

Jeroen Gertjan Koopman

Computation of Mass Spectra Using Quantum Chemical Methods



Dissertation

Computation of Mass Spectra Using Quantum Chemical Methods

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Jeroen Gertjan Koopman
aus
Siegburg

Bonn, 20.06.2022

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen
Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Stefan Grimme
2. Gutachter: Prof. Dr. Thomas Bredow
Tag der Promotion: 06.09.2022
Erscheinungsjahr: 2022

Affirmation in Lieu of an Oath

I, Jeroen Gertjan Koopman, hereby declare that I am the sole author of this dissertation.

The doctoral thesis I submitted is my own work and was prepared without unauthorized outside assistance. I have not included text passages, graphics or other materials from third parties or my own examination papers without identifying them. Only the sources and resources that I have indicated were used. All verbatim and non-verbatim citations from other works are identified in accordance with the citation rules for academic writing. The thesis that I submitted has not yet been published, or has been published in full or in part other than the locations indicated. The thesis that I submitted has not yet been submitted in any form as part of an examination/qualification course. I prepared the doctoral thesis that I submitted in accordance with the principles of good research practice. I am aware of the significance and criminal consequences of a false affirmation in lieu of an oath.

My statements are true to the best of my knowledge and belief.

Bonn, 20. June 2022

Place and Date

Jeroen Koopman

Signature

“There is no single ‘golden rule’ in approaching the wide field of mass spectrometry.”

– Jürgen H. Gross –

Abstract

Mass spectrometry (MS) is widely used for the structural elucidation of compounds in a vast variety of scientific research. Unfortunately, many methods that aim to predict mass spectra often lack the capability to elucidate the complete fragmentation pathways of unknown compounds. For this reason, the QCEIMS program was developed in 2013, which combines quantum chemical (QC) methods with statistical models to calculate standard 70 eV Electron Ionization (EI) mass spectra. In this thesis, the QCEIMS software was improved to compute more reliable EI spectra and extended to enable calculations of collision-induced dissociation (CID) processes. This development led to the renaming of the program to QCxMS ($x = \text{EI, CID}$), reflecting the greater general applicability of the software. For a perspective overview of the subject, the [introduction](#) to this thesis provides an outline of the operating principles of mass spectrometry and highlights the differences between the EI and CID methods. Existing computational and chemoinformatic methods for obtaining MS references are listed, and their shortcomings are discussed, resulting in the motivation for the development of the QCxMS approach.

The [theoretical background](#) for the underlying concepts of calculating mass spectra using quantum chemical methods are broken down along with the working mechanisms of the QCxMS software. Since the software is based on calculating a statistically relevant number of dissociation processes through massively parallel molecular dynamics (MD) simulations, the use of fast and reliable QC methods is mandatory. In particular, semiempirical quantum mechanical (SQM) methods are well suited for this purpose, making the GFN n -xTB ($n = 1, 2$) methods essential in this context. They allow fast and accurate calculations of dissociation processes for molecules composed of atoms up to Radon ($Z = 86$). The implementation of GFN2-xTB in QCxMS and the [calculation of EI mass spectra with both GFN \$n\$ -xTB methods](#) are evaluated for various compounds consisting of organic, inorganic main-group, and transition metal elements.

In many MS applications, ionization is accomplished by protonation. However, unlike EI, the low ionization energy achieved in this process is insufficient to initiate fragmentation simultaneously. Instead, collisions between the ionized analyte and neutral gas atoms are used to activate the species. The implementation of this principle in QCxMS leads to [a tool for routine calculations of CID mass spectra](#). The most important aspects in this regard are implemented in different run types, allowing the software to calculate slow heating and collision processes. Tests on various benchmark structures are shown.

The relevance of QCxMS for the chemical/pharmaceutical industry is demonstrated by the application of the CID mode to calculate the [spectra of drug molecules](#). For the first time, the unknown fragmentation pathways of the two large molecules nateglinide and zopiclone are fully elucidated by combining generic rules and calculated MD trajectories. Unusual decompositions and unintuitive protonation sites are found and discussed.

Developments for the CID module are concluded by [removing the charge constraints](#) on the molecular

ion when calculating mass spectra. The considerations that went into the development of this approach are discussed, and the weak points are presented. Negatively charged spectra of deprotonated organic compounds are calculated, and the influence of the deprotonation sites on the computed spectrum analyzed. In addition, mass spectra of multiply charged compounds are calculated fully automatically, which was not possible by other methods so far.

Finally, the most important details of this work are [summarized, and an outlook](#) on future developments that could improve the general applicability and the accuracy is given. Although not explicitly reported, the program was released open-source and comprehensive documentation is provided for free on the Internet.

Kurzzusammenfassung

Die Massenspektrometrie (MS) wird für die Strukturaufklärung von Verbindungen in einer Vielzahl von wissenschaftlichen Forschungsgebieten verwendet. Leider sind viele Methoden zur Vorhersage von Massenspektren oft nicht in der Lage, die vollständigen Fragmentierungswege unbekannter Verbindungen aufzuklären. Aus diesem Grund wurde 2013 das Programm QCEIMS entwickelt, das quanten-chemische Methoden (QC) mit statistischen Modellen kombiniert, um Standardisierte 70 eV Elektronenionisations (EI) Massenspektren zu berechnen. In dieser Arbeit wurde die QCEIMS-Software verbessert, um zuverlässigere EI-Spektren zu berechnen, und erweitert, um die Berechnungen von kollisionsinduzierten Dissoziationsprozessen (CID) zu ermöglichen. Diese Entwicklung führte zur Umbenennung des Programms in QCxMS ($x = \text{EI, CID}$), was der größeren allgemeinen Anwendbarkeit der Software Rechnung trägt.

Für einen perspektivischen Überblick über das Thema bietet die [Einleitung](#) zu dieser Arbeit eine Darstellung der Funktionsprinzipien der Massenspektrometrie und zeigt die Unterschiede zwischen der EI- und der CID-Methode auf. Bestehende computergestützte und chemoinformatische Methoden zur Gewinnung von MS-Referenzen werden aufgelistet und ihre Unzulänglichkeiten diskutiert, woraus sich die Motivation für die Entwicklung des QCxMS-Ansatzes ergibt.

Um den theoretischen Hintergrund für die zugrundeliegenden Konzepte zu verstehen die für die Berechnung von Massenspektren mit quantenchemischen Methoden erforderlich sind, werden im folgenden die Arbeitsmechanismen der QCxMS-Software aufgeschlüsselt. Da die Software auf der Berechnung einer statistisch relevanten Anzahl von Dissoziationsprozessen durch massiv parallele molekular-dynamische (MD) Simulationen basiert, ist die Verwendung schneller und zuverlässiger QC-Methoden zwingend erforderlich. Vor allem semiempirische quantenmechanische (SQM) Methoden sind für diesen Zweck gut geeignet, insbesondere sind die GFN n -xTB ($n = 1, 2$) Methoden in diesem Zusammenhang von großer Bedeutung. Sie ermöglichen schnelle und genaue Berechnungen von Dissoziationsprozessen für Moleküle, die aus Atomen bis einschließlich Radon ($Z = 86$) zusammengesetzt sein können. Die Implementierung von GFN2-xTB in QCxMS und die [Berechnung von EI-Massenspektren mit beiden GFN \$n\$ -xTB-Methoden](#) werden für verschiedene Verbindungen evaluiert, die aus organischen, Hauptgruppen anorganischen, sowie Übergangs-Metallatomen bestehen.

Bei vielen MS-Anwendungen erfolgt die Ionisierung durch Protonierung. Anders als bei EI reicht die bei diesem Verfahren erzielte niedrige Ionisierungsenergie jedoch nicht aus, um gleichzeitig eine Fragmentierung zu initiieren. Stattdessen werden Kollisionen zwischen dem ionisierten Analyten und neutralen Gasatomen zur Aktivierung der Spezies genutzt. Die Umsetzung dieses Prinzips in QCxMS führt zu [einem Werkzeug zur routinemäßigen Berechnung von CID-Massenspektren](#). Die wichtigsten Aspekte in diesem Bezug sind in verschiedenen "Run-Types" implementiert, die die Berechnung von langsamen Aufheizvorgängen und Kollisionsprozesse mit der Software ermöglichen. Tests mit verschiedenen Strukturen werden gezeigt.

Die Relevanz von QCxMS im chemisch-pharmazeutischen Bereich wird durch die Anwendung des

CID-Modus zur Berechnung von [Arzneimittelmolekülspektren](#) demonstriert. Zum ersten Mal werden die unbekanntenen Fragmentierungswege der beiden großen Moleküle Nateglinid und Zopiclon durch eine Kombination aus generischen Regeln und berechneten MD-Trajektorien vollständig aufgeklärt. Dabei werden ungewöhnliche Zersetzungen und unintuitive Protonierungsstellen gefunden und diskutiert.

Die Entwicklungen für das CID-Modul werden durch die [Aufhebung der Ladungsvorgaben](#) des Molekülions bei der Berechnung von Massenspektren abgerundet. Negativ geladene Spektren von deprotonierten organischen Verbindungen werden berechnet und der Einfluss der Deprotonierungsstellen auf die berechneten Spektren diskutiert. Darüber hinaus werden Massenspektren von mehrfach geladenen Verbindungen vollautomatisch berechnet, was bisher mit keiner anderen Methode möglich ist. Die Aspekte, die bei der Entwicklung der Methode berücksichtigt wurden, werden gezeigt und die Schwachstellen dargestellt.

Schließlich werden die wichtigsten Details dieser Arbeit [zusammengefasst und ein Ausblick](#) auf zukünftige Entwicklungen, die die allgemeine Anwendbarkeit sowie die Genauigkeit verbessern könnten, gegeben. Obwohl in dieser Arbeit nicht speziell darüber berichtet wird, wird das Programm Open-Source angeboten und eine umfassende Dokumentation wird kostenlos im Internet zur Verfügung gestellt.

Contents

1	Introduction	1
2	Theoretical Background	7
2.1	Electronic Structure Methods	7
2.1.1	The Electronic Hamiltonian	7
2.1.2	Hartree–Fock Theory	8
2.1.3	Kohn–Sham Density Functional Theory	9
2.1.4	Basis Sets	11
2.1.5	Extended Tight-Binding Methods	11
2.2	Computation of Mass Spectra	13
2.2.1	Potential Energy Surface and Molecular Dynamics	13
2.2.2	General Setup of QCxMS	14
2.2.3	Computing Electron Ionization Mass Spectra	16
2.2.4	Computing Collision Induced Dissociation Mass Spectra	17
2.2.5	Protonation Tool and Thermochemistry	18
3	Overview of “Calculation of Electron Ionization Mass Spectra with Semiempirical GFNn-xTB methods”	21
4	Overview of “From QCEIMS to QCxMS: A Tool to Routinely Calculate CID Mass Spectra Using Molecular Dynamics”	23
5	Overview of “Quantum Chemistry-based Molecular Dynamics Simulations as a Tool for the Assignment of ESI-MS/MS Spectra of Drug Molecules”	25
6	Overview of “Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules”	27
7	Conclusion and Outlook	29
A	Calculation of Electron Ionization Mass Spectra with Semiempirical GFNn-xTB Methods	35
A.1	Introduction	36
A.2	Methodology	36
A.2.1	QCEIMS	37
A.2.2	Extended Tight-Binding Methods	38
A.2.3	Technical Details	39

A.3	Results and Discussion	39
A.3.1	Benchmark Set	39
A.3.2	Timings	41
A.3.3	Stability	42
A.3.4	Comparison of Experimental and Theoretical Spectra	42
A.4	Conclusions	53
A.5	Supporting Information	55
A.5.1	Details of computational demands and method stability	55
A.5.2	Calculated spectra using DFT for IP calculations	57
B	From QCEIMS to QCxMS: A Tool to Routinely Calculate CID Mass Spectra Using Molecular Dynamics	59
B.1	Introduction	60
B.2	Theoretical Background	61
B.2.1	Ionization and Internal Energy	61
B.2.2	Protonation Sites	62
B.2.3	Collision Events	62
B.2.4	Analysis	63
B.3	Methodology	63
B.3.1	Production Runs and General Collision Simulation	64
B.3.2	Multiple Collisions	66
B.3.3	Technical Details	67
B.4	Results and Discussion	68
B.4.1	Collision Energetics	69
B.4.2	Single Collisions	70
B.4.3	Protonation sites and proton mobility	72
B.4.4	Multiple collisions	73
B.5	Conclusion and Outlook	79
B.6	Supporting Information	80
B.6.1	Instrumentation details of the experiments	80
B.6.2	Relative energy ranking of protonated structures	81
B.6.3	Calculated alternative protomer spectra	82
B.6.4	Effects of the <i>general activation</i> run-type settings	86
B.6.5	Additional calculated spectra	89
C	Quantum Chemistry-based Molecular Dynamics Simulations as a Tool for the Assignment of ESI-MS/MS Spectra of Drug Molecules	91
C.1	Introduction	92
C.2	Methodology	93
C.2.1	Benchmark Molecules	93
C.2.2	Experimental Details	94
C.2.3	Ranking of Difficulty by Common Fragmentation Pathways	94
C.2.4	Computational Details	94
C.2.5	Discrepancies Between Calculations and Experiments	95

C.3	Results and Discussion	96
C.3.1	Nateglinide	96
C.3.2	Zopiclone	100
C.4	Conclusion and Outlook	104
C.5	Supporting Information	106
C.5.1	Calculated Spectra Using QCxMS	106
C.5.2	Mass accuracy of the measured fragments	108
D	Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules	111
D.1	Introduction	112
D.2	Theoretical Background	113
D.2.1	Ionization	113
D.2.2	Method	114
D.2.3	Charge Assignment	115
D.2.4	Plotting Spectra	115
D.3	Technical Details	116
D.3.1	Benchmark Molecules	116
D.3.2	Computational and Technical Details	117
D.3.3	Differences Between Experiment and Theory	117
D.4	Results and Discussion	118
D.4.1	Negative Charges	118
D.4.2	Multiple Charges	124
D.5	Conclusion and Outlook	130
D.6	Supporting Information	132
D.6.1	Protomer rankings	132
D.6.2	Additional calculated Spectra	134
	Bibliography	141
	List of Figures	161
	List of Tables	167
	Acknowledgements	169

Introduction

The high-precision analytical results achieved by mass spectrometry (MS) have become indispensable in many areas of natural sciences. Besides its operation for structure elucidation in various (in)organic chemical areas, its routine application in food,^{1,2} environmental,^{3,4} and petroleum⁵ chemistry is of great importance. Modern material science,⁶ military,⁷ medicinal,⁸ and the “-omics” (genomics, metabolomics, lipidomics, etc.)^{9,10} areas all profit from the accuracy achieved by using MS with only small quantities of an examined analyte.

The success of the method is based on the simplicity of the concept: the composition of an analyte can be determined by the mass of its molecular or atomic components.¹¹ Information about the mass provides a means for labeling the elements and thus the structural formula of the analyte. By fragmentation of the compound, greater insight into the structural properties can be gained. The dissociation patterns produce an improved understanding of the exact arrangements and connections between individual atoms or functional groups. Decomposition is induced by increasing the internal energy of a molecule until dissociation of its bonds occurs, causing it to break down.

Charge plays an essential role in measuring the (un)fragmented compounds. The working mechanics of an MS instrument are based on manipulating the trajectory of a charged species by magnetic and electric fields, allowing it to be accelerated, deflected, and directed towards a mass detector. Arrived, the masses of the compounds are measured according to their mass-to-charge ratio (m/z) and expressed in relation to their quantity in a mass spectrum.¹¹

Over the last decades, various methods to induce charge by ionizing the molecule were developed, each with its advantages and disadvantages.¹² A general distinction is made between “hard” and “soft” ionization methods, which refers to the internal energy acquired by the molecule during ionization.¹³ Hard ionization can be obtained by using an electron beam (electron ionization (EI)) that removes an electron from a molecule, leading to a radical cation,¹⁴ also called the “molecular ion” $[M]^{\bullet+}$. The amount of energy introduced into the system by the ionization process is usually large. Hence, simultaneous activation of the analyte occurs, which results in high fragmentation rates of the produced molecular ions. Because ionization is performed in the gas phase, instruments are often coupled to gas chromatographs (GC-EI-MS).¹⁵ The method is best suited for volatile, preferably non-polar analytes of molecular masses up to 800 u.¹¹ A schematic depiction of the EI ionization and activation step is displayed in figure 1.1 at the top.

The downside of using GC-EI-MS is that the molecular ion often dissipates completely, whereby any information about its mass is lost.¹¹ Alternatively, the development of “soft-ionization” methods has

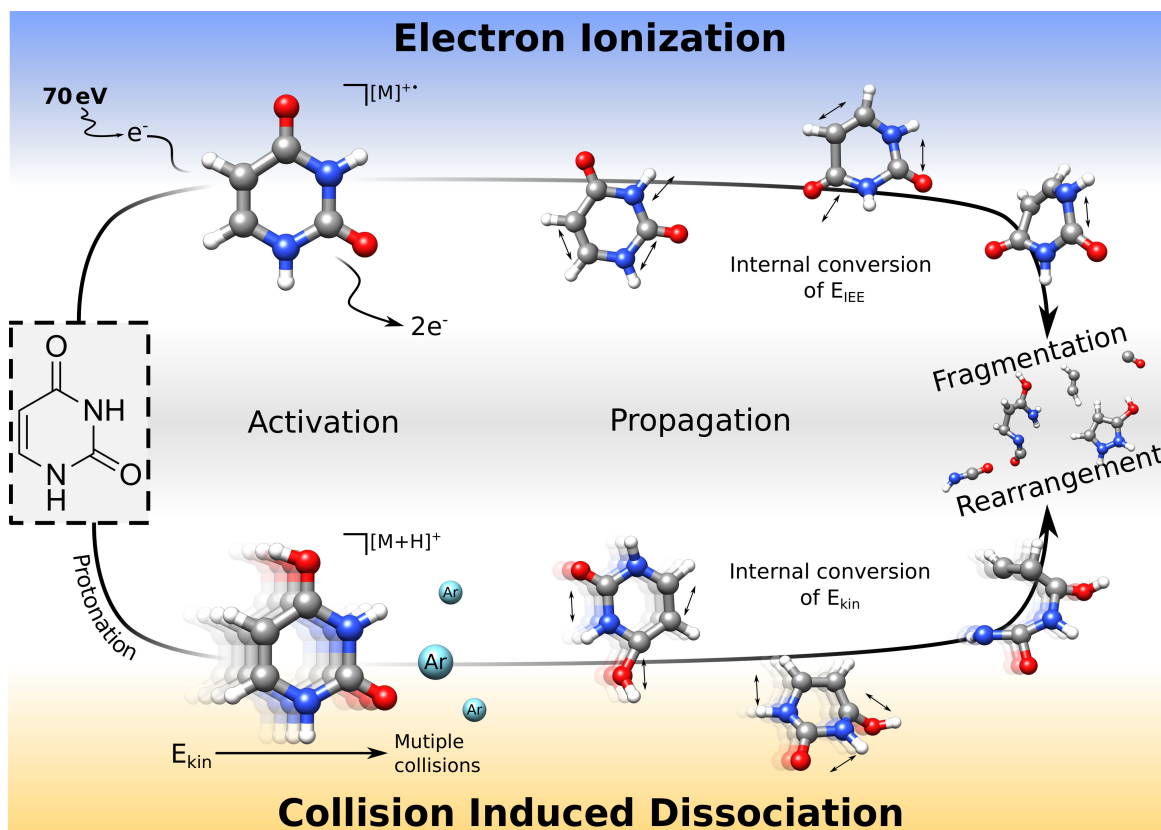


Figure 1.1: Schematic comparison between the concepts of Electron Ionization (top) and Collision Induced Dissociation (bottom). While activation, energy is introduced into the system. For EI: Impact Excess Energy (E_{IEE}) after electron impact; for CID: Kinetic Energy (E_{kin}). Internal conversion leads to rearrangement and fragmentation.

led to experiments with much lower fragmentation rates. Techniques like Field Ionization/Desorption (FI/FD)¹⁶ and Chemical Ionization (CI)¹⁷ at Atmospheric Pressure (API/APCI)¹⁸ introduce ionization by (de-)protonation of the analyte. A standard API method is electrospray ionization (ESI),¹⁹ which enables the transfer of ions from solution into the gas phase. It is commonly coupled to liquid chromatographs LC-ESI-MS,²⁰ so non-volatile and predominantly polar compounds can be analyzed,^{21,22} which makes it complementary to the EI method. ESI can be used to produce positive and negative ions and routinely introduces multiple charges into the analyte.²³ This has made the method extremely useful for the structure elucidation of large molecules,²⁴ which is especially relevant for biology and biomedical sciences, making ESI the most frequently employed ionization technique nowadays.^{25–27}

However, ionization using soft methods often results in low to non-existing fragmentation rates of the molecular ion $[M+H]^+$. To enforce dissociation of $[M+H]^+$, collisional activation (collision-induced dissociation, CID) is routinely applied. The molecular ion is accelerated in an electric field and fed into a collision cell, where it encounters neutral collision gas atoms (usually He, N₂, Ar). Figure 1.1 (bottom) shows a schematic depiction of the CID concept. Depending on the pressure and temperature of the collision gas, as well as the sizes of the collision partners, $[M+H]^+$ undergoes a multitude of

collisions in which it is stepwise activated until fragmentation occurs. The use of ionization and activation instruments in succession is usually accompanied by a second mass spectrometric analysis, known as *tandem MS* or *MS/MS*.^{28,29} The possibilities for instrumental setups are manifold, and various examples are found in the literature.^{11,30}

With the measurement of the exact masses, determining the chemical formula of the molecular ion and all produced fragments is possible. However, resolving the structural properties only by information about the mass is complicated. Many different approaches were developed to assign a measured spectrum to its chemical structure. Solving this problem is considered the “holy grail” of mass spectrometry. Over the last decades, it was found that structures with comparable chemical groups dissociate similarly, so empirical rules^{31–34} were developed that are applied for the retrospective reconstruction of the molecular ion from the measured fragments. Chemoinformatic approaches emerged, in which tabulated fragmentation rules and master equations were used to determine typical decomposition pathways automatically.^{35,36} But with growing molecular size, the application of such approaches becomes very complex.^{37,38} Over time, the number of generally applicable rules has shrunk in proportion to the number of rule exceptions.³⁹ Considering only known decomposition pathways, the results obtained by these methods are biased towards the tabulated reactions and rearrangements. Uncharacteristic and unknown bond breakage, as well as unreported rearrangement reactions, cannot be considered by these empirical approaches.

Alternatively, a measured spectrum can be compared to reference spectra of known chemicals. If the fragmentation patterns are the same, the structure was already measured and can be assigned. A huge number of references are stored in broadly available spectral databases,^{40–44} and fast screening of these references is routinely conducted using a wide variety of machine learning (ML) algorithms.^{45–48} These algorithms are trained on the existing data and quickly provide a selection of possible candidate structures for an input spectrum with a matching score of agreement between measurement and reference. However, as the chemical space is too large to acquire mass spectra of all possible chemical compounds, databases will remain incomplete,^{45,49} and algorithms can solely be trained on partial data, introducing a bias towards the training set. The addition of references using experimental measurements is often time-consuming, costly, and can even be dangerous, depending on the chemical class under investigation. As ML approaches do not compute the fragmentation pathways, their results lack interpretability, and reactions essential for correctly describing an unknown spectrum cannot be accounted for, e.g., not previously measured rearrangement reactions.

All in all, the available methods strongly rely on the completeness of the underlying information, but references are lacking. Solving this issue by creating a method that can compute mass spectra as references without the need to rely on experimental data is the goal of this thesis. A workflow benefiting from computational reference generation is schematically depicted in figure 1.2.

To this end, *ab-initio* quantum chemical (QC) methods produce unbiased and inherently “black-box” results. Based on this, Grimme developed the quantum chemistry electron ionization mass spectrometry (QCEIMS) program⁵⁰ in 2013, which routinely computes standard 70 eV EI mass spectra of (previously unknown) compounds. It utilizes Born-Oppenheimer molecular dynamics (BO-MD), which enables the detailed elucidation of a molecular structure dissociation and rearrangement reaction “on-the-fly”. Important reactions like the famous McLafferty rearrangement are computed, which is an important feature of the method. The application of the QCEIMS routine was the topic of various publications,^{51–54} showing that the results produced are reasonable when compared to reference spectra. A more detailed picture of how the software computes EI-MS is provided in Section 2.2.3 and the original publication [50].

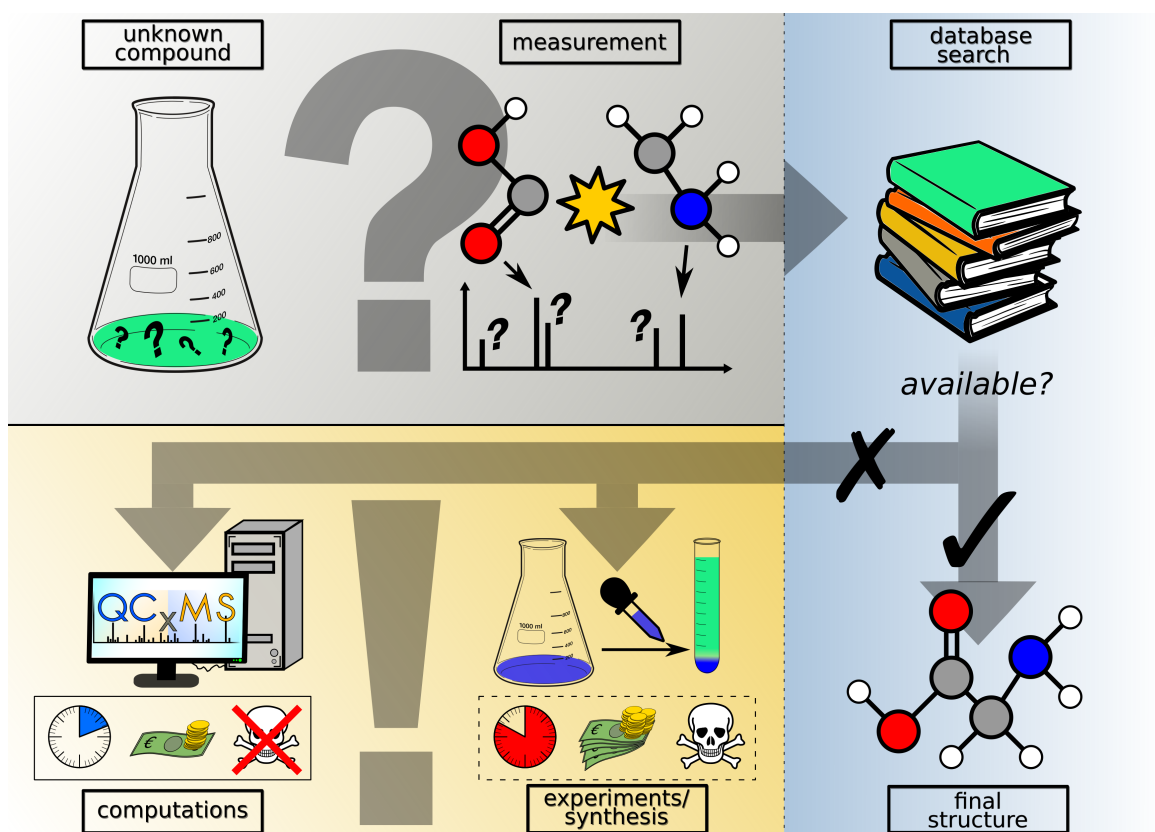


Figure 1.2: The motivation of designing the QCxMS workflow. The measurement of an unknown compound can be compared to database entries. If these are not available, either cost expensive experiments or the *ab-initio* QCxMS program can be used as reference.

Unfortunately, the computation of an entire mass spectrum using QC methods requires hundred to thousands of BO-MD trajectories over several picoseconds of simulation time and, therefore, considerable computational resources. (Post-)Hartree-Fock (HF) and density functional theory (DFT) methods may have sufficient accuracy but are computationally too demanding. To reduce the computational timings, simplifications to HF and DFT were developed into semiempirical quantum mechanical (SQM) methods that are predestined to provide the necessary computational performance for long-term molecular dynamics.⁵⁵ Most promising in this respect was the development of the extended tight-binding method GFN-xTB, which is parametrized for geometries, frequencies, and non-covalent interactions. It was implemented into QCEIMS and tested on organic, main-group inorganic, and transition-metal compounds.⁵⁶

The overall good performance of GFN-xTB to compute mass spectra raised the question if its more sophisticated successor, GFN2-xTB, could even improve upon the results produced. The implementation of GFN2-xTB is the first topic discussed in this thesis. Performance of the two GFN n -xTB ($n = 1, 2$) methods is compared in Chapter 3 on a benchmark of various molecules containing organic, main-group inorganic, and transition-metal atoms. Computing ionization potentials (IPs) with the two methods is assessed in detail, and differences in timings are analyzed. Section 2.1.5 provides a short description of the working mechanisms of the GFN n -xTB methods for better

understanding.

As mentioned before, the usability of GC-EI-MS is limited to certain molecular groups of defined sizes and polarity. Fueled by the convincing results of QCEIMS, developing a method to routinely compute the frequently used LC-ESI-MS/MS technique (abbreviated as “CID” in the following) is the focus point of Chapter 4. Various run-types are designed that account for thermal (slow-heating) and collisional activation of the molecular ion. Single and multiple collision processes are inspected, and the associated collision dynamics are highlighted. Singly protonated positive ions are selected, and their mass spectra computed. The generated spectra are discussed in the context of the initial protonation sites and mobile protons. Section 2.2.4 gives an overview and a general remark on the activation processes in CID. An insight into the concept of the fully automatic protonation procedure and the free-energy ranking of the produced structures is provided in Section 2.2.5. The greater general applicability of the software is illustrated by its renaming from QCEIMS to QCxMS, in which the “x” is the placeholder for the distinct MS modes (EI, CID, and future developments).

Utilizing the newly developed CID run-mode, the previously unknown fragmentation pathways of the two large drug molecules zopiclone and nateglinide are computed in Chapter 5. Since the unidentified decomposition routes could not be elucidated using classical fragmentation rules alone, QCxMS is utilized to compute the dissociation and rearrangement reactions required to describe the signals produced in the corresponding experiment. The complexity of reconstructing the signals by the generic rules is ranked, thereby finding dissociation and rearrangement reactions that are considered untypical regarding the rules.

All previous chapters assumed singly charged structures. As ionization of the analyte depends on its structural properties, some soft ionization techniques produce singly or multiply, positively or negatively charged molecular ions. In Chapter 6, the extension of QCxMS to compute CID MS without charge restriction of the molecular ion is presented. Especially for negative ion fragmentation pathway interpretation, data to create empirical routes are scarce.³³ For this reason, QCxMS is tested on a benchmark of four negatively charged organic molecules, and the computed fragmentation patterns are discussed at length. The influence of the deprotonation site is investigated, and interesting connections between functional groups and mobile protons are discovered. Additionally, the mass spectra of two multiple positively charged structures are computed for the first time from scratch. Therefore, a new algorithm is developed that allows distributing multiple charges on single fragments by ionization potential calculations. This expansion enables QCxMS to compute CID spectra of large organic complexes and proteins.

Theoretical Background

This chapter presents an overview of the concepts that form the basis of how to compute mass spectra using quantum chemical methods. Section 2.1 introduces the quantum-mechanical concepts used throughout this work by providing the description of the Hamiltonian (Section 2.1.1), how the Hartree-Fock (HF) method is set up (Section 2.1.2), introducing the concepts of density functional theory (DFT, Section 2.1.3), and finally explaining the details of the GFN n -xTB methods that are used primarily in this work (Section 2.1.5). Following this, Section 2.2 focuses on the mass spectrometry aspects. As QCxMS is heavily based on molecular dynamics (MD) simulations, an introduction to the technical features of MDs is given in Section 2.2.1. A general overview of the QCxMS method is shown in 2.2.2, and the subsequent sections focus on the specifics for computing EI- (2.2.3) and CID- (2.2.4) based mass spectra.

2.1 Electronic Structure Methods

2.1.1 The Electronic Hamiltonian

In classical mechanics, the total energy E of a system is described by its kinetic (T) and potential (V) energy, which can be expressed in terms of the Hamiltonian mechanics (Eq. 2.2).

$$E = T + V \quad (2.1)$$

$$\mathcal{H} = \frac{p^2}{2m} + V(r) \quad (2.2)$$

with the classical Hamiltonian \mathcal{H} , p being the momentum, m the mass, and r the space coordinate. For its use in quantum mechanics, the Hamilton operator \hat{H} can be formulated similarly.

$$\hat{H} = \hat{T} + \hat{V} \quad (2.3)$$

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(r) \quad (2.4)$$

with \hbar the reduced Planck constant and the Laplace operator ∇^2 . For simplification, atomic units are

used from here onwards.

To obtain the energy of a molecular system with N_e electrons e and N_n nuclei n , the Hamilton operator can be formulated according to equation 2.5.

$$\hat{H} = \underbrace{-\sum_{i=1}^{N_e} \frac{\nabla_i^2}{2}}_{\hat{T}_e} - \underbrace{\sum_{A=1}^{N_n} \frac{\nabla_A^2}{2M_A}}_{\hat{T}_n} - \underbrace{\sum_{i=1}^{N_e} \sum_{A=1}^{N_n} \frac{Z_A}{r_{iA}}}_{\hat{V}_{ne}} + \underbrace{\sum_{i=1}^{N_e} \sum_{\substack{j=1 \\ j < i}}^{N_e} \frac{1}{r_{ij}}}_{\hat{V}_{ee}} + \underbrace{\sum_{A=1}^{N_n} \sum_{\substack{B=1 \\ B < A}}^{N_n} \frac{Z_A Z_B}{r_{AB}}}_{\hat{V}_{nn}} \quad (2.5)$$

with i, j being the indices of the electrons, and A, B the indices for the nuclei. r is the distance between the respective coordinates, and Z is the nuclear charge. M represents the ratio of the mass of a nucleus to the mass of an electron.

As nuclei are much heavier than electrons, it can be assumed that the nuclear motion is stationary in comparison to the movement of the electrons. This simplification is known as the Born–Oppenheimer approximation,⁵⁷ which enables the formulation of an electronic Hamiltonian \hat{H}_{el} with the electronic kinetic energy \hat{T}_e , the electron-nuclei interaction \hat{V}_{ne} , and the electron-electron interaction \hat{V}_{ee} in equation 2.6.⁵⁸

$$\hat{H}_{\text{el}} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ne} \quad (2.6)$$

With the time-independent electronic Schrödinger equation (SE) 2.7,⁵⁹ the electronic energy E_{el} in dependence of a given wave function Ψ is obtained by the eigenvalue equation 2.7.

$$\hat{H}_{\text{el}}\Psi = E_{\text{el}}\Psi \quad (2.7)$$

The equation can only be solved for a few systems exactly, e.g., the hydrogen atom. For systems with multiple electrons, approximations have to be made.

2.1.2 Hartree–Fock Theory

To solve the SE for a many-particle system, a wave function has to be constructed. A possible approach is the formulation of a single Slater determinant (SD), in which one-electron orbitals ϕ_i are used to construct a multi-electron molecular wave function Ψ^{SD} for a system with N_e electrons.

$$\Psi^{\text{SD}}(1, 2, \dots) = \frac{1}{\sqrt{N_e!}} \begin{vmatrix} \phi_1(1) & \phi_2(1) & \cdots & \phi_{N_e}(1) \\ \phi_1(2) & \phi_2(2) & \cdots & \phi_{N_e}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(N_e) & \phi_2(N_e) & \cdots & \phi_{N_e}(N_e) \end{vmatrix} \quad (2.8)$$

The advantage of using an SD is the fulfillment of the Pauli principle of antisymmetry, as well as the indistinguishability of electrons. The energy of a normalized wavefunction can be calculated as the expectation value of the electronic Hamiltonian operator, resulting in the Hartree–Fock energy expression (Eq. 2.9).

$$E_{\text{el}} = \langle \Psi^{\text{SD}} | \hat{H}_{\text{el}} | \Psi^{\text{SD}} \rangle = \sum_{i=1}^{N_e} h_i + \frac{1}{2} \sum_{i=1}^{N_e} \sum_{j=1}^{N_e} (J_{ij} - K_{ij}) \quad (2.9)$$

The individual energy terms can be expressed using the Dirac notation⁶⁰ in a one-electron term (Eq. 2.10) and the two-electron Coulomb (Eq. 2.11) and exchange (Eq. 2.12) interactions.

$$h_i = \left\langle \phi_i \left| \frac{-\nabla_i^2}{2} - \sum_{A=1}^{N_n} \frac{Z_A}{r_{iA}} \right| \phi_i \right\rangle \quad (2.10)$$

$$J_{ij} = \left\langle \phi_i \phi_j \left| \frac{1}{r_{ij}} \right| \phi_i \phi_j \right\rangle \quad (2.11)$$

$$K_{ij} = \left\langle \phi_i \phi_j \left| \frac{1}{r_{ij}} \right| \phi_j \phi_i \right\rangle \quad (2.12)$$

By applying the variational principle, a set of orbitals is determined that minimize the energy of the wavefunction. Orthonormality of the orbitals is ensured by the use of Lagrange multipliers ϵ_i , which yields the Hartree–Fock equation 2.13.

$$\underbrace{\left(h_i + \sum_j^{N_e} (J_j - K_j) \right)}_{F_i} \phi_i = \epsilon_i \phi_i \quad (2.13)$$

The effective one-electron Fock operator F_i is obtained, in which the kinetic energy and electron-nuclear interaction are expressed by h_i , and repulsion to all other electrons (*via* Coulomb (J_j) and exchange (K_j)) is treated by a mean-field approximation. However, through the mean-field treatment, correlation effects of electrons with opposite spin are not considered. The energy retained by HF is an upper bound to the exact solution of the Schrödinger equation in the Born–Oppenheimer approximation. The difference in energy between HF and the exact solution is the correlation energy.

2.1.3 Kohn–Sham Density Functional Theory

Density functional theory (DFT)⁶¹ is based on a one-to-one mapping between the electronic energy E_{el} and a functional of the electron density $F[\rho]$, which is described by the first Hohenberg–Kohn (HK) theorem.⁶² The advantage of the DFT approach is that the calculation of $3N$ spatial variables of the electrons is reduced to a single three-dimensional problem of the electron density. Under the Born–Oppenheimer approximation, the electronic Hamiltonian can be formulated in analogy to equation 2.6, with the kinetic energy of the electrons $T_e[\rho]$, the nuclei-electron interaction $V_{ne}[\rho]$ and the electron-electron interaction $V_{ee}[\rho]$. The latter can be decomposed into its Coulomb ($J[\rho]$) and exchange ($K[\rho]$) part in equation 2.15.

$$E_{\text{el}}[\rho] = T_e[\rho] + V_{ne}[\rho] + V_{ee}[\rho] \quad (2.14)$$

$$= T_e[\rho] + V_{ne}[\rho] + (J[\rho] + K[\rho]) \quad (2.15)$$

For $V_{ne}[\rho]$ and $J[\rho]$, classical expressions can be expressed in equations 2.16 and 2.17.

$$V_{ne}[\rho] = \sum_{A=1}^{N_n} Z_A \int \frac{\rho(r_i)}{r_{Ai}} d^3 r_i \quad (2.16)$$

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(r_i)\rho(r_j)}{r_{ij}} d^3 r_i d^3 r_j \quad (2.17)$$

The kinetic energy of the electrons $T_e[\rho]$ cannot be formulated classically. By adoption of the orbital picture used in the HF theory, Kohn–Sham orbitals can be introduced, and the kinetic energy expressed as

$$T_e^{\text{KS}} = \sum_{i=1}^{N_e} \left\langle \phi_i \left| -\frac{1}{2} \nabla^2 \right| \phi_i \right\rangle \quad (2.18)$$

Using orbitals, the electronic one-body density can be obtained from a single Slater determinant in analogy to the HF approach.

$$\rho(r) = \sum_i^{N_e} |\phi_i|^2 \quad (2.19)$$

However, T_e^{KS} is calculated under the assumption of non-interacting electrons. The energy difference to the exact kinetic energy, as well as the electron correlation and exchange, are summed up in equation 2.20 to the exchange-correlation term $E_{xc}[\rho]$.

$$E_{xc}[\rho] = (T_e[\rho] - T_e^{\text{KS}}[\rho]) + (V_{ee}[\rho] - J[\rho]) \quad (2.20)$$

This leads to the KS-DFT total electronic energy expression $E_{\text{el}}^{\text{KS}}[\rho]$.

$$E_{\text{el}}^{\text{KS}}[\rho] = T_e^{\text{KS}}[\rho] + V_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \quad (2.21)$$

As the exact expression of the exchange-correlation functional $E_{xc}[\rho]$ is unknown, various approximations were designed to solve this issue. In the local density approximation (LDA), the density ρ is treated as a uniform electron gas. The generalized gradient approximation (GGA) includes the gradient $\nabla\rho$ of the density to better account for systems with inhomogeneous electron densities, i.e. molecules. Including the second derivation of the gradient $\nabla^2\rho$ is the basis of meta-GGA functionals. A significant improvement of the functional approximation is obtained by hybrid functionals that treat non-local electron exchange through Fock-exchange contributions (see Eq. 2.12). Better treatment of electron correlation can be introduced by including virtual orbitals in the calculations, leading to double-hybrid functionals. The functionals can be sorted by the quality of their results, that increase with the mathematical complexity from LDA \rightarrow GGA \rightarrow meta-GGA \rightarrow hybrid \rightarrow double-hybrid.⁶³

Semi-local functionals cannot describe long-range correlation effects, and thus non-covalent interactions are often not sufficiently described.⁶⁴ The attractive part of the van der Waals interaction potentials between atoms and molecules can be accounted for by the established DFT-D3⁶⁵⁻⁶⁷ and DFT-D4^{68,69} dispersion correction schemes.

2.1.4 Basis Sets

The unknown molecular wave functions ϕ used in the SD can be approximated in the linear combination of atomic orbitals (LCAO) approach using a linear combination of m atomic orbitals (or basis functions) χ and coefficients c .

$$\phi_i = \sum_{\alpha}^m c_{\alpha i} \chi_{\alpha} \quad (2.22)$$

Variational minimization of the energy expressions of HF or DFT yields the Roothaan–Hall equations in matrix notation.

$$\mathbf{FC} = \mathbf{SC}\epsilon \quad (2.23)$$

with the \mathbf{S} matrix containing the overlap between the basis functions, the \mathbf{F} matrix representing the Fock operator (equation 2.13), and ϵ the diagonal matrix of the orbital energies ϵ_i . The molecular orbital coefficient matrix \mathbf{C} is varied using the iterative self-consistent field (SCF) method until a set of MOs is found that most accurately describes the system. Slater functions can correctly describe the atomic orbitals, but due to the lower computational effort, a linear combination of several Gaussian-type functions is more commonly used. An exponential term describes the expansion of an orbital in space. Choosing larger exponents leads to less expansion, while small exponents describe diffuse functions. The accurate description of charged systems can be influenced by the choice of the functions, e.g., for negatively charged ions, the loosely bound valence electrons can be best represented using diffuse functions. As mainly charged species are relevant in MS, the influence of basis functions has to be considered for calculations.

2.1.5 Extended Tight-Binding Methods

The semi-empirical quantum mechanics (SQM) GFN n -xTB ($n = 1, 2$) methods⁷⁰⁻⁷² find their origin in the density functional tight-binding approach (DFTB). Similar to other SQM approaches, only valence shell electrons are considered, which are computed in a small basis set with an empirical Hamiltonian. Both GFN n -xTB methods use a global, element-specific parameter approach focusing on molecular geometries, vibrational frequencies, and non-covalent interaction energies for elements up to radon ($Z = 86$).

To obtain an energy expression, equation 2.21 is expanded in a Taylor series. The GFN n -xTB methods employ a truncation after the third-order term, as shown in equation 2.24.

$$E[\rho] = E^{(0)}[\rho_0] + E^{(1)}[\rho_0, \delta\rho] + E^{(2)}[\rho_0, (\delta\rho)^2] + E^{(3)}[\rho_0, (\delta\rho)^3] + \dots \quad (2.24)$$

with a fixed reference density ρ_0 and the electronic structure calculations are conducted in respect to density fluctuations $\delta\rho$.

The corresponding energy expressions for GFN1-xTB are displayed in equation 2.25 and for GFN2-xTB in equation 2.26.

$$E_{\text{GFN1-xTB}} = E_{\text{rep}}^{(0)} + E_{\text{EHT}}^{(1)} + E_{\gamma}^{(2)} + E_{\Gamma}^{(3)\text{GFN1}} + E_{\text{disp}}^{D3} + G_{\text{Fermi}} + E_{\text{XB}} \quad (2.25)$$

$$E_{\text{GFN2-xTB}} = E_{\text{rep}}^{(0)} + E_{\text{EHT}}^{(1)} + E_{\gamma}^{(2)} + E_{\Gamma}^{(3)\text{GFN2}} + E_{\text{disp}}^{D4} + G_{\text{Fermi}} + E_{\text{AES}} + E_{\text{AXC}}. \quad (2.26)$$

The energy terms are grouped according to their contribution order. The repulsion energy (Eq. 2.27) is of zeroth-order and includes contributions such as the Coulomb repulsion between nuclei.

$$E_{\text{rep}}^{(0)} = \frac{1}{2} \sum_{A,B} \frac{Z_A^{\text{eff}} Z_B^{\text{eff}}}{R_{AB}} e^{-\sqrt{\alpha_A \alpha_B} (R_{AB})^{k_f}} \quad (2.27)$$

with the element-specific constants Z^{eff} defining the magnitude for the repulsion energy and the α exponents are element-specific parameters. $k_f = 3/2$ is a global parameter, which is $k_f^{\text{H,He}} = 1$ for first row elements in GFN2-xTB.

The extended-Hückel-Type (EHT) energy includes the main first-order contribution in equation 2.28.

$$E_{\text{EHT}}^{(1)} = \sum_{\mu\nu} P_{\mu\nu} H_{\mu\nu}^{\text{EHT}} (R_{AB})^{k_f} \quad (2.28)$$

with the valence electron density matrix $P_{\mu\nu}$ given in the non-orthogonal AO basis, and the Hückel–Hamiltonian matrix $H_{\mu\nu}^{\text{EHT}}$ obtained from solving the Roothaan–Hall equations 2.23.

Second-order exchange–correlation and isotropic electrostatic energy contributions are included in E_{γ} .

$$E_{\gamma}^{(2)} = \frac{1}{2} \sum_{A,B}^{N_{\text{atoms}}} \sum_{l \in A} \sum_{l' \in B} q_l q_{l'} \gamma_{AB, ll'} \quad (2.29)$$

with the Mulliken partial shell charges q_l . The Coulomb interaction energy $\gamma_{AB, ll'}$ includes damping functions and atomic hardness parameters.

In the third-order term, the exchange–correlation and isotropic energy corrections include shell-resolved partial charge q_A and an element-specific atomwise Hubbard parameter Γ .

$$E_{\Gamma}^{(3)} = \frac{1}{3} \sum_A^{N_{\text{atoms}}} \sum_{l \in A} q_{A,l}^3 \Gamma_{A,l} \quad (2.30)$$

For GFN2-xTB, a shell-specific global parameter K_l^{Γ} is added to equation 2.30.

A comparison of equations 2.25 and 2.26 indicate the more sophisticated approach used in the GFN2-xTB Hamiltonian. The GFN2-xTB method was developed to include multipole electrostatics (E_{AES}) and anisotropic exchange–correlation (E_{AXC}) contributions. In contrast to GFN1-xTB, no halogen or hydrogen bond parametrization (E_{XB}) is needed, and instead of the DFT-D3 dispersion correction (E_{disp}^{D3}), the more sophisticated DFT-D4 (E_{disp}^{D4}) correction is used. Similarities of the methods are found in the repulsion energy term 2.27 and the description of the extended Hückel term

2.28.

An important aspect in both GFN n -xTB methods is the finite electron temperature approach (so-called ‘‘Fermi–Smearing’’) ⁷³ for static correlation treatment, as shown in equation 2.31.

$$G_{\text{Fermi}} = k_{\text{B}} T_{\text{el}} \sum_{\sigma=\alpha\beta} \sum_i [n_{i\sigma} \ln(n_{i\sigma}) + (1 - n_{i\sigma}) \ln(1 - n_{i\sigma})] \quad (2.31)$$

with k_{B} being the Boltzmann constant, T_{el} the electronic temperature, and $n_{i\sigma}$ the fractional occupation number of the spin-MO $\phi_{i\sigma}$. The latter is obtained by the Fermi-distribution

$$n_{i\sigma} = \frac{1}{e^{(\epsilon_i - \epsilon_F^\sigma)/(k_{\text{B}} T_{\text{el}})} + 1} \quad (2.32)$$

with the energy ϵ_i of orbital ϕ_i and ϵ_F^σ the Fermi level in the respective spin-orbital space (α or β). A variational solution to fractional orbital occupations is introduced, which provides numerical exact gradients of equations 2.25 and 2.26. The approach leads to robust convergence and treatment of static correlation, making its application particularly suitable in high-temperature MD simulations. For this reason, the finite temperature approach is thoroughly used for calculating mass spectra, as explained in Section 2.2.

The basis set used in GFN n -xTB is treated in the LCAO approach (see Section 2.1.4). It consists of minimal-valence, atom-centered Gaussian functions, which are contracted from three to six primitives m to approximate Slater-type functions (STO- m G). For higher row elements (third row etc.), d-polarization functions are employed. ⁷⁰

2.2 Computation of Mass Spectra

The computation of mass spectra requires the capability of a method to account not only for all possible decomposition reactions but ultimately resolve rearrangement reactions necessary for recovering the entirety of relevant fragments. This is especially crucial when considering unknown rearrangements for uncharted chemical compounds. The most pristine method to examine fragmentation reactions occurring during an MS experiment is using MD simulations, which are introduced in Section 2.2.1. They are able to calculate the processes taking place in a close-to-the-experiment manner.

2.2.1 Potential Energy Surface and Molecular Dynamics

The electronic energy of a molecule depends on the position of its atoms in space, which can be calculated by the methods described in Sections 2.1.2, 2.1.3, and 2.1.5. The relation between the molecular geometry and its energy yields a potential energy hypersurface, commonly referred to the potential energy surface (PES).

In Born–Oppenheimer molecular dynamics (BO-MD), a nucleus A moves classically in the potential V_e^{BO} of its electrons e (Eq. 2.33). ⁷⁴

$$M_A \frac{d^2 R_A}{dt^2} = -\nabla_A V_e^{\text{BO}}(\{R_A(t)\}) \quad (2.33)$$

$$= -\underbrace{\nabla_A E}_{F_A} \quad (2.34)$$

with M_A being the mass of the nucleus A , R_A its position at time t , and F_A the force acting on it. In the leap-frog algorithm, the forces acting on the nuclei are calculated “on-the-fly” by solving the static electronic structure stepwise for fixed nuclear positions with their negative gradient (Eq. 2.34). The position R_A and velocity v_A of all nuclei are alternately evaluated with the time step δt .

$$R_A(t + \delta t) = R_A(t) + v_A(t + \frac{\delta t}{2})\delta t \quad (2.35)$$

$$v_A(t + \frac{\delta t}{2}) = v_A(t - \frac{\delta t}{2}) + a_A \delta t \quad (2.36)$$

with the acceleration of the nuclei given by $a_A = F_A/M_A$. To sample the forces on all nuclei correctly, the MD timestep has to be set to be an order of magnitude smaller than the fastest nuclear motion, which often corresponds to the hydrogen vibrational oscillation.

The conditions under which MD simulations are performed can be distinguished into microcanonical and canonical ensembles. In the microcanonical ensemble (NVE), the number of particles N , volume V , and total energy E of the system remain constant. The canonical ensemble (NVT) is used for kinetic energy rescaling of a molecule/ensemble to a constant temperature T . A commonly used model for heat exchange simulation in MDs is the Berendsen thermostat,⁷⁵ which methodically relaxes a system to a target temperature. Fluctuations during the scaling of the kinetic energy are reduced by a time constant τ (Eq. 2.37).

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \quad (2.37)$$

2.2.2 General Setup of QCxMS

QCxMS was developed in such a way that the EI and CID modules can be used with the same setup. An overview of the workflow is displayed in figure 2.1 and can generally be described by the following steps:

1. Equilibration
2. Ensemble generation
3. Ionization
4. Activation
5. Mean-free-path sampling

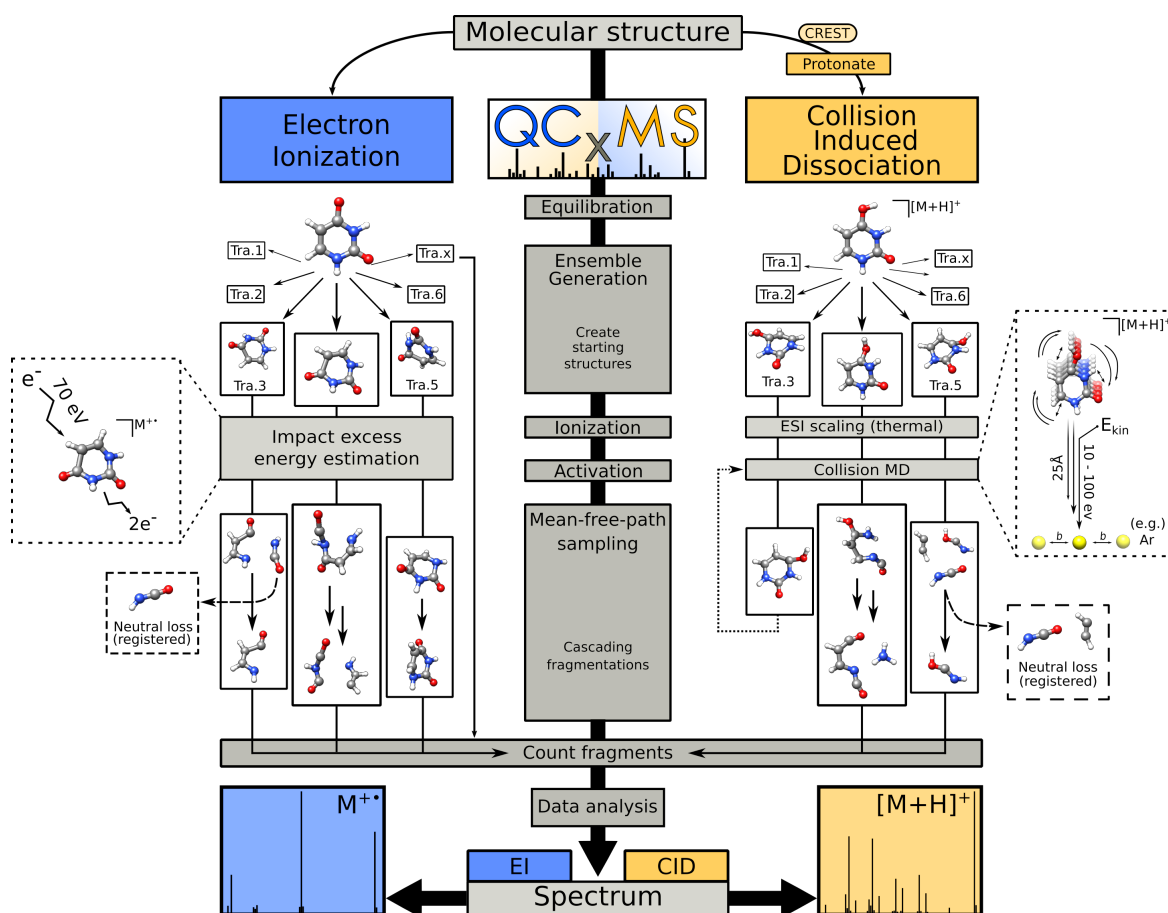


Figure 2.1: Flowchart of the QCxMS protocol for the EI (left) and CID (right) run-modes.

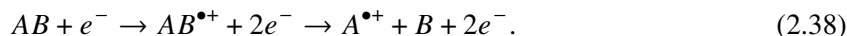
6. Data analysis

Equilibration of the molecule (Step 1) is followed by the sampling of an MD trajectory at constant energy (NVE), from which snapshot geometries are taken (Step 2) that represent an ensemble of random starting points for the fragmentation runs. For each snapshot geometry, ionization and activation simulations are conducted (steps 3 and 4) that are distinct for the EI and CID run modes (see below). Following the activation, the mean-free-path (MFP) is sampled (Step 5), which represents the reaction time of the ion to undergo fragmentation. As soon as fragmentation occurs, the delta self-consistent field (Δ SCF) approach⁷⁶ is used to compute the ionization potential (IP) (in positive ion mode) or the electron affinity (EA) (in the negative ion mode) for the produced fragments. Fractional charge is allocated to the fragments by Boltzmann distribution. Adapted from the Stevenson's rule,⁷⁷ the fragment with the largest charge is taken for extended simulations, while the other parts are counted according to their partial charge. If the internal energy (E_{int}) is sufficient, cascading fragmentation events for the most charged fragment are accounted for by consecutive MD simulations. Finally, the fragments are counted and plotted (Step 6) using the external Plot Mass Spectrum (PlotMS) program, which is part of the QCxMS software package. Details on the exact working mechanics are provided in various earlier publications,^{50,54} as well as in chapters 3, 4, and 6.

2.2.3 Computing Electron Ionization Mass Spectra

This section provides an overview of the specifics required to compute EI-MS. The application of the method is shown in chapter 3.

A typical EI process is described by



The energy needed to remove an electron e^- from a molecule AB is expressed as the ionization energy. The energy excess not used for ionization is defined as the impact excess energy (E_{IEE}), which is stored in the internal energy of the ion. As the transition of e^- through AB occurs faster than any nuclei motion, a vertical ionization of AB can be considered according to the Franck–Condon principle.⁷⁸ From this principle it can be assumed that the ionization is completed before the nuclei of AB move to a new equilibrium position. The resulting electronically excited ion states relax by internal conversion (IC) to their vibrationally excited (or “hot”) ion ground states.

The implementation into QCxMS can be seen on the left-hand side of figure 2.1. In the program, E_{IEE} is determined from the molecular orbital (MO) levels of AB . The probability P of a fragmentation run to have an E_{IEE} equal to E is calculated stochastically by a Poisson-type distribution (Eq. 2.39).

$$P(E) = \frac{e^{cE(1+\ln(b/cE))-b}}{\sqrt{aE+1}} \quad (2.39)$$

with the empirically determined parameters $a \approx 0.2$ eV, $b \approx 1$, and $c = 1/aN_e$, with N_e being the number of valence electrons. E_{IEE} is transferred into E_{int} of the molecule *via* scaling the nuclear velocities in Step 4. The IC time τ_{IC} is calculated from the differences in energy between the ionized MO i to the higher lying MOs j up to the highest occupied MO (HOMO) by the energy-gap law.⁷⁹

$$\tau_{\text{IC}} = \sum_{j>i}^{\text{HOMO}} \frac{k_h}{N_{el}} e^{\alpha(\epsilon_i - \epsilon_j)} \quad (2.40)$$

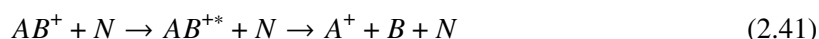
with $\alpha = 0.5$ eV⁻¹ and $k_h = 2$ ps, and ϵ_i the orbital energy of the i -th orbital.

Several hundreds of production runs have to be performed to correctly account for the statistical distribution of E_{IEE} . The number of runs scales with the molecular ion’s degrees-of-freedom (DOF). The finite temperature method is utilized to promote convergence of the electronic wave function (Eq. 2.31). This method accounts for the partial multireference electronic structure of the ions and simultaneously represents the electronically excited states of the ion. If E_{IEE} is sufficiently large enough, dissociation of $AB^{\bullet+}$ into $A^{\bullet+} + B$ or $A^+ + B^{\bullet}$ can occur. The statistical charge of each fragment is considered a factor for its abundance in the final spectrum. Alternatively, the rate constants by which the reactions occur can be calculated classically using the Rice–Ramsperger–Kassel–Marcus (RRKM)^{80–82} or the Quasi–Equilibrium–Theory (QET)⁸³ approach. An overview of these methods compared to QCEIMS is given in reference [54]. More details on the EI run-mode can be found in the original publication [50].

2.2.4 Computing Collision Induced Dissociation Mass Spectra

A significant part of this thesis comprises the development of the CID method for QCxMS. The right-hand side of Figure 2.1 displays the general workflow of how the method was implemented into QCxMS. The detailed development of this method is the content of Chapter 4.

In general, CID can be seen as a two-step process, represented in equation 2.41.¹¹ First, an ion AB^+ is activated by a collision with a neutral atom N . This converts the collision energy E_{coll} of AB^+ into its internal energy $E_{\text{int}}^{AB^+}$. In the second step, the now activated species AB^{+*} undergoes fragmentation. The internal energy of the activated species $E_{\text{int}}^{AB^{+*}}$ determines the possible fragmentation pathways the structure can follow and is determined by equation 2.42.



$$E_{\text{int}}^{AB^{+*}} = E_{\text{int}}^{AB^+} + E_{\text{coll}} \quad (2.42)$$

Before the collision, $E_{\text{int}}^{AB^+}$ is generally lower than E_{coll} and depends on the amount of energy gained during the ionization process. The maximum E_{coll} that can be obtained by a collision is called the center-of-mass energy E_{COM} and is defined by the mass of the collision gas atom m_g , the mass of the precursor ion m_p , and laboratory frame of reference energy E_{LAB} (Eq. 2.43).⁸⁴

$$E_{\text{COM}} = \frac{m_g}{m_g + m_p} E_{\text{LAB}} \quad (2.43)$$

Other factors such as the collision angles and hardness of the collision partners can reduce this energy. The more flexible AB^+ is and the more DOF it has, the more E_{coll} can be absorbed until fragmentation occurs.

During the development of the CID mode, many different aspects had to be considered, as various factors influence the collision procedure and the internal energy of the ion in an experiment. Ionization processes, collision gas pressure, collision cell parameters, detection methods, and even the instrument itself significantly influence the resulting spectrum. Three run-types were developed that represent the most accurate methods for calculating a CID mass spectrum. A schematic depiction of the run-types is shown in figure 2.2. With these run-types, the program enables the computation of CID spectra using slow heating (*thermal activation*), as well as single and multiple collision (*forced activation*) simulations, but these approaches depend on various settings and require trial-and-error computations. The most significant development of this thesis is the construction of the *general activation* run-type, which combines the other two run-types to enable the automatic computation of mass spectra without setting many parameters. A detailed discussion of these run-types, their application, and their advantages and disadvantages, is provided in Section 4.

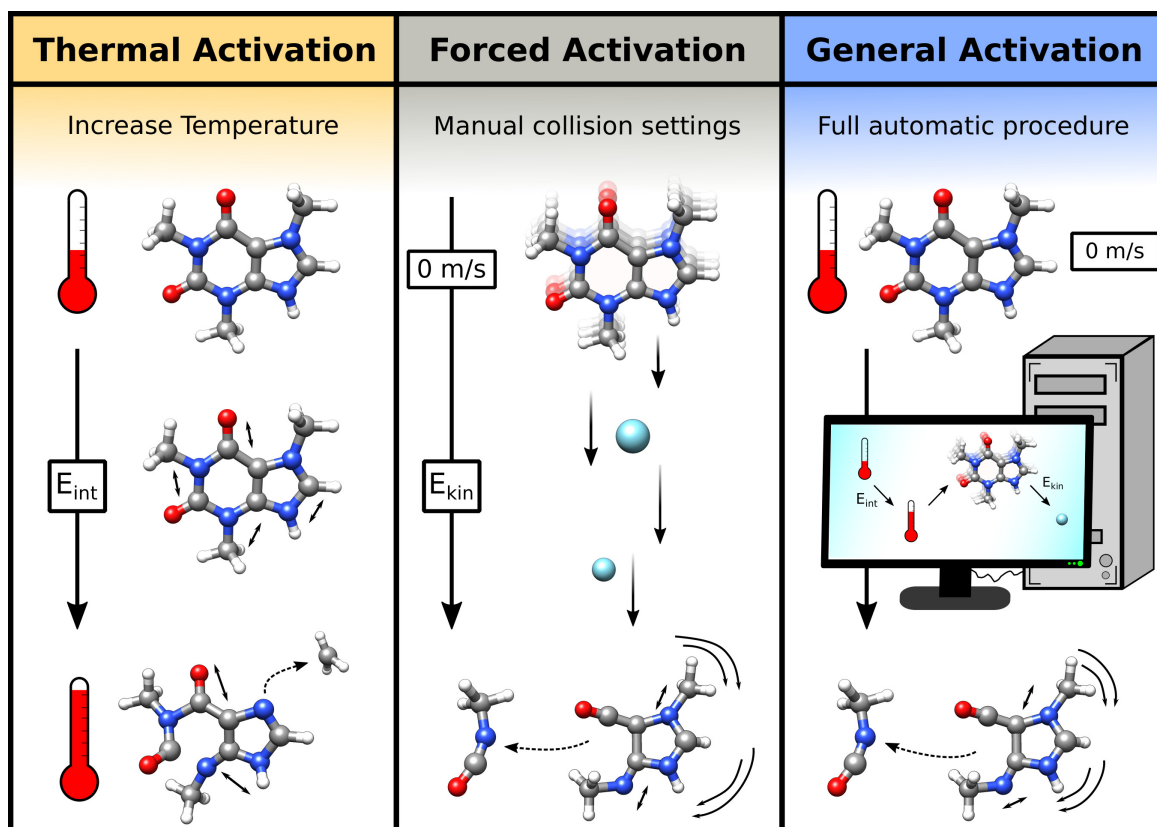


Figure 2.2: The three run-types developed for the CID mode of QCxMS.

2.2.5 Protonation Tool and Thermochemistry

As described in chapter 1, ionization for CID MS is obtained *via* (de-)protonation of the molecule. Throughout this thesis, the fully automated tool developed by Pracht *et al.*⁸⁵ was used to determine all possible (de-)protonated structures (called “protomers”) of an input molecule before CID calculations. The program computes the localized MOs (LMO) of a molecular structure and locates its lone pairs (LP) and π centers. Consecutively, a proton is added along the axis of the LMO charge center. The structures are then energetically ranked after optimization. Deprotonation is conducted by removing all H atoms of the structure and consecutive sorting the optimized protomers. The protonation tool is implemented in Conformer Rotamer Ensemble Search Tool (CREST),⁸⁶ which uses GFN2-xTB as default for the optimization and energy computations. Because the correct free energy ranking of the protomers is essential to determine the actual structure for the mass spectrum calculations, the protomers are routinely re-ranked throughout this thesis on the DFT level using the Command-line Energetic SORTing Algorithm (CENSO) developed by Grimme *et al.*⁸⁷ The program determines the protomer populations by sorting the geometries according to their differences in Gibbs free energy ΔG . While MS is conducted in the gas phase, the influence of solvation can be substantial considering LC-MS experiments. The free energy is computed at a given temperature T according to equation 2.44.

$$G(T) = E_{el} + \delta G_{solv}(T) + G_{TRV}(T) \quad (2.44)$$

with E_{el} being the molecular energy obtained from the electronic Hamiltonian, δG_{solv} being the solvation free energy at given temperature T , and G_{TRV} the translational, rotational, and vibrational free energies, including zero-point vibration and volume work terms. The thermostatical contributions are thereby calculated using the modified rigid-rotor-harmonic-oscillator (mRRHO) approximation.⁸⁸ Population analysis is done in equation 2.46 by Boltzmann statistics, with probability p_i and the free energy G_i of state i .

$$G = \sum_i^N p_i G_i \quad (2.45)$$

$$p_i = \frac{e^{-G_i/kT}}{\sum_i^N e^{-G_i/k_B T}} \quad (2.46)$$

Overview of “Calculation of Electron Ionization Mass Spectra with Semiempirical GFN n -xTB methods”

Jeroen Koopman,¹ and Stefan Grimme¹

Received: 02 July 2019,

Published: 05 September 2019

Reprinted (adapted) with permission² from
Koopman, J.; Grimme, S. *ACS Omega* **2019**, *4*, (12), 15120–15133.

— Copyright © 2019 American Chemical Society.

DOI [10.1021/acsomega.9b02011](https://doi.org/10.1021/acsomega.9b02011)

Own manuscript contribution

- Performing all calculations
- Interpretation of the results
- Writing the manuscript

¹ Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Berlingstraße 4, 53115 Bonn, Germany

² Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

In this initial work, the semiempirical GFN2-xTB⁷¹ method is implemented into the QCxMS program and compared to the performance of its predecessor GFN1-xTB⁷⁰ in terms of computing electron ionization mass spectra (EI-MS). As the results of GFN1-xTB excelled other quantum chemical methods in accuracy and computational costs for this purpose,⁵⁶ the question arose how the more sophisticated method would perform in terms of the calculation of realistic fragmentation patterns. An overview of the GFN n -xTB methods is provided in Section 2.1.5, and considerations concerning the computation of EI-MS are shown in Section 2.2.3. The full text of this work⁸⁹ can be found in its entirety in Appendix A.

For an unbiased evaluation, the default settings of the QCxMS program in the EI mode are used, and neither of the GFN n -xTB methods ($n = 1, 2$) is modified. A wide variety of smaller and larger organic, transition-metal, and main-group inorganic systems are computed to obtain an overview of the general applicability of the methods. Validation is obtained by comparison of the results produced by the two methods with experimental references taken from the NIST⁴⁰ and SDBS⁹⁰ databases. The computational demands and the stability of combinations of the GFN n -xTB methods for PES and IP calculations are discussed to obtain insights into their performance.

The first test set includes six organic molecules that vary in structure, size, and functionality, e.g., uracil, testosterone, and sucrose. Results using GFN2-xTB show better agreement with the experiment for testosterone, but no general improvement in comparison to GFN1-xTB is observed. Notably, GFN1-xTB performs better for hexafluorobenzene, and both methods yield qualitatively less accurate spectra for sucrose than for the other organic molecules. The second benchmark includes three molecules containing transition metal atoms. For this group of compounds, failure rates and computational demands are higher with GFN2-xTB in comparison to GFN1-xTB, and the newer method fails to predict important signals of the bis(benzene)chromium fragmentation pathways. Nevertheless, for the IP calculations of this set of molecules, GFN2-xTB can be recommended. The best overall agreement with the experimental spectra is obtained by combining GFN1-xTB for PES and GFN2-xTB for IP calculations. The third test group is composed of inorganic main-group molecules. The two GFN n -xTB methods perform similarly regarding timings, stability, and the overall agreement with the experiment. Both GFN n -xTB methods are recommended for MS computations of these systems. However, the accuracy for the MS computation of ferrocene improves significantly when using GFN2-xTB for IP calculations. Remarkably, the GFN2-xTB method completely fails for octasulfur, whereas the GFN1-xTB method performs very well.

To sum up, utilizing the two GFN n -xTB methods for EI-MS computations leads in the majority of instances to comparable decomposition pathways, but signal intensities often differ. Overall, the results from both methods compare reasonably well with the experimental results, but GFN2-xTB produces fewer artifacts in the spectra. Cross-checking the results using DFT for IP calculations can improve the quality of the obtained spectra but is computationally very demanding. Similar to the experimental variations between MS instruments, intensities, missing signals, and survival rates obtained by the calculations depend on technical settings related to internal energy distribution and the ionization/heating procedure. The settings can be individually modified, thereby significantly improving the simulated spectra, and can e.g., account for too intense $[M]^{•+}$ signals. Nevertheless, the work concludes that with default settings in QCxMS, the two tight-binding methods GFN1-xTB and GFN2-xTB broaden the applicability of the software for calculating EI mass spectra, producing an accurate matching to experimental results without the need for third-party software. The implementation of the GFN2-xTB method overall improves the elucidation of the structural compositions and fragmentation paths.

Overview of “From QCEIMS to QCxMS: A Tool to Routinely Calculate CID Mass Spectra Using Molecular Dynamics”

Jeroen Koopman,¹ and Stefan Grimme¹

Received: 16 March 2021,

Published: 03 June 2021

Reprinted (adapted) with permission² from
Koopman, J.; Grimme, S. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1735–1751.

— Copyright © 2021 American Chemical Society.

DOI [10.1021/jasms.1c00098](https://doi.org/10.1021/jasms.1c00098)

Own manuscript contribution

- Development of the concept
- Performing all calculations
- Interpretation of the results
- Writing the manuscript

¹ Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Berlingstraße 4, 53115 Bonn, Germany

² Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

The scope of this work is to develop a method that enables the computation of mass spectra based on activation by collision-induced dissociation (CID) reactions. Section 2.2.4 provides an overview of the theoretical considerations concerning CID processes. As an alternative to EI, the ionization of the molecules is considered to be obtained *via* protonation,¹¹ and the nature of the molecular ions between these approaches differ (usually open-shell (unpaired electron) vs. closed-shell (paired electron) electronic structure). Automated protonation and energetic ranking of the molecule are done as explained in Section 2.2.5. The greater generalizability of the software is represented by renaming it from QCEIMS to QCxMS.

The software development starts by designing collision MDs for any arbitrary ion structure with user-set input conditions, considering collisional energy in the laboratory (E_{LAB}) and the center-of-mass (E_{COM}) frame of reference. The rotational energy and sampling temperature of the molecular ion are automatically scaled to obtain the correct starting conditions. Multiple collisions are implemented by repeated simulation of collision MDs and interposed mean-free-path simulations. From this, the collision kinetics of single and multiple collisions are computed. Statements about the molecular stability and survival yields are gained, allowing for a detailed description of the amount of energy taken up by the molecular ion through each collision and showing the decrease of E_{COM} and E_{LAB} in the process. More details on the working mechanism can be found in Appendix B, in which the corresponding publication⁹¹ is presented.

The major problem encountered in this work is that calibration of a theoretical model to experimental conditions is generally not possible due to the considerable variation between instruments and a lack of insight in “typical” MS settings. For this reason, different run-types are developed to consider various experimental settings. These are depicted schematically in Figure 2.2.

The *thermal activation* run-type (Figure 2.2, left) yields fragmentation without simulation of direct collisions. This slow-heating process facilitates fast computations, as no actual collisions are computed. However, only statistical dissociation processes are considered, and the scaling can lead to under- or over-fragmentation. The *forced activation* run-type (Figure 2.2, middle) enables the calculation of any given number of ion–gas collisions. The simulation of ion–gas collisions accounts for direct fragmentation processes. However, the exact number of collision events that take place inside an MS instrument cannot be established in general. Finally, the automated *general activation* run-type (Figure 2.2, right) is developed. The internal energy of the protonated system is increased, followed by the simulation of multiple collisions between the precursor ion and neutral gas atoms. The number of collisions are estimated by the kinetic gas theory,⁸⁴ in which the pressure and temperature of the collision gas, as well as the collisional cross-section of the collision partners, approximate the conditions in the experiment. This estimation is also applied to collisions between produced fragments and neutral gas atoms.

The different run-types are tested on a benchmark of six diverse organic molecules. An interesting insight into the influences of protonation sites on the computed spectrum is gained, which is discussed in the scope of “mobile protons”. – the protomers are not in thermal equilibrium and the most populated structures do not necessarily account for the main fragmentation pathways, so kinetically controlled fragmentation reactions are observed. The GFN2-xTB Hamiltonian,⁷¹ which was implemented into QCxMS in Chapter 3, enables excellent ionic PES and IP calculations and produces good agreement with experimental spectra. The final software is the first quantum chemistry-based protocol that automatically computes unbiased, singly positively charged ion CID mass spectra without the need for pre-tabulated, database-driven algorithms or third-party software.

Overview of “Quantum Chemistry-based Molecular Dynamics Simulations as a Tool for the Assignment of ESI-MS/MS Spectra of Drug Molecules”

Romina Schnegotzki^{1,2}, Jeroen Koopman,^{3,2} Stefan Grimme³ and Roderich D. Süssmuth¹

Received: 05 February 2022,

Published: 02 March 2022

Reprinted (adapted) with permission⁴ from

Schnegotzki, R.; Koopman, J.; Grimme, S., Süssmuth, R.D. *Eur. J. Chem.* **2022**, *28*, e202200318.

— Copyright © 2022 Wiley-VCH GmbH.

DOI [10.1002/chem.202200318](https://doi.org/10.1002/chem.202200318)

Own manuscript contribution

- Performing all calculations
- Interpretation of the results
- Writing the manuscript

¹ Institut für Chemie, Technische Universität Berlin, Straße des 17.Juni 124,10623 Berlin, Germany

² These authors contributed equally to this work

³ Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany

⁴ Permission requests to reuse material from this chapter should be directed to the Wiley-VCH GmbH.

In this work, the two drug molecules nateglinide and zopiclone are computed by QCxMS in single positive ion mode. It is shown how the automatism implemented in QCxMS combined with the fast calculations conducted with GFN1-xTB⁷⁰ and GFN2-xTB⁷¹ are used to elucidate the fragmentation pathways of these complex molecules completely. The investigated compounds are of typical size for common bioactive molecules (45 and 51 atoms) and contain functional groups that are representatives of several drug classes. The spectra are computed using the *general activation* method, as introduced in the preceding work of Chapter 4. The complete work⁹² can be found in Appendix C.

As no information on the protonation sites is available, computations using the automated protonation tool of CREST⁸⁶ and thermodynamic ranking of the structures with CENSO⁸⁷ are conducted, as described in Section 2.2.5. The most populated protomers within a defined free energy window of 20 kcal/mol are used as starting points to calculate corresponding fragment spectra, allowing for multiple reaction pathways. It is found that the spectra for the different protomers do not significantly differ for either of the two structures, which can be traced back to the close vicinity of heteroatoms that enhance proton rearrangement reactions between the structures *via* mobile protons before fragmentation occurs. The computed signals are validated by comparison to experimental spectra measured on an Orbitrap Fusion ESI-MS/MS instrument. Excellent coverage of the computed to the measured signals of > 90 % is achieved when only the signal existence is considered. The agreement between signal intensities of the computations and measurements is satisfactory.

In the field of applied mass spectrometry, molecule fragmentation mechanisms of $[M+H]^+$ ions are typically elucidated by classical generic formalisms that are based on standard empirical rules originating from electron migration protocols.³¹⁻³⁴ However, fragmentation patterns are regularly observed that the generic rules cannot straightforwardly explain, and competing decomposition reactions can lead to uncertainty about the correct spectral assignment, so the final fragment structures are often approximated.^{31,93,94} As a solution to reduce the uncertainty, the trajectories calculated by the unbiased MD simulations of QCxMS are used to obtain detailed insights into the dissociation pathways of the ions. The fast and easy access to the trajectories is used to complement the classical interpretation of the final fragmentation pathway, simultaneously significantly shortening the time for structure elucidation. Categorization of the level of complexity for describing the dissociation pathways solely with classical fragmentation rules, the formation of the signals are sorted into “complexity levels”. Utilizing this merger of classical and quantum chemical methods, uncommon fragmentation pathways of four non-evident fragment ion structures are identified that contrast the accepted generic formulations. Furthermore, a non-negligible number of fragmentation reactions can only be assembled in retrospect by the classic rules after inspection of the trajectories obtained by the MD simulations of QCxMS.

In summary, this work shows how QCxMS can be used as a valuable tool to facilitate the detailed illustration of fragmentation mechanisms or even enables a description of decomposition pathways in the first place. While the differences between computations and measurements are evident, the discrepancies between these approaches are discussed and put into perspective. It is established that the comprehensive applicability of combining the results obtained using QCxMS in combination with the knowledge of generic rules allows the description of the vast majority of different molecular classes. Given the enormous structural diversity, a thorough interpretation of fragment spectra without the aid of QC encourages further investigation and might lead to an expansion of what to consider “typical fragmentation pathways”.

Overview of “Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules”

Jeroen Koopman,¹ and Stefan Grimme¹

Preprint submitted: 11 May 2022

Koopman, J.; Grimme, S. *ChemRxiv* **2022**

DOI [10.26434/chemrxiv-2022-w5260](https://doi.org/10.26434/chemrxiv-2022-w5260)

Manuscript received: 22 July 2022,

Manuscript published: 25 October 2022

Reprinted (adapted) with permission² from

Koopman, J.; Grimme, S. *J. Am. Soc. Mass Spectrom.* **2022**

— Copyright © 2022 American Chemical Society.

DOI [10.1021/jasms.2c00209](https://doi.org/10.1021/jasms.2c00209)

Own manuscript contribution

- Development of the concept
- Performing all calculations
- Interpretation of the results
- Writing the manuscript

¹ Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Berlingstraße 4, 53115 Bonn, Germany

² Permission requests to reuse material from this chapter should be directed to the American Chemical Society.

The final work in this thesis concludes the development of the CID module for the QCxMS software by enabling the calculations of mass spectra independently from the charge state of the molecular ion. In Chapters 4 and 5, single positively charged structures are calculated. The work presented here makes the *ab-initio* computation of structures carrying positive, negative, single, or multiple charges possible. This is especially important because commonly used LC-MS/MS methods routinely produce ions with multiple positive or negative charges,^{95,96} depending on the analyte and the ionization mode used. Details on the implementation are provided in full in Appendix D.

The negative ion mode is tested on a benchmark set of four organic molecules of small size, for which reference spectra are taken from standard databases.^{97,98} Computations are conducted using the CID *general activation* run-type with GFN2-xTB. The accuracy of the results is cross-examined using DFT methods with minimally augmented basis sets.⁹⁹ For PES calculations, PBE¹⁰⁰/ma-def2-SV(P) is used, while electron affinity (EA) computations, that enable the charge assignment after fragmentation, are handled using PBE0¹⁰¹/ma-def2-TZVP.

First, the molecular deprotonation sites and their influence on the computed fragment signals are investigated. Comparable to the finding for positive ion spectra computed in Chapter 3, it is found that the concept of “mobile protons” is significant for deprotonated structures. Free energy calculations of the protomers, computed at the r²SCAN-3c¹⁰² level of theory (see Section 2.2.5), indicate that the main fragmentation pathways are not necessarily created from the most populated structures.

Since fragmentation pathway analysis for negative ions with classical methods often lacks references,³³ the decomposition reactions of the benchmark molecules are examined in detail. Using the GFN2-xTB level of theory, the majority of signals observed in the database references are reproduced. A higher agreement with the database spectra signal intensities is reported for full-DFT computations. However, the increased computational cost by three orders of magnitude renders these computations not affordable for routine computations or large molecules. Combining the QC methods by using GFN2-xTB for PES and PBE0/ma-def2-TZVP for EA calculations shows only a minor increase in accuracy. It can be considered unimportant concerning the cost-to-accuracy ratio. From this, it can be established that the computations on the GFN2-xTB level of theory for negative ion mass spectra produce good results.

In the second part of this work, the mass spectra of two multiple positively charged molecular ions are computed at GFN2-xTB level of theory and compared to spectra taken from the literature.^{103,104}

For the doubly charged crizotinib, the comparison demonstrates that by the usage of the Δ SCF method (see Section 2.2.2), QCxMS correctly assigns multiple charges to single fragments. Through this, highly accurate coverage of all experimentally measured signals is obtained. It is shown that all fragments are produced from the doubly charged molecular ion instead solely from its singly charged counterpart, which was suggested in the literature.¹⁰³ The second ion is a triply charged lysine derivate, for which the fragment charge assignment is complicated. Five out of nine reported fragments are calculated correctly. However, the flexibility of the structure and the low level of theory used for the computations lead to some cases of wrong charge assignment. Cross-checking either of the structures by computation of the IP values using PBE0/def2-TZVP does not influence the outcome significantly.

Overall, this work shows how the expansion of the CID run mode enables computations of negatively and multiply charged compounds, thereby increasing its applicability for a broad range of chemical structures. For anions produced by deprotonation, the computations of decomposition channels can help to generate empirical rules for “classic” generic fragmentation formalisms. Furthermore, QCxMS can compute the mass spectra of compounds carrying multiple charges sufficiently.

Conclusion and Outlook

In this thesis, the computation of mass spectra using quantum chemical approaches is presented. To this end, the well-established QCEIMS program was used, which runs molecular dynamics simulations to sample the decomposition pathways of an input analyte structure. The PES is computed on-the-fly, allowing for close-to-the-experiment dissociation and rearrangement reactions. By sampling hundreds of fragmentation MDs in parallel, a statistically relevant number of decomposition processes are obtained that produce a virtually “black-box” mass spectrum. The method was originally designed to calculate EI processes, so a particular focus in this work was set to establish a CID pendant for this run mode. It was implemented into the QCEIMS software and enables the computation of soft-ionization-based mass spectra, which are produced through protonation and subsequent collisional activation. The specific name of QCEIMS was changed to QCxMS to account for the more general approach, in which the “x” stands for the MS types that can be used ($x = \text{EI, CID}$). The particular steps that contributed to the extension of the program are described throughout this thesis in successive chapters, for which brief summaries are given in the following.

Implementation and application of the GFN2-xTB method into QCxMS provided accurate computations of the PES and IP to compute EI mass spectra, which showed an improvement over its predecessor GFN1-xTB. Calculations of organic, main-group inorganic, and transition-metal-containing molecules showed excellent results for creating spectra with the new SQM method, and comparison to experimental results showed high agreement. The two GFN n -xTB methods were comparably robust in terms of stability. When comparing the computational timings, GFN2-xTB was slightly slower. Improvements over GFN1-xTB were observed for the different molecules. Especially for ferrocene, the use of GFN2-xTB for IP calculations enabled the correct computation of the mass spectrum, which otherwise failed when using the IPEA-xTB method. Vice versa, the computation of octasulfur could be accurately described by GFN1-xTB, which was not possible by its successor. Combining GFN2-xTB for IP with GFN1-xTB for PES calculations showed an overall improvement in accuracy and timings, and the combination is generally recommended. Overall, the semi-empirical tight-binding methods GFN1-xTB and GFN2-xTB proved to be versatile tools for fast and reliable computations for EI mass spectra.

In order to simulate LC-ESI-MS/MS experiments, a method had to be established and implemented into the available software from scratch. As the exact conditions of the experiments could not be established, the development of a general working mechanism was not straightforward. Different run-types for

molecular ion activation were developed to overcome this obstacle, allowing the automatic calculation of CID mass spectra. Dissociation reactions obtained by the *thermal activation* run-type showed a lack of direct fragmentations, usually induced through high-energy collisions. Computations are fast, but missing information on exact energy distributions of typical experiments lead to trial-and-error runs for accurate reproduction of the references. Fragmentation through high energy collision events can be simulated by the *collisional activation* run-type. However, as the number of collisions and the impact energy are very flexible parameters, it was not possible to establish general settings. The produced signal intensities showed strong dependence on the molecular ion properties, like size or structural flexibility. The *general activation* run-type was developed as a solution that combines thermal and collisional activation for automatic computations. The established run modes were used to successfully compute the mass spectra of various sample molecules at the GFN2-xTB level of theory. As the thermodynamically most populated structures were not always the ones that accounted for the most pronounced fragmentation pathways, it was found that the protonation site of the molecular ion had a significant influence on the outcome. In the literature, this is known as the “mobile proton” theory, for which the working mechanisms are still not entirely understood. Besides the computations of entire spectra, the CID run mode was able to provide insights into the collision kinetics, through which effects of single and multiple collisions on the survival yield of the molecular ion could be explored. The goal to create an applicable workflow to automatically compute singly positively charged CID mass spectra was achieved.

The ability of QCxMS to assist in complete structure elucidation of unknown fragmentation pathways was tested on the two large drug molecules nateglinide (45 atoms) and zopiclone (51 atoms). The *general activation* run-type of the CID mode was used with GFN1-xTB and GFN2-xTB to compute the decomposition routes of the two structures in singly protonated, positive ion mode. After determining all available protonation sites, the computations found that the different protonated molecular ion structures produced comparable dissociation patterns, from which it was determined that proton rearrangement occurred between the protonated structures before fragmentation. This was an important finding with regard to the mobile proton effect. The computed trajectories were used for retrospective assembly of the fragmentation pathways by empirical rules that are typically applied for mass spectral elucidation. The results were categorized into classes that reflected the level of complexity for applying the rules. It was found that two fragment structures for each of the two molecules could not be recreated *via* classical fragmentation pathways. They were instead only elucidated by the QCxMS computations. To conclude, the work demonstrated the usability of the CID mode to compute complicated decomposition reactions of unknown compounds for the complete structure elucidation of large molecules.

Finally, the CID module was extended to enable the calculations of CID mass spectra without charge restrictions. The method was tested on negatively and multiple positively charged molecules. For the negative ion mode, the PES of small sample molecules were calculated on SQM and DFT level of theory after initial deprotonation. EA computations were carried out on the GFN2-xTB level and cross-checked by PBE0/ma-def2-TZVP calculations. A considerable increase in accuracy by using DFT methods was not found. However, significant differences between the computed mass spectra of various protomer structures showed a severe influence of the mobile proton concept, even for deprotonated ions. Apart from singly charged structures, *ab-initio* methods were used to automatically calculate the mass spectra of two multiply charged structures for the first time. The computations for the doubly protonated crizotinib revealed a serious influence of solvation effects on the thermodynamic Boltzmann population rankings of the protomers. Fragmentation reactions displayed reasonable

agreement with the experimental reference, showing that the QC approach can compute multiply charged fragments correctly. The agreement was lower for the triply protonated lysine-derivate because the charge assignment failed for triply charged fragments. Hydrogen atom rearrangements typical for heterolytic fragmentation reactions were not sufficiently accounted for in the calculations, leading to false charge assignment after (consecutive) decomposition of the structure. Nevertheless, five out of nine reported signals could successfully be accounted for.

Overall, this work demonstrated the ability of QCxMS to correctly compute deprotonated and multiply protonated structures without the need for pre-tabulated fragmentation channels or database entries.

This thesis aimed to design a method that can create reference spectra for the unbiased structure elucidation of experimentally measured GC-EI-MS and LC-ESI-MS/MS mass spectra. Developing a concept that was not based on tabulated fragmentation pathways or machine learning of stored database information was essential. In summary, this goal was achieved through the development of QCxMS. While there are still aspects to explore considering the computation of mass spectra, the program produces “black-box” results by only using the structure of a molecule (covalent bonding topology, i.e., chemical formula).

Various relevant insights concerning soft-ionization methods and collisional activation were obtained by application of the software on diverse molecules. The correct treatment of protonation sites was found to be of major importance, and a deeper insight into the concept of the not yet fully conceptualized mobile proton theory on the dissociation behavior of molecules was gained. MS experiments are found to be very dynamic processes in which the ion structures are in a constant state of rearrangement. This makes spectrum prediction dependent on the ability to consider ever-changing processes, for which MD simulations are predestined. Direct fragmentation reactions obtained by simulating actual collision simulations are found to be crucial to obtaining the correct signal intensity, while thermal scaling only accounts for statistical fragmentation. Collisions between created fragments and neutral gas atoms are of significant relevance for the final spectrum, introducing sufficient internal energy for subsequent dissociation reactions.

Nevertheless, there is still room for improvement of the QCxMS working mechanisms. Dissociation pathways often lack the characteristics of heterolytic bond breakage, which are typical for soft ionization experiments. This is especially critical for multiply charged structures, as false charge assignment is the direct consequence of wrongly protonated fragments. Relying solely on Stevenson’s rule for charge assignment has weaknesses in this regard. Implementing proton affinity computations could be the first step in approaching this problem.

An increase in the applicability of QCxMS can be achieved by implementing surface-induced dissociation ($x = \text{SID}$) methods. Instead of a collision gas atom, a wall potential can be applied to simulate the molecule collision.

In order to reach the goal of creating an *in-silico* library, high-throughput computations of mass spectra must be established for all sorts of chemical compounds, which requires a drastic decrease in the computational resources needed for the simulations. Some considerations to obtain this goal are listed in the following:

1. Reduction of the number of protomers for CID computations – development of an efficient presorting algorithm for mobile proton rearrangements, e.g., by screening of the reaction barriers.

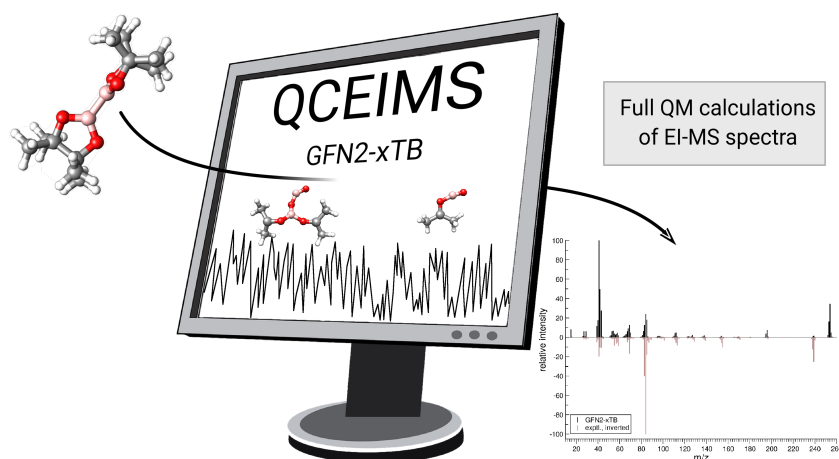
2. Shortening of the ensemble generation by sampling of the conformational space – e.g., using CREST instead of MD trajectories.
3. Using graph theory to pre-screen typical fragmentation reactions so fewer MD simulations have to be conducted for typical fragmentation reactions. Smaller number of MDs are then used to sample untypical dissociation and rearrangement reactions.

Apart from using the concepts of QCxMS, an algorithm can be developed that completely forgoes MD simulations. All possible fragments must be created that sample the entire reaction space of an MS experiment. Consecutively, the reaction barriers between the molecular ion and fragments are calculated, leading to a statistical result for the fragment signal intensities. This approach could decrease the computational resources and improve the accuracy by routine DFT computations for the reaction energies.

Within this work, many aspects needed to effectively calculate EI and CID mass spectra with QC methods using MD simulations were examined. It is clear that the highly complex subject tackled by QCxMS has many more facets than have been discussed here. During development, a variety of different features concerning the accurate description of MS processes were tested meticulously, e.g., other CID run-types, variation of starting internal energy, electronic temperature, and much more. However, only the elements that had a tremendous influence on the final result were presented. Everything that did not lead to an improvement of the algorithm was discarded. The contributions from this thesis and the development of QCxMS enable the routine structure elucidation of EI and CID mass spectrometry processes. To allow future developments, the software and documentation of its use are made available on the internet open-source¹⁰⁵ in the sense of scientific availability without restrictions.

Appendix

Publication corresponding to chapter 3: Calculation of Electron Ionization Mass Spectra with Semiempirical GFNn-xTB Methods



Abstract In this work, we have tested two different extended tight-binding methods in the framework of the quantum chemistry electron ionization mass spectrometry (QCEIMS) program to calculate electron ionization mass spectra. The QCEIMS approach provides reasonable, first-principles computed spectra, which can be directly compared to experiment. Furthermore, it provides detailed insight into the reaction mechanisms of mass spectrometry experiments. It sheds light upon the complicated fragmentation procedures of bond breakage and structural rearrangements that are difficult to derive otherwise. The required accuracy and computational demands for successful reproduction of a mass spectrum in relation to the underlying quantum chemical method are discussed. To validate the new GFN2-xTB approach, we conduct simulations for 15 organic, transition-metal, and main-group inorganic systems. Major fragmentation patterns are analyzed, and the entire calculated spectra are directly compared to experimental data taken from the literature. We discuss the computational costs and the robustness (outliers) of several calculation protocols presented. Overall, the new, theoretically more sophisticated semiempirical method GFN2-xTB performs well and robustly for a wide range of organic, inorganic, and organometallic systems.

A.1 Introduction

Nowadays, structure elucidation of molecules or condensed phases is one of the key ingredients in everyday work in chemistry and related sciences. Over the past decades, several excellent experimental methods, namely NMR, IR, Raman, and UV-vis spectroscopy, have been developed enhancing the facility of solving molecular structures tremendously. To date, it has become computationally affordable to use quantum chemical (QC) methods to calculate the properties that are needed to predict such spectra.¹⁰⁶ Another extremely important analytic tool for various areas in the organic and bioorganic chemistry is electron ionization mass spectrometry (EI-MS).^{11,14} Its daily application, e.g., in the field of forensic drug testing¹⁰⁷ or pharmacokinetics,¹⁰⁸ requires continuous investigation of many new substances and their structure-spectrum relationship. In practice, compound identification is often assisted using chemoinformatic approaches^{35,45,109} or database-driven programs.^{110,111} The ongoing development of computer-based neural networks aims to ease this procedure.⁴⁷ However, these approaches lack the basic physics and chemistry of the EI-MS process and do not have the ability to determine the basic reaction mechanisms leading to the observed spectra. While it is possible to predict previously unknown structures of molecules by these methods³⁵ the agreement with the experiments is often low.^{3,112}

Unfortunately, the straightforward computation of the required properties to generate accurate EI-MS is not possible with standard theoretical techniques. To tackle this problem, statistical methods based upon the quasi-equilibrium theory (QET)⁸³ or the Rice-Ramsperger-Kassel method (RRKM)^{80-82,113} have been developed. Downsides when using these methods arise for larger molecules, leading to computationally very demanding procedures, which are difficult to generalize. To overcome this problem, our group proposed to compute mass spectra using on-the-fly computed potential energy surfaces with Born–Oppenheimer *ab-initio* molecular dynamics (BO-AIMD).⁷⁴ Based on this idea, a widely applicable protocol for predicting mass spectra termed as the quantum chemistry electron ionization mass spectrometry (QCEIMS)^{50,105} method has been developed. It is, to our knowledge, the first attempt to use BO-AIMD to calculate EI-MS in a “close to the experiment” manner, without relying on any database or pretabulated fragmentation rules.⁵⁴ QCEIMS yields standard 70 eV EI-MS for organic and inorganic molecules that agree reasonably well with corresponding experimental data and gives an unprecedented insight into the reaction mechanisms. Different QC methods for the calculation of MS have already been tested in the past for various molecules.⁵⁰⁻⁵⁴ It has been shown that at least Kohn–Sham density functional theory (KS-DFT) with small basis sets or alternatively semiempirical quantum mechanical (SQM) methods like DFTB3^{114,115} or OM2/OM3^{116,117} have to be used to gain an acceptable accuracy-to-cost ratio. With this in mind, we have implemented the GFN1-xTB and IPEA-xTB methods into the program, and this combination has outperformed other methods for predicting EI mass spectra with QCEIMS.⁵⁶ Very recently, we have implemented its successor GFN2-xTB⁷¹ into the QCEIMS code. The improved physics of this method should increase the quality of the calculations, while the computational demands stay low.

A.2 Methodology

In an EI-MS experiment, a molecule is hit by a beam of high-kinetic-energy electrons. The impact of the accelerated electrons ejects a valence electron from the targeted molecule and, in positive ionization mode, creates a radical cation, as well as two out-going electrons with continuous energy, a

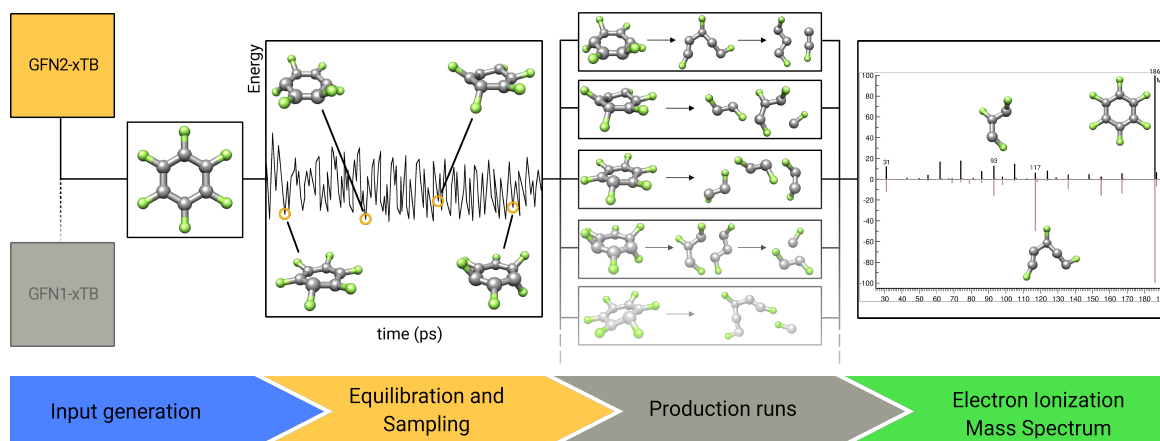


Figure A.1: Flowchart of the QCEIMS protocol.

so-called 1e-2e process. The deposited internal energy, if averaged over many molecules, is called the internal excess energy (IEE) and follows a complicated distribution for which we make reasonable assumptions.⁵⁰ Its value determines which reaction channels are eventually possible. After the electron impact, the IEE is distributed in the vibrational modes of the molecule through internal conversion (IC) and converted to the nuclear kinetic energy of the corresponding atoms. The vibration can cause bond breakage or other chemical reactions, which mostly lead to fragmentation into smaller molecules. The potential energy surface (PES) of the ions, which can be calculated, determines the most energetically favorable reaction pathways. For a more detailed description of these processes and a discussion of other important details that have to be considered in a theoretical EI-MS experiment, e.g., the question where the charge remains after the fragmentation process, the reader is referred to the original publication⁵⁰ and the textbooks.^{11,14}

A.2.1 QCEIMS

In the following section, we will briefly discuss the working principle of the program. For more details, the reader is referred to the original publication.⁵⁰

The prediction of an EI mass spectrum by QCEIMS proceeds in three steps (see Figure A.1):

1. **Equilibration and conformer sampling.** Equilibration and conformer sampling: An initial guess of the neutral molecular starting structure will be equilibrated in the first MD run, and a predefined number of snapshots are randomly selected and saved to obtain starting coordinates. For complicated cases, a preceding detailed conformational analysis should be conducted and the entire QCEIMS procedure is then started separately for various conformers.
2. **Assignment of IEE and IC.** For each snapshot geometry, the molecular orbital spectrum is calculated by a single-point calculation after which a Mulliken population analysis¹¹⁸ is performed. With this information, the internal excess energy (IEE) and internal conversion (IC) time are estimated and assigned to all starting geometries. The IC time is calculated by the energy-gap law.⁷⁹

- 3. Production runs.** The snapshot structures are instantaneously (valence) ionized and independently propagated in time on a QC PES until a reaction occurs in the simulation. The ionization potential (IP) of the so-created fragments is calculated and used to determine the statistical charge of these fragments. The fragment with the highest statistical charge is selected for further propagation in a cascade. It can again undergo fragmentation until either no internal energy is left or the fragment gets too small. All charged fragments are counted and stored. Taking together all production runs allows the program to compute the mass spectrum. The natural isotope ratios are implied in a postsimulation treatment.

The calculations done in the program are basically first-principles and fully theoretical, i.e., not based upon any experimental results. The EI-MS process is based upon a simplified theoretical model, and the PES is approximated by quantum chemical methods. Hence, the underlying QC method has a significant impact on the quality of the simulated spectrum. Furthermore, the number of production runs and the maximum simulation time can considerably alter the results. More subtle effects, like wrong assignments of ionization potentials to the fragments or the nature of the IEE distribution, can lead to false intensities or even signals from unphysical fragmentation (artifacts). A more detailed discussion of these influences can be found in the original publication.⁵⁰

When computed spectra of this unbiased approach used by QCEIMS are compared to those from rule-based, chemoinformatic programs (*e.g.* refs.^{35,45,47,109}), it is to be kept in mind that this “black-box” method may give rise to results with lower accuracy. However, QCEIMS can predict EI spectra of unknown chemical compounds and is able to retrace the composition of the fragments created during the process from the MD trajectory. For this reason, the program allows the discussion of the computed spectra in a detailed way, from fragmentation patterns to recombinations and rearrangements occurring in the experiment. At no point, intermediate structures have to be guessed or altered.

A.2.2 Extended Tight-Binding Methods

In 2017, a special-purpose SQM method called GFN-xTB⁷⁰ was published as a variant to the well-established tight-binding DFTB3^{114,115} scheme. Recently, our group has developed a second variant, termed GFN2-xTB, (30) that includes anisotropic second-order density fluctuation effects via short-range damped interactions of cumulative atomic multipole moments. Both extended tight binding (xTB) methods are designed to account for properties around the energetic minimum, such as geometries, vibrational frequencies and non-covalent (GFN) interactions and are parametrized for elements with atomic numbers up to $Z = 86$. Interestingly, the methods show an overall good performance and great robustness also for electronically complicated situations, including covalent bond breaking.

In GFN2-xTB, the dispersion interactions are treated by means of a self-consistent variant of the D4 dispersion model⁶⁸ instead of D3(BJ),⁶⁵⁻⁶⁷ used in GFN-xTB, and furthermore, the description of electrostatic interactions has been greatly improved. For a more detailed discussion of the differences between both methods, please refer to ref. [71].

The originally proposed method is from here on called GFN1-xTB for a better distinction between both schemes.

The computed ionization potentials with GFN1-xTB are not sufficiently accurate. To remedy this, a special-purpose IPEA-xTB method has been developed. It is a reparametrization of GFN1-xTB and uses additional $(n + 1)$ s basis functions to better represent the electron affinities. Unfortunately, the

spectra of some molecules containing transition metals were not described well with the IPEA-xTB method, which partially could be traced back to wrong charge assignments. The errors made on these systems can be fixed by calculating the ionization potentials at the hybrid DFT (PBE0¹⁰¹/def2-SV(P)¹¹⁹) level, for which the computational time is proportionally large in comparison to semiempirical methods. For GFN2-xTB, we did not find the need for a reparametrization because the IPs are being calculated qualitatively correctly with this method and are sufficiently accurate for the charge assignment of the fragments; thus, it is likely that use of this new method will achieve the proper results for these systems with less computational effort. To validate this statement and to test the performance of this method, we have tested combinations of GFN1- and GFN2-xTB using IPEA- and GFN2-xTB for IP calculations. To present the performance of the two methods more clearly, semiempirical OM3-D3 calculations were performed for organic molecules. AM1 and PM3 calculations were omitted because of the bad performance of these methods with QCEIMS (see ref. [50]). DFT was used to cross-check molecules involving transition metal atoms. The combinations are noted as *Method/IP-Method*.

A.2.3 Technical Details

The calculations in this work were executed on Intel[®] XEON[®] E5-2660 2.00 GHz cores. Computations were performed using QCEIMS version 3.8. For OM3-D3 calculations, MNDO2005 version 7.0¹²⁰ was used and DFT calculations were gained using the ORCA 4.0.1.2. suite of programs.^{121–123} Statistically converged results are obtained for 1000 production trajectories that were carried out for each molecule and method with a maximum MD simulation time of 10 ps. Each production run required about 10 000 – 20 000 QC calls. We did not alter any settings in QCEIMS, nor did we modify the tight-binding methods for this work. The results presented here therefore do not show the full capability of QCEIMS, which may be improved by choosing different simulation conditions in the program for different compound classes.

A.3 Results and Discussion

A.3.1 Benchmark Set

To test the performance of the methods, test molecules for benchmarking were chosen, which vary in structure, size, and chemical functionality. They are designed to involve commonly known molecules and inherit various elements across the periodic table and were inspired by our previous work using GFN1-xTB for EI-MS calculations.⁵⁶ In Figure A.2, we display a selection of 15 different molecules, sorted into three groups:

The organic molecule group includes 1-butanol (**1**), hexafluorobenzene (**2**), uracil (**3**), testosterone (**4**), sucrose (**5**), and leucylglycylglycine (**6**).

The transition-metal group includes bis(benzene)chromium (**7**), zirconocene dichloride (**8**), and nickel(II)bis(diphenyl-acetylacetonate) (**9**).

The main-group inorganic group contains bis(pinacolato)diboron (**10**), chinalphos (**11**), triphenylstibine (**12**), dichloro(ethyl)aluminum (**13**), 2-(dimethyl-(naphthalen-1-yl)silyl)-phenyl)methanol (HOMSi, (**14**)), and octasulfur (**15**).

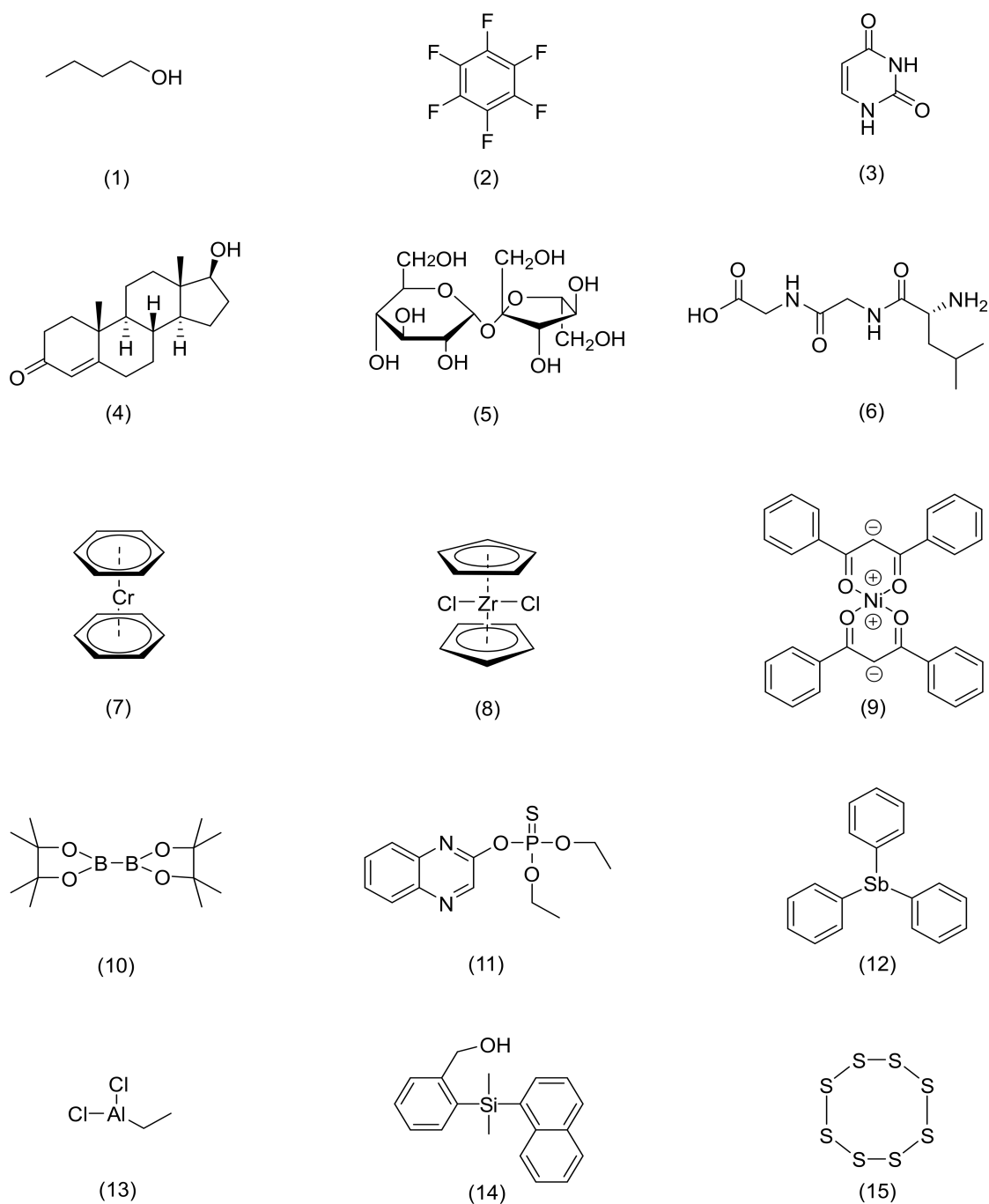


Figure A.2: The benchmark set. Molecules (1) – (6) represent the organic group, molecules (7) – (9) represent the transition-metal group and molecules (10) – (15) represent the main-group inorganic group.

A.3.2 Timings

The computation time of all production runs is summed up to gain the total time for creating a full simulated spectrum. Because the trajectories run independently, the wall timings in actual projects can be reduced by parallel runs, i.e., they can be divided by the number of available computer cores. Outliers will be discussed in Section A.3.4.

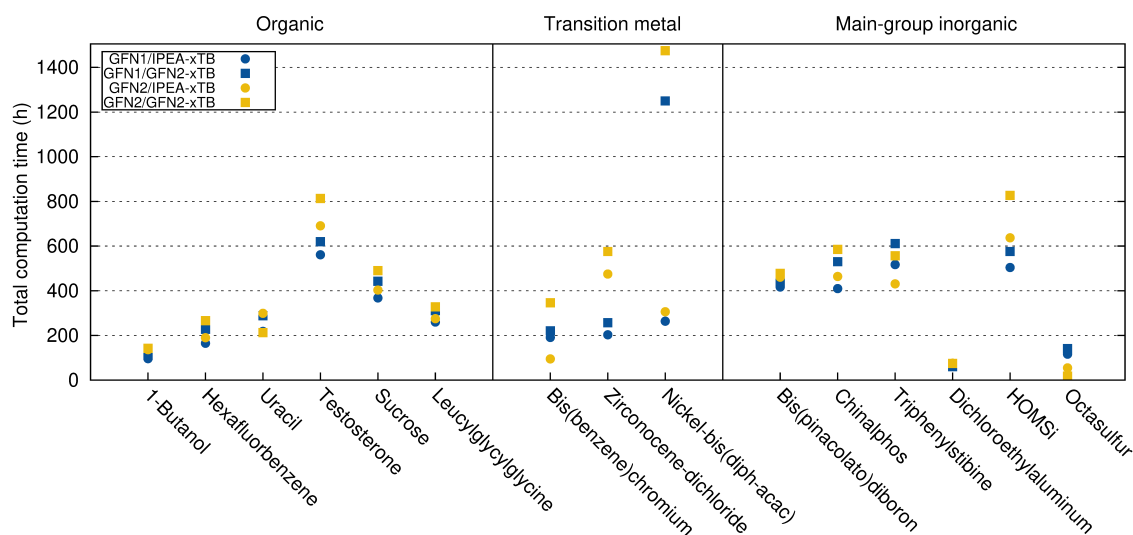


Figure A.3: Total calculation time of all test-set molecules in hours. The molecules are grouped in their corresponding categories. GFN1-xTB results are shown in blue, and GFN2-xTB results are shown in yellow.

For small organic molecules (**1-3**), the total calculation time for a single spectrum (see Figure A.3) lies between 100 and 250 h for all combinations of GFN methods. For testosterone (**4**), computation times increase from 600 h with GFN1/IPEA-xTB to 700 h with GFN2/IPEA-xTB and up to 800 h with GFN2/GFN2-xTB. The total calculation times of the organic molecules **5** and **6** are between 300 and 400 h. For comparison, the organic molecules were calculated with the semiempirical OM3-D3 method. The computational demands for these calculations are in the range of the GFN calculations, with slightly shorter running times for molecules **1**, **5**, and **6** and slightly longer running times for molecules **2**, **3**, and **4**. These results can be found in the Supporting Information (SI) (see Section A.5).

For transition-metal-containing molecules (**7-9**), calculations of a total spectrum using different GFN1-xTB combinations take on average 200 h to complete, while the calculations with GFN2-xTB for molecules **7** and **8** take between 100 and 550 h to complete. Using GFN2-xTB for ionization potential calculations, the timings increase. For molecule **9**, both GFN methods take more than 1200 h for the complete calculation of the spectra. Use of DFT for this system increases the consumed time by a factor of 4 so that the overall timing increases up to 5000 computational hours. For a better overview in the figures, the DFT results are omitted in Figures A.3 and A.4. These results can be found in the SI (see Section A.5).

The time per calculation with GFN1/IPEA-xTB for the main-group inorganic molecule group (**10-15**) averages for all four combinations of methods between 400 and 600 h, where use of GFN2-xTB for IP evaluation takes longest. This excludes the outlier dichloro-(ethyl)aluminum **13**, for which the timing is in the range of the small organic molecules (e.g. 1-butanol (**1**)). For the molecule **15**, timings

with the GFN2-xTB combinations are lower than expected, but this is due to technical failures in the calculations, as discussed in the corresponding results Section A.3.4.

A.3.3 Stability

The stability of the GFN1- and GFN2-xTB methods is evaluated by the number of successful production runs (see Figure A.4). For both methods, the majority of runs complete properly, leading to an excellent average failure rate of less than 1%. For the GFN1/IPEA-xTB calculations, the largest failure rates are produced for hexafluorobenzene (**2**, 3.6%), bis(benzene)chromium (**7**, 3.8%), and octasulfur (**15**, 4.9%). For the GFN2-xTB method, failure rates are comparable, except for the transition metal involving molecule bis(benzene)-chromium (**7**), having a failure rate of 11.2%, and octasulfur (**15**) (>15%). The effect on the calculated spectrum is discussed in the corresponding results, Section A.3.4.

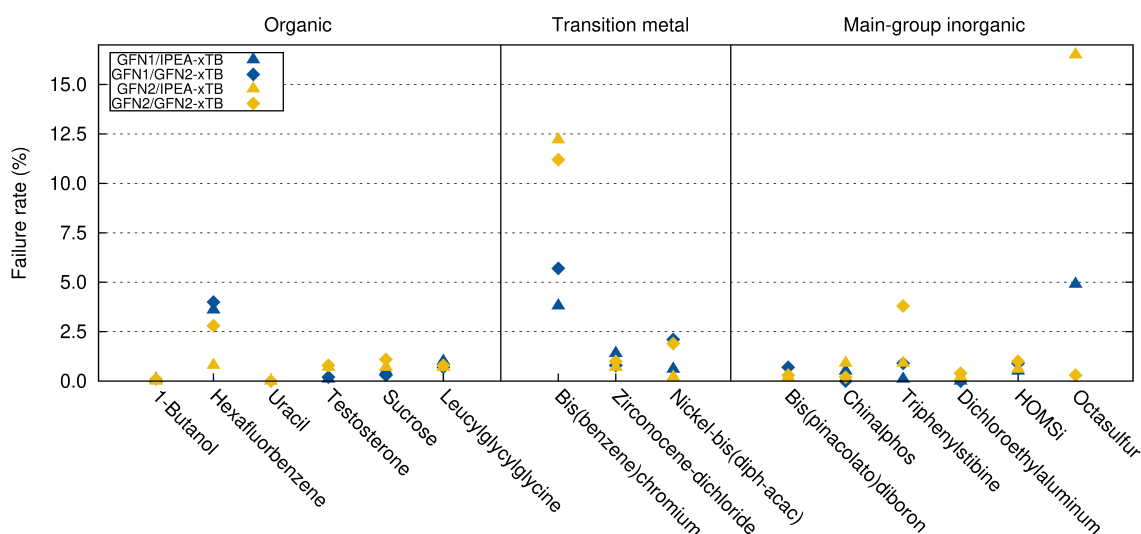


Figure A.4: Average failure percentage of all test-set molecules. The molecules are grouped in their corresponding categories. GFN1-xTB results are shown in blue, and GFN2-xTB results are shown in yellow.

In conclusion, low failure rates and acceptable timings indicate good applicability of the new GFN2-xTB method. This could not be expected because GFN2-xTB is inherently more involved mainly due to the additional multipole electrostatic treatment. Thus, GFN2/IPEA-xTB and GFN2/GFN2-xTB both can be used in QCEIMS as an alternative to the GFN1/IPEA-xTB method without significant restrictions in the computational demands or robustness.

A.3.4 Comparison of Experimental and Theoretical Spectra

In the following section, we present the calculated QCEIMS spectra using GFN1/IPEA-xTB and GFN2/GFN2-xTB. The computed spectra are directly compared with their corresponding experimental EI-MS obtained from the NIST⁴⁰ or SDBS⁹⁰ databases, and the agreement between theory and experiment is determined by a composite matching score^{51,111} with a range of values between 0 (no match) and 1000 (perfect match). Matching scores of organic molecules obtained with OM3-D3 are being listed for the purpose of validation of the quality of spectra gained using the GFN methods. The

main differences between the results of the two GFN methods are being discussed, with a focus on the presence of important m/z signals as well as corresponding variation in signal intensity. We discuss determinative peak-series and point out interesting or important structures of the obtained fragments, which are shown explicitly as insets in the spectra. Major differences in signals or intensities between GFN2/IPEA- and GFN2/GFN2-xTB calculations were not observed or were of minor influence for the resulting spectra and are therefore not discussed further.

Organic Group

Small organic molecules have already been investigated in former studies^{50,54,56} and are only briefly considered here.

For 1-butanol (Figure A.5a), the main fragmentation pathways result from the loss of alkyl groups. These are reproduced well by both methods with a satisfying agreement of the simulation with the experiment. However, the survival rate of the precursor ion is too high in the simulations, meaning it does not decompose accordingly under given simulation conditions. This effect is due to the IEE distribution, which is of a Poisson-type variant and not obtained *ab-initio* and thus can lead to a bad description of unusual electronic situations. This can partially be alleviated by applying higher IEE values and/or longer simulation times. These and other effects of various simulation conditions are discussed in the original publication.⁵⁰ Differences between the methods are found in the intensities of the signals. For GFN2-xTB, especially the peaks of $C_3H_3^+$ at m/z 39 and of $C_4H_9^+$ at m/z 56 are in better agreement with the experiment. Through the high survival rate of the precursor ion, the matching scores between the experiment and calculations are 225 for GFN1-xTB and 223 for GFN2-xTB. OM3-D3 calculations account better for the precursor ion signal, which leads to a matching score of 530.

Hexafluorobenzene (Figure A.5b) represents an interesting case. The fragment pattern is dominated by the ring breakage products. In the experiment, the dominant path forms $C_3F_3^+$ (m/z 117) with the remainder being CF_2 (m/z 50) and a single fluorine atom (m/z 19). The loss of a single (neutral) fluorine atom from the precursor ion can be observed by the appearance of the signal at m/z 167. Furthermore, the breakage of the precursor into two units of $C_3F_3^+$ yields the signal at m/z 93. Both GFN methods provide similarly good results in this respect. However, the methods differ in the calculated intensities, as GFN2-xTB overestimates some of the signals, especially for the fragments $C_6F_4^+$ at m/z 148, $C_6F_2^+$ at m/z 110, and $C_4F_2^+$ at m/z 86. Matching scores are 734 for GFN1-xTB, 647 for GFN2-xTB, and 687 for OM3-D3.

An earlier QCEIMS work has been conducted for four different nucleobases.⁵³ These were calculated at the semiempirical levels OM2-D3^{116,117} and DFTB3-D3.^{114,115} As an example molecule from this series, we present uracil (Figure A.5c), computed with the two GFN methods. We find that the spectra produced with both GFN methods are in good agreement with the experiment, although some intensities of the calculated signals are either over- or underestimated when directly compared to the measured signals. Compared to the results obtained in the earlier work, the spectra calculated by the GFN methods are of better quality as the spectra are produced using DFTB3-D3 or OM2-D3. The fragmentation proceeds via the bond breakage of the precursor ion into units of $HCNO^+$ (m/z 43) and $HNC_3H_2O^+$ (m/z 69) and the subsequent dissociation into the fragments $HNCH^+$ and CO^+ at m/z 28. Between the calculated spectra of the two GFN methods, only minor differences in the intensities are observed. Matching scores are 780 for GFN1-xTB, 745 for GFN2-xTB, and 711 for OM3-D3.

The experimental spectrum of testosterone (Figure A.6a) contains a large number of signals, which

Appendix A Calculation of Electron Ionization Mass Spectra with Semiempirical GFNn-xTB Methods

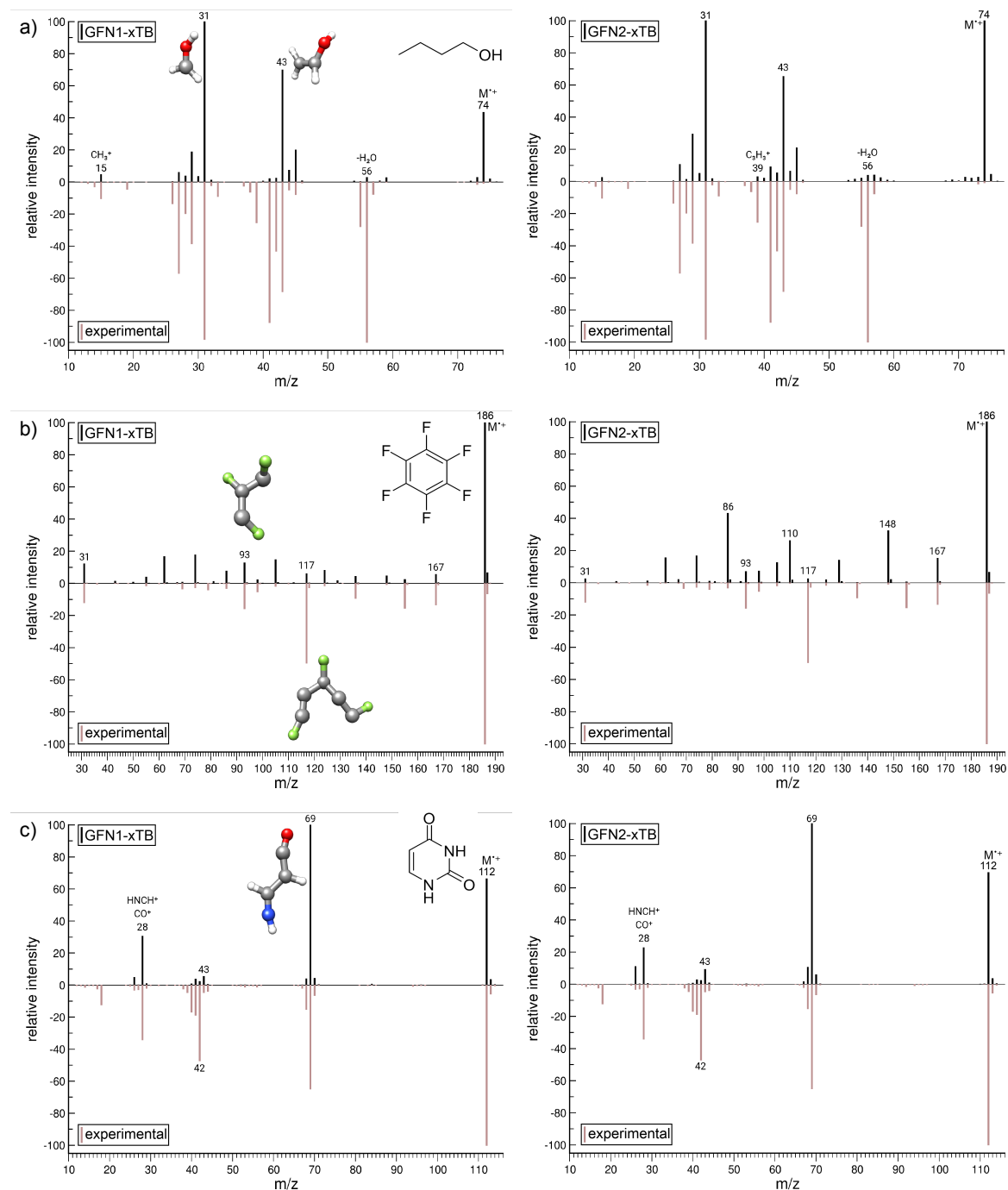


Figure A.5: Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the organic compounds (a) 1-butanol, (b) hexafluorobenzene, and (c) uracil. The structures of the precursor ion (denoted $M^{+\bullet}$) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.

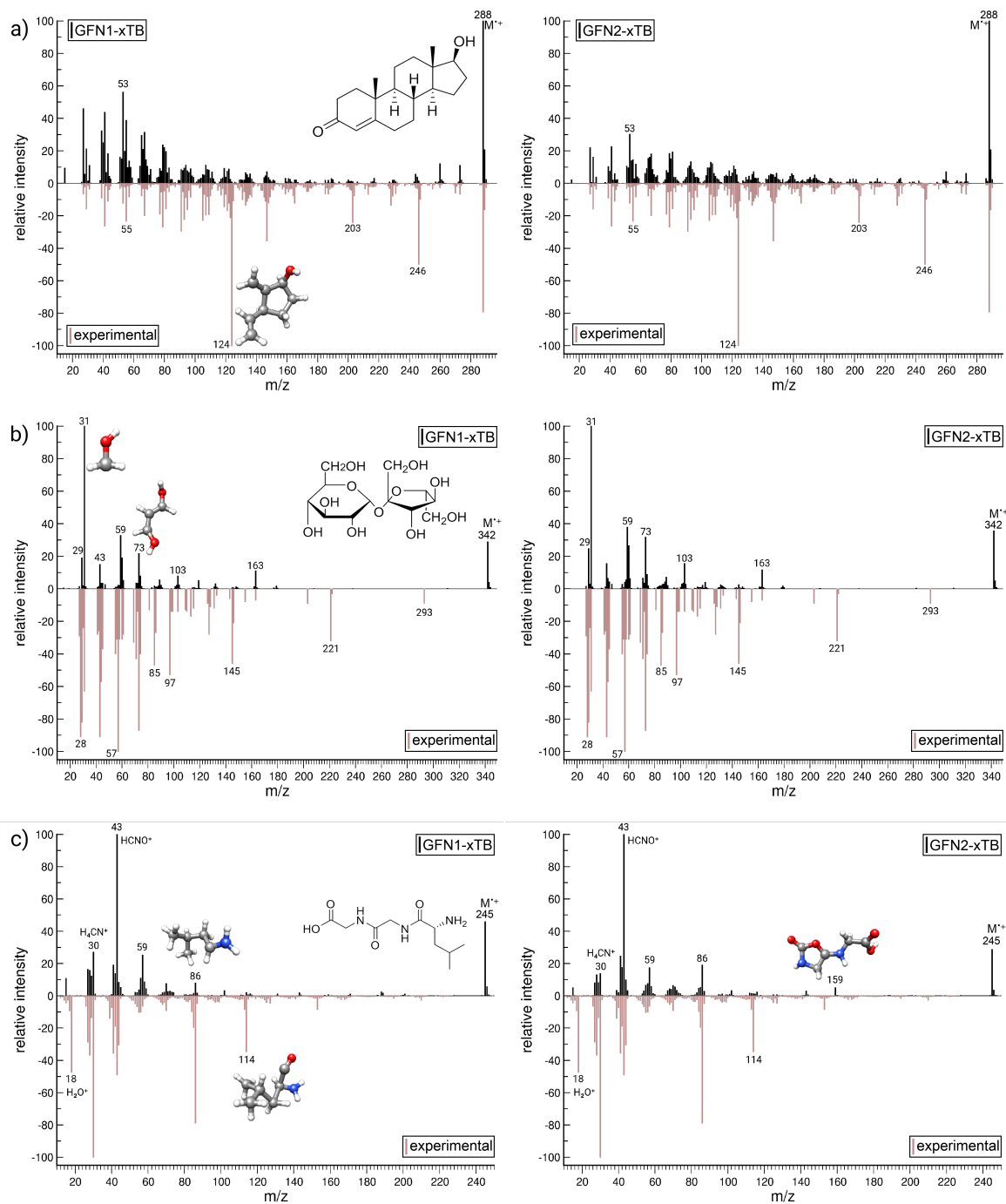


Figure A.6: Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the organic compounds (a) testosterone, (b) sucrose, and (c) leucylglycylglycine. The structures of the precursor ion (denoted $M^{\bullet+}$) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.

are overall reproduced well by the two GFN methods. Especially, the peak series of the lower and medium mass fragments (m/z 30–150) are replicated very accurately by both tight-binding methods when compared with the experiment. However, neither of the two computed spectra correctly reproduce the intense signals displayed in the experimental spectrum at m/z values 124, 203, and 246. Furthermore, the GFN-based calculations favor the formation of the fragment $C_4H_5^+$ at m/z 53 over the experimental found fragment $C_4H_7^+$ at m/z 55. Comparing the results produced by the two GFN methods with each other, the spectrum simulated with GFN1-xTB overestimates the intensities of the lower mass signals in the area of about m/z 30–90, while the spectrum calculated with the GFN2-xTB method is in overall better agreement with the experimental results and also accounts better for signals in the area between m/z 145 and 200. Matching scores are 532 for GFN1-xTB, 545 for GFN2-xTB, and 592 for OM3-D3.

The fragmentation scheme of sucrose (see Figure A.6b) produces a considerable amount of highly intense signals in the experimental spectrum, notably in the area between m/z values 28 and 73. Large discrepancies between simulations and experiment can be observed in the number of hydrogen atoms bound to the fragments. Especially, the peaks from the experiment at m/z 28, belonging to CO^+ , compared to the calculated signal at m/z 29 of HCO^+ , as well as the experimental signal of $HC_2O_2^+$ at m/z 57 in contrast to the simulated peak of $H_3C_2O_2^+$ at m/z 59, are typical examples for this divergence. Furthermore, various signals that can be noticed in the experimental spectrum, e.g., m/z values 97, 221, and 293, and some of the less intense peaks in between, are not well recreated by the calculations using the GFN methods. A comparison of the simulated spectra of the two tight-binding methods with each other reveals no significant differences. Matching scores are 201 for GFN1-xTB, 217 for GFN2-xTB, and 169 for OM3-D3.

Leucylglycylglycine (see A.6 c) is composed of one l-leucine and two glycine residues. In the EI-MS experiment, the main decomposition pathway leads to the fragments $H_6C_3N_2O^+$ (m/z 86), $H_3C_2O_2^+$ (m/z 59), $H_9C_4^+$ (m/z 57), and $HCNO^+$ (m/z 43). In a second step, $H_6C_3N_2O^+$ dissolves into $H_2C_2NO^+$ (m/z 56) and H_4CN^+ (m/z 30). The simulated spectra produced using either GFN1-xTB or GFN2-xTB account for the majority of these signals in accordance with the experiment. However, in the simulations, the survival rate of the precursor ion is too high so that less intense signals measured in the experiment do not appear in the calculated spectra. When comparing the GFN methods with each other, some of the intensities of various signals differ, e.g., peaks at m/z values 59, 86, and 159. The latter signal belongs to an intermediate product that gains stability through a ring formation of the peptide. The signal of H_2O^+ at m/z 18 in the experimental spectrum is probably measured due to the presence of water in the ionization chamber during the experiment and is therefore not produced by the simulations. Matching scores are 192 for GFN1-xTB, 224 for GFN2-xTB, and 123 for OM3-D3.

Transition-Metal Group

For transition metal molecules, the ferrocene system has to be discussed explicitly. The IPEA-xTB method yields wrong IP values and therefore leads to false signals in the calculated EI-MS spectrum. This is visualized by the red circle in Figure A.7, where the ionized iron cation is not being charged in the simulated spectrum (red, inverted) and thus the experimentally found signal of Fe^+ is missing. Calculating the IPs with DFT or GFN2-xTB instead solves this problem. It is therefore recommended to use the GFN1/GFN2-xTB, GFN2/GFN2-xTB or the GFN1/DFT, GFN2/DFT combination for transition-metal-containing molecules. The results presented in this section for transition-metal-containing molecules were calculated with the GFN1/GFN2-xTB and

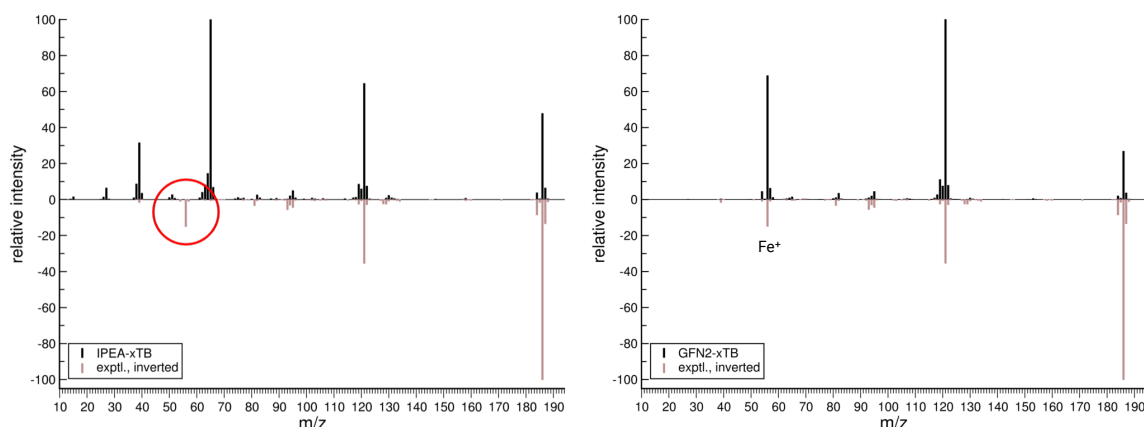


Figure A.7: The EI-MS of ferrocene as an example of the impact of falsely calculated ionization potentials. The red circle marks the signal of Fe^+ that is not being reproduced by GFN1/IPEA-xTB (left). The usage of GFN1/GFN2-xTB (right) accounts for the correct signal.

GFN2/GFN2-xTB combinations.

For bis(benzene)chromium (Figure A.8a), both GFN simulated spectra are in satisfying agreement with the experimental spectrum. A detailed comparison between the calculated and the experimental spectra reveals some divergences in the intensities, e.g., the simulated signal at m/z 77 and the experimental peak at m/z 78, resulting from a discrepancy between the simulated and the measured number of hydrogen atoms bound to fragmented benzene rings. Failure rates by both methods of the GFN-xTB family were high (see Section A.3.3), especially for calculations with GFN2-xTB. This is most likely due to the wrong description of the Cr^+ fragment, for which the corresponding signal at m/z 52 is almost completely missing in the calculated spectrum of this method. Instead, the protonated form HCr^+ is preferred and simultaneously the simulation fails to account for signals with low mass-to-charge values between 20 and 40. Both calculations conducted with the GFN methods account for the signal at m/z 104, which is created by a fragment in which a H_2Cr molecule is bound to a C_4H_2 chain. Matching scores are 677 for GFN1-xTB and 688 for GFN2-xTB.

The Kaminsky catalyst¹²⁴ can contain the group 4 metal components Ti, Zr, or Hf. We have decided to test the two GFN methods for EI-MS calculations of zirconocene dichloride (Figure A.8b). The experimental spectrum reveals a variety of fragmentation pathways, from single chloride loss of the precursor ion (resulting in a signal at m/z 256) to the cyclopentadienyl dissociation (signals at m/z values 162 and 227). The EI-MS created using the GFN1-xTB method is in very good agreement with the experimental one. In contrast, the spectrum created using the GFN2-xTB method shows an overestimation of almost all intensities of the produced signals in comparison with those from the experiment. Furthermore, the new approach creates various artifacts that lead to a significant increase in the computational demands for the calculation of this system (see Section A.3.2). However, the signals computed by GFN2-xTB around m/z 201 are in surprisingly good accordance with the experiment, which corresponds to the subsequent loss of C_2H_2 from the cyclopentadienyl fragment (m/z 227). These signals are not simulated well by GFN1-xTB. Matching scores are 671 for GFN1-xTB and 680 for GFN2-xTB.

Quantum chemical calculations of the mass spectrum of nickel(II)bis(diphenyl-acetyl-acetonate)

Appendix A Calculation of Electron Ionization Mass Spectra with Semiempirical GFNn-xTB Methods

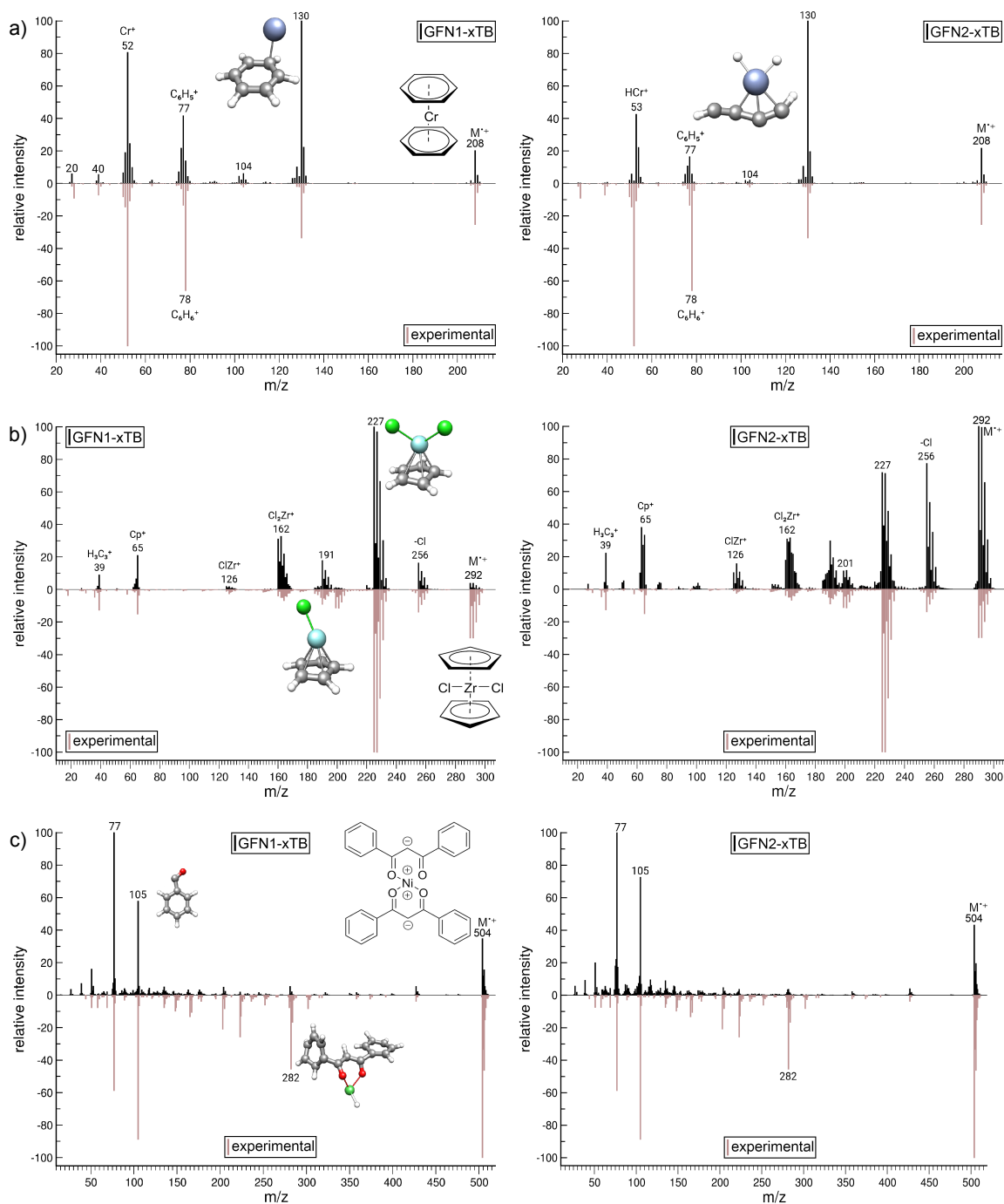


Figure A.8: Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the transition metal compounds (a) bis(benzene)chromium, (b) zirconocene dichloride, and (c) nickel(II)bis(diphenyl-acetylacetonate). The structures of the precursor ion (denoted M^{*+}) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.

(Figure A.8c) have already been conducted in an earlier publication, for which the ionization potentials have been determined using a hybrid DFT method (PBE0/def2-SV(P)).⁵⁶ In this work, we have used the IPEA- and GFN2-xTB methods for IP calculations instead, which show great robustness and low computational demands, as demonstrated in Sections A.3.2 and A.3.3. The simulated spectra created using GFN1- and GFN2-xTB are in good agreement with the experimental ones, regardless of which method was used to obtain the ionization potentials, including DFT. The experimental main peaks are being reproduced accordingly, and the intensities for the dehydrogenated phenol (m/z 105) and benzene (m/z 77) displayed in the experiment are reconstructed well by both methods. The simulations fail to recreate the peak intensities as measured in the experiment for the signal at m/z 282 and the precursor ion m/z 504. Some artifacts are found in both simulated spectra, although we note a somewhat better performance of GFN2-xTB compared to its predecessor. Matching scores are 805 for GFN1-xTB and 830 for GFN2-xTB.

Main-Group Inorganic Group

Bis(pinacolato)diboron (Figure A.9a) is an interesting compound related to Suzuki coupling reactions.¹²⁵ In the EI-MS simulations, the fragmentation patterns can be retraced convincingly, as indicated by the good agreement between experimental and calculated spectra. The main fragmentation pathway is characterized by the cleavage of CH_3^+ and C_3H_6^+ from the alkane groups of the molecule. These fragments themselves emerge in the lower mass area of the spectrum at m/z values 15 and 42 with various constellations in the number of hydrogen atoms bound to these fragments. This fragmentation is followed by various rearrangement reactions: In the first step, one of the C_6H_{12} (m/z 84) side chains splits off from the molecule, which leads to a rearrangement of the remaining fragment, creating a $(\text{O}_2\text{-})\text{B-O-B(-O)}$ chain (m/z 169). In the second rearrangement step, the remote O-B-O unit substitutes an oxygen atom at one of the side chains, which leads to the decomposition of the structure into two units of $\text{O-B-O-C-(CH}_3)_2^+$, displayed by the signal at m/z 84. The detailed description of this process shows the outstanding capabilities of QCEIMS to analyze rearrangement procedures during EI-MS experiments, demonstrating its usefulness in elucidating structures and complex reaction mechanisms. Overall, we find the agreement between calculated spectra using the GFN methods and the experimentally measured spectrum to be very satisfying, as the two GFN methods account for the majority of signals. GFN2-xTB seems to perform better here since the agreement with the intensities measured in the experiment is slightly better than in the spectrum created with GFN1-xTB. Both methods overestimate the signals at m/z values 41 and 42, as well as the survival rate of the precursor ion. Matching scores are 245 for GFN1-xTB and 250 for GFN2-xTB.

The simulated MS of the insecticide chinalphos (Figure A.9b) matches the experimental spectrum badly. The computed signal at m/z 145 is generated by the fragmentation of the precursor ion into a $\text{SP}(\text{OC}_2\text{H}_5)_2^+$ fragment, which differs from the experimental signal at m/z 146 due to a hydrogen atom bonded less to the fragment. The signal at m/z 157, which is almost totally missing in the simulation, is generated due to the rearrangement of a CH_3 fragment (which dissociates from the $\text{SP}(\text{OC}_2\text{H}_5)_2^+$ group) between the benzene ring and one of the nitrogen atoms. This rearrangement is rarely reproduced in the simulation since the required migration of the methane group can progress into all spatial directions and is not often aligned at the ring system. In contrast to its predecessor, the simulation with the GFN2-xTB method displays a higher probability that in a first step the sulfur atom breaks from the precursor ion, which resolves in an intense fragment signal at m/z 266. This is followed by the decomposition of this fragment into the $\text{PS}(\text{OC}_2\text{H}_5)_2^+$ fragment, generating the signal

Appendix A Calculation of Electron Ionization Mass Spectra with Semiempirical GFNn-xTB Methods

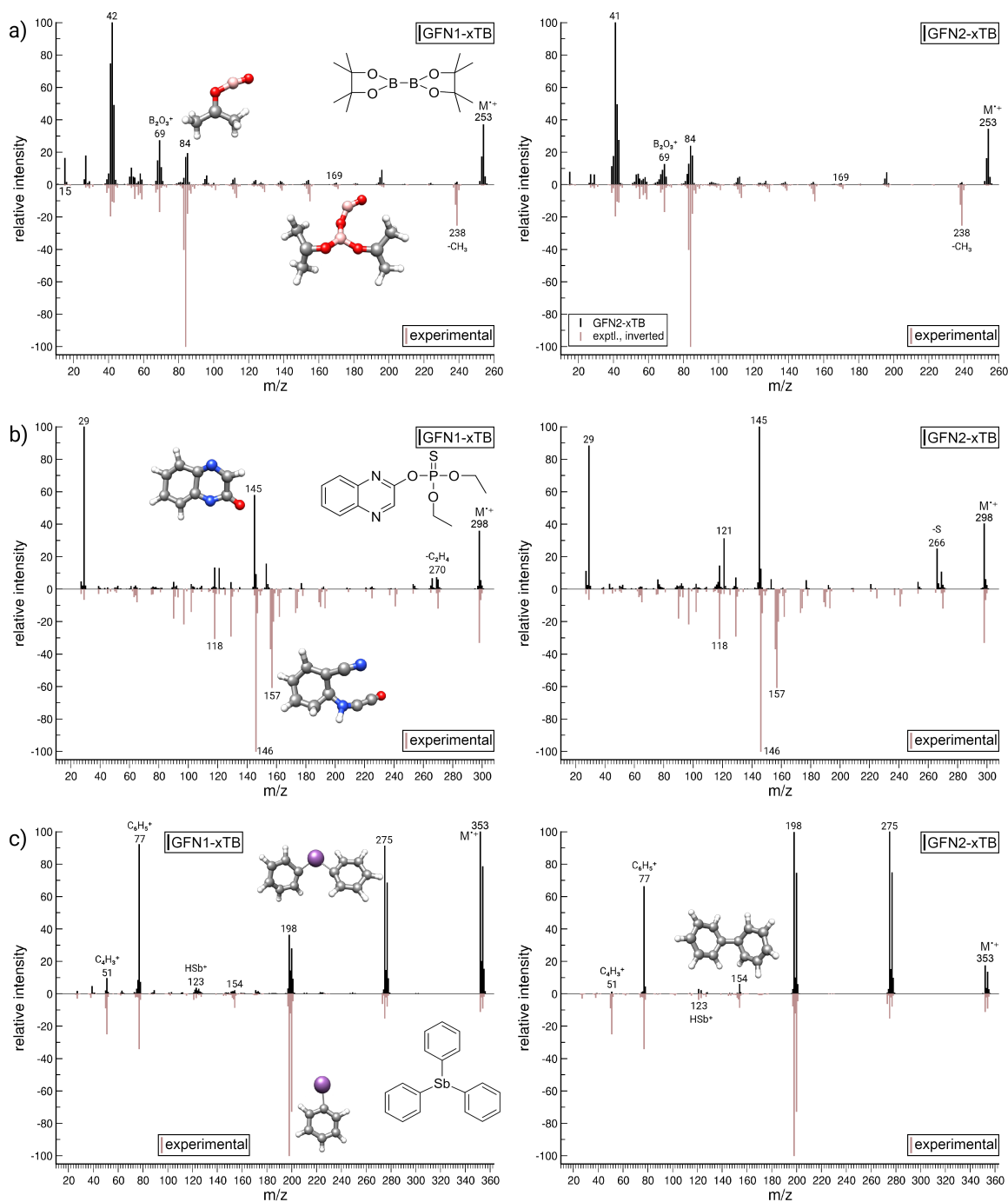


Figure A.9: Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the main-group inorganic compounds (a) bis(pinacolato)diboron, (b) chinalphos, and (c) triphenylstibine. The structures of the precursor ion (denoted M^+) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.

at m/z 121. Matching scores are 598 for GFN1-xTB and 510 for GFN2-xTB.

In the spectrum of triphenylstibine (Figure A.9c), the dominant fragmentation pattern corresponds to the cleavage of the phenyl groups and their subsequent dissociation, which has already been reported in an earlier publication using GFN1-xTB for structure elucidation.⁵⁶ We find that the simulated spectra using both GFN methods are in good agreement with the experimental data, but simulations using the GFN1-xTB method yield small artifacts throughout the spectrum. In contrast, the spectrum created using GFN2-xTB does not display any artifacts, and the method reduces the survival rate of the precursor ion to match the experiment almost perfectly. Furthermore, use of GFN2-xTB improves the intensities of the signals at m/z values 198 and 154 when compared to those from the experiment, where the latter signal describes the bond formation between two phenyl groups. The agreement between the simulated and the experimental signal of HSb^+ at m/z 123 is better when using the GFN2-xTB method instead of GFN1-xTB. Matching scores are 642 for GFN1-xTB and 664 for GFN2-xTB.

The spectrum of dichloro(ethyl)aluminum (Figure A.10a) indicates competing fragmentation reactions. The signals at m/z values 29 and 97 describe the loss of the ethane group, forming Cl_2Al^+ and C_2H_5^+ , while the signals at m/z values 36 and 91 are generated by the loss of the chlorine atoms. The simulated spectrum of the GFN2-xTB method nearly matches all of these findings and recreates the intensities of the signals in the experimental spectrum in better agreement than its predecessor GFN1-xTB. The substitution of the carbon–aluminum bond by a single chlorine is reproduced according to the experiment, generating the signal at m/z 64 in the GFN2-xTB-created spectrum. Matching scores are 514 for GFN1-xTB and 507 for GFN2-xTB.

In Hiyama cross-coupling reactions, so-called HOMSi reagents¹²⁶ are used for carbon–carbon coupling reactions. As an example for this group of reagents, we have chosen to compare the simulated spectra created with GFN1- and GFN2-xTB to the experimental spectrum of (2-(dimethyl(naphthalen-1-yl)silyl)phenyl)methanol (Figure A.10b). The EI-MS spectrum measured in the experiment contains highly intense peaks in the larger mass region between m/z values 215 and 277 and less intense signals for small to medium mass regions with m/z values 20–200. The calculated spectra of the two GFN methods show a good agreement to the experiment in the lower and medium mass regions, but the main peaks at m/z values 215 and 259 and the survival rate of the precursor ion are displayed poorly with both methods compared to experiment. In the simulation, first, a rearrangement of the hydroxide molecule from the benzyl alcohol group ($\text{C}_7\text{H}_7\text{OH}$) to the silicon atom takes place. The hydroxide molecule substitutes one of the ethane groups bound to the silicon atom, and the resulting structure creates a signal at m/z 277. From here on, the simulated fragmentation patterns do not seem to correspond to the bond breakage scheme of the experiment. While in simulation one methane group is cleaved off (forming the signal at m/z 261), in the experiment, a H_2 molecule is additionally dissociated during this fragmentation process and the resulting fragment creates the signal at m/z 259. Subsequently, this fragment either decomposes into HOSi^+ , giving rise to the signal at m/z 45 and which has a dominating intensity in the simulated spectrum using the GFN1-xTB method, or, in contrast, forms the CH_3Si^+ fragment displayed at m/z 43, which is favored in the spectrum computed with the GFN2-xTB method. Both of these fragment signals are displayed in the experimental spectrum. Further differences between the simulated spectra of the GFN methods are visible in the overall signal intensity, as the lesser intense peaks created by the GFN2-xTB method produce a “cleaner” spectrum. However, some of the experimentally found signals are under-represented in the spectrum at this level. Matching scores are 315 for GFN1-xTB and 282 for GFN2-xTB.

The simulated spectra of octasulfur (Figure A.10c) strongly depend on the method used. Calculations

Appendix A Calculation of Electron Ionization Mass Spectra with Semiempirical GFNn-xTB Methods

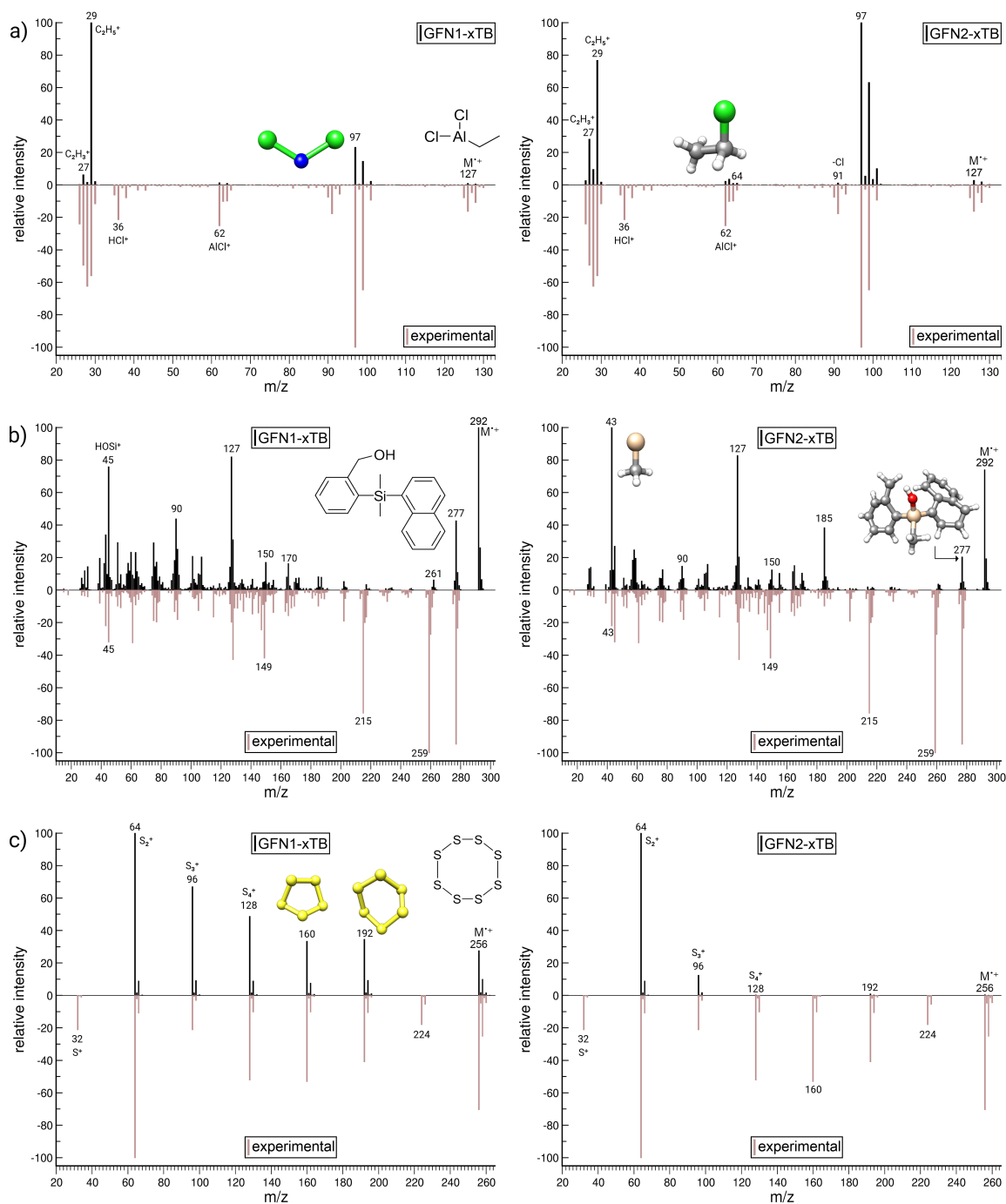


Figure A.10: Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the organic compounds (a) dichloro(ethyl)aluminum, (b) 2-(dimethyl(naphthalen-1-yl)silyl)phenylmethanol, and (c) octasulfur. The structures of the precursor ion (denoted $M^{+\bullet}$) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.

with the GFN2-xTB method result in a spectrum in which nearly all signals are missing. The majority of production runs show the immediate bond breakage of the sulfur ring into S_2^+ and S_3^+ fragments. Hereby, the intermediate products at m/z values 128 and 192 are strongly under-represented and the signals at m/z values 32, 160, and 224 do not appear. On the contrary, the spectrum simulated with GFN1-xTB is in very good agreement with the experimental results and reproduces all signals and their intensities satisfactorily, except for those signals corresponding to S_7^+ at m/z 224 and S^+ at m/z 32. Matching scores are 821 for GFN1-xTB and 213 for GFN2-xTB.

A.4 Conclusions

We have implemented the new semiempirical, special-purpose method GFN2-xTB into the QCEIMS program. We have tested its ability for the calculation of EI mass spectra and realistic fragmentation patterns and compared the results to spectra produced by using its predecessor GFN1-xTB. For an unbiased evaluation, neither the methods were modified for this purpose nor did we adjust any parameters or settings in the QCEIMS program to influence its behavior. For validation, we have compared the computed spectra of the two methods to experimental results gained from the NIST and SDBS databases. A wide variety of smaller and larger organic, transition-metal, and main-group inorganic systems have been studied. The computational demands and the stability of various method combinations also for the ionization potential calculations have been discussed. Most of the calculations were conducted using the GFN1/IPEA-xTB or the GFN2/GFN2-xTB combination, which were found to be robust and accurate.

In previous work, it has been shown that the performance of the GFN1-xTB method excels other quantum chemical methods in accuracy and low computational costs for the calculation of EI mass spectra.⁵⁶ One main point of this work was the question if this can even be improved by the new GFN2-xTB method featuring a better underlying quantum mechanical description.

For organic molecules, both methods produce qualitatively good results in comparison to the experimental spectra. However, GFN2-xTB does not generally improve the calculated spectra, e.g., for hexafluorobenzene (**2**), GFN1-xTB performs better, while the opposite holds for testosterone. Notably, both methods yield somewhat less accurate spectra for sucrose (**5**) than for other organic molecules.

For the calculations of the ionization potentials of transition-metal-containing molecules, GFN2-xTB can be recommended. The best overall agreement with the experimental spectra is obtained with the GFN1/GFN2-xTB combination. Failure rates and computational demands with GFN2/GFN2-xTB were higher, and the combination failed to predict important signals, e.g., the missing Cr^+ ion in the bis(benzene)chromium (**7**) fragmentation. For critical cases, it is advised to cross-check the results with a DFT IP calculation. This can improve the quality of the obtained spectra but is computationally very demanding.

Regarding timings, stability, and the overall agreement with the experiment, the two GFN methods perform similarly for the inorganic main-group molecules. It is recommended to apply both GFN methods for such systems, as it is not clear which of the methods will deliver the better result in the end. Up to this point, the only molecule for which the new GFN2-xTB method completely fails is octasulfur (**15**), where the GFN1-xTB method performs very well.

Differences between the GFN1-xTB and GFN2-xTB methods are mostly observed for the calculated intensities that sensitively depend on details of the computed PES. The new GFN2-xTB method produces fewer artifacts in the spectra but eventually misses important signals. Nevertheless, the

results from both methods compare reasonably well to the experimental results. It is to be kept in mind that deviations like intensities, missing signals, and survival rates of the precursor ion depend on technical settings related to internal energy distribution and the ionization/heating procedure. They can be individually improved by changing the default settings, thereby significantly improving the simulated spectra.

In conclusion, the good quality of the two tight-binding methods GFN1-xTB and GFN2-xTB broadens the applicability of the QCEIMS program for calculating electron ionization mass spectra. The quality of the predicted fragmentation patterns and the elucidation of the structural compositions have been improved by the new GFN2-xTB method. This enables users of the program to gain a more detailed look into the EI-MS process without the need for prior knowledge of the reaction pathways involved.

The expansion of the QCEIMS program to involve collision-induced dissociation (CID) mass spectrometry techniques is currently being developed in our laboratory.

A.5 Supporting Information

A.5.1 Details of computational demands and method stability

Timings

To gather an overview of the time consumed by the different combinations of methods, the computation time of all production runs is summed up to gain the total time for creating a full simulated spectrum. For OM3-D3 calculations of the organic molecules, MNDO2005 Version 7.0 was used. The DFT calculations have been conducted using the ORCA 4.0.1.2. suite of programs at PBE0/def2-SV(P) level of theory only for molecules containing transition-metal atoms.

Table A.1: Total computational time (in hours) of all production runs summed up for different types of molecules and methods.

Method:	GFN1			GFN2			OM3-D3
IP-Method:	IPEA	GFN2	DFT	IPEA	GFN2	DFT	OM3-D3
1-Butanol	96	110	-	136	142	-	51
Hexafluorbenzene	165	228	-	190	266	-	190
Uracil	218	288	-	299	213	-	260
Testosterone	561	620	-	691	813	-	999
Sucrose	368	443	-	403	490	-	332
Leucylglycylglycine	260	311	-	275	328	-	240
Bis(benzene)chromium	191	221	507	95	346	222	-
Zirconocene-dichloride	203	257	358	475	576	679	-
Nickel-bis(diph-acac)	264	1250	4664	306	1475	5113	-
Bis(pinacolato)diboron	417	435	-	461	478	-	-
Chinalphos	410	530	-	464	585	-	-
Triphenylstibine	517	611	-	431	557	-	-
Dichloroethylaluminum	69	60	-	76	75	-	-
HOMSi	504	576	-	637	827	-	-
Octasulfur	116	141	-	55	18	-	-

Stability

To gather an overview of the stability using the different combinations of methods, the number of successful production runs are displayed. For each system, 1000 production runs were started. The DFT calculations have been conducted using the ORCA 4.0.1.2. suite of programs at PBE0/def2-SV(P) level of theory only for molecules containing transition-metal atoms.

Table A.2: Number of successful production runs. For each system there have been a total of 1000 parallel production runs.

Method: IP-Method:	GFN1			GFN2		
	IPEA	GFN2	DFT	IPEA	GFN2	DFT
1-Butanol	999	1000	-	999	1000	-
Hexafluorbenzene	964	960	-	992	972	-
Uracil	1000	1000	-	1000	1000	-
Testosterone	1000	999	-	993	992	-
Sucrose	995	997	-	993	989	-
Leucylglycylglycine	990	993	-	993	992	-
Bis(benzene)chromium	962	943	981	878	888	900
Zirconocene-dichloride	986	992	993	993	990	995
Nickel-bis(diph-acac)	994	979	1000	998	964	996
Bis(pinacolato)diboron	998	993	-	998	997	-
Chinalphos	995	1000	-	991	998	-
Triphenylstibine	999	991	-	991	962	-
Dichloroethylaluminum	1000	1000	-	998	996	-
HOMSi	995	991	-	994	990	-
Octasulfur	951	997	-	835	997	-

A.5.2 Calculated spectra using DFT for IP calculations

The following spectra of the three transition metal containing molecules bis(benzene)chromium, zircocene-dichloride and nickel(II)bis(diphenyl-acetylacetonate) have been calculated using the ORCA 4.0.1.2. suite of programs at PBE0/def2-SV(P) level of theory for IP calculations.

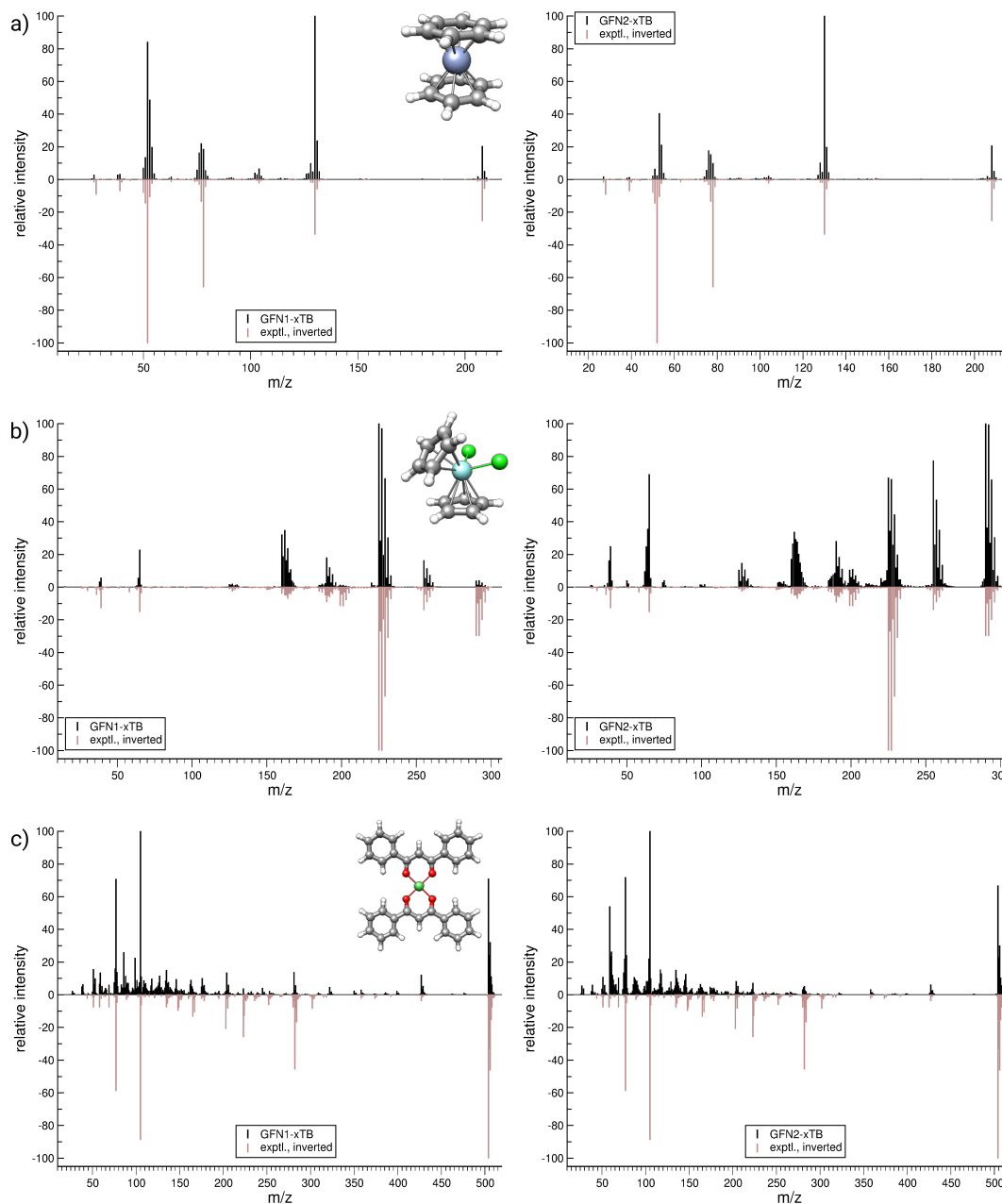
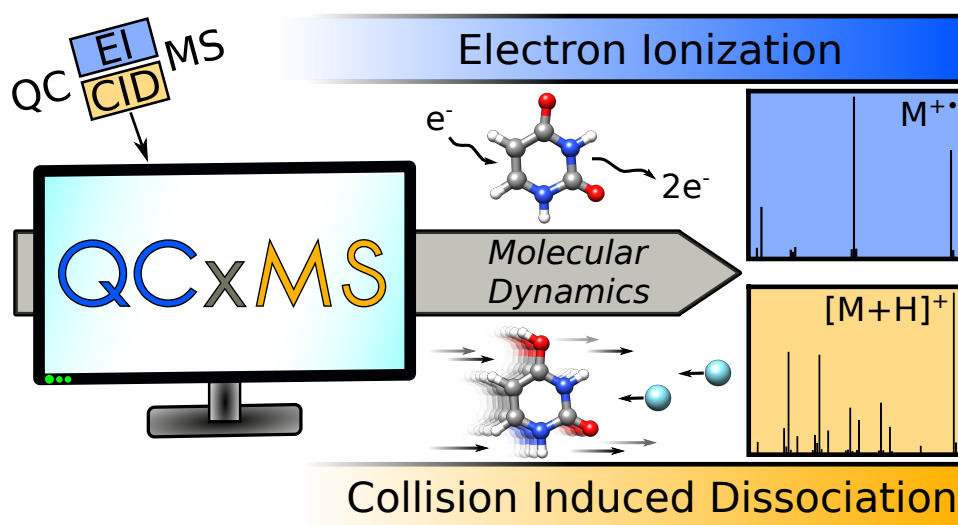


Figure A.11: The EI-MS of bis(benzene)chromium (a), zircocene-dichloride (b) and nickel(II)bis(diphenyl-acetylacetonate)(c). The parent ion has been superimposed onto the corresponding spectrum.

Publication corresponding to chapter 4: From QCEIMS to QCxMS: A Tool to Routinely Calculate CID Mass Spectra Using Molecular Dynamics



Abstract Mass spectrometry (MS) is a powerful tool in chemical research and substance identification. For the computational modeling of electron ionization MS, we have developed the quantum-chemical electron ionization mass spectra (QCEIMS) program. Here, we present an extension of QCEIMS to calculate collision-induced dissociation (CID) spectra. The more general applicability is accounted for by the new name QCxMS, where “x” refers to EI or CID. To this end, fragmentation and rearrangement reactions are computed “on-the-fly” in Born–Oppenheimer molecular dynamics (MD) simulations with the semiempirical GFN2-xTB Hamiltonian, which provides an efficient quantum mechanical description of all elements up to $Z = 86$ (Rn). Through the explicit modeling of multicollision processes between precursor ions and neutral gas atoms as well as temperature-induced decomposition reactions, QCxMS provides detailed insight into the collision kinetics and fragmentation pathways.

In combination with the CREST program to determine the preferential protonation sites, QCxMS becomes the first standalone MD-based program that can predict mass spectra based solely on molecular structures as input. We demonstrate this for six organic molecules with masses ranging from 159 to 296 Da, for which QCxMS yields CID spectra in reasonable agreement with experiments.

B.1 Introduction

Mass spectrometry (MS) is applied commonly in everyday analytical routines, e.g., for drug testing in sports,^{107,108} explosives testing at airports,¹²⁷ or for substance analysis in the chemical laboratory.¹²⁸ In order to be able to identify substances quickly and reliably, authentic references must be generated and stored for comparison. For this purpose, spectral libraries^{40–44,90} were created. However, the maintenance or expansion of the databases is time-consuming, expensive, chemically complex, and could even be dangerous.¹²⁹ Alternatively, computer-aided methods are considered,^{130,131} since data is easily generated, and molecular properties can be calculated inexpensively without endangering the user.

In the last two decades, many theoretical methods have been developed with the aim to simplify the interpretation of mass spectra. A fast approach to screen thousands of possible structures in seconds is to use machine learning algorithms^{45,46} or neural networks⁴⁷ that compare spectra with large amounts of stored data from databases. The disadvantage of this methodology is that databases are incomplete due to the sheer number of possible molecular structures,^{46,49} so structure elucidation of new and unlearned compounds becomes unreliable. Also, these methods do not provide any insight into the fragmentation processes occurring in the experiment. Mathematical algorithms³⁶ or rule-based techniques³⁵ require transition state analysis for *a-priori*-defined decomposition channels that are often unknown for uncommon molecules. Due to the use of predefined fragmentation schemes of typical reactions, the results obtained are biased in one way or another.

The most general approach is to use quantum chemistry (QC)-based methods for the simulation of the basic chemical reactions, because no prior fragmentation schemes or measured spectra are needed to set up the calculations. The Quantum Chemistry Electron Ionization Mass Spectra (QCEIMS) program^{50,105} can automatically calculate standard 70 eV electron ionization (EI) mass spectra using Born–Oppenheimer ab initio molecular dynamics (BO-AIMD)⁷⁴ simulations in positive⁵⁰ and negative-ion mode (dissociative electron attachment, DEA).¹³² The program provides reasonable EI mass spectra for organic, inorganic, and transition-metal-containing molecules that agree well with experimental spectra.^{51–54,89,132} Because the method is based on MD simulations, the resulting trajectories provide detailed insights into fragmentation processes and rearrangement reactions that are difficult to derive otherwise. It was shown that the program is able to identify semiochemicals and metabolites of prior unknown compounds better than the commonly used CFM-ID program.^{4,112}

As an alternative to EI, the ionization of molecules can be obtained through protonation. Since the nature of the primary ions between these approaches differs (usually an open-shell vs closed-shell electronic structure), the underlying potential energy surfaces (PESs) are (at least initially) different as well. Chemical activation of the ions can be achieved *via* collisions of the precursor ions with neutral gas atoms,²⁹ either in-source¹³³ or in tandem (MS/MS).^{134–136} The resulting collision-induced dissociation (CID) allows controllable fragmentation rates and detailed structural characterization of the species under study.

MD methods to simulate CID experiments have been first proposed by Hase *et. al.*^{137–144} and carried

out on many systems by Spezia *et. al.*^{37,38,145–152} using the VENUS program package.¹⁵³ Single collisions between precursor ions and neutral gas atoms were analyzed, and multiple collisions were approximated by thermal activation. For a recent review, see ref. [154]. To our knowledge, multiple collisions by repeated explicit collisions between ions and neutral gas atoms were not considered in prior work. Furthermore, no automated software has been developed that is specifically designed to routinely calculate CID mass spectra based on MD simulations.

To this end, the QCEIMS program has been extended to automatically calculate CID spectra in positive-ion mode. To avoid confusion with the specific name used so far, we decided to rename the method QCxMS, where “x” refers to EI or CID. Since many processes and technical details of a CID experiment can not be fully elucidated, it is difficult to simulate the exact conditions in a collision chamber. Many parameters influence the calculated spectrum, so the main goal of this work was to develop a method that can reliably and routinely calculate CID spectra with only a few varying parameters. This initial work focuses on collisions of singly protonated ions $[M + H]^+$ with neutral gas atoms comparable to the conditions in low-energy electrospray ionization (ESI) CID-MS/MS experiments. Therefore, run-types have been developed to simulate single and multiple collisions between precursor ions and neutral gas atoms under various experimental conditions as well as thermal activation by simply increasing the internal energy.

For testing, six molecules composed of 16 to 34 atoms were considered. The set consists of organic compounds (H, C, N,O) with cases including chlorine or sulfur atoms. The calculated results were compared against experimental spectra taken from common databases.^{41,42} Basic theoretical concepts and technical details of the new CID module as well as observed fragmentation patterns and collision energetics are discussed.

B.2 Theoretical Background

B.2.1 Ionization and Internal Energy

The internal energy of an ion is the most important factor for the molecular fragmentation processes in an MS experiment. In EI, the targeted molecule is ionized by an electron beam with a well-defined energy (standardized 70 eV) that creates an open-shell radical ion.¹¹ By this “hard” ionization process, the precursor ion gains a large internal energy, which leads to a high degree of cascading fragmentation events. In contrast, the average internal energy distribution of the precursor ion after “soft” ionization is lower.^{13,14} Typical methods like atmospheric pressure chemical ionization (APCI),¹⁸ electrospray ionization (ESI),^{19,155–157} or matrix-assisted laser desorption ionization (MALDI)^{158,159} lead to ionization through (de)protonation of the target molecule such that commonly a closed-shell ion (even number of electrons) is formed. Depending on the chosen ionization method, the analyte undergoes heating, acceleration through strong electric fields, collisions with gas molecules, and large pressure changes that modify the internal energy of the ion during the ionization. The exact distribution of the internal energy gained by the ionization process can differ significantly depending on the experimental setup and has to be determined under strict conditions for different instrumentation using thermometer ions.^{160,161} For simplicity, ESI is henceforth chosen as the reference ionization technique, even if the mechanisms in ESI are not completely clarified.^{162–167} While ESI often produces highly charged samples, only singly charged ions are considered in this work.

B.2.2 Protonation Sites

Fragmentation of the precursor ion often occurs charge-directed, meaning that the initial location of the additional proton (charge) has a significant influence on the resulting spectrum.^{168,169} However, it was found that protons can transfer between spatially close-lying heteroatoms before fragmentation occurs, which is known as the “mobile proton model”.^{170–172} It states that due to high internal energy, the initially protonated structure, further called a “protomer”,¹⁷³ is not necessarily in thermal equilibrium. This can lead to kinetically promoted reactions of protomers that are higher in energy and have lower dissociation barriers than the energetically lowest protomer structure. Thus, the spectrometrically “active” protomer cannot be determined solely on the relative free energy values of equilibrium ground state structures.¹⁷⁰ Although the mobile proton model was originally formulated for peptides, studies show that it may also hold for other, small organic molecules.^{174–177} If the internal energy after ionization is too large, the molecular ions will fragment before detection, while too low energies prevent any fragmentation.¹³ This limits the energy available for rearrangements to a small number of candidate protomer structures. With increasing molecular size or many heteroatoms (atoms like N or O with lone pairs), the effort to study all possible protomers becomes tedious and can easily lead to a steep increase of the computational cost. To simplify this process, a generally applicable and automated protonation protocol was developed recently⁸⁵ and implemented into the Conformer Rotamer Ensemble Sampling Tool (CREST).^{86,178,179} This computational protocol automatically analyzes lone pairs and π orbitals of a molecule (i.e., possible protonation sites) and generates an energy-ranked ensemble of protomers. To ensure a correct ordering, the ensembles generated in this work were initially obtained at the GFN2-xTB⁷¹ level and refined at the density functional theory (DFT) level. Details on the used settings can be found in Section B.3.3.

B.2.3 Collision Events

The internal energy (E_{int}) of a molecule ionized by soft ionization methods is low so that the precursor ion $[M+H]^+$ has to be activated by other means than the ionization itself. In collision activated/induced dissociation (CAD/CID) experiments, the precursor ion is accelerated in an electric field and subsequently brought to collision with neutral gas atoms. Depending on the experimental conditions like pressure and collision energy (CE), precursor ions can undergo multiple collision events in the collision cells of quadrupole instruments running in the low-energy regime^{13,14,38,180,181} (up to 100 eV in the laboratory frame E_{LAB} ¹⁸⁰). In order to reproduce the correct energy transfer in the experiment, the number of collisions has to be known. Applying the kinetic gas theory, the average number of collisions can be approximated by⁸⁴

$$Z = L/\lambda, \tag{B.1}$$

$$\lambda = \frac{k_B T_g}{\sigma P_g}, \tag{B.2}$$

where Z is the number of collisions, L is the length of the collision chamber, λ is the mean free path, P_g is the pressure of collision gas, T_g is the absolute temperature, k_B is the Boltzmann constant, and σ is the collision cross section.

With each collision, the kinetic energy (E_{kin}) of the accelerated ion is transferred into its rotational and vibrational modes. If a specific energy threshold is reached, bond-breaking of the ion occurs

(“statistical fragmentation process”). At high collision energies, a collision can cause a large local increase in energy that leads to “nonstatistical” or “direct bond dissociation” at the impact site.^{147,182} The maximum amount of energy that can be transformed from E_{kin} into E_{int} per collision is defined as the center-of-mass energy (E_{COM}).³⁶

$$E_{COM} = \frac{m_g}{m_g + m_p} E_{kin}, \quad (\text{B.3})$$

where m_g is the mass of the collision gas atom and m_p is the mass of the ion.

After each collision, the amount of kinetic energy available for the subsequent collision is reduced, depending on the in-elasticity η of the collision event³⁶. This effect is also known as “collisional cooling”. When averaging over all possible scattering angles, the loss in kinetic energy through a collision is given by

$$\Delta E_{kin} = -E_{COM} \frac{2m_p + \eta m_g}{m_p + m_g}. \quad (\text{B.4})$$

B.2.4 Analysis

A common method to investigate the internal energy accumulation is to study the survival yield (SY) of the precursor ion

$$SY = \frac{I_p}{I_p + \sum I_{F,i}}, \quad (\text{B.5})$$

where I_p is the ratio of the precursor signal intensity to the sum of fragment signals intensities $I_{F,i}$. The SY provides a means to analyze the kinetics of CID processes.^{13,183} It includes various effects, such as collision energy and number of collisions on the stability of the precursor ion, which is discussed in Section B.4.1. The CE50 value¹⁸⁴ indicates the threshold at which the internal energy is large enough to yield a ratio between $[M+H]^+$ and fragment signals of 1:1.

B.3 Methodology

The basic structure of QCxMS has been maintained during the development of the CID approach so that all the various modules (EI, DEA, CID) run with the same basic architecture. Details on the other modules can be found in previous publications.^{50–54,56,89,132}

First, the molecule under study has to be protonated. For large or flexible molecules, it is advised to perform a conformational search in advance. Both tasks can be conducted with the CREST program, as described in B.2.2. QCxMS runs in three steps:

1. **Equilibration and sampling:** the equilibration of $[M+H]^+$ is conducted using an MD simulation at a constant temperature to ensure canonical ensemble conditions (NVT). This is followed by a sampling run in the micro-canonical ensemble (NVE), in which snapshots of the structure are saved. These generate different starting geometries for the highly parallel production runs in step 3.

2. **Setting up production runs:** the geometries and input information are set up for the following production runs. In contrast to the EI mode of QCxMS, no impact excess energy distribution or relaxation time have to be computed.
3. **Production runs:** the collision events are simulated by MD. The runs are done massively parallel by executing each simulation on a single computer core. Due to the many technical aspects involved in these calculations, a detailed discussion of this step is given in Section B.3.1.

B.3.1 Production Runs and General Collision Simulation

Each production run is a sequence (cascade) of independent MD simulations of a single ion undergoing collisions. The general setup was taken from the work of Hase and Spezia^{37,38,137–152} but was adjusted to allow simulating single and multiple collisions as well as a thermal activation in a universally applicable workflow. Likewise, some initial starting parameters were adopted, but others are automatically determined by the program to ensure a general applicability. The details on each unique collision simulation are as follows (see Figure B.1):

1. The ionization MD: Scaling E_{int} of the precursor ion prior to collision simulations. The energy is provided using the Berendsen thermostat⁷⁵ with a Box–Muller standard distributed¹⁸⁵ threshold. The length of the MD simulation depends on the difference between the equilibration temperature and the target temperature. This ensures that the scaling of the molecular temperature is done uniformly and the thermal energy is distributed correctly.
2. The precursor ion is randomly rotated around its Euler axis to consider all potential impact directions between $[M+H]^+$ and the collision partner. For each principle moment of inertia, an internal rotation energy of $k_B T/2$ is given.
3. The neutral gases helium, neon, argon, krypton, xenon, and dinitrogen are available for the simulation. The collision gas atom/molecule is placed along a Cartesian axis at a distance of 25 Å away from the molecular ion's center-of-mass (COM) and is considered to be stationary. The exact position of the atom is displaced perpendicular to this vector by a random value to account for a varying scattering angle, better known as the impact parameter b . To guarantee a collision in the simulation, the maximum value of b is the radius of $[M+H]^+$. In experiments, near-hits and high-angle collisions are possible, but due to the low transfer of kinetic energy, nondirect impacts are not significant. Omitting these in the simulation leads to a significant reduction of the computational cost. A more detailed work on the effect of the impact parameter can be found in the literature.³⁸
4. The velocity of the precursor ion after the acceleration in the electric field is calculated from the provided E_{LAB} or E_{COM} values and alternated randomly for each independent production run by a Box–Muller distribution to account for the variations in the experimental setup. The velocity is then added uniformly onto the atoms of the ion and oriented along a vector between the COM and the collision gas atom.

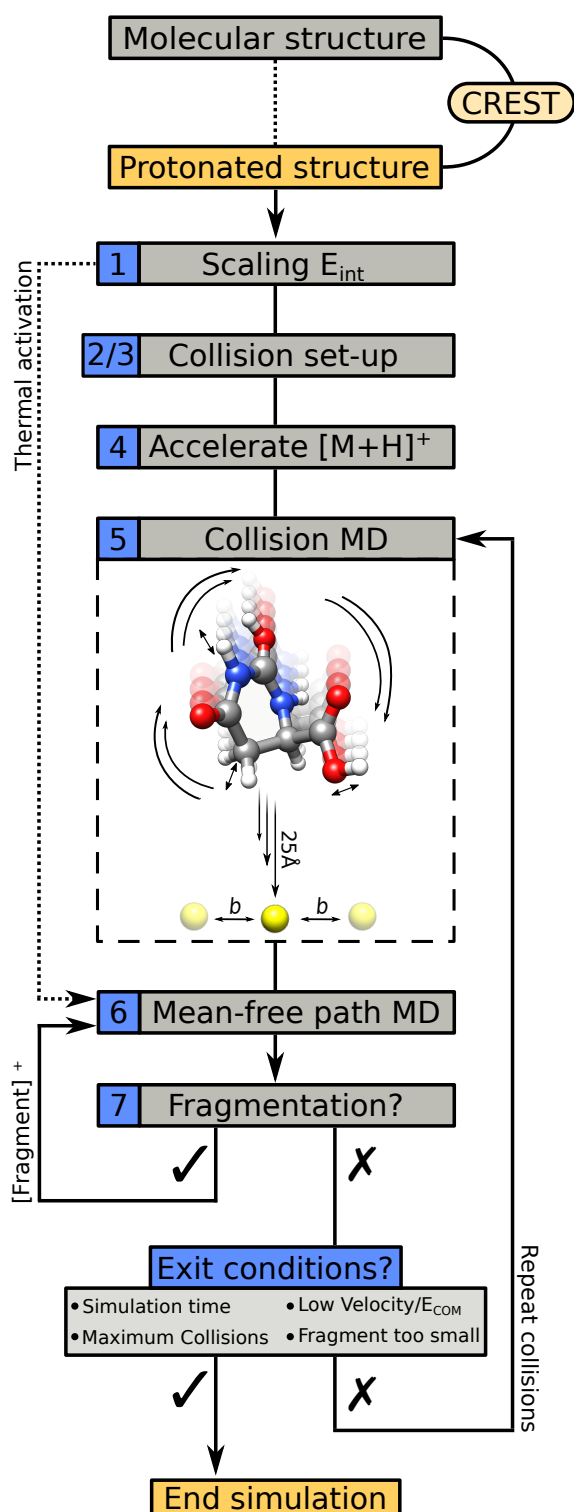


Figure B.1: Flowchart of the steps in the CID module.

5. The collision MD run: After the program concludes a collision event, the simulation runs for 800 more MD time steps to guarantee that the interactions between the collision partners decayed.

6. The mean-free-path MD run: The ion is propagated in time to sample the energy redistribution of the impact energy into the rovibrational modes of the ion. This step significantly influences the overall simulation time. The default value between multiple collisions is set to 5 ps. To adjust the total simulation time, single-collision and thermal activation simulations require longer propagation times.

7. If the ion undergoes fragmentation, the vertical ionization potentials of the fragments are calculated by the Δ SCF (self-consistent field) method. The fragment with the highest statistical charge is propagated further in a subsequent MD simulation to account for complete energy distribution. If this again leads to fragmentation, consecutive MD simulations are conducted until the fragmentation cascade stops. The fragments with lower statistical charge are not considered for further MD simulations but are instead counted and stored. A detailed description on how the charge is counted can be found in a previous publication.⁵⁶ By the automatic assignment of the charge to the fragments, all the information needed to generate a CID spectrum is provided.

Four generally valid conditions are defined that conclude the simulations after step 7:

- if the maximum simulation time has been exceeded;
- if the number of collisions is reached;
- if the amount of transferred energy gets too low,

i.e., if the ion velocity decreases below 800 m/s or E_{COM} drops below 0.4 eV; and

- if the ion gets too small, i.e., there are less than five atoms left or the total mass is lower than a given value.

At each time step, information on the geometries, velocities, and energetics of the ions are stored by the program for analysis purposes. An external script that provides the natural isotope ratios computes and visualizes the resulting spectrum. The script is distributed with the main program¹⁰⁵ and was used throughout this work. Computational details on the MD runs can be found in Section B.3.3.

B.3.2 Multiple Collisions

The steps mentioned in this section refer to the workflow in section B.3.1.

Multiple collisions of the precursor ion with a collision partner can be simulated by repetition of the setup step 3 followed by the sequence of steps 5 to 7. Overall, the simulated CID spectrum depends strongly on three parameters:

1. the starting velocity of the precursor ion;
2. the number of collisions the ion undergoes; and
3. the internal energy the precursor ion has prior to the first collision.

Roughly speaking, increasing the velocity or the number of collisions are different ways to increase internal energy. Various methods were tested that either explicitly simulate collisions or imply collisions by increasing the internal energy of the precursor ion. Three of the most useful approaches are discussed in the following.

Forced Fragmentation by Multiple Collisions

In the *forced activation* run-type, the collisions between the precursor ion and the collision gas atoms are repeated up to a predefined value. This run-type was used to simulate all single, double, or higher multiple collisions shown in this work. It is the most straightforward method to obtain a correlation between the number of collisions and the fragmentation behavior of the molecular ion. The simulation of actual collision processes between the collision partners renders this run-type an “explicit collision” model. Various settings can be selected, e.g., switching on collisions between fragments and neutral gas atoms (fragment–gas collisions (*fgc*)). The number of these collisions is randomly distributed between zero and one-tenth of the number of atoms the fragment is composed of. This size-dependence helps to counteract overfragmentation and decreases simulation time but is based on empirical experience rather than physical principles. More details on all settings can be found in the QCxMS manual.¹⁰⁵

Thermal Activation by Increase of Internal Energy

As suggested in previous work,^{146–148} multiple collisions can be approximated by increasing the E_{int} of the molecular ion to a fixed value. Because no “real” collisions are simulated, this *thermal activation* run-type is an “implicit collision” model. First, the precursor ion is heated up to a predefined internal energy in step 1. To account for variations in the experiment, the value can be varied randomly by

a Box–Muller distribution. The run-type bypasses steps 2 to 5 and simulates the mean-free-path MD (step 6). The scaled energy is distributed into the vibrational modes of the ion, which leads to statistical fragmentation. Cascading fragmentation events are accounted for by subsequent MD simulations, as described in step 7.

General Scheme by Combining Thermal Activation and Multiple Collisions

The *general activation* run-type first scales the internal energy of the precursor ion in step 1 to mimic the ESI process. The energy scaling value is randomly varied for each individual production run to account for variations in the experiment. To prevent fragmentation in this step, the scaling is done in an empirically determined energy range that depends on the molecular size. Optimally, this allows mobile proton transfer to occur. Consecutively, multiple collision simulations are performed, in which the number of collisions is calculated according to eq B.1. The inclusion of actual collisions renders this an “explicit collision” model. For collisions between fragments and gas atoms (*fgc*), the radius of the fragment is automatically determined, and the number of possible collisions with neutral gas atoms is calculated via eq B.1. The number of *fgc* is reduced depending on the number of collisions the precursor ion already conducted, to account for the way the ion traveled along the collision chamber.

B.3.3 Technical Details

Experimental Details

The experimental spectra were taken from the Human Metabolome Database (HMDB)^{97,186–188} and the MassBank Europe database.^{98,189} The provided spectra are typically measured with collision energies between 10 and 50 eV. The instruments documented on the Web sites^{41,42} are the Waters Micromass Quattro Triple Quadrupole mass spectrometer (LC-ESI/APCI-QQQ), the Bruker Maxis Impact mass spectrometer (LC-ESI-QTOF), and the Applied Biosystems API3000 mass spectrometer (LC-ESI-QQ). For details on the individual experimental settings and specifications, see the Supporting Information (SI) (see Section B.6). Standardized and informative documentation of the experiments is only partially provided. The experiments conducted on the HMDB Web site are documented by the standard operating protocol (SOP) #21 Version #2, while MassBank spectra were analyzed under nonstandardized, independent experimental conditions.¹⁸⁹ For none of the experiments was information on the type of collision gas, collision gas pressure, or collision cell length available. The experimental E_{LAB} collision energy refers to an additive electric field acceleration, but the initial velocity of the ions is not provided. No information on the initial internal energy of the molecular ions was provided. The design of MS/MS instruments can differ fundamentally so that, e.g., the detection limit and reaction time frame of the instrument can lead to discrepancies between experiments. While this results in low reproducibility of the intensities of the signals produced by ESI-MS experiments, the fragmentation patterns between instruments should be almost identical.^{189,190}

Computational Details

All calculations in this work have been conducted on Intel Xeon E3-1270 3.60 GHz CPU cores. A new QCxMS version 5.0¹⁰⁵ was developed in the course of this work and used throughout. The automatic protonation tool of CREST version 2.11 was used to create the protomer ensembles in an energy window of 30 kcal/mol. The ENantiomer SORTing algorithm (ENSO)¹⁹¹ version 2.0.2 was

used for the reranking at DFT level. The DFT calculations were conducted with the ORCA^{121–123} suite of programs version 4.2.1 at the PBEh-3c¹⁹² level of theory.

In earlier work on CID processes,^{37,38,145–152} calculations were conducted using the semiempirical AM1 and PMx methods,¹⁹³ with and without the DFT-D2 method.¹⁹⁴ However, in various publications^{195–199} it was shown that the GFNn-xTB^{70–72} (n = 1,2) methods outperform other semiempirical quantum mechanical (SQM) methods for a wide range of applications. The GFNn-xTB schemes can properly dissociate bonds mainly due to the use of the finite electronic temperature (Fermi smearing) model. Therefore, xtb version 5.8.1 was implemented into QCxMS, and the GFNn-xTB methods were successfully tested for calculating of EI mass spectra,^{56,89} making the program independent from any third-party software. The QC calculations in this work were done with GFN2-xTB. It uses the DFT-D4^{56,89} model to account for the inter- and intramolecular dispersion interactions.

MD simulations were performed through the integration of Newton's equation of motion using the leapfrog algorithm with a time step of 0.5 fs. To obtain statistically converged results, the number of trajectories for each calculation was set to 25 times the number of atoms in the ion. The equilibration MD and the snapshot sampling were set to 50 fs times the number of trajectories. The electronic temperature for the ground state sampling was set to 298 K and increased to 5000 K for the production runs. The desolvation temperature of 600 K reported in SOP #21 Version #2 was used as initial MD temperature. Argon was employed as the collision gas throughout.

The activation time of single collisions in the low-energy regime of CID experiments is in the range of picoseconds and increases with multiple collisions depending on the mean free path between the collisions.²⁰⁰ Collision cells in a triple-quadrupole setup are often tens of centimeters long, so an ion needs several milliseconds to cover this distance.¹⁸⁰ However, affordable MD simulations are in the range of tens of picoseconds. This may cause the survival rates of short- and long-lived fragments to be significantly different than in experiments and may lead to incomplete accounting of long-term dissociation or rearrangement reactions. Calculations on this time scale are only feasible with rather approximate SQM methods, and their errors may appear in the underlying PES leading to wrong dissociation and rearrangement reactions or weird fragment structures. Furthermore, inaccurate IP calculations can cause faulty charge assignment, and thus, following fragmentation cascades might occur for the wrong fragment.

B.4 Results and Discussion

The influence of the collision energy, number of collisions, and increase of internal energy on the SY of the molecular ion is discussed in Section B.4.1. Single collisions with increasing acceleration energies are presented in Section B.4.2, and Section B.4.3 deals with protonation sites and their influence on the calculated spectrum. In Section B.4.4, the effects of multiple collisions on a calculated spectrum are discussed, and Section B.4.4 shows results of the *general activation* run-type. Many aspects need to be considered when comparing calculated to experimental spectra, especially the type of and settings in the experimental instrument. However, QCxMS only simulates the collision dynamics and mean free paths of the ions under study. While some aspects of the collision cell can be adapted, the simulations can not account for special features of various instruments, so a comprehensive discussion of the differences between calculated and experimental results would go beyond the scope of this initial work. Further results can be found in the SI, where spectra of various protomers, as well as

spectra measured at different experimental settings, are shown and compared.

B.4.1 Collision Energetics

As an example for a typical organic compound, calculations on caffeine were conducted for a different number of collisions and impact/internal energies. The thermally most populated protomer at 600 K was used, and 100 trajectories were calculated for each plot in Figure B.2. The *forced activation* run-type was used, but the collision energies were not randomly distributed in these calculations. This guarantees that the exact energy values at each given velocity can be calculated.

The effects of the collision energy on the SY of the precursor ion are demonstrated in Figure B.2 (a). The calculations were conducted from 40 to 120 eV E_{LAB} for one, two, and three collisions. The SY for a single collision starts decreasing at 60 eV E_{LAB} and reaches 50% SY (CE50) at 104 eV (~ 17 eV E_{COM}). With two collisions, CE50 is reached at 76 eV, and 10% of the precursor survives at 120 eV E_{LAB} . For three collisions, the CE50 value is achieved with 62 eV. Simulating multiple collisions efficiently decreases the collision energy needed to reach the CE50 threshold.

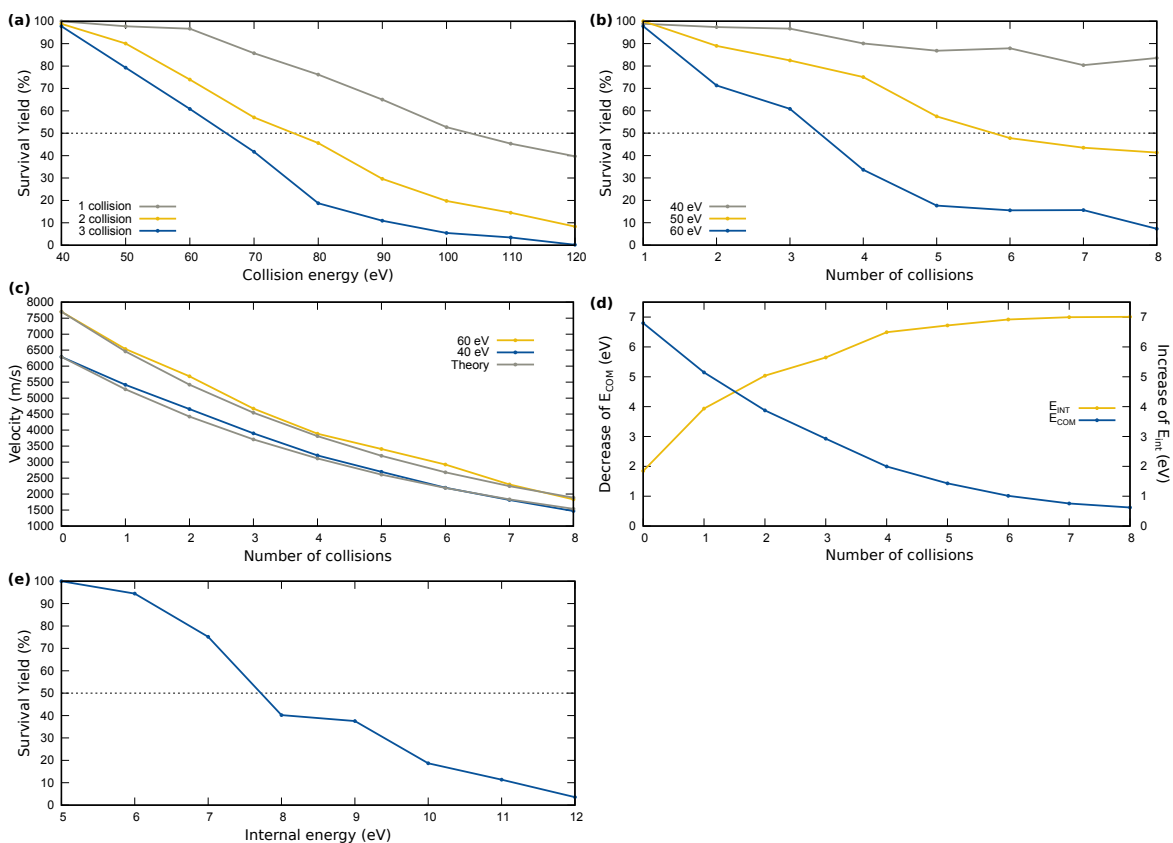


Figure B.2: Collision kinetics for the caffeine precursor ion equilibrated at 600 K. (a) SY at 40 to 120 eV E_{LAB} with one, two, and three collisions. (b) SY after 1 to 8 collisions events at 40, 50, and 60 eV E_{LAB} . (c) Velocity of $[M+H]^+$ after 1 to 8 collisions at 40 and 60 eV E_{LAB} compared to theory.⁸⁴ (d) Correlation between E_{COM} and E_{int} after 1 to 8 collisions at 40 eV E_{LAB} . (e) SY after internal energy scaled from 5 to 12 eV.

To emphasize this correlation, the precursor SY was calculated for one to eight collisions at

acceleration energies of 40, 50, and 60 eV E_{LAB} . The results are plotted in Figure B.2 (b). The change of the SY at 40 eV E_{LAB} stagnates after five collisions and reaches a minimum SY of 80% in total. This means that the $[M+H]^+$ signal remains apparent in the spectrum and does not diminish by more collisions. At 50 eV E_{LAB} , the CE50 rate is reached between five and six collisions, and the curve flattens to an SY of 38%. The 60 eV E_{LAB} curve crosses the CE50 threshold between three and four collisions, and the SY reaches 8% after eight collisions. The reason for the flattening of the curves is due to the decrease of the ion velocity after each collision, i.e., the CE for subsequent ion–gas interactions is substantially lower than the initial CE.

The averaged velocity of $[M+H]^+$ as a function of the number of collisions is shown in Figure B.2 (c) for 40 and 60 eV E_{LAB} . The offset between the two plots demonstrates that the ion with the higher starting acceleration energy retains more velocity after each collision. To validate the performance of QCxMS, the results are compared to the theoretical considerations of Section B.2.3, in which eq B.4 was calculated with a collision in-elasticity of 0.5. Overall, the velocities obtained by QCxMS reproduce the results calculated by the theoretical kinetic gas theory in good agreement. Since the program calculates the kinetic and internal energy of any created fragment, a detailed analysis of the collision dynamics in CID processes can be conducted with the new QCxMS module.

At low velocities, the energy transferred by the collisions becomes insignificant. Hence, a threshold must be defined at which the simulation is no longer economical. In Figure B.2 (d), the average decrease in E_{COM} and the respective average increase of E_{int} were calculated at 40 eV E_{LAB} . E_{int} refers to the averaged thermal energy of the ideal gas. At 600 K, the molecular ion possesses an average initial E_{int} of 1.9 eV. In the first four collisions, the ion obtained a total of 4.6 eV E_{int} . After four collisions, E_{int} increased by 0.2 eV. After six collisions, E_{int} increased 0.08 eV. Considering an E_{COM} of 7.0 eV at the beginning of the simulation and a total increase in internal energy of 5.1 eV after eight collisions, only 73% of the available energy was transmitted in the process. This demonstrates the effect of the in-elasticity η of the collision event.

In Figure B.2 (e), the SY is plotted against increased initial E_{int} values (thermal activation). Because no mean free path between the collisions were calculated, the mean free path MD was set to 15 ps to account for the difference in simulation time. At an E_{int} of about 8 eV, 50% of the precursor ion was decomposed. A comparison to the collision simulations indicates that this condition is fulfilled at about 60 eV E_{LAB} (10.2 eV E_{COM}) acceleration energy after three to four collisions, as can be seen in Figure B.2 (b). An increase to 12 eV E_{int} lowers the SY to 5%.

B.4.2 Single Collisions

As a first example, tetrahydrofuran (THF) was protonated at the oxygen atom and a single collision between the precursor ion and a neutral argon atom was calculated. This system was chosen because of the small molecular size and the inability of intramolecular proton transfer prior to fragmentation, i.e., excluding “mobile proton” effects.

The calculations were conducted for 350 trajectories along a 15.6 ps MD trajectory at 20, 30, 40, 50, and 60 eV E_{LAB} collision energies, with an initial temperature of 600 K. The spectra as a function of increasing impact energy (E_{LAB}) are illustrated in Figure B.3 (a,1–5). After a collision at 20 eV E_{LAB} , the molecular ion retains an average E_{COM} of 2.8 eV and possesses an average 3 eV E_{int} . This is sufficient to form two of the most concise signals of the spectrum, m/z 31 and 55 (Figure B.3 (a,1)). Increasing the collision energy to 30 eV (Figure B.3 (a,2)) increases E_{int} to 3.9 eV, and the third-most prominent signal m/z 45 appears. The molecular ion peak decreases significantly with

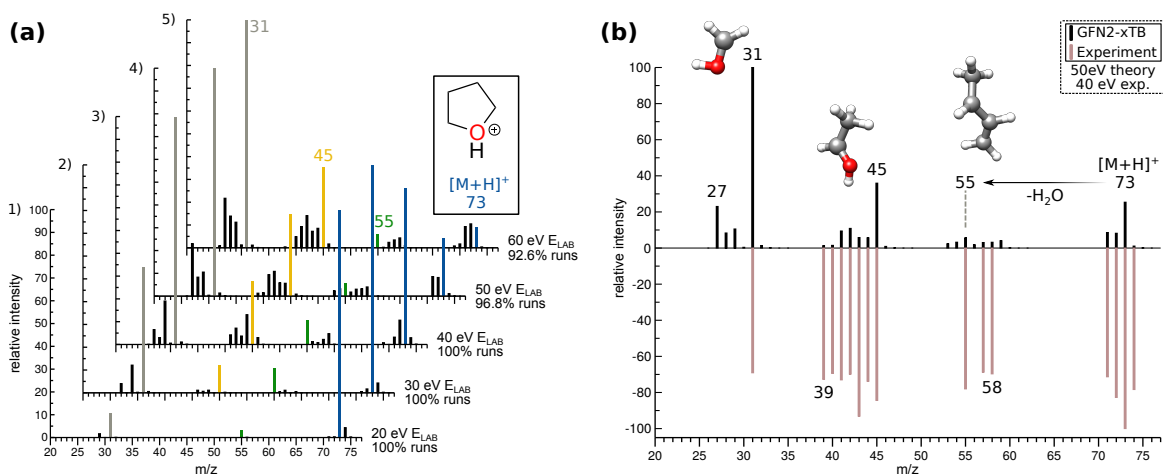


Figure B.3: Calculated spectra of tetrahydrofuran using single collisions at (a) (1) 20 eV, (2) 30 eV, (3) 40 eV, (4) 50 eV, and (5) 60 eV E_{LAB} . The most pertinent fragments m/z 31, 45, 55, and 73 are colored for better traceability. (b) Comparison of the 50 eV calculated (black) to the 40 eV literature spectrum (red, inverted).⁴²

collision energies higher than 30 eV E_{LAB} , and signal m/z 31 becomes the base peak of the spectrum. It is formed by direct fragmentation of $[M+H]^+$ into HOCH_3 (m/z 31) and C_3H_6 (m/z 42). The signals between m/z 39 and 44 increase with higher collision energies, while the intensity of signal m/z 45 stagnates at a relative intensity of 37% after 40 eV E_{LAB} is reached. The signal m/z 55 is formed through rearrangement of a hydrogen atom to the protonated oxygen atom and subsequent H_2O loss. Since this is not a proton transfer between heteroatoms, it is not necessarily described by the “mobile proton model”. With increasing collision energy, direct fragmentation is favored, and rearrangement reactions before bond-breaking are less likely. As a result, signal m/z 55 decreases at higher collision energies.

Experimental spectra are often measured over a range of energies (usually 10 to 50 eV) to visualize differences in signal strength, i.e., analyte lability, and signal occurrence between instruments. Figure B.3(b) shows the theoretical spectrum calculated at 50 eV compared to a literature spectrum⁹⁷ measured at 40 eV E_{LAB} . Comparison to literature spectra measured at different energies can be found in the SI. The qualitative overall agreement is good, and all signals displayed in the experiment can be found in the theoretical spectrum. Signals m/z 32, 33, 46, 47, 53, 54, 59, 60, and 70 produced by the calculations are not measured, and the intensities between calculation and experiment differ. These discrepancies can have different reasons, which are discussed in Section B.3.3. Furthermore, no signals with $m/z < 31$ are observed in the literature spectrum, indicating a (technical) mass cutoff.

The calculations show that single-collision simulations can be used to analyze the energy needed to form fragments and if their formation is due to statistical or direct fragmentation. However, the calculation of CID mass spectra by single collisions is difficult to generalize, because molecules of the same size can require different amounts of collision energy to induce fragmentation, depending on the functional groups present.¹⁷⁰ Furthermore, larger structures need more E_{LAB} collision energy to produce sufficient fragmentation. This can bear some problems: (i) the high impact energy favors fragments that are created through the scattering of the precursor ion, and thus, statistical fragmentation is not sufficiently accounted for. Depending on the structure, this creates lots of small fragments and leads to a loss of information on the fragmentation mechanism; (ii) increasing velocities may lead to

more complicated electronic structures and SCF convergence failure. For THF, failure rates increased from 3.2% at 50 eV to 27.4% at 80 eV E_{LAB} . The impact event sampling is too sudden, and the drastic conversion of the energy cannot be sampled by the MD simulation. Decreasing the time step to 0.1 fs did not solve this problem. However, simulating multiple collisions at low velocities ameliorate these issues.

B.4.3 Protonation sites and proton mobility

The influence of different protonation sites was investigated for the example of 2,6-dichlorobenzamide (BAM). The free energy rankings of all protomers in a 30 kcal/mol energy window do not change for temperatures up to 1200 K. The calculated values can be found in the SI. In Figure B.4, the structures and their respective relative energies at 600 K are depicted in the insets of the corresponding spectra. Calculations were performed with a single collision at 100 eV E_{LAB} (~ 17 eV E_{COM}). The resulting spectra were compared to an experimental spectrum measured at 40 eV E_{LAB} ²⁰¹ (mass cutoff at $m/z < 100$).

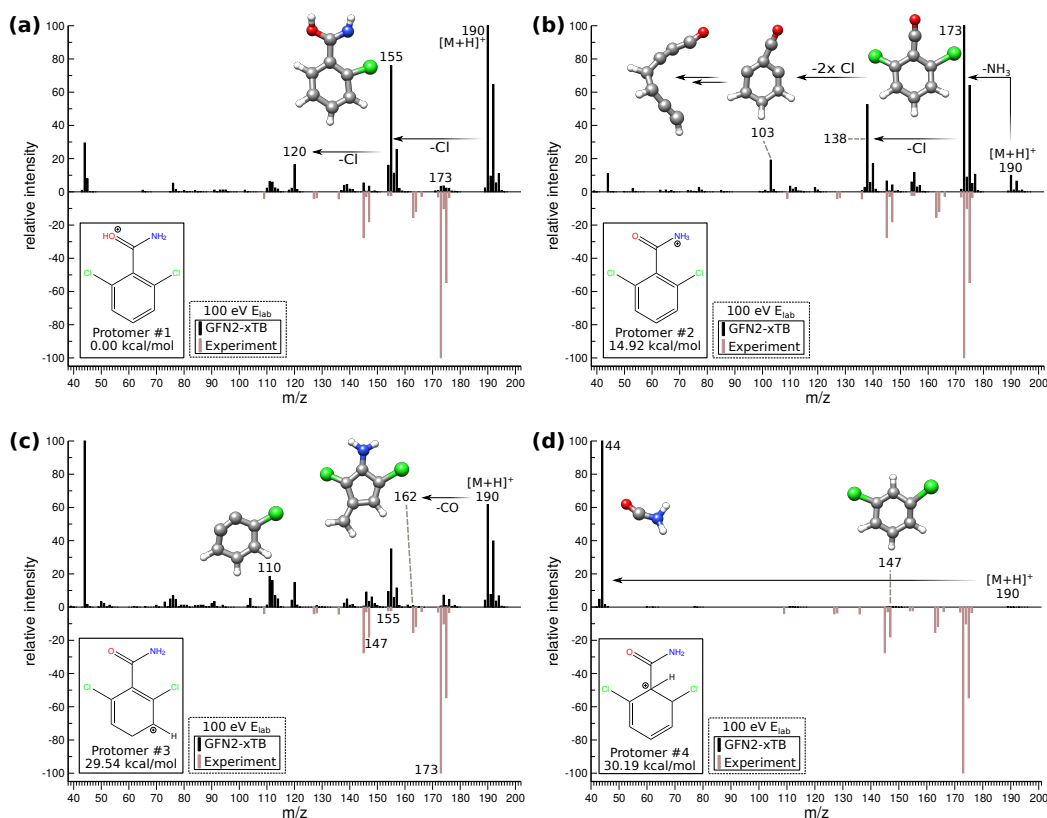


Figure B.4: Calculated spectra at 100 eV with a single collision of the four energetically lowest protomers of 2,6-dichlorobenzamide (black, top) compared to experimental spectra taken from the literature²⁰¹ at 40 eV (red, inverted). The protomer structures are superimposed at the corresponding spectrum with their respective difference in total energy.

Figure B.4(a) shows the calculated spectrum of protomer #1. The structure is protonated at the

oxygen atom of the amide group. The main fragmentation cascade involves the two chlorine atoms, creating the signals m/z 155 and 120. After each separate chlorine atom detachment, proton transfer from the amine group to the unoccupied carbon atoms of the benzene ring occurs. The survival rate of the molecular ion at this collision energy is high. The overall agreement between the experimental and theoretical spectrum is not satisfactory, because the base peak m/z 173 in the experiment is not recreated sufficiently. Sporadic mobile proton transfer from the protonated oxygen atom to the neighboring nitrogen atom occurs before fragmentation. This leads to the separation of NH_3 , and the remaining fragment forms signal m/z 173. The simulation conditions do not significantly favor this rearrangement before the fragmentation occurs, because the high velocity of the precursor ion promotes direct fragmentation through the collision event.

In Figure B.4(b), the results for the thermodynamically second-most stable protomer #2 are displayed. It is protonated at the nitrogen atom of the amide group as in the rearranged structure described before. The free energy difference to protomer #1 at 600 K is 14.9 kcal/mol, indicating that it is not significantly populated. However, the calculated spectrum for this structure is in good agreement with experiment. NH_3 is directly detached from the starting structure and does not first require rearrangement, forming a base peak at m/z 173. Subsequent fragmentation of the chlorine atoms leads to the signals m/z 138 and 103, and ring-opening of the latter structure happens occasionally.

Protomer #3 is protonated at the benzene ring and is high in free energy (29.5 kcal/mol at 600 K). The spectrum is shown in Figure B.4(c). In addition to the signals already discussed, signals around m/z 110 become more intense. The peak at m/z 110 is formed after the fragmentation of the amide group directly from the precursor ion and subsequent HCl loss. A low number of counts for signal m/z 162 was observed in the simulation. This signal derives from the separation of CO from the amide group, while the amine group rearranges simultaneously to the benzene ring. The recyclization of benzene into a five-membered ring with a bound methyl group is energetically favorable, but however rare, maybe due to entropic reasons.

For protomer #4, the protonation of the carbon atom connecting the dichlorobenzene structure to the amide group favors the splitting at these two groups, forming signals m/z 44 and 147 (see Figure B.4(d)). However, the program ascribes the entire charge to the H_2NCO fragment. This decomposition is already observed at very low collision energies (20 eV E_{LAB} ; 3 eV E_{COM}) and demonstrates the influence of charge-directed fragmentation.

Overall, it is evident that the protomer equilibrium populations based on free energies are not sufficient to explain the fragmentation patterns or signal intensities. Dynamical proton transfer between heteroatoms occurring before fragmentation seems to play an essential role for the correct description of fragmentation paths. Different starting protomer structures should be considered when calculating theoretical spectra to account for such effects. Structures in a roughly 30 kcal/mol free energy window may be used for a first evaluation. In this example, however, the two protomers that are close to this value are of little relevance for a proper simulation.

B.4.4 Multiple collisions

Forced activation and thermal activation run-types

The approaches discussed in Sections B.3.2 and B.3.2 were tested on the most populated protomer of caffeine at 600 K. The structure of this protomer is shown in the inset of Figure B.5(a). The spectra of other populated protomers and their energetic ranking can be found in the SI. Since this section

Appendix B From QCEIMS to QCxMS: A Tool to Routinely Calculate CID Mass Spectra Using Molecular Dynamics

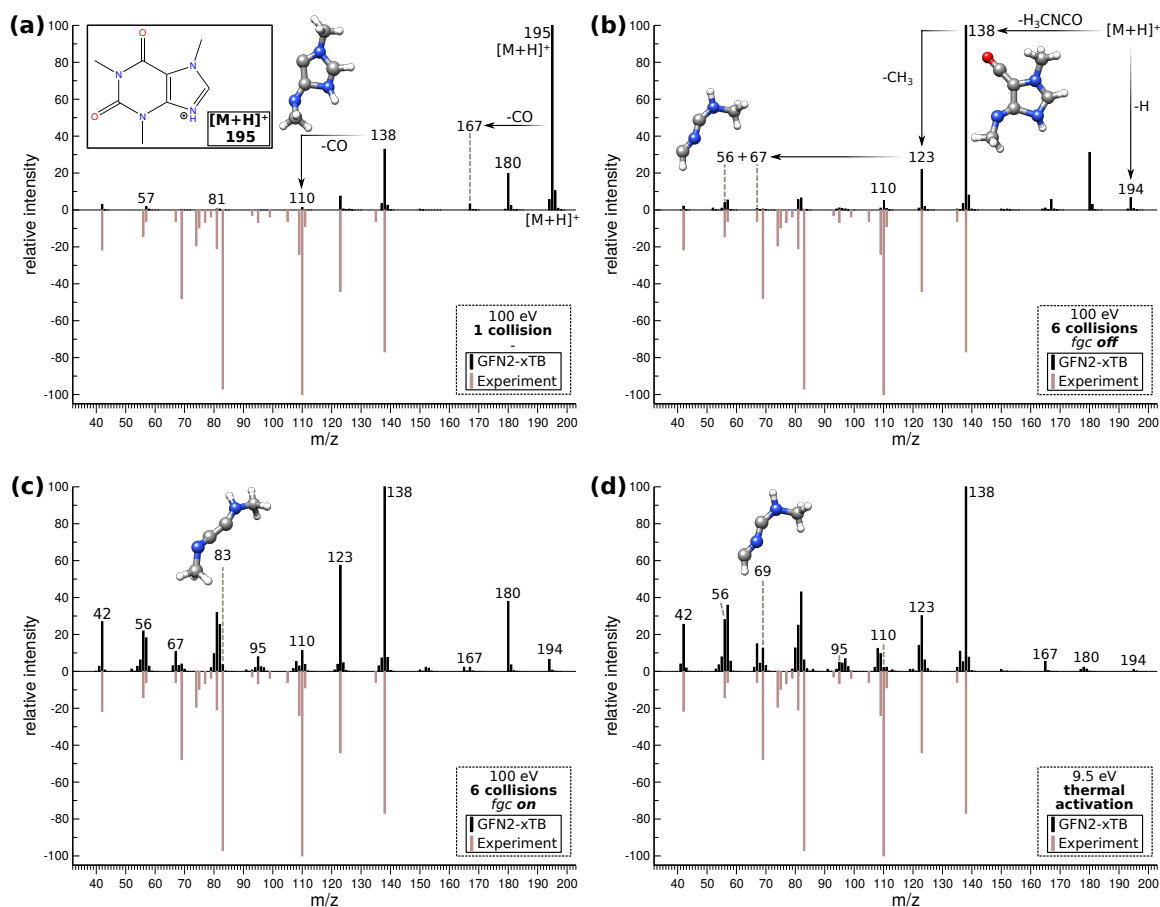


Figure B.5: Comparison of calculated spectra (black) vs. 40 eV literature spectra²⁰² (red, inverted) of the most populated caffeine protomer at 600 K. (a) Single collision at 100 eV E_{LAB} ; (b) six collisions at 100 eV E_{LAB} , *fgc off*; (c) six collisions at 100 eV E_{LAB} *fgc on*. (d) scaling of the internal energy to 10 eV.

focuses on the results gained by the different run-types, a comparison of the different protomer spectra is omitted, and only the effect of multiple collisions is discussed for this example.

The collision-based simulations were calculated with the *forced activation* run-type with a maximum of six collisions between the precursor ion and the neutral gas atom. $[M+H]^+$ was accelerated with an energy of 100 eV E_{LAB} (~ 17 eV E_{COM}), and the mean free paths between the collisions were sampled with a 5 ps MD trajectory. For the *thermal activation* run-type, the target value was set to 9.5 eV E_{int} (~ 3100 K). The total simulation time was set to 15 ps to compensate for not simulating the collision MD. Each spectrum was calculated from 625 trajectories and is compared to experimental spectra²⁰² documented at 40 eV E_{LAB} collision energy (mass cutoff at m/z 32).

The single-collision spectrum is shown in Figure B.5(a) for comparison. The spectrum in Figure B.5(b) was calculated with the forced activation run-type, but subsequent collisions between formed fragments and the collision gas atoms were not considered (*fgc off*). The main difference to the single-collision simulation is that the $[M+H]^+$ signal is not observed. The molecular ion is entirely fragmented after two to three collisions. The base peak of the spectrum is the signal m/z 138. It results

from the separation of H_3CNCO from the six-membered ring. Consecutive fragmentation of a methyl group forms signal m/z 123 and further decomposes into m/z 67 and 56.

The spectrum in Figure B.5(c) was calculated by including subsequent collisions between formed fragments and neutral gas atoms (*fgc on*). In comparison to Figure B.5(b), the relative intensities of fragments with masses $m/z < 138$ are increased, because subsequent fragmentation is amplified. This leads to a better overall agreement with the literature spectrum. Signals m/z 180 and 167 are emphasized by direct fragmentation, in which the impact gas knocks CH_3 or CO directly out of the molecular ion. In a subsequent fragmentation, the imidazole ring remains intact and generates signal m/z 110. Likewise, signal m/z 194, which is a neutral (radical) hydrogen atom loss, is formed preferably by direct fragmentation.

The spectrum in Figure B.5(d) was calculated with the thermal activation run-type. The fragmentation patterns and intensities are comparable to the results produced in Figure B.5(c). The overall agreement with the literature spectrum is good. Because this is an implicit collision model, signals m/z 194, 180, and 167 are formed statistically and thus are less pronounced than in Figure B.5(c). This in turn leads to lesser intensity of signal m/z 110. Signals at $m/z < 90$ are more distinct.

For validation, some of the calculated fragments were compared to structures proposed in a study conducted on deuterated isotopomers of caffeine.²⁰³ The fragments in question are displayed in picture B.5. The fragment structures are largely identical to those reported previously, and a comparison of the fragmentation pathways showed good agreement. Furthermore, signals between m/z 195 and 160 were documented in the experiment, so the signals calculated in this mass range are reasonable.

The results obtained with the forced activation run-type compare best when subsequent fragment-gas collisions are included (*fgc on*). As described in Section B.3.2, the collision number is randomized. It is intended that the random collision numbers lead to runs in which fragments sometimes collide rarely or not at all. This improves the survival rate of fragile fragments that would otherwise decompose by too many repeated collisions. The disadvantage is that this may lead to a high survival rate of fragments that should dissociate with higher internal energies. Nevertheless, this approach produces better results than other protocols that have been tested in the course of this work. Another promising and less random-based procedure to handle fragment-gas collisions is used in the general activation run-type, which is discussed in Sections B.3.2 and B.4.4.

In the *forced activation* run-type, the ions lose considerable amounts of kinetic energy after each collision. As a result, subsequent collisions transfer less energy into the ion (see Section B.4.1). Increasing the starting velocity leads to a larger amount of kinetic energy remaining on the formed fragments, but scattering of the precursor ion and the number of unsuccessful calculations increase, as examined in Section B.4.2. In contrast, fragments created in the *thermal activation* run-type retain considerable energy quantities, because the entire reaction energy is converted into the vibrational energy of the ion. Neither kinetic energy loss by conversion into rotational energy nor collision in-elasticity is taken into account. For this reason, the scaled energy values are not directly comparable to experimental E_{LAB} or E_{COM} data. Adding large amounts of energy in this way is similar to the approach used in the EI method, but the energy in CID experiments is not well-defined, and overfragmentation is often observed.

General Activation Run-Type

The *general activation* run-type was developed to enable a fine adjustment of the theoretical collision conditions to match the experimental ones and thereby generates a tool that can calculate spectra

without the need for excessive trial and error runs. To reduce the collision energy and the number of collisions needed, the internal energy of the precursor ion is increased in an ionization MD prior to the collision runs. As described in Section B.3.2, the heating is done over an empirically determined, molecular-size-dependent energy range. The number of collisions is made dependent on the characteristics of the collision cell described by eq B.1 in Section B.2.3, which takes different properties of instruments into account. For the default setting discussed in this work, an experimental setup with multiple collisions was taken as the reference.⁸⁴ The neutral gas pressure inside the collision cell was set to 0.132 Pa with a chamber length of 0.25 m. The effects of collision cell lengths and ionization MD energies on the calculated spectra are demonstrated in the SI. Although the collision gas is considered static in the calculations, the temperature of the gas affects the number of collisions in the experiment, as can be seen in eq B.1. The collision gas temperature was set to 300 K.

The approach was tested on 3-methylhistidine, phenylephrine, and acetyl-sulfamethoxazole, and the results are compared to experimental spectra measured at different collision energies. The spectra of the protomer structures that matched best with the corresponding experimental spectrum are displayed. The calculated spectra of other protomers and their energetic ranking can be found in the SI. For the collision energies, E_{LAB} values were chosen, so a comparison to experimental values taken from the MassBank⁹⁸ database is appropriate. To increase comparability, the spectra were cut off at lowest m/z signal measured.

For 3-methylhistidine, 600 trajectories were calculated. In the ionization MD, the E_{int} of the precursor ion was increased to 2–3 eV. The collision cell settings led to an average of eight collisions between the precursor ion and the neutral gas atoms. The spectrum in Figure B.6(a) was calculated with 50 eV E_{LAB} collision energy (~ 9.5 eV E_{COM}) and compared to a spectrum measured at 30 eV E_{LAB} collision energy, taken from the literature.²⁰⁴ The spectra are in good overall agreement. The precursor intensity is well-reproduced, and no significant overfragmentation occurs. The main fragmentation path leads to m/z 96 ($C_5N_2H_8$) and 74 ($C_2NO_2H_4$). The former fragment decomposes into m/z 68 ($C_3N_2H_4$) through rearrangement and subsequent ethylene loss, while the latter refers to a neutral loss. However, the neutral structure appears in the spectrum in low yield because of inaccurate IP calculations (wrong charge assignment). Direct fragmentation of $[M+H]^+$ promotes the formation of signal m/z 83 ($C_4N_2H_7$).

The spectrum calculated at 60 eV E_{LAB} collision energy (~ 11.4 eV E_{COM}) is shown in Figure B.6(b). It is compared to a spectrum measured at 40 eV E_{LAB} taken from ref [205]. In comparison to Figure B.6(a), the precursor ion signal is almost nonexistent, and the intensities of low mass signals increase. It is noticeable that signal m/z 95 intensifies strongly in the experiment, which is not observed to this extent in the calculations.

The spectrum of phenylephrine was calculated from 650 trajectories. The ionization MD was simulated for values between 3 and 5 eV E_{int} . An average of eight collisions between the precursor ion and the neutral gas was determined. In Figure B.6(c), the calculations refer to a collision energy of 60 eV E_{LAB} (~ 11.6 eV E_{COM}) and are compared to a 40 eV E_{LAB} experimental spectrum.²⁰⁶ Separation of the terminal H_2NCH_3 group results in rearrangement and creates signal m/z 137. H_2 and neutral H atom loss form signal m/z 134. From signal m/z 137, competing reactions lead to fragments H_7C_7O (m/z 107, loss of HCOH) and H_5C_6O (m/z 93, loss of H_4C_2O). For the latter ion, the program calculates a maximum of six fragment-gas collisions due to its size. Nevertheless, the yield of the ion remains high. Small amounts of H_2 loss form the fragment at signal m/z 91, which further decomposes into differently sized hydrocarbon structures.

The effect of larger collision energies on the spectrum is depicted in Figure B.6(d). The collision

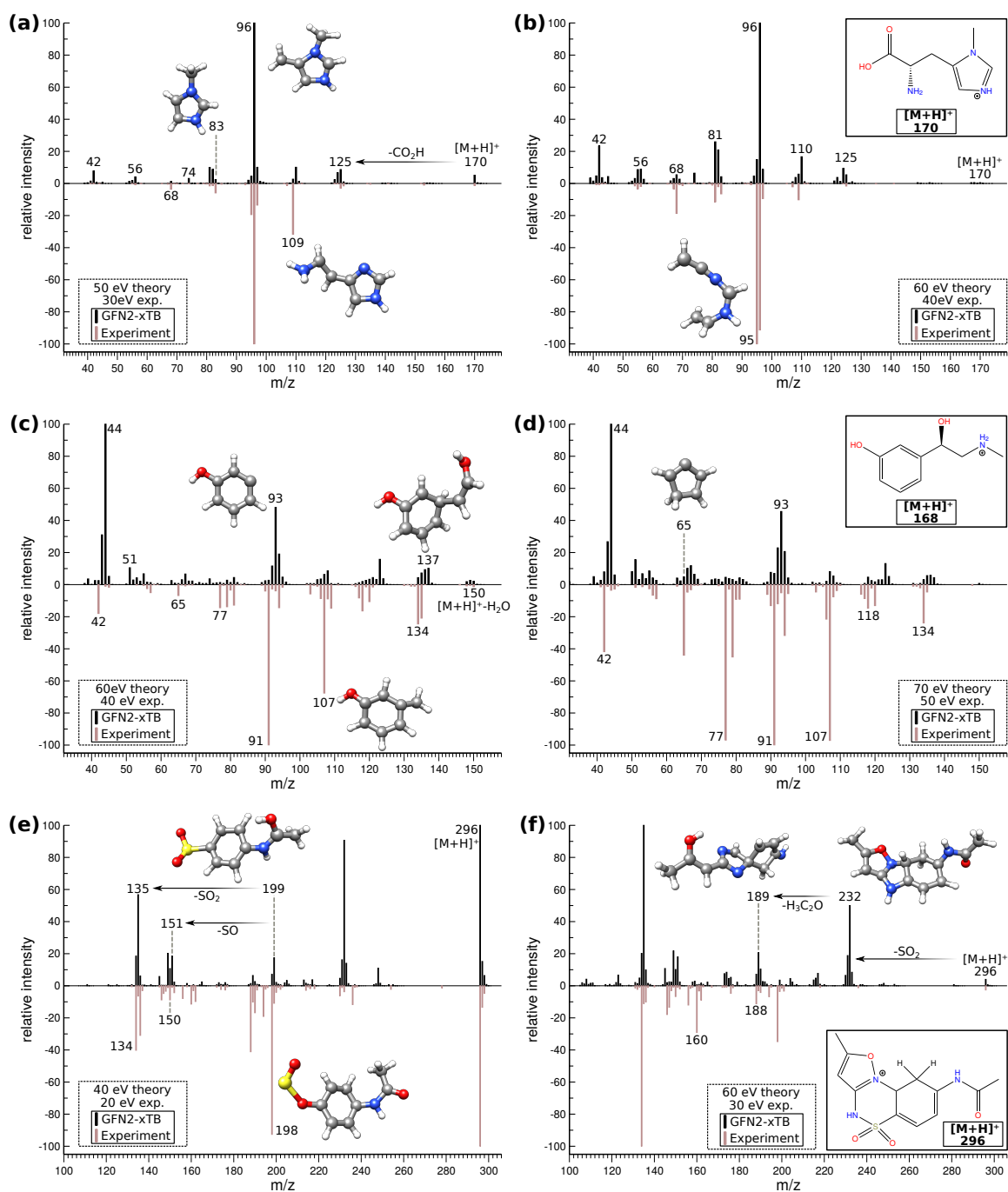


Figure B.6: Calculated spectra (black) compared to experiment⁹⁸ (red, inverted) with collision energies in E_{LAB} in parentheses (theor./expr.). (a) 3-Methylhistidine (50/30), (b) 3-methylhistidine (60/40), (c) phenylephrine (60/40), (d) phenylephrine (70/50), (e) acetyl-sulfamethoxazole (40/20), (f) acetyl-sulfamethoxazole (60/30).

energy for the calculations was set to 70 eV E_{LAB} (~ 13.4 eV E_{COM}) and compared to a literature spectrum²⁰⁷ measured at 50 eV E_{LAB} . In the experimental spectrum, signals appear in the small mass range that were not observed with lower energy. The majority of signals were reproduced by the calculations, but the intensities differ. Hydrogen loss is more dominant in the experiment and cannot be reproduced to the same extent by theory. Reasons for this are discussed in Section B.3.3.

For acetyl-sulfamethoxazole, the third lowest free energy protomer #3 at 600 K provided the most interesting fragments. The calculations were conducted for 850 trajectories. In the ionization MD, E_{int} was increased to values in a range between 3 and 5 eV. The collision chamber settings led to an average of 10 collisions between precursor ion and neutral gas atoms. No experimental signals $m/z < 119$ were reported. Figure B.6(e) shows the spectrum calculated at 40 eV E_{LAB} collision energy (~ 4.7 eV E_{COM}) compared to an experimental spectrum²⁰⁸ measured at 20 eV E_{LAB} . The structure is protonated at the benzene ring, which results in a bond formation between benzene and the isoxazole ring. The structure is shown in the inset of Figure B.6(f). Dissociation of the S–N bond leads to signal m/z 198, and SO_2 remains bound to the six-membered ring structure. Subsequent decomposition of SO or SO_2 leads to signals m/z 150 and 134. Alternatively, signals m/z 160 and 136 are formed by S–C bond-breaking. Here, the SO_2 moiety remains on the methyl-isoxazole group.

In Figure B.6(f), the acetyl-sulfamethoxazole protomer is accelerated to a collision energy of 60 eV E_{LAB} (~ 7.1 eV E_{COM}) and compared to a database spectrum²⁰⁹ measured at 30 eV E_{LAB} . The increase of the collision energy leads to diminished $[M+H]^+$ and m/z 232 signals, while lower mass signals increase in intensity. The latter signal is formed by fragmentation of the SO_2 group from the molecular ion, and the resulting fragment is stabilized through recyclization. The program concludes a maximum of 10 fragment–gas collisions due to the size of the fragment, but the fragment survival rate remains high. Dissociation of the amide group (H_3C_2O) yields signal m/z 189. The benzene and the isoxazole rings undergo rearrangement into a [4.4]-spiro composite.

In protomer #1, the nitrogen atom of the isoxazole ring is protonated, and no cyclization of the precursor ion occurs. Fragmentation leads to dissociation of the N–S bond, the proton remains on the leaving group, and the resulting spectrum consists solely of the signals m/z 198, 150, and 134. Combining and weighing of the two calculated spectra will probably increase the overall agreement between theory and experiment. The spectrum of protomer #1 can be found in the SI.

Considering the many unknown aspects of the experiments, the computed spectra using the general activation run-type agree reasonably well with the reference spectra. The calculated results were more consistent with experiment when a minimum acceleration threshold was considered. This offset can be explained by the initial velocity of the structure gained in the ionization process (see Section B.3.3). The combination of prescaling the internal energy and simulating multiple collisions lowers the required collision energy needed to induce fragmentation and simultaneously accounts for direct fragmentation in the simulation. Different instruments can require different settings, which can be manipulated in the program accordingly; see the QCxMS manual.¹⁰⁵

An important factor is the computational cost of the implemented run-types. Caffeine was calculated with the default settings of each run-type. The *forced activation* run-type (*fgc on*) finished after an average of 32 min per production run. The *general activation* run-type used an average of 21 min per production run, while the *thermal activation* run-type is the fastest approach and takes only 10 min. It has to be kept in mind, however, that the simulation times can vary significantly depending on the settings, the number of subsequent fragmentation and *fgc* events, as well as the size of the precursor ion.

B.5 Conclusion and Outlook

The scope of the QCxMS program was extended by developing and implementing positive-ion collision-induced dissociation (CID) run-modes. It is the first automatic quantum-chemistry-based protocol to compute unbiased EI and CID mass spectra without the need for pretabulated or database-driven algorithms. Utilizing the fast and accurate built-in semiempirical GFN2-xTB Hamiltonian enables calculations on molecules that are composed of elements with atomic numbers up to $Z = 86$ without the need of any third-party software. The capability of the program to describe extended fragmentation pathways and complicated rearrangement reactions was demonstrated on six diverse organic molecules. The new CID mode provides detailed insight into the physics of collision processes and can be used to analyze molecular stability and survival yields.

The main aim of this work was to set up a run-type that is capable to automatically calculate CID spectra comparable to experimentally measured spectra taken from standard databases. A precise one-to-one mapping of the experimental collision/thermal activation conditions and the theoretical setup could not be determined due to the lack of insight into the details of typical experiments. For this reason, the CID extension was developed in such a way that various experimental conditions can be considered. Therefore, different run-types were established.

The *forced activation* run-type was implemented to calculate a controllable amount of ion-gas collisions. A good approach for the calculation of single collisions was the stepwise increase of the collision energy. This procedure reveals the energy range where specific signals appear. The simulation of multiple collisions yields more fragments, requiring lower collision energies than with single collisions. Setting a high number of collisions enforces fragmentation, so the method generates useful spectra even if the starting energy is not spot-on. Taking into account further collisions between a fragment and neutral gas atoms, the intensity of subsequently formed fragments can be increased.

The *thermal activation* run-type yields fragmentation without simulation of explicit collisions. With this method, computational costs were low, but only statistical processes are considered, and the E_{int} scaling value has to be specifically determined in order to reduce under- or overfragmentation. Nevertheless, this method is recommended for a quick and cheap initial assessment when a proper energy scaling value can be estimated.

Finally, the automated *general activation* run-type was developed. In the first step, the internal energy of the protonated system is increased in an energy range that depends on the molecule size. This is followed by the simulation of multiple collisions between the precursor ion and neutral gas atoms. The number of collisions depends on the collision chamber properties, which also applies to the fragment–gas collisions. These features enable theoretical spectra to be computed that are in satisfying agreement to experimental spectra without the need for adjusting many technical simulation parameters. The combination of E_{int} scaling and explicit collisions provides a good accuracy-to-cost ratio, and hence, this setup can be generally recommended.

We conclude that the QCxMS program is able to calculate EI and positive-ion CID spectra based solely on the molecule input structure. Calculations for adducts like $[M+Na]^+$, $[M+H_2O]^+$, and $[M+NH_4]^+$ are possible. However, this was not yet tested. Calculations for multiple charged cations are also feasible, and the corresponding implementations are planned. An extension for the calculation of negative-ion CID spectra is straightforward and is currently being tested. The QCxMS software and documentation are available free of charge.¹⁰⁵

B.6 Supporting Information

B.6.1 Instrumentation details of the experiments

Table B.1: Individual experimental settings and specifications.

Chapter:	Molecule	Instrument	Type	Mode
4.2	Tetrahydrofuran	Waters Micromass Quattro Triple Quadrupole	LC-ESI-QQQ	MS2
4.3	2,6-Dichlorobenzamide	Bruker maXis Impact	LC-ESI-QTOF	MS2
4.4.1	Caffeine	API3000, Applied Biosystems	LC-ESI-QQ	MS2
4.4.2	3-Methylhistidine	API3000, Applied Biosystems	LC-ESI-QQ	MS2
4.4.2	Phenylephrine	API3000, Applied Biosystems	LC-ESI-QQ	MS2
4.4.2	Acetyl-sulfamethoxazole	Bruker maXis Impact	LC-ESI-QTOF	MS2

B.6.2 Relative energy ranking of protonated structures

Table B.2: Relative G_{tot} [kcal/mol] of protonated structures to most stable structure protomer #1 (0.0 kcal/mol) calculated using ENSO at PBEh-3c level. The ensembles were sorted according to the simulation temperature at 600 K. Identical and improper structures were sorted out.

Molecule	Protomer	300 K	600 K	1200 K
2,6-Dichlorobenzamide	# 2	15.36	14.92	14.22
	# 3	30.30	29.54	28.10
	# 4	33.63	33.20	32.36
Caffeine	# 2	6.74	6.73	6.35
	# 3	8.74	9.75	12.15
	# 4	16.98	17.08	17.44
3-Methylhistidine	# 2	20.54	20.44	20.83
	# 3	19.21	20.62	24.05
	# 4	24.72	24.39	23.94
	# 5	26.20	25.83	25.30
Phenylephrine	# 2	26.79	26.64	26.36
	# 3	27.85	27.75	27.53
Acetyl-sulfamethoxazole	# 2	7.52	9.72	13.79
	# 3	13.69	15.82	20.18
	# 4	17.52	18.02	19.88
	# 5	19.76	21.59	25.79
	# 6	22.46	23.87	26.28

B.6.3 Calculated alternative protomer spectra

Caffeine

Caffeine was calculated using the *general activation* run-type. All protomer structures provide distinct signal intensities under the same conditions.

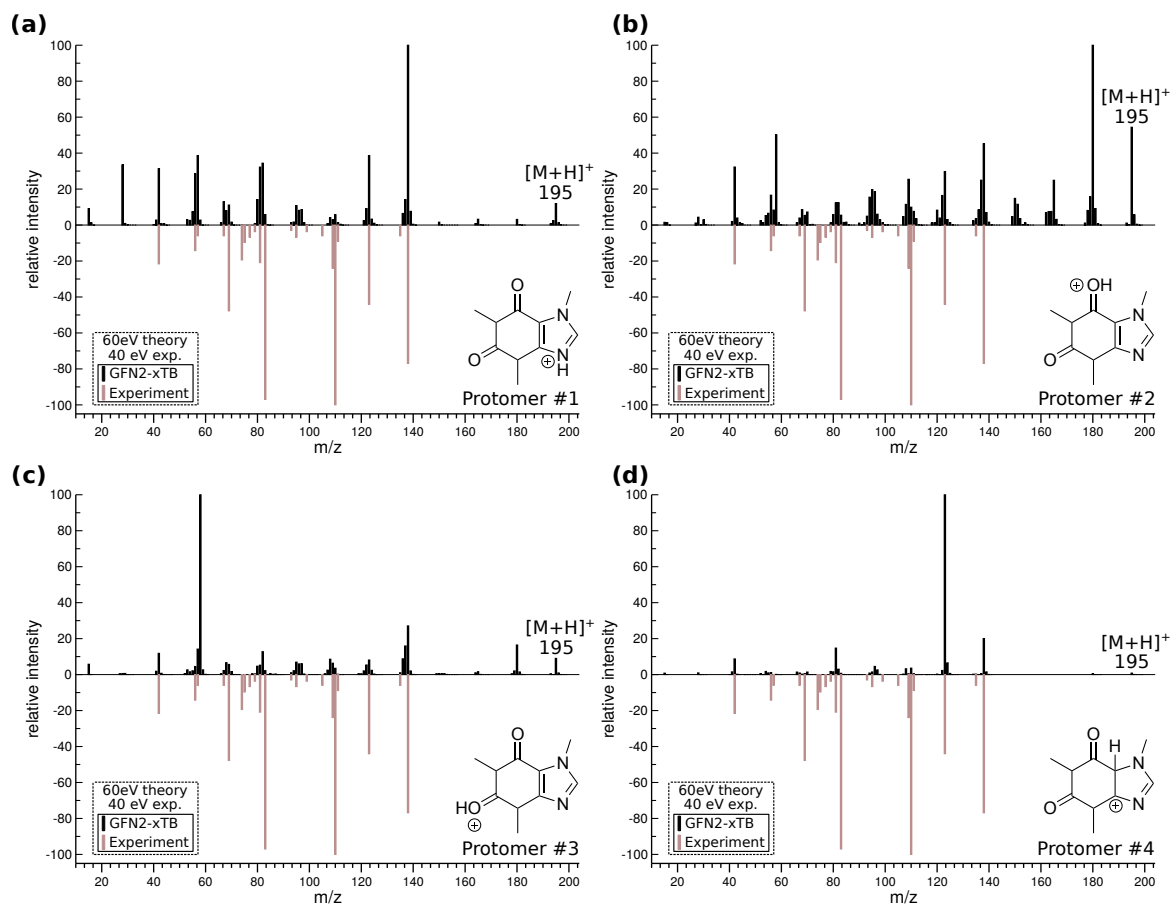


Figure B.7: Protomer spectra of caffeine calculated at 60 eV E_{LAB} using the *general activation* run-type (black) vs. 40 eV E_{LAB} literature spectra (red, inverted). Energy rankings can be found in table B.2.

3-Methylhistidine

Protomer #2 (figure B.8 (b)) and protomer #3 (figure B.8 (c)) produce almost the same spectrum due to the proximity of the neighboring nitrogen atom of the amine group and the carbon atom in the imidazole ring, promoting the proton transfer between these atoms during the equilibration of the structures at 600 K.

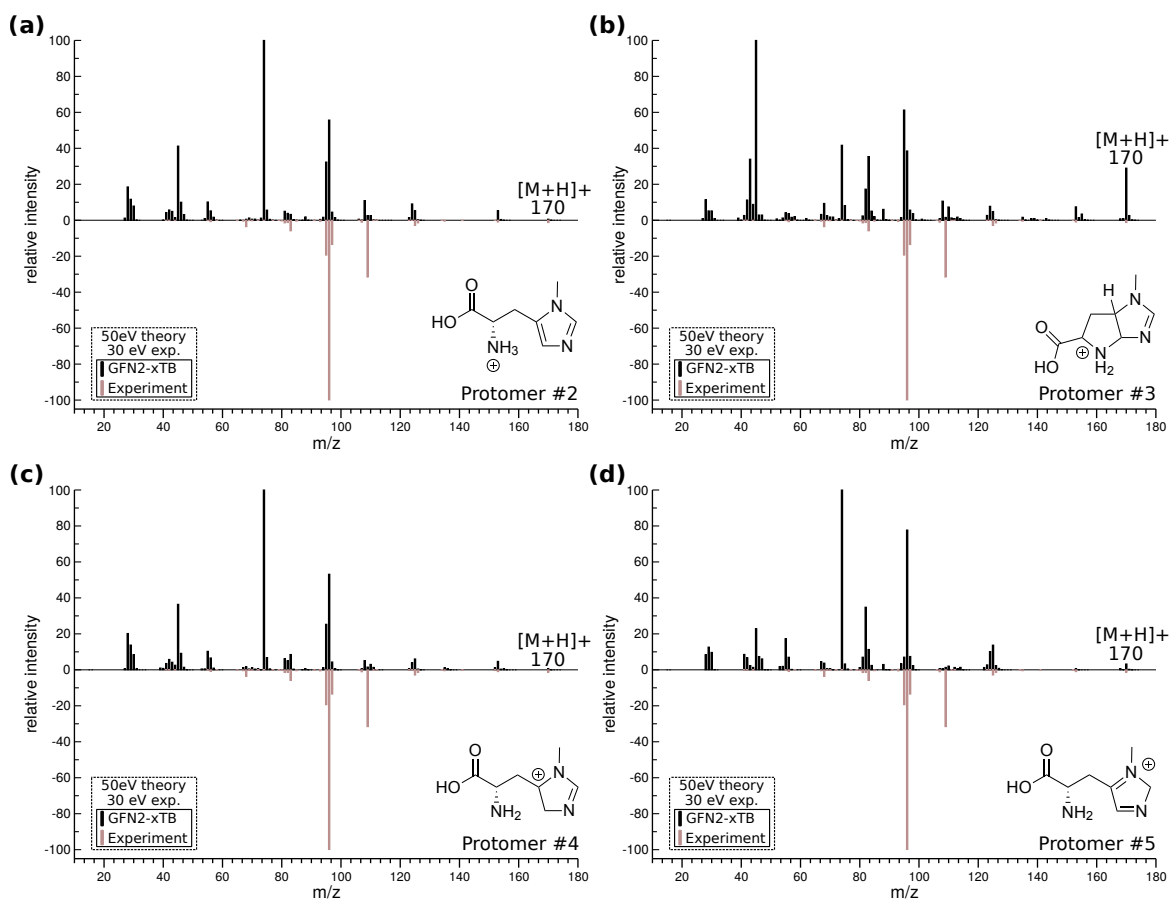


Figure B.8: Protomer spectra of 3-methylhistidine at 50 eV E_{LAB} using the *general activation* run-type (black) vs. 30 eV E_{LAB} literature spectra (red, inverted). Energy rankings can be found in table B.2.

Phenylephrine

Protomer #2 (figure B.9 (a)) rearranges into protomer #1 during the equilibration of the structure at 600 K, transferring the proton onto the nitrogen atom. The calculated spectra of both protomers are therefore identical. Protomer #3 (figure B.9 (b)) fragments almost exclusively into NC_2H_6^+ , which is assigned the entire charge.

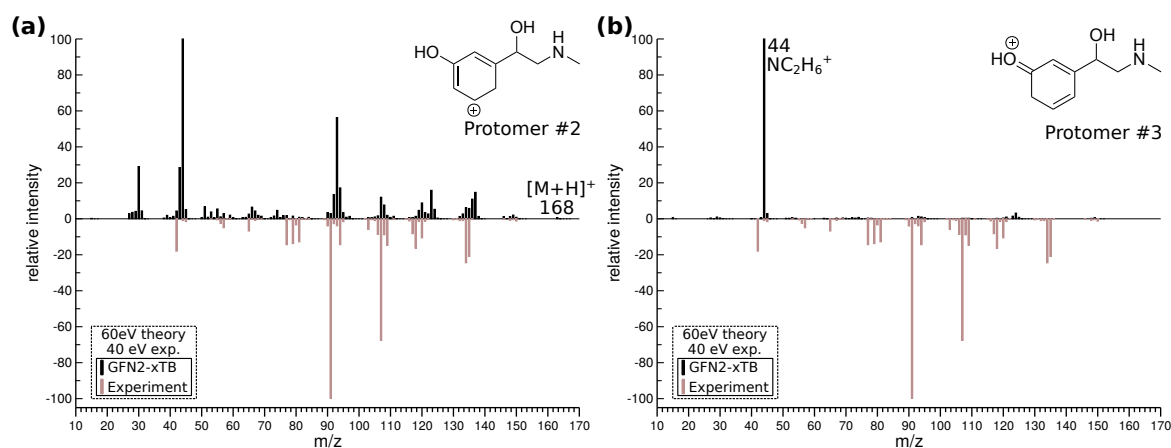


Figure B.9: Protomer spectra of phenylephrine at 60 eV E_{LAB} using the *general activation* run-type (black) vs. 40 eV E_{LAB} literature spectra (red, inverted). Energy rankings can be found in table B.2.

Piperidin

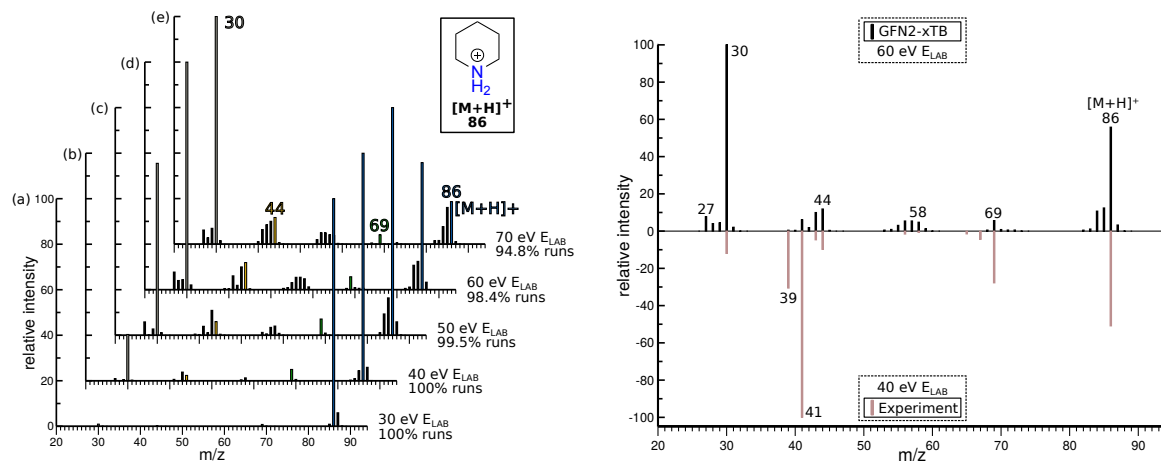


Figure B.10: (a) Calculated spectra of piperidin using single collisions at (1) 30 eV, 2) 40 eV, 3) 50 eV, 4) 60 eV, and 5) 70 eV E_{LAB} . The most concise fragments m/z 30, 44, 69, and 86 are colored for better traceability. (b) Calculated spectra of piperidin using single collision at 60 eV E_{LAB} vs. 40 eV E_{LAB} literature spectrum (red, inverted)²¹⁰. Instrument: API3000, Applied Biosystems.

Acetyl-sulfamethoxazole

The protomer structures #1, #2 and #4, #5 produce the same fragmentation patterns, demonstrated by the spectra of protomer #1 in figures B.11 (a), (b) and (d). This occurs either because the protomer structures rearrange into the structure of protomer #1 prior to fragmentation or the main fragmentation pathways produce fragments with the same mass (m/z 134, 150 and 198). The main bond fragmentation occurs between the S and N atoms. The proton for protomer #6 is on the side that remains charged, so here signals m/z 135 and 161 appear.

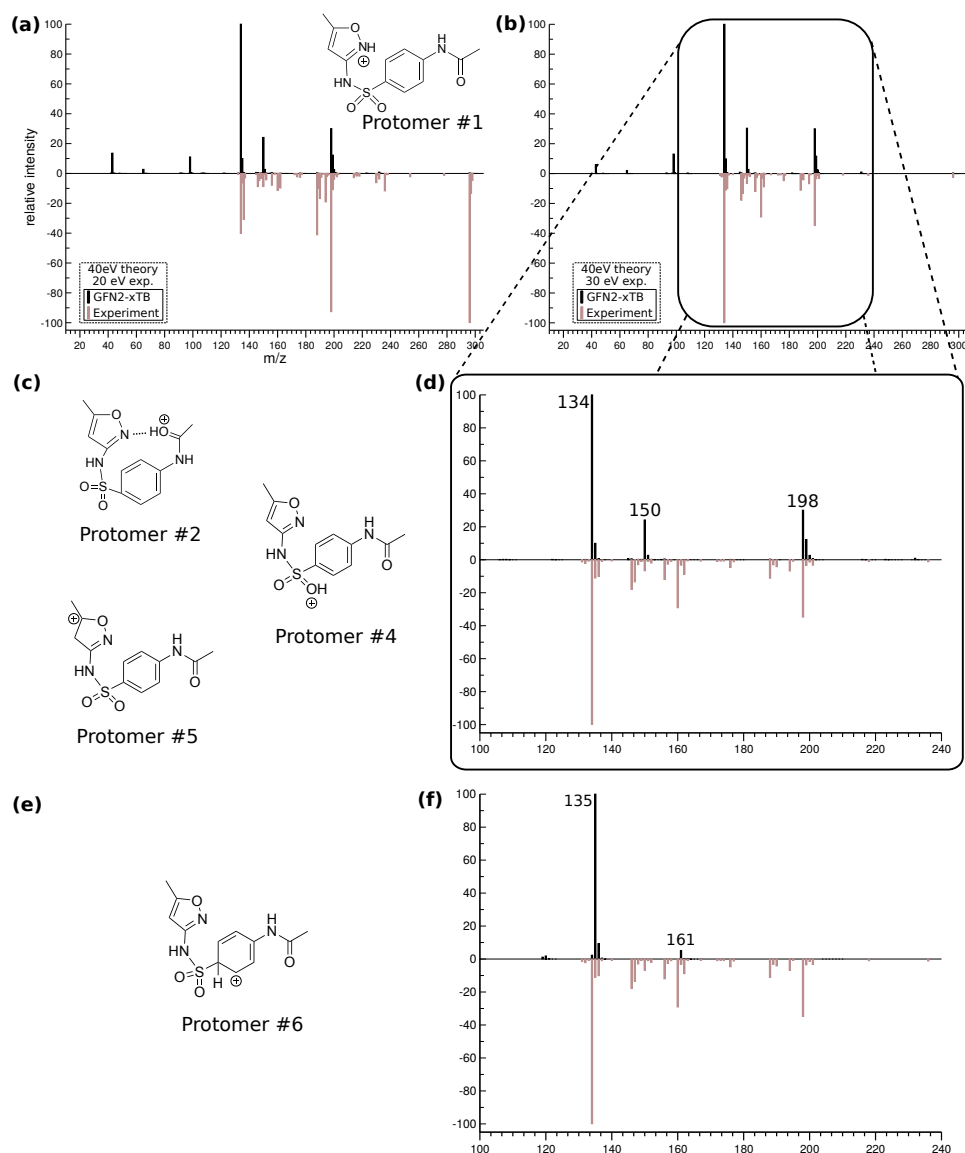


Figure B.11: Calculated (40 eV E_{LAB}) of the most populated protomer #1 compared to literature spectra at (a) 20 eV E_{LAB} and (b) 30 eV E_{LAB} of acetyl-sulfamethoxazole at 600 K. (c) Lewis structures of protomers #2, #4, and #5. The energy rankings can be found in table B.2. (d) Zoomed into the spectrum in between m/z 100 and m/z 240. (e) Lewis structure of protomer #6. (f) The spectrum of protomer #6 between m/z 100 and m/z 240.

B.6.4 Effects of the *general activation* run-type settings

Collision cell length

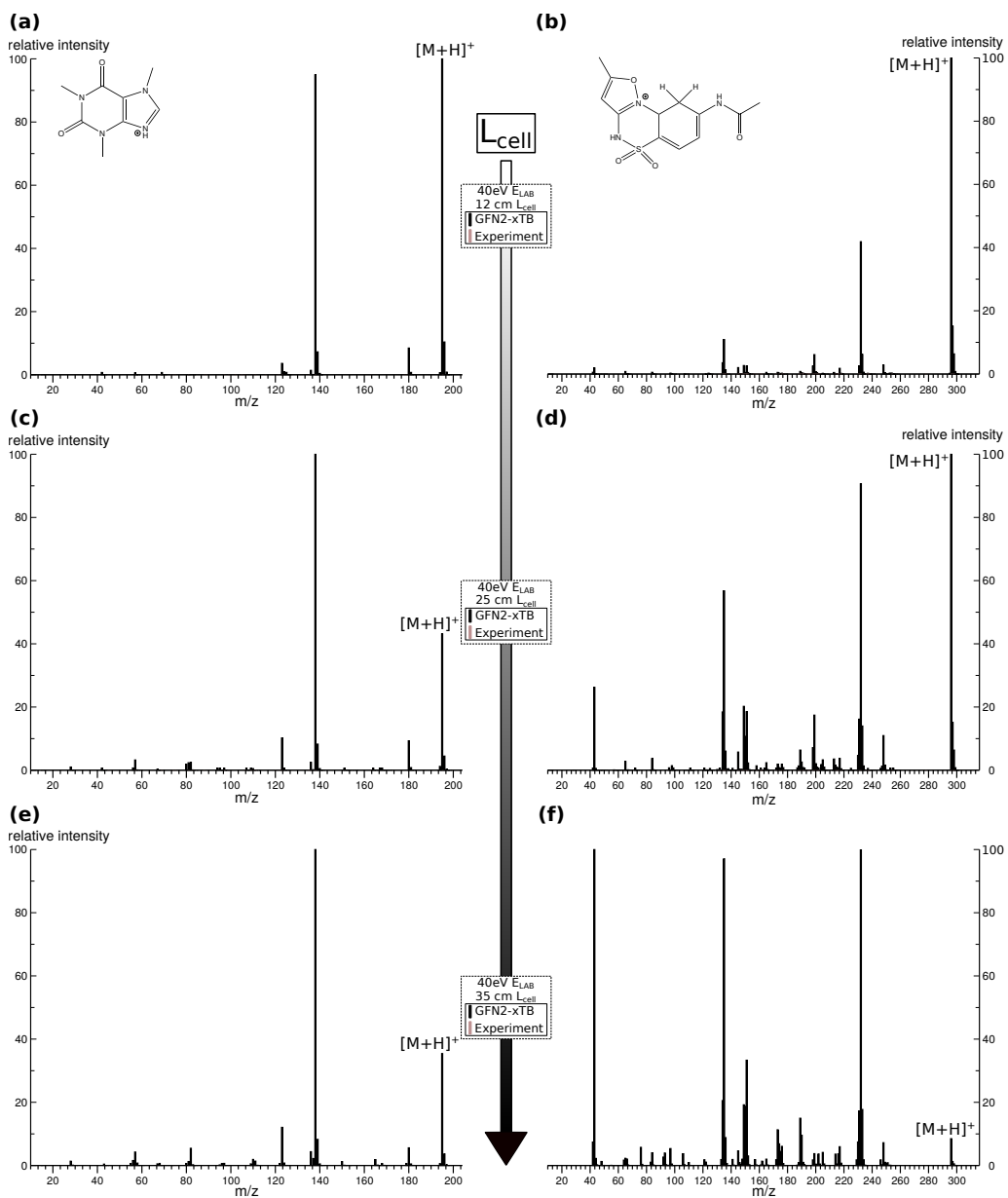


Figure B.12: Calculated spectra at 40 eV E_{LAB} with increasing collision cell length. Test structures were protomer #1 of caffeine ((a),(c) and (e)) and protomer #3 of acetyl-sulfamethoxazole ((b),(d) and (f)). (a),(b) $L_{Coll.cell}$ 0.12 m; (c),(d) $L_{Coll.cell}$ 0.25 m; (e),(f) $L_{Coll.cell}$ 0.35 m.

For a molecule in the size of caffeine, the increase in collision cell size between 0.25 m (figure B.12 (c)) and 0.35 m (figure B.12 (e)) becomes insignificant, as the velocity of the precursor ion and corresponding fragments becomes too low after a certain amount of collisions. Longer cells and thus

more collisions do not lead to a different result. Because larger and heavier structures retain more kinetic energy after a collision, the effect of collision cell length is more apparent in the spectra of acetyl-sulfamethoxazole. Figures B.12 (d) and (f) therefore demonstrate the effect the collision cell length has on the amount of subsequent fragment-gas-collisions, leading to an increased fragmentation into lower mass fragments.

E_{int} pre-scaling (ESI MD)

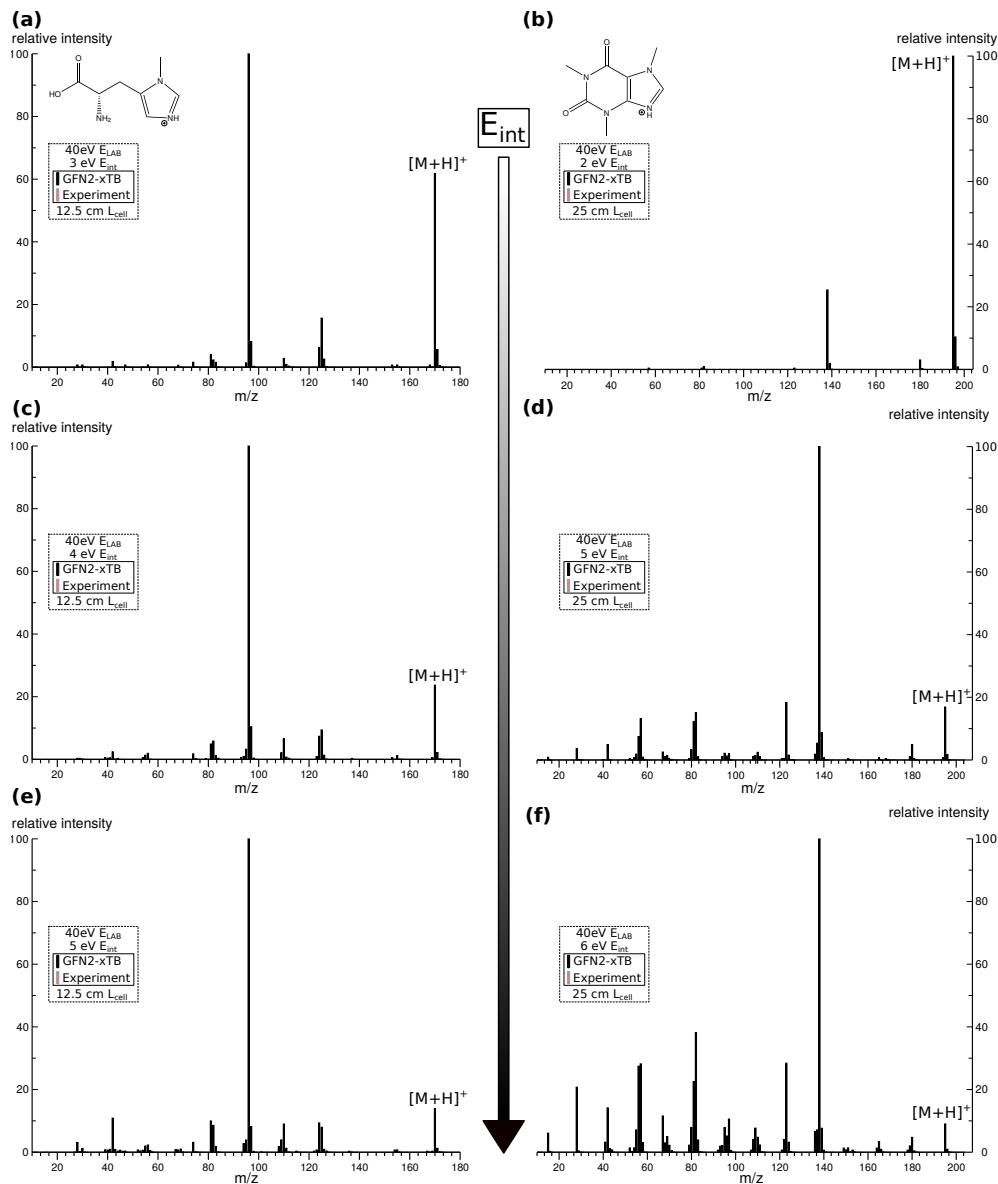


Figure B.13: Calculated spectra at 40 eV E_{LAB} with increasing E_{int} pre-scaling. Test structures were protomer #1 of 3-methylhistidine ((a),(c) and (e); $L_{Coll.cell}$ 12.5 cm) and protomer #1 of caffeine ((b),(d) and (f); $L_{Coll.cell}$ 25 cm). (a) E_{int} 3 eV; (b) E_{int} 2 eV; (c) E_{int} 4 eV; (d) E_{int} 5 eV; (e) E_{int} 5 eV; (f) E_{int} 6 eV.

The E_{int} threshold at which the precursor ions undergo fragmentation starts at 5 eV for 3-methylhistidine and 6 eV for caffeine. Greater E_{int} values will induce fragmentation behaviour comparable to results obtained with the *thermal activation* run-type. For this reason, the internal energy scaling is done depending on the precursor ion size.

B.6.5 Additional calculated spectra

Tetrahydrofuran

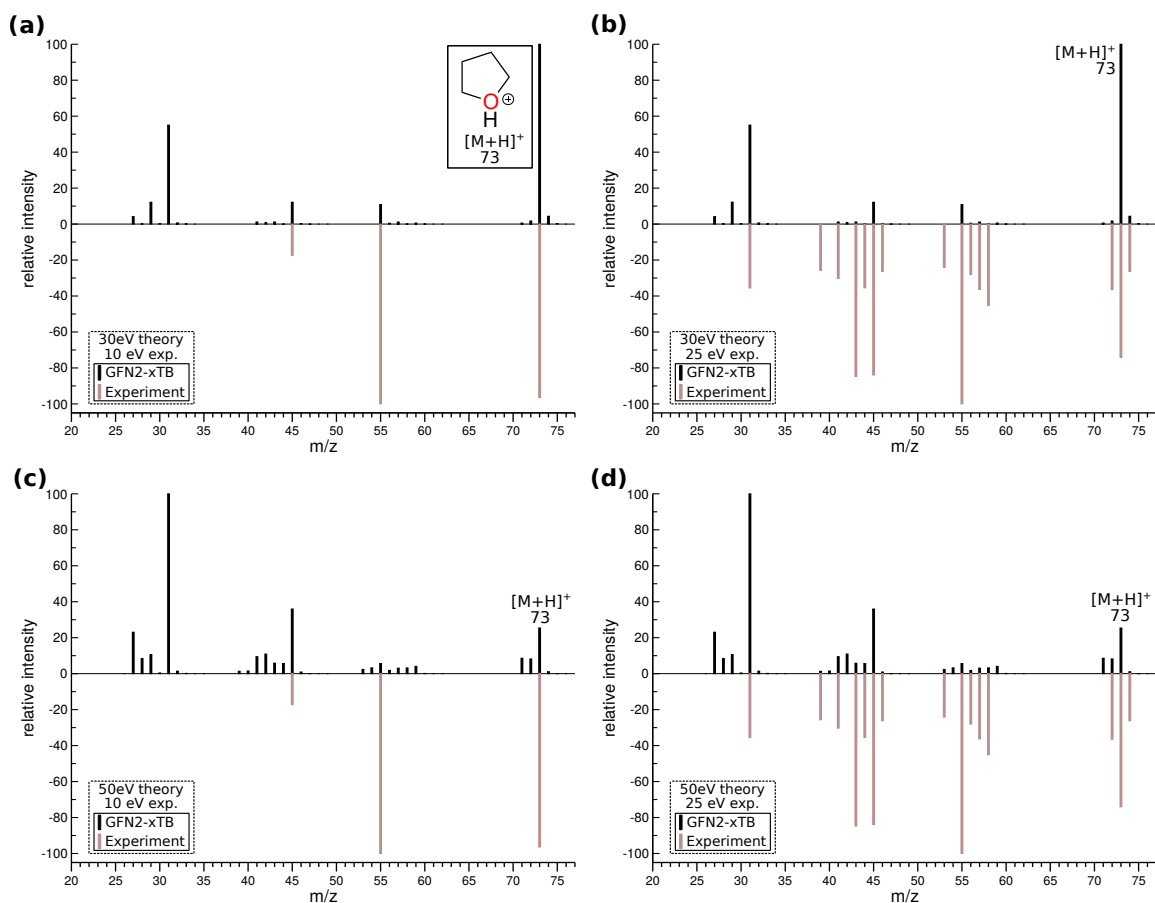


Figure B.14: Comparison of calculated single collision spectra (black) vs. literature spectra (red, inverted) of tetrahydrofuran. (a) 30 eV vs. 10 eV⁹⁷; (b) 30 eV vs. 25 eV⁹⁷; (c) 50 eV vs. 10 eV⁹⁷; (d) 50 eV vs. 25 eV⁹⁷.

4.5-Dihydroorotic Acid

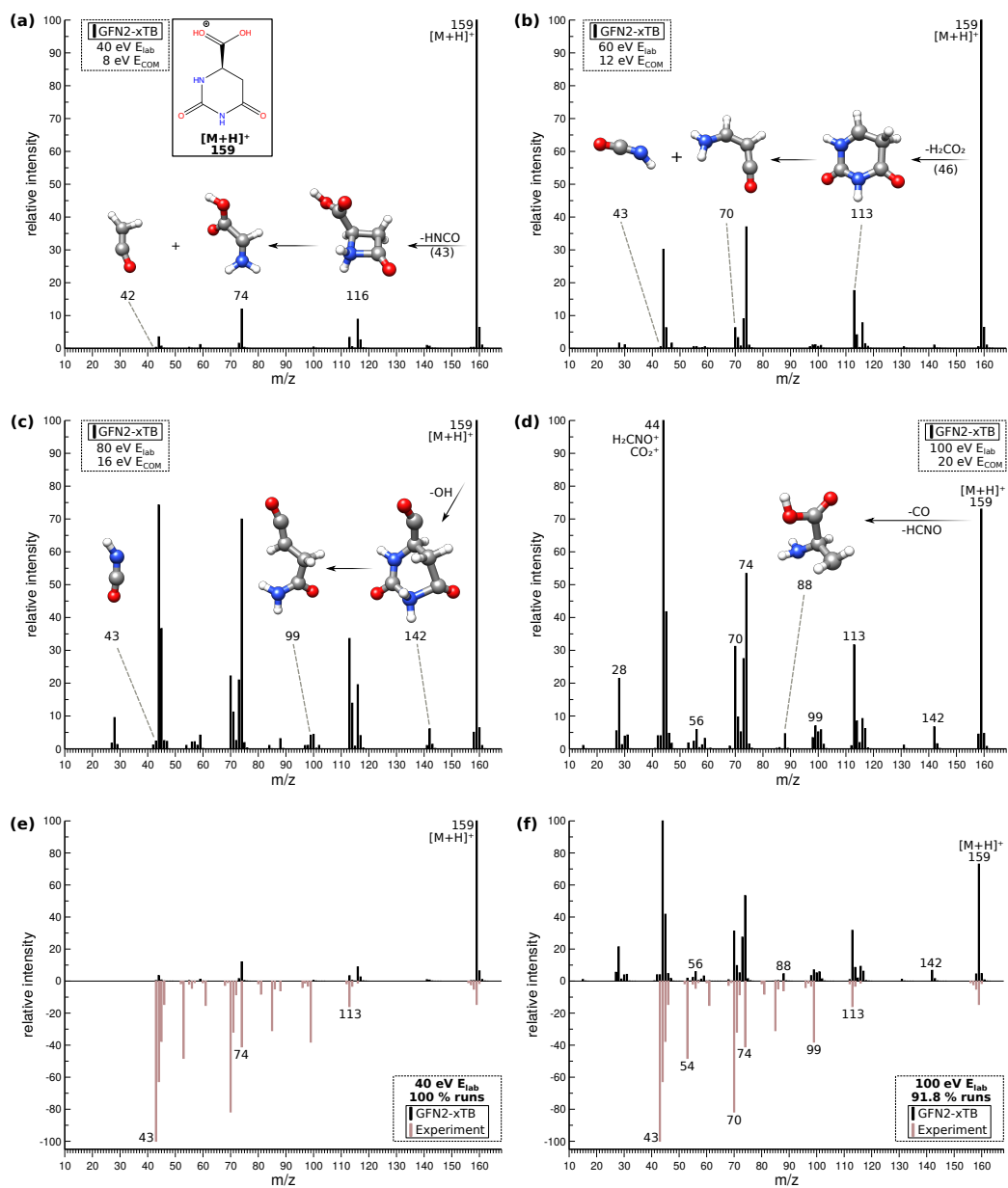
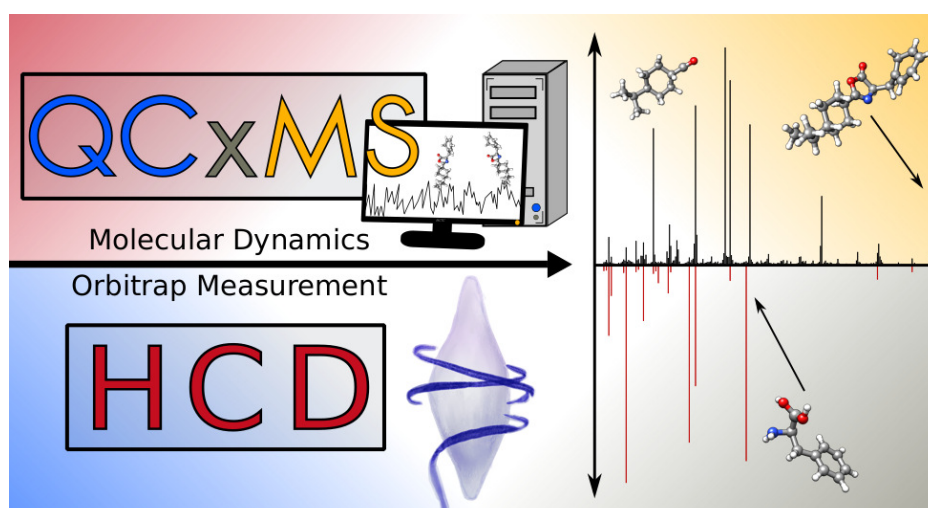


Figure B.15: Calculated spectrum of 4,5-Dihydroorotic-acid using single collisions at (a) 40 eV, (b) 60 eV, (c) 80 eV, and (d) 100eV E_{LAB} . (e) Comparison of the 40eV calculated (black) to the 40 eV literature spectrum (red,inverted)⁹⁷; (f) Comparison of the 100eV calculated (black) to the 40 eV literature spectrum (red,inverted).⁹⁷

Publication corresponding to chapter 5: Quantum Chemistry-based Molecular Dynamics Simulations as a Tool for the Assignment of ESI-MS/MS Spectra of Drug Molecules



Abstract In organic mass spectrometry, fragment ions provide important information on the analyte as a central part of its structure elucidation. With increasing molecular size and possible protonation sites, the potential energy surface (PES) of the analyte can become very complex, which results in a large number of possible fragmentation patterns. Quantum chemical (QC) calculations can help here, enabling the fast calculation of the PES and thus enhancing the mass spectrometry-based structure elucidation processes. In this work, the previously unknown fragmentation pathways of the two drug molecules nateglinide (45 atoms) and zopiclone (51 atoms) were investigated using a combination of generic formalisms and calculations conducted with the Quantum Chemical Mass Spectrometry (QCxMS) program. The computations of the *de novo* fragment spectra were conducted with the semi-empirical GFNn-xTB (n=1, 2) methods and compared against Orbitrap measured electrospray

ionization (ESI) spectra in positive ion mode. It was found that the unbiased QC calculations are particularly suitable to predict non-evident fragment ion structures, sometimes contrasting the accepted generic formulation of fragment ion structures from electron migration rules, where the “true” ion fragment structures are approximated. For the first time, all fragment and intermediate structures of these large-sized molecules could be elucidated completely and routinely using this merger of methods, finding new undocumented mechanisms, that are not considered in common rules published so far. Given the importance of ESI for medicinal chemistry, pharmacokinetics, and metabolomics, this approach can significantly enhance the mass spectrometry-based structure elucidation processes and contribute to the understanding of previously unknown fragmentation pathways.

C.1 Introduction

Mass spectrometry (MS) is a highly sensitive core analytical technique for researchers of various disciplines ranging from organic chemistry, medicinal chemistry to biochemistry and includes a vast amount of pharmaceutical, environmental, and forensic applications. The central analytical parameters are the structure and the amount of the analyte, reflecting qualitative and quantitative aspects of the analytics. Key to the fragmentation process is the nature of the analyte and the ionization technique applied.

Electron ionization (EI) is an ionization technique applied to volatile, preferably non-polar analytes. The process of the formation of the odd-electron (OE) radical cations $[M]^{•+}$ (with commonly 70 eV kinetic energy of the bombarding electron) leads to the subsequent fragmentation of the molecule in a highly reproducible manner.^{211–213} Hence, virtually classical fragmentation rules have been derived¹¹ and the fragment-rich mass spectra commonly are deposited in databases for straightforward compound identification.¹³⁰

Over the past decades, electrospray ionization (ESI) combined with collision induced dissociation (CID)²¹⁴ has evolved into the most commonly applied analytical ionization technique in medicinal chemistry and pharmacokinetics/pharmacodynamics (PK/PD) applications.^{215,216} This can be attributed to the predominantly polar nature of the analytes, which makes ESI highly complementary to EI.

The fragment assignment in the measured spectra and the resulting structure elucidation is commonly based on empirical rules, in which the (de-)protonated molecule is subsequently fragmented either by charge-migration (CMF) or charge retention fragmentation (CRF).^{31–34} However, while these rules can yield satisfying fragment ion assignments, fragmentation patterns are often observed that cannot straightforwardly be explained by CMF or CRF.^{31,93,94} Furthermore, competing fragmentation pathways can increase the level of difficulty for describing the “real” fragmentation process, leading to uncertainty about the correct spectral assignment.^{37,38} In earlier studies, molecule fragmentation mechanisms of $[M+H]^+$ ions were investigated either based on literature and personal experience, rather than experimental evidence.^{32,217,218} Due to the associated uncertainties resulting from using the generic rules, theoretical methods have to be developed that are able to support the interpretation of ESI-tandem mass spectrometry (MS/MS) fragmentation.

To date, it has become computationally affordable to use quantum chemical (QC) methods to calculate mass spectra. Most recently, quantum mechanical calculations found their way into the prediction of EI fragment spectra.⁵⁰ The structural assignment of fragment ions is based on molecular dynamics (MD) simulations in which the fragmentation of the molecular ion is calculated along multiple, cascading trajectories. Unlike to already established computational approaches,^{47,129,219–223}

the on-the-fly quantum chemical calculation of the potential energy surface (PES) enables an unbiased determination of the composition of fragments and intermediate structures and does not depend on already known fragmentation mechanisms or database spectra.^{50-54,56,89}

The usage of MDs to simulate collision induced dissociation (CID) reactions has also been investigated in other contributions.^{37,38,145,150,224,225} However, no direct comparison between measured and calculated signals was conducted, so that an overall agreement between experiment and theory could not be illustrated. Furthermore, the systems under consideration were rather small (size < 20 atoms; molecular mass < 170 Da), so that their complexity does not represent most common drug or macro-molecules.

In this report, the mass spectra of the two drug molecules nateglinide and zopiclone were investigated using the quantum chemical mass spectrometry program QCxMS (x = EI, CID) in positive ion CID mode. The calculated results were compared to experimental measurements produced with an Orbitrap Fusion ESI-MS/MS instrument leading to the complete fragmentation schemes of both drug molecules. In earlier work with QCxMS and its predecessor QCEIMS, it was shown that semi-empirical quantum mechanical (SQM) methods, especially the GFN1-xTB⁷⁰ and GFN2-xTB⁷¹ methods, can successfully be applied to calculate theoretical mass spectra, that agree reasonably well with database spectra.^{89,91} Using the automatic computation of fragment ion structures demonstrates the potential of QC calculations to become an important standard in matching experimental data and how to use this tool for fragmentation pathway interpretation.

C.2 Methodology

In the following considerations, the protonated molecules $[M+H]^+$ selected for collision-induced fragmentation are referred to as precursor ions and their fragments as product ions. For electrospray ionization (ESI), relative low internal (thermal excitation) energies are utilized, usually leading to ions with paired electrons referred to as even-electron (EE) or closed-shell ions. The resulting precursor ion is commonly fragmented in CID experiments.²¹⁴ On that account, the ionization and activation processes have to be treated separately. For the interpretation of the fragmentation routes, bond fissions are categorized as homolytic and heterolytic. In ESI-MS/MS, heterolytic cleavage is observed almost exclusively.²²⁶ The charge either remains on the initial atom or is transferred to the cleaved fragment, respectively.²²⁷

C.2.1 Benchmark Molecules

Two different drug classes were considered: the hypoglycemic agent nateglinide and the sedative zopiclone. Nateglinide (M = 317.4 Da, C₁₉H₂₇NO₃) is an anti-diabetic drug from the class of hypoglycemic agents, which lowers the glucose levels in the blood for the treatment of Diabetes mellitus. Zopiclone (M = 388.1 Da, C₁₇H₁₈ClN₆O₃) acts as an agonist for the neurotransmitter gamma-aminobutyric acid (GABA) receptor and works as a sedative. Both molecules differ in size and contain functional groups, which are representative of various drug molecules. Given that for the compound identification the existence of signals is of greater importance than the relative and absolute ion intensities, this work solely focused on comparing signals that exist in both, calculated and measured spectra, rather than the discussion of ion counts or each discrepancy between the theory and experiment. Details on the possible reasons for the diverging of some calculated signals are

discussed later.

C.2.2 Experimental Details

All MS experiments were performed using an Orbitrap Fusion ETD mass spectrometer coupled to an UltiMate 3000 RSLC nano LC system (Thermo Fisher Scientific GmbH, Bremen, Germany). The samples were ionized using an electrospray needle with a voltage of 3800 V (ES+) and a sheath gas pressure of 4 Arb. The vaporizer temperature was adjusted to 35 °C. The precursor ions were mass-selected using a linear ion trap and allowed to collide in an HCD collision cell with N₂ in a stepped collision energy mode at HCD collision energies of 15, 60, 120 %. The isolation width was set to m/z 0.7 with an AGC target of 2.0⁴ and a resolution of 60,000. Based on the S/N ratio, only counted fragments with a relative intensity > 1.5 % were considered. The measured compounds were isolated from film-coated tablets (Zopiclon-ratiopharm, Starlix Novartis) by crushing the solid and subsequent extraction with methanol for nateglinide and chloroform for zopiclone.

C.2.3 Ranking of Difficulty by Common Fragmentation Pathways

The degree of difficulty to describe the observed fragments was categorized by the feasibility to explain the shown pathways (Schemes C.1 – C.3) on rule-based fragmentation descriptions known from literature³¹ (Table C.1 & C.2, column 4). Of the illustrated fragments, the experimentalist was able to infer those assigned as *feasible* with reasonable efforts based on the common rules. A comparison of the proposed fragmentation reactions to the simulated trajectories confirmed the described pathways. For fragmentation reactions with higher complexity, the elaboration of rearrangements and cleavages occurring in CID experiments can be very time consuming and requires vast experience. In that regard, the possibility to utilize theoretical trajectories was highly expedient, so these fragments were categorized as *with QCxMS*. In retrospect, we were then able to corroborate the compliance of these fragmentation pathways with the classical fragmentation rules. The third category contains fragments designated as *only QCxMS*. The fragmentation pathways could solely be described with the help of the simulated trajectories. Their rearrangements and cleavages are untypical and differ from the common rules in the literature.³¹

C.2.4 Computational Details

Before the QCxMS simulations were conducted, protonation of the targeted species was achieved using the automated QC-based protonation protocol⁸⁵ of CREST⁸⁶ at the GFN2-xTB level of theory. The most populated protonated structures inside a 20 kcal/mol energy range were re-optimized using density functional theory (DFT) at the PBEh-3c level¹⁹² to guarantee the correct energy ranking. To provide the reader with a better understanding of the QCxMS protocol, the mechanics are discussed in short in the following. For an in-depth discussion of the QCxMS program and its working mechanics, the reader is referred to the original publication.⁹¹

Basically, QCxMS runs in multiple steps.

1. **Ground state sampling:** the input structure is equilibrated at 600 K on a 15 ps MD trajectory with a timestep of 0.5 fs.

2. **Set-up:** 1,250 structural snapshots were taken along a 30 ps MD trajectory for sampling of the conformational space.
3. **Production runs:** conducting massively parallel calculations with the snapshot structures as starting geometries.

In the production run step, the collisional activation is simulated. The collision simulations imply the following conditions for each starting geometry:

- rotation along the Euler axes to guarantee different impact sites.
- adding rotational energy of $k_B T/2$ per principal axis, with k_B for the Boltzmann constant and T as the temperature of the ion.
- scaling of the internal energy prior to the collision along a 1 ps MD trajectory to standard distributed values between 4 and 8 eV.
- velocity scaling of $[M+H]^+$ according to the acceleration potential of 10 eV E_{COM} .

Subsequently, the collisions between $[M+H]^+$ and neutral He gas atoms with a randomized collision angle (impact parameter b) are simulated for each product ion run. The collisions transform kinetic into the internal energy of the ion. If the critical energy E_0 is reached, statistical and non-statistical fragmentation of the molecular ion occurs. To induce sufficient dissociation of $[M+H]^+$, the collision process had to be repeated multiple times.

Unfortunately, the exact number of collisions in the experiment cannot be determined. For the best reproduction of the measured spectrum, an automatic run-mode was developed in QCxMS, that circumvents a tedious trial-and-error approach to determine the correct number of collisions and corresponding collision energies (see the original publication⁹¹). It uses the kinetic gas theory as an indication to solve this problem, in which the number of collisions can be calculated through the collision cell length, collision gas pressure and collisional cross-section. For the calculations conducted in this work, the parameters were set to 1.25 cm cell length with 0.132 Pa collision gas pressure at a temperature of 300 K, which are in agreement with the program's default values. These values provide an estimate of the number of collisions, but do not affect any other simulation condition.

The calculations were conducted by QCxMS version 5.0 using the xTB version 5.8.1 on Intel Xeon E3-1270@v5 3.6 GHz computer cores. The forces were calculated with GFN1-xTB for zopiclone and GFN2-xTB for nateglinide and applied with a timestep of 0.5 fs in the MD simulations. CREST version 2.8.1 using GFN2-xTB was employed for the protonation of the molecule. Automatic re-ranking was done by the ENSO¹⁹¹ script version 2.0.2. DFT calculations were performed using ORCA¹²¹⁻¹²³ version 4.2.1.

C.2.5 Discrepancies Between Calculations and Experiments

Some considerations have to be taken into account when comparing computed spectra with experimental measured results.

Depending on the design of the instrument used, the amount of internal energy of the precursor ion after ionization is commonly unknown. While efforts can be undertaken to determine this value, calibration and measurement need a lot of work and are thus not routinely conducted. The ionization

source can influence the overall fragmentation process, depending on the “hardness” of the ionization, even sometimes leading to in-source fragmentation. Further contributing factors involve details like the setup of the used quadrupoles, collision cell, and means of detection. Generalizability and reproducibility depend strongly on these factors, making a direct comparison even between different instruments complicated.

From the computational point of view, reasonable MD simulations run over shorter time scales (picoseconds) than experimental measurements (milliseconds). To conduct appropriate MD simulations on this time scale, semi-empirical methods must be employed, sacrificing the accuracy of calculations for the sake of the computational cost. This can lead to the wrong description of the PES and thus false fragment structures and/or charge assignments. Furthermore, decomposition effects that occur after a long timescale can be missed by the simulations, while short-lived fragments might be over-represented.

In the current version of QCxMS, the precursor ions are accelerated once at the beginning of the simulation, so the loss of kinetic energy in multiple collisions is not compensated for, as it is done in modern MS instruments (here: Orbitrap Fusion). It was found that the agreement to measured spectra is increased drastically when subsequent collisions between created fragments and neutral gas atoms (fragment-gas-collisions (*fgc*)) were considered. Not accounting for re-acceleration thus might lower the *fgc* collision energies in the calculations, leading to an underrepresentation of lower mass fragments.

C.3 Results and Discussion

C.3.1 Nateglinide

The most populated protomers of nateglinide ($M = 317.4$ Da, $C_{19}H_{27}NO_3$) were determined using ENSO at 600 K. By the QCxMS calculations, it was found, that the protomers in the lowest 20 kcal/mol energy window contributed most to the final spectrum, while other protomer spectra did not provide additional information on the overall fragmentation behavior. The structures considered are displayed in Figure C.1.

Using the classical protonation formalism, structures I and III can be created, while the structure of protomer II is non-intuitive. The structure of protomer II was re-optimized at PBEh-3c and PBE0/def2-TZVP-D4 levels to gain further insight. It was found that after protonation of the carboxylic acid group, a ring formation occurred. Single point calculations at PW6B95-D3/QZVPP level confirmed that the ring formation stabilizes this structure by 16.2 kcal/mol and is thus populated in the given energy window. However, by increasing the internal energy through collisions, higher energy structures can become populated and a ring-opening occurs before the

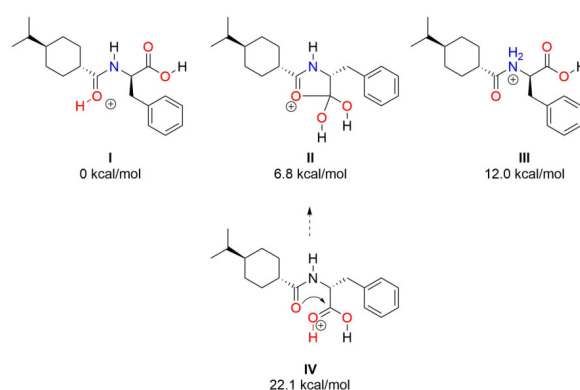


Figure C.1: Protomers I–IV of nateglinide and their relative energies (kcal/mol) related to the protomer with the lowest protonation energy as obtained by PBEh-3c DFT calculations.

fragmentation takes place. Since the “mobile proton theory” gives strong evidence that protomers can easily rearrange into each other in the high-temperature regime of mass spectrometry experiments, and several fragments originate from more than one protomer^{172,228,229}, it is likely that the protomer structures displayed in Figure C.1 can simultaneously be present in the measurements. Through the close vicinity of the heteroatoms in this structure, proton migration at high internal energy is promoted and the individually simulated spectra of the different protomers I, II, and III provided similar fragments with varying intensities (see Figure S1–3).

In Figure C.2, the measured spectrum is compared to the calculated spectrum of nateglinide. The calculated spectrum is composed of the combined and normalized results from the calculations of all three protomer structures.

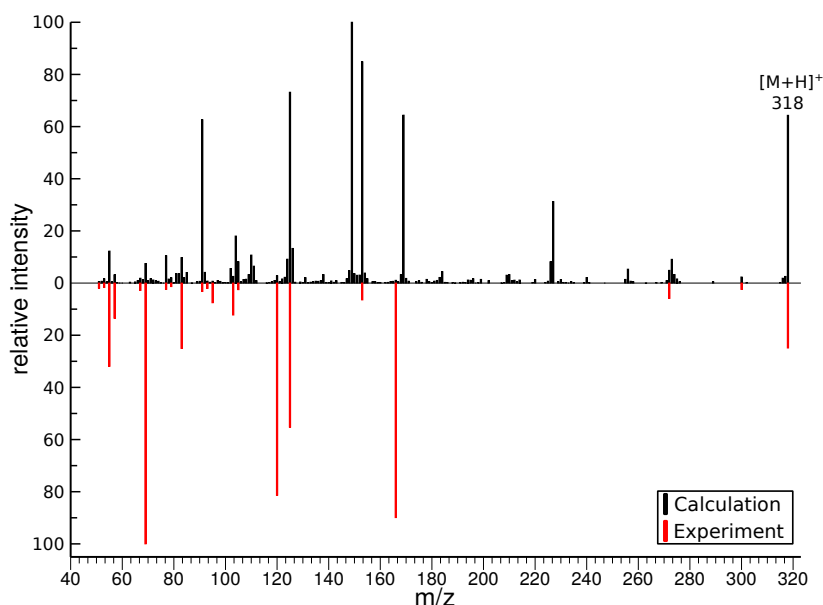


Figure C.2: Comparison of the simulated ESI-MS/MS spectrum computed by QCxMS@GFN2-xTB with the experimental spectra (red, inverted) of nateglinide. The spectra for all three considered protomer structures are averaged with equal weight.

A detailed comparison of the signals showed that the measured fragments could be reproduced with 100% agreement by the calculations. The corresponding complexities, as described earlier, are displayed in Table C.1. The comparison of the simulated spectra to the measured results display significantly more peaks, including differences in intensities. The additional peaks are no isotopologs. As mentioned before, differences between calculated and experimental spectra are to be expected. For the calculation of large molecules, a factor to consider is that the distribution of the collision energy into the ion’s internal energy can take a long time due to the high number of degrees of freedom. Due to the shorter simulation times than reaction times and the use of SQM methods for the calculations, an underrepresentation of slower dissociation events and an overestimation in survival rates of non-physical artifacts can be observed in the calculated spectrum. This also causes short-lived fragments and intermediate structures in the simulations, leading to the small, unmatched signals in the theoretical spectrum. Using a higher level of QC theory, deficiencies in the PES calculation could be alleviated; however, this is currently not feasible for molecules of this size. Nevertheless, due to the fragment coverage of the calculation (Table C.1), the molecule’s fragment structures can be unequivocally assigned.

A formulation of a classical fragmentation scheme would start from the protomers formed by the protonation of heteroatoms bearing non-binding electron pairs. The outlined fragmentation pathways for nateglinide are shown in Scheme C.1. The CMF and CRF mechanisms of the three protomer

Table C.1: Fragment list of nateglinide. Fragments **measured** by ESI-MS/MS with intensity >1.5% are listed according to their molecular mass. Fragments **simulated** by QCxMS are marked. Classification of the **rule-based fragmentation** ranking as described before.

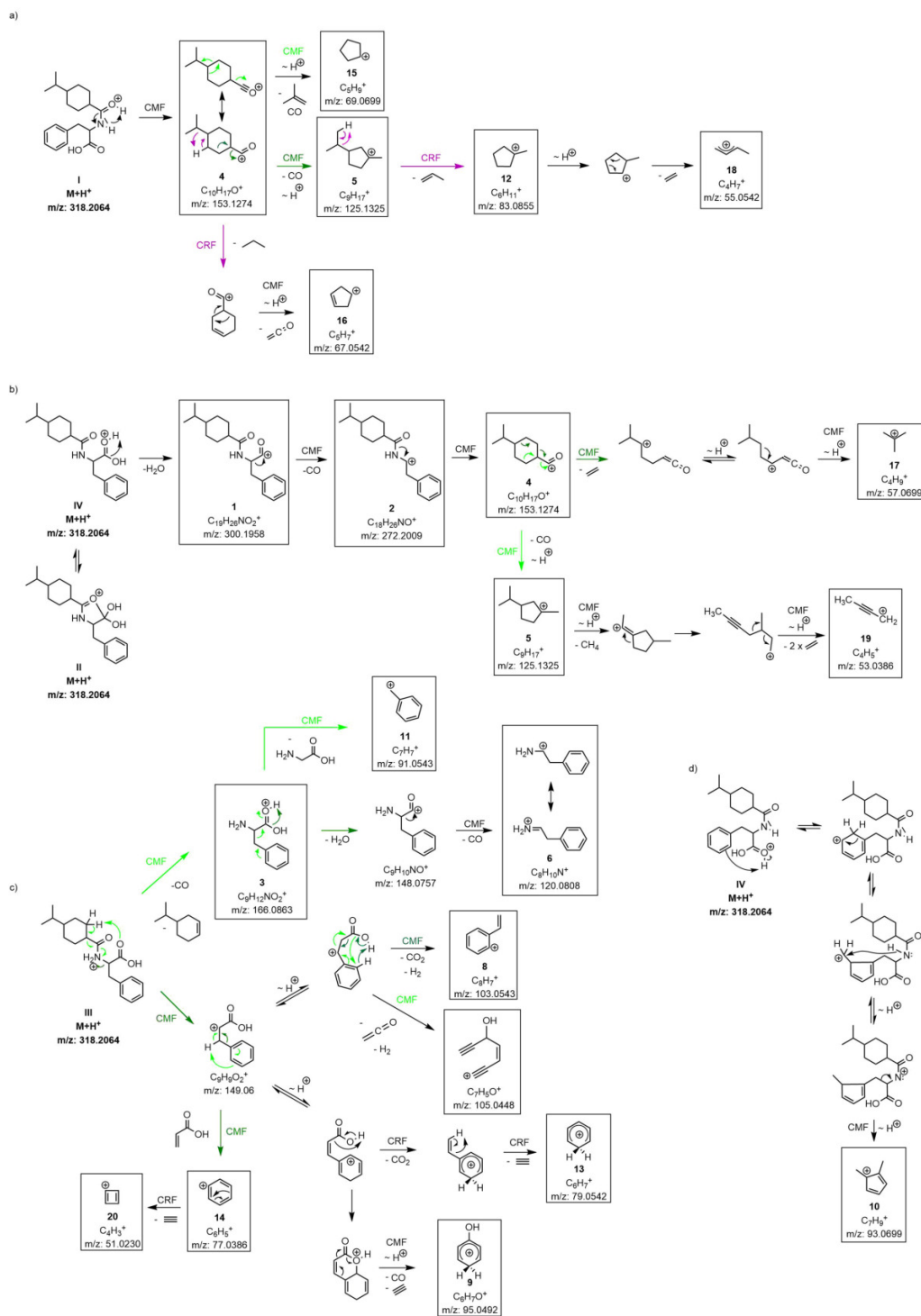
Fragment No.	Measured fragments (m/z)	Simulated	Ruled-based fragmentation
1	300.1962	yes	feasible
2	272.2011	yes	feasible
3	166.0862	yes	feasible
4	153.1274	yes	feasible
5	125.1325	yes	feasible
6	120.0808	yes	feasible
7	105.0448	yes	feasible
8	103.0543	yes	only QCxMS
9	95.0492	yes	with QCxMS
10	93.0700	yes	only QCxMS
11	91.0543	yes	feasible
12	83.0857	yes	feasible
13	79.0544	yes	feasible
14	77.0387	yes	feasible
15	69.0700	yes	feasible
16	67.0544	yes	feasible
17	57.0700	yes	feasible
18	55.0544	yes	feasible
19	53.0387	yes	with QCxMS
20	51.0231	yes	feasible

structures (I–III) were drawn to retrace every step of fragment formation.

The description of the fragmentation routes for difficult fragments (with QCxMS) proved to be particularly challenging and heavily relied on the support of the calculated trajectories. For this reason, the fragmentation routes for the two fragments **8** (m/z 103) and **10** (m/z 93) could not be derived from common mechanisms published so far. With the help of the calculated trajectories, the untypical reaction pathways (only QCxMS) could now be elucidated. Ion **8** (m/z 103) undergoes a 1,6-elimination of H_2 , resolving the aromaticity of the benzene ring (Scheme C.1c). This reaction mechanism is not part of the classic fragmentation rules; mostly 2,4-eliminations of H_2 are observed. The trajectory can be found in the Supporting Information as “*elimination_reaction_ion8.mp3*”.

Fragment **10** (m/z 93) is obtained in the calculations via proton migration of protomer II from the protonated oxygen to the aromatic ring. Again, aromaticity is resolved, which according to generic rules is not favored and therefore would not be formulated (Scheme C.1d). The corresponding trajectory can be found in the Supporting Information as “*elimination_reaction_ion10.mp3*”.

The breakdown of the aromatic compounds may be related to the high temperatures under which the reactions take place. The available energy makes reactions possible which, under “normal” circumstances, would have too high of a reaction barrier. These findings should be considered when establishing fragmentation patterns of aromatic compounds.



Scheme C.1: Proposed fragmentation pathways of nateglinide. Fragmentation according to a one-channel fragmentation (black arrows) or branched fragmentation pathways (highlighted in coloured arrows; CMF green, CRF purple). $\sim \text{H}^+$ means proton migration. Boxed fragments were experimentally detected. Unboxed structures are “snapshots” on calculated trajectories and are displayed for clearer retracing of the reaction pathways; they are not global minima on the potential energy surface. **For matters of clarity, the neutral fragment has not been depicted in all cases.**

C.3.2 Zopiclone

For zopiclone ($M = 388.1$ a, $C_{17}H_{18}ClN_6O_3$), CREST computed five protomer structures that were populated in the 20 kcal/mol free energy range at 600 K (see Figure C.3).

The combined spectrum of the protomers of zopiclone calculated with QCxMS is compared to the experimentally measured spectrum in Figure C.4. The QC calculations identified 23 out of the 25 fragments generated by the mass spectrometer, resulting in a coverage of the fragment pattern of 92 % (Table C.2, column simulated). To illustrate the mechanistic details of the proposed fragmentation pathways for zopiclone, common fragmentation rules were used (Schemes C.2 & C.3).

Most of the observed fragments of zopiclone are formed by highly complex rearrangements and proton migrations (marked as *with QCxMS*, Table C.2). To determine a detailed illustration of these pathways by classical fragmentation schemes, considerable amounts of time and experience are required. These efforts extend even further if multiple protomer structures have to be considered. Due to the increased internal energy that the ions receive from the collision processes, the system is in a high-temperature regime. With high internal energies, reaction barriers between the protomers can be exceeded which leads to proton mobility between the different starting structures.

The exact population in this temperature regime is unknown and lacks detailed research, so it is not straightforward to weigh the influences of the protomers on the final spectrum based on free energies alone. Calculations on the different protomer structures revealed similarities in the main fragmentation behavior, thus it is to be expected that zopiclone creates a tautomeric network between the protonation sites. However, the simulated fragmentation pathways of protomer VI showed the best conformity with the generic rules, which allowed for the construction of the majority of fragmentation pathways based on this protomer. The CID of zopiclone showed two dissociations that do not undergo common CRF or CMF reactions 6, 9 (marked as *only QCxMS*, Table C.2). The first example is complex **21**. It is formed by the ionic fragment and the zwitterionic piperazine fragment derived from the same precursor ion, which was stable enough to be detectable (m/z 345, Scheme C.3d). Such ion-dipole complex formations in the gas phase were described in several studies^{230–235} and occur when the energy threshold for the direct precursor decomposition is not reached. The complex formation enables reorientations, thereby allowing for transfers or reactions between parts of the molecule otherwise not possible due to the remoteness in the precursor molecule.²³² The capability of QCxMS to indicate ion-dipole complexes is of great value when interpreting and elucidating fragmentation pathways.

To confirm this finding, several independent MDs were calculated in which the two fragment units were placed close to each other and propagated over a 5 ps trajectory. The calculations were performed at a BLYP-D3/def2-SVP level at a temperature of 500 K. It was found that different H-bond formations are possible – the actual formation depends strongly on the protonation site and the spatial arrangement

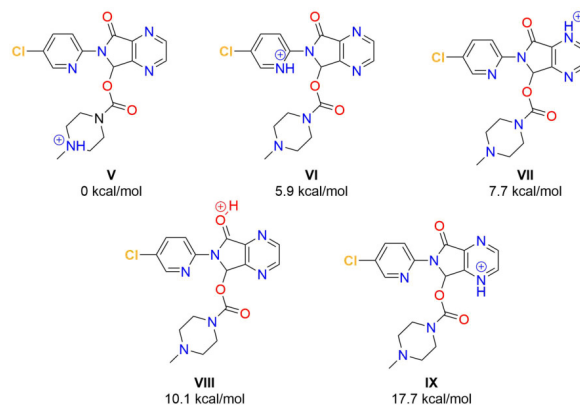
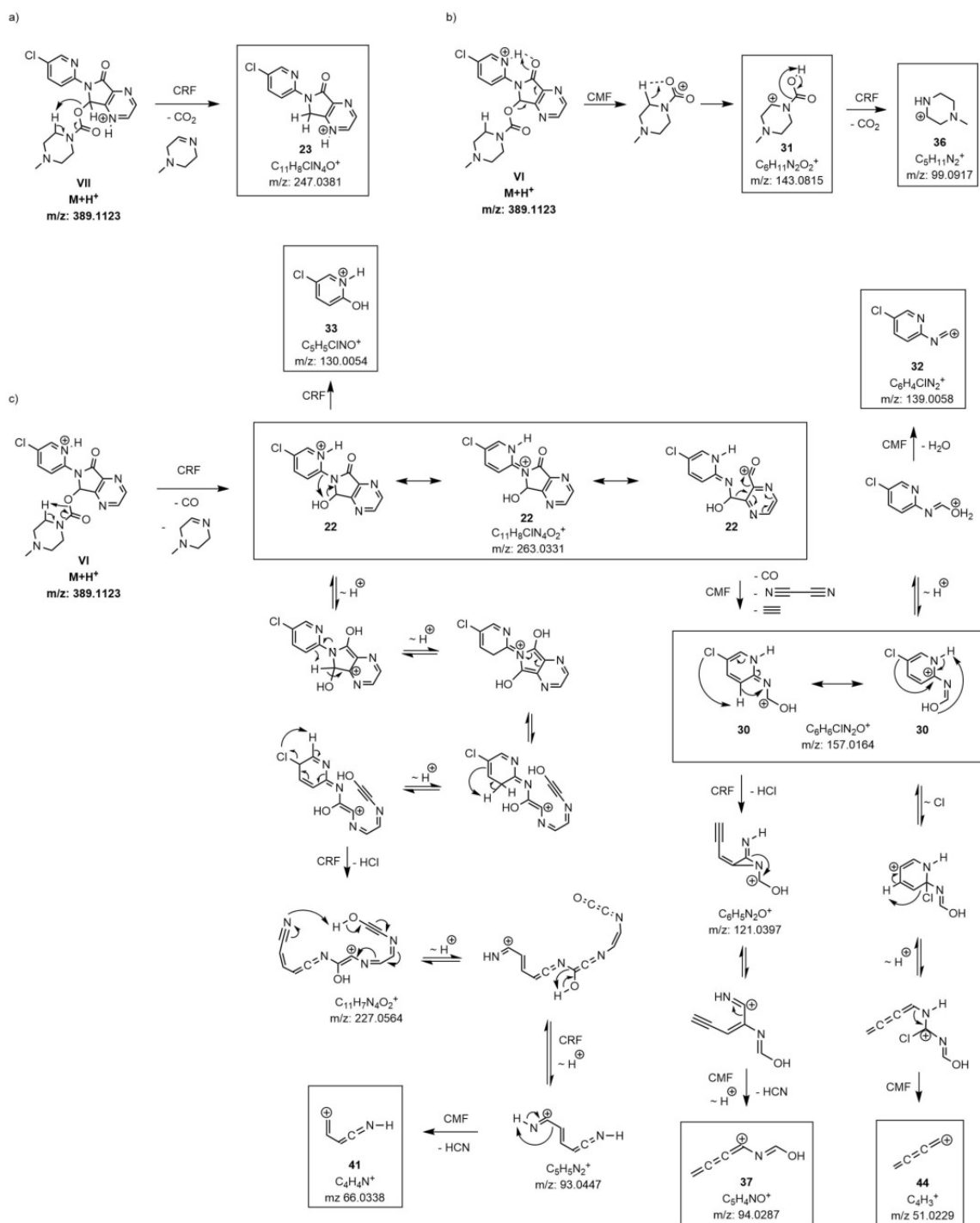
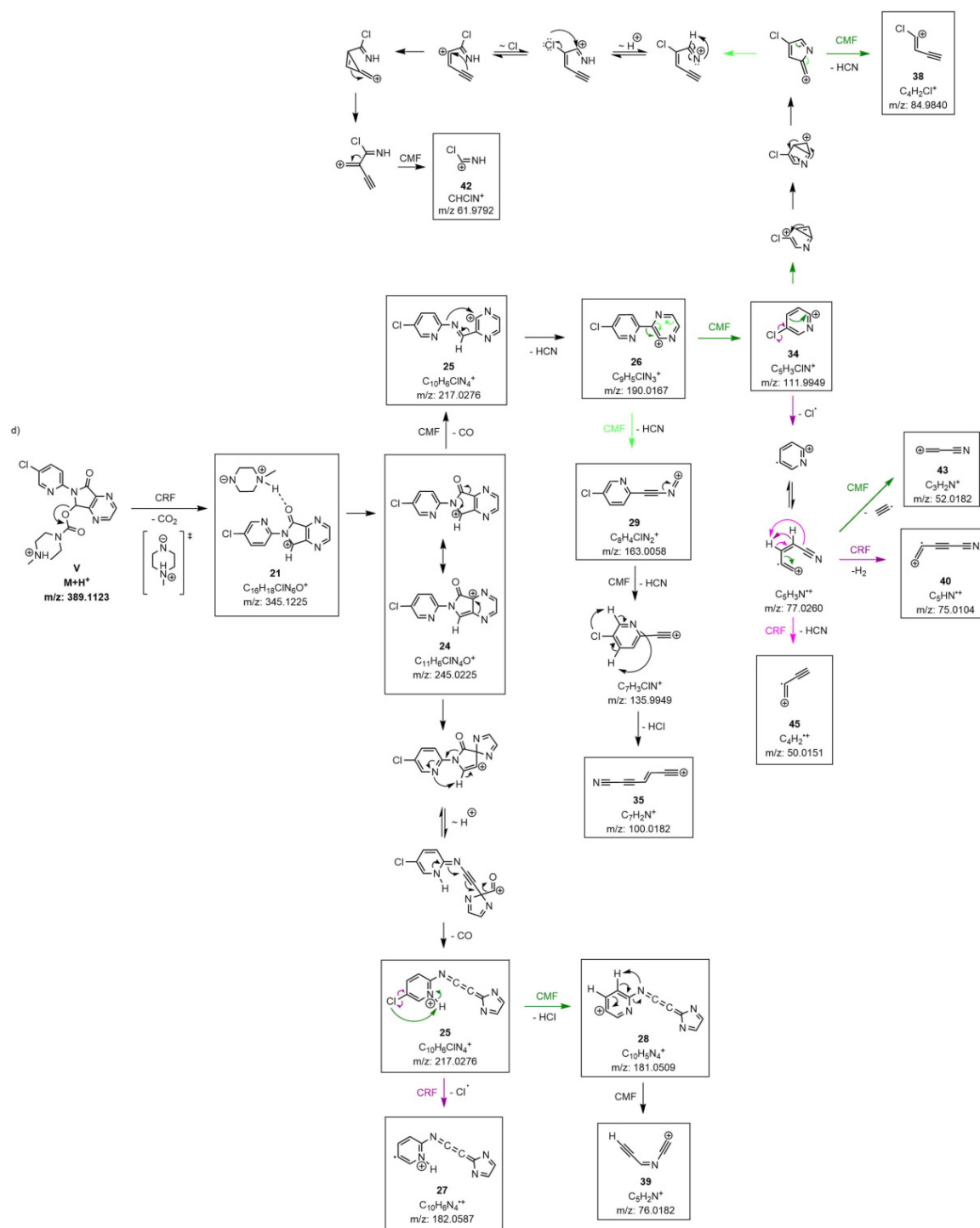


Figure C.3: Protomers of zopiclone and their energies (kcal/mol) related to the protomer with the lowest protonation energy as obtained by PBEh-3c DFT calculations.



Scheme C.2: Proposed fragmentation pathways of zopiclone – part 1. Fragmentation according to a one-channel fragmentation (black arrows) or branched fragmentation pathways (highlighted in coloured arrows; CMF green, CRF purple). $\sim \text{H}^+$ means proton migration. Boxed fragments were experimentally detected. Unboxed structures are “snapshots” on calculated trajectories and are displayed for clearer retracing of the reaction pathways; they are not global minima on the potential energy surface. **For matters of clarity, the neutral fragment has not been depicted in all cases.**

Appendix C Quantum Chemistry-based Molecular Dynamics Simulations as a Tool for the Assignment of ESI-MS/MS Spectra of Drug Molecules



Scheme C.3: Proposed fragmentation pathways of zopiclone – part 2. Fragmentation according to a one-channel fragmentation (black arrows) or branched fragmentation pathways (highlighted in coloured arrows; CMF green, CRF purple). $\sim H^+$ means proton migration. Boxed fragments were experimentally detected. Unboxed structures are “snapshots” on calculated trajectories and are displayed for clearer retracing of the reaction pathways; they are not global minima on the potential energy surface. **For matters of clarity, the neutral fragment has not been depicted in all cases.**

of the reaction partners to one another at the start of the simulation. Calculations of the H-bond binding energies might reveal the preferred binding sites and the binding situation in this context.

Overall, the underlying formation of fragment **21** can be described using quantum-mechanical principles, but due to the high number of possible binding situations and the extension of the underlying work, a detailed description of the exact binding situation is beyond the scope of this publication.

The second fragment formation addressed as *only* QCxMS (**31**, m/z 143, Scheme C.2b) is another example indicating ion-dipole complex formation. To describe the observed structure by common reaction mechanisms, a nucleophilic attack of a piperazine-hydrogen on the oxygen of the carbamate

group must occur, which results in a neutral cleavage of the main part of zopiclone. The attack of a hydride on a heteroatom, bearing already a negative partial charge, is contradictory. However, it can be resolved by looking into the simulated trajectories with their corresponding calculated energies, showing hydrogen transfers most likely induced by ion-dipole interactions.

Remarkably, the theoretical calculations for fragments derived from **34** and **26**, predict the occurrence of the radical cations **45** and **27**, respectively. This seems plausible as the experimental occurrence of radicals has been described for heteroatoms of higher atomic numbers, e.g. SO_2CH_3 and Cl .²²⁶ It is important to notice, that the fragmentation pathway described in Scheme C.2c is not sufficiently reproduced by the initial simulations. The relatively short simulation times and high collisional energies led to an under-representation of the corresponding signals in the calculated spectrum. However, the proposed fragmentation pathway could be confirmed by using fragment **22** as starting structure for a separate QCxMS calculation. Here, all displayed dissociation events except for **33** (m/z 130, Scheme C.2c) and **40** (m/z 75, Scheme C.3) were described by the calculations and the resulting fragments could be generated sufficiently. The rearrangement reactions of fragment **33** and fragment **40** could not be calculated by the MD simulation. The formation of fragment **33** involves a substitution of a C–N bond with a C–OH bond, most probably involving an intermediate four-ring formation between the neighboring C–N–C–O atoms. Most likely the underlying QC method might not be suitable to calculate this specific rearrangement reaction. Increasing the level of theory might solve this problem, but was not investigated in the course of this work. The same holds for fragment **40**, but in addition, the involved H rearrangement must succeed over a large distance, which reduces

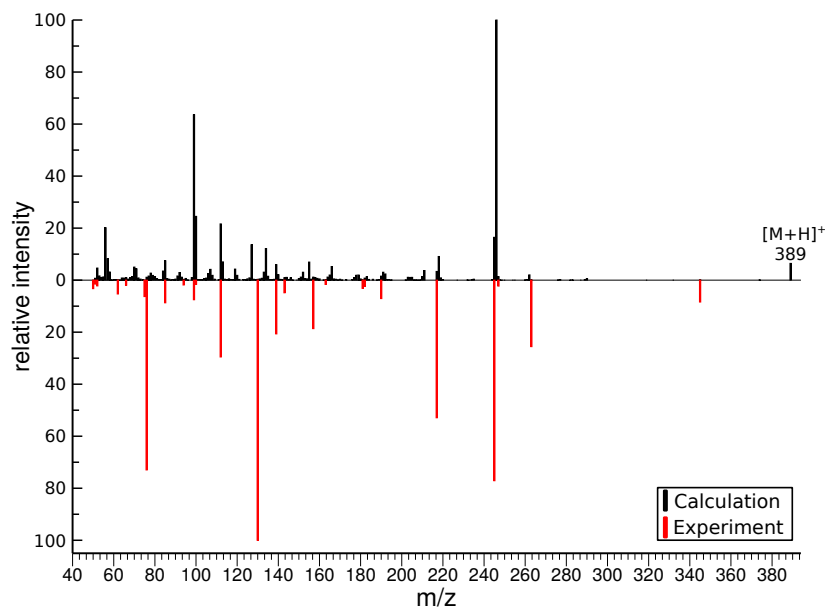


Figure C.4: Comparison of the simulated ESI-MS/MS spectrum computed by GFN1-xTB (energy window 20 kcal/mol) with the experimental spectra (red, inverted) of zopiclone. The spectra for all five considered protomer structures are averaged with equal weight.

Table C.2: Fragment list of zopiclone. Fragments **measured** by ESI-MS/MS with intensity >1.5% are listed according to their molecular mass. Fragments **simulated** by QCxMS are marked. Classification of the **rule-based fragmentation** ranking as described before.

Fragment No.	Measured fragments (m/z)	Simulated	Ruled-based fragmentation
21	345.1227	yes	only QCxMS
22	263.0333	yes	feasible
23	247.0384	yes	feasible
24	245.0228	yes	feasible
25	217.0276	yes	feasible
26	190.0167	yes	feasible
27	182.0587	yes	with QCxMS
28	181.0509	yes	with QCxMS
29	163.0057	yes	feasible
30	157.0163	yes	with QCxMS
31	143.0815	yes	only QCxMS
32	139.0058	yes	feasible
33	130.0054	no	feasible
34	111.9949	yes	feasible
35	100.0183	yes	with QCxMS
36	99.0918	yes	feasible
37	94.0288	yes	with QCxMS
38	84.9841	yes	feasible
39	76.0183	yes	with QCxMS
40	75.0105	no	feasible
41	66.0340	yes	with QCxMS
42	61.9793	yes	with QCxMS
43	52.0183	yes	with QCxMS
44	51.0231	yes	with QCxMS
45	50.0152	yes	with QCxMS

its possibility to occur in the simulations. Although fragments **33** and **40** could not be retraced by the simulations, the high congruence of the fragment patterns allows for a clear assignment.

C.4 Conclusion and Outlook

In this work, the previously unknown collision-induced fragmentation pathways of the two drug molecules nateglinide and zopiclone have been completely described by using a combination of generic formalisms and quantum chemical calculations conducted with the Quantum Chemical Mass Spectrometry (QCxMS) program in positive ion mode. The most populated protomers within a defined energy window of 20 kcal/mol were used as starting points to calculate corresponding fragment spectra, allowing for multiple reaction pathways. Utilizing molecular dynamics simulations, the MS/MS spectra could be calculated and compared to the experimental spectrum, achieving an excellent coverage > 90 % of the measured signals. Fast and easy access to the calculated trajectories was an important complementation to the classical interpretation of the final fragmentation pathway, which substantially supported the experimentalist's interpretation of the CID spectra, while simultaneously

reducing the time of structure elucidation significantly. Uncommon fragmentation pathways were identified which encourage further investigation and might lead to an expansion of what to consider as “typical fragmentation pathways”.

The investigated structures were of typical size for common bioactive molecules and contained functional groups which are representative for several drug classes. In that context, it is worth mentioning, that the complexity of the fragment spectra rapidly increases with the molecular mass and number of heteroatoms. Given the enormous structural diversity, e.g. of drug-like synthetic molecules, a thorough interpretation of fragment spectra without the aid of QC calculations becomes unrealistic and is eventually based on improper assumptions. For the first time, the presented work showed how the automatism implemented in QCxMS combined with the fast calculations conducted with GFN-xTB can be used to completely elucidate molecules of this size and complexity. In summary, QCxMS proved to be a valuable tool to facilitate the detailed illustration of fragmentation mechanisms or even enables a description of pathways in the first place. The comprehensive applicability of this method will allow the description of a vast majority of different molecular classes, which will be topic of future investigations.

Other chemical dynamics simulations for CID fragmentation analysis published so far were created via licensed programs, which limits its use to a restricted group of researchers. Therefore, a guiding principle for the development of QCxMS was its open accessibility to foster scientific exchange across disciplines. Furthermore, it may support the research and teaching segment in the understanding of fragmentation pathways and the distribution of this knowledge.

C.5 Supporting Information

C.5.1 Calculated Spectra Using QCxMS

Nateglinide

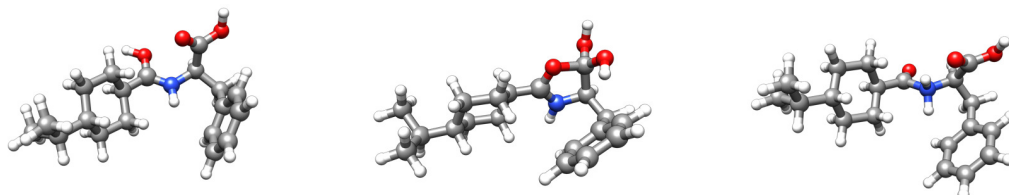


Figure C.8: 3D structures of the different nateglinide protomers I (left), II (middle) and III (right).

The spectra of the three different protomers of nateglinide show the same fragments with differing intensities.

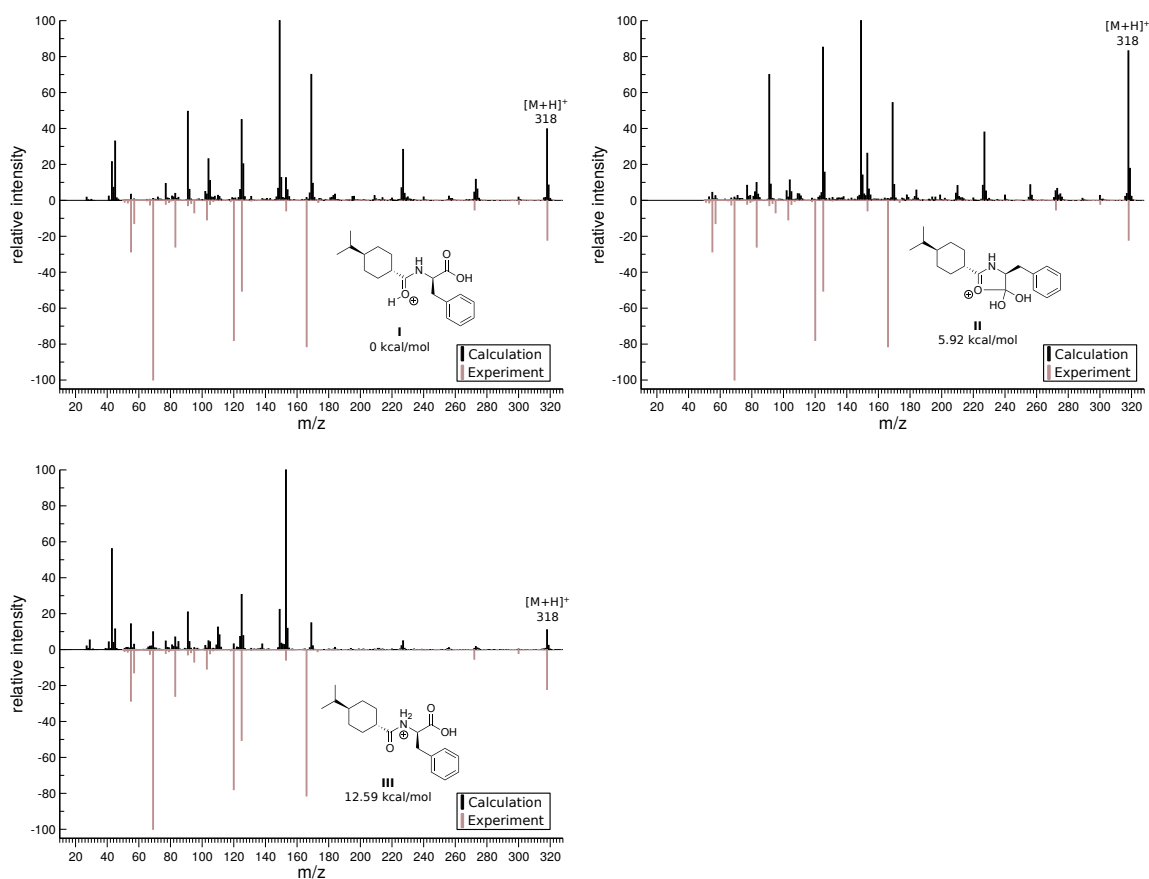


Figure C.9: QCxMS simulated spectra of the three nateglinide protomers I–III.

Zopiclone

The spectra of the five different protomers of zopiclone show the same fragments with differing intensities.

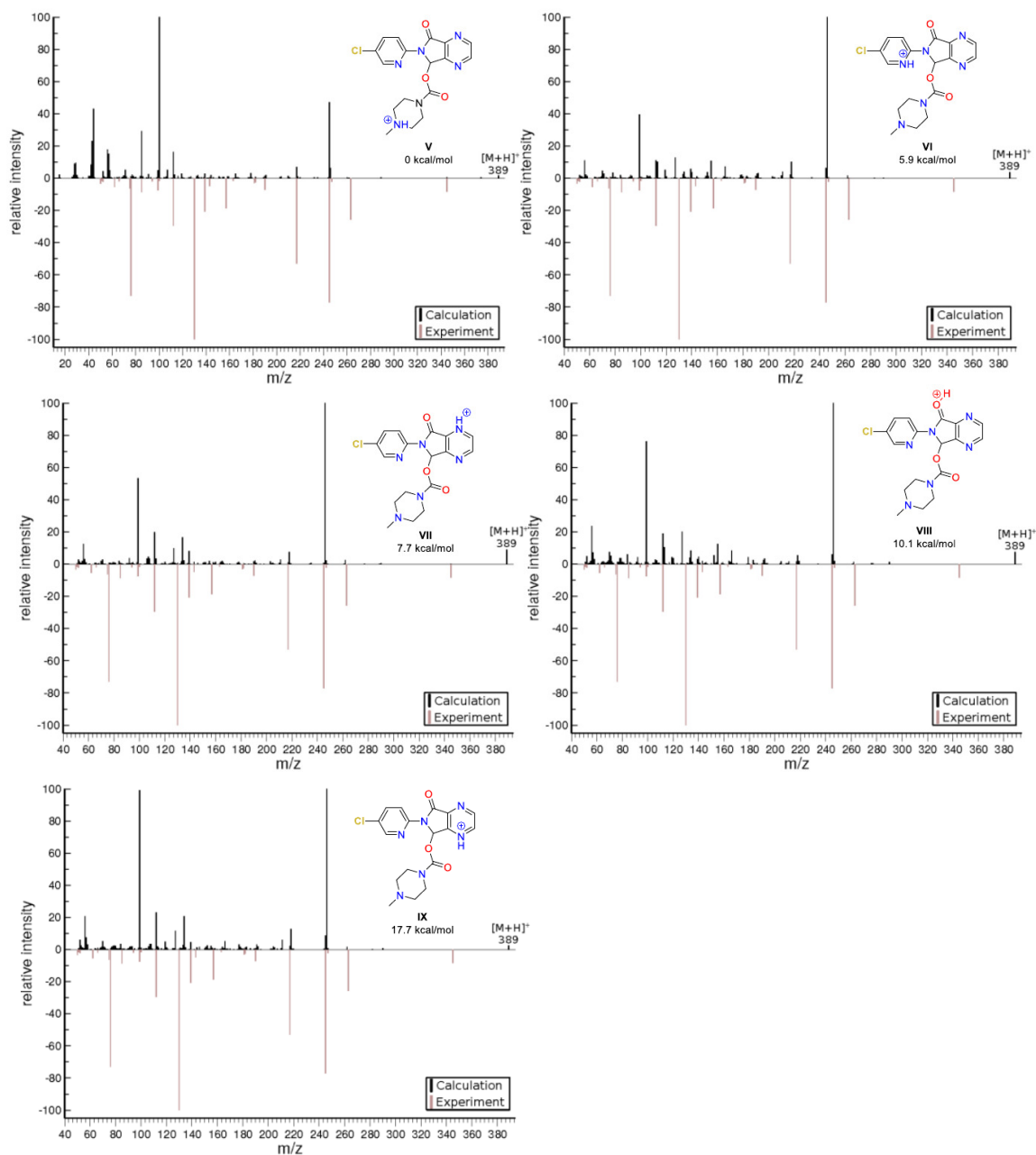


Figure C.10: QCxMS simulated spectra of the five zopiclone protomers V-IX.

C.5.2 Mass accuracy of the measured fragments

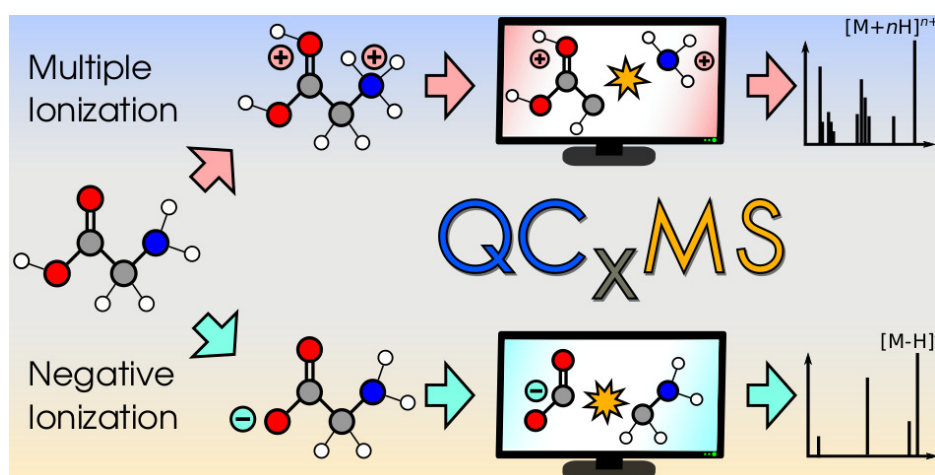
Table C.3: Comparison of the m/z values of the measured fragment of nateglinide with the calculated values based on the sum formula.

Fragment	m/z of measured fragments	Calculated m/z	Δ mass	Sum formula
1	300.1962	300.1958	0.0004	$C_{19}H_{26}NO_2^+$
2	272.2011	272.2009	0.0002	$C_{18}H_{26}NO^+$
3	166.0862	166.0863	0.0001	$C_9H_{12}NO_2^+$
4	153.1274	153.1274	0	$C_{10}H_{17}O^+$
5	125.1325	125.1325	0	$C_9H_{17}^+$
6	120.0808	120.0808	0	$C_8H_{10}N^+$
7	105.0448	105.0335	0.0113	$C_7H_5O^+$
8	103.0543	103.0543	0	$C_8H_7^+$
9	95.0492	95.0492	0	$C_6H_7O^+$
10	93.0700	93.0699	0.0001	$C_7H_9^+$
11	91.0543	91.0543	0	$C_7H_7^+$
12	83.0857	83.0855	0.0002	$C_6H_{11}^+$
13	79.0544	79.0542	0.0002	$C_6H_7^+$
14	77.0387	77.0386	0.0001	$C_6H_5^+$
15	69.0700	69.0699	0.0001	$C_5H_9^+$
16	67.0544	67.0542	0.0002	$C_5H_7^+$
17	57.0700	57.0699	0.0001	$C_4H_9^+$
18	55.0544	55.0542	0.0002	$C_4H_7^+$
19	53.0387	53.0386	0.0001	$C_4H_5^+$
20	51.0231	51.0230	0.0001	$C_4H_3^+$

Table C.4: Comparison of the m/z values of the measured fragment of zopiclone with the calculated values based on the sum formula.

Fragment	m/z of measured fragments	Calculated m/z	Δ mass	Sum formula
21	345.1227	345.1225	0.0002	$C_{16}H_{18}ClN_6O^+$
22	263.0333	263.0331	0.0002	$C_{11}H_8ClN_4O_2^+$
23	247.0384	247.0381	0.0003	$C_{11}H_8ClN_4O^+$
24	245.0228	245.0225	0.0003	$C_{11}H_6ClN_4O^+$
25	217.0276	217.0276	0	$C_{10}H_6ClN_4^+$
26	190.0167	190.0167	0	$C_9H_5ClN_3^+$
27	182.0587	182.0587	0	$C_{10}H_6ClN_4^{*+}$
28	181.0509	181.0509	0	$C_{10}H_5N_4^+$
29	163.0057	163.0058	0.0001	$C_8H_4ClN_2^+$
30	157.0163	157.0164	0.0001	$C_6H_6ClN_2O^+$
31	143.0815	143.0815	0	$C_6H_{11}ClN_2O_2^+$
32	139.0058	139.0058	0	$C_6H_4ClN_2^+$
34	111.9949	111.9949	0	$C_5H_3ClN^+$
35	100.0183	100.0182	0.0001	$C_7H_2N^+$
36	99.0918	99.0917	0.0001	$C_5H_{11}N_2^+$
37	94.0288	94.0287	0.0001	$C_5H_4NO^+$
38	84.9841	84.9840	0.0001	$C_4H_2Cl^+$
39	76.0183	76.0182	0.0001	$C_5H_2N^+$
40	75.0105	75.0104	0.0001	C_5HN^{*+}
41	66.0340	66.0338	0.0002	$C_4H_4N^+$
42	61.9793	61.9793	0.0001	$CHClN^+$
43	52.0183	52.0183	0.0001	$C_3H_2N^+$
44	51.0231	51.0229	0.0002	$C_4H_3^+$
45	50.0152	50.0152	0.0001	$C_4H_2^{*+}$

Publication corresponding to chapter 6: Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules



Abstract 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, but 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.

D.1 Introduction

Given the enormous variety of possible compositions of chemicals^{46,49}, 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 deduced.³¹⁻³⁴ But with an increasing number of atoms and functional groups in a molecule, assignment by this method can become exhaustively complicated.^{37,38} 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^{45,47,219}, but their applicability strongly depends on the existence of high-quality training data and the results often lack interpretability.²³⁶ Tabulation of typical fragmentation patterns^{35,36} 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)¹⁰⁵ was developed, which can operate in x = electron ionization (EI)⁵⁰, dissociative electron attachment (DEA)¹³² and collision induced dissociation (CID)⁹¹ run modes. The effectiveness of QCxMS to successfully generate in-silico spectra in its EI mode is well documented⁵¹⁻⁵⁴ and has been demonstrated recently by its use for extension of mass spectra databases^{43,237,238}. Detailed fragmentation pattern analysis using the EI, DEA, and positive ion CID modes have successfully been conducted earlier.^{56,89,92,132} 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^{18,19,158,239} can produce ions with multiple positive or negative charges.^{95,96} 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.^{171,228,230,240} 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^{70,71}, 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.

D.2 Theoretical Background

D.2.1 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.^{241,242} 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.²⁴³ Also, more information on the fragmentation mechanisms can be gained, because multiply charged fragments produce distinctive isotope patterns.^{244,245} In the most commonly used soft ionization methods^{158,159,162,239,246}, 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”¹⁷³, 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.^{247,248} Especially pH^{249–251} and the influence of different solvents^{241,252} were studied and it was found that the preferred protonation site is significantly diverse when the ion is in gas or liquid phase.^{176,177,253,254} 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.^{255–257}

This is known as the so-called mobile-proton effect,^{171,228,230,240} which states that the energetically lowest structure must not necessarily yield the dominant fragmentation pattern.

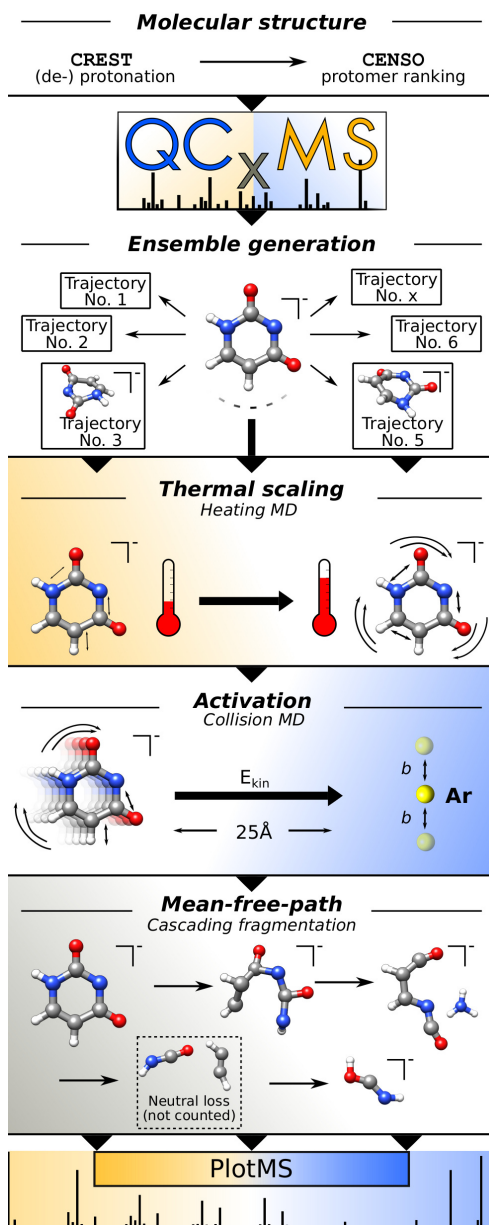


Figure D.1: Schematic diagram of the QCxMS workflow. First, an ensemble of MD snapshot structures is created. Consecutive simulations first increase the temperature, then simulate the collision and the mean-free-path. After fragmentation, the spectrum is plotted with PlotMS.

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 protocol⁸⁵, that is implemented in the Conformer Rotamer Ensemble Sampling Tool (CREST)^{86,178,179}. 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),⁸⁷ in which gas-phase computations can be compared to liquid phase results calculated with different solvents.

D.2.2 Method

Activation of a molecular ion after “soft-ionization” is commonly achieved using CID^{180,214}, 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 D.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 D.1. More details on the implementa-

tion are given in the original publication.⁹¹

D.2.3 Charge Assignment

Only charged structures can be measured in MS experiments. In QCxMS, the delta self-consistent field (Δ SCF)⁷⁶ 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 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 D.1.

Table D.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 \rightarrow 1	1.0	#1	0 \rightarrow 1	1.0
	0 \rightarrow 2	3.0		0 \rightarrow 2	2.0
#2	0 \rightarrow 1	1.5	#2	0 \rightarrow 1	1.5
	0 \rightarrow 2	3.5		0 \rightarrow 2	3.5
		Σ			Σ
		1-1	2.5		
		2-0	3.0		
		0-2	3.5		
			1-1	2.5	
			2-0	2.0	
			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.^{50,54}

D.2.4 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.¹¹¹

D.3 Technical Details

D.3.1 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 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-ketobutyric acid (**1**), 3-ureidopropionic acid (**2**), ascorbic acid (**3**), and tryptophan (**4**), which are shown in Figure D.2. Reference spectra were taken from the human metabolome (HMDB)^{42,97,186–188} and the MassBank^{98,189} 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.²⁵⁸ Such systems were thoroughly studied²⁵⁹, but exclusively dissociate into singly charged fragments. In large molecules, multiple (de-)protonated fragments can be obtained in high abundance.⁹⁶ However, running MD simulations for large structures can become computationally expensive. The fragmentation pathways of doubly protonated crizotinib (**5**)¹⁰³ and a triply protonated lysine derivate (**6**)¹⁰⁴ 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 affordable.

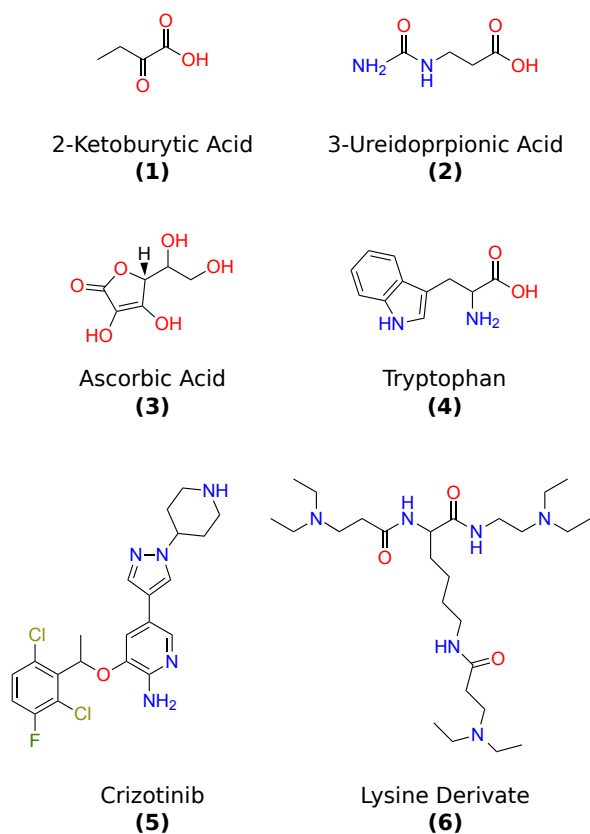


Figure D.2: Benchmark set of molecules for testing different charge states with QCxMS.

D.3.2 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.¹⁰⁵ 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^{86,178,179} version 2.11.3. Free energy ranking of the ensemble was done using CENSO⁸⁷ version 1.2.0 at the r²SCAN-3c¹⁰² 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^{260,261}. 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) (see Section D.6). DFT calculations were executed either using ORCA^{121,262} version 5.0.3 or TURBOMOLE²⁶³ version 7.5.1. The MD calculations in QCxMS are mostly based on the GFN2-xTB method^{71,72} 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)¹³² 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 D4⁶⁸ dispersion corrected PBE¹⁰⁰ and PBE0¹⁰¹ functionals with def2-SV(P)¹¹⁹, def2-TZVP²⁶⁴, as well as the minimal augmented ma-def2-XVP⁹⁹ (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).

D.3.3 Differences Between Experiment and Theory

The various reasons for observed differences between experimental and calculated spectra were discussed in detail in earlier work.^{91,92} In experiments, the “hardness” of the ionization process influences the degree of fragmentation.^{160,265} The conditions in the collision cells are device-specific and cannot exactly be reproduced by the simulation.¹⁹⁰ 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 severe 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.

D.4 Results and Discussion

D.4.1 Negative Charges

Ketobutyric Acid

The smallest benchmark structure, 2-ketobutyric 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 D.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.⁹⁷ A computed spectrum for protomer #2 is provided in the SI.

In all database spectra for 2-ketobutyric 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 D.3 a) describe the dissociation of neutral CO and CO₂ and produce fragments at signals m/z 73.028 and 57.034. The latter fragment dissociates further by H₂ loss, creating signal m/z 55.018. Signal m/z 44.998 (HCO₂⁻) is produced from protomer #2 (see SI).

The spectrum in Figure D.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 CO₂⁻ (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 D.3 c). The same collision energy used before leads to stronger dissociation of [M-H]⁻ in this spectrum. The structure dissociates more frequently into CO₂⁻ (m/z 43.989), CO and C₂H₅⁻ (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 D.3 a) and b).

In Figure D.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 D.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.

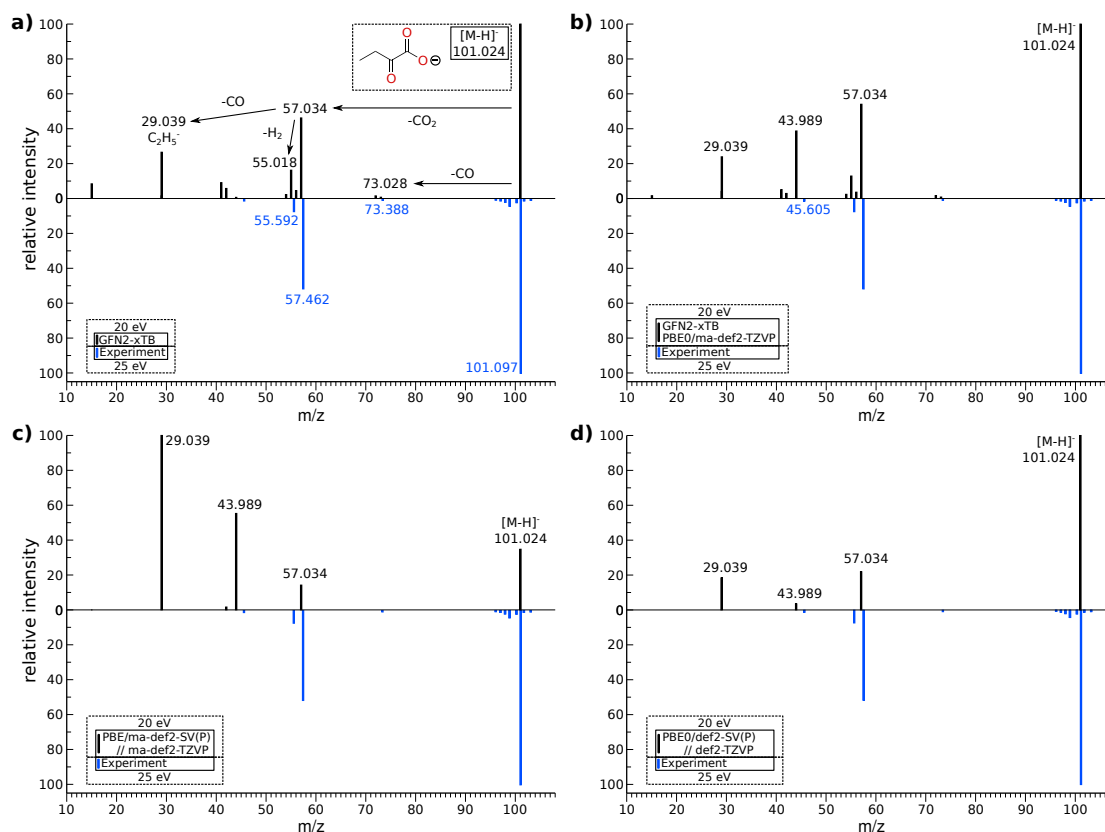


Figure D.3: Calculated spectra of 2-Ketobutyric acid (black, top) computed at 20 eV E_{LAB} compared to a measured spectrum (LC-ESI-QQQ) at 25 eV E_{LAB} (blue, inverted). a) GFN2-xTB//GFN2-xTB, b) GFN2-xTB//PBE0/ma-def2-TZVP, c) PBE/ma-def2-SV(P)//ma-def2-TZVP, d) PBE0/def2-SV(P)//PBE0/def2-TZVP.

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 reference⁹⁷ as shown in Figures D.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 ($\text{H}_3\text{N}_2\text{CO}^-$). 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 $\text{H}_2\text{C}_2\text{H}_4\text{CO}_2^-$ and neutral $\text{HN}=\text{C}=\text{O}$, which is described sufficiently by protomer #3 in Figure D.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

Appendix D Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules

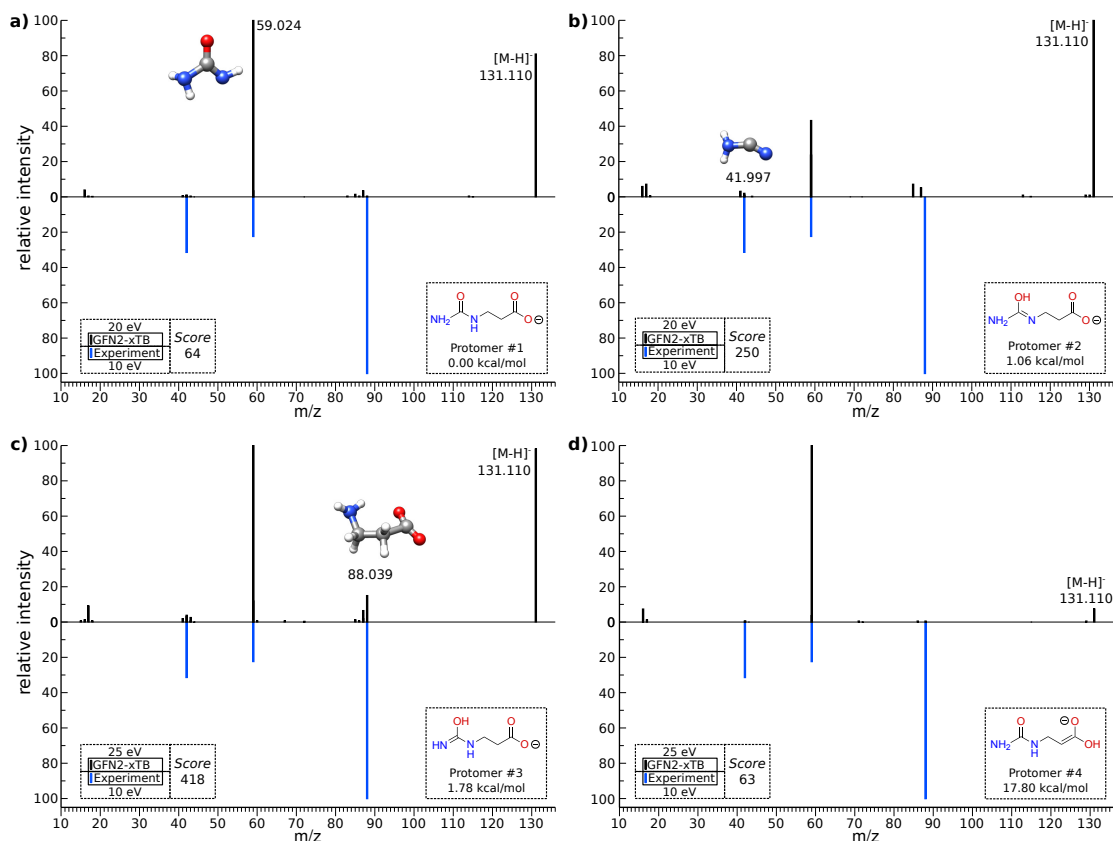


Figure D.4: Calculated spectra (black, top) of the four protomers of ureidopropionic acid using GFN2-xTB//GFN2-xTB compared to reference (LC-ESI-QTOF) at 10 eV E_{LAB} (blue, inverted). Protomer structures, relative free energies, and spectral matching scores are added to their respective spectrum.

different matching scores for the structures in Figure D.4.

In Figure D.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 D.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 D.4 c).

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 D.6 alongside their computed mass spectra. Calculations were performed at GFN2-xTB//GFN2-xTB level and the results are compared to a database spectrum.⁴² 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 ($\text{HO-C}_2\text{H}_4\text{-OH}$) and the negatively charged 2,3-hydroxyfuran fragment (m/z

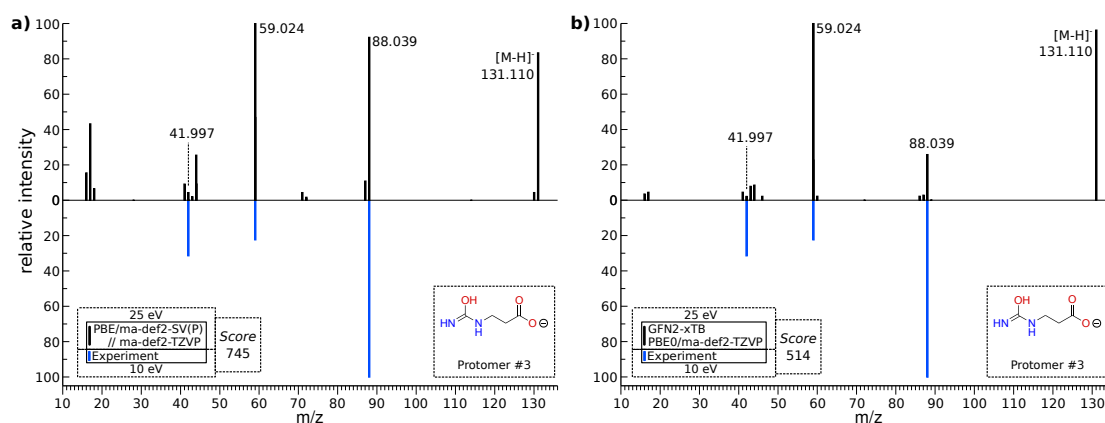


Figure D.5: Calculated spectrum (black, top) of ureidopropionic acid protomer #3 compared to reference (LC-ESI-QTOF) at 10 eV E_{LAB} (blue, inverted). a) PBE/ma-def2-SV(P)//PBE/ma-def2-TZVP. b) GFN2-xTB//PBE0/ma-def2-TZVP. Protomer structures, relative free energies, and matching scores are given with their respective spectrum.

115.003). For a better distinction in the following, ethanediol is called “side-chain” and 2,3-hydroxyfuran is called “backbone”.

The computed mass spectrum and structure of protomer #1 is displayed in Figure D.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 D.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 D.6 c)) and #5 (Figure D.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.

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 D.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 spectrum²⁶⁶ in Figures D.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 D.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

Appendix D Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules

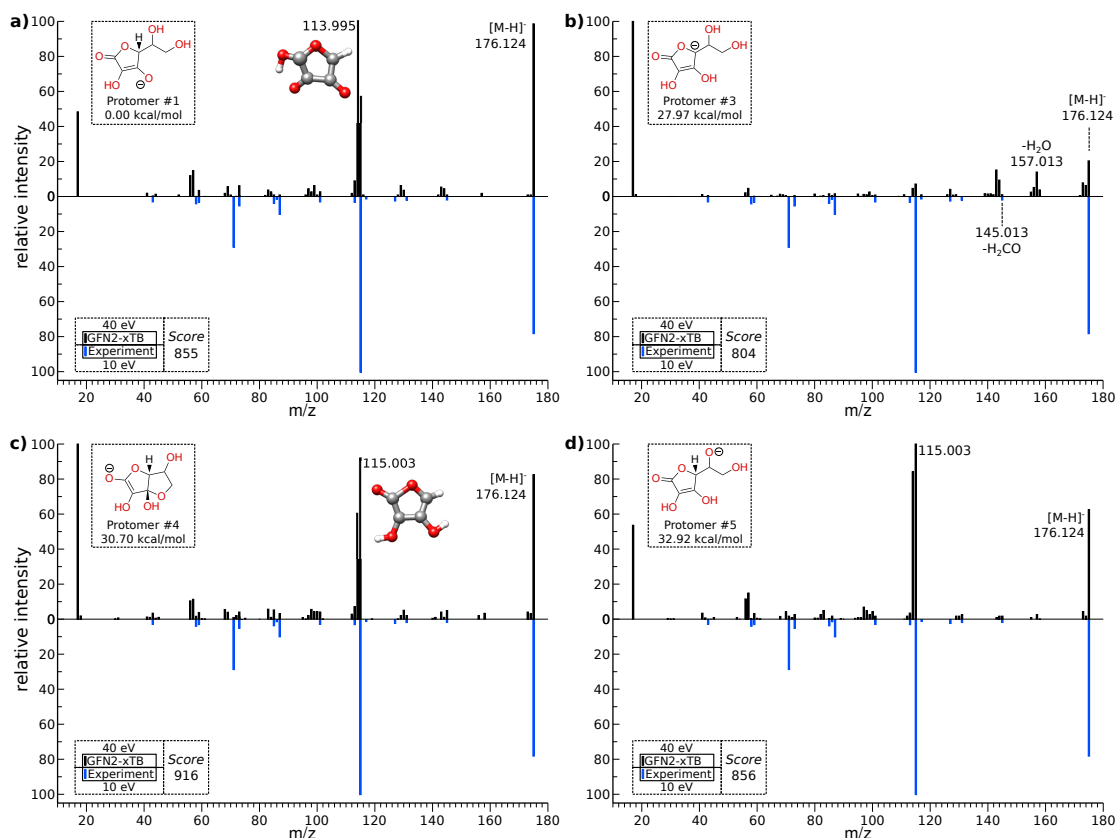


Figure D.6: GFN2-xTB//GFN2-xTB calculated spectra (black, top) at 40 eV E_{LAB} of ascorbic acid compared to measured spectrum (LC-ESI-QQQ) at 10 eV E_{LAB} (blue, inverted). Protomer structures, relative free energies, and matching scores are added to their respective spectrum. The structures attributed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

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 D.7 b). Three main fragmentation reactions of $[\text{M}-\text{H}]^-$ 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 ($\text{C}_2\text{H}_2\text{-NH}_2\text{-CO}_2\text{H}$) 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 $\text{NH}_2\text{-CH-COOH}$ glycine derivate, while a CH_2 group remains bound to the deprotonated indole fragment (see Figure D.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 D.7 c) and d) display considerable discrepancies to the experimental spectrum and low populations render these structures irrelevant for the observed spectrum.

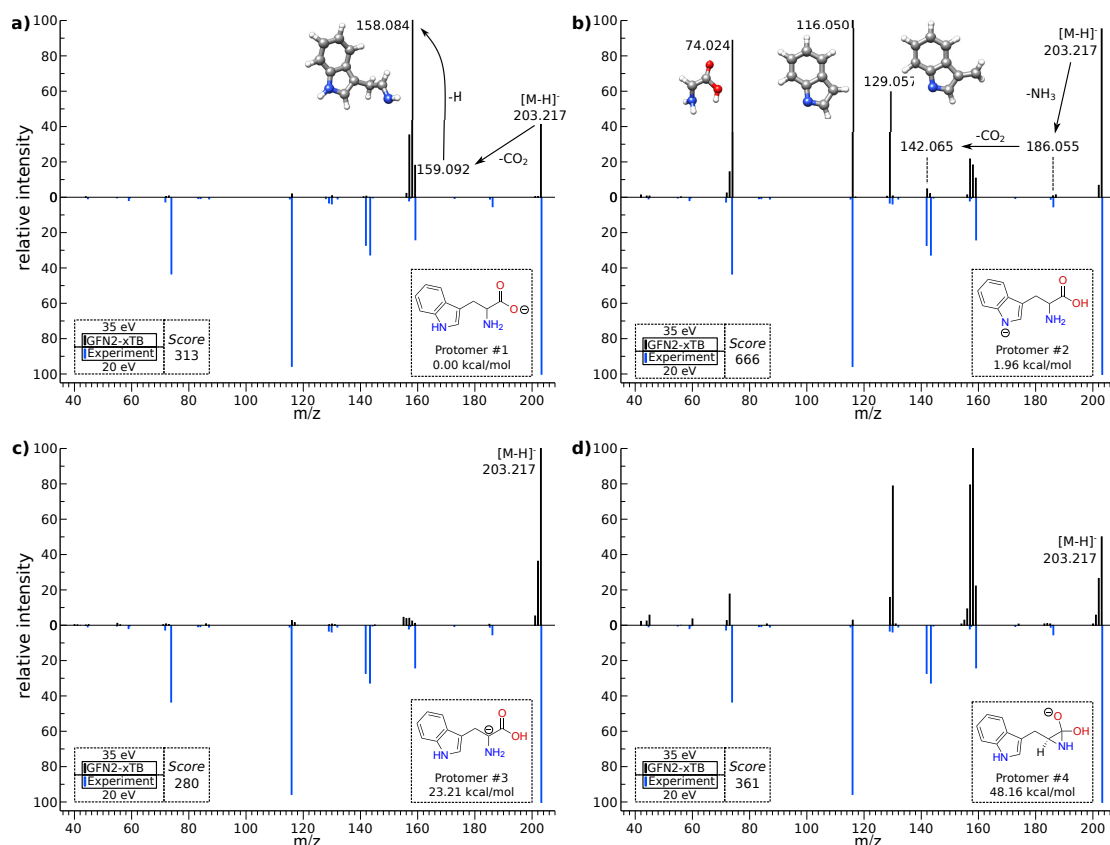


Figure D.7: Spectra calculated at GFN2-xTB/GFN2-xTB (black, top) of the four protomers of tryptophan at 35 eV E_{LAB} compared to a measured spectrum (LC-ESI-QQ) at 20 eV E_{LAB} ²⁶⁶ (blue, inverted). Protomer structures, relative free energies, and matching scores are added to their respective spectrum.

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 D.2), timings of the different [PES]/[EA] combinations are provided in table D.2.

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 rule²⁶⁷, 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 DEA¹³² which follows Stevenson's rule⁷⁷ for hard-ionization-based methods. Overall, the good performance

Appendix D Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules

Table D.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	–	–

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.

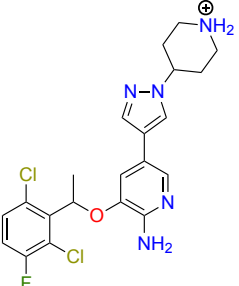
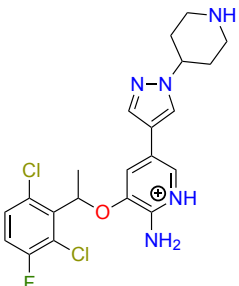
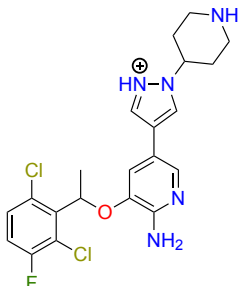
D.4.2 Multiple Charges

Crizotinib

The fragmentation pathways of doubly protonated crizotinib (**5**) were studied previously.¹⁰³ Figure D.9 a) shows the measurement of the singly positive charged molecular ion ($[M+H]^+$, m/z 450.1266). Figure D.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 D.8. The black roman numerals denote the reactions described in the literature, while gray numerals indicate alternative fragmentation pathways computed by QCxMS.

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 D.3 with their free energy ranking in water, methanol, and the gas phase at 300 K.

Table D.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

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 D.9 b). All experimental fragments were calculated correctly, however with a

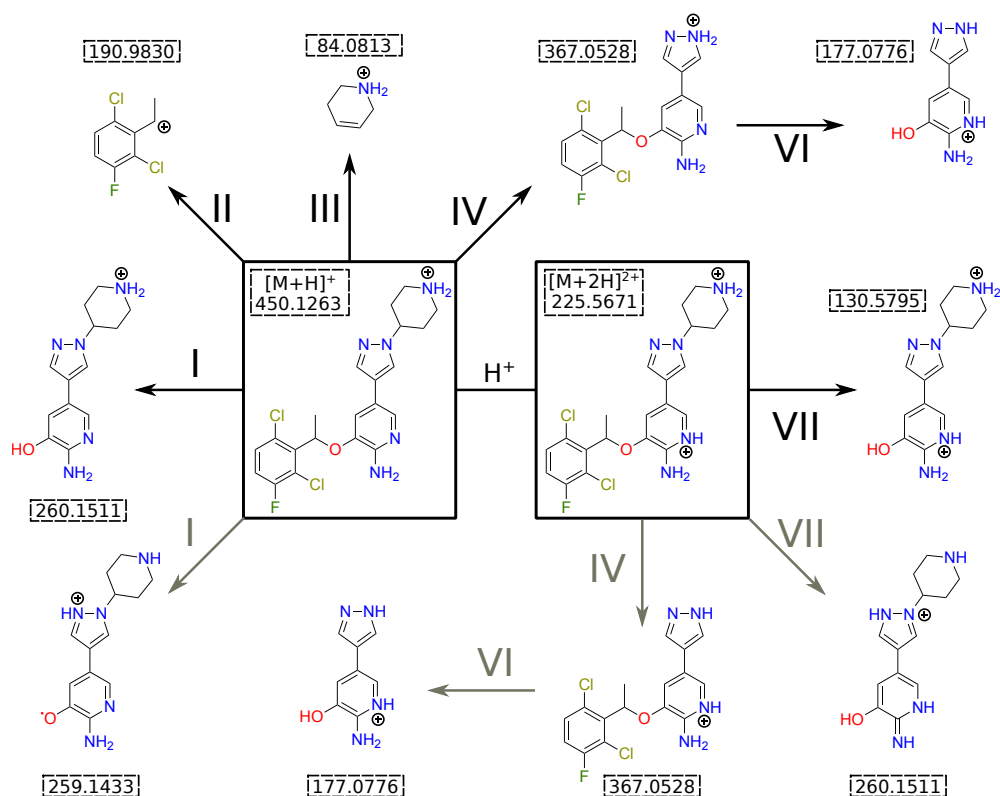


Figure D.8: Dissociation reactions as proposed in ref.¹⁰³ are marked with black roman numerals and arrows. Alternative fragmentation reactions computed with QCxMS are indicated with gray roman numerals and arrows.

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 D.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 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 (see Section D.6).

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 D.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 D.2.4, 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 D.9 b) and D.10 b). $\Delta m/z$ of 1 indicates a single charge present in the signal, while $\Delta m/z$ of 0.5 indicates two charges. $\Delta m/z$ 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, $\Delta m/z$ 1) in Figure D.9 b) to $[M+2H]^{2+}$ (m/z 225.5671, $\Delta m/z$ 0.5) in Figure D.10 b), it is evident that the software is able to

Appendix D Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules

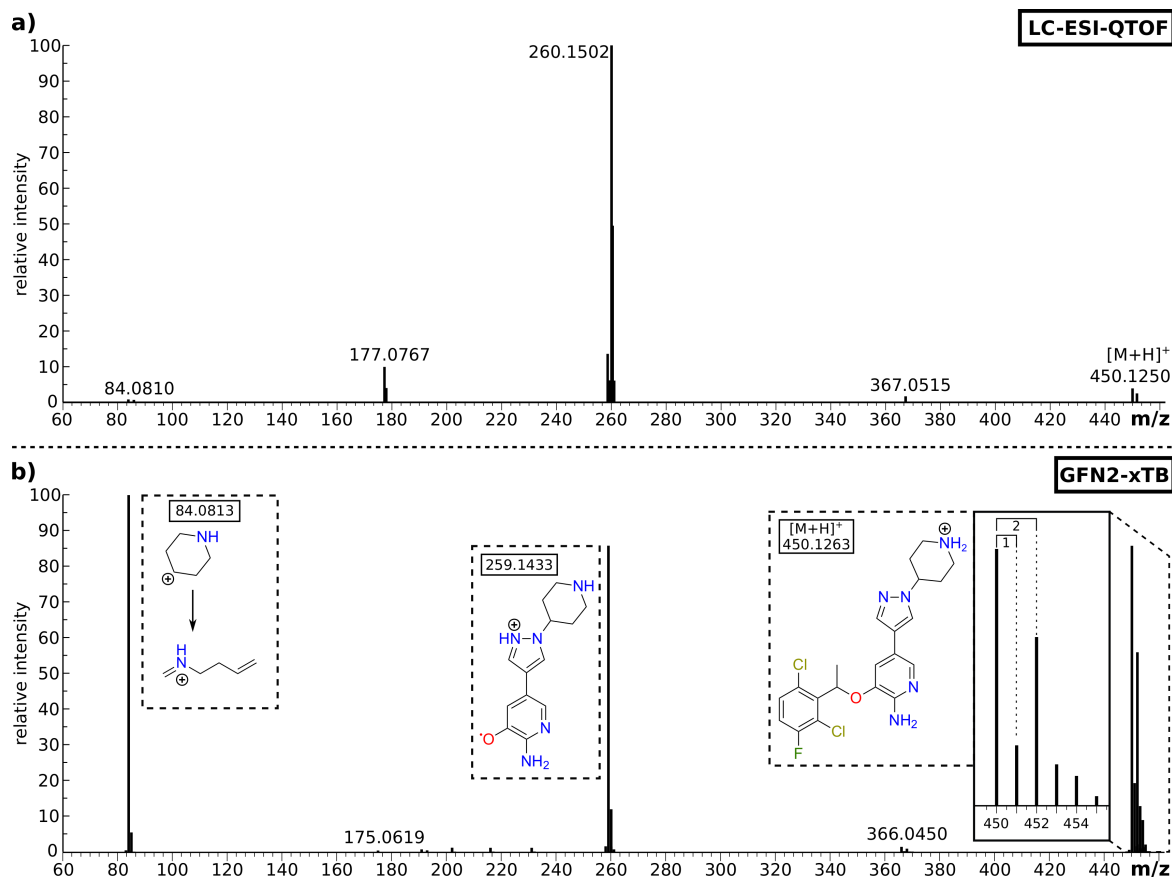


Figure D.9: a) measured spectrum of singly protonated crizotinib (LC-ESI-QTOF; 23 eV E_{LAB})¹⁰³. b) Computed spectrum (GFN2-xTB//GFN2-xTB at 50 eV E_{LAB}) of singly protonated crizotinib. Isotope patterns computed with PlotMS are enhanced for specific signals.

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 D.10 b), bottom left), as $\Delta m/z$ between the signals is 0.5. This proves that our approach is able to compute and assign multiply charged fragments with the Δ SCF method, for which details were described in Section D.2.3. In table D.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 D.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 D.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 reported signals for this compound. Using IP calculations to correctly allocate multiple charges to single fragments was accurate. Appropriate

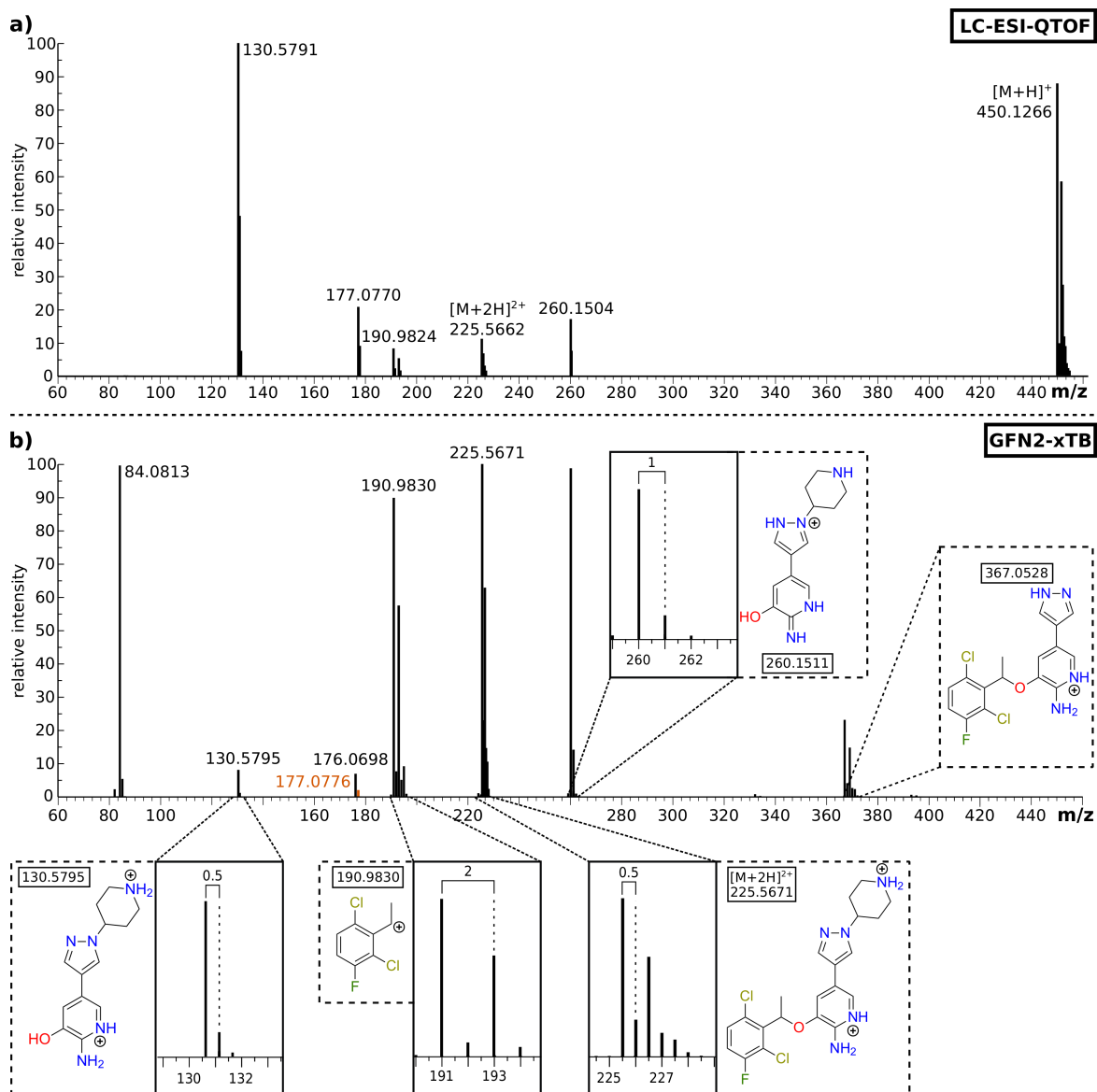


Figure D.10: a) Full scan MS of crizotinib (LC-ESI-QTOF; 25 eV E_{LAB})¹⁰³.
 b) Computed spectrum (GFN2-xTB//GFN2-xTB at 50 eV E_{LAB}) of doubly protonated crizotinib. Isotope patterns computed with PlotMS are enhanced for specific signals.

proton assignment after fragmentation was crucial for recreating the right signals, which was accounted for by the 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 D.3), which is not accounted for in the reference fragmentation scheme (Figure D.8). Signal intensities between measured and calculated spectrum are different, which is directly related to the discrepancies between experiment and calculation discussed in Section D.3.3. Increasing the level of theory for IP calculations at DFT level did not increase the

Appendix D Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules

Table D.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

overall accuracy (see SI).

Derivatized Lysine

In the literature,¹⁰⁴ 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 D.11. The proposed fragmentation pathway from the literature is displayed in Figure D.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 D.5.

Table D.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+

QCxMS computes four out of nine reported structures correctly and one correct fragment with

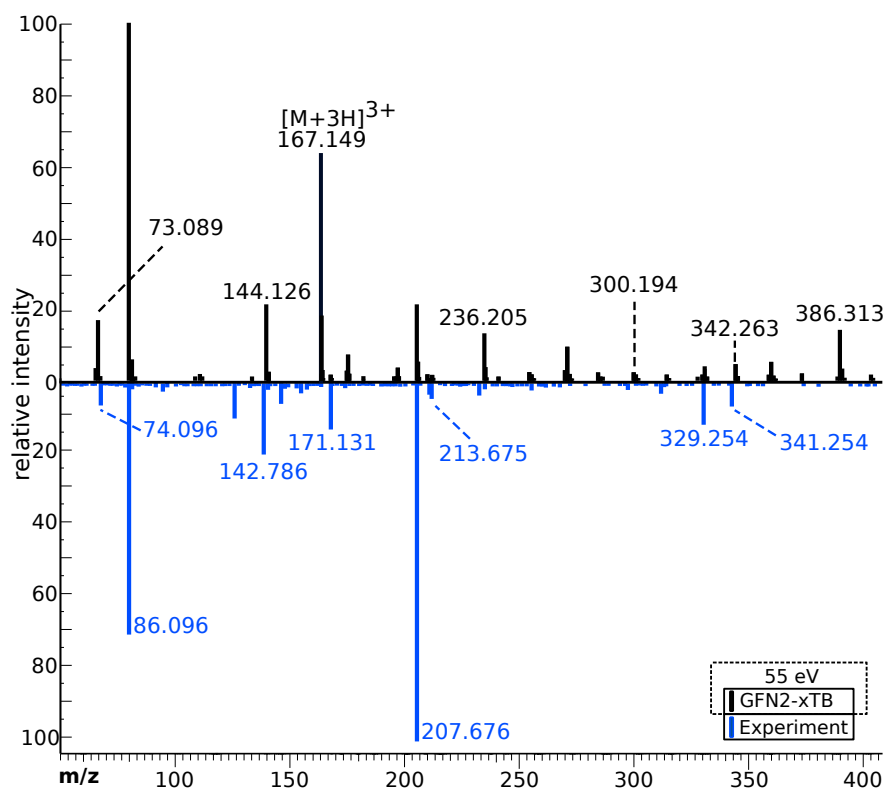


Figure D.11: Calculated spectrum (black, top) using GFN2-xTB//GFN2-xTB at 55 eV E_{LAB} of the triply charged lysine derivate compared to experimental spectrum (blue, inverted) taken from the literature.¹⁰⁴

wrong charge assignment. Signal m/z 142.786 (III) carries a 3+ charge in the experiment, but is 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+) \rightarrow 207.675 (2+) + 73.089 (1+) \rightarrow 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 D.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.

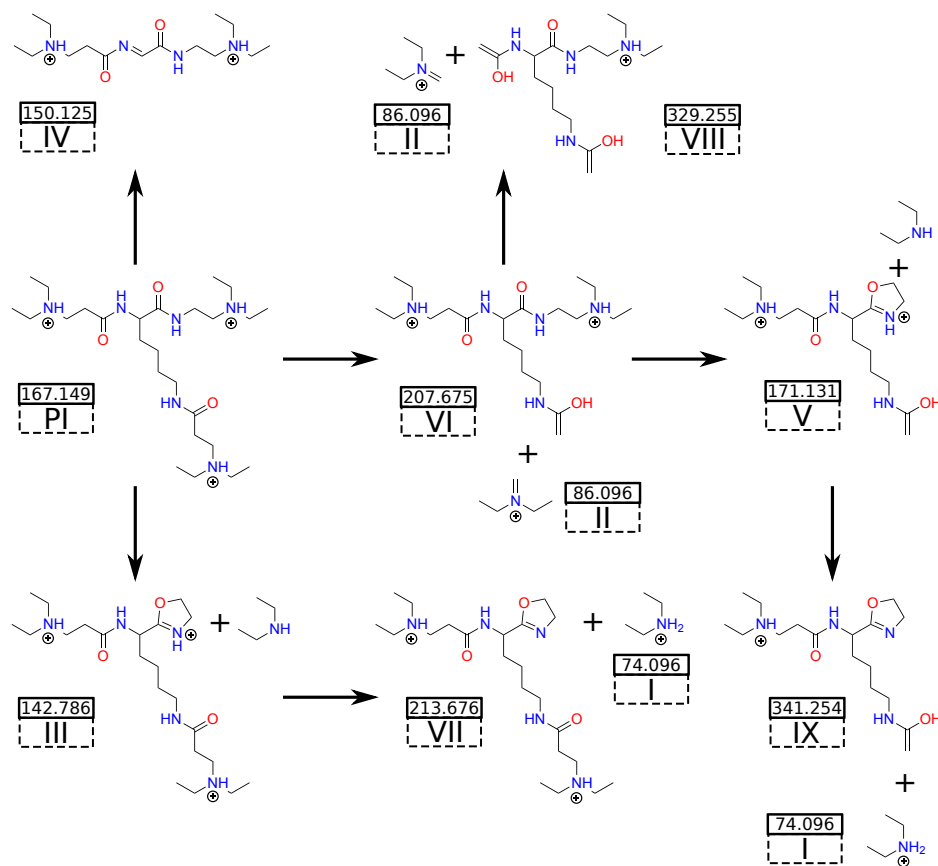


Figure D.12: CID fragmentation analysis proposed in the reference¹⁰⁴. Roman numerals were used for the product ions of spectrum in Figure D.11. PI is the precursor ion.

Overall, more investigation into the details of the charge assignment has to be conducted. An adapted implementation of Field's rule²⁶⁷ into QCxMS is planned, which renders a fragment with higher proton affinity more abundant.

D.5 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-ketobutyric 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-source¹⁰⁵ 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.

D.6 Supporting Information

D.6.1 Protomer rankings

Negative charge

In table D.6, the relative free energies of the benchmark structures (1) – (4) were computed using CENSO⁸⁷ with the composite method r²SCAN-3c¹⁰². Calculations were done in the gas phase and in solvation (methanol) using the implicit solvation model COSMO-RS.^{260,261}

Table D.6: Relative free energies of the negative ion benchmark structures in the gas phase and in methanol computed with CENSO at different temperatures. [kcal/mol]

Benchmark molecule	Protomer #	Gas	Solv	Gas	Solv	Gas	Solv
2-Ketobutyric Acid (1)	1	0.00	0.00	0.00	0.00	0.00	0.00
	2	20.50	17.00	21.31	21.58	22.26	22.69
3-Ureidopropionic Acid (2)	1	0.76	0.00	0.00	0.00	0.00	0.00
	2	0.00	15.58	1.06	10.79	4.26	11.66
	3	0.69	14.95	1.78	10.30	5.42	11.33
	4	18.13	26.82	17.80	24.21	17.77	22.76
Ascorbic Acid (3)	1	0.00	0.00	0.00	0.00	0.00	0.00
	2	13.17	5.97	12.82	10.68	12.06	11.59
	3	29.61	17.42	28.85	27.40	26.87	28.02
	4	30.78	24.81	31.23	25.56	32.02	25.42
	5	33.74	23.76	34.79	28.32	36.57	30.41
Tryptophan (4)	1	0.00	0.00	0.00	0.00	0.00	0.00
	2	2.04	16.74	1.96	11.45	1.58	9.74
	3	24.51	32.91	23.21	30.11	19.70	27.03
	4	47.86	44.78	48.16	47.32	47.87	48.27

Multiple charge

In table D.7, the relative free energies of the crizotinib (**5**) were computed using CENSO⁸⁷ with the composite method r²SCAN-3c¹⁰². Calculations were done in the gas phase and in solvation (water) using the implicit solvation model COSMO-RS.^{260,261} For the first protonation, the three most populated structures in the reported 12 kcal/mol energy range are marked. For the second protonation, the most populated structure (**3**) in solution was used, in accordance with the results reported in the literature and calculations.

Table D.7: Relative free energies of the first and second protonation of crizotinib in the gas phase and in water computed with CENSO at 300 K. [kcal/mol]

Protonation state	Protomer #	300K	
		Gas	Solv
1st Protonation	1	0.00	2.06
	2	7.04	10.13
	3	11.23	0.00
	4	15.35	27.18
	5	19.05	30.89
	6	16.01	13.21
	7	19.54	30.43
	8	30.34	40.50
2nd Protonation	1	0.00	0.00
	2	26.19	10.34
	3	38.65	27.67
	4	26.88	29.58
	5	34.15	31.28
	6	35.65	31.33
	7	41.03	34.37
	8	38.23	37.37
	9	30.08	39.08
	10	33.21	40.48
	11	26.18	43.80

D.6.2 Additional calculated Spectra

2-Ketobutyric Acid

The second protomer #2 of 2-Ketobutyric acid is computed in figure D.13. Figure D.13 a) displays the GFN2-xTB//GFN2-xTB computations, in figure D.13 b) the mixed GFN2-xTB//PBE0-ma-def2-TZVP results are shown. Overall agreement is acceptable. Especially signal m/z 44.998 is created by protomer #2, as the fragment CO_2H^- is formed.

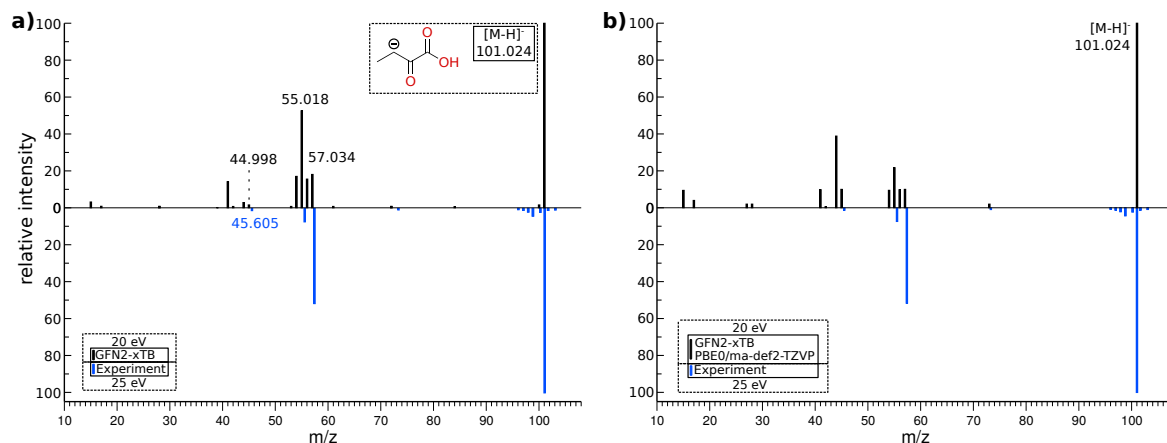


Figure D.13: Calculated spectrum of protomer #2 of 2-Ketobutyric acid (black, top) computed at 20 eV E_{LAB} compared to a measured spectrum (LC-ESI-QQQ) at 25 eV E_{LAB} (blue, inverted). The protomer structure, was added to the spectrum. a) GFN2-xTB//GFN2-xTB. b) GFN2-xTB//PBE0-ma-def2-TZVP.

Ascorbic acid

Spectra of protomers #1, #2, and #4 were calculated using GFN2-xTB//PBE0-ma-def2-TZVP and are compared to measurements⁹⁷ (LC-ESI-QQQ) in figure D.14. No significant differences to GFN2-xTB//GFN2-xTB results could be observed and matching scores are similar.

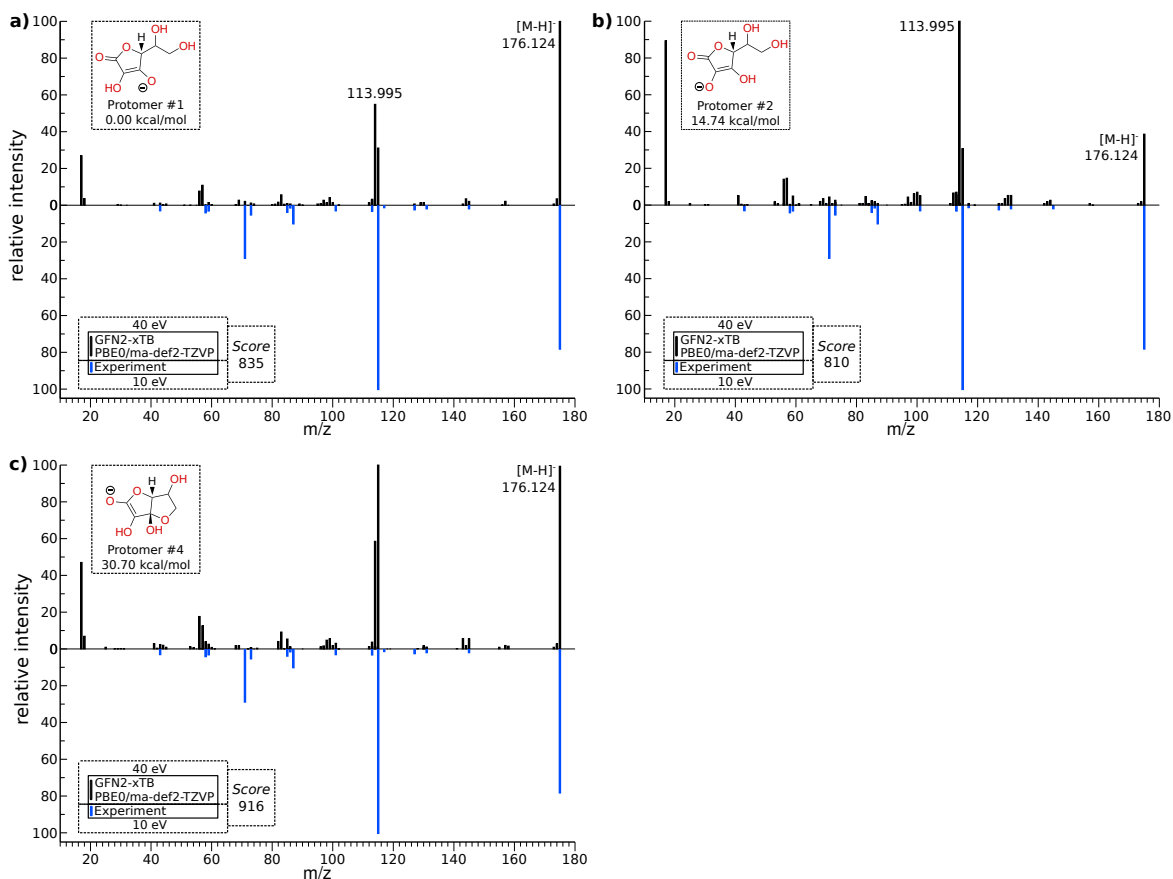


Figure D.14: GFN2-xTB//PBE0-ma-def2-TZVP calculated spectra (black, top) at 40 eV E_{LAB} of ascorbic acid compared to measured spectrum (LC-ESI-QQQ) at 10 eV E_{LAB} (blue, inverted). Protomer structures, relative free energies, and matching scores are added to their respective spectrum.

Tryptophan

The spectra of protomers #1 and #2 were calculated at 35 eV E_{LAB} using GFN2-xTB//PBE0/ma-def2-TZVP level. Comparison was done in figure D.15 to measurements (LC-ESI-QQ) at 20eV E_{LAB} . There were no significant improvements observable from the use of the DFT method for EA calculations.

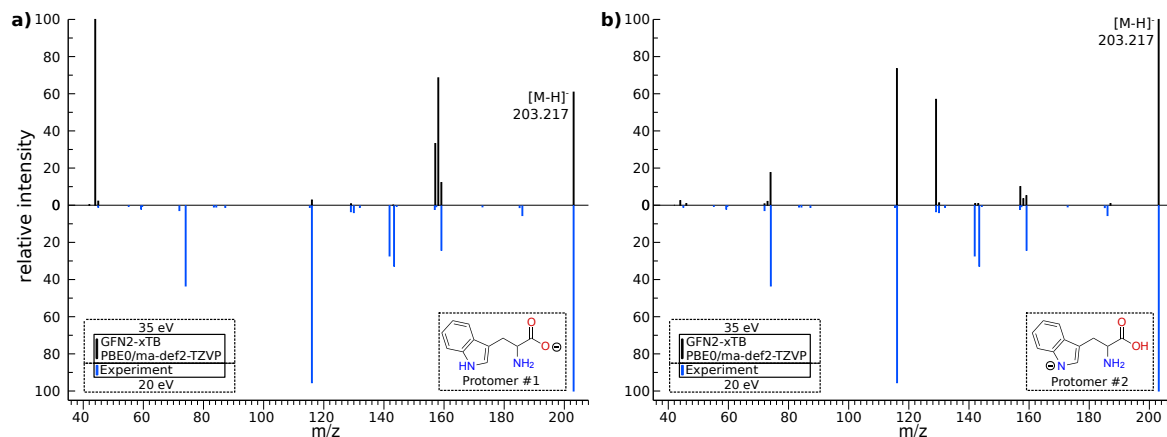


Figure D.15: Calculated spectra (black, top) of four protomers #1 and #2 of tryptophan at 35 eV E_{LAB} using GFN2-xTB//PBE0/ma-def2-TZVP compared to measured spectra (LC-ESI-QQ) at 20eV E_{LAB} (blue, inverted). Protomer structures, relative free energies, and matching scores are added to their respective spectrum.

Crizotinib

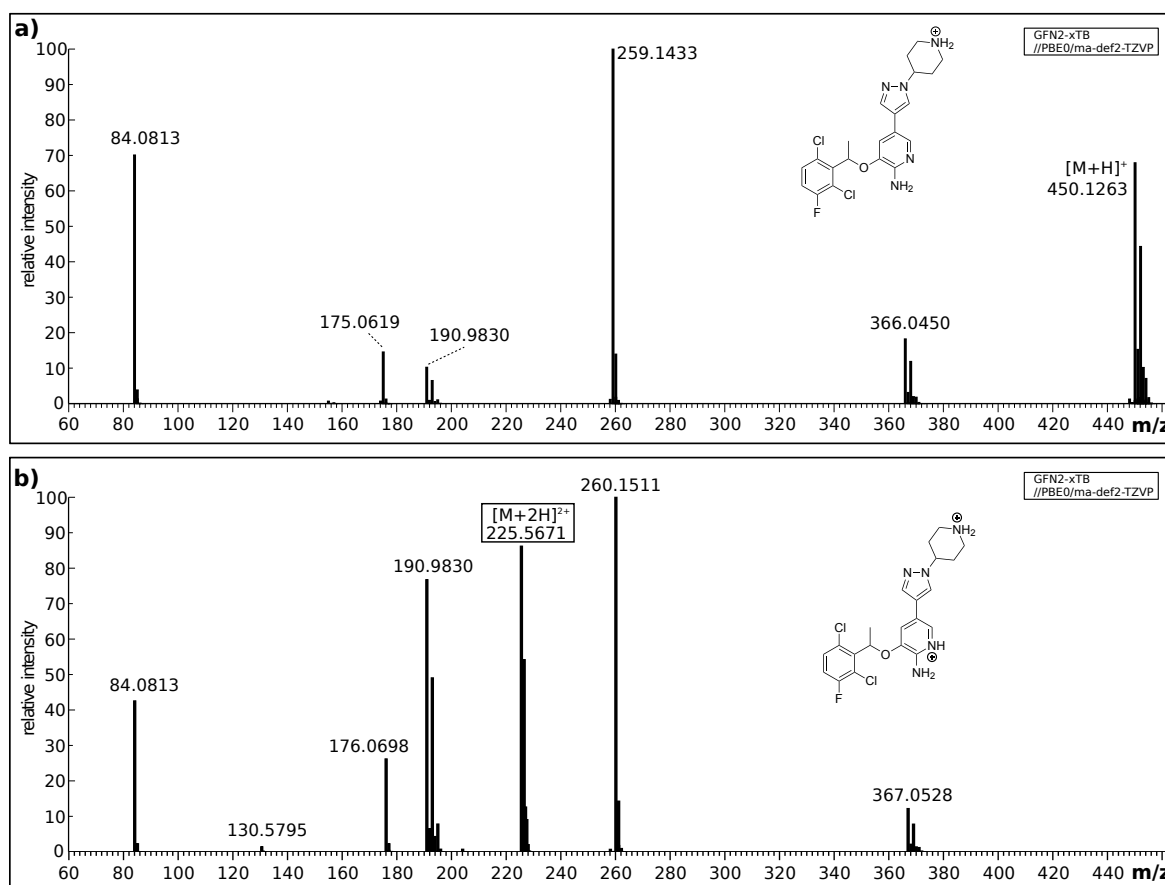


Figure D.16: Computed spectra of crizotinib using GFN2-xTB//PBE0/def2-TZVP at 50 eV E_{LAB} with isotope patterns computed with PlotMS. a) Singly protonated protomer #1. b) Doubly protonated protomer #1.

In figure D.16, singly a) and doubly b) protonated protomers #1 in water at 300 K were calculated using GFN2-xTB//PBE0/def2-TZVP at 50 eV E_{LAB} . Isotope patterns were computed with the external PlotMS program. Increase of the signal strength of peaks m/z 175.0619/176.0698 and decrease of intensity of signal m/z 84.0813 is observed. The DFT method ran for ~ 3347 minutes per fragmentation MD, while GFN2-xTB//GFN2-xTB computations were finished after 100 minutes.

The fragmentation patterns calculated using GFN2-xTB//GFN2-xTB at 50 eV E_{LAB} of the singly protonated structures protomers #2 and #3 are displayed in figure D.18 a) and b). The free energy values are provided in table D.7. The overall fragmentation patterns do not differ tremendously, however, the fragmentation intensity is diverging.

The second most populated structure in water at 300 K regarding the second protonation is shown in figure D.18 c). besides missing signal m/z 130, the spectrum looks similar to the spectrum of protomer #1. The lacking of signal m/z 130 is due to missing the correct proton transfer reaction.

In the literature¹⁰³, the fragment of signal m/z 130.5794 was mass selected and used for consecutive MS/MS dissociation reactions. The structure was also used for a QCxMS simulation and is depicted in figure D.17. The main fragmentation reaction of this structure dissociates into m/z 177.0776 and

Appendix D Calculation of Mass Spectra with the QCxMS Method for Negatively and Multiply Charged Molecules

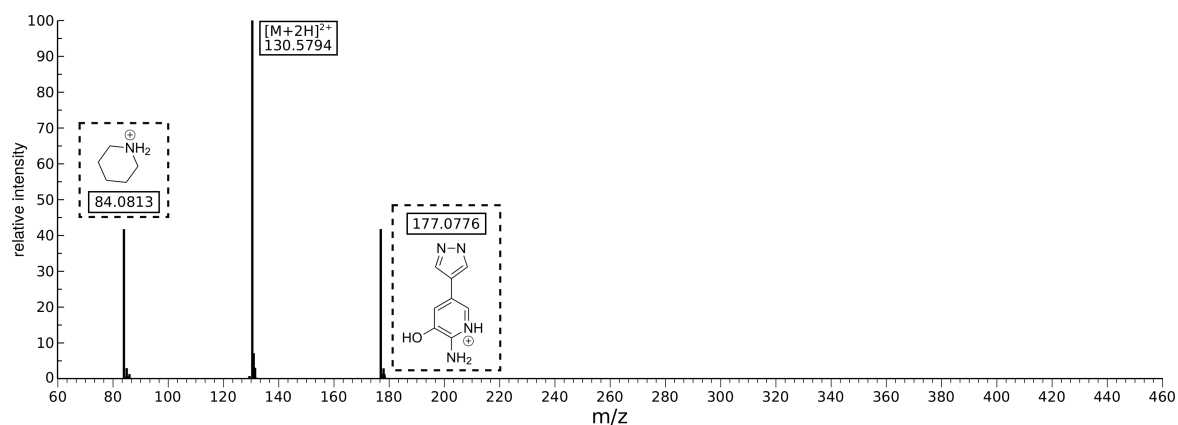


Figure D.17: Computed spectrum of crizotinib m/z 130 as precursor ion. Computations conducted at GFN2-xTB//GFN2-xTB level with 70 eV E_{LAB} .

84.0813, as expected. Further subsequent dissociation reactions from m/z 177.0776 as reported in the literature could not be observed. However, this structure can be taken as a precursor itself to enforce fragmentation and get the reported fragments. This was not computed.

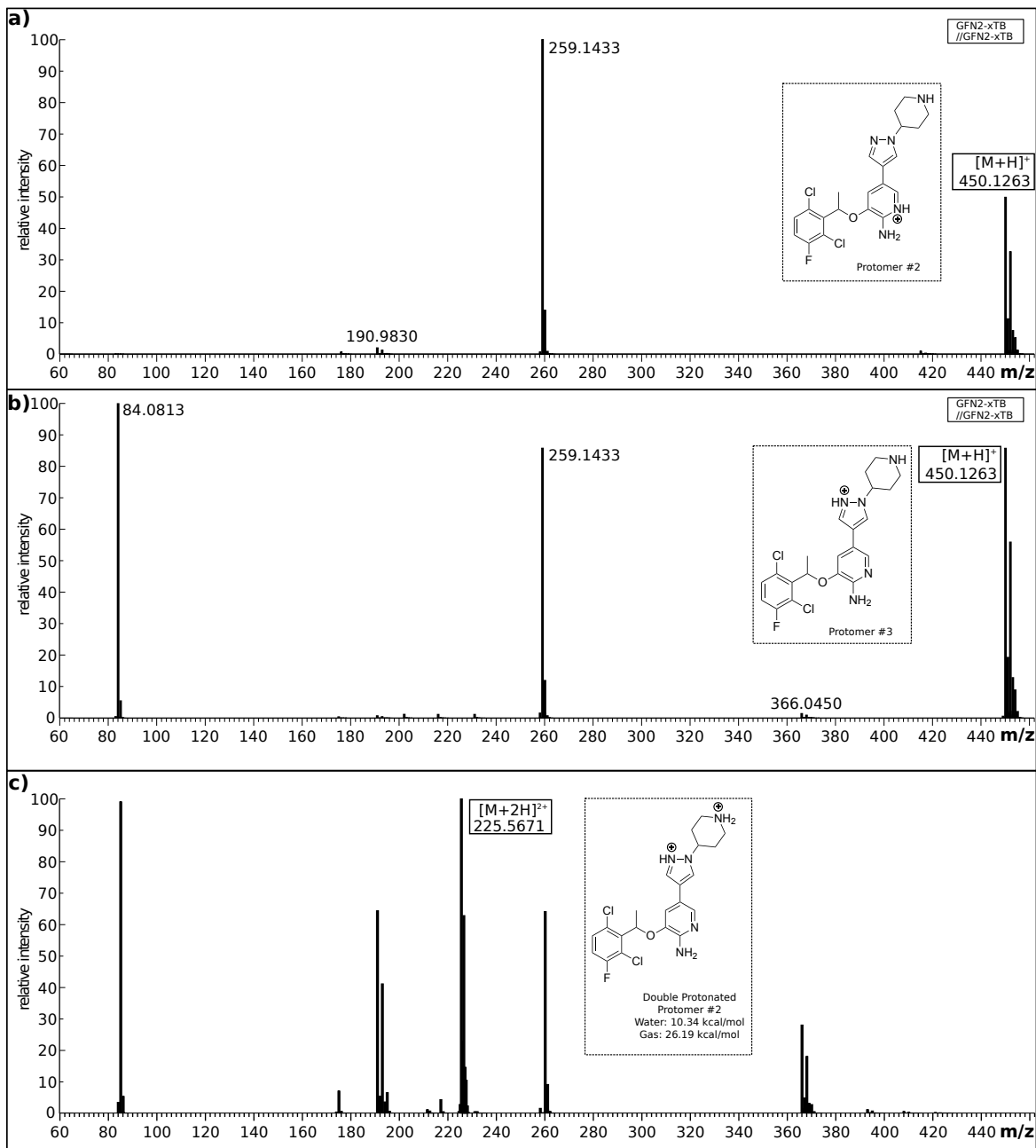


Figure D.18: Computed spectra of crizotinib using GFN2-xTB//GFN2-xTB at 50 eV E_{LAB} with isotope patterns computed with PlotMS. a) Protomer #2. b) Protomer #3. c) Doubly protonated protomer #2.

Bibliography

- [1] F. He, C. L. Hendrickson and A. G. Marshall, *Baseline Mass Resolution of Peptide Isobars: A Record for Molecular Mass Resolution*, *Anal. Chem.* **73** (2001) 647.
- [2] H. J. Cooper and A. G. Marshall, *Electrospray Ionization Fourier Transform Mass Spectrometric Analysis of Wine*, *J. Agric. Food Chem.* **49** (2001) 5710.
- [3] R. A. Hites and K. J. Jobst, *Response to "Letter to the Editor: Optimism for Nontarget Analysis in Environmental Chemistry"*, *Environ. Sci. Technol.* **53** (2019) 5531.
- [4] S. A. Schreckenbach, J. S. M. Anderson, J. Koopman, S. Grimme, M. J. Simpson and K. J. Jobst, *Predicting the Mass Spectra of Environmental Pollutants Using Computational Chemistry: A Case Study and Critical Evaluation*, *J. Am. Soc. Mass Spectrom.* **32** (2021) 1508.
- [5] C. A. Hughey, R. P. Rodgers and A. G. Marshall, *Resolution of 11 000 Compositionally Distinct Components in a Single Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrum of Crude Oil*, *Anal. Chem.* **74** (2002) 4145.
- [6] M. Kasperkowiak, M. Kurowska, M. Zalas and R. Frański, *Laser Desorption/Ionization Mass Spectrometry as a Potential Tool for Evaluation of Hydroxylation Degree of Various Types of Titanium Dioxide Materials*, *Materials* **14** (2021).
- [7] W. A. Harris, P. T. A. Reilly and W. B. Whitten, *Detection of Chemical Warfare-Related Species on Complex Aerosol Particles Deposited on Surfaces Using an Ion Trap-Based Aerosol Mass Spectrometer*, *Anal. Chem.* **79** (2007) 2354.
- [8] K. Vékey and A. Telekes, "Chapter 2 - Basics of analytical chemistry and mass spectrometry for medical professionals", *Medical Applications of Mass Spectrometry*, ed. by K. Vékey, A. Telekes and A. Vertes, Amsterdam: Elsevier, 2008 7.
- [9] V. G. Zaikin and R. S. Borisov, *Spectrometry as a Crucial Analytical Basis for Omics Sciences*, *J. Anal. Chem.* **76** (2021) 1567.
- [10] A. Vertes, "Chapter 8 - Mass spectrometry in proteomics", *Medical Applications of Mass Spectrometry*, ed. by K. Vékey, A. Telekes and A. Vertes, Amsterdam: Elsevier, 2008 173.
- [11] J. Gross, *Mass Spectrometry - A Textbook*, Weinheim: Springer-Verlag, 2017.

- [12] G. Münzenberg, *Development of mass spectrometers from Thomson and Aston to present*, *Int. J. Mass Spectrom.* **349-350** (2013) 9, 100 years of Mass Spectrometry.
- [13] K. Vékey, *Internal Energy Effects in Mass Spectrometry*, *J. Mass Spectrom.* **31** (1996) 445.
- [14] F. McLafferty and Tureček, *Interpretation of Mass Spectra*, Sausalito, California: University Science Books, 1993.
- [15] K. Maštovská and S. J. Lehotay, *Practical approaches to fast gas chromatography–mass spectrometry*, *J. Chromatogr. A* **1000** (2003) 153, A Century of Chromatography 1903-2003.
- [16] H. D. Beckey, *Massenspektrometrische Untersuchungen mit Hilfe einer Feldemissions-Ionenquelle*, *Z. Naturforsch. A* **14** (1959) 712.
- [17] V. L. Tal'roze and A. K. Ljubimova, *Secondary processes in the ion source of a mass spectrometer (presented by academician N.N. Semenov 27 viii 1952)—reprinted from report of the Soviet Academy of Sciences, Volume LXXXVI, -n5 (1952)*, *J. Mass Spectrom.* **33** (1998) 502.
- [18] E. C. Horning, M. G. Horning, D. I. Carroll, I. Dzidic and R. N. Stillwell, *New picogram detection system based on a mass spectrometer with an external ionization source at atmospheric pressure*, *Anal. Chem.* **45** (1973) 936.
- [19] M. Yamashita and J. B. Fenn, *Electrospray ion source. Another variation on the free-jet theme*, *J. Phys. Chem.* **88** (1984) 4451.
- [20] S. C.A., *Liquid Chromatography-Mass Spectrometry – An Introduction*. Chichester: Wiley, 2003.
- [21] M. H. Amad, N. B. Cech, G. S. Jackson and C. G. Enke, *Importance of gas-phase proton affinities in determining the electrospray ionization response for analytes and solvents*, *J. Mass Spectrom.* **35** (2000) 784.
- [22] S. Cristoni and L. R. Bernardi, *Development of new methodologies for the mass spectrometry study of bioorganic macromolecules*, *Mass Spectrom. Rev.* **22** (2003) 369.
- [23] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong and C. M. Whitehouse, *Electrospray Ionization for Mass Spectrometry of Large Biomolecules*, *Science* **246** (1989) 64.
- [24] J. B. Fenn, *Electrospray Wings for Molecular Elephants (Nobel Lecture)*, *Angew. Chem. Int. Ed.* **42** (2003) 3871.
- [25] E. J. Want, B. F. Cravatt and G. Siuzdak, *The Expanding Role of Mass Spectrometry in Metabolite Profiling and Characterization*, *Chem. Bio. Chem.* **6** (2005) 1941.
- [26] M. Lehmann, D. Gustav and C. G. Galizia, *The early bee catches the flower - circadian rhythmicity influences learning performance in honey bees, Apis mellifera*, *Behav. Ecol. Sociobiol.* **65** (2011) 205.

-
- [27] B. Pramanik, A. Ganguly and M. Gross, *Applied Electrospray Mass Spectrometry: Practical Spectroscopy Series Volume 32*, Practical Spectroscopy Bd. 32, CRC Press, 2002.
- [28] R. A. Yost and C. G. Enke, *Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation*, *Anal. Chem.* **51** (1979) 1251.
- [29] L. Sleno and D. A. Volmer, *Ion activation methods for tandem mass spectrometry*, *J. Mass Spectrom.* **39** (2004) 1091.
- [30] J. V. Johnson, R. A. Yost, P. E. Kelley and D. C. Bradford, *Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles and quadrupole ion traps*, *Anal. Chem.* **62** (1990) 2162.
- [31] D. P. Demarque, A. E. M. Crotti, R. Vessecchi, J. L. C. Lopes and N. P. Lopes, *Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products*, *Nat. Prod. Rep.* **33** (3 2016) 432.
- [32] W. Niessen, *Fragmentation of toxicologically relevant drugs in positive-ion liquid chromatography–tandem mass spectrometry*, *Mass Spectrom. Rev.* **30** (2011) 626.
- [33] W. Niessen, *Fragmentation of toxicologically relevant drugs in negative-ion liquid chromatography–tandem mass spectrometry*, *Mass Spectrom. Rev.* **31** (2012) 626.
- [34] C. Cheng and M. L. Gross, *Applications and mechanisms of charge-remote fragmentation*, *Mass Spectrom. Rev.* **19** (2000) 398.
- [35] Mistrik, *Mass FrontierTM 3.0*, <http://www.highchem.com>.
- [36] L. Drahos and K. Vékely, *MassKinetics: a theoretical model of mass spectra incorporating physical processes, reaction kinetics and mathematical descriptions*, *J. Mass Spectrom.* **36** (2001) 237.
- [37] D. Ortiz, J.-Y. Salpin, K. Song and R. Spezia, *Galactose-6-Sulfate collision induced dissociation using QM+MM chemical dynamics simulations and ESI-MS/MS experiments*, *Int. J. Mass Spectrom.* **358** (2014) 25.
- [38] E. R. Molina, D. Ortiz, J.-Y. Salpin and R. Spezia, *Elucidating collision induced dissociation products and reaction mechanisms of protonated uracil by coupling chemical dynamics simulations with tandem mass spectrometry experiments*, *J. Mass Spectrom.* **50** (2015) 1340.
- [39] M. Hesse, H. Meier and B. Zeeh, *Spektroskopische Methoden in der organischen Chemie*, Stuttgart: Thieme, 2005 242.
- [40] *NIST website, U.S. Secretary of Commerce on behalf of the United States of America.*
- [41] *Mass Bank Europe website, Helmholtz centre for environmental research, accessed Apr.05, 2022.*
- [42] *Human Metabolome DataBase website, The metabolomics innovation center, accessed Apr.05, 2022.*
- [43] *MassBank of North America website, MassBank of North America, accessed Apr.05, 2022.*

- [44] *Drugbank website, University of Alberta.*
- [45] F. Allen, A. Pon, M. Wilson, R. Greiner and D. Wishart, *CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra*, *Nucleic Acids Res.* **42** (2014) W94.
- [46] F. Allen, A. Pon, R. Greiner and D. Wishart, *Computational Prediction of Electron Ionization Mass Spectra to Assist in GC/MS Compound Identification*, *Anal. Chem.* **88** (2016) 7689.
- [47] J. N. Wei, D. Belanger, R. P. Adams and D. Sculley, *Rapid Prediction of Electron Ionization Mass Spectrometry Using Neural Networks*, *ACS Cent. Sci.* **5** (2019) 700.
- [48] U. W. Liebal, A. N. T. Phan, M. Sudhakar, K. Raman and L. M. Blank, *Machine Learning Applications for Mass Spectrometry-Based Metabolomics*, *Metabolites* **10** (2020).
- [49] C. M. Dobson, *Chemical space and biology*, *Nature* **432** (2004) 824.
- [50] S. Grimme, *Towards First Principles Calculation of Electron Impact Mass Spectra of Molecules*, *Angew. Chem. Int. Ed.* **52** (2013) 6306.
- [51] C. A. Bauer and S. Grimme, *Elucidation of Electron Ionization Induced Fragmentations of Adenine by Semiempirical and Density Functional Molecular Dynamics*, *J. Phys. Chem. A* **118** (2014) 11479.
- [52] C. A. Bauer and S. Grimme, *First principles calculation of electron ionization mass spectra for selected organic drug molecules*, *Org. Biomol. Chem.* **12** (43 2014) 8737.
- [53] C. A. Bauer and S. Grimme, *Automated Quantum Chemistry Based Molecular Dynamics Simulations of Electron Ionization Induced Fragmentations of the Nucleobases Uracil, Thymine, Cytosine, and Guanine*, *Eur. J. Mass Spectrom.* **21** (2015) 125.
- [54] C. A. Bauer and S. Grimme, *How to Compute Electron Ionization Mass Spectra from First Principles*, *J. Phys. Chem. A* **120** (2016) 3755.
- [55] A. S. Christensen, T. Kubař, Q. Cui and M. Elstner, *Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications*, *Chem. Rev.* **116** (2016) 5301.
- [56] V. Ásgeirsson, C. A. Bauer and S. Grimme, *Quantum chemical calculation of electron ionization mass spectra for general organic and inorganic molecules*, *Chem. Sci.* **8** (7 2017) 4879.
- [57] M. Born and R. Oppenheimer, *Zur Quantentheorie der Molekeln*, *Ann. Phys.* **389** (1927) 457.
- [58] A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*, Courier Corporation, 2012.
- [59] E. Schrödinger, *An Undulatory Theory of the Mechanics of Atoms and Molecules*, *Phys. Rev.* **28** (6 1926) 1049.

-
- [60] P. A. M. Dirac, *A new notation for quantum mechanics*, *Math. Proc. Camb. Philos. Soc.* **35** (1939) 416.
- [61] R. G. Parr, *Density functional theory of atoms and molecules*, Springer, 1980.
- [62] P. Hohenberg and W. Kohn, *Inhomogeneous electron gas*, *Phys. Rev.* **136** (1964) B864.
- [63] J. P. Perdew and K. Schmidt, *Jacob's ladder of density functional approximations for the exchange-correlation energy*, vol. 577, AIP Conference Proceedings, 2001.
- [64] S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Dispersion-Corrected Mean-Field Electronic Structure Methods*, *Chem. Rev.* **116** (2016) 5105.
- [65] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu*, *J. Chem. Phys.* **132** (2010) 154104.
- [66] S. Grimme, S. Ehrlich and L. Goerigk, *Effect of the damping function in dispersion corrected density functional theory*, *J. Comput. Chem.* **32** (2011) 1456.
- [67] A. D. Becke and E. R. Johnson, *A density-functional model of the dispersion interaction*, *J. Chem. Phys.* **123** (2005) 154101.
- [68] E. Caldeweyher, C. Bannwarth and S. Grimme, *Extension of the D3 dispersion coefficient model*, *J. Chem. Phys.* **147** (2017) 034112.
- [69] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth and S. Grimme, *A generally applicable atomic-charge dependent London dispersion correction*, *J. Chem. Phys.* **150** (2019) 154122.
- [70] S. Grimme, C. Bannwarth and P. Shushkov, *A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements ($Z = 1-86$)*, *J. Chem. Theory Comput.* **13** (2017) 1989.
- [71] C. Bannwarth, S. Ehlert and S. Grimme, *GFN2-xTB - An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions*, *J. Chem. Theory Comput.* **15** (2019) 1652.
- [72] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Extended tight-binding quantum chemistry methods*, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **11** (2021) e1493.
- [73] N. D. Mermin, *Thermal properties of the inhomogeneous electron gas*, *Phys. Rev.* **137** (1965) A1441.
- [74] D. Marx and J. Hutter, *Ab Initio Molecular Dynamics*, Cambridge: Cambridge University Press, 2009.
- [75] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *Molecular dynamics with coupling to an external bath*, *J. Chem. Phys.* **81** (1984) 3684.

- [76] R. O. Jones and O. Gunnarsson, *The density functional formalism, its applications and prospects*, *Rev. Mod. Phys.* **61** (3 1989) 689.
- [77] D. Stevenson, *Ionization and dissociation by electronic impact. The ionization potentials and energies of formation of sec.-propyl and tert.-butyl radicals. Some limitations on the method*, *Discuss. Faraday Soc.* **10** (1951) 35.
- [78] E. Condon, *A Theory of Intensity Distribution in Band Systems*, *Phys. Rev.* **28** (6 1926) 1182.
- [79] N. Turro, *Modern Molecular Photochemistry*, Menlo Park, California: The Benjamin/Cummings Publishing Company Inc., 1978.
- [80] O. Rice and H. Ramsperger, *Theories of Unimolecular Gas Reactions at Low Pressures*, *J. Am. Chem. Soc.* **49** (1927) 1617.
- [81] L. S. Kassel, *Studies in Homogeneous Gas Reactions. II. Introduction of Quantum Theory*, *J. Phys. Chem.* **32** (1927) 1065.
- [82] R. A. Marcus and O. K. Rice, *The Kinetics of the Recombination of Methyl Radicals and Iodine Atoms*, *J. Phys. Chem.* **55** (1951) 894.
- [83] H. Rosenstock, M. Wallenstein, A. Wahrhaftig and H. Eyring, *Absolute Rate Theory for Isolated Systems and the Mass Spectra of Polyatomic Molecules*, *Proc. Natl. Acad. Sci. U.S.A.* **38** (1952) 447.
- [84] Á. Kuki, G. Shemirani, L. Nagy, B. Antal, M. Zsuga and S. Kéki, *Estimation of activation energy from the survival yields: Fragmentation study of leucine enkephalin and polyethers by tandem mass spectrometry*, English, *J. Am. Soc. Mass Spectrom.* **24** (2013) 1064.
- [85] P. Pracht, C. A. Bauer and S. Grimme, *Automated and efficient quantum chemical determination and energetic ranking of molecular protonation sites*, *J. Comput. Chem.* **38** (2017) 2618.
- [86] P. Pracht, F. Bohle and S. Grimme, *Automated exploration of the low-energy chemical space with fast quantum chemical methods*, *Phys. Chem. Chem. Phys.* **22** (14 2020) 7169.
- [87] S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher and M. Stahn, *Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules*, *J. Phys. Chem.* **125** (2021) 4039.
- [88] S. Grimme, *Supramolecular binding thermodynamics by dispersion-corrected density functional theory*, *Chem. Eur. J.* **18** (2012) 9955.
- [89] J. Koopman and S. Grimme, *Calculation of Electron Ionization Mass Spectra with Semiempirical GFNn-xTB Methods*, *ACS Omega* **4** (2019) 15120.
- [90] *SDBS website, National Institute of Advanced Industrial Science and Technology.*

-
- [91] J. Koopman and S. Grimme, *From QCEIMS to QCxMS: A Tool to Routinely Calculate CID Mass Spectra Using Molecular Dynamics*, *J. Am. Soc. Mass Spectrom.* **32** (2021) 1735.
- [92] R. Schnegotzki, J. Koopman, S. Grimme and R. D. Süssmuth, *Quantum Chemistry-based Molecular Dynamics Simulations as a Tool for the Assignment of ESI-MS/MS Spectra of Drug Molecules*, *Chem. Eur. J.* **28** (2022) e202200318.
- [93] K. Chen, N. S. Rannulu, Y. Cai, P. Lane, A. L. Liebl, B. B. Rees, C. Corre, G. L. Challis and R. B. Cole, *Unusual odd-electron fragments from even-electron protonated prodiginine precursors using positive-ion electrospray tandem mass spectrometry*, *J. Am. Soc. Mass Spectrom.* **19** (2008) 1856.
- [94] S. Beuck, T. Schwabe and S. Grimme, *Unusual mass spectrometric dissociation pathway of protonated isoquinoline-3-carboxamides due to multiple reversible water adduct formation in the gas phase*, *J. Am. Soc. Mass Spectrom.* **20** (2009) 2034.
- [95] S. A. McLuckey and J. L. Stephenson, *Ion/ion chemistry of high-mass multiply charged ions*, *Mass Spectrom. Rev.* **17** (1998) 369.
- [96] K. Vékey, *Multiply charged ions*, *Mass Spectrom. Rev.* **14** (1995) 195.
- [97] 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. Nazzyrova, R. Shaykhtudinov, L. Li, H. J. Vogel and I. Forsythe, *HMDB: a knowledgebase for the human metabolome*, *Nucleic Acids Res.* **37** (2008) D603.
- [98] *MassBank consortium and its contributors; MassBank/MassBank-data: Release version 2020.06*, version 2020.06.
- [99] J. Zheng, X. Xu and D. G. Truhlar, *Minimally augmented Karlsruhe basis sets*, *Theor. Chem. Acc.* **128** (2011) 295.
- [100] J. P. Perdew, M. Ernzerhof and K. Burke, *Rationale for mixing exact exchange with density functional approximations*, *J. Chem. Phys.* **105** (1996) 9982.
- [101] C. Adamo and V. Barone, *Toward reliable density functional methods without adjustable parameters: The PBE0 model*, *J. Chem. Phys.* **110** (1999) 6158.
- [102] S. Grimme, A. Hansen, S. Ehlert and J.-M. Mewes, *r2SCAN-3c: A "Swiss army knife" composite electronic-structure method*, *J. Chem. Phys.* **154** (2021) 064103.
- [103] J. R. Joyce and D. S. Richards, *Kinetic Control of Protonation in Electrospray Ionization*, *J. Am. Soc. Mass Spectrom.* **22** (2011) 360.
- [104] T. Huang, J. M. Rabus, B. J. Bythell and J. L. Edwards, *Fragmentation of Multi-charged Derivatized Lysine Using Nanospray CID Tandem Mass Spectrometry*, *J. Am. Soc. Mass Spectrom.* **30** (2019) 1158.

- [105] *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.*
- [106] J. Grunenberg, *Computational Spectroscopy: Methods, Experiments and Applications*, Weinheim: Wiley-VCH, 2010 1.
- [107] R. L. Foltz, D. M. Andrenyak and D. J. Crouch, *Forensic Science, Applications of Mass Spectrometry*, 2nd, Oxford: Academic Press, 1999 693.
- [108] M. Tsivou, N. Kioukia-Fougia, I.-P. Lyris, M. Stamou, M.-H. Spyridaki and C. Gerorgakopolous, *An overview of the doping control analysis during the Olympic Games of 2004 in Athens, Greece*, *Anal. Chim. Acta.* **555** (2006) 1.
- [109] F. Rasche, A. Svatos, R. Maddula, C. Böttcher and S. Böcker, *Computing Fragmentation Trees from Tandem Mass Spectrometry Data*, *Anal. Chem.* **83** (2011) 1243.
- [110] F. W. McLafferty, M.-Y. Zhang, D. B. Stauffer and S. Y. Loh, *Comparison of algorithms and databases for matching unknown mass spectra*, *J. Am. Soc. Mass Spectrom.* **9** (1998) 92.
- [111] S. E. Stein and D. R. Scott, *Optimization and testing of mass spectral library search algorithms for compound identification*, *J. Am. Soc. Mass Spectrom.* **5** (1994) 859.
- [112] P. R. Spackman, B. Bohman, A. Karton and D. Jayatilaka, *Quantum chemical electron impact mass spectrum prediction for de novo structure elucidation: Assessment against experimental reference data and comparison to competitive fragmentation modeling*, *Int. J. Quantum Chem.* **118** (2018) e25460.
- [113] R. A. Marcus, *Unimolecular Dissociations and Free Radical Recombination Reactions*, *J. Phys. Chem.* **20** (1952) 359.
- [114] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties*, *Phys. Rev. B* **58** (11 1998) 7260.
- [115] M. Gaus, Q. Cui and M. Elstner, *DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB)*, *J. Chem. Theory Comput.* **7** (2011) 931.
- [116] W. Weber and W. Thiel, *Orthogonalization corrections for semiempirical methods*, *Theor. Chem. Acc.* **103** (2000) 495.
- [117] P. O. Dral, X. Wu, L. Spörkel, A. Koslowski, W. Weber, R. Steiger, M. Scholten and W. Thiel, *Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Theory, Implementation, and Parameters*, *J. Chem. Theory Comput.* **12** (2016) 1082.
- [118] R. S. Mulliken, *Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I*, *J. Chem. Phys.* **23** (1955) 1833.

-
- [119] A. Schäfer, H. Horn and R. Ahlrichs, *Fully optimized contracted Gaussian basis sets for atoms Li to Kr*, *J. Chem. Phys.* **97** (1992) 2571.
- [120] W. Thiel, *MNDO2005 program, version 7.0*, Max-Planck-Institut für Kohlen-forschung, Mülheim, Germany, 2005.
- [121] F. Neese, *The ORCA program system*, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2** (2012) 73.
- [122] F. Neese, *ORCA - an Ab initio, Density Functional and Semiempirical SCF-MO Package, Ver.4.0*, 2016.
- [123] F. Neese, *Software update: the ORCA program system, version 4.0: Software update*, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8** (2017) e1327.
- [124] W. Kaminsky, *Highly active metallocene catalysts for olefin polymerization*, *J. Chem. Soc., Dalton Trans.* **9** (1998) 1413.
- [125] N. Miyaura, K. Yamada and A. Suzuki, *A new stereospecific cross-coupling by the palladium-catalyzed reaction of 1-alkenylboranes with 1-alkenyl or 1-alkynyl halides*, *Tetrahedron Lett.* **20** (1979) 3437.
- [126] Y. Nakao, H. Imanaka, A. K. Sahoo, A. Yada and T. Hiyama, *Alkenyl- and Aryl[2-(hydroxymethyl)phenyl]dimethylsilanes: An Entry to Tetraorganosilicon Reagents for the Silicon-Based Cross-Coupling Reaction*, *J. Am. Chem. Soc.* **127** (2005) 6952.
- [127] N. R. Council, *Opportunities to Improve Airport Passenger Screening with Mass Spectrometry*, Washington, DC: The National Academies Press, 2004.
- [128] P. L. Urban, *Quantitative mass spectrometry: an overview*, *Philos. Trans. R. Soc. A* **374** (2016) 20150382.
- [129] F. C. Chernicharo, L. Modesto-Costa and I. Borges Jr., *Molecular dynamics simulation of the electron ionization mass spectrum of tabun*, *J. Mass Spectrom.* **55** (2020) e4513.
- [130] K. Scheubert, F. Hufsky and S. Böcker, *Computational mass spectrometry for small molecules*, *J. Cheminformatics* **5** (2013) 12.
- [131] T. Kind and O. Fiehn, *Advances in structure elucidation of small molecules using mass spectrometry*, *Bioanal. Rev.* **2** (2010) 23.
- [132] V. Ásgeirsson, C. A. Bauer and S. Grimme, *Unimolecular decomposition pathways of negatively charged nitriles by ab initio molecular dynamics*, *Phys. Chem. Chem. Phys.* **18** (45 2016) 31017.
- [133] J. F. Parcher, M. Wang, A. G. Chittiboyina and I. A. Khan, *In-source collision-induced dissociation (IS-CID): Applications, issues and structure elucidation with single-stage mass analyzers*, *Drug Test. Anal.* **10** (2018) 28.
- [134] F. W. McLafferty and T. A. Bryce, *Metastable-ion characteristics: characterization of isomeric molecules*, *Chem. Commun. (London)* (23 1967) 1215.

- [135] T. L. Kruger, J. F. Litton, R. W. Kondrat and R. G. Cooks, *Mixture analysis by mass-analyzed ion kinetic energy spectrometry*, *Anal. Chem.* **48** (1976) 2113.
- [136] F. W. McLafferty and F. M. Bockhoff, *Separation/identification system for complex mixtures using mass separation and mass spectral characterization*, *Anal. Chem.* **50** (1978) 69.
- [137] P. de Sainte Claire and W. L. Hase, *Thresholds for the Collision-Induced Dissociation of Clusters by Rare Gas Impact*, *J. Phys. Chem.* **100** (1996) 8190.
- [138] O. Meroueh and W. L. Hase, *Collisional Activation of Small Peptides*, *J. Phys. Chem. A* **103** (1999) 3981.
- [139] J. Liu, K. Song, W. L. Hase and S. L. Anderson, *Direct dynamics study of energy transfer and collision-induced dissociation: Effects of impact energy, geometry, and reactant vibrational mode in H_2CO^+ -Ne collisions*, *J. Chem. Phys.* **119** (2003) 3040.
- [140] W. H. Miller and S. J. Cotton, *Classical molecular dynamics simulation of electronically non-adiabatic processes*, *Faraday Discuss.* **195** (0 2016) 9.
- [141] S. Pratihar, X. Ma, Z. Homayoon, G. L. Barnes and W. L. Hase, *Direct Chemical Dynamics Simulations*, *J. Am. Chem. Soc.* **139** (2017) 3570.
- [142] M. Gu, J. Zhang, W. L. Hase and L. Yang, *Direct Dynamics Simulations of the Thermal Fragmentation of a Protonated Peptide Containing Arginine*, *ACS Omega* **5** (2020) 1463.
- [143] E. Martínez-Núñez, S. A. Vázquez and J. M. C. Marques, *Quasiclassical trajectory study of the collision-induced dissociation of CH_3SH^+ +Ar*, *J. Chem. Phys.* **121** (2004) 2571.
- [144] E. Martínez-Núñez, A. Fernández-Ramos, S. A. Vázquez, J. M. C. Marques, M. Xue and W. L. Hase, *Quasiclassical dynamics simulation of the collision-induced dissociation of $Cr(CO)_6^+$ with Xe*, *J. Chem. Phys.* **123** (2005) 154311.
- [145] R. Spezia, J.-Y. Salpin, M.-P. Gageot, W. L. Hase and K. Song, *Protonated Urea Collision-Induced Dissociation. Comparison of Experiments and Chemical Dynamics Simulations*, *J. Phys. Chem. A* **113** (2009) 13853.
- [146] R. Spezia, A. Martin-Somer, V. Macaluso, Z. Homayoon, S. Pratihar and W. L. Hase, *Unimolecular dissociation of peptides: statistical vs. non-statistical fragmentation mechanisms and time scales*, *Faraday Discuss.* **195** (0 2016) 599.
- [147] A. Martin-Somer, J. Martens, J. Grzetic, W. L. Hase, J. Oomens and R. Spezia, *Unimolecular Fragmentation of Deprotonated Diproline [Pro_2 -H] - Studied by Chemical Dynamics Simulations and IRMPD Spectroscopy*, *J. Phys. Chem. A* **122** (2018) 2612.
- [148] A. Carrá, V. Macaluso, P. W. Villalta, R. Spezia and S. Balbo, *Fragmentation Spectra Prediction and DNA Adducts Structural Determination*, *J. Am. Soc. Mass Spectrom.* **30** (2019) 2771.

-
- [149] R. Spezia, J. Martens, J. Oomens and K. Song, *Collision-induced dissociation pathways of protonated Gly₂NH₂ and Gly₃NH₂ in the short time-scale limit by chemical dynamics and ion spectroscopy*, *Int. J. Mass Spectrom.* **388** (2015) 40.
- [150] R. Spezia, S. B. Lee, A. Cho and K. Song, *Collision-induced dissociation mechanisms of protonated penta- and octa-glycine as revealed by chemical dynamics simulations*, *Int. J. Mass Spectrom.* **392** (2015) 125.
- [151] R. Rodríguez-Fernández, S. A. Vázquez and E. Martínez-Núñez, *Collision-induced dissociation mechanisms of [Li(uracil)]⁺*, *Phys. Chem. Chem. Phys.* **15** (20 2013) 7628.
- [152] A. Martin Somer, V. Macaluso, G. L. Barnes, L. Yang, S. Pratihar, K. Song, W. L. Hase and R. Spezia, *Role of Chemical Dynamics Simulations in Mass Spectrometry Studies of Collision-Induced Dissociation and Collisions of Biological Ions with Organic Surfaces*, *J. Am. Soc. Mass Spectrom.* **31** (2020) 2.
- [153] X. Hu, W. L. Hase and T. Pirraglia, *Vectorization of the general Monte Carlo classical trajectory program VENUS*, *J. Comput. Chem.* **12** (1991) 1014.
- [154] A. Carrá and R. Spezia, *In Silico Tandem Mass Spectrometer: an Analytical and Fundamental Tool*, *Chemistry–Methods* **1** (2021) 123.
- [155] M. Dole, H. L. Cox and J. Gieniec, “Electrospray Mass Spectroscopy”, *Polymer Molecular Weight Methods*, 1973, chap. 7 73.
- [156] J. Fenn, M. Mann, C. Meng, S. Wong and C. Whitehouse, *Electrospray ionization for mass spectrometry of large biomolecules*, *Science* **246** (1989) 64.
- [157] P. Kebarle and U. H. Verkerk, *Electrospray: From ions in solution to ions in the gas phase, what we know now*, *Mass Spectrom. Rev.* **28** (2009) 898.
- [158] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida and T. Matsuo, *Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry*, *Rapid Commun. Mass Spectrom.* **2** (1988) 151.
- [159] F. Hillenkamp, M. Karas, R. C. Beavis and B. T. Chait, *Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers*, *Anal. Chem.* **63** (1991) 1193A.
- [160] V. Gabelica and E. De Pauw, *Internal energy and fragmentation of ions produced in electrospray sources*, *Mass Spectrom. Rev.* **24** (2005) 566.
- [161] R. Rahrt, T. Auth, M. Demireva, P. B. Armentrout and K. Koszinowski, *Benzhydrylpyridinium Ions: A New Class of Thermometer Ions for the Characterization of Electrospray-Ionization Mass Spectrometers*, *Anal. Chem.* **91** (2019) 11703.
- [162] S. Nguyen and J. B. Fenn, *Gas-phase ions of solute species from charged droplets of solutions*, *Proc. Natl. Acad. Sci. U.S.A.* **104** (2007) 1111.

- [163] M. Dole, L. L. Mack, R. L. Hines, R. C. Mobley, L. D. Ferguson and M. B. Alice, *Molecular Beams of Macroions*, *J. Chem. Phys.* **49** (1968) 2240.
- [164] L. L. Mack, P. Kralik, A. Rheude and M. Dole, *Molecular Beams of Macroions. II*, *J. Chem. Phys.* **52** (1970) 4977.
- [165] J. V. Iribarne and B. A. Thomson, *On the evaporation of small ions from charged droplets*, *J. Chem. Phys.* **64** (1976) 2287.
- [166] B. A. Thomson and J. V. Iribarne, *Field induced ion evaporation from liquid surfaces at atmospheric pressure*, *J. Chem. Phys.* **71** (1979) 4451.
- [167] J. Fernandez de la Mora, *Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism*, *Anal. Chim. Acta* **406** (2000) 93.
- [168] A. L. McCormack, A. Somogyi, A. R. Dongré and V. H. Wysocki, *Fragmentation of protonated peptides: surface-induced dissociation in conjunction with a quantum mechanical approach*, *Anal. Chem.* **65** (1993) 2859.
- [169] J. Graton, M. Berthelot, J.-F. Gal, S. Girard, C. Laurence, J. Lebreton, J.-Y. Le Questel, P.-C. Maria and P. Nauš, *Site of Protonation of Nicotine and Nornicotine in the Gas Phase: Pyridine or Pyrrolidine Nitrogen?*, *J. Am. Chem. Soc.* **124** (2002) 10552.
- [170] A. R. Dongré, J. L. Jones, Á. Somogyi and V. H. Wysocki, *Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model*, *J. Am. Chem. Soc.* **118** (1996) 8365.
- [171] G. E. Reid, R. J. Simpson and R. A. J. O'Hair, *A mass spectrometric and ab initio study of the pathways for dehydration of simple glycine and cysteine-containing peptide $[M+H]^+$ ions*, *J. Am. Soc. Mass Spectrom.* **9** (1998) 945.
- [172] I. P. Csonka, B. Paizs, G. Lendvay and S. Suhai, *Proton mobility in protonated peptides: a joint molecular orbital and RRKM study*, *Rapid Commun. Mass Spectrom.* **14** (2000) 417.
- [173] 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, *Protomers: formation, separation and characterization via travelling wave ion mobility mass spectrometry*, *J. Mass Spectrom.* **47** (2012) 712.
- [174] K. Peckelsen, J. Martens, L. Czympiel, J. Oomens, G. Berden, D. Gründemann, A. J. H. M. Meijer and M. Schäfer, *Ergothioneine and related histidine derivatives in the gas phase: tautomer structures determined by IRMPD spectroscopy and theory*, *Phys. Chem. Chem. Phys.* **19** (34 2017) 23362.
- [175] M. J. van Stipdonk, M. J. Kullman, G. Berden and J. Oomens, *Infrared multiple-photon dissociation spectroscopy of deprotonated 6-hydroxynicotinic acid*, *Rapid Commun. Mass Spectrom.* **28** (2014) 691.
- [176] M. Almasian, J. Grzetic, J. van Maurik, J. D. Steill, G. Berden, S. Ingemann, W. J. Buma and J. Oomens, *Non-Equilibrium Isomer Distribution of the Gas-Phase Photoactive Yellow Protein Chromophore*, *J. Phys. Chem. Lett.* **3** (2012) 2259.

-
- [177] J. D. Steill and J. Oomens, *Gas-Phase Deprotonation of p-Hydroxybenzoic Acid Investigated by IR Spectroscopy: Solution-Phase Structure Is Retained upon ESI*, *J. Am. Chem. Soc.* **131** (2009) 13570.
- [178] S. Grimme, *Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations*, *J. Chem. Theory Comput.* **15** (2019) 2847.
- [179] 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>.
- [180] J. Mitchell Wells and S. A. McLuckey, "Collision-Induced Dissociation (CID) of Peptides and Proteins", *Biological Mass Spectrometry*, vol. 402, Methods in Enzymology, Academic Press, 2005 148.
- [181] I. V. Chernushevich, A. V. Loboda and B. A. Thomson, *An introduction to quadrupole-time-of-flight mass spectrometry*, *J. Mass Spectrom.* **36** (2001) 849.
- [182] S. O. Meroueh, Y. Wang and W. L. Hase, *Direct Dynamics Simulations of Collision- and Surface-Induced Dissociation of N-Protonated Glycine. Shattering Fragmentation*, *J. Phys. Chem. A* **106** (2002) 9983.
- [183] P. B. Armentrout, "Threshold Collision-Induced Dissociations for the Determination of Accurate Gas-Phase Binding Energies and Reaction Barriers", *Modern Mass Spectrometry*, ed. by C. A. Schalley, Berlin, Heidelberg: Springer Berlin Heidelberg, 2003 233.
- [184] T. M. Kertesz, L. H. Hall, D. W. Hill and D. F. Grant, *CE50: Quantifying collision induced dissociation energy for small molecule characterization and identification*, *J. Am. Soc. Mass Spectrom.* **20** (2009) 1759.
- [185] G. E. P. Box and M. E. Muller, *A Note on the Generation of Random Normal Deviates*, *Ann. Math. Statist.* **29** (1958) 610.
- [186] 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, *HMDB: the Human Metabolome Database*, *Nucleic Acids Res.* **35** (2007) D521.
- [187] 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. Bjorn Dahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner and A. Scalbert, *HMDB 3.0 - The Human Metabolome Database in 2013*, *Nucleic Acids Res.* **41** (2012) D801.
- [188] 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, *HMDB 4.0: the human metabolome database for 2018*, *Nucleic Acids Res.* **46** (2017) D608.

- [189] 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, *MassBank: a public repository for sharing mass spectral data for life sciences*, *J. Mass Spectrom.* **45** (2010) 703.
- [190] K. Volná, M. Holčapek, L. Kolářová, K. Lemr, J. Čáslavský, P. Kačer, J. Poustka and M. Hubálek, *Comparison of negative ion electrospray mass spectra measured by seven tandem mass analyzers towards library formation*, *Rapid Commun. Mass Spectrom.* **22** (2008) 101.
- [191] *Find the code on the ENSO GitHub Repository:*
<https://github.com/grimme-lab/enso>
and the documentation at:
https://xtb-docs.readthedocs.io/en/latest/enso_doc/enso.html.
- [192] S. Grimme, J. G. Brandenburg, C. Bannwarth and A. Hansen, *Consistent structures and interactions by density functional theory with small atomic orbital basis sets*, *J. Chem. Phys.* **143** (2015) 054107.
- [193] T. Husch, A. C. Vaucher and M. Reiher, *Semiempirical molecular orbital models based on the neglect of diatomic differential overlap approximation*, *Int. J. Quantum Chem.* **118** (2018) e25799.
- [194] S. Grimme, *Accurate description of van der Waals complexes by density functional theory including empirical corrections*, *J. Comput. Chem.* **25** (2004) 1463.
- [195] H. Neugebauer, F. Bohle, M. Bursch, A. Hansen and S. Grimme, *Benchmark Study of Electrochemical Redox Potentials Calculated with Semiempirical and DFT Methods*, *J. Phys. Chem. A* **124** (2020) 7166.
- [196] M. Bursch, L. Kunze, A. M. Vibhute, A. Hansen, K. M. Sureshan, P. G. Jones, S. Grimme and D. B. Werz, *Quantification of Non-covalent Interactions in Azide-Pnictogen, -Chalcogen, and -Halogen Contacts*, *Chem. Eur. J.* **27** (2021) 4627.
- [197] M. Bursch, A. Hansen and S. Grimme, *Fast and Reasonable Geometry Optimization of Lanthanoid Complexes with an Extended Tight Binding Quantum Chemical Method*, *Inorg. Chem.* **56** (2017) 12485.
- [198] S. Dohm, M. Bursch, A. Hansen and S. Grimme, *Semiautomated Transition State Localization for Organometallic Complexes with Semiempirical Quantum Chemical Methods*, *J. Chem. Theory Comput.* **16** (2020) 2002.
- [199] S. Spicher, M. Bursch and S. Grimme, *Efficient Calculation of Small Molecule Binding in Metal–Organic Frameworks and Porous Organic Cages*, *J. Phys. Chem. C* **124** (2020) 27529.
- [200] S. A. McLuckey and D. E. Goeringer, *SPECIAL FEATURE: TUTORIAL Slow Heating Methods in Tandem Mass Spectrometry*, *J. Mass Spectrom.* **32** (1997) 461.

-
- [201] N. Alygizakis and N. Thomaidis, *2,6-Dichlorobenzamide; LC-ESI-QTOF; MS2; CE: 40 eV; R=35000; [M+H]⁺*, Copyright (C) 2015 Department of Chemistry, University of Athens, 09.12.2015.
- [202] Y. Kakazu and H. Horai, *Caffeine; LC-ESI-QQ; MS2; CE:40 V; [M+H]⁺*, 19.01.2019.
- [203] D. Bier, R. Hartmann and M. Holschbach, *Collision-induced dissociation studies of caffeine in positive electrospray ionisation mass spectrometry using six deuterated isotopomers and one N1-ethylated homologue*, *Rapid Commun. Mass Spectrom.* **27** (2013) 885.
- [204] Y. Kakazu and H. Horai, *L-3-Methylhistidine; LC-ESI-QQ; MS2; CE:30 V; [M+H]⁺*, 19.01.2019.
- [205] Y. Kakazu and H. Horai, *L-3-Methylhistidine; LC-ESI-QQ; MS2; CE:40 V; [M+H]⁺*, 19.01.2019.
- [206] Y. Kakazu and H. Horai, *(R)-(-)-Phenylephrine; LC-ESI-QQ; MS2; CE:40 V; [M+H]⁺*, 19.01.2019.
- [207] Y. Kakazu and H. Horai, *(R)-(-)-Phenylephrine; LC-ESI-QQ; MS2; CE:50 V; [M+H]⁺*, 19.01.2019.
- [208] N. Alygizakis, K. Galani and N. Thomaidis, *Acetyl-sulfamethoxazole; LC-ESI-QTOF; MS2; CE: 20 eV; R=35000; [M+H]⁺*, Copyright (C) 2019 Department of Chemistry, University of Athens, 30.05.2019.
- [209] N. Alygizakis, K. Galani and N. Thomaidis, *Acetyl-sulfamethoxazole; LC-ESI-QTOF; MS2; CE: 30 eV; R=35000; [M+H]⁺*, Copyright (C) 2019 Department of Chemistry, University of Athens, 30.05.2019.
- [210] Y. Kakazu and H. Horai, *Piperidine; LC-ESI-QQ; MS2; CE:40 V; [M+H]⁺*, accessed: 08.March 2021, 19.01.2016.
- [211] E. Champarnaud and C. Hopley, *Evaluation of the comparability of spectra generated using a tuning point protocol on twelve electrospray ionisation tandem-in-space mass spectrometers*, *Rapid Commun. Mass Spectrom.* **25** (2011) 1001.
- [212] T. De Vijlder, D. Valkenburg, F. Lemière, E. P. Romijn, K. Laukens and F. Cuyckens, *A tutorial in small molecule identification via electrospray ionization-mass spectrometry: The practical art of structural elucidation*, *Mass Spectrom. Rev.* **37** (2018) 607.
- [213] P. Lecchi, J. Zhao, W. S. Wiggins, T.-H. Chen, P. F. Yip, B. C. Mansfield and J. M. Peltier, *A method for monitoring and controlling reproducibility of intensity data in complex electrospray mass spectra: A thermometer ion-based strategy*, *J. Am. Soc. Mass Spectrom.* **20** (2009) 398.
- [214] K. Levsen and H. Schwarz, *Gas-phase chemistry of collisionally activated ions*, *Mass Spectrom. Rev.* **2** (1983) 77.
- [215] M. M. Siegel, *Early Discovery Drug Screening Using Mass Spectrometry*, *Curr. Top. Med. Chem.* **2** (2002) 13.

- [216] E. Werner, J.-F. Heilier, C. Ducruix, E. Ezan, C. Junot and J.-C. Tabet, *Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends*, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **871** (2008) 143, Hyphenated Techniques for Global Metabolite Profiling.
- [217] M. Holčapek, R. Jirásko and M. Lísa, *Basic rules for the interpretation of atmospheric pressure ionization mass spectra of small molecules*, *J. Chromatogr. A* **1217** (2010) 3908, Mass Spectrometry: Innovation and Application. Part VI.
- [218] A. Weissberg and S. Dagan, *Interpretation of ESI(+)-MS-MS spectra—Towards the identification of “unknowns”*, *Int. J. Mass Spectrom.* **299** (2011) 158.
- [219] F. Allen, R. Greiner and D. Wishart, *Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification*, *Metabolomics* **11** (2015) 98.
- [220] M. Heinonen, A. Rantanen, T. Mielikäinen, J. Kokkonen, J. Kiuru, R. Ketola and J. Rousu, *FiD: A software for ab initio structural identification of product ions from tandem mass spectrometric data*, English, *Rapid Commun. Mass Spectrom.* **22** (2008) 3043.
- [221] C. Ruttkies, E. Schymanski, S. Wolf, J. Hollender and S. Neumann, *MetFrag relaunched: incorporating strategies beyond in silico fragmentation*, *Cheminform.* **8** (2016).
- [222] K. Dührkop, H. Shen, M. Meusel, J. Rousu and S. Böcker, *Searching molecular structure databases with tandem mass spectra using CSI:FingerID*, *Proc. Natl. Acad. Sci. U.S.A.* **112** (2015) 12580.
- [223] B. F. Minaev, M. I. Shafranyosh, Y. Y. Svida, M. I. Sukhoviya, I. I. Shafranyosh, G. V. Baryshnikov and V. A. Minaeva, *Fragmentation of the adenine and guanine molecules induced by electron collisions*, *J. Chem. Phys.* **140** (2014) 175101.
- [224] D. Ortiz, P. Martin-Gago, A. Riera, K. Song, J.-Y. Salpin and R. Spezia, *Gas-phase collision induced dissociation mechanisms of peptides: Theoretical and experimental study of N-formylalanyl amide fragmentation*, *Int. J. Mass Spectrom.* **335** (2013) 33.
- [225] E. Rossich Molina, A. Eizaguirre, V. Haldys, D. Urban, G. Doisneau, Y. Bourdreux, J.-M. Beau, J.-Y. Salpin and R. Spezia, *Characterization of Protonated Model Disaccharides from Tandem Mass Spectrometry and Chemical Dynamics Simulations*, *Chem. Phys. Chem.* **18** (2017) 2812.
- [226] E. M. Thurman, I. Ferrer, O. J. Pozo, J. V. Sancho and F. Hernandez, *The even-electron rule in electrospray mass spectra of pesticides*, *Rapid Commun. Mass Spectrom.* **21** (2007) 3855.
- [227] A. Steckel and G. Schlosser, *An Organic Chemist's Guide to Electrospray Mass Spectrometric Structure Elucidation*, *Molecules* **24** (2019) 611.
- [228] R. S. Johnson, D. Krylov and K. A. Walsh, *Proton mobility within electrosprayed peptide ions*, *J. Mass Spectrom.* **30** (1995) 386.

-
- [229] B. Paizs and S. Suhai, *Fragmentation pathways of protonated peptides*, *Mass Spectrom. Rev.* **24** (2005) 508.
- [230] P. Longevialle and R. Botter, *Evidence for intramolecular interaction between ionic and neutral fragments in the mass spectrometer*, *J. Chem. Soc., Chem. Commun.* (17 1980) 823.
- [231] P. Longevialle and R. Botter, *Electron impact mass spectra of bifunctional steroids. The interaction between ionic and neutral fragments derived from the same parent ion*, *Org. Mass Spectrom.* **18** (1983) 1.
- [232] P. Longevialle, *Ion-neutral complexes in the unimolecular reactivity of organic cations in the gas phase*, *Mass Spectrom. Rev.* **11** (1992) 157.
- [233] J. Terlouw, W. Heerma, G. Dijkstra, J. Holmes and P. Burgers, *Characterization of ion-dipole complexes by collisional activation and collisional ionization spectrometry*, *Int. J. Mass Spectrom. Ion Phys.* **47** (1983) 147.
- [234] C. A. Schalley, D. Schröder and H. Schwarz, *Intermediacy of Proton-Bound Dimers and Ion/Dipole Complexes in the Unimolecular Decompositions of Dialkyl-Peroxide Radical Cations: Evidence for a coupled proton and hydrogen-atom transfer*, *Helv. Chim. Acta* **78** (1995) 1999.
- [235] H. Friedrichs, G. A. McGibbon and H. Schwarz, *Double hydrogen atom transfer in lactamide radical cations via ion-neutral complexes*, *Int. J. Mass Spectrom. Ion Processes* **152** (1996) 217.
- [236] Z. C. Lipton, *The Mythos of Model Interpretability: In Machine Learning, the Concept of Interpretability is Both Important and Slippery.*, *Queue* **16** (2018) 31.
- [237] J. Lee, T. Kind, D. J. Tantillo, L.-P. Wang and O. Fiehn, *Evaluating the Accuracy of the QCEIMS Approach for Computational Prediction of Electron Ionization Mass Spectra of Purines and Pyrimidines*, *Metabolites* **12** (2022) 68.
- [238] S. Wang, T. Kind, P. L. Bremer, D. J. Tantillo and O. Fiehn, *Quantum Chemical Prediction of Electron Ionization Mass Spectra of Trimethylsilylated Metabolites*, *Anal. Chem.* **94** (2022) 1559.
- [239] M. Barber, R. Bordoli, R. Sedgwick and A. N. Tyler, *Fast atom bombardment of solids as an ion source in mass*, *Nature* **293** (1981) 270.
- [240] A. G. Harrison and T. Yalcin, *Proton mobility in protonated amino acids and peptides*, *Int. J. Mass Spectrom. Ion Processes* **165-166** (1997) 339.
- [241] N. B. Cech and C. G. Enke, *Practical implications of some recent studies in electrospray ionization fundamentals*, *Mass Spectrom. Rev.* **20** (2001) 362.
- [242] P. Liigand, K. Kaupmees, K. Haav, J. Liigand, I. Leito, M. Girod, R. Antoine and A. Kruve, *Think Negative: Finding the Best Electrospray Ionization/MS Mode for Your Analyte*, *Anal. Chem.* **89** (2017) 5665.
- [243] A. T. Iavarone, J. C. Jurchen and E. R. Williams, *Supercharged protein and peptide ions formed by electrospray ionization*, *Anal. Chem.* **73** (2001) 1455.

- [244] B. L. Frey, D. T. Ladrer, S. B. Sondalle, C. J. Krusemark, A. L. Jue, J. J. Coon and L. M. Smith, *Chemical derivatization of peptide carboxyl groups for highly efficient electron transfer dissociation*, *J. Am. Soc. Mass Spectrom.* **24** (2013) 1710.
- [245] A. T. Iavarone and E. R. Williams, *Mechanism of charging and supercharging molecules in electrospray ionization*, *J. Am. Chem. Soc.* **125** (2003) 2319.
- [246] S. Banerjee and S. Mazumdar, *Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte*, *Int. J. Anal. Chem.* (2012) 1.
- [247] G. J. Van Berkel, F. Zhou and J. T. Aronson, *Changes in bulk solution pH caused by the inherent controlled-current electrolytic process of an electrospray ion source*, *Int. J. Mass Spectrom.* **162** (1997) 55.
- [248] J. A. Loo, H. R. Udseth, R. D. Smith and J. H. Futrell, *Collisional effects on the charge distribution of ions from large molecules, formed by electrospray-ionization mass spectrometry*, *Rapid Commun. Mass Spectrom.* **2** (1988) 207.
- [249] S. Zhou, A. G. Edwards, K. D. Cook and G. J. Van Berkel, *Investigation of the electrospray plume by laser-induced fluorescence spectroscopy*, *Anal. Chem.* **71** (1999) 769.
- [250] S. Zhou, B. S. Prebyl and K. D. Cook, *Profiling pH changes in the electrospray plume*, *Anal. Chem.* **74** (2002) 4885.
- [251] M. Girod, X. Dagany, R. Antoine and P. Dugourd, *Relation between charge state distributions of peptide anions and pH changes in the electrospray plume. A mass spectrometry and optical spectroscopy investigation*, *Int. J. Mass Spectrom.* **308** (2011) 41.
- [252] S. Zhou and K. D. Cook, *Protonation in electrospray mass spectrometry: wrong-way-round or right-way-round?*, *J. Am. Soc. Mass Spectrom.* **11** (2000) 961.
- [253] B. M. Ehrmann, T. Henriksen and N. B. Cech, *Relative importance of basicity in the gas phase and in solution for determining selectivity in electrospray ionization mass spectrometry*, *J. Am. Soc. Mass Spectrom.* **19** (2011) 719.
- [254] M. Koné, B. Illien, C. Laurence, J.-F. Gal and P.-C. Maria, *Are nicotinoids protonated on the pyridine or the amino nitrogen in the gas phase?*, *J. Phys. Org. Chem.* **19** (2006) 104.
- [255] V. Vais, A. Etinger and A. Mandelbaum, *Intramolecular proton transfers in stereoisomeric gas-phase ions and the kinetic nature of the protonation process upon chemical ionization*, *J. Mass Spectrom.* **34** (1999) 755.
- [256] V. Vais, A. Etinger and A. Mandelbaum, *Steric hindrance and the kinetic nature of the protonation process in 4-amino-1-methoxycyclohexanes upon chemical ionization*, *Eur. J. Mass Spectrom.* **5** (1999) 449.
- [257] Y.-P. Tu, *Dissociative protonation sites: reactive centers in protonated molecules leading to fragmentation in mass spectrometry*, *J. Org. Chem.* **71** (2006) 5482.

-
- [258] J. Zhao, C.-K. Siu, T. Shi, A. C. Hopkinson and K. W. M. Siu, *Abundant Dipositively Charged Protonated a2 and a3 Ions from Diproline and Triproline*, *J. Phys. Chem. B* **113** (2009) 4963.
- [259] K. Vékey, A. G. Brenton and J. H. Beynon, *Electron capture induced decomposition of the benzene C₆H₆²⁺ ion*, *J. Phys. Chem.* **90** (1986) 3569.
- [260] A. Klamt, *Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena*, *J. Phys. Chem.* **99** (1995) 2224.
- [261] A. Klamt, V. Jonas, T. Bürger and J. C. W. Lohrenz, *Refinement and Parametrization of COSMO-RS*, *J. Phys. Chem. A* **102** (1998) 5074.
- [262] F. Neese, *Software update: The ORCA program system—Version 5.0*, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **n/a** () e1606.
- [263] *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>.*
- [264] F. Weigend, M. Häser, H. Patzelt and R. Ahlrichs, *RI-MP2: optimized auxiliary basis sets and demonstration of efficiency*, *Chem. Phys. Lett.* **294** (1998) 143.
- [265] C. Collette and E. De Pauw, *Calibration of the internal energy distribution of ions produced by electrospray*, *Rapid Commun. Mass Spectrom.* **12** (1998) 165.
- [266] H. H. Kakazu Y, *L-Tryptophan; LC-ESI-QQ; MS2; CE:20 V; [M-H]⁻*, 19.01.2016.
- [267] M. S. B. Munson and F. H. Field, *Chemical Ionization Mass Spectrometry. I. General Introduction*, *J. Am. Chem. Soc.* **88** (1966) 2621.

List of Figures

1.1	Schematic comparison between the concepts of Electron Ionization (top) and Collision Induced Dissociation (bottom). While activation, energy is introduced into the system. For EI: Impact Excess Energy (E_{IEE}) after electron impact; for CID: Kinetic Energy (E_{Kin}). Internal conversion leads to rearrangement and fragmentation.	2
1.2	The motivation of designing the QCxMS workflow. The measurement of an unknown compound can be compared to database entries. If these are not available, either cost expensive experiments or the <i>ab-initio</i> QCxMS program can be used as reference.	4
2.1	Flowchart of the QCxMS protocol for the EI (left) and CID (right) run-modes.	15
2.2	The three run-types developed for the CID mode of QCxMS.	18
A.1	Flowchart of the QCEIMS protocol.	37
A.2	The benchmark set. Molecules (1) – (6) represent the organic group, molecules (7) – (9) represent the transition-metal group and molecules (10) – (15) represent the main-group inorganic group.	40
A.3	Total calculation time of all test-set molecules in hours. The molecules are grouped in their corresponding categories. GFN1-xTB results are shown in blue, and GFN2-xTB results are shown in yellow.	41
A.4	Average failure percentage of all test-set molecules. The molecules are grouped in their corresponding categories. GFN1-xTB results are shown in blue, and GFN2-xTB results are shown in yellow.	42
A.5	Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the organic compounds (a) 1-butanol, (b) hexafluorobenzene, and (c) uracil. The structures of the precursor ion (denoted $M^{\bullet+}$) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.	44
A.6	Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the organic compounds (a) testosterone, (b) sucrose, and (c) leucylglycylglycine. The structures of the precursor ion (denoted $M^{\bullet+}$) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.	45
A.7	The EI-MS of ferrocene as an example of the impact of falsely calculated ionization potentials. The red circle marks the signal of Fe^+ that is not being reproduced by GFN1/IPEA-xTB (left). The usage of GFN1/GFN2-xTB (right) accounts for the correct signal.	47

A.8	Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the transition metal compounds (a) bis(benzene)chromium, (b) zirconocene dichloride, and (c) nickel(II)bis(diphenylacetylacetonate). The structures of the precursor ion (denoted M^{*+}) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.	48
A.9	Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the main-group inorganic compounds (a) bis(pinacolato)diboron, (b) chinalphos, and (c) triphenylstibine. The structures of the precursor ion (denoted M^{*+}) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.	50
A.10	Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the organic compounds (a) dichloro(ethyl)aluminum, (b) 2-(dimethyl(naphthalen-1-yl)silyl)phenylmethanol, and (c) octasulfur. The structures of the precursor ion (denoted M^{*+}) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.	52
A.11	The EI-MS of bis(benzene)chromium (a), zirconocene-dichloride (b) and nickel(II)bis(diphenylacetylacetonate)(c). The parent ion has been superimposed onto the corresponding spectrum.	57
B.1	Flowchart of the steps in the CID module.	65
B.2	Collision kinetics for the caffeine precursor ion equilibrated at 600 K. (a) SY at 40 to 120 eV E_{LAB} with one, two, and three collisions. (b) SY after 1 to 8 collisions events at 40, 50, and 60 eV E_{LAB} . (c) Velocity of $[M+H]^+$ after 1 to 8 collisions at 40 and 60 eV ELAB compared to theory. ⁸⁴ (d) Correlation between E_{COM} and E_{int} after 1 to 8 collisions at 40 eV ELAB. (e) SY after internal energy scaled from 5 to 12 eV.	69
B.3	Calculated spectra of tetrahydrofuran using single collisions at (a) (1) 20 eV, (2) 30 eV, (3) 40 eV, (4) 50 eV, and (5) 60 eV E_{LAB} . The most pertinent fragments m/z 31, 45, 55, and 73 are colored for better traceability. (b) Comparison of the 50 eV calculated (black) to the 40 eV literature spectrum (red, inverted). ⁴²	71
B.4	Calculated spectra at 100 eV with a single collision of the four energetically lowest protomers of 2,6-dichlorobenzamide (black, top) compared to experimental spectra taken from the literature ²⁰¹ at 40 eV (red, inverted). The protomer structures are superimposed at the corresponding spectrum with their respective difference in total energy.	72
B.5	Comparison of calculated spectra (black) vs. 40 eV literature spectra ²⁰² (red,inverted) of the most populated caffeine protomer at 600 K. (a) Single collision at 100eV E_{LAB} ; (b) six collisions at 100eV E_{LAB} , <i>fgc off</i> ; (c) six collisions at 100eV E_{LAB} , <i>fgc on</i> . (d) scaling of the internal energy to 10 eV.	74
B.6	Calculated spectra (black) compared to experiment ⁹⁸ (red, inverted) with collision energies in E_{LAB} in parentheses (theor./expr.). (a) 3-Methylhistidine (50/30), (b) 3-methylhistidine (60/40), (c) phenylephrine (60/40), (d) phenylephrine (70/50), (e) acetyl-sulfamethoxazole (40/20), (f) acetyl-sulfamethoxazole (60/30).	77

B.7	Protomer spectra of caffeine calculated at 60 eV E_{LAB} using the <i>general activation</i> run-type (black) vs. 40 eV E_{LAB} literature spectra (red,inverted). Energy rankings can be found in table B.2.	82
B.8	Protomer spectra of 3-methylhistidine at 50 eV E_{LAB} using the <i>general activation</i> run-type (black) vs. 30 eV E_{LAB} literature spectra (red,inverted). Energy rankings can be found in table B.2.	83
B.9	Protomer spectra of phenylephrine at 60 eV E_{LAB} using the <i>general activation</i> run-type (black) vs. 40 eV E_{LAB} literature spectra (red,inverted). Energy rankings can be found in table B.2.	84
B.10	(a) Calculated spectra of piperidin using single collisions at (1) 30 eV, 2) 40 eV, 3) 50 eV, 4) 60 eV, and 5) 70 eV E_{LAB} .The most concise fragments m/z 30, 44, 69, and 86 are colored for better traceability. (b) Calculated spectra of piperidin using single collision at 60 eV E_{LAB} vs. 40 eV E_{LAB} literature spectrum (red,inverted) ²¹⁰ . Instrument: API3000, Applied Biosystems.	84
B.11	Calculated (40 eV E_{LAB}) of the most populated protomer #1 compared to literature spectra at (a) 20 eV E_{LAB} and (b) 30 eV E_{LAB} of acetyl-sulfamethoxazole at 600 K. (c) Lewis structures of protomers #2, #4, and #5. The energy rankings can be found in table B.2. (d) Zoomed into the spectrum in between m/z 100 and m/z 240. (e) Lewis structure of protomer #6. (f) The spectrum of protomer #6 between m/z 100 and m/z 240.	85
B.12	Calculated spectra at 40 eV E_{LAB} with increasing collision cell length. Test structures were protomer #1 of caffeine ((a),(c) and (e)) and protomer #3 of acetyl-sulfamethoxazole ((b),(d) and (f)). (a),(b) $L_{Coll.cell}$ 0.12 m; (c),(d) $L_{Coll.cell}$ 0.25 m; (e),(f) $L_{Coll.cell}$ 0.35 m.	86
B.13	Calculated spectra at 40 eV E_{LAB} with increasing E_{int} pre-scaling. Test structures were protomer #1 of 3-methylhistidine ((a),(c) and (e); $L_{Coll.cell}$ 12.5 cm) and protomer #1 of caffeine ((b),(d) and (f); $L_{Coll.cell}$ 25 cm). (a) E_{int} 3 eV; (b) E_{int} 2 eV; (c) E_{int} 4 eV; (d) E_{int} 5 eV; (e) E_{int} 5 eV; (f) E_{int} 6 eV.	88
B.14	Comparison of calculated single collision spectra (black) vs. literature spectra (red,inverted) of tetrahydrofuran. (a) 30 eV vs. 10 eV ⁹⁷ ; (b) 30 eV vs. 25 eV ⁹⁷ ; (c) 50 eV vs. 10 eV ⁹⁷ ; (d) 50 eV vs. 25 eV ⁹⁷	89
B.15	Calculated spectrum of 4.5-Dihydroorotic-acid using single collisions at (a) 40 eV, (b) 60 eV, (c) 80 eV, and (d) 100eV E_{LAB} . (e) Comparison of the 40eV calculated (black) to the 40 eV literature spectrum (red,inverted) ⁹⁷ ; (f) Comparison of the 100eV calculated (black) to the 40 eV literature spectrum (red,inverted). ⁹⁷	90
C.1	Protomers I–IV of nateglinide and their relative energies (kcal/mol) related to the protomer with the lowest protonation energy as obtained by PBEh-3c DFT calculations.	96
C.2	Comparison of the simulated ESI-MS/MS spectrum computed by QCxMS@GFN2-xTB with the experimental spectra (red, inverted) of nateglinide. The spectra for all three considered protomer structures are averaged with equal weight.	97

C.1	Proposed fragmentation pathways of nateglinide. Fragmentation according to a one-channel fragmentation (black arrows) or branched fragmentation pathways (highlighted in coloured arrows; CMF green, CRF purple). $\sim H^+$ means proton migration. Boxed fragments were experimentally detected. Unboxed structures are “snapshots” on calculated trajectories and are displayed for clearer retracing of the reaction pathways; they are not global minima on the potential energy surface. For matters of clarity, the neutral fragment has not been depicted in all cases.	99
C.3	Protomers of zopiclone and their energies (kcal/mol) related to the protomer with the lowest protonation energy as obtained by PBEh-3c DFT calculations.	100
C.2	Proposed fragmentation pathways of zopiclone – part 1. Fragmentation according to a one-channel fragmentation (black arrows) or branched fragmentation pathways (highlighted in coloured arrows; CMF green, CRF purple). $\sim H^+$ means proton migration. Boxed fragments were experimentally detected. Unboxed structures are “snapshots” on calculated trajectories and are displayed for clearer retracing of the reaction pathways; they are not global minima on the potential energy surface. For matters of clarity, the neutral fragment has not been depicted in all cases.	101
C.3	Proposed fragmentation pathways of zopiclone – part 2. Fragmentation according to a one-channel fragmentation (black arrows) or branched fragmentation pathways (highlighted in coloured arrows; CMF green, CRF purple). $\sim H^+$ means proton migration. Boxed fragments were experimentally detected. Unboxed structures are “snapshots” on calculated trajectories and are displayed for clearer retracing of the reaction pathways; they are not global minima on the potential energy surface. For matters of clarity, the neutral fragment has not been depicted in all cases.	102
C.4	Comparison of the simulated ESI-MS/MS spectrum computed by GFN1-xTB (energy window 20 kcal/mol) with the experimental spectra (red, inverted) of zopiclone. The spectra for all five considered protomer structures are averaged with equal weight.	103
C.8	3D structures of the different nateglinide protomers I (left), II (middle) and III (right).	106
C.9	QCxMS simulated spectra of the three nateglinide protomers I–III.	106
C.10	QCxMS simulated spectra of the five zopiclone protomers V–IX.	107
D.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.	114
D.2	Benchmark set of molecules for testing different charge states with QCxMS.	116
D.3	Calculated spectra of 2-Ketobutyric 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.	119
D.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.	120

D.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.	121
D.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.	122
D.7	Spectra calculated 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}^{266} (blue, inverted). Protomer structures, relative free energies, and matching scores are added to their respective spectrum.	123
D.8	Dissociation reactions as proposed in ref. ¹⁰³ are marked with black roman numerals and arrows. Alternative fragmentation reactions computed with QCxMS are indicated with gray roman numerals and arrows.	125
D.9	a) measured spectrum of singly protonated crizotinib (LC-ESI-QTOF; 23 eV E_{LAB}) ¹⁰³ . 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.	126
D.10	a) Full scan MS of crizotinib (LC-ESI-QTOF; 25 eV E_{LAB}) ¹⁰³ . 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.	127
D.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. ¹⁰⁴	129
D.12	CID fragmentation analysis proposed in the reference ¹⁰⁴ . Roman numerals were used for the product ions of spectrum in Figure D.11. PI is the precursor ion.	130
D.13	Calculated spectrum of protomer #2 of 2-Ketobutyric acid (black,top) computed at 20 eV E_{LAB} compared to a measured spectrum (LC-ESI-QQQ) at 25 eV E_{LAB} (blue, inverted). The protomer structure, was added to the spectrum. a) GFN2-xTB//GFN2-xTB. b) GFN2-xTB//PBE0-ma-def2-TZVP.	134
D.14	GFN2-xTB//PBE0-ma-def2-TZVP 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.	135
D.15	Calculated spectra (black, top) of four protomers #1 and #2 of tryptophan at 35 eV E_{LAB} using GFN2-xTB//PBE0/ma-def2-TZVP compared to measured spectra (LC-ESI-QQ) at 20eV E_{LAB} (blue, inverted). Protomer structures, relative free energies, and matching scores are added to their respective spectrum.	136
D.16	Computed spectra of crizotinib using GFN2-xTB//PBE0/def2-TZVP at 50 eV E_{LAB} with isotope patterns computed with PlotMS. a) Singly protonated protomer #1. b) Doubly protonated protomer #1.	137
D.17	Computed spectrum of crizotinib m/z 130 as precursor ion. Computations conducted at GFN2-xTB//GFN2-xTB level with 70 eV E_{LAB}	138

List of Figures

D.18 Computed spectra of crizotinib using GFN2-xTB//GFN2-xTB at 50 eV E_{LAB} with isotope patterns computed with PlotMS. a) Protomer #2. b) Protomer #3. c) Doubly protonated protomer #2. 139

List of Tables

A.1	Total computational time (in hours) of all production runs summed up for different types of molecules and methods.	55
A.2	Number of successful production runs. For each system there have been a total of 1000 parallel production runs.	56
B.1	Individual experimental settings and specifications.	80
B.2	Relative G_{tot} [kcal/mol] of protonated structures to most stable structure protomer #1 (0.0 kcal/mol) calculated using ENSO at PBEh-3c level. The ensembles were sorted according to the simulation temperature at 600 K. Identical and improper structures were sorted out.	81
C.1	Fragment list of nateglinide. Fragments measured by ESI-MS/MS with intensity >1.5% are listed according to their molecular mass. Fragments simulated by QCxMS are marked. Classification of the rule-based fragmentation ranking as described before.	98
C.2	Fragment list of zopiclone. Fragments measured by ESI-MS/MS with intensity >1.5% are listed according to their molecular mass. Fragments simulated by QCxMS are marked. Classification of the rule-based fragmentation ranking as described before.	104
C.3	Comparison of the m/z values of the measured fragment of nateglinide with the calculated values based on the sum formula.	108
C.4	Comparison of the m/z values of the measured fragment of zopiclone with the calculated values based on the sum formula.	109
D.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 $\sum 2-0 < \sum 1-1 < \sum 0-2$	115
D.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	124
D.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)	124
D.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)	128

List of Tables

D.5	Signals found in the experiment and the calculations with their corresponding charge state	128
D.6	Relative free energies of the negative ion benchmark structures in the gas phase and in methanol computed with CENSO at different temperatures. [kcal/mol]	132
D.7	Relative free energies of the first and second protonation of crizotinib in the gas phase and in water computed with CENSO at 300 K. [kcal/mol]	133

Acknowledgements

A lot has happened throughout the entire course of my study and I would like to thank all the people who supported me in not giving up, making this dissertation possible in the first place.

First, I like to thank Stefan Grimme for believing in me and allowing me to work on this complex, but interesting, frustrating, but rewarding topic. I can't imagine having found any other subject even remotely as exciting. I would also like to thank him for the financial and thematic support he has given me over the years.

I like to thank is my second reviewer, Thomas Bredow, who already supported me throughout my bachelor's and master's thesis and was the reason I started pursuing a career in quantum chemistry in the first place. I hope his guidance and patience with his students will never be lost on him.

Furthermore, my thanks go out to all my colleagues at the Mulliken Center who have always supported me in the development, presentation, and debugging of my work. I will definitely miss the relaxed working atmosphere and the cheerful conversations on the roof terrace. I would like to thank a few of my colleagues in particular, without whom my work might not ever come to a conclusion. Hagen Neugebauer and Sebastian Ehlert have always supported me when I needed any kind of help, be it thematic, programming technical or just to lighten my mood. Thank you both so much, I don't think I could have done it without you guys! Also, I like to thank Dr. Markus Bursch, who guided a lot of my work in the right direction and always succeeded to increase my effort while decreasing my stress levels. For relaxed conversations and some special memories, I would like to thank especially Joachim Laun, Sebastian Spicher, Julius Stückrath, Dr. Jan-Michael Mewes, and Christian Hölzer. Furthermore, I would like to thank Johannes Gorges, Sarah Löffelsender, Eike Calderweyher, Fabian Bohle, Sebastian Dohm, Philipp Pracht, Christoph Bannwarth, Marcel Stahn, Marcel Müller, Thomas Rose, Thomas Gasevic, Christoph Plett, Andreas Hansen, and Albert Katbashev for all the scientific, non-scientific and sometimes strange discussions. I really enjoyed all of it! Special thanks for proof-reading this thesis goes out to Julius Stückrath, Hagen Neugebauer, Sebastian Ehlert, Christian Hölzer, and Johannes Gorges. For administrative and technical support I would like to thank Claudia Kronz and Jens Meikelburger.

Besides the academic community, there are a lot of people that supported me throughout the years. The majority of my academic life, I was supported by my band members, who helped me through some tough episodes. I like to thank Jonas Schmitz and Sebastian Rami for their moral support and their uplifting persona, helping me focus on the good parts of life. A truly great anchor in the most difficult times for me was Christoph Hager, who was there when others left. Thank you for your constant support and keep on being that interested and positive person! No less important for me was Moritz Ziegert, who made difficult situations feel like the best of times. Thanks to you, I was able to cope with major upheavals and small pandemics, and I would like to thank you for that and so much more! Also, I would like to thank Anna Fischer and Beatrice Karduck for believing in me and for their

Acknowledgements

constant support and uplifting words. You were a great motivational factor for not giving up when frustration set in. I thank Charlotte Zerfin, Sebastian Buchholz, Frank Göring, and Malte Laub for being such amazing long-time companions, it is always good to know that I can count on you. I thank the Fischer family and Simon Schmitz for their great support and good times. Last, I want to thank my family, some members more than others.