Interlaboratory development and proposition for a new quality control sample for chemical forensics analysis of chemical warfare agents

Säde, S.; Delaporte, G.; Fraga, C.G.; Hakulinen, H.; Holmgren, K.H.; Spiandore, M.; Åstot, C.; Akmeemana, A.; de Bruin-Hoegée, M.; Doward, J.; Jacques, K.; Kesah, S.A.; See, S.; Thomson, S.; Flinck, J.; Kalliovirta, L.; Vanninen, P.

DOI
10.1016/j.talo.2023.100249

Publication date
2023

Document Version
Final published version

Published in
Talanta Open

Talanta Open 8 (2023) 100249

Contents lists available at ScienceDirect

Talanta Open

journal homepage: www.sciencedirect.com/journal/talanta-open

Interlaboratory development and proposition for a new quality control sample for chemical forensics analysis of chemical warfare agents☆

Solja Säde [a,*], Grégoire Delaporte [b,1], Carlos G. Fraga [c,1], Hanna Hakulinen [a,1], Karin Höjer Holmgren [d,1], Marie Spiandore [b,1], Crister Åstot [d,1], Anuradha Akmeemana [e,2], Mirjam de Bruin-Hoegée [f,g,2], Justin Doward [h,2], Karen Jacques [i,2], Saiful Anuar Kesah [j,2], Samantha See [k,2], Stuart Thomson [l,2], Jens Flinck [m], Leena Kalliovirta [m], Paula Vanninen [a]

[a] Finnish Institute for Verification of the Chemical Weapons Convention (VERIFIN), Department of Chemistry, University of Helsinki, P.O. Box 55, FI-00014, Helsinki, Finland
[b] Analytical Chemistry Department, DGA CBRN Defence, 91710, Vert-Le-Petit, France
[c] Air Force Research Laboratory, 10 E. Saturn Blvd., Edwards Air Force Base, CA, 93524, USA
[d] Swedish Defence Research Agency (FOI), CBRN Defence and Security, Cementvägen 20, 90182, Umeå, Sweden
[e] National Center for Forensic Science, University of Central Florida, Orlando, FL, 32826, USA
[f] Van 't Hoff Institute for Molecular Sciences, Faculty of Science, University of Amsterdam, P.O. Box 94157, 1090GD, Amsterdam, the Netherlands
[g] TNO Defence, Safety and Security, Department of CBRN Protection, Lange Kleiweg 137, 2288GJ, Rijswijk, the Netherlands
[h] Defence Science and Technology Group, 506 Lorimer Street, Fishermans Bend, 3207, Victoria, Australia
[i] Dstl Porton Down, Salisbury, Wiltshire, SP4 0JQ, UK
[j] Department of Chemistry Malaysia, Jalan Sultan, 46661, Petaling Jaya, Selangor, Malaysia
[k] DSO National Laboratories, 12 Science Park Drive, Singapore, 118225
[l] Organisation for the Prohibition of Chemical Weapons (OPCW), The Hague, the Netherlands
[m] Department of Mathematics and Statistics, University of Helsinki, Pietari Kalmin katu 5, FI-00014, Helsinki, Finland

ARTICLE INFO

Keywords:
Chemical forensics
Quality control
Gas chromatography–mass spectrometry
Interlaboratory study
Chemometrics
Chemical forensic signatures

ABSTRACT

A new quality control (QC) test sample for gas chromatography–mass spectrometry (GC–MS) was created and analysed to test the comparability and repeatability of chemical forensics results within the Organisation for the Prohibition of Chemical Weapons (OPCW)–designated laboratories. The QC test sample was designed in collaboration between four laboratories and consists of 27 compounds which evaluate the performance of GC–MS instruments. This solution was analysed with GC–MS(EI) in 11 laboratories, seven of which were OPCW designated. The participating laboratories analysed the sample multiple times on consecutive days, as well as after the analysis of a set of complex matrix samples. Retention times, retention indices, peak areas, peak tailing values, signal-to-noise ratios, and isotope ratios were extracted from the GC–MS data, and statistical multivariate analyses with principal component analysis and Hotelling's T^2-tests were conducted. The results from these analyses indicate that differences between GC–MS analyses by multiple laboratories were not statistically significant at the 5% level, as the approximate *p*-value for the null hypothesis of "no differences between the runs" was 0.69. However, similar data processing methods and data normalisation are essential for enabling the reliable comparison of chemical fingerprints between laboratories. A composition for the QC sample and criteria for acceptable GC–MS performance for chemical forensics are proposed. The composition and criteria differ from the currently used chemical weapons verification analysis QC sample by e.g. broadening the range for retention

Abbreviations: CWA, chemical warfare agent; CWC, Chemical Weapons Convention; IIT, Investigation and Identification Team; OPCW, Organisation for the Prohibition of Chemical Weapons; OCAD, OPCW Central Analytical Database; RI, retention index; PCA, principal component analysis; ANOVA, analysis of variance; UV scaling, scaling to unit variance; C26, hexacosane; C28, octacosane; C30, triacontane; TEPP, tetraethyl pyrophosphate; TMS, trimethylsilyl; TBDMS, *tert*-butyldimethylsilyl.

☆ The views expressed are those of the authors and do not reflect the official views of the United States Air Force, nor the Department of Defense. Mention of trade names, commercial products, or organizations do not imply endorsement by the U.S. Government.

* Corresponding author at: VERIFIN, University of Helsinki, A.I.Virtasen aukio 1, 00560, Helsinki, Finland.
E-mail address: solja.sade@helsinki.fi (S. Säde).

[1] These authors contributed equally to this work.
[2] These authors contributed equally to this work.

https://doi.org/10.1016/j.talo.2023.100249
Received 31 May 2023; Received in revised form 14 July 2023; Accepted 19 July 2023
Available online 20 July 2023

index calculations by addition of new alkane compounds, including new chemicals with concentrations close to the limit of detection (10–100 ng/ml), and including compounds with higher polarity to emulate real-life forensic samples. The proposed criteria include monitoring of retention indices, isotope ratios, peak tailing, signal-to-noise ratios, peak height, mass spectra, and sensitivity of the instrument. The new compounds and criteria will be the subject of future confidence building exercises to validate their relevancy on a large scale.

1. Introduction

The use of chemical warfare agents (CWAs) remains a current issue even 25 years after the Chemical Weapons Convention (CWC) entered into force in 1997. Sarin and chlorine attacks in the Syrian Arab Republic in 2013–2018 led to the injury or death of hundreds of people [1]. These attacks, for which no party claimed responsibility, led to the establishment of the Investigation and Identification Team (IIT) by the Organisation for the Prohibition of Chemical Weapons (OPCW) in 2018 [2]. The IIT investigates suspected chemical weapon use in the Syrian civil war with a broader forensic scope than the verification analysis element of the OPCW. The need for holding the responsible parties accountable for the attacks highlights the importance of applying chemical forensics methodologies on laboratory investigations on the use of chemical weapons.

One method for chemical forensics investigations is source determination by chemical fingerprint analysis of samples obtained from crime scenes. A chemical fingerprint of a sample can be a list of compounds (impurities, synthesis by-products, degradation products, etc.) found in a sample, their relative abundances, and any other data that can be extracted. The fingerprint can also contain information about the isotopic ratios of the detected molecules. Those markers can then be associated with a certain synthetic pathway, type of equipment, reaction condition, or source of reagents [3]. Chemical fingerprints have been successfully utilised previously in various other fields of forensics, such as food fraud investigations [4] and drug profiling [5], including methods for source determination of cocaine [6,7]. These methods have been increasingly implemented on studies of CWAs and chemical threat agents as well, such as sarin [8,9], VX [10–13], fentanyl [14–19], sulphur mustard [20,21], calcium ammonium nitrate [22], pesticides [23], *N,N*-dimethylphosphoramidic dichloride [24], pepper spray [25], cyanides [26], and isopropyl bicyclophosphate [3]. Considering the increasing number of CWA related chemicals, the research methods and the comparability and reproducibility of results must be examined further. In theory, to get a reliable estimation of its source, a given sample must produce a similar fingerprint regardless of the laboratory analysing the sample. Recent papers on interlaboratory results of the chemical fingerprint of methylphosphonic dichloride (a nerve agent precursor) by Höjer Holmgren et al. [27] and Fraga et al. [28] show promising results on the comparability of chemical fingerprints and highlight the need for further research, specifically on quality control (QC) for chemical forensics.

The aim of this work was to investigate the reproducibility of results between multiple laboratories, and develop a new QC sample for chemical forensics analyses of CWAs with gas chromatography–mass spectrometry (GC–MS). The new QC sample is based on the Provisional Technical Secretariat Interlaboratory Comparison Test [29], in which GC–MS analyses and QC methods were evaluated. The current QC sample is utilised by the OPCW and its designated laboratories to ensure the acceptable performance of the GC–MS instruments used in verification analyses [30]. The new QC sample is developed specifically for chemical forensics, and with changes in the mixture composition and QC criteria, it aims to improve the comparability and consistency of chemical fingerprints obtained from multiple laboratories.

2. Materials and methods

2.1. QC test mixture

Participants from DGA CBRN Defence, The Swedish Defence Research Agency (FOI), Pacific Northwest National Laboratory (PNNL), and the Finnish Institute for Verification of the Chemical Weapons Convention (VERIFIN) chose 27 compounds for the QC test mixture based on the compounds' properties. These properties included the ability to test the condition of the GC (e.g. column activity), having poor elution (to test if differences between laboratories would be observed), ability to be used for RI calculation, which aids in target compound identification (alkane sequence), or usage for isotopic ratio calculations (analysis of $^{37}Cl/^{35}Cl$ from 5-chloro-2-methylaniline and $^{34}S/^{32}S$ from dibenzothiophene). The compounds are listed in Table 1 along with relevant information. Initially, the compounds were analysed separately with GC–MS by VERIFIN, after which a solution was prepared. There was no indication of reactivity between the compounds, as no product formation was observed. The solution was determined stable at ambient temperature after one month as the *p*-value from an analysis of variance (ANOVA) for the null hypothesis of "no difference between the analyses" was 0.99.

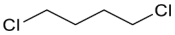
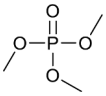
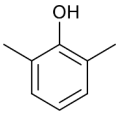
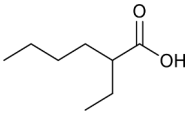
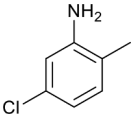
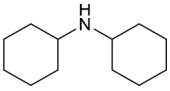
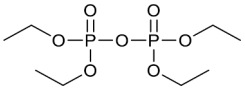
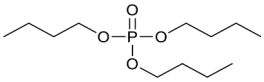
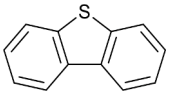
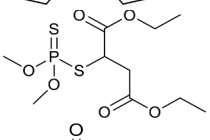
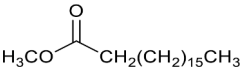
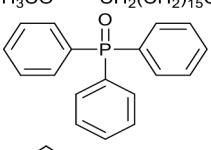
2.2. Sample preparation

The QC test mixture was prepared at concentrations of 1, 5, or 10 μg/ml for each compound, as specified in Table 1. The compounds were weighed (0.010 g, 0.050 g, or 0.100 g) into separate vials, and five stock solutions (100 ml) were prepared. Each stock solution (1 ml) was combined to form the final QC test mixture (100 ml). Dichloromethane (DCM) was used as a solvent. The sample preparation is described in more detail in the supplementary information (S01). Measurement error in sample preparation u_c was calculated to be 0.12% (S02), and the homogeneity of samples was evaluated for six replicates produced from the stock solution in the same way as the samples sent to the laboratories. The QC test mixture was packaged in vials, which had been tested for leakage and impurity contamination. Three ml of the QC solution was sent to each of the participating laboratories (Table 2). Vials were sent through a courier service at ambient temperature, and they arrived at the laboratories between 4 and 24 days after despatch.

2.3. GC–MS analysis

Laboratories cleaned their GC–MS instruments and replaced the column, septum, deactivated liner, and bottom plate (when relevant) prior to the analyses. The analysis conditions followed the OPCW Central Analytical Database (OCAD) guidance [31]. As some laboratories conducted additional analyses with different parameters, the analyses are coded with letters to indicate the laboratory and numbers to indicate additional analyses, e.g. A1, A2, B, C1, C2… The different analyses will later be referred to as data sets (e.g. data set A1). The GC–MS instrument models and specific analysis conditions for each data set are described in the supplementary information (Tables S02, S03, and S04). The predominant analysis conditions are described below: a non-polar, Agilent DB-5 ms or corresponding, column (30 m, ID 0.25 mm, 0.25 μm) was used for compound separation. Analyses were conducted with 1 μl splitless injection at 250 °C, with helium as the carrier gas with a constant flow of 1.0 ml/min and solvent delay of 2.5 min. The temperature

Table 1
Compound name, CAS-number, structure, supplier, purity, and concentration in the QC test mixture. Purity provided by supplier for all compounds excluding tetraethyl pyrophosphate, for which the purity was determined with NMR by VERIFIN.

Compound name	CAS number	Structure	Supplier[c]	Purity (%)	Conc. (μg/ml)
Heptane	142–82–5	$CH_3(CH_2)_5CH_3$	Supelco, Merck	99.5	10
Octane[a]	111–65–9	$CH_3(CH_2)_6CH_3$	Fluka	> 99.7	5
Decane[a]	124–18–5	$CH_3(CH_2)_8CH_3$	Acros	> 99.0	5
Dodecane[a]	112–40–3	$CH_3(CH_2)_{10}CH_3$	Acros	> 99.0	5
Tetradecane[a]	629–59–4	$CH_3(CH_2)_{12}CH_3$	Acros	> 99.0	5
Hexadecane[a]	544–76–3	$CH_3(CH_2)_{14}CH_3$	Acros	> 99.0	5
Octadecane[a]	593–45–3	$CH_3(CH_2)_{16}CH_3$	Acros	> 99.0	5
Eicosane[a]	112–95–8	$CH_3(CH_2)_{18}CH_3$	Lancaster	99.0	5
Docosane[a]	629–97–0	$CH_3(CH_2)_{20}CH_3$	Sigma Aldrich	99.0	5
Tetracosane[a]	646–31–1	$CH_3(CH_2)_{22}CH_3$	Fluka	> 99.0	5
Hexacosane	630–01–3	$CH_3(CH_2)_{24}CH_3$	Alfa Aesar	99.0	10
Octacosane	630–02–4	$CH_3(CH_2)_{26}CH_3$	Alfa Aesar	99.0	10
Triacontane	638–68–6	$CH_3(CH_2)_{28}CH_3$	Alfa Aesar	> 98.0	10
1,4-Dichlorobutane[a]	110–56–5		Sigma Aldrich	99.0	10
Trimethyl phosphate[a]	512–56–1		Acros	99 (p.a.[b])	5
Trimethylphosphine oxide	676–96–0		Alfa Aesar	99.0	10
1,4-Dithiane	505–29–3		Sigma Aldrich	≥ 97.0	10
2,6-Dimethylphenol[a]	576–26–1		Acros	99 (p.a.[b])	5
2-Ethylhexanoic acid	149–57–5		Sigma Aldrich	99.0	10
5-Chloro-2-methylaniline[a]	95–79–4		Alfa Aesar	99.0	5
Dicyclohexylamine	101–83–7		Fluka	> 99.5	10
Tetraethyl pyrophosphate (TEPP)	107–49–3		Synthesis from Spiez Laboratory (year 1994)	72*	10
Tributyl phosphate[a]	126–73–8		Sigma Aldrich	≥ 99.0	5
Dibenzothiophene[a]	132–65–0		Sigma Aldrich	≥ 99.0	1
Malathion[a]	121–75–5		European Pharmacopoeia	99.4	5
Methyl stearate[a]	112–61–8		Fluka	> 99.0	5
Triphenylphosphine oxide	791–28–6		Sigma Aldrich	98.0	10
Dichloromethane[a]	75–09–2		SupraSolv, Merck	≥ 99.8	Solv.

[a] Included also in the currently used verification analysis QC sample.
[b] Pro analytical. *Standard included 27% of hydrolysis product diethyl phosphate, not detected with GC–MS.

program started at 40 ℃ (1 min hold), continuing with 10 ℃/min to 300 ℃ with a 7 min hold, resulting in a total run time of 34 min. The MS was operated in electron ionisation (EI) mode, with scan range of *m/z* 30–550.

Acceptable performance of the GC–MS instruments was first tested by each laboratory according to their current quality control measures. The QC test mixture sample was then analysed three times in a row on three consecutive days, resulting in nine analyses. In addition, six runs of a complex matrix sample, such as soil, were analysed with the instrument, followed by three runs of the QC test mixture sample. The complex matrix samples were used to investigate any possible effects of matrix compounds on the instrument and the QC sample.

To test the performance of the QC compounds and the reproducibility of results between laboratories in lower concentrations, dilutions (Table S05) of the original QC sample were analysed with four instruments (A, D, I1, and I2). As the undiluted QC test mixture included compounds in three different concentrations (1, 5, and 10 μg/ml), the dilutions did as well. The samples were prepared in each laboratory from the originally sent QC test mix and analysed in sequence from lowest to highest concentration. Additionally, a sample of the QC test mix was analysed undiluted. All raw data was analysed at VERIFIN.

2.4. Data analysis

The raw GC–MS data (Fig. 1) was analysed by each laboratory, and the following parameters were extracted from the data: retention time, retention index (RI), peak area, peak tailing, signal-to-noise (S/N) ratio, and isotope ratios. The results for the 27 compounds were determined from the total ion chromatograms (TIC). The isotope ratios for 5-chloro-2-methylaniline (*m/z* 143/141, $^{37}Cl/^{35}Cl$) and dibenzothiophene (*m/z* 186/184, $^{34}S/^{32}S$) were calculated from spectral abundances.

Laboratories provided the raw GC–MS data and the processed data (later referred to as *original data*) for analysis at VERIFIN. Original data processing was performed by the laboratories mainly with the Automatic Mass Spectral Deconvolution and Identification System [AMDIS, National Institute of Standards and Technology (NIST)] for the original data sets A1, B, D, E1, E2, G, H, I1, I2, J (RI and peak tailing only), and K, but MassHunter software Qualitative Analysis and Quantitative Analysis (Agilent, Santa Clara, CA, USA) were utilised as well [original data sets A2, C1, C2, and J (for retention time, peak area, and S/N ratio)]. AMDIS is a piece of data processing software that utilises a deconvolution algorithm to extract all component spectra and then identifies compounds present in the sample through library comparison. RI can be calculated using an alkane calibration file. MassHunter software integrate the chromatogram, deconvolute the ions from the compound spectra, and match them with library spectra, identifying the compounds. Raw data processing with AMDIS or MassHunter software provided the numerical values for retention time, RI, peak area, peak tailing, and S/N ratio for each compound. Retention time data was not included in the following data analyses, as RIs are commonly used instead.

SIMCA 17.0 (Sartorius Stedim Biotech Goettingen, Germany) was utilised to analyse the data with principal component analysis (PCA) and Hotelling's T^2-test. PCA is a multivariate statistical method which can be used to reduce the dimensionality of large datasets, enabling a quick visualisation of the result distribution. Hotelling's T^2 measures the distance between sample mean and each observation. PCA results are employed in ANOVA in SIMCA, which then yields approximate *p*-values.

The Department of Mathematics and Statistics at the University of Helsinki consulted on the execution of the statistical analyses. The statistical analysis started with the original data; this was first analysed by VERIFIN with SIMCA as received, comparing the results from all data sets with each other but including only the first nine analyses and excluding the three complex matrix runs. The data were combined to include the following results: all original data sets, all compounds, RI, peak area, peak tailing, and S/N ratio parameters. It was observed that the covariance matrix based PCA becomes biased due to differences in the scales of results. The differences in scale can be explained by two factors. Firstly, the numerical values of RIs, peak areas, peak tailing, and S/N ratios were on different scales (RI = 700–3000, peak area = 1000,000–1000,000,000, peak tailing = 0.3–15, and S/N ratio = 50–1500), and this distorted the analysis. Secondly, the numerical values of the results were on different scales between data sets, especially for peak area results. Original data set F had approximately 100–1000 times higher peak area values due to the high resolution mass spectrometry (HRMS) instrument used (Fig. S01). Original data set I2 had higher values from the others as well, although not as high as original data set F.

To remove the bias in PCA, the data was normalised. A convenient normalisation method was available for the data because hexadecane yields stable and consistent results in all the analyses. Normalisation was done by dividing the value of each parameter with the value of hexadecane from the same GC–MS run. At first, only the results for peak tailing, peak area, and S/N ratio were normalised; RI values were left unnormalised, as they are in themselves a normalised version of retention times. However, due to the scale differences still visible in the results, the RI data needed to be normalised with hexadecane as well to enable reliable PCA for all data. Therefore, normalisation permitted the covariance matrix based PCA. With this normalisation of data for all data sets, the compounds and related parameters became comparable, and the significance of the differences between data sets could be tested for reliably.

As an alternative method to hexadecane-normalisation, scaling to unit variance (UV scaling) was tested. In UV scaling, a single mean and a single variance are computed for each compound–parameter combination using all observations, i.e. all runs in all data sets. The mean is then used to centre the compound–parameter combinations, followed by rescaling the centred value by dividing with the square root of variance [32]. Thus, the use of UV scaling would have required the exclusion of data sets containing results in a different scale from others. Without this exclusion, UV scaling would distort the analysis by moving the mean value disproportionally towards the outlying data sets and affect results for other data sets as well. In addition, even with the results on similar scales, UV scaling can decrease the significance of separating results, and thus "hide" outliers in data. In UV scaling, the results of one data set affect the results of all others, whereas with hexadecane-normalisation, each run is separately scaled with run-specific information, and so the

Table 2
Laboratories participating in the study, and their state party. *Organisations participating in QC mixture development in addition to the interlaboratory study.

Participating laboratories	State party
Analytical Chemistry Department, DGA CBRN Defence*	France
Defence Science and Technology Group	Australia
Department of Chemistry Malaysia	Malaysia
DSO National Laboratories	Singapore
Defence Science and Technology Laboratory, Porton Down	UK
Finnish Institute for Verification of the Chemical Weapons Convention, VERIFIN*	Finland
National centre for Forensic Science, University of Central Florida	USA
Organisation for the Prohibition of Chemical Weapons, OPCW	Intergovernmental
Pacific Northwest National Laboratory, PNNL*	USA
The Swedish Defence Research Agency, FOI*	Sweden
TNO Defence, Safety and Security	The Netherlands

[c] Supplier information: Merck (Merck KGaA, Munich, Germany), Fluka (Fluka analytical, St. Louis, USA), Acros (Acros Organics N.V., New Jersey, USA), Lancaster (Lancaster Synthesis, Morecambe, England), Sigma Aldrich (Merck KGaA, Munich, Germany), Alfa Aesar (Thermo Fisher GmbH, Kandel, Germany), European Pharmacopoeia (European Directorate for the Quality of Medicines & Healthcare EDQM, Council of Europe, Strasbourg, France).

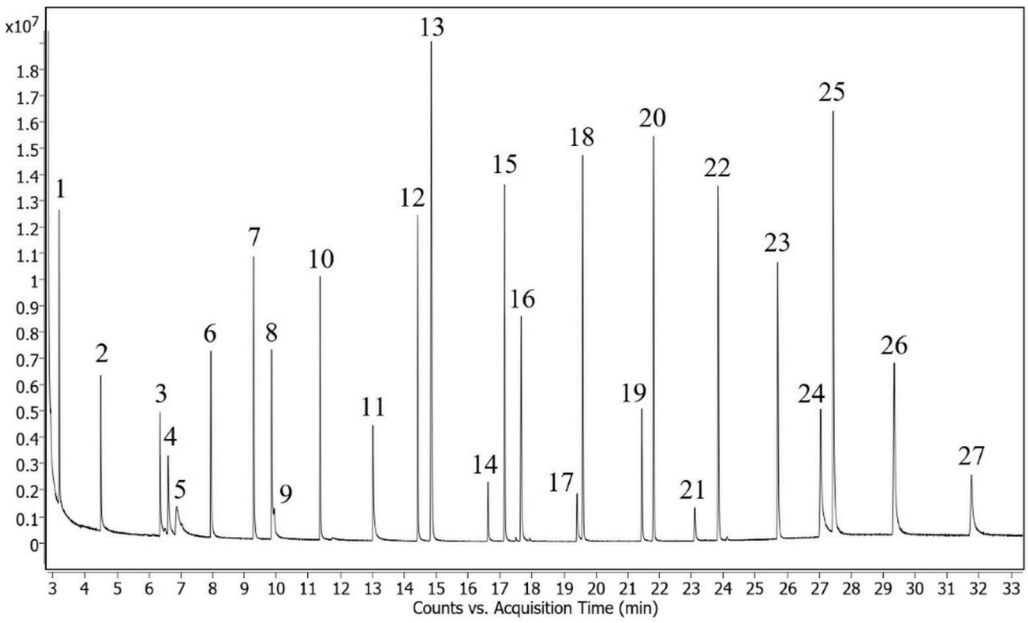


Fig. 1. A chromatogram of the QC test mixture. Compounds: 1) heptane, 2) octane, 3) 1,4-dichlorobutane, 4) trimethyl phosphate, 5) trimethylphosphine oxide, 6) decane, 7) 1,4-dithiane, 8) 2,6-dimethylphenol, 9) 2-ethylhexanoic acid, 10) dodecane, 11) 5-chloro-2-methylaniline, 12) tetradecane, 13) dicyclohexylamine, 14) tetraethyl pyrophosphate, 15) hexadecane, 16) tri-*n*-butyl phosphate, 17) dibenzothiophene, 18) octadecane, 19) malathion, 20) eicosane, 21) methyl stearate, 22) docosane, 23) tetracosane, 24) triphenylphosphine oxide, 25) hexacosane, 26) octacosane, 27) triacontane.

results from data sets have no effect on one another.

Differences in data processing methods between data sets affected the results. Therefore, all raw GC–MS data was processed again by one operator, using the AMDIS software and a single processing method at VERIFIN. This data will later be referred to as *newly processed data*. The newly processed data was normalised with hexadecane for all parameters and compounds. Laboratory F utilised a GC–MS instrument from a different manufacturer and was excluded from this analysis due to software incompatibility. As original data sets A1 and A2 differed only in the data processing software used (AMDIS and MassHunter Qualitative Analysis, respectively) but not in the actual raw data, the newly processed data was given the code A.

The data from different data sets are compared in chapter 3.1 Interlaboratory comparison and the compounds in the QC test mix in chapter 3.2 Quality control sample development. In chapter 3.1, separate models were built for the comparison of data sets with all parameters combined and for the comparison of data sets within only one parameter, such as peak tailing. Similarly, in chapter 3.2, models were built to compare the compounds with all parameters, in addition to separate models for each parameter. Between chapters 3.1 and 3.2, the data was transposed to allow the examination of compounds instead of data sets.

3. Results and discussion

3.1. Interlaboratory comparison

The interlaboratory comparison was conducted by having the laboratories analyse the QC sample three times on three consecutive days, which resulted in nine analyses. Thus, the variability of results between runs and analysis days could be accounted for in addition to interlaboratory variability, which was tested for with the statistical analyses PCA and Hotelling's T^2-tests. These tests determined if the differences between data sets were statistically significant, and would therefore interfere with the reliable comparison of chemical fingerprints obtained from different laboratories.

3.1.1. Comparison of combined parameters

To investigate the interlaboratory reproducibility of results, PCA and Hotelling's T^2-tests were conducted. First, the original data was hexadecane-normalised and analysed with PCA as seen in Fig. 2A. To allow for a comprehensive comparison of the data in one analysis, all 27 compounds and four parameter (RI, peak tailing, peak area, and S/N ratio) combinations over the first nine runs for each data set were included. The PCA was executed with the first two principal components

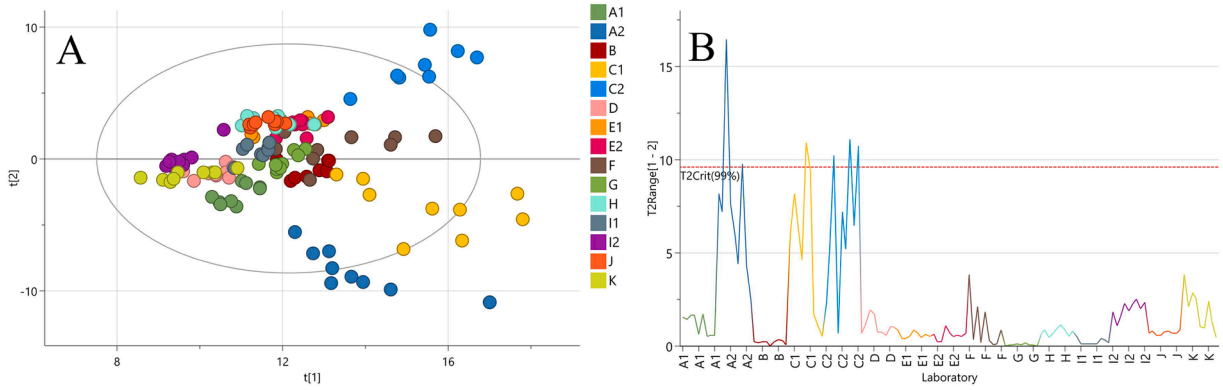


Fig. 2. A) PCA for the original and hexadecane-normalised data with a critical limit of 95%. Each dot represents a GC–MS run for one data set; therefore, each data set has nine dots. Each colour represents one data set. B) Hotelling's T^2 graph for hexadecane-normalised original data from all data sets with a critical limit of 99%. Different colours represent different data sets, and separate peaks within one data set are the nine GC–MS runs.

because they accounted for 87% of the parameter–compound variability. Principal component values for all PCAs can be found in Table S06. This ensured the sufficiency of the first two principal components for the analysis. The analysis displays the separation of data sets A2, C1, and C2 from the others, as they exceeded the 95% critical limit represented by the ellipse in Fig. 2A. This limit indicates the difference between an individual data set and the other data sets at 5% significance level. However, it should be noted that the 5% significance level, or the 95% critical limit, is not appropriate at all in this analysis due to the high number of hypotheses being tested simultaneously, with each hypothesis stating that an observation has a mean different from all other observations. A 99% limit that corresponds to the 1% significance level would be closer to the suitable criteria for statistically significant differences in this analysis, and it is utilised in the subsequent Hotelling's T^2 analyses. The 95% critical limit is shown in the PCA plots instead of the 99% limit due to restricted settings in the software.

The Hotelling's T^2 analysis shown in Fig. 2B examined the distance of each GC–MS run to the sample mean through all model components utilising the same data set as the PCA in Fig. 2A. In both analyses (PCA and Hotelling's T^2), original data sets A2, C1, and C2 had runs that statistically significantly differed from the average results. For these three data sets, six crossings of the 99% significance critical limit are observed, which corresponds to an approximate *p*-value of around 0.003, and therefore the null hypothesis of "no differences between the runs" is rejected. If four or more crossings of the 99% significance critical limit are observed, it can be inferred that the differences are statistically significant at the 5% level because the approximate *p*-value is around 0.04. The rejected analyses of data sets A2, C1, and C2 were expected to originate from the choice of the data processing software, as data sets A1 and A2 used the same GC–MS raw data, but processed the data with AMDIS (A1) and MassHunter Qualitative Analysis (A2), and the results for A1 were not rejected. MassHunter Quantitative Analysis was used for data sets C1 and C2, while AMDIS was predominantly chosen for the other data sets. For original data set J, retention time, peak area, and S/N ratio were calculated using the MassHunter Quantitative Analysis software, but RI and peak tailing using AMDIS. As the rejection of A2, C1, and C2 was largely affected by peak tailing values, the results for original data set J were not rejected.

The loadings plot in Fig. S02 displays the contribution of each variable to the first two principal components. The peak tailing results for trimethylphosphine oxide, trimethyl phosphate, triphenylphosphine oxide, octacosane (C28), triacontane (C30), and hexacosane (C26) had the biggest impact on the rejection of original data sets A2, C1, and C2. The peak tailing values for trimethyl phosphate and triphenylphosphine oxide were disproportionally large in the results for original data set A2 in comparison with results from other data sets and results from other compounds within data set A2. The numerical values for trimethyl phosphate peak tailing in original data set A2 varied between 7 and 16, whilst within original data set A1, they varied between 3 and 5. The significant difference in results for original data set C1 occurred due to higher peak tailing values for trimethylphosphine oxide, C26, C28, C30, and triphenylphosphine oxide; and for data set C2 it was mainly due to trimethylphosphine oxide, but trimethyl phosphate contributed as well. The differences observed were presumed to come from the choice of data processing software, and thus the raw GC–MS data from all laboratories was processed again by one operator at VERIFIN using AMDIS, resulting in the *newly processed data*.

The results of the PCA (Fig. 3A) obtained with the newly processed data (only AMDIS data processing) display decreased overall variability compared to the results obtained with the original data. As the original data sets A2, C1, and C2 were newly processed (resulting in newly processed data sets A, C1, and C2) they were no longer separated from the other data sets. Therefore, the reason for their initial difference appears to lie in the data processing method. In the Hotelling's T^2-test (Fig. 3B), only one GC–MS run exceeded the 99% critical limit. The probability for at least one crossing is very large, 0.69. This is due to the fact that the Hotelling's T^2 in Fig. 3B performs 117 tests overall, so it is very likely that one of the tests rejects even when the hypothesis of no differences holds. Thus, it can be concluded that no statistically significant differences between the newly processed data sets exist. The rejection of the run of newly processed data set E2, shown in Fig. 3B, is the result of higher peak areas for triphenylphosphine oxide, C26, C28, and C30. These peak areas are approximately three times higher than results for newly processed data set E1. Data sets E1 and E2 were analysed with the same instrument, but for data set E1, only basic maintenance was carried out; for E2, the instructed thorough cleaning was done. It is possible that the new column, liner, or other consumables affected the increased variability in the results, due to e.g. manufacturing or maintenance differences.

3.1.2. Separate parameter comparison

The data for RI, peak tailing, peak area, and signal-to-noise ratio were analysed separately to look for more specific variation in the results. RIs were generally comparable, and the variability of results within each data set was minimal. The original data processing performed by laboratories was done automatically, but manual enhancements have possibly been done to correct errors in automatic processing. The new data processing by one operator at VERIFIN was conducted only automatically, which could have led to limited detection of some of the peaks. In the analysis of the original data RIs, data set C1 showed clear data processing differences, such as C30 getting an RI value of 3600, when it should be 3000. This difference might have resulted from faulty calibration or peak identification by the data processing software. Therefore, when calculating the ranges for results (maximum value –

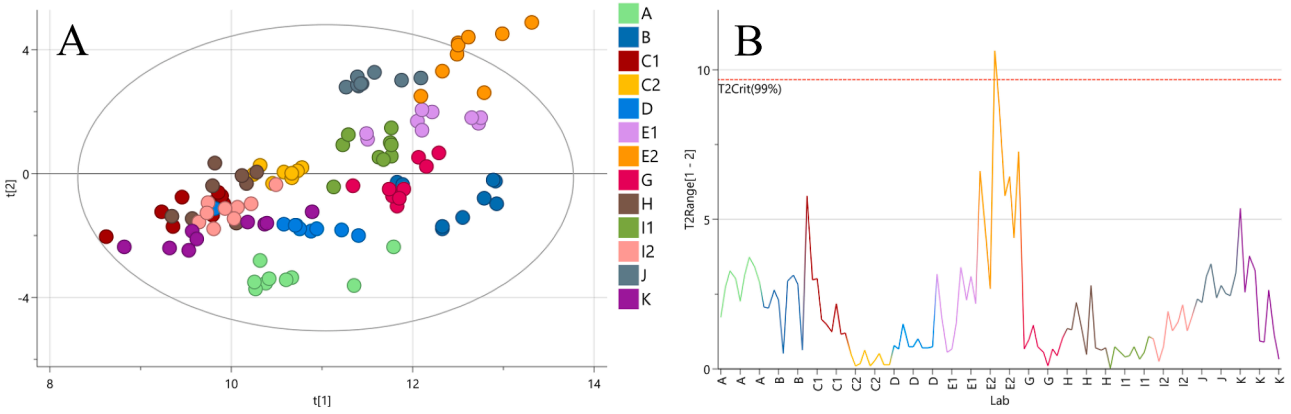


Fig. 3. A) PCA for the newly processed and normalised data with a critical limit of 95%. B) Hotelling's T^2-test for the newly processed and normalised data with a critical limit of 99%.

minimum value), an additional version was calculated without C1. General OCAD criteria instruct that RI values are acceptable if they are ±20 units from the library value. As all of the studied QC compounds are not included in the OCAD library, and therefore do not have a reference RI value against which to make comparisons, a general 40-unit range of results was examined here. Table 3 lists the RI results for all compounds and data sets, the mean value for each compound, and the range of results with and without C1. Each RI value listed is the average value for the nine GC–MS analyses. The range of results without C1 shows 21 of the 27 compounds receiving a value below 40. From the six compounds exceeding the 40-unit limit, docosane (C22) and C30 could give acceptable results (2200 and 3000 respectively) if RI calibration is done correctly. The other four compounds exceeding the limit are trimethyl phosphate, trimethylphosphine oxide, methyl stearate, and triphenylphosphine oxide. In the results for data set H, trimethylphosphine oxide was either not detected, or it eluted after decane and 1,4-dithiane, unlike with other data sets.

The peak tailing values exhibited large differences based on the calculation algorithm and software used. A general formula for the calculation, used for example in Agilent Quantitative Analysis [33], is shown in Fig. S03. This method of calculation compares the distances between the front and back sides of the peak, measured at 5 or 10% of peak height, whilst the method used in AMDIS compares the peak areas of the front and back of the peak [34]. These differences in calculation algorithms are presumably the cause for result differentiation between data processing software.

Analyses of peak tailing in the original data processed by the laboratories shows the separation of data sets A2, C1 and C2, similarly to Fig. 2B. Data set F had similar results to others in the original data. However, in the results for the newly processed data (Fig. S04), the values for different data sets are very similar to each other, and none of the data sets are close to the 99% critical value. This strengthens the hypothesis that differences in data processing software cause the differentiation of data.

The peak area results show that data set E2 is separated from the rest of the group for both the original data (Fig. S05) and the newly processed data (Fig. S06). The reason for this is the same as in the analysis with all parameters in chapter 3.1.1 Comparison of combined parameters: higher peak areas for triphenylphosphine oxide, C26, C28, and C30. Data set E2 differs from the rest also due to its disproportionally higher area values for dicyclohexylamine, triphenylphosphine oxide, C26, and C28.

In the analyses for signal-to-noise ratio, data sets F and J are separated in the Hotelling's T^2 analysis for the original data (Fig. S07), and data set E2 is separated in the Hotelling's T^2 analysis for the newly processed data (Fig. S08). Original data set F has up to seven times higher values for dicyclohexylamine than other data sets and slightly higher values for alkane chains. Original data set J has two GC–MS runs above the 99% critical limit, and these runs seem to have slightly higher values for multiple compounds compared to the other data sets. However, the S/N ratio value for hexadecane for these two runs was slightly lower than with their other runs, which could affect the results after normalisation. In the newly processed data, the results for hexadecane for these two runs were not lower than for the other runs, and hence no separation from other data sets was observed. When comparing the original data, laboratory J processed the S/N ratios with the MassHunter software instead of AMDIS, and this could have affected the results. However, similar deviation was not observed in other data sets where MassHunter was used. In the newly processed data analyses, data set E2 differed from the rest due to same compounds as with peak area comparisons: triphenylphosphine oxide, C26, C28, and C30.

The isotopic ratios were evaluated from spectral abundances and according to the guidelines set in OCAD. These guidelines include an acceptance range of 33 ± 3.0%-unit for 5-chloro-2-methylaniline ($^{37}Cl/^{35}Cl$) (m/z 143/141), and 5.3 ± 1.0%-unit for dibenzothiophene ($^{34}S/^{32}S$) (m/z 186/184). The results for the original data sets fell within the limits of acceptance for all but one data set and one compound (Figs. S09 and S10). For the analysis of dibenzothiophene, original data set F showed results below the acceptance range. This might have occurred due to the differences in analysis with the high resolution instrument that laboratory F used. However, acceptable results for data set F were acquired for 5-chloro-2-methylaniline.

3.1.3. Low concentration analyses

The reproducibility of results for concentrations close to the detection limit were studied separately for four data sets. Three dilutions of the original QC sample were analysed in this study to investigate each compound's detection limit. The concentrations of compounds in the solutions ranged from 10 μg/ml in the original solution to 1 ng/ml in the most diluted solution. The results in Table 4 represent the lowest concentrations in which each compound was observed in each data set. The data analysis and processing were done by one operator at VERIFIN. The peaks were visually evaluated from the TIC, with a minimum S/N ratio of 3, as the current verification analysis QC criteria are also monitored from the TIC. However, as there is a possible need for low concentration analyses for chemical fingerprints, the raw GC–MS data was additionally processed with AMDIS to evaluate the detection of compounds from the extracted ion chromatograms (EICs) with deconvolution software.

The results were similar for most of the compounds, with a maximum of a 10-fold difference in detected concentrations. For C26 and the alkanes eluting after it, up to a 100-fold difference in detected concentrations was observed. Alkane sequence compounds from octane to eicosane were predominantly detected at a 0.05 μg/ml concentration level, lower than other compounds. Some other compounds (1,4-dichlorobutane, dibenzothiophene, and 1,4-dithiane) were detected at a 0.1 μg/ml concentration level. As expected, compounds with a low S/N ratio in the original analyses, such as trimethylphosphine oxide, 2-ethylhexanoic acid, and methyl stearate, were generally not detected in the diluted samples. With visual evaluation from the TIC, no compounds were detected with concentrations of 10, 5, or 1 ng/ml. However, AMDIS deconvolution could identify 1,4-dithiane, 1,4-dichlorobutane, 2,6-dimethylphenol, and decane at a 5–10 ng/ml level.

3.2. Quality control sample development

The quality control sample for chemical forensics was developed on the basis of the current QC sample used by OPCW designated laboratories in the verification analysis of CWAs [1]. From the 27 compounds in the new test mixture, 16 are also present in the verification analysis QC sample. The performance and properties of all the compounds were examined manually from the raw GC–MS data and statistically with PCA. Variability in or rejection of results for a compound may indicate its suitability for the QC sample if the changes in results originate from altered performance of the GC–MS instrument. Compounds with acceptable results can also be included in the QC sample if their inclusion serves a purpose, such as alkanes for RI calculation, dibenzothiophene for isotope measurements, or 2,6-dimethylphenol for column basicity testing. If a compound is concluded to be of no benefit to the solution, it can be excluded from the sample.

The data used for interlaboratory comparison was transposed to enable evaluation of the compounds. The hexadecane-normalised analyses were conducted separately for the original data and the newly processed data. The performance of the compounds was additionally examined with the complex matrix runs included.

3.2.1. Complex matrix samples' influence on the QC sample

In order to examine the effect of complex matrices, such as soil, to the QC-sample analysis results obtained from the GC–MS analyses, the laboratories analysed complex matrix samples followed by the QC-samples. These experiments were designed to reveal the effect of the challenging matrix to the GC–MS condition. The complex matrices used in the analyses included soil, hydroalcoholic solution, solvent mix,

Table 3
Mean RI values for each data set and compound separately. Original values obtained by the laboratories are shown. Ranges of results with and without data set C1, and standard deviations for each compound are displayed. OCAD library values are listed for compounds included in the library. *Values exceeding the 40 unit range.

Compound	A1	A2	B	C1	C2	D	E1	E2	F	G	H	I1	I2	J	K	Range of results	Range without C1	Experimental mean RI value	Standard deviation	OCAD value
Heptane					717		700	700	700	701	701	717	724		699	24	24	706	9.8	700
Octane	801	800	800	816	816	800	800	800	800	800	800	800	800	800	800	16	16	802	5.6	800
1,4-Dichlorobutane	912	911	903	927	926	909	907	905	907	913	912	908	908	908	911	24	23	911	6.8	
Trimethyl phosphate	929	928	918	939	939	924	932	930	932	927	967	923	923	923	928	49*	49*	931	11.6	938
Trimethylphosphine oxide	973	975	933	954	959	948	944	939	951	954		939	939	937	957	42*	42*	950	12.9	
C10	1000	1000	998	1015	1015	1000	1000	997	1000	1000	1000	1000	1001	1000	1003	18	18	1002	5.5	1000
1,4-Dithiane	1083	1083	1072	1093	1091	1079	1070	1067	1072	1086	1078	1078	1079	1076	1076	26	24	1079	7.4	1071
2,6-Dimethylphenol	1115	1116	1108	1125	1125	1111	1114	1109	1112	1117	1121	1111	1111	1110	1109	18	18	1114	5.6	1112
2-Ethylhexanoic acid	1121	1121	1120	1133	1134	1117	1137	1133	1120	1121	1138	1117	1116	1120	1114	24	24	1124	8.3	
C12	1200	1200	1200	1214	1214	1200	1200	1200	1200	1200	1200	1200	1200	1200	1200	14	14	1202	4.9	1200
5-Chloro-2-methylaniline	1314	1314	1302	1320	1319	1307	1310	1306	1309	1317	1322	1307	1308	1305	1304	20	20	1311	6.3	1308
C14	1400	1400	1400	1413	1413	1400	1400	1400	1400	1400	1400	1399	1400	1400	1400	14	14	1402	4.6	1400
Dicyclohexylamine	1433	1433	1428	1448	1443	1433	1433	1429	1429	1436	1436	1432	1431	1431	1428	19	15	1433	5.5	
TEPP	1565	1566	1559	1573	1571	1562	1581	1581	1579	1567	1595	1561	1562	1562	1562	35	35	1570	10.2	
C16	1600	1600	1600	1612	1612	1600	1600	1600	1600	1600	1601	1599	1600	1600	1600	13	13	1602	4.2	1600
Tributyl phosphate	1645	1646	1640	1653	1651	1642	1654	1654	1653	1645	1663	1641	1642	1642	1642	23	23	1648	6.7	1655
Dibenzothiophene	1788	1789	1768	1797	1798	1779	1773	1767	1779	1795	1796	1783	1785	1772	1777	31	31	1783	10.6	1774
C18	1800	1800	1800	1810	1810	1800	1800	1800	1800	1800	1800	1799	1800	1800	1800	11	11	1801	3.6	1800
Malathion	1972	1972	1961	1976	1974	1965	1984	1983	1985	1974	1997	1967	1967	1964	1964	36	36	1974	10.0	1986
C20	2000	2000	2000	2009	2009	2000	2000	2000	2000	2000	2001	2000	2000	2000	2000	9	9	2001	3.2	2000
Methyl stearate	2130	2130	2126	2134	2072	2127	2130	2130	2130	2130	2135	2127	2127	2127	2125	63*	63*	2125	15.0	2130
C22	2200	2200	2200	2207	2153	2200	2200	2200	2200	2200	2201	2200	2200	2200	2200	54*	48*	2197	12.4	2200
C24	2400	2400	2400	2407	2407	2400	2400	2400	2400	2400	2401	2400	2400	2400	2400	8	8	2401	2.4	2400
Triphenylphosphine oxide	2561	2561	2526	2683	2562	2545	2552	2546	2561	2570	2595	2543	2544	2535	2541	156*	69*	2562	37.4	
C26	2599	2600	2600	2805	2606	2604	2600	2600	2600	2600	2601	2585	2586	2600	2600	219*	21	2612	53.6	2600
C28	2800	2800	2786	3105	2804	2795	2800	2797	2800	2792	2801	2783	2791	2800	2794	321*	21	2817	80.0	2800
C30	3000	3000	2969	3604	3003	3014	3059	2999	3000	3000	2978	3033	3051	3000	3000	635*	90*	3047	155.8	3000

Table 4
Low concentration analysis results with visual evaluation from the TIC. Result concentrations are in μg/ml scale and the colour scale is used to represent the concentrations. Undetected compounds are indicated with a blank (white) cell.

Compound	A	D	I1	I2
Heptane		1		
C8	0.05	0.05	0.05	0.05
1,4-Dichlorobutane	0.1	0.1	0.1	0.1
Trimethyl phosphate	0.5	0.5	0.5	0.5
Trimethylphosphine oxide	10	10	10	10
C10	0.05	0.05	0.05	0.05
1,4-Dithiane	0.1	0.1	0.1	0.1
2,6-Dimethylphenol	0.05	0.05	0.05	0.05
2-Ethylhexanoic acid	10	10	10	1
C12	0.05	0.05	0.05	0.05
5-Chloro-2-methylaniline	0.5	0.5	0.05	0.5
C14	0.05	0.05	0.05	0.05
Dicyclohexylamine	0.1	1	0.1	0.1
Tetraethyl pyrophosphate	1		10	1
C16	0.05	0.05	0.05	0.05
Tributyl phosphate	0.5	0.5	0.5	0.5
Dibenzothiophene	0.1	0.1	0.1	0.1
C18	0.5	0.05	0.05	0.05
Malathion	0.5	0.05	0.5	0.5
C20	0.5	0.05	0.05	0.05
Methyl stearate	5	5	5	5
C22	0.5	0.05	0.5	0.5
C24	5	0.05	0.5	0.5
Triphenylphosphine oxide	1	10	1	1
C26	10	0.1	1	1
C28	10	1	1	1
C30	10	1	1	1

diesel, and sediment, but the most common matrix was soil. In addition, different sample preparation and derivatisation methods were used. Clear separation can be seen between the analyses before and after the complex matrix runs for peak area data for data sets I1, J, and K (Fig. 4). Differences in peak area result analyses could as well be observed with original data sets B, C2, D, F, and I2.

Peak area increase and decrease after complex matrix runs depended on the instrument used. The peak areas of original data set I1 decreased after complex matrix runs, while those of original data set I2 increased. Data sets I1 and I2 were obtained using the same complex matrix sample, and the instruments were from Agilent, albeit different models (Table S02). The RI of trimethylphosphine oxide increased by 20 units in original data set I1's analyses after the complex matrix samples, and slight RI changes in four other original data sets took place as well.

Phosphorous containing compounds TEPP, trimethylphosphine oxide, and triphenylphosphine oxide experienced disproportionally high increases in peak areas and occasionally in peak height. The peak height

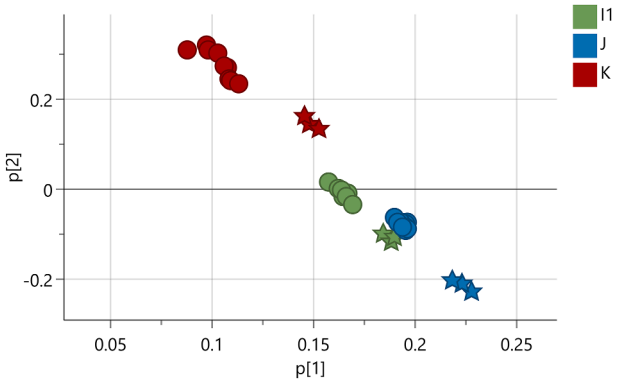


Fig. 4. Loadings plot of analyses after the complex matrix samples, and clean runs for peak area data from three data sets (chosen for visual presentation). Data sets are indicated with colours, clean runs with circles, and runs after the complex matrix samples with stars.

of dicyclohexylamine decreased to nearly undistinguishable from the baseline in data sets I1 and I2 after the complex matrix runs. Additionally, in analyses of data set E1, which were conducted only after basic maintenance, the peak height of dicyclohexylamine was less than 50% of that of tetradecane. In analyses of data set E2, which were conducted after complete cleaning and maintenance, the peak height of dicyclohexylamine was approximately 90% of that of tetradecane. Due to its activity, dicyclohexylamine needs to remain included in the QC mix and its criteria. The criteria include the assessment of the peak height of dicyclohexylamine against the peak height of tetradecane, since the peak of tetradecane is stable, and therefore can be used as a standard to measure against. Similarly, the peak height of trimethyl phosphate is assessed against the stable peak of dibenzothiophene, as it has successfully been in the verification analysis QC sample previously.

3.2.2. *Compound comparison*

PCA (Fig. 5A) was conducted on the newly processed data combining all parameters but excluding the complex matrix sample runs. In these analyses, any increased differentiation of a compound is investigated to study trends in the results, regardless of whether the critical limits are exceeded. The first principal component accounted for 86% of result variability, and the second for 5.9%. The loadings plot (Fig. 5B) displays the first two principal components separating the peak tailing, peak area, and S/N ratio parameters nearly completely, with a clear definition in peak tailing results. Peak area and S/N ratio results give roughly the same results, which can be expected due to their usual correlation. The contribution of RIs to the variability was lower than other parameters', and it is not shown in these principal components.

Observing the PCA of data combining all parameters (Fig. 5A), trimethylphosphine oxide was the most separated from the other compounds, as it exceeded the 95% critical limit, while 2-ethylhexanoic acid, trimethyl phosphate, and C26 produced values resulting in a slight difference as well. 2-Ethylhexanoic acid was separated due to its low peak tailing values of approximately 0.4, compared to approximately 1–1.5 for other compounds. 2-Ethylhexanoic acid's reactivity and coelution with 2,6-dimethylphenol would suggest its removal from the QC mixture. However, as polar compounds such as acids will inevitably exist in forensic samples, we propose a substitutive acid with a different RI to be included in the QC mix. This way, interlaboratory differences in elution of very polar compounds will be monitored, their reflection on forensic samples considered, and the coelution of two QC compounds will be avoided. This acid substitution will be done after additional testing in the future.

The loadings plot (Fig. 5B) shows that the separation of trimethylphosphine oxide and trimethyl phosphate might mainly occur due to peak tailing, and the separation of late eluting compounds (triphenylphosphine oxide, C26, C28, and C30) due to peak area. It is important to include C26, C28, and C30 in the QC mix for RI calculations and monitoring of late eluting compounds. Triphenylphosphine oxide can be included in the QC mix for monitoring the changes in peak area, especially after complex matrix samples, and due to its late elution, which is beneficial for the verification of RI calibration.

In addition to the combined analysis for parameters, all parameters except RI were analysed separately to identify more specific differences in the data. The comparison of compounds with RIs is shown in Table 3.

Peak tailing results are similar to the combined results of all parameters, as the combined results were significantly affected by the peak tailing values. Trimethylphosphine oxide, trimethyl phosphate, and 2-ethylhexanoic acid are separated from the other compounds, the first two due to high peak tailing values, and 2-ethylhexanoic acid due to low values. Trimethyl phosphate is included in the QC mix as it is already in the verification analysis QC and its peak tailing indicates column activity. Trimethylphosphine oxide routinely yields results with higher values of peak tailing, and no results with a better peak shape were obtained. However, some variability in the detection of the compound occurred, and therefore its inclusion in the QC mix is proposed for

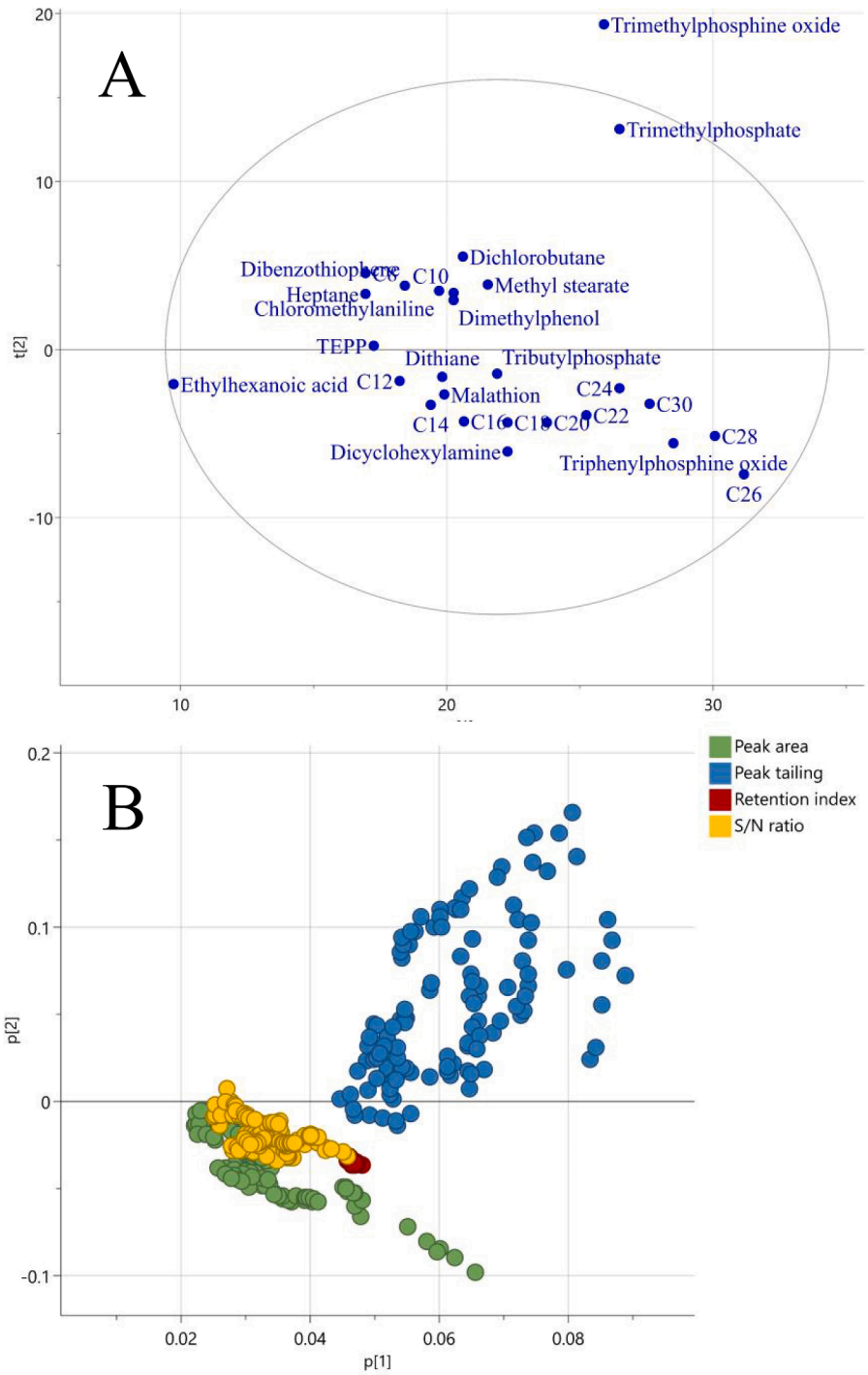


Fig. 5. A) PCA for newly processed combined parameters data with a critical limit of 95%. B) Loadings plot for all parameters combined and newly processed data.

further examination. The analysis including the complex matrix samples gave similar results but on a higher scale, i.e. trimethylphosphine oxide got values of 48 on the first principal component t[1] without the complex matrix samples (Fig. S11), and values of 63 with the complex matrix samples (Fig. S12).

Peak area results displayed the separation of dicyclohexylamine, triphenylphosphine oxide, C26, C28, and C30 in analyses with and without complex matrix samples. This separation was caused by data sets F and E2 obtaining higher peak area values for these compounds compared to other data sets. Similarly to the complex matrix data's effect on peak tailing, the result distribution scale is increased when analyses following complex matrix samples are included (Figs. S13 and S14).

The results for signal-to-noise ratios showed the separation of dicyclohexylamine in the original data. This separation occurred due to data set F obtaining significantly higher results. The inclusion of analyses that followed complex matrix samples increased the scale (i.e. created more separation) compared to the original, but only slightly (the loading of dicyclohexylamine for the first principal component increased from 14 to 16, Figs. S15 and S16).

TEPP could be removed from the QC sample due to its lack of instrument performance–testing properties, such as column activity measurement or RI calculation, not already present in other QC compounds. In addition, heptane was observed only in 60% of the data sets

due to its coelution with the solvent peak, and therefore its detection cannot be included in the QC criteria. However, it could be included for RI calculation if detected, and as an additional marker for instrument performance not counting towards data acceptance.

The compounds chosen for the new QC mixture and their corresponding concentrations are listed in Table S07.

3.2.3. Quality control analysis criteria

QC criteria are established for the evaluation of GC–MS instrument performance. If the set criteria are not passed, the analyses cannot be started, and corrective actions must be taken before continuing. Implementation of the criteria will allow the monitoring of the comparability of results between separate laboratories. Criteria listed in Table 5 are a proposition on which the final criteria will be built after further confidence building tests. The proposition includes the criteria used in the current QC for verification analysis and additional new criteria based on the results of this study. The new criteria focus on low concentration compounds, properties of newly added compounds, and specifications for different data processing software.

As 1,4-dichlorobutane and 1,4-dithiane were successfully detected in low concentrations by most of the laboratories, they are proposed to be added to the QC-mixture to measure the capability of the instrument to detect low concentrations. The proposed concentrations for these compounds are 50 and 100 ng/ml, respectively. The addition of pentadecane is proposed for the measurement of the 10 ng/ml concentration, as alkanes were detected easily in low concentrations, but are used for the calculation of RIs and must therefore have a 5–10 μg/ml concentration in the QC sample. Additionally, pentadecane's RI would suggest no coelution with other compounds in the QC mix. The suitability of pentadecane to the QC sample will be tested in future confidence building exercises. Detection of the low concentration compounds would not be set as a mandatory criterion yet, but instead the compounds would provide additional information for instrument operators on the ability of the GC–MS to detect compounds at low concentrations. A general understanding of instruments' detection capabilities would aid in the description of chemical fingerprints.

3.2.4. Other compounds detected in the analyses

It is important to consider the other compounds present in the QC mix as well. Reactivity of QC compounds can result in impurities and affect the passing of QC criteria. Reaction products of two QC compounds, dicyclohexylamine and 2-ethylhexanoic acid, were observed in the chromatograms of some of the data sets (Table S08). Two reaction products of dicyclohexylamine, *N,N*-dicyclohexylmethylamine and *N*-ethyldicyclohexylamine, were both detected in six data sets. For 2-ethylhexanoic acid, trimethylsilyl (TMS) and *tert*-butyldimethylsilyl (TBDMS) reaction products were detected, as well as 2-ethylhexanoic anhydride. 2-Ethylhexanoic anhydride was detected in 12 data sets, the TMS by-product in two data sets in the analyses preceding the complex matrix samples, and the TBDMS by-product in three data sets in the analyses following the complex matrix runs. These reaction products for both compounds occurred most likely due to their acidic and basic properties. The TMS and TBDMS reactions could result from contamination of the liner with traces of a silylation agent. The amount of these and other impurities was approximately 200 in all data sets together, but the impurities did not affect the QC compounds or the criteria due to their low amounts.

Table 5
Proposed QC criteria for chemical forensics analyses of CWAs.

Parameter	Compound and criteria
Isotope ratios	Isotope ratios of 5-chloro-2-methylaniline ($^{37}Cl/^{35}Cl$) (*m/z* 143/141) must be 33 ± 3.0%. Isotope ratios of dibenzothiophene ($^{34}S/^{32}S$) (*m/z* 186/184) must be 5.3 ± 1.0%.
Peak height	Trimethyl phosphate must be at least 20% of that of dibenzothiophene (if concentrations are equal). Dicyclohexylamine must be at least 50% of that of C14 when concentrations are 10 and 5 μg/ml, respectively.
S/N ratio (TIC)	Over 10:1 for all compounds with concentrations of 5–10 μg/ml when evaluated visually When calculated with AMDIS, over 150 for all compounds with concentrations of 5–10 μg/ml, except methyl stearate, for which the S/N ratio should be over 100.
Peak tailing	Trimethylphosphine oxide's peak tailing manually calculated < 6. All other compounds AMDIS calculated ≤ 4.
Retention index	Calibration from C8 to C30; from C7 if observed. All compounds ±20 units from their OCAD specified value for those listed in OCAD.
Mass spectra	The *m/z* values of major ions for each test chemical are correct and no extra ions above 5% are present.
Additional future criteria	
Low concentration analyses	Pentadecane detected at concentration 10 ng/ml.
	1,4-Dichlorobutane detected at concentration 50 ng/ml.
	1,4-Dithiane detected at concentration 100 ng/ml.

4. Conclusions

The results of the interlaboratory comparison analyses suggest that there is no statistically significant difference between the laboratories in GC–MS QC-analyses. To minimise variability in results, all analysis methods including data processing must be synchronised. Identical results cannot be achieved, but different instruments can obtain results similar enough for reliable comparison when normalised with run specific information. HRMS instruments are sufficiently comparable with unit resolution instruments after normalisation; however, comparisons within the same mass resolution would be preferable in isotope ratio analyses.

The comparison of compounds in the QC sample was conducted for the original and newly processed data with PCA and visually from the TIC. The low concentration analyses were additionally deconvoluted with AMDIS to compare the results from the TIC to the EICs. The analyses on all data resulted in the identification of compounds with differing properties, such as the high peak tailing values for trimethylphosphine oxide or an increased peak area for phosphorous containing compounds. With the help of these analyses and the additional low concentration analyses, new QC criteria were proposed. The new criteria listed in Table 5 include measurements of isotope ratios for 5-chloro-2-methylaniline and dibenzothiophene, peak height for trimethyl phosphate and dicyclohexylamine, S/N ratio for all compounds with visual assessment and AMDIS processing, peak tailing for all compounds with AMDIS and trimethylphosphine oxide manually, RI with a range from 700 to 3000, and mass spectra for accurate *m/z* values. In addition, pentadecane (10 ng/ml), 1,4-dichlorobutane (50 ng/ml), and 1,4-dithiane (100 ng/ml) were added for low concentration analyses, since the formation of a chemical fingerprint often relies on compounds with low concentrations. However, the detection of these compounds does not guarantee the detection of all compounds with the same concentration, as the properties of each compound can affect their detection limit substantially. The negative effects of the complex matrix samples on the condition of the GC–MS instrument demonstrated the need for sufficient cleaning of the instrument before the analyses. Additionally, to avoid contamination or undesired reactions in the inlet, frequent replacement of the injector liner and other consumables is recommended.

These analyses demonstrate the adequate similarity of interlaboratory analysis results with GC–MS(EI) in addition to the continuous need for development and testing of quality control methods for chemical fingerprint analysis. With the rapid progress in the field of chemical forensics for CWAs, it is essential to gain a deeper understanding of the formation process of chemical fingerprints to prevent underlying sources of error. Future confidence building tests will continue this work to improve QC measures for GC–MS and aid in the implementation of

chemical forensics methodologies.

CRediT authorship contribution statement

Solja Säde: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Project administration. **Grégoire Delaporte:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing. **Carlos G. Fraga:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing. **Hanna Hakulinen:** Conceptualization, Methodology, Resources, Writing – review & editing. **Karin Höjer Holmgren:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing. **Marie Spiandore:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing. **Crister Åstot:** Conceptualization, Methodology, Investigation, Resources, Writing – review & editing. **Anuradha Akmeemana:** Investigation, Resources, Writing – review & editing. **Mirjam de Bruin-Hoegée:** Investigation, Resources, Writing – review & editing. **Justin Doward:** Investigation, Resources, Writing – review & editing. **Karen Jacques:** Investigation, Resources, Writing – review & editing. **Saiful Anuar Kesah:** Investigation, Resources, Writing – review & editing. **Samantha See:** Investigation, Resources, Writing – review & editing. **Stuart Thomson:** Investigation, Resources, Writing – review & editing. **Jens Flinck:** Methodology, Formal analysis, Resources, Writing – review & editing. **Leena Kalliovirta:** Methodology, Writing – review & editing. **Paula Vanninen:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Lina Mörén (Swedish Defence Research Agency) and Ivana Sofrenić (University of Belgrade, Faculty of Chemistry) are acknowledged for their advice on the SIMCA software.

The authors acknowledge the following funders: Ministry for Foreign Affairs of Finland (receiver: S. Säde), The National Centre for Forensic Science, a Florida SUS recognized research centre at the University of Central Florida (receiver: A. Akmeemana), The Dutch Ministry of Defence (receiver: M. de Bruin-Hoegée), United States State Department Bureau of Arms Control, Verification and Compliance (AVC) Key Verification Assets Fund (receiver: C. Fraga).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.talo.2023.100249.

References

[1] T. Schneider, T. Lütkefend, Nowhere to Hide: The Logic of Chemical Weapons Use in Syria, Global Public Policy Institute, 2019. https://gppi.net/2019/02/17/the-logic-of-chemical-weapons-use-in-syria (accessed 3 October 2022).
[2] OPCW, Conference of the State Parties, Addressing the Threat from Chemical Weapons Use, C-SS-4/DEC.3, 27 June 2018.
[3] C.L. Mazzitelli, M.A. Re, M.A. Reaves, C.A. Acevedo, S.D. Straight, J.E. Chipuk, A systematic method for the targeted discovery of chemical attribution signatures: application to isopropyl bicyclophosphate production, Anal. Chem. 84 (2012) 6661–6671, https://doi.org/10.1021/ac300859j.
[4] A. González-Pereira, P. Otero, M. Fraga-Corral, P. Garcia-Oliveira, M. Carpena, M. A. Prieto, J. Simal-Gandara, State-of-the-art of analytical techniques to determine food fraud in olive oils, Foods 10 (2021) 484, https://doi.org/10.3390/foods10030484.
[5] M. Morelato, A. Beavis, M. Tahtouh, O. Ribaux, P. Kirkbride, C. Roux, The use of forensic case data in intelligence-led policing: the example of drug profiling, Forensic Sci. Int. 226 (2013) 1–9, https://doi.org/10.1016/j.forsciint.2013.01.003.
[6] J.R. Mallette, J.F. Casale, J. Jordan, D.R. Morello, P.M. Beyer, Geographically sourcing cocaine's origin – delineation of the nineteen major coca growing regions in South America, Sci. Rep. 6 (2016) 23520, https://doi.org/10.1038/srep23520.
[7] J.R. Ehleringer, J.F. Casale, M.J. Lott, V.L. Ford, Tracing the geographical origin of cocaine, Nature 408 (2000) 311–312, https://doi.org/10.1038/35042680.
[8] C.G. Fraga, G.A. Pérez Acosta, M.D. Crenshaw, K. Wallace, G.M. Mong, H. A. Colburn, Impurity profiling to match a nerve agent to its precursor source for chemical forensics applications, Anal. Chem. 83 (2011) 9564–9572, https://doi.org/10.1021/ac202340u.
[9] R.L. Webster, S.P.B. Ovenden, L.J. McDowall, G.H. Dennison, M.J. Laws, N. W. McGill, J. Williams, S.D. Zanatta, Chemical forensic profiling and attribution signature determination of sarin nerve agent using GC–MS, LC–MS and NMR, Anal. Bioanal. Chem. 414 (2022) 3863–3873, https://doi.org/10.1007/s00216-022-04027-1.
[10] S.P.B. Ovenden, R.L. Webster, E. Micich, L.J. McDowall, N.W. McGill, J. Williams, S.D. Zanatta, The identification of chemical attribution signatures of stored VX nerve agents using NMR, GC-MS, and LC-HRMS, Talanta 211 (2020), 120753, https://doi.org/10.1016/j.talanta.2020.120753.
[11] K. Höjer Holmgren, C.A. Valdez, R. Magnusson, A.K. Vu, S. Lindberg, A. M. Williams, A. Alcaraz, C. Åstot, S. Hok, R. Norlin, Part 1: tracing Russian VX to its synthetic routes by multivariate statistics of chemical attribution signatures, Talanta 186 (2018) 586–596, https://doi.org/10.1016/j.talanta.2018.02.104.
[12] D. Jansson, S.W. Lindström, R. Norlin, S. Hok, C.A. Valdez, A.M. Williams, A. Alcaraz, C. Nilsson, C. Åstot, Part 2: forensic attribution profiling of Russian VX in food using liquid chromatography-mass spectrometry, Talanta 186 (2018) 597–606, https://doi.org/10.1016/j.talanta.2018.02.103.
[13] A.M. Williams, A.K. Vu, B.P. Mayer, S. Hok, C.A. Valdez, A. Alcaraz, Part 3: solid phase extraction of Russian VX and its chemical attribution signatures in food matrices and their detection by GC-MS and LC-MS, Talanta 186 (2018) 607–614, https://doi.org/10.1016/j.talanta.2018.03.044.
[14] L. Mörén, J. Qvarnström, M. Engqvist, R. Afshin-Sander, X. Wu, J. Dahlén, C. Löfberg, A. Larsson, A. Östin, Attribution of fentanyl analogue synthesis routes by multivariate data analysis of orthogonal mass spectral data, Talanta 203 (2019) 122–130, https://doi.org/10.1016/j.talanta.2019.05.025.
[15] B.P. Mayer, A.J. DeHope, D.A. Mew, P.E. Spackman, A.M. Williams, Chemical attribution of fentanyl using multivariate statistical analysis of orthogonal mass spectral data, Anal. Chem. 88 (2016) 4303–4310, https://doi.org/10.1021/acs.analchem.5b04434.
[16] M. de Bruin-Hoegée, D. Kleiweg, D. Noort, A.C. van Asten, Chemical attribution of fentanyl: the effect of human metabolism, Forens. Chem. 24 (2021), 100330, https://doi.org/10.1016/j.forc.2021.100330.
[17] L. Mörén, P. Lindén, J. Qvarnström, M. Engqvist, M. Carlsson, R. Afshin Sander, S. Lindberg, A. Larsson, A. Östin, Classification of carfentanil synthesis methods based on chemical impurity profile, Foren. Chem. 26 (2021), 100355, https://doi.org/10.1016/j.forc.2021.100355.
[18] S.P.B. Ovenden, L.J. McDowall, H.E. McKeown, N.W. McGill, O.A.H. Jones, J. R. Pearson, M. Petricevic, M.L. Rogers, T.J. Rook, J. Williams, R.L. Webster, S. D. Zanatta, Investigating the chemical impurity profiles of fentanyl preparations and precursors to identify chemical attribution signatures for synthetic method attribution, Forens. Sci. Int. 321 (2021), 110742, https://doi.org/10.1016/j.forsciint.2021.110742.
[19] B.P. Mayer, C.A. Valdez, A.J. DeHope, P.E. Spackman, A.M. Williams, Statistical analysis of the chemical attribution signatures of 3-methylfentanyl and its methods of production, Talanta 186 (2018) 645–654, https://doi.org/10.1016/j.talanta.2018.02.026.
[20] K. Höjer Holmgren, L. Mörén, L. Ahlinder, A. Larsson, D. Wiktelius, R. Norlin, C. Åstot, Route determination of sulfur mustard using nontargeted chemical attribution signature screening, Anal. Chem. 93 (2021) 4850–4858, https://doi.org/10.1021/acs.analchem.0c04555.
[21] K. Höjer Holmgren, S. Hok, R. Magnusson, A. Larsson, C. Åstot, C. Koester, D. Mew, A.K. Vu, A. Alcaraz, A.M. Williams, R. Norlin, D. Wiktelius, Synthesis route attribution of sulfur mustard by multivariate data analysis of chemical signatures, Talanta 186 (2018) 615–621, https://doi.org/10.1016/j.talanta.2018.02.100.
[22] C.G. Fraga, A.V. Mitroshkov, N.S. Mirjankar, B.P. Dockendorff, A.M. Melville, Elemental source attribution signatures for calcium ammonium nitrate (CAN) fertilizers used in homemade explosives, Talanta 174 (2017) 131–138, https://doi.org/10.1016/j.talanta.2017.05.066.
[23] E.D. Strozier, D.D. Mooney, D.A. Friedenberg, T.P. Klupinski, C.A. Triplett, Use of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometric detection and random forest pattern recognition techniques for classifying chemical threat agents and detecting chemical attribution signatures, Anal. Chem. 88 (2016) 7068–7075, https://doi.org/10.1021/acs.analchem.6b00725.
[24] X. Lu, Z. Zhang, H. Liu, H. Tang, R. Gao, C. Pei, H. Wang, J. Xiao, Forensic signatures of a chemical weapon precursor DMPADC for determination of a synthetic route, Talanta 232 (2021), 122476, https://doi.org/10.1016/j.talanta.2021.122476.
[25] L. Mörén, S. Jonsson, T. Tengel, A. Östin, Origin identification of homemade pepper spray by multivariate data analysis of chemical attribution signatures, Forens. Sci. Int. 304 (2019), 109956, https://doi.org/10.1016/j.forsciint.2019.109956.
[26] N.S. Mirjankar, C.G. Fraga, A.J. Carman, J.J. Moran, Source attribution of cyanides using anionic impurity profiling, stable isotope ratios, trace elemental analysis and

chemometrics, Anal. Chem. 88 (2016) 1827–1834, https://doi.org/10.1021/acs.analchem.5b04126.
[27] K. Höjer Holmgren, H. Hakulinen, R. Norlin, M. de Bruin-Hoegée, M. Spiandore, S. Qi Shu See, R. Webster, K.L. Jacques, L. Mauravaara, L. Hwi Ang, C.P. Evans, S. Ovenden, D. Noort, G. Delaporte, J. Dahlén, C.G. Fraga, P. Vanninen, C. Åstot, Interlaboratory comparison study of a chemical profiling method for methylphosphonic dichloride, a nerve agent precursor, Forens. Chem. 33 (2023), 100473, https://doi.org/10.1016/j.forc.2023.100473.
[28] C.G. Fraga, A. Breton-Vega, K. Höjer Holmgren, C. Åstot, P.S. Vanninen, H. Hakulinen, G. Delaporte, M. Spiandore, J. Ng, L.Hwi Ang, Evaluation of Inter-Laboratory Variability in Chemical Threat Agent Impurity Profiling by Gas Chromatography /Mass Spectrometry, Submitted 2023.
[29] M. Rautio (Ed.), Interlaboratory Comparison Test Coordinated by the Provisional Technical Secretariat for the Preparatory Commission For the Organisation for the Prohibition of Chemical Weapons, 1st ed., the Ministry for Foreign Affairs of Finland, Helsinki, 1994. ISBN: 951-724-009-0 (paperback).
[30] P. Vanninen (Ed.), Recommended Operating Procedures For Analysis in the Verification of Chemical Disarmament, 2017 Edition, University of Helsinki, Helsinki, 2017. ISBN: 978-951-51-3916-0 (paperback), ISBN 978-951-51-3917-7 (pdf).
[31] OPCW, Note by the Technical Secretariat, Procedures for the Evaluation of Data to be Included in the OPCW Central Analytical Database, S/1686/2018, 13 November 2018.
[32] Sartorius, SIMCA 15 User Guide. https://www.sartorius.com/download/544940/simca-15-user-guide-en-b-00076-sartorius-data.pdf, 2020 (accessed 15 February 2023).
[33] Agilent Technologies, Agilent MassHunter Workstation Software Quantitative Analysis. https://www.agilent.com/cs/library/usermanuals/public/G3335-90000%20QuantitationDataSet.pdf, 2017 (accessed 13 April 2023).
[34] NIST, AMDIS - User Guide. https://chemdata.nist.gov/mass-spc/amdis/docs/amdis.pdf, 1997 (accessed 13 April 2023).