Universiteit Antwerpen

This item is the archived peer-reviewed author-version of:

Mass spectrometry-based structure elucidation of small molecule impurities and degradation products in pharmaceutical development

Mass spectrometry-based structure elucidation of small molecule impurities and degradation products in pharmaceutical development

*Youzhong Liu[1,2], Edwin P. Romijn[3], Guido Verniest[3], Kris Laukens[1,2] & Thomas De Vijlder[3, *]*

[1] Department of Mathematics and Computer Science, ADREM data lab, University of Antwerp, Antwerp, Belgium

[2] Biomedical Informatics Network Antwerp (Biomina), University of Antwerp, Antwerp, Belgium

[3] Small Molecule Pharmaceutical Development, Janssen Research & Development, Beerse, Belgium

* Corresponding author: Turnhoutseweg 30, 2340 Beerse, Belgium
Email address: TDEVIJLD@its.jnj.com

Abstract

Mass spectrometry (MS) has played a critical role in the identification of unknown small molecules for many years. As such, it is also a pivotal technique in the structure elucidation of drug impurities and degradation products during pharmaceutical development. In this review we discuss the regulatory expectations and some general practices regarding the identification of process-related impurities and degradation products of small molecule therapeutics; with an emphasis on the role of MS in this process. Recent advances in instrumentation, facilitating MS-based structure elucidation are covered. Over the last few years, many software tools have been developed that aid in the annotation and identification of small molecules by their precursor and product ion spectra. This evolution is promising since it is key to alleviate the current data interpretation bottleneck. The main approaches used in these bioinformatics tools are critically surveyed. This review aims to give meaningful insight in the current state of MS-based structure elucidation of small molecule impurities and degradation products.

Keywords

Mass spectrometry; Structure elucidation; Process-related impurities; Degradation products; Pharmaceutical development; Bioinformatics tools.

Abbreviations

APCI   Atmospheric pressure chemical ionization
API   Active Pharmaceutical Ingredient
ASAP   accelerated stability test
CAD   Charged aerosol detector
CCS   collision cross section
CID   Collision-induced dissociation
CQA   Critical quality attributes
DART   Direct analysis in real-time
DDA   Data-dependent acquisition
DESI   Desorption electrospray ionization
DIA   Data-independent acquisition
DP   Drug product
DPRD Degradation product
DS   Drug substance
EI   Electron ionization
ELSD   Evaporative light scattering detector
EMA   European medicines agency
ESI   Electrospray ionization
fMDiN Functional mass difference network

FAIMS   Field asymmetric waveform ion mobility spectrometry
FT   Fragmentation tree
FT-ICR   Fourier-transform ion cyclotron resonance
FWHM   Full width at half maximum
GC-MS   Gas chromatography-mass spectrometry
H/D   Hydrogen–deuterium
HILIC   Hydrophilic interaction chromatography
ICH   International Council for Harmonisation of Technical Requirements for Pharmaceuticals for
Human Use
IM   Ion mobility
IMP-DPRD Impurity and/or degradation product
IMP-DPRDs Impurities and/or degradation products
MALDI   Matrix assisted laser desorption ionisation
MS/MS Tandem mass spectrometry
MSI   Mass spectrometry imaging
NDMA   N-nitrosodimethylamine
NME   New molecular entity
NMR   Nuclear magnetic resonance spectroscopy
QToF   Quadrupole time-of-flight
PEG   Polyethylene glycol
RDB   Rings-and-double-bond equivalents
RP   Reversed phase chromatography
SIMS   Secondary ion mass spectrometry
SPME   Solid phase microextraction
SWATH Sequential windowed acquisition of all theoretical fragment ion
TWIMS Travelling wave ion-mobility spectrometry
UHPLC High/Ultra-high-performance liquid chromatography
USFDA   Food and drug administration
UV/vis   Ultraviolet–visible spectroscopy
UVPD   Ultraviolet photo-dissociation

1. Introduction

The pharmaceutical development of new small molecule medicines is highly regulated and comes with an ever-increasing cost [1]. From the declaration of a new molecular entity (NME) and the start of clinical trials, until the approval of a new

drug application (NDA), the safety and efficacy of the candidate therapeutic must be demonstrated. In 1995, the first versions of guidelines Q3A and Q3B were adopted by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)[2–4]. These guidelines address the aspects of impurities in new drug substances (DS), which in most cases is composed of only the active pharmaceutical ingredient(s) (API), and drug products (DPRD) (i.e. the formulated dosage form), respectively. More recently, in 2014, guideline M7 on the identification and assessment of mutagenic impurities was published by the ICH [5]. These and other guidelines emphasize

the importance of identifying impurities and degradation products (IMP-DPRDs) of investigational new drugs. Mass Spectrometry (MS) is a pivotal technique in this process of structure elucidation [6].

In pharmaceutical development, a staged-approach for structure elucidation is often applied. An illustration of such an approach, where MS plays a central role, is shown in Fig. 1. Usually, novel IMP-DPRDs from different sources are first detected using chromatography (LC) coupled to UV/vis detectors and/or MS or via an LC on coupled to an open-access MS system. In both cases, low resolution mass spectrometers are often applied. These routine analyses are typically executed in quality control laboratory or another analytical laboratory that is closely associated with chemical or pharmaceutical development. In the second stage, samples containing IMP-DPRDs that need identification are usually handed over to a more specialized structure elucidation laboratory where they are further characterized by high-resolution mass analyzers (QToF, orbitrap) equipped with MS/MS capabilities. This typically occurs when IMP-DPRDs approach or exceed the reporting thresholds described by ICH guidelines. The methodologies used in this stage are the focus of current review since high-resolution MS and MS/MS data play a central role to establish a molecular formula and a structure proposal. The final stage consists of full structure elucidation and utilizes complementary techniques. In this regard, it is important to realize that MS alone is not sufficient for absolute structure elucidation of pharmaceutical IMP-DPRDs. In general, NMR spectroscopy is necessary to unequivocally establish the stereochemistry and to discriminate between potential regioisomers or rotamers in a final stage of structure elucidation [7,8]. Structure elucidation by NMR is typically pursued, when IMP-DPRDs exceed the identification threshold imposed by ICH guidelines during pharmaceutical development. It must be noted that the approach described above is also dependent on the stage of development of the pharmaceutical asset and can vary between different pharmaceutical companies.

In the first part of the review (sections 2, 3 and 4), we will give a brief overview of regulatory expectations and general practices regarding structure elucidation in small molecule pharmaceutical development. Process-related impurities and degradation products will be discussed in section 2 and 3, respectively, both with an emphasis on the role of MS. Section 4 tackles additional challenges when identifying unknown impurities and DPRDs in complex matrices such as marketed drugs.

For the second part (sections 5 and 6), we will discuss recent advances in the MS-based methodology and their impact on structure elucidation in the context of pharmaceutical drug development. Advances in MS hardware and strategies of data acquisition will be elaborated in section 5. Section 6 will be devoted to MS data analysis because the spectra acquired need to be deciphered to unravel the underlying chemical structures. While thousands of compounds can be analyzed by LC-MS within hours (e.g. in metabolomics or proteomics experiments), manual processing of resulting data may take hours or even days for each IMP-DPRD. Therefore, instead of discussing classical expert-driven spectra interpretation, we are going to explore computational procedures that will potentially facilitate and improve structure elucidation. Since some related software, databases and pipelines are currently rarely used for the identification of IMP-DPRDs, this part of the review will be written in a tutorial style with a practical example [9].

We note that the metabolic conversion of small molecule pharmaceuticals is also frequently studied by MS [10,11], but is out of scope in this review. Also for the identification of inorganic impurities (e.g. trace elements), residual solvents, leachables and extractables derived from packaging and other contact material, we refer the reader elsewhere [12].

2. Process-related impurities

The identification of synthesis impurities at the level of the API is generally carried out continuously during pharmaceutical drug development. Also, the strategies for the timing and level (see Fig 1) of structure elucidation are often company and product-dependent. Consequently, literature about these strategies is scarce [13]. In all assessments and strategies, patient safety is the central factor. In early stages of drug development, impurities are generally regarded as safe if they are not exceeding levels detected in preclinical toxicological safety studies. In later stages however, impurities (and DPRDs) are considered as critical quality attributes (CQA) by the ICH guideline Q8 [14] and should be identified when exceeding the identification thresholds defined by ICH. Sometimes, potential impurities that are likely to be present in the drug substance and were assessed to be mutagenic in accordance to ICH guideline M7, might need to be controlled at lower levels [15]. In these specific cases, both the initial identification and quantification of the potential impurities are predominantly performed by MS-based techniques.

Apart from patient safety, the identification of impurities is crucial for acquiring thorough understanding of the API synthesis process. Changes in the synthetic route or downstream processing can have major effects on the impurity profile [16]. Therefore, it is important to not only investigate the impurity profile of the final API, but of every (critical) step in the synthetic route. In this case, it might be appropriate to mainly apply MS-based structure elucidation techniques (Fig. 1, Stage 2) during the early stages of drug development to increase process understanding and proceed for full structure elucidation (Fig. 1, Stage 3) after the commercial manufacturing route has been locked.

Impurities have to be identified not only in the DS but also in (regulatory) starting materials and intermediates [17]. Changes in the impurity profile of starting materials (e.g. when changing suppliers) can impact the formation of impurities in intermediates or even at the level of the API [18]. Next to that, changes in (the purity of) solvents, reagents or catalysts are also plausible culprits whenever novel impurities appear. Detailed prior information on and chemical knowledge of the chemical reactions performed and on the purity of the reaction starting materials can therefore be pivotal to facilitate MS-based structure elucidation.

In most cases, LC-MS using ESI is the primary method-of-choice for MS-based structure elucidation due to the physicochemical properties of the majority of small molecule APIs. For smaller, more volatile starting materials or low molecular weight (mutagenic) impurities that lack a chromophore, GC-MS with EI ionization and or chemical ionization are alternatives that need to be considered. In these cases, (e.g. for sulfonate esters) derivatization of the presumed impurity might also be needed prior to analysis [19]. Determination of stereochemical purity is expected by regulatory authorities but can generally not be

obtained by MS-based techniques. These necessitate the use of complementary techniques (e.g. chiral LC, NMR, etc).

3. Degradation products

The evaluation of the stability of APIs and the identification of DPRDs is generally, like process-related impurities, not associated to a specific time point in drug development but is rather a continuous effort from early development through to approval of the marketing registration for the DP. In fact, the degradation of APIs never stops and can also impact the environment once the API is released from the body and reaches the waste water treatment plants [20]. Evidently, the level of identification of degradants and the associated level of advanced analytical equipment and techniques, rises throughout the process and is governed by regulatory compliance [15]. While up to Phase I and II of clinal trial process, degradant profiling and identification is recommended but strictly not obliged, stability studies are required in Phase III [13].

To predict the stability of APIs and DPs and to develop stability indicating methods, the compounds (or compound mixtures) are subjected to forced degradation tests with stress durations ranging from minutes to a few weeks. Accelerated stability tests, usually under conditions less severe than stress testing, are performed to monitor the degradant profile over longer time periods. The stress conditions to be used in the initial forced degradation studies to reach a relevant level of degradation are not carved in stone but regulatory guidelines (USFDA, ICH, WHO, EMA) recommend evaluating a set of stress factors including acidic and alkaline hydrolysis, oxidation, thermolysis and photolysis [21,22]. Many reviews and monographs are available describing the current most commonly used stress factors and conditions [23–26], but also alternative techniques are being explored to either speed up, miniaturize or automate the degradation and analysis process. Examples include the use of microwave irradiation or sonication to increase degradation rates [27–29], or using degradation in levitated droplets combined with nano-ESI-MS analysis [30].

Although a one-parameter-at-a-time experimental setup is common practice in forced degradation studies and following method development, also multivariate and DoE approaches have been described [31–33]. For drug combinations consisting of multiple APIs, the drug-drug compatibility must be shown and the development of stability indicating methods can sometimes be challenging in terms of the selection of the most appropriate wavelength for UPLC-MS analysis [23,34]. Indeed, by far the mostly used analysis methods for the quantification and identification of degradants formed in forced degradation studies rely on RP or HILIC H/UPLC with UV diode array detection coupled with ESI-MS. In case non-UV active degradation is expected or observed, ELSD or the newer CAD detection is often used [34]. Mapping the organic chemistry of the degradation pathways via forced degradation studies and by using predictive software, insights in degradant types can give direction to the choice of analytical techniques to be used in follow-up studies [35]. Although LC with ESI detection is often chosen as first approach, some volatiles and reactive intermediates formed are better analyzed using headspace GC-MS, sometimes preceded by SPME [36]. An example is given in Fig. 2 where the photochemical

6

degradation of ranitidine gives rise to reactive volatiles (e.g. acetaldehyde oxime and amino) that were assessed via SPME-GC-MS [37,38]. An example is given in Fig. 2 where the photochemical degradation of ranitidine hydrochloride (**1**) as a solid after ICH light irradiation at 500W/m2 for 0.5 to 48 hours was analyzed. These conditions gave rise to the formation of reactive degradant **3** via radical α-cleavage and to volatile and reactive degradants **4** and **5** which are challenging to analyze using standard UPLC-MS. However, in this case the use of SPME-GC-MS allowed the detection of these reactive degradants **3, 4** and **5** amongst a plethora of other photodegradation products [37,38].

In practice, detection of reactive degradants often requires chemical derivatization prior to GC or LC analysis. In this field further advancements in the development of simple and sensitive methods are still needed [39]. Another common challenge is the investigation of configurational lability of drugs with stereogenic centers. For instance, in peptide degradation studies, locating the center where epimerization occurs can be challenging with standard UPLC-MS analysis [40]. Single-enantiomer drugs that are prone to racemization evidently require chiral methods to assess the potential formation of the other enantiomer [41].

4. Analytical challenges of identification in pharmaceutical dosage forms

Structure elucidation of IMP-DPRDs by MS-based techniques can be harder in DPs (dosage forms) compared to the DSs. This is mainly because of the excipients (polymers, disintegrants, lubricants, surfactants etc.) that are used to improve API performance. The presence of these excipients results in a more complex matrix that can hamper detection and thus identification of IMP-DPRDs that might be present. Sample preparation (e.g. liquid-liquid extraction, solid-phase extraction) approaches can be optimized to remove the excipients prior to analysis. If this removal is not possible, more elaborate separation and/or detection strategies need to be developed.

Novel MS techniques, such as IM and MSI, were explored to investigate active pharmaceutical ingredients and their IMP-DPRDs in situ, i.e. in pharmaceutical dosage forms. IM-MS has proven to be a valuable tool to screen small molecule pharmaceuticals in formulations. IM separates ions in the gas phase based on their size and shape. Already in 2005, Weston and colleagues showed that travelling wave ion-mobility spectrometry (TWIMS) can effectively separate chlorhexidine from PEG-containing excipients [42]. Eckers et al. used TWIMS to facilitate detection of drug-related compounds in dissolved Combivir tablets [43]. As shown in figure 3, TWIMS has been used in authors' laboratory to detect IMP-DPRDs of propiconazole in a polysorbate 20 containing dosage form). Polysorbate 20 (also known as Tween 20) is a frequently used polymer excipient, and its molecular heterogeneity hampers the detection of low-level IMP-DPRDs by LC-MS. This can be observed from the complex chromatographic profile in fig. 3a and the complex mass spectrum in figure 3 (top spectrum). It was however noted that propiconazole, eluting at a retention time of 4.25 min, had a distinct IM drift time compared to the polysorbate-related ions eluting at a similar retention time (fig 3b). The raw mass spectra (fig 3c, top spectrum) were therefore filtered on the IM drift time region of the API (fig 3c, bottom spectrum). Doing so a dechlorinated impurity with m/z 308 can be immediately discerned from the polysorbate-related

ions at a retention time of 3.95 min in the filtered spectrum. This example clearly shows that TWIMS can be used as an orthogonal separation technique to discern ions originating from IMP-DPRDs from excipient-related ions in complex mass spectra. Other types of IM, such as field asymmetric waveform ion mobility spectrometry (FAIMS), also have been used to separate pharmaceutical excipients prior to MS detection [44]. Direct analysis in real-time (DART) was previously used in various pharmaceutical formulations such as powders and tablets [45]. Belu et al. used mass spectrometry imaging (MSI) based on ToF-SIMS to image multilayer drug beads used in controlled-release drug delivery [46]. Ambient ionization technologies such as DESI and DART have also been used for MSI of powders and tablets, next to MALDI-based imaging [47]. Nevertheless, we agree with others that these techniques have yet to see mainstream use in pharmaceutical development [48], at least with regard to structure elucidation of IMP-DPRDs.

On the other hand, the presence of excipients can mediate or catalyze degradation of DSs. Although most resulting DPRDs can be identified a priori via forced degradation studies, it becomes much less predictable when an API reacts with an excipient or with an impurity from the excipient, since the newly formed compounds cannot be identified directly by comparing their spectra with reference degradation data collected for the DS alone [49].

5. Technological advancements facilitating data acquisition

Mass spectrometry is a cornerstone technique in structure elucidation of small molecules. As such, it has become inevitable during pharmaceutical development for the identification of process IMP-DPRDs. Over the last decades, and especially since the development of electrospray that facilitated coupling with LC, MS techniques have evolved rapidly. For a long time, HRMS was performed on magnetic sector or Fourier-transform ion cyclotron resonance (FT-ICR) spectrometers. These instruments were expensive and maintenance intensive. Therefore, the introduction of easier-to-use HRMS instruments, first QToF and later orbitrap mass analyzers, heralded a new era in MS-based structure elucidation.

The high resolving power and coinciding mass accuracy of these HRMS instruments greatly facilitated the elemental formula determination of most small molecule ions. The accuracy of these instruments is usually in the low (or sub-) ppm range, corresponding to low (or sub-) mDa mass deviations. Nevertheless, accurate mass alone does not guarantee the correctness of elemental composition, not even if the mass accuracy would be 1 ppm [50]. Therefore, other types of available information like the ratio of the different isotope clusters ,the number of rings-and-double-bonds (RDB) or double bond equivalents (DBE), and the nitrogen rule (see figure 1) do need to be considered. Kind and Fiehn (2007) formulated the "seven golden rules" that should be taken into account in order to determine the correct elemental formula [51] (Stage 2 in Fig. 1). These rules are suitable for pharmaceutical compounds up to 2000 Da consisting of the elements C, H, N, S, O, P, F, Cl and Br. In instances where the unknown analyte is related to a compound with known structure, such as drug IMP-DPRDs, their structural relationship can be exploited to determine the correct elemental composition.

More recently, FT-ICR and orbitrap mass analyzers with "ultra"-high resolving power became available. In 2011, the high-field orbitrap was commercialized with a resolving power of 240,000 FWHM [52],

which was further improved to 1,000,000 FWHM in 2015 [53]. These instruments can resolve the isotopic fine-structure of small molecules, a feature that is extremely powerful in structure elucidation. The resolution of the isotopes of the most commonly occurring atoms (e.g. ^{18}O, ^{15}N, ^{34}S, ^{37}Cl) from the more ubiquitous ^{13}C isotopes can provide a unique isotopic "fingerprint". This fingerprint helps to unambiguously assign the correct elemental composition to an unknown IMP-DPRD [54].

Next to the improvements in resolving power, new mass spectrometers tend to become increasingly sensitive. For the identification of drug IMP-DPRDs, ample material is usually available. So, while very welcome in other areas of MS research (e.g. proteome analysis, bioanalysis of drugs), sensitivity is usually not the biggest challenge in the structure elucidation of drug IMP-DPRDs. With recent instrumentation, it is not very difficult to reach the identification thresholds that are currently described in the regulatory guidelines. Nevertheless, improved sensitivity can be useful in when dealing with IMP-DPRDs that do not ionize very well in the standard ionization modes (usually ESI and/or APCI). New developments in ionization sources (e.g. Unispray [55]) can further help in this regard. High sensitivity can be necessary in other investigational research, such as the screening for potential mutagenic impurities. A recent example highlighting the importance of this research is the detection low levels of the mutagenic NDMA and NDEA impurities in tetrazole containing drugs such as valsartan [56]. Also, highly-conjugated IMP-DPRDs at levels below the identification threshold can cause discoloration of DS or DP material [57,58].

As mentioned above, in some cases it can be more practical to turn to GC-MS instead of LC-MS analyses. However, over the last decade, the progress in HRMS instruments that can be coupled to GC has lagged behind compared to that seen in LC-HRMS instrumentation. On the other hand, the developments in Atmospheric Pressure Gas Chromatography (APGC) [59] and the recent introduction of an GC-compatible orbitrap mass analyzer seem to turn this trend. The second was already applied in a pharmaceutical setting beyond determination of residual solvents [60]. The combination of these new instruments with the advantage of electron ionization (EI) and concurrent spectral library searching has the potential to introduce a paradigm shift in the choice of MS inlet system (LC versus GC).

As stated above, the exact mass measurement of an unknown IMP-DPRD and the contemporary determination of its correct elemental formula is not sufficient to unravel its identity. Therefore, new developments in MS fragmentation techniques and MS/MS data acquisition strategies are equally important for the structure elucidation of these small molecules. Modern mass spectrometers allow for diverse MS/MS methods. Next to targeted MS/MS analysis of the IMP-DPRD precursor ions that were detected in a preceding LC-MS run, one can distinguish between data-dependent acquisition (DDA) and data-independent acquisition (DIA) strategies. In DDA methods, a product ion scan (MS/MS) of a particular precursor ion is automatically acquired once this precursor reaches a certain intensity limit [61]. In DIA methods, mass spectrometers with rapid duty cycles acquire MS/MS scans without selecting specific precursor ions. Here, ions in the entire mass range (e.g. "all-fragment-ion") or a larger part of the mass range (e.g. SWATH) are subjected to fragmentation and a "mixed" product ion spectrum is acquired [62]. In a comprehensive study, Zhu and colleagues compared DDA and DIA approaches for the identification of drug metabolites [63]. They concluded that DDA outperformed DIA in terms of overall spectral quality but some of the metabolite ions did not trigger a MS2 spectrum in the DDA

methodology, especially with increasing matrix complexity. Therefore, DDA and DIA are complementary for structure elucidation challenges that involve complex matrices.

Typically, fragmentation of small molecule ions in a MS/MS experiment is accomplished by collision-induced dissociation (CID) in either a collision cell or an ion trap. The differences between collision cell and ion trap fragmentation are discussed elsewhere [64]. In 2017, an orthogonal fragmentation technique, ultraviolet photodissociation (UVPD), became commercially available. Here, a 213 nm UV laser is used to activate ions via absorption of photons. Since double bonds or aromatic systems are more prone to photon absorption, cleavage will predominantly occur at these sites. This feature makes UVPD highly complementary to CID where, in most cases, cleavage at double bonds or in aromatic systems is precluded. UVPD can provide additional structure information. For example, Ryan and colleagues recently used UVPD to improve the characterization and disentanglement of different classes of sphingolipids [65].

Besides complementary fragmentation techniques, the adoption of multistage fragmentation (MS^n) is also a very valuable development in the identification of drug IMP-DPRDs. Here mass spectral trees (ion trees) are generated by sequential ion isolation and fragmentation events, generally performed in (linear) ion traps. Since all product ions are linked to specific precursor ions in MS^n experiments, the reconstruction of fragmentation pathway can give additional information on the genealogy of the ions and thus the potential identity of precursor (sub)structures. Moreover, since MS^n can yield additional diagnostic ions, it is often used to differentiate constitutional isomers that are not distinguishable by MS/MS [66–68]. A drawback of MS^n-based strategies is that the accumulation of product ion spectra is time-consuming and therefore difficult to combine with rapid UHPLC methods that result in narrow elution windows. This limits the depth of the mass spectral tree that can be recorded. The use of sensitive ion traps with fast cycle times can potentially solve this issue.

Next to mass spectrometric data, mass spectrometers equipped with ion mobility devices provide researchers with orthogonal information that can be exploited for structure elucidation. For example, IM-MS may be used to discern isomers of small molecules [69]. Moreover, IM devices can be used to obtain collision cross sections (CCS) of unknown analytes. These experimental CCS can be used as complementary information in a structure elucidation workflow by comparing them to reference CCS values in databases or theoretical CCS values obtained through quantum chemical calculations [70]. These strategies were already implemented to identify small molecules in metabolomics studies [71], pesticide analysis [72] and in pharmaceutical development [73]. For a general overview of IM-MS applications for small molecules, the reader is referred to the review by Lapthorn et al. [74]

6. Technological advancements assisting data interpretation

In practice, data analysis and interpretation are the most time-consuming steps in MS-based impurity identification. Assigning reliable structures to mass spectra of analytes still requires considerable manual intervention by highly trained MS experts. This might increase the possibility for human errors in the interpretation process [75]. In small molecule pharmaceutical development, the data analysis pipeline is comparable to identification or replication procedures in non-targeted metabolomics because of the

10

similarity in MS instruments, data acquisition and target compound classes (i.e. small molecules). Although metabolomics focuses on biological samples, both fields face similar data interpretation challenges, ranging from elemental formula determination and putative annotation to *de novo* identification [9,76–78]. Surprisingly however, bioinformatic tools (software, databases and pipelines) developed for metabolite identification are rarely applied in MS-based structure elucidation of pharmaceutical products.

In the following section, we will discuss a typical impurity identification pipeline relying on expert knowledge and vendor software. The example is about the structure elucidation for trace-level process-related impurities of Doravirine [79]. The drug solution was profiled with a Waters QToF coupled to a Waters Acquity UPLC system. MS/MS data was generated for Doravirine along with five major impurity peaks, noted as A, B, C, D and E. The vendor software Waters MassLynx was used for raw data processing (denoising, calibration, etc) and for supporting the interpretation of selected MS1 and MS2 scans.

Their pipeline started with molecular formula determination based on accurate precursor masses and isotope patterns of impurity precursor ion spectra [79]. This step was automated by the MassLynx module i-Fit and already led to putative structures for A, B and D based on expert knowledge. Hypothetic structures were validated by product ion spectra via fragmentation pathway reconstruction. Identification of unknowns C and E was more challenging. Authors performed de novo identification based on MS2 spectral similarity with the parent compound. In this study, pathway reconstruction and de novo identification were fully expert-driven. Following identification tools are introduced as alternatives to automate or improve the presented pipeline.

6.1. Precursor formula annotation

The precursor formula determination is the first, and also important step towards impurity identification. In most commercial calculators, such as i-Fit, a function is provided for scoring all possible molecular formulas given an accurate mass, by comparing their simulated isotope patterns with the observed one. However, MS1 spectra alone can often lead to false positive assignment. As for the unknown E in our example, the authors found that the candidate scored in second place ($C_{21}H_{14}N_6O_3F_6Cl$) explained best the product ions in the MS2 spectrum.

In practice, smaller product ions are indicators of a correct formula assignment since they are associated with a smaller number of potential formulas. Based on this feature, Rasche et al. introduced the concept of spectral fragmentation trees (FT) to simultaneously examine formula annotation of precursor and product ions [80]. Combined scoring via isotope and FT analysis is available in the open-source software SIRIUS [81,82]. When we submitted the spectra of E on SIRIUS (version 4.0.1), the joint MS1-MS2 analysis correctly ranked $C_{21}H_{14}N_6O_3F_6Cl$ as the best candidate (Fig. 4A, Fig. S1). The corresponding FT assigned elemental formulas to 6 out of 8 major fragments. SIRIUS also succeeded to compute molecular formulas of four other impurities.

6.2. Structural hypotheses generation

The most common way to generate structural hypotheses in metabolomics is to search the molecular formula against structure databases such as PubChem Compound [83](over 96 million compounds, September 2019). Although these massive databases contain diverse marketed drugs and new chemical entities, they usually have a poor coverage of drug impurities, making this approach less efficient. In fact, searching the keyword "impurity" in PubChem Compound only generates 3381 items (September, 2019), and to our knowledge, drug-specific databases such as DrugBank [84] do not contain impurities or degraded drugs. In our example, by searching PubChem Compound, we obtained putative annotations for compounds A, B and D only, while no structural candidates were found by using DrugBank as search engine.

Recently, a new concept called functional mass difference network (fMDiN) was introduced for automatic hypotheses generation from MS1 data [85]. In a fMDiN, two mass signals (nodes) are connected if their elemental difference (edge) matches one of user-defined biochemical links. Such network eases spectra interpretation by exploring all potential transformations between detected drug API and impurities (Fig. 4B). Given the starting molecule (e.g. drug API) and transformations, potential impurity structures can be deduced with software for organic reactivity prediction [86–89]. However, this approach does not guarantee that all masses are annotated since some impurities (e.g. C and E) are not connected to other compounds via known chemical reactions.

6.3. Structure validation

To validate a putative annotation from product-ion spectra, its theoretical spectrum is compared with measured MS2 spectra. Reference spectra might be found in a spectral library. However, drug impurities are rarely present in public spectra libraries such as HMDB [90], METLIN [91] or GNPS [92]. Therefore, the best alternative to obtain theoretical spectra is to use in silico fragmentation tools such as MetFrag [93], CFM-ID [94] and MSFinder [95]. These tools are key to recent progress in structure identification, at least in metabolomics.

We tested MetFrag for the structure validation of A, B and D. Metfrag first generates a complete graph of substructures for the candidate compound and then prunes the graph to obtain the optimal subset of the substructures that best matches the query spectrum. This process assigns substructures to most product ions (example for B in Fig. 4C). The links between substructures and with the intact compound can help experts reconstruct the fragmentation pathway and validate the structure easily. We note that Metfrag and similar tools do not depend on chemical rules of fragmentation often used by experts. Therefore, they can better capture complex fragmentation reactions and rearrangement.

6.4. De novo identification

Until now, no structure hypotheses were formulated for C and E. In metabolomics-based structure elucidation, de novo identification of such unknowns is considered a challenging task. A recent approach called MS2LDA is possibly useful for the challenge of impurity identification [96]. In fact, impurities and degraded drugs can share substructures between them or with the drug API, which can yield similar product ions in their MS/MS spectra. MS2LDA extracts common patterns of fragments and neutral losses (Mass2Motifs or M2M) from a collection of MS2 spectra. For instance, a M2M containing one

fragment and two neutral losses was found in drug API and impurity E, indicating that they shared substructures (Fig. 4D). Although the structure interpretation requires the intervention by an expert, M2Ms allow easy knowledge transfer from known to unknown spectra. In addition, annotated M2Ms can be stored in a database to be used for identifying new spectra.

We have shown through a typical pharmaceutical identification pipeline the clear utility of metabolite identification tools. Compared to vendor software, these tools are free, polyvalent and open-source, compatible to most instruments after format conversion [64]. The major advantage of these tools is to provide an objective ranking of formula or structural candidates, thus reducing expert intervention time and avoiding human error. Like in other fields, although the structural hypotheses generation step can be fully automated, the final decision will be made by experts possibly in combination with other analytical techniques. On the other hand, the application of some tools is currently limited by the lack of compound databases and spectral libraries that contain drug IMP-DPs. Therefore, a comprehensive evaluation on prediction quality and reliability of structural hits (i.e. false positive rate) [97] is needed to provide clear guidelines for applying such tools in a pharmaceutical context.

7. Conclusions and trends

Throughout pharmaceutical development, ensuring the safety of patients and safeguarding the efficacy of small molecule medicines is of utmost importance. In this regard, structure elucidation of drug IMP-DPRDs is a pivotal aspect and an absolute requirement enforced by regulatory agencies through global and local guidelines. As a result, and also driven by the increasing cost of drug development, there is a need to both improve and accelerate the process of structure elucidation. First, it is crucial to develop a thorough understanding of the specific background of the samples in which certain IMP-DPRDs are observed. Full identification usually combines several different techniques, including chromatography, MS, and NMR as each technique sheds light on different aspects of the chemical structure. As MS-based techniques play a central role in structure elucidation, the recent developments in this area are described here. The advances in MS hardware (e.g. increasing resolving power) and data acquisition techniques, allow to speed up the acquisition enormous amounts of high-quality data. Some papers describe the *in-situ* identification of drugs and their IMP-DPRDs in pharmaceutical dosage forms, but this remains a niche activity. To our opinion, the current bottleneck is in the analysis of this data. Therefore, we embrace the development of novel bioinformatics tools as described here and argue, through the use of a representative example, that tools that were developed for metabolomics and metabolite identification can be adopted in a pharmaceutical setting to facilitate structure elucidation of IMP-DPRDs, despite some specific limitations. As such, we expect that the use of these software tools will be a major area of improvement in the following years and will be used to help the expert in their structure elucidation work.

Reference

[1] A. Mullard, New drugs cost US$2.6 billion to develop, Nature Reviews Drug Discovery. 13 (2014) 877. doi:10.1038/nrd4507.

[2] Abraham - 2010 - International Conference On Harmonisation Of Techn.pdf, (n.d.). http://academy.gmp-compliance.org/guidemgr/files/ICH%20Q3B%20R2.PDF (accessed September 18, 2019).

[3] ICH Topic Q 2 (R1) Validation of Analytical Procedures: Text and Methodology. Step 5 - PDF, (n.d.). https://docplayer.net/21824247-Ich-topic-q-2-r1-validation-of-analytical-procedures-text-and-methodology-step-5.html (accessed September 18, 2019).

[4] ich-q-3-r2-impurities-new-drug-substances-step-5_en.pdf, (n.d.). https://www.ema.europa.eu/en/documents/scientific-guideline/ich-q-3-r2-impurities-new-drug-substances-step-5_en.pdf (accessed September 18, 2019).

[5] M7_R1_Addendum_Step_4_31Mar2017.pdf, (n.d.). http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M7/M7_R1_Addendum_Step_4_31Mar2017.pdf (accessed April 25, 2019).

[6] S. Görög, C. Szántay, Spectroscopic Methods in Drug Quality Control and Development, in: J.C. Lindon, G.E. Tranter, D.W. Koppenaal (Eds.), Encyclopedia of Spectroscopy and Spectrometry (Third Edition), Academic Press, Oxford, 2017: pp. 178–187. doi:10.1016/B978-0-12-803224-4.00066-2.

[7] R.M. Maggio, N.L. Calvo, S.E. Vignaduzzo, T.S. Kaufman, Pharmaceutical impurities and degradation products: Uses and applications of NMR techniques, Journal of Pharmaceutical and Biomedical Analysis. 101 (2014) 102–122. doi:10.1016/j.jpba.2014.04.016.

[8] C. Szántay, Z. Béni, G. Balogh, T. Gáti, The changing role of NMR spectroscopy in off-line impurity identification: A conceptual view, TrAC Trends in Analytical Chemistry. 25 (2006) 806–820. doi:10.1016/j.trac.2006.06.006.

[9] I. Blaženović, T. Kind, J. Ji, O. Fiehn, Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics, Metabolites. 8 (2018). doi:10.3390/metabo8020031.

[10] T. Erkens, R. Bueters, M. van Heerden, F. Cuyckens, R. Vreeken, N. Goeminne, L. Lammens, Translational safety biomarkers of colonic barrier integrity in the rat, Journal of Applied Toxicology. 38 (2018) 1282–1292. doi:10.1002/jat.3639.

[11] 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.

[12] V.P. Sica, K.L. Krivos, D.E. Kiehl, C.J. Pulliam, I.D. Henry, T.R. Baker, The role of mass spectrometry and related techniques in the analysis of extractable and leachable chemicals, Mass Spectrom Rev. (2019). doi:10.1002/mas.21591.

[13] B.A. Olsen, A. Sreedhara, S.W. Baertschi, Impurity investigations by phases of drug and product development, TrAC Trends in Analytical Chemistry. 101 (2018) 17–23. doi:10.1016/j.trac.2017.10.025.

[14] J. Abraham, International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use, in: A. Brouder, C. Tietje (Eds.), Handbook of Transnational

Economic Governance Regimes, Brill, 2009: pp. 1041–1054. doi:10.1163/ej.9789004163300.i-1081.897.

[15] J. Harvey, A. Fleetwood, R. Ogilvie, A. Teasdale, P. Wilcox, S. Spanhaak, Management of organic impurities in small molecule medicinal products: Deriving safe limits for use in early development, Regul. Toxicol. Pharmacol. 84 (2017) 116–123. doi:10.1016/j.yrtph.2016.12.011.

[16] D.D. Wirth, M.S. Miller, S.K. Boini, T.M. Koenig, Identification and Comparison of Impurities in Fluoxetine Hydrochloride Synthesized by Seven Different Routes, Org. Process Res. Dev. 4 (2000) 513–519. doi:10.1021/op000212g.

[17] M.D. Argentine, P.K. Owens, B.A. Olsen, Strategies for the investigation and control of process-related impurities in drug substances, Adv. Drug Deliv. Rev. 59 (2007) 12–28. doi:10.1016/j.addr.2006.10.005.

[18] E.M. Sheldon, J.B. Downar, Development and validation of a single robust HPLC method for the characterization of a pharmaceutical starting material and impurities from three suppliers using three separate synthetic routes, Journal of Pharmaceutical and Biomedical Analysis. 23 (2000) 561–572. doi:10.1016/S0731-7085(00)00333-2.

[19] A. Teasdale, E.J. Delaney, S.C. Eyley, K. Jacq, K. Taylor-Worth, A. Lipczynski, W. Hoffmann, V. Reif, D.P. Elder, K.L. Facchine, S. Golec, R. Schulte Oestrich, P. Sandra, F. David, A Detailed Study of Sulfonate Ester Formation and Solvolysis Reaction Rates and Application toward Establishing Sulfonate Ester Control in Pharmaceutical Manufacturing Processes, Org. Process Res. Dev. 14 (2010) 999–1007. doi:10.1021/op900301n.

[20] B. Petrie, E.J. McAdam, M.D. Scrimshaw, J.N. Lester, E. Cartmell, Fate of drugs during wastewater treatment, TrAC Trends in Analytical Chemistry. 49 (2013) 145–159. doi:10.1016/j.trac.2013.05.007.

[21] A.C. Moraes do Carmo, R.S. Pereira, T. Gratieri, Brazilian requirements for stability indicating methods, TrAC Trends in Analytical Chemistry. 98 (2018) 58–63. doi:10.1016/j.trac.2017.10.017.

[22] P. Sengupta, B. Chatterjee, R.K. Tekade, Current regulatory requirements and practical approaches for stability analysis of pharmaceutical products: A comprehensive review., Int J Pharm. 543 (2018) 328–344. doi:10.1016/j.ijpharm.2018.04.007.

[23] S.W. Baertschi, K.M. Alsante, R.A. Reed, Pharmaceutical Stress Testing: Predicting Drug Degradation, Second Edition, CRC Press, 2016.

[24] K.M. Alsante, A. Ando, R. Brown, J. Ensing, T.D. Hatajik, W. Kong, Y. Tsuda, The role of degradant profiling in active pharmaceutical ingredients and drug products, Adv. Drug Deliv. Rev. 59 (2007) 29–37. doi:10.1016/j.addr.2006.10.006.

[25] S. Görög, Critical review of reports on impurity and degradation product profiling in the last decade, TrAC Trends in Analytical Chemistry. 101 (2018) 2–16. doi:10.1016/j.trac.2017.09.012.

[26] S. Singh, M. Junwal, G. Modhe, H. Tiwari, M. Kurmi, N. Parashar, P. Sidduri, Forced degradation studies to assess the stability of drugs and products, TrAC Trends in Analytical Chemistry. 49 (2013) 71–88. doi:10.1016/j.trac.2013.05.006.

[27] P. Patel, P. Mehta, Impact of Microwave Assisted Heating on Hydrolytic Forced Degradation Study of Pharmaceuticals, Current Microwave Chemistry. 4 (2017) 152–157. doi:10.2174/2213335603666161024145503.

[28] G. Shankaraiah, P. Saritha, N.V. Pedamalla, D. Bhagawan, V. Himabindu, Degradation of Rabeprazole-N-oxide in aqueous solution using sonication as an advanced oxidation process, Journal of Environmental Chemical Engineering. 2 (2014) 510–515. doi:10.1016/j.jece.2013.10.007.

[29] M. Blessy, R.D. Patel, P.N. Prajapati, Y.K. Agrawal, Development of forced degradation and stability indicating studies of drugs-A review, J Pharm Anal. 4 (2014) 159–165. doi:10.1016/j.jpha.2013.09.003.

[30] Y. Li, Y. Liu, H. Gao, R. Helmy, W.P. Wuelfing, C.J. Welch, R.G. Cooks, Accelerated Forced Degradation of Pharmaceuticals in Levitated Microdroplet Reactors, Chemistry – A European Journal. 24 (2018) 7349–7353. doi:10.1002/chem.201801176.

[31] M. Kurmi, S. Kumar, B. Singh, S. Singh, Implementation of design of experiments for optimization of forced degradation conditions and development of a stability-indicating method for furosemide, J Pharm Biomed Anal. 96 (2014) 135–143. doi:10.1016/j.jpba.2014.03.035.

[32] M.N. Patel, C.S. Kothari, Review on Implementation of Multivariate Approach for Forced Degradation Study and Impurity Profiling with Regulatory Considerations, Chromatographia. 81 (2018) 105–125. doi:10.1007/s10337-017-3393-0.

[33] M. Hadzieva Gigovska, A. Petkovska, J. Acevska, N. Nakov, P. Antovska, S. Ugarkovic, A. Dimitrovska, Comprehensive Assessment of Degradation Behavior of Simvastatin by UHPLC/MS Method, Employing Experimental Design Methodology, Int J Anal Chem. 2018 (2018) 7170539. doi:10.1155/2018/7170539.

[34] R.M. Maggio, S.E. Vignaduzzo, T.S. Kaufman, Practical and regulatory considerations for stability-indicating methods for the assay of bulk drugs and drug formulations, TrAC Trends in Analytical Chemistry. 49 (2013) 57–70. doi:10.1016/j.trac.2013.05.008.

[35] A.D.C. Parenty, W.G. Button, M.A. Ott, An expert system to predict the forced degradation of organic molecules, Mol. Pharm. 10 (2013) 2962–2974. doi:10.1021/mp400083h.

[36] M. Jamrógiewicz, K. Pieńkowska, Recent breakthroughs in the stability testing of pharmaceutical compounds, TrAC Trends in Analytical Chemistry. 111 (2019) 118–127. doi:10.1016/j.trac.2018.12.007.

[37] D.E. Latch, B.L. Stender, J.L. Packer, W.A. Arnold, K. McNeill, Photochemical Fate of Pharmaceuticals in the Environment: Cimetidine and Ranitidine, Environ. Sci. Technol. 37 (2003) 3342–3350. doi:10.1021/es0340782.

[38] M. Jamrógiewicz, B. Wielgomas, Detection of some volatile degradation products released during photoexposition of ranitidine in a solid state, J Pharm Biomed Anal. 76 (2013) 177–182. doi:10.1016/j.jpba.2012.12.019.

[39] K. Zhang, J.D. Pellett, A.S. Narang, Y.J. Wang, Y.T. Zhang, Reactive impurities in large and small molecule pharmaceutical excipients – A review, TrAC Trends in Analytical Chemistry. 101 (2018) 34–42. doi:10.1016/j.trac.2017.11.003.

[40] M. D'Hondt, N. Bracke, L. Taevernier, B. Gevaert, F. Verbeke, E. Wynendaele, B. De Spiegeleer, Related impurities in peptide medicines, Journal of Pharmaceutical and Biomedical Analysis. 101 (2014) 2–30. doi:10.1016/j.jpba.2014.06.012.

[41] B. Nickerson, J.J. Salisbury, J.W. Harwood, Enantioselective analysis for L-pidolic acid in ertugliflozin drug substance and drug product by chiral gas chromatography with derivatization, J Pharm Biomed Anal. 159 (2018) 212–216. doi:10.1016/j.jpba.2018.06.042.

[42] D.J. Weston, R. Bateman, I.D. Wilson, T.R. Wood, C.S. Creaser, Direct analysis of pharmaceutical drug formulations using ion mobility spectrometry/quadrupole-time-of-flight mass spectrometry combined with desorption electrospray ionization, Anal. Chem. 77 (2005) 7572–7580. doi:10.1021/ac051277q.

[43] C. Eckers, A.M.-F. Laures, K. Giles, H. Major, S. Pringle, Evaluating the utility of ion mobility separation in combination with high-pressure liquid chromatography/mass spectrometry to facilitate detection of trace impurities in formulated drug products, Rapid Communications in Mass Spectrometry. 21 (2007) 1255–1263. doi:10.1002/rcm.2938.

[44] L.J. Brown, R.W. Smith, D.E. Toutoungi, J.C. Reynolds, A.W.T. Bristow, A. Ray, A. Sage, I.D. Wilson, D.J. Weston, B. Boyle, C.S. Creaser, Enhanced analyte detection using in-source fragmentation of field asymmetric waveform ion mobility spectrometry-selected ions in combination with time-of-flight mass spectrometry, Anal. Chem. 84 (2012) 4095–4103. doi:10.1021/ac300212r.

[45] J. Srbek, B. Klejdus, M. Douša, J. Břicháč, P. Stasiak, J. Reitmajer, L. Nováková, Direct analysis in real time--high resolution mass spectrometry as a valuable tool for the pharmaceutical drug development, Talanta. 130 (2014) 518–526. doi:10.1016/j.talanta.2014.07.007.

[46] A.M. Belu, M.C. Davies, J.M. Newton, N. Patel, TOF-SIMS characterization and imaging of controlled-release drug delivery systems, Anal. Chem. 72 (2000) 5625–5638.

[47] T. Greer, R. Sturm, L. Li, Mass spectrometry imaging for drugs and metabolites, J Proteomics. 74 (2011) 2617–2631. doi:10.1016/j.jprot.2011.03.032.

[48] I.D. Campuzano, J.L. Lippens, Ion mobility in the pharmaceutical industry: an established biophysical technique or still niche?, Current Opinion in Chemical Biology. 42 (2018) 147–159. doi:10.1016/j.cbpa.2017.11.008.

[49] C. Pan, F. Liu, M. Motto, Identification of pharmaceutical impurities in formulated dosage forms, Journal of Pharmaceutical Sciences. 100 (2011) 1228–1259. doi:10.1002/jps.22376.

[50] 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.

[51] T. Kind, O. Fiehn, Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, BMC Bioinformatics. 8 (2007) 105. doi:10.1186/1471-2105-8-105.

[52] S. Eliuk, A. Makarov, Evolution of Orbitrap Mass Spectrometry Instrumentation, Annu Rev Anal Chem (Palo Alto Calif). 8 (2015) 61–80. doi:10.1146/annurev-anchem-071114-040325.

[53] E.D. Hoegg, S. Godin, J. Szpunar, R. Lobinski, D.W. Koppenaal, R.K. Marcus, Coupling of an atmospheric pressure microplasma ionization source with an Orbitrap Fusion Lumos Tribrid 1M mass analyzer for ultra-high resolution isotopic analysis of uranium, J. Anal. At. Spectrom. (2019). doi:10.1039/C9JA00154A.

[54] T. Nagao, D. Yukihira, Y. Fujimura, K. Saito, K. Takahashi, D. Miura, H. Wariishi, Power of isotopic fine structure for unambiguous determination of metabolite elemental compositions: in silico evaluation and metabolomic application, Anal. Chim. Acta. 813 (2014) 70–76. doi:10.1016/j.aca.2014.01.032.

[55] A. Lubin, R. De Vries, D. Cabooter, P. Augustijns, F. Cuyckens, An atmospheric pressure ionization source using a high voltage target compared to electrospray ionization for the LC/MS analysis of pharmaceutical compounds, J Pharm Biomed Anal. 142 (2017) 225–231. doi:10.1016/j.jpba.2017.05.003.

[56] N.A. Charoo, A.A. Ali, S.K. Buha, Z. Rahman, Lesson Learnt from Recall of Valsartan and Other Angiotensin II Receptor Blocker Drugs Containing NDMA and NDEA Impurities, AAPS PharmSciTech. 20 (2019) 166. doi:10.1208/s12249-019-1376-1.

[57] S.M. Blaug, W.T. Huang, Browning of dextrates in solid-solid mixtures containing dextroamphetamine sulfate, J Pharm Sci. 63 (1974) 1415–1418. doi:10.1002/jps.2600630917.

[58] P.A. Harmon, W. Yin, W.E. Bowen, R.J. Tyrrell, R.A. Reed, Liquid Chromatography–Mass Spectrometry and Proton Nuclear Magnetic Resonance Characterization of Trace Level Condensation Products Formed Between Lactose and the Amine-Containing Diuretic Hydrochlorothiazide, Journal of Pharmaceutical Sciences. 89 (2000) 920–929. doi:10.1002/1520-6017(200007)89:7<920::AID-JPS9>3.0.CO;2-W.

[59] T. Bristow, M. Harrison, M. Sims, The application of gas chromatography/atmospheric pressure chemical ionisation time-of-flight mass spectrometry to impurity identification in Pharmaceutical Development, Rapid Communications in Mass Spectrometry. 24 (2010) 1673–1681. doi:10.1002/rcm.4557.

[60] S. Baldwin, T. Bristow, A. Ray, K. Rome, N. Sanderson, M. Sims, C. Cojocariu, P. Silcock, Applicability of gas chromatography/quadrupole-Orbitrap mass spectrometry in support of pharmaceutical research and development, Rapid Communications in Mass Spectrometry. 30 (2016) 873–880. doi:10.1002/rcm.7505.

[61] H.P. Benton, J. Ivanisevic, N.G. Mahieu, M.E. Kurczy, C.H. Johnson, L. Franco, D. Rinehart, E. Valentine, H. Gowda, B.K. Ubhi, R. Tautenhahn, A. Gieschen, M.W. Fields, G.J. Patti, G. Siuzdak, Autonomous Metabolomics for Rapid Metabolite Identification in Global Profiling, Analytical Chemistry. 87 (2015) 884–891. doi:10.1021/ac5025649.

[62] K. Arnhard, A. Gottschall, F. Pitterl, H. Oberacher, Applying 'Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra' (SWATH) for systematic toxicological analysis with liquid chromatography-high-resolution tandem mass spectrometry, Analytical and Bioanalytical Chemistry. 407 (2015) 405–414. doi:10.1007/s00216-014-8262-1.

[63] X. Zhu, 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) 1202–1209. doi:10.1021/ac403385y.

[64] T. De Vijlder, D. Valkenborg, F. Lemière, E.P. Romijn, K. Laukens, F. Cuyckens, A tutorial in small molecule identification via electrospray ionization-mass spectrometry: The practical art of structural elucidation, Mass Spectrom Rev. 37 (2018) 607–629. doi:10.1002/mas.21551.

[65] E. Ryan, C.Q.N. Nguyen, C. Shiea, G.E. Reid, Detailed Structural Characterization of Sphingolipids via 193 nm Ultraviolet Photodissociation and Ultra High Resolution Tandem Mass Spectrometry, J. Am. Soc. Mass Spectrom. 28 (2017) 1406–1419. doi:10.1007/s13361-017-1668-1.

[66] A. Vaniya, O. Fiehn, Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics, Trends Analyt Chem. 69 (2015) 52–61. doi:10.1016/j.trac.2015.04.002.

[67] T. Fouquet, H. Sato, Y. Nakamichi, A. Matsushika, H. Inoue, Electrospray multistage mass spectrometry in the negative ion mode for the unambiguous molecular and structural characterization of acidic hydrolysates from 4-O-methylglucuronoxylan generated by endoxylanases, J Mass Spectrom. 54 (2019) 213–221. doi:10.1002/jms.4321.

[68] P. Shi, Q. He, Y. Song, H. Qu, Y. Cheng, Characterization and identification of isomeric flavonoid O-diglycosides from genus Citrus in negative electrospray ionization by ion trap mass spectrometry and time-of-flight mass spectrometry, Analytica Chimica Acta. 598 (2007) 110–118. doi:10.1016/j.aca.2007.07.027.

[69] J.L. Kaszycki, A. La Rotta, B. Colsch, F. Fenaille, C. Dauly, A. Kamleh, C. Wu, Separation of biologically relevant isomers on an Orbitrap mass spectrometer using high-resolution drift tube ion mobility and varied drift gas mixtures, Rapid Commun. Mass Spectrom. (2019). doi:10.1002/rcm.8414.

[70] S.M. Colby, D.G. Thomas, J.R. Nunez, D.J. Baxter, K.R. Glaesemann, M. Brown, M.A. Pirrung, N. Govind, J.G. Teeguarden, T.O. Metz, ISiCLE: A molecular collision cross section calculation pipeline for establishing large in silico reference libraries for compound identification, (n.d.) 18.

[71] G. Paglia, J.P. Williams, L. Menikarachchi, J.W. Thompson, R. Tyldesley-Worster, S. Halldórsson, O. Rolfsson, A. Moseley, D. Grant, J. Langridge, B.O. Palsson, G. Astarita, Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications, Anal. Chem. 86 (2014) 3985–3993. doi:10.1021/ac500405x.

[72] L. Bijlsma, R. Bade, A. Celma, L. Mullin, G. Cleland, S. Stead, F. Hernandez, J.V. Sancho, Prediction of Collision Cross-Section Values for Small Molecules: Application to Pesticide Residue Analysis, Anal. Chem. 89 (2017) 6583–6589. doi:10.1021/acs.analchem.7b00741.

[73] I. Campuzano, M.F. Bush, C.V. Robinson, C. Beaumont, K. Richardson, H. Kim, H.I. Kim, Structural Characterization of Drug-like Compounds by Ion Mobility Mass Spectrometry: Comparison of Theoretical and Experimentally Derived Nitrogen Collision Cross Sections, Anal. Chem. 84 (2012) 1026–1033. doi:10.1021/ac202625t.

[74] C. Lapthorn, F. Pullen, B.Z. Chowdhry, Ion mobility spectrometry-mass spectrometry (IMS-MS) of small molecules: Separating and assigning structures to ions, Mass Spectrometry Reviews. 32 (2013) 43–71. doi:10.1002/mas.21349.

[75] C.S. Jr, Anthropic Awareness: The Human Aspects of Scientific Thinking in NMR Spectroscopy and Mass Spectrometry, Elsevier, 2015.

[76] T.A. Gillespie, B.E. Winger, Mass spectrometry for small molecule pharmaceutical product development: A review, Mass Spectrometry Reviews. 30 (2011) 479–490. doi:10.1002/mas.20289.

[77] D.H. Nguyen, C.H. Nguyen, H. Mamitsuka, Recent advances and prospects of computational methods for metabolite identification: a review with emphasis on machine learning approaches, Brief Bioinform. (n.d.). doi:10.1093/bib/bby066.

[78] F. Hufsky, S. Böcker, Mining molecular structure databases: Identification of small molecules based on fragmentation mass spectrometry data, Mass Spectrometry Reviews. 36 (2017) 624–633. doi:10.1002/mas.21489.

[79] L.-K. Zhang, R. Yang, H. Sheng, R. Helmy, J. Zheng, Y. Cao, D.R. Gauthier, Characterization of impurities of HIV NNRTI Doravirine by UHPLC-high resolution MS and tandem MS analysis, Journal of Mass Spectrometry. 51 (2016) 959–968. doi:10.1002/jms.3807.

[80] F. Rasche, A. Svatos, R.K. Maddula, C. Böttcher, S. Böcker, Computing fragmentation trees from tandem mass spectrometry data, Anal. Chem. 83 (2011) 1243–1251. doi:10.1021/ac101825k.

[81] S. Böcker, M.C. Letzel, Z. Lipták, A. Pervukhin, SIRIUS: decomposing isotope patterns for metabolite identification, Bioinformatics. 25 (2009) 218–224. doi:10.1093/bioinformatics/btn603.

[82] K. Dührkop, K. Scheubert, S. Böcker, Molecular Formula Identification with SIRIUS, Metabolites. 3 (2013) 506–516. doi:10.3390/metabo3020506.

[83] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B.A. Shoemaker, P.A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E.E. Bolton, PubChem 2019 update: improved access to chemical data, Nucleic Acids Res. 47 (2019) D1102–D1109. doi:10.1093/nar/gky1033.

[84] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res. 46 (2018) D1074–D1082. doi:10.1093/nar/gkx1037.

[85] F. Moritz, M. Kaling, J.-P. Schnitzler, P. Schmitt-Kopplin, Characterization of poplar metabotypes via mass difference enrichment analysis, Plant Cell Environ. 40 (2017) 1057–1073. doi:10.1111/pce.12878.

[86] H. T'jollyn, K. Boussery, R.J. Mortishire-Smith, K. Coe, B. De Boeck, J.F. Van Bocxlaer, G. Mannens, Evaluation of three state-of-the-art metabolite prediction software packages (Meteor, MetaSite, and StarDrop) through independent and synergistic use, Drug Metab. Dispos. 39 (2011) 2066–2075. doi:10.1124/dmd.111.039982.

[87] K.M. Alsante, S.W. Baertschi, M. Coutant, B.L. Marquez, T.R. Sharp, T.C. Zelesky, 3 - Degradation and Impurity Analysis for Pharmaceutical Drug Candidates, in: S. Ahuja, S. Scypinski (Eds.), Separation Science and Technology, Academic Press, 2011: pp. 59–169. doi:10.1016/B978-0-12-375680-0.00003-6.

[88] L. Leclercq, R.J. Mortishire-Smith, M. Huisman, F. Cuyckens, M.J. Hartshorn, A. Hill, IsoScore: automated localization of biotransformations by mass spectrometry using product ion scoring of virtual regioisomers, Rapid Commun. Mass Spectrom. 23 (2009) 39–50. doi:10.1002/rcm.3854.

[89] C.W. Coley, R. Barzilay, T.S. Jaakkola, W.H. Green, K.F. Jensen, Prediction of Organic Reaction Outcomes Using Machine Learning, ACS Cent. Sci. 3 (2017) 434–443. doi:10.1021/acscentsci.7b00064.

[90] D.S. Wishart, Y.D. Feunang, A. Marcu, A.C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, A. Scalbert, HMDB 4.0: the human metabolome database for 2018, Nucleic Acids Res. 46 (2018) D608–D617. doi:10.1093/nar/gkx1089.

[91] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, METLIN: a metabolite mass spectral database, Ther Drug Monit. 27 (2005) 747–751.

[92] M. Wang, J.J. Carver, V.V. Phelan, L.M. Sanchez, N. Garg, Y. Peng, D.D. Nguyen, J. Watrous, C.A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A.V. Melnik, M.J. Meehan, W.-T. Liu, M. Crüsemann, P.D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R.D. Kersten, L.A. Pace, R.A. Quinn, K.R. Duncan, C.-C. Hsu, D.J. Floros, R.G. Gavilan, K. Kleigrewe, T. Northen, R.J. Dutton, D. Parrot, E.E. Carlson, B. Aigle, C.F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B.T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R.A. Keyzers, A.C. Sims, A.R. Johnson, A.M. Sidebottom, B.E. Sedio, A. Klitgaard, C.B. Larson, C.A. Boya P, D. Torres-Mendoza, D.J. Gonzalez, D.B. Silva, L.M. Marques, D.P. Demarque, E. Pociute, E.C. O'Neill, E. Briand, E.J.N. Helfrich, E.A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J.J. Kharbush, Y. Zeng, J.A. Vorholt, K.L. Kurita, P. Charusanti, K.L. McPhail, K.F. Nielsen, L. Vuong, M. Elfeki, M.F. Traxler, N. Engene, N. Koyama, O.B. Vining, R. Baric, R.R. Silva, S.J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P.G. Williams, J. Dai, R. Neupane, J. Gurr, A.M.C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B.M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J.E. Kyle, T.O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K.M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P.R. Jensen, B.Ø. Palsson, K. Pogliano, R.G. Linington, M. Gutiérrez, N.P. Lopes, W.H. Gerwick, B.S. Moore, P.C. Dorrestein, N. Bandeira, Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking, Nature Biotechnology. 34 (2016) 828–837. doi:10.1038/nbt.3597.

[93] S. Wolf, S. Schmidt, M. Müller-Hannemann, S. Neumann, In silico fragmentation for computer assisted identification of metabolite mass spectra, BMC Bioinformatics. 11 (2010) 148. doi:10.1186/1471-2105-11-148.

[94] F. Allen, R. Greiner, D. Wishart, Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification, Metabolomics. 11 (2015) 98–110. doi:10.1007/s11306-014-0676-4.

[95] H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihira, W. Tanaka, T. Cajka, K. Saito, O. Fiehn, M. Arita, Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software, Anal. Chem. 88 (2016) 7946–7958. doi:10.1021/acs.analchem.6b00770.

[96] J.J.J. van der Hooft, J. Wandy, M.P. Barrett, K.E.V. Burgess, S. Rogers, Topic modeling for untargeted substructure exploration in metabolomics, PNAS. 113 (2016) 13738–13743. doi:10.1073/pnas.1608041113.

[97] I. Blaženović, T. Kind, H. Torbašinović, S. Obrenović, S.S. Mehta, H. Tsugawa, T. Wermuth, N. Schauer, M. Jahn, R. Biedendieck, D. Jahn, O. Fiehn, Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: database boosting is needed to achieve 93% accuracy, J Cheminform. 9 (2017). doi:10.1186/s13321-017-0219-x.

Figure Captions

Fig. 1. General, staged scheme of structure elucidation in small molecule pharmaceutical development with an emphasis on the role of MS-based techniques.
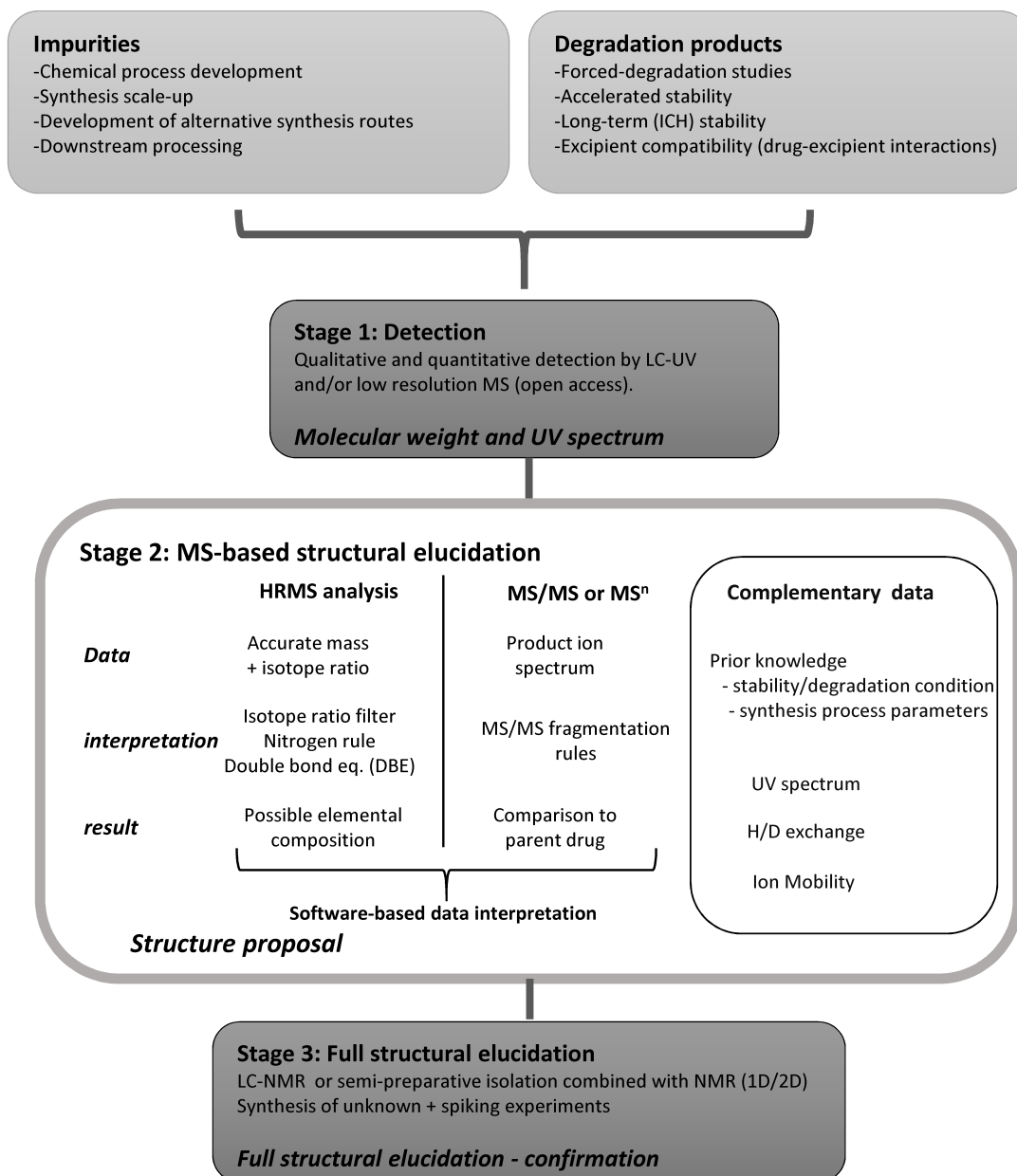
Fig. 2. A selection of reactive and low molecular weight photochemical degradants formed by irradiation of ranitidine at 500W/m2 for 0.5h to 48h at 45°C [37,38].
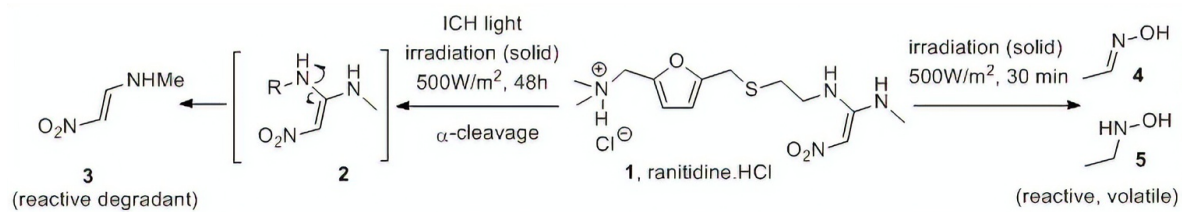
Fig. 3. Ion mobility-assisted detection of a DPRD of propiconazole a) Total ion current chromatogram of dosage-form propiconazole. A complex chromatogram profile is observed due to the presence of

polysorbate 20. The drug API is eluted at a retention time of 4.25 min. b) Two-dimensional retention time - drift time plot of the DP. A DPRD is expected at a retention time of 3.95 min, and it should have a similar drift time (around 50 ms) as the API. c) The raw spectrum of the DPRD extracted at 3.95 min (before drift-time filtration) contains numerous polysorbate-derived mass peaks. d) The drift-time filtered spectrum reveals clear signals of the dechlorinated impurity of propiconazole.

Fig. 4. Bioinformatics tools and concepts to support chemical identification of IMP-DPRDs. A) Fragmentation tree built by SIRIUS for impurity **E** B) fMDiN of precursor ions reveals potential biochemical links between compounds C) MetFrag assigns substructures (highlighted in green) to major fragments of **B** D) MS2LDA discovers a motif from spectra of Doravirine and **E**. Red peak and arrows indicate the common fragment and neutral losses.
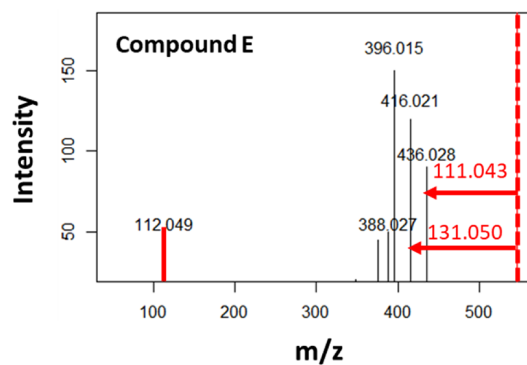
Impurities
-Chemical process development
-Synthesis scale-up
-Development of alternative synthesis routes
-Downstream processing

Degradation products
-Forced-degradation studies
-Accelerated stability
-Long-term (ICH) stability
-Excipient compatibility (drug-excipient interactions)

Stage 1: Detection
Qualitative and quantitative detection by LC-UV
and/or low resolution MS (open access).

Molecular weight and UV spectrum

Stage 2: MS-based structural elucidation

	HRMS analysis	MS/MS or MS^n	Complementary data
Data	Accurate mass + isotope ratio	Product ion spectrum	Prior knowledge - stability/degradation condition - synthesis process parameters
interpretation	Isotope ratio filter Nitrogen rule Double bond eq. (DBE)	MS/MS fragmentation rules	UV spectrum
result	Possible elemental composition	Comparison to parent drug	H/D exchange Ion Mobility

Software-based data interpretation

Structure proposal

Stage 3: Full structural elucidation
LC-NMR or semi-preparative isolation combined with NMR (1D/2D)
Synthesis of unknown + spiking experiments

Full structural elucidation - confirmation

a

b

API

c Raw spectrum ?

d Drift-time filtered spectrum

MW 307.1

A)

Compound E
C21H13ClF6N6O3
547.0717 Da

C4H5N3O

C17H8ClF6N3O2
436.0279 Da

FH

C17H7ClF5N3O2
416.0213 Da

CO

C16H7ClF5N3O
388.0266 Da

FH

C17H6ClF4N3O2
396.0153 Da

FH

C17H5ClF3N3O2
376.0091 Da

CO

C16H5ClF3N3O
248.0145 Da

B)

Doravirine
$C_{17}H_{11}ClF_3N_5O_3$

i) +H_2O iii) - $C_4H_5N_3O$

A $C_{17}H_{13}ClF_3N_5O_4$ D $C_{13}H_6ClF_3N_2O_2$

ii) − O + S

B $C_{17}H_{13}ClF_3N_5O_3S$

i)   Hydrolysis
ii)   Exchange of O with S
iii) Cytosine loss

C)

Compound B

331.976
112.049
443.019
399.046

D)

Doravirine
315.014
112.05
111.043
260.059 295.008 131.050

Compound E
396.015
416.021
436.028
111.043
112.049
388.027 131.050

m/z

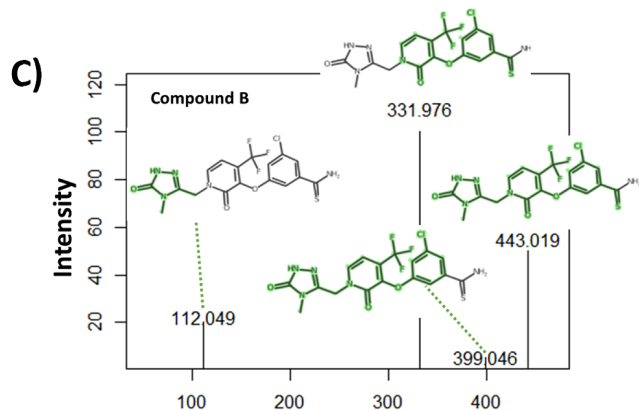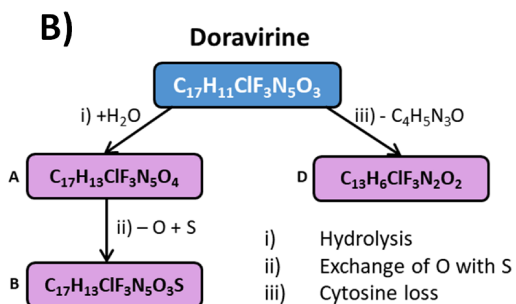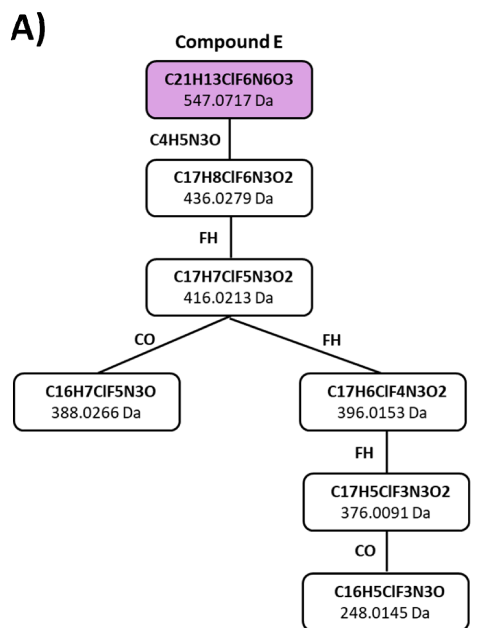**Highlights**

- Structural elucidation of impurities and degradation products of small molecules is a regulatory requirement.
- Mass spectrometry plays a critical role in the structural elucidation process.
- Recent advances in MS instrumentation facilitate the identification of unknown compounds.
- Bioinformatics tools will become critical to alleviate the bottleneck of data interpretation.