International Journal of Mass Spectrometry 299 (2011) 158–168

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Interpretation of ESI(+)-MS-MS spectra—Towards the identification of "unknowns"

Avi Weissberg*, Shai Dagan

Israel Institute for Biological Research (IIBR) P.O.B. 19, Ness Ziona, Israel

ARTICLE INFO

Article history:
Received 18 July 2010
Received in revised form 17 October 2010
Accepted 20 October 2010
Available online 28 October 2010

Keywords:
Unknown
Fragmentation rules
ESI(+)-MS-MS
EI-MS
Proposed structures

ABSTRACT

We report a systematic empirical investigation of ESI(+)-MS-MS dissociation pathways of over 1000 spectra of small organic compounds, containing more than 30 chemical functional groups. The dissociation processes of the protonated molecular ions were explored and interpretated. We derived typical basic fragmentation channels for individual functional groups and established a unified set of fragmentation rules. Multiple bond cleavages of molecules containing single and multiple functional groups were explored as well and the corresponding fragmentation rules were derived. Applying these rules enabled to match between proposed chemical structures and an ESI(+)-MS-MS spectrum of an "unknown". Comparison to EI fragmentation routes was also carried out. Despite the general dissimilarity between ESI(+)-MS-MS and EI-MS spectra, we exploit the minor similarities between the spectra, and utilizing NIST-EI database and search option, can be successfully reduced the number of proposed structures. The two step methodology developed here is demonstrated and evaluated in the identification of various "unknowns".

1. Introduction

The identification of an "unknown" is a crucial task difficult to tackle, due to lack of authentic standards and scarce information available. Volatile and semi-volatile compounds are commonly identified by comparison of electron ionization (EI) spectra with reference databases. EI-MS libraries have become large, increasing their effectiveness for identifying "unknowns". For instance, The National Institute of Standards and Technology library (NIST 08) incorporates 220,460 spectra of 192,108 different compounds.

On the other hand, non-volatile, thermally instable, or highly polar compounds are usually not compatible with gas chromatography and widely determined by high performance liquid chromatography (HPLC) utilizing electrospray ionization mass spectrometry (ESI-MS) [1].

ESI-MS-MS produces informative spectra, however, databases based on ESI-MS-MS are still relatively small (for example, 14,802 spectra of only 5,308 compounds in NIST 2008). Therefore, other approaches are required to identify unexpected compounds analyzed by ESI-MS-MS.

The most common approach to determine the structure of an unknown small organic molecule uses liquid chromatography electrospray time-of-flight mass spectrometry (LC–QTOF-MS) for accurate mass measurement of the molecular ions of interest followed by searching databases such as Chemindex-(77,000 compounds), Merck index (on CD-10,000 compounds), NIST 08 library-(192,108 compounds), or even Scifinder (millions of compounds) for possible structures. ESI-MS-MS data of the unknown is then processed by fragment ion identification using a chemical drawing software and comparison with accurate-mass ion fragments or making tentative identification in order to establish relationships between fragments and the molecular ion [2–6]. The final step is verification with authentic standards, if available. Reconstruction of the parent compound by determination of the structure of each of the individual fragment ions following collision-induced dissociation (CID) is usually a difficult process because multiple structures can be constructed for the same monoisotopic molecular weight (MIMW) and not all portions of the original structure are usually exhibited as fragments in the CID mass spectra.

General rules for mass spectral fragmentation of mostly electron-impact-produced ions have been established [7]. Various algorithms that evaluate the structures of organic compounds and predict mass spectral fragments based on these rules have been developed (Mass Frontier, Mass Fragmenter, Advanced Chemistry Development, EPIC, MOLGEN-MS [8–11]). The limitations of some of the commercial software packages have been addressed by Schymanski et al. [12] in a comparative study of EI fragmentation.

An important contribution to mass spectral simulations was reported by Gasteiger et al. The system MASSIMO [13] developed by

* Corresponding author. Tel.: +972 8 9381687; fax: +97289381688.
E-mail addresses: aviwe@iibr.gov.il (A. Weissberg), shaid@iibr.gov.il (S. Dagan).

1387-3806/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2010.10.024

him, is capable of predicting the mass spectra for given molecular structures. In addition, a mass spectrum simulation system (MASSIS) [14], based on reduced and concentrated knowledge databases extracted from the literature and obtained from data mining of spectral databases was developed. A software, based on *ab initio* structural identification of product ions from tandem mass spectrometric data was also developed. This software "Fragment Identificator" (FID) [15] conducts a combinatorial search over all possible fragmentation paths and outputs a ranked list of alternative structures.

One of the major limitations of structure elucidation of small organic molecules originates from the limited understanding of the fragmentation rules in ESI-MS-MS of protonated molecules, even though the structure–fragmentation relation and fragmentation mechanisms have been widely studied. Niessen illustrated the four-center fragmentation mechanism for ethers, amines, and amides [16]. Smith studied and reviewed the ESI-MS-MS behavior of selected drugs with amine-containing side chains, drugs with N-containing saturated ring structures, drugs with N-containing unsaturated ring structures and quaternary ammonium drugs [17,18]. A study of the fragmentation and ion formation of three major families of pesticides (herbicides, insecticides and fungicides) as well as emerging contaminants in the environmental field was carried out mainly by Thurman et al. [19–21]. Structure elucidation for aromatic compounds was studied by Levsen et al. [22].

Although there have been several publications addressing ESI-MS-MS of specific structures and functional groups, no comprehensive and general rule set for the fragmentation of even electron ions has been generated yet.

This led us to systematically study ESI(+)-MS-MS fragmentation spectra of a broad range of molecules with various functional groups with a view to establishing a set of general fragmentation rules that can be used for characterization of unknowns. More than 30 major chemical families (amines, carbamates, phosphates, etc.) contained in over 1000 ESI-MS-MS library spectra were explored. The dissociation processes for the protonated molecular ions were elucidated, supported by proposed mechanisms and compared to typical EI fragmentation routes. A comprehensive fragmentation rule database, with over 60 channels is reported herein. In addition, we studied the fragmentation behavior of molecules bearing multiple functional groups and their mutual effects on the fragmentation. Applying these rules enabled us to match between proposed chemical structures and ESI-MS-MS degradation pathways which is the main step in the identification of an unknown.

In case where too many structures are proposed for the potential empirical formulas, we successfully reduced the number by utilizing the large NIST-EI database with its constrains functionality, despite the well known dissimilarity between ESI/MS/MS and EI/MS spectra [23].

Our proposed process for the identification of unknowns includes the following steps:

- Detection of a possible unknown, using accurate mass measurement generating possible empirical formulas, as well as performing an MS/MS experiment.
- Searching databases for possible structures matching the suggested empirical formula(s).
- Data reduction, minimizing the number of proposed structures.
- Applying the ESI-MS-MS fragmentation rules to each of the proposed structures and predicting their fragmentation routes.
- Comparing between the predicted fragmentation of the suspected structure and the measured ESI-MS-MS spectrum of the unknown.
- Ranking the structure by the number of matching fragments.

2. Materials and methods

ESI(+)-MS-MS data were retrieved from three major commercial databases:

- NIST/EPA/NIH Mass Spectral library, 2005 (5191 entries).
- The spectra in this database originate from various instruments. For most of the compounds, multiple spectra at several CID energies are available.
- BFR pesticide database [24] (1368 entries). The spectra in this database originate from triple quadrupole instruments.
- QSTAR drug database-ABI (Sciex) (1323 entries). The spectra in this database originate from QTrap instruments.

EI-MS data was retrieved from NIST 2005 (190,825 entries) and 2008 (220,460 entries) libraries.

More than 30 functional chemical groups were defined, and spectra of more than 1000 molecules containing these groups were chosen and manually processed. All the fragments in each spectrum were examined and served for building the rule tables presented here. In NIST database, fragmentation was processed for spectra with various CID energies available per compound so that a full fragmentation pattern could be studied.

3. Results and discussion

3.1. ESI fragmentation rules of selected functional groups

We have explored ESI-MS-MS dissociation processes of the protonated molecular ions of various chemical families and constructed a set of fragmentation rules. The entire fragmentation rule database is presented herein, in Table 1, divided into 25 sub-tables for each functional group or several groups containing similar structures. The most left column represents the rule number, the second describes the functional group (the cleavage position is marked by a dashed line), following is the fragment mass, then the observed EI fragment, and remarks. In the last column, several compounds that exemplify the rules are mentioned. The functional groups examined included amines, halides, acids, amides, carbamates, ketones, ethers, phosphates, phosphonates, sulfides, sulfones and many more. In the following paragraphs we discuss several representative functional groups and their ESI-MS-MS dissociation processes.

3.1.1. Amines

The ESI-MS-MS spectra of more than 200 compounds containing amines (primary, secondary, tertiary, quaternary and cyclic) were explored and 3 fragmentation channels involving the amine group were characterized.

3.1.1.1. Amines—Rule A (3 in Table 1). Molecules containing an amine group with (at least) an ethylene or substituted ethylene group separating the nitrogen atom from other functional groups, will cleave and lose the nitrogen atom as the corresponding amine, and will form the deaminated ion. Smyth reported this rule for drugs with amine-containing side chains [17,18]. Protonation occurs at the nitrogen followed by C–N cleavage. Thus, the calculated fragment mass will simply be the molecular weight minus the mass of the amine leaving group. This is exemplified by the ESI-MS-MS behavior of imipramine in which such cleavage occurs in both nitrogens, and peaks at m/z 238 and 86 are observed.

3.1.1.2. Amines—Rule B (4 in Table 1). This rule applies mainly to secondary amines. We found that the cleavage occurs at a carbon atom attached to a nitrogen, (similar to amine rule 3), but the charge is now located on the amine moiety which possesses 2 addi-

Table 1
Database of the ESI fragmentation rules for the various functional groups. Columns from left to right: chemical structure, calculation of the derived fragment ion mass, the observed EI analogue (if available), remarks and examples of compounds that obey the fragmentation rule.

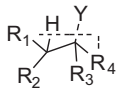
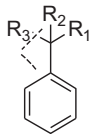
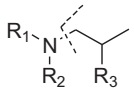
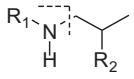
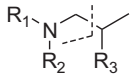
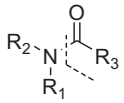
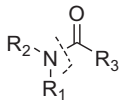
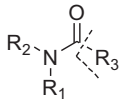
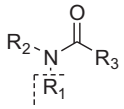
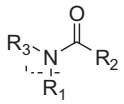
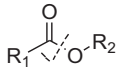
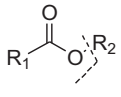
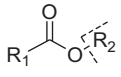
Rule	Scheme	ESI fragment mass	EI analogue	Remarks	Examples
Elimination					
1		Common, occurs mostly prior to the fragmentation process $F^+ = MW + 1\text{-}HY$	Much less common	Y = OH, Cl, Br ESI and EI spectra are completely different	Metoprolol[N] Fenoterol[N] Fluoxetine[N] Trichlorfon[B] Procyclidine[N]
Formation of stabilized cations					
2		Common $F^+ = MW - LG(R_3) +$ $F^+ = MW - LG(R_3) - LG(R_2) + 1$	Common		Verapamil[N] Imazalil[B] Diphenylhydramine[N] Fendiline[N] Proadifen[N]
Amines, hydrazines					
3		$F^+ = MW - LG$	Uncommon	Spacer of at least two methylene groups between nitrogen and the attached functional group	Chloroquine[N] Moclobemide[N] Benzedrex[N+Q] Phenalzine[N] Selegiline[N] Imipramine[N]
4		$F^+ = MW - LG+2$	Uncommon	See remark 3	Fenoterol[N] Chloroquine[N] Fendiline[N]
5		$F^+ = MW - LG$	Common		Lidocaine[N+Q]
Amides					
6		$F^+ = MW - LG$	EI = ESI EI = ESI-2	Followed by CO release where $F^+ = R_3$ (*a* ion in peptides)	Carboxin[B] LSD[N], *b* ion in peptides
7		$F^+ = MW - LG+2$	EI = ESI-1 EI = ESI-2		Phenacetin[N] Capsaicin[N] *y* ion in peptides
8		$F^+ = MW - LG$	EI = ESI		Bupivacaine[N+Q] Capsaicin[N] Ropivacaine[Q]
9		$F^+ = MW - LG$			Zolpidem[N]
10		$F^+ = MW - LG+2$		R_1 = Alkyl > Me	
Esters					
11		$F^+ = MW - LG$	EI = ESI	Common	Procaine[N] Ecgonine methyl ester[N]
12		$F^+ = MW - LG+2$	EI = ESI	R_2 = Alkyl > Me Followed by loss of H_2O	Proline butyl ester[N]
13		$F^+ = MW - LG$	EI = ESI EI = ESI-1		Cocaine[N] Procaine[N]

Table 1 (*Continued*)

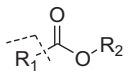
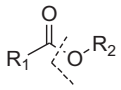
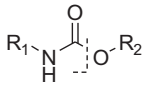
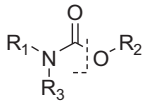
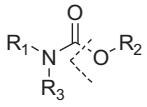
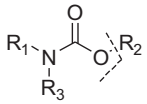
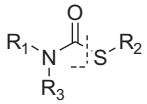
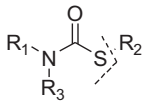
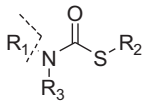
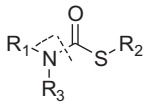
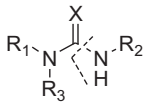
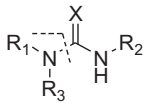
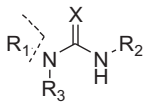
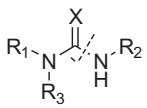
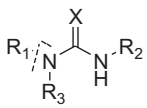
Rule	Scheme	ESI fragment mass	EI analogue	Remarks	Examples
14		$F^+ = MW - LG$	EI = ESI		Meperidine[N]
15		$F^+ = MW - LG+2$	EI = ESI-1 EI = ESI-2	R_2 = Contains an amine	Scopolamin[Q] Oxybutynin[Q]
Carbamates					
16A		$F^+ = MW - LG$	EI = ESI-1		Albendazole[N] Carbedazim[N]
16B		$F^+ = MW - LG$	EI = ESI	$R_1, R_3 \neq H$	Pirimicarb[B] Isoproturon[N]
17		$F^+ = MW - LG+2$	EI = ESI-1		Carbaryl[N+B] Propoxur[N+B] Carbofuran[N+B]
18		$F^+ = MW - LG+2$	EI = ESI-1	R_2 = Alkyl > Me	Isopropyl-N-phenyl-carbamate[N]
Thiocarbamates					
19		$F^+ = MW - LG$	EI = ESI	Like rule 16A	Butylate[N+B]
20		$F^+ = MW - LG+2$	EI = ESI	R_2 = Alkyl > Me, like rule 18	Butylate[N+B] Molinate[B] Dimepiperate[N]
21		$F^+ = MW - LG$	EI = ESI		Butylate[N+B]
22		$F^+ = MW - LG$	EI = ESI		Molinate[B]
Ureas, guanidines, thiourea					
23		$F^+ = MW - LG$	EI = ESI	$R_1, R_3 \neq H$	Diuron[N+B] Diafenthiuron[N]
24		$F^+ = MW - LG+2$	EI = ESI-1		Diafenthiuron[N] Tebuthiuron[N]
25		$F^+ = MW - LG$		X = NH (Guanidine), O (Urea), S (Thiourea)	Diafenthiuron[N]
26		$F^+ = MW - LG$	EI = ESI		Diuron[N+B]
27		$F^+ = MW - LG+2$		R_1 = Alkyl > Me	Siduron[B]

Table 1 (*Continued*)

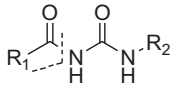
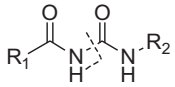
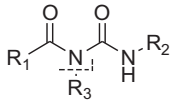
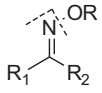
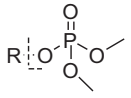
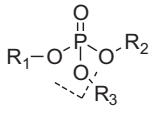
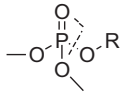
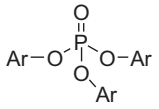
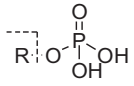
Rule	Scheme	ESI fragment mass	EI analogue	Remarks	Examples
Acyl ureas					
28		$F^+ = MW - LG$	EI = ESI		Difluron[N]
29		$F^+ = MW - LG+2$			Difluron[N]
30		$F^+ = MW - LG+2$	EI = ESI	R_3 = Alkyl > Me	Bromazil[N+B]
Ketones					
31		$F^+ = MW - LG$	EI = ESI	Most common	Haloperidol[N] Lobeline[N]
32		$F^+ = MW - LG$	EI = ESI		
Ethers					
33		$F^+ = MW - LG$	EI = ESI		Alachlor[B] Viloxazine[N] Clemastine[N]
34		$F^+ = MW - LG+2$	EI = ESI-2 EI = ESI	R_1 Contains an amine, R_2 does not contain an amine and $R_2 \neq$ Me	Clemastine[N] Cinchocaine[N] Pyriproxifen[B]
35		$F^+ = MW - LG$		R_2 Contains any nitrogen group	Pyriproxen[B] Terconazole[B] Fenoxanil[B]
Oxime-ethers					
36		$F^+ = MW - LG$			Prednisolone,3,20 bisethoximes[N] 2-Hydroxyestrone-3-methylethermethyloxime[N]
Phosphates					
37A		$F^+ = MW - LG+2$	EI = ESI	R > Me	Crotoxyphos[B] Phosphamidon[B] Naled[B]
37B		$F^+ = MW - LG+2$	EI = ESI	R_{1-3} = alkyl > Me, Multiple cleavages occur, $F^+ = MW - LG+1$ + number of cleavage positions	TEPP[B]
38		$F^+ = MW - LG$	EI = ESI	In addition, loss of MeOH	Naled[B] Dichlorovos[B]
39				Apply rules 37, 38	Tritolyl phosphate[B]
40		$F^+ = MW + 1 - 98$		H_3PO_4 loss	Phosphoserine-containing peptides, Pyridoxal-5 phosphate[N]

Table 1 (*Continued*)

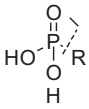
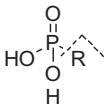
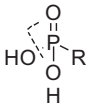
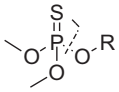
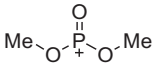
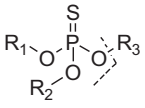
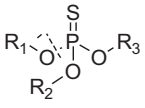
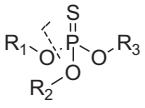
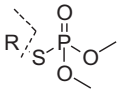
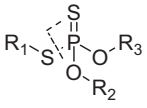
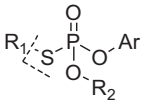
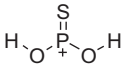
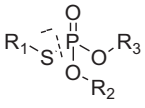
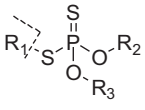
Rule	Scheme	ESI fragment mass	EI analogue	Remarks	Examples
Phosphonic acids					
41		F^+ = MW – LG+2		Followed by loss of H_2O	Ethylmethylphosphonic acid, F^+ = 97
42		F^+ = MW – LG+2		R > Me	Ethylphosphonic acid[N]
43		F^+ = MW – LG			Propylphosphonic acid[N]
44		F^+ = MW – LG		Loss of H_2O	Methylphosphonic acid[N]
Thiophosphate A					
45		F^+ = MW – LG	ESI = EI	R = Ar, Alkyl (like rule 38) ESI Common fragment is 	Etrimfos[B] Fenitrothion[B]
46		F^+ = MW – LG+2	ESI = EI Common fragment: 125, 109, 97, 81	R_{1-3} = alkyl > Me, Multiple cleavages occur, F^+ = MW – LG+1 + number of cleavage positions	Coumaphos[B] Diazinon[B]
47		F^+ = MW – LG+2	EI = ESI-1	ESI Common fragment: $R_1SH_2^+$	Triazophos[B] Quinalphos[B]
48		F^+ = MW – LG	EI = ESI-1	See remark 46	Triazophos[B]
Thiophosphate B					
49		F^+ = MW – LG	ESI = EI	Common EI fragments 125–  and 109	Azamethiphos[B] Vamidothion[B] Demeton-S-methyl[B]
50		F^+ = MW – LG+2	ESI = EI	See remark 46. Common EI fragments 97–  and 125	Profenofos[B]
51		F^+ = MW – LG	EI = ESI		
Dithiophosphates					
52		F^+ = MW – LG	ESI = EI	Like rule 49	Phosmet[B] Methidathion[B] Dioxathion[B] Ethion[B]

Table 1 (*Continued*)

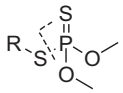
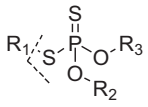
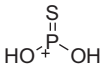
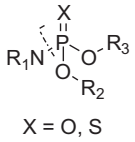
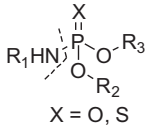
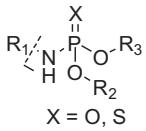
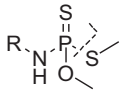
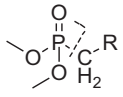
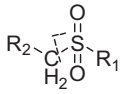
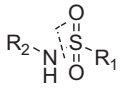
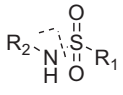
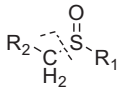
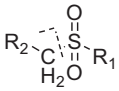
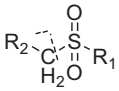
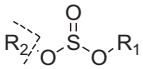
Rule	Scheme	ESI fragment mass	EI analogue	Remarks	Examples
53		$F^+ = MW - LG$	ESI = EI	Like rule 38	Dimethoate[B]
54		$F^+ = MW - LG$	EI = ESI	See remark 46. Common fragment is 	Phosalone[B] Azinphos-ethyl[B]
Aminophosphates/aminothiophosphates					
55		$F^+ = MW - LG$	ESI = EI	See remark 46	Isofenphos[B] Fenamiphos[B] Propetamphos[B]
56		$F^+ = MW - LG+2$	ESI = EI	$F^+ = MW - LG+2$	Fenamiphos[B]
57		$F^+ = MW - LG+2$	EI = ESI	In addition see remark 46	Fenamiphos[B]
Aminodithiophosphate					
58		$F^+ = MW - LG$	ESI = EI	P–N, P–O cleavages. In addition see remark 46.	Methamidophos[B]
Phosphonates					
59		$F^+ = MW - LG$	ESI = EI		Trichlorfon[B]
Thiophosphonates					
60		$F^+ = MW - LG$		 Common ESI fragment is in addition see remark 46	EPN[B] Cyanofenfos[B]
Sulfides					
61		$F^+ = MW - LG$			Ethiofencarb[B]
62		$F^+ = MW - LG+2$			Cimetidine[N]
Sulfoneamide/aminosulphate					
63		$F^+ = MW - LG$	EI = ESI	Followed by SO_2 release, $F^+ = R_1$	Sulfisoxazole[N+Q]
64		$F^+ = MW - LG+2$			Sulfisoxazole[N+Q]
Sulfoxides					
65		$F^+ = MW - LG$	EI = ESI-1		Sulfinpyrazone[N+Q]
Sulfones					

Table 1 (*Continued*)

Rule	Scheme	ESI fragment mass	EI analogue	Remarks	Examples
66		F^+ = MW – LG			Decanenitrile, 10-(methyl sulfonyl)[N]
67		F^+ = MW – LG+2		Followed by SO_2 release, $F^+ = R_1$	1H-benzimidazol-2-amine, 5-(propylsulfonyl)-[N]
Sulfites					
68		F^+ = MW – LG	EI = ESI		Aramite[B]

LG = leaving group; LG+2 = leaving group plus two mass units; MW = molecular weight (non-protonated); F = fragment; Y = OH or Cl or Br; Ar = aryl; ESI = electrospray ionization; EI = electron impact; [N] = spectrum retrieved from NIST library; [B] = spectrum retrieved from BFR database; [Q] = spectrum retrieved from QSTAR library.

tional mass units. This is termed the "+2" rule. Again, protonation on the amine is followed by C–N cleavage, forming a free amine and the deaminated ion. Then, proton transfer occurs to form the protonated amine and an olefin. This is exemplified in Fig. 1 by the ESI-MS-MS behavior of fendiline when a peak at *m*/*z* 212 that obeys rule 4 was observed, in addition to the peak at *m*/*z* 105 which obeys rule 3. The "+2" rule may obey a four-center fragmentation mechanism that has been proposed for several compound classes such as ethers, amides, etc. [15].

3.1.1.3. Amine—Rule C (5 in Table 1). If there is only one methylene attached to the amine (primary, secondary, tertiary, quaternary or cyclic), the molecule is most likely to cleave at the carbon, located at β position to the nitrogen, whereas the charge is located on the amine. This is exemplified by the ESI-MS-MS behavior of lidocaine where the peak at *m*/*z* 86 obeys rule 5. This type of cleavage is very common in EI-MS of amines.

3.1.2. Amides

The ESI-MS-MS spectra of a few dozen compounds containing amides were explored. 5 fragmentation channels involving the amide group were characterized and two of them are described here.

3.1.2.1. Amides—Rule A (6 in Table 1). The most commonly observed CID products of protonated amides are formed by cleavage at the amide bond with the charge remaining on the carbonyl group. The fragment ion mass is the molecular weight minus the leaving group mass. Protonation occurs at the nitrogen, followed by neutral loss (R_1NH_2 for alkylamides and R_1NHR_2 for dialkylamides). The observed acyl ion tends to release CO. This is exemplified by the ESI-MS-MS behavior of LSD in which both cleavages occur and peaks at *m*/*z* 251 and 223 are observed. This cleavage is dominant in peptides, leading to the well known formation of *b* and *a* ions correspondingly [25].

3.1.2.2. Amides—Rule B (7 in Table 1). In some cases, cleavage still occurs at the C–N bond, but the charge is now located on the amine moiety which possesses 2 additional mass units ("+2" rule). Again, protonation on the amine is followed by C–N cleavage, forming a free amine and the acyl ion. Then, proton transfer occurs to form the protonated amine. This is exemplified by the ESI-MS-MS behavior of phenacetin where a peak at *m*/*z* 138 obeys this rule. This cleavage is common in peptides, leading to the formation of *y* ion [25].

3.1.3. Carbamates

The ESI-MS-MS spectra of a few dozen compounds containing carbamates were explored.

3.1.3.1. Carbamates—Rule A (16A in Table 1). Secondary and tertiary carbamates rarely dissociate at the C–N bond and are most likely to dissociate at the C–O bond with the charge remaining on the carbonyl group. The fragment ion mass is the molecular weight minus the leaving group mass. Protonation occurs at the ether oxygen, followed by an uncharged alcohol release. This is exemplified in Fig. 2 by the ESI-MS-MS behavior of albendazole (secondary carbamate) where a peak at *m*/*z* 234 obeys this rule. In EI-MS, when cleavage occurs at the same position, the observed fragment mass ion is lower by one mass unit. However, the fragment mass of tertiary carbamates in EI-MS is identical to ESI-MS-MS.

3.1.3.2. Carbamates—Rule B (17 in Table 1). In some cases, secondary carbamates dissociate at the C–O bond but the charge is located on the alcoholic moiety. This fragment mass is higher by two mass units, and the fragment mass will be the molecular weight minus the leaving group mass plus two mass units ("+2" rule). The same cleavage occurs as in rule 16A, followed by proton transfer to form the hydronium ion. Again, the fragment mass in EI-MS is lower by one mass unit. This is exemplified by the ESI-MS-MS behavior of carbofuran where the *m*/*z* 165 peak obeys this rule.

3.1.4. Phosphates

3.1.4.1. Phosphate—Rule A (37A in Table 1). Phosphates are most likely to dissociate retaining the phosphate moiety, thus C–O cleavage is favored over P–O, and the charge is stabilized on the phosphate group. The fragment ion mass is the molecular weight

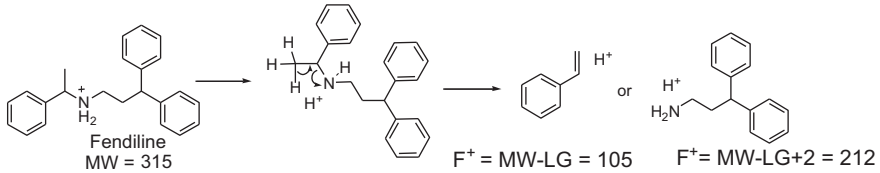


Fig. 1. Demonstration of the "+2" rule in fendiline (an amine, rule 4 in Table 1) assuming a four-center fragmentation mechanism. The fragment ion (F^+ = 212) obtained following C–N cleavage, contains 2 additional mass units of one proton from the system and one hydrogen transferred in the ionization process.

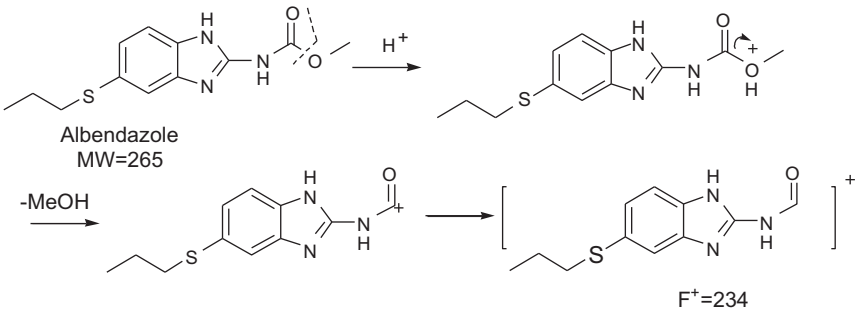


Fig. 2. A carbamate rule (rule 16A in Table 1) demonstrated by the ESI-MS-MS behavior of Albendazole. The C–O bond is dissociated (MeOH release) and the fragment ion obtained is the molecular weight minus the leaving group.

minus the leaving group mass plus two mass units ("+2" rule). The first proton originates from the system and the second from proton transfer. This is exemplified by the ESI-MS-MS behavior of Naled where the peak at *m*/*z* 127 obeys this rule.

3.1.4.2. Phosphate—Rule B (38 in Table 1). In some cases, P–O cleavage of phosphates is observed and the charge is stabilized on the phosphorus moiety. The fragment ion mass is the molecular weight minus the leaving group mass. This is exemplified by the ESI-MS-MS behavior of dichlorovos where the *m*/*z* 109 ion obeys this rule. EI-MS behavior was found to be identical to ESI-MS-MS.

In Table 1 rules are presented for phosphates, thiophosphates and dithiophosphates. Analogies between thiophosphate and dithiophosphate rules to phosphate rules are indicated.

3.1.5. Phosphonic acids

3.1.5.1. Phosphonic acids—Rule A (41 in Table 1). Phosphonic acids are most likely to dissociate at the O–alkyl bond in case it is larger than methyl. The charge is stabilized on the phosphonate group. The fragment mass is the molecular weight minus the leaving group mass plus two mass units ("+2" rule). This is exemplified by the ESI-MS-MS behavior of ethyl methylphosphonic acid (EMPA), where the peak at (*m*/*z* 97) obeys this rule. This fragment ion tends to further lose water to form a peak at (*m*/*z* 79).

3.2. Generalization of fragmentation rules—the "+2" rule

From this study, some general rules can be formulated with respect to ESI-MS-MS cleavages of heteroatom-carbon bonds, when the charge is stabilized on the heteroatom cleaved moiety. In many cases, no proton transfer occurs and the fragment ion mass is simply the molecular weight minus the leaving group. However, in some cases the fragment ion mass contains additional two mass units. This phenomenon strongly depends on the nature of both the carbon containing group and the heteroatom residue and is complicated to predict. Based on all the "+2" observations we have come upon, we established a general rule as follows:

1. When an electronegative heteroatom (O, N, S), contained in any functional group bound to an alkyl chain (N–R, O–R, S–R) is cleaved and the charge is stabilized on the heteroatom moiety, the fragment mass is the molecular weight minus the leaving group mass plus two mass units.
2. The rule above is valid also for N–R′ bonds (when R′ = COAlkyl-amides, R′ = CONHAlkyl-ureas) and O–R′ bonds (when R′ = CONHAlkyl-carbamates).

It should be noted, that when the bond between an electronegative heteroatom (O, N, S) and electropositive atom (C, P) is cleaved and charge is stabilized on the electropositive moiety (C, P), the fragment mass is simply the molecular weight minus the leaving group mass.

3.3. Multiple cleavages

Sequential product ion fragmentations occur often, which may provide further useful information on the structure of the unknown molecule. Therefore, we explored multiple cleavages behavior and present here the rules assigned.

3.3.1. Rule MC1 (multiple cleavages)

For a molecule that contains two C–X bonds (X = O, N), where the corresponding fragment ion formed contains the two X atoms and each cleavage obeys the "+2" rule, the product ion will have the mass: fragment (F^+) = molecular weight (MW) – leaving group 1 (LG1) – leaving group 2 (LG2) + 3 mass units. In case there are *n* cleaved bonds, the product ion will have the mass: fragment (F^+) = molecular weight (MW) – leaving group 1 (LG1) – leaving group 2 (LG2) – (LG*n*) + (*n* + 1) mass units.

3.3.2. Rule MC2 (multiple cleavages)

For a molecule that contains two C–X bonds (X = O, N) or a C–N bond and a C–C bond and the corresponding formed fragment ion contains only one X atom which obeys the "+2" rule, the product ion will have the mass: fragment (F^+) = molecular weight (MW) – leaving group 1 (LG1) – leaving group 2 (LG2) + 1 mass unit.

We demonstrate the use of the above fragmentation rules with N-acetyltyrosine, butyl ester spectrum retrieved from NIST 08 MS-MS database (illustrated in Fig. 3).

This compound contains three different functional groups (ester, amide and benzyl moiety). Applying the esters rules (11–15), the amides rules (6–10) and the benzyl rule (2), we find 4 fragments matching our predictions. Three esters rules (fragments in *m*/*z* 225, 206, 178), and one amide rule (fragment in *m*/*z* 238) are obeyed. In addition, multiple cleavages are observed which obey rule MC 1 (*m*/*z* 182) and (*m*/*z* 164, 136) fragments that obey rule MC 2. Thus, in this case, interpretation is provided for all 7 fragments (a mechanism is proposed in supplementary Information 1).

3.4. Data reduction

Applying the fragmentation rules for a potential empirical formula in case too many structure candidates are proposed, is time consuming, especially when the prediction and processing is done manually. Thus, we utilized the NIST-EI library, which contains a large mass spectral database of small molecules (192,108 compounds), to reduce the number of proposed structures prior to applying the fragmentation rules. Using the NIST-EI library enabled us to compare between ESI-MS-MS and EI-MS fragmentation routes. Despite the general dissimilarity between ESI-MS-MS and EI-MS, one can take advantage of the minor similar parts. We

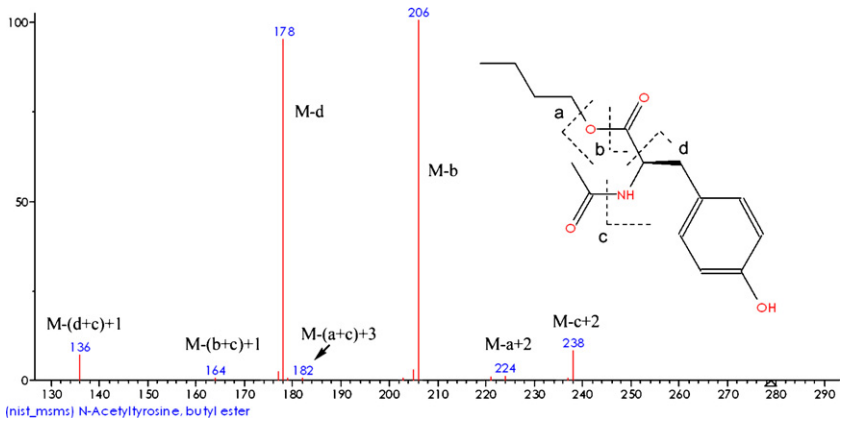


Fig. 3. Multiple cleavages (MC). Demonstration of the fragmentation in N-acetyltyrosine, butyl ester (NIST number 1007002) that contains three different functional groups. Three ester rules (fragments at *m*/*z* 224, 206, 178), and one amide rule (fragment at *m*/*z* 238) are shown (each is a single cleavage). In addition, double cleavages are obtained which obey rule MC 1 (fragment at *m*/*z* 182) and rule MC 2 (fragments at *m*/*z* 164, 136). Thus, interpretation is provided for all 7 fragments.

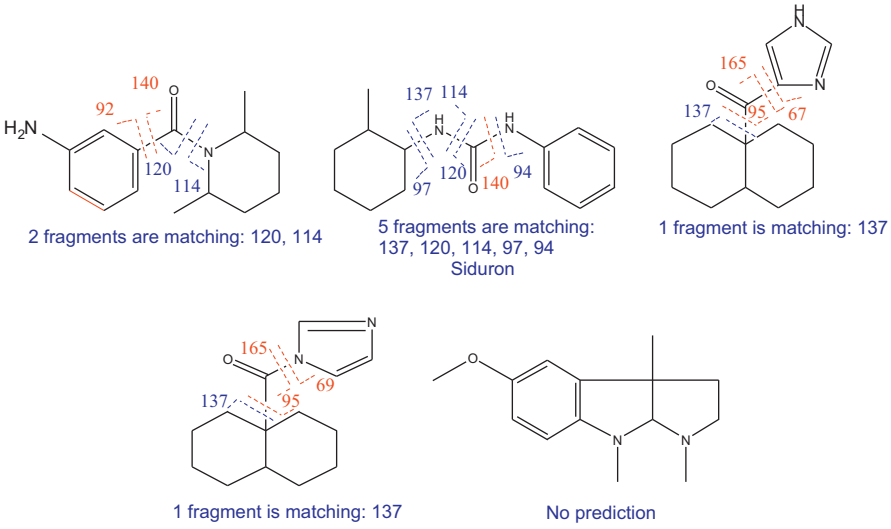


Fig. 4. Proposed structures retrieved from NIST-EI 2005 database for $C_{14}H_{20}N_2O$, after data reduction from 24 to only 7 candidates. 5 out of those 7 structures are presented. Applying the fragmentation rules, followed by comparison between the predicted fragmentation of the proposed structures and the ESI-MS-MS spectrum, reveals only one compound (siduron) with 5 fragments (based on urea rules) that fit the spectrum, where in other structures no more than 2 predicted fragments fit the spectrum.

noticed that when dissociation occurs at the same position in EI-MS and ESI-MS-MS, the fragment ion mass observed with EI was identical or lower by one mass unit than with ESI, ensuring that either the EI fragment or its (M + 1) isotopic peak will fit the ESI fragment. This was observed by Levsen et al. as well in its systematic study of the neutral species lost [22]. We applied this observation and used the NIST-EI-MS database with its "constrains" search option to reduce the proposed structure list, based on requiring (only) >1% abundance of ESI-MS-MS ions at the range of *m*/*z* > 90 in the EI-MS database spectra of a proposed structure. The search was for all ions with *m*/*z* > 90, to avoid false matches at the very crowded low mass range of EI. The candidate list was then ranked by the number of similar fragments to EI, and in most cases (>80%) the correct structure was ranked at the top of the list. The list of proposed structures can now be reduced by seeking only these structures with an EI-MS spectrum having similar fragments. The reduced structure list should next be examined by the fragmentation rules. For example, EI-MS and ESI-MS-MS spectra of procaine are substantially different although there are several identical fragments, some of them are isotopic peaks in the EI spectrum. In this particular case ESI-MS-MS peaks at 164, 120, 100 *m*/*z* were examined. By searching the NIST-EI database, we successfully reduced the list by a factor of 5, which means that out of 20 candidates, only 4 compounds contained all three requested fragments and procaine was one of them. Although over 1% abundance is considered to be a very modest threshold level, it appeared to be sufficient for effective data reduction, while maintaining the correct structure in the list.

3.5. The entire process—an example

The following example demonstrates the entire process. Given an "unknown" ESI-MS-MS spectrum and a proposed structure list of isomers of the same empirical formula, the list is first reduced by comparison to NIST-EI database. The remaining structures are then analyzed to predict fragmentation and compared to the ESI-MS-MS data of the unknown.

Assuming that the *m*/*z* value of the protonated molecular ion in the ESI-MS-MS spectrum is 233.158, combined with isotope ratios measured, a few (sometimes only one) molecular formulas can be proposed. To demonstrate the entire process we chose for example one of the generated molecular compositions that is $C_{14}H_{20}N_2O$. The following process should be repeated for each proposed molecular formula separately. Searching $C_{14}H_{20}N_2O$ in the NIST-EI 2005 database results in 24 proposed structures. Data reduction is applied based on requiring ≥1% presence of *m*/*z* 137,

120, 97, 94 ESI-MS-MS peaks in the EI library spectrum. Then, ranking the structures by the number of similar fragments. Five candidates contained all 4 fragments (the correct structure is herein) and 2 candidates contained 3 fragments, an overall reduction from 24 down to 7. Matching between the structures and the ESI-MS-MS spectrum was carried out by applying the fragmentation rules for each structure and examining how many predicted fragments are present in the ESI-MS-MS spectrum. 5 out of 7 structures are presented in Fig. 4. Only in one structure (siduron-the correct compound), 5 predicted cleavages fit the measured ESI-MS-MS degradation pathway, applying the urea rules (m/z 137-rule number 27 in Table 1, m/z 114, 94-rule number 24 (both are "+2 rule"), m/z 120-rule number 26, and m/z 97-rule number 25). A mechanism in proposed in supplementary information 2. The second best structure has only two matching fragments. For the right most bottom structure (Esermethole) no valid rule was determined so far and therefore no prediction could be made.

3.6. Method evaluation

To evaluate the usefulness of our method for identifying structures of unknown chemical compounds, we performed a test, where we randomly retrieved 35 ESI-MS-MS spectra of mostly drugs and pesticides with their molecular formulas from the three mass spectra databases. All the spectra were processed manually according to our method, in a "blind" manner. For 20 molecular formulas, containing either Cl, P, or S atoms, the number of candidate structures in the NIST database (EI + ESI) was less than 5. For the other 15 formulas, containing only CHNO atoms, a large number of structures (27 in average) was proposed per molecular formula. We therefore applied the data reduction procedure to those 15 formulas using the NIST-EI database and search option and reduced the candidate structure number from 27 to 6–7 averagely, requiring at least 2 fragments common to ESI-MS-MS and EI. Matching between the proposed structures and the ESI-MS-MS spectrum was now carried out by applying the fragmentation rules for each structure and examining how many predicted fragments appear in the ESI-MS-MS spectrum. All together, fragmentation was predicted for more than 150 proposed structures, reduced from an initial list of 441. In 30 out of the 35 cases, the correct structure had at least 2 (up to 5) predicted fragments matching the spectrum. In 33 out of 35 cases the correct compound had the highest number of matching cleavages, 26 of them, solely. In the remaining 2 cases, fragmentation could not be predicted.

4. Conclusions

A methodology for the identification of "unknown" small organic compounds by matching experimental ESI-CID data to predicted fragmentation was developed. The construction of a comprehensive empirical set of ESI-MS-MS fragmentation rules based on this systematic study enables to match between chemical structures and ESI-MS-MS degradation pathways. The set of fragmentation rules allows a better prediction of the more likely fragments, while ignoring other theoretical cleavages, thus focusing on realistic routes. Several routes are proposed for some of the functional groups, where a single rule is proposed for others.

The topic of interpretation of atmospheric pressure ionization mass spectra of small molecules has been addressed in several publications in the last decade [17–19,21,22], including a recent review by Holcapek et al. [26]. To the best of our knowledge, many of the rules listed here, including the set of "+2" rules detailed in Table 1 and in Section 3.3, are reported here for the first time.

In case too many structure candidates are proposed for an empirical formula, a protocol which allows reducing the number of potential chemical structures for an empirical formula is proposed here. The protocol is based on the use of the large EI-MS database and addressing the minor similarities between ESI-MS-MS and EI spectra.

The new approach was evaluated by us in a "blind" test, addressing 35 "unknown" spectra and possible empirical formulas. The combination of NIST-EI-based data reduction and ESI-MS-MS interpretation proved to be effective with no false negatives.

All the study and evaluation reported here were performed manually. While automation of the whole process is feasible and desirable for multi compound identification, we note that manual processing is possible and relatively rapid in a matter of a few minutes per structure.

Further improvements can be made by exploring additional functional groups as well as including prediction for the complex cyclic systems. Automation of the reduction process, as well as the fragmentation matching algorithms should be addressed.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ijms.2010.10.024.

References

[1] L. Wenta, M.D. Williams, P.S. Kusum, Anal. Chem. 80 (2008) 7765.
[2] L. Mingxiang, L. Min, A. Rustum, J. Pharm. Biol. Anal. 45 (2007) 747.
[3] I. Bobeldijk, J.P.C. Vissers, G. Kearney, H. Major, J.A. Van Leerdam, J. Chromatogr. A 929 (2001) 63.
[4] J.F. Garcia-Reyes, A. Molina-Diaz, A.R. Fernandez-Alba, Anal. Chem. 79 (2007) 307.
[5] M. Lbanez, J.V. Sancho, O.J. Pozo, W. Niessen, F. Hernandez, Rapid Commun. Mass Spectrom. 19 (2005) 169.
[6] E.M. Thurman, I. Ferrer, A. Fernandez-Alba, J. Chromatogr. A 1067 (2005) 127.
[7] F.W. McLafferty, F. Turecek, Interpretation of Mass Spectra, 4th ed., University Science Books, Sausalito, CA, 1993.
[8] D.W. Hill, T.M. Kertesz, D. Fontaine, R. Friedman, D.F. Grant, Anal. Chem. 80 (2008) 5574.
[9] A. Pelander, E. Tyrkko, I. Ojanpera, Rapid. Commun. Mass. Spectrom. 23 (2009) 506.
[10] Available at http://www.acdlabs.com/.
[11] A. Kerber, R. Laue, M. Meringer, K. Varmuza, Adv. Mass Spectrom. 15 (2001) 939.
[12] E.L. Schymanski, M. Meringer, W. Brack, Anal. Chem. 81 (2009) 3608.
[13] J. Gasteiger, W. Hanebeck, K.P. Schulz, J. Chem. Inf. Comput. Sci. 32 (1992) 264.
[14] B. Fan, H. Chen, M. Petitjean, A. Panaye, J.P. Doucet, H. Xia, S. Yuan, Spect. Lett. 38 (2005) 145.
[15] M. Heinonen, A. Rantanen, T. Mielikainen, J. Kokkonen, J. Kiuru, R.A. Ketola, J. Rousu, Rapid Commun. Mass Spectrom. 22 (2008) 3043.
[16] W.M.A. Niessen, Analysis 28 (2000) 885.
[17] W.F. Smyth, J. Chromatogr. B 824 (2005) 1.
[18] C. Joyce, W.F. Smith, V.N. Ramachandran, E. O'kane, D.J. Coulter, J. Pharm. Biol. Anal. 36 (2004) 465.
[19] E.M. Thurman, I. Ferrer, O.J. Pozo, J.V. Sancho, F. Hernandez, Rapid Commun. Mass Spectrom. 21 (2007) 3855.
[20] I. Ferrer, E.M. Thurman, Trends Anal. Chem. 22 (2003) 750.
[21] I. Ferrer, E.M. Thurman, Liquid chromatography mass spectrometry, MS/MS and time-of-flight MS: analysis of emerging contaminants, in: American Chemical Society Symposium Series, vol. 850, New York, 2003.
[22] K. Levsen, H.M. Schiebel, J.K. Terlouw, K.J. Jobst, M. Elend, A. Preib, H. Thiele, A. Ingendoh, J. Mass Spectrom. 42 (2007) 1024.
[23] B.L. Milman, Trends Anal. Chem. 24 (2005) 493.
[24] A. Lutz, G. Kerstin, K. Gunther, V. Barbel, Mass Spectrom. Rev. 25 (2006) 838.
[25] T. Matsuo, R.M. Caprioli, M.L. Gross, Y. Seyama, Biological Mass Spectrometry—Present and Future, Wiley, Chichester, 1994.
[26] M. Holcapek, R. Jirasko, M. Lisa, J. Chromatogr. A 1217 (2010) 3908.