HBM4EU science and policy for a healthy future

HORIZON2020 Programme
Contract No. 733032 HBM4EU

Annotation framework

AD16.4

WP16 - Emerging Substances

Deadline: December 2018

Upload by Coordinator: 20 December 2018

Entity	Name of person responsible	Short name of institution	Received [Date]
Coordinator	Marike Kolossa- Gehring	UBA	10/12/2018
Grant Signatory	Robert Barouki	INSERM	07/12/2018
Pillar Leader	Robert Barouki	INSERM	10/12/2018
Work Package Leader	Jean-Philippe Antignac	INRA	07/12/2018
Task leader	Adrian Covaci	UA	07/12/2018

Responsible author	Herbert Oberacher	E-mail	herbert.oberacher@i-med.ac.at
Short name of institution	MUI	Phone	+43-512-9003-70639
Co-authors	Adrian Covaci (UA), Noelia Caballero-Casero (UA), Jean-Philippe Antignac (INRA), Laurent Debrauwer (INRA), Martin Krauss (UFZ), Tobias Schulze (UFZ), Marja Lamoree (VU&EH), Annelaure Damont (CEA), François Fenaille (CEA)		

Table of contents

1 Authors and acknowledgements

Lead authors

Herbert Oberacher, MUI
Adrian Covaci, UAntwerpen

Contributors

Noelia Caballero-Casero, UAntwerpen
Jean-Philippe Antignac. INRA
Laurent Debrauwer, INRA
Martin Krauss, UFZ
Tobias Schulze, UFZ
Marja Lamoree, VU
Annelaure Damont (CEA)
François Fenaille (CEA)

2 Glossary

Annotation/identification: Capability to assign a signal detected by non-targeted or suspect screening (i.e. a spectrometric descriptor) to a chemical with a given confidence level, by means of a reference library and/or structural elucidation work. Annotation is the act of linking a detected mass spectrometric feature with a chemical. Identification is the act of proving to be the same.

Non-targeted LC-MS: Analytical process for gathering comprehensive information on the composition of a sample. Workflows involve different steps of sample collection, sample preparation, data acquisition and data mining. The fraction of compounds accessible by a certain workflow depends on the characteristics of the individual steps applied. Data-dependent or data-independent acquisition techniques are employed for data acquisition. Detected features are characterised by retention time, MS, and, where possible, MS/MS information to enable annotation.

Targeted LC-MS: Analytical process for gathering specific information on the composition of a sample. Workflows involve different steps of sample collection, sample preparation, data acquisition and data mining. The steps were optimised for a preselected number of molecules. Often selected reaction monitoring techniques are employed for data acquisition. Furthermore, target screening usually involves a reference standard measured in-house under the same analytical conditions such that retention time, MS, and, where possible, MS/MS information is available for identification and confirmation.

Tandem mass spectrometry: Tandem mass spectrometry, also known as MS/MS or MS^2, involves multiple steps of mass spectrometry selection, with some form of fragmentation occurring in between the stages. Multiple stages of mass analysis separation can be accomplished with individual mass spectrometer elements separated in space or using a single mass spectrometer with the MS steps separated in time.

Target: A compound that is expected to be included in a sample and of which full mass spectrometric reference data, including MS/MS fragmentation data, is available to enable annotation. The reference data is usually acquired with certified reference standards, and is stored in tandem mass spectral databases. The mass spectrometric data is often accompanied by metadata.

Suspect: A compound that is expected to be included in a sample. Typically, the available mass spectrometric data is incomplete and does not allow unequivocal annotation. Often, information on MS/MS fragmentation and retention time is missing or has only been predicted with computational tools.

Known: A detected signal that was annotated to a suspect or target at a certain confidence level.

Unknown: An unannotated signal or feature.

Identification level: An approach for communicating identification confidence. A commonly used classification system was proposed by Schymanski [2]. It includes five levels: exact mass – unequivocal molecular formula – tentative candidate – probable structure – confirmed structure.

Tandem mass spectral database: An organised collection of tandem mass spectral data which comes bundled with a management system. The database management system is a software application that interacts with the user, other applications, and the database itself to capture and analyse data. Tandem mass spectra are typically acquired from certified reference compounds. Spectral information is processed prior to storage in a library.

Tandem mass spectral library: A curated and annotated collection of mass spectra acquired from certified reference compounds. Curation efforts may include manual inspection of mass spectra by mass spectrometry experts, noise and artifact removal, recalibration of spectra and peak annotations, as well as inter-library comparisons. The mass spectrometric data is often accompanied by metadata.

3 Introduction

3.1 Detection, annotation, and identification

The goal of suspect and non-targeted analysis is providing extensive qualitative information on the chemical composition of a sample. The emergence of such approaches is supported by currently available technologies and instrumentation that can generate large sets of chemical information from a low sample amount. Mass spectrometry (especially at high resolution, HRMS) appears as one of these cutting edge technologies for large scale and high throughput profiling. However, the process to assign chemical identities from a set of MS signals is not trivial and requires especially a strongly consolidated QA/QC procedure. Particularly the confidence level associated with this annotation process appears as a crucial issue. Since the presence of as many as possible compounds should be unambiguously confirmed, adequate strategies have to be used, especially in the context of regulatory use of the obtained results (support to policy), but also for research. While QA/QC aspects are well established in the field of targeted methods, these are currently less developed for non-targeted analyses, even if some significant work has to be acknowledged and exploited from the metabolomics and environmental communities. In the frame of the HBM4EU WP16, special attention was paid to a consolidated annotation, also as compared to already existing initiatives or approaches that still have gaps with regard to confidence levels associated to biomarker annotation.

Detection can be seen as the collection of compound-specific data by instrumental analysis. For chromatography coupled to mass spectrometry, collected data may include retention times, *m/z*-values of molecular ions, adducts and possibly fragment ions as well as relative abundances of fragment ions and isotopologues.

Annotation is defined as the act of linking a detected feature with a chemical, taking into account the above mentioned chromatographic and spectrometric characteristics, by comparison with the same characteristics originating from an authentic standard.

Identification is the process of proving that a compound is the very same that it is alleged or reputed to be. While the identity is collected from a number of indications, the identification is accomplished by comparing measured data sets. One set of features is obtained from the analysis of an unknown compound; the other one from a reference standard of known identity. In this context, a challenging task is the definition of objective metrics of identity. Proper settings may reduce the number of false positives, as well as false negative identifications to a minimum, ideally to zero.

To enable definitive identification, the availability of a sufficient amount of analytical data is mandatory. Like a fingerprint, these data should act as a unique identifier excluding all other chemical entities from being the compound analysed. In the case of a MS based characterisation, such a chemical fingerprint can partly be created *in silico* (e.g., *m/z*-values of molecular ions, relative abundances of fragment ions and isotopologues on the basis of a given elemental composition) and/or by analysing reference standards (e.g. *m/z*-values of fragment ions). Fingerprints are often stored in databases.

From an analytical point of view and depending on the available information for annotation, chemicals can be divided into two categories (see Figure 1):

"Suspects" are known compounds that are expected to be present in a sample. Nevertheless, due to the availability of an incomplete set of mass spectrometric data, unequivocal annotation is not possible. Typically, information on MS/MS fragmentation and retention time is missing or has only been predicted with computational tools.

"Targets" are compounds that are expected to be included in a sample and of which full mass spectrometric reference data, including MS/MS fragmentation, is available for annotation. The reference data is usually acquired with certified reference standards, and is stored in tandem mass spectral databases.

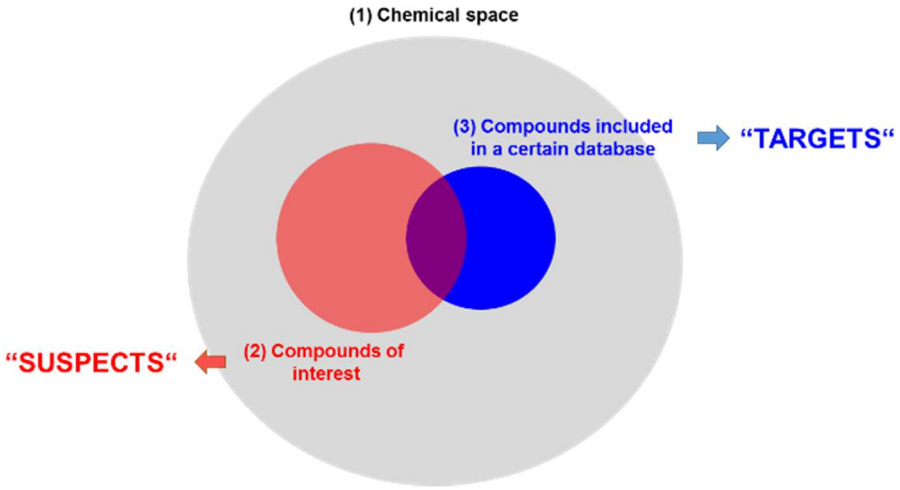

Figure 1: Difference between targets and suspects

Both, suspects and targets, represent subsets of the entire chemical space, defined as chemical domain of the entire method used for sampling, sample preparation, separation and acquisition (Figure 1). Suspects can be converted into targets by collecting comprehensive mass spectrometric reference data that enables unequivocal identification of this compound. This step usually requires the availability of reference standard compounds.

Non-targeted LC-MS or GC-MS techniques produce a large amount of detected mass spectrometric features. These features are characterised by retention time, *m/z*-values of molecular ions and/or adducts, *m/z*-values of fragment ions, and/or relative abundances of fragment ions and isotopologues. Features that are annotated to a certain compound are called "knowns". The remaining unidentified features are "unknowns".

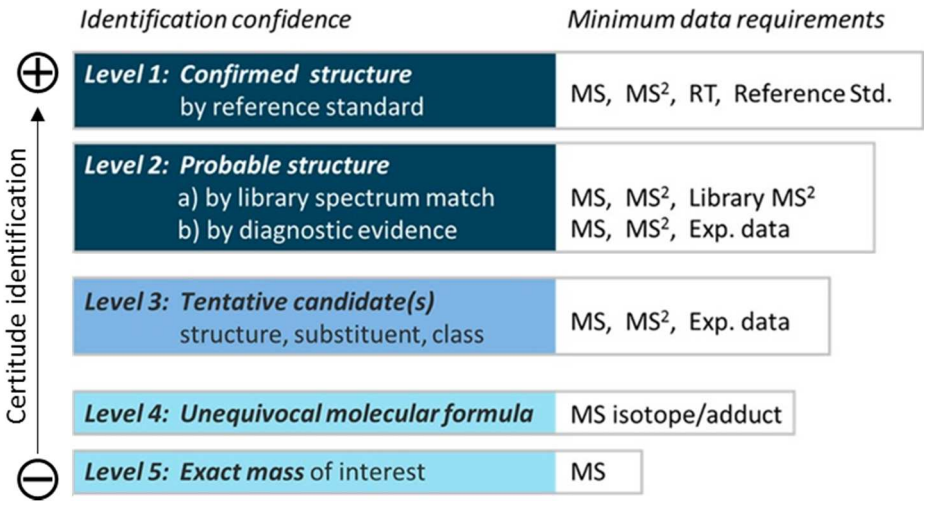

Figure 2: Identification confidence levels proposed by Schymanski et al. [1]

From the classification initially proposed by Sumner for metabolomics [2], a more refined classification system was proposed by Schymanski [1]. “Knowns” were divided into two subclasses. A “confirmed structure” (level 1) has been verified via the appropriate measurement of a reference standard with MS, MS/MS and retention time matching. If possible, an orthogonal method should also be used. A “probable structure” (level 2) is obtained by unambiguous matching literature or library data. “Unknowns” are referred to as level 4 or level 5 identifications. “Tentative candidates” (level 3) are somewhere between “knowns” and “unknowns”. Typically, the available data provide evidence for possible or likely structure(s), but insufficient information for one exact structure only (e.g., positional isomers).

This reference classification (Figure 2) is proposed to be a basis of our annotation framework within HBM4EU WP16 to accurately document the level of confidence associated to the measurement of each reported putative or identified marker from any non-targeted screening of human specimen.

3.2 Current status of tandem mass spectral databases

Tandem mass spectral databases are indispensable tools for compound annotation in non-targeted MS workflows based on soft-ionisation mass spectrometry (typically LC-ESI-MS). Several reviews are available, in which progress in development and application of tandem mass spectral databases has been described [3-9]. Such a database represents an organised collection of tandem mass spectral data which comes bundled with management systems. The database management system is a software application with which the user, other applications, and the database itself interact to capture and analyse data.

MS/MS databases are acquired by analysis of reference standards. Since a fragmentation spectrum can look different depending on the excitation process (e.g., resonant vs. non resonant) as well as the collision energy applied to the parent ion, state-of-the-art databases include sets of compound-specific spectra that were acquired by applying different collisions energy settings as well as different instruments. Usually, obtained spectral information is processed prior to storage in a library. Curation efforts may include manual inspection of mass spectra by experienced mass spectrometry analysts, noise and artefact removal, recalibration of spectra and peak annotations, as well as inter-library comparisons.

Currently, liable databases contain MS/MS spectra acquired on diverse instruments with multiple collision energies. According to definitions setup by the instrument manufacturers, fragmentation is typically accomplished by collision-induced dissociation (CID) or so-called higher-energy collisional dissociation (HCD). Experimental data is often complemented with *in silico* generated spectra.

The overlap of compounds with MS/MS spectra from authentic reference standards in most public and commercial databases has been evaluated by Vinaixa et al. [6]. A total of 27,622 unique compounds were present across all databases. Among the four open databases (HMDB, MassBank, GNPS, and ReSpect) totaling 7127 compounds, only 18 compounds (<1%) have some types of spectral data in all databases. When comparing all combined open databases versus four commercial ones, only 225 compounds out of 27,622 (<1%) have some types of spectral data in all databases. Typically, the ratio of compounds in each database with any type of spectral data in two or more databases is >50%, with the exception of Agilent METLIN, and GNPS, which only overlap with other databases in approximately 35% of their compounds. Obviously, there is a relatively low overlap of compounds among existing spectral databases due to the specific scientific background underlying each library, which explains and justifies why most scientists currently use multiple databases. Some of those libraries contain mass spectral records of other databases. For example, MassBank records are also included in HMDB, MassBank of North America and GNPS.

The SPectraL hASH (SPLASH) initiative was developed in order to identify similar mass spectra across libraries (Wohlgemuth et al., 2016). The SPLASH is available in all freely accessible mass spectral libraries including mzCloud and will be established in a future release of the NIST library.

Conceptually, the premise of spectral library searching is very simple: the fragmentation pattern of a molecule under some fixed conditions is a reproducible fingerprint of that molecule, such that unknown spectra acquired under similar conditions can be identified by spectral matching [10].

Automated spectral library searching involves software with search algorithms tailor-made for tandem mass spectral databases [11-14]. The obtained search score after a database search represents the likelihood of a searched spectrum to correspond to a reference spectrum in a mass spectral reference database. A low score indicates that the compounds' fragmentation pattern has low similarity to any stored reference spectrum. A high score indicates significant spectral overlap and, consequently, structural similarity or even identity between the analyte and the reference compound. Library search should be sensitive and specific, producing as low as possible false negative and false positive results. Ideally, the obtained scores can be used to sort out false positive matches from correct positive matches [3]. If a comparison is made with historical targeted methods applied in a regulatory context (e.g., food safety), the primary objective of a screening method is to limit the risk of false negatives (e.g., 1%) and to keep an acceptable risk of false positives (e.g., 5%). The latter should be further reduced by confirmatory analyses using reference standard compounds. QA/QC consolidated non-targeted methods thus have to consider and document these issues during development and evaluation of method performances, and also during reporting of results.

There is also a vital discussion about robustness and transferability of tandem mass spectral libraries. For a long time, the predominant opinion was that due to the limited reproducibility of tandem mass spectra, libraries will only be useful on the same model of instrument used to acquire reference spectra. This situation has changed thanks to progress in instrument technologies and informatics tools. Databases combining advanced library designs with tailor-made search algorithms have been shown to enable reliable compound identification with spectra acquired in different labs with various instruments and different instrument settings [3]. From a previous situation where a pre-acquisition harmonisation of analytical procedures was researched which faced a number of severe difficulties for laboratories [15], the current trend is more to look for a post-acquisition flexibility of the MS reference library and associated matching algorithms to compose (as much as possible) with the diversity of imported experimental data without sacrificing the ambitioned confidence level in terms of correct annotation.

Over the last ten years, we have seen substantial progress in the quality of tandem mass spectral databases. Acquisition of reference spectra is today accomplished regularly on high-resolution instrumentation (i.e., QqTOF, Orbitrap) employing multiple collision energies for fragmentation to comprehensively cover the breakdown curves of reference compounds. Besides the protonated and deprotonated molecular ions, adduct ions, in-source fragments as well as isotopologues are commonly selected as precursor ions. Furthermore, to improve spectral quality, only curated spectra are stored in databases which come bundled with improved search algorithms. However, some of these semi-automated tools remains somewhat obscure in term of precise algorithms and parametrisation. Overall, we are getting closer to achieve the ambition of having available a universal tandem mass spectral database. However, this ambition is requiring to define and implement some common procedures to ensure the reliability and robustness of the generated data both stored in the desired MS reference library and generated from each experimental sample to be characterised.

In conclusion, a number of reference MS and MS/MS databases exist that have to be considered in the frame of our HBM4EU WP16 work plan in terms of structuration and/or content. These should serve as a basis to avoid unnecessary time spent to re-implement existing and reliable elements, and to ensure a coherence of the ambitioned outputs with potential established standards. However, an obvious lack of high level QA/QC consolidation appears within most of these existing databases (e.g., percentage of erroneous information, insufficient or non-adequately documented confidence level associated to the compound identification), together with some necessary adjustments for the specific HBM application field (e.g., human metabolites of contaminants are not well represented as compared to parent compounds).

4 Objective

The aim of the present deliverable is to describe a harmonised methodology for collecting and generating the necessary reference MS data to be used for annotating non-targeted chemical profiles generated from human matrices in the scope of identifying exposure markers.

5 Annotation framework

5.1 Starting point: “List of Known Emerging Chemicals”

Various initiatives and/or sources of information are available that have proposed lists of emerging compounds of interest in various contexts, mainly in the field of environment (e.g., the NORMAN suspect list exchange, https://www.norman-network.com/?q=node/236 and https://www.norman-network.com/nds), the US EPA Chemistry Dashboard (https://comptox.epa.gov/dashboard (Williams et al., 2017) but also through toxicological databases (e.g., TEXD database). These sources are overlapping to some extent, but also provide complementary information. In that context, one action of HBM4EU WP16 was to provide an overview of the existing lists or databases related to emerging chemicals at an international level. An aggregation of these different existing public inventories has been done, leading to more than 70,000 indexed compounds.

Then a first level of curation of this list was performed to clean and consolidate this database, including the consolidation of unambiguous and unique compound identifiers (i.e., simplified molecular-input line entry specification (SMILES) and International Chemical Identifier (InChI) or InChIKey, its hashed format, and MS-relevant information (e.g., elemental composition, target m/z signals). Exposure/source related information is also included as additional criterion for prioritisation. At the moment a second level of curation is on-going, in order (1) to discard strictly environmental markers irrelevant for screening in human matrices, (2) to conversely introduce main biotransformation products as relevant target markers in HBM matrices, and (3) to introduce toxicologically related information/criteria as additional prioritisation keys. This consolidated list will be used as a guiding rail for the suspect screening activities conducted in WP16.

The long-term goal of WP16 in terms of screening capabilities (i.e., availability of appropriate large scale screening methods with corresponding necessary MS libraries for annotation) would be to cover progressively a large part of this list. Finally, the compounds included in this “List of Known Emerging Chemicals” (cf. D16.1 HBM4EU_EmergScreenDB) are considered to represent suspects. From that point, the ambitious goal of WP 16 is to convert as many suspects as possible into targets. This goal can only be achieved by collecting comprehensive mass spectrometric reference data, including MS/MS fragmentation data to enable reliable and unequivocal identification of compounds.

The strategy for accomplishing the envisioned conversion involves (1) customisation of already available tandem mass spectral data and (2) acquisition of new reference spectra (Figure 3).

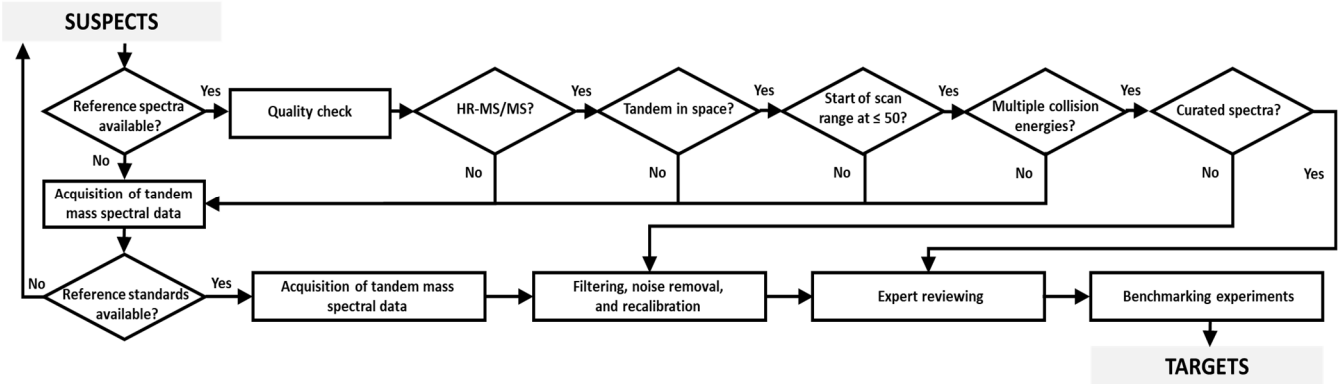

Figure 3: Outline of the proposed strategy to convert suspects into targets.

As outlined above, we are expecting that only for a limited number of suspects tandem mass spectral fragmentation information will already be available in databases. For the majority of compounds, this data has to be newly generated. Acquisition of tandem mass spectral data requires the physical availability of reference substances and sufficient experimental capacities for acquiring fragment ion mass spectra. Even by joining forces, the acquisition of reference data for ten thousands of compounds is a multi-annual project requiring significant resources. To provide fragmentation information in the near term, tandem mass spectral data will be produced by computational tools [16]. *In silico* fragmentation tools have made a remarkable progress over the last year and have been shown to enable prediction or explanation of experimental fragmentation reactions. Nevertheless, *in silico* databases are not considered to represent a full substitute for experimental data. Despite considerable advances in improving the quality and reliability of computationally generated fragment ion mass spectra, matches to *in silico* databases might produce a higher portion of false positive and false negative results than matches to experimental databases usually do [17,18]. Finally, both approaches (experimental and computational) have to be considered as complementary for a pragmatically balanced work plan, but again with a special emphasis on appropriate QA/QC measures to ensure reliable results. This work plan should be co-ordinated with the NORMAN Association and the Metabolomics Society in order to avoid double efforts and to maximise the outcomes.

5.2 Consolidation of available tandem mass spectral information

Tandem mass spectral databases accessible to the members of the HMB4EU consortium (e.g., MassBank, “The Wiley Registry of Tandem Mass Spectral Data” (WRTMD), in-house collections) will be surveyed to identify already available reference spectra. These reference spectra will be submitted to a quality check. The quality check is a two-step process.

Firstly, the data should pass the following quality criteria:

(1) The data must have been acquired on high-resolution instrumentation (e.g. QqTOF, Orbitrap). Such instruments typically provide a minimum resolution of 5000 in MS/MS mode, and the *m/z* error is usually lower than 10 ppm.

(2) Tandem-in-space techniques are primarily used by HBM4EU WP16 partners. To ensure a high degree of spectral similarity, fragmentation should have been accomplished either by Orbitrap-HCD or QqTOF-CID.

(3) Different ionisation techniques can be used (ESI, APCI or APPI), and positive and negative ionisation modes are considered.

(4) Ideally, information on the lower limit of the applied scan range should be available. For many compounds, low *m/z* fragment ions would hardly be needed to ensure unequivocal identification. Nevertheless, as higher starting values might lead to truncated spectral information, which could have a negative impact on spectral similarity and match scores, the scan range should preferentially start at ≤ *m/z* 50.

(5) Compound-specific breakdown curves should be covered by spectra acquired at several multiple collision energies (minimum of 3 values) over a meaningful energy range (e.g., 5-60 eV).

(6) Spectra should have been curated, which includes multiple steps of centroiding, filtering, noise removal, and recalibration (see section 4.5).

(7) Ideally, spectral series should be reviewed by an expert to identify issues like artefacts, improper noise removal, or truncated spectra.

Checking the quality of already available spectral collections reveals that spectral collections part of MassBank as well as the WRTMD would substantially satisfy the proposed quality criteria, and thus are submitted to the second step of the process.

Secondly, each collection of reference spectra will be submitted to benchmarking experiments with suitable collections of high-quality tandem mass spectral data (e.g., WRTMD). Appropriate benchmarking procedures have been described recently [19]. These involve the identification of positive controls and negative controls, and matching these data sets to the tested libraries.

Positive controls are particularly suitable for testing the quality and comparability of databases. Matching positive controls is of importance to identify reliable sources of tandem mass spectral data that can find immediate application for compound identification. Ideally, the obtained sensitivity values (= true positive rates) should be close to 100%.

Negative controls are used to test the specificity (= true negative rate) of a database.

An example of successful benchmarking was elaborated recently [20]. It involved the Eawag collection part of MassBank. MassBank is an important collection of reference tandem mass spectra [21]. Currently, 35 groups are contributing to MassBank (https://massbank.eu/MassBank/RecordIndex.html) with more than 30,000 tandem mass spectra representing thousands of compounds. In terms of compound coverage, there is significant overlap between MassBank and the WRTMD that can be used to create sets of positive controls and negative controls for testing sensitivity and specificity of the databases. For instance, the Eawag set part of MassBank contains 233 compounds that are included in the WRTMD. The corresponding spectra were used as positive controls for benchmarking. Of particularly interest is the fact that the Eawag spectra were acquired with an Orbitrap instrument, whereas the WRTMD spectra were acquired on a QqTOF. Benchmarking revealed that spectra in the range of collision energy 20-50 eV on the QqTOF and 30-60 nominal collision energy units on the Orbitrap provided optimal library matching results with sensitivity-values 95.1-98.4% [20].

Benchmarking of the Eawag collection and the WRTMD provided evidence that both collections enable reliable compound identifications, and that they are ready for use in suspected screening applications.

Thorough quality control of already existing spectral collections will identify libraries for immediate application to compound identification in suspect screening. The tandem mass spectral data will be converted in different formats to enable import into instrument vendor and third party software. Both steps are of importance to ensure immediate and widespread use of created reference data. Data conversion and exchange will be accomplished and/or coordinated by MUI and UFZ. Libraries will be stored at a central server, for instance the MassBank repository (https://github.com/MassBank/MassBank-data). All WP16 members will get access rights for downloading libraries.

5.3 Compound prioritisation

Definitively, consolidation of already available tandem mass spectral data will provide a valuable resource of reference spectra which will cover thousands of compounds that are of importance for human biomonitoring. Nevertheless, our efforts will also identify existing gaps in compound coverage which need to be closed by joined efforts within the consortium.

As the acquisition of reference data for ten thousands of compounds will involve multiple research groups, proper project planning will be obligatory to ensure efficient use of available resources. An integral step will represent the prioritisation of compounds.

Prioritisation criteria might include (1) access to reference standards, (2) inclusion in monitoring programs, (3) occurrence in biological samples, (4) documented human consumption or exposure, (5) industrial use, (6) toxicity or (7) occurrence in environmental samples.

A major hurdle will be the availability of reference standards. Chemical companies are continuously progressing with regard to the commercial availability of reference exposure markers, they represent the first source for such supporting material in a medium and long term perspective. *De novo* chemical synthesis is another option that could be envisaged for specific confirmatory purposes, but it has only a low throughput, and elevated cost. Globally, the available funding will hardly enable the purchase of all compounds of interest. There is a need to find alternative sources, and these might include chemical storages of consortium members or collaborations outside the frame of the HBM4EU project.

A further challenge will be the acquisition of reference spectra for the biotransformation products of suspects. The majority of metabolites have been predicted by *in silico* tools only. Their occurrence in human specimens is awaiting verification. Again, the availability of corresponding standards is unlikely, and these ones need to be produced experimentally employing different kinds of *in vitro* or *in vivo* approaches, or by chemical synthesis. *In vitro* or *in vivo* biological production is the most probable option, but without large capacities.

At the current stage of the project, the further prioritisation strategy needed to organise this building of reference MS library remains complex and difficult to accomplish at the short term. Nevertheless, there is a consensus that the availability of reference standards to HBM4EU WP16 partners will represent the primary selection criteria at the initial phase of the project. Starting with this criterion is rational, and this will give us sufficient time to work on more specific and comprehensive guidance.

5.4 Acquisition of new tandem mass spectral data

In the scope of reaching a high level of interlaboratory comparability of the newly generated reference data, laboratories that are willing to provide experimental spectra must first participate in a harmonisation study. The harmonisation study is intended to verify that any participating lab is applying experimental settings and workflows that ensure such compatibility and transferability of newly acquired tandem mass spectral data with already available reference spectra collections. The intended harmonisation is then considered to represent a prerequisite for the creation of robust tandem mass spectral identification tools.

One way to characterise this interlaboratory comparability is to introduce a range of predefined known compounds as QA/QC criteria that must be recovered and successfully identified with a given instrumentation and related procedure to validate the method appropriateness and reliability.

A set of such reference compounds may be proposed (Table 1), which were already found suitable for testing this comparability of tandem mass spectral data acquired in different laboratories with various available instruments and procedures. These reference compounds have extensively been used to test the transferability of the WRTMD [3,11,13,22,23].

Table 1: Test compounds used to test the comparability of tandem mass spectral data acquired in different labs with available instruments and procedures.

Compound	Molecular formula	*m/z* $[M+H]^+$	CAS
Amiloride	$C_6H_8ClN_7O$	230.0551	2609-46-3
Nylidrin	$C_{19}H_{25}NO_2$	300.1958	447-41-6
Dibucaine	$C_{20}H_{29}N_3O_2$	344.2332	85-79-0
Cyclizine	$C_{18}H_{22}N_2$	267.1855	82-92-8
Desipramine	$C_{18}H_{22}N_2$	267.1855	50-47-5
Dosulepin	$C_{19}H_{21}NS$	296.1467	113-53-1
Dixyrazine	$C_{24}H_{33}N_3O_2S$	428.2366	2470-73-7
Ethambutol	$C_{10}H_{24}N_2O_2$	205.1910	74-55-5
Etilefrine	$C_{10}H_{15}NO_2$	182.1176	709-55-7
Etofylline	$C_9H_{12}N_4O_3$	225.0982	519-37-9
Mefruside	$C_{13}H_{19}ClN_2O_5S_2$	383.0496	7195-27-9
Metoclopramide	$C_{14}H_{22}ClN_3O_2$	300.1474	364-62-5
Antipyrine	$C_{11}H_{12}N_2O$	189.1022	60-80-0
Phentermine	$C_{10}H_{15}N$	150.1277	122-09-8
Phenytoin	$C_{15}H_{12}N_2O_2$	253.0971	57-41-0
Sulfamethoxazole	$C_{10}H_{11}N_3O_3S$	254.0593	723-46-6
Sulfamoxole	$C_{11}H_{13}N_3O_3S$	268.0750	729-99-7
Sulthiame	$C_{10}H_{14}N_2O_4S_2$	291.0467	61-56-3
Tetracycline	$C_{22}H_{24}N_2O_8$	445.1605	60-54-8

As a proof of concept, a first interlaboratory study was organised with six participating laboratories involved in the HBM4EU WP16 consortium. An overview of the labs, the applied instrumentation as well as the applied fragmentation technique is provided in Table 2.

The interlaboratory study clearly demonstrated that the participating labs are able to acquire high-quality reference spectra for building databases. It provided further evidence that Orbitrap-HCD and QqTOF-CID introduce quite similar fragmentation reactions. Thus, databases produced on these types of instruments will offer complementary identification possibilities.

Of utmost importance was the observation that there is a significant overlap of the compound-specific collision energy ranges between instruments. Thus, databases that contain series of multiple spectra acquired on one instrument will enable reliable compound identifications by entering spectra from all other instruments as queries. Clearly, databases produced in different labs will offer complementary identification possibilities.

The interlaboratory study also highlighted some limitations. These are mainly connected with the use of tandem-in-time fragmentation in the ion trap of LIT-Orbitrap.

In contrast to quadrupole collision or HCD cells, ion traps are able to produce fragmentation trees beyond MS^2. Generally, these fragmentation trees cover the full range of possible fragmentation pathways, and are therefore specific identifiers for the corresponding molecules, which can be stored in databases (e.g. mzCloud). With ion trap MS^2, even by applying high collision energies only parts of the entire range of possible fragmentation reactions are covered. Such spectra match well the low energy part of spectral series acquired on tandem-in-space instruments (Orbitrap-HCD

and QqTOF-CID). There is, however, limited overlap with spectra acquired at high collision energies.

Another problem of ion trap fragmentation is related to the “low mass cut-off”, or the so-called “1/3 rule”. This means that fragment ions with an *m/z*-value below 1/3 of the *m/z*-value of the precursor ion are not efficiently trapped under normal operation conditions allowing a high sensitivity and get lost. Thus, a considerable part of fragment ions that are observed with high collision energy fragmentation on tandem-in-space instruments is not detectable with IT fragmentation. In comparison to Orbitrap-HCD and QqTOF-CID, Orbitrap-CID spectra appear to be truncated. This truncation can seriously hamper compound identification if abundant fragment ions are missing. One such example is desipramine. At low collision energies, this compound has one abundant fragment ion at *m/z* 72.08. This ion was not observed in the LIT-Orbitrap spectra. Accordingly, spectral match gave a low score.

Table 2. Overview of the labs participating in the interlaboratory study dedicated to evaluating the degree of comparability and transferability of acquired reference spectra.

Lab	Instrument type	Brand	Fragmentation technique
CEA	Q-Orbitrap-LIT	Orbitrap Fusion, Thermo	HCD
INRA (Toulouse)	LIT-Orbitrap	LTQ Orbitrap XL, Thermo	HCD
INRA (Nantes)	Q-Orbitrap	Q Exactive Orbitrap, Thermo	HCD
MUI	QqTOF	Qstar XL, Sciex	CID
	QqTOF	TripleTOF 5600+, Sciex	CID
UA	Q-TOF	6530, Agilent	CID
UFZ	LIT-Orbitrap	LTQ Orbitrap XL, Thermo	CID and HCD

The results of the interlaboratory study formed the basis for drafting and adopting the following recommendations:

(1) High-resolution instrumentation (i.e. QqTOF, Orbitrap) should be used for the acquisition of the new reference tandem mass spectra.
(2) Instruments should be properly tuned and calibrated. Ideally, the instrument status should be checked on a daily basis or before starting a batch of analyses. Furthermore, high mass accuracy should be maintained by use of a lock mass system or similar disposition. The instrument should provide a minimum resolution of 5000 in MS/MS mode, and the *m/z* error should be lower than 10 ppm.
(3) Reference standards with certified compound identity should be used.
(4) Samples may be introduced by direct infusion, flow injection or chromatography. A special caution should be paid to the minimal number of acquisition points (related to dwell time values and scan speed capabilities) to ensure a sufficient number of spectra to be averaged. Background interferences should also be eliminated.
(5) If the acquisition of reference spectra is accomplished in batches of reference compounds, then isobaric compounds must not be processed in consecutive runs. Otherwise, carryover effects might produce chimeric spectra.
(6) Primarily, protonated or deprotonated molecules will represent the selected precursor ions. Some molecules will also show abundant signals corresponding to in-source fragments, isotopic peaks or other related species. They could be considered as additional precursor ions.
(7) Fragmentation should be accomplished by tandem-in-space techniques (e.g., HCD for Orbitrap, CID for QqTOF).

(8) The lower limit of the applied scan range should be at ≤ *m/z* 50. Larger values will only be accepted if the lower limit of the applied scan range is specified within the stored metadata. The lower limit of the applied scan range should not exceed 100.
(9) Compound-specific breakdown curves should be covered by spectra acquired at multiple collision energies. Instrument-specific collision energy ranges will be deduced form the outcome of the harmonisation study. A spectral series should contain at least 5-10 fragment ion spectra acquired at sufficiently different collision energies within the defined range.
(10) Sample concentration should be sufficiently high to produce fragment ion mass spectra with signal-to-noise ratios >100. In this way, the reliable acquisition of low-abundant fragment ions should be enabled. In most cases, the signal-to-noise ratios could be calculated using the software tools provided by the manufacturers.
(11) For fragment ions, detector saturation should be avoided. Saturation would only be acceptable for precursor ions, if the artefacts are removed during curation.
(12) Fragment ion mass spectra should be acquired in centroid mode or centroided during export and curation (see section 4.5).
(13) Spectra should be curated, which includes multiple steps of filtering, noise removal, and recalibration (see section 4.5).
(14) Spectral series should be reviewed by an expert to identify issues like artefacts, improper noise removal, or truncated spectra (see section 4.5).

5.5 Curation of acquired tandem mass spectral data

Curation of acquired tandem mass spectral data is of utmost importance to obtain a high quality database. Curation efforts may include noise and artifact removal, recalibration of spectra and peak annotations, manual inspection of mass spectra by experts, as well as inter-library comparisons.

Acquired tandem mass spectral data will be extracted from raw data and converted into appropriate ASCII-formats to enable processing which includes multiple steps of centroiding, noise and artifact removal, recalibration of spectra and peak annotations. The preferred software package for accomplishing the described spectral curation will be RMassBank [24]. Other software and workflows will be accepted if they provide similar functionalities and performance.

Removal of noise during data processing may lead to losses of spectral information of compounds. Accordingly, processed spectra will be reviewed by mass spectrometry experts to check the integrity of data with a special focus on the occurrence of artifacts and processing errors.

Ideally, compound-specific reference spectra should be acquired in two different labs using two different types of instrumentation (i.e., QqTOF and Orbitrap) to consolidate the reliability of the generated data. As this is connected with immense costs and efforts, this strategy will only be used at the initial stage of the project and will be omitted for laboratories with a proven record of success in acquiring reference spectra.

Finally, each reference spectrum should be complemented with analytical information of the compound (e.g., names, unique chemical identifiers, molecular formula, type of the precursor ion, *m/z* of the precursor ion) and the applied mass spectral conditions (e.g., mobile phase composition, mode of ionisation, collision energy, CID/HCD, product scan range) to facilitate library searching and identification steps.

5.6 Storage of acquired tandem mass spectral data

The tandem mass spectral data created by the participating consortium members will remain their intellectual property. The groups involved are willing to share the libraries among each other. Curated tandem mass spectral data will be converted in different formats to enable on the one hand upload to open-access repositories such as MassBank (https://github.com/MassBank/MassBank-data) and on the other hand import into instrument vendor and third party software. Both steps are of importance to ensure immediate and widespread use of created reference data. The MassBank consortium is developing tools to convert MassBank record files into the NIST format (in co-operation with NIST). The availability of such general purpose libraries would be of great advantage as most of the vendors' software is able to handle NIST libraries. Importing tools for MassBank records are integrated in the Bruker and Thermo software. The University of Tübingen programmed a converter for MassBank records for direct import in the MassHunter library format. However, the status of the latter is unknown and it is not (yet) included in MassHunter.

Data conversion and exchange will be accomplished and/or coordinated by MUI and UFZ. The latest versions of the developed libraries will be stored at a central server. All WP16 members will get access rights to enable for downloading libraries.

6 Conclusions

The purpose of the present component of the HBM4EU WP16 work plan is to provide a guideline for collecting and generating reference tandem MS spectra to be used for annotating suspect compounds in high-resolution mass spectrometry data for finally identifying a maximum number of exposure markers present in given human biological samples. Most existing databases lack a high level QA/QC consolidation (e.g., % of erroneous information, insufficient or non-adequately documented confidence level associated with the compound identification). Furthermore, some adjustments for the application in HBM have to be made (e.g. inclusion of human metabolites as compared to parent compounds). Therefore, the present initiative is proposing to build a highly qualitative framework shared between the main current actors in the field of suspect screening at the European scale. The proposed roadmap has to be seen certainly as a medium to long term process, but is considered as the necessary investment for finally reaching an appropriate level of quality and reliability for compound identification. The proposed approach represents a first step that will merit further elaboration and harmonisation, mainly because some of the underlying issues remain a matter of research, for which the continuous progress in technology and computational capability has to be considered. Given the diverse instrumental platforms used for acquiring tandem mass spectral data, the elaboration of a unique reference SOP would be then clearly premature and counterproductive at this stage, and is not the objective of this work. This document is intended as a first stone to pave the way to a sustainable partnership permitting to build this ambitioned European capacity of high level and high throughput large scale screening of chemical exposure markers in human matrices.

7 References

1. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, Hollender J (2014) Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol 48:2097-2098.
2. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reily MD, Thaden JJ, Viant MR (2007) Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics 3:211-221.
3. Oberacher H, Arnhard K (2016) Current status of non-targeted liquid chromatography-tandem mass spectrometry in forensic toxicology. TrAC-Trend Anal Chem 84:94-105.
4. Oberacher HA, K. (2015) Compound identification in forensic toxicological analysis with untargeted LC/MS-based techniques. Bioanalysis 7:2825-2840.
5. Kind T, Tsugawa H, Cajka T, Ma Y, Lai Z, Mehta SS, Wohlgemuth G, Barupal DK, Showalter MR, Arita M, Fiehn O (2018) Identification of small molecules using accurate mass MS/MS search. Mass Spectrom Rev 37:513-532.
6. Vinaixa M, Schymanski EL, Neumann S, Navarro M, Salek RM, Yanes O (2016) Mass spectral databases for LC/MS- and GC/MS-based metabolomics: State of the field and future prospects. TrAC-Trend Anal Chem 78:23-35.
7. Blazenovic I, Kind T, Ji J, Fiehn O (2018) Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. Metabolites 8:31.
8. Milman BL, Zhurkovich IK (2016) Mass spectral libraries: A statistical review of the visible use. Trac-Trend Anal Chem 80:636-640.
9. Frainay C, Schymanski EL, Neumann S, Merlet B, Salek RM, Jourdan F, Yanes O (2018) Mind the gap: mapping mass spectral databases in genome-scale metabolic networks reveals poorly covered areas. Metabolites 8:51.
10. Lam H (2011) Building and searching tandem mass spectral libraries for peptide identification. Mol Cell Proteomics 10:R111.008565.
11. Oberacher H, Pavlic M, Libiseller K, Schubert B, Sulyok M, Schuhmacher R, Csaszar E, Kofeler HC (2009) On the inter-instrument and the inter-laboratory transferability of a tandem mass spectral reference library: 2. Optimization and characterization of the search algorithm. J Mass Spectrom 44:494-502.
12. Pavlic M, Libiseller K, Oberacher H (2006) Combined use of ESI-QqTOF-MS and ESI-QqTOF-MS/MS with mass-spectral library search for qualitative analysis of drugs. Anal Bioanal Chem 386:69-82
13. Oberacher H, Whitley G, Berger B, Weinmann W (2013) Testing an alternative search algorithm for compound identification with the 'Wiley Registry of Tandem Mass Spectral Data, MSforID'. J Mass Spectrom 48:497-504.
14. Mylonas R, Mauron Y, Masselot A, Binz PA, Budin N, Fathi M, Viette V, Hochstrasser DF, Lisacek F (2009) X-Rank: a robust algorithm for small molecule identification using tandem mass spectrometry. Anal Chem 81:7604-7610.
15. Ichou F, Schwarzenberg A, Lesage D, Alves S, Junot C, Machuron-Mandard X, Tabet JC (2014) Comparison of the activation time effects and the internal energy distributions for the CID, PQD and HCD excitation modes. J Mass Spectrom 49:498-508.
16. Hufsky F, Scheubert K, Bocker S (2014) Computational mass spectrometry for small-molecule fragmentation. TrAC-Trend Anal Chem 53:41-48.

17. Hill DW, Kertesz TM, Fontaine D, Friedman R, Grant DF (2008) Mass spectral metabonomics beyond elemental formula: Chemical database querying by matching experimental with computational fragmentation spectra. Anal Chem 80:5574-5582.
18. Kaufmann A, Butcher P, Maden K, Walker S, Widmer M (2017) Using in silico fragmentation to improve routine residue screening in complex matrices. J Am Soc Mass Spectr 28:2705-2715.
19. Oberacher H, Weinmann W, Dresen S (2011) Quality evaluation of tandem mass spectral libraries. Anal Bioanal Chem 400:2641-2648
20. Oberacher H, Reinstadler V, Kreidl M, Stravs MA, Hollender J, Schymanski EL (2018) Annotating non-targeted LC-MS/MS data with two complementary tandem mass spectral libraries. Submitted to Metabolites.
21. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Tanaka K, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T (2010) MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom 45:703-714.
21. Oberacher H, Pitterl F, Siapi E, Steele BR, Letzel T, Grosse S, Poschner B, Tagliaro F, Gottardo R, Chacko SA, Josephs JL (2012) On the inter-instrument and the inter-laboratory transferability of a tandem mass spectral reference library. 3. Focus on ion trap and upfront CID. J Mass Spectrom 47:263-270
22. Oberacher H, Pavlic M, Libiseller K, Schubert B, Sulyok M, Schuhmacher R, Csaszar E, Kofeler HC (2009) On the inter-instrument and inter-laboratory transferability of a tandem mass spectral reference library: 1. Results of an Austrian multicenter study. J Mass Spectrom 44:485-493
23. Stravs MA, Schymanski EL, Singer HP, Hollender J (2013) Automatic recalibration and processing of tandem mass spectra using formula annotation. J Mass Spectrom 48:89-99.