Drug Testing and Analysis

Systematic optimization of ethyl glucuronide extraction conditions from scalp hair by design of experiments and its potential effect on cut-off values appraisal

Journal:	*Drug Testing and Analysis*
Manuscript ID	DTA-18-0070.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Alladio, Eugenio; Università di Torino , Department of Chemistry; Centro Regionale Antidoping Biosa, Giulia; Università di Torino , Department of Chemistry Di Corcia, Daniele; Centro Regionale Antidoping Seganti, Fabrizio; Centro Regionale Antidoping Salomone, Alberto; Centro Regionale Antidoping Vincenti, Marco; Università di Torino , Department of Chemistry; Centro Regionale Antidoping Baumgartner, Markus; University of Zurich, Zurich Institute of Forensic Medicine
Keywords:	Ethyl glucuronide, Hair EtG, Design of experiments, Hair analysis, Cut-off
Abstract:	The quantitative determination of ethyl glucuronide (EtG) in hair samples is consistently used throughout the world to assess chronic excessive alcohol consumption. For administrative and legal purposes, the analytical results are compared with cut-off values recognized by regulatory authorities and scientific societies. However, it has been recently recognized that the analytical results depend on the hair sample pretreatment procedures, including the crumbling and extraction conditions. A systematic evaluation of the EtG extraction conditions from pulverized scalp hair was conducted by design of experiments (DoE) considering the extraction time, temperature, pH, and solvent composition as potential influencing factors. It was concluded that an overnight extraction at 60°C with pure water at neutral pH represents the most effective conditions to achieve high extraction yields. The absence of differential degradation of the internal standard (isotopically-labeled EtG) under such conditions was confirmed and the overall analytical method was validated according to SGWTOX and ISO17025 criteria. Twenty real hair samples with different EtG content were analyzed with three commonly accepted procedures: (a) hair manually cut in snippets and extracted at room temperature; (b) pulverized hair extracted at room temperature; (c) hair treated with the optimized method. Average increments of EtG concentration around 69% (from a to c) and 29% (from b to c) were recorded. In light of these results, the authors urge the scientific community to undertake an inter-laboratory study with the aim of defining more in detail the optimal hair EtG detection method and verifying the corresponding cut-off level for legal

	enforcements.

Systematic optimization of ethyl glucuronide extraction conditions from scalp hair by design of experiments and its potential effect on cut-off values appraisal

Eugenio Alladio[a,b], Giulia Biosa[b], Fabrizio Seganti[a],Daniele Di Corcia[a], Alberto Salomone[a], Marco Vincenti[a,b,*], Markus R. Baumgartner[c]

a Centro Regionale Antidoping ''A. Bertinaria'', Regione Gonzole 10, 10043 Orbassano, Turin, Italy

b Dipartimento di Chimica, Universita` degli Studi di Torino, via P. Giuria 7, 10125 Turin, Italy

c Zurich Institute of Forensic Medicine, University of Zurich, Kurvenstrasse 17, 8006 Zurich, Switzerland

Corresponding author:

Prof. Marco Vincenti

Dipartimento di Chimica, Università degli Studi di Torino,

Via Pietro Giuria, 7 – 10125 Torino, Italy

Phone: +390116705264 – Mobile: +393474198878

E-mail: marco.vincenti@unito.it

Short title: Dependence of hair EtG cut-off appraisal on the extraction conditions

Keywords: Hair EtG, Ethyl glucuronide, Design of experiments, Hair analysis, Cut-off

Abstract

The quantitative determination of ethyl glucuronide (EtG) in hair samples is consistently used throughout the world to assess chronic excessive alcohol consumption. For administrative and legal purposes, the analytical results are compared with cut-off values recognized by regulatory authorities and scientific societies. However, it has been recently recognized that the analytical results depend on the hair sample pretreatment procedures, including the crumbling and extraction conditions. A systematic evaluation of the EtG extraction conditions from pulverized scalp hair was conducted by design of experiments (DoE) considering the extraction time, temperature, pH, and solvent composition as potential influencing factors. It was concluded that an overnight extraction at 60°C with pure water at neutral pH represents the most effective conditions to achieve high extraction yields. The absence of differential degradation of the internal standard (isotopically-labeled EtG) under such conditions was confirmed and the overall analytical method was validated according to SGWTOX and ISO17025 criteria. Twenty real hair samples with different EtG content were analyzed with three commonly accepted procedures: (a) hair manually cut in snippets and extracted at room temperature; (b) pulverized hair extracted at room temperature; (c) hair treated with the optimized method. Average increments of EtG concentration around 69% (from a to c) and 29% (from b to c) were recorded. In light of these results, the authors urge the scientific community to undertake an inter-laboratory study with the aim of defining more in detail the optimal hair EtG detection method and verifying the corresponding cut-off level for legal enforcements.

Introduction

Ethyl glucuronide (EtG) is a minor phase II metabolite of ethanol normally produced after consumption of alcoholic beverages. It can be analytically detected in urine, blood, and oral fluid in order to ascertain recent alcohol intake[1]. Moreover, EtG is nowadays extensively used as an effective biomarker to assess prolonged abstinence or chronic excessive alcohol consumption, provided that its determination is made on a keratin matrix, typically scalp hair[2]. The applications of EtG determination in hair range from compliance to driving regulation[3] to workplace testing[4,5] and many other usages of clinical and forensic interest[6]. Hair samples different from scalp hair can cautiously be analysed, in case scalp hair is not available or is degraded[7–9]. In practice, the effectiveness of EtG determination in hair as a biomarker for chronic excessive alcohol consumption outperforms all the other alcohol biomarkers[10–12] to the extent that is frequently used as a unique laboratory testing. However, several studies have shown that bias can be induced by cosmetic treatments[13] exposure to chlorinated water[14], external contamination by EtG-containing lotions[15] and many others[16]. The practical convenience of hair EtG as a biomarker explains its widespread use, which account for its hundreds of thousands of yearly determinations worldwide, mainly used for driving license renewal and rehabilitation, and workplace testing.

The Society of Hair Testing (SoHT) established cut-off values for hair EtG concentration that supports judgments of chronic excessive alcohol consumption (30 pg/mg) and non-contradiction with self-reported abstinence (7 pg/mg). These cut-off values were originally determined on the basis of several prevalence and observational studies, meta-analyses, and prudential considerations[2]. Previous SoHT consensus documents[17] are regularly updated and, in the most recent issue, the SoHT also recommends to ‘‘powder hair prior to the extraction of EtG’’[18,19]. Several studies supported the conclusion that higher EtG extraction yields are obtained if the hair aliquot is pulverized in a mill instead of being manually cut into small snippets[20–23]. Recently, another study proved that also the choice of the extraction solvent and temperature significantly affected the EtG extraction yield from hair and its detected concentration[24]. In particular, the study demonstrated that EtG extraction with water is more effective than with methanol and conducting the extraction at 60°C provides more exhaustive recovery that at ambient temperature[24]. Notably, previous SoHT consensus documents prescribed well-defined cut-off values and criteria for their interpretation, but provided very little hints about the analytical and instrumental methods, leaving to the specialist the choice and demonstration of equivalence with the best practices. Clearly, if the extraction yield, and consequently the EtG detected concentration, depend on the hair sample pretreatment, then also the cut-off values becomes questionable[23].

In the present study, we took up the proposal of Mueller and coworkers[24] of using multifactorial experimental design to investigate the dependence of EtG extraction on several experimental factors and expanded it further, with closer sampling of the experimental domain and consecutive modelling. In general, Design of Experiment (DoE) strategies allows to reduce the experimental effort and simultaneously increase the quality of obtained information[25–27]. In our case, DoE was exploited to obtain robust interpretation of the factors that impact on EtG extraction yield and their reciprocal interactions. Then, the optimized analytical method was validated with a stepwise, analyst-independent protocol. Lastly, we measured the combined effect of hair milling and optimal extraction conditions with respect to previous analytical procedures on a series of real hair samples and discuss the consequences in the forensic toxicology context.

Materials and Methods

Analytical Method

The determination of EtG in hair samples was initially performed via a UHPLC–MS/MS method (a) that was validated according ISO/IEC 17025 criteria[28], accredited in 2013[29], and subsequently revised in 2016[23] (subsequently referred to as method (b)), when the hair sample pre-treatment procedure was modified according to the superior efficiency of the milling technique[23] with respect to the previous method (a) of cutting hair into small segments before the extraction step[29]. In summary, our initial analytical method (b) – which was submitted to systematic optimization in the present study - applied the following steps:

1. the collected hair samples (about 40-50 mg) corresponding to the proximal 0–3 cm segments were weighted and then washed twice using methylene chloride and methanol in sequence;
2. the dried hair were pulverized in a Polypropylen Co-Polymer (PPC) tube using a metal beads mill Precellys 24 Tubes Homogenizer (Bertin Pharma, France), equipped with six 2.8 mm metal beads;
3. internal standard (IS, EtG-D_5 at 100 pg/mg final concentration) was added;
4. EtG extraction was performed overnight at room temperature (ca. 20°C) with a 35:1 water:methanol (v/v) mixture;
5. lastly, the sample was sonicated and an aliquot of the liquid phase was transferred into a vial for UHPLC–MS/MS analysis, performed by injecting 3 µL of hair extract into a Shimadzu Nexera 30 UHPLC-system (Shimadzu, Duisburg, Germany) interfaced to an AB Sciex API 5500 triple quadrupole mass spectrometer (AB Sciex, Darmstadt, Germany).

At the end of the optimization work, the extraction conditions were modified as follows (method (c) – as subsequently referred to):

4. EtG extraction was performed with 500 µL of pure water, overnight at 60°C;

In the present study, each hair batch was extensively mixed, homogenized, decontaminated, and dried, following the procedure described above. The milling procedure was executed on about 50 mg of sample. Then, 500 µL of different extraction solvents/mixtures (according to the planned DoE) and 5 µL of internal standard EtG-D_5 (using a working solution of 1 ng/µL in methanol) were added. A short centrifugation (1.5 min, 13300 rpm, 17000 x g – VWR Micro Star, Leuven, Belgium) was executed to completely submerge the hair material within the extraction solvent and to remove air bubbles. The extractions were performed within a laboratory stove at different temperatures and extraction times in accordance with the drafted DoE plans. The stove temperature was controlled immediately before and after the extraction period. Lastly, an ultra-sonication of 1.5 hours was performed and 100 µL of the liquid phase was transferred into a clean vial to be analysed by UHPLC-MS/MS. Further details about instrumental conditions are available in our previous publications[23,29].

Hair specimens

The hair samples used in this study were collected from anonymized residual specimens, stored and available in our laboratory, formerly belonging to individuals who underwent hair analysis before October 2015, with resulting EtG values higher than the validated limit of quantitation (LOQ) of the UHPLC-MS/MS method (i.e., 1 pg/mg). Samples with EtG values higher than 12 pg/mg were specifically selected, in order to avoid any misinterpretation of DoE results due to the higher uncertainty of the lower points of the calibration curves. In particular, DoE was performed on two large batches of hair samples: the first one (A) contained only hair locks with EtG concentrations comprised between 13 pg/mg and 20 pg/mg, while the second one (B) contained the specimens with EtG values comprised between 40 pg/mg and 100 pg/mg (namely, the ones collected from subjects identified as excessive alcohol drinkers). The reason to consider two batches at different EtG concentrations in the DoE plans was to evaluate the results at EtG levels considered respectively above and below the 30 pg/mg cut-off suggested by the Society of Hair Testing[17,18]. All hair samples used in the present study had been originally analysed before the introduction of the milling protocol in our laboratory (i.e., October 2015). The concentration intervals indicated above refer to the original pre-treatment procedure[29].

Design of Experiments (DoE)

According to a previous study[24], the choice of the extraction solvent and temperature turned out as the most significant factors that influence the EtG extraction yield from hair. The first DoE (i.e. Preliminary DoE) set-up examined in detail these two parameters. A second DoE (i.e. Optimization DoE) set-up also considered the pH of the solvent and the extraction time as valuable factors to be examined. In detail, the first DoE consisted in a 2-factors full-factorial design performed by varying the extraction temperature (the first factor, T) and the composition of the extraction mixture (the second factor). Four levels were selected for the first factor, namely 20°C, 32°C, 45°C, and 57°C, and three compositions (levels) for the second factor, respectively distilled water, water/methanol 35:1 (v/v) and water/methanol 17:1 (v/v). Since four and three levels were evaluated for extraction temperature and extraction solvent, respectively, and each experiment was replicated three times, a total number of 36 experiments ($4 \times 3 \times 3 = 36$) was executed on each batch (A and B), and 72 experiments overall. A geometric representation of the Preliminary DoE is shown in Figure 1a, where each point on the square represents one experiment. The levels of the evaluated extraction temperatures were coded from −1.5 up to +1.5 (i.e., −1.5, −0.5, +0.5, +1.5 for 20°C, 32°C, 45°C, and 57°C, respectively), while the codes for the different extraction mixture were −1, 0, +1 for pure distilled water, water/methanol 17:1, and water/methanol 35:1, respectively. The experiments were executed in random order, and coded as reported in Table S1 of the Supplementary Material.

The Optimization DoE was planned with a face-centred central composite design (corresponding to 15 different experimental conditions), where extraction temperature, extraction time and solvent pH (distilled water) were selected as variable factors. Three levels were chosen for all factors following the results obtained from the first DoE: (i) 45°C, 54°C and 63°C for the extraction temperature (T), (ii) 1 h, 8 h and 16 h (overnight) hours for the extraction time (t), and (iii) 5.5, 7.0 and 8.5 for the pH of the extraction solvent (water). Acidic and basic pH values were obtained by adding HCl and NaOH 0.1 M, respectively, and the pH constancy at the end of the extraction was positively verified. Since each experiment was performed in triplicate for both batch A and B, a total number of 90 experiments ($15 \times 3 \times 2 = 90$) was executed in the second DoE. All levels were coded from −1 to +1. A geometric representation of the performed DoE is shown in Figure 1b. Again, the experiments were performed in random order and the respective codes are reported in Table S2 of the Supplementary Material. For validation purposes, 10 replicates were completed for both batches at the end of the study, at the experimental conditions identified as optimal according to the response surfaces of a Multiple Linear Regression (MLR) model.

Differential degradation of the internal standard

Since the second DoE tested extraction conditions at relatively high temperatures (i.e., much higher than the traditional room temperature), the occurrence of differential degradation of the internal standard (EtG-D_5) was evaluated at the new experimental settings. The experiments were carried out by adding 5 µL of EtG-D_5 to 500 µL of pure water (the novel extraction solvent). Afterwards, different combinations of five extraction temperatures (room temperature, 35°C, 45°C, 55°C and 65°C) and three extraction time (1, 8 and 16 hours) were tested in triplicate, for an overall of 45 experiments. Boxplots and Kernel Density Estimation (KDE) curves were calculated to interpret the results.

Method validation

The new analytical method was validated using a stepwise, analyst-independent protocol that required the preparation of seven independent calibration curves[30,31], prepared in three different days, at seven calibration levels: 2, 5, 10, 30, 50, 100, and 300 pg/mg. Most validation parameters were determined from these data, including linearity range, limit of detection (LOD), limit of quantification (LOQ), selectivity, specificity, trueness, accuracy, repeatability and carry-over effect, in accordance with ISO/IEC 17025 and SWGTOX requirements[28,32]. At first, the linearity parameter was investigated by initially evaluating the homo-/heteroscedasticity of the data, followed by the estimation of the order (linear or quadratic), and weight (1, 1/x or $1/x^2$) of the calibration curve. The linearity was checked by lack-of-fit and Mandel tests[33,34]. Determination coefficient (R^2), relative standard deviation of the slope, normality of the standardized residuals, and deviation from back-calculated concentrations were also evaluated using in-house spreadsheets, package mvtnorm[35,36], and the routines developed by B. Desharnais et al.[30,31]. LOD and LOQ were estimated by the Hubaux-Vos algorithm[37].

Specificity was assessed by analysing seven blank head samples from acknowledged teetotaller individuals; in particular, the presence/absence of interfering ions on each single-ion chromatograms was evaluated, with reference to EtG pure standard. Then, the data collected for the preparation of the seven calibration curves were used to evaluate selectivity, accuracy, trueness, intra-assay precision, and repeatability. In particular, the data collected for a specific calibration curve were quantified by using a different calibration curve, prepared the same day or in a previous day (i.e., simulating our routine approach to test the method prior of a working session). This

procedure allowed us to manage each set of data as independent. Therefore, 7 samples (from 7 batches) per each calibration level were utilized to evaluate the validation parameters previously cited. The consistency of EtG retention time was successfully verified for all 49 (7×7) samples, as well as the relative intensities of the characteristic ions. Trueness, and intra-assay precision were estimated as percent bias and CV%, respectively. Satisfactory results were expected to be within ±15% for the 1st and 4th calibration levels (i.e., 2 and 30 pg/mg) and within ±20% for the the 7th calibrator (i.e., 300 pg/mg). The repeatability was determined at the 1st, 4th and 7th calibration levels; moreover, Shapiro-Wilk, Dixon, and Grubbs tests were performed to investigate the Gaussian distribution of the data and the occurrence of outliers. Finally, the occurrence of carry-over effect was tested by injecting one distilled water sample after the highest point of each calibration curve (i.e., 300 pg/mg), for seven times; the appearance of unintended EtG signal was supposed not to exceed the 10% of the signal of the lowest calibrator (i.e., 2 pg/mg).

Matrix effect

A final investigation was conducted on matrix effects in order to evaluate possible differences between manual cutting and mill-pulverisation of the hair specimen. Matrix effect was evaluated from six replicates by comparing the experimental results from neat aqueous solutions spiked with EtG at three concentration levels (low level = 10 pg/mg, mid-level = 50 pg/mg, and high-level = 300 pg/mg), with the data obtained from negative hair samples (collected from 1-3 years old children) that underwent milling or manual cutting procedures, then spiked at the same levels after the extraction step. The matrix effect for each pre-treatment procedure was expressed as the percentage ratio between the measured concentrations.

Comparison of real samples

At the end of the optimization process, the new protocol was compared with the previous validated method[23], involving the manual cutting of hair locks into 2-3 mm snippets on a set of real hair samples. Two batches of 10 hair samples each were analysed: the first batch (C) contained only samples with EtG values originally detected in the range between 20 pg/mg and 31 pg/mg, while batch D included specimens with EtG values above 60 pg/mg. Afterwards, boxplots, t-test and ANOVA test were performed to compare the analytical results.

Software

All statistical analyses were conducted using the software R Studio version 1.0.153[38], while DoE interpretation was performed with an R package developed by the Italian Group of Chemometrics of the Italian Society of Chemistry (SCI), freely available on internet[39], in the 3.1.0 version[40].

Results and Discussion

Design of Experiment – preliminary plan

The goal of the first DoE plan was to test our validated analytical method in the light of the results obtained by Mueller et al.[24], who suggested to carry out the EtG extraction on pulverised hair at 60 °C (instead of room temperature) with pure water as the extraction solvent. Taking advantage of their differentiation between influencing and non-influencing experimental factors[24], we decided to focus the study on the significant ones (temperature and solvent) and develop a DoE based on a more detailed tuning of these selected parameters. In the same time, we verified the inter-laboratory repeatability of their conclusions. In particular, four temperature levels were checked (from ambient to 57 °C) and three solvent compositions, in which methanol is used at low percentages (0%, 2.7%, and 5.6%) as a modifier of the prevalent aqueous constituent.

From the 2 × 36 experiments carried out within the first DoE, the ratio of the target analyte area to the IS provided the corresponding "extracted" EtG concentrations, assuming that the recovery of the analyte from the real matrix was variable while that of the spiked EtG-D_5 was complete. All the analytical results are reported in Table S1 of the Supplementary Material for both batches A and B. According to the full-factorial design, a multiple linear regression (MLR) model was calculated with reference to the following equation:

$$y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2$$

where y represents the concentration of the extracted EtG (pg/mg), b_i represent the regression coefficient (0 = intercept, 1 = coefficient relative to the extraction temperature, 2 = coefficient relative to the extraction solvent) and x_i stands for the evaluated parameters (1 = the extraction temperature, 2 = the extraction solvent). The quadratic terms for both temperature (x_1^2) and solvent composition (x_2^2) were evaluated, too. From the model regression, Figure 2 displays the values of b coefficients and the relative significance of each factor, as determined by a t-test. The coefficient plot reported in Figure 2 is relative to the batch A (13-20 pg/mg) and shows that the extraction temperature is the most significant factor that affects the EtG extraction (p-value < 0.001), with a

positive trend (i.e., the higher the temperature, the higher the extraction yield). Analogous conclusions were drawn from the coefficient plot relative to the batch B (40-100 pg/mg) reported in Figure 1 of the Supplementary Material), although the significance level of the temperature factor is lower (p-value < 0.05). The latter diagram also shows a significant negative influence of the quadratic term relative to the extraction temperature (x_{11}), which suggests, at the current stage, to regulate the extraction temperature at a relatively high but not extreme value. The two-dimensional response surfaces reported in Figure 2 (Batch A) and Figure S1 of the Supplementary Material (Batch B) confirms the conclusions of Mueller and co-workers[24]: the maximum response value for both batches A and B was observed in the lower-right sections of the graph, indicating that the highest extraction yield was reached when (i) only water was employed as the extraction solvent (coded as -1.0 on the y-axis), along with (ii) an extraction temperature setting in the range between 45°C and 57°C (coded as 0.5 and 1.5 on the x-axis).

Design of Experiment – optimization plan

The need of accurate temperature adjustment together with the opportunity to reduce the extraction time inspired the second DoE scheme, which also investigated the solvent pH as a potential influencing factor. According to the conclusions reported above, the second DoE was planned using water as the extraction solvent, and a restricted interval of extraction temperatures was tested, ranging from 45°C to 63°C, with a 9°C interval step. The pH of the aqueous solvent was varied from slightly acidic to slightly basic (5.5, 7.0, and 8.5) while the extraction time tested both day-time and overnight conditions (1 h, 8 h, and 16 h) for practical reasons. In the present case, a face-centred central composite design was selected because its experimental space covered a wide range of useful setting, including the extreme conditions for all parameters. To interpret the data, optimization of a MLR model with the following formula is suggested:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2$$

where y and b_i have the same meaning as in the preceding section, and x_i stands for the evaluated parameters (1 = extraction temperature, 2 = extraction time, 3 = solvent pH). The quadratic terms for all factors, namely the extraction temperature (x_1^2), time (x_2^2) and pH (x_3^2) were also evaluated. Figure 3 reports the histogram values of b coefficients and their significance: in the present case, both the extraction temperature and time proved highly significant for the extraction efficiency of EtG (p-value < 0.001), with positive trends, while the pH of the solvent does not appear to have an influence, at least within the tested range. Likewise, the interaction and quadratic terms of the

model show relatively large variance and consequently no significance. The overall conclusion is that increasing both extraction temperature and time - independently from one another - results into an increase of the extraction yield. Analogous deductions were drawn from the coefficient plot relative to the batch B, reported in Figure S2 of the Supplementary Material. It must be noted that a significant, positive contribution of the quadratic x_{11} term (i.e. the extraction temperature) has been observed for batch B. Nevertheless, we do not accredit substantial importance to this positive contribution (partly evident also for batch A) because quite large variance is associated to both the interaction and quadratic terms and an opposite - yet not significant – negative contribution was detected in the preliminary DoE.

Maximum response value for both batches A and B was observed in the upper-right parts of the response surface graphs, at the point encoded as [1, 1, 0] (see Figure 3 for batch A, and Figure S2 of the Supplementary Material for batch B). This means that the highest extraction yield for EtG was reached when simultaneously (i) the extraction temperature was set at 63°C (coded as 1.0 on the x-axis), and (ii) the extraction time was set at 16 hours/overnight (coded as 1.0 on the y-axis). In contrast, all the response surfaces provided similar behaviour at any pH value tested (all the response surfaces are reported in Figure S3 of the Supplementary Material), confirming pH as a non-significant factor. In the subsequent experiments, neutral pH of the aqueous solvent was consistently used, for simplicity.

At the end of the second DoE plan, both the observed and estimated results indicated that the optimization of the extraction conditions produced a significantly higher EtG extraction yield. Consequently, the original method was modified by substituting pure distilled water as the extraction solvent in place of a 35:1 water:methanol (v/v) mixture, and employing an overnight extraction temperature of 60°C, instead of room temperature. Nevertheless, the final experimental setting had not been directly tested within the DoE plan, but rather high extraction time and temperature were tested at both acidic and basic conditions. Therefore, ten replicates were executed at the point encoded [1, 1, 0] (i.e., T = 63°C, t = 16 hours and pH = 7) for both batches in order to validate the model. The ten replicates for batch A gave an estimated average value of 42 pg/mg, with an estimate standard deviation of 5 pg/mg and a coefficient of variation (CV%) of 12%. Then, the estimate of the experimental response at the tested point was calculated via the following formula:

$$\bar{y} \pm \frac{t \cdot s}{\sqrt{n}}$$

where $\bar{y}$ represents the estimated average of the ten replicates, t is the tabulated Student's t-value at the 0.05 significance level (95%, d.f. = 9, t = 2.262), s is the estimate of the experimental standard deviation, and n is the number of replicates. The resulting EtG concentration in batch A is calculated as 42 ± 4 pg/mg. On the other hand, the estimated EtG value that was calculated by the MLR model at the same experimental point [1, 1, 0] is equal to 48 pg/mg, with an experimental uncertainty of 10 % (48 ± 5 pg/mg). Similarly for batch B (that provided an estimated average value of 154 pg/mg, with an estimate standard deviation of 25 pg/mg and a coefficient of variation (CV%) of 16%), the extracted EtG concentration was equal to 154 ± 18 pg/mg, while the MLR model estimate was calculated as 140 ± 14 pg/mg with positive overlapping of the intervals. Since the experimental values were not significantly different from the predicted concentrations for both batches A and B, the model was validated and could be applied in the entire experimental domain. Notably, both average experimental concentrations for batches A and B largely exceed the values originally determined for the single hair samples that form the batches.

Test on the internal standard

The optimization of the analytical method led to increase the extraction temperature from ambient to 63 °C, while keeping the extraction time fixed at 16 h (overnight). While it was experimentally verified that no solvent evaporation occurred from the sealed vial used for the extraction, one can doubt that the apparent increase of the extracted EtG concentration may actually arise from partial degradation of the EtG-D_5 internal standard, whose chromatographic peak area is used as the measurement unit for the analyte concentration computation. In practice, an artificial decrease of the EtG-D_5 concentration would result in an over-estimation of the extracted EtG, leading to artificially increased concentrations. To check the absence of differential degradation of the EtG-D_5 internal standard, 45 experiments were carried out at five temperatures (room temperature, 35°C, 45°C, 55°C, and 65°C) and three extraction time (1 h, 8 h, and 16 h). Figure 4 shows the results in the form of boxplots for the time variable (a) and the temperature variable (b). No significant variations of the EtG-D_5 areas were observed at the different levels of both extraction temperature and time. Possibly, a slight non-significant increase of the EtG-D_5 extraction is observed by increasing the extraction time from 1 h to 16 h. These results, together with the limited number of outliers in the boxplots – with the exception of T = 35°C graph – confirm the absence of any differential degradation of EtG-D_5 over the entire experimental domain. Incidentally, also the robustness of EtG under the tested experimental conditions is confirmed. It can be concluded that the changes of the measured EtG concentration observed under different experimental settings in

DoE experiments are actually due to the different extraction yields of the analyte. Despite the response surfaces of the Optimization DoE (Figure 3 and S2) suggested the possibility to raise the extraction temperature even more, no further DoE were performed in order not to damage the hair matrix during the extraction process. In our opinion, the current extraction temperature of 63°C represent a robust compromise between the extraction yield of EtG and the feasibility of the analytical methodology.

Method validation

The stepwise systematic method proposed by Desharnais et al.[30,31] was used to select the most appropriate calibration model and validate the choice. The first step of the procedure involved the evaluation of data heteroscedasticity by means of F-test; then different statistical tests were executed, including lack-of-fit and normality testing, in order to choose the model order, either linear or quadratic, that best fitted the experimental calibration points (7 levels × 7 replicates), and the corresponding weighting. According to this procedure, the data proved to be heteroscedastic, and a linear model involving the use of $1/x^2$ weighting turned to be the most appropriate for calibration purposes. All the results of significance tests are reported in the Supplementary Material, together with the information about the the slope and the intercept of the tested calibration model, and its determination coefficient, in the output format provided by the R codes developed and made available by Desharnais and coworkers[30,31]. The whole procedure was repeatedly tested on a lower number of the already prepared calibration curves (i.e., including 4 or 5 replicates only) to test the model robustness and similar results were obtained.

From the final calibration model, LOD and LOQ values were calculated following the Hubaux-Vos' algorithm[37], which yielded the following values: LOD = 0.8 pg/mg and LOQ = 1.7 pg/mg. The latter concentration is lower than the first calibration level, which was experimentally verified (see below). Selectivity and specificity of the method were confirmed, as no interfering signals were detected at the retention times of the target analytes, and the retention time precision proved satisfactory, as the deviations from the expected retention times were largely below 1%. The relative abundancies of the characteristic ions of EtG were positively evaluated. Trueness and accuracy data at the 1st, 4th and 7th calibration levels (2, 30, and 300 pg/mg) turned out adequate, as the percent bias and CV% values were lower than 15% at all concentration levels. Moreover, repeatability was tested at the 1st, 4th and 7th calibration levels giving satisfactory results since all the performed significance tests were passed (i.e., Shapiro-Wilk, Dixon, and Grubbs tests). Lastly, no carry-over effect was observed.

Matrix effect

In our previous study[23], it was concluded that – on average – significantly higher EtG extraction yields were obtained if the hair samples were primarily pulverized with a ball mill, instead of cutting them manually in small snippets. However, the differences for individual samples were highly variable and, for a few hair samples, even higher EtG concentrations were measured after applying the cutting pre-treatment than after milling. These rare events could be explained by considering that relatively high random variability is generally associated to incomplete extraction yields. An alternative explanation, that we intended to verify in the present study, was that different matrix effects may be produced by the two pre-treatment procedures, as a consequence of dissimilar abundance of interfering substances.

Comparison of the matrix effect induced by the two procedures was made by analysing three neat aqueous solutions spiked with EtG at 3 concentration levels and comparing their results with those obtained from negative hair samples that underwent milling or manual cutting procedures and spiked after the extraction step. The use of childhood hair as negative samples may limit the general legitimacy of the comparison, since it implies that the matrix components are the same in the childhood and adult age, but represented a practical way to obtain a mixed batch of several unquestionably negative hair samples. The six replicated determinations at three concentration levels produced very limited variability (CV% = 7-9 for milling and CV% = 4-7 for cutting experiments). The measured average matrix effect was equal to –6.9%, –6.6%, and –6.4% at the three concentration level when the milling procedure was applied, while was equal to –7.4%, –6.1%, and –7.2% when the hair was manually cut into snippets. For all these data reporting the signal decrease due to matrix effects, the t-test yielded statistical significance at 90% confidence level or above. The differences between the milling and the manual cutting procedures was never statistically significant. If the data were corrected by the contribution of the internal standard, the measured matrix effect was calculated as +2.0%, +1.4, +2.6% and –1.8%, +3.4, +7.0, respectively. The latter results are affected by larger uncertainty as a result of the added contribution of the internal standard variability.

The overall results confirm that limited and substantially equal matrix effect is observed no matter what pre-treatment procedure is used to crumble the hair samples and assures the compliance of both pre-treatment strategies with respect to the modest impact of matrix component on EtG quantitation.

Comparison of real samples

20 hair samples belonging to batches C and D (see Experimental), originally analysed with method (a) involving manual cutting of the hair lock[29], were analysed again using method (b), namely the accredited method presently in use in our laboratory[23] that involves the pulverisation of the hair lock, and also with the new method (c) optimized by DoE involving extraction with pure water at a temperature of 63°C. The summary and detailed results are reported in Table 1 and Figure 5.

The data confirm our previous conclusion[23] that the average increase of the extraction efficiency when the hair matrix is pulverized with a mill rather than manually cut into snippets exceeds 30%. Remarkably, seven samples out of ten (batch C) exhibited EtG concentrations above the 30 pg/mg cut-off when they were analysed with method (b), whereas this occurred with only one out of ten samples with method (a). Of course, it should be reminded that specific selection of samples with EtG close to the cut-off was performed. All ten samples exceeded the cut-off when they were analysed with method (c). By comparing method (b) with method (c), another 30% average increase of extraction efficiency is inferred from the data. The comparison made on parallel determinations on aliquots of the same real hair samples unequivocally shows that modifying the extraction temperature from ambient to 63 °C increased the detected EtG concentration for all 20 samples, most likely because more exhaustive extraction is achieved. This improvement is recorded in both the medium and high EtG concentration levels to a comparable extent. The boxplots represented in Figure 5 gives a clear graphical evidence of the dependence of the quantitative results from the experimental conditions adopted for sample treatment.

ANOVA and unpaired (two-sided) t-test expressed in a quantitative way the significance level of the differences observed between the data obtained from the novel and the old conditions of EtG extraction. In the comparison between methods (b) and (c), the t-test yielded p-values of 1.5×10^{-5} and 9.7×10^{-5} for batches C and D, respectively, rejecting the null hypothesis of no difference. In the same comparison, ANOVA test gave p-values of 1.3×10^{-4} and 2.6×10^{-3} for batches C and D, respectively, rejecting the null hypothesis (samples are not different).

Conclusions

The present study supports the conclusions of Mueller and coworkers, with more detailed investigation of the experimental domain, that (i) pure water represents the best solvent to extract

EtG from scalp hair, (ii) an extraction temperature above 60 °C achieves more exhaustive EtG recovery. We also verified that extending the extraction time overnight allows safe and reproducible recovery conditions without implying any risk of analyte and internal standard decomposition. Furthermore, it was observed that limited changes of the pH (from slightly acidic to slightly basic) had no impact on the extraction yield.

The major achievement of the present study is the unequivocal demonstration that several experimental parameters strongly influence the results of the analysis on each tested hair sample. These include particularly the hair crumbling method and the extraction conditions. It is highly plausible that the different analytical results are due to a dissimilar extraction efficiency. Moreover, it is well known that the recovery variability generally depends on its absolute value and that maximizing the extraction yield reduces its variability[41–43]. It can be deduced that achieving the most exhaustive extraction of EtG represents a valuable goal in order to obtain stable and reliable analytical results.

In most clinical determinations, each laboratory has its own reference population on which normality ranges are calculated, that depend on the experimental method. On the other hand, in forensic toxicology, fixed cut-off values are commonly defined for general use worldwide, due to the legal consequences of the analytical determinations. Unlike most drugs of abuse, one has to put high requirements on the quantitative determination of EtG in hair samples in order to discriminate the different consumption profiles, which in turn are defined by the use of cut-off values. Therefore, a significant effort should be made in the future to define cut-off values based on the most effective operating conditions and possibly on controlled administration studies, although within the inherent ethical constrains[44]. The authors strongly suggest that several institutions should collaborate within an inter-laboratory comparison in order to amend the forthcoming issues of consensus documents.

References

[1] Høiseth G, Yttredal B, Karinen R, Gjerde H, Mørland J, Christophersen A. Ethyl glucuronide concentrations in oral fluid, blood, and urine after volunteers drank 0.5 and 1.0 g/kg doses of ethanol. *J Anal Toxicol* 2010;34(6):319-24. https://www.ncbi.nlm.nih.gov/pubmed/20663284 Accessed February 12, 2018.

[2] Pragst F. Alcohol Biomarkers in Hair, in: Kintz P, Salomone A, Vincenti M, *Hair Analysis in Clinical and Forensic Toxicology*, 1st ed. San Diego, CA, USA: Academic Press, Elsevier Inc; 2015: 71–139.

[3] Agius R, Nadulski T, Kahl HG, Dufaux B. Ethyl glucuronide in hair - A highly effective test for the monitoring of alcohol consumption. *Forensic Sci Int* 2012;218(1-3):10-4. doi: 10.1016/j.forsciint.2011.10.007.

[4] Salomone A, Tsanaclis L, Agius R, Kintz P, Baumgartner MR. European guidelines for workplace drug and alcohol testing in hair. *Drug Test Anal* 2016;8(10):996-1004. doi: 10.1002/dta.1999.

[5] Hastedt M, Herre S, Pragst F, Rothe M, Hartwig S. Workplace alcohol testing program by combined use of ethyl glucuronide and fatty acid ethyl esters in hair. *Alcohol Alcohol* 2012;47(2):127-32. doi: 10.1093/alcalc/agr148.

[6] Kintz P, Salomone A, Vincenti M. *Hair Analysis in Clinical and Forensic Toxicology*, 1st ed. San Diego, CA, USA: Academic Press, Elsevier Inc; 2015.

[7] Pirro V, Di Corcia D, Pellegrino S, Vincenti M, Sciutteri B, Salomone A. A study of distribution of ethyl glucuronide in different keratin matrices. *Forensic Sci Int* 2011;210(1-3):271-7. doi: 10.1016/j.forsciint.2011.03.026.

[8] Pianta A, Liniger B, Baumgartner MR. Ethyl Glucuronide in Scalp and Non-head Hair: An Intra-individual Comparison. *Alcohol Alcohol* 2013;48(3):295-302. doi: 10.1093/alcalc/agt012.

[9] Cappelle D, Neels H, De Keukeleire S, et al. Ethyl glucuronide in keratinous matrices as biomarker of alcohol use: A correlation study between hair and nails. *Forensic Sci Int* 2017;279:187-191. doi: 10.1016/j.forsciint.2017.08.022.

[10] Pirro V, Valente V, Oliveri P, De Bernardis A, Salomone A, Vincenti M. Chemometric evaluation of nine alcohol biomarkers in a large population of clinically-classified subjects: pre-eminence of ethyl glucuronide concentration in hair for confirmatory classification. *Anal Bioanal Chem* 2011;401(7):2153-64. doi: 10.1007/s00216-011-5314-7.

[11] Kharbouche H, Faouzi M, Sanchez N, et al. Diagnostic performance of ethyl glucuronide in

hair for the investigation of alcohol drinking behavior: a comparison with traditional biomarkers. *Int J Legal Med* 2012;126(2):243-50. doi: 10.1007/s00414-011-0619-9.

[12] Marques PR, Tippetts AS, Yegles M. Ethylglucuronide in Hair Is a Top Predictor of Impaired Driving Recidivism, Alcohol Dependence, and a Key Marker of the Highest BAC Interlock Tests. *Traffic Inj Prev* 2014;15(4):361-9. doi: 10.1080/15389588.2013.824569.

[13] Petzel-Witt S, Pogoda W, Wunder C, Paulke A, Schubert-Zsilavecz M, Toennes SW. Influence of bleaching and coloring on ethyl glucuronide content in human hair. *Drug Test Anal* 2018;10(1):177-183. doi: 10.1002/dta.2206.

[14] Luginbühl M, Nussbaumer S, Weinmann W. Decrease of ethyl glucuronide concentrations in hair after exposure to chlorinated swimming pool water. *Drug Test Anal* 2017 Aug 30. doi: 10.1002/dta.2295.

[15] Sporkert F, Kharbouche H, Augsburger MP, Klemm C, Baumgartner MR. Positive EtG findings in hair as a result of a cosmetic treatment. Forensic Sci Int 2012;218(1-3):97-100. doi: 10.1016/j.forsciint.2011.10.009.

[16] Vincenti M, Kintz P. New challenges and perspectives, in: Kintz P, Salomone A, Vincenti M, *Hair Analysis in Clinical and Forensic Toxicology*, 1st ed. San Diego, CA, USA: Academic Press, Elsevier Inc; 2015: 337–368.

[17] Kintz P. 2014 consensus for the use of alcohol markers in hair for assessment of both abstinence and chronic excessive alcohol consumption. *Forensic Sci Int* 2015;249:A1-2. doi: 10.1016/j.forsciint.2014.11.001.

[18] Society of Hair Testing. 2016 Consensus for the Use of Alcohol Markers in Hair for Assessment of both Abstinence and Chronic Excessive Alcohol Consumption 2016. http://www.soht.org/images/pdf/Revision%202016_Alcoholmarkers.pdf Accessed February 12, 2018.

[19] Pragst F, Suesse S, Salomone A, Vincenti M, Cirimele V, Hazon J, Tsanaclis L, Kingston R, Sporkert F, Baumgartner MR. Commentary on current changes of the SoHT 2016 consensus on alcohol markers in hair and further background information. *Forensic Sci Int* 2017;278:326-333. doi: 10.1016/j.forsciint.2017.07.023.

[20] Albermann ME, Musshoff F, Aengenheister L, Madea B. Investigations on the influence of different grinding procedures on measured ethyl glucuronide concentrations in hair determined with an optimized and validated LC-MS/MS method. *Anal Bioanal Chem* 2012;403(3):769-76. doi: 10.1007/s00216-012-5926-6.

[21] Mönch B, Becker R, Nehls I. Quantification of Ethyl Glucuronide in Hair: Effect of Milling on Extraction Efficiency. *Alcohol Alcohol* 2013;48(5):558-63. doi: 10.1093/alcalc/agt059.

[22] Kummer N, Wille SM, Di Fazio V, et al. Impact of the Grinding Process on the Quantification of Ethyl Glucuronide in Hair Using a Validated UPLC–ESI–MS-MS Method. *J Anal Toxicol* 2015;39:17-23. doi: 10.1093/jat/bku108.

[23] Salomone A, Baumgartner MR, Lombardo T, Alladio E, Di Corcia D, Vincenti M. Effects of various sample pretreatment procedures on ethyl glucuronide quantification in hair samples: Comparison of positivity rates and appraisal of cut-off values. *Forensic Sci Int 2016;267:60-65. doi: 10.1016/j.forsciint.2016.08.012*.

[24] Mueller A, Jungen H, Iwersen-Bergmann S, Raduenz L, Lezius S, Andresen-Streichert H. Determination of ethyl glucuronide in human hair samples: A multivariate analysis of the impact of extraction conditions on quantitative results. *Forensic Sci Int* 2017;271:43-48. doi: 10.1016/j.forsciint.2016.12.011.

[25] Brereton RG. *Applied Chemometrics for Scientists*, John Wiley & Sons, Ltd., Chichester, UK, 2017.

[26] Leardi R. Experimental design in chemistry: A tutorial. *Anal Chim Acta* 2009;652:161-172. doi: 10.1016/j.aca.2009.06.015

[27] Hibbert DB. Experimental design in chromatography: A tutorial review. *J Chromatogr B Anal Technol Biomed Life Sci* 2012;910:2-13. doi: 10.1016/j.jchromb.2012.01.020.

[28] International Organisation for Standardisation. *ISO/IEC 17025:2017, General Requirements for the Competence of Testing and Calibration Laboratories* 2017:1-30.

[29] Pirro V, Di Corcia D, Seganti F, Salomone A, Vincenti M. Determination of ethyl glucuronide levels in hair for the assessment of alcohol abstinence. *Forensic Sci Int* 2013;232:229-36. doi: 10.1016/j.forsciint.2013.07.024.

[30] Desharnais B, Camirand-Lemyre F, Mireault P, Skinner CD. Procedure for the Selection and Validation of a Calibration Model I—Description and Application. *J Anal Toxicol* 2017;41(4):261-268. doi: 10.1093/jat/bkx001.

[31] Desharnais B, Camirand-Lemyre F, Mireault P, Skinner CD. Procedure for the Selection and Validation of a Calibration Model II—Theoretical Basis. J Anal Toxicol 2017;41(4):269-276. doi: 10.1093/jat/bkx002.

[32] Scientific Working Group for Forensic Toxicology (SWGTOX). Standard Practices for Method Validation in Forensic Toxicology. *J Anal Toxicol* 2013;37:452-474. doi: 10.1093/jat/bkt054.

[33] Raposo F. Evaluation of analytical calibration based on least-squares linear regression for instrumental techniques: A tutorial review. *TrAC Trends Anal Chem* 2016;77:167-185. doi: 10.1016/j.trac.2015.12.006.

[34] Olivieri AC. Practical guidelines for reporting results in single- and multi-component analytical calibration: A tutorial. *Anal Chim Acta* 2015;868:10-22. doi: 10.1016/j.aca.2015.01.017

[35] Genz A, Bretz F. *Computation of Multivariate Normal and T Probabilities*, 1st ed. Berlin, Heidelberg, Germany: Springer; 2009.

[36] Genz A, Bretz F, Miwa T, Hothorn T. mvtnorm: Multivariate Normal and t Distributions. R package version 1.0-6. 2017. https://cran.r-project.org/web/packages/mvtnorm/index.html Accessed February 12, 2018.

[37] Hubaux A, Vos G. Decision and detection limits for calibration curves. *Anal Chem* 1970;42:849-855. doi: 10.1021/ac60290a013.

[38] RStudio Team. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com Accessed February 12, 2018.

[39] Gruppo di Chemiometria della Società Chimica Italiana (SCI). http://gruppochemiometria.it/index.php/software Accessed February 12, 2018.

[40] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, 2014 Vienna, Austria. URL http://www.R-project.org/. Accessed February 12, 2018.

[41] González O, Blanco ME, Iriarte G, Bartolomé L, Maguregui MI, Alonso RM. Bioanalytical chromatographic method validation according to current regulations, with a special focus on the non-well defined parameters limit of quantification, robustness and matrix effect. *J Chromatogr A* 2014;1353:10-27. doi: 10.1016/j.chroma.2014.03.077.

[42] Hubert P, Chiap P, Crommen J, et al. The SFSTP guide on the validation of chromatographic methods for drug bioanalysis: From the Washington Conference to the laboratory. *Anal Chim Acta* 1999;391(2):135-148. doi10.1016/S0003-2670(99)00106-3.

[43] Shah VP, Midha KK, Dighe S, et al. Analytical methods validation: Bioavailability, bioequivalence, and pharmacokinetic studies. *Eur J Drug Metab Pharmacokinet* 1991;16(4):249-55. https://www.ncbi.nlm.nih.gov/pubmed/1823867 Accessed February 12, 2018.

[44] Kronstrand R, Brinkhagen L, Nyström FH, Ethyl glucuronide in human hair after daily consumption of 16 or 32 g of ethanol for 3 months. *Forensic Sci Int* 2012;215:51-55. doi: 10.1016/j.forsciint.2011.01.044.

Figure captions

Figure 1. (a) Geometric representation of the first DoE (full-factorial design) showing the extraction temperature on the x-axis and the solvents on the y-axis. The levels of extraction temperatures were coded from -1.5 up to +1.5 (i.e., -1.5, -0.5, +0.5, +1.5 for 20°C, 32°C, 45°C, and 57°C, respectively), while the codes for the different extraction mixture were -1, 0, +1 for pure distilled water, water/methanol 35:1, and water/methanol 17:1, respectively. (b) Geometric representation of the second DoE (face-centred central composite design) showing the extraction temperature on the x-axis, the extraction time on the y-axis and the pH on the z-axis. The levels were coded from -1 to +1 (i.e. -1, 0, +1) representing (i) 45°C, 54°C and 63°C for the extraction temperature, (ii) 1, 8 and 16 hours for the extraction time, and (iii) 5.5, 7.0 and 8.5 for the pH of the extraction solvent (water).

Figure 2. Graphs of the full-factorial DoE related to batch A. (left) Coefficients plot showing that the extraction temperature (i.e. x1) was the most significant factor influencing the EtG extraction (p-value < 0.001). In particular, the significance level is indicated in the plot according to the following convention: * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. (right) Bidimensional response surface plot with contour lines outlining equal EtG concentrations (numerical values, pg/mg) showing that the highest EtG concentrations were observed between the temperatures encoded as +0.5 and +1.5 (i.e. 45°C and 57°C) and using the solvent encoded as -1.0 (i.e. only pure water). The values reported within the blue lines represent the EtG concentration for each response surface (i.e., the higher the value, the better the extraction yield).

Figure 3. Graphs of the face-centred central composite DoE related to batch A. (Left): coefficients plot showing that the extraction temperature (x1) and time (x2) were the most significant factors that affected the EtG extraction (p-value < 0.001). In particular, the significance level is indicated in the plot according to the following convention: * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. (Right): two-dimensional response surface plot with Y=EtG concentration (pg/mg) showing that the highest EtG concentration was observed at the temperature encoded as +1.0 (63°C) and at the extraction time encoded as +1.0 (16 hours). The graph was calculated at pH 7.0 of the extraction solvent (only water).

Figure 4. (a-b) Boxplots of the areas of the EtG-D_5 at the different tested levels of extraction time (left) and temperature (right).

Figure 5. Boxplots of the extracted EtG concentration for 10 hair samples (left = batch C; right = batch D) analysed by the three tested methods. The red boxes represent the EtG concentration values provided by method (a) that involved manual hair cutting, the yellow boxes indicate the EtG results from the method (b) that employed hair milling, while the green boxes show the EtG values provided by the new method optimized by DoE.

Table 1. EtG concentration values (pg/mg) of 10 hair samples for batches C (i.e., with EtG values originally detected in the range between 20 pg/mg and 31 pg/mg) and D (i.e. with EtG values originally detected above 60 pg/mg). The hair specimens were analysed by three method, as follows: (a) involving manual cutting only; (b) using the pulverisation of the hair lock; (c) using pulverisation of the hair lock and the new extraction protocol optimized by DoE. Positive percentage differences among the average results provided by each method are reported.

Batch C	Method (EtG concentration in pg/mg)			Batch D	Method (EtG concentration in pg/mg)		
Sample n°	(a)	(b)	(c)	Sample n°	(a)	(b)	(c)
1	25	33	39	11	71	95	141
2	23	30	41	12	109	140	196
3	26	36	43	13	111	146	192
4	23	32	45	14	100	134	144
5	24	30	37	15	103	131	181
6	25	32	38	16	93	129	139
7	20	27	52	17	84	105	133
8	25	33	44	18	81	104	136
9	31	40	43	19	91	117	142
10	26	34	36	20	65	87	138
Average	24.8	32.7	41.8	Average	90.8	118.8	154.2
Positive differences	(b-a)/a	(c-b)/b	(c-a)/a	Positive differences	(b-a)/a	(c-b)/b	(c-a)/a
	+31.9%	+27.8%	+68.5%		+30.8%	+29.8%	+69.8

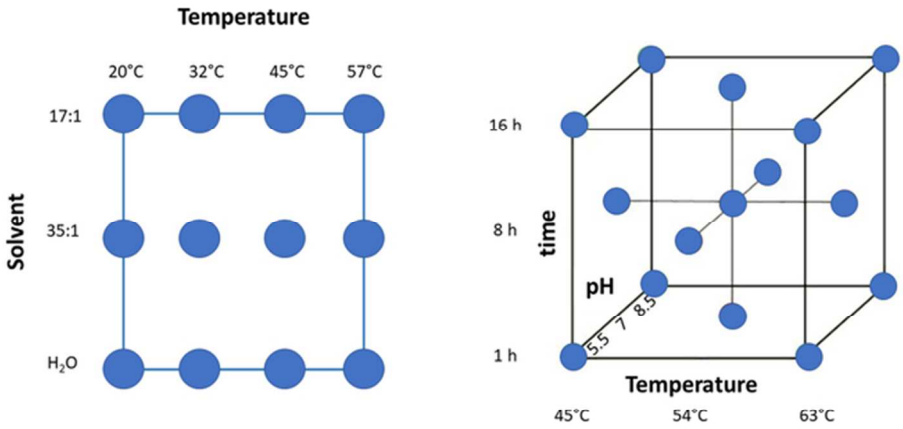

Figure 1. (a) Geometric representation of the first DoE (full-factorial design) showing the extraction temperature on the x-axis and the solvents on the y-axis. The levels of extraction temperatures were coded from -1.5 up to +1.5 (i.e., -1.5, -0.5, +0.5, +1.5 for 20°C, 32°C, 45°C, and 57°C, respectively), while the codes for the different extraction mixture were -1, 0, +1 for pure distilled water, water/methanol 35:1, and water/methanol 17:1, respectively. (b) Geometric representation of the second DoE (face-centred central composite design) showing the extraction temperature on the x-axis, the extraction time on the y-axis and the pH on the z-axis. The levels were coded from -1 to +1 (i.e. -1, 0, +1) representing (i) 45°C, 54°C and 63°C for the extraction temperature, (ii) 1, 8 and 16 hours for the extraction time, and (iii) 5.5, 7.0 and 8.5 for the pH of the extraction solvent (water).

30x17mm (600 x 600 DPI)

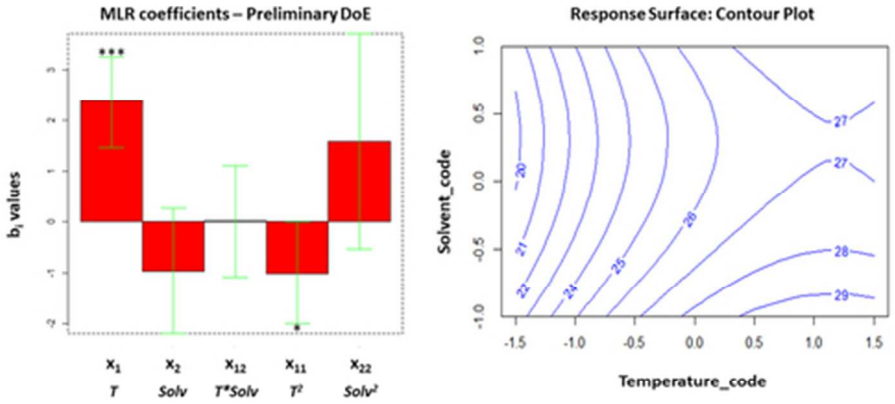

Figure 2. Graphs of the full-factorial DoE related to batch A. (left) Coefficients plot showing that the extraction temperature (i.e. x1) was the most significant factor influencing the EtG extraction (p-value < 0.001). In particular, the significance level is indicated in the plot according to the following convention: * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. (right) Bidimensional response surface plot with contour lines outlining equal EtG concentrations (numerical values, pg/mg) showing that the highest EtG concentrations were observed between the temperatures encoded as +0.5 and +1.5 (i.e. 45°C and 57°C) and using the solvent encoded as -1.0 (i.e. only pure water). The values reported within the blue lines represent the EtG concentration for each response surface (i.e., the higher the value, the better the extraction yield).

45x25mm (300 x 300 DPI)

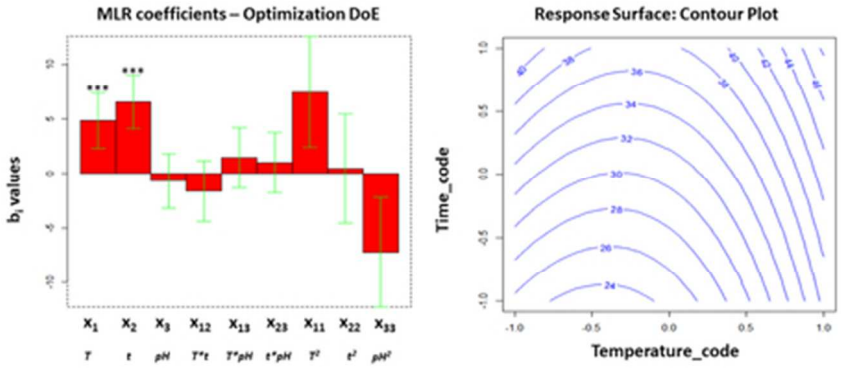

Figure 3. Graphs of the face-centred central composite DoE related to batch A. (Left): coefficients plot showing that the extraction temperature (x1) and time (x2) were the most significant factors that affected the EtG extraction (p-value < 0.001). In particular, the significance level is indicated in the plot according to the following convention: * = p-value < 0.05, ** = p-value < 0.01, *** = p-value < 0.001. (Right): two-dimensional response surface plot with Y=EtG concentration (pg/mg) showing that the highest EtG concentration was observed at the temperature encoded as +1.0 (63°C) and at the extraction time encoded as +1.0 (16 hours). The graph was calculated at pH 7.0 of the extraction solvent (only water).

45x25mm (300 x 300 DPI)

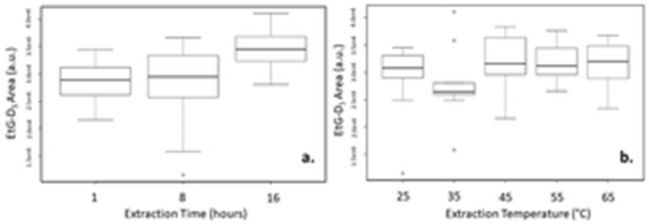

Figure 4. (a-b) Boxplots of the areas of the EtG-D5 at the different tested levels of extraction time (left) and temperature (right).

29x10mm (300 x 300 DPI)

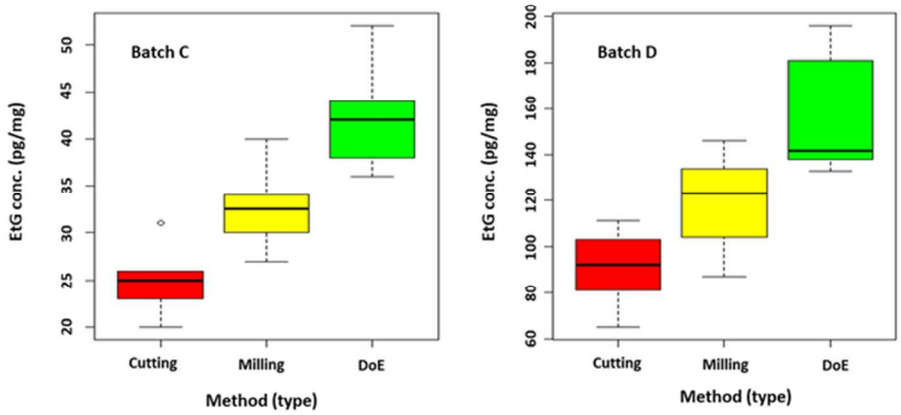

Figure 5. Boxplots of the extracted EtG concentration for 10 hair samples (left = batch C; right = batch D) analysed by the three tested methods. The red boxes represent the EtG concentration values provided by method (a) that involved manual hair cutting, the yellow boxes indicate the EtG results from the method (b) that employed hair milling, while the green boxes show the EtG values provided by the new method optimized by DoE.

30x17mm (600 x 600 DPI)