

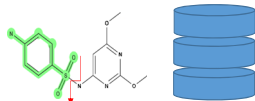
## RESEARCH ARTICLE

# Using In Silico Fragmentation to Improve Routine Residue Screening in Complex Matrices

Anton Kaufmann,<sup>id</sup> Patrick Butcher, Kathryn Maden, Stephan Walker, Mirjam Widmer

Official Food Control Authority of the Canton of Zurich, Fehrenstrasse 15, 8032, Zürich, Switzerland

[DIA spectra]  
+  
[Molecular structure database]  
=  
[Compound name]



**Abstract.** Targeted residue screening requires the use of reference substances in order to identify potential residues. This becomes a difficult issue when using multi-residue methods capable of analyzing several hundreds of analytes. Therefore, the capability of in silico fragmentation based on a structure database (“suspect screening”) instead of physical reference substances for routine targeted residue screening was investigated. The detection of fragment ions that can be predicted or explained by in silico software was utilized to reduce the number of false positives. These “proof of principle” experiments were done with a tool that is integrated into a commercial MS vendor instrument operating software (UNIFI) as well as with a platform-independent MS tool (Mass Frontier). A total of 97 analytes belonging to

different chemical families were separated by reversed phase liquid chromatography and detected in a data-independent acquisition (DIA) mode using ion mobility hyphenated with quadrupole time of flight mass spectrometry. The instrument was operated in the MS<sup>E</sup> mode with alternating low and high energy traces. The fragments observed from product ion spectra were investigated using a “chopping” bond disconnection algorithm and a rule-based algorithm. The bond disconnection algorithm clearly explained more analyte product ions and a greater percentage of the spectral abundance than the rule-based software (92 out of the 97 compounds produced  $\geq 1$  explainable fragment ions). On the other hand, tests with a complex blank matrix (bovine liver extract) indicated that the chopping algorithm reports significantly more false positive fragments than the rule based software.

**Keywords:** In silico fragmentation, Molecular structure database, High resolution mass spectrometry, Suspect, Screening, Residue, Veterinary drug

Received: 19 June 2017/Revised: 28 August 2017/Accepted: 28 August 2017/Published Online: 12 September 2017

## Introduction

Modern multi-residue methods for veterinary drugs routinely screen for more than 100 compounds [1, 2], whereas pesticide methods have been developed to cover more than 400 compounds [3, 4]. Their detection principle commonly relies on monitoring MS/MS transitions or accurate masses in combination with the retention time of the targeted compounds. A compound is considered to be identified when the MS, MS/MS, and liquid chromatographic properties of the suspected peak correspond closely to that of a reference compound. This strategy requires the physical availability of a reference substance. Producing a mixed reference compound solution containing 400 compounds can take a week, while solubility issues as well as chemical and physical stabilities of the individual compounds make it difficult to produce and to maintain such a solution. The shelf life of such a mixed solution

is defined by the most unstable compound. This may be prolonged by freezing the solution, yet there are issues of selective analyte precipitation. Reference compounds can be expensive and are sometimes out of stock. In addition, the physical handling of reference substances can lead to contamination and carryover issues during sample processing or in the analytical instrument (e.g., auto sample loop). Additionally, analytical standards are either not available or prohibitively expensive in the case of marine toxins or pyrrolizidine alkaloids. There may also be a motivation to avoid the physical contact with highly toxic or otherwise legally regulated compounds.

This paper investigated the use of in silico fragmentation algorithms as an alternative to physical reference substances. In silico algorithms are either used to predict the product ion spectrum of a compound based on its molecular structure or to explain the experimentally obtained product ion spectrum of an unknown or suspected compound. Most current published works aim at annotating an unknown chromatographic peak by searching the best fitting candidate present in a molecular

Correspondence to: Anton Kaufmann; e-mail: anton.kaufmann@klzh.ch

structure database [5]. This paper uses in silico fragmentation for a different purpose. It proposes a suspect screening strategy [6, 7] that investigates the precursor and all product ions (as generated by data-independent acquisition; DIA) for the presence of accurate masses that correspond to the in silico fragments of targeted compounds. It is certainly true that searching based on a spectral library (especially if the library spectra were obtained with the same instrument as used for the screening purpose) will produce superior results than can be expected from a theoretical in silico algorithm. Yet, there are a number of reasons why in silico fragmentation-based screening may be of future interest.

A generic approach using in silico fragmentation to detect a large list of suspect (target) compounds could easily be extended to detect similar or derived “non-targeted” compounds. There is a wide demand for such applications in routine analysis. The proof that a certain drug has indeed been taken by a human, or medicated feed has been given to an animal, may be more successful or sensitive by monitoring the metabolites that are frequently present at higher concentration than the parent drug [8]. Unfortunately, such metabolites are rarely available as commercial reference substances. Monitoring a metabolite in addition to the parent drug increases the quality of an analytical finding, as this indicates that the substance has been metabolized (e.g., drug abuse, animal medication) and does not arise from contamination, during sampling, or the laboratory workflow. Finally, forensic chemists rarely have access to designer drug reference materials. Hence, being able to utilize newly reported chemical structures or even designing modifications of currently used drugs would empower forensic chemistry. An illustrative example is the monitoring of sildenafil (Viagra) derivatives, where an ever increasing number of illegal, pharmacologically untested analogues have appeared on the market [9].

There have been a number of excellent papers reviewing the current state of mining molecular structure databases to identify unknown or suspected compounds (see e.g., [5, 10–15]). The oldest approach, which is still commonly used, is rule-based fragmentation spectrum prediction. This methodology relies on an extensive knowledge base obtained from generic mass spectrometry rules as well as published fragmentation reactions, reported rearrangements, and neutral losses. Software such as MassFrontier and ACD/MS Fragmenter [10, 11] can produce fragmentation spectra for any molecular structures provided (e.g., mol files that are freely available from public web-based databases such as ChemSpider).

Another approach is combinatorial fragmentation [16–19], also known as “chopping” or “bond disconnection” algorithms. These approaches “chop” the chemical bonds in a given molecular structure and calculate the exact mass of the resulting fragments. Heuristic rules are used to rank the resulting fragments (often hundreds), for example by penalizing fragmentation in an aromatic ring compared with a C–N bond. Many mass spectrometrists may consider these approaches less sophisticated than the rule-based approach, but these approaches are often complementary [20]. The software verifies if the

experimentally measured product ion spectrum contains peaks that correspond (considering a certain mass error margin) to any of these in silico postulated product ions. In other words, the software performs a post hoc rationalization of the experimental data by systematic bond disconnection based on a scoring function.

There are a number of newer algorithms that use fragmentation trees [11, 21], fingerprint prediction [12, 21], and machine learning (CFM-ID, CSI:FingerID, IOKR) as well as chemical interpretation (MS-FINDER). Such emerging strategies appear promising, yet many were developed and trained on metabolomics data such that it is unclear if they can be applied on the wide range of compounds of interest here. These algorithms are also not yet integrated into vendor software. In a routine laboratory setting, vendor approaches are often easier to integrate into daily routines. Hence, the focus here is on two commercial approaches, complementing recent (extensive) evaluations of openly available approaches. It was beyond the scope here to evaluate additional approaches.

As mentioned before, the conventional use of in silico techniques as reported in a great number of papers starts from an experimentally measured spectrum, which can be more or less reliably linked to a particular molecular structure present in a database. This is different from the screening approach as proposed in this paper. Using an in silico fragmentation algorithm for screening purposes requires the reliable prediction or explanation of at least one product ion for every analyte of interest. If this is not the case, a false negative screening result will be obtained. Hill et al. [22] used rule-based algorithm for screening purposes. They observed at least one predicted product ion for all of their compounds of interest. In addition, they compared the ranking of the true match versus a large number of near isobaric structures downloaded from PubChem; 65 analytes were indeed listed at the top rank (rank no.1). Yet strychnine produced more than 100 experimental fragments of which only two were explained by the software and consequently the rank was 378 out of 664. A different approach based on Bayesian statistic has recently been published [23], which yields the probability of the compound of interest being present/absent in the sample. More frequently, screening was performed by looking for the mass and relative isotopic abundance of the precursor ions. The fragments found were investigated afterwards using in silico-assisted human expert knowledge (e.g., [24]).

The focus of this paper was the evaluation of a technique that may be used in a “routine” residue analysis laboratory. This implies that the algorithm must be sufficiently fast and robust to be capable of evaluating a large number of DIA sample sets consistently. In addition, individual optimization for a particular set of analytes should be kept at a minimum so that the analyst can rely on a generic methodology. Last but not least, the software should be well integrated into the routine LC-MS data processing, preferably obtainable or at least well supported by the instrument vendor. These considerations all led to the investigation of two software: the MassFragment algorithm, as integrated in the UNIFI (Waters) software, which

also controls the ion mobility-time of flight instrument used. On the other hand, the rule-based algorithm (Mass Frontier) approach is not limited to data from Thermo instruments, but is broadly applicable to different instruments, mass accuracies, and fragmentation types via adjustable settings.

The experimental investigation focused on the capabilities of the software to predict or explain experimentally observed product ions for a large number of different compounds with a wide range of structural variety. The absence of false negatives (inability to predict or explain any fragment ion) was investigated. In addition, the appearance of false positives (the software wrongly annotating a matrix-related fragment) when analyzing a complex blank matrix (bovine liver extract) was investigated.

## Materials and Methods

### *Standards and Solutions*

All veterinary drug reference substances were obtained from various sources and were of the highest available purity. Acriflavine A, acriflavine B, albendazole aminosulfone, albendazole sulfone, albendazole sulfoxide, ampicillin, azithromycin, cephalexin, cefalonium, cefaperazone, cefapirin, cefazolin, cefquinom, ceftiofur, chlortetracycline, ciprofloxacin, clenbuterol, clindamycin, clotrimazol, cloxacillin, danofloxacin, dapson, demeclocycline, dicloxacillin, difloxacin, dimetronidazol, doxycycline, enoxacin, enrofloxacin, erythromycin A, fenbendazole sulfone, fleroxacin, flumequine, HMMNI, hydroxyflubendazole, hydroxymebendazole, ipronidazole, ipronidazole OH, josamycin, lincomycin, lomefloxacin, marbofloxacin, mebendazole amine, metronidazole OH, minocycline, nafcillin, nalidixinic acid, norfloxacin, ofloxacin, oleanomycin, orbifloxacin, oxacillin, oxolinic acid, oxytetracycline, penicilline V, penicilline G, pirlimycin, piromidinic acid, praxiquanantel, pyrimethamin, rifamixin, rifampicin, roxithromycin, salinomycin, sarafloxacin, sparfloxacin, spiramycin 1, sulfabenzamide, sulfacetamide, sulfachlorpyrazine, sulfachlorpyridazine, sulfadiazine, sulfadimethoxine, sulfadimidine, sulfadoxine, sulfaguandine, sulfamerazine, sulfameter, sulfamethizole, sulfamethoxazole, sulfamethoxypridazine, sulfamonomethoxine, sulfamoxole, sulfapyridine, sulfaquinoxaline, sulfasalazine, sulfathiazole, sulfisomidine, sulfisoxazole, tetracycline, tiamulin, tilmicosin, tinidazol, triclabendazole sulfone, triclabendazole sulfoxide, trimethoprim, tylosin A were obtained from Sigma-Aldrich (Buchs, Switzerland) and Dr. Ehrenstorfer (Wesel, Germany). The compounds were dissolved individually in various solvents and mixed solutions were prepared. The final mixed spiking solution contained  $0.1 \text{ mgL}^{-1}$  of the analytes. Leucine enkephalin for the lock mass was purchased from Sigma-Aldrich (Buchs, Switzerland). Formic acid, acetonitrile, and dimethylsulfoxide were of analytical grade and obtained from VWR (Darmstadt, Germany). Ammonium hydroxide was from Scharlau (Barcelona, Spain), and the purified water was made

in the lab with a lab-water unit from Labtec (Wohlen, Switzerland).

Mobile phase A was made by adding 50 mL acetonitrile and 3 mL formic acid to a 1000 mL volumetric flask and adjusting to the mark with purified water. Mobile phase B was made by adding 50 mL purified water and 3 mL of formic acid to a 1000 mL volumetric flask and adjusting to the mark with acetonitrile.

A bovine liver was obtained from the local Swiss market and was extracted and processed according to a published method [2]. Briefly, the sample was extracted with a mixture of acetonitrile and aqueous ammonium sulfate. After centrifugation, the organic layer was evaporated and the remaining aqueous layer processed with a HLB reversed phase solid phase extraction cartridge. The eluate was evaporated and reconstituted.

## Apparatus and Measurements

### *LC-IMS-TOF*

The standards and spiked liver extracts were analyzed using a Vion IMS QToF (Manchester, UK) equipped with an electrospray (ESI) ion source, which was coupled to a Waters Acquity UPLC.

A Kinetex-C18 column,  $2.6 \mu\text{m}$ ,  $2.1 \times 150 \text{ mm}$ —from Phenomenex (Schlieren, Switzerland) was utilized at  $25 \text{ }^\circ\text{C}$ , while the injection volume was  $10 \mu\text{L}$ . The gradient was 0–2 min at  $0.4 \text{ mL/min}$ , with 0% B, 2.0–7.0 min at  $0.4 \text{ mL/min}$  with 0–30% B, 7.0–11.0 min at  $0.4 \text{ mL/min}$  with 30%–100% B, 11.0–11.1 min at  $0.8 \text{ mL/min}$  with 100% B, 11.1–12.5 min at  $0.8 \text{ mL/min}$  with 100% B, 12.5 – 12.51 min at  $0.4 \text{ mL/min}$  with 100 – 0% B, 12.51 – 14.0 min at  $0.4 \text{ mL/min}$  with 0% B. The capillary voltage of the ESI interface was set to  $+0.8 \text{ kV}$ . The source temperature was set to  $120 \text{ }^\circ\text{C}$  and the desolvation temperature to  $550 \text{ }^\circ\text{C}$ , whereas the gas flows were set to  $20 \text{ L/h}$  (cone) and  $800 \text{ L/h}$  (desolvation). The experimental settings were as follows: high definition  $\text{MS}^E$ , The scan range was 50–1000  $m/z$  and the scan time was set to 0.2 s. For the collision energy,  $4.0 \text{ eV}$  was chosen for the low setting and the high energy ramp was set to  $4.0\text{--}50.0 \text{ eV}$ . The lock correction settings were: single reference mass (leucine enkephalin):  $556.2766 \text{ } m/z$ ; interval: 1 min.

CCS and mass calibration were done with the Waters calibration mix “major mix IMS-TOF calibration.”

## Data Processing

### *Bond Disconnection Algorithm (MassFragment)*

Data was processed by UNIFI software (v.1.7) using the componentization data extraction and processing mode. Data was extracted by using the noise background filter setting “High”. The intensity threshold for features extraction from the high energy scans was set to 20 counts, the corresponding parameter for low energy scans to 50 counts. The total number

of extracted peaks per channel was 1,000,000. Componentization of detected peaks was based on the maximum charge of a cluster of 1 and (due to the small number of chlorinated veterinary drugs) the maximum number of isotopes was also set to 1.

The obtained features were filtered by the following criteria. The mass deviation from measured and theoretical mass was limited to 3 mDa. The setting “allow scores below” was set to “8”. The setting “keep all fragments” was “yes”. The current software version does not permit the definition of additional parameter to control or direct the in silico fragmentation process (e.g., penalty scores). The mol files used for the bond disconnection algorithm-based in silico fragmentation were downloaded from [chemspider.com](http://chemspider.com).

### Rule-Based Algorithm (Mass Frontier)

Both the “General Fragmentation Rules” and the “Fragmentation Library” were used. The maximum number of reaction steps was limited to 50 and the value of the Reactions Limit to 20,000. The ionization mode was  $[M+H]^+$  protonation (ESI, APCI). Resonance reactions included electron sharing and charge stabilization. H-rearrangements in the even electron ion settings incorporated  $\alpha,\beta,\gamma$  hydrogen transfer and charge remote rearrangements. Otherwise, the default settings as available in MassFrontier 7.0 were used.

## Results and Discussion

### Set-Up of the Comparison

An important aspect of this study was the question whether an in silico algorithm could be used to help reduce the false positives in suspect screening without significantly increasing the false negatives. Hence, a large set of chemically diverse

compounds was investigated. The compounds were mostly regulated, but also included banned veterinary drugs that are commonly analyzed in animal-based food matrices. The compounds represented a number of different chemical families, covered a relevant mass range ( $m/z$  140–850), and were distinct from the substances commonly used to assess such algorithms. The compounds were present in a mixed standard solution at concentrations of 0.1 mg/L. Representative structures given in Figure 1 show the chemical diversity of the compound set. A compound fragment was considered to be “recognized” if the mass deviation between the experimental and theoretical fragment was less than 5 ppm and the ion abundance of the experimental fragment was more than 1% of the base peak. These parameters were deduced from the mass resolving power of the instrument and the selectivity as provided by the deconvolution and the drift time filtering. The use of lower resolving systems may require higher mass deviations or a combination of absolute and relative mass deviations.

As mentioned above, the rule-based prediction algorithm predicts product ions based on an extensive set of rules that were obtained from mass spectrometry literature. The applied rules can be traced back to the literature source where this fragment or fragmentation reaction was initially published. Yet, a number of user definable parameters are available to control the in silico fragmentation process.

The combinatorial “chopping” algorithm (bond disconnection algorithm) explains experimentally observed fragments. Fragments can only be explained when they represent a substructure of the precursor ion since the algorithm simply “chops” the molecular structure and does not account for rearrangements. The algorithm will only report theoretical fragment structures that can be used to explain the experimentally observed product ions. As a matter of fact, the algorithm seems to propose a large number of “useless” fragments, but it only reports the “good” ones (those that match an observed

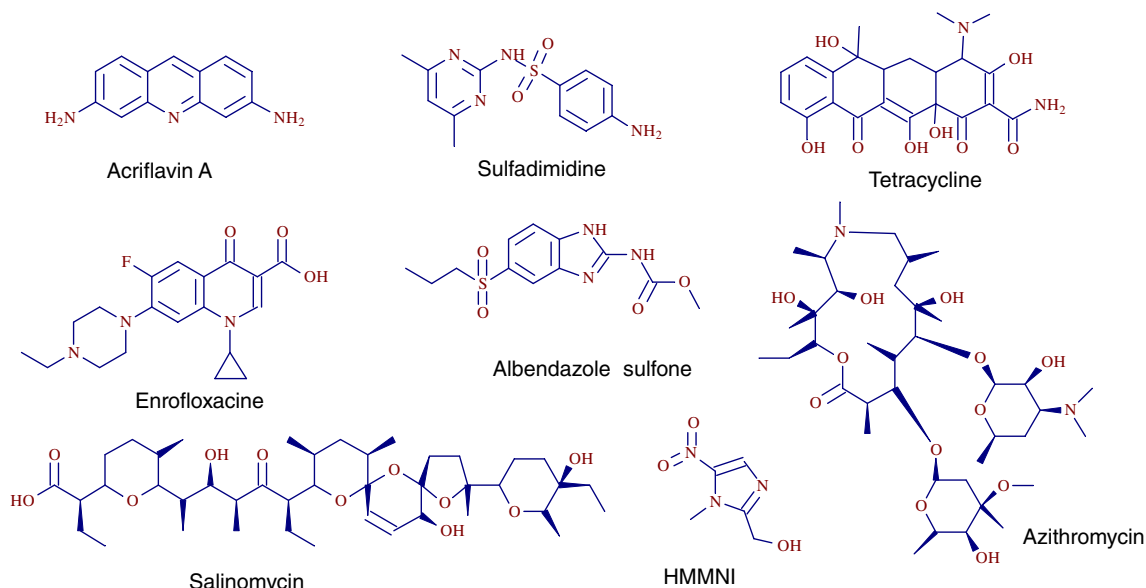


Figure 1. The structural variety of the selected compounds shown with representative structures

**Table 1.** Indicating the Performance of the Tested Software in Detecting the Analytes of Interest. Given are the Number of Correctly Recognized Ions within the Measured Product Ion Spectrum of Each Compound. In Addition, the Summed Ion Abundance of these Correctly Recognized Fragment Ions in Relation (%) to the Total Fragment Ion Abundance is Listed. Data are Given for Both Tested Algorithm (Prediction Performance of the Rule-Based Algorithm and Explanation Performance of the Bond Disconnection Algorithm)

Analyt	Rule-based algorithm		Bond dissociation algorithm	
	no. of fragments	% of abundance	no. of fragments	% of abundance
Acriflavin A	0	0	0	0
Acriflavin B	0	0	0	0
Albendazol aminosulfon	1	51	2	92
Albendazol sulfon	3	87	3	87
Albendazol sulfoxid	4	82	6	86
Ampicillin	5	41	9	76
Azithromycin	3	72	4	75
Cefalexin	3	15	3	22
Cefolonium	6	70	7	76
Cefaperazon	0	0	4	13
Cefapirin	10	46	11	40
Cefazolin	1	24	1	24
Cefquinom	4	62	7	67
Ceftiofur	3	41	10	75
Chlortetracycline	4	54	10	66
Ciprofloxacin	3	44	6	71
Clenbuterol	2	44	5	51
Clindamycin	3	90	3	90
Clotrimazol	1	38	5	65
Cloxacillin	5	32	6	36
Danofloxacin	0	0	3	54
Dapson	0	0	0	0
Demeclocycline	9	48	11	50
Dicloxacillin	3	100	2	87
Difloxacin	1	2	4	83
Dimetronidazol	1	59	1	59
Doxycycline	3	71	6	78
Enoxacin	1	40	2	41
Enrofloxacin	0	0	3	84
Erythromycin A	9	63	17	79
Fenbendazol sulfon	1	81	2	82
Fleroxacin	1	2	5	86
Flumequine	2	77	2	77
HMMNI	1	100	1	100
Hydroxyflubendazol	1	81	2	85
Hydroxymebendazol	1	82	2	84
Ipronidazol	2	55	3	93
Ipronidazol OH	1	81	2	91
Josamycin	3	92	4	100
Lincomycin	3	91	3	91
Lomefloxacin	1	2	5	81
Marrbofloxacin	1	75	3	90
Mebendazol amin	2	100	2	100
Metronidazol OH	3	78	3	78
Minocyclin	4	63	8	75
Nafcillin	3	51	3	51
Nalidixinic acid	2	86	2	86
Norfloxacin	1	1	5	47
Ofloxacin	2	48	4	89
Oleanomycin	2	65	6	94
Orbifloxacin	0	0	2	84
Oxacillin	3	17	5	40
Oxolinsäure	2	79	1	48
Oxytetracycline	7	58	13	67
Penicillin V	2	10	0	0
Penicillin G	4	28	2	19
Pirlimycin	4	66	5	77
Piromidinic acid	2	83	2	83
Praxiquanтел	3	85	3	85
Pyrimethamin	0	0	1	42
Rifamixin	2	49	5	54
Rifampicin	2	51	7	81
Roxithromycin	2	55	6	69
Salinomycin	4	49	9	76

Table 1 (continued)

Analyt	Rule-based algorithm		Bond dissociation algorithm	
	no. of fragments	% of abundance	no. of fragments	% of abundance
Sarafloxacin	0	0	2	74
Sparfloxacin	0	0	3	80
Spiramycin 1	6	16	10	59
Sulfabenzamide	2	59	2	59
Sulfacetamide	0	0	0	0
Sulfachlorpyrazine	2	22	5	65
Sulfachlorpyridazine	1	9	2	9
Sulfadiazine	2	19	3	20
Sulfadimethoxine	1	12	3	67
Sulfadimidine	2	9	4	51
Sulfadoxine	0	0	4	57
Sulfaguanidine	1	55	1	55
Sulfamerazine	2	25	4	71
Sulfamerter	2	27	4	39
Sulfamethizole	2	10	2	10
Sulfamethoxazole	2	29	2	29
Sulfamethoxypyridazine	2	20	4	48
Sulfamonomethoxine	2	28	4	49
Sulfamoxole	2	12	3	14
Sulfapyridine	2	29	3	34
Sulfaquinoxaline	2	38	5	62
Sulfasalazine	1	55	2	60
Sulfathiazole	2	57	2	57
Sulfisomidine	0	0	2	93
Sulfisoxazole	2	60	3	72
Tetracycline	5	46	10	51
Tiamulin	3	89	3	89
Tilmicosin	3	9	4	56
Tinidazol	2	11	3	15
Triclabendazol sulfon	0	0	1	11
Triclabendazol sulfoxid	0	0	2	9
Trimethoprim	1	10	4	57
Tylosin A	2	83	2	83
Total	218	4057	387	5840
Average	2.2	41.8	4.0	60.2
No. of undetected analytes	15		5	

fragment) to the user. The MassFragment bond disconnection algorithm is part of the UNIFI software. The user has access to a number of parameters (e.g., maximum number of bonds to be broken, hydrogen tolerance, etc.) when focusing on single spectra investigation. Unfortunately, undisclosed default settings are enforced when MassFragment is used as a part of the data processing software (e.g., when automatically analyzing all the features as obtained by DIA).

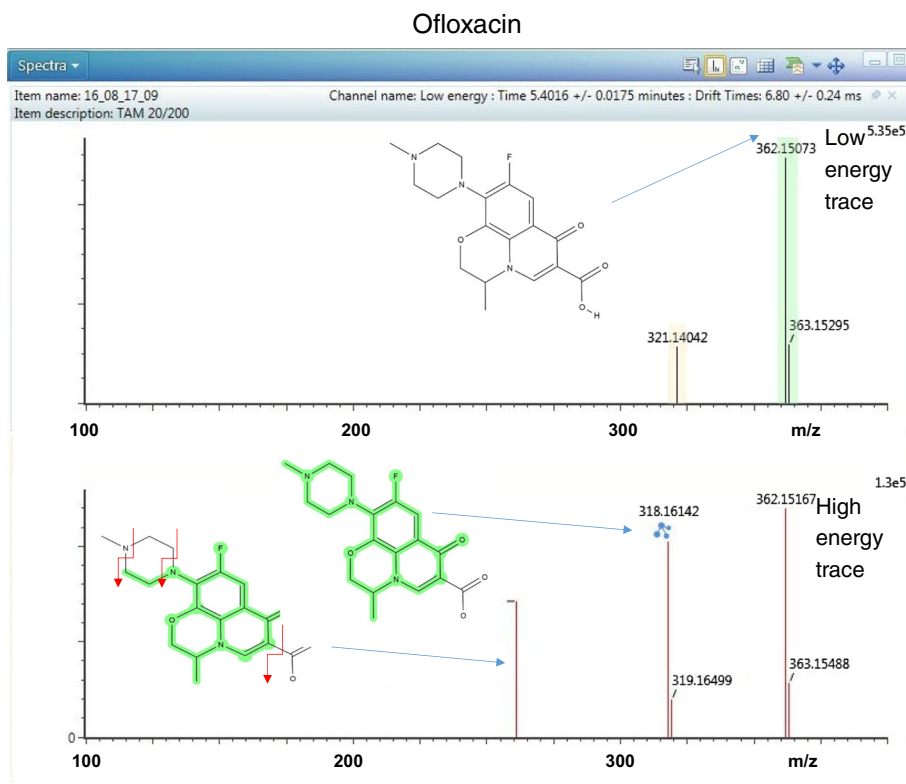
The specific particularities of rule-based algorithm and bond disconnection algorithm discussed above make a fair comparison extremely difficult. Yet, it was the aim of this paper to make a comparison as generic as possible. The question was not how can expert knowledge-guided fine tuning be used to improve the performance for a particular compound or set of similar compounds, but instead what output can an average routine user expect when he/she uses one or the other algorithm for any given small molecule. A special focus was on the number of false positive and false negative findings. Therefore, the default settings were selected for rule Mass Frontier, but both rule-based algorithm knowledge bases (General Fragmentation Rules and Fragmentation Library) were activated, since this significantly improved the number of correctly proposed

fragments. The only chosen deviation from the default settings was the selection of a higher number of permitted reaction steps and reaction limits. It was reported previously for Mass Frontier on EI-MS data, where using fewer fragmentation steps were shown to be more selective for the correct structure [25]. Yet, such settings may cause an increase of the number of false negative findings.

Although it was initially planned to compare the number of correctly and incorrectly proposed fragments, this was not possible since MassFragment only reports the explained fragments. Therefore, false positive findings had to be evaluated in a different way here. A complex blank sample (bovine liver extract) was analyzed and the top 20 “false positive” hits were investigated in detail to compare the number and abundance of the wrongly explained ions obtained by the two in silico algorithms.

#### *Prediction and Explanation Capability of In Silico Fragmentation (False Negatives)*

Table 1 lists the number of predicted and explained fragments as well as the derived total percentage of predicted and



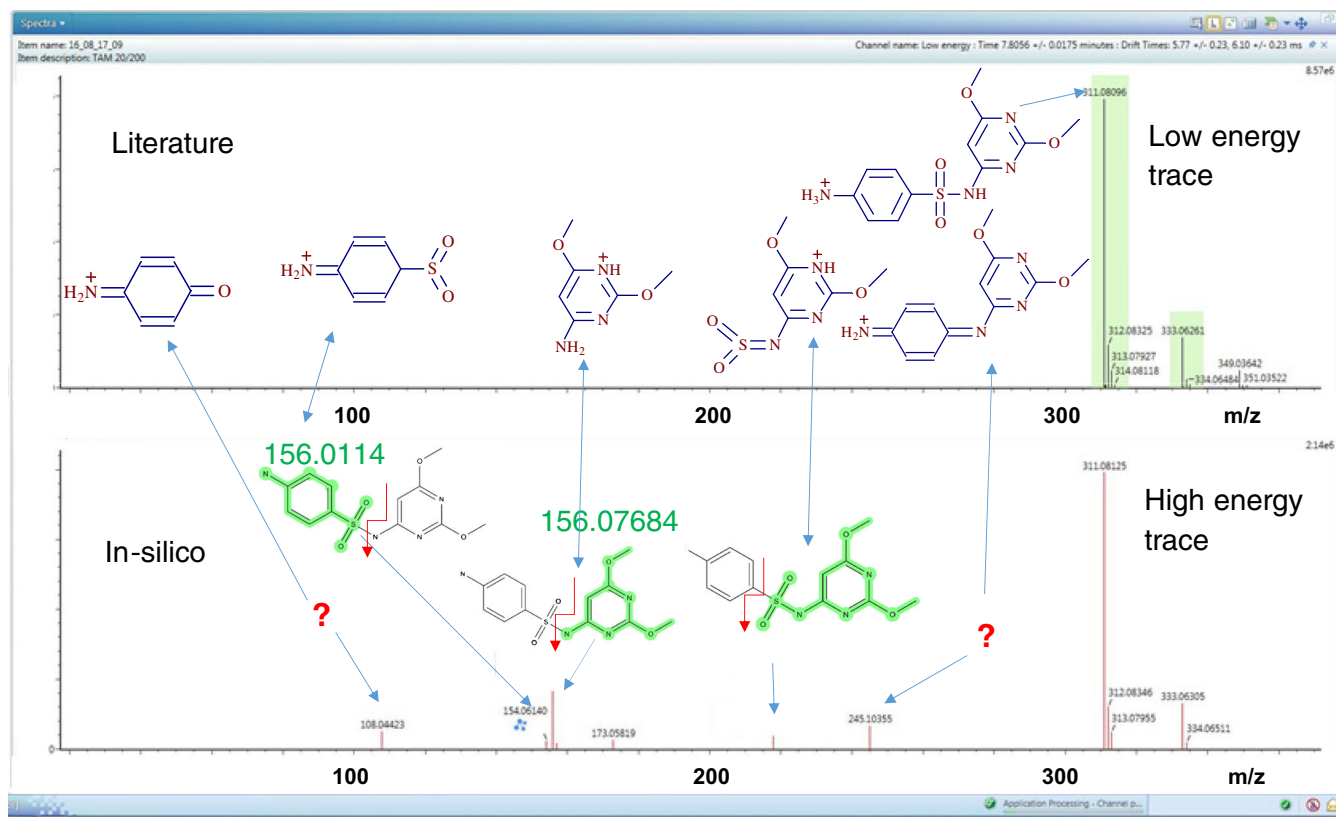
**Figure 2.** Spectra of ofloxacin as obtained by IMS-MS<sup>E</sup>. The top plot shows the low energy trace; the high energy trace is below. The three structures drawn were automatically annotated with the Bond disconnection algorithm. These structures could not be explained by rule-based algorithm

explained product ion spectral abundance. On average, the rule-based algorithm predicted 2.2 fragments per analyte. This average of 2.2 ions accounted to 41.8% of the experimentally observed total fragment ion abundance. The bond disconnection algorithm explained 4.0 fragments per analyte on average, corresponding to an average 60.2% of explained total fragment ion abundance. Owing to the chemical variety of the investigated compounds and the generic, ramped collision energy, the number of detectable fragments (including isotopic contributions) per compound varied between 1 and 83. Even more relevant is the fact that rule-based algorithm failed to predict any correct fragments for 15 compounds, whereas the bond disconnection algorithm only failed to explain any fragment for five compounds. These five compounds were not detectable using this approach for screening purposes due to a lack of detectable fragments or precursors. Sulfacetamide and penicillin V were detected as the sodium adduct, which did not produce fragments with the collision energy regime used. The hydrogen adduct of penicillin V escaped detection because the labile precursor ion fragmented already in the interface region. Acriflavin A and B as well as Dapson (highly aromatic compounds) did not produce any relevant fragments with the applied generic collision energy regime. The failure of the rule-based algorithm to predict at least one correct fragment was the reason behind the inability of the rule-based algorithm to explain the remaining 10 analytes, not the absence of precursor or product ions. Five of these compounds for which no product

ion could be proposed by the rule-based algorithm belong to the class of quinolones. The others belonged to different chemical families (see Table 1). Quinolones are important veterinary drugs and should not be overlooked by a multi-residue screening technique. Therefore, this limitation was investigated further. Quinolones can easily be fragmented and produce intense fragments (see Figure 2). Generally, the carboxylic group disappears due to the neutral loss of carbon dioxide during the initial fragmentation step. This is normally followed by the loss of sections of the ion not directly adjacent to the carboxylic function. As the rule-based algorithm (using the positive ionization mode) never proposed a carbon dioxide loss, it was also not able to propose the second and third generation fragments. Although this comparison was performed using the “default settings”, attempts to modify the available settings to explain this neutral loss remained unsuccessful. Therefore, including this rule (loss of carbon dioxide) would be a valuable addition to utilized rule-based algorithm.

As expected, the “chopping” bond disconnection algorithm was never able to explain true rearrangements (e.g., McLafferty rearrangements). On the other hand, rearrangements were not frequently observed among the studied compounds. The majority of observed fragments were caused by rather simple bond breakage. The fragmentation of sulfadimethoxine serves as an interesting example (Figures 3 and 4). The rule-based algorithm predicted only the ion  $m/z = 156.0114$ . The bond disconnection algorithm explained three ions (as shown in Figures 3 and 4 at

## In Silico-Fragmentation



**Figure 3.** Spectra of sulfadimethoxine as obtained by IMS-MS<sup>E</sup>. The top plot shows the low energy trace; the high energy trace is below. The three structures drawn (high energy trace) were automatically annotated by Bond disconnection algorithm. The drawn structures (low energy trace) were obtained from literature

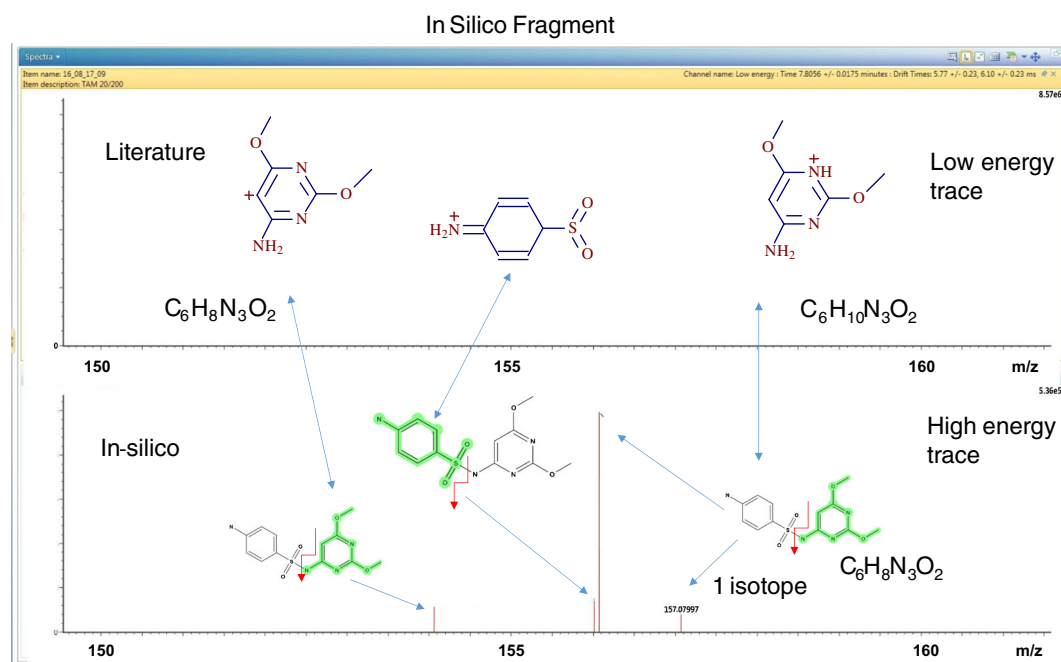
the bottom trace). The top trace in Figures 3 and 4 shows the “true” structure as listed in the scientific literature [26–28]. The two rearrangements ( $m/z$  108 and  $m/z$  245) initiated by the neutral loss of sulfur dioxide were neither predicted by rule-based algorithm nor explained by bond disconnection algorithm. Yet, as expected, the bond disconnection algorithm was able to explain the unique intramolecular isobaric fragmentation of sulfadimethoxine, as described in a paper from Thurman et al. [27]. These two ions ( $m/z$  = 156.0114 and 156.0768) are more clearly visible in the amplified spectrum in Figure 4. Figure 4 shows another interesting feature. The ions  $m/z$  = 154 and 157 were found by bond disconnection algorithm. They are true fragments of the investigated compounds, but wrong structures are proposed (see bottom trace of Figure 4). Note the algorithm bond dissociation algorithm draws identical structures for the ion  $m/z$  = 154 and 157. The top trace shows again the structures as published in the literature [26–28]. This strange observation highlights the way the “chopping” algorithm works. The bond disconnection algorithm initially only considers heavy atoms (hydrogens are excluded). It will propose the breakage of the S–N bond. The remaining part of the molecule  $[C_6H_8N_3O_2]$  must not only be charged  $[C_6H_8N_3O_2+H]^+$  but would have an unpaired electron. The “chopping” algorithm is not able to explain how an ion with

an unpaired electron will be stabilized (e.g., obtaining a hydrogen from the other part of the precursor ion, or producing a double bond etc.). Being aware of this limitation, programmers included the “Maximum H difference” parameter. If the accurate mass of  $[C_6H_8N_3O_2+H]^+$  cannot be found, the masses corresponding to  $[C_6H_8N_3O_2+H_2]^+$ ,  $[C_6H_8N_3O_2+H_3]^+$ ,  $[C_6H_8N_3O_2]^+$ ,  $[C_6H_8N_3O_2-H_1]^+$  will be searched. A penalty may be added to account for the addition or subtraction of more H’s from the fragment formula. While this may look like a “cheap fix,” it significantly improves the performance of the algorithm and has also been applied in other bond disconnection approaches [17, 19]. Certainly this feature will also increase the number false positive hits. Hence, the incorporation of hydrogen rearrangement rules [29] into the bond dissociation software would reduce the number of such false positive hits.

### Overprediction and Overexplanation of In Silico Fragmentation (False Positives)

The proposed screening strategy, to be successful, should lead to a marginal number of false negative findings. The inability to explain a single mass of a targeted analyte will inevitably lead to a false negative screening finding, irrespective of the analyte concentration present in a sample. Hence, a low false negative





**Figure 4.** An expanded section of Figure 3. Note the explained isobaric fragmentation ( $m/z = 156$ ). In addition, the two ions ( $m/z = 154$  and  $157$ ), which the Bond disconnection algorithm annotated with the identical structure are visible. The top trace (low energy trace) contains the structures as reported in literature

rate is of highest importance. Although false positive findings can still be reduced or eliminated by orthogonal filtration criteria or by time-intensive human expert knowledge, it is still important to keep the false positive findings below a certain level. Therefore, the performance of both algorithms was investigated regarding the likelihood that an unrelated ion is

wrongly considered to be an analyte product ion (which may happen often in DIA analysis of complex matrices). As explained above, the total number of proposed fragment formulas or  $m/z$  values was not available from the bond disconnection algorithm. Hence, a simple comparison of the number of true hits versus not existing fragments is not possible.

**Table 2.** False Positive Hits Obtained by Analyzing a Blank Bovine Liver Extract. This Lists the 20 Most Intense Features that were Obtained when Matching the Precursor Mass and One of the Ionized Structures in the Database within 3 mDa. The First Column gives the Name of the Wrongly Assigned Compound

Analyt	Signal abundance	Rule-based algorithm		Bond dissociation algorithm	
		no. of fragments	% of abundance	no. of fragments	% of abundance
Marbofloxacin	370000	0	0	3	11.9
Ciprofloxacin	75000	3	1.4	3	1.3
Sarafloxacin	69000	0	0	1	4.1
Marbofloxacin	57000	0	0	1	2.84
Salinomycin	44000	1	0	1	0.2
Sulfadimidine	38000	0	0	0	0
Tiamulin	37000	2	0.1	13	0.6
Sulfaisomidine	29630	0	0	0	0
Clenbuterol	27000	2	38.8	1	5.3
Sulfadimidine	23384	0	0	0	0
Oxytetracyclin	23000	0	0	7	2.6
Tilmicosin	22000	7	0.2	53	103
Azithromycin	20000	0	0	3	0.1
Penicillin G	19628	9	31.53	8	43.2
Spiramycin 3	19220	3	0.39	67	148.9
Sulfadimidine	15077	0	0	0	0
Penicillin G	14754	1	0.96	3	23.75
Nafcillin	14311	0	0	9	7.72
Tiamulin	14001	0	0	4	4.04
Sulfadimidine	13219	0	0	0	0
Total		28	73	177	360
Average		0.3	3.7	1.8	18.0
No. of false positives		8		15	

Therefore, a very complex blank matrix sample (bovine liver extract) was chromatographed and analyzed by the described DIA technique (a targeted method was used beforehand to ensure that the sample was indeed free of the veterinary drugs of interest). The DIA signals were deconvoluted and componentized by the UNIFI software. The used target compound list (molecular structure library) contained the analytes listed in Table 1. A “hit” was recorded for any extracted feature that showed a precursor mass located within a mass window of 3 mDa. The features were ranked according to their abundance (peak area) and the top 20 of these false positive hits were analyzed for the number of “wrongly” detected product ion fragments. The abundance (peak area) of all wrongly explained and predicted fragments were summed and compared with the total abundance (peak area) of all the signals present in the deconvoluted and componentized feature (product ion spectrum). This data is given in Table 2. This shows that a greatly varying number of wrongly detected fragments and corresponding summed ion abundance was detected, depending on the feature. Yet, the rule-based algorithm shows a significantly lower rate of false positive findings.

As mentioned earlier, the first and most important capability of a screening methodology is the absence of false negative findings. Yet, the routine use of a screening technique is only feasible if the number of false positive findings can be kept within a certain extent. Hence, it will be important to investigate orthogonal post-extraction data processing algorithm that can be used to improve the ranking of the hits produced by in silico-based screening. This is the topic of a recent paper [30] published by the same authors, which investigated the practical application of the concept proposed within this “proof of principle” paper. Ion mobility was not only found to be able to separate precursor ions within the first DIA dimension, but in addition, ion mobility permits filtration tools that significantly reduce false positive findings. Unlike other tested filtration strategies, ion mobility-based methodologies are capable of eliminating false positive findings [30] where very low ion abundance is involved. It is also conceivable to couple alternative approaches such as MetFrag, CFM-ID, Finger ID, or the investigated MassFrontier in post-processing the spectra to further reduce the number of false positives.

## Conclusions

The investigation showed that using in silico fragmentation is a feasible way to reduce the number of false positive matches in suspect screening. Yet there are clear performance differences between the two tested methodologies. The “chopping” algorithm gave a lower number of false negative findings than the rule-based algorithm. A chopping algorithm is probably better suited to explain fragments of a large variety of molecular structures than a rule-based algorithm, as the rule-based algorithms only propose fragmentation reactions that have been explicitly programmed into the system. The failure to predict useful fragmentation trees (initial neutral loss of carbon

dioxide) for quinolones may be an example for this limitation. On the other hand, the rule-based algorithm may be better suited to post processing instead of candidate selecting.

The “chopping” bond disconnection algorithm is significantly faster than rule-based algorithm (when utilizing the rule-based algorithm library rules) and it is integrated into the UNIFI data processing algorithm. This is relevant for the intended application, where the software has to do an automatic search of every feature extracted out of the DIA data by a deconvolution algorithm. The good performance of the basic chopping algorithm compared with the rule-based algorithm has to be seen in the light of the number of wrongly explained and predicted fragments. This was shown by investigating a complex blank matrix. This shows that the “chopping” algorithm produces only few false negative findings. On the other hand, the output has to be further orthogonally filtered to reduce the number of false positive findings. Such “cleaning” can be obtained by detection selectivity (e.g., DIA like SWATH or IMS as well as by highly resolving liquid chromatography) but also by additional filter criteria (e.g., relative ion abundance etc.) or even including a more sophisticated in silico spectral interpretation of the remaining candidates. In addition, the use of generic (known) fragments of particular compound families may be utilized.

It was the aim of this paper to show the proof of principle that in silico fragmentation can explain and predict virtually every ionizable small molecule and thus be used to direct suspect screening results in an efficient way in routine analysis. This seems to be the case when a sufficiently wide collision energy ramp is used. A recently available paper [30] investigated ways to reduce the still significant number of reported false positive hits when analyzing low residue levels in complex matrices.

## Acknowledgments

The authors acknowledge the interesting discussions with Emma Schymanski, Eawag, Dübendorf, Switzerland, and the time she spent and effort made in discussing and proofreading of the manuscript.

## References

1. León, N., Pastor, A., Yusà, V.: Target analysis and retrospective screening of veterinary drugs, ergot alkaloids, plant toxins and other undesirable substances in feed using liquid chromatography-high resolution mass spectrometry. *Talanta*. **149**, 43–52 (2016)
2. Kaufmann, A., Butcher, P., Maden, K., Walker, S., Widmer, M.: Development of an improved high resolution mass spectrometry based multi-residue method for veterinary drugs in various food matrices. *Anal. Chim. Acta*. **700**, 86–94 (2011)
3. Wang, Z., Chang, Q., Kang, J., Cao, Y., Ge, N., Fan, C., Pang, G.-F.: Screening and identification strategy for 317 pesticides in fruits and vegetables by liquid chromatography-quadrupole time-of-flight high resolution mass spectrometry. *Anal. Methods*. **7**, 6385–6402 (2015)
4. Sapozhnikova, Y., Lehotay, S.: Multi-class, multi-residue analysis of pesticides, polychlorinated biphenyls, polycyclic aromatic hydrocarbons, polybrominated diphenyl esters, and novel flame retardants in fish using

- fast, low-pressure gas chromatography-tandem mass spectrometry. *Anal. Chim. Acta.* **758**, 80–92 (2013)
- Dührkop, K., Shen, H., Meusel, M., Rousu, J., Böcker, S.: Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *PNAS.* <https://doi.org/10.1073/pnas.1509788112>
  - Krauss, M., Singer, H., Hollender, J.: LC-high resolution MS in environmental analysis: from targeted screening to the identification of unknowns. *J. Anal. Bioanal. Chem.* **397**, 943–951 (2010)
  - Schymanski, E., Singer, H., Slobodnik, J., Ipolyi, I., Oswald, P., Krauss, M., Schulze, T., Haglung, P., Letzel, T., Grosse, S.: Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. *J. Anal. Bioanal. Chem.* **407**, 6237–6255 (2015)
  - Kaufmann, A., Butcher, P., Maden, K., Widmer, M.: Ultra-performance liquid chromatography coupled to time of flight mass spectrometry (UPLC-TOF): a tool for multiresidue screening of veterinary drugs in urine. *Anal. Chim. Acta.* **586**, 13–21 (2007)
  - Jeong, J.H., Lee, J.-H., Kim, H., Park, H., Hwang, I.-S., Han, K.-M., Yoon, C.-Y., Cho, S., Kim, W.-S.: LC-ESI-MS/MS analysis of phosphodiesterase-5 inhibitors and their analogues in foods and dietary supplements. *Food Addit. Contam. Part B.* **9**, 1–8 (2016)
  - Kind, T., Fiehn, O.: Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal. Rev.* **2**, 23–0 (2010)
  - Scheubert, K., Hufsky, F., Böcker, S.: Computational mass spectrometry for small molecules. *J. Cheminformatics.* **5**, 12 (2013)
  - Hufsky, F., Böcker, S.: Mining molecular structure databases: identification of small molecules based on fragmentation mass spectrometry data. *Mass Spectrom Rev.* (2016). <https://doi.org/10.1002/mas.21489>
  - Cai, T., Guo, Z.-Q., Xu, X.-Y., Wu, Z.-J.: Recent (2000–2015) developments in the analysis of minor unknown natural products based on characteristic fragment information using LC-MS. *Mass Spectrom. Rev.* (2016). <https://doi.org/10.1002/mas.21514>
  - Gago-Ferrero, P., Schymanski, E.-L., Hollender, J., Thomaidis, N.-S.: Nontarget analysis of environmental samples based on liquid chromatography coupled to high resolution mass spectrometry (LC-HRMS). *Compr. Anal. Chem.* Elsevier. <https://doi.org/10.1016/bs.coac.2016.01.012> (2016)
  - Yi, L., Dong, N., Yun, Y., Deng, B., Ren, D., Liu, S., Liang, Y.: Chemometric methods in data processing of mass spectrometry-based metabolomics: a review. *Anal. Chim. Acta.* **914**, 17–34 (2016)
  - Hufsky, F., Scheubert, K., Böcker, S.: Computational mass spectrometry for small-molecule fragmentation. *Trends Anal. Chem.* **53**, 41–48 (2014)
  - Wolf, S., Schmidt, S., Müller-Hausmann, M., Neumann, S.: In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinforma.* **11**, 148 (2010)
  - Hill, A.W., Mortishire-Smith, R.-J.: Automated assignment of high-resolution collisionally activated dissociation mass spectra using a systematic bond disconnection approach. *Rapid Commun. Mass Spectrom.* **19**, 3111–3118 (2005)
  - Ruttkies, C., Schymanski, E., Wolf, S., Hollender, J., Neumann, S.: MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J. Cheminformatics.* **8**, 3 (2016)
  - Schymanski, E., Gallapois, C., Krauss, M., Meringer, M., Neumann, S., Schulze, S., Wolf, S., Brack, W.: Consensus structure elucidation combining GC/EI-MS, structure generation, and calculated properties. *Anal. Chem.* **84**, 3287–3295 (2012)
  - Hufsky, F., Scheubert, K., Böcker, S.: New kids on the block: novel informatics methods for natural product discovery. *Nat. Prod. Rep.* **31**, 807–817 (2014)
  - Hill, D.-W., Kertesz, T., Fontaine, D., Friedmann, R., Grant, D.-F.: Mass spectral metabonomics beyond elemental formula: chemical database querying by matching experimental with computational fragmentation spectra. *Anal. Chem.* **80**, 5574–5582 (2008)
  - Woldegebriel, M., Vivó-Truyols, G.: A new Bayesian approach for estimating the presence of a suspected compound in routine screening analysis. *Anal. Chem.* **88**, 9843–9849 (2016)
  - Gago-Ferrero, P., Schymanski, E.-L., Bletsou, A.-A., Aalizadeh, R., Hollender, J., Thomaidis, N.-S.: Extended suspect and non-target strategies to characterize emerging polar organic contaminants in raw wastewater with LC-HRMS/MS. *Environ. Sci. Technol.* **49**, 12333–12341 (2015)
  - Schymanski, E., Meringer, M., Brack, W.: Matching structure to mass spectra using fragmentation patterns: are the results as good as they look? *Anal. Chem.* **81**, 3608–3617 (2009)
  - Gehrig, T., Rushing, L., Churchwell, M., Doerge, D., McErlane, K., Thompson, H.: HPLC determination of sulfadiazine residues in Coho salmon (*Oncorhynchus kisutch*) with confirmation by liquid chromatography with atmospheric pressure chemical ionization mass spectrometry. *J. Agric. Food Chem.* **44**, 3164–3169 (1996)
  - Thurman, E.-M., Ferrer, I., Benotti, M., Heine, C.-E.: Intramolecular isobaric fragmentation: a curiosity of accurate mass analysis of sulfadimethoxine in pond water. *Anal. Chem.* **76**, 1228–1235 (2004)
  - Barry, S., Wolff, J.-C.: Identification of isobaric amino-sulfonamides without prior separation. *Rapid Commun. Mass Spectrom.* **26**, 419–429 (2012)
  - Tsugawa, H., Kind, T., Nakabayashi, R., Yukihiro, D., Tanaka, W., Cajka, T., Saito, K., Fiehn, O., Arita, M.: Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidating using MS-FINDER software. *Anal. Chem.* **88**, 7946–7958 (2016)
  - Kaufmann, A., Butcher, P., Maden, K., Walker, S., Widmer, M.: Practical application of in silico fragmentation based residue screening with ion mobility high-resolution mass spectrometry. *Rapid Commun. Mass Spectrom.* **31**, 1147–1157 (2017)