International Journal of Mass Spectrometry 443 (2019) 77–85

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

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

Reference-standard free metabolite identification using infrared ion spectroscopy

Rianne E. van Outersterp [a], Kas J. Houthuijs [a], Giel Berden [a], Udo F. Engelke [b], Leo A.J. Kluijtmans [b], Ron A. Wevers [b], Karlien L.M. Coene [b], Jos Oomens [a, c], Jonathan Martens [a, *]

[a] Radboud University, Institute for Molecules and Materials, FELIX Laboratory, Toernooiveld 7c, 6525ED, the Netherlands
[b] Department of Laboratory Medicine, Translational Metabolic Laboratory, Radboud University Medical Center, Geert Groote Plein Zuid 10, 6525 GA, Nijmegen, the Netherlands
[c] van't Hoff Institute for Molecular Sciences, University of Amsterdam, 1098XH, Amsterdam, Science Park 908, the Netherlands

ARTICLE INFO

Article history:
Received 15 April 2019
Received in revised form
24 May 2019
Accepted 24 May 2019
Available online 27 May 2019

Keywords:
Infrared ion spectroscopy
Metabolite identification
Computational chemistry
Hydrophilic interaction liquid chromatography

ABSTRACT

Liquid chromatography-mass spectrometry (LC-MS) is, due to its high sensitivity and selectivity, currently the method of choice in (bio)analytical studies involving the (comprehensive) profiling of metabolites in body fluids. However, as closely related isomers are often hard to distinguish on the basis of LC-MS(MS) and identification is often dependent on the availability of reference standards, the identification of the chemical structures of detected mass spectral features remains the primary limitation. Infrared ion spectroscopy (IRIS) aids identification of MS-detected ions by providing an infrared (IR) spectrum containing structural information for a detected MS-feature. Moreover, IR spectra can be routinely and reliably predicted for many types of molecular structures using quantum-chemical calculations, potentially avoiding the need for reference standards. In this work, we demonstrate a workflow for reference-free metabolite identification that combines experiments based on high-pressure liquid chromatography (HPLC), MS and IRIS with quantum-chemical calculations that efficiently generate IR spectra and give the potential to enable reference-standard free metabolite identification. Additionally, a scoring procedure is employed which shows the potential for automated structure assignment of unknowns. Via a simple, illustrative example where we identify lysine in the plasma of a hyperlysinemia patient, we show that this approach allows the efficient assignment of a database-derived molecular structure to an unknown.

1. Introduction

Due to its ultra-high sensitivity and selectivity, mass spectrometry (MS), often combined with separation techniques such as liquid chromatography (LC), gas chromatography (GC) or capillary electrophoresis (CE), is currently the method of choice for the analysis of low-abundance compounds in samples of biological origin [1]. Although the number of compounds that can be detected from a sample in a single MS experiment is unmatched by alternative analytical techniques, the identification of the detected molecular features on the basis of their (exact) mass and retention time information often remains a significant challenge [2,3]. While several modern MS platforms provide sufficient resolving power to assign a unique chemical formula to each detected molecular feature, the subsequent assignment of a full molecular structure is not straightforward. Frequently, a tandem MS (MS/MS) experiment, most often collision-induced dissociation (CID), is used to generate the fragmentation mass spectrum of an unidentified molecular feature of interest and can be compared to databases of known fragmentation spectra measured from reference standards. However, in an untargeted analytical workflow it is often the case that many of the detected molecular features have never been previously identified and are therefore by definition not present in any available database. In that case, the structure of the unknown has to be inferred from the (fragmentation) mass data directly. As the gas-phase ion chemistry determining the fragmentation behavior of ions in an MS experiment is difficult to predict with certainty,

* Corresponding author.
E-mail address: jonathan.martens@ru.nl (J. Martens).

https://doi.org/10.1016/j.ijms.2019.05.015
1387-3806/

deriving molecular structure on this basis is highly challenging and is an insufficient basis for reliable identification. Therefore, identification based on LC-MS often results in ambiguously or partially resolved structures and alternative techniques, such as nuclear magnetic resonance (NMR) spectroscopy that often require extensive chromatography-based purification steps, are needed to confidently resolve full molecular structures.

An alternative MS-based structural characterization technique is infrared ion spectroscopy (IRIS), which generates IR spectra for mass-selected ions. As an IR spectrum often contains a plethora of information related to molecular structure, IRIS is now frequently used for the characterization of structures and conformations of gas-phase ions in MS studies [4–14]. As IRIS is able to provide an IR spectrum for essentially any ion detected in an ion trap mass spectrometer, the potential to become a powerful technique for the structural identification of detected features in analytical MS workflows is currently being realized [15-17,23]. Several recent studies from our group have demonstrated this potential by highlighting the IRIS-based identification of metabolites from body fluids with and without the use of LC [18-20]. Cryogenic IR techniques employing tagging experiments are as well being recognized as potentially interesting for small molecule identification [15,21-23]. While these techniques provide higher resolution IR spectra, the home-built cryogenic MS instrumentation currently needed for these experiments lacks the sensitivity and efficiency needed to probe low-abundance metabolites in biological samples.

One of the intrinsic advantages of metabolite identification using IRIS lies in the fact that IR spectra of gas-phase species, in contrast to CID fragmentation patterns, can be routinely and reliably predicted for many types of molecular structures using quantum-chemical calculations. This offers the opportunity to screen a large number of candidate molecular structures by matching an experimental IR spectrum to computed IR spectra of the candidates, without the need for time-consuming and expensive synthesis of reference standards. Although this strategy has great potential, spectral matching to computational spectra of candidate structures is not always conclusive, especially when identifying larger molecules. This is predominantly related to the larger conformational phase space and potentially overlapping IR bands, which can lead to spectral broadening. However, as IRIS can effectively be performed on any ion population generated in MS, IR spectra can also be obtained for the MS/MS generated fragments of an unknown. CID fragments are often smaller and more rigid structures with lower conformational flexibility as compared to their precursor ions, especially when cyclization rearrangements accompany fragmentation, which often makes their IR spectra sharper and simpler to interpret [20]. Thus, the structural identification of fragment ions can both add confidence to the assignment of a precursor structure and, in the case that the molecular formula of interest is not present in databases, aid in bottom-up structure elucidation.

In this work, we demonstrate an approach to small-molecule identification which involves the separation of complex mixtures using high-pressure liquid chromatography (HPLC) and the subsequent analysis using (tandem) MS and IRIS on both a precursor and one of its fragment ions. We demonstrate that using the combination of IRIS and quantum-chemically predicted IR spectra of candidate molecular structures, detected features can be identified without the use of reference standards. Candidate structures can be gathered from databases of known molecules, or more specifically metabolites, such as the Human Metabolome Database (HMDB) [24-27], METLIN [28] or ChemSpider [29], which often leads to a large set of isomeric molecular structures. By comparing the IR spectrum of an unknown molecular feature to the predicted IR spectra of a large set of candidate structures, there is additionally the potential to extract molecular structure information even in cases when the exact molecular structure is not present in the set of candidate structures. The correlation of matching and mismatching IR spectral features between an IR spectrum of an unknown and predicted IR spectra of candidate structures provides suggestions towards the presence/absence of specific functional groups.

As a simple illustrative example, we focus in this work on the identification of lysine in plasma from a patient with an inborn error of lysine metabolism. In the field of inborn errors of metabolism (IEMs), including in our own laboratory [30], untargeted metabolomics strategies are now routinely used in order to discover new metabolic biomarkers whose deviating levels in comparison to controls indicate the presence of both known (clinical diagnostics) and yet unknown IEMs. Reliable identification of these known metabolites as well as detected but unidentified molecular features is crucial for their incorporation in newborn screening programs and for a pathophysiological understanding of these diseases. As these untargeted screening methods are often based on LC-MS, molecular structure identification forms one of the main bottlenecks in these strategies.

The fragmentation chemistry of peptides and their amino acid building blocks, including lysine, has been extensively studied previously, providing a firm foundation for a validation of the workflow. Protonated lysine at *m/z* 147 is known to fragment mainly via loss of ammonia, resulting in a fragment at *m/z* 130 [31-33]. This fragment has been suggested to be L-pipecolic acid, which is also biochemically one of the downstream metabolites of lysine [34]. Moreover, gas-phase IR spectra of protonated lysine have previously been recorded and compared to predicted IR spectra [35-37], which revealed that the preferred protonation site is the ε-amino nitrogen, resulting in two hydrogen bonds with the α-amino nitrogen and the carbonyl-group.

2. Methods

2.1. Chemicals

Methanol, ethanol, water and formic acid (LC-MS grade) used during the sample preparation procedure was obtained from Sigma-Aldrich (St. Louis, USA). Mobile-phase solvents were prepared with LC-MS grade water and methanol obtained from Merck (Darmstadt, Germany). LC-MS grade formic acid and ammonium formate were obtained from VWR International (Leuven, Belgium) and Fisher Scientific (Geel, Belgium), respectively. L-Lysine and L-pipecolic acid reference standards were obtained from Sigma-Aldrich (St. Louis, USA).

2.2. Sample preparation

The concentration of lysine in the plasma sample was determined via a standard amino acid quantitation method in our lab which is based on ion-exchange chromatography and ninhydrin derivatization. For the other experiments, the plasma sample was prepared following a procedure described previously [30]. The sample, which was stored at −80 °C, was thawed at 4 C. 400 µl of ice-cold methanol/ethanol 50:50 [v/v] was added to 100 µl of plasma and mixed with a vortex mixer for 15 s. Next, these dilutions were incubated for 20 min at 4 °C and subsequently centrifuged at 18600 g for 15 min at 4 °C. 350 µl of the supernatant was transferred into new tubes and dried in a centrifugal vacuum evaporator (Eppendorf). The sample was reconstituted in 100 ml 0.1% formic acid in deionised water/methanol 90:10 [v/v], mixed with a vortex mixer for 15 s and centrifuged at 18600 g for 15 min at room temperature. 90 µl of the supernatant was transferred to an autosampler vial and used for LC-MS analysis.

2.3. Separations

LC separations were performed with a Bruker Elute SP HPLC system consisting of a binary pump, cooled autosampler and column oven. The outlet of the column was coupled to a quadrupole ion trap mass spectrometer (Bruker, AmaZon Speed ETD). The separations were performed with a Waters Acquity amide column (100 × 2.1 mm i.d., 1.7 μm particles, 130 Å pore size) held at 30 °C using a mobile phase consisting of 10 mM ammonium formate and 0.125% (v/v) formic acid in 95:5 (v/v) acetonitrile:water (mobile phase A) and 50:50 (v/v) acetonitrile:water (mobile phase B). A flow-rate of 0.5 ml/min was used. After an initial time of 1 min at 100% A, a gradient was run to 100% B in 10 min followed by a hold at 100% B of 1 min. An equilibration time of 8 min was used, leading to a total analysis time of 21 min. For initial experiments an injection-volume of 4 μl was used. For the fraction collection experiments, an injection-volume of 4 μl was used. The eluent was collected in two 96-well plates using a Foxy R2 fraction collector. The accurate mass of all detected molecular features was determined with an Agilent 1290 ultra-high-performance (UHP) LC system connected to a Agilent 6540 QTOF mass spectrometer as described elsewhere [30].

2.4. Infrared ion spectroscopy

IRIS experiments were performed in a quadrupole ion trap mass spectrometer (Bruker, AmaZon Speed ETD) modified for spectroscopy. Details of the hardware modifications and synchronisation of the experiment with the infrared laser are described elsewhere [38]. The collected fraction and solutions of reference compounds (~10^{-7} M in 50:50 acetonitrile:water) were introduced at 80 μl/h flow rates to the electrospray source (+ESI). The ions of interest were mass-isolated and subjected to IR analysis. IR spectra were recorded using the FELIX infrared free electron laser, which was set to produce IR radiation in the form of ~10 μs macropulses of 50–150 mJ at a 10 Hz repetition rate (bandwidth ~0.4% of the centre frequency).

When the laser is resonant with a vibrational transition of the ions this leads to absorption of the IR photons, producing an increase in the ions internal energy and eventually leading to photodissociation. Thus, IR absorption can be observed by recording a fragmentation MS spectrum. Plotting the amount of fragmentation (IR yield = Σ*I*(fragment ions)/Σ*I*(parent + fragment ions)) as a function of IR laser frequency produces an IR spectrum. In our experiments, the yield at each point is typically calculated from 4 to 8 averaged fragmentation mass spectra. The IR frequency is calibrated using a grating spectrometer, and the IR yield is linearly corrected for frequency-dependent variations in the laser pulse energy.

2.5. Computational procedure

The SMILES structure format [39,40] of selected HMDB entries was used as starting 2D-structure for the workflow using the cheminformatics toolbox RDKit [41] and all oxygen and nitrogen atoms were considered as possible protonation sites. Some selected HMDB entries contained an unspecified stereocentre. These centres were randomly assigned to be R or S, as both would result in enantiomeric structures having the same IR spectrum. A conformation search was performed for each protonation isomer using a distance geometry algorithm, which yielded 500 random 3D-conformations, which were minimized using a classical forcefield (MMFF94) [42-47]. Of these 500 structures a maximum of 10 unique conformations was selected after clustering, or fewer if conformations were too similar (as determined by a root means squared deviation (RMSD) threshold of 1.4 Å) [48]. These protonated 3D-conformations were then submitted to Gaussian 16 for geometry optimisations and frequency calculations using the semi-empirical PM6 level [49]. By comparing the relative energies (electronic and thermal) of the resulting optimised geometries, unfavourable protonation sites and conformations were filtered by using a relative energy cut-off of 40 kJ/mol. Additionally, similar geometries were filtered based on (close to) identical calculated frequencies and corresponding intensities. After these filtering steps, the remaining structures were reoptimized using the B3LYP density functional and 6-311+G(d,p) basis set and thereafter a frequency calculation was performed. Harmonic vibrational frequencies were scaled by 0.975. To aid comparison to experimental spectra, Gaussian broadening (20 cm^{-1} at full width half maximum) was applied to the calculated vibrational lines. To obtain reliable energies the thermal energy of the frequency calculation was combined with the Møller-Plesset second order correction to the Hartree Fock electronic energy, again using the 6-311+G(d,p) basis set.

For the comparison of computational spectra to the experimental spectrum only the lowest energy protonated structures were considered. An objective score for spectral similarity is obtained by employing the cosine similarity score, which expresses the similarity S between two spectra A and B by their normalized Euclidean dot product according to:

$$S = \cos(\theta) = \frac{\boldsymbol{A}\cdot\boldsymbol{B}}{|\boldsymbol{A}||\boldsymbol{B}|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

Here a score closer to 1 indicates greater similarity. The intensity values of the computational spectra are evaluated at the same wavenumber points as the experimental spectrum, in order to ensure a common x-axis. The above equation is slightly modified to make the similarity S less sensitive to intensity deviations and thus more sensitive to frequency overlap [50]. This is achieved by scaling both the experimental and computational spectra to 1 and then taking the logarithm:

$$A_i^{transformed} = log\left(\frac{A_i}{A_{max}} + c\right)$$

where c is a constant that is identical for both A and B. The value of c determines the extent to which deviations in peak intensity are allowed, and was set to the value of 10^{-8}, as this gave the best result for a small set of test experimental and computational spectra.

3. Results & discussion

Hyperlysinemia is an autosomal recessive inborn error of lysine metabolism which leads to high lysine levels in plasma (quantified in the sample used in this study at 1177 μM, see method section). In a previous proof-of-principle experiment, the IR spectrum of a targeted metabolite was recorded via direct injection of the diluted plasma sample [18]. However, in many cases this approach is not feasible when analysing body fluids due to ion suppression effects, the presence of other isobaric metabolites and reduced overall sensitivity. Therefore, separation of the body fluid samples using HPLC before the IRIS measurements is employed here. Using collected LC fractions containing the molecular feature of interest, IRIS measurements are conducted by direct infusion electrospray ionization (ESI). Although reversed-phase liquid chromatography (RPLC) is the most common HPLC technique, very polar metabolites, including lysine, are poorly retained. Therefore, hydrophilic interaction liquid chromatography (HILIC) was used instead, retaining highly polar compounds while still maintaining good MS-compatibility.

Panel (a) of Fig. 1 shows the total ion chromatogram (TIC) and extracted ion chromatogram (EIC) of *m/z* 147 measured during the separation of a hyperlysinemia plasma sample as compared to the chromatograms resulting from the separation of a control plasma sample. It can be seen that both samples contain an *m/z* 147 ion that elutes between 8.0 and 8.5 min but the levels in the hyperlysinemia sample are significantly elevated as compared to the control sample. In order to obtain structural information about the *m/z* 147 ion, the fraction of eluent containing the feature of interest was collected and injected into the MS instrument at a lower flow rate,

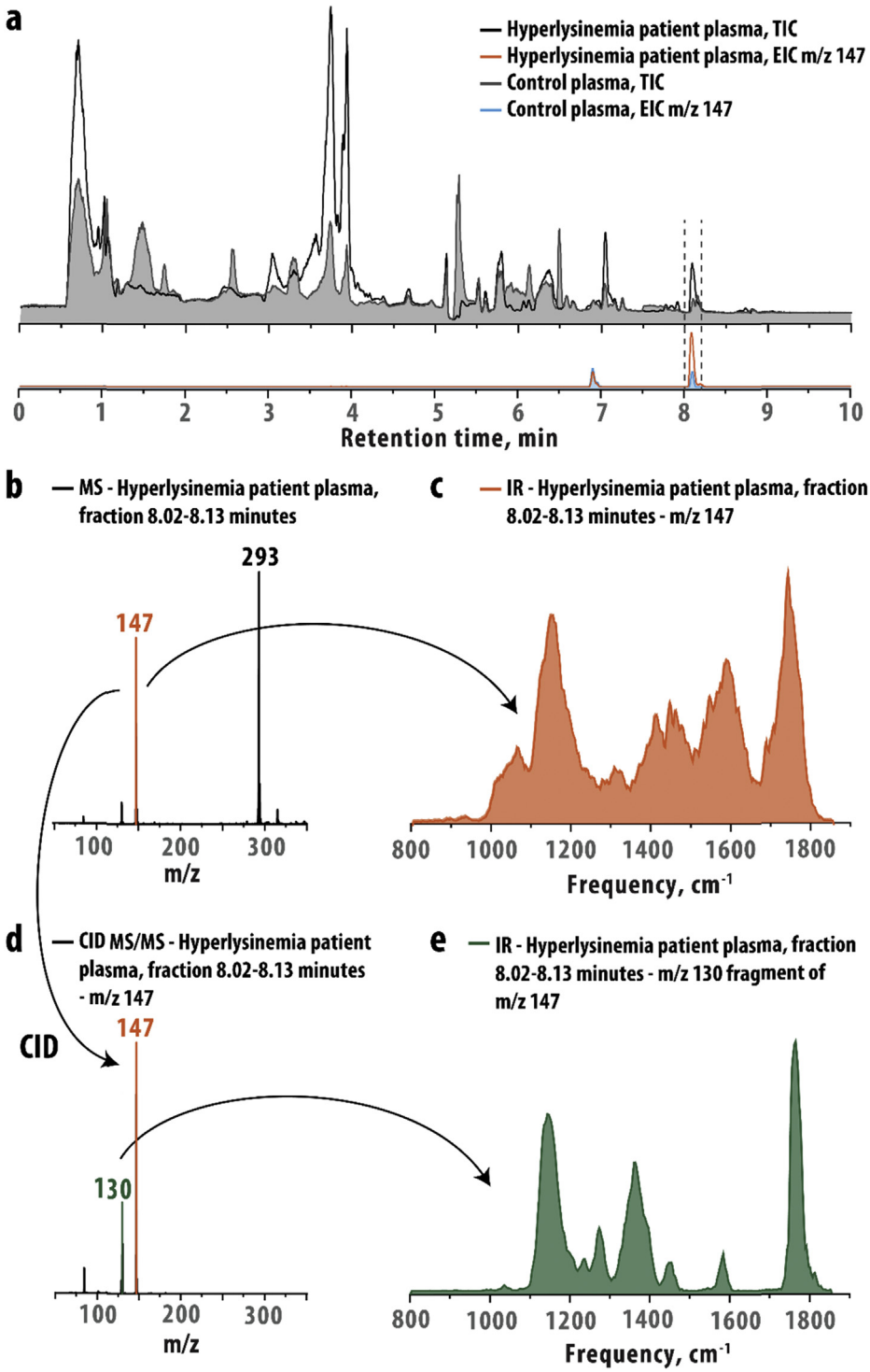

Fig. 1. Overview of experimental results. (a) Total ion chromatograms (TIC) (black and grey traces) and extracted ion chromatograms (EIC) of *m/z* 147 (blue and orange traces) resulting from the separation of the patient and control plasma sample. The dashed vertical lines indicate the collected fraction of eluent. (b) MS spectrum of the collected fraction. The observed *m/z* 147 ion is isolated and its IR spectrum measured using IRIS, resulting in the spectrum shown in panel (c). Alternatively, the isolated *m/z* 147 ion is fragmented first using CID, resulting in the MS/MS spectrum shown in panel (d). The observed *m/z* 130 fragment is isolated and its IR spectrum measured using IRIS, giving the IR spectrum shown in panel (e).

to accommodate for the longer time needed to acquire an IR spectrum (1 µl min^{-1} vs. 400 µl min^{-1}). The *m/z* 147 ion was observed (see panel (b)), isolated, and subjected to IR spectroscopy measurements, resulting in the IR spectrum shown in panel (c). Additionally, the *m/z* 147 ion was fragmented using collision induced dissociation (CID), giving the fragmentation spectrum shown in panel (d). The *m/z* 147 ion fragments mainly via the loss of ammonia (−17) resulting in a fragment with *m/z* 130, which is in agreement with the fragmentation of lysine studied in previous work [31-33]. The fragment is re-isolated and its IR spectrum measured, yielding the IR spectrum shown in panel (e). The spectra of the precursor and fragment ion show a relatively high degree of similarity, which is to be expected as they continue to share several functional groups after the small neutral loss of ammonia. However, the spectrum of the fragment ion shows sharper, better defined IR bands in comparison to the precursor spectrum. This is likely related to the reduced conformational flexibility and the presence of fewer hydrogen bonds, as will be discussed further below.

The IR spectra shown in Fig. 1 both show a band just below 1800 cm^{-1}, consistent with the presence of a carboxylic acid C=O group. The band just below 1600 cm^{-1} is consistent with the presence of an NH_2-group. A common approach to generate a list of candidate structures in studies that use (high resolution) mass spectrometry for small molecule identification is to search the molecular formula of the observed ion against a database containing known or hypothesized candidate molecules. This approach, however, often generates several isomeric candidate structures. For instance, running the molecular formula of lysine ($C_6H_{14}N_2O_2$) against the human metabolome database (HMDB) results in three different candidate structures, which are shown in Fig. 2. All three candidate structures share the functional groups indicated by the IR spectrum in panel (c) of Fig. 1, which would in this case make selection from those candidate structures on this basis impossible.

Quantum-chemical calculations, here based on density functional theory (DFT) at the B3LYP/6-311+G(d,p) level, were used to obtain the optimised structures and IR spectra of the candidate structures for $C_6H_{14}N_2O_2$ presented in Fig. 2. All tautomeric forms resulting from multiple possible protonation sites and different conformations of the same molecular system have been considered. However, in the comparisons of the computational IR spectra to the experimental spectra, only the most energetically favourable structures for each HMDB entry are presented (see methods section for details). In this comparison we employ a scoring algorithm which aids in identifying the best matching computational spectrum. This algorithm calculates the cosine similarity of a computed spectrum and the experimental spectrum, after logarithmic transformation of both spectra to make the score more sensitive to frequency overlap and less sensitive to intensity deviations [50]. The similarity scores allow for automatic exclusion of structures with poorly matching predicted spectra. The remaining structures with closely matching spectra can then be manually inspected.

Fig. 2 contains the comparisons of the computed IR spectra of the three candidate structures derived from the HMDB to the experimental IR spectrum of the *m/z* 147 feature with the scores

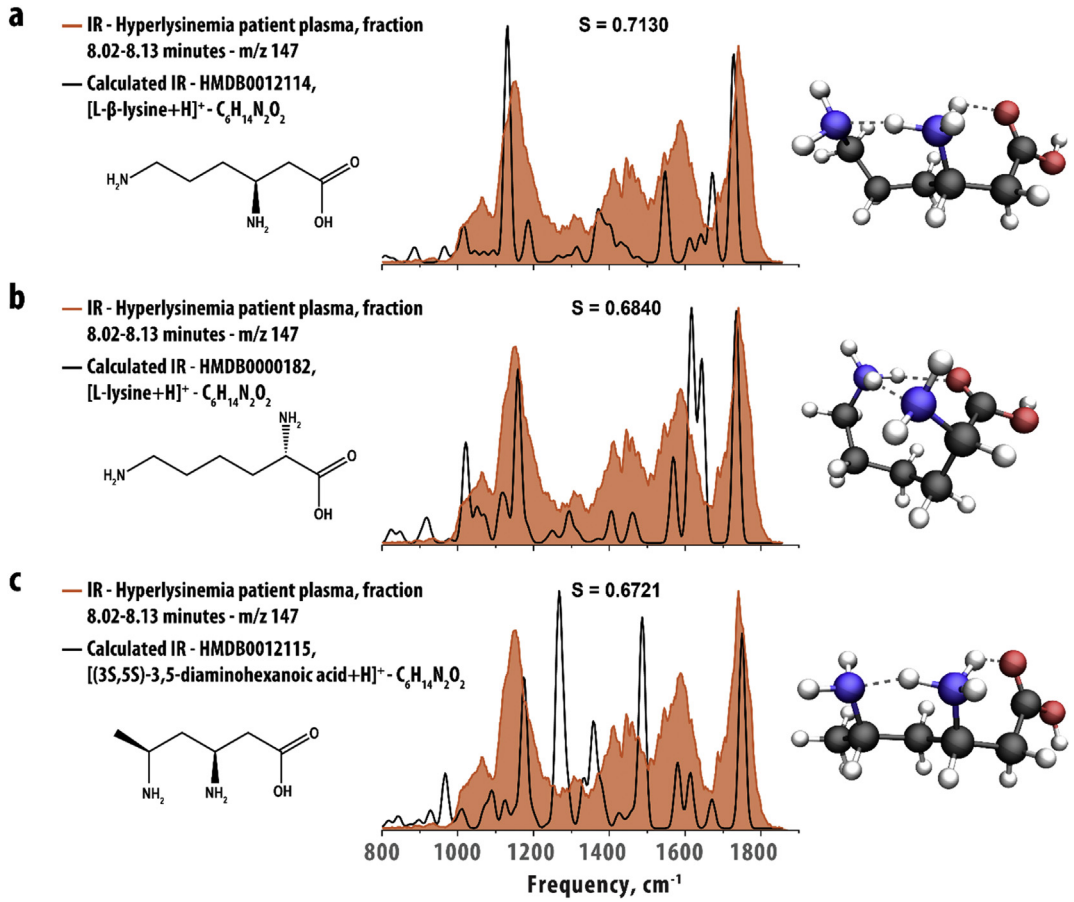

Fig. 2. Experimental IR spectrum recorded for the unknown *m/z* 147 ion (shaded orange spectra) in comparison to predicted IR spectra for (a) protonated L-β-lysine, (b) L-lysine and (c) (3S,5S)-3,5-diaminohexanoic acid (black traces). Input structures and the conformations resulting from the quantum-chemical calculations are inlayed in each panel.

obtained from the automated scoring procedure inlayed in each panel. All three predicted spectra give a relatively similar score as the molecular structures are very similar. The inlayed 3-dimensional structures show several strong hydrogen bonds, possibly explaining the fact that bands in the experimental spectrum are relatively broad [20]. The assignment of the molecular structure is therefore challenging in this case. A direct visual inspection suggests that the predicted spectra in panel (a) and (b) give a better match with experiment than the predicted spectrum in panel (c). For instance, the predicted spectrum in panel (c) shows three intense bands around 1273, 1358 and 1484 cm^{-1} that coincide with dips in the experimental IR spectrum, whereas the predicted spectra in panels (a) and (b) show a better match in this region. However, a distinction between L-lysine and L-β-lysine is difficult and relying on the relative scores would even lead to the assignment of the incorrect structure, L-β-lysine. It should be noted that the assignment is partly impeded by the significant broadening of the experimental spectrum. This broadening can be explained by the predicted three-dimensional structure of L-lysine inlayed in panel (b) of Fig. 2, which shows strong hydrogen bonding, including a shared proton between the two amino groups.

We next aimed at assigning the molecular structure of the *m/z* 130 fragment generated by CID from *m/z* 147. In order to obtain possible candidate structures for the *m/z* 130 ion, we performed a similar search of the HMDB using the molecular formula of the *m/z* 130 fragment ion, which is obtained by expulsion of a neutral ammonia molecule from the precursor ion. In general, the use of a more general database such as ChemSpider would be a better

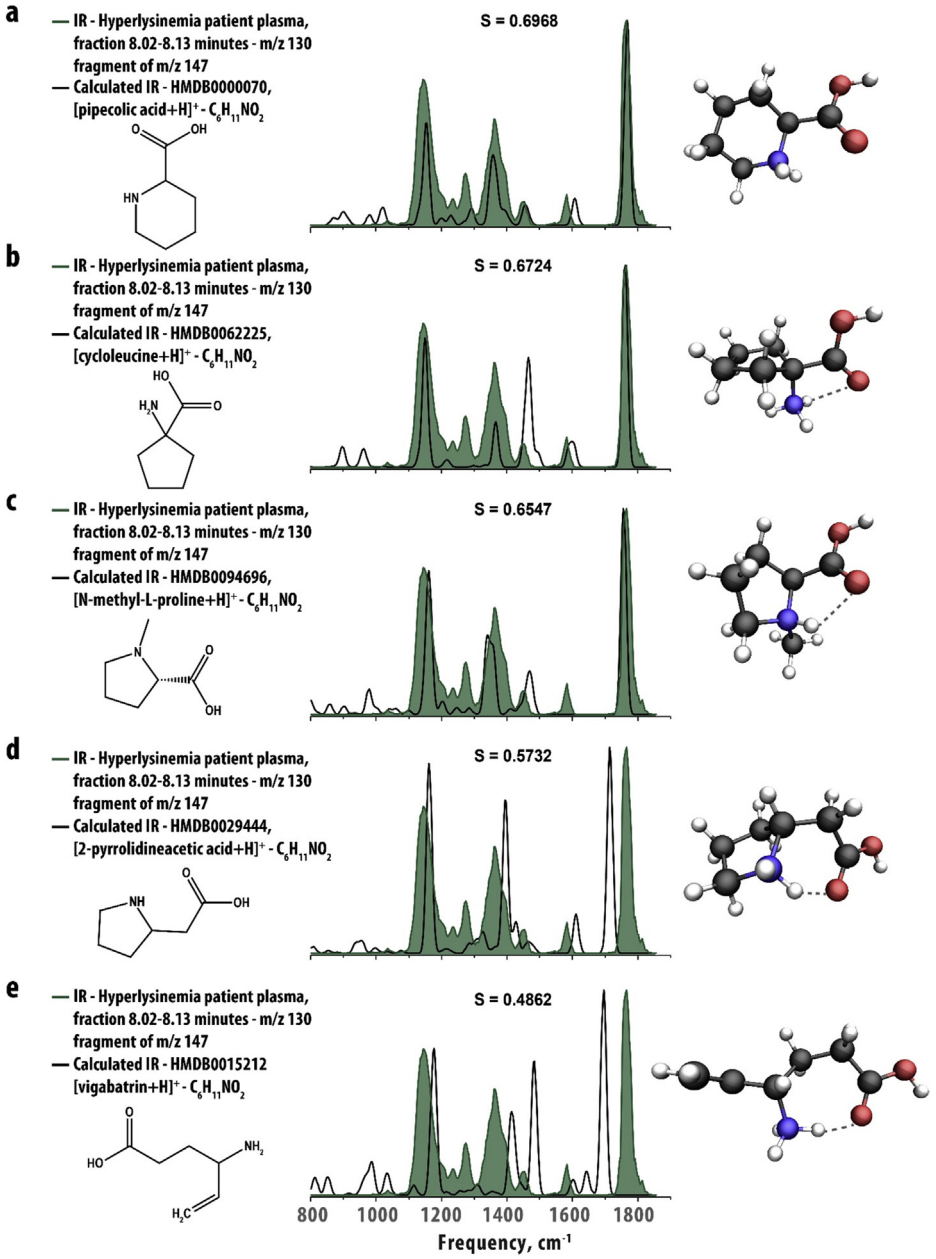

Fig. 3. Experimental IR spectrum recorded for the *m/z* 130 CID fragment (shaded green) compared to predicted IR spectra for protonated (a) pipecolic acid, (b) cycloleucine, (c) *N*-methyl-L-proline, (d) 2-pyrrolidineacetic acid and (e) vigabatrin (black traces). Input structures and the conformations resulting from the quantum-chemical calculations are inlayed in each panel.

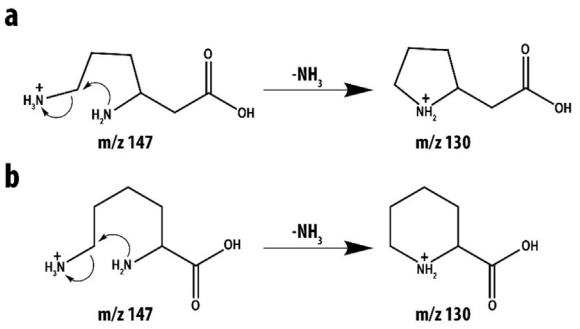

Fig. 4. Proposed mechanisms for the gas-phase deamination reaction of (a) protonated L-β-lysine and (b) protonated L-lysine via loss of ammonia, leading to a fragment of *m/z* 130.

choice to search for candidate structures for CID fragment ions of metabolites, but for the demonstrative purpose of the current study we restricted ourselves to the molecular entries in the HMDB. Searching $C_6H_{11}NO_2$ against the HMDB results in nine isomeric candidate structures. We applied the same procedure as described above to obtain their predicted IR spectra and a comparison between the computed IR spectra of five of the candidate structures, including the highest and lowest scoring structure, to the experimental spectrum of the *m/z* 130 fragment is shown in Fig. 3. The remaining structures and their predicted IR spectra are shown in Fig. S1. The three highest scoring structures are all cyclic structures having a carboxylic acid group (COOH) directly connected to one of the ring-carbons. This carboxylic acid group gives rise to a C=O stretching feature just below 1800 cm^{-1} which, as is seen in Fig. 3, corresponds well to one of the observed IR bands of the *m/z* 130 ion. On the other hand, Fig. S1 shows that the lower scoring IR spectra all show a mismatch in this region which is due to the fact that their corresponding structures contain a C=O group that is either connected more remotely to one of the ring-carbons or to the nitrogen-atom of the ring. The predicted spectra shown in panels (a) and (b) both contain a band that explains the observed band around 1600 cm^{-1}. This band is related to a vibration of the positively charged NH_3- or NH_2-group, and is missing in the predicted spectrum in panel (c), because the corresponding structure is N-functionalized. Finally, both by considering the matching scores of the remaining candidates in panels (a) and (b) and by visual inspection, we can clearly assign the *m/z* 130 feature as protonated pipecolic acid.

Next, the knowledge on the structure of the fragment ion can be used to identify the precursor structure by considering the possible gas-phase fragmentation chemistry that would lead to the identified fragment structure. Noting that the computational procedure above was unable to distinguish between two possible candidate structures for the *m/z* 147 precursor ion, we can now check which of these structures is likely to yield the pipecolic acid structure upon CID. Fig. 4 shows possible mechanisms for the loss of ammonia from the candidate *m/z* 147 structures. Fragmentation could obviously also occur via nucleophilic attack of the ε-amino-group but this would lead to the same fragment structures (see Fig. S2). Isotope-labelling experiments on lysine fragmentation have provided evidence for the formation of pipecolic acid via loss of the ε-amino group (shown in Fig. 4) [51]. Analogous experiments in our lab confirmed this conclusion (see Fig. S3). The fragmentation mechanisms in Fig. 4 suggest that L-lysine (panel (b)) fragments to L-pipecolic acid, while L-β-lysine would be expected to yield 2-pyrrolidineacetic acid due to the different relative positions of the amino-groups. Therefore, by combining the information derived from the comparison with computed IR spectra with knowledge on gas-phase ion chemistry we could identify the *m/z* 147 feature as being L-lysine, which fragments to L-pipecolic acid.

In order to verify the conclusions, we obtained reference standards for L-lysine and L-pipecolic acid and acquired reference IRIS spectra of their protonated ions. Fig. 5 shows an overlay of these reference spectra and the spectra of the *m/z* 147 ion measured from plasma (a) and its *m/z* 130 fragment (b). The spectra show an excellent overlap, showing lysine as the *m/z* 147 ion and pipecolic acid as its main fragment. The L-lysine reference standard was also subjected to CID and an IR spectrum was obtained for its *m/z* 130 fragment. This spectrum also overlaps favourably with the pipecolic acid reference spectrum (see Fig. S4), confirming the identity of the L-lysine fragment proposed in literature [31-33]. An IR spectrum was further obtained for α-^{15}N-labelled lysine and its labelled pipecolic acid fragment, where the labelling was found to produce no significant spectral differences (see Fig. S5). We note that the IRIS spectrum of protonated lysine presented here is very similar to that published previously [36] and the calculated 3D-conformation assigned here (see Fig. 2) is also in agreement with that study. Additionally, the other low-energy tautomers and conformations that were reported, including the lysine zwitterion, were also found by the computations performed here with similar relative energies.

4. Conclusions

We demonstrate the potential for a comprehensive workflow based on HPLC, MS, IRIS and quantum-chemical calculations to aid in the identification of detected molecular features, possibly

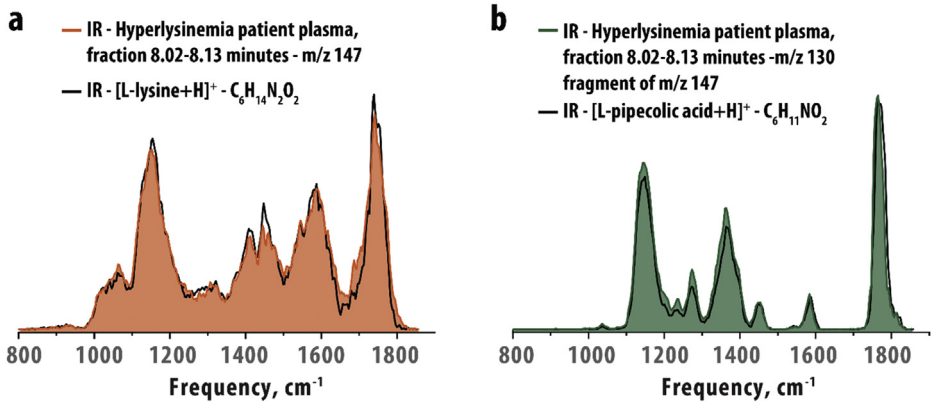

Fig. 5. Experimental IR spectra recorded for (a) the *m/z* 147 ion measured from plasma (shaded orange spectrum) and (b) its *m/z* 130 fragment (shaded green spectrum), in comparison with reference spectra recorded for protonated L-lysine and protonated L-pipecolic acid (black traces), respectively.

without the need for reference standards. The workflow presented here was specifically tailored for the identification of polar analytes in plasma samples, but can be adapted to identify any other small molecule that can be detected by HPLC-MS based analytical workflows. Currently, our lab employs several RPLC- and HILIC-based chromatography methods in order to identify metabolites in urine, blood plasma and cerebrospinal fluid. In the current work, we demonstrated that identifying MS/MS fragments of a precursor ion of interest can aid the identification of the precursor structure by looking back at the gas-phase fragmentation reactions that would lead to the fragment ion structure. Here, we focused on the identification of one fragment but, naturally, the procedure can be extended to multiple CID fragments as well for bottom-up structural identification. As well, we demonstrate that combining quantum-chemical calculations on database-derived structures with a scoring procedure leads to an efficient interpretation of experimental IRIS spectra. In an ongoing effort we are working towards optimizing the scoring procedure to provide a more confident and automatic assignment of candidate structures. Here, we focussed on the HMDB but we envision that using other databases such as METLIN [28] or ChemSpider [29], which in general contain more candidates, can further enhance identification of many types of metabolites.

Acknowledgements

We gratefully acknowledge the Nederlandse Organisatie voor Wetenschappelijk Onderzoek for the support of the FELIX Laboratory. NWO-VICI (724.011.002), NWO-TTW (15769), and NWO-TKI-LIFT (731.017.419) grants allowed us to develop the LC-MS infrastructure used in this study. We also thank NWO Physical Sciences (EW) and the SurfSARA Supercomputer Centre for providing the computational time and resources (grant 17603).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijms.2019.05.015.

References

[1] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics, Mass Spectrom. Rev. 26 (1) (2007) 51–78.
[2] T. De Vijlder, D. Valkenborg, F. Lemière, E.P. Romijn, K. Laukens, F. Cuyckens, A tutorial in small molecule identification via electrospray ionization-mass spectrometry: the practical art of structural elucidation, Mass Spectrom. Rev. 37 (5) (2018) 607–629.
[3] T. Kind, O. Fiehn, Advances in structure elucidation of small molecules using mass spectrometry, Bioanal. Rev. 2 (1) (2010) 23–60.
[4] J. Lemaire, P. Boissel, M. Heninger, G. Mauclaire, G. Bellec, H. Mestdagh, A. Simon, S.L. Caer, J.M. Ortega, F. Glotin, P. Maitre, Gas phase infrared spectroscopy of selectively prepared ions, Phys. Rev. Lett. 89 (27) (2002) 273002.
[5] J. Oomens, B.G. Sartakov, G. Meijer, G. von Helden, Gas-phase infrared multiple photon dissociation spectroscopy of mass-selected molecular ions, Int. J. Mass Spectrom. 254 (1) (2006) 1–19.
[6] L.M. Aleese, A. Simon, T.B. McMahon, J.-M. Ortega, D. Scuderi, J. Lemaire, P. Maître, Mid-IR spectroscopy of protonated leucine methyl ester performed with an FTICR or a Paul type ion-trap, Int. J. Mass Spectrom. 249–250 (2006) 14–20.
[7] T.D. Fridgen, Infrared consequence spectroscopy of gaseous protonated and metal ion cationized complexes, Mass Spectrom. Rev. 28 (4) (2009) 586–607.
[8] N.C. Polfer, J. Oomens, Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance, Mass Spectrom. Rev. 28 (3) (2009) 468–494.
[9] J. Martens, J. Grzetic, G. Berden, J. Oomens, Structural identification of electron transfer dissociation products in mass spectrometry using infrared ion spectroscopy, Nat. Commun. 7 (2016) 11754.
[10] N.C. Polfer, Infrared multiple photon dissociation spectroscopy of trapped ions, Chem. Soc. Rev. 40 (5) (2011) 2211–2221.
[11] N.C. Polfer, J.J. Valle, D.T. Moore, J. Oomens, J.R. Eyler, B. Bendiak, Differentiation of isomers by wavelength-tunable infrared multiple-photon dissociation-mass spectrometry: application to glucose-containing disaccharides, Anal. Chem. 78 (3) (2006) 670–679.
[12] J.L. Rummel, J.D. Steill, J. Oomens, C.S. Contreras, W.L. Pearson, J. Szczepanski, D.H. Powell, J.R. Eyler, Structural elucidation of direct analysis in real time ionized nerve agent simulants with infrared multiple photon dissociation spectroscopy, Anal. Chem. 83 (11) (2011) 4045–4052.
[13] J. Martens, G. Berden, J. Oomens, Structures of fluoranthene reagent anions used in electron transfer dissociation and proton transfer reaction tandem mass spectrometry, Anal. Chem. 88 (12) (2016) 6126–6129.
[14] A.P. Cismesia, L.S. Bailey, M.R. Bell, L.F. Tesler, N.C. Polfer, Making mass spectrometry see the light: the promises and challenges of cryogenic infrared ion spectroscopy as a bioanalytical technique, J. Am. Soc. Mass Spectrom. 27 (5) (2016) 757–766.
[15] A.P. Cismesia, M.R. Bell, L.F. Tesler, M. Alves, N.C. Polfer, Infrared ion spectroscopy: an analytical tool for the study of metabolites, Analyst 143 (7) (2018) 1615–1623.
[16] F. Berthias, B. Maatoug, G.L. Glish, F. Moussa, P. Maitre, Resolution and assignment of differential ion mobility spectra of sarcosine and isomers, J. Am. Soc. Mass Spectrom. 29 (4) (2018) 752–760.
[17] B. Schindler, G. Laloy-Borgna, L. Barnes, A.-R. Allouche, E. Bouju, V. Dugas, C. Demesmay, I. Compagnon, Online separation and identification of isomers using infrared multiple photon dissociation ion spectroscopy coupled to liquid chromatography: application to the analysis of disaccharides regio-isomers and monosaccharide anomers, Anal. Chem. 90 (20) (2018) 11741–11745.
[18] J. Martens, G. Berden, R.E. van Outersterp, L.A.J. Kluijtmans, U.F. Engelke, C.D.M. van Karnebeek, R.A. Wevers, J. Oomens, Molecular identification in metabolomics using infrared ion spectroscopy, Sci. Rep. 7 (1) (2017) 3363.
[19] J. Martens, V. Koppen, G. Berden, F. Cuyckens, J. Oomens, Combined liquid chromatography-infrared ion spectroscopy for identification of regioisomeric drug metabolites, Anal. Chem. 89 (8) (2017) 4359–4362.
[20] J. Martens, G. Berden, H. Bentlage, K.L.M. Coene, U.F. Engelke, D. Wishart, M. van Scherpenzeel, L.A.J. Kluijtmans, R.A. Wevers, J. Oomens, Unraveling the unknown areas of the human metabolome: the role of infrared ion spectroscopy, J. Inherit. Metab. Dis. 41 (3) (2018) 367–377.
[21] A.B. Wolk, C.M. Leavitt, E. Garand, M.A. Johnson, Cryogenic ion chemistry and spectroscopy, Acc. Chem. Res. 47 (1) (2014) 202–210.
[22] T.R. Rizzo, O.V. Boyarkin, Cryogenic methods for the spectroscopy of large, biomolecular ions, in: A.M. Rijs, J. Oomens (Eds.), Gas-phase IR Spectroscopy and Structure of Biological Molecules, Springer International Publishing, Cham, 2015, pp. 43–97.
[23] O. Gorlova, S.M. Colvin, A. Brathwaite, F.S. Menges, S.M. Craig, S.J. Miller, M.A. Johnson, Identification and partial structural characterization of mass isolated valsartan and its metabolite with messenger tagging vibrational spectroscopy, J. Am. Soc. Mass Spectrom. 28 (11) (2017) 2414–2422.
[24] D.S. Wishart, D. Tzur, C. Knox, R. Eisner, A.C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.-A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D.D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G.E. Duggan, G.D. MacInnis, A.M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B.D. Sykes, H.J. Vogel, L. Querengesser, HMDB: the human metabolome database, Nucleic Acids Res. 35 (suppl_1) (2007) D521–D526.
[25] A. Nazyrova, A.C. Guo, A. De Souza, A. Lewis, A. Zuniga, B. Gautam, C.A. Sobsey, C. Knox, D. Tzur, D.D. Hau, D. Cheng, D. Clive, E. Dong, E. Lim, H.J. Vogel, I. Forsythe, I. Sinelnikov, J. Xia, J.A. Cruz, J. Peng, L. Jia, L. Li, L. Fang, M. Dawe, M. Clements, N. Young, N. Psychogios, P. Huang, P. Liu, R. Eisner, R. Mandal, R. Greiner, R. Shaykhutdinov, R. Fradette, S. Shrivastava, S. Bouatra, Y. Xiong, D.S. Wishart, HMDB: a knowledgebase for the human metabolome, Nucleic Acids Res. 37 (suppl_1) (2008) D603–D610.
[26] A.C. Guo, A. Scalbert, C. Knox, D. Arndt, E. Dong, F. Yallou, F. Aziat, F. Allen, I. Sinelnikov, J. Xia, M. Wilson, P. Liu, R. Perez-Pineiro, R. Eisner, R. Mandal, R. Greiner, S. Bouatra, T. Jewison, T. Bjorndahl, V. Neveu, Y. Djoumbou, Y. Liu, D.S. Wishart, HMDB 3.0—the human metabolome database in 2013, Nucleic Acids Res. 41 (D1) (2012) D801–D807.
[27] D.S. Wishart, Y.D. Feunang, A. Marcu, A.C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, A. Scalbert, HMDB 4.0: the human metabolome database for 2018, Nucleic Acids Res. 46 (D1) (2017) D608–D617.
[28] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, METLIN: a metabolite mass spectral database, Ther. Drug Monit. 27 (6) (2005) 747–751.
[29] H.E. Pence, A. Williams, ChemSpider: an online chemical information resource, J. Chem. Educ. 87 (11) (2010) 1123–1124.
[30] K.L.M. Coene, L.A.J. Kluijtmans, E. van der Heeft, U.F.H. Engelke, S. de Boer, B. Hoegen, H.J.T. Kwast, M. van de Vorst, M.C.D.G. Huigen, I.M.L.W. Keularts, M.F. Schreuder, C.D.M. van Karnebeek, S.B. Wortmann, M.C. de Vries, M.C.H. Janssen, C. Gilissen, J. Engel, R.A. Wevers, Next-generation metabolic screening: targeted and untargeted metabolomics for the diagnosis of inborn errors of metabolism in individual patients, J. Inherit. Metab. Dis. 41 (3) (2018) 337–353.
[31] T. Yalcin, A.G. Harrison, Ion chemistry of protonated lysine derivatives, J. Mass Spectrom. 31 (11) (1996) 1237–1243.
[32] C. Bleiholder, B. Paizs, Competing gas-phase fragmentation pathways of asparagine-, glutamine-, and lysine-containing protonated dipeptides, Theor.

Chem. Acc. 125 (3) (2010) 387–396.
[33] P.I. Shek, J. Zhao, Y. Ke, K.M. Siu, A.C. Hopkinson, Fragmentations of protonated arginine, lysine and their methylated derivatives: concomitant losses of carbon monoxide or carbon dioxide and an amine, J. Phys. Chem. A 110 (27) (2006) 8282–8296.
[34] E.A. Struys, C. Jakobs, Metabolism of lysine in α-aminoadipic semialdehyde dehydrogenase-deficient fibroblasts: evidence for an alternative pathway of pipecolic acid formation, FEBS (Fed. Eur. Biochem. Soc.) Lett. 584 (1) (2010) 181–186.
[35] T.D. Vaden, T.S.J.A. de Boer, N.A. MacLeod, E.M. Marzluff, J.P. Simons, L.C. Snoek, Infrared spectroscopy and structure of photochemically protonated biomolecules in the gas phase: a noradrenaline analogue, lysine and alanyl alanine, Phys. Chem. Chem. Phys. 9 (20) (2007) 2549–2555.
[36] R. Wu, T.B. McMahon, An investigation of protonation sites and conformations of protonated amino acids by IRMPD spectroscopy, ChemPhysChem 9 (18) (2008) 2826–2835.
[37] R. Wu, R.A. Marta, J.K. Martens, K.R. Eldridge, T.B. McMahon, Experimental and theoretical investigation of the proton-bound dimer of lysine, J. Am. Soc. Mass Spectrom. 22 (9) (2011) 1651.
[38] J. Martens, G. Berden, C.R. Gebhardt, J. Oomens, Infrared ion spectroscopy in a modified quadrupole ion trap mass spectrometer at the FELIX free electron laser laboratory, Rev. Sci. Instrum. 87 (10) (2016) 103108.
[39] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1) (1988) 31–36.
[40] D. Weininger, A. Weininger, J.L. Weininger, SMILES. 2. Algorithm for generation of unique SMILES notation, J. Chem. Inf. Comput. Sci. 29 (2) (1989) 97–101.
[41] G. Landrum, RDKit: Open-Source Cheminformatics, 2006.
[42] T.A. Halgren, Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94, J. Comput. Chem. 17 (5-6) (1996) 490–519.
[43] T.A. Halgren, Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions vol. 17 (5-6) (1996) 520–552.
[44] T.A. Halgren, Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94, J. Comput. Chem. 17 (5-6) (1996) 553–586.
[45] T.A. Halgren, R.B. Nachbar, Merck molecular force field. IV. conformational energies and geometries for MMFF94, J. Comput. Chem. 17 (5-6) (1996) 587–615.
[46] T.A. Halgren, Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules, J. Comput. Chem. 17 (5-6) (1996) 616–641.
[47] P. Tosco, N. Stiefl, G. Landrum, Bringing the MMFF force field to the RDKit: implementation and validation, J. Cheminf. 6 (1) (2014) 37.
[48] J.-P. Ebejer, G.M. Morris, C.M. Deane, Freely available conformer generation methods: how good are they? J. Chem. Inf. Model. 52 (5) (2012) 1146–1158.
[49] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A.V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian 16, Revision A.03, Gaussian, Inc., Wallingford CT, 2016.
[50] L.J.M. Kempkes, J. Martens, G. Berden, K.J. Houthuijs, J. Oomens, Investigation of the position of the radical in z3-ions resulting from electron transfer dissociation using infrared ion spectroscopy, Faraday Discuss. (2019 Apr 23), https://doi.org/10.1039/c8fd00202a.
[51] N.N. Dookeran, T. Yalcin, A.G. Harrison, Fragmentation reactions of protonated α-amino acids, J. Mass Spectrom. 31 (5) (1996) 500–508.