Calculation of mass spectra with the QCxMS method for negatively and multiply charged molecules

Jeroen Koopman and Stefan Grimme

Mulliken Center for Theoretical Chemistry, Institute for Physical and Theoretical Chemistry,
University of Bonn,
Beringstr. 4, 53115 Bonn, Germany
E-mail:grimme@thch.uni-bonn.de

Detailed information about the structural composition of an unknown chemical analyte can be obtained routinely and reliably by using mass spectrometry (MS). Analysis and validation of an MS experiment are usually performed by comparison to reference spectra, which are stored in databases that contain a large number of entries for common molecules. This procedure relies on the quality and completeness of the entries and if structures (classes) are missing, measured spectra cannot be properly matched. To close this gap, and to enable detailed mechanistic analysis, the Quantum Chemical Mass Spectrometry (QCxMS) program has been developed. It enables fully automatic calculations of electron ionization (EI), dissociative electron attachment (DEA), and positive ion collision induced dissociation (CID) mass spectra of singly charged molecular ions. In this work, the extension to negative and multiple ion charge for the CID run mode is presented. QCxMS is now capable of calculating structures carrying any charge, without the need for pre-tabulated fragmentation pathways or machine-learning of database spectra. Mass spectra of four single negatively charged, as well as two multiple positively charged organic ions with molecular sizes ranging from 12 to 92 atoms were computed and compared to reference spectra taken from the literature. The underlying Born-Oppenheimer molecular dynamics (MD) calculations were conducted using the extended tight-binding semi-empirical quantum mechanical GFN2-xTB method while for some small molecules, ab-initio DFT-based MD simulations were performed. Detailed insights into the fragmentation pathways were gained and the effects of the computed charge assignments on the resulting spectrum are discussed. Especially for the negative ion mode, the influence of the deprotonation site to create the anion was found to be substantial. Doubly charged fragments could successfully be calculated for the first time while higher charged structures introduced severe assignment problems. Overall, this extension of the QCxMS program further enhances its applicability and underlines its value as a sophisticated toolkit for CID-based tandem MS structure elucidation.

Introduction

Given the enormous variety of possible compositions of chemicals[1,2], there is a great need for methods that can establish an unambiguous assignment of substances to their chemical structure. Various analytical methods, like NMR, IR, or UV-Vis spectroscopy have been developed that enable structural assignment of unknown compounds. Another accurate and universally applicable technique is mass spectrometry (MS). The method enforces chemical fragmentation of an analyte and measures the mass of its fragmented, as well as non-fragmented components. Evaluation of the fragmentation pathways has led to the development of empirical rules, from which the chemical structure can be deducted.[3–6] But with an increasing number of atoms and functional groups in a molecule, assignment by this method can become exhaustively complicated.[7,8] Alternatively, a measured spectrum can be matched against database stored references with known structure. However, if an entry is missing, a structure cannot be unambiguously assigned. At this point, computational approaches can help to overcome experimental limitations and imperfections as well as provide detailed insight into fragmentation processes. Machine learning (ML) approaches are used for this task in many variations[9–11], but their applicability strongly depends on the existence of high-quality training data and the results often lack interpretability.[12] Tabulation of typical fragmentation patterns[13,14] can be used as a substitute, but this approach lacks flexibility for unknown or untypical dissociation or rearrangement processes.

A way out of this dilemma are computationally affordable quantum chemical (QC) methods. They are generally applicable, do not rely on predefined empirical rules or experimental data, and avoid molecule specific training step as required for ML-type approaches. Based on these ideas, the Quantum Chemical Mass Spectrometry program (QCxMS)[15] was developed, which can operate in x = electron ionization (EI)[16], dissociative electron attachment (DEA)[17] and collision induced dissociation (CID)[18] run modes. The effectiveness of QCxMS to successfully generate in-silico spectra in its EI mode is well documented[19–22] and has been demonstrated recently by its use for extension of mass spectra databases[23–25]. Detailed fragmentation pattern analysis using the EI, DEA, and positive ion CID modes have successfully been conducted earlier.[17,26–28] In this work, an extension of the CID run mode is presented, in which the charge state of the molecular ion is no longer restricted to single positive values so that computations of negatively and multiply charged molecular ions are now possible. This improvement is important because common experimental ionization techniques used in tandem with CID[29–32] can produce ions with multiple positive or negative charges.[33,34] The new charge unrestricted CID mode was tested on a benchmark set of molecules, for which the most apparent fragmentation pathways are discussed in detail.

Four typical representatives of negatively charged metabolites were computed using semi-empirical quantum mechanics (SQM) as well as density functional theory (DFT) based calculations. As QC computations for negatively charged systems require a better description of the more diffuse molecular orbitals, the influence of the level of theory used to calculate the fragmentation reactions and charge assignments was examined thoroughly. The effects of the deprotonation sites of the molecular ion were analyzed and put into perspective regarding the so-called mobile proton theory.[35–38] In addition, two multiply positively charged mass spectra were calculated. To our knowledge, QCxMS is the only freely available program able to compute such spectra without the need for any information other than the geometry (covalent bonding topology, i.e., chemical formula) and charge of the input molecule. In combination with the implemented GFNn-xTB (n=1,2) methods[39,40], the program runs efficiently for any structure consisting of atoms up to radon (Z=86) without the need for third-party-software.

Accounting for the correct dissociation mechanisms of multiple charged structures is of high importance, as it determines if either a single fragment obtains the entire charge or multiple fragments receive separated charges. The effects of multiple charged fragments on computed spectra and possible problems with the correct charge assignment are discussed in detail. This includes the extension of the PlotMS software tool (as part of the QCxMS package) to correctly display the distinct isotope patterns of multiply charged structures.

1 Theoretical Background

1.1 Ionization

In experiments, the negative ion mode is chosen when the analyte cannot be charged positively or if improvements in ionization efficiency and detection limits are required.[41,42] Multiply charged systems typically occur when heteroatoms, such as oxygen and nitrogen, are present in the molecule. This is important in large structures, as manifold charges can reduce the mass-to-charge signal (m/z) of the molecular ion into a measurable area.[43] Also, more information on the fragmentation mechanisms can be gained, because multiply charged fragments produce distinctive isotope patterns.[44,45]

In the most commonly used soft ionization methods[30,31,46–48], the molecular ion is obtained via (de-)protonation of the molecule under consideration, creating positive or negative closed-shell (even-electron) ions. Determining the true (de-)protonated structure of a molecule, from here on called 'protomer'[49], can quickly become a very complex problem. The number of heteroatoms in the molecule determines the number of possible (de-)protonation sites and the most favored one is influenced by various effects. Because the analyte is often ionized from solution after the use of liquid chromatography (LC), solvation effects on the observed (de-)protonation site were investigated.[50,51] Especially pH[52–54] and the influence of different solvents[41,55] were studied and it was found that the preferred protonation site is significantly diverse when the ion is in gas or liquid phase.[56–59]. However, a general statement

about which phase is mostly relevant in standard MS measurements was not provided. Another effect influencing the favored (de-)protonation site is that upon activation of the molecular ion during the CID process, reaction barriers between different protomers can be overcome. This leads to proton transfer between the structures so that thermodynamically less favored protomers can undergo rapid, kinetically favored fragmentation and hence are observed in the spectrum.[60–62] This is known as the so-called mobile-proton effect,[35–38] which states that the energetically lowest structure must not necessarily yield the dominant fragmentation pattern. In QCxMS, the CID run mode requires a (de-)protonated molecular ion as starting structure. All possible protomers in a user-defined energy range can be calculated by a generally applicable (de-)protonation protocol[63], that is implemented in the Conformer Rotamer Ensamble Sampling Tool (CREST)[64–66]. This procedure can be applied iteratively, so that multiply (de-)protonated molecular ions are obtained. Ranking of the protomers with their relative free energies can be conducted by using the Command-line ENsemble SOrting algorithm (CENSO),[67] in which gas-phase computations can be compared to liquid phase results calculated with different solvents.

1.2 Method

Activation of a molecular ion after 'soft-ionization' is commonly achieved using CID[68,69], in which the analyte is forced to undergo (multiple) collisions with a neutral gas, e.g., argon or dinitrogen. In QCxMS, this process was implemented by performing consecutive Born-Oppenheimer molecular dynamics (MD) simulations. The underlying potential energy surfaces (PESs) are calculated 'on-the-fly' using QC methods. A simplified flowchart of the workflow is provided in figure 1.

Preparation of the input geometry is recommended as described before by using CREST and CENSO. For the (de-)protonated structure, an ensemble of snapshot geometries is taken along a sampling MD and used as starting geometries for the following fragmentation simulations (production runs). The sampling of the CID process is done using the *general activation* run type, in which fragmentation is induced by thermal heating, collisional activation, and consecutive mean-free-path MD simulations. The individual steps of the *general activation* run type are color coded in figure 1. More details on the implementation are given in the original publication.[18]

1.3 Charge assignment

Only charged structures can be measured in MS experiments. In QCxMS, the delta self-consistent field (ΔSCF)[70] method is used to determine the charge distribution between the created fragments after dissociation. The difference in energy between the neutral and ion fragment structure is the ionization potential (IP) for positive ions and electron affinity (EA) for negative ions, respectively. The statistical charge q for each fragment i is determined using Boltzmann statistics with the IPs (or EAs) and the average internal molecular temperature T as factors, with k_B
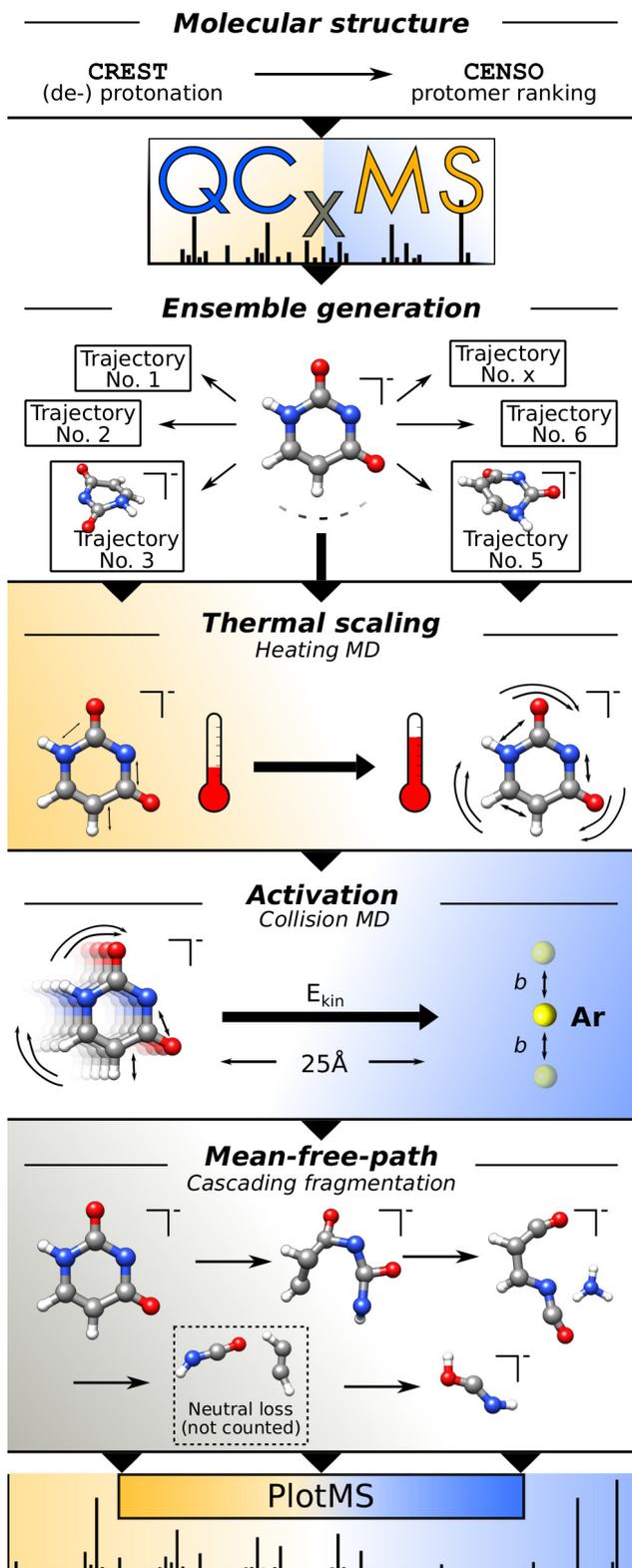
Figure 1 Schematic diagram of the QCxMS workflow. First, an ensemble of MD snapshot structures is created. Consecutive simulations first increase the temperature, then simulate the collision and the mean-free-path. After fragmentation, the spectrum is plotted with PlotMS.

being the Boltzmann constant according to

$$q_i = \frac{exp(-\frac{IP_i}{k_B T})}{\sum_{j=1}^{n} exp(-\frac{IP_j}{k_B T})}.$$

For systems with multiple charges, the IP/EA of each individual fragment is computed with all possible combinations of charge states. For a fragment to obtain several charges, the IP/EA of the fragment must be lower than the combined IPs/EAs of its singly charged counterparts. An example calculation is given in table 1.

Table 1 IP calculations for dividing two charges between two example fragments 1 & 2. Summation of the IPs in the left case assigns a single charge on each fragment, as $\sum$ 1-1 $<$ $\sum$ 2-0 $<$ $\sum$ 0-2. In the right case, adding the IPs assigns both charges on fragment #1, as $\sum$ 2-0 $<$ $\sum$ 1-1 $<$ $\sum$ 0-2

Fragment	Charge	IP (eV)	Fragment	Charge	IP (eV)
#1	$0 \to 1$	1.0	#1	$0 \to 1$	1.0
	$0 \to 2$	3.0		$0 \to 2$	2.0
#2	$0 \to 1$	1.5	#2	$0 \to 1$	1.5
	$0 \to 2$	3.5		$0 \to 2$	3.5
	$\sum$			$\sum$	
	1-1	**2.5**		1-1	2.5
	2-0	3.0		2-0	**2.0**
	0-2	3.5		0-2	3.5

If the fragments from a dissociation event have sufficient internal energy, consecutive dissociation cascades can occur. In QCxMS, the fragment with the **lowest IP** or **highest EA** is taken to undergo subsequent fragmentation MDs. When the IP/EA values for two fragments are similar, the size of the fragments is taken as the decisive factor. Because larger molecules are more likely to undergo consecutive dissociation, the statistical charge of each fragment is multiplied by its number of atoms. In the current form, QCxMS is only capable to compute one fragmentation cascade for each fragment. For multiple charged structures, only the fragment with the highest charge and largest size is used for the cascade. Other fragments are just stored and do not undergo further fragmentations. The computed fractional Boltzmann charges are used as in the original QCEIMS algorithm.[16,22]

1.4 Plotting spectra

The counting of the fragments and plotting of the theoretical spectrum is done by the external Plot Mass Spectrum (PlotMS) program. For this work, it was enhanced to compute 'exact' masses instead of integer masses. Furthermore, the isotope patterns of multiply charged structures can now be calculated. This is important, because the charge z of a fragment can be deduced experimentally from its isotope pattern, as the spacing of the isotope peaks decreases with $1/z$. For an overall assessment of the agreement between experiment and theory, the program yields a (weighted dot product) spectral matching score.[71]

2 Technical details

2.1 Benchmark Molecules

For the negative ion mode, typical metabolite structures were tested including linear and cyclic functional groups, as well as different heteroatoms. To enable extensive testing, the structures were chosen to be small with molecular sizes of $13 - 26$ atoms after deprotonation. However, database entries for small

2-Ketoburytic Acid
(1)

3-Ureidoprpionic Acid
(2)

Ascorbic Acid
(3)

Tryptophan
(4)

Crizotinib
(5)

Lysine Derivate
(6)

Figure 2 Benchmark set of molecules for testing different charge states with QCxMS.

negatively charged molecules often only display the molecular ion peak. The availability of fragment signals was a determining factor for adding a structure to the benchmark. The final set consists out of the organic molecules 2-ketoburytic acid **(1)**, 3-ureidopropionic acid **(2)**, ascorbic acid **(3)**, and tryptophan **(4)**, which are shown in figure 2. Reference spectra were taken from the human metabolome (HMDB)[72–76] and the MassBank[77,78] databases.

For spectra of multiply charged species, computing fragments that retain more than a single charge is of interest. In small molecules, the Coulomb repulsion of many charges leads to fast dissociation and thus low signal abundance.[79] Such systems were thoroughly studied[80], but exclusively dissociate into singly charged fragments. In large molecules, multiple (de-)protonated fragments can be obtained in high abundance.[34] However, running MD simulations for large structures can become computationally expensive. The fragmentation pathways of doubly protonated crizotinib **(5)**[81] and a triply protonated lysine derivate **(6)**[82] were studied and reported multiply charged fragment signals. With 54 and 93 atoms after protonation, MD simulations at the semi-empirical quantum mechanical (SQM) level were af-

fordable.

2.2 Computational and technical details

All calculations in this work were performed on Intel ® Xeon ® E3-1270 3.60GHz CPU cores. The QCxMS code version 5.2 was used throughout which is available open-source.[15] For CID calculations, the *general activation* run type was utilized. Argon was used as the collision gas with a pressure of 0.132 Pa and a collision cell length of 0.25 meters.

The automatic (de-)protonation runs were conducted using the CREST[64–66] version 2.11.3. Free energy ranking of the ensemble was done using CENSO[67] version 1.2.0 at the r^{2}SCAN-3c[83] DFT composite level of theory. A structure is ranked by its free energy difference (ΔG) to the most populated protomer (usually protomer #1). The relative free energies were calculated at temperatures of 300 K, 600 K, and 1200 K in the gas phase. If not stated otherwise, 600 K was used as the default. Solvation effects on the rankings were investigated with the implicit solvation model COSMO-RS[84,85]. Methanol was used to represent a protic solvent commonly used in LC-MS experiments. For all structures, the relative free energy values in gas and solvent can be found in the supporting information (SI). DFT calculations were executed either using ORCA[86,87] version 5.0.3 or TURBOMOLE[88] version 7.5.1. The MD calculations in QCxMS are mostly based on the GFN2-xTB method[40,89] in combination with the finite electronic temperature (Fermi smearing) model. It is set to a temperature of 298 K in the ground state sampling step and increased to 5000 K in the production runs. MD steps were carried out using the leap-frog algorithm with a time step of 0.5 fs. Reasonable statistical convergence of the spectral results was obtained by calculating a number of trajectories equal to 25 times the number of atoms per molecule.

Earlier work on dissociative electron attachment (DEA)[17] showed that calculation of the EA values on the DFT level can improve the computed spectra. Negatively charged anions require the inclusion of diffuse basis functions to correctly describe the more loosely bound outer valence electrons. For cross-checking, the D4[90] dispersion corrected PBE[91] and PBE0[92] functionals with def2-SV(P)[93], def2-TZVP[94], as well as the the minimal augmented ma-def2-XVP[95] (X=S,TZ) basis sets were used. The latter increases accuracy for EA calculations. In the following, combinations of QC levels for PES and IP/EA calculations are written as [PES method]//[IP/EA method]. Calculations of the multiply charged structures were done using GFN2-xTB//GFN2-xTB.

Reference spectra taken from the literature were measured between 10 and 50 eV laboratory frame collision energy (E$_{LAB}$) using the *Waters Micromass Quattro Triple Quadrupole* mass spectrometer (LC-ESI-QQQ), the *Bruker Maxis Impact* mass spectrometer (LC-ESI-QTOF) or the *Applied Biosystems API3000* mass spectrometer (LC-ESI-QQ).

2.3 Differences between experiment and theory

The various reasons for observed differences between experimental and calculated spectra were discussed in detail in ear-

lier work.[18,28] In experiments, the 'hardness' of the ionization process influences the degree of fragmentation.[96,97] The conditions in the collision cells are device-specific and cannot exactly be reproduced by the simulation.[98] Thus, collision cell settings used in QCxMS are determined empirically and do not necessarily reflect the instrumental specifics. Discrepancies in activation time, ionization energy, and velocity can lead to deviating fragmentation behavior. Other experimental conditions (e.g., cooling effects, photon excitation, etc.) are not accounted for in the simulations. Furthermore, using SQM methods for the MD simulations can introduce severed errors in the underlying PES, leading to artifacts or incorrect signal intensities. On the theoretical side, this is presumably the most important factor.

When multiple charges are present, the electric field acceleration of an ion is greater by the factor of its charge. In the current version of QCxMS, the velocity of the ion is scaled uniformly and does not account for the molecular charge.

Most tandem MS instruments do not have the resolution to display isotope patterns. For better comparison of experiment and theory, isotope pattern calculations with PlotMS were switched off in the computations on negative ion mass spectra in the following part.

3 Results and discussion

3.1 Negative Charges

3.1.1 Ketoburytic acid

The smallest benchmark structure, 2-ketoburytic acid **(1)**, is deprotonated either at the carboxylic acid group (protomer #1) or less likely at the α-alkyl carbon (protomer #2, $\Delta G > 20$ kcal/mol).

In figure 3, spectra of protomer #1 computed using different combinations of QC methods for [PES]/[EA] calculations are shown. The used levels of theory are depicted at the corresponding spectrum. For validation, the results were compared to a database reference.[74] A computed spectrum for protomer #2 is provided in the SI.

In all database spectra for 2-ketoburytic acid, the experimental m/z peaks are shifted compared to the values at which these signals should appear based on their mass. In the example used here, signals m/z 57.462, 55.576, and 45.605 are unexpected, as there is no combination of atoms available that would sum up to these values. These differences originate from inaccuracies in the signal resolution of the instrument. Signals lower than $\sim$ m/z 45 are not measured in the experiment, apparently due to a mass cutoff.

Calculations using GFN2-xTB//GFN2-xTB shown in figure 3 a) describe the dissociation of neutral CO and CO_2 and produce fragments at signals m/z 73.028 and 57.034. The latter fragment dissociates further by H_2 loss, creating signal m/z 55.018. Signal m/z 44.998 (HCO_2^-) is produced from protomer #2 (see SI).

The spectrum in figure 3 b) was calculated with the GFN2-xTB//PBE0/ma-def2-TZVP method combination. The most pronounced difference to the full SQM approach shown in a) is the higher abundance of CO_2^- (m/z 43.989). Otherwise, the use of DFT for EA calculations does not significantly change the spectrum.

A full-DFT calculation was conducted at the PBE/ma-def2-SV(P)//PBE/ma-def2-TZVP level, which is displayed in figure 3 c). The same collision energy used before leads to stronger dissociation of [M-H]$^-$ in this spectrum. The structure dissociates more frequently into CO_2^- (m/z 43.989), CO and $C_2H_5^-$ (m/z 29.038). Signal m/z 55.018 is missing. Overall, fewer fragmentation pathways are computed, leading to a lower variety of signals when compared to figures 3 a) and b).

In figure 3 d), a full-DFT calculation with the more sophisticated PBE0/def2-SV(P)//PBE0/def2-TZVP hybrid DFT level was conducted. The fragmentation pathways are similar to those displayed in figure 3 c). However, the fragment signal intensity is lower, indicating lower [M-H]$^-$ dissociation rates.

Overall, the experimental spectrum is well reproduced by all theory levels.

3.1.2 Ureidopropionic acid

For ureidopropionic acid **(2)**, the three most populated structures are deprotonated at the carboxylic acid group and form a tautomeric network at the diamide functional group. Free energy differences are small in the gas phase ($\Delta G < 2$ kcal/mol), but more distinct in solution ($\Delta G > 10$ kcal/mol, see SI). Another protomer #4 is formed by removal of the α-hydrogen atom, but the structure is not significantly populated.

Mass spectra of all protomers were calculated using GFN2-xTB//GFN2-xTB and compared to the experimental reference[74] as shown in figures 4 a) - d). Protomer structures and free energy differences relative to the most populated protomer #1 are given in addition to their respective mass spectrum.

All simulated spectra show good a representation of signal m/z 59.024 ($H_3N_2CO^-$). However, the influence of the protomer structures is significant for producing signal m/z 88.03. Intramolecular proton transfer to the secondary amine leads to the formation of fragment $H_2C_2H_4CO_2^-$ and neutral HN=C=O, which is described sufficiently by protomer #3 in figure 4 c). While such a transfer is also possible from the other heteroatoms, the mobile proton is more likely to relocate from the neighboring hydroxyl group than from the terminal amine or carboxyl groups. This is reflected by the different matching scores for the structures in figure 4.

In figure 5 a), the spectrum of protomer #3 was computed using PBE/ma-def2-SV(P)//PBE/ma-def2-TZVP. The result is in excellent agreement with the experiment. Signal m/z 88.039 is observed in high abundance, which indicates a good description of the above described mobile proton transfer. A mixed GFN2-xTB//PBE0/ma-def2-TZVP approach was used to produce the spectrum in figure 5 b). The resulting matching score of 514 indicates a slight improvement to the score of 418 obtained by the SQM//SQM calculations in figure 4 c).
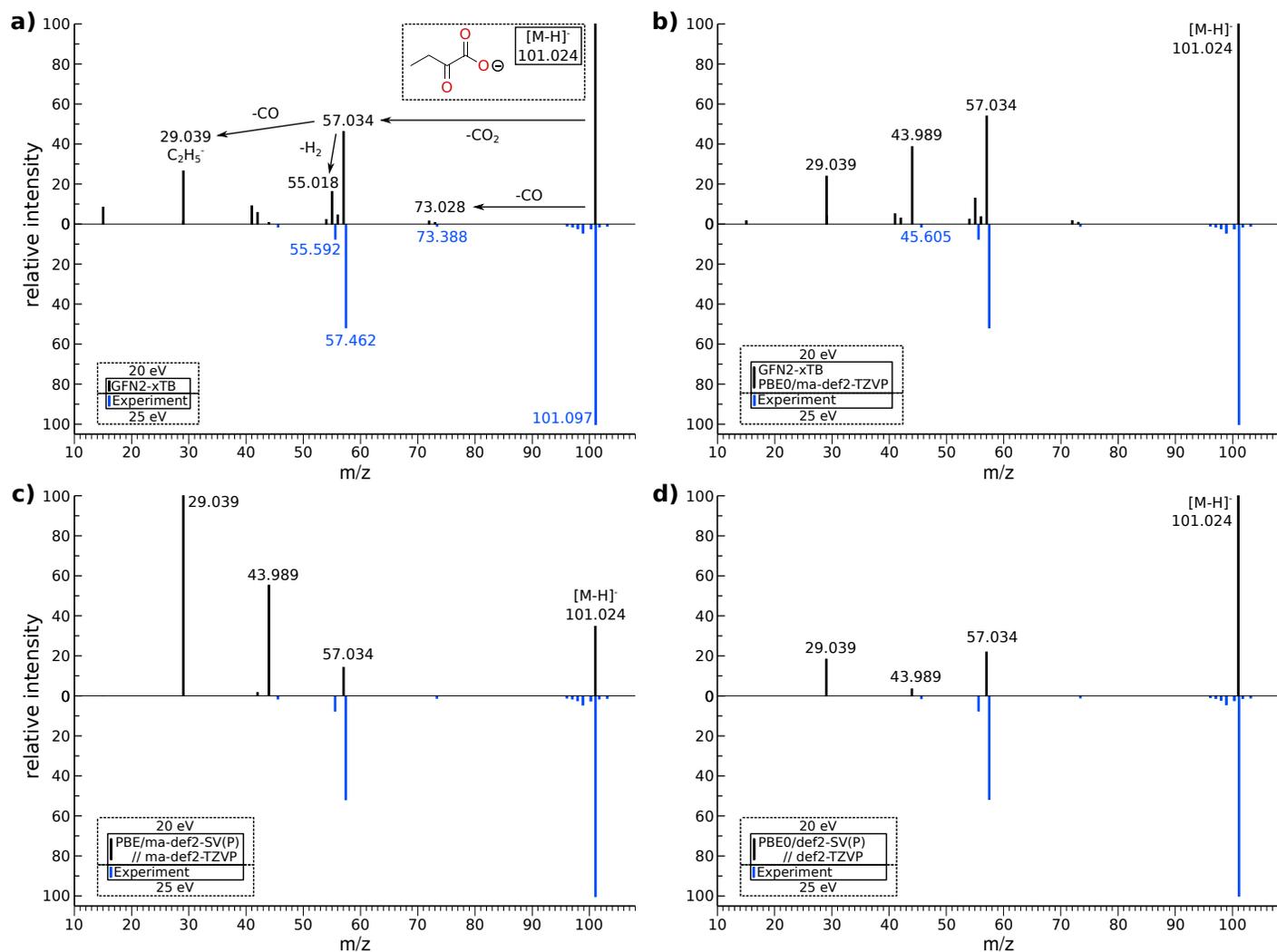
Figure 3 Calculated spectra of 2-Ketoburytic acid (black,top) computed at 20 eV E_{LAB} compared to a measured spectrum (LC-ESI-QQQ) at 25 eV E_{LAB} (blue, inverted). a) GFN2-xTB//GFN2-xTB, b) GFN2-xTB//PBE0/ma-def2-TZVP, c) PBE/ma-def2-SV(P)//PBE/ma-def2-TZVP, d) PBE0/def2-SV(P)//PBE0/def2-TZVP.

3.1.3 Ascorbic acid

Deprotonation of ascorbic acid **(3)** leads to five protomers with up to 35 kcal/mol difference in free energy. Protomer structures are depicted in figure 6 alongside their computed mass spectra. Calculations were performed at GFN2-xTB//GFN2-xTB level and the results are compared to a database spectrum.[72] The use of DFT for EA calculations did not significantly improve the overall accuracy and these spectra can be found in the SI.

The base peak of the experimental spectrum is produced by bond cleavage between the neutral ethanediol fragment (HO-C_2H_4-OH) and the negatively charged 2,3-hydoxyfuran fragment (*m/z* 115.003). For a better distinction in the following, ethanediol is called 'side-chain' and 2,3-hydoxyfuran is called 'backbone'.

The computed mass spectrum and structure of protomer #1 is displayed in figure 6 a). The molecular ion is deprotonated at the backbone. Fragmentation between backbone and side-chain produces a high abundance of signal *m/z* 113.995. The corresponding fragment structure was added to the figure. The spectrum of protomer #2 is similar to that of protomer #1 and can

be found in the SI.

The spectrum of protomer #3 shown in figure 6 b) has the lowest matching score. It is deprotonated at the carbon atom connecting the backbone to the side-chain. The resulting destabilization of the bond between the structures leads to high fragmentation rates already at low collision energies.

Protomers #4 (see figure 6 c)) and #5 (figure 6 d)) are deprotonated at either of the two hydroxyl groups of the side-chain. In protomer #4, deprotonation of the outermost hydroxyl group leads to a ring formation and a shift of the double bond inside the backbone. Because the backbone is not deprotonated, the experimental base peak is reproduced and matching scores are high. However, protomers #4 and #5 are not significantly populated, neither in gas nor when solvation effects are included (see SI). This indicates rearrangement reactions via mobile protons between the protomers before dissociation takes place.
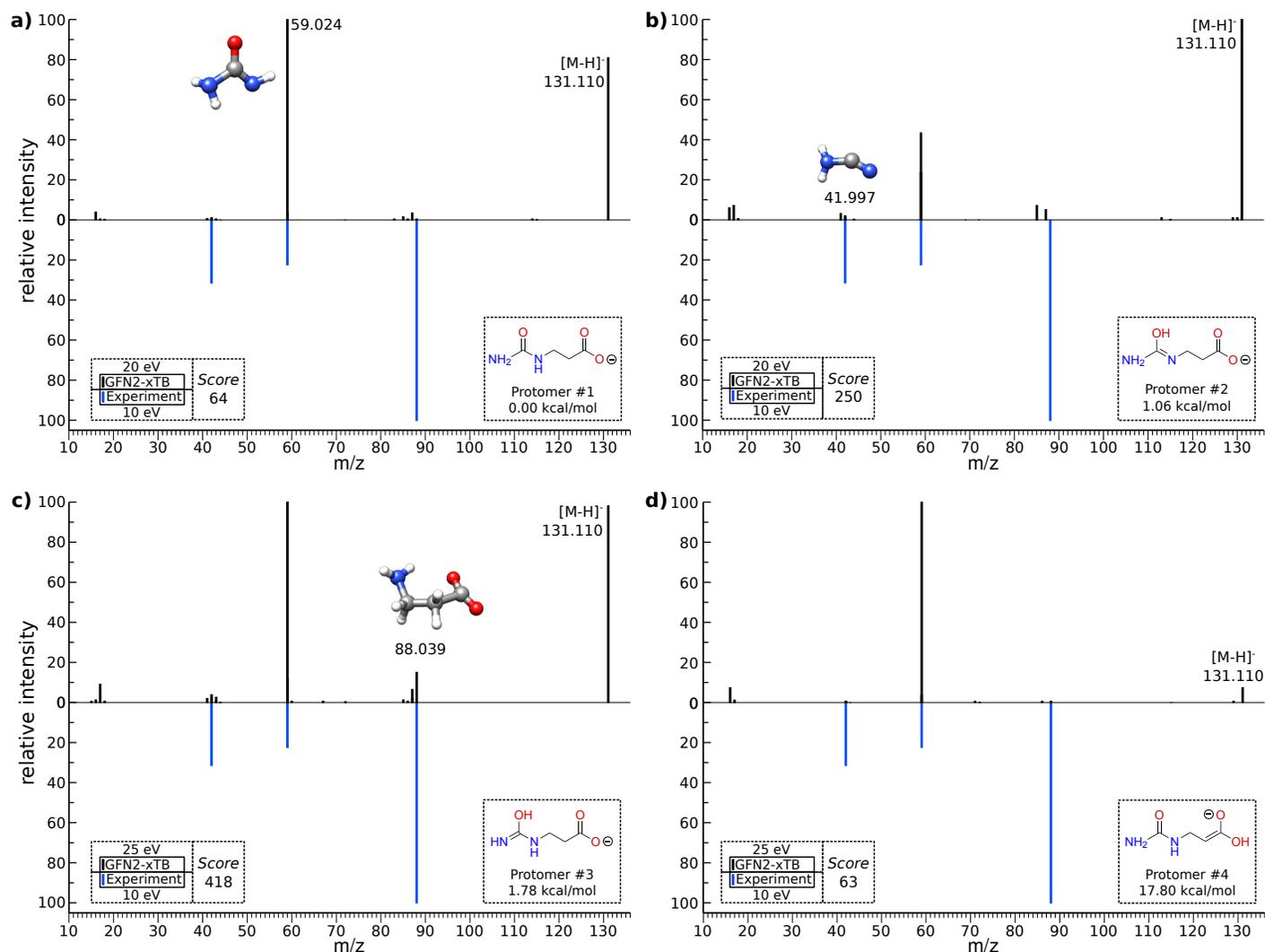
Figure 4 Calculated spectra (black, top) of the four protomers of ureidopropionic acid using GFN2-xTB//GFN2-xTB compared to reference (LC-ESI-QTOF) at 10 eV E$_{LAB}$ (blue, inverted). Protomer structures, relative free energies, and spectral matching scores are added to their respective spectrum.
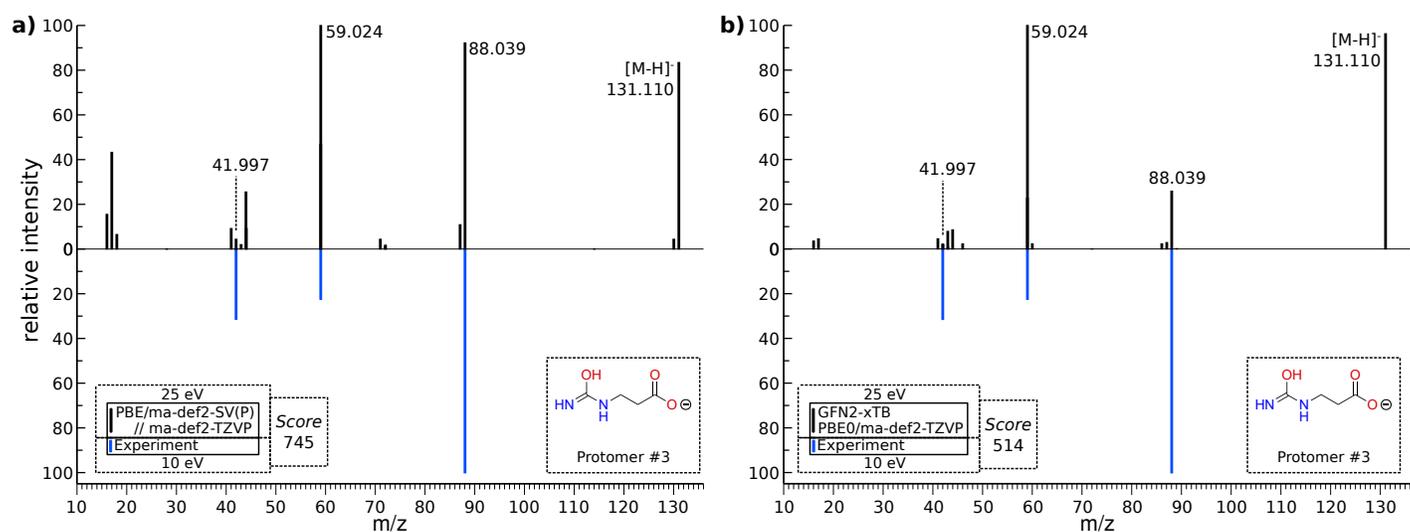
Figure 5 Calculated spectrum (black, top) of ureidopropionic acid protomer #3 compared to reference (LC-ESI-QTOF) at 10 eV E$_{LAB}$ (blue, inverted). a) PBE/ma-def2-SV(P)//PBE/ma-def2-TZVP. b) GFN2-xTB//PBE0/ma-def2-TZVP. Protomer structures, relative free energies, and matching scores are given with their respective spectrum.

7
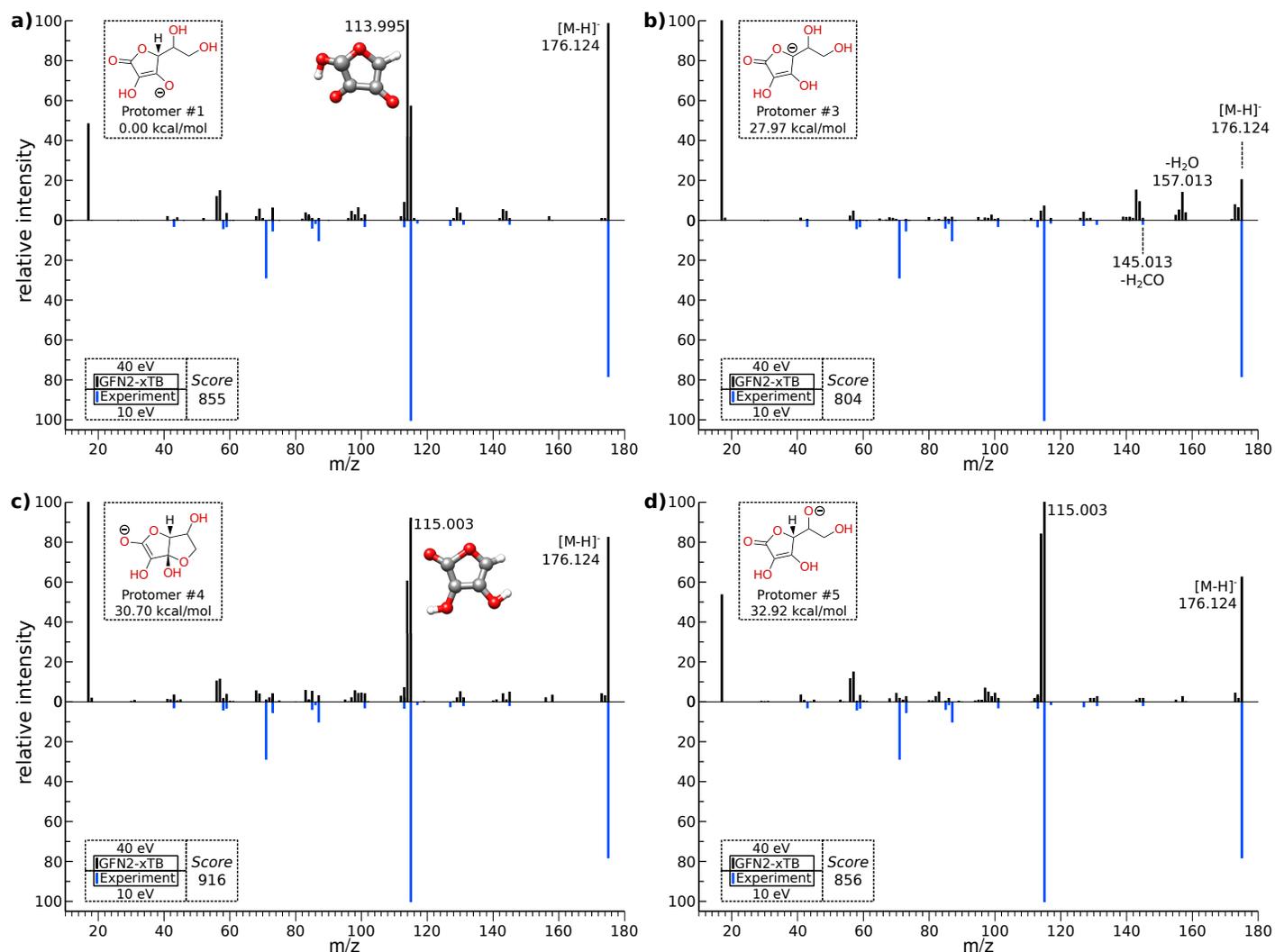
Figure 6 GFN2-xTB//GFN2-xTB calculated spectra (black, top) at 40 eV E_{LAB} of ascorbic acid compared to measured spectrum (LC-ESI-QQQ) at 10 eV E_{LAB} (blue, inverted). Protomer structures, relative free energies, and matching scores are added to their respective spectrum. The structures attributed to signals *m/z* 113.995 and 115.003 are added for reference. The spectrum of protomer #2 can be found in the SI.

3.1.4 Tryptophan

Tryptophan (4) has four protomers in a 50 kcal/mol free energy window. The structures and the relative free energies are depicted in figure 7. Protomers #1 and #2 are similarly populated in the gas phase, while the ranking in solvation is more distinct (see SI). QCxMS calculations on all protomer structures were conducted at GFN2-xTB//GFN2-xTB level and compared to a database spectrum[99] in figures 7 a)–d). Using DFT methods for EA calculations did not significantly improve the results and the spectra can be found in the SI.

Protomer #1 is deprotonated at the carboxyl acid group. The structure and calculated spectrum are shown in figure 7 a). Apparently, the simulated spectrum shows only a weak match to the experimental, which is reflected by the matching score of 313. Direct dissociation of [M-H]$^-$ by neutral CO_2 loss produces signal *m/z* 159.092. In a second step, single hydrogen atom separation forms the most abundant signal *m/z* 158.084, which is not present in the experimental spectrum.

Protomer #2 is formed through the deprotonation of the pyrrole nitrogen. It has the highest matching score with 666. The good agreement between the calculated and the experimental spectra is depicted in figure 7 b). Three main fragmentation reactions of [M-H]$^-$ were observed. First, proton transfer from the carboxyl acid to the neighboring amine group leads to NH_3 elimination (*m/z* 186.055) and subsequent CO_2 dissociation (*m/z* 142.065). Second, heterolytic fragmentation of the side-chain (C_2H_2-NH_2-CO_2H) leads to the deprotonated indole fragment (*m/z* 116.050). Third, signals *m/z* 129.057 and 74.024 are formed by homolytic dissociation. Part of the side-chain dissociates as an NH_2-CH-COOH glycine derivate, while a CH_2 group remains bound to the deprotonated indole fragment (see figure 7). Both fragments obtain statistical charge, while signal *m/z* 74.024 is more pronounced in the calculations.

The computed spectra of protomers #3 and #4 (figures 7 c) and d)) display considerable discrepancies to the experimental spectrum and low populations render these structures irrelevant for the observed spectrum.
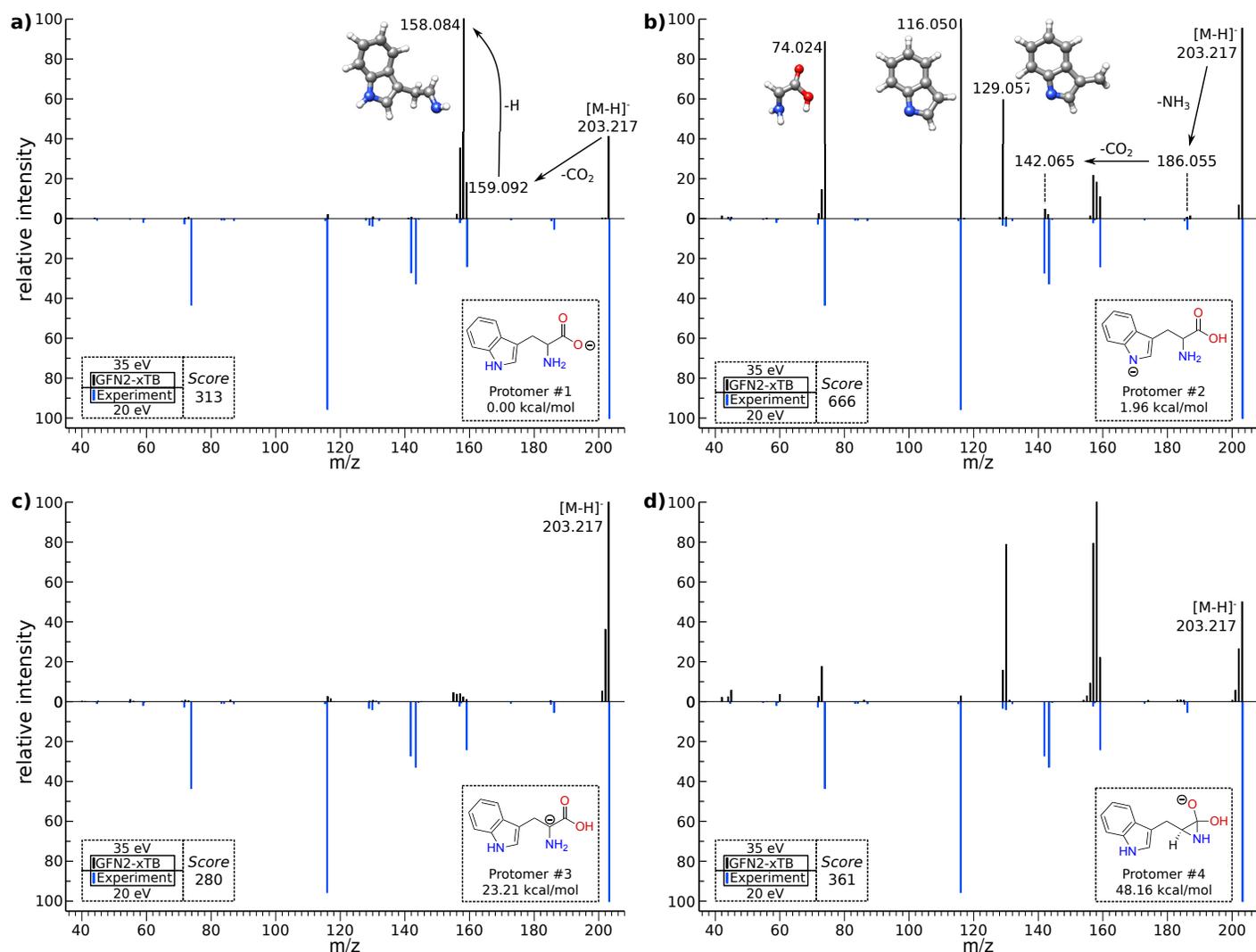
Figure 7 Spectra calculated at at GFN2-xTB//GFN2-xTB (black, top) of the four protomers of tryptophan at 35 eV E_{LAB} compared to a measured spectrum (LC-ESI-QQ) at 20eV E_{LAB} [99] (blue, inverted). Protomer structures, relative free energies, and matching scores are added to their respective spectrum.

3.1.5 Level of theory for negative charge CID

Computational cost is a significant factor in choosing the level of theory for the computations. For the benchmark molecules **(1)** – **(4)** (see figure 2), timings of the different [PES]/[EA] combinations are provided in table 2.

Table 2 Average timings [min] per fragmentation MD for mass spectrum calculations of the benchmark molecules **(1)**–**(4)**. Different QC level combinations for PES and EA calculations were used when affordable

PES level	EA level	time [min]			
		(1)	**(2)**	**(3)**	**(4)**
GFN2-xTB	GFN2-xTB	1.5	3.5	4	17
GFN2-xTB	PBE0/ma-def2-TZVP	10.5	14	156	335
PBE/ma-def2-SVP	PBE/ma-def2-TZVP	742	4340	–	–

Computations using the full GFN2-xTB//GFN2-xTB method take on average between 1-2 and 17 minutes for a single fragmentation MD. The more sophisticated GFN2-xTB//PBE0/ma-def2-TZVP approach increases the computation times dramatically to up to 5 hours for **(4)**. With over three days (4340 minutes) computation time for a fragmentation MD for **(2)**, the full-DFT PBE/ma-def2-SVP//PBE/ma-def2-TZVP approach is three orders of magnitude more expensive than using GFN2-xTB.

The observations made in the calculation of negative ion mass spectra presented here indicate that the choice of the initial protomer structure is of greater importance in describing the correct fragmentation pathways than investing in EA computations at DFT levels. This conclusion is in accordance with Field's rule [100], which states that for soft-ionization-based methods the protonation state of a fragment is of major relevance for its signal intensity. This contrasts earlier work on DEA [17] which follows Stevenson's rule [101] for hard-ionization-based methods. Overall, the good performance of QCxMS in combination with the implemented GFN2-xTB Hamiltonian seems to provide fast and reliable results for CID mass spectra in negative ion mode for common organic molecules.

3.2 Multiple Charges

3.2.1 Crizotinib

The fragmentation pathways of doubly protonated crizotinib **(5)** were studied previously.[81] Figure 9 a) shows the measurement of the singly positive charged molecular ion ([M+H]$^+$, *m/z* 450.1266). Figure 10 a) depicts the experimental full MS/MS scan that includes the singly protonated [M+H]$^+$ (*m/z* 450.1266) and doubly protonated [M+2H]$^{2+}$ (*m/z* 225.5662) species. The proposed fragmentation scheme by Joyce and Richards is displayed in figure 8. The black roman numerals denote the reactions described in the literature, while gray numerals indicate alternative fragmentation pathways computed by QCxMS.
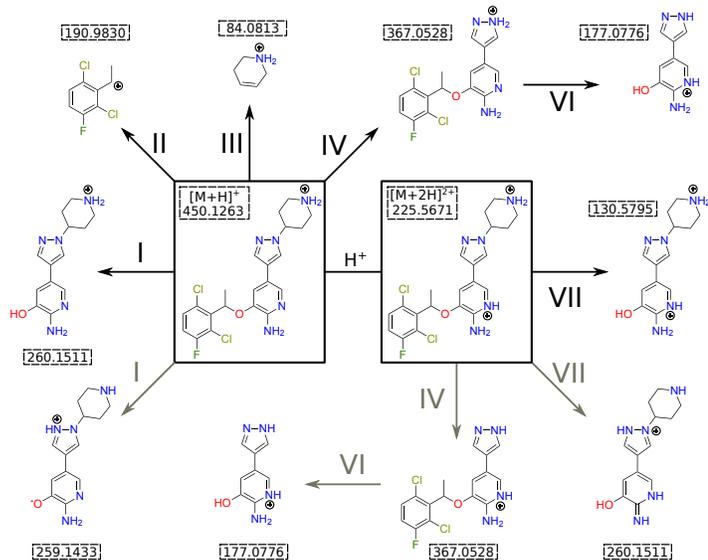
Figure 8 Dissociation reactions as proposed in ref.[81] are marked with black roman numerals and arrows. Alternative fragmentation reactions computed with QCxMS are indicated with gray roman numerals and arrows.

In the literature, the most populated protomer was determined by the computation of the most basic sites of the neutral structure in aqueous media. For validation, CREST and CENSO were utilized in this work to verify the reported findings. The three most populated protomers computed here are displayed in table 3 with their free energy ranking in water, methanol, and the gas phase at 300 K. The most populated protomer #1 in water and methanol is the same as reported in the literature. In the gas phase, protomer #1 is not significantly populated.

The calculated spectrum of protomer #1 is displayed in figure 9 b). All experimental fragments were calculated correctly, however with a mass discrepancy. The computations produce signals *m/z* 366.0450, 259.1433, and 175.0619 instead of the measured signals *m/z* 367.0515, 260.1502, and 177.0767. In the scheme of figure 8, reaction I is proposed in the experiment as a heterolytic dissociation of [M+H]$^+$, producing signal *m/z* 260.1502. In contrast, the QCxMS calculations favor homolytic dissociation and a radical ion fragment (*m/z* 259.1433) is formed. Interestingly, various homolytic dissociation reactions were reported in the literature, that are produced by other protomers. Signal *m/z*

Table 3 Free energy differences of the three most populated structures of crizotinib in water, methanol, and in the gas phase. Computed at 300 K with CENSO (ΔG in kcal/mol)

CENSO	#1	#2	#3
Water	0.00	2.06	10.13
Methanol	0.00	1.87	10.64
Gas	11.23	0.00	7.04

259.1433 was described to have an odd electron structure, supporting the findings of QCxMS. However, the computed spectra of protomers #2 and #3 are similar and display the same overall fragmentation patterns as protomer #1. These spectra can be found in the SI.

Using CREST and CENSO, the second protonation in water at a temperature of 300 K leads to the same doubly charged protomer #1 as proposed in the literature. The computed spectrum is depicted in figure 10 b). A spectrum of the second most populated protomer #2 (ΔG 10 kcal/mol) can be found in the SI.

As mentioned in section 1.4, the charge of an ion is reflected by its isotope pattern. Using PlotMS, the isotope patterns of all signals were calculated for the theoretical spectrum and are shown enlarged in figures 9 b) and 10 b). Δ*m/z* of 1 indicates a single charge present in the signal, while Δ*m/z* of 0.5 indicates two charges. Δ*m/z* of 2 is the isotope abundance of the chlorine atoms. By the comparison of the computed isotope patterns of [M+H]$^+$ (*m/z* 450.1263, Δm/z 1) in figure 9 b) to [M+2H]$^{2+}$ (*m/z* 225.5671, Δm/z 0.5) in figure 10 b), it is evident that the software is able to distinguish multiple charged signals from single charged ones. More complicated, fragment signals *m/z* 260.1511 and 130.5791 are of interest concerning the capabilities of QCxMS and PlotMS. The structures are similar, but the latter carries one proton more and is thus charged twice. It is visible that the latter fragment was correctly assigned a twofold charge with QCxMS (figure 10 b), bottom left), as Δ*m/z* between the signals is 0.5. This proofs that our approach is able to compute and assign multiply charged fragments with the Δ SCF method, for which details were described in section 1.3. In table 4, the calculated IP values of fragments *m/z* 260.1511 and 130.5791 are listed.

Because QCxMS is currently only able to calculate a spectrum of either single, doubly,..., multiply protonated species at a time, the signal of [M+H]$^+$ (*m/z* 450.1266) is not present in figure 10 b). In contrast to the literature, signals *m/z* 367.0528 and 260.1511 are formed from the doubly protonated molecular ion and not from the singly charged protomer (see figure 8 IV and VII). In the computed spectrum, signal *m/z* 177.0776 is underrepresented and signal *m/z* 176.0698 is more abundant. The latter is an odd electron structure formed through homolytic bond dissociation and described in the literature as an impurity.

Overall, QCxMS successfully matches all experimentally re-
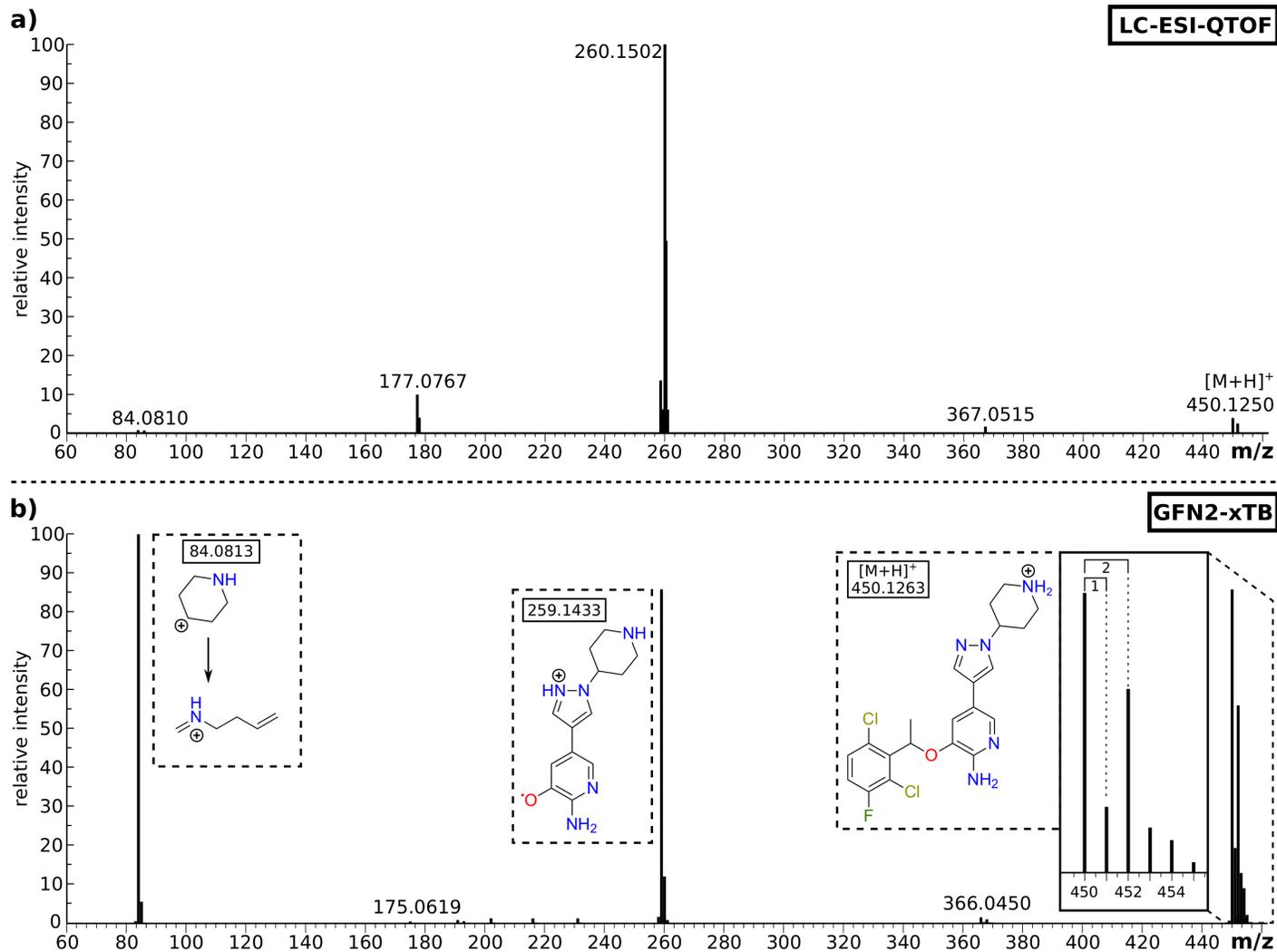
Figure 9 a) measured spectrum of singly protonated crizotinib (LC-ESI-QTOF; 23 eV E_{LAB}) [81].
b) Computed spectrum (GFN2-xTB//GFN2-xTB at 50 eV E_{LAB}) of singly protonated crizotinib. Isotope patterns computed with PlotMS are enhanced for specific signals.

Table 4 Ionization potential calculations on the fragments m/z 190.983 and 260.151 compared to fragments m/z 189.975 and 261.158. Summation of the potentials show that in the first case the charge is split between the two fragments (Sum 1-1), while in the latter two charges remain on fragment m/z 261.158, leading to signal m/z 130.579 (Sum 0-2)

m/z	Charge	IP (eV)	m/z	Charge	IP (eV)
190.983	1	**11.65**	189.975	1	13.13
	2	29.10		2	31.04
260.151	1	**8.55**	261.158	1	7.77
	2	22.51	(130.579)	2	**19.49**
	Σ			Σ	
	1-1	**20.20**		1-1	20.90
	2-0	29.10		2-0	31.04
	0-2	22.51		0-2	**19.49**

GFN2-xTB calculations. However, the protonation sites of the fragments are different in our computations than reported in the literature. This is due to the equilibration of the structure to its gas phase geometry (see protomer #2 in table 3), which is not accounted for in the reference fragmentation scheme (figure 8). Signal intensities between measured and calculated spectrum are different, which is directly related to the discrepancies between experiment and calculation discussed in section 2.3. Increasing the level of theory for IP calculations at DFT level did not increase the overall accuracy (see SI).

ported signals for this compound. Using IP calculations to correctly allocate multiple charges to single fragments was accurate. Appropriate proton assignment after fragmentation was crucial for recreating the right signals, which was accounted for by the
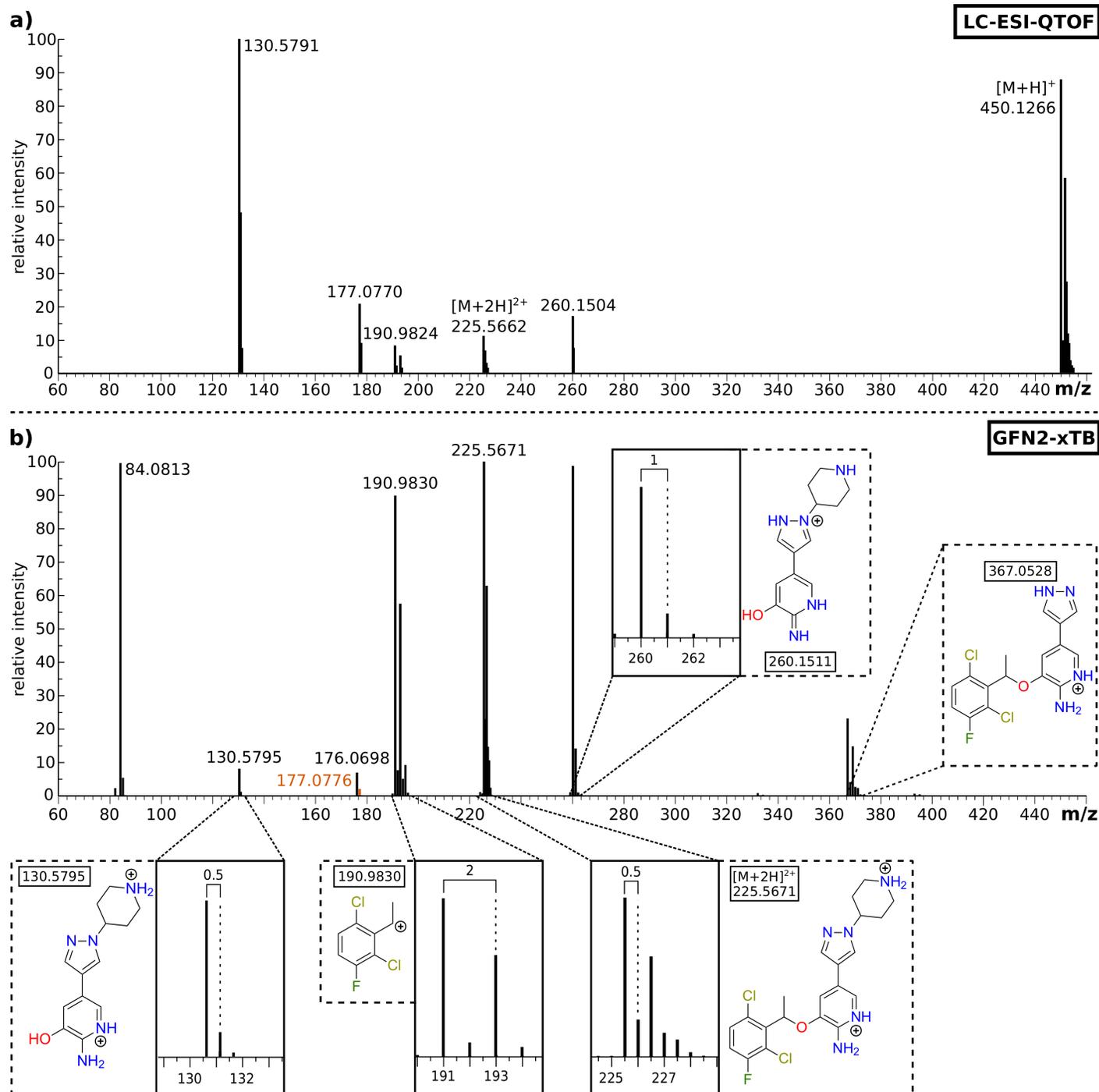
Figure 10 a) Full scan MS of crizotinib (LC-ESI-QTOF; 25 eV E_{LAB}) [81].
b) Computed spectrum (GFN2-xTB//GFN2-xTB at 50 eV E_{LAB}) of doubly protonated crizotinib. Isotope patterns computed with PlotMS are enhanced for specific signals.

3.2.2 Derivatized Lysine

In the literature, [82] three lysine molecules were connected via peptide bonds and protonated thrice at the respective tertiary amine groups. The computed spectrum of this structure **(6)** and a comparison to the experimental spectrum is provided in figure 11. The proposed fragmentation pathway from the literature is displayed in figure 12, in which roman numerals are used for fragment assignment.

$[M+3H]^{3+}$ at m/z 167.149 is not marked in the experimental spectrum, most likely due to its low signal abundance. An overview of all experimental and computed signals with their respective charged states is provided in table 5.

QCxMS computes four out of nine reported structures correctly and one correct fragment with wrong charge assignment. Signal m/z 142.786 (III) carries a 3+ charge in the experiment, but is

12
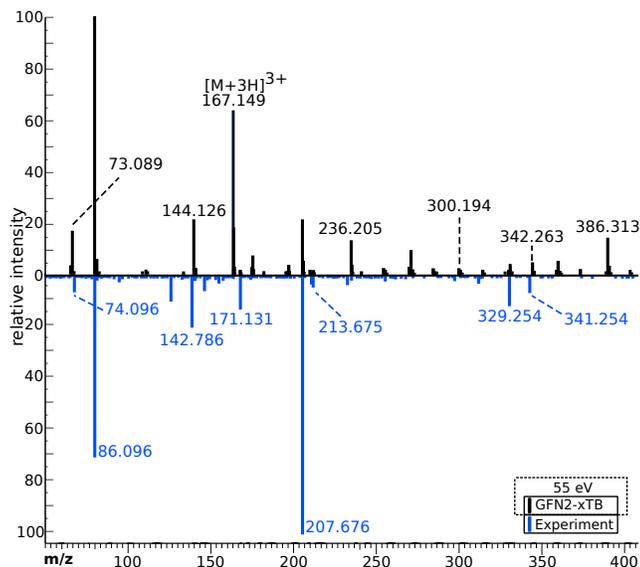
Figure 11 Calculated spectrum (black, top) using GFN2-xTB//GFN2-xTB at 55 eV E_{LAB} of the triply charged lysine derivate compared to experimental spectrum (blue, inverted) taken from the literature.[82]

Figure 12 CID fragmentation analysis proposed in the reference[82]. Roman numerals were used for the product ions of spectrum in figure 11. PI is the precursor ion.

Table 5 Signals found in the experiment and the calculations with their corresponding charge state

Fragment	exp. m/z	exp. charge	comp. m/z	comp. charge
—	—	—	386.313	1+
—	—	—	342.263	1+
IX	341.254	1+	—	—
VIII	329.254	1+	329.254	1+
—	—	—	300.194	1+
—	—	—	236.205	2+
VII	213.675	2+	—	—
VI	207.675	2+	207.675	2+
—	—	—	178.661	2+
V	171.131	2+	171.131	2+
V	171.131	2+	342.262	1+
IV	150.125	2+	—	—
—	—	—	144.126	1+
III	142.786	3+	214.180	2+
II	86.096	1+	86.096	1+
I	74.096	1+	—	—
—	—	—	73.089	1+

assigned a charge of 2+ in the calculations, creating signal m/z 214.180. Here, QCxMS assigns a charge to the neutral fragment at m/z 73.089. The described rearrangement into the ring structure III and consecutive fragmentation into VII (m/z 213.676) is not computed.

Dissociation of the tertiary amine groups is otherwise adequately described. Single amine dissociation forms the doubly charged structure VI (m/z 207.676). Depending on the protonation state of the leaving amine group, consecutive tertiary amine fragmentation creates either 1) signals m/z 329.255 (VIII) and the base peak m/z 86.096 (II) or 2) the doubly charged m/z 171.131 (V) and neutral m/z 73.089. In a competitive fragmentation pathway, the latter dissociation reaction is computed with a single charge on both fragments: m/z 167.149 (3+) → 207.675 (2+) + 73.089 (1+) → 342.263 (1+) + 73.089 (1+). The formally neutral leaving group (m/z 73.089) is again wrongfully charged.

Two factors are of significance when computing this lysine derivate structure. First, the molecule consists of 92 atoms after protonation and has large, flexible side chains. Second, the rearrangement reactions and proton transfers described in figure 12 are significant in the correct portrayal of the fragmentation reactions. To account for these factors, a good description of the underlying PES is needed. However, due to the high computational cost, the MD simulations of this system can only be carried out with SQM methods. Using low-level QC methods limits the accuracy of the computations and the MD simulations are unlikely to account for all consecutive rearrangement reactions in high yield.

Overall, more investigation into the details of the charge assignment has to be conducted. An adapted implementation of Field's rule[100] into QCxMS is planned, which renders a fragment with higher proton affinity more abundant.

4 Conclusion and outlook

The collision induced dissociation (CID) run mode of QCxMS was successfully extended to calculate mass spectra independent of the charge state of the molecular ion. Technically, arbitrary charge states can be investigated while actual computations reveal, that practically — depending on the size and flexibility of the molecules — only a few charges can be treated reliably.

The negative ion mode was tested on a benchmark set of 2-ketoburytic acid **(1)**, 3-ureidopropionic acid **(2)**, ascorbic acid **(3)**, and tryptophan **(4)**. Good agreement with experimental database spectra when using GFN2-xTB//GFN2-xTB for potential energy surface (PES) and electron affinity (EA) calculations renders the method highly satisfactory for computing CID mass spectra in negative ion mode. A mixed variant using GFN2-xTB//PBE0/ma-def2-TZVP did not significantly improve the results. Using full-DFT at PBE/ma-def2-SV(P)//PBE/ma-def2-TZVP levels improved the agreement with the experiment, but increased the

computational costs by three orders of magnitude.

The mass spectrum of doubly positive charged crizotinib was successfully computed, covering all experimentally reported signals. It was shown that all fragments result from the doubly charged molecular ion, instead solely from its singly charged counterpart as reported in the literature. For a triply charged lysine derivate, the fragment charge assignment is more complicated. The flexibility of the structure and the low level of theory used for the computations led to some cases of wrong charge assignment. Nevertheless, five out of nine reported fragments were computed correctly using QCxMS.

Overall, it is demonstrated that QCxMS is a valuable, freely available open-source[15] tool for the unbiased and 'black-box' elucidation of dissociation reactions occurring in various mass spectrometry experiments. It is the first program able to compute spectra of unknown compounds carrying multiple positive and negative charges. In combination with the PlotMS tool, plotting of accurate masses and isotope patterns of multiple charged fragments is routinely possible. Using the build-in GFN2-xTB Hamiltonian, the program is independent of any third-party software. Nevertheless, QC software like ORCA or TURBOMOLE can be used for DFT-based MS calculations as well.

An interesting case for future applications is the calculation of multiply deprotonated structures. Further run modes, like the surface-induced dissociation (SID) method, are currently being realized.

Acknowledgements

Notes and references

[1] F. Allen, A. Pon, R. Greiner and D. Wishart, *Anal. Chem.*, 2016, **88**, 7689–7697.

[2] C. M. Dobson, *Nature*, 2004, **432**, 824–828.

[3] D. P. Demarque, A. E. M. Crotti, R. Vessecchi, J. L. C. Lopes and N. P. Lopes, *Nat. Prod. Rep.*, 2016, **33**, 432–455.

[4] W. Niessen, *Mass Spectrom. Rev.*, 2011, **30**, 626–663.

[5] W. Niessen, *Mass Spectrom. Rev.*, 2012, **31**, 626–665.

[6] C. Cheng and M. L. Gross, *Mass Spectrom. Rev.*, 2000, **19**, 398–420.

[7] E. R. Molina, D. Ortiz, J.-Y. Salpin and R. Spezia, *J. Mass Spectrom.*, 2015, **50**, 1340–1351.

[8] D. Ortiz, J.-Y. Salpin, K. Song and R. Spezia, *Int. J. Mass Spectrom.*, 2014, **358**, 25 – 35.

[9] F. Allen, A. Pon, M. Wilson, R. Greiner and D. Wishart, *Nucleic Acids Res.*, 2014, **42**, W94–W99.

[10] F. Allen, R. Greiner and D. Wishart, *Metabolomics*, 2015, **11**, 98–110.

[11] J. N. Wei, D. Belanger, R. P. Adams and D. Sculley, *ACS Cent. Sci.*, 2019, **5**, 700–708.

[12] Z. C. Lipton, *Queue*, 2018, **16**, 31–57.

[13] L. Drahos and K. Vékey, *J. Mass Spectrom.*, 2001, **36**, 237–263.

[14] Mistrik, *Mass FrontierTM 3.0*, `http://www.highchem.com`.

[15] *To obtain QCxMS, download it at* `https://github.com/qcxms` *or use* `conda-forge`. *The manual and further information can be found at the xtb docs* `https://xtb-docs.readthedocs.io/en/latest/qcxms_doc/qcxms.html`. *Further information can be found at the homepage of the Mulliken Center for Theoretical Chemistry*, `http://www.thch.uni-bonn.de/tc/`.

[16] S. Grimme, *Angew. Chem. Int. Ed.*, 2013, **52**, 6306 – 6312.

[17] V. Ásgeirsson, C. A. Bauer and S. Grimme, *Phys. Chem. Chem. Phys.*, 2016, **18**, 31017–31026.

[18] J. Koopman and S. Grimme, *J. Am. Soc. Mass Spectrom.*, 2021, **32**, 1735–1751.

[19] C. A. Bauer and S. Grimme, *J. Phys. Chem. A*, 2014, **118**, 11479–11484.

[20] C. A. Bauer and S. Grimme, *Org. Biomol. Chem.*, 2014, **12**, 8737–8744.

[21] C. A. Bauer and S. Grimme, *Eur. J. Mass Spectrom.*, 2015, **21**, 125–140.

[22] C. A. Bauer and S. Grimme, *J. Phys. Chem. A*, 2016, **120**, 3755–3766.

[23] J. Lee, T. Kind, D. J. Tantillo, L.-P. Wang and O. Fiehn, *Metabolites*, 2022, **12**, 68.

[24] S. Wang, T. Kind, P. L. Bremer, D. J. Tantillo and O. Fiehn, *Analytical Chemistry*, 2022, **94**, 1559–1566.

[25] *MassBank of North America website, MassBank of North America, accessed Apr.05, 2022*, `https://mona.fiehnlab.ucdavis.edu`.

[26] V. Ásgeirsson, C. A. Bauer and S. Grimme, *Chem. Sci.*, 2017, **8**, 4879–4895.

[27] J. Koopman and S. Grimme, *ACS Omega*, 2019, **4**, 15120–15133.

[28] R. Schnegotzki, J. Koopman, S. Grimme and R. D. Süssmuth, *Chemistry – A European Journal*, **n/a**, e202200318.

[29] M. Yamashita and J. B. Fenn, *J. Phys. Chem.*, 1984, **88**, 4451–4459.

[30] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida and T. Matsuo, *Rapid Commun. Mass Spectrom.*, 1988, **2**, 151–153.

[31] M. Barber, R. Bordoli, R. Sedgwick and A. N. Tyler, *Nature*, 1981, **293**, 270–275.

[32] E. C. Horning, M. G. Horning, D. I. Carroll, I. Dzidic and R. N. Stillwell, *Anal. Chem.*, 1973, **45**, 936–943.

[33] S. A. McLuckey and J. L. Stephenson, *Mass Spectrometry Reviews*, 1998, **17**, 369–407.

[34] K. Vékey, *Mass Spectrometry Reviews*, 1995, **14**, 195–225.

[35] P. Longevialle, and R. Botter, *J. Chem. Soc., Chem. Commun.*, 1980, 823–825.

[36] A. G. Harrison and T. Yalcin, *Int. J. Mass Spectrom. Ion Processes*, 1997, **165-166**, 339–347.

[37] G. E. Reid, R. J. Simpson and R. A. J. O'Hair, *J. Am. Soc. Mass Spectrom.*, 1998, **9**, 945–956.

[38] R. S. Johnson, D. Krylov and K. A. Walsh, *J. Mass Spectrom.*, 1995, **30**, 386–387.

[39] S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009.

[40] C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.

[41] N. B. Cech and C. G. Enke, *Mass Spectrometry Reviews*, 2001, **20**, 362–387.

[42] P. Liigand, K. Kaupmees, K. Haav, J. Liigand, I. Leito, M. Girod, R. Antoine and A. Kruve, *Analytical Chemistry*, 2017, **89**, 5665–5668.

[43] A. T. Iavarone, J. C. Jurchen and E. R. Williams, *Analytical chemistry*, 2001, **73**, 1455–1460.

[44] B. L. Frey, D. T. Ladror, S. B. Sondalle, C. J. Krusemark, A. L. Jue, J. J. Coon and L. M. Smith, *Journal of the American Society for Mass Spectrometry*, 2013, **24**, 1710–1721.

[45] A. T. Iavarone and E. R. Williams, *Journal of the American Chemical Society*, 2003, **125**, 2319–2327.

[46] S. Nguyen and J. B. Fenn, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 1111–1117.

[47] S. Banerjee and S. Mazumdar, *Int. J. Anal. Chem.*, 2012, **2012**, 1–40.

[48] F. Hillenkamp, M. Karas, R. C. Beavis and B. T. Chait, *Anal. Chem.*, 1991, **63**, 1193A–1203A.

[49] P. M. Lalli, B. A. Iglesias, H. E. Toma, G. F. de Sa, R. J. Daroda, J. C. Silva Filho, J. E. Szulejko, K. Araki and M. N. Eberlin, *J. Mass Spectrom.*, 2012, **47**, 712–719.

[50] G. J. Van Berkel, F. Zhou and J. T. Aronson, *International Journal of Mass Spectrometry and Ion Processes*, 1997, **162**, 55–67.

[51] J. A. Loo, H. R. Udseth, R. D. Smith and J. H. Futrell, *Rapid Communications in Mass Spectrometry*, 1988, **2**, 207–210.

[52] S. Zhou, A. G. Edwards, K. D. Cook and G. J. Van Berkel, *Analytical Chemistry*, 1999, **71**, 769–776.

[53] S. Zhou, B. S. Prebyl and K. D. Cook, *Analytical chemistry*, 2002, **74**, 4885–4888.

[54] M. Girod, X. Dagany, R. Antoine and P. Dugourd, *International Journal of Mass Spectrometry*, 2011, **308**, 41–48.

[55] S. Zhou and K. D. Cook, *Journal of the American Society for Mass Spectrometry*, 2000, **11**, 961–966.

[56] M. Almasian, J. Grzetic, J. van Maurik, J. D. Steill, G. Berden, S. Ingemann, W. J. Buma and J. Oomens, *J. Phys. Chem. Lett.*, 2012, **3**, 2259–2263.

[57] J. D. Steill and J. Oomens, *J. Am. Chem. Soc.*, 2009, **131**, 13570–13571.

[58] B. M. Ehrmann, T. Henriksen and N. B. Cech, *Journal of the American Society for Mass Spectrometry*, 2011, **19**, 719–728.

[59] M. Koné, B. Illien, C. Laurence, J.-F. Gal and P.-C. Maria, *Journal of physical organic chemistry*, 2006, **19**, 104–114.

[60] V. Vais, A. Etinger and A. Mandelbaum, *Journal of mass spectrometry*, 1999, **34**, 755–760.

[61] V. Vais, A. Etinger and A. Mandelbaum, *European Mass Spectrometry*, 1999, **5**, 449–454.

[62] Y.-P. Tu, *The Journal of Organic Chemistry*, 2006, **71**, 5482–5488.

[63] P. Pracht, C. A. Bauer and S. Grimme, *J. Comput. Chem.*, 2017, **38**, 2618–2631.

[64] P. Pracht, F. Bohle and S. Grimme, *Phys. Chem. Chem. Phys.*, 2020, **22**, 7169–7192.

[65] S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 2847–2862.

[66] *Find the code on the CREST GitHub Repository: https://github.com/grimme-lab/crest and the documentation at https://xtb-docs.readthedocs.io/en/latest/crest.html*.

[67] S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher and M. Stahn, *J. Phys. Chem.*, 2021, **125**, 4039–4054.

[68] K. Levsen and H. Schwarz, *Mass Spectrom. Rev.*, 1983, **2**, 77–148.

[69] J. Mitchell Wells and S. A. McLuckey, *Biological Mass Spectrometry*, Academic Press, 2005, vol. 402, pp. 148–185.

[70] R. O. Jones and O. Gunnarsson, *Rev. Mod. Phys.*, 1989, **61**, 689–746.

[71] S. E. Stein and D. R. Scott, *J. Am. Soc. Mass Spectrom.*, 1994, **5**, 859–866.

[72] *Human Metabolome DataBase website, The metabolomics innovation center, accessed Apr.05, 2022*, `https://hmdb.ca`.

[73] 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 and L. Querengesser, *Nucleic Acids Res.*, 2007, **35**, D521–D526.

[74] D. S. Wishart, C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J. A. Cruz, E. Lim, C. A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H. J. Vogel and I. Forsythe, *Nucleic Acids Res.*, 2008, **37**, D603–D610.

[75] D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert, *Nucleic Acids Res.*, 2012, **41**,

D801–D807.

[76] 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 and A. Scalbert, *Nucleic Acids Res.*, 2017, **46**, D608–D617.

[77] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T. Nishioka, *J. Mass Spectrom.*, 2010, **45**, 703–714.

[78] *MassBank consortium and its contributors; MassBank/MassBank-data: Release version 2020.06*, https://doi.org/10.5281/zenodo.3903207.

[79] J. Zhao, C.-K. Siu, T. Shi, A. C. Hopkinson and K. W. M. Siu, *The Journal of Physical Chemistry B*, 2009, **113**, 4963–4969.

[80] K. Vekey, A. G. Brenton and J. H. Beynon, *The Journal of Physical Chemistry*, 1986, **90**, 3569–3577.

[81] J. R. Joyce and D. S. Richards, *Journal of the American Society for Mass Spectrometry*, 2011, **22**, 360–368.

[82] T. Huang, J. M. Rabus, B. J. Bythell and J. L. Edwards, *Journal of the American Society for Mass Spectrometry*, 2019, **30**, 1158–1162.

[83] S. Grimme, A. Hansen, S. Ehlert and J.-M. Mewes, *The Journal of Chemical Physics*, 2021, **154**, 064103.

[84] A. Klamt, *The Journal of Physical Chemistry*, 1995, **99**, 2224–2235.

[85] A. Klamt, V. Jonas, T. Bürger and J. C. W. Lohrenz, *The Journal of Physical Chemistry A*, 1998, **102**, 5074–5085.

[86] F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012,

2, 73–78.

[87] F. Neese, *WIREs Computational Molecular Science*, **n/a**, e1606.

[88] *TURBOMOLE V7.5.1 2020, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2020, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.*

[89] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2021, **11**, e1493.

[90] E. Caldeweyher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2017, **147**, 034112.

[91] J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982–9985.

[92] C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.

[93] A. Schäfer, H. Horn and R. Ahlrichs, *The Journal of Chemical Physics*, 1992, **97**, 2571–2577.

[94] F. Weigend, M. Häser, H. Patzelt and R. Ahlrichs, *Chemical Physics Letters*, 1998, **294**, 143–152.

[95] J. Zheng, X. Xu and D. G. Truhlar, *Theor. Chem. Acc.*, 2011, **128**, 295–305.

[96] C. Collette and E. De Pauw, *Rapid Commun. Mass Spectrom.*, 1998, **12**, 165–170.

[97] V. Gabelica and E. De Pauw, *Mass Spectrom. Rev.*, 2005, **24**, 566–587.

[98] K. Volná, M. Holčapek, L. Kolářová, K. Lemr, J. Čáslavský, P. Kačer, J. Poustka and M. Hubálek, *Rapid Commun. Mass Spectrom.*, 2008, **22**, 101–108.

[99] H. H. Kakazu Y, *L-Tryptophan; LC-ESI-QQ; MS2; CE:20 V; [M-H]-*, 19.01.2016, https://massbank.eu/MassBank/RecordDisplay?id=KO001868.

[100] M. S. B. Munson and F. H. Field, *Journal of the American Chemical Society*, 1966, **88**, 2621–2630.

[101] D. Stevenson, *Discussions of the Faraday Society*, 1951, **10**, 35–45.