



Exploring the chemistry of complex samples by tentative identification and semi-quantification: a food contact material case

Pieke, Eelco Nicolaas; Smedsgaard, Jørn; Granby, Kit

Published in:
Journal of Mass Spectrometry

Link to article, DOI:
[10.1002/jms.4052](https://doi.org/10.1002/jms.4052)

Publication date:
2018

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Pieke, E. N., Smedsgaard, J., & Granby, K. (2018). Exploring the chemistry of complex samples by tentative identification and semi-quantification: a food contact material case. *Journal of Mass Spectrometry*, 53(4), 323-335. <https://doi.org/10.1002/jms.4052>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Exploring the chemistry of complex samples by tentative identification and semi-quantification: a food contact material case

Eelco N. Pieke*, Jørn Smedsgaard, Kit Granby

Technical University of Denmark, National Food Institute, Research Group for Analytical Food Chemistry, Kemitorvet Building 202, Kgs. Lyngby, DK-2800, Denmark

Abstract

In fields such as food safety and environmental chemistry, ensuring safety is greatly challenged by large numbers of unknown substances occurring. Even with current state of the art mass spectrometers, dealing with non-identified substances is a very laborious process as it includes structure elucidation of a vast number of unknowns, of which only a fraction may be relevant. Here, we present an exploration and prioritization approach based on high resolution mass spectrometry. The method uses algorithm-based precursor/product-ion correlations on Quadrupole-Time of Flight (Q-TOF) MS/MS data to retrieve the most likely chemical match from a structure database. In addition, TOF-only data is used to estimate analyte concentration via semi-quantification. The method is demonstrated in recycled paper food contact material (FCM). Here, 585 chromatographic peaks were discovered, of which 117 were unique to the sample and could be tentatively elucidated via accurate mass, isotopic pattern, and precursor/product-ion correlations. Nearly 85% of these 117 peaks were matched with database entries, which provided varying certainty of information about the analyte structure. Semi-quantitative concentration ranges of investigated compounds were between $0.7 \mu\text{g dm}^{-2}$ and $1600 \mu\text{g dm}^{-2}$. With this data, a subgroup of chemicals was risk-categorized and prioritized using the most likely candidate structure(s) obtained. Prioritization based on expected health impact was possible using the tentatively assigned data. Overall, the described method is a valuable chemical exploration tool for non-identified substances, but also may be used as a preliminary prioritization tool for substances expected to have the highest health impact, for example in FCMs.

Keywords: structure assessment; mass spectrometry; semi-quantification; exploration; food contact materials

* Corresponding author; mail: enpi@food.dtu.dk, phone: +45 518 642 05

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/jms.4052

1. Introduction

The increasing awareness of adverse effects of chemicals in our environment, especially in food, has led to ever more demanding questions about occurrence of unwanted chemicals. Traditionally, analytical methods are designed to detect specific groups of chemicals using targeted methods.

However, these methods are limited to the original method scope so fail to detect chemicals outside this scope. Fortunately, newer different analytical methods may address the challenge of unknown compounds which are not known to be present or where the structure is not initially expected to be present. This challenge is prevalent in many areas, e.g., metabolomics [1], food safety [2,3], and environmental research [4,5]. Unknown chemicals shift the research question from “can *this* substance be found in this sample?” towards “what *else* can be found in this sample?” While targeted analyses are known for high sensitivity and useful for routine analysis [6], they can similarly provide a falsified argument of safety: absence of evidence is not equal to evidence of absence. Food scandals, notably the melamine case [7,8], have proved that targeted methodology cannot always ensure safety due to the large number of substances not included.

When one takes into account the non-targeted substances, orders of 100,000s of chemical substances may be found in food contact materials, which may have the potential to migrate to food in concentrations of concern [9]. Research is often directed according to the “popularity” of a substance: much for well-studied compounds, but few or none for unknown compounds. The lack of knowledge on unexpected compounds in food and environment has driven a need to look past the targeted methodology. Yet, even untargeted methods still have a too strong dependence on pre-existing knowledge [10], and the available data sources require extensive cleaning and pruning prior to use [11]. Dealing with complex samples, especially to ensure human safety, requires a level of discovery that targeted or even some untargeted analysis do not currently provide [12].

Modern accurate MS provides the capability to predict molecular structures from isotopic patterns, MS-MS fragmentation, retention behavior, and elemental composition [13,14]. Even so, MS is rarely sufficient for a full identification, and authentic standards or other elucidation methods are still required, e.g., purification followed by analysis by Nuclear Magnetic Spectroscopy (NMR) [15–17]. Recently, Kind and Fiehn [18] reviewed unknown elucidation, and summarized promptly that there is currently no software or tool that can truly predict an unknown structure from MS spectra. Hence, the unknown chemicals tend to direct to the void between targeted analysis (not included in method) and untargeted analysis (unable to be identified or quantified).

Food science (flavor, safety, etc.), metabolomics, and environmental chemistry often use MS combined with reference libraries of known chemicals [19–22]. Food safety, risk assessment relies

on the potential hazard and the exposure to these compounds [23], hence the challenge is in two questions: *what* is present in the sample and *how much*. We recently showed that quantifications could be decoupled from identification [24], and this allows for rough concentration estimates of virtually an unlimited amount of unidentified substances. Furthermore, studies by Bengtström *et al.* [25] and Rosenmai *et al.* [26] showed that untargeted screening can be used in a bio-guided strategy to find hazardous compounds present in extracts from cardboard food packaging materials. This approach aims to avoid wasting effort identifying and quantifying irrelevant compounds by employing prioritization [27,28]. Yet, ideally elucidation, quantification, and prioritization should be performed on *any* present substance to assess which substances are risk-relevant or risk-irrelevant. It may be viable to adapt a strategy similar to recent literature [25,26] and combine it with the latest semi-quantification approach [24] to obtain a chemical exploration tool that can prioritize chemical substances based on adjustable endpoints.

In this paper, we present a fast workflow based on high resolution mass spectrometry to combine structure prediction methodology with semi-quantitative methodology to generate a profile of potential hazardous compounds in complex samples. To demonstrate the feasibility of such a method, we apply it in a food contact material (FCM) setting. The investigated workflow provides a much-needed tool for complex sample exploration: the abundances and chemical profiles that can be used for further investigations.

2. Materials and methods

2.1 Software

Marvin (16.6.13.0, 2016, ChemAxon, <http://www.chemaxon.com>) and Calculator Plugins were used for drawing, displaying and characterizing chemical structures, calculating properties, and converting names. JChem Standardizer (JChem 16.5.2.0, 2016, ChemAxon) was used for structure cleanup and pruning. JChem for Office (16.1.1100.489, 2016, ChemAxon) was used for chemical reporting in Excel. Agilent MassHunter Molecular Structure Correlator (8.0.36.0, 2016, Agilent Technologies, Santa Clara, CA, USA) was used to correlate accurate mass MS/MS fragment ions to databases to search for a most likely match. Agilent MassHunter Acquisition (B.06.01, 2016, Agilent Technologies) was used for data acquisition. Agilent MassHunter Qualitative Analysis (B.07.0.7024.29, 2016, Agilent Technologies) was used for data analysis via in-built Molecular Feature Extraction (MFE), and further data treatment of TOF MS and Q-TOF MS/MS data.

2.2 Instrumental setup

2.2.1 Electrospray Ionization Quadrupole-Time of Flight mass spectrometry

Mass analysis was performed using an Agilent 6550 Quadrupole-Time of Flight (Q-TOF) mass spectrometer (Agilent Technologies, Santa Clara, CA, USA) equipped with Agilent JetStream (AJS) electrospray ionization (ESI) interface. Equipment setup and optimization follows recent work [24]. The optimized parameters for the ESI source are shown in Table 1. The Q-TOF MS was auto-tuned for mass accuracy and resolution between m/z 100 to m/z 3200 prior to each sequence, and mass axes were 2-point calibrated real-time using supplied ESI calibration mixture.

Mass spectral experimental were run two modes: TOF only (MS) and Q-TOF (MS/MS). In TOF mode, acquisition was performed in the mass range 50 – 2500 m/z at 1.3 Hz with no collision gas or collision energy. During product-ion acquisition, the mass range was reduced to 50 – 2400 m/z at 2.5 Hz with the quadrupole isolation width set to “Narrow” (1.3 m/z width) and at varying collision energy voltages: 10V, 20V, and 40V. The Auto-MS/MS feature was used to isolate precursor ions dynamically in real-time. Common background ions were added to the Auto-MS/MS “Exclusion” list to improve spectral quality. Other settings of the Auto-MS/MS were tuned for this specific instrument and the samples.

2.2.2 Liquid chromatography

UHPLC-MS was performed on an Agilent 1290 system (Agilent, Waldron, Germany). For separation, a separation strategy was adopted as described in [24] using two UHPLC columns connected serially to improve separation power and selectivity.. Both columns were held isothermally at 50°C. Injection volume was 5 μ l. Eluent flow was 0.2 ml min⁻¹. The first column was a Phenomenex Luna Omega Polar C18 (100 Å, 1.6 μ m, 100 x 2.1 mm) (Phenomenex, Denmark). The second column was a Waters ACQUITY UPLC CSH C18 (130 Å, 1.7 μ m, 100 x 2.1 mm) (Waters, Denmark). Columns were protected using the Phenomenex SecurityGuard™ ULTRA (Phenomenex, Denmark) fitted with Omega Polar C18 phase. Eluent A was MilliQ-grade (Millipore Corp., Bedford, MA, USA) water, and eluent B was TOF-grade methanol (Fluka, Denmark). Both eluent A and B contained 2 mM ammonium formate buffer (pH 3.5).

The columns were selected for improved selectivity towards polar compounds (Luna Omega) and free acids (ACQUITY) while retaining high capacity C18-like separation. Methanol was chosen for compatibility with buffer salts, a favorable interaction with sugars, and being a generally weaker eluent allowing a smoother gradient. Eluent gradient was as follows: after injection gradient was held for 3 min at 0% B, then increased to 90% B at 25 min, then ramped to 99% B at 33 min, and

finally set to 100% B at 33.01 min, performing isocratic elution for 5 min until 38 minutes. Post-time cleaning (4 min) involved flushing the injection loop and column with 100% A.

2.3 Sample preparation

2.3.1 Sampling

Samples were taken from an unused recycled food cardboard item (pizza-box) used in retail applications. The sample was chosen as recent work has shown substances of concern can migrate from these types of samples [29]. From the center of the bottom of the box, a 1 dm² board piece was cut out using a cleaned surgical knife. Sample cutout was resized to four identically sized strips of 2.5 cm x 10 cm. Before extraction, all non-disposable glassware, e.g. glass extractor vessel, glass vials (PYREX), and measuring glasses were cleaned and heated to 400°C for a minimum of 12 hours. All plastic equipment was soaked in 95% ethanol for 24 hours before use. LC-MS vials (Agilent Technologies) and Pasteur pipettes were used as received.

Four sample strips were placed in an extraction vessel, and 100 mL of warm (40 – 50°C) extraction solvent was added. The extraction vessels were capped and sealed with a metal clamp, transferred into a thermally sealed airtight bag, and placed in an oven for 24 hours at 40°C. After 24 hours, the vessels were removed from the oven, and the liquid content was poured into 100 mL glass bottles where it was cooled down to room temperature. A blank extract was prepared by following the extraction procedure without sample.

The resulting extract was filtered in triplicate and transferred to LC-MS compatible vials for analysis. Each sample was filtered using a plastic 2 mL BD Discardit™ syringe (Becton Dickinson, USA) and plastic-PTFE 0.2 µm Phenex™ filters (Phenomenex, Denmark), thoroughly rinsed with 50% ethanol. The syringe was used in a way that ensured there was an air cavity between plunger and solvent to prevent contamination. First, 0.9 mL extract was filtered to waste followed by 0.5 mL of extract filtered into the LC-MS vial. 1.0 mL of quantification marker (QM) mixture was added to each LC-MS vial. The contents of the QM mixture are described in Supplementary Material Table A.2.1.

2.4 Molecular Structure Correlation

The Agilent MassHunter Molecular Structure Correlator (MSC) program proposes a structure suggestion of an analyte of interest (AOI) by comparing MS/MS data containing accurate mass, fragment ions, and isotopic patterns with local or online databases of available chemical structures. Each potential molecular formula of an AOI is compared against a database of theoretically fragmented chemical structures as described by Hill and Mortishire-Smith [30]. As experimental spectra are compared to theoretical spectra, a matching score is calculated based on fragment

matching, mass accuracy matching of precursor and fragment ions, and the overall percentage of fragment ions that can be plausibly explained [31]. As a result the MSC reports the most plausible chemical structure inside the consulted database that fits best the experimental MS/MS data.

2.5 Database preparation

Seven data sources (DS) were used to generate a set of five databases. Three DS were chosen for high specificity to sample types. DS #1 and #2: compiled list from research regarding contaminants in paper and board material [29,32]. DS #3: the US FDA database 'Indirect Additives used in Food Contact Substances' [33]. Two DS were selected for moderate relevance to the field of usage. DS #4: REACH Annex III database containing chemicals used in industry that have some associated safety concerns [34]. DS #5: Agilent MassHunter Extractables and Leachables library, containing extractable and leachable compounds found in components of (mostly plastic) food and drug packaging and medical devices [35]. Finally, two large DS were consulted online. DS #6: ChemSpider (<http://www.chemspider.com/>). DS #7: PubChem [36].

Before use as database, DS #1 to #4 were made converted to a structure format and pruned. In addition, DS #1 - #3 were merged into a single database (PBDB) due to large overlap. DS #5, #6, and #7 needed no further modification. The conversion and pruning for each DS firstly involved converting entries (CAS or name) to *simplified molecular-input line-entry system* (SMILES) by using a combination of ChemAxon software and in-house Excel scripts. After conversion, the SMILES database was pruned for duplicate, erroneous, or empty entries. The conversion process is described in greater detail in Supplementary Material A.1 and Figure A.1.1.

3. Results and discussions

3.1 Databases and pruning performance

The mechanism of MSC is designed to propose the best-matching structure inside a database of structures, thereby critically dependent on the appropriateness, quality, and size of the databases. Hence, the *database scope* and *database specificity* are decisive quality factors in structural elucidation. In Table 2 a quality assessment of different databases is presented. Here, the *scope* of database is proportional to the database size and uniqueness: a larger scope increases the likelihood of obtaining a positive or false positive match from a mass spectrum. The *specificity* of the database is based on the relevance of database to the sample type, i.e., whether the substances in the database are representative to the analytical problem and sample types. Database *authenticity* is a qualitative parameter based on the *size* and *specificity*. Hence, authenticity is a measure of the trustworthiness of a MSC prediction from that database.

In Table 2, the PBDB and the REACH database scored well on authenticity due to a mid-sized scope along with a high specificity. The Agilent PCDL had diminished use due to a relatively small scope, but also a low specificity as it is designed primarily for plastic substances. The ChemSpider and PubChem databases could not be considered very authentic despite the largest scope, as these databases included structures from a highly diverse pool of possible applications. Here, structural correlations may shift to overfitting data: small measurement errors in the data can be fitted on the near-unlimited number of theoretical spectra present in these databases. This does not imply correlation results from ChemSpider and PubChem are not usable, but it does imply that the user needs to be critical with the matching results.

3.2 The role of databases in structure correlation

Databases PBDB, REACH, and AgiEL are strongly relevant to the investigated samples, and the structure predictions can be used without reassessment. However, for CS and PC, the predictions require manual assessment of the proposed structure, as these predictions were found to contain anomalies, e.g., excess triple bonds, prevalent radicals, or carbon rings of $n < 5$. Entries containing said anomalies required reassessment because the presence in the REACH Annex III database, representative of a large proportion of industrial chemicals in use, was uncommon: only 0.64% of the entries contained a single triple bond, while 0.037% contained a ring with $n = 3$, and 0.019% a ring with $n = 4$. Thus, if the best-scoring hit in CS or PC contained no structural anomalies it was accepted, but if anomalies were present the next-best hit with no anomalies was accepted.

The issue with anomalies entries highlights a somewhat larger issue in using untargeted predictions for structural elucidation. Firstly, available databases are used *as-is*: the search range is limited by the content of the database. Therefore, chemical structures that are not originally found in the database cannot be proposed. Consequently, if the actual Analyte of Interest (AOI) structure is not in any database then no exact structure is obtained and at most a similar structure may be found.

Secondly, fitting experimental data to nonspecific large databases involves a high risk of false positive results. The nonspecific public online databases (such as ChemSpider and PubChem) are useful for elucidating unknown chemicals, but there also is a need for specific databases. Specific databases require extensive knowledge on the sample type, but are relatively easy to construct, inexpensive, and unlikely to comprise proprietary materials as the only input required are molecular structures. Hence, specific databases might be constructed in collaboration with industrial companies that often have deep knowledge on the samples.

3.3 Using structural elucidation in untargeted work

3.3.1 Data-dependent acquisition

Multi-stage mass spectrometry (e.g., QqQ, Q-TOF, or ion traps) are often associated with targeted analyses: the mass-to-charge (m/z) of AOI is preselected, and fragmentation spectra are collected only for preselected fragments (Selected- or Multiple Reaction Monitoring (SRM or MRM)). Because of preselection, SRM/MRM results in product ion spectra of sufficient abundance and high quality. However, SRM/MRM is not well-suited for any sort of discovery that involves unknowns (Figure 1a). In fact, SRM/MRM is blind to any mass features not included in preselection as illustrated in Figure 1b.

Due to the limitation of SRM as required for untargeted analyses, alternative data acquisition methods are in development such as data-dependent acquisition (DDA) [37,38]. In DDA, real-time data is evaluated for precursor ions based on TOFMS data, and an algorithm selects relevant precursors for fragmentation. This permits the investigation of unknown analytes, as no prior definition is needed. Another methodology is data-independent acquisition (DIA) [39,40], in which all ions are fragmented simultaneously without selection. DIA results in complex product-ion spectra and data analysis is challenging, while spectral quality is not optimal [41,42]. This study exclusively uses DDA due to the added complexity of performing DIA.

A limitation associated with DDA is the compromise needed for complex samples containing many unknown substances [43]. The number of precursor ions to be analyzed by fragmentation is limited by time required per scan cycle, so only a limited number of precursors can be included within a given time range [41]. If a large number of viable precursors are discovered, those with highest abundance are preferentially selected while the remainder is discarded (Figure 1c), which is problematic for chromatographic areas with many eluting substances, or where background peaks are dominant.

To control the selection of ions in DDA, preference/exclusion monitoring based on a preset list of priority ions (preference list) was used. The design of a preference list requires TOF-only (MS) analysis to identify the time and m/z of interesting peaks, and this data was already available from the quantification experiments. Preference monitoring ensures certain ions and/or fragments have fragmentation spectra collected, which is beneficial for investigating low-abundance analytes. Unfortunately, designing the preference list is a manual task and, for large chromatograms, can be time-intensive. Consequently, the investigative power of the method is enhanced by combining real-time targeting with a pre-selected list of ions of interest.

The experimental results indicated that DDA deals poorly with different adducts generated by the ESI process. The DDA algorithm cannot unambiguously identify the adduct type of the precursor, and the most abundant adduct ion did not consistently produce high-quality fragmentation spectra. In Figure 2, the product-ion spectrum for three different adducts for a single structure in positive ESI are shown. Fragmentation spectra based on protonated ($[M+H]^+$), deprotonated ($[M-H]^-$) or ammoniated ($[M+NH_4]^+$) adducts were generally of high quality, while adducts based on alkaline salts (Na^+ or K^+) produced little to no fragmentation. This signifies a flaw in DDA: some compounds ionize dominantly as salt adducts which consecutively are selected as precursor ion due to the high intensity. Yet, in order for fragmentation to be effective, the precursor ion must not be sodium ($[M+Na]^+$), potassium ($[M+K]^+$), or formate ($[M-COO]^-$) adducts. Ensuring proper adduct selection can partly be achieved by using DDA preference monitoring to select optimal adducts for fragmentation spectra despite having a lower initial abundance..

3.4 FCM analysis: identification and quantification

3.4.1 Analysis workflow

To evaluate the method performance on a real sample, a recycled cardboard pizza-box was investigated. Pizza-boxes contain both virgin and recycled board fibers, so they are expected to contain a diverse set of chemicals with a potentially large probability of containing unknowns [44,45]. Figure 3 shows the simplified workflow from sample workup to quantification and elucidation. The extracted sample was analyzed in triplicate (A1 to A3) in both positive and negative mode ESI. Due to the size of the dataset produced, only a single sample in positive ESI (A1) is discussed here to demonstrate the method.

After sample preparation, sample and blanks were run in TOF mode. TOF MS data was analyzed by the Molecular Feature Extraction (MFE) algorithm in MassHunter Qualitative Analysis software, and the results were overlaid with the blank extraction. A DDA preference monitoring list was generated by recording the retention time and precursor mass of analytes which were distinct in the sample compared to the blank. When several adducts were observed for a preferred compound, a maximum two precursor ions were selected with a priority order of: $[M+H]^+ > [M+NH_4]^+ > [M+Na]^+ > [M+K]^+$. Adduct prioritization was not necessary in negative ESI mode due to the lower number of possible adducts.

Semi-quantification (SQ) was limited to the largest 1200 MFE groups (i.e., possible compounds) per sample to reduce computation times for the MFE algorithm. First, mass spectrometric data was investigated for the MFE groups of the quantification markers (QM). MFE groups consist of multiple ion traces originating from the same analyte. For each compound, the summed group ion traces

were recognized as being the true abundance of the analyte in the sample. Actual SQ was achieved by comparing each analyte to its nearest QM [24].

Finally, compounds detected in both Q-TOF (identification) and TOF (quantification) were coupled by using Excel VBA scripts. The coupling occurred only when both methods resulted in similar retention time and similar m/z of principal ions.

3.4.2 Identification

Triplicate analysis of the pizza box sample (A1, A2, and A3) resulted in 133, 131, and 131 compounds, respectively. To facilitate structure assignment, MSC predictions were categorized as either authentic predictions, based on PBDB, AgiEL or REACH, or non-authentic predictions, based on PubChem or ChemSpider. Two simplified rules were used to decide whether a structure assignment by MSC could be passed as relevant:

1. A threshold average prediction score was above 70: at least 70% of the observed Q-TOF spectrum was explained by the database structure.
2. A compound required a match from at least 2 databases, and at least one match originated from an 'authentic' source (PBDB, REACH, or AgiEL).

The performance of the assignments was assessed by the total prediction statistics as summarized in Figure 4 for sample A1.

Assessing the quality of the predictions revealed that nearly 85% of the predictions (Figure 4A) were of sufficient quality usable for interpretation (score > 70). The average prediction score was 78; thus, on average more than three-quarters of the fragmentation spectrum could be explained. A substantial part of the predictions scored between 70 and 80 (37.6%), but the majority scored 80 or above (47.4%). Some predictions were considered of insufficient quality (score < 70) for interpretation (15%), and most of these prediction scores were between 50 and 70 (11.3%).

When looking at the prediction quantity n (Figure 4B), more than 83% of possible AOI had $n \geq 3$ predictions that could support structure assignment. For $n = 2$, these predictions were exclusively from CS and PC. As discussed, CS and PC are probable to contain false positive predictions, so the MSC results were evaluated manually. However, a lack of authentic predictions may indicate that these were truly unknown compounds not yet available in any database. Cases where insufficient prediction reliability ($n = 0$ or $n = 1$) occurred were rare: only in 3 out of 133 cases.

3.4.3 Quantification

To simplify the description and tables in these results, only data from A1 is used here. In sample A1, 585 potential peaks were semi-quantified. Only compounds for which the concentration could be estimated in at least two out of three samples were included. A coefficient of variation (CV) was calculated for 109 compounds using replicate experiments A1, A2, and A3. The median CV was 10% while the average CV was 17%. However, compounds eluting before 3.5 minutes retention (gradient start) contributed more to the variation, if these were excluded, the median CV was similarly 10%, but the average CV reduced to 13%.

Figure 5 summarizes the semi-quantitative results. The range of concentrations calculated by semi-quantification was about five orders of magnitude between $2.7 \text{ nmol liter}^{-1}$ and $159 \text{ } \mu\text{mol liter}^{-1}$, which approximately corresponded to a range from 470 ppb to 17 ppm. The median of concentration estimates was $68 \text{ nmol liter}^{-1}$. The range of concentrations of entries that were successfully linked between quantification and identification (marked in Figure 5 by filled points) was between $19 \text{ nmol liter}^{-1}$ and $159 \text{ } \mu\text{mol liter}^{-1}$. As the masses of the linked compounds are known from the identification, the concentration could be expressed as mass unit per dm^2 : $0.7 \text{ } \mu\text{g dm}^{-2}$ to $1653 \text{ } \mu\text{g dm}^{-2}$.

3.4.4 Combining semi-quantification and identification data

The 585 semi-quantifications were compared with the 133 structure elucidations, which resulted in 117 compounds with both structure and concentration assignment. The unlinked 468 semi-quantifications, marked in Figure 5 by transparent points, could not be coupled to existing MS/MS data. The 16 identified but not quantified compounds were probably due to a targeted ion being of such low abundance that the data processing did not detect the principal ion.

There appeared to be a trend in the coupling of semi-quantification with identification where the successful links are often at the higher range of concentrations. This effect was expected: DDA can only simultaneously evaluate a certain quantity of ions, and during saturation the higher abundance signals take priority. This indicates the balance needed between operating in MS mode and MS/MS mode: fragmentation spectra (MS/MS) be used to evaluate new candidate ions, but shortening the acquisition time in MS/MS reduces the quality of the spectra due to lower abundance. Hence, to collect spectra from sub-ppm level compounds or those with elution overlap, the precursor ions need to be part of the DDA preference list.

Alternatively, the elution time window could be broader so that DDA can utilize the longer time to evaluate more ions. A broader elution window could be beneficial as the DDA algorithm excludes peaks after a set number of MS/MS spectra; so a longer time would free up the queue for lower

abundance precursor ions. However, broader peaks are also inversely related to abundance and generally unfavorable as the peak overlap increases significantly thus reinforcing the initial problem.

3.5 Assessment of quantification and identification of a paper and board FCM

In risk studies, the most prominent decision is whether the presence of a certain substance is acceptable (i.e. safe and/or legal), or unacceptable (i.e. present a risk and/or is illegal) and needs to be considered for further study. Targeted methods are more sensitive than untargeted screening [20], so highly potent toxic substances are not properly assessed via the latter methodology since they require sufficient sensitivity [46]. Assessing risk and/or legal status for the remaining chemical substances requires a combination of knowledge about hazard characterization and structure: to assess health effects, or determine if it is a regulated substance; and about the concentration: to determine if the human exposure poses a risk, or is above the legal threshold.

Initially, the number of discovered compounds in this study is not as high as literature might suggest [9]. The real number of compounds is most likely higher if the DDA were optimized to include more low-abundance peaks. However, also a change in analytical methodology will increase the number and diversity of chemicals detected, e.g., different HPLC columns or different ionization principles. In this case study, analytes that are strong nonpolar (e.g. polycyclic aromatic hydrocarbons, mineral oils) are excluded by the use of ESI, while those that are very small or polar (e.g. ionic substances, monomers, solvents) could be excluded by using C₁₈ liquid chromatography.

Not all substances are discussed or reported in this paper due to the size of the total dataset (133 compounds in triplicate); however, illustrative examples are presented to document the method and output. A subset (25) of compounds was investigated sorted by concentration, of which 20 have the highest estimated concentration, and 5 have the lowest estimated concentration. The results (Table 4) show the estimated concentration (without the 3-fold error from [24]) in both $\mu\text{mol liter}^{-1}$ and $\mu\text{g dm}^{-2}$, together with the structures obtained from MSC. Empty structures imply no structure was predicted using this database.

3.5.1 Case: insufficient information

The evaluation of data was somewhat complicated when structure predictions contained little agreement and/or low scores. There is not sufficient evidence to prioritize further studies if databases predicted inconsistent or low-score structures, or if the estimation of concentration is unavailable. This is the case for compounds C.109 & C.110, C.65, C.39, C.34, C.36, and C.14, which are all at relatively low concentration with ambiguous structure assignment. The most straightforward option will be to treat the results as tentative and conclude no consensus can be reached. This solution, however, is controversial as there is no prioritization, and it implies

postponing the assessment. A second approach is to reanalyze the sample under more specific targeting conditions. However, this may not actually lead to a better result, especially if the actual chemical simply is not available in the databases and/or as reference standard. In light of a safety assessment, a more practical approach may be to use the available structures, in context of the worst-possible application. As human safety is the top concern, one could consider the worst-case chemical proposed as the leading case when in doubt, and then proceed to evaluate the worst-case scenario. This would often provide an over-estimation of risk, but most risk-based evaluation consider that the preferred method. Unfortunately, there is not yet a solution if there is lacking or conflicting data, and it remains the decision of the user in light of the application to decide the best course of action.

3.5.2 Case: low expected risk

A number of found chemicals were classified as low risk. For example, the compounds with the highest estimated concentration C.3 and C.2 appear to be an amino-ethanol structure, with high concentrations (including uncertainty) in the interval from 0.15 to 4.5 mg dm⁻². The actual structure appears fairly certain across database predictions. These chemicals are often used in the production processes, and have very little known toxicological effects attributed to them [47]. The LD50 for rats is remarkably high at 2.46 mL kg⁻¹ [48]. Similarly, C.115 appears to be an aliphatic carboxylic amide, like stearamide, between 20 and 150 µg dm⁻². These type of compounds are also known to have a relative low toxicity [49]. Other examples of low-risk substances are compounds C.10, C.11, and C.18, which indicate polyethylene glycol (PEG) at relatively high concentrations between 185 and 1630 µg dm⁻². PEG is historically known to have few issues associated with its intake [50,51]. Finally, compound C.6 seems to be a sugar-like compound that is unlike to be a risk.

3.5.3 Case: possible or uncertain expected risk

Some chemicals cannot be directly classified as a low or high risk. An example is compound C.16 where structure prediction suggests dehydroacetic acid, which is a fungicide and common industrial chemical, or maltol acetate, which also is an industry chemical, at concentrations from 25 to 225 µg dm⁻². The assigned structures to C.16 did not have high scores, so the suggested and actual structure is uncertain. Dehydroacetic acid is permitted in squash at maximum 65 ppm [52] and as adhesive in packaging [53] by the U.S. FDA, but dehydroacetic acid was recently discovered in coatings of cheese in the EU which caused a withdrawal [54]. There are no known regulations for maltol acetate. Because the identification did not have a high score and the use of the possible chemicals is debatable, C.16 is an example of a possible but no certain risk. Another case in this category is compound C.13. The structure proposed by one authentic database is ethyl N-benzyl-β-alaninate at 50 to 450 µg dm⁻², and PC predicted a similar compound, while CS proposed a lower-score

alternative. Unfortunately, neither have toxicological data or known uses for the proposed structures while the concentration is high (2.5 to 22 $\mu\text{mol liter}^{-1}$), which merit further studies. As discussed before, in cases of uncertainty the safe approach is more appropriate, which marks classifying C.13 as possible risk.

3.5.4 Case: certain expected risk

Only a few of the predicted structures seem to call for immediate attention. A notably entry is C.118, that all predictions indicate as a phthalate-like substance with good prediction score (>80), levels between 70 and 600 $\mu\text{g dm}^{-2}$, and good agreement between structure predictions. Due to the large number of rather similar phthalates, determining the exact structure is not possible. Phthalates are found frequently in food contact materials due to world-wide use and migration of phthalates into food from plastics, but to a lesser extent also from a paper-based material, e.g. recycled paper and board [55]. In addition, it has been argued that the continuous exposure to phthalates is often related to adverse health effects in humans [56,57], so some risk may be expected. A final case discussed here is compound C.9, where the structure assignment points to methylisothiazolinone based on authentic database predictions with scores > 80 , supported by non-authentic predictions showing highly similar structures. This substance is frequently used in cosmetics, water tanks, and paper mills as biocide or preservative, and is considered to cause skin sensitization [58,59] and other effects [60]. It is among the substances with higher concentration (40 to 360 $\mu\text{g dm}^{-2}$), yet very little testing on chronic dietary exposure has been performed, so the presence in food contact materials could be of concern.

4. Conclusion

In this study we investigated combining semi-quantification with structural elucidation as a tool for exploration of chemical risk. We have shown that an exploration tool based on elucidative and semi-quantification principles can be used on complex samples to give an essential overview of present chemical substances and concentration levels occurring, here demonstrated using a paper and board food contact material extract.. The method is designed to be applicable on most samples analyzed by LC-MS since it does not rely on availability of a large number of standards. Instead, the method is capable to operate on structure libraries without extensive MS/MS data, and these can be created inexpensively as long as information on the sample type is available, for example via collaboration with product manufacturers or authorities. In essence, the method permits chemical exploration of the world of chemically diverse and complex samples with simplicity, speed, and only few requirements for prior experiments or standardized methods.

Evidently, no “one size fits all” methodology for comprehensive unknown screening can be provided yet. Instead, exploration should be performed with the aim to obtain as much and as broad information as possible, followed by selecting the relevant information to evaluate further. Indeed, an indication of concentration and identity are already useful data to any field dealing with unknown substances. With both quantitative data and indications of chemical structures available, decisions can be taken at a much greater knowledge level than possible with the lack of knowledge that analysis of unknowns currently imposes.

The dependence on standards or existing knowledge on the sample is not completely eliminated because relevant structure databases are essential for results, and this marks one of the method limitations. Substances that are not (yet) in a database are hard to retrace, and predicting true unknowns with the current data tools is not possible. However, this is a first step towards a simplification of identification databases, having already reduced the need for fully recorded MS/MS spectra and instead relying on structure-only databases, but MS/MS spectra are still the preferred source of data. Ideally, future software tools will be able to understand MS data better, and synthesize new structures based purely on mass spectra. This means that truly unknown substances could be eligible for elucidation. In addition, it would be valuable to see data-independent acquisition (DIA) applied in this methodology for its improved precursor throughput, although the issues with low-quality fragment spectra would need to be resolved. Finally, the use of explorative data in different fields or in risk assessments needs to be evaluated.

5. Acknowledgements

The authors would like to thank Signe Helle Ptak for her contributions to data acquisition. The authors report no conflict of interest. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The authors would like to thank the Technical University of Denmark (DTU) and the Danish Veterinary and Food Authority (DVFA) for financial support to enable this study.

6. References

- [1] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics, *Mass Spectrom. Rev.* 26 (2007) 51–78. doi:10.1002/mas.20108.
- [2] A.K. Malik, C. Blasco, Y. Picó, Liquid chromatography-mass spectrometry in food safety, *J. Chromatogr. A.* 1217 (2010) 4018–4040. doi:10.1016/j.chroma.2010.03.015.
- [3] M. Biedermann, K. Grob, Assurance of safety of recycled paperboard for food packaging through comprehensive analysis of potential migrants is unrealistic, *J. Chromatogr. A.* 1293 (2013) 107–119. doi:10.1016/j.chroma.2013.04.009.
- [4] M.R. Viant, U. Sommer, Mass spectrometry based environmental metabolomics: A primer and review,

- Metabolomics. 9 (2013) 144–158. doi:10.1007/s11306-012-0412-x.
- [5] S.D. Richardson, Environmental mass spectrometry: Emerging contaminants and current issues, *Anal. Chem.* 84 (2012) 747–778. doi:10.1021/ac202903d.
- [6] K.L. Thoren, J.M. Colby, S.B. Shugarts, A.H.B. Wu, K.L. Lynch, Comparison of information-dependent acquisition on a tandem quadrupole TOF vs a triple quadrupole linear ion trap mass spectrometer for broad-spectrum drug screening, *Clin. Chem.* 62 (2016) 170–178. doi:10.1373/clinchem.2015.241315.
- [7] J. Chen, A worldwide food safety concern in 2008--melamine-contaminated infant formula in China caused urinary tract stone in 290,000 children in China., *Chin. Med. J. (Engl.)* 122 (2009) 243–4. doi:10.3760/cma.j.issn.0366-6999.2009.03.001.
- [8] C.E. Handford, K. Campbell, C.T. Elliott, Impacts of Milk Fraud on Food Safety and Nutrition with Special Emphasis on Developing Countries, *Compr. Rev. Food Sci. Food Saf.* 15 (2016) 130–142. doi:10.1111/1541-4337.12181.
- [9] K. Grob, Work plans to get out of the deadlock for the safety assurance of migration from food contact materials? A proposal, *Food Control.* 46 (2014) 312–318. doi:10.1016/j.foodcont.2014.05.044.
- [10] E.L. Schymanski, H.P. Singer, J. Slobodnik, I.M. Ipolyi, P. Oswald, M. Krauss, T. Schulze, P. Haglund, T. Letzel, S. Grosse, N.S. Thomaidis, A. Bletsou, C. Zwiener, M. Ibáñez, T. Portolés, R. De Boer, M.J. Reid, M. Onghena, U. Kunkel, W. Schulz, A. Guillon, N. Noyon, G. Leroy, P. Bados, S. Bogialli, D. Stipanich, P. Rostkowski, J. Hollender, Non-target screening with high-resolution mass spectrometry: Critical review using a collaborative trial on water analysis, *Anal. Bioanal. Chem.* 407 (2015) 6237–6255. doi:10.1007/s00216-015-8681-7.
- [11] C.G. Bologa, M.M. Olah, T.I. Oprea, Chemical database preparation for compound acquisition or virtual screening., *Methods Mol. Biol.* 316 (2006) 375–88. doi:10.1385/1-59259-964-8:375.
- [12] J. Muncke, T. Backhaus, B. Geueke, M. V Maffini, O.V. Martin, J.P. Myers, A.M. Soto, L. Trasande, X. Trier, M. Scheringer, Scientific Challenges in the Risk Assessment of Food Contact Materials, *Environ. Health Perspect.* 125 (2017) 1–9. doi:10.1289/EHP644.
- [13] M. Ibáñez, J. V. Sancho, Ó.J. Pozo, W. Niessen, F. Hernández, Use of quadrupole time-of-flight mass spectrometry in the elucidation of unknown compounds present in environmental water, *Rapid Commun. Mass Spectrom.* 19 (2005) 169–178. doi:10.1002/rcm.1764.
- [14] T. Kind, O. Fiehn, Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm, *BMC Bioinformatics.* 7 (2006) 234. doi:10.1186/1471-2105-7-234.
- [15] C. Steinbeck, Recent developments in automated structure elucidation of natural products., *Nat. Prod. Rep.* 21 (2004) 512–518. doi:10.1039/b400678j.
- [16] C. Foti, K. Alsante, G. Cheng, T. Zelesky, M. Zell, Tools and workflow for structure elucidation of drug degradation products, *TrAC - Trends Anal. Chem.* 49 (2013) 89–99. doi:10.1016/j.trac.2013.06.005.
- [17] M. Elyashberg, K. Blinov, S. Molodtsov, Y. Smurnyy, A.J. Williams, T. Churanova, Computer-assisted methods for molecular structure elucidation: Realizing a spectroscopist's dream, *J. Cheminform.* 1 (2009) 1–26. doi:10.1186/1758-2946-1-3.
- [18] T. Kind, O. Fiehn, Advances in structure elucidation of small molecules using mass spectrometry, *Bioanal. Rev.* 2 (2010) 23–60. doi:10.1007/s12566-010-0015-9.
- [19] A.A. Bletsou, J. Jeon, J. Hollender, E. Archontaki, N.S. Thomaidis, Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment, *TrAC - Trends Anal. Chem.* 66 (2015) 32–44. doi:10.1016/j.trac.2014.11.009.
- [20] M. Krauss, H. Singer, J. Hollender, LC-high resolution MS in environmental analysis: From target screening to the identification of unknowns, *Anal. Bioanal. Chem.* 397 (2010) 943–951. doi:10.1007/s00216-010-3608-9.
- [21] T. Hoffmann, D. Krug, S. Hüttel, R. Müller, Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow, *Anal. Chem.* 86 (2014) 10780–10788. doi:10.1021/ac502805w.
- [22] J.F. García-Reyes, M.D. Hernando, A. Molina-Díaz, A.R. Fernández-Alba, Comprehensive screening of target, non-target and unknown pesticides in food by LC-TOF-MS, *TrAC Trends Anal. Chem.* 26 (2007) 828–841. doi:10.1016/j.trac.2007.06.006.
- [23] J. Muncke, Endocrine disrupting chemicals and other substances of concern in food contact materials: An updated review of exposure, effect and risk assessment, *J. Steroid Biochem. Mol. Biol.* 127 (2011) 118–127. doi:10.1016/j.jsbmb.2010.10.004.

- [24] E.N. Pieke, K. Granby, X. Trier, J. Smedsgaard, A framework to estimate concentrations of potentially unknown substances by semi-quantification in liquid chromatography electrospray ionization mass spectrometry, *Anal. Chim. Acta.* 975 (2017) 30–41. doi:10.1016/j.aca.2017.03.054.
- [25] L. Bengtström, A.K. Rosenmai, X. Trier, L.K. Jensen, K. Granby, A.M. Vinggaard, M. Driffield, J. Højslev Petersen, Non-targeted screening for contaminants in paper and board food-contact materials using effect-directed analysis and accurate mass spectrometry, *Food Addit. Contam. Part A.* 33 (2016) 1080–1093. doi:10.1080/19440049.2016.1184941.
- [26] A.K. Rosenmai, L. Bengtström, C. Taxvig, X. Trier, J.H. Petersen, T. Svingen, M.-L. Binderup, B.M.A. van Vugt-Lussenburg, M. Dybdahl, K. Granby, A.M. Vinggaard, An effect-directed strategy for characterizing emerging chemicals in food contact materials made from paper and board, Submitted. (2017). doi:10.1016/j.fct.2017.05.061.
- [27] M. Ibáñez, J. V. Sancho, F. Hernández, D. McMillan, R. Rao, Rapid non-target screening of organic pollutants in water by ultraperformance liquid chromatography coupled to time-of-flight mass spectrometry, *TrAC - Trends Anal. Chem.* 27 (2008) 481–489. doi:10.1016/j.trac.2008.03.007.
- [28] E.L. Schymanski, H.P. Singer, P. Longrée, M. Loos, M. Ruff, M.A. Stravs, C. Ripollés Vidal, J. Hollender, Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry., *Environ. Sci. Technol.* 48 (2014) 1811–8. doi:10.1021/es4044374.
- [29] L. Bengtström, Chemical identification of contaminants in paper and board food contact materials, Technical University of Denmark, 2014.
- [30] A.W. Hill, R.J. Mortishire-Smith, Automated assignment of high-resolution collisionally activated dissociation mass spectra using a systematic bond disconnection approach, *Rapid Commun. Mass Spectrom.* 19 (2005) 3111–3118. doi:10.1002/rcm.2177.
- [31] Agilent Technologies Inc., Agilent MassHunter Molecular Structure Correlator (MSC) Software, (2011).
- [32] K. Pivnenko, E. Eriksson, T.F. Astrup, Waste paper for recycling: Overview and identification of potentially critical substances, *Waste Manag.* 45 (2015) 134–142. doi:10.1016/j.wasman.2015.02.028.
- [33] U.S. Food and Drug Administration, List of indirect additives used in food contact substances, (2012). <https://www.accessdata.fda.gov/scripts/fdcc/?set=IndirectAdditives>.
- [34] European Chemicals Agency, Preparation of an inventory of substances suspected to meet REACH Annex III criteria, (2016). <https://echa.europa.eu/information-on-chemicals/annex-iii-inventory> (accessed April 25, 2017).
- [35] Agilent Technologies Inc., MassHunter Extractables and Leachables PCDL, (2015).
- [36] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B.A. Shoemaker, J. Wang, B. Yu, J. Zhang, S.H. Bryant, PubChem substance and compound databases, *Nucleic Acids Res.* 44 (2016) D1202–D1213. doi:10.1093/nar/gkv951.
- [37] M.C. Thomas, S.R. Dunn, J. Altwater, S.G. Dove, G.W. Nette, Rapid identification of long-chain polyunsaturated fatty acids in a marine extract by HPLC-MS using data-dependent acquisition., *Anal. Chem.* 84 (2012) 5976–83. doi:10.1021/ac3006523.
- [38] D. Schwudke, G. Liebisch, R. Herzog, G. Schmitz, A. Shevchenko, Shotgun lipidomics by tandem mass spectrometry under data-dependent acquisition control., *Methods Enzymol.* 433 (2007) 175–91. doi:10.1016/S0076-6879(07)33010-3.
- [39] L.C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis., *Mol. Cell. Proteomics.* 11 (2012) O111.016717. doi:10.1074/mcp.O111.016717.
- [40] J.D. Egerton, B. MacLean, R. Johnson, Y. Xuan, M.J. MacCoss, Multiplexed peptide analysis using data-independent acquisition and Skyline, *Nat. Protoc.* 10 (2015) 887–903. doi:10.1038/nprot.2015.055.
- [41] X. Zhù, Y. Chen, R. Subramanian, Comparison of information-dependent acquisition, SWATH, and MS(All) techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry., *Anal. Chem.* 86 (2014) 1202–9. doi:10.1021/ac403385y.
- [42] J.F. Xiao, B. Zhou, H.W. Resson, Metabolite identification and quantitation in LC-MS/MS-based metabolomics, *TrAC - Trends Anal. Chem.* 32 (2012) 1–14. doi:10.1016/j.trac.2011.08.009.
- [43] G. Hopfgartner, D. Tonoli, E. Varesio, High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices., *Anal. Bioanal. Chem.* 402 (2012) 2587–96.

doi:10.1007/s00216-011-5641-8.

- [44] N.A. Suci, F. Tiberto, S. Vasileiadis, L. Lamastra, M. Trevisan, Recycled paper-paperboard for food contact materials: Contaminants suspected and migration into foods and food simulant, *Food Chem.* 141 (2013) 4146–4151. doi:10.1016/j.foodchem.2013.07.014.
- [45] A. Ozaki, Y. Yamaguchi, T. Fujita, K. Kuroda, G. Endo, Chemical analysis and genotoxicological safety assessment of paper and paperboard used for food packaging, *Food Chem. Toxicol.* 42 (2004) 1323–1337. doi:10.1016/j.fct.2004.03.010.
- [46] K. Grob, M. Biedermann, E. Scherbaum, M. Roth, K. Rieger, Food Contamination with Organic Materials in Perspective: Packaging Materials as the Largest and Least Controlled Source? A View Focusing on the European Situation, *Crit. Rev. Food Sci. Nutr.* 46 (2006) 529–535. doi:10.1080/10408390500295490.
- [47] U.S. Environmental Protection Agency, 2-[2-(Dimethylamino)ethoxy]ethanol, Chem. Dashboard. (n.d.). <https://comptox.epa.gov/dashboard/DTXSID1027427> (accessed July 22, 2017).
- [48] C.P. Carpenter, C.S. Weil, H.F. Smyth, Range-finding toxicity data: list 8., *Toxicol. Appl. Pharmacol.* 28 (1974) 313–9. <http://www.ncbi.nlm.nih.gov/pubmed/4854023> (accessed June 22, 2017).
- [49] F.A. Patty, *Industrial hygiene and toxicology*. Vol. II. 2nd rev. ed. Frank A. Patty, Editor. John Wiley & Sons, Inc., 605 Third Ave., New York 16, N. Y., 1963. xxxi + 1546 pp. 15 1/2 × 24 cm. Price \$40.00, 1963. doi:10.1002/jps.2600520932.
- [50] P.K. Working, M.S. Newman, J. Johnson, J.B. Cornacoff, Safety of poly(ethylene glycol) and poly(ethylene glycol) derivatives, *ACS Symp. Ser.* 680 (1997) 45–57. doi:10.1021/bk-1997-0680.ch004.
- [51] H.F. Smyth, C.P. Carpenter, C.S. Weil, The Toxicology of the Polyethylene Glycols*, *J. Am. Pharm. Assoc. (Scientific Ed.)*. 39 (1950) 349–354. doi:10.1002/jps.3030390615.
- [52] Federal Register, Food additives permitted for direct addition to food for human consumption, 21 C.F.R. § 172.130, United States of America, 1988.
- [53] Federal Register, Indirect Food Additives: Adhesives and Components of Coatings. Subpart B - Substances for Use Only as Components of Adhesives, 21 C.F.R. § 175.105, United States of America, 2017.
- [54] European Commission, 2017.0798: unauthorised food additive E 265 - dehydroacetic acid (presence /kg) in external coating for cheeses from Spain, 2017. https://webgate.ec.europa.eu/rasff-window/portal/index.cfm?event=notificationDetail&NOTIF_REFERENCE=2017.0798.
- [55] M.L. Binderup, G. a Pedersen, a M. Vinggaard, E.S. Rasmussen, H. Rosenquist, T. Cederberg, Toxicity testing and chemical analyses of recycled fibre-based paper for food contact., *Food Addit. Contam.* 19 Suppl (2002) 13–28. doi:10.1080/02652030110089878.
- [56] U. Heudorf, V. Mersch-Sundermann, J. Angerer, Phthalates: toxicology and exposure., *Int. J. Hyg. Environ. Health.* 210 (2007) 623–34. doi:10.1016/j.ijheh.2007.07.011.
- [57] J. Muncke, Phthalates, *Food Packag. Forum.* (2012). <http://www.foodpackagingforum.org/food-packaging-health/phthalates> (accessed February 7, 2017).
- [58] K. Yazar, M.D. Lundov, A. Faurshou, M. Matura, A. Boman, J.D. Johansen, C. Lidén, Methylisothiazolinone in rinse-off products causes allergic contact dermatitis: a repeated open-application study., *Br. J. Dermatol.* 173 (2015) 115–22. doi:10.1111/bjd.13751.
- [59] M. Isaksson, I. Hauksson, M. Hindsén, A. Pontén, C. Svedman, M. Bruze, Methylisothiazolinone contact allergy is rising to alarming heights also in southern Sweden., *Acta Derm. Venereol.* 95 (2015) 31–4. doi:10.2340/00015555-1844.
- [60] T.H. Connor, P.G. Tee, M. Afshar, K.M. Connor, Mutagenicity of cosmetic products containing Kathon®, *Environ. Mol. Mutagen.* 28 (1996) 127–132. doi:10.1002/(SICI)1098-2280(1996)28:2<127::AID-EM9>3.0.CO;2-C.

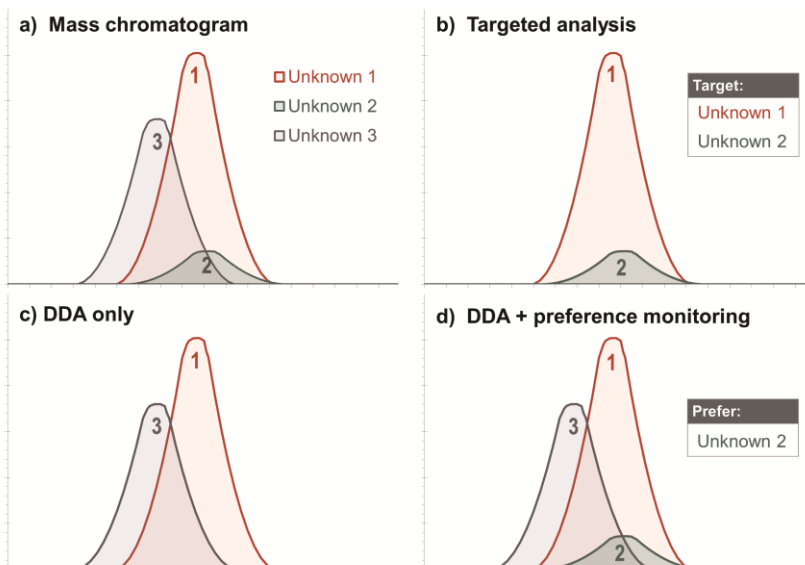


Figure 1. Mass chromatogram ion traces demonstrating preselection and active selection acquisition. a) Full mass chromatogram illustrating three different overlapping peaks of varying abundance. b) Analysis via targeted methods exclusively fragments the preselected precursor ions (1 and 2), but excludes all ions not listed (3). c) Analysis via DDA results in product-ion spectra via dynamic precursor selection, but excludes low-abundance ions (2) by preferentially selecting high-abundance ions (1 and 3) in case there are large numbers of ions. d) Analysis via DDA augmented with preference monitoring permits dynamic precursor selection without sacrificing low-abundance ions, provided these are included in the preference list.

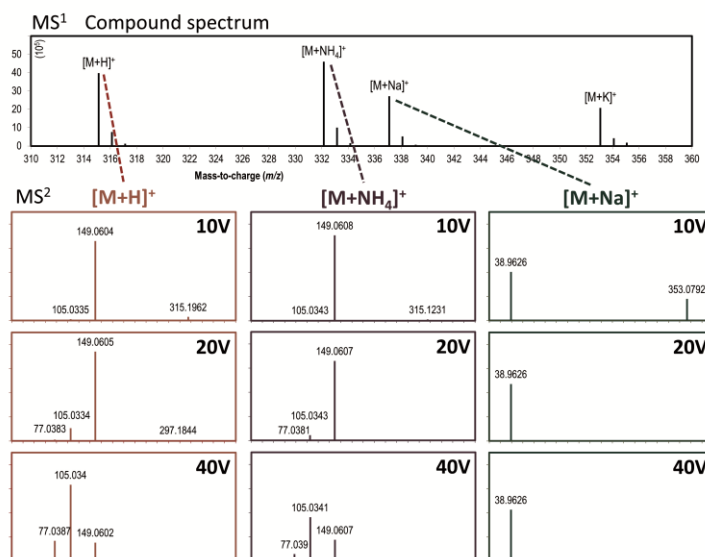


Figure 2. The fragmentation spectrum of an unknown compound demonstrates the adverse effect of alkaline ion adducts on spectral quality. The original mass spectrum is shown at the top for phthalate-like compound with four different precursor adducts. In MS/MS, fragmentation of the molecular adducts $[M+H]^+$ and $[M+NH_4]^+$ resulted in high quality spectra, containing numerous fragments of varying intensity depending on the energy applied. Fragmentation of molecular adducts $[M+Na]^+$ and $[M+K]^+$ (data not shown) resulted in spectra with little fragmentation unchanged by varying collision energies, thereby containing little structural information on the compound.

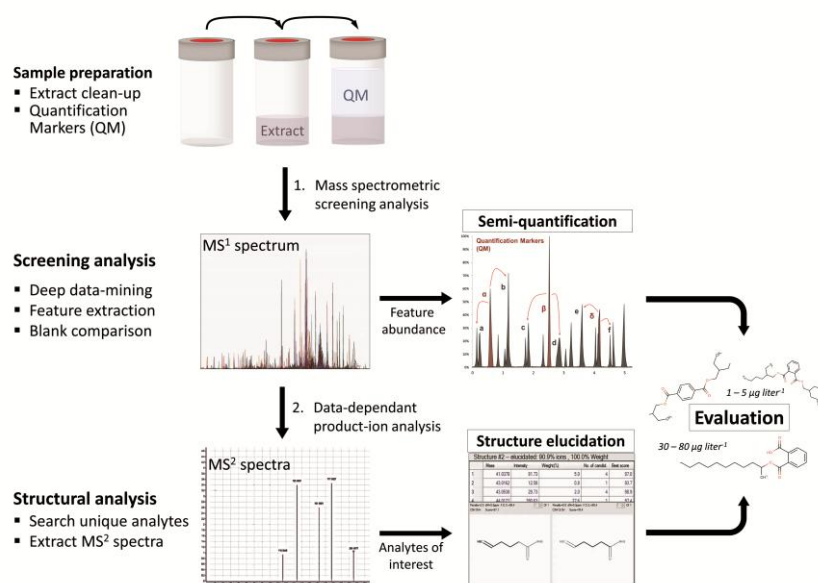


Figure 3. Workflow for structure elucidation and for semi-quantification. First, sample cleanup is required before most LC-MS analysis. During this step, the quantification markers (QM) are added to all samples. Semi-quantification data and preference lists are retrieved via screening analysis. Structural prediction is done by structure correlations on product-ion spectra of listed and detected analyte in the sample. The semi-quantification and structural prediction are combined into an evaluation of the substance.

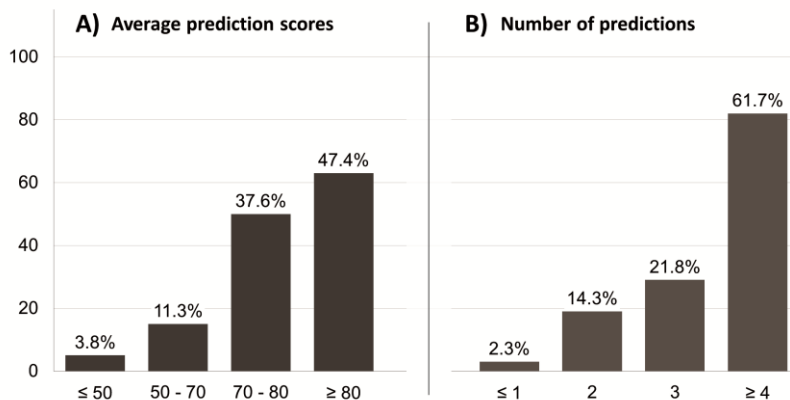


Figure 4. Graph indicating the distribution of prediction scores and number of effective predictions. Data is shown for sample A1, a recycled pizza box with total number of compounds $n = 133$. A) Distribution of prediction matching scores. The scores were divided into four bins centered on the acceptance threshold of score ≥ 70 . B) Distribution of the number of viable predictions from various databases (max. 5). At minimum two predictions were needed to consider the elucidation as adequate for further work, where one prediction must have originated from a high-specificity database.

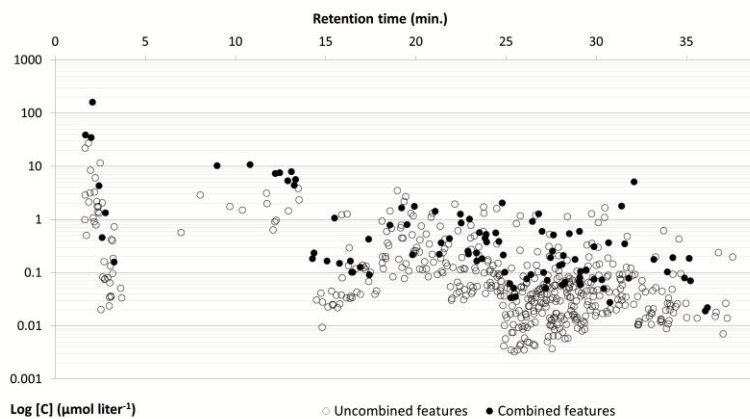


Figure 5. Retention time and semi-quantified concentration of 585 chromatographic peaks in a recycled board sample. Peaks include those also found in the blank. Of these 585 peaks, 117 peaks were linked to elucidation data in subsequent MS/MS experiments, which are marked as filled points. The remaining unlinked peaks (blank points) either were considered irrelevant in the MS/MS experiments or did not have suitable fragmentation spectra recorded.

Table 1. Optimized settings employed in the Electrospray Ionization module. Settings are given for both positive mode electrospray (ESI+) and negative mode electrospray (ESI-).

Setting	Unit	ESI+	ESI-
Nozzle Voltage	Volt (V)	1000	1000
Sheath Gas Temperature	°C	350	350
Sheath Gas Flow	L min ⁻¹	12	10
Capillary Voltage	Volt (V)	2000	2000
Nebulizer Pressure	psi	25	40
Drying Gas Temperature	°C	150	150
Drying Gas Flow	L min ⁻¹	14	11

Accepted Article

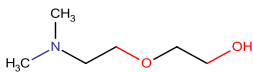
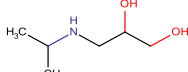
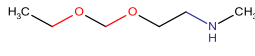
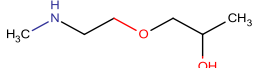
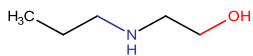
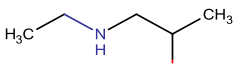
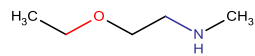
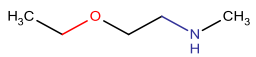
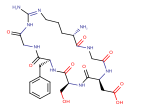
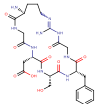
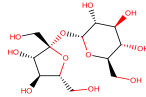
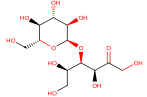
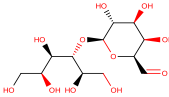
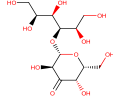
Table 2. Qualitative assessment of databases: scope, specificity, and authenticity. Specificity (spec.) is the relevance of the database to the field used; scope is based on the uniqueness and size of the database. Authenticity (auth.) is the overall reliability of predictions using this database. Database properties were qualitative marked; ++: very good, +: good, ±: moderate, -: low, --: very low.





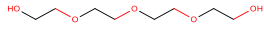
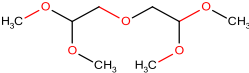
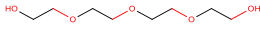
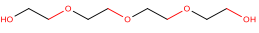
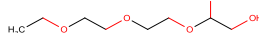
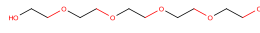
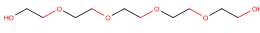
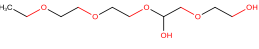
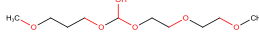
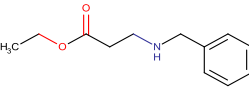
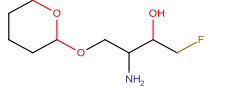
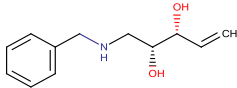

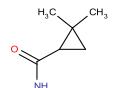

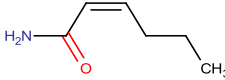
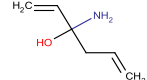
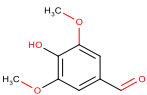
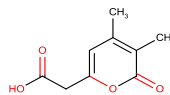

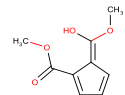

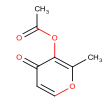


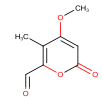
Database	Number	Locality	# Entries	Spec.	Scope	Auth.
Paper & Board Database (PBDB)	#1 - #3	Local	4,353	++	+	++
REACH	#4	Local	55,648	+	+	+
Agilent Extractables & Leachables (AgEL)	#5	Local	1,840	±	-	±
ChemSpider (CS)	#6	Online	> 58 million [†]	--	++	--
PubChem (PC)	#7	Online	> 91 million [†]	--	++	--

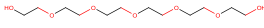
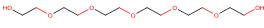
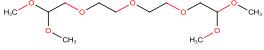

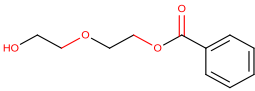
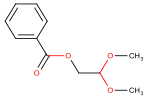

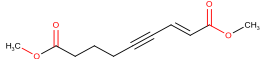

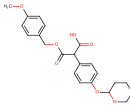
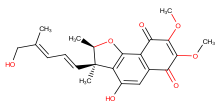
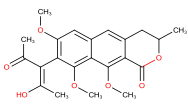



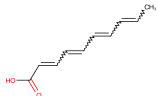
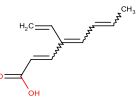
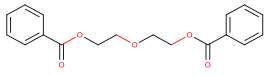
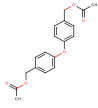
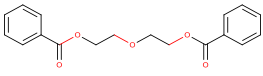
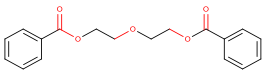
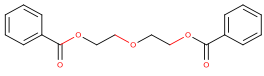
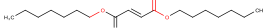
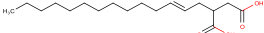
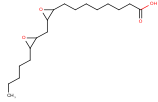
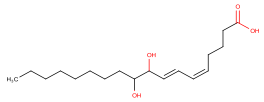
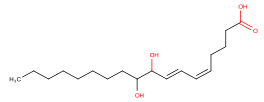
[†] Online as of June, 2017

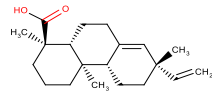
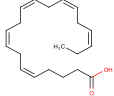
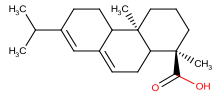
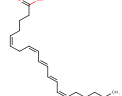
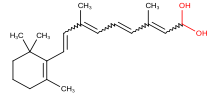
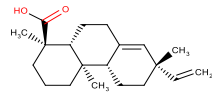
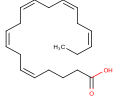
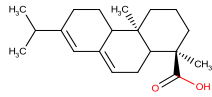
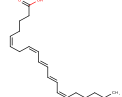
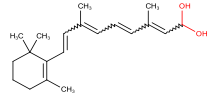
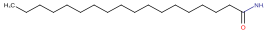
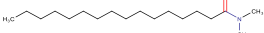
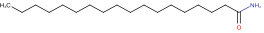
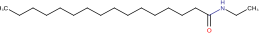
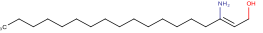
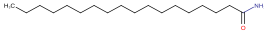
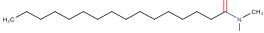
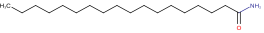
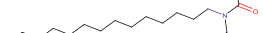

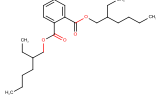
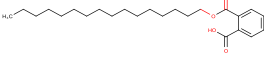
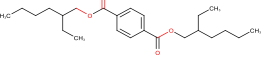
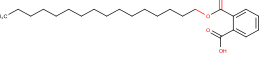
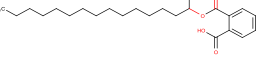
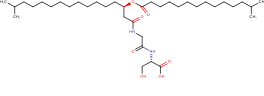
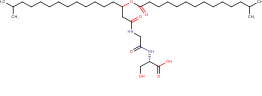
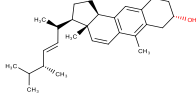
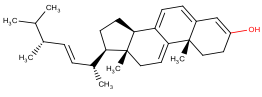
Accepted Article

Table 3. Structural identifications via MSC combined with semi-quantitative data for a selection of compounds in a board sample. The top 20 highest concentrations and 5 lowest concentrations are included in the table. For each compound, its serial number, retention time (RT), accurate mass, and estimated concentration are reported. In addition, structure predictions for each database are given along with the prediction score (scale 0 to 100). Entries with no suitable prediction are marked with n.d.. Entries discussed in the text are marked with an asterisk (*).

C.	RT (min)	Mass (amu)	Level ($\mu\text{g dm}^{-2}$)	Level (μM)	PBDB (score)	REACH (score)	Agilent E&L (score)	ChemSpider (score)	PubChem (score)
1	1.696	344.8755	1336	38.7	n.d.	n.d.	n.d.	n.d.	n.d.
2*	2.003	134.1177	460	34.3	 (93.2)	 (93.7)	n.d.	 (93.8)	 (93.8)
3*	2.091	104.1074	1653	159	 (93.7)	 (93.7)	n.d.	 (93.9)	 (93.9)
4	2.247	620.2773	265	4.28	n.d.	n.d.	n.d.	 (77.7)	 (77.7)
6*	2.799	360.1506	48	1.33	 (89.1)	 (89.5)	n.d.	 (91.5)	 (91.7)

C.	RT (min)	Mass (amu)	Level ($\mu\text{g dm}^{-2}$)	Level (μM)	PBDB (score)	REACH (score)	Agilent E&L (score)	ChemSpider (score)	PubChem (score)
9*	8.989	116.016	118	10.1	 (81.5)	 (81.5)	n.d.	 (90.2)	 (87.8)
10*	10.831	195.1231	207	10.6	 (94.3)	 (91.4)	 (94.3)	 (94.0)	 (94.0)
11*	12.218	239.1496	173	7.24	n.d.	 (93.8)	 (93.8)	 (94.2)	 (94.2)
13*	12.524	208.1334	156	7.48	n.d.	 (79.7)	n.d.	 (74.1)	 (81.9)
14	12.923	114.0916	60	5.29	 (66.2)	 (80.3)	 (66.2)	 (80.8)	 (81.6)
15	13.105	183.0655	143	7.82	 (57.3)	 (75.6)	n.d.	 (80.5)	 (80.5)
16*	13.27	169.0497	74	4.37	 (57.3)	 (75.6)	 (66.2)	 (80.5)	 (80.5)

C.	RT (min)	Mass (amu)	Level ($\mu\text{g dm}^{-2}$)	Level (μM)	PBDB (score)	REACH (score)	Agilent E&L (score)	ChemSpider (score)	PubChem (score)
					(77.2)	(77.6)	(77.2)	(79.5)	(79.4)
18*	13.347	300.2018	169	5.61	 (88.1)	 (88.1)	n.d.	 (88.3)	 (88.3)
34	18.986	211.0971	35	1.63	 (86.7)	 (86.9)	 (80.3)	 (87.6)	 (88.2)
36	19.375	401.1598	70	1.74	n.d.	 (82.5)	n.d.	 (87.1)	 (88.1)
39	21.088	165.0914	23	1.41	 (70.4)	 (74.6)	 (72.2)	 (75.2)	 (75.4)
61	24.822	332.1507	67	2.03	 (75.3)	 (66.9)	 (75.0)	 (75.0)	 (75.0)
65	25.283	330.2643	1.1	0.03	 (27.2)	 (42.9)	 (39.0)	 (44.2)	 (44.2)

C.	RT (min)	Mass (amu)	Level ($\mu\text{g dm}^{-2}$)	Level (μM)	PBDB (score)	REACH (score)	Agilent E&L (score)	ChemSpider (score)	PubChem (score)
109	30.386	303.2326	0.8	0.03	 (75.5)	 (79.9)	 (75.5)	 (79.4)	 (79.6)
110	30.521	303.2331	0.8	0.03	 (85.8)	 (91.0)	 (85.7)	 (91.5)	 (91.8)
115*	31.474	284.2956	51	1.78	 (93.6)	 (93.7)	 (93.6)	 (93.7)	 (93.7)
116*	30.662	284.295	51	1.78	 (80.8)	 (81.0)	 (80.8)	 (80.4)	 (80.8)
118*	32.171	391.2857	198	5.07	 (81.6)	 (81.6)	 (81.6)	 (81.5)	 (81.5)
133	36.2	655.5247	1.4	0.02	n.d.	n.d.	n.d.	 (85.0)	 (85.0)
132	36.131	393.3159	0.7	0.02	n.d.	n.d.	n.d.		

C.	RT (min)	Mass (amu)	Level ($\mu\text{g dm}^{-2}$)	Level (μM)	PBDB (score)	REACH (score)	Agilent E&L (score)	ChemSpider (score)	PubChem (score)
								(88.1)	(88.6)