OV1701 column (86% polydimethylsiloxane, 7% phenyl, 7% cyanopropyl) (2m × 0.1 mm dc, 0.10 μm df), from J&W (Agilent, Little Falls, DE, USA).

The n-alkanes solution for $I^S_T$ determination was analyzed with a split/splitless injector in split mode, split ratio 1:50, injector temperature 250°C, and injection volume 2 μL.

2.5 Data acquisition and processing

GC×GC-TOF MS data were acquired by TOF-DS software (Markes International, Llantrisant, UK) and processed using GC Image GC×GC Software, version 2.8 (GC Image, LLC, Lincoln NE, USA). GC-MS data were collected by GCMS Solution 2.5SU1 software (Shimadzu, Milan, Italy).

2.6 Analytes identification and chemometrics

Targeted analysis was focused on about 70 compounds identified by matching their EI-MS fragmentation patterns (NIST MS Search algorithm, version 2.0, National Institute of Standards and Technology, Gaithersburg, MD, USA, with Direct Matching threshold 900 and Reverse Matching threshold 950) with those stored in commercial (NIST2014 and Wiley 7n) and in-house databases. Linear retention indices ($I^S_T$) were taken as a further parameter to support identification, and experimental values were compared to tabulated units. Principal Component Analysis (PCA), Partial Least Square Discriminant Analysis (PLS-DA) and regression analysis were performed with Pirouette® (Comprehensive Chemometrics Modeling Software, version 4.5-2014) (Infometrix, Inc. Bothell, WA).

Heat-map was implemented in Morpheus (https://software.broadinstitute.org/morpheus/) while the Kruskall-Wallis test was performed with XLstat (version 16.05) (Addinsoft, New York, NY USA).
3 RESULTS AND DISCUSSION

The chemical complexity of the volatilome of cocoa beans and cocoa liquor depends on the many chemical reactions occurring during the early stages of processing, most of which are catalyzed by specific enzymes (endogenous or exogenous from moulds, yeasts and bacteria) (Ohene Afoakwa, Paterson, Fowler, & Ryan, 2008). A top-down approach was here adopted to define the characteristic markers of smoky off-flavor (Konieczka & Namieśnik, 2018). This approach exploits the possibility of capturing the necessary information from highly-informative fingerprints of non-smoky and smoky samples, with sophisticated and powerful techniques. This information is then exploited to develop a 1D-GC-MS method with suitable informative potential, that offers reliability appropriate for the needs of a routine laboratory. The differential compositional characteristics highlighted by GC×GC-MS, between smoky and non-smoky samples, are used in 1D-GC-MS in combination with chemometrics, to discriminate samples and to obtain robust evidence of the markers related to the defect. Once their informative role is confirmed, some of the representative analytes are submitted to accurate quantitation, to set the limits of acceptability for incoming samples. Quantitation of smoky compounds was then done on cocoa beans.

The following sections illustrate: (a) the chemical complexity of the volatile fraction of cocoa samples and the information deriving from comparative analysis between smoky and non-smoky samples, as revealed by the untargeted-targeted investigation; (b) the optimization of the sampling procedure to improve the analytical response from analytes related to defective samples in the 1D-GC-MS method, (c) the role of unsupervised and supervised approaches in supporting the selection of informative chemicals, whose quantitation provides a reliable range of sample acceptability (Sgorbini et al., 2019).
3.1 Reveiling smoky odorant patterns by GC×GC-TOF MS

The sensory description of the smoky off-flavor has driven the search for the compound(s) that may be related to the smoky note. The smoked flavor is variously described as smoky, ashy, woody, musty/dusty, musty/earthy, burnt, acrid, pungent, petroleum-like, creosote/tar, cedar, bitter, metallic and sour (Jaffe et al., 2017). Several phenolic compounds, such as 2,6-dimethoxyphenol, 4-ethyguaiacol, thymol, guaiacol, and carvacrol, have been indicated as chemicals potentially associated with smoky aroma in foods. The smoky note is one of the positive sensory attributes for different foods, such as coffee, cocoa, ham and fish, and it is also used as artificial smoky flavorings (Frauendorfer & Schieberle, 2006; Janairo & Amalin, 2018; Marušić Radovčić, Vidaček, Janči, & Medić, 2016; Wang et al., 2018). Conversely, smoky notes may also be considered as negative, e.g. in wine, where they are associated with volatile and glycoconjugated phenols, and their removal considerably reduces smoke taint (Krstic, Johnson, & Herderich, 2015).

Smoky off-flavor in cocoa and chocolate liquor was studied by Lehrian et al. in 1978; they proposed a colorimetric method to measure phenols associated with the off-flavor, without offering any specific chemical speciation (Lehrian et al., 1978). Serra Bonvehí et al. identified 3-methylphenol (m-cresol), 2,3-dimethylphenol (2,3-xylenol), 3-ethylphenol, and 4-ethylphenol as discriminant markers of the smoky note, after hydro-distillation followed by solvent extraction and GC-MS of cocoa powders (Serra Bonvehí et al., 1998). Misnawi et al. in 2011 suggested that the smoky odor of cocoa liquor analyzed by HS-SPME-GC-O was associated with the presence of α-ethyldiene-benzeneacetaldehyde, trimethyl pyrazine, and 2,3-dimethyl-trans-oxirane (Misnawi & Ariza, 2011).

The smoky note has been correlated with several volatiles of different natures and chemical structures, unlike the case of other sensory defects, such as the musty-earthy note imparted by haloanisoles, methylisoborneol, or geosmin. This lack of specific information is a challenge that has here been taken up through a multi-approach strategy.
An initial detailed “screening” was applied to reveal compositional differences on the volatile fractions of smoky and non-smoky beans. Analyses were carried out by HS-SPME-GC×GC-TOF MS on a sub-set of samples, and resulted in 2D-peak patterns described by an average of 230 peak-regions, corresponding to both untargeted and targeted analytes. This approach, known as combined untargeted and targeted (UT) fingerprinting, is based on the template matching strategy and enables direct comparison of peak patterns through their specific metadata (i.e., retention times and MS fragmentation patterns) (Magagna et al., 2017). Reliable correspondences are established between the same chemical entities across multiple chromatograms, thanks to analytes’ relative positions (i.e., retention time correspondences) and MS spectral similarity (Direct Match Factor above 800). The output of UT fingerprinting is a data matrix of aligned 2D peaks and peak-regions, that can be used for comparative purposes (Cordero et al., 2010). Figures 1 A-D show the 2D-patterns of volatiles from a non-smoky (1A) and a smoky (1B) sample, produced by/as they emerged from a polar × semi-polar column combination. Enlarged areas in 1C and 1D show in detail the region where some aromatic compounds elute. In particular, Fig. 1C provides a comparative visualization of the difference between a non-smoky sample (reference image) and a smoky sample (analyzed image) rendered as colorized fuzzy ratio. Green areas correspond to analytes (known or unknown) with a higher relative detector response in the smoky (analyzed) sample. Fig.1D shows, for the smoky sample, the elution region of interest for aromatics and phenol derivatives.

The average percent difference between smoky and non-smoky samples was calculated, to locate the most informative 2D peak-regions (known or unknown) describing the smoky pattern. The analyte response from non-smoky samples was taken as reference, and an arbitrary cut-off of 100 was set, to include or otherwise the feature in the final list. Of the most informative compounds, 56 were found to be more abundant in smoky samples; however, only a few of them could be correlated with the smoky note. These were naphthalene, 2-phenylethyl acetate, ethyl-4-ethoxy
benzoate, methoxy-4-propylphenol (dihydroeugenol), phenylethyl alcohol, 2-phenoxyethanol, m-

cresol, phenol, p-cresol and 3-ethylphenol.

These results are in agreement with other reports (Serra Bonvehí & Ventura Coll, 1998; H. Wang et

al., 2018); they were therefore taken into account in the next step, in which a 1D-GC approach was

applied to screen these targeted odorants.

3.2 Transfer to 1D-GC-MS analysis: improving method sensitivity toward phenolic compounds

When the HS-SPME-GC×GC-TOF MS method was transferred to HS-SPME-GC-MS, smoky markers

gave poor signals; in particular, the areas of the analytes with a relatively high boiling point were

very small, in some cases below the method’s limit of detection. Moreover, phenolic derivatives

were connoted by high hydrophobicity, thus showing rather high affinity for the highly abundant

cocoa fatty matrix (45-53%), while not being readily releases from ground beans. (Kopjar, Andriot,

Saint-Eve, Souchon, & Guichard, 2010).

Careful tuning of the sample preparation step was therefore necessary to improve the recovery of

these compounds, so as to obtain information about the volatiles whose average percentages varied

significantly between the two sets of samples. A compromise was also sought between the need to

increase the extraction rate and the need to adopt a sampling method that is easy to automate and

to combine on-line with the analytical instrumentation for routine controls.

The following sampling conditions/variables were investigated in this perspective:

a) Sample amount (1 – 3g);

b) Sampling temperature (50 and 80°C);

c) SPME fiber coatings and composition (PDMS/DVB, DVB/CAR/PDMS, PDMS; and PDMS/DVB

coated with PDMS for in-solution sampling)
**3.3 Chemometric-driven approach to select informative markers of the smoky note**

The HS-SPME-GC-MS profiles obtained under the optimized sampling conditions (see paragraph 2.2) on the bean and liquor samples under study, and detailed in Table 1, are shown in the heat-map of Figure 2. The rows indicate the investigated samples (beans and liquors) and the columns the targeted analytes by HS-SPME-GC-MS. The color scale varies from blue (low abundance) to red (high abundance). Hierarchical cluster analysis (HCA) of both rows and columns, by Spearman rank correlation through the average linkage method, shows a different distribution of the volatiles based on their normalized response across samples. HCA shows a clear separation between beans and liquors. Analytes are ordered according to their inter-class variance.

Cocoa beans display higher relative abundances for several targets, as a reflection of quantitative changes of acids, esters, alcohols and ketones (Table 2) and in particular for methyl and ethyl acetates (green-fruity), 2-phenylethylacetate (flowery), 2-methyl-1-propanol, 2-heptanol (citrusy), 2,3-butandiol (fruity/creamy), 3-hydroxy-2-butanone (acetoin-buttery) and 2-pentanone (fruity) (Figure 2). In particular, acetic acid is the most abundant volatile and, when present in high amounts in beans, it gives an intense vinegar-like odor that can affect the cocoa aroma quality (Frauendorfer...
et al., 2006). However, during cocoa processing, undesired volatiles with low boiling points are removed or drastically decreased in concentration (up to 70% for acetic acid) (E. Ohene Afoakwa et al., 2008). Ethyl and methyl esters, in particular acetates, derive from amino acids and are typical aroma components in unroasted beans. 2-Phenylethyl acetate has been found in unroasted and roasted cocoa, and it can also be formed through yeast metabolism. Alcohols, aldehydes and ketones result from microbial activity during fermentation but, during roasting, aldehydes (in particular) are significantly reactive, also taking part in the formation of heterocyclic compounds (pyrazines), while the alcohol concentration decreases, negatively affecting the aroma because their presence is desirable to obtain sweet and floral notes in finished cocoa products (Aprotosoaie et al., 2016; Misnawi & Ariza, 2011; Ziegleder G., 2009).

Conversely, liquors contain higher amounts of 2,3,5-trimethylpyrazines and tetramethylpyrazines (cocoa/nutty/musty notes), acetophenone, benzaldheyde and furfural (almond/sweety), 2-butanone and 2-nonanone (sweety/fruity), 4-hydroxybutanoate (fruity), guaiacol and phenol (phenolic/smoky) Table 2 and Figure 2. Phenolic compounds are key-odorants formed during roasting in relatively small amounts (Frauendorfer & Schieberle, 2006; Rychlik, Schieberle, & Grosch, 1998). They are present in both non-smoky and smoky liquors, although large amounts can be formed because of incorrect drying or storage processes. Their level can also increase during bean roasting, which is generally between 110°C and 140°C. The clusters and the red and green right-hand-side bars highlight the smoky and non-smoky bean and liquor samples. The heat-map highlights the volatiles virtually linked to these clusters, including 1,2-dimethoxybenzene, guaiacol derivatives (smoky/phenolic/spice), p-cresol (phenolic/pungent), naphtalene (pungent), phenol (phenolic/rubbery), 2,6-dimethoxyphenol (sweet/smoky/medicinal), 2-methoxy-4-methylphenol (sweet/smoky/medicinal), and 3-ethylphenol (musty).
Unsupervised pattern recognition through Principal Component Analysis (PCA) was applied to the targeted data matrix for beans and liquors, to explore the conformation (groups) of samples and to localize informative chemicals responsible for discrimination. PCA in Figure 3, referred to bean volatiles, makes a clear distinction between smoky (pink) and non-smoky (blue) samples with an explained variance of 69.14% on the first 3 PCs, regardless of the origin of the samples (Figure 3A). The loading plot Figure 3B shows that smoky samples are described by most of the volatiles mentioned above, confirming initial observations derived from the heat-map (Figure 2). Within beans, older samples (*_old harvested in 15/16 and 16/17 in table 1) are recognizable on PC1, and are characterized by a relatively high abundance of hexanoic acid and 1-H-pyrrole-carboxyaldhyde (Figure 3A and 3B).

PCA obtained by extrapolating only those volatiles related to the discrimination of smoky samples still shows a coherent distribution by smoky and non-smoky, at 79.44% of explained variance (data not shown).

Supervised pattern recognition via PLS-DA on both beans and liquors provides a coherent classification by beans or liquors, and by smoky or non-smoky samples Figure 3C. PLS-DA was done on logarithm (Log10) transformed data, pre-processed by auto scaling and cross validated (5 CV). The total classification rate was 97%, in particular the classification model built up showed a 100% ability for beans, and a slightly lower one for liquor (92%). The correlation spectrum is a useful function to exclude x variables (e.g. volatiles) that correlate weakly with the qualitative y variable (e.g. smoky/non-smoky liquors and beans) (Figure 3D). The correlation spectrum facilitates the identification of the closest smoky-correlated analytes, i.e. the ten components in the HS-SPME-GC-MS pattern (highlighted at top right of the graph in Figure 3D). These components can discriminate smoky from non-smoky samples independently of the origin or processing step considered (raw cocoa beans or liquors). Significance analysis on all analytes was carried out through the non-
parametric Kruskall-Wallis test. The \( p \)-values of bean and liquor volatiles are reported in Table 2. Naphthalene, guaiacol, 2-methoxy-4-methylphenol, phenol, 1H-Pyrrole-2-carboxaldehyde, \( p \)-ethylguaiacol, \( p \)-cresol, 3-ethylphenol, 2,6-dimethoxyphenol, 4-methyl-2,6-dimethoxyphenol differed significantly between smoky and non-smoky samples, in both beans and liquors, although to differing extents. These analytes were therefore submitted to accurate quantification (3.4). Other components show significant variations in beans, but their variation may also be influenced by the year of harvest, and thus by the “age” of the samples (Table 1 and 2). Table 2 also shows other volatiles significantly varying in smoky and non-smoky liquors, including 2-butanone, 3-methylbutanal, 2 and 3-methyl-ethyl butanoate, hexanal, 1,3-dimethyl-benzene, benzaldehyde, 2,3-butanediol and hexanoic acid. Further investigations will be required to define their roles.

3.4 Quantitation of the selected marker compounds

Cross-sample comparisons through relative quantitation, based on Peak Area % or on Internal Standard normalization, may be inaccurate or misleading if taken as analyte(s) concentration indicators, in particular when the aim of profiling is to correlate chemical composition with sensory properties, or process kinetics. However, absolute quantitation of solid matrices is complex, since they are characterized by a heterogeneous composition and structure, in which volatiles can be retained and released into the HS in different ways (Sgorbini et al., 2019). Multiple Headspace Extraction (MHE) is one of the approaches to quantifying solid samples, enabling the matrix effect to be overcome (Kolb & Ettre, 2006). More recently, its use has successfully been extended to HS-SPME, also known as MHS-SPME (Bicchi et al., 2011; Sgorbini et al., 2015 and references cited therein).
MHS-SPME is based on stepwise dynamic gas extraction of the investigated analyte/s from a solid or liquid sample. It comprises three main steps:

Step 1. Exhaustive extraction of analytes from samples to define HS linearity boundaries;

Step 2. Application of the MHE procedure to the samples of interest;

Step 3. Exhaustive extraction of analytes from calibration solutions, in a range of concentrations matching real-sample concentrations.

Steps 1 and 2 define the total peak area obtained from a series of consecutive and exhaustive extractions; it is directly related to the total amount of analyte in the matrix. The analyte peak area decreases exponentially with the number of extractions, provided that a suitable amount of matrix is processed. The cumulative instrumental response is obtained from the following equation:

\[
AT = \sum_{i=1}^{\infty} Ai = A1 \left( \frac{1}{1-e^{-q}} \right) = \frac{A1}{1-\beta} \quad \text{(Eq. 1)}
\]

where \( AT \) is the total estimated area, \( A1 \) is the area detected with the first extraction, and \( q \) is a constant describing the exponential decay of the area through successive extractions. Quantitation is achieved by external standard calibration with a standard solution of the analyte(s) investigated, subjected to the same MHE conditions as the real sample.

Six of the ten markers selected from the above procedure were quantified by MHS-SPME in smoky and in non-smoky beans; 3-ethylphenol was not quantified at this stage of the study because the standard was not available commercially, while 2,6-dimethoxyphenol, 4-methyl-2,6-dimethoxyphenol and 1-H-pyrrole-2-carboxyaldheyde were excluded because they were outside the HS linearity boundaries related to the sample amount chosen for the MHE procedure.
The results indicated that the average amounts of the investigated markers in the smoky samples were between 7 and 125 times higher than in non-smoky beans ranging from 8 ng/g for \( p \)-ethylguaiacol to 482 ng/g for phenol (Table 3). 2-Methoxy-4-methylphenol, \( p \)-ethylguaiacol and \( p \)-cresol are detectable but below the limit of quantitation in non-smoky samples. The high standard deviation for phenol and \( p \)-cresol is probably due to the different seasonality of the investigated samples (crop of 2 years). An operative limit below 10 ng/g of the selected smoky compounds can thus generally be adopted in acceptance of incoming bean samples (Table 3).

4. Conclusions

The intrinsic information potential of the cocoa volatile fraction has been shown to be diagnostic to discriminate defective from non-defective samples, for both beans and liquors. The top-down strategy employed successfully defined the cocoa aroma components related to smoky off-flavor, and led to a routine method for their detection. The informative power of GC×GC-TOF MS analysis, combined with advanced fingerprinting (i.e., UT fingerprinting), was used for a preliminary investigation on a limited but significant selection of smoky and non-smoky samples. This step enabled the chromatographic peak-regions (features) discriminating off-flavor samples from the rest to be detected. The chief compounds identified related to the smoky note are phenolic in nature, and are present either as minor components or in larger amounts than in the non-smoky samples. The results of this first step were used to develop a simple and automatic routine HS-SPME-GC-MS method combined with multivariate statistics, for discrimination and classification of beans and liquors. HS-SPME sampling was chosen because of the nature of the cocoa matrix, the low target analyte concentration, and method's high concentration capability and reliability in quantitation. Quantitation of the selected markers allows adopting an operative limit below 10 ng/g for the acceptation of the incoming cocoa beans.
However, some aspects related to quantitation (e.g. HS linearity ranges for markers) merit investigation in greater depth, because of both the relatively small amounts of components related to the smoky off-odor, and the rheological complexity of the cocoa matrix. In particular, quantitation of smoky markers in liquor is still challenging, because the modification of the lipid crystalline structure during processing can influence their release to the headspace, and affect not only the HS linearity ranges with MHE (Nicolotti et al., 2013) but also the definition of their chemical limits of acceptability in compliance with the sensory perception.

Aknowledgments

The research was financially supported by Soremartec Italia s.r.l Alba (CN), Italy.
Table captions

Table 1 Cocoa samples analyzed listed by type (beans and liquors), origin and year of harvesting.

Table 2 List of volatiles targeted by HS-SPME-GC-TOF MS, together with their experimental and reference $I^T_s$, volatiles’ normalized responses both in beans and liquors, $p$-value through the Kruskall-Wallis test (alpha=0.05), and odor descriptors as reported in the reference literature (Frauendorfer & Schieberle, 2006; Rychlik et al., 1998; http://www.thegoodsentscompany.com/).

Abbreviations: A: target analytes identified by means of authentic standards, MS: analytes tentatively identified on MS fragmentation patterns available in commercial libraries, and RI: Linear Retention Indices ($I^T_s$) available in Nist (https://webbook.nist.gov/).

Table 3 Amounts of selected smoky markers in cocoa beans (smoky and non-smoky) with standard deviation, LOQ and LOD.

Figure captions

Figure 1 GC×GC-TOF MS patterns of volatiles for non-smoky A) and smoky samples B); in the colorized fuzzy differences, brilliant green represents the positive differences in component abundances of the smoky vs. non-smoky samples; C) enlargement of the aromatic and phenol region, with tentative identification.

Figure 2 Heat-map of the HS-SPME-GC-MS volatile profiling of bean and liquor samples, and hierarchical cluster analysis of rows and columns by Spearman rank correlation, with the average linkage method.
Figure 3 PCA scores A) and loadings B) plots of bean samples on the first 3 PCs; C) PLS-DA class prediction: 1 smoky liquors (red), 2 non-smoky liquors (green), 3 smoky beans (cyan), 4 non-smoky beans (blue); D) PLS-DA correlation spectrum. Data matrix was transformed by Log10 and pre-processed through auto scaling both in PCA and in PLS-DA.
References


https://doi.org/10.1111/joss.12262


Table 1

<table>
<thead>
<tr>
<th>Harvesting Year</th>
<th>Smoky</th>
<th>Non-Smoky</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15/16</td>
<td>16/17</td>
</tr>
<tr>
<td>Origin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beans</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Camerun</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Ecuador</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Nigeria</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Liquors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Camerun</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Nigeria</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Identified compounds</td>
<td>Compounds Confirmation</td>
<td>Calc $I^*$</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Methyl Acetate</td>
<td>A; RI; MS</td>
<td>836</td>
</tr>
<tr>
<td>2-Methylfuran</td>
<td>A; RI; MS</td>
<td>866</td>
</tr>
<tr>
<td>Ethyl Acetate</td>
<td>A; RI; MS</td>
<td>878</td>
</tr>
<tr>
<td>2-Butanone</td>
<td>A; RI; MS</td>
<td>886</td>
</tr>
<tr>
<td>2-Methyl-butanal</td>
<td>A; RI; MS</td>
<td>898</td>
</tr>
<tr>
<td>3-Methyl-butanal</td>
<td>RI; MS</td>
<td>902</td>
</tr>
<tr>
<td>2,4,5-Trimethyl-1,3-Dioxolane</td>
<td>RI; MS</td>
<td>927</td>
</tr>
<tr>
<td>2-Pentanone</td>
<td>A; RI; MS</td>
<td>957</td>
</tr>
<tr>
<td>Isobutyl Acetate</td>
<td>RI; MS</td>
<td>992</td>
</tr>
<tr>
<td>α-Pinene</td>
<td>A; RI; MS</td>
<td>1006</td>
</tr>
<tr>
<td>2-methyl-ethyl butanoate</td>
<td>RI; MS</td>
<td>1034</td>
</tr>
<tr>
<td>3-methyl-ethyl butanoate</td>
<td>RI; MS</td>
<td>1050</td>
</tr>
<tr>
<td>2-Pentanol acetate</td>
<td>RI; MS</td>
<td>1055</td>
</tr>
<tr>
<td>Hexanal</td>
<td>A; RI; MS</td>
<td>1060</td>
</tr>
<tr>
<td>2-Methyl-1-propanol</td>
<td>A; RI; MS</td>
<td>1081</td>
</tr>
<tr>
<td>3-Methyl-2-butanol</td>
<td>RI; MS</td>
<td>1107</td>
</tr>
<tr>
<td>1-Butanol-3-methyl-acetate</td>
<td>A; RI; MS</td>
<td>1107</td>
</tr>
<tr>
<td></td>
<td>RI; MS</td>
<td>1118</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>1,3-Dimethyl-benzene</td>
<td>A; RI; MS</td>
<td>1130</td>
</tr>
<tr>
<td>β-Mycene</td>
<td>A; RI; MS</td>
<td>1137</td>
</tr>
<tr>
<td>2-Heptanone</td>
<td>A; RI; MS</td>
<td>1157</td>
</tr>
<tr>
<td>Limonene</td>
<td>A; RI; MS</td>
<td>1166</td>
</tr>
<tr>
<td>3-Ethyl-1-butanol</td>
<td>RI; MS</td>
<td>1187</td>
</tr>
<tr>
<td>cis β-Ocimene</td>
<td>RI; MS</td>
<td>1218</td>
</tr>
<tr>
<td>3-Hydroxy-2-butanoone</td>
<td>A; RI; MS</td>
<td>1256</td>
</tr>
<tr>
<td>2-Heptanol</td>
<td>A; RI; MS</td>
<td>1294</td>
</tr>
<tr>
<td>2-Nonanone</td>
<td>RI; MS</td>
<td>1363</td>
</tr>
<tr>
<td>2,3,5-Trimethylpyrazine</td>
<td>RI; MS</td>
<td>1380</td>
</tr>
<tr>
<td>Acetic Acid</td>
<td>A; RI; MS</td>
<td>1410</td>
</tr>
<tr>
<td>Furfural</td>
<td>A; RI; MS</td>
<td>1432</td>
</tr>
<tr>
<td>Tetramethyl-Pyrazine</td>
<td>RI; MS</td>
<td>1442</td>
</tr>
<tr>
<td>Benzaldehyde</td>
<td>A; RI; MS</td>
<td>1483</td>
</tr>
<tr>
<td>2,3-Butanediol</td>
<td>A; RI; MS</td>
<td>1507</td>
</tr>
<tr>
<td>1-Methoxy-2-propyl acetate</td>
<td>RI; MS</td>
<td>1532</td>
</tr>
<tr>
<td>2-Ethoxy-propane</td>
<td>RI; MS</td>
<td>1544</td>
</tr>
<tr>
<td>4-Hydroxy-butanoate</td>
<td>RI; MS</td>
<td>1581</td>
</tr>
<tr>
<td>Acetophenone</td>
<td>A; RI; MS</td>
<td>1606</td>
</tr>
<tr>
<td>Pentanoic acid</td>
<td>A; RI; MS</td>
<td>1642</td>
</tr>
<tr>
<td>Compound</td>
<td>A; RI; MS</td>
<td>RI</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------</td>
<td>----</td>
</tr>
<tr>
<td>2-Phenylethyl acetate</td>
<td>1773 1785</td>
<td></td>
</tr>
<tr>
<td>Hexanoic acid</td>
<td>1814 1816</td>
<td></td>
</tr>
<tr>
<td>Phenylethyl Alcohol</td>
<td>1863 1912</td>
<td></td>
</tr>
<tr>
<td>Ethylbenzoate</td>
<td>1626 1641</td>
<td></td>
</tr>
<tr>
<td>1,2-Dimethoxybenzene</td>
<td>1687 1699</td>
<td></td>
</tr>
<tr>
<td>Naphthalene</td>
<td>1689 1707</td>
<td></td>
</tr>
<tr>
<td>Isoamylbenzoate</td>
<td>1798 1921</td>
<td></td>
</tr>
<tr>
<td>Guaiacol</td>
<td>1812 1823</td>
<td></td>
</tr>
<tr>
<td>2-Methoxy-4-methylphenol</td>
<td>1906 1956</td>
<td></td>
</tr>
<tr>
<td>Phenol</td>
<td>1958 1965</td>
<td></td>
</tr>
<tr>
<td>1H-Pyrrole-2-carboxaldehyde</td>
<td>1968 -</td>
<td></td>
</tr>
<tr>
<td>p-Ethylguaiacol</td>
<td>1979 2032</td>
<td></td>
</tr>
<tr>
<td>p-Cresol</td>
<td>2040 -</td>
<td></td>
</tr>
<tr>
<td>2-Phenoxyethanol</td>
<td>2085 2087</td>
<td></td>
</tr>
<tr>
<td>3-Ethylphenol</td>
<td>2123 -</td>
<td></td>
</tr>
<tr>
<td>2,6-Dimethoxyphenol</td>
<td>2204 2269</td>
<td></td>
</tr>
<tr>
<td>4-Methyl-2,6-dimethoxyphenol</td>
<td>2289 -</td>
<td></td>
</tr>
<tr>
<td>Compounds</td>
<td>smoky</td>
<td>non-smoky</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td>Average (ng/g)</td>
<td>s.d.</td>
</tr>
<tr>
<td>Naphthalene</td>
<td>32.5</td>
<td>10.7</td>
</tr>
<tr>
<td>Guaiacol</td>
<td>68.7</td>
<td>25.0</td>
</tr>
<tr>
<td>2-Methoxy-4-methylphenol</td>
<td>63.8</td>
<td>20.1</td>
</tr>
<tr>
<td>Phenol</td>
<td>721.8</td>
<td>482.2</td>
</tr>
<tr>
<td>p-Ethylguaiacol</td>
<td>83.0</td>
<td>8.0</td>
</tr>
<tr>
<td>p-Cresol</td>
<td>143.0</td>
<td>47.9</td>
</tr>
</tbody>
</table>
Figure 2
Perotti P

Beans

Liquors
Figure 3
Perotti P
A careful tuning of the sample preparation conditions was required to improve informative analytes recovery. Results of the experimental design for HS-SPME sampling optimization are reported below together with the experimented parameters.

The following sampling conditions were tested (Kolb & Ettre, 2006):

a) Sampling temperature (50 and 80°C)

b) Evaluation of different sample amount (1-3 g) and polymer coatings (PDMS/DVB, DVB/CAR/PDMS, PDMS, and PDMS/DVB coated with PDMS for in-solution sampling)

c) Modifying analytes’ solubility

*Increase of the sampling temperature*: the sampling temperature was increased from 50°C to 80°C keeping constant the other conditions (1 g sampled for 40 min) to improve the transfer of the low volatility compounds to the headspace. The bar chart (Figure 1S a) shows the increase of the phenolic derivatives signals due to the effect of the temperature on the analytes.

Evaluation of different sample amounts and polymer coatings: different coating fibers and amount of sample were tested in consideration of the nature of the investigated markers. As expected (Figure 1S b), DVB/CAR/PDMS fiber shows the highest analyte recovery. On the other hand, an increase of sample amount does not improve signals for PDMS/DVB coating, as already observed by Mejias et al. with other SPME coatings (Castro Mejías et al., 2003; Jelen, 2006).

c) Changing analytes solubility: analyte solubility in a food matrix can also influence their matrix/headspace partition coefficient, as well as their recoveries. Cocoa beans are solid where analytes are adsorbed, the strength of their interaction can therefore be modified through: 1) Salting out applied to the suspended matrix: suspension was made by adding a displacer (2 mL of water solution with 30% of NaCl) to 1 g of grinded beans. Suspension in a high ionic strength solution should afford both the displacement of the retained analytes in the liquid phase and a better partition between the displacer and headspace Figure 1S c); this approach does not result in an increased abundance of the investigated analytes with the DVB/CAR/PDMS fiber.
2) Salting out applied to water suspended samples in combination with direct immersion SPME: 1 g of beans was suspended in 20 mL of a saturated NaCl water solution (Figure 1S d) and sampled with different polymeric coatings (PDMS/DVB, PDMS/DVB coated with PDMS). Salting out of the suspended sample displays recoveries not comparable to those with untreated grounded beans at 80°C for almost all compounds investigated. Direct immersion shows a very poor recovery. Figure 1S d shows a comparison of the recovery of analytes under investigation with the different sampling approaches using PDMS/DVB coating.

References


Figure S1.