Christoph Bauer

Quantum Chemical Calculation of Electron Ionization Mass Spectra

Theoretische Chemie

Quantum Chemical Calculation of Electron Ionization Mass Spectra

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften im Fachbereich Chemie
der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Christoph Bauer
aus Baden bei Wien, Österreich
– Bonn, 2017–

Dekan:	Prof. Dr. Johannes Beck
Erster Gutachter:	Prof. Dr. Stefan Grimme
Zweite Gutachterin:	Prof. Dr. Barbara Kirchner
Tag der Mündlichen Prüfung:	24.01.2018
Erscheinungsjahr:	2018

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

Statement of Authorship

I, Christoph Bauer, hereby declare that I am the sole author of this thesis. The ideas and work of others, whether published or unpublished, have been fully acknowledged and referenced in my thesis.

Bonn, July 25, 2017

Publications

Parts of this thesis have already been published in international, peer-reviewed journals:

1. C. A. Bauer, A. Hansen and S. Grimme, "The Fractional Occupation Number Weighted Density as a Versatile Analysis Tool for Molecules with a Complicated Electronic Structure", Chem.–Eur. J., **2017**, *23*, 6150–6164.

2. C. A. Bauer and S. Grimme, "How to Compute Electron Ionization Mass Spectra from First Principles", J. Phys. Chem. A, **2016**, *120*, 3755–3766.

3. C. A. Bauer and S. Grimme, "First principles calculation of electron ionization mass spectra for selected organic drug molecules", Org. Biomol. Chem., **2014**, *12*, 8737–8744.

4. 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, **2014**, *118*, 11479–11484.

5. 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., **2015**, *21*, 125–140.

6. V. Ásgeirsson, C. A. Bauer and S. Grimme, "Quantum chemical calculation of electron ionization mass spectra for general organic and inorganic molecules ", Chem. Sci., **2017**, *8*, 4879–4895

Further publications:

1. V. Ásgeirsson, C. A. Bauer and S. Grimme, "Unimolecular decomposition pathways of negatively charged nitriles by ab initio molecular dynamics", Phys. Chem. Chem. Phys., **2016**, *18*, 31017–31026.

2. M. Engeser, C. Mundt, C. Bauer and S. Grimme, "N-Methylimidazolidin-4-one Organocatalysts: Gas-phase Fragmentations of Radical Cations by Experiment and Theory", J. Mass Spectrom., **2017**, *52*, 452–458

Presentations:

1. Poster: "First Principles Calculation of Electron Ionization Mass Spectra", *Symposium on Theoretical Chemistry*, **2014**, Vienna, Austria

2. Talk: "First Principles Calculation of Electron Ionization Mass Spectra", *Annual Meeting of the German Society for Mass Spectrometry*, **2015**, Wuppertal, Germany

3. Talk: "Electron Ionization Induced Fragmentations of Organometallic Compounds from First Principles", *Annual Meeting of the German Society for Mass Spectrometry*, **2016**, Hamburg, Germany

4. Poster: "Benchmarking Quantum Chemical Methods for Radical Cation Fragmentation Reactions", *Symposium on Theoretical Chemistry*, **2015**, Potsdam, Germany

5. Poster: "On the Importance of Statistical Charge Assignment for the Computation of Electron Ionization Mass Spectra", *AIP Future in Chemical Physics Meeting*, **2016**, Oxford, United Kingdom

6. Poster: "Application of the Fractional Occupation Number Weighted Density Analysis", *Symposium on Theoretical Chemistry*, **2016**, Bochum, Germany

7. Poster: "The Fractional Occupation Number Weighted Density (FOD) Analysis for the Modeling of Molecules with a Complicated Electronic Structure", *Young Chemists' Forum Spring Symposium*, **2017**, Mainz, Germany

Abstract

This thesis reports the computation of electron ionization (EI) mass spectra using a method that combines statistical theory and molecular dynamics. Due to the complexity of the unimolecular reaction space, not all competing fragmentation pathways can be fully treated in an *ab initio* way using a purely statistical framework. The main idea behind the present simulation protocol is to use approximate quantum chemical potential energy surfaces and simple internal energy distributions to discover the reaction pathways and barriers, and thus the relative rate constants automatically. This idea was proposed, implemented and published in late 2013 by my thesis supervisor Stefan Grimme, and termed QCEIMS.

The first part of this thesis gives a brief overview over the physical chemistry of EI mass spectrometry and the most important theoretical methods that I have used. These involve finite-temperature density functional theory and the semi-empirical Geometries, Frequencies and Noncovalent Interaction eXtended Tight Binding Hamiltonian (GFN-xTB). The energies and forces computed at these levels of theory are the input for the subsequent Born-Oppenheimer molecular dynamics simulations.

The second part deals with the application of finite-temperature density functional theory. The results show that the fractional occupation number weighted density ρ^{FOD} can be used as a measure for static electron correlation in biradicals and related systems, and that the fractional occupation numbers can be useful for the first guess at a multiconfigurational wave function. Furthermore, potential energy surfaces along model reaction coordinates are explored and the transferability of the ρ^{FOD} concept to semi-empirical quantum chemistry is shown.

The third part shows the main results of this work related to EI mass spectrometry. In Chapter 4, the literature is reviewed and the "Quantum Chemistry Electron Ionization Mass Spectra" (QCEIMS) method is presented. It is then evaluated concerning the assignment of the charge to a fragment using a series of ethanol homologues. A small mass spectrometric benchmark study is also included, showing that isomers can be distinguished by QCEIMS predicted EI mass spectra, provided their fragmentation pathways differ substantially.

In Chapters 5, 6, and 7 QCEIMS applications to large drug molecules, the nucleobase adenine and other nucleobases, are presented. For each case, the fragmentation pathways are analyzed, thereby elucidating the structures of the fragment ions.

Finally, in Chapter 8, predicted EI mass spectra for 23 compounds across the whole periodic table are shown. This has been made possible by V. Ásgeirsson's implementation of GFN-xTB into QCEIMS. This robust and efficient method performs remarkably well for organic molecules as well as organometallic compounds and main group inorganic systems while reducing the computational cost by a factor of 1,000 when compared to hybrid density functional calculations.

Zusammenfassung

In dieser Doktorarbeit berichte ich über die Berechnung von Elektronimpakt (EI)-Massenspektren. Dazu habe ich einen kombinierten statistischen und molekulardynamischen Ansatz verwendet. Aufgrund der Komplexität des unimolekularen Reaktionsraums können nicht alle konkurrierenden Fragmentationspfade vollkommen *ab initio* und im Rahmen der statistischen Theorie behandelt werden. Die Hauptidee hinter dem vorliegenden Simulationsprotokoll zur Berechnung von EI-Massenspektren ist die Verwendung einer genäherten quantenchemischen Potentialhyperfläche und einer modellhaften internen Energieverteilung zur automatischen Bestimmung der Reaktionsbarrierend und damit der relativen Reaktionsraten. Diese Idee wurde von meinem Doktorarbeitsbetreuer Stefan Grimme 2013 implementiert und publiziert.
Im ersten Teil dieser Doktorarbeit gebe ich einen kurzen Überblick über die physikalische Chemie der EI-Massenspektrometrie und stelle die wichtigsten theoretischen Methoden, die ich verwendet habe, vor. Diese umfassen die Dichtefunktionaltheorie unter Einsatz einer endlichen elektronischen Temperatur und die in unserem Arbeitskreis entwickelte semiempirische Methode für Geometrien, Frequenzen und nicht-kovalente Wechselwirkungen ("GFN-xTB"). Die Energien und Kräfte, die auf diesen Niveaus berechnet worden sind, sind dann in den nachfolgenden Born-Oppenheimer Molekulardynamiksimulationen verwendet worden.
Der zweite Teil setzt sich mit der Anwendung der Dichtefunktionaltheorie bei endlicher Temperatur, spezifisch mit der "Fractional occupation number weighted density" (ρ^{FOD}), auseinander. Die Ergebnisse zeigen, dass ρ^{FOD} als Maß für statische Elektronenkorrelation etwa in Biradikalen und ähnlichen Systemen verwendet werden kann. Zudem können gebrochene Orbitalbesetzungszahlen nützlich bei der ersten Erstellung einer Multikonfigurationswellenfunktion sein. Weiterhin werden einige mit Dichtefunktionaltheorie bei endlicher Temperatur berechnete Potentialhyperflächen entlang von Modellreaktionskoordinaten gezeigt und die Übertragbarkeit des ρ^{FOD}-Modells auf semiempirische Elektronenstrukturmethoden erläutert.
Der dritte Teil zeigt die Hauptergebnisse dieser Arbeit. In Kapitel 4 wird ein Literaturüberblick gegeben und die "Quantum Chemistry Electron Ionization Mass Spectra" (QCEIMS)-Methode präsentiert. Anschließend wird diese im Bezug auf die Verteilung der Ladung auf die Fragmente evaluiert, indem eine Reihe Ethanolhomologer untersucht wird. Eine kleine massenspektrometrische Vergleichsstudie wird ebenfalls präsentiert, welche zeigt, dass Isomere anhand von mit QCEIMS vorhergesagten Massenspektren voneinander unterschieden werden

können, wenn ihre Fragmentationspfade substantiell verschieden sind.
Kapitel 5, 6 und 7 behandeln jeweils die Anwendung von QCEIMS auf relativ große Arnzneimittelmoleküle, die Nukleobase Adenin sowie weitere Nukleobasen. In allen Fällen werden die Fragmentationspfade analysiert, wodurch die Fragmentionenstrukturen aufgeklärt werden können.
Abschließend werden in Kapitel 8 vorhergesagte EI-Massenspektren von Molekülen, die aus verschiedensten Atomen des Periodensystems der Elemente bestehen, gezeigt. Dies wurde durch die Implementierung von GFN-xTB in QCEIMS durch V. Ásgeirsson ermöglicht. Diese robuste und effiziente Methode lässt die Vorhersage von EI-Massenspektren von bemerkenswerter Qualität für organische, organometallische und Hauptgruppenelementsysteme zu, wobei die Rechenkosten um einen Faktor 1.000 (verglichen mit Hybrid-Dichtefunktionalberechnungen) geringer sind.

Contents

Part I.

Introduction

1. Concepts of Electron Ionization Mass Spectrometry

Mass Spectrometry (MS) is one of the work-horses of modern analytical chemistry.[1,2] The combination of Gas Chromatography and Electron Ionization Mass Spectrometry (GC/EI-MS) has been a scientific and commercial success for many decades,[3,4,5] with a wide range of applications, *e.g.*, in forensics[3] and doping control.[6] GC/EI-MS has even been used in astrochemistry to analyze the atmospheres of Venus[7] and Titan.[8] The main strength of GC/EI-MS that has brought about its success and the feature that unites all of the aforementioned fields is the identification of (unknown) compounds. The measured EI mass spectra are often compared to large spectral library databases, using a variety of matching score methods.[9] Therefore, a library of *predicted* EI mass spectra is highly desirable. However, the *ab initio* route to the calculation of EI mass spectra has been one of the most challenging tasks in computational chemistry. This is due to the complexity of the unimolecular chemical reaction space and the difficult estimation of the available internal energy. For this reason there has not been one consistent protocol to routinely compute the EI mass spectrum from first principles until 2013,[10] while there is a continuing effort to compute EI mass spectra to assist in GC/MS compound identification by rule-based, chemoinformatic approaches.[11,12] In contrast to those methods, our methodology[10] is based on Born-Oppenheimer Molecular Dynamics (BO-MD)[13] to elucidate the unimolecular fragmentation reactions in an automatic and unbiased way. It is based on the robust finite-temperature (FT) quantum chemical computation of the energy and the forces acting on molecules (see Chapter 2 for a brief description on FT-DFT, and Chapter 3 for FT-DFT applications).

In this Chapter, I will present the physicochemical concepts of EI mass spectrometry in a nutshell. Excellent further reading is provided in Chapter 7 of McLafferty's book.[1]

In EI-MS, molecules are hit with an electron (e) beam of high energy. In an (e,2e) process, a radical cation $M^{\bullet+}$ is formed from a neutral, closed-shell molecule. Under the EI-MS high vacuum conditions, only unimolecular reactions are possible. I will describe only the Quasi-Equilibrium Theory (QET)[14] case, which assumes that the $M^{\bullet+}$ relaxes to the ground state, thereby re-distributing the ionization and excitation energy into internal energy (E) statisti-

cally.* Depending on the amount of E available to the molecule, the $M^{\bullet+}$ will fragment and thus give rise to a mass spectrum. Consider a molecular ion $ABC^{\bullet+}$. It can fragment for example in the following ways:

$$ABC^{\bullet+} \rightarrow AB^+ + C^\bullet \tag{1.1}$$

$$ABC^{\bullet+} \rightarrow AC^+ + B^\bullet \tag{1.2}$$

$$ABC^{\bullet+} \rightarrow A^+ + BC^\bullet \tag{1.3}$$

$$ABC^{\bullet+} \rightarrow A^+ + B + C^\bullet. \tag{1.4}$$

Within the bounds of QET, the ionization potentials (IPs) of the fragments will determine which fragment is charged (Stevenson's rule)[1]. For $\Delta IP > 0.3$ eV, only the fragment with the lower IP will be charged.[15] As shown in Chapter 4, the charge may also be distributed to more than one fragment in a statistical way if ΔIP is small. Every reaction has an activation barrier E_0 to overcome. Figure 1.1 sketches some possible unimolecular reactions of $ABC^{\bullet+}$ on a model Potential Energy Surface (PES). As depicted, $ABC^{\bullet+}$ may easily rearrange to $ACB^{\bullet+}$, but a steep barrier (E_0^{AC+}) has to be overcome, thus making the back-reaction to $ABC^{\bullet+}$ and dissociation to $AB^+ + C^\bullet$ the only energetically available pathway for decomposition when $E_0^{AB+} < E < E_0^{AC+}$.

A probability function $P(E)^\dagger$ describes the distribution of E to the initial $M^{\bullet+}$ ensemble. Figure 1.2 shows a model $P(E)$ for our hypothetical $ABC^{\bullet+}$. For $E < E_0^{AB+}$, only $M^{\bullet+}$ will be produced. For $E_0^{AB+} < E < E_0^{AC+}$, two scenaria are possible: Either $M^{\bullet+}$ survives kinetically as a metastable ion, or it fragments into $AB^+ + C^\bullet$. The metastable ion case is omitted for simplicity in Figure 1.2. Similarly, when $E > E_0^{AC+}$, more AC^+ will be produced, but it is also possible that AB^+ is produced, depending on the relative rate constants $k(E)$, see below. At very high E, $M^{\bullet+}$ will fragment into $A^+ + B + C^\bullet$, provided A has the lowest IP of all the fragments. For all molecules, $P(E)$ has a low-energy part and a high-energy tail. Thus, a mass spectrum may have high abundances of $M^{\bullet+}$ as well as peaks of small fragment ions. Importantly, Figure 1.2 shows only initial ion abundances, and the fragment ions AB^+ and AC^+ may react further, depending on their own $P(E)$ and the associated rate constants $k(E)$. For the estimation of $P(E)$, the photoelectron spectrum has been used in the literature.[1,16] This is, however, an approximation. The correct physics is measured by electron momentum spectroscopy.[17,18] Unfortunately, only a few such experiments on small molecules are available.[19,20,21]

*Non-ergodic (*i.e.* non-statistical) dissociations are very rare. They mostly appear in mass spectra of small molecules with a low density of electronic states. One of the larger systems where a non-ergodic dissociation takes place is the enol of acetone. Such processes cannot be described by QET.

†$P(E)$ is not a probability in a mathematical sense, but a distribution function that can be normalized to 1. It has, however, only values ≥ 0.

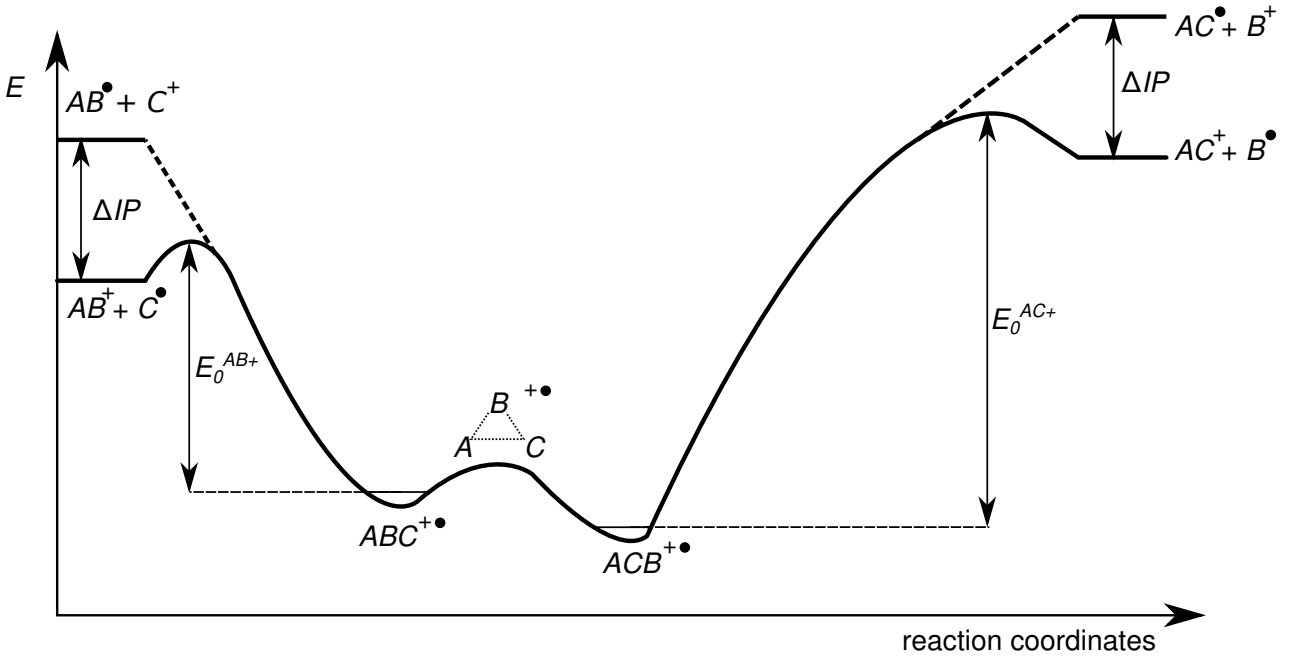

Figure 1.1.: Schematic view of selected unimolecular fragmentations on the $ABC^{\bullet+}$ PES.

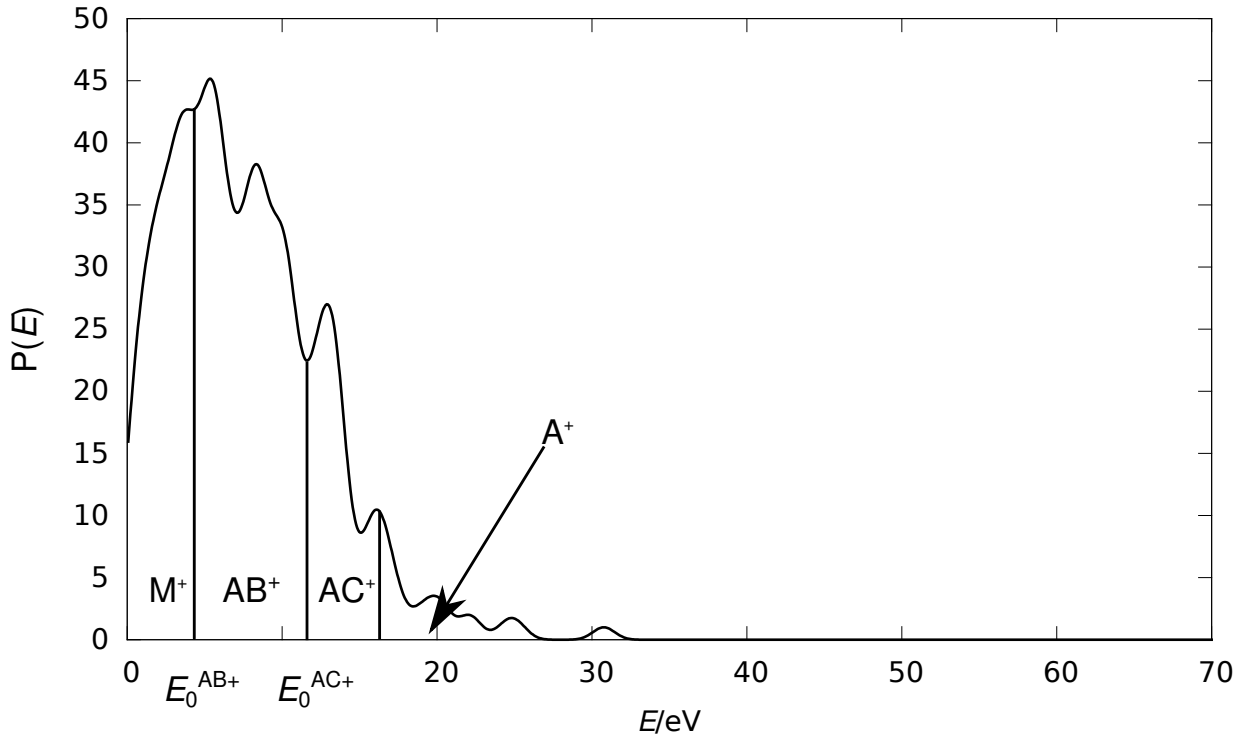

Figure 1.2.: Model $P(E)$ and indication of ion abundances. The 0 of the abscissa corresponds to the IP of the neutral molecule.

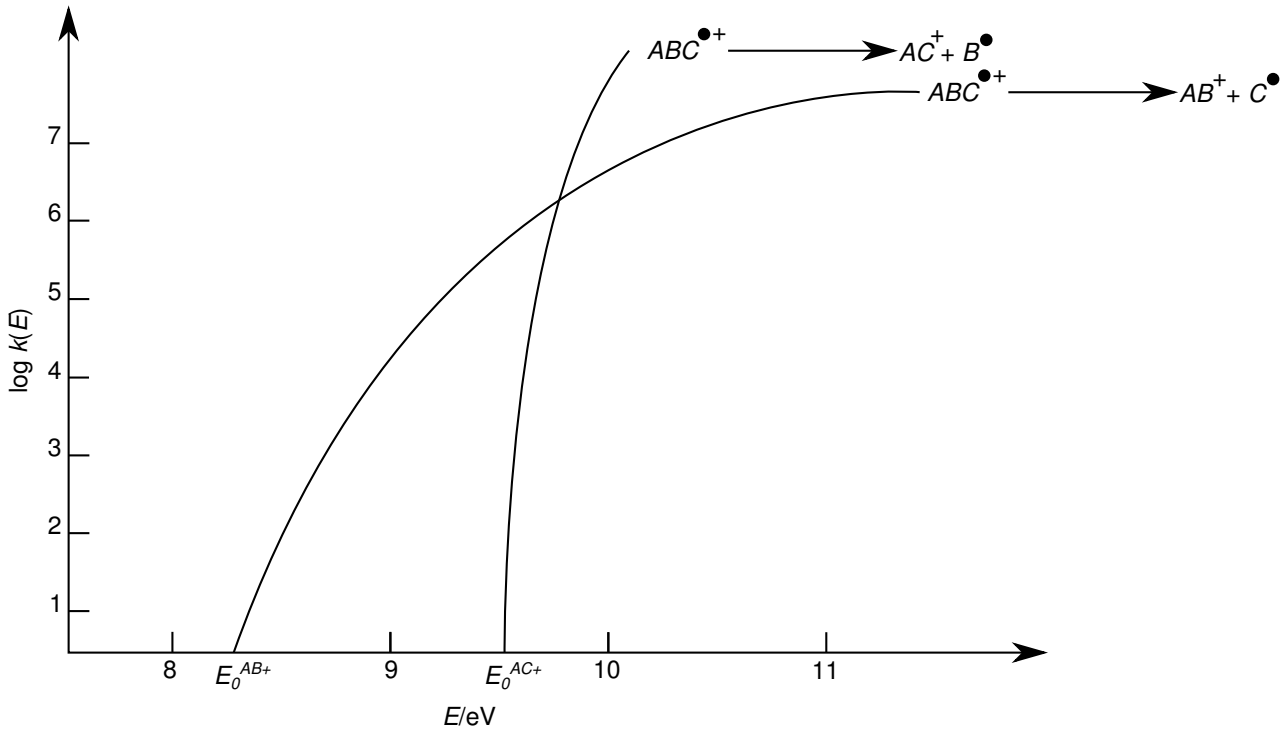

Figure 1.3.: Model $k(E)$ functions in RRKM theory.

The rate constants $k(E)$ are calculated by the following expression, which results from the Rice-Ramsperger-Kassel-Marcus (RRKM) theory.[22,23,24,25]

$$k(E) = \frac{\sigma N^{\ddagger}(E - E_0)}{h\rho(E)}. \tag{1.5}$$

Here, σ is the reaction path degeneracy, $N^{\ddagger}(E - E_0)$ is the transition state sum of states, h is Planck's constant, and $\rho(E)$ is the vibrational density of states. Figure 1.3 shows a typical RRKM diagram for the competing reactions $ABC^{\bullet+} \rightarrow AB^+ + C^{\bullet}$ and $ABC^{\bullet+} \rightarrow AC^+ + B^{\bullet}$. At higher energies, $k^{AC+}(E)$ grows larger than $k^{AB+}(E)$, thereby increasing the abundance of AC^+ in the mass spectrum. Given the tabulated thermochemical data of many small ions (including IP),[26] the photoelectron or electron momentum spectrum of a molecule ($P(E)$), and assuming that the activation energies $E_{0,i}$ can be calculated from transition state theory, it should therefore be possible to predict the EI mass spectrum of a molecule. However, as Eyring and co-workers note in their original publication,[14] the choice of the activated complex (and therefore E_0) is arbitrary. Moreover, the number of possible unimolecular reactions grows extremely fast with the number of atoms. Even more, the photoelectron and electron momentum spectra of molecules are often not available.

In spite of all these challenges, we aim at the routine calculation of the full EI mass spectrum. Figure 1.4 sums up the goal of this thesis exemplified for the limonene molecule. It shows

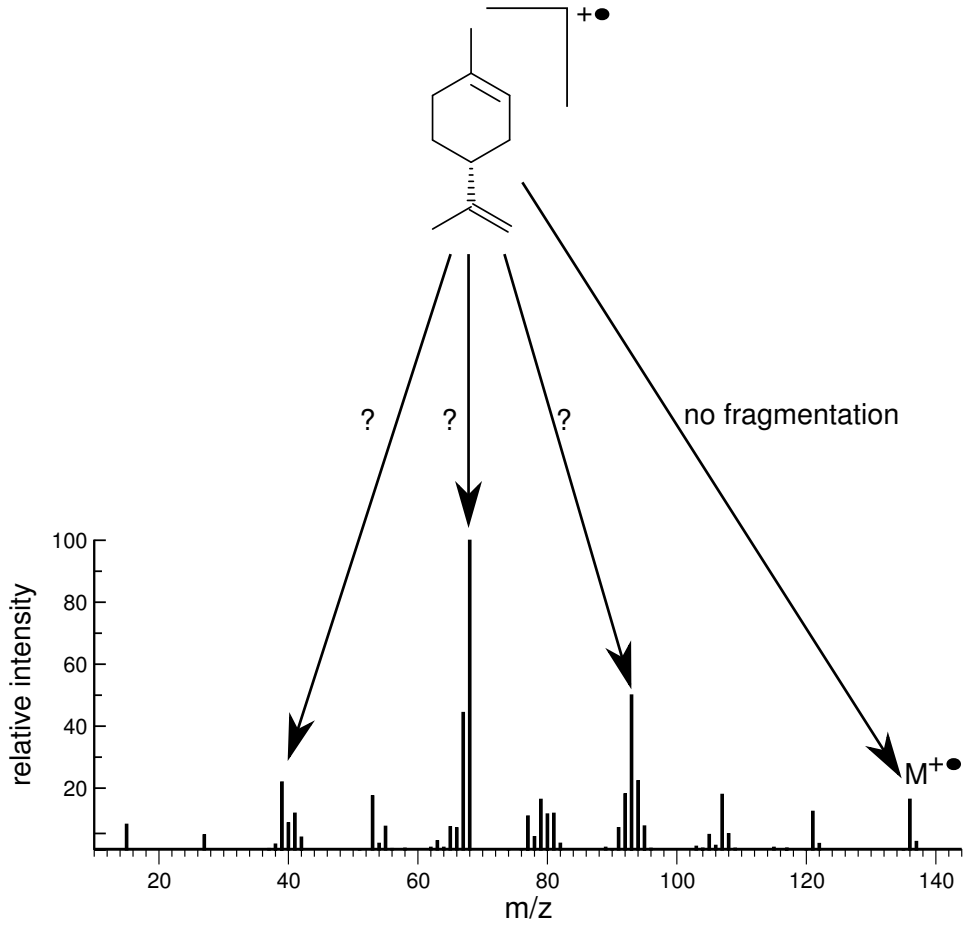

Figure 1.4.: Graphical summary of the goal of this thesis: Computation of the full EI mass spectrum, including the reaction pathways that lead to fragment ion peaks of high intensities. The full experimental EI mass spectrum of the limonene molecule is shown above.

the experimental mass spectrum, and indicates reaction pathways which lead to the fragment ion peaks of the highest intensities. Importantly, not only these processes are to be computed, but the full mass spectrum, including the survival rate of $M^{\bullet+}$. To achieve this goal while circumventing the problems of statistical theory, Grimme has proposed a combination of BO-MD[13] and statistical theory to calculate EI mass spectra, dubbed QCEIMS.[10] The PES on which these simulations are carried out are calculated on the fly by Finite-Temperature Density Functional Theory / Semiempirical Quantum Chemistry (FT-DFT/FT-SQC).[27] The concepts in FT-DFT/FT-SQC and BOMD are described briefly in Chapter 2. The aim and the scope of this thesis has been twofold: (i), to apply QCEIMS to organic molecules ranging in size from 10–100 atoms, and (ii) to extend the capabilities of QCEIMS to organometallic systems by using appropriate FT-quantum chemical methods. By establishing thoroughly analyzed results in both fields, it can be argued that the calculation of EI mass spectra from first priciples will have become close to a routine procedure.

2. Theoretical Methods

In the following, I will give a short recapitulation of the quantum chemical methods which I have used to simulate the unimolecular fragmentation reactions that take place after electron ionization. In Linear Combination of Atomic Orbitals Molecular Orbitals (LCAO-MO) theory, MOs (φ_i) are constructed from AO-like basis functions (χ_μ), which shall be real-space functions:

$$\varphi_i(\vec{r}) = \sum_\mu c_{i\mu}\chi_\mu(\vec{r}), \tag{2.1}$$

where $c_{i\mu}$ are the AO coefficients, and χ_μ shall be Gaussian-Type AOs (GTOs).[28] The orbitals φ_i are then used to construct a many-electron wave function as a Slater determinant $|\Psi\rangle$:

$$|\Psi\rangle = \frac{1}{\sqrt{N!}}\begin{vmatrix} \varphi_1(\vec{r}_1) & \varphi_2(\vec{r}_1) & \dots & \varphi_n(\vec{r}_1) \\ \varphi_1(\vec{r}_2) & \varphi_2(\vec{r}_2) & \dots & \varphi_n(\vec{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(\vec{r}_n) & \varphi_2(\vec{r}_n) & \dots & \varphi_n(\vec{r}_n) \end{vmatrix} \tag{2.2}$$

Here, N is the normalization constant. The electronic energy E_{el} can be calculated according to the electronic Schrödinger equation (the Bra-Ket notation is used for integrals)[28]:

$$\langle\Psi|H_{\mathrm{el}}|\Psi\rangle = E_{\mathrm{el}}, \tag{2.3}$$

where the electronic Hamiltonian H_{el} is the sum of electronic kinetic energy (T_{el}) and potential energy (V_{el}):

$$H_{\mathrm{el}} = T_{\mathrm{el}} + V_{\mathrm{el}}. \tag{2.4}$$

Using the LCAO-MO expansion, the general eigenvalue problem of the Roothan-Hall equations[29,30] can be formulated:

$$\mathbf{FC} = \mathbf{SC}\varepsilon. \tag{2.5}$$

The Fock matrix $\mathbf{F}$ is the representation of H_{el}, the coefficient matrix $\mathbf{C}$ comprises the AO coefficients, the overlap matrix $\mathbf{S}$ is composed of the AO overlap integrals, and ε is a diagonal matrix holding the eigenvalues.

In the following, two approximate ways of building the restricted closed-shell Fock matrix

will be presented:

1. The Hamiltonian arising from Kohn-Sham Density Functional Theory (KS-DFT)

2. The semi-empirical Geometry, Frequency and Noncovalent- eXtended Tight-Binding (GFN-xTB) Hamiltonian

The finite temperature "Fermi smearing" procedure will then be introduced, and finally, I will briefly describe Born-Oppenheimer Molecular Dynamics (BO-MD).

2.1. Kohn-Sham DFT

The following is a brief summary of KS-DFT[31,32] and the D3 dispersion correction[33,34,35] to account for long-range electron correlation effects. An excellent introduction to DFT and its historical predecessors is given by Parr and Yang in their classic 1989 book.[36] In KS-DFT, the Fock matrix elements $F_{\mu\nu}^{\mathrm{KS}}$ in atomic units are given as follows:

$$F_{\mu\nu}^{\mathrm{KS}} = \underbrace{T_{\mu\nu}^{KS} + V_{\mu\nu}^{\mathrm{ext}}}_{H_{\mu\nu}^{\mathrm{core}}} + J_{\mu\nu} + a_X K_{\mu\nu} + F_{\mu\nu}^{XC}, \tag{2.6}$$

where a_X denotes the fraction of Fock exchange that is to be computed in hybrid functionals. The one-electron $H_{\mu\nu}^{\mathrm{core}}$ matrix elements are computed as in Hartree-Fock (HF) theory.[28] The first term is the kinetic energy integral

$$T_{\mu\nu}^{KS} = \left\langle \chi_\mu \left| -\frac{1}{2}\nabla^2 \right| \chi_\nu \right\rangle. \tag{2.7}$$

This expression of the kinetic energy is the characterizes KS-DFT. The electron-nuclear attraction integrals $V_{\mu\nu}^{\mathrm{ext}}$ are given as:

$$V_{\mu\nu}^{\mathrm{ext}} = \left\langle \chi_\mu \left| -\sum_A \frac{Z_A}{r_A} \right| \chi_\nu \right\rangle, \tag{2.8}$$

where Z_A is the charge of nucleus A and r_A the distance between the nucleus A and the aufpunkt of the respective AO. In HF theory, the two-electron part of $\mathbf{F}$, $G_{\mu\nu}$, containing the Coulomb part $J_{\mu\nu}$ and the exchange part $K_{\mu\nu}$, is defined as:[28]

$$G_{\mu\nu} = \sum_{\lambda\sigma} P_{\lambda\sigma}[\underbrace{(\mu\nu|\lambda\sigma)}_{\rightarrow J_{\mu\nu}} - \frac{1}{2}\underbrace{(\mu\lambda|\nu\sigma)}_{\rightarrow K_{\mu\nu}}], \tag{2.9}$$

where the two electron AO integrals are given in Mulliken notation. The density matrix $\mathbf{P}$ is constructed from $\mathbf{C}$, where n_a is the occupation number of the a^{th} orbital:

$$P_{\lambda\sigma} = \sum_a n_a \sum_{\lambda\sigma} C_{\lambda a} C_{\sigma a}. \tag{2.10}$$

In pure KS-DFT ($a_X = 0$), the Fock exchange part $K_{\mu\nu}$ of $G_{\mu\nu}$ is replaced by the one-electron exchange-correlation matrix elements $F_{\mu\nu}^{\text{XC}}$:

$$F_{\mu\nu}^{\text{XC}} = \left\langle \chi_\mu | v^{XC} | \chi_\nu \right\rangle. \tag{2.11}$$

These integrals, requiring some approximate exchange-correlation expression v^{XC}, have to be evaluated numerically. The exchange-correlation energy E^{XC} is then calculated separately by another numerical integration. Once $\mathbf{F}$ is computed, the self-consistent field (SCF) iterations are performed as in standard HF implementations, *i.e.*, the optimized AO coefficients are variationally obtained to yield the lowest E_{el}:

$$E_{\text{el}} = \frac{1}{2} \sum_{\mu\nu} P_{\mu\nu} (H_{\mu\nu}^{\text{core}} + J_{\mu\nu} + a_X K_{\mu\nu}) + E^{\text{XC}}, \tag{2.12}$$

where the E^{XC} is obtained from the KS-density $\rho(\vec{r})$ as follows:

$$E^{\text{XC}} = \int \frac{\delta E^{\text{XC}}[\rho(\vec{r})]}{\delta \rho(\vec{r})} \rho(\vec{r}) d\vec{r} \tag{2.13}$$

The total energy of the molecular system is then obtained by adding the nuclear repulsion energy:

$$E_{\text{tot}} = E_{\text{el}} + \frac{1}{2} \sum_{A \neq B} \frac{Z_A Z_B}{R_{AB}}, \tag{2.14}$$

with the distances R_{AB} between the nuclei A and B.

The D3 Dispersion Correction The D3 dispersion correction is a post-SCF energy correction.[33] It models the stabilizing long-range electron correlation effects, which are not usually included in the approximate E^{XC} expressions of KS-DFT. The pair-wise term of 6^{th} and 8^{th} orders for the nuclei A, B at internuclear distances R_{AB} using a Becke-Johnson (BJ)[34] damping scheme reads as follows:

$$E_{\text{disp}}^{\text{D3(BJ)}} = -\frac{1}{2} \sum_{A \neq B} \left(s_6 \frac{C_6^{AB}}{R_{AB}^6 + (a_1 R_{AB}^0 + a_2)^6} + s_8 \frac{C_8^{AB}}{R_{AB}^8 + (a_1 R_{AB}^0 + a_2)^8} \right). \tag{2.15}$$

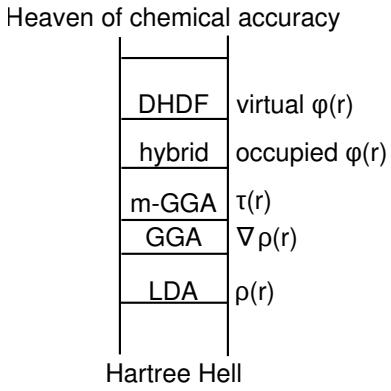

Figure 2.1.: Jacob's Ladder in KS-DFT according to Perdew. The hybrid functional rung is also called hyper-GGA by Perdew.[37]

The dispersion coefficients C_6^{AB} and C_8^{AB} are pre-computed on hydride model systems with different coordination numbers for each element, thus taking into account the chemical environment and the hybridization state. R_{AB}^0 is calculated from the dispersion coefficients,[35] and the global scaling parameter $s_6 = 1$ by definition, which leaves only three adjustable parameters, s_8, a_1, a_2. These are available for around 100 functionals in KS-DFT as well as for HF and semi-empirical tight binding methods. The DFT-D3(BJ) total energy expression reads:

$$E_{\text{tot}}^{\text{DFT-D3(BJ)}} = E_{\text{el}} + \frac{1}{2}\sum_{A\neq B}\frac{Z_A Z_B}{R_{AB}} + E_{\text{disp}}^{\text{D3(BJ)}}. \tag{2.16}$$

Jacob's Ladder There are many flavors of KS-DFT. Perdew has classified them by the information that is used for the computation of the exchange-correlation energy E^{XC} in a Jacob's Ladder scheme,[37] which is sketched in Figure 2.1 In Local Density Approximation (LDA)-KS-DFT, E^{XC} is only dependent on the local density $\rho(\vec{r})$. The next rung, the Generalized gradient approximation (GGA), takes into account the first derivative of the density $\vec{\nabla}\rho(\vec{r})$. Functionals that take into account higher-order derivatives of the density (denoted τ) are called meta-GGAs. In hybrid DFT, the Fock exchange is computed from the KS-DFT occupied orbitals φ_i, and $0 < a_X < 1$, cf. Equation 2.6. Finally, in double-hybrid density functionals (DHDF), the virtual orbitals are also taken into account, *e.g.*, by an MP2-like expression.[38]

The rungs of the ladder symbolize the accuracy that can be expected from a molecular calculation. One ascends from Hartree Hell, where there is no E^{XC} to the Heaven of chemical accuracy, which is typically defined as 1 kcal mol^{-1}. Note that the step size between GGA and m-GGA is smaller than the other step sizes. This hierarchy of functionals is generally reflected by large-scale benchmark studies such as the GMTKN30 database.[39] Modern developments have led to m-GGA functionals that can compete with hybrid DFT.[40]

2.2. The GFN-xTB Hamiltonian

The semi-empirical GFN-xTB Tight Binding Hamiltonian[41] speeds up the calculation of the approximate E_{el} by 2–3 orders of magnitude compared to GGA KS-DFT computations. It makes use of a slightly modified minimal GTO valence-only LCAO basis set (STO-mG),[42] where additional s functions are placed on H atoms and d functions on heavier main group elements to model hypervalent and hydrogen bonding.[41] The total energy expression is given by:

$$E_{\mathrm{tot}}^{\mathrm{GFN\text{-}xTB}} = E_{\mathrm{el}} + E_{\mathrm{rep}} + E_{\mathrm{disp}}^{\mathrm{D3(BJ)}} + E_{\mathrm{XB}}. \tag{2.17}$$

The GFN-xTB Fock matrix elements read:

$$F_{\mu\nu}^{\mathrm{GFN\text{-}xTB}} = H_{\mu\nu}^{0} + \frac{1}{2} S_{\mu\nu} \sum_{C} \sum_{l''} (\gamma_{AC,ll''} + \gamma_{BC,l'l''}) p_{l''}^{C} + \frac{1}{2} S_{\mu\nu} (q_A^2 \Gamma_A + q_B^2 \Gamma_B) \tag{2.18}$$

$$(\mu \in l(A), \nu \in l^{'}(B)).$$

The $H_{\mu\nu}^{0}$ elements are determined by the Hückel constants, the effective atomic energy levels, the electronegativities of nuclei A, B and two global scaling constants.[41] The second order contributions involve the monopole electrostatics approximate expressions γ and the shell charges $p_{l''}^{C}$, where l'' runs over the angular momenta and C runs over the nuclei. The third order contributions depend on the Mulliken partial charges q_I and the charge derivative of the Hubbard parameter, Γ.* The latter two terms are scaled by the atomic overlap $S_{\mu\nu}$. GFN-xTB is a Self-Consistent Charge (SCC) Hamiltonian.[43] The electronic energy is finally evaluated by:

$$E_{\mathrm{el}} = \sum_{i} n_i \langle \varphi_i | H_0 | \varphi_i \rangle + \frac{1}{2} \sum_{A,B} \sum_{l(A)} \sum_{l'(B)} p_l^A p_{l'}^B \gamma_{AB,ll''} + \frac{1}{3} \sum_{A} \Gamma_A q_A^3 - T_{\mathrm{el}} S_{\mathrm{el}}. \tag{2.19}$$

The last term in equation 2.19 is zero for an electronic temperature of 0 K. The finite electronic temperature case is discussed in section 2.3. $E_{\mathrm{disp}}^{\mathrm{D3(BJ)}}$ has already been described above. The repulsion energy in Equation 2.17 is given by:

$$E_{\mathrm{rep}} = \sum_{AB} \frac{Z_A^{\mathrm{eff}} Z_B^{\mathrm{eff}}}{R_{AB}} \mathrm{e}^{-(\alpha_A \alpha_B)^{1/2} (R_{AB})^{k_f}}. \tag{2.20}$$

It is an atom-pairwise expression with the fitted element-specific parameters Z^{eff} and α, and the global parameter k_f. The halogen bond correction E_{XB} is of a modified Lennard-Jones

*The Hubbard parameter models the chemical hardness of an atom.

form:

$$E_{\mathrm{XB}} = \sum_{\mathrm{XB}} f_{\mathrm{damp}}^{\mathrm{AXB}} k_X \left(1 + \left(\frac{R_{\mathrm{cov,AX}}}{R_{\mathrm{AX}}}\right)^{12} - k_{X2}\left(\frac{R_{\mathrm{cov,AX}}}{R_{\mathrm{AX}}}\right)^{6}\right) / \left(\frac{R_{\mathrm{cov,AX}}}{R_{\mathrm{AX}}}\right)^{12}, \tag{2.21}$$

where $f_{\mathrm{damp}}^{\mathrm{AXB}}$ is designed such that the correction vanishes for non-linear AXB arrangements. $R_{\mathrm{cov,AX}}$ are effective covalent distances and k_{X2} is another global parameter. A full implementation of the GFN-xTB Hamiltonian including many functionalities such as orbital localization is available in the `xtb` program.* Parts of the code for the computation of $E_{\mathrm{tot}}^{\mathrm{GFN\text{-}xTB}}$ and the corresponding forces have been adapted and transferred by V. Ásgeirsson for the implementation in the `qceims` program, see Chapter 8.

2.3. Finite Temperature DFT and TB

The concept of finite temperature DFT (FT-DFT), which introduces an electronic temperature T_{el} was proposed by Mermin in 1964.[27] In its LCAO-MO/KS-DFT/GFN-xTB implementation, the SCF/SCC procedures are modified such that the electronic free energy is minimized:

$$G_{\mathrm{el}} = E_{\mathrm{el}} - T_{\mathrm{el}} S_{\mathrm{el}}, \tag{2.22}$$

where S_{el} denotes the electronic entropy. In order to achieve this minimization, fractional orbital occupations are introduced by distributing the orbital occupation numbers according to a Fermi-Dirac distribution ("Fermi smearing"):

$$n_i = \frac{1}{\mathrm{e}^{(\varepsilon_i - E_F)/\mathrm{k}T_{\mathrm{el}}} + 1}, \tag{2.23}$$

where k is the Boltzmann constant, and E_F is the Fermi level. The electronic temperature T_{el} is a parameter, which is dependent on the fraction of Fock exchange a_X:[10]

$$T_{\mathrm{el}} = 20000\,\mathrm{K} \times a_X + 5000\,\mathrm{K}. \tag{2.24}$$

FT-DFT procedures have been used to converge the SCF procedure in difficult cases.[44,45] Applications are presented in Chapter 3 of this thesis. Additionally, all simulations in Part III of this thesis have been performed either by FT-DFT or FT-semi-empirical LCAO computations, including the FT treatment of the GFN-xTB Hamiltonian (see equation 2.19) for which $T_{\mathrm{el}} = 5000$ K.

*The `xtb` program is available on request from xtb@thch.uni-bonn.de.

2.4. Born-Oppenheimer Molecular Dynamics

The Born-Oppenheimer approximation assumes that the coupling between nuclear and electronic motion is negligible. Given the KS-DFT/GFN-xTB expressions for E_{tot}, the negative gradient with respect to the nuclear positions, *i.e.*, the forces may be computed. In BO-MD, it is equal to mass times acceleration:[13]

$$M_I \frac{d^2 \vec{R}_I}{dt^2} = -\underbrace{\vec{\nabla}_I E_{\text{tot}}}_{\vec{F}_I}. \tag{2.25}$$

M_I is the mass of nucleus I, $\vec{R}_I$ its position, and $\vec{F}_I$ the force acting on it. The right hand side of equation 2.25 is known from¡ an electronic structure calculation. The nuclear positions at a new point in time $t + \Delta t$ have to be computed numerically, using a finite time step Δt, which is governed by the fastest nuclear motions.* In practice, this can be performed by the leap-frog algorithm. The nuclear position and velocity ($\vec{v}_I$) updates are computed as follows:

$$\vec{R}_I(t) = \vec{R}_I(t - \Delta t) + \vec{v}_I(t - \frac{\Delta t}{2})\Delta t \tag{2.26}$$

$$\vec{v}_I(t + \frac{\Delta t}{2}) = \vec{v}_I(t - \frac{\Delta t}{2}) + \frac{\vec{F}_I}{M_I}\Delta t. \tag{2.27}$$

By the interleaving evaluations of positions and velocities, the algorithm gains numerical stability concerning the conservation of energy. The internal energy E occurring in mass spectrometry simulations is pumped into the system by scaling $\vec{v}_I$ during the BO-MD procedure.

2.5. Overview of the Methodology

The graphical overview in Figure 2.2 shows the general simulation procedure that I have used. First, the choice of an appropriate (FT) electronic structure method has to be made. The electronic and total energies are then calculated along with the forces. The application of FT-DFT and FT-GFN-xTB methods in static calculations is presented in Chapter 3. To simulate processes that occur in EI mass spectrometry experiments, the total energy in the BO-MD simulations, which use the computed energy and the forces, is scaled up to the internal energy E, which is then conserved, leading to very high kinetic energies ("hot" molecular ions). The model trajectory in Figure 2.2 shows some fragmentation. The last snapshot depicts a neutral loss, indicated also by the fact the LCAO-HOMO coefficients are small on the neutral fragment. The quantitative way of determining the distribution of the charge to the fragments is discussed in Chapter 4.

*In the simulations described below, the time step is always 0.5 fs.

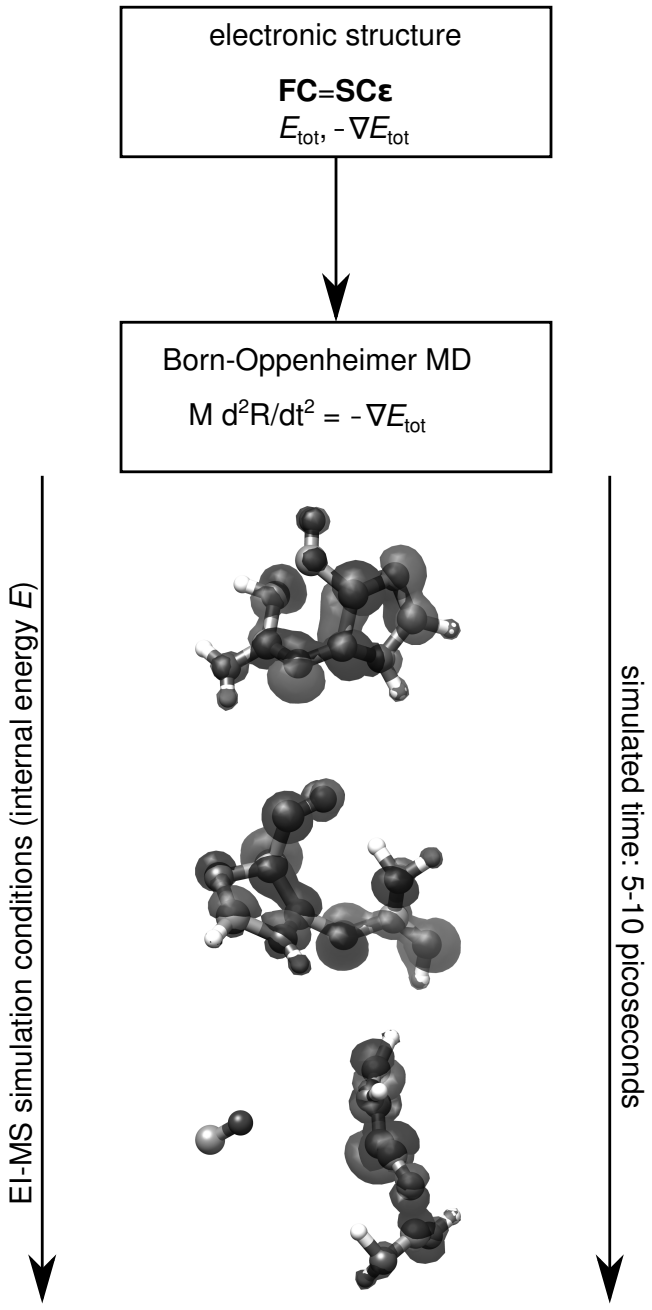

Figure 2.2.: Overview of the simulation methodology. The GFN-xTB HOMO is depicted on snapshots of a model BO-MD trajectory at some internal energy E, cf. Chapter 1.

Part II.

Application of Finite Temperature DFT

This part of the thesis presents the application of FT-DFT and FT-GFN-xTB. The FT treatment has been shown to describe static electron correlation.[46] This is shown qualitatively for the H_2 dissociation curve at the FT-PBE/def2-SVP level of theory at $T_{el} = 5000$ K in Figure 2.3. The full configuration interaction solution, which takes into account all configuration state functions, is exact and goes asymptotically to -1 E_h (the exact absolute energy of two separated H atoms). The HF solution is known to have the wrong dissociation limit because it incorporates 50 % ionic terms in the wave function. The PBE functional also overestimates the dissociation energy, but it is clear that some static electron correlation energy is recovered by the exchange-correlation functional. The FT-PBE curve incorporates even more static correlation, and the dissociation energy is lowered accordingly, which is effectively a correction in the right direction for the simplest static electron correlation test case.

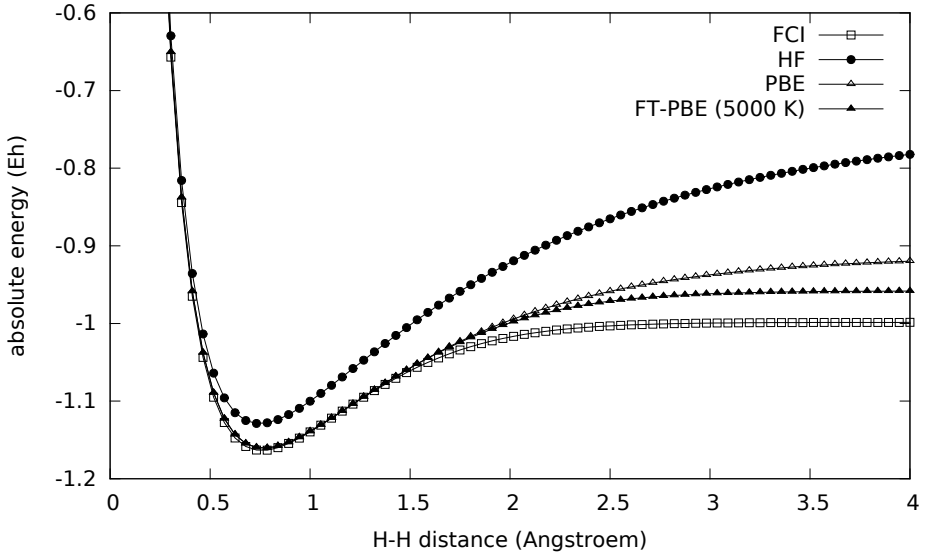

Figure 2.3.: H_2 dissociation curve for HF, PBE,[47] and FT-PBE in comparison with the exact configuration interaction (FCI) solution using a def2-SVP[48] basis set.

This example shows that the fractional occupation numbers introduced in the FT-DFT treatment may qualitatively ameliorate the results in the presence of static electron correlation. Note that the energy correction of FT-PBE in the equilibrium distance of H_2 is negligible because of a large orbital energy gap, and that the PBE energy is close to the FCI energy. The occupation numbers are only smeared over the orbitals when the HOMO-LUMO gap becomes sufficiently small. In Chapter 3, four more complex, chemical static electron correlation cases are explored using a variety of density functionals as well as FT-GFN-xTB.

3. The Fractional Occupation Number Weighted Density as a Versatile Analysis Tool for Molecules with a Complicated Electronic Structure

Christoph Alexander Bauer, Andreas Hansen and Stefan Grimme*
Received 5th of October 2016, Published online 16th of January 2017

The article is dedicated to Prof. Gerhard Erker on the occasion of his 70th birthday.

Reproduced (adapted) from
Christoph Alexander Bauer, Andreas Hansen and Stefan Grimme Chem.– Eur. J. **2017**, *23*, 6150–6164

DOI 10.1002/chem.201604682

Own manuscript contribution

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

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

Abstract The fractional occupation number weighted density (FOD) analysis is explored as a general theoretical diagnostic for complicated electronic structures. Its main feature is to provide robustly and fast the information where 'hot' (strongly correlated and chemically active) electrons are localized in a molecule. We demonstrate its usage in four different prototypical applications: (i) As a new and fast measure for the biradical character of polycyclic aromatic hydrocarbons, (ii) for the selection of active orbital spaces in multiconfigurational or complete active space self consistent field (MCSCF/CASSCF) treatments, (iii) as a possibility to consistently describe molecular energy landscapes in regions with varying biradical character as exemplified by partial double bond torsions, and (iv) as a powerful visualization method for static electron correlation effects in large biomolecules in connection with an efficient semi-empirical tight-binding molecular orbital scheme. The later application opens a full quantum mechanical, unbiased route to the automatic detection of errors in experimental protein X-ray structures such as false protonation states or misplaced atoms. In a first example, the complete (unfragmented) quantum chemical calculation of the FOD for an entire metallo-protein with more than 7,500 atoms is described.

3.1. Introduction

A qualitative analysis and classification of the electronic structure of molecules is a fundamental and longstanding question in theoretical chemistry. More specifically, one may ask how 'difficult' a particular electronic structure is, which is important for the selection of an appropriate quantum chemical method and a reasonable computational description. Moreover, this question is related to the very basic properties of chemical reactivity and thermodynamic stability. It is commonly assumed that simple electronic structures as, e.g., saturated alkane chains for example are unreactive, of low electronic energy and have usual (classical) geometric structures. The 'difficulties' which are the topic of this work may arise inherently from the chemical composition and bonding pattern (e.g. metallic, biradical or excited compounds) or artificially. The latter case commonly occurs in badly resolved protein structures which are often taken from experiment as starting point for further theoretical studies. Here, due to low resolution or other experimental problems, atoms can be missing or even their nuclear charge can be incorrect. Such failures which prevent any reasonable theoretical treatment are difficult to detect visually and require laborious manual analysis, particularly for large molecules. A difficult electronic structure mostly arises from so-called static (near-degeneracy) electron correlation (SEC) effects while the remaining dynamic electron correlation can be handeled quite efficiently by various 'low-cost' quantum chemical methods like single-reference (SR) perturbation theory[49] or dispersion-corrected density functional theory (DFT-D)[50]. Electron correlation as a many-body effect has cooperative as well as non- and anti-cooperative contri-

butions. In a standard perturbative scheme, pair-wise electron correlations are strictly additive in second-order while (mostly destabilizing) pair-correlations appear in third-order perturbation theory[49]. Explicit triple electron collisions (excitations) are covered for example in the CCSD(T) quantum chemistry 'gold standard' and can be cooperative or anti-cooperative.[51] Unfortunately, SEC effects require sophisticated and computationally involved multiconfigurational treatments. In practice, the distinction between a SEC dominated and a simple SR system is often unclear and mixed cases occur commonly. Recently, Grimme and Hansen introduced the Fractional Occupation number weighted Density (FOD) as a real-space measure for SEC.[46] The special density ρ^{FOD} is obtained by performing a computationally cheap Finite-Temperature DFT (FT-DFT)[27] computation. FT-DFT procedures have been used to accelerate and enable self-consistent field convergence[44], and to perform *ab initio* molecular dynamics simulations.[10,52,53] In this technique the electrons are self-consistently smeared over the molecular orbitals according to a Fermi-Dirac distribution. In the FOD method,[46] the resulting fractional occupation numbers f_i ($0 \leq f_i \leq 1$) serve as the input for the calculation of ρ^{FOD}

$$\rho^{\mathrm{FOD}}(\vec{r}) = \sum_{i}^{N} (\delta_1 - \delta_2 f_i) |\varphi_i^2(\vec{r})|, \tag{3.1}$$

where the δ functions are chosen such that only the fractionally occupied orbitals φ contribute.[46] The key variable is the electronic temperature, T_{el} and its dependence on the included fraction of Fock exchange, a_X, has been derived in previous studies.[10,46] For more details, see the original publication[46]. An FT-DFT treatment corresponds to an ensemble of many (not explicitly specified) determinants describing a singlet state if it is employed in the spin-restricted formalism, while it loses any spin-multiplicity information in the unrestricted case.

Herein we aim at exploring the usefulness of the FT-DFT procedure and the ρ^{FOD} measure to identify, analyze, and describe the presence of SEC effects for some general chemistry problems. We have applied the FOD analysis to the following areas of contemporary research:

1. Organic biradical systems. Biradicals are perhaps the most prominent example of SEC effects in molecular chemistry,[54] where the mean-field SCF procedure without any further modifications very often provides unphysical results. In the first part of the next section, we show the application of the FOD analysis on such systems, and discuss the correlation of the N^{FOD} measure, i.e., the number of 'hot' electrons (i.e., spatially integrated ρ^{FOD}) in the system, to the experimentally derived biradical character y.

2. The use of FT-DFT results as the input for active space selection in multiconfigurational/complete active space (MCSCF/CASSCF) computational procedures. Once the presence of SEC effects has been identified (e.g., by an FOD analysis), one may wish

to compute a MCSCF/CASSCF wave function, which describes these effects in an *ab initio* way. Numerous procedures exist to either simplify such calculations[55], or to rigorously determine the active space.[56,57,58] Here we demonstrate the performance of the FT-DFT/FOD procedure for an efficient and robust active space selection.

3. The exploration of FT-DFT potential energy surfaces (PES) for double bond twists. The PES for those torsions often feature biradical character even in the electronic ground state wave function. In the third part of the results section, we use the FOD analysis to probe the SEC effects for double bond rotations in C_2H_4, tetracyanoethylene, retinal, and examine the transition state of a photoactivatable molecular switch molecule. It is shown that the FT-DFT procedure provides not only reasonable fractional occupations but also improves the energetic description considerably.

4. Biomolecular simulations. The FOD analysis, based on semi-empirical tight-binding based molecular orbital computations, is applied to two metalloproteins comprising several thousands of atoms and tested as a sanity check for faulty or possibly wrongly prepared input geometries derived from experimental crystal structures.[59] It is shown that the FOD method is robust and can be applied routinely and consistently also for very large molecular structures.

3.2. Computational details

If not otherwise specified, structures were optimized at the TPSS[60]-D3[33,34,35]/def2-TZVP[48] level of theory. Finite temperature-DFT (FT-DFT) calculations were conducted at the TPSS/def2-TZVP level with a default electronic temperature T_{el} of 5000 K. T_{el} was 15000 K for FT-BH-LYP[61]/def2-TZVP calculations, see Ref 4 for a discussion of the chosen temperature in relation to the amount of Fock exchange included in the density functional. The default isocontour value of the ρ^{FOD} plots is 0.005 e Bohr^{-3}. All FT-DFT calculations and DFT structure optimizations were performed with TURBOMOLE.7.0.2[62,63]. For the retinal system, we used the domain based local pair natural orbital coupled cluster singles and doubles with perturbative triple excitations (DLPNO-CCSD(T)) method employing the default threshold values and the linear scaling implementation based on sparse maps.[64,65] All wave function theory calculations (including CASSCF[66] and NEVPT2[67,68,69] calculations) were performed with the ORCA suite of programs, development version of ORCA 4.0.[70,71] We use a modified version of the extended tight binding (xTB)[72] Hamiltonian (termed GFN-xTB) as the basis of the FOD analysis for the biomolecular systems.[41]. We employed chimera 1.1[73] for the visualization of the FOD plots, using the cube file format as calculated either by Multiwfn 3.3.7[74] or by our in-house xTB program. Visualizations of proteins were produced with VMD 1.9.[75]

3.3. Results and Discussion

3.3.1. FOD as a Tool to Gauge Biradical Character

We examine the use of the FOD analysis for a quantitative description of the open shell singlet biradical character y of ground state organic molecules. These systems (specifically polycyclic aromatic hydrocarbons, PAHs) have recently sparked great research interest.[54,76,77] They are useful in the design and production of functional materials, e.g., for organic field effect transistors (OFETs)[78,79], in energy storage[80,81], and in non-linear optics research.[82] An experimental route to gauge the biradical character has been obtained from Two-Photon Absorption (TPA) spectroscopy[82,83]

$$y = 1 - \sqrt{1 - \left(\frac{E_{S_{1u},S_{1g}} - E_{T_{1u},S_{1g}}}{E_{S_{2g},S_{1g}}} \right)^2}. \tag{3.2}$$

The relevant excitation energies $E_{S_{1u},S_{1g}}$ and $E_{S_{2g},S_{1g}}$ can be obtained from one- and two-photon absorption measurements, respectively, and $E_{T_{1u},S_{1g}}$ is accessible *via* phosphorescence and ESR measurements.[83] A common computational approach to y is based on UHF natural orbital occupation numbers (UNOs)[84]:

$$y = 1 - \frac{4(n_{\mathrm{HOMO}} - n_{\mathrm{LUMO}})}{4 + (n_{\mathrm{HOMO}} - n_{\mathrm{LUMO}})^2}. \tag{3.3}$$

From Eq. 3.3, one can easily see that $y = 1$ when $n_{\mathrm{HOMO}} = n_{\mathrm{LUMO}}$, which occurs in perfect biradicals. Since the calculated value of y depends critically on the molecular structure[85], all molecular geometries should be optimized at the same level of theory (here TPSS-D3/def2-SV(P), except where noted otherwise). This ensures that trends are reproduced reliably, and that one can deduce quantitative correlations. As a meta-GGA type density functional, TPSS includes already some SEC effects, and therefore a UKS broken-symmetry open shell singlet solution often does not exist. Instead, we will demonstrate the use of FT-TPSS ($T_{\mathrm{el}} = 5000$ K) on these structures as an alternative to quantify the biradical character.

We start our exemplary discussion with two well-known organic biradicaloid structures, namely Thiele's hydrocarbon (HC) and Tschitschibabin's HC.[86,87,88] We have calculated y at the UHF/6-31G** level of theory and found for Thiele's HC $y = 0.41$, and for Tschitschibabin's HC $y = 0.69$ on their TPSS-D3/SV(P)-optimized geometries, indicating the larger biradical character of the latter molecule. Fig. 3.1 a) shows the closed-shell and biradical Lewis resonance structures of these systems. The FOD plots in Figure 3.1 provide b) are in line with this concept because the ρ^{FOD} appears at the same carbon atoms where the radicals are located in the open-shell resonance structures. These examples may serve as the proof of principle that FT-DFT computations can identify the 'hot' electrons of these textbook biradical molecules –

or at least that it is conceptually consistent with chemical intuition. Having already studied model PAH systems in the initial FOD analysis publication,[46] we discuss PAHs with known biradical character from the recent literature in the next paragraph.

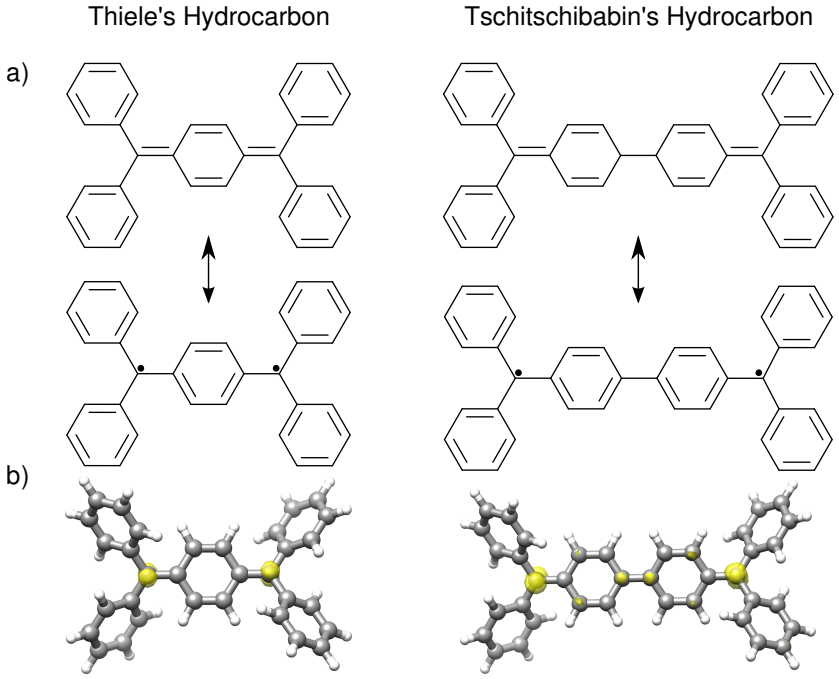

Figure 3.1.: a) Lewis resonance structures of Thiele's HC and Tschitschibabin's HC. b) FOD plots (FT-TPSS/def2-SV(P), T_{el} = 5000 K) at an isocontour value of $\sigma = 0.005$ e Bohr^{-3} .

The Relation of the FOD Analysis to Clar's Sextet Rule

Clar's sextet rule states that the resonance structure with the highest number of benzene-like moieties within a PAH will dominate the electronic ground state.[89,90] The compounds DBHZ1 and DBHZ2[91] (see Fig. 3.2) follow Clar's sextet rule. For DBHZ1, one can write two open-shell resonance structures with three benzenoid moieties each, and for DBHZ2, there are two open-shell resonance structures, one with three and one with four benzenoid rings, respectively.[91] Therefore, DBHZ2 exhibits the larger biradical character. Figure 3.2 shows the structures and the FOD plots of the two systems. We have calculated the N^{FOD} and y values (at the UHF/6-31G** level of theory as recommended[83]) based on the TPSS-D3/def2-SV(P) optimized structures of DBHZ1 and DBHZ2. For DBHZ1, $y = 0.657$ and $N^{\mathrm{FOD}} = 1.678$. For DBHZ2, we find $y = 0.780$ and $N^{\mathrm{FOD}} = 1.937$. Hence, the FOD analysis of DBHZ1 and DBHZ2 is in accordance with Clar's sextet rule and the experimental findings as it clearly reproduces the trend.

Taken from the recent literature, tetrabenzo[a,f,j,o]perylene (**TBP1**)[92] is an example of a PAH with a biradical ground state. Figure 3.3 shows that ρ^{FOD} is again in accordance

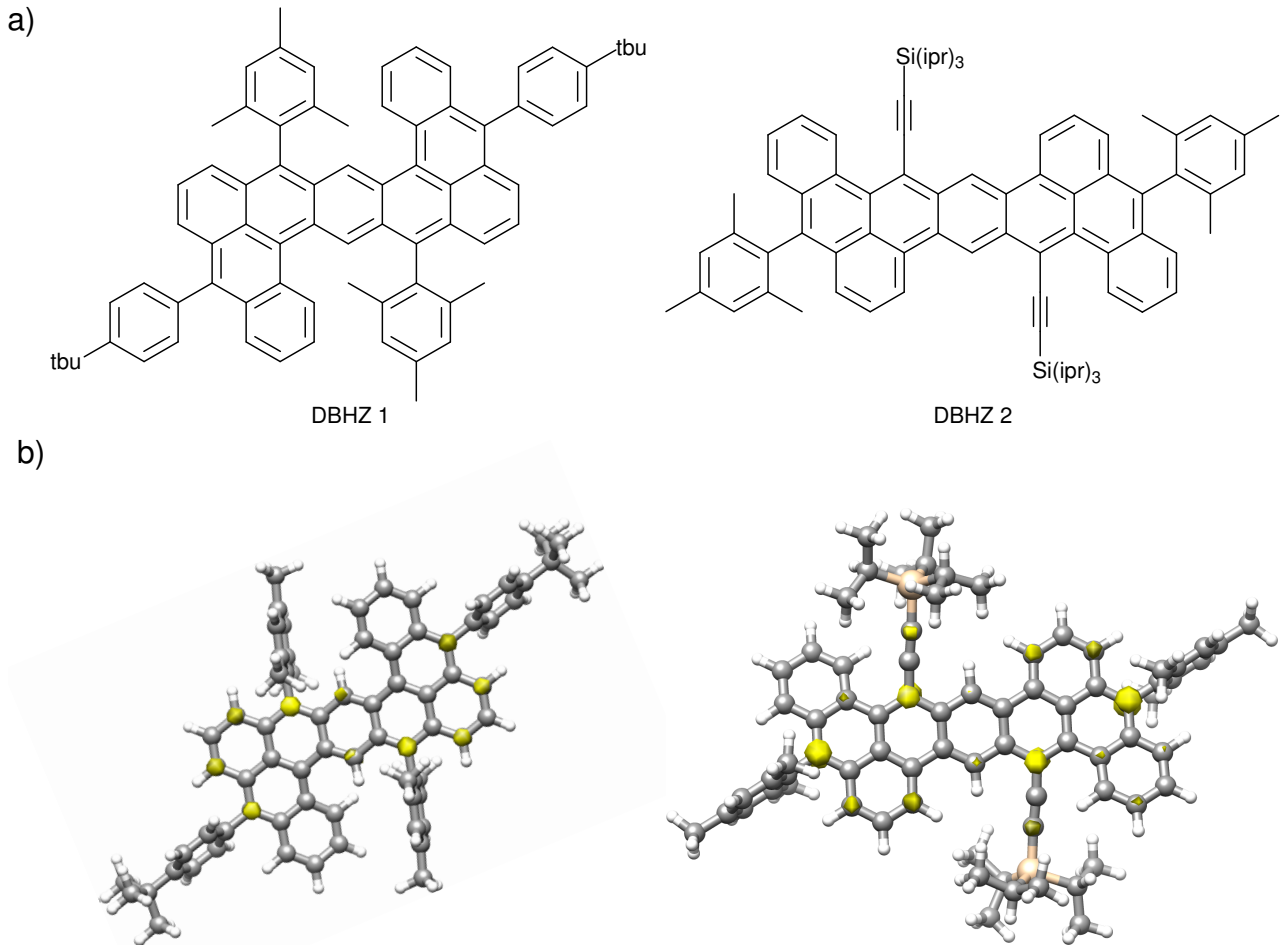

Figure 3.2.: a) Closed shell Lewis resonance structures of DBHZ1 and DBHZ2. b) FOD plots of DBHZ1 and DBHZ2[91] (FT-TPSS/def2-SV(P), T_{el} = 5000 K) at an isocontour value of σ = 0.005 e Bohr^{-3}.

with Clar's sextet[89] rule as both models identify the same carbon atoms as unpaired spin centers. We therefore argue that the FOD analysis leads to a view on organic biradicals that is consistent with resonance structures and Clar's sextet rule.

At this point, it is worth noting that related methods have been reported: Nakano et al. have analyzed the properties of biradicals based on the odd-electron density from long-range corrected density functional theory, which also provides a spatially resolved image of the unpaired electron density in open-shell singlet systems.[93] Lischka and co-workers have studied analogous bi- and polyradicaloid systems using multi-reference *ab initio* methods.[94] Their unpaired electron density also identifies the spin centers and yields valuable information about the electronic structure of PAHs. However, the advantages of the FOD analysis are that it is much faster to compute, and second that it works as a black box and does not require laborious multireference treatments.

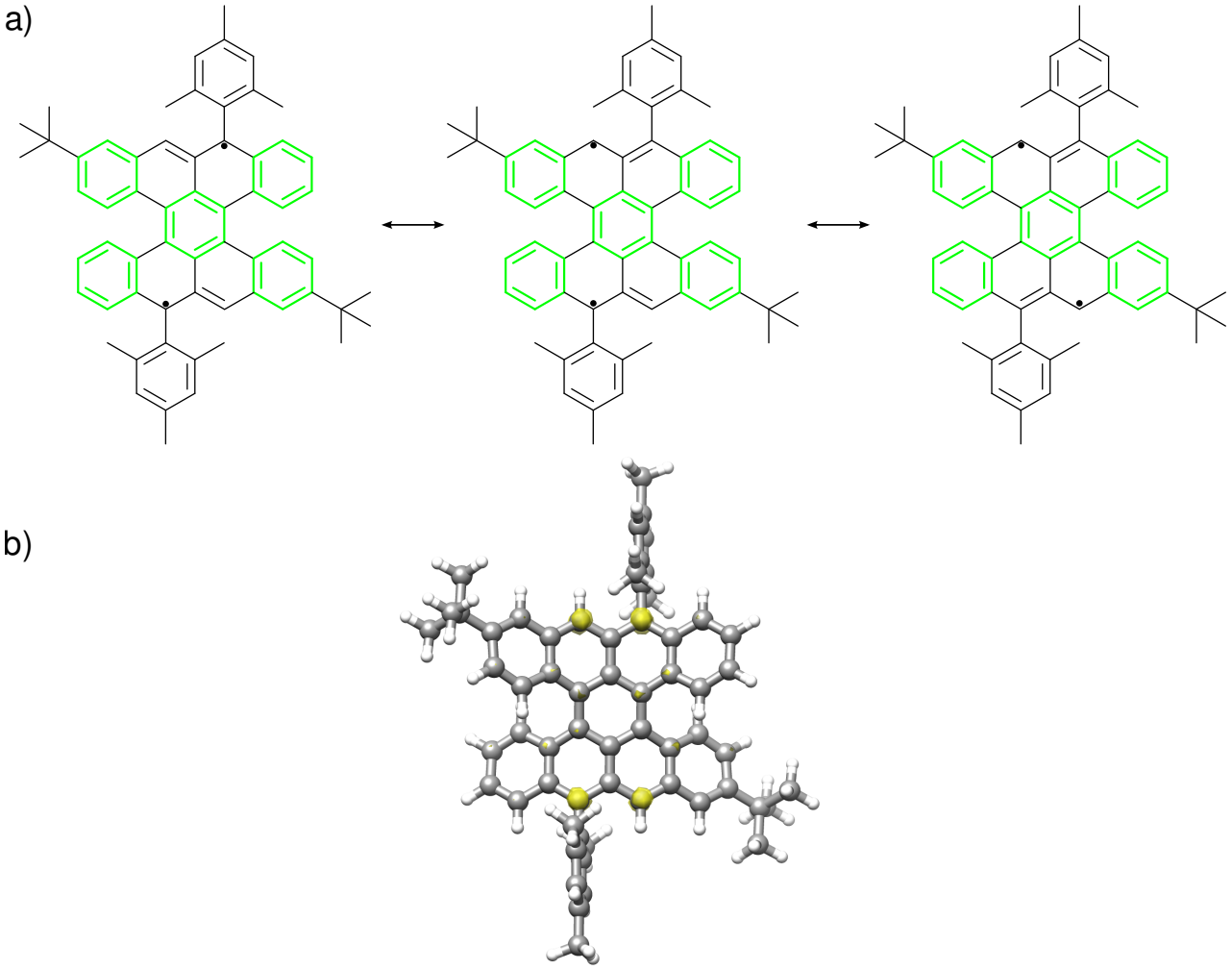

Figure 3.3.: a) Resonance structures of **TBP1** containing five Clar sextets, highlighted in green. b) FOD plot at $\sigma = 0.005$ e Bohr^{-3} (TPSS/def2-SV(P), $T_{\mathrm{el}} = 5000\,\mathrm{K}$)

The Relation of N^{FOD} to the Biradical Character

As a last point in this subsection, we show the correlation of N^{FOD} to the experimental and calculated biradical character y. To this end, we have investigated nine compounds with a reported experimental y value, see Figure 3.4. The compounds **1-7** have been taken from the report by Kamada.[83] Compound **8**, which is known to have a biradical singlet ground state, has been taken from a study of Wu and co-workers.[95] Compound **9** is a meso-diketo hexaphyrin with 26 π electrons and a singlet biradical ground state.[96]

We have calculated the number of 'hot' electrons N^{FOD} by the FOD analysis at the FT-TPSS/def2-SV(P) ($T_{\mathrm{el}} = 5000$ K) level of theory. These calculations come at a moderate cost of minutes to hours per molecule including structure optimization, depending on the available number of CPU cores. The FT-TPSS/SV(P) single point calculations take a maximum of a few minutes for the presented PAH systems. The results in Table 3.1 reveal that there is a clear correlation between the experimentally obtained biradical character and the value of N^{FOD}

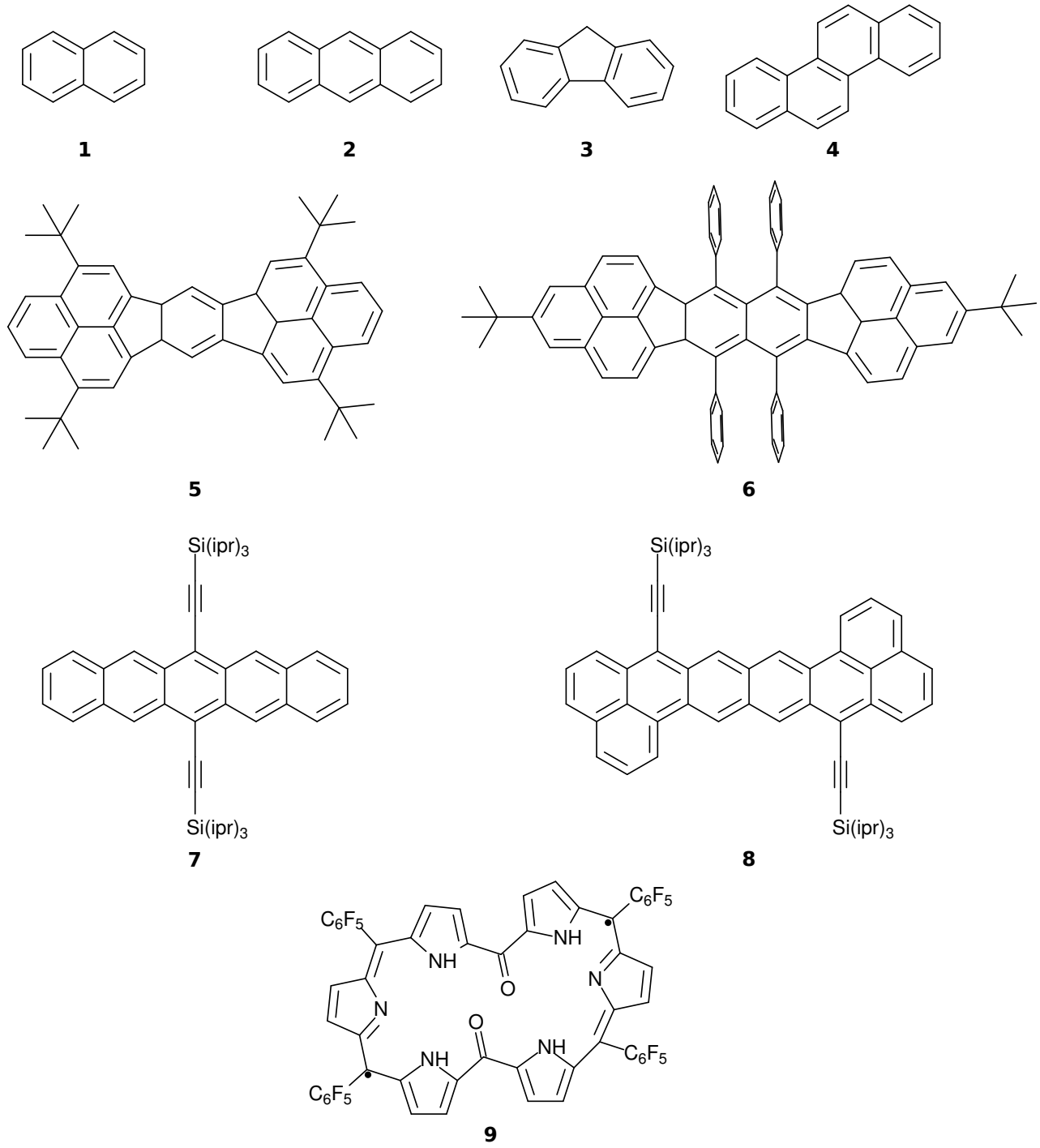

Figure 3.4.: Set of molecules with experimentally determined y values.[83,95,96]

Table 3.1.: Experimental biradical character (y) and calculated (TPSS/def2-SV(P), $T_{\mathrm{el}} = 5000$ K) N^{FOD} values of compounds **1-9**, see Fig. 3.4.

compound	y	N^{FOD}
1	0.02[83]	0.19
2	0.06[83]	0.29
3	0.04[83]	0.22
4	0.04[83]	0.36
5	0.34[83]	1.15
6	0.43[83]	1.61
7	0.15[83]	0.79
8	0.56[95]	1.28
9	0.63[96]	1.77

given by the FOD analysis. As shown in Figure 3.5, with a correlation coefficient $R = 0.95$, our calculated N^{FOD} values correlate very well with the experimental data. We therefore argue that the FOD analysis can be used in an *in silico* screening procedure to find compounds with a desired low, moderate, or high biradical character. The prerequisite for such a procedure is a moderate-cost level of DFT such as the TPSS/def2-SV(P) model chemistry we have used in this study. As shown below, even faster access to an FOD at only small loss of accuracy is provided by a semi-empirical tight binding model. The advantages of the $\rho^{\mathrm{FOD}}/N^{\mathrm{FOD}}$ model over the UNO approach are twofold: (i) Its technical robustness in terms of SCF convergence (because SEC effects are accounted for) and its speed, and (ii) its simple usage, as it avoids the calculation of a broken-symmetry UHF determinant entirely.

3.3.2. Selection of Active Spaces for Multiconfigurational Wave Functions

In this section, we present the FOD analysis as an efficient approach to choose the active space for a CASSCF wave function. We use FT-DFT computations at the BH-LYP/def2-TZVP level with $T_{\mathrm{el}} = 15000$ K, and the fractional occupation numbers to devise an active space.

We have chosen the BH-LYP functional because of its relatively large fraction of Fock exchange. It is known that Hartree-Fock lacks static correlation effects by construction, and that (meta-)GGA functionals already account for some of those by the locality of the exchange-correlation hole.[97,98] Hence, FT-DFT based on a (meta-)GGA partially double counts the SEC effects in a similar way as unrestricted DFT. With 50 % Fock exchange this problem is alleviated, and we consider BH-LYP to be a good compromise. The cutoff for the fractional occupation numbers is 1.98 for occupied orbitals, and 0.02 for virtual orbitals, respectively, meaning that orbitals with smaller/larger values are included in the CAS. Similar values have

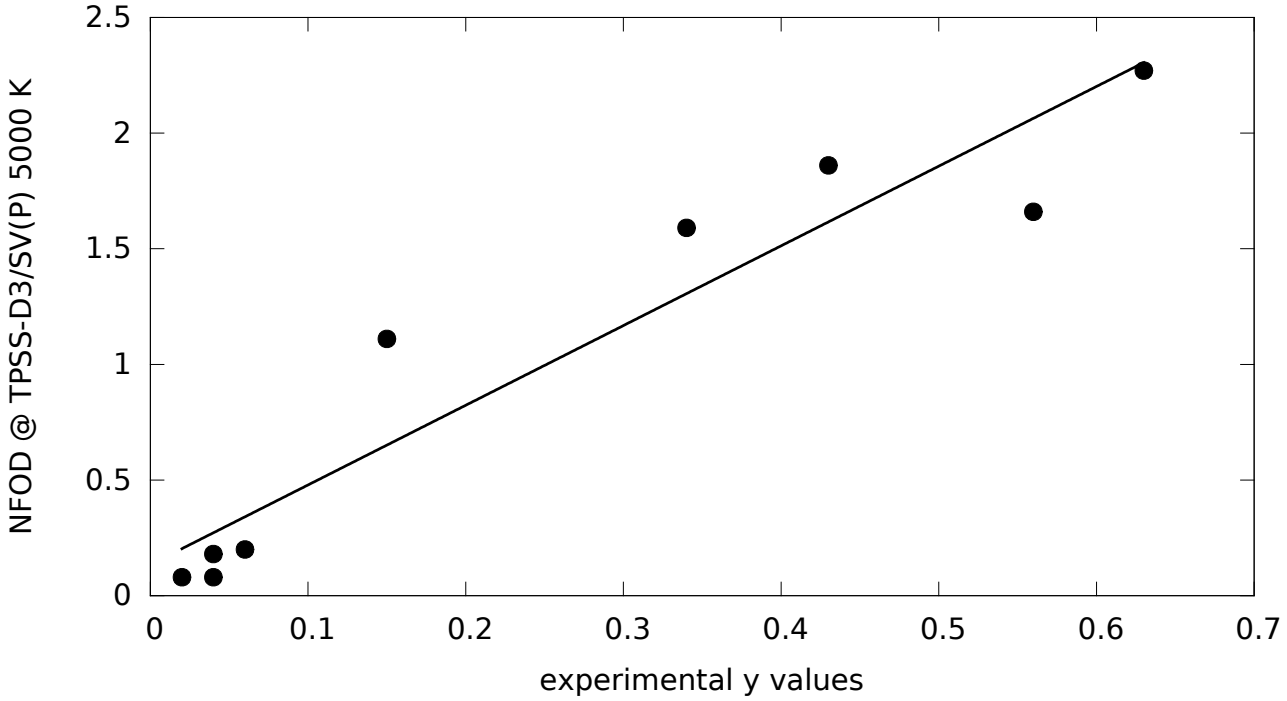

Figure 3.5.: Comparison of N^{FOD} (TPSS/SV(P), $T_{el} = 5000$ K) and experimentally derived y values. The linear regression correlation coefficient is 0.95.

been used in related studies.[56,99] First, we discuss organic molecules with a weight of the Hartree-Fock reference wave function below about 0.7 in the final CASSCF. These systems have a multiconfigurational ground state and should not be treated with single-reference quantum chemical methods. Our three examples are cyclobutadiene (D_{4h} symmetry), p-benzyne, and a structure of retinal twisted around the double bond at position 11, see Figure 3.6. Cyclobutadiene in D_{4h} symmetry is not a minimum on the molecular PES.[100] A large, delocalized FOD is visible (Figure 3.6 a), and the occupation numbers of the frontier orbitals are 0.974 each. We therefore compute a CASSCF(2,2) wave function, which features exactly equal weights of the reference wave function and the doubly excited configuration state function (CSF). Consequently, the CASSCF natural orbital occupation numbers (NOONs) are 1.000 for the frontier orbitals. We draw a similar picture for p-benzyne (Figure 3.6 b). This molecule, which has been discussed in the original FOD publication already,[46] is known to have a multiconfigurational singlet ground state. Based on the FT-BH-LYP (canonical) orbital occupation numbers, we compute a CASSCF(6,6) wave function. The CASSCF(6,6) NOONs compare quite well to the FT-DFT occupation numbers. The weight of the reference wave function is 0.65, and the weight of the doubly excited CSF is 0.27. A number of previous studies used a MCSCF(8,8) or CASSCF(8,8) wave function for *p*-benzyne.[101,102,103] While we agree with Lischka and co-workers that a CAS(2,2) is too small for this system[103], a smaller CAS(6,6) space for ground state or vertical singlet-triplet gap computations (which is beyond

the scope of this paper) seems to be sufficient on the basis of the FOD analysis. A twisted structure of retinal where the dihedral angle around the double bond in position 11 is set to 90° is our third example. The multiconfigurational character of the model structure is clear from Figure 3.6 c). Based on our cutoff values for FT-BH-LYP occupation numbers, we determined the active space to comprise six electrons in six orbitals, which proved to be a reasonable and stable choice for the CASSCF calculation. The CASSCF(6,6) wave function features two leading CSFs (the reference wave function and the doubly excited CSF with weights of 0.44 each), and one CSF consisting of two single excitations to different orbitals. A detailed investigation of the isomerization coordinate of retinal is given in the following section.

Now, we turn to three linear acene molecules of increasing size, anthracene, pentacene, and heptacene. These molecules served already as examples in the original FOD publication.[46] The FOD plots in Figure 3.7 indicate the growing multiconfigurational character of the singlet ground state with an increasing number of rings. Table 3.2 provides data to quantify this trend and demonstrate our CAS-selection procedure.

Choosing the active space size by the FT-BHLYP occupation numbers (same cutoffs as above), we arrive at a CAS(6,6) for anthracene, CAS(8,8) for pentacene, and CAS(12,12) for heptacene. The increasing values of N^{FOD} and the decreasing values of the weight of the restricted Hartree-Fock reference determinant ($w(|0\rangle)$) reflect the growing multiconfigurational character from anthracene to heptacene. Apart from the frontier orbitals, where the FT-DFT occupation numbers are quite smeared across the Fermi level, the FT-BH-LYP occupation numbers are very similarly to the CASSCF NOONs. We therefore argue that the FT-BH-LYP ($T_{\mathrm{el}} = 15000\,\mathrm{K}$) occupation numbers are a reasonable guideline for choosing the size of the active space. In 2012, Aiga[104] chose much larger active spaces for structure optimizations of acenes, and got sensible agreement with experiment. However, this might be due to partially accounting for dynamic correlation by choosing large active spaces. In our view the better strategy is to obtain a CASSCF wave function accounting mainly for SEC effects and subsequently add dynamic correlation by, e.g., perturbative schemes.

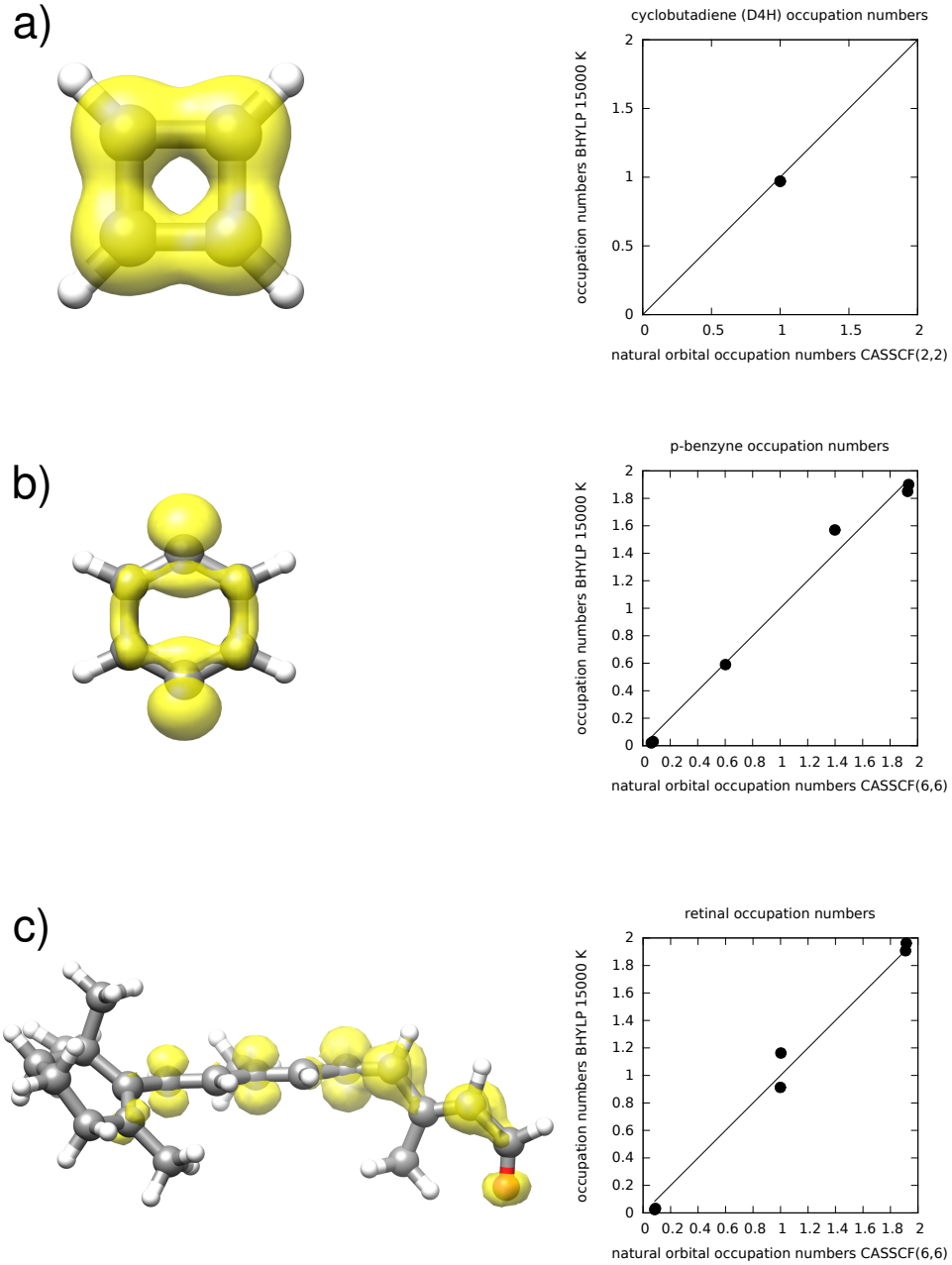

Figure 3.6.: FOD plots at $\sigma = 0.005$ e Bohr^{-3} (BHLYP/def2-TZVP, $T_{\mathrm{el}} = 15000\,\mathrm{K}$), and comparison of FT-DFT and CASSCF orbital occupation numbers of a) cyclobutadiene (D_{4h}), b) p-benzyne c) a perpendicular retinal structure. The occupation numbers are given in the supporting information in appendix A.

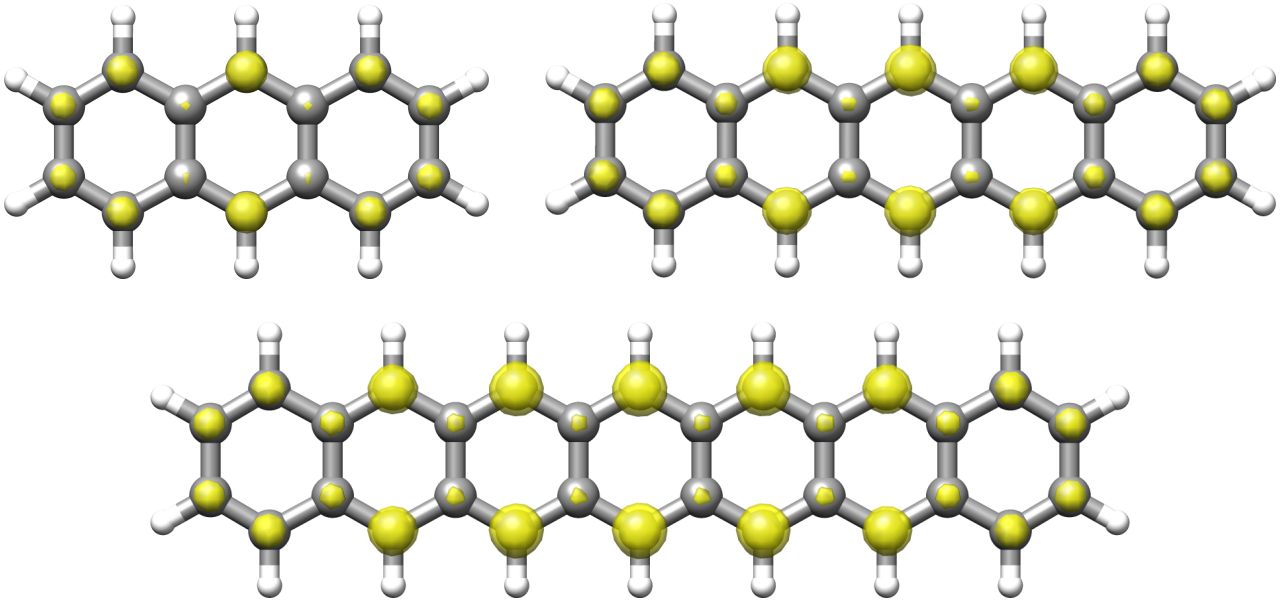

Figure 3.7.: FOD plots at $\sigma = 0.005$ e Bohr^{-3} (BHLYP/def2-TZVP, $T_{\mathrm{el}} = 15000\,\mathrm{K}$ of anthracene, pentacene, and heptacene.

Table 3.2.: Orbital occupations from FT-BH-LYP/def2-TZVP ($T_{\mathrm{el}} = 15000\,\mathrm{K}$) and CASSCF/def2-TZVP calculations. The gray cells mark the suggested choice of the active orbitals. CASSCF occupation numbers are NOONs in the active space, canonical orbital occupations (restricted Hartree-Fock wave function) otherwise. $w(|0\rangle)$ is the weight of the reference wave function.

	FT-BH-LYP occupations			CASSCF occupations		
index	anthracene	pentacene	heptacene	anthracene	pentacene	heptacene
LUMO+5	0.00	0.01	0.02	0.00	0.00	0.05
LUMO+4	0.00	0.01	0.02	0.00	0.00	0.05
LUMO+3	0.01	0.02	0.07	0.00	0.03	0.07
LUMO+2	0.02	0.06	0.07	0.03	0.05	0.08
LUMO+1	0.05	0.11	0.25	0.05	0.06	0.14
LUMO	0.24	0.52	0.73	0.07	0.10	0.23
HOMO	1.75	1.48	1.29	1.93	1.90	1.78
HOMO-1	1.94	1.88	1.73	1.94	1.94	1.86
HOMO-2	1.97	1.93	1.92	1.97	1.94	1.92
HOMO-3	2.00	1.97	1.93	2.00	1.97	1.93
HOMO-4	2.00	1.99	1.97	2.00	2.00	1.95
HOMO-5	2.00	2.00	1.98	2.00	2.00	1.95
N^{FOD}	1.08	2.08	3.16	–	–	–
$w(\|0\rangle)$	–	–	–	0.92	0.88	0.72

As a final batch of examples how the FOD can be used for CAS selection, we present results on three transition metal complexes with multiconfigurational ground states. The

FOD plots clearly indicate their complicated electronic structure, see Figure 3.8. The first

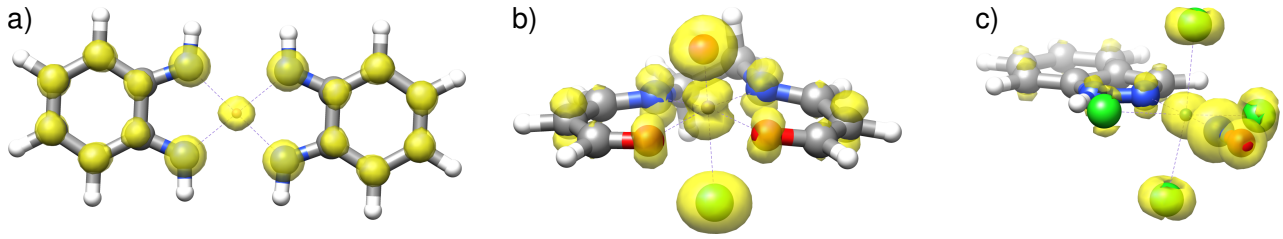

Figure 3.8.: FOD plots at $\sigma = 0.005$ e Bohr^{-3} (BHLYP/def2-TZVP, $T_{\mathrm{el}} = 15000\,\mathrm{K}$ of a) bis-*o*-phenylene Cobalt ([$\mathrm{Co^{II}}(^2L_{\mathrm{NN}}^{\bullet 1-})$]), b) oxo-Mn salen c) [$\mathrm{RuCl_4}$(Hind)(NO)]$^-$.

example is bis-*o*-phenylene cobalt ([$\mathrm{Co^{II}}(^2L_{\mathrm{NN}}^{\bullet 1-})$], which has been studied previously by Bill *et al.*[105] This complex has a $S = \frac{1}{2}$ ground state with strongly anti-ferromagnetically coupled electrons. Table 3.3 gives the unrestricted FT-BH-LYP and CASSCF occupation numbers for this system. From the FT-U-BH-LYP occupation numbers, one can either choose a CAS(11,8) (a conservative choice, with the cutoffs as above) or a CAS(7,6) (discarding the orbitals 65 and 66 with β occupation numbers > 0.95). We have computed both the CASSCF(7,6) and CASSCF(11,8) wave functions. As one can see from the occupations and the weights of the respective reference wave functions, CAS(7,6) appears to be large enough to describe the SEC effects in the ground state. The orbitals 65 and 66, inactive in the CASSCF(7,6) wave function, are nearly doubly occupied in the CASSCF(11,8) wave function. This example already shows that FT-U-BH-LYP occupation numbers and the FOD plot/N^{FOD} measures can provide valuable and cost-efficient input for the choice of a proper multireference wave function.

Next, we investigate oxo-Mn salen, which is a catalyst for enantioselective olefin expoxidation[106,107], and has been the subject of many *ab initio* studies in the last decades.[57,108,109,110,111,112] The most recent article on this system recommends an active space of 26 electrons in 21 orbitals, providing a balanced treatment of the singlet, triplet, and quintet states.[57] Our FT-BH-LYP occupation numbers for the singlet state (occupation of orbital 64: 1.98; occupation of orbital 73: 0.03) suggest a choice of a CAS(12,10). This CAS would clearly be too small to provide an consistent description of all the spin states. However, we suggest it as starting point for higher level *ab initio* calculations as Reiher and co-workers use a CASSCF(10,10) wave function to initiate their DMRG(26,21) computations.[57]

The final transition metal example is [$\mathrm{RuCl_4}$(Hind)(NO)]$^-$, which is a candidate and model for an anticancer drug in the context of delivering NO molecules in a biochemical environment.[113] Recently, Freitag *et al.* have studied the S_0 and T_1 states of this compound in depth using the CASSCF and DMRG methods.[114] The FT-BH-LYP occupation numbers show significant fractional occupations (1.98 ¡ n ¡ 0.02) for orbitals 72 (1.984) to 87 (0.021). This suggests a CAS(20,16), which differs from the active space that Freitag *et al.* used in their

Table 3.3.: Orbital occupations of $[Co^{II}(^2L^{\bullet 1-}_{NN})]$ from FT-UBH-LYP/def2-TZVP (T_{el} = 15000 K) and CASSCF/def2-TZVP calculations. CASSCF occupation numbers are NOONs in the active space, canonical orbital occupations otherwise. $w(|0\rangle)$ is the weight of the reference wave function.

	FT-UBH-LYP occupations			CASSCF occupations	
index	α	β	$\alpha+\beta$	CAS(7,6)	CAS(11,8)
72	0.01	0.01	0.02	0.03	0.04
71	0.47	0.01	0.48	0.63	0.60
70	0.67	0.62	1.27	1.00	1.01
69	0.92	0.65	1.57	1.37	1.40
68	0.94	0.84	1.78	1.98	1.98
67	1.00	0.93	1.93	1.99	1.98
66	1.00	0.97	1.97	2.00	1.99
65	1.00	0.98	1.98	2.00	1.99
N^{FOD}		2.56		–	–
$w(\|0\rangle)$		–		0.67	0.68

DRMG(18,18)-SCF calculations. However, applying just slightly less conservative cutoff values (1.975 ¡ n ¡ 0.025) – which appears to be reasonable for transition metal compounds (see the $[Co^{II}(^2L^{\bullet 1-}_{NN})]$ example above) – leads to a CAS(18,15). Considering that the DRMG(18,18)-SCF wave function of Freitag *et al.* contains three extra d orbitals (second d shell), which the authors deem negligible in terms of SEC, the FOD procedure provides a sensible choice here as well.

3.3.3. FOD Analysis for Rotations around Double Bonds

We now examine FT-DFT potential energy surfaces (PES) to investigate the effect of the Fermi-Smearing procedure on the shape of molecular energy landscapes. We choose the example of double bond rotations which cannot be treated with single-reference methods since there is a significant biradical character at the twisted transition state.[115] Many systems feature a conical intersection along the photochemical isomerization coordinate, e.g., butadiene.[116] We have carried out FT-DFT calculations on the model systems of ethylene and tetracyanoethylene (TCNE), and retinal molecules. The equilibrium geometries of ethylene and TCNE, and the *trans* isomer of retinal have been optimized at the PBE0[47,117]-D3[33,34,35]/def2-TZVP[48] level of theory. In the following, each reaction coordinate has been obtained by performing a rigid scan along the dihedral angles of ethylene, TCNE, and retinal. Note that these coordinates do not represent the true isomerization pathways, and in general they do not hit the conical intersections of the respective molecules because the intermediate structures are not optimized. However, these model structures are sufficient in order to elucidate the SEC effects, and to

compare between the different theoretical methods.
First, we discuss ethylene (C_2H_4), which exhibits strong biradical character in its perpendicular structure. Therefore, single-reference methods do not provide reliable results. In Figure 3.9, we show relative energies and diagnostics obtained from (FT-)DFT, CASSCF(2,2), and CCSD(T) calculations. Figure 3.9 a) displays the PES along the rigid scan coordinate with (FT-)TPSS. The TPSS PES shows a kink where the dihedral angle of ethylene approaches 90° whereas the FT-TPSS PES is smooth. The relative energy is lowered by the FT-DFT procedure at the points along the coordinate which exhibit substantial SEC effects. The relative energy at 90° is reasonable with 70 kcal mol^{-1}. The respective value of N^{FOD} at that point is 2.00, and the ρ^{FOD} plot clearly indicates the presence of SEC. In contrast, no ρ^{FOD} is visible at the equilibrium geometry displayed in Figure 3.9 a). We compare these results with those we have obtained from wave function theory in Figure 3.9 b). The CASSCF(2,2) PES resembles remarkably the FT-DFT PES, and the relative energy of perpendicular ethylene is 77 kcal mol^{-1}. Adding dynamic correlation by the NEVPT2 method yields very similar energetics, meaning that in this particular case, dynamic correlation lowers the total energy of each structure equally. The T1 diagnostic curve from CCSD(T) follows the curve of N^{FOD} and thereby indicates the usefulness of N^{FOD} as a multireference diagnostic. The CASSCF(2,2) wave function shows equal weights of the doubly excited CSF and the reference (Hartree-Fock) wave function. Both CASSCF(2,2) active natural orbitals are therefore singly occupied. We argue that FT-DFT and its static correlation measures ρ^{FOD} and N^{FOD} can describe these effects, which manifests itself in the smooth PES of FT-DFT.

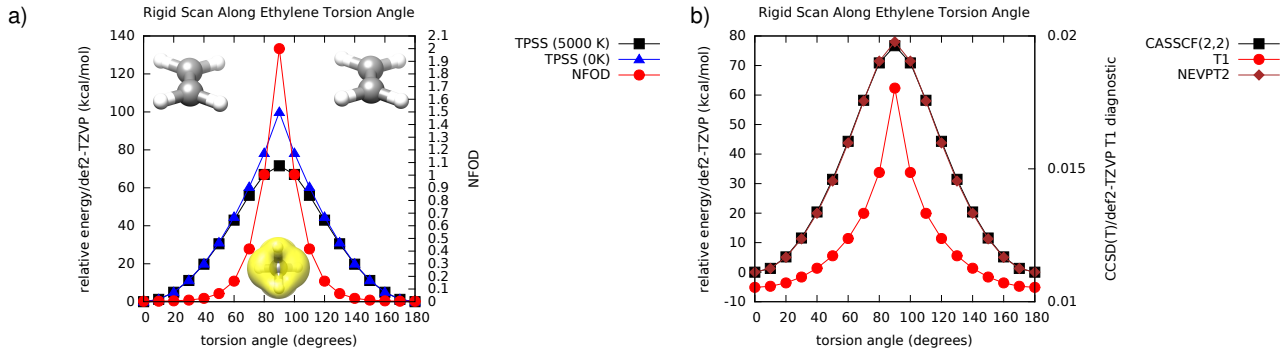

Figure 3.9.: Rigid scan along the ethylene torsion angle. a) TPSS and FT-TPSS ($T_{\mathrm{el}} = 5000$ K) b) CASSCF(2,2) and NEVPT2 energies, and CCSD(T) T1 diagnostic. In a) the FOD is plotted as well.

As a second example, we investigate tetra-cyanoethene (TCNE). Figure 3.10 a) shows the TPSS and FT-TPSS PES, and the value of N^{FOD} along the rigidly scanned PES. As in the case of ethylene, the TPSS PES has a kink at 90°, and a large, delocalized ρ^{FOD} is visible (inset in Figure 3.10 a). The FT-TPSS relative energy at this perpendicular structure is 37

kcal mol^{-1}, which is much lower than the respective TPSS energy of 73 kcal mol^{-1}. The shapes of the CASSCF(2,2), the NEVPT2, and the FT-DFT PES in Figure 3.10 b) are all smooth. In this case, the addition of dynamic correlation by the NEVPT2 method based on the CASSCF(2,2) wave function, lowers the energy of the perpendicular structure by a few kcal mol^{-1}. The curve of the T1 diagnostic from the CCSD(T) calculations follows the curve of the N^{FOD} along the same coordinate. Both curves have a sharp peak at 90°. Analogous to the case of ethylene, the CASSCF(2,2) wave function comprises two CSFs with equal weights at 90°, and the two active natural orbitals are singly occupied. The corresponding value of $N^{\mathrm{FOD}}(90°)$ is 2.00.

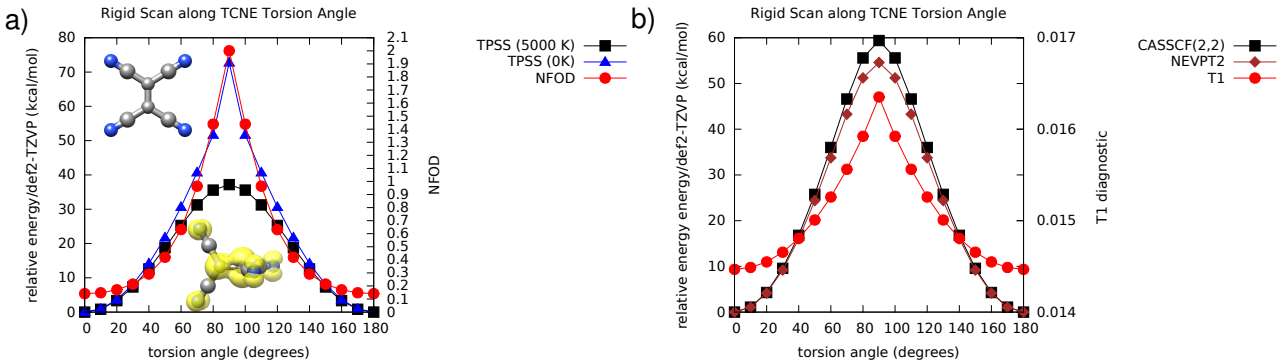

Figure 3.10.: Rigid scan along the TCNE torsion angle. a) TPSS and FT-TPSS ($T_{\mathrm{el}} = 5000$ K) b) CASSCF(2,2) and NEVPT2 energies, and CCSD(T) T1 diagnostic. In a) the FOD is plotted as well.

Next, we model the isomerization of retinal. The protonated Schiff base derivative of retinal is the chromophore involved in the vision process of seeing animals.[118] The corresponding photoisomerization pathways of retinal, its Schiff base derivatives, and model compounds have been studied thoroughly by multiconfigurational and multireference wave function methods.[119,120,121,122,123] Gozem et al. have investigated the isomerization pathways of a retinal chromophore model extensively using a multireference wavefunction methodology.[124,125] They found drastic effects of the dynamic electron correlation on the potential energy surfaces of that compound. They concluded that these may alter the interpretation of the simulations on photochemical and thermal isomerization pathways. Since there is a certain interest in thermal isomerization of retinal and its derivatives, we have carried out FT-DFT and DLPNO-CCSD(T) calculations on the retinal molecule. Figure 3.11 a) shows the model of the isomerization coordinate from 11-cis to all-trans retinal. The FT-DFT PES in Figure 3.11 b) is smooth along this rigid scan coordinate, with the 11-cis isomer being 20 kcal mol^{-1} higher in energy in the gas phase than the all-trans isomer. At 90° (inset in Figure 3.11 b), a large and delocalized ρ^{FOD} is observed, indicating the presence of SEC. The 11-*cis* and all-*trans* structures have only a small and localized ρ^{FOD}, and the approximate barrier height is 25 kcal

mol^{-1} which is in good agreement with the NEVPT2 value of 30 kcal mol^{-1}. Figure 3.11 b) shows that the CASSCF(6,6) PES (where we have chosen the size of the active space by the procedure outlined above) is smooth along the coordinate. Adding dynamic correlation by NEVPT2 significantly lowers the relative energies of the 11-*cis* structure and the (approximate) barrier height. Additionally, like the N^{FOD} curve, the DLPNO-CCSD(T) T1 diagnostic spikes sharply at 90°. While these model calculations do not bear much biochemical meaning, they show that FT-DFT is useful for exploring the ground state PES of retinal and its derivatives including regions with a large biradical character, and hence, significant SEC effects.The larger difference between FT-DFT and NEVPT2 results in the case of TCNE (underestimation of the barrier) compared to ethylene and retinal may be attributed to the self interaction error in the more polar system.

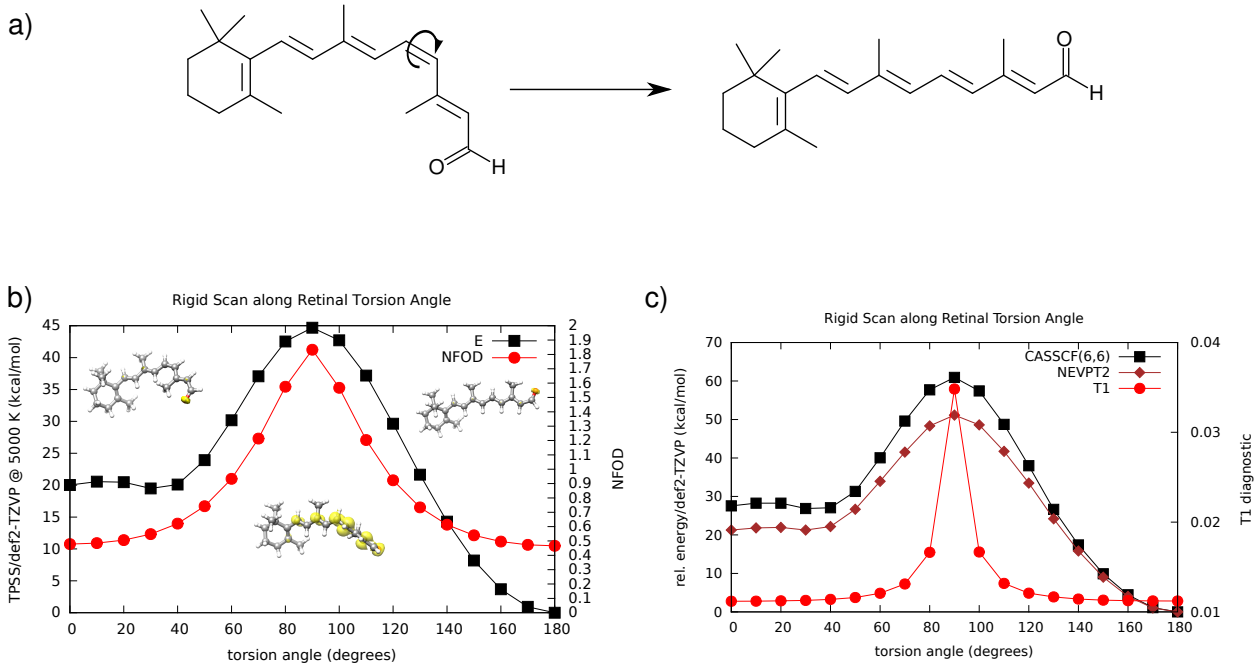

Figure 3.11.: a) The isomerization coordinate from 11-cis retinal to all-trans retinal. b) Rigid scan along the ethylene torsion angle at the FT-TPSS/def2-TZVP ($T_{\mathrm{el}} = 5000$ K) level of theory c) CASSCF(6,6)/NEVPT2 energies, and DLPNO-CCSD(T)/def2-TZVP T1 diagnostic along the same coordinate. In a) the FOD is plotted as well.

Furthermore, rotations around CC double bonds are also of importance in the design of molecular switches.[126] The goal here is to control the isomerization reactions, which may ultimately lead to control over molecular machines. One recent literature example of such a switch molecule investigates the collision- and light-inducible double bond rotation of a merocyanine dye.[127] Figure 3.12 shows its protonated form in its TTC rotamer (nomenclature taken from the reference) and the corresponding transition state. A large and delocalized

ρ^{FOD} is visible at the transition state (Figure 3.12 b, $N^{\mathrm{FOD}} = 2.00$), confirming the authors' statement that single-reference methods are inadequate for the calculation of the torsional barrier. This is another example where one can benefit from the FOD analysis in order to choose a proper quantum chemical method to calculate reliable barrier heights, and identify the 'hot' electrons in the molecule of interest. Very recently, Zeng and co-workers have performed calculations on the protoisomerization of indigo and its derivatives,[128] where they chose a multiconfigurational approach to calculate the barrier height. Following the PES results above, such processes could also easily be investigated by FT-DFT.

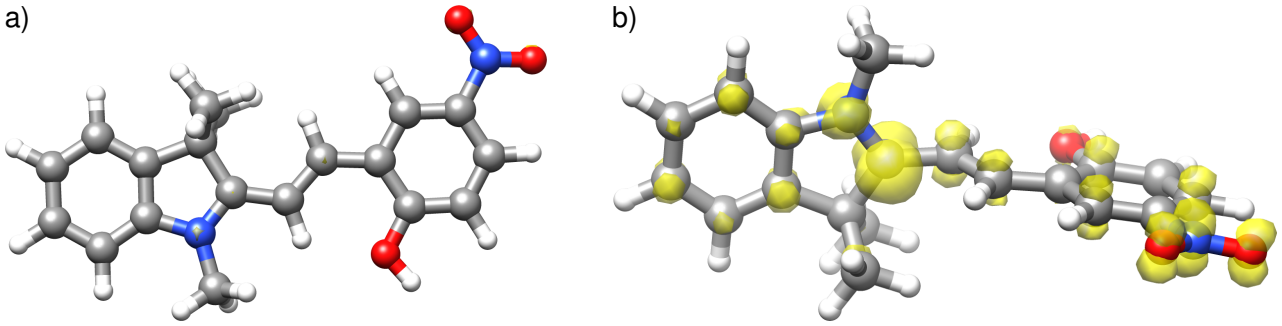

Figure 3.12.: FOD plots of a) protonated merocyanine (TTC rotamer) and b) transition state for double bond rotation. The geometries are taken from Markworth *et al.*[127]

3.3.4. FOD as an Indication for Static Electron Correlation in Large Biochemical Systems

Finally, we show that the FOD analysis can robustly visualize local SEC effects in extended molecular systems. We have chosen a specimen of the cytochrome P450 family[129] as a first example. These enzymes catalyze the oxidation of organic molecules, including alkanes *in vivo*. The H-rebound mechanism has been discussed extensively in the literature, both from the experimental and the theoretical points of view.[130,131] There is a consensus that the active species, termed Cpd I[131,132] is an iron-oxo compound, which arises by the oxidation of the ferrous iron in the active heme center of the protein by dioxygen. Schlichting *et al.* have managed to crystallize a P450 enzyme with a conspicuously short Fe–O distance of about 1.62 Å.[133] This crystal structure, PDB entry 1DZ9, appears to be the reactive complex between that molecule and a camphor ligand. With more than 15,000 atoms after the addition of H atoms to the structure performed by the protein preparation wizard program within the Schrödinger software suite[134,135], such a large molecule cannot be treated with standard DFT methods. We have restricted our calculation to chain A of the protein where a short Fe–O distance is observed. However, around 7,500 atoms are still too large for performing a reasonable DFT calculation. We have therefore resorted to a self-consistent charge (SCC) tight-binding (TB)

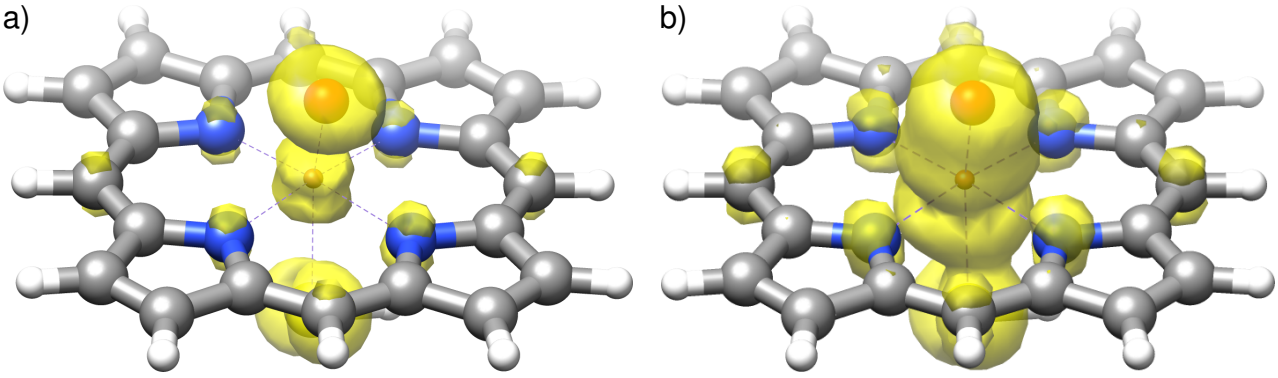

Figure 3.13.: FODs plots at $\sigma = 0.005$ e Bohr^{-3} of a minimal model for Cpd I. a) TPSS-D3/def2-TZVP, $T_{el} = 5000$ K b) GFN-xTB, $T_{el} = 4000$ K.

Hamiltonian related to the DFTB method.[43,136] The difference to the SCC-DFTB method lies in the parametrization and setup of the Fock matrix elements. Parameters will be available for the whole periodic table.[†] This semi-empirical quantum chemical method termed GFN-xTB, which will be published separately, provides DFT quality structures, electron densities, and similar, albeit slightly smaller HOMO-LUMO gaps than a GGA functional. We therefore choose $T_{el} = 4000$ K in the FOD calculations. The applied Fermi-smearing procedure is equivalent to that used in our DFT protocol.

First, we examine a minimal model of Cpd I, shown in Figure 3.13.[137] This model consists of a porphyrine molecule with an iron-oxo moiety in its center. The coordinating cysteine residue is modeled by a simple HS^- fragment, leading to a net neutral molecule with an odd number of electrons. It is well-known that an unpaired electron located on the porphyrine ring is coupled antiferromagnetically to the intermediate-spin iron-oxo core, leading to a doublet ground state.[137,138] In Figure 3.13 a), the FOD plot obtained at the DFT level of theory shows a large and delocalized FOD, which indicates the multiconfigurational ground state character. The FOD plot obtained by GFN-xTB is quite comparable to the DFT result, with a slightly larger degree of delocalization, owed mostly to a very small HOMO-LUMO gap of only around 0.1 eV. Nevertheless, the FOD plots obtained by both methods lead to the same conclusion that electronic structure of Cpd I can only be discussed in a meaningful way with quantum chemical methods that account for SEC effects.

In Figure 3.14, we visualize chain A of the whole protein. Our TB Hamiltonian, which we use in combination with an implicit solvation model (details on this Generalized Born/solvent accessible surface[139] type solvation model will also be presented elsewhere), gives the same FOD plot as for the small model system (cf. Figure 3.13 b). Figure 3.14 a) depicts the whole chain A of the protein (coordinates from PDB entry 1DZ9). Significant ρ^{FOD} is only visible

[†]The xTB program can be downloaded from the website of our resarch group, http://www.thch.uni-bonn.de/tc/downloads/. It includes the FOD analysis feature described in this chapter.

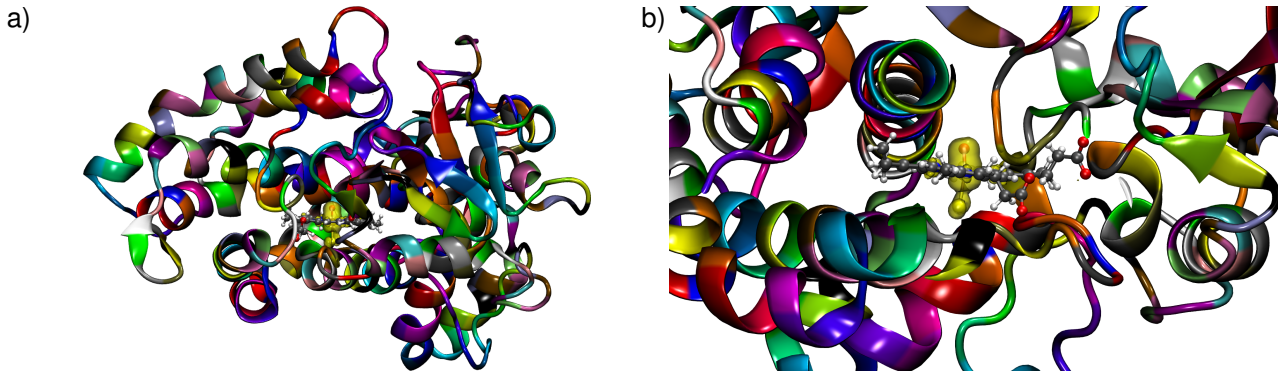

Figure 3.14.: FOD plot at $\sigma = 0.005$ e Bohr^{-3} (GFN-xTB Hamiltonian, $T_{\mathrm{el}} = 4000\,\mathrm{K}$) of cytochrome P450 iron-oxo reaction complex (PDB entry 1DZ9)[133]. a) chain A of the protein b) zoom into the active center.

in the active center containing the Cpd I structure. Figure 3.14 b) contains a zoomed view on the active center. To our knowledge, these are the first full QM based electronic structure calculations on thousands of atoms with a focus on SEC effects and multiconfigurational ground state character. The fact that the computed ρ^{FOD} from the model compound and the computation on the macromolecule give a consistent picture of the electronic structure is encouraging and suggests that carefully parametrized semi-empirical FT-TB Hamiltonians could be used in biomolecular full QM calculations to elucidate e.g., enzymatic mechanisms. As another example, we turn to the Cubredoxin protein (PDB entry 5K49)[140], which belongs to the class of artificial metalloproteins.[141] In Figure 3.15 we compare the FOD plots of the copper containing active site (coordinates are taken from the PDB structure and saturated with hydrogen atoms) obtained at DFT and semi-empirical levels of theory. The FOD plot obtained with FT-DFT (Figure 3.15 a) shows a smaller ρ^{FOD} than the one obtained by the GFN-xTB Hamiltonian (Figure 3.15 b). However, the degree of delocalization is comparable, as both methods identify the presence of SEC effects.

In Figure 3.16 we show that the GFN-xTB based FOD analysis GFN-xTB is able to identify the SEC in the active Cu site also in the full protein, and that the FOD plot is comparable to the one of the isolated ligand shown Figure 3.15 b). Additionally, there is a significant ρ^{FOD} at the C terminus of the structure, which has been wrongly saturated in the automatic preparation of the calculations: Instead of a COO^- group, a single hydrogen atom has been placed there, creating an aldehyde. This saturation is chemically meaningful, but biochemically wrong, and may lead to unintended consequences in the simulations.

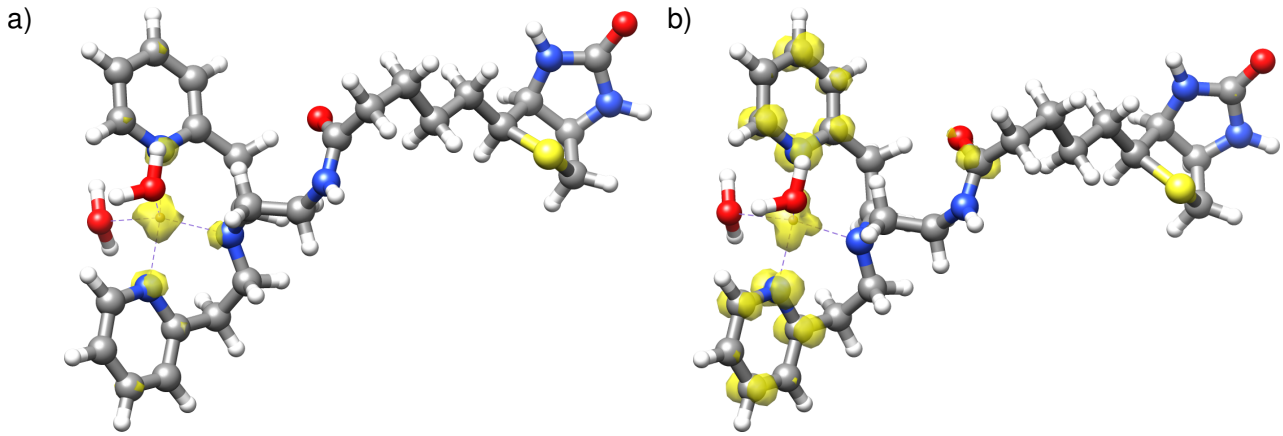

Figure 3.15.: FOD plots at $\sigma = 0.005$ e Bohr^{-3} of the active Cu site of an artificial cubredoxin,[140] PDB entry 5K49. a)TPSS-D3/def2-TZVP, $T_{el} = 5000\,K$, b) Tight-Binding Hamiltonian, $T_{el} = 4000\,K$.

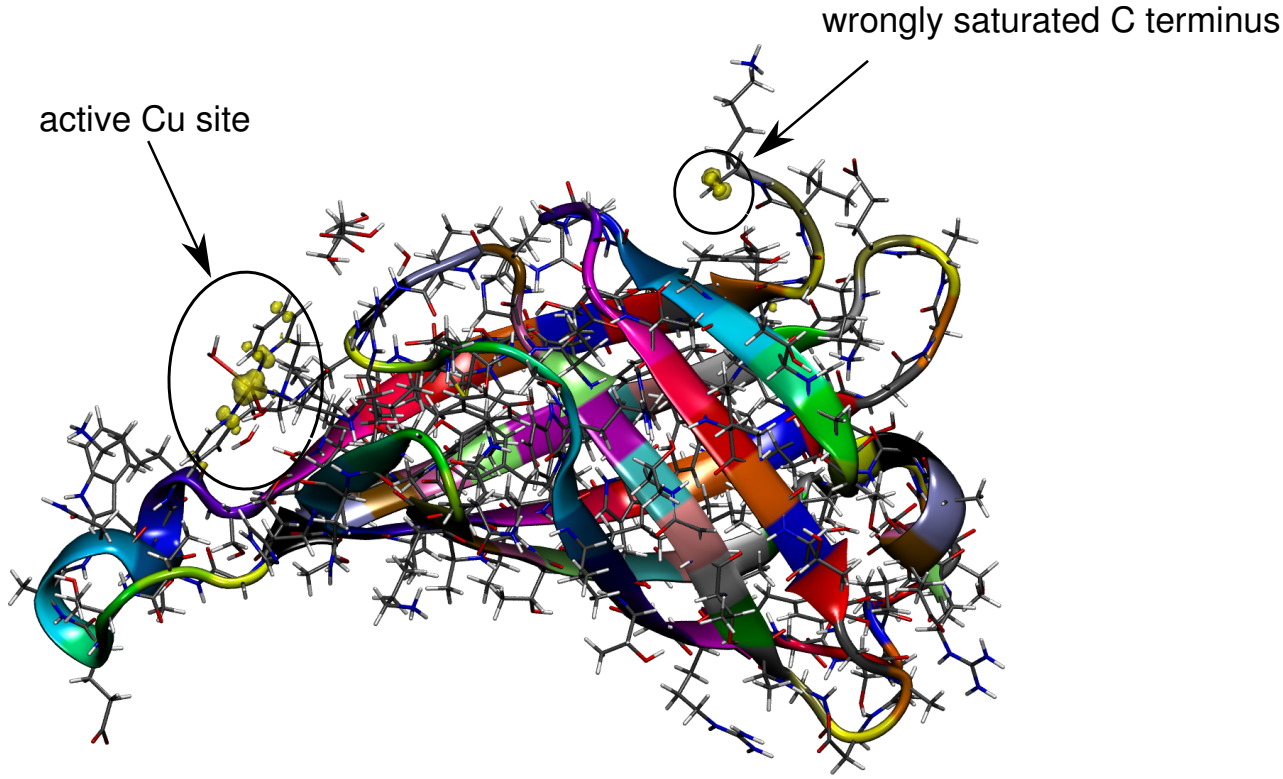

Figure 3.16.: FOD plot (GFN-xTB Hamiltonian, $T_{el} = 4000\,K$) at $\sigma = 0.005$ e Bohr^{-3} of an artificial cubredoxin,[140] PDB entry 5K49. The FOD identifies the active Cu site and the wrongly saturated C terminus.

The final example concerns the structure of a human artemin crystal structure, PDB code 2ASK.[142] This crystal structure contains an unusually long disulfide bond between the cysteine residues 135 and 199, with a sulfur-sulfur distance of 2.76 Å. The FOD plot in Figure 3.17 a) reveals that there is a significant (but unexpected) ρ^{FOD} at that disulfide bridge. All other

visualized instances of ρ^{FOD} are negligible for this structure. Figure 3.17 b) shows that all other disulfide bridges do not display any visible ρ^{FOD}, as the sulfur-sulfur distances are around 2.25 Å. In the case of artemins, the disulfide bridge networks are characteristic, and the large S–S distance shrinks upon structure minimization because the established preparation and force-field algorithms for biomolecular simulations automatically treat this elongated disulfide bridge with the appropriate bond terms. However, in general, there could be cases where it is not obvious whether a structure contains a disulfide bridge or whether two cysteine residues are merely spatially close. In these cases, the FOD analysis may give an indication on problematic sites, and may suggest whether it is more sensible to place protons or an S–S bond there.

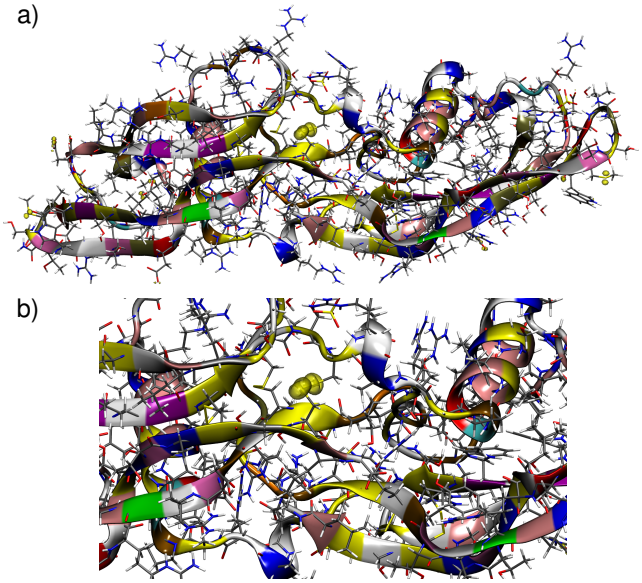

Figure 3.17.: FOD plot (GFN-xTB Hamiltonian, $T_{\mathrm{el}} = 4000\,\mathrm{K}$) at $\sigma = 0.005$ e Bohr^{-3} of a human artemin crystal structure,[142] PDB entry 2ASK. a) total view b) zoom on the disulfide bridge between residues 135 and 199. The ρ^{FOD} indicates the unusually long sulfur-sulfur distance.

Based on the above experience with the FOD analysis for protein systems, we suggest it as a new tool to find subtle errors in protein structures. It is essential to identify these errors prior to running large-scale biomolecular simulations.[59]

3.4. Conclusions

We have presented prototypical applications of the computational FOD analysis in four fields of chemistry. In each field, we have focused on the presence of static electron correlation (SEC) effects leading to systems with complicated electronic structures. SEC introduces unforeseeable non-additive effects into electronic wave functions and derived properties and hence, their detection and analysis is of high relevance. Based on the results presented above

for a wide range of different exemplary cases, we conclude the following:

1. The FOD analysis is a cost-efficient, simple to use computational method to gauge the biradical character of organic molecules like PAHs. The FOD analysis reproduces the correct trends at a reasonable computational cost, and N^{FOD} correlates will with the experimentally obtained biradical character. The FOD analysis may therefore be used for *in silico* screening procedures to identify compounds with a desired low, moderate, or high biradical character. The ρ^{FOD} identifies the same spin centers as qualitative open-shell resonance structures, and therefore provides a consistent picture of organic biradicals.

2. FT-DFT occupation numbers, computed using the BH-LYP functional ($T_{\mathrm{el}} = 15000$ K) serve as a good initial guess for the selection of active orbital spaces for MCSCF/CASSCF calculations. While the occupation number thresholds for occupied and virtual orbitals are empirical, choosing the size of an active space in this way is efficient, and requires no further input. ρ^{FOD} additionally indicates the localization and the extent of the SEC effects.

3. The PESs for double bond twists calculated by FT-DFT are smooth and parallel to CASSCF/NEVPT2 surfaces, as demonstrated on the examples of C_2H_4, TCNE, and retinal.
   Indeed, the robustness of FT-DFT for PES computations has recently been the subject of several studies using methods related to the FOD analysis[143,144]. It has been compared to standard single-reference methods in the exploration of the PES for the molecular dynamics based prediction of mass spectra, which due to the many bond breaking events requires definitely going beyond single-reference methods, see Ref. 9. A study on the quantitative accuracy of FT-DFT involving benchmark results on mass spectrometrically relevant reactions is underway and will be presented elsewhere.

4. The FOD analysis is extendable to large, biomolecular systems. A semi-empirical, yet fully quantum mechanical treatment of protein structures has revealed the following: (i) The FOD analysis is robust with respect to the level of theory, as both DFT and tight-binding methods can be used for the underlying electronic structure computations. This has enabled us to apply the FOD analysis to structures of more than 7,500 atoms of a cytochrome P450 protein where ρ^{FOD} indicates the presence of SEC effects in its active center. (ii) The FOD analysis may detect faulty structures. We have shown that ρ^{FOD} appears at sites of a wrongly saturated C terminus in a protein, and at an elongated disulfide bridge. The FOD analysis is therefore able to indicate problematic sites in biomolecular structures, which may occur even after routine preparation by biomolecular software.

These findings are encouraging, and we hope that the FOD analysis will become a staple for the unraveling of (anti)cooperative electronic many-body and SEC effects. The FOD analysis will be implemented in the next release of the ORCA program (ORCA V. 4.0), and may be used with the TURBOMOLE suite of programs (version 6.7 or higher) in conjunction with additional tools, which are available at our website, see http://http://www.thch.uni-bonn.de/tc/software/FODplot/. The GFN-xTB method including the FOD analysis is implemented in a stand-alone code which can be freely downloaded from our website.

Part III.

Prediction of Electron Ionization Mass Spectra

Chapter 3 has shown the versatility of FT-DFT and FT-GFN-xTB approaches. In this part of the thesis, FT-DFT and semiempirical FT quantum chemical calculations such as FT-GFN-xTB are used to compute the energy and the forces acting on molecules, which are then used in BO-MD simulations (cf Fig. 2.2). This part includes the main body of results presented in this work. It deals with the prediction of EI mass spectra for a broad range of molecules. Chapter 4 first presents a literature overview on studies that have been conducted to compute EI mass spectra or specific features thereof such as branching ratios and base peaks. RRKM studies are reviewed as well as non-statistical quantum chemistry studies and MD-based simulations. Then the QCEIMS methodology is introduced in detail and evaluated on a small benchmark.

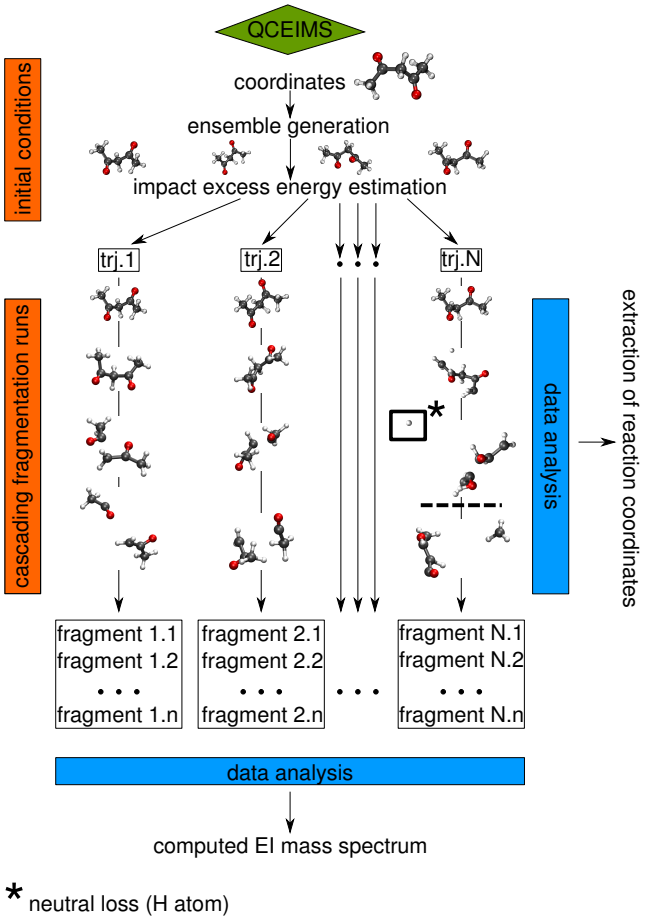

Figure 3.18.: Overview of the QCEIMS work-flow. Reproduced from Chapter 4.

Chapters 5 – 7 present applications of the QCEIMS method on large organic drug molecules and nucleobase systems, and Chapter 8 presents the first GFN-xTB computed EI mass spectra for compounds composed of 24 different elements.

4. How to Compute Electron Ionization Mass Spectra from First Principles

Christoph Alexander Bauer and Stefan Grimme*
Received 21st of March 2016, Published online 3rd of May 2016

Reproduced (adapted) with permission from The American Chemical Society
Christoph Alexander Bauer and Stefan Grimme J. Phys. Chem. A **2016**, *120*, 3755–3766.
 DOI 10.1021/acs.jpca.6b02907

Own manuscript contribution

- Performance of all calculations
- Interpretation of data
- Writing the manuscript

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

4. How to Compute Electron Ionization Mass Spectra from First Principles

Abstract The prediction of Electron Ionization (EI) mass spectra from first principles has been a major challenge for quantum chemistry (QC). The unimolecular reaction space grows rapidly with increasing molecular size. On the one hand, statistical models like Eyring's Quasi-Equilibrium Theory (QET) and Rice-Ramsperger-Kassel-Marcus (RRKM) theory have provided valuable insight, and some predictions and quantitative results can be obtained from such calculations. On the other hand, Molecular Dynamics (MD)-based methods are able to explore automatically the energetically available regions of phase space and thus yield reaction paths in an unbiased way. We describe in this feature article the status of both methodologies in relation to mass spectrometry for small to medium sized molecules. We further present results obtained with the "Quantum Chemistry Electron Ionization Mass Spectra" (QCEIMS) program developed in our laboratory. Our method, which incorporates stochastic and dynamic elements, has been a significant step towards the reliable routine calculation of EI mass spectra.

4.1. Introduction

The calculation of Electron Ionization (EI) mass spectra is a hard problem for chemical theory. While mass spectral interpretation and prediction algorithms based on empirical chemical rules and cheminformatics have had considerable success[145,146,147,148], they are only able to describe fragmentation pathways that are already known. We focus in this article on the prediction of EI mass spectra from first principles. This task demands the knowledge of the Ionization Excess Energy (IEE) distribution and the description of all the energetically accessible regions of reaction space. The challenge is twofold for the computation of EI mass spectra:

1. It is presently not feasible to determine the IEE distribution for molecular ions from first principles. The experimental route to the IEE distribution, electron momentum spectroscopy,[18] is a specialized, comparatively rarely applied technique. Data are available only for small molecules, e.g. small hydrocarbons[19].

2. The number of possible fragmentation pathways grows rapidly with molecular size.[149] All attempts at the prediction of a full mass spectrum run into the problem of innumerable fragmentation pathways. It is by no means easy to evaluate the activation (free) energies of these pathways.

Advances in computing power and electronic structure theory have brought about a remedy, and the combination of statistical theories and MD-based approaches has appeared as the most promising choice of a simulation protocol to treat huge unimolecular reaction spaces. We briefly review the progress in the field of mass spectral prediction since the 1950s and put our own contribution to this field into its this perspective. We also analyze in some detail the redistribution of the molecular charge upon fragmentation in our model which has hitherto not

been discussed thoroughly. Finally, we present recent results obtained with the QCEIMS[10] approach and give an outlook on possible applications.

4.1.1. Statistical and Non-Dynamic Approaches

RRKM/QET Theory

The first approach capable of computing mass spectra is based on statistical theory. It is known as Quasi Equilibrium Theory (QET)[14] and Rice-Ramsperger-Kassel-Marcus (RRKM) theory[22,23,24,25]. Both theories are so well-established that they have been included in the IUPAC's definitions of terms relating to mass spectrometry[150]. Statistical theory works under the assumptions that only the statistical redistribution of the energy E to the internal degrees of freedom determines the rate constant[2,151]

$$k(E) = \frac{\sigma N^{\ddagger}(E - E_0)}{h\rho(E)}, \tag{4.1}$$

where σ is the reaction path degeneracy, $N^{\ddagger}(E - E_0)$ is the transition state sum of states, E_0 is the activation energy, h is Planck's constant, and $\rho(E)$ is the density of states. Usually, $\rho(E)$ is taken to be the vibrational density of states. The rotatory and translatory degrees of freedom are often neglected,[151] and the vibrational states are usually calculated by invoking the harmonic oscillator approximation. $\rho(E)$ may be calculated by using a direct count of states algorithm.[152] For the effect of anharmonicity, see a study by Yao and co-workers on the dissociation of ethylene.[153] In RRKM theory, one easily obtains the qualitative result that the rate constant increases with increasing energy and decreases with the number of vibrational degrees of freedom (contained in $\rho(E)$). E_0 can either be calculated or taken from experimental data.

This theory was expanded to non-equilibrium situations by Drahos and Vékey.[154] Their method, called "Mass Kinetics" takes into account effects of acceleration, collisions and photon exchange by so called master equations. Additionally, the differences between the original RRK theory and RRKM theory were surveyed by Ervin.[155] The landmarks in the theory of mass spectra have been reviewed by Lorquet[156], and the reader is referred to that review article regarding the (sometimes problematic) marriage of quantum chemistry, (quantum) transition state theory,[157] and statistical theory. We refer to an extensive review on the use of statistical theory for Collision-Induced Dissociation (CID) processes by Armentrout, Ervin and Rodgers published in this journal's centennial feature issue.[158] Armentrout and co-workers have since been prolific in studying CID processes of amino acid derivatives, e.g. protonated glycine (Gly) and Gly-Gly.[159] Below, we give an overview of what are in our view the most important as well as some very recent applications of QET/RRKM theory in relation to EI mass spectrometry and the fragmentations of radical cations.

In their seminal 1952 paper, Rosenstock et al. calculated the EI mass spectrum of propane based on experimental appearance energies of the fragment ions and a simple distribution function for the IEE of up to 12 eV.[14] They found that there was an excellent agreement of the calculated ratio of fragment ions with the experimental ratios but noticed at the same time the arbitrariness of their calculation because there was a dependence on the choice of the activated complex.

RRKM theory has since then been applied to many long-standing questions in organic mass spectrometry. The most prominent example is the formation of the tropylium ion from the toluene radical cation, addressed in a famous paper by Lifshitz.[160] In that article it is concluded that RRKM theory employing *ab initio* results[161,162] predicts the correct tropylium/benzylium abundances. Although some details of Lifshitz's findings have been called into question[163], and the potential energy surface has been reexamined at a higher level of theory[164], her main conclusions still hold. Choe and co-workers have consecutively carried out RRKM studies on benzylium vs. tropylium formation from the ethyl- and propylbenzene radical cations,[165] where they have noted an increase of benzylium abundance with increasing length of the alkyl substituent. For the chlorotoluene[166], and the bromo- and iodotoluene radical cations[167], their conclusion has been that the benzylium ion forms preferentially below a certain energy threshold.

In 1996, Vékey calculated the fragment ion yields of butyl benzene with different internal energy distributions and showed the effect on the resulting mass spectrum.[168] Knyazev and Stein also modeled the dissociations of butyl benzene.[169] In their 1997 review of statistical theory in relation to mass spectrometry, Baer and Mayer pointed out the usefulness of RRKM calculations and gave the example of chlorobenzene dissociation.[151] The tropylium/benzylium dichotomy has continued to attract the interest of physical chemists recently[170]. Further work on aromatic compounds was presented by Halbert and Bouchoux, who scrutinized the decomposition of the butyl benzene radical cation. They found a molecular thermometer in the branching ratio of the reaction pathways leading to $m/z\,92$ and $m/z\,91$.[171] Muntean and Armentrout modeled the loss of CO from the phenol cation[172] and the dissociation of n-butylbenzene[173] by RRKM theory. Numerous aromatic compounds including thiophenol,[174] and picoline[175] were also studied by Choe et al.

Wolken et al. conducted a study on the decomposition of ionized cytosine.[176] That report concluded that several ionized tautomers of cytosine could equilibrate prior to dissociation based on RRKM calculations. Such computations were also applied to describe competing reaction channels of triphenyl methanol, which was cooled by helium nanodroplets prior to dissociation.[177] Vallejo Navaez et al. used RRKM theory to describe the fragmentations of substituted $N-$(2-methylfuryl)anilines.[178] They found excellent agreement between the experimental fragment ion yields and the calculated branching ratios.

The fragmentation of peptide radical cations has also been an active area of research, and RRKM theory is routinely applied there to explain branching ratios for competing reaction paths. Tureček and co-workers have extensively reviewed this field.[179] One study that appeared after that review on the dissociation of the Tyr-Gly-Gly radical cation was published by Lai et al.[180]

Sztáray, Bodi and Baer described RRKM modeling of photoelectron-photoion coincidence spectroscopy (PEPICO) experiments in 2010.[181] Sztáray and Baer had successfully used PEPICO in conjunction with RRKM theory to characterize the dissociation dynamics of an organometallic Cobalt complex.[182] Rennie et al. calculated the RRKM breakdown diagrams of methyl tert-butyl ether and methyl trimethylsilyl ether and compared them to experimental data.[183] Bodi, Stevens and Baer used high-level composite methods to calculate the activation energies of dichloro-ethylene ion dissociation reactions that were experimentally described by imaging photoelectron-photoion coincidence spectroscopy (iPEPICO).[184,185] Mayer and co-workers studied the dissociation of polycyclic aromatic hydrocarbons in joint experimental and iPEPICO studies.[186,187,188]

In 2013, statistical theory was used to calculate fragmentations of interstellar C_n molecules.[189] In that study, the authors successfully brought a concept originating from the field of mass spectrometry to the field of astrochemistry. Tsyshevsky et al. conducted RRKM calculations with energy evaluations at the CCSD(T) level of theory to describe the main fragmentation pathways of 1-nitropropane.[190]

In one of the most recent RRKM studies, Solano and Mayer report an extensive scrutiny of the PES of naphthalene$^{\bullet+}$.[191] They predict the formation of pentalene$^{\bullet+}$ after C_2H_2 loss, which has very recently been confirmed experimentally.[192] This goes to show that RRKM theory has continually served as the model of choice for the fragmentation reactions of molecular ions – as long as the number of possible reaction pathways allows for the sufficient exploration of phase space.

Non-Dynamic Approaches without the Use of RRKM Theory

A second line of quantum chemical calculations predicting mass spectral fragmentations has appeared since the 1990s. These studies make no use of statistical theory. Instead, the authors carry out QC calculations of stationary points on the PES, and analyze the electronic and molecular structures and reaction energies in order to predict the main decomposition pathways of molecular ions qualitatively.

Mayer and Gömöry postulated in 1994 that the base peaks in the EI mass spectra of simple organic compounds could be rationalized by calculating bond orders and energy partitions of (semi-empirical) Hamiltonians.[193,194] A related approach by the same authors was published in 2001[195], after those kinds of calculations were applied to explain the fragmentations of norbornane- and norbonene-fused heterocyclic molecules,[196] cyclopropyl silanes[197], and pep-

tide fragmentations[198]. Zayed and co-workers have published numerous studies using this method, e.g. for codeine.[199]

In 2000, Improta, Scalmani and Barone used similar reasoning to argue that their DFT calculations indicated the main fragmentation paths of the nucleobases by comparing the optimized geometries of the neutral species and the radical cations.[200] Arani et al. rationalized the fragmentation paths of the uracil[201], cytosine, adenine, and guanine[202] radical cations by analyzing DFT reaction free energies. Similar studies were quite recently conducted by Minaev et al. for adenine and Dawley et al. for adenine and hypoxanthine.[203,204] The fragmentation pathways of the guanine radical cation were computed by Cheng and co-workers.[205]

The aforementioned reports concern relatively small molecules of biological interest where the number of reasonable reaction pathways is still manageable. The non-dynamic QC description of the chemical reaction space becomes more and more tedious with increasing number of nuclear degrees of freedom, and therefore, it is limited. The same holds for RRKM/QET applications, which provide a deeper insight into the reaction kinetics than the non-statistical, non-dynamic QC predictions of molecular decomposition reactions. These studies include no general method of automatically and safely recognizing the important degrees of freedom along which a bond fission or rearrangement will take place. Moreover, the activation energy must be computed on a high level of theory to obtain reliable results.[206] One therefore has to predefine sensible reaction coordinates using chemical intuition, which may bias the result of the calculations and lead to an incomplete picture of the decomposition pathways.

4.1.2. Molecular Dynamics-based Approaches

With the rise of computer power and the advent of efficient QC implementations, it has become possible to simulate chemical dynamics directly. In order to treat chemical reactions accurately, one often turns to Born-Oppenheimer Molecular Dynamics (BOMD)[13,207] (sometimes also called "direct dynamics"[208]). In such simulations, the nuclei move classically on a PES determined on the fly by a QC method. Quantum effects like vibrational zero-point energy and tunneling are ignored due to the prohibitive cost of taking them into account in this method. Reactive trajectories that result from BOMD calculations can be analyzed with regard to reaction coordinates and reaction times. The reaction rate is therefore a direct result of BOMD calculations, and anharmonicity effects are automatically included.

All the research reviewed below has been carried out by programs employing the same strategy (shown schematically in Figure 4.1): First, a set of initial conditions is statistically generated for a number of trajectories. This becomes necessary because every individual molecular ion has a different internal energy. The number of trajectories is an important parameter to converge the results of the calculations because one reaction trajectory is not representative of the whole ensemble. Second, the BOMD production runs are set up by sampling from

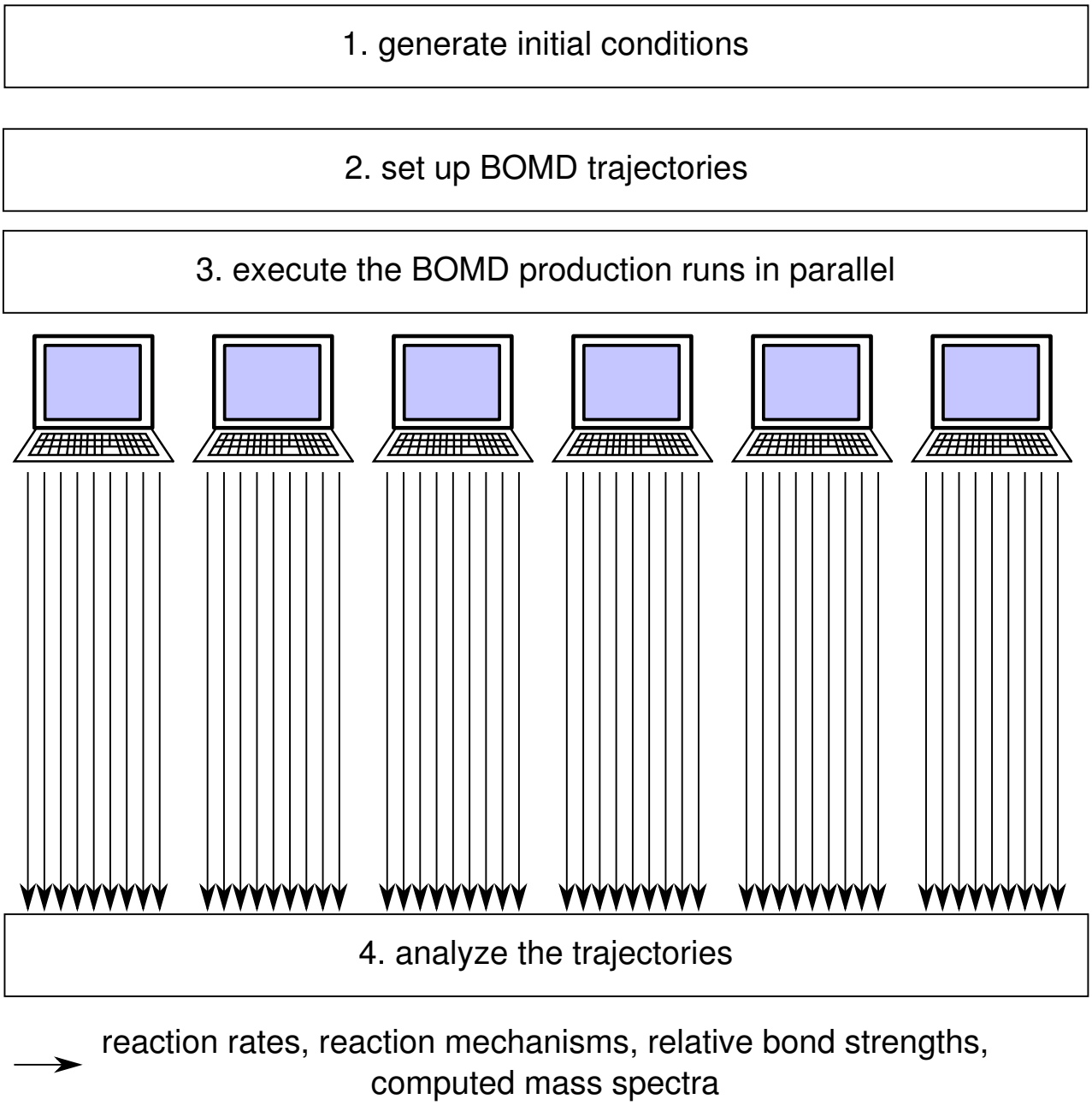

Figure 4.1.: The BOMD strategy to simulate unimolecular (fragmentation) reactions.

the initial conditions. Third, they are run in parallel for a simulation time on the order of picoseconds. The perfect parallelization of this strategy is crucial for the technical feasibility of BOMD calculations, as the calculations of the individual BOMD trajectories may become expensive depending on the underlying quantum chemical method. Fourth, the results from all production runs are gathered and analyzed. Our own contribution to the field, which is described in more detail below, uses the same recipe, which has been used generally for many purposes including EI mass spectra[10] and CID mass spectra[209].

Significant advances in the field of unimolecular reaction BOMD-simulations were made by the group of Hase using their VENUS[210] program. In 1994, a study on the dissociation of formaldehyde (H_2CO) was published, and the energy distributions in the fragment molecules were investigated.[211] In 2003, Hase and co-workers followed up on this subject and simulated the collision-induced dissociation of H_2CO^+.[209]. In 2004, González-Vázquez et al. used a combination of RRKM theory and BOMD (at the MP2/6-31G** level of theory) to describe the H_2 elimination from vinyl chloride.[212] In 2005, Martínez-Núñez et al. studied the CID of $Cr(CO)_6^+$ with Xenon using a precalculated PES.[213] Recent BOMD calculations of CID mass spectra of protonated biomolecules were performed by Spezia and co-workers.[214,215,216,217,218,219] These authors also included RRKM studies to compare to their BOMD results. BOMD simulations were also performed to gain a deeper understanding of dissociative electron attachment (DEA) processes, which may play a major role in biochemical systems.[220,221] Flosadóttir et al. analyzed the fragmentation reactions subsequent to DEA in amino acids[222] and nucleosides[223] using a DFT PES. Omarsson et al. reported the observation of molecular rearrangements through BOMD simulations for the radical anions of pentaflourinated aromatic compounds.[224]

The number of BOMD studies on unimolecular dissociations is far smaller than the number of RRKM calculations. However, BOMD has also been in use for decades, and valuable mechanistic insights into fragmentation and rearrangement reactions have been reported. BOMD simulations also have the power to capture unintuitive reaction pathways such as the roaming of dissociated H atoms[225], which cannot be discovered using stationary point optimizations on a molecular PES.* BOMD calculations can also describe partially dissociated ion-molecule complexes,[226] which are bound by weak molecular interactions, and give rise to prominent rearrangement peaks, e.g. in the radical cation of *n*-propyl phenyl ether.[227] However, these rearrangement peaks are often underrepresented in our scheme because they depend critically on the internal energy, and the competition of rearrangement and translational diffusion of such complexes is hard to simulate accurately.†

*A trajectory video showing such a process simulated by QCEIMS is found at www.thch.uni-bonn.de/tc/software/movies/h2_loss_in_octane.mpg

†A trajectory video showing a CH_3 rearrangement in methyl sulfonamide simulated by QCEIMS is found at www.thch.uni-bonn.de/tc/software/movies/me-sulfonamide-31.avi. In the computed EI mass spectrum, the corresponding peak is underrepresented, but clearly, such reaction mechanisms can by captured by

4.2. Results and Discussion

In this section, we present our own research regarding quantum chemical calculation of EI mass spectra. Our approach has been dubbed Quantum Chemical Electron Ionization Mass Spectrometry (QCEIMS).[10] First, we provide a short technical overview of QCEIMS. We describe the studies that we have conducted thus far, and put them in the perspective of the current literature. Second, we scrutinize the distribution of the positive charge upon fragmentation. This is the penultimate step in our mass spectral prediction routine, and a crucial part of our model. We choose three ethanol derivatives in a case study to show that statistical charges are a key to the accurate quantum chemical prediction of EI mass spectra. Third, we give an outlook on the potential of QCEIMS as a computational mass spectra library generator. We have compiled three small benchmark sets, and analyze the ranking of the computed mass spectra against the experimental ones by a common mass spectral matching score. The results enable us to make conclusions about the accuracy and potential of QCEIMS in the MS informatics context.

4.2.1. Overview of the QCEIMS Method

In Figure 4.2 we present a graphical work-flow of the QCEIMS program. It combines elements of statistical theory with MD. The input for QCEIMS is just a good guess of the molecular structure (Cartesian coordinates). The prediction of an EI mass spectrum by QCEIMS proceeds in four steps:[10]

1. Generation of an ensemble of molecular geometries and nuclear velocities by a ground state MD trajectory for a desired number of (reactive) production runs.

2. Assignment of an IEE for each production run by a Poisson distribution and generation of nuclear coordinates and velocities for the ion state from the ground state MD.

3. MD production runs on the ion state on a QC PES generated on the fly for a chosen maximum simulation time.

4. Generation of the spectrum by counting the fragment ions from the production runs. Isotopic distributions are taken into account after the BOMD procedure.

Each MD production run is carried out independently from all the others, enabling a perfectly parallel distribution of computational tasks. Importantly, a production run may consist of cascading trajectories where the fragment of the highest statistical charge (*vide infra*) is followed in a subsequent trajectory, and the other fragments are directly counted with their (lower) statistical charge. An example is depicted in Figure 4.2 where in the N^{th} production

BOMD.

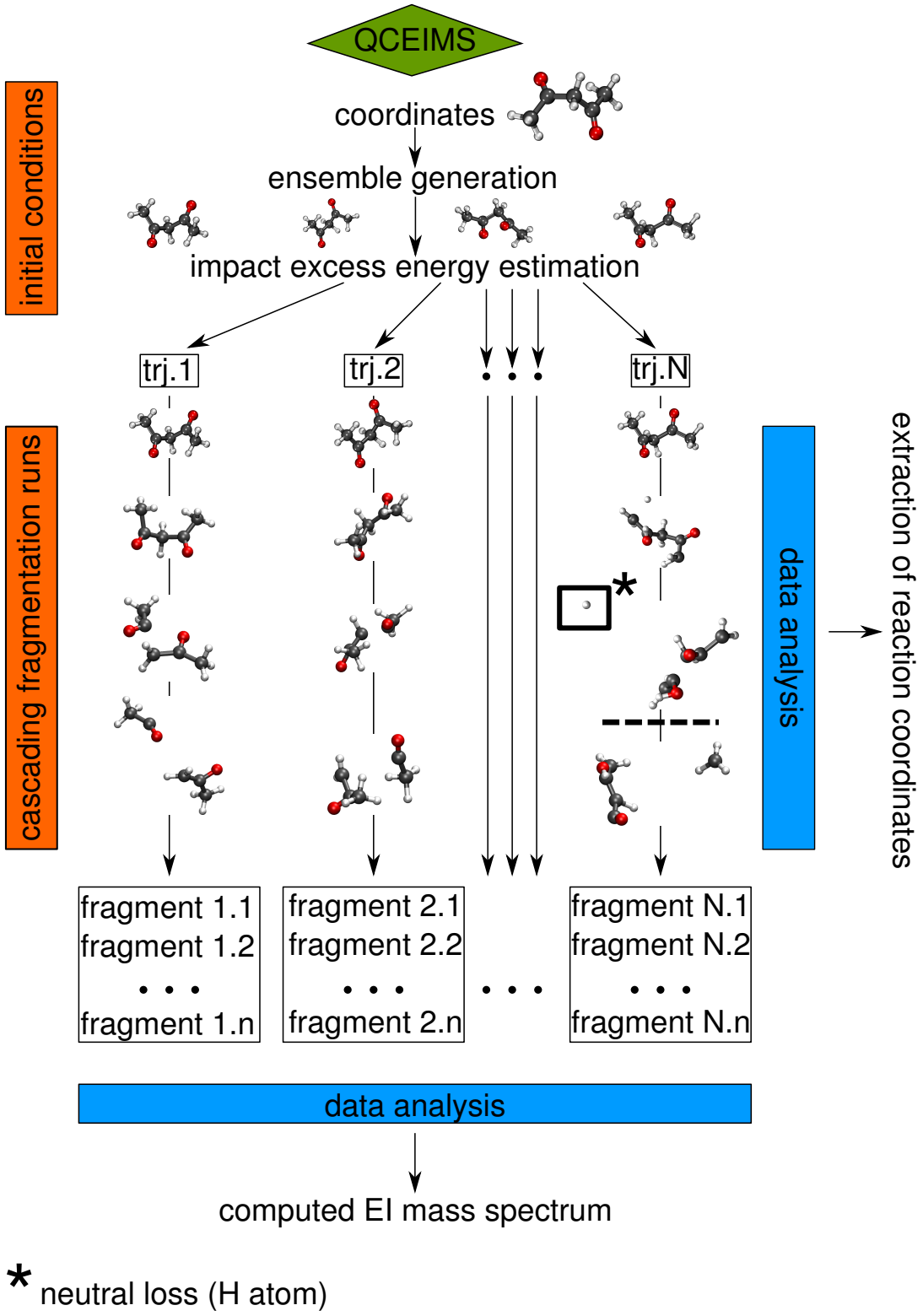

Figure 4.2.: Overview of the QCEIMS work-flow

run, a hydrogen atom dissociates in the first trajectory of the cascade. This neutral loss is recorded in the output. The second trajectory in the cascade with the surviving ion is then started, indicated by a dashed line in Figure 4.2. This procedure adds to the efficiency and robustness of QCEIMS because it (i) reduces the computational cost of each trajectory in the cascade by discarding neutral fragments (ii) stabilizes the electronic structure calculations because large inter-fragment distances may lead to severe self-consistent field (SCF) convergence problems. At the end of all production runs, the generated fragments are counted with their statistical charge, which after renormalization with respect to the base peak gives the calculated EI mass spectrum. In runs, where no fragmentation is detected, a count for the molecular ion $\mathbf{M}^{\bullet+}$ is registered. Since we do not want to duplicate the discussion of the technical details such as the implications of the maximum simulation time, the time step, the chosen IEE distribution, the initial temperature, the electronic temperature, etc. we refer the reader to the supporting information of the original publication,[10] where one of the authors (SG) has provided ample information. So far, QCEIMS has shown promising results using three different quantum chemical methods: the semi-empirical OM2[228]-D3[33,34,35] and DFTB3[136,229,230]-D3 methods, and the PBE0[47,117]-D3-gCP[231]/SVx[232] (DFT-D3) PES. It should be noted that one typical QCEIMS production run for a mid-sized organic molecule (50 atoms) takes minutes to hours for the OM2-D3 and DFTB3-D3 methods and days to weeks when using DFT-D3. We have been able to show the versatility of QCEIMS by calculating the EI mass spectra of large organic drug molecules[233], and the nucleobases adenine[234] and uracil, thymine, cytosine, and guanine[235]. In the case of the organic drug molecules, we have noted the following limitations of QCEIMS: (i) only semi-empirical QC calculations are feasible for such large molecules in a reasonable amount of time, and these methods may introduce a significant error resulting in artifacts or missing peaks in the computed EI mass spectra, and (ii) the IEE distribution in its current state is effectively an empirical tool by which the time scale and the likelihood of reactions is greatly affected.[233] On the other hand, the same study has shown that QCEIMS can be applied to organic molecules consisting of 100 atoms. There are currently no other methods which can predict the EI mass spectra of such drug molecules. The comparison of OM2-D3, DFTB3-D3, and DFT-D3 results for adenine shows that the calculated EI mass spectrum at the higher DFT-D3 theoretical level has resembles the experimental mass spectrum more closely than the other two[234]. The analysis of the DFT-D3 production runs for adenine leads to the important conclusion that the ion m/z 108 is formed by the loss of H-C2-N1, which is in agreement with a previous experimental study.[236] For cytosine and guanine, we have been able to show that their experimental mass spectra are best explained when taking into account their populated tautomers.[235]

4.2.2. Statistical Charges – The Right Model for the Prediction of Relative Intensities

An important step in the calculation of EI mass spectra is the assignment of charges to the fragments. This is related to Stevenson's Rule[15,237,238] which is defined in the IUPAC recommendations relating to mass spectrometry as "the rule stating that in competing fragmentations the product ion formed from its neutral species counterpart with the lower ionization energy will usually be the more abundant."[150] A quantitative version and in fact extension and generalization of this rule has been implemented in QCEIMS. At the end of each trajectory when QCEIMS has detected a fragmentation (based on geometry and standard covalent bond distances), the following procedure is initiated:

1. Assignment of each atom to a fragment.
2. Computation of average fragment geometries from the last 50 MD time steps.
3. ΔSCF computation of the fragment ionization potentials (IPs) at these geometries.
4. Computation of the statistical charge using Boltzmann factors with the IPs and the actual (average) internal temperature.
5. Addition of the statistical charges to the calculated fragment ion (m/z) counts.

The statistical charge q_i of each fragment evaluates to the following expression

$$q_i = \frac{\mathrm{e}^{-\frac{IP_i}{\mathrm{k_B}T}}}{\sum_{j=1}^{n} \mathrm{e}^{-\frac{IP_j}{\mathrm{k_B}T}}}, \tag{4.2}$$

where $\mathrm{e}^{-\frac{IP_i}{\mathrm{k_B}T}}$ is the Boltzmann factor for the ionization potential of fragment i, $\mathrm{k_B}$ is the Boltzmann constant, T is the actual temperature, and the denominator is the partition function regarding the IPs of all fragments generated in one trajectory. The temperature is dependent on the kinetic energy E_{kin} and the number of atoms N according to

$$T = \frac{E_{\mathrm{kin}}}{\mathrm{k_B} * 3N}. \tag{4.3}$$

The kinetic energy of the nuclei is evaluated classically, and is governed by the initial temperature (set to 500 K by default in QCEIMS) and the IEE (see above). Typical internal temperatures at dissociation events are 2000-4000 K. The predictive power of this statistical charge distribution model can be seen from a small homologue series. The ethanol derivatives 2-amino-ethanol, 2-mercapto-ethanol and 2-chloro-ethanol undergo the dissociation of the C–C bond upon ionization, yielding two principal fragments, as summarized in Table 4.1.

The experimental adiabatic IPs of these fragments are known[239,240,241,242], and the Boltzmann population of the ions is readily evaluated. We choose T to be 2820 K, which corresponds to an IEE of 0.6 eV per atom (5.4 eV to 6.6 eV for the three homologue molecules including an initial temperature of 500 K before ionization).

Table 4.1.: Overview of ethanol derivatives, and the adiabatic IPs (in eV) of principal fragments. BPR= Boltzmann population ratio at T =2820 K

R1–R2	IP(R1)	IP(R2)	BPR($R1^+$)	BPR($R2^+$)
$HOCH_2-CH_2NH_2$	7.56[241]	6.20[239]	0.00	1.00
$HOCH_2-CH_2SH$	7.56[241]	7.54[242]	0.47	0.53
$HOCH_2-CH_2Cl$	7.56[241]	8.75[240]	1.00	0.00

For 2-amino ethanol, we thus expect to observe the $CH_2NH_2^+$ fragment ion (m/z 30). For 2-mercapto ethanol, both CH_2OH^+ (m/z 31) and CH_2SH^+ (m/z 47 and 49) are expected, and for 2-chloro ethanol exclusively CH_2OH^+ should be observed. Figure 4.3 shows the QCEIMS calculated EI mass spectra of these compounds, and reveals that these expectations are indeed met. For $HOCH_2-CH_2SH$ both signals are found in the simulation although the fragment ion CH_2SH^+ (m/z 47 and 49) is more abundant in the experimental spectrum of 2-mercapto ethanol (Figure 4.3 b) than the ion CH_2OH^+ (m/z 31), whereas the QCEIMS result is apparently qualitatively wrong. This is related not primarily to inaccurate IPs, but rather to error in the PES and initial conditions leading to too much fragmentation and overestimation of side reaction channels. The calculated spectra of 2-amino ethanol and 2-chloro ethanol compare quite well to the experiment.[243,244] The missing molecular ion for 2-amino ethanol is probably related to a combination of PES inaccuracies (i.e., a too low dissociation barrier for the C–C bond in this case) and too high internal energy.

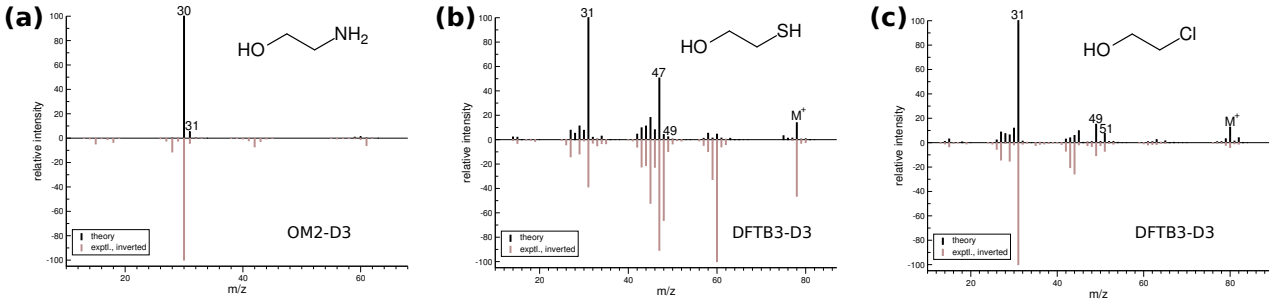

Figure 4.3.: Calculated EI mass spectra of ethanol derivatives. The default settings of QCEIMS[10] and the indicated QC methods were used.

Summarizing these observations we conclude that the statistical distribution of the charge to fragments based on their ground state IPs is a very reasonable model. It allows for an estimate

of the order of magnitude of the IEE and therefore the temperature of the ions. The reasonable agreement between theory and experiment in such cases where the charge distribution is crucial (see Ref.[10] for another example) indirectly supports the IEE distribution we use through the temperatures that appear in the Boltzmann factor. So far, we have not observed a significant contribution from low lying excited states of the fragments to the partition function in equation 2 as indicated e.g. by incorrect charge assignments. Such effects may become non-negligible in compounds with many states in a narrow energetic window, e.g. transition metal complexes. The assignment of the right statistical charges to fragment ions is included in any QET/RRKM approach by construction, so long as one evaluates the full reaction space including all processes where two fragments compete for the charge (the difference of the adiabatic IPs is just $\Delta\Delta E$ for these).

4.2.3. A Small Performance Test

We have implemented a mass spectral matching score[9,234] to quantify the similarity between the computed and experimental spectra. The score ranges from zero (no similarity at all) to 1000 (identical spectra). The introduction of this measure has enabled us to rate the quality of our predictions and conduct assessments more relevant to the MS community. We report the results of a small QCEIMS benchmark in the following. If one were in possession of an EI mass spectrum of an unknown species and looking to assign it correctly by QCEIMS, one would proceed by computations for several structural candidates. The three sets below have been designed to test whether QCEIMS could deliver useful results in the identification of unknown molecular structures.

The molecules **1-12** are shown in Figure 4.4. **1-4** are C_6H_{12} isomers ($\mathbf{M}^+$ at m/z 84, cyclohexane [**1**], methyl-cyclopentane [**2**], ethyl-cyclobutane [**3**], 1-hexene [**4**]). They all have one double bond equivalent, either in the form of a double bond (**4**) or in the form of a ring (**1-3**). **5-8** are $C_4H_{10}O$ isomers ($\mathbf{M}^+$ at m/z 74, 1-butanol [**5**], 2-butanol [**6**], isobutyl alcohol [**7**], and tert-butyl alcohol [**8**]). They are all alcohols with different alkyl chains. **9-12** are C_4H_7NO isomers ($\mathbf{M}^+$ at m/z 85, 2-pyrrolidinone [**9**], methyl-acrylamide [**10**], acetone-cyanohydrine [**11**], 1-isocyanato-propane [**12**]). **9-12** differ in their functional groups, **9** is a lactim, **10** an open-chained amide, **11** a cyanohydrine, and **12** an alkyl isocyanate. The computed mass spectra in Figures 4.5-4.7 below are the result of 300 production runs for each molecule at the OM2-D3 level of theory, employing all the default settings of QCEIMS[10].

First, we turn to the results for the C_6H_{12} isomers. Figure 4.5 shows that all four molecules exhibit very similar EI mass spectra, indicating that the fragmentation pathways of **1-4** are also very similar. The peak series m/z 84 ($\mathbf{M}^+$), 69 ($[\mathbf{M}\text{-}CH_3]^+$), 56 ($C_4H_8{}^+$), and 41 ($C_3H_5{}^+$) appears in each spectrum with varying relative intensities. QCEIMS gives the wrong relative

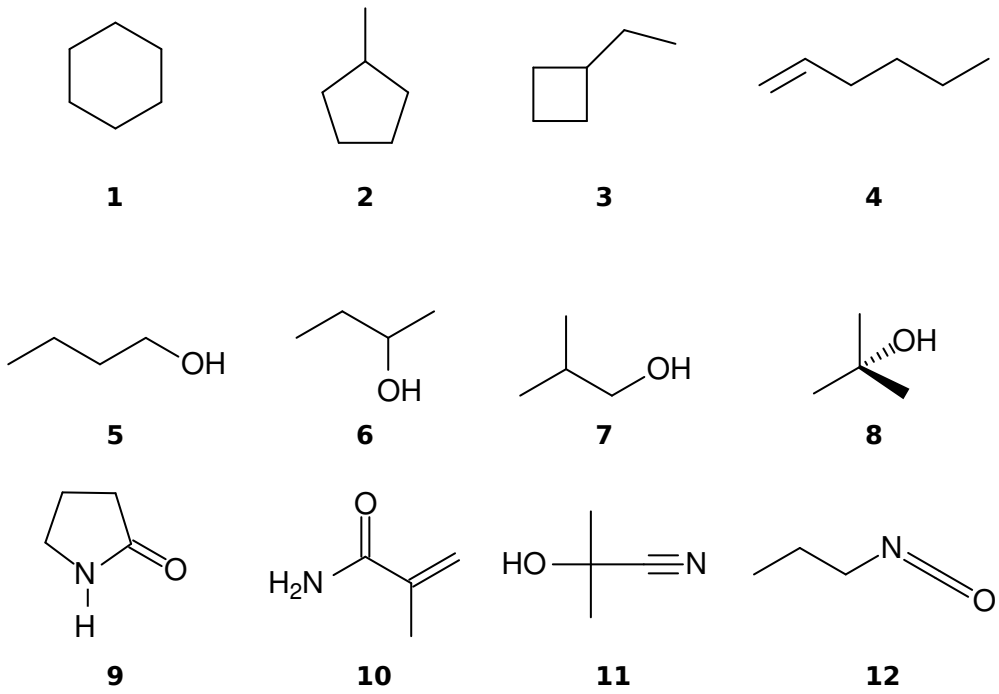

Figure 4.4.: C_6H_{12}, $C_4H_{10}O$, and C_4H_7NO isomers chosen for the performance test.

intensities for this peak series. However, most experimental signals are accounted for in the simulated EI mass spectra in Figure 4.5, with the best agreement for **3**.

We discuss the usefulness of the results for **1-4** by interpreting the matching score matrix for **1-4**, see Table 4.2. The diagonal elements are the scores for the computed spectrum versus its experimental counterpart, the off-diagonal elements correspond to the scores of the computed spectrum versus the other molecules. Therefore, the assignment of a computed spectrum to a compound is only correct when the diagonal element in each row has the highest value. Table 4.2 reveals that for compounds with very similar EI mass spectra, QCEIMS cannot always provide the right assignments, although the quality of the computed mass spectra is adequate (we consider the diagonal element scores between 539 and 717 satisfactory). This failure is expected, and it shows the intricacy of the problems in MS informatics and mass spectral prediction.

The computed EI mass spectra of the alcohols **5-8** in Figure 4.6 may seem somewhat disappointing at first because QCEIMS misses some of the major peaks. The base peak in the spectrum of **5** is m/z 56 ($[\mathbf{M}\text{-}H_2O]^+$). The corresponding fragmentation pathway is underrepresented in the QCEIMS simulations of $\mathbf{5}^{\bullet+}$. In addition, some of the fragment ions with lower mass are missing in the QCEIMS spectra of **5-8**. We note, however, that the ions m/z 31 (CH_2OH^+) and m/z 43 ($C_3H_7{}^+$) are correctly predicted for **5** and **7**. The base peaks of **6** (m/z 45, $C_2H_5O^+$) and **8** (m/z 59, $[\mathbf{M}\text{-}CH_3]^+$) are also reproduced by the QCEIMS simulations. Clearly, the molecules **5-8** have easily distinguishable fragmentation reactions that lead to different peaks in their EI mass spectra. Even the spectra of the two primary

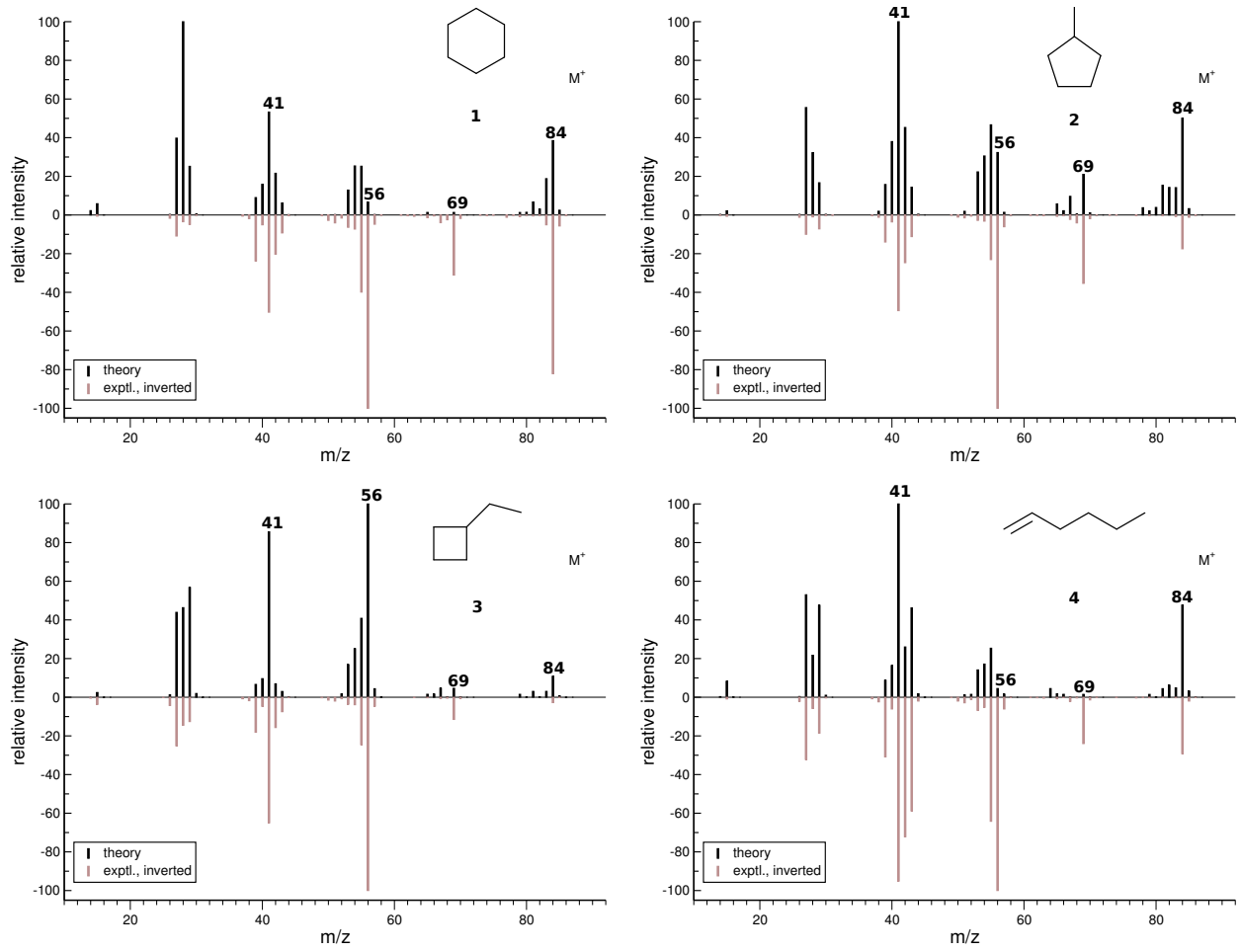

Figure 4.5.: Computed vs. experimental EI mass spectra of **1-4**. The marked peaks are discussed in the text.

alcohols **5** and **7** differ significantly.

Therefore, the score matrix in Table 4.3 shows that the assignment of the species **5-8** by QCEIMS is correct for **5**, **6**, and **8** because none of the elements in those rows have a higher value than the diagonal element. The computed spectra of **5** and **7** are tied for the highest score against the experimental spectrum of **7**, which leads to an ambiguous assignment.

The comparison of the computed EI mass spectra for **9-12** (see Figure 4.7) to the experimental spectra reveals that QCEIMS is able to predict some of the main peaks correctly, while other peaks are missing. In the case of **9**, the relative intensity of the $\mathbf{M}^{\bullet+}$ peak is too small compared to experiment. The ion m/z 56 stems from the loss of $H_2C{=}NH$ according to the analysis of the trajectories, and the ion m/z 42 is identified as the propenyl radical cation $C_3H_6{}^{\bullet+}$, which is formed by the excision of HNCO from $\mathbf{9}^{\bullet+}$. Concerning **10**, we find that QCEIMS also underestimates the stability of $\mathbf{M}^{\bullet+}$. The ion m/z 44 is the product of α cleavage, H_2NCO^+, the neutral loss being the allyl radical, $C_3H_5{}^{\bullet}$. The ion m/z 41 is the allyl cation $C_3H_5{}^+$ that arises by the losses of $NH_2{}^{\bullet}$ (nominal mass of 16 u) and CO (28 u).

Table 4.2.: Score matrix of computed vs. experimental spectra. The diagonal (gray) cells correspond to the correct assignments. The red cells correspond to the wrong assignments, i.e the scores are higher than for the diagonal elements.

	Computed spectra			
exptl. spectra	**1**	**2**	**3**	**4**
1	539	606	619	569
2	404	558	607	445
3	351	549	717	364
4	443	580	618	543

Table 4.3.: Score matrix of computed vs. experimental spectra. The diagonal (gray) cells correspond to the correct assignments. The red cells correspond to the wrong assignments, i.e., the scores are higher than for the diagonal elements.

	Computed spectra			
exptl. spectra	**5**	**6**	**7**	**8**
5	516	128	368	89
6	321	612	275	291
7	483	208	482	123
8	182	236	270	481

This result further underlines the importance of the fragment IPs, which are vastly different for the fragments listed above, and therefore determine the computed relative peak intensities of m/z 41, 28 and 16 in QCEIMS. The interpretation of the computed EI mass spectrum of **11** is easy, as there is only one main peak, m/z 70, which results from methyl radical loss from $\mathbf{11}^{\bullet+}$. This is also the experimentally obtained base peak. The other fragment ions are not reproduced by QCEIMS using the OM2-D3 PES. The computed fragmentation pathways of $\mathbf{12}^{\bullet+}$ show that the two main fragment ions, m/z 56, and m/z 29 are $H_2C{=}N{=}C{=}O^+$, and $C_2H_5{}^+$, respectively. These two fragments have very similar computed IPs, and therefore similar relative intensities. QCEIMS predicts mainly the formation of $C_2H_5{}^+$, which may be the result of a shortcoming regarding the computation of very similar IPs.

Notwithstanding all the deficits, errors, and shortcomings of QCEIMS regarding the prediction of the EI mass spectra of **9-12**, Table 4.4 demonstrates that the assignment of the computed EI mass spectrum to the experimental spectrum based on the matching score is fully correct, with differences in scores of at least 100 points to the next most likely candidate. We therefore argue that QCEIMS could be used to reduce the number of possible isomers in the search for structures of unknown compounds – provided their fragmentation pathways are sufficiently different. Databases of computed mass spectra could be generated for a specific set of isomers in order to assign the right chemical structure.

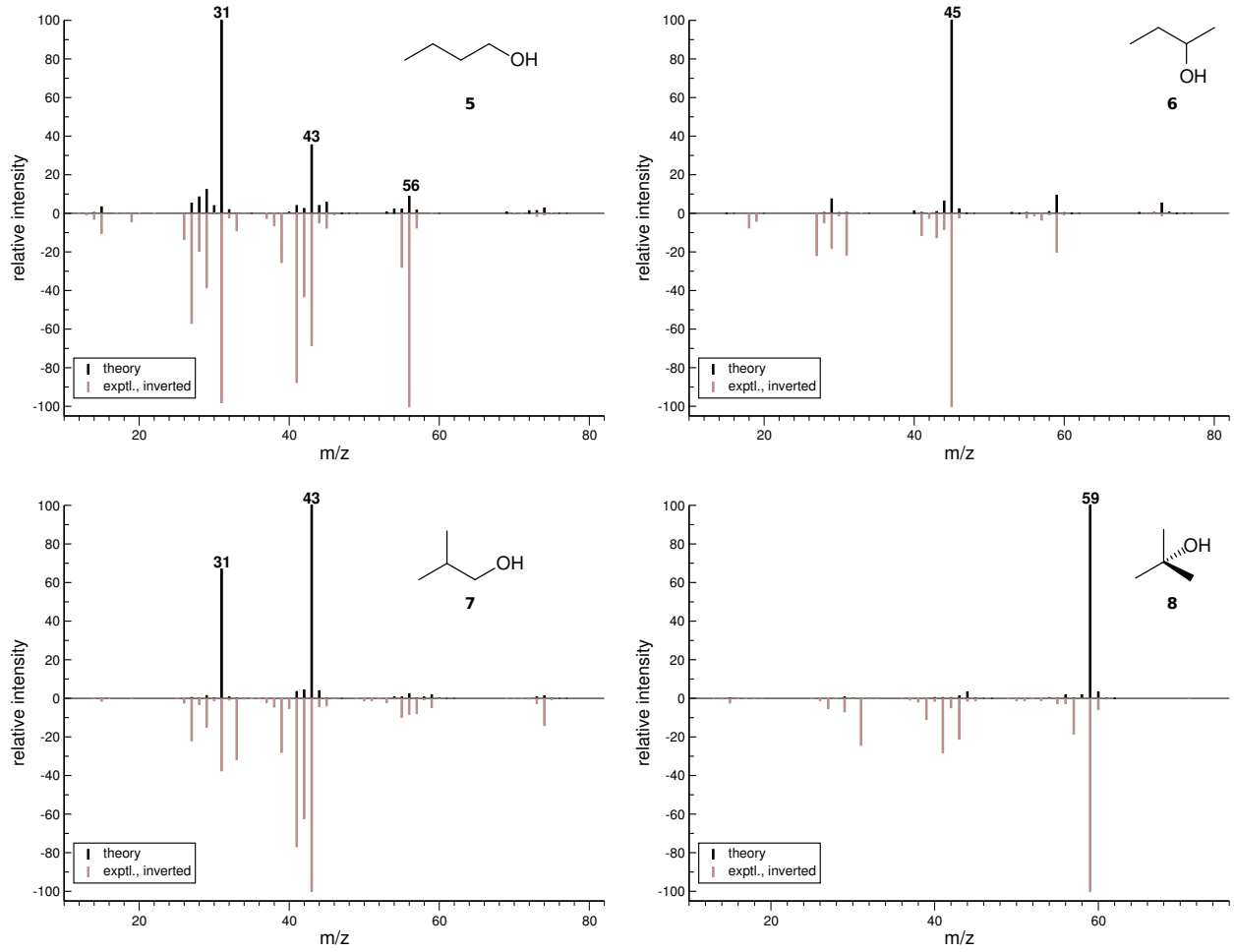

Figure 4.6.: Computed vs. experimental EI mass spectra of **5-8**. The marked peaks are discussed in the text.

Table 4.4.: Score matrix of computed vs. experimental spectra. The diagonal (gray) cells correspond to the correct assignments.

	Computed spectra			
exptl. spectra	**9**	**10**	**11**	**12**
9	644	441	90	463
10	505	629	123	304
11	128	134	600	179
12	442	192	90	563

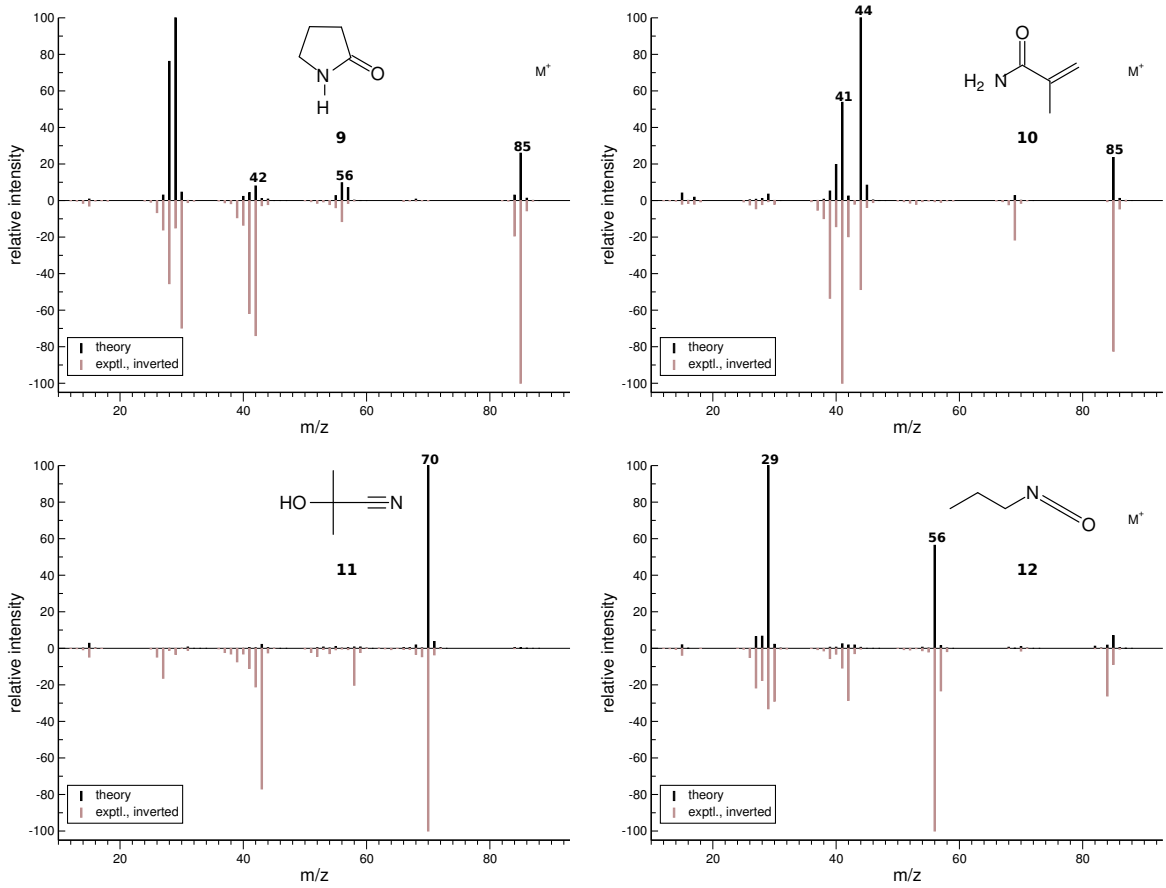

Figure 4.7.: Computed vs. experimental EI mass spectra of **9-12**. The marked peaks are discussed in the text.

4.3. Conclusions

We have surveyed some historical as well as the most recent endeavors to calculate EI mass spectra from first principles. On the one hand, one can use statistical (QET/RRKM) theory to predict branching ratios for competing fragmentation reactions. The drawback of QET/RRKM theory is the required knowledge (or at least an educated guess) of these pathways prior to the calculations. On the other hand, BOMD based methods may elucidate new fragmentation paths as they lack the bias of pre-defined reaction coordinates. The drawback of MD based methods is the requirement to run a large number of of long calculations, which may be moderated by parallel execution thereof. Since both methods have their indisputable advantages, QET/RRKM theory and BOMD based methods to calculate the fragmentation pathways of molecular ions will continue to coexist.

We have also presented our own BOMD based QCEIMS approach, which incorporates statistical theory by the assignment of a statistical charge to a fragment. It is based non-

empirically on the Boltzmann factors for the calculated ionization potentials and the average internal temperature. We have shown that the statistical charges model is consistent with experimental results. Using QCEIMS in combination with a semi-empirical PES, one can obtain computed EI mass spectra of organic compounds in a reasonable time if one has access to a moderately sized computing cluster. Finally, we propose that QCEIMS may be used in the future for purposes of structural assignments of unknown compounds by building a database of computed EI mass spectra, against which unknown compounds may be checked.

The `qceims` executable and additional tools for conducting QCEIMS calculations are available from the corresponding author by request. `qceims` presently has interfaces for the `MNDO99`,[245] `DFTB+`,[246] `ORCA`,[70,71] and `TURBOMOLE`[62,63] programs, which are used to generate the quantum chemical PES on the fly. Development of `qceims` is ongoing in our laboratory to extend it to other mass spectral methods. The development of a special-purpose semi-empirical method specifically parametrized for fragmentation reactions is a long-term goal. Several video examples (reactive trajectories) for educational purposes can be downloaded from the website of the Mulliken Center for Theoretical Chemistry, http://www.thch.uni-bonn.de/tc/downloads/movies/

5. First Principles Calculation of Electron Ionization Mass Spectra for Selected Organic Drug Molecules

Christoph Alexander Bauer and Stefan Grimme*
Received 5th of August 2014, Published online 8th of September 2014

Reproduced (adapted) by permission of The Royal Society of Chemistry
Christoph Alexander Bauer and Stefan Grimme Org. Biomol. Chem. **2014**, *12*, 8737–8744.
 DOI 10.1039/C4OB01668H

Own manuscript contribution

- Performance of all calculations.
- Interpretation of data
- Writing the manuscript

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

Abstract This study presents a showcase for the novel Quantum Chemistry Electron Ionization Mass Spectrometry (QCEIMS) method on five FDA-approved drugs. The method allows a first-principles electronic structure-based prediction of EI mass spectra in principle for any molecule. The systems in this case study are organic substances of nominal masses between 404 and 853 atomic mass units and cover a wide range of functional groups and organic molecular structure motifs. The results demonstrate the widespread applicability of the QCEIMS method for the unbiased computation of EI mass spectra even for larger molecules. Its strengths compared to standard (static) or data base driven approaches in such cases are highlighted. Weak points regarding the required computation times or the approximate character of the employed QC methods are also discussed. We propose QCEIMS as a viable and robust way of predicting EI mass spectra for sizeable organic molecules relevant to medicinal and pharmaceutical chemistry.

5.1. Introduction

Modern quantum chemistry (QC) methods have made it possible to routinely compute and predict spectral properties of reasonably sized chemical compounds.[247] Today, excitation energies (related to UV-Vis spectra), vibrational frequencies (IR and Raman spectra) and nuclear magnetic resonance chemical shifts (NMR spectra) of many organic substances can even be calculated on low-cost computers with appropriate, mostly density functional theory (DFT) based, methods. While not resulting from electromagnetic radiation-induced transitions, mass spectrometry (MS), especially electron ionization mass spectrometry (EI-MS), is an extremely important analytic method in organic chemistry.[1,2] Thus far, the prediction of EI mass spectra without relying on existing spectral databases or pre-tabulated fragmentation rules has been based on Quasi Equilibrium Theory (QET)[14] or Rice-Ramsperger-Kassel-Marcus (RRKM) theory.[24,25] However, even the most sophisticated attempts within these frameworks[248] have found no application on a regular basis.

The Quantum Chemistry Electron Ionization Mass Spectrometry (QCEIMS) method[10] has recently been presented as an attempt to fill this gap in theoretical spectra prediction. It is to our knowledge the first comprehensive attempt based on Born-Oppenheimer *ab initio* Molecular Dynamics (BO-AIMD)[13] to compute the fragmentation patterns that arise by bombarding molecules with electrons in the gas phase. The approach is 'brute force' in the sense that the EI-MS experiment is represented as closely as possible in a theoretical and computational model. Figure 5.1 shows a schematic summary of the QCEIMS procedure (the number of trajectories X is typically on the order of 10^2-10^3). The result depicted in this figure is purely illustrative as the quality of our simulations regarding small organic molecules has already been discussed in the original paper.[10]

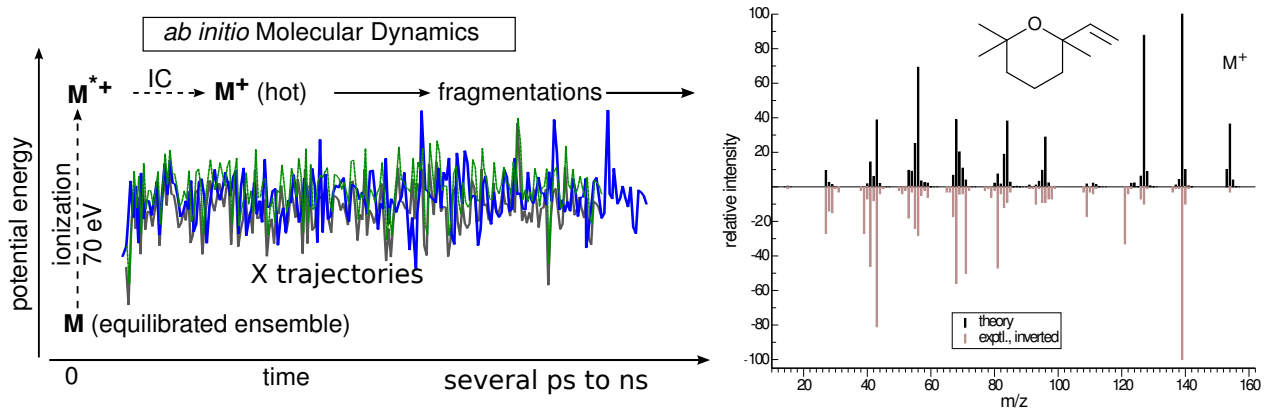

Figure 5.1.: Schematic description of the QCEIMS procedure together with an illustrative result for a small organic molecule.

In the present study, we have taken our novel method to its current limits in order to assess whether it is workable for medicinally and pharmaceutically relevant organic compounds thereby becoming a new routine tool in theoretical chemistry. To this end five organic drugs (for structures, see Figure 5.2) have been selected as realistic examples to undergo the QCEIMS procedure. These are valsartan (**1**), erythromycin (**2**), taxol (**3**, also known as paclitaxel), lovastatin (**4**) and simvastatin (**5**). Systematic names (IUPAC nomenclature) of these molecules are supplied in the Electronic Supplementary Information (ESI). In the following, we briefly introduce the five chosen substances.

Valsartan ($C_{24}H_{29}N_5O_3$, nominal mass 435 u, **1**) is an angiotensin II receptor antagonist indicated e.g., against hypertension.[249] It is considered as the 1-*H*-tetrazol tautomer as found in the mass spectral database.[244] Erythromycin ($C_{37}H_{67}NO_{13}$, nominal mass 733 u, **2**)[250] is a macrolide antibiotic[251] with a 14-membered macro-cycle and the two sugar moieties cladinose and desosamine (an amino sugar). Taxol ($C_{47}H_{51}NO_{14}$, nominal mass 853 u, **3**) is an anticancer agent, which can be isolated from the pacific yew tree.[252] It is active against various cancer cell types and its mechanism of action is based on the promotion of microtubule assembly within the cancer cells.[253] Lovastatin ($C_{24}H_{36}O_5$, nominal mass 404 u, **4**) and simvastatin $C_{25}H_{38}O_5$, nominal mass 418 u, **5**) belong to a highly profitable class of cholesterol-lowering drugs, the statins.[254] Note that not all statins are as structurally similar to each other as **4** and **5**.

As can be seen from the structures in Figure 5.2, the five drugs have been selected to cover a wide range of functional groups, from tetrazol moieties over (bridged) macro-cycles to open-chained as well as cyclic esters and amides.* Therefore, their experimental EI mass spectra (with the exception of the intentionally chosen nearly homologous statins) reflect a multitude of different decomposition pathways. In essence, it is this complex reactive labyrinth against which QCEIMS was tested. The systems where picked more or less randomly from intensive

*A secondary motivation was the availability of experimental EI mass spectral data.

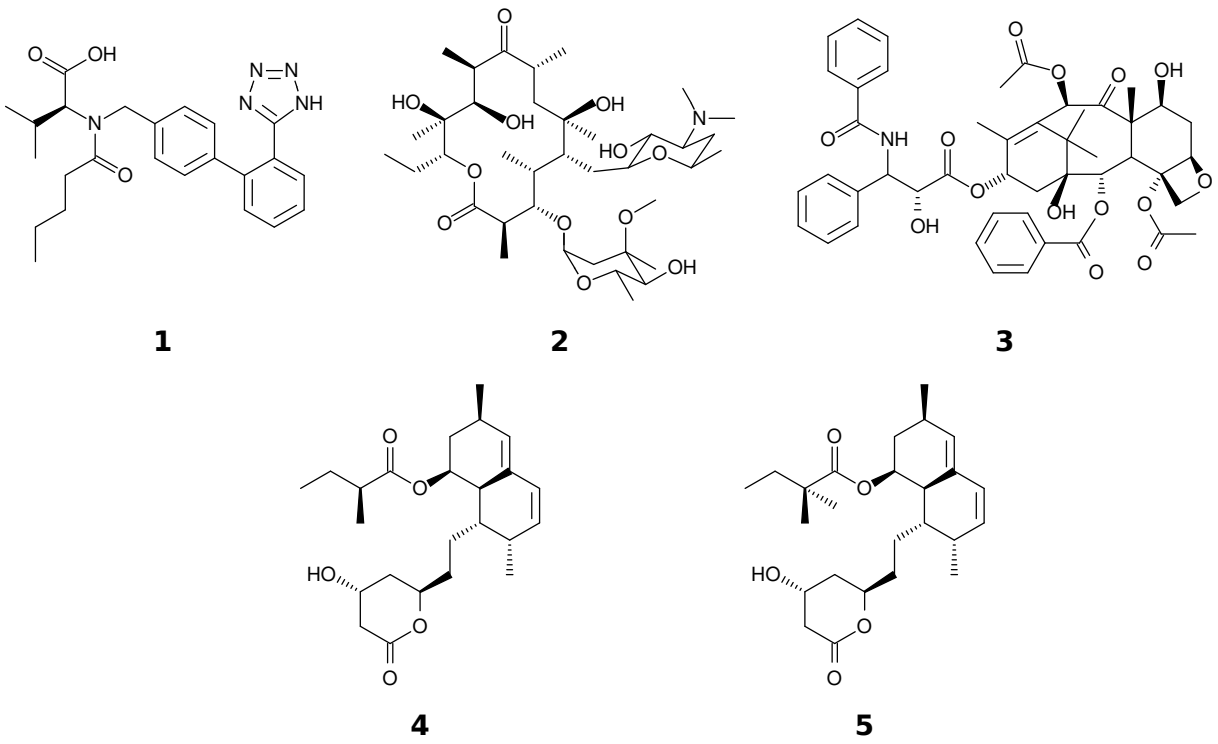

Figure 5.2.: Formulas of the drug molecules chosen as examples for QCEIMS.

literature searches without any presumptions except that of a reasonable molecule size in order to keep the computational resources within our limits. The results are discussed in the following section. The details of the QCEIMS procedure, which only requires a ground state molecular structure and the impact energy (usually 70 eV) as input are described in every detail in the ESI of the original publication[10] and hence not repeated here.

5.2. Results and Discussion

Comparisons between simulated and experimentally obtained EI mass spectra for compounds **1-5** are shown below. Since it would take too much space to exhaustively cover every detail of each spectrum, the discussion is limited to the main peaks and to a selected few other illustrative fragmentations, bearing in mind that the purpose of this article is to highlight strengths and weaknesses of our method and to provide an overview.

By carrying out a large number of BO-AIMD fragmentation runs for each compound the QCEIMS procedure directly mimics the experiment. In this study, 1,000 automatically randomized molecular geometries for each structure were instantly ionized and allowed to de-

compose. Because of the use of unbiased, on-the-fly computed potential energy surfaces, this automatically involves various processes such as simple homolytic and heterolytic bond cleavages, multiple (complex) fragmentations, and unimolecular rearrangements, according to the QC propagation method during the specified simulation time. Every run was completely independent of all other runs, as every molecule in the gas phase is independent of all other molecules and fragmentations detected in mass spectrometry are in essence unimolecular gas phase reactions. The base peak in the computed spectra typically translates to a few hundred counts of one main fragment ion, whereas in the experiment, many more counts are registered. However, as can be seen from the computed spectra below, the important relative number of counts in the experimental spectra is reflected astonishingly well by our simulations. Convergence for a QCEIMS spectrum is reached when its overall shape does not change significantly by adding more fragmentation runs. A summarizing discussion of the results presented below is given in section 5.2.5.

5.2.1. Valsartan

Choosing a system such as valsartan with its electronically relatively complicated, disintegration-prone tetrazol moiety could be considered as a daring choice. Yet, the direct comparison of experimental and QCEIMS spectra in Figure 5.3 reveals surprisingly good results.

While admittedly the molecular ion, which survives some of the fragmentation runs in our simulations, should not give any significant signal and does come out as small but false positive from our calculations, a significant number of peaks have been predicted correctly. Especially the experimental base peak at m/z 178 is almost reproduced by QCEIMS. A few of the $C_{14}H_{10}^{+}$ isomers responsible for the m/z 178 peak are schematically depicted in Figure 5.4. It is important to note that several different fragment isomers contribute to the same peak. This mechanistic information is difficult to obtain experimentally for larger systems. The fragmentation mechanism leading to the $C_{14}H_{10}^{+}$ fragments involves the splitting of the C–N tertiary amine bond and decomposition of the tetrazol ring to two N_2, which may happen in any order. In our simulations, it took four to six consecutive fragmentation runs (a cascade) to arrive at these m/z 178 structures. The QCEIMS code automatically takes this into account by further propagation of hot daughter ions until their internal energy has decreased below the dissociation threshold.

Additionally, several other peaks of valsartan were assigned based on an analysis of the fragmentation trajectories. Figure 5.5 depicts the m/z range of the computed and experimental spectra of **1** from 160 to 210. The structures assigned to peaks at m/z 165 ($C_{13}H_9^{+}$), 179 ($C_{14}H_{11}^{+}$), and 192 ($C_{14}H_{10}N^{+}$) are clearly related to each other; they all appear after splitting the tertiary amine bond and decomposition of the tetrazol ring. In $C_{14}H_{10}N^{+}$, one nitrogen atom of the tetrazol moiety is still left forming a benzonitrile moiety. The geometries shown

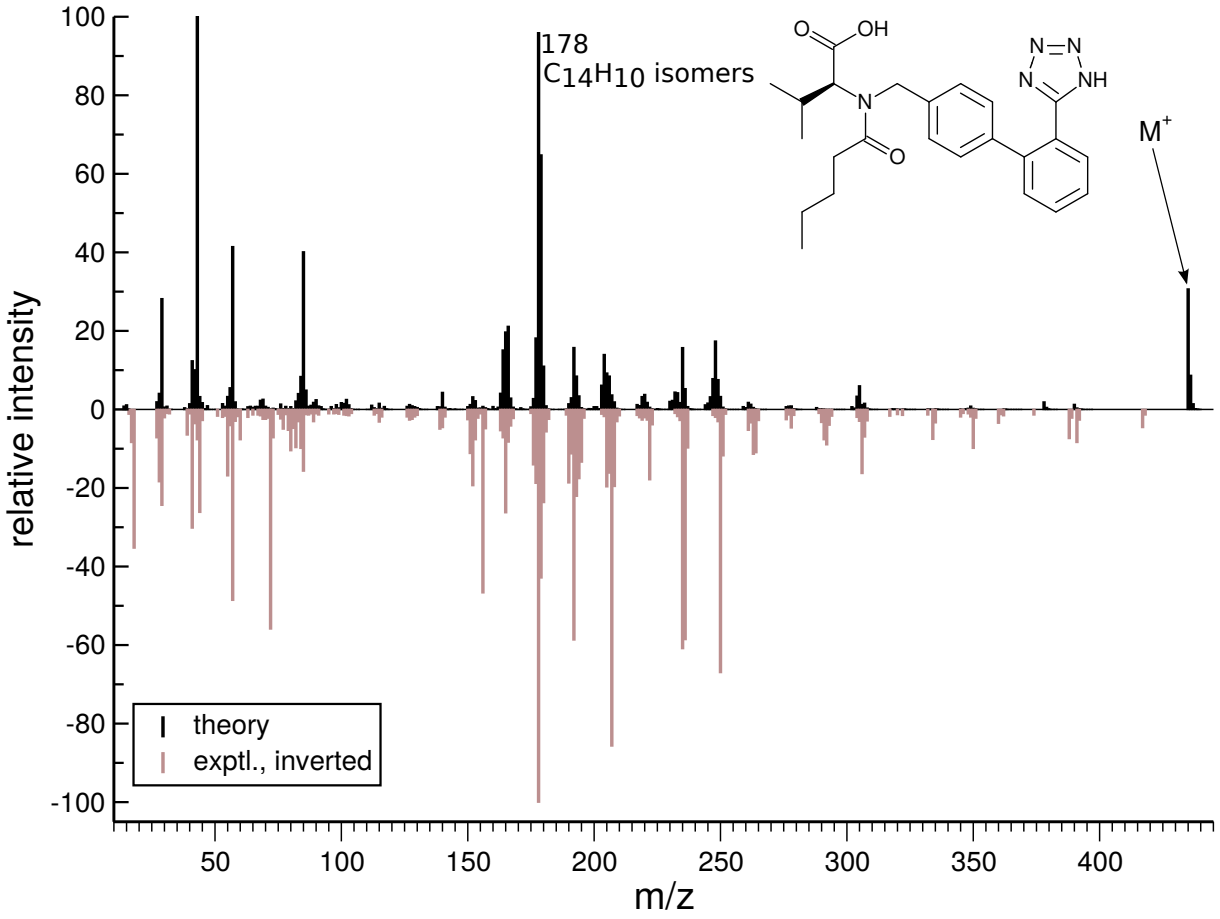

Figure 5.3.: Calculated mass spectrum of valsartan in comparison with the experimental spectrum. The indicated m/z value is discussed in the text among others.

in Figure 5.5 were taken directly from the QCEIMS output. For the ions at m/z 178 and m/z 179, two constitutional isomers out of many, one cyclic, one partly open-chained, are shown as examples. The peak at m/z 207 is not easily assignable due to the low count of ions in the QCEIMS simulations, which infers unreliable statistics.

Predicting almost correctly the main fragments $C_{14}H_{10}^{+}$ by our method is a very positive result, which means that an important dissociative pathway including the corresponding reaction barrier heights have been modeled accurately. The decomposition of the tetrazol ring in multiple reaction steps is certainly also nontrivial. Moreover, special intramolecular rearrangement reactions such as the formation of new 5- and 7-membered rings as shown by the schematic drawings in Figure 5.4 are also taken into account. Lastly, the ability of the highly conjugated $C_{14}H_{10}^{+}$ fragment to retain the charge (in this case as a radical cation), which in organic chemistry is often ascribed to resonance, is reflected well, too.* Note that QCEIMS

*A note has to be given on the electronic structures of fragments **1a-1c**. These are naturally very complicated and most certainly not accurately described by either the drawings in Figure 5.4 or the computational

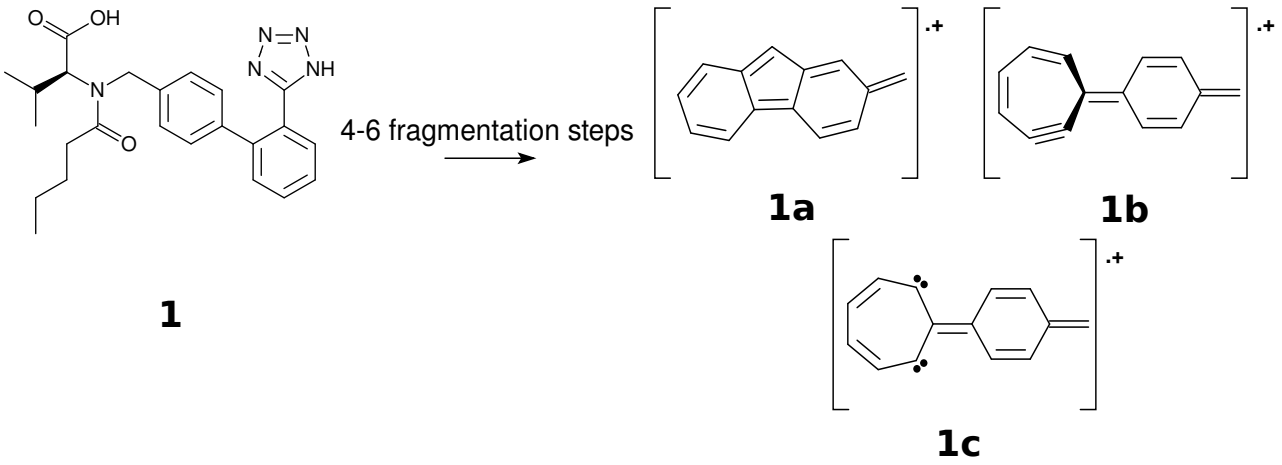

Figure 5.4.: Examples of calculated $C_{14}H_{10}^{+}$ (m/z 178) Valsartan fragment isomers.

automatically derives from a very reasonable QC computation and the actual effective temperature the distribution of charge between fragments (assuming Boltzmann statistics) and based upon that decides which fragmentation cascade to follow. Undoubtedly, there are some deficiencies as well: false positives and negatives and wrong peak intensities are all clearly visible in Figure 5.3. Before jumping too hastily to conclusions we first examine the other simulated spectra in order to get a more comprehensive view.

5.2.2. Erythromycin

Figure 5.6 shows that QCEIMS is able to predict nearly all major peaks of the experimental spectrum correctly. The m/z series 58, 71, 86, 99 may be explained by the rationale in Figure 5.7. The difference of 28 m/z units between 58 and 86 (and 71 and 99, respectively) is connected to a formal loss of ethylene (C_2H_4). This is of course to be taken with care as already seen from the fragment structures in Figure 5.7, which resulted from different individual fragmentation runs. The main peak of the experimental record, m/z 158, most likely results from the desaminosyl unit ($C_8H_{16}NO_2^{+}$) of erythromycin.

This is only partially reflected by the computed spectrum, which suffers from some over-fragmentation. Apparently, in this case the default settings in the 'hot' ion preparation procedure of our algorithm put too much energy into the molecule, leading to further and further fragmentation to a greater degree than observed in the experiment. This is also supported by the inspection of individual computational fragmentation runs wherein the $C_8H_{16}NO_2^{+}$ fragment itself often decomposes to the N-containing fragments in Figure 5.7. Note that we have not made any attempts for improving or fine-tuning of the calculations for the individual

method. Especially the spin states are ill-defined as indicated by $\langle \hat{S}^2 \rangle$ values of 1.8 to 1.9 (the exact value for a doublet is 0.75). However, as already discussed in the original publication [10], this may not be crucial.

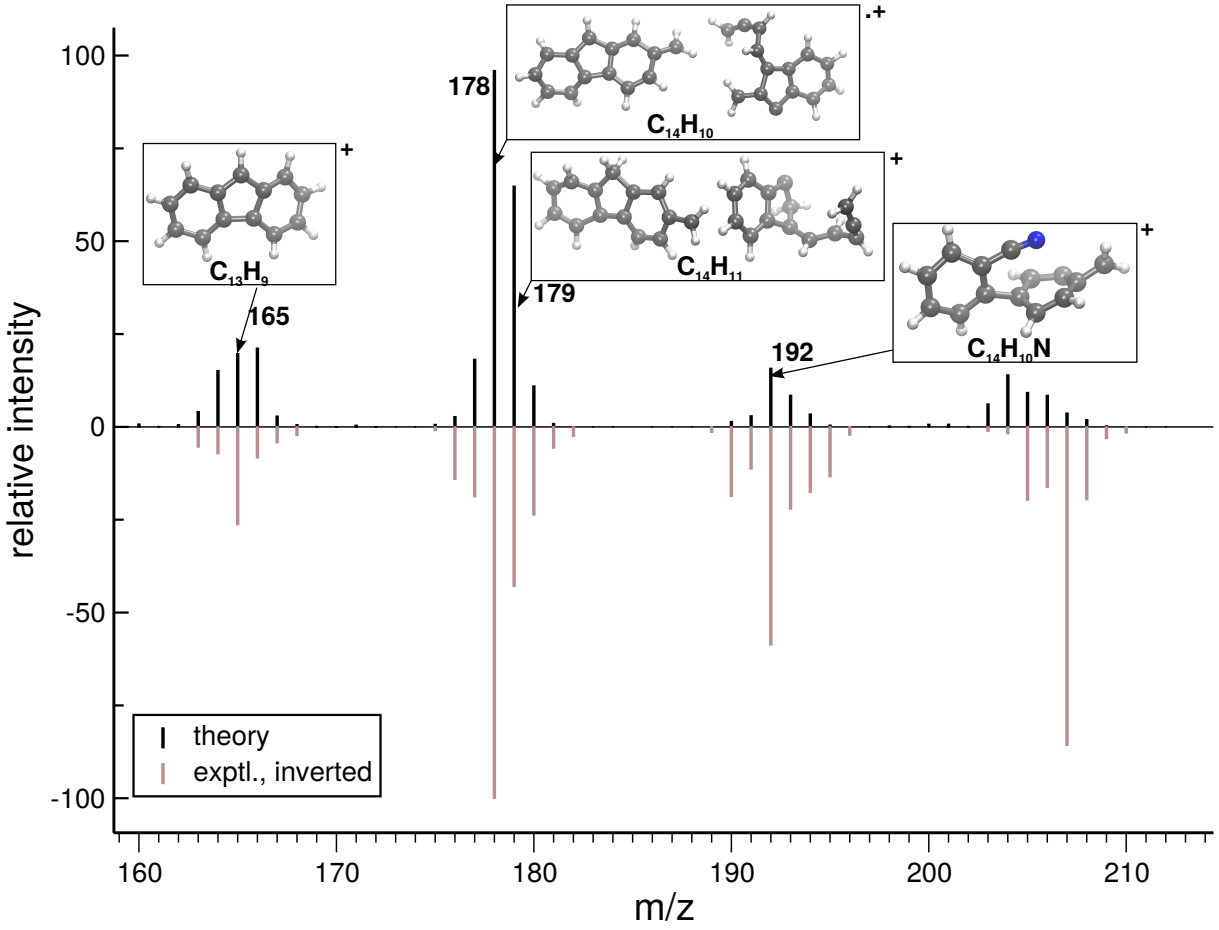

Figure 5.5.: Comparison of experimental and computed spectrum of valsartan in the m/z range 160-210. Additionally, selected average fragment structures are displayed as given by the QCEIMS computations, and assigned to peaks in the mass spectrum of valsartan.

examples and that the theoretical spectra have been obtained always under the same computational conditions. While apparently a change of internal parameters of QCEIMS (mostly a single one which changes the internal excess energy in the molecule upon ionization) could lead to a better predicted spectrum for **2**, at this point we refuse to do any further empirical modifications in order to keep one consistent 'first principles' protocol. Note that there is currently no practical theory available to non-empirically estimate the critical internal energy of the ionized molecule in an EI-(2e,e) process.

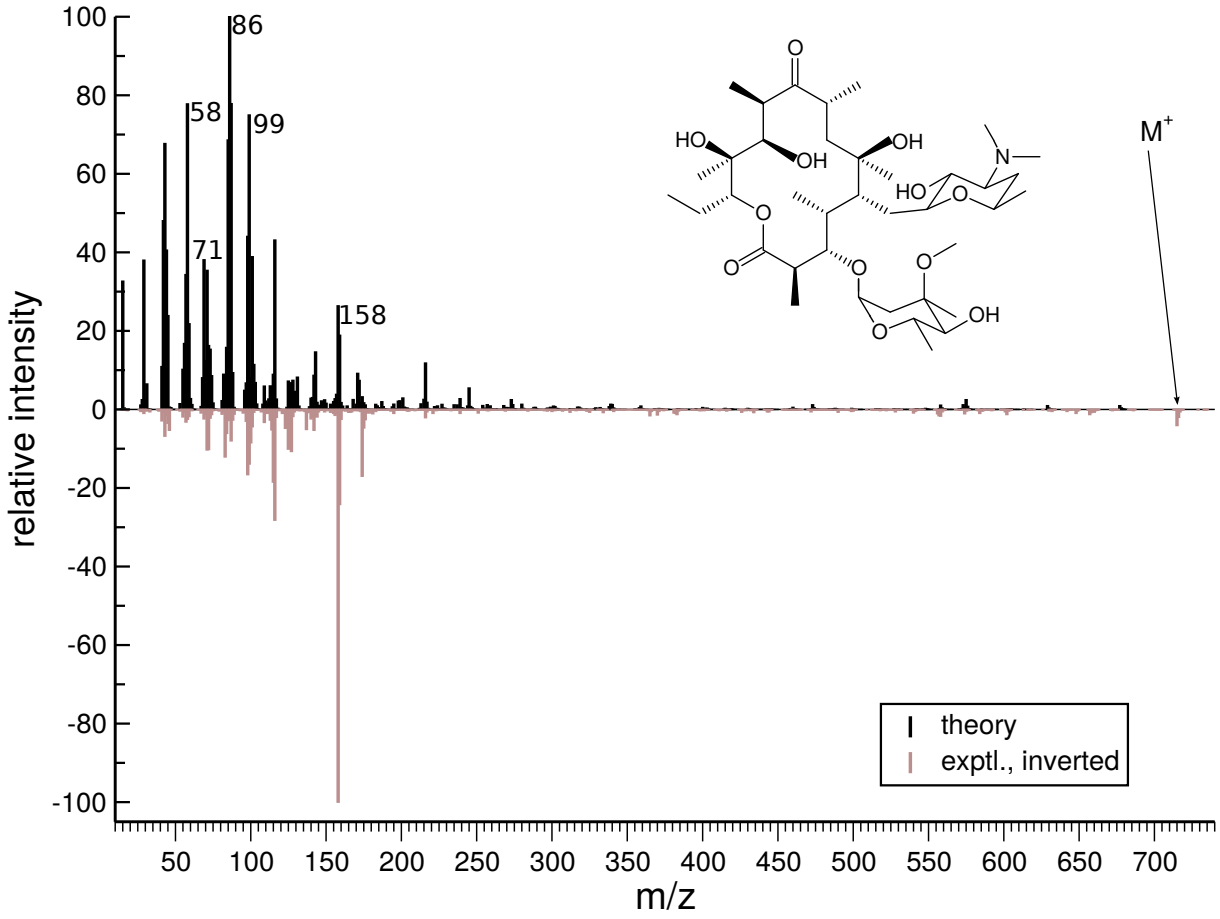

Figure 5.6.: Calculated mass spectrum of erythromycin in comparison with the experimental spectrum. Indicated m/z values are discussed in the text. Note that the molecular ion gives only a very weak signal at m/z 733 in the experimental spectrum and none in the computed.

5.2.3. Taxol

Figure 5.8 shows the results for taxol, which is the heaviest of the example molecules. The fragmentation pattern is reproduced to a remarkable degree. As in the experimental record, there are hardly any ions heavier than m/z 350 and there is no signal at all by the molecular ion. Some peaks with high relative intensities are accounted for particularly well by the simulation. The m/z 43 peak belongs to an acetyl moiety (H_3CCO^+) that can be cleaved off at two positions in the taxol structure, which may also explain the high probability of producing this signal. The m/z 77 peak results from the phenyl cation $C_6H_5^+$, which may also be produced at various positions of the taxol frame. The base peak (both theoretically and experimentally) at m/z 105 stems from a benzoyl group ($C_6H_5CO^+$), which again may dissociate from the parent molecular ion at two different positions. The signal at m/z 210 is

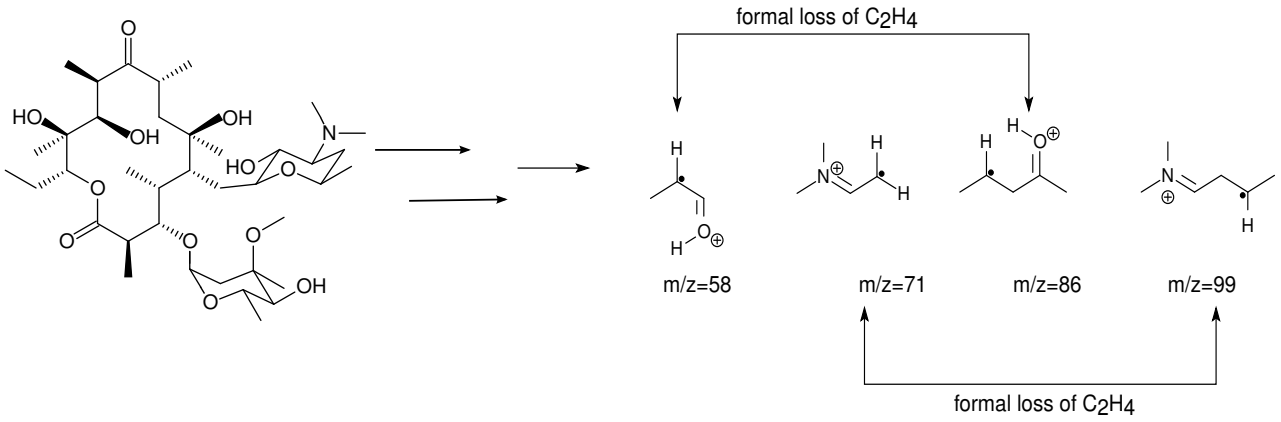

Figure 5.7.: Calculated fragments of erythromycin yielding a series of peaks.

one of the more 'diagnostic' peaks of the spectrum as it results from the cation of the N-benzyl benzamide moiety (molecular formula $C_{14}H_{12}NO^+$, see drawing in Figure 5.8), which is also reflected by the calculations. However, there are different contributing ions with m/z 210 from our calculations, which are not $C_{14}H_{12}NO^+$ isomers. This perhaps indicates that there are competing processes that both lead to this signal. In order to resolve this issue one would need highly resolved experimental data likely involving isotopic substitution, which are not at our disposal.

Among the most prominent peaks that are missing in the simulated spectrum are m/z 51, m/z 60, and m/z 91. Possible reasons for these false negatives are discussed in section 5.2.5.

5.2.4. Statins

Figure 5.9 shows the calculated and experimental spectra of lovastatin (**4**) and simvastatin (**5**). Calculations on these two molecules may be viewed as a miniature homology series. The computed spectra for **4** and **5** have similar overall 'shape', much like their experimental counterparts. Small (local) chemical modifications in a remote part of a molecule should have and have only a minor impact on the computed spectra. This indicates good internal consistency and sufficient sampling in the QCEIMS method. Moreover, many experimental peaks were predicted correctly, albeit with somewhat incorrect relative intensities.

This is mainly related to an excess of *sec*-butylium (m/z 57, $C_4H_9{}^+$) and *tert*-pentylium (m/z 71, $C_5H_{11}{}^+$) signals for **4** and **5**, respectively. These groups are in α position of the open-chained ester moieties of **4** and **5**. The spectral prediction for **5** is worse than for **4**, caused mainly by the (inaccurately) immensely high count of *tert*-pentylium cations as depicted in Figure 5.9b. A similar problem of exaggerated alkyl loss has also been discussed in the original paper.[10] The reasons for this problem are not yet fully understood but likely result from inaccurate potential energy surfaces calculated by the approximate semi-empirical QC method, a field of ongoing development and testing.

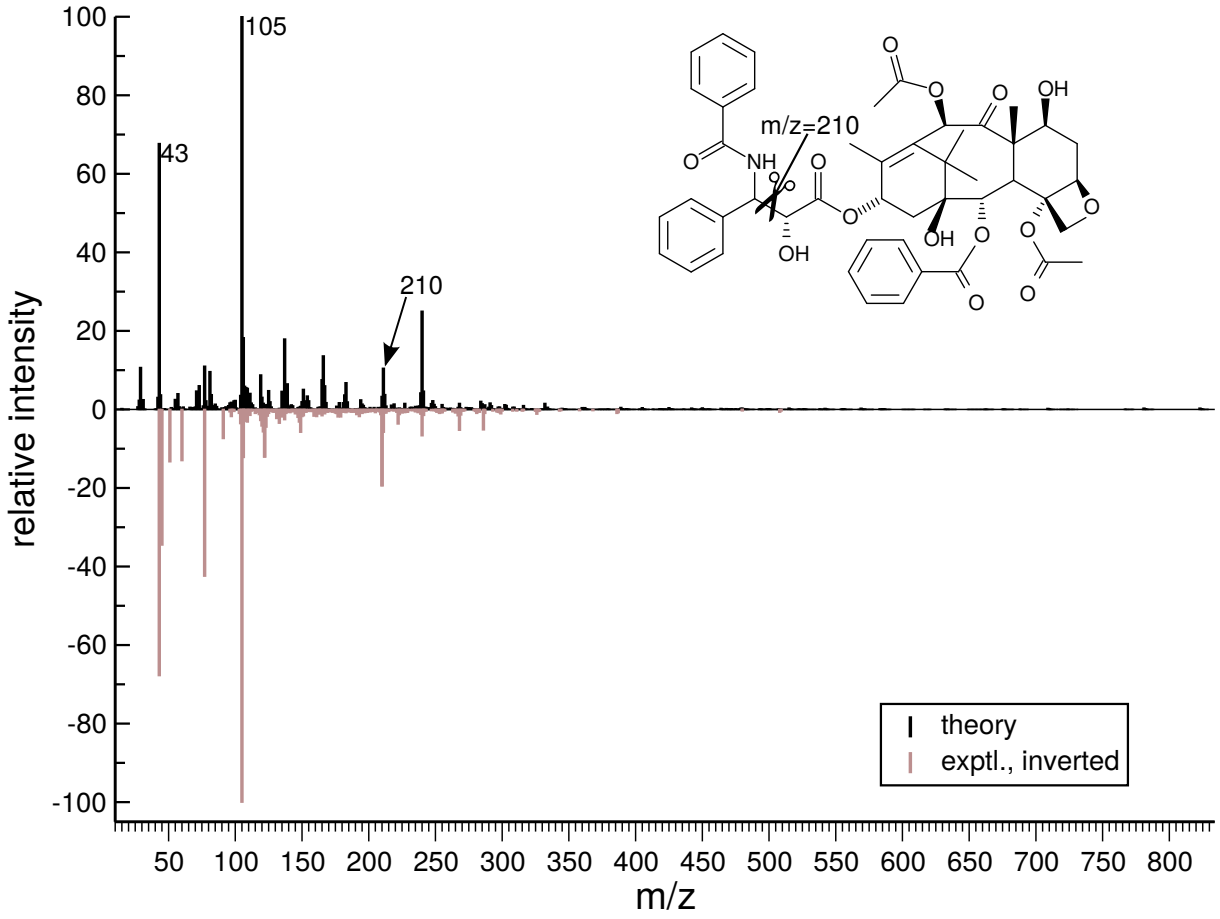

Figure 5.8.: Calculated mass spectrum of taxol in comparison with the experimental spectrum. Indicated m/z values are discussed in the text. Note that the molecular ion (m/z 853) is missing in both spectra.

5.2.5. Strengths and Weaknesses of the Approach

From the data presented above one may assess advantages and disadvantages of our method. There are a few very strong points which are summarized first. QCEIMS is a reliable and robust method. A single theoretical calculation protocol was used (vide infra) to compute spectra of good quality. The relatively accurate computation of EI mass spectra for comparatively large molecules in a highly parallel and nearly automated fashion is unprecedented. Molecules of relatively large sizes at the edge of the applicability of experimental EI-MS itself (given mostly by the limitation by vapor pressure and related thermal decomposition processes) can be treated by our MD protocol. This stands in stark contrast to conventional approaches employing static quantum chemistry, which rely on priory knowledge of decomposition pathways. This information is often not available and impossible to comprehensively obtain for the here considered large molecules in light of the immeasurable complexity of reaction space for even

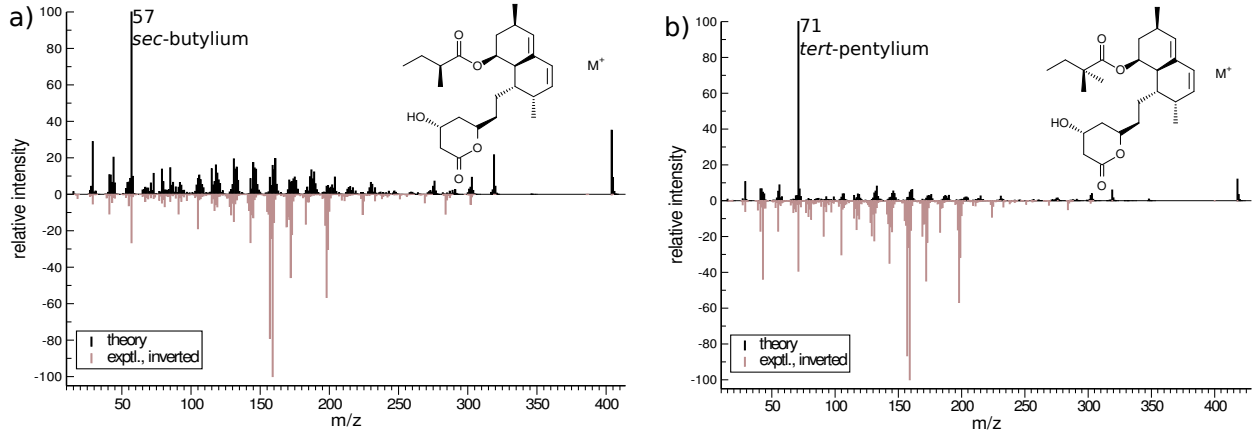

Figure 5.9.: Calculated mass spectra of a) lovastatin and b) simvastatin in comparison with the experimental spectra. Marked peaks are discussed in the text.

a handful of heavy atoms. Fragmentation pathways up to a user-defined recursive depth of cascade reactions come out naturally from our simulations. By employing MD, vibrational, thermal and anharmonic effects are naturally taken into account and fragmentation or reaction mechanisms can derived from the molecular trajectories and translated to conventional formula 'language'. A more detailed statistical analysis and automated procession of the thousands of MD runs is planned in the future.

On the other hand, QCEIMS is not perfect and does produce false positive and false negative signals. There are three basic reasons for this: (i) There is a problem with the potential energy surface (PES), i.e., the dissociation energies or barriers as computed by the chosen quantum chemical method are in error. This problem can only be addressed by applying a more accurate quantum chemistry which is difficult in practice for larger molecules with the current computational resources. (ii) The fragments are produced in reactions that take longer than the maximum simulation time, such as certain types of rearrangement reactions. This can be checked in principle by simulating longer (e.g., to the 0.1-1 ns range), which at the moment turned out to be somewhat too costly for routine treatments, at least for such large cases as studied here. (iii) The energy distribution of the ionization excess energy (IEE) in the parent ion is in error or reactions occur from electronically excited ion states. Practically nothing is known here for large molecules and one can only speculate how big these effects are. For the larger compounds considered in this work we had to made some changes compared to the original ansatz to distribute the IEE as discussed in appendix B. The original algorithm localized the impact energy too much in parts of the molecule which lead to very unrealistic, too fast fragmentations. Further work to understand this part of the theory better is under way. False positive signals are produced rarely by QCEIMS. Such errors may be traced back to an overestimated IEE or an inaccurate assessment of ionization potentials (IPs) because of

heavily distorted fragment geometries. Most mass spectral search algorithms used in conjunction with MS databases will treat such artifacts in much the same manner as impurities in an experimental mass spectrum of an unknown compound. Therefore, false positives - so long as there are not too many of them - are the lesser problem when compared to missing fragment peaks.
Lastly, there can be a problem with the general usability of the QC method. For compounds containing third-row elements there are currently no parameters available for the semi-empirical OM2 Hamiltonian used. For systems containing the elements H, B, C, N, O, F, P, S, and Cl, OM2-D3 may be substituted by DFTB3-D3,[136,229] which delivers only slightly worse results at even lower computational cost, see ESI or the original paper for examples. For molecules with less common elements the use of standard DFT is always a fall-back option but as mentioned before, this is at least with the current hard- and software capabilities computationally too demanding for compounds with more than 20-30 atoms.

5.3. Conclusions

By the QCEIMS method we were able to reproduce EI mass spectra of medium-sized to large organic molecules relevant in medicinal chemistry to a satisfactory degree. Unimolecular decomposition and rearrangement reactions are described rather well by QCEIMS, and peak assignments as well as fragmentation paths can be extracted from our simulations. No molecule-specific empiricism was applied and solely the molecular structure was used as input. Despite the drawbacks and possible shortcomings mentioned in the discussion, we suggest QCEIMS as a sound new approach that could potentially be used as a 'black box' tool in order to routinely compute EI mass spectra of organic compounds. This claim is supported by the fact that one consistent protocol based on semi-empirical QC and DFT methods has proven to be more than adequate to reproduce EI mass spectra of sizeable drug molecules. There is ongoing work in our laboratory to achieve the following mid to long-term goals: (i) to make nanosecond simulation timescales routinely accessible by developing even more efficient computational methods. (ii) to include organometallic compounds which at the moment cannot be treated by the semi-empirical methods used and (iii) to increase user-friendliness of our program in order to make it available to a wider community.

5.4. Computational Details

The neutral ground state structures of the molecules **1-5** were optimized using dispersion-corrected DFT at the TPSS[60]-D3[33,34,35]/def2-TZVP[48] level as implemented in Turbomole 6.5.[63] The nature of the stationary point on the PES was confirmed to be a (local) minimum

by calculating the harmonic vibrational frequencies.*

The QCEIMS program was used with an impact energy of 70 eV. The IEE distribution was computed according to a Poisson energy distribution, with the greatest possible IEE being 70 eV - $\varepsilon_{\mathrm{HOMO}}$, where the orbital energy $\varepsilon_{\mathrm{HOMO}}$ was computed at the PBE12[47,117]/SVx[232†]//TPSS-D3/def2-TZVP level of theory. Starting from their optimized geometries, the systems were equilibrated and a randomized geometry/nuclear velocity ensemble was generated for each case by running an MD trajectory of the respective molecular ground state, wherein the PES was generated 'on the fly'. The quantum chemical method for this purpose was OM2[228] with the D3 dispersion correction.[33,35] The initial temperature for each trajectory was set at 500 K, which is the default parameter (and sufficiently close to 250°C $\widehat{=}$ 523 K, which was given as the source temperature in the experimental records for **1**, **2**, and **4**). The number of production runs performed for each spectrum was set to 1,000 for all cases studied, as the simulated spectra showed convergence even at this low number of runs. The maximum number of cascading runs in order to track down secondary, tertiary etc. fragmentations was seven. The starting point was always the lowest electronic radical cation state of the molecular ion, with the geometry and nuclear velocities taken from the ground state ensemble. Unrestricted SCF calculations were performed in all fragmentation runs. In order to achieve SCF convergence and to partially account for the multiconfigurational character of the electronic structure of the electronic state(s) involved, the Fermi 'smearing' technique was used.[27,44,45] The vibronically 'hot' ensemble was created by scaling the nuclear velocities along the nuclear degrees of freedom uniformly until the - internal - kinetic energy was equal to the IEE. This is a deviation from the protocol used in the original work, where velocity scaling was dependent on the localization of molecular orbitals to be ionized selected at random. The effects of this modification in the algorithm are discussed in appendix B. The statistical fragment charge assignment algorithm used the Boltzmann factor for ionization potentials (IPs), $\mathrm{e}^{-\frac{\Delta \mathrm{IP}}{\mathrm{k}T}}$, where k is the Boltzmann constant and T is the current vibronic temperature at the fragmentation event. The IPs were computed at the OM2//'average fragment geometry' level of theory. This is justified by demonstration of difference spectra between IP calculations at semi-empirical and DFT levels, see ESI. The maximum simulation time for the initial trajectories was set to 5 ps. Depending on the number of secondary runs performed, the actual maximum simulated time reached times between 5 and 10 ps in some individual runs. At the very end, all fragments were counted according to their statistical (Boltzmann) weight, yielding the theoretical EI mass spectrum.

*This in some cases yielded one imaginary mode with a very small negative eigenvalue, which may be neglected in our case since it is only important to provide a geometry close enough to a local minimum in order to create a randomized ensemble of starting geometries.

†PBE12 is the same as PBE0, only with a Hartree-Fock exchange coefficient of 0.5. The SVx basis set is the same as SV(P), only without d-polarization functions at the carbon atoms.

For OM2 calculations, the MNDO program[245] was called and the DFT calculations were carried out by the ORCA suite of programs.[70,71]
Experimental spectra for comparison were downloaded from mass spectral databases available on-line.[243,244]

6. Elucidation of Electron Ionization Induced Fragmentations of Adenine by Semi-Empirical and Density Functional Molecular Dynamics

Christoph Alexander Bauer and Stefan Grimme*
Received 24th of September 2014, Published online 13th of November 2014

Own manuscript contribution

- Performance of all calculations
- Interpretation of data
- Writing the manuscript

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

Abstract The gas phase fragmentation pathways of the nucleobase adenine upon 70 eV electron ionization are investigated by means of a combined stochastic and first principles based molecular dynamics approach. We employ no pre-conceived fragmentation channels in our calculations, which simulate standard electron ionization mass spectrometry (EI-MS) conditions. The reactions observed compare well to a wealth of experimental and theoretical data available for this important nucleic acid building block. All significant peaks in the experimental mass spectrum of adenine are reproduced. Additionally, the fragment ion connectivities obtained from our simulations at least partially concur with results from previous experimental studies on selectively isotope labeled adenines. Moreover, we are able to assign non-cyclic structures which are entropically favored and have not been proposed in non-dynamic quantum chemical studies before to the decomposition products, which result automatically from our molecular dynamics procedure. From simulations under various conditions it is evident that most of the fragmentation reactions even at low internal excess energy (<10 eV) occur very fast within a few ps.

6.1. Introduction

The question of fragmentation pathways of the nucleobase adenine ($C_5H_5N_5$, **A**, for structure and atom labels see Figure 6.2) has been keeping the scientific community busy for nearly 50 years. During the 1990s and 2000s, this topic again attracted attention due to the discovery free-electron interaction with nucleic acids in the aftermath of radiation-induced events in the cell.[220,255,256] Fragmentations of $\mathbf{A}^+$ occurring in the context of EI-MS were first discussed comprehensively by Rice and Dudek in 1967[257], revealing the sequential loss of hydrogen cyanide (HCN) units. In the following years, mechanistic studies involving selectively isotope-labeled **A** were conducted[236,258,259,260] and subsequently the fragmentation patterns of $\mathbf{A}^+$, $\mathbf{HA}^+$, and $\mathbf{A}^{2+}$ using various ionization techniques were reported.[261,262,263,264,265,266] The protonated and doubly charged species showed similar reactivity patterns to the singly charged one. Jochims et al. determined appearance energies of certain decomposition products of **A** by photo-ionization mass spectrometry and assigned fragment structures entirely without theoretical support.[267] Attempts to rationalize fragmentation channels of purines based on non-dynamic quantum chemistry (QC) have been scarce[200,268], which is undoubtedly related to the complexity of the problem. Recently, a combined experimental and theoretical study on the fragmentation of **A** was published, in which the authors rationalize dissociation events and intramolecular rearrangements after EI by standard QC calculations.[203]

Incited by the re-kindled interest in this subject, we have computed the unimolecular decomposition reactions of $\mathbf{A}^+$ using a novel approach based on Born-Oppenheimer Quantum Chemical Molecular Dynamics (BO-QCMD)[13,207], dubbed 'Quantum Chemistry Electron Ion-

ization Mass Spectra' (QCEIMS), where the QC method may in principle be chosen freely (provided there are affordable, preferably analytic nuclear forces available), see below. This method employs an automatic, black-box type procedure to compute EI mass spectra from first principles.[10] One main advantage of dynamic procedures is that the prior knowledge of reactive pathways is not necessary. The simulation is instead guided by the gradient of the potential energy surface (PES), generated "on the fly" by a semi-empirical QC (for cost reasons) or density functional theory (DFT) method, rather than by "chemical intuition". The QCEIMS method aims at a complete and realistic simulation of the MS experiment and directly provides relative reaction rates for various fragmentation processes rather accurately. In addition it yields semi-quantitative information about absolute time scales which are not easily accessible otherwise. The output of the stochastic procedure after hundreds of fragmentation MD runs is a simulated mass spectrum which can be directly compared to the experimental one.

The trajectories obtained in the output also provide information about the time scales and mechanisms of the unimolecular decomposition reactions, rearrangements etc. On the one hand, one may obtain knowledge about a given fragmentation path by inspecting visually the trajectories given by our program. A link to a selection of trajectory videos is given in appendix C. On the other hand, the runs may be quantitatively analyzed with regard to geometric parameters such as bond lengths, angles, and dihedral angles. Since in this case, we are particularly interested in bond breaking events, we have analyzed interatomic distances of given nuclear coordinates (i.e., atoms that are bonded to each other in the ground state geometry of **A**) along the purine frame, see results below.

6.2. Theoretical and Computational Details

All calculations herein were carried out for the 9-H tautomer of **A**, which is the only isomer significantly populated in the gas phase, even at elevated temperatures.[269] The ground state geometry was optimized at the DFTB3-D3 level of theory.[33,34,35,136,229,230] A low-level optimized ground state structure is fully sufficient as input for our simulations. The QCEIMS program framework[10] employs a combined stochastic and dynamic algorithm which is described in detail in the original publication.[10]. Here, we provide only the main features of the general procedure with its three distinctive steps:

1. An ensemble of randomized molecular structures was obtained by sampling a (neutral) ground state BO-QCMD-trajectory on the semi-empirical OM2-D3[228] PES. Newton's equations of motion were integrated numerically in an NVE ensemble (where the conserved quantities N, V and E are the number of particles, the volume the energy, respectively) using a Velocity-Verlet algorithm[270] with a time step of 0.5 fs. After 6 ps

of equilibration, a 12 ps production run was conducted from which the starting structures for the fragmentation cascades were generated. This rather short simulation time is justified by the lack of conformational degrees of freedom for the nearly planar **A** molecule.

2. An ionization excess energy (IEE) was assigned to each randomly chosen starting geometry based upon a Poisson distribution:

$$P(E) = \frac{\exp[cE(1 + ln(b/cE)) - b]}{\sqrt{(aE + 1)}}, \tag{6.1}$$

where $P(E)$ is the probability to have an IEE of E in the ion, $b \approx 1$, $c = \frac{1}{aN_{\mathrm{el}}}$ with $a \approx 0.2$ eV and N_{el} the number of valence electrons. To further clarify this, a depiction of two different IEE distributions, with IEE/atom 0.6 eV and 0.3 eV, respectively, is shown in Figure 6.1 below, where the zero of energy corresponds to the ionization potential of the molecule and the structure in the distributions results from the stochastic character of the procedure (selection of orbital energy levels, see below and Ref.[10] for details). In this study a value of 0.6 eV/atom was taken corresponding for **A** to a total, average IEE of 9 eV. This value has been adopted from the original publication where it was determined the best choice in order to reproduce EI mass spectra of a selection organic compounds.[10]

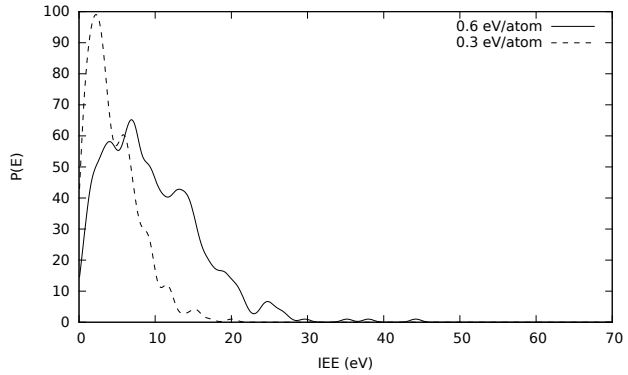

Figure 6.1.: $P(E)$ for **A** with two different settings. The maxima of $P(E)$ are located at 4.5 eV (0.3 eV/atom), and 9 eV (0.6 eV/atom).

3. Simulation of the unimolecular ion fragmentations: The molecule was assumed to be instantaneously ionized formally to the ground state of $\mathbf{A}^+$. Internal energy conversion and its consequence, vibrational heating, was simulated by scaling the nuclear velocities up to the previously assigned IEE. The time constant τ_{h} over which this was done varied

between 0.2 and 2 ps, as approximated by equation 6.2 (energy gap law[271])

$$\tau_{\mathrm{h}} = \sum \frac{\mathrm{k_h}}{N_{\mathrm{el}}} \exp[-\alpha(\varepsilon_i - \varepsilon_j)], \tag{6.2}$$

where the rate constant $\mathrm{k_h}$=2 ps, $\alpha = 0.5$ eV^{-1} (a parameter), and $\varepsilon_i - \varepsilon_j$ is the energy gap (in eV) between the two Kohn-Sham molecular orbitals i (the one which is ionized) and the higher lying orbitals j up to the highest occupied molecular orbital (HOMO), to which the system relaxes electronically. The energy gap between two states is represented here in the single-particle picture. The orbital energies ε_i were taken from a standard DFT calculation for the ground state (same level of theory as for the DFT PES used during the dynamics, *vide infra*). Because of the heating process (transfer of electronic to thermal energy), the total energy was only formally but not numerically preserved in the first part of the MD run.

The PES for the fragmentation runs was computed on the fly on three different levels of theory: (i) semi-empirical OM2[228], together with the D3 dispersion correction[33,34,35] (OM2-D3) (ii) the DFTB3-D3 (as used for ground state sampling and geometry optimization, respectively), and (iii) PBE0[47,117]/SVx[232*]-gCP[231]-D3 (DFT-D3). The number of cascading fragmentation runs was 500. These were carried out in parallel using the chosen QC method. The maximum simulation time for the initial production run was 5 ps but overall run times including cascade fragmentations can reach up to 10 ps, and longer simulation times have also been tested, see below and appendix C. Unrestricted (spin-polarized) self-consistent field (SCF) calculations were carried out throughout, and the fractional occupation number (FON) technique was used to ensure SCF convergence in electronically complicated situations[27,44,45] which often occur in the MD runs. Statistical charges were assigned to the fragments (detected automatically by the program) according to the Boltzmann factor based on differences of computed ionization potentials (IPs), $\exp[\Delta IP/(\mathrm{k}T)]$, where T is the current vibronic temperature at the time of the fragmentation event, and k is the Boltzmann constant. These were obtained at the same DFT level used for the trajectories on average fragment geometries from the last 100 steps (corresponding to 50 fs) of the fragmentation runs.

The global run parameters were the primary ionization energy of 70 eV and an assumed initial source temperature of 500 K. A more detailed description of the QCEIMS algorithm and minor technical details has been given and discussed in the original paper[10] and its SI. No modifications to the original procedure specifically for the here investigated molecule have been made. The computed EI mass spectra were generated by adding up all calculated fragment ions weighted by their statistical charges. Natural isotope ratios were considered implicitly.

*The SVx basis set is the same as SV(P), only without d-polarization functions at the carbon atoms.

The numbers of occurrences of the molecular ion in the theoretical base peak of the presented spectra were 146 (OM2-D3), 172 (DFTB3-D3), and 230 (DFT-D3), respectively. For all OM2 calculations, the MNDO program[245] was called, DFTB3 calculations were carried out by `dftb+`[246], and all DFT calculations were carried out using the ORCA program system.[70,71]

6.3. Results and Discussion

The computed mass spectra ordered by increasing quality of the corresponding QC PES are shown in Figure 6.2. All of them agree very well with the experimental spectrum, taken from the NIST database[243]. To the best of our knowledge, this is the first computational account of an EI mass spectrum including spectral topology with regard to such a relatively complex heterocyclic compound as **A**. The first significant result of our study is the observation that mass spectral prediction improves when going from semi-empirical methods (OM2, DFTB3) to the DFT level. This amelioration is already qualitatively apparent in Figure 6.2. A quantitative measure is provided by a composite spectral matching score,[9] which shows significant improvement when comparing the DFT-D3 calculated mass spectrum to the two spectra determined at semi-empirical QC levels. We arrive at this score by a calculation of the overlap of two spectral vectors. 0 means that the two vectors are orthogonal and 1 means that they are identical. For this article, the computed score was multiplied by 1,000, yielding a number between 0 and 1,000. The technical details of how this number was calculated are found in appendix C.

Longer test calculations (simulated time of 100 ps) at the OM2-D3 and DFTB3-D3 levels of theory reveal no significant changes of the quality of the computed spectra, which indicates that the majority of fragmentations is fast or even ultra-fast, i.e., occurring within < 1 ps. The IEE distribution was also changed for these testing purposes to simulate "milder" conditions and to check whether fragmentation paths become accessible only after a longer simulated time, e.g., after 10 ps. This was hardly the case in the computations reported here. The match scores of the calculated spectra produced with a simulation time of up to 100 ps are quite comparable to the ones shown in Figure 6.2 (541 for OM2-D3 and 536 for DFTB3-D3, respectively). The pertaining figures (additional computed spectra with different IEE distributions and maximum simulation time of 100 ps) are found in appendix C.

Notably, all experimentally observed major peaks are found in the theoretical mass spectra, regardless of the relative quality of the underlying QC method. This provides some confidence that the extracted fragmentation paths from the MD simulations are realistic and allow a detailed elucidation of the corresponding mechanism. To this end, the QCEIMS production runs at the OM2-D3, DFTB3-D3 and DFT-D3 levels of theory were inspected, and an analysis of the fragment ion structures and their relative abundances was performed. Moreover, the interatomic distances along the purine frame were investigated for the individual runs in order

to reveal possible bond ruptures. Detailed analysis of this kind is reported for the first time in the QCEIMS framework.

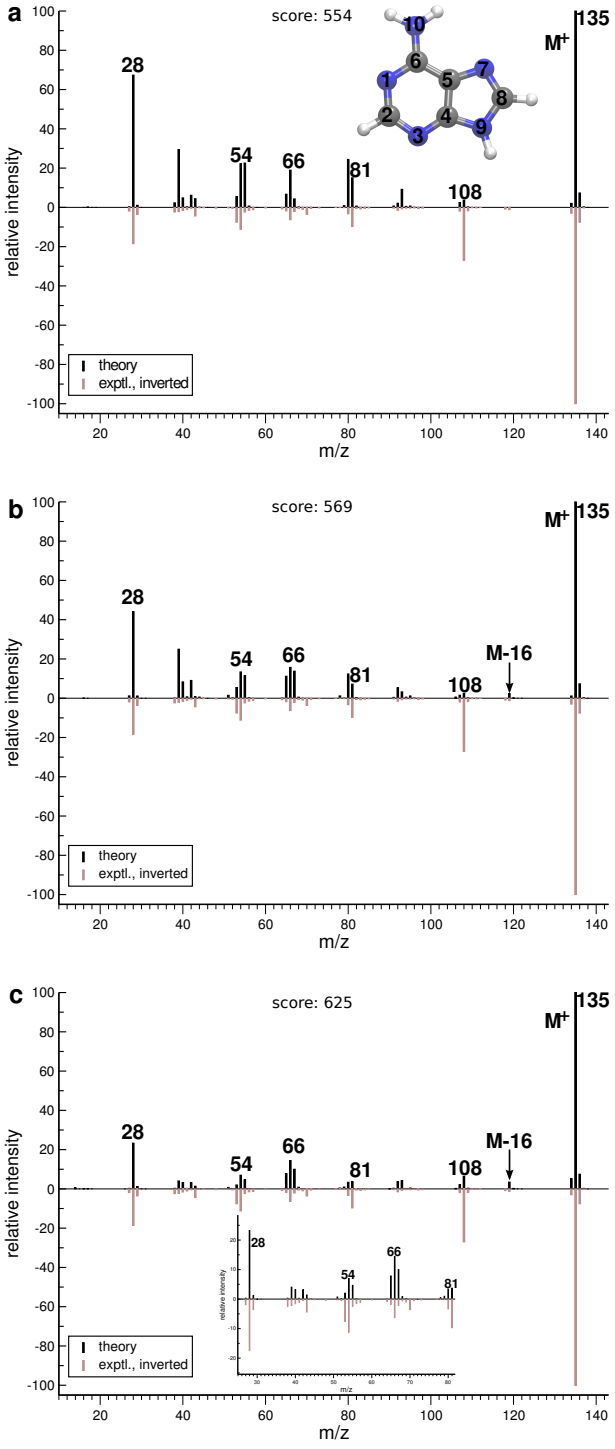

Figure 6.2.: Calculated vs. experimental EI mass spectra of **A** with at different levels of theory with corresponding match scores to the experimental data (see text and SI). **a**: OM2-D3 **b**: DFTB3-D3 **c**: DFT-D3 with zoom on m/z range 25 to 81 as insert. The peaks indicated by their m/z values are discussed in the text.

First of all, in the relative majority (29 % for OM2-D3, 34 % for DFTB3-D3, and 46 %

for DFT-D3) of all fragmentation runs, no bond splitting events occur within the finite time window of 5 ps at all, which translates to a correct molecular ion base peak ($\mathbf{M}^+$, m/z 135, in Figure 6.2 a-c) in the computed spectra. Secondly, the small $(\mathbf{M}\text{-}1)^+$ peak can be explained by loss of the H atom originally bound to C2. This result is to be taken with care because of a somewhat unreliable statistics regarding such rare events in our calculations. The expulsion of the NH_2 group is also very scarce in our QCEIMS simulations, and no $(\mathbf{M}\text{-}16)^+$ fragment is observed in our OM2-D3 simulations (and only very few counts are registered at DFTB3-D3 and DFT-D3 levels, respectively), which concurs with experimental results[243,257].

More intricate decompositions of $\mathbf{A}^+$ require the disintegration of the purine ring system. The bond distances along the purine frame were investigated to this end during the initial trajectories of the fragmentation runs. More details on this analysis (including graphical representations of interatomic distances in a 1-5 ps time window) are presented in appendix C. Figure 6.3 shows the percentage of broken bonds along the ring perimeters of $\mathbf{A}$ after the initial MD fragmentation runs on all three levels of theory employed. The largest differences between the different PES occur regarding the most frequently cleaved bonds, which are the C4–C5 and C5–C6 bonds. These are a lot more prone to splitting events on the OM2-D3 and DFTB3-D3 PES than on the DFT-D3 PES. While C4–C5 is still the most vulnerable bond even on the DFT-D3 PES (opened in 27 % of all trajectories on this level of theory), the distribution of bond cleavage events is more uniform at this level of theory. However, with the exception of C4–C5 and C5–C6 (and N9–C8 to a lesser extent), the results for the three different PES compare rather well, e.g., the N7–C5 bond is the most stable one for all QC methods. Dissociations along the purine frame of $\mathbf{A}^+$ lead eventually to de-anellated intermediates, which may be meta-stable before further reactions occur.

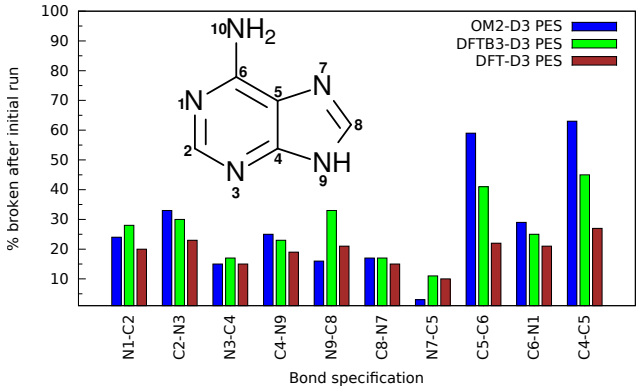

Figure 6.3.: Percentage of broken bonds (inter-atomic distance ≥ 2 Å) along the ring perimeters of $\mathbf{A}$ after the initial MD fragmentation runs.

We have compiled one facet of the decomposition network of $\mathbf{A}^+$ as taken from our simulations on the OM2-D3 PES in Figure 6.4. From the molecular ion $C_5H_5N_5{}^+$, HCN units are

cleaved off sequentially as expected, but in various places during different individual trajectories, shown by the (exemplary) connectivities of the "survivor ions". All depicted fragments are open-chained. While the fragment at m/z 108 ($C_4H_4N_4^+$) serves in some simulation runs as the parent ion of m/z 54 ($C_2H_2N_2^+$), the production of $C_2H_2N_2^+$ ions by this pathway is rather unlikely. For one alternative pathway, see snapshots of an exemplary fragmentation trajectory in appendix C. We also predict that the formation of the $C_3H_3N_3^+$ ion (m/z 81) involves the splitting of different bonds than for the depicted $C_4H_4N_4^+$ ion. An additional neutral loss that may occur apart from HCN is the geminal amino nitril H_2N10−C6≡N1, whose dissociation together with H–C2–N3 is the prevalent computed pathway for the appearance of the $C_3H_2N_2^+$ species (m/z 66). Finally, the computed peak at m/z 28 stems from $HCNH^+$, which is most easily formed at positions HC8–N9H, where no hydrogen rearrangements are necessary prior to the formation of this fragment. This peak is somewhat too intense in our computed OM2-D3 and DFTB3-D3 mass spectra, which indicates that the C4–N9 and C8–N7 bonds are split too easily on the semi-empirical PES (which occurs less often on the DFT-D3 PES, see Figure 6.3).

At this point, we must also refer to the potential weaknesses of our simulation protocol: Firstly, the quality of the quantum chemical PES is not perfect, and neither can it be at this point due to cost reasons. Although the semi-empirical methods and the hybrid DFT method with a small basis set used here have so far tested well for the computational reproduction of EI mass spectra, there is certainly a need for a thorough benchmarking investigation in order to provide more confidence that the picture drawn from our MD simulations is realistic. Work in this direction is underway in our laboratory. From the differences between the three PES in Figure 6.3 concerning especially the C4–C5 and C5–C6 bonds, the connectivities of the fragments depicted below are not 100 % conserved over all simulations (at different levels of theory) for all of the fragments. Especially hydrogen positions may vary as it is well-known that they are scrambled during the fragmentation processes.[272]. It follows from these considerations that the fragmentation network in Figure 6.4 resulting from our simulations is intrinsically linked to the underlying QC method that is used for on the fly PES generation. Nevertheless, we are confident that a more detailed overview of the dissociative processes of $\mathbf{A}^+$ may be obtained by our simulations based on comparisons to a number of previous experimental and theoretical studies, see below. A problem of a different sort arises from the fact that our protocol is stochastic: For some of the peaks with low intensities (indicating a small number of fragment ion counts), the reported structures may not be wholly representative because of insufficient statistical treatment. In order to get reliable and significant statistics for these less accessible decomposition pathways, one would have to increase the number of trajectories by one order of magnitude which was not possible with our current computational resources.

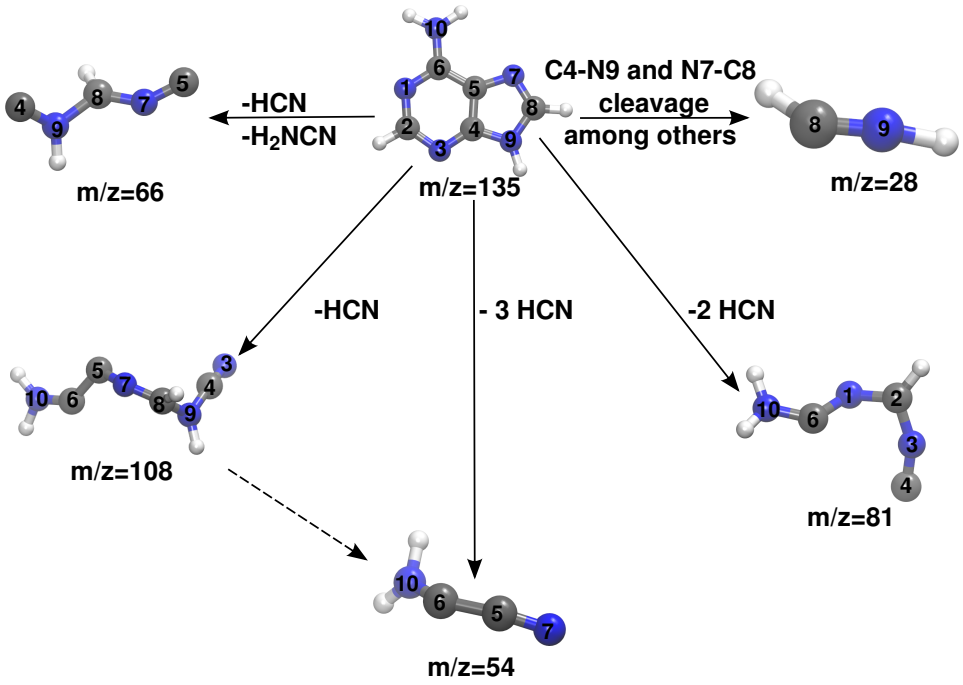

Figure 6.4.: One facet of the $\mathbf{A}^+$ fragmentation network including exemplary average fragment geometries on the OM2-D3 PES. The dashed arrow from m/z 108 to m/z 54 represents a very low number of runs.

Having indicated possible weak points of our novel method, we now compare our results (from the trajectories on the OM2-D3 PES) to those obtained from isotope-labeling EI-MS experiments and from static QC (DFT) methods conducted previously by other working groups. From the former type of study it was concluded that the first HCN expulsion from $\mathbf{A}^+$, leading to $C_4H_4N_4^+$ (m/z 108) involves principally H–C2–N1,[236,260] while C8 was mostly retained in this fragment.[258,259] This indeed concurs with our results (see structure of ion at m/z 108 in Figure 6.4, whose connectivity is consistent throughout our calculations). However, the absolute count of $C_4H_4N_4^+$ resulting from the here reported simulations (OM2-D3 PES) is only five and therefore, this positive finding is to be taken with caution. More importantly, McCloskey and co-workers stated based on their experiments that the fragment at m/z 81 ($C_3H_3N_3^+$) is probably not derived from the one at m/z 108 ($C_4H_4N_4^+$), and that many intermediate structures appear to be non-cyclic[236], which is in full agreement with our results, see Figure 6.4.

Our BO-QCMD results compare well to some non-dynamic DFT results by Improta et al. and Minaev et al.[200,203]. They both note an increase in bond length for C4–C5 as well as C5–C6 when comparing the optimized geometries for $\mathbf{A}$ and $\mathbf{A}^+$. These are indeed the bonds most likely to be broken in our calculations, although this finding is more pronounced for the semi-empirical OM2-D3 and DFTB3-D3 PES than for the presumably more accurate DFT-D3 PES. It is worth noting that our findings are in disagreement with previous suggestions for

fragment structures of **A**, which depicted the 5-membered ring intact and did not discuss the possibility of early C4–C5 bond cleavage[203,267], nor the formation of linear fragments. In our view it is questionable to use 0 K equilibrium structures as models for vibronically (and probably electronically) very "hot" meta-stable intermediates occurring during EI-MS experiments. Based on our computations, we suggest that the majority of decomposition products of **A** detected in these measurements are non-cyclic. This conclusion is supported by the argument of more conformational degrees of freedom for non-cyclic structures compared to annellated ones, and therefore a favorable entropy gain for ring-opening reactions at high temperatures. The typical vibrational temperature of $\mathbf{A}^+$ when these processes occur in our simulation lies between 500 K (initial source temperature) and about 5000 K.

6.4. Conclusions

We have provided new insights into the fragmentation network for the molecular ion of the free nucleobase adenine by employing mixed quantum-classical dynamics. We were able to (i) reproduce in a basically non-empirical manner remarkably well the 70 eV EI mass spectrum of adenine, (ii) assign sensible, open-chained fragment structures likely responsible for the "daughter ion" peaks detected in experiments, and (iii) provide evidence that most of the fragmentation reactions even at low internal excess energy (<10 eV) occur very fast in the ps or even sub-ps time regime. Notably, the improvement of the quality of the underlying PES by higher level quantum chemical methods lead to better agreement of experimental and simulated spectra, which further supports the basic theoretical assumptions in the QCEIMS model. However, the efficient semi-empirical methods also yield very reasonable results and still seem to provide an overall useful accuracy level which is a prerequisite for similar studies on larger systems. Our results are partially in agreement with conclusions based on static, zero Kelvin quantum chemistry investigations but also seem to indicate that a more comprehensive and accurate picture of EI-MS requires a dynamic, "high-energy" theoretical treatment as provided by the QCEIMS procedure. Further work is in progress on the other four nucleobases in our laboratory, including the effect of different tautomers for cytosine and guanine. The pertaining results will be presented soon.

7. Automated Quantum Chemistry Based Molecular Dynamics Simulations of Electron Ionization Induced Fragmentations of the Nucleobases Uracil, Thymine, Cytosine, and Guanine

Christoph Alexander Bauer and Stefan Grimme*
Received 15th of December 2014, Published online 20th of March 2015

Reproduced (adapted) by permission of SAGE Publications, Ltd.
Christoph Alexander Bauer and Stefan Grimme Eur. J. Mass Spectrom. **2015**, *21*, 125–140

Own manuscript contribution

- Performance of all calculations
- Interpretation of data
- Writing the manuscript

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

AbstractThe gas phase decomposition pathways of the electron ionization (EI)-induced radical cations of the nucleobases uracil, thymine, cytosine, and guanine are investigated by means of mixed quantum-classical molecular dynamics. No preconceived fragmentation channels are used in the calculations. The results compare well to a plethora of experimental and theoretical data for these important biomolecules. By our combined stochastic and dynamic approach, one can access in an unbiased way energetically available decomposition mechanisms. Additionally, we are able to separate the EI mass spectra of different tautomers of of cytosine and guanine. Our method (termed previously Quantum Chemistry-Electron Ionization Mass Spectra, "QCEIMS") reproduces free nucleobase experimental mass spectra well and provides detailed mechanistic insight into high-energy unimolecular decomposition processes.

7.1. Introduction

Free radical-induced oxidative damage in the DNA has important biological consequences. DNA lesions may arise by chemical reactions of nucleobase radical cations, e.g., ring opening reactions or formation of various adducts[273]. Such damage may then either be repaired enzymatically or lead to mutations or cell death.[274,275] The source of free radicals in the cell can either be ionizing radiation or a chemical reaction, possibly subsequent to a radiation-induced event. During the last decades, DNA damage by free electrons has been reported as well.[220,255,276] These ballistic particles may destroy the chemical structure of DNA by either dissociative electron attachment (DEA) or - provided their energy is high enough - by inducing radical cations of the nucleobases in a (e,2e) process.

The latter possibility and its dissociative consequences have been studied by electron ionization mass spectrometry (EI-MS) for the free nucleobases in the gas phase.[236,257,260,277,278,279,280] Additionally, the fragmentations of nucleobase radical cations have been investigated by photionization mass spectrometry (PI-MS).[267] From these experimental data, fragmentation pathways of the nucleobases have been deduced.[257,267,277] Due to the small number of atoms in pyrimidine nucleobases, several research groups have tried to rationalize the mass spectra of uracil (unprotonated and protonated)[201,281,282,283] and cytosine[176,284] by means of static quantum chemistry (QC). There has also been an attempt to predict fragmentation channels for the radical cations of the purine and pyrimidine nucleobases by analyzing their optimized structures[200]. Cheng et al. have rationalized fragmentation channels of the guanine radical cation by collision induced dissociation (CID) measurements, combined with static QC.[205] Very recently, Minaev et al. and Dawley et al. conducted new experiments and calculations (with predefined fragmentation paths) on adenine, hypoxanthine, and guanine.[203,204]

A novel stochastic and dynamic approach based on mixed quantum-classical dynamics (QC-MD), see e.g. Refs.[13,207], dubbed Quantum Chemistry Electron Ionization Mass Spectra, QCEIMS, has been presented recently in order to calculate EI mass spectra routinely.[10]

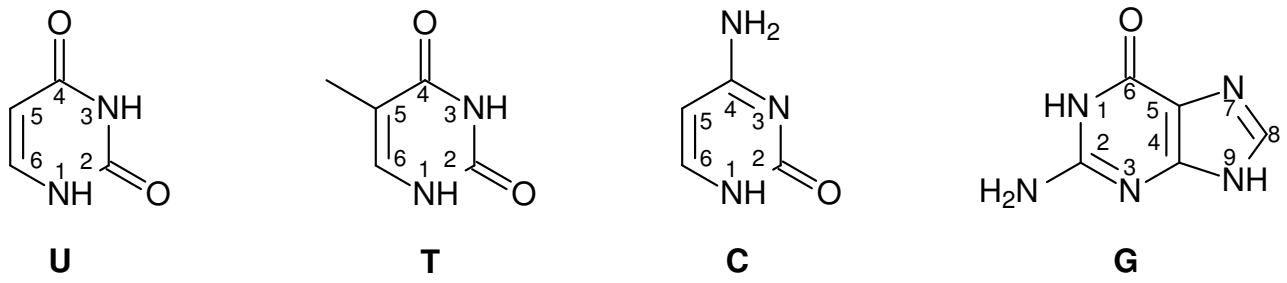

Figure 7.1.: Structures of the four nucleobases uracil (**U**), thymine (**T**), cytosine (**C**), and guanine (**G**).

The nuclei are propagated classically on a Born-Oppenheimer potential energy surface (PES) as determined by the chosen QC method. A great advantage of this dynamic method is that prior knowledge of decomposition pathways is not needed. The simulation is guided by the gradient of the PES, generated "on the fly" by a QC method, rather than by "chemical intuition". The trajectories obtained by such calculations elucidate energetically accessible paths through a vast and complicated chemical reaction space (for a related approach to form molecules theoretically, see Ref.[285]). There has been a substantial effort by the theoretical and computational chemistry community to shed light by applying QC-MD on unimolecular decomposition reactions of neutral molecules[286,287] and radical anions (for DEA[222,223]). However, these attempts have been limited to small molecules, whereas we have shown that organic molecules with about 100 atoms can be treated by QCEIMS.[233] In the present work we consider the fragmentation channels of the four nucleobases displayed in Figure 7.1 (with nominal masses between 111 and 151 atomic mass units [u]). The results obtained from our calculations are then compared to the fragmentation patterns of the nucleobases, which have been published earlier.[257,277] To the best of our knowledge, the results presented below are the first comprehensive theoretical MD-based assessment of the dissociation routes of the radical cations of the nucleobases **C**, **G**, **T**, and **U**. We have taken into account the populated low-energy gas phase tautomers of **C** and **G**, see below. For an analogous study on the nucleobase adenine see Ref[234].

7.1.1. Nucleobase Tautomerism

Although there exist numerous previous experimental[288,289,290,291,292] and computational[269,293,294,295,296] studies, we have re-computed the gas phase Boltzmann population ratio (BPR) of several tautomers of **C** and **G** at $T = 500$ K (for technical details, see below). The three tautomers of lowest energy of **C** and the four tautomers of lowest energy of **G** (structures in Figure 7.2) were considered in our simulations. Some rare high-energy gas phase tautomers of **G** have been observed experimentally[297], but their role in EI-MS measurements is likely unimportant due to their low population.

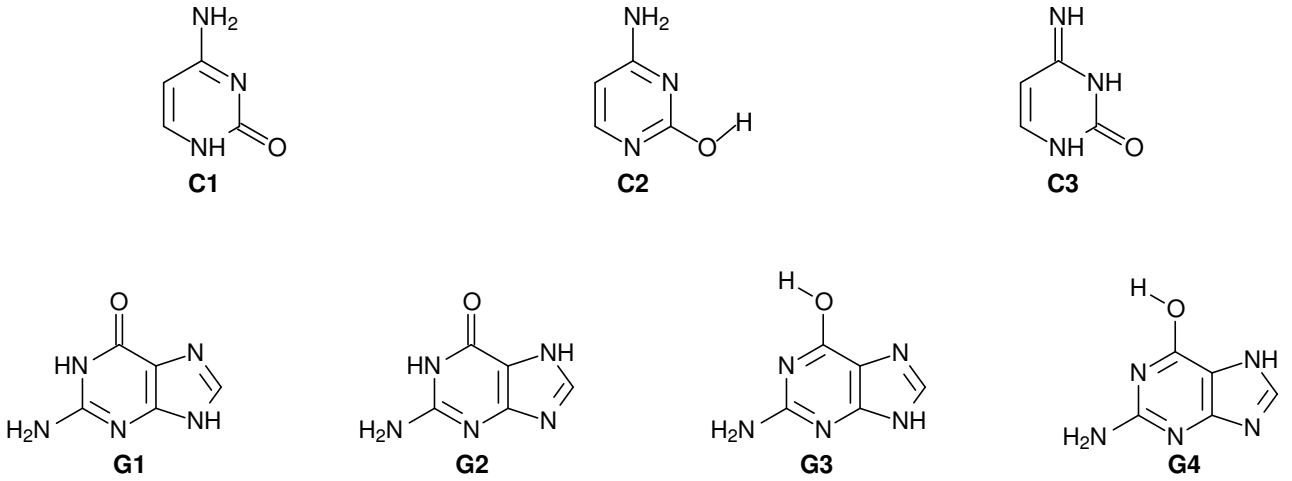

Figure 7.2.: Considered tautomers of **C** and **G**.

For cytosine, we considered the keto- (**C1**), enol- (**C2**), and imine (**C3**) tautomers. There are two rotamers of **C2**, which cannot be distinguished by EI-MS experiments. Thus, only the rotamer shown in Figure 7.2 was considered further. For guanine, we took into account the 9*H*- and 7*H*- keto (**G1** and **G2**) as well as the 9*H*- enol and 7*H*- enol isomers (**G3**/**G4**, one rotamer each as shown in Figure 7.2). While it is established that the 1*H*- keto form of cytosine (**C1**) is the prevalent form in solution[298], the enol form (**C2**) is more populated in the gas phase. The gas phase BPRs in % for the **C** tautomers are as follows: **C1:C2:C3** = 34:56:10 (at T=500 K). This compares well to previous studies, which were conducted for a temperature of 450 K.[290,293] For **G**, we computed a BPR of **G1:G2:G3:G4** = 27:47:24:2 (at T=500 K) Again, our finding largely concurs with a previous study, where the BPR was computed for T=600 K.[291] For **U** and **T** we investigate the tautomer depicted in Figure 7.1, based primarily on a comprehensive theoretical study[269], which computed other tautomers at approximately 10 kcal/mol higher in energy in agreement with experimental studies.[299,300,301] We will show below that each tautomer exhibits its specific decomposition reaction cascades.

7.2. Computational Details

Structures of **U**, **T**, and the tautomers of low energy of **C** and **G** were optimized using dispersion-corrected density functional theory (DFT) at the TPSS[60]-D3[33,34,35]/def2-TZVP[48] level as implemented in Turbomole 6.5.[62,63] The stationary point on the PES was confirmed to be a (local) minimum by calculating the harmonic vibrational frequencies. The gas phase Boltzmann distributions for $T = 500$ K were calculated according to the following procedure: CCSD(T)/CBS single-point energies were computed on the TPSS-D3/def2-TZVP optimized geometries. The complete basis set (CBS) estimate was reached by extrapolation using the MP2 correlation energy according to a scheme devised by Helgaker and co-workers, slightly

modified by Neese and Valeev.[302,303] Specifically, T→Q extrapolation was performed using the cc-pVTZ and cc-pVQZ basis sets.[304] These calculations were carried out using the ORCA program system.[70,71] Corrections to free energy G for $T = 500$ K are based on unscaled harmonic vibrational frequencies at the TPSS-D3/def2-TZVP level of theory.

The QCEIMS approach[10] was used in the default setup with 70 eV ionization energy and an initial temperature of 500 K. The number of production runs performed for each spectrum was 300 for **U**, 325 for **C**, 375 for **T**, and 400 for **G**, yielding theoretical base peaks with > 100 counts. As the technical details concerning the implementation of our algorithm as well as some more theoretical background has been provided in our previous studies[10,234], we repeat it here only briefly. First, a randomized ensemble of neutral ground state structures was obtained by sampling one QC-MD-trajectory on the semi-empirical OM2-D3[33,35,228] PES, where a Velocity-Verlet algorithm[270] was used for integration with a time step of 0.5 femtoseconds (fs). An ionization excess energy (IEE) was then assigned to each randomly chosen start structure based upon a Poisson distribution

$$P(E) = \frac{\exp[cE(1 + ln(b/cE)) - b]}{\sqrt{(aE + 1)}}, \tag{7.1}$$

where $P(E)$ is the probability to have an IEE of E in the ion, $b \approx 1$, $c = \frac{1}{aN_{\mathrm{el}}}$ with $a \approx 0.2$ eV and N_{el} the number of valence electrons. The distribution was chosen such that the maximum IEE probability is at 0.6 eV/atom. The PES for the fragmentation runs was computed on the fly on two different semiempirical levels of theory: (i) OM2-D3 (ii) DFTB3-D3[136,229,230]. The molecule was assumed to be instantaneously ionized formally to the ground state of $\mathbf{M}^{\bullet+}$. The time constant τ_{h} over which internal conversion was simulated varied between 0.2 and 2 ps, as approximated by an exponential energy gap law[271]. Because of the heating process (transfer of electronic to internal vibrational energy), the total energy was only formally but not numerically preserved in the first part of the MD run. The maximum number of cascading runs in order to track down secondary, tertiary etc. fragmentations was automatically determined to be three. Unrestricted SCF calculations in conjunction with a fractional occupation number (Fermi 'smearing') technique[27,44,45] were used throughout to ensure SCF convergence. Fragmentation events were automatically detected by QCEIMS, and statistical fragment charges on average fragment structures (last 100 steps of a given trajectory) were assigned according to the Boltzmann factor based on differences of computed ionization potentials (IPs), $\exp[\Delta\mathrm{IP}/(\mathrm{k}T)]$. The maximum simulation time for the initial trajectories was set to 5 ps and was reduced for secondary and tertiary fragmentation runs. Finally, all fragments were counted according to their statistical weight, yielding the calculated EI mass spectra. For OM2 calculations, the MNDO program[245] was used, DFTB3 calculations were carried out by `dftb+`[246], and the necessary DFT calculations (see Refs.[10,234] for details)

were performed by ORCA.[70,71]

Experimental spectra for comparison were taken from the NIST mass spectral database freely available on-line[243]. We also used the experimental EI mass spectra of Rice[257,277] for comparison. A standard mass spectral matching score[9] was employed to examine the quality of the computed mass spectra. The score measures the overlap between the computed and experimental spectra. A value of zero means that the two compared spectra have nothing in common (are orthogonal) and the maximum value of 1000 refers to two identical spectra. For details, see Ref.[9] and the supporting information of our previous publication.[234]

7.3. Results and Discussion

7.3.1. Evaluation of Computed Mass Spectra

Since the aim is to gain insight into the fragmentation paths of the nucleobases, we have chosen the level of theory that has given the best agreement with the experiment based on the match score. For **C** and **G**, the computed spectra for the individual tautomers and their scores are reported in the respective subsections. Table 7.1 shows that the simulations on the OM2-D3 level of theory lead to the best computed spectra for **U**, **T** and **C**, while for **G** the DFTB3-D3 simulations gave the best results.

Table 7.1.: Mass spectral scores for the given levels of theory. The marked entries were the basis of the computational fragmentation paths given in the following sections.

molecule	OM2-D3	DFTB3-D3
U	691	574
T	523	412
C	681	636
G	541	595

The first main result of this work is that all major experimentally observed peaks are found in the computed spectra of **C**, **G**, **T** and **U**, hence allowing us to analyze the MD simulations and extract the fragmentation pathways from them. The spectra are shown below in the subsections on the individual molecules. Since score differences of ≈ 50 points signify a visible difference in the quality of a computed spectrum, we chose to analyze the OM2-D3 simulations in the cases of **U**, **T** and **C**, and the DFTB3-D3 simulations for **G**.

In some instances substantial differences appear in different experimental EI-MS spectra for one molecule with the same ionization energy of 70 eV and even the base peak may be different. For example, Rice reports the m/z 55 peak as the main peak in the spectrum of thymine at 70 eV,[277] whereas the spectrum for thymine from the NIST database has the molecular ion (m/z 126) as the main peak. Therefore, we primarily focus on whether a

peak appears in the simulated spectrum and in the experimental one with an appreciable intensity and then proceed to assign it by analyzing the corresponding MD trajectories. This experimental uncertainty and its possible reasons as discussed previously[10] should be kept in mind when the accuracy of our theory is judged in comparison to experiment.

7.3.2. Main Fragmentation Pathways

In the following four subsections, the computed fragmentation paths are discussed for the four nucleobases in detail. We start with the pyrimidine bases **U**, **T**, and **C**. The findings presented below were obtained by inspection of the QCEIMS production runs and analysis of the fragment ion counts. A detailed analysis concerning important internal coordinates linked to bond cleavages, was also performed. To this end, the first of the cascading fragmentation trajectories, comprising the initial time frame until the first recorded fragmentation event (ranging from a few fs to 5 ps), were investigated. We focus in the discussion on three main points: (i) the comparison between calculated and experimental EI mass spectra, (ii) the MD analysis with respect to bond ruptures along the pyrimidine and purine ring systems in order to unveil significant contributions to decomposition reaction coordinates, and (iii) the display of the principal dissociation routes as given by the simulations on the chosen PES.

We then compare with the decomposition routes of the molecular ions of **U**, **T**, **C**, and **G** as they were discussed in the experimental literature[257,267,277] and discuss the agreements and differences. Selected fragmentation trajectory videos are available at our web-site, *vide infra*, and the pertaining reactions are indicated below.

Uracil

Figure 7.3 demonstrates the large overlap between calculated (OM2-D3) and experimental EI mass spectra of **U**. The base peak in the calculated spectrum is m/z 69, whereas in the experimental one, $\mathbf{M}^{\bullet+}$ forms the base peak. This error in spectral prediction is most probably related to the OM2-D3 PES, and indicates that our method produces slightly more fragmentations in the standard setup compared to the experiment.

32 % of the MD production runs show no fragmentation of $\mathbf{U}^{\bullet+}$. This accounts for the large signal by the molecular ion $\mathbf{M}^{\bullet+}$, m/z 112, found in both simulation and experiment. More than half of the runs show ejection of isocyanic acid (HNCO), a neutral loss of 43 u in a retro Diels-Alder (rDA) reaction[305,306], as the first fragmentation step. Figure 7.4 a-f shows the time series of bond distances along the pyrimidine ring of $\mathbf{U}^{\bullet+}$. The curves oscillate around the original bond lengths (circa 1.4 Å) when no cleavages happen. The pertaining runs translate to $\mathbf{M}^{\bullet+}$ in the computed spectrum. When a bond is broken, the distance between the two atoms grows rapidly. The crucial coordinates for the rDA reaction, namely, the N1–C2 (Figure 7.4 a) and N3–C4 (Figure 7.4 c) bonds, break nearly simultaneously in many runs.

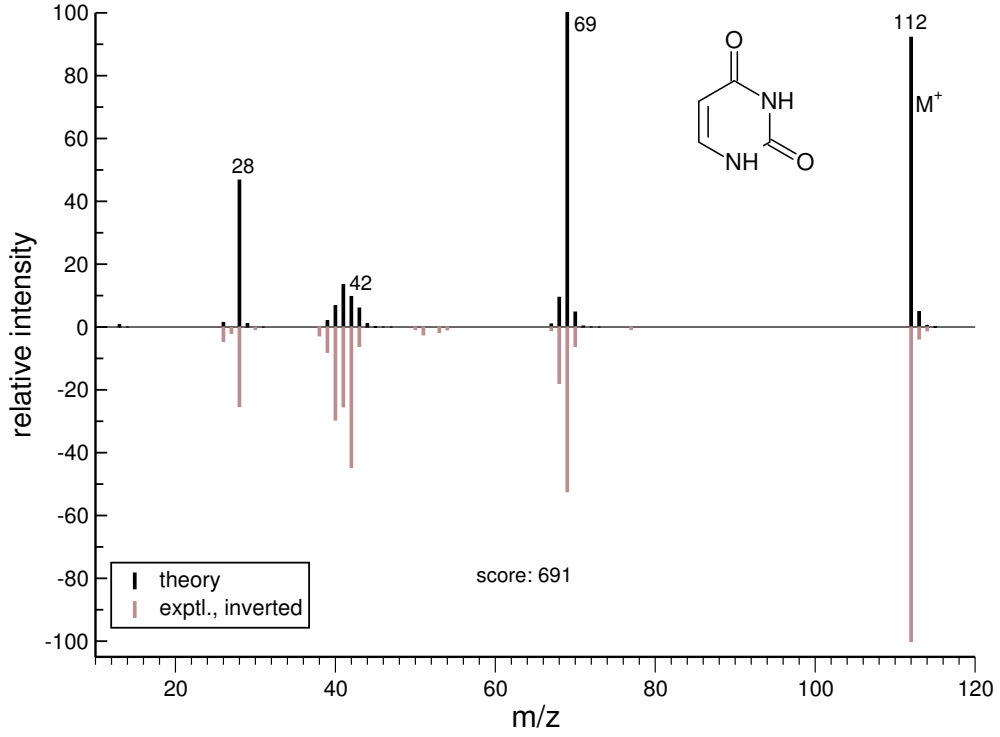

Figure 7.3.: Calculated (OM2-D3) vs. experimental EI mass spectrum of **U**. The peaks labeled by their m/z values are discussed in the text.

While there are a few outliers, we conclude that the rDA reaction happens essentially as a concerted process. The excision of HCNO at these positions is in accord with the experimental results[277]. The other bonds along the ring perimeter of **U** break less often (Figure 7.4 b, d, e, f). Another important finding is that many primary fragmentation events happen within the first ps, underpinning our assumption that longer simulation times are not necessary.

Figure 7.5 shows the main fragmentation paths of $\mathbf{U}^{\bullet+}$ with probabilities as extracted from the OM2-D3 fragmentation trajectories. The rDA reaction yields the radical cation m/z 69. 35 % of all runs lead to this signal (cf. a supporting example trajectory video available for download at our website). When a secondary fragmentation step happens, the two main neutral losses are $HCCO^{\bullet}$ (41 u) and hydrogen isocyanide (HNC, 27 u). The former process yields the HNCN cation, m/z 28, the latter the ketene radical cation $H_2C{=}C{=}O^{\bullet+}$, m/z 42. According to the experimental spectrum, both processes appear to have about the same probability, whereas in our simulations the m/z 28 signal is significantly stronger (16 % of all runs vs. 3 % of all runs). There could also be a loss of carbon monoxide (CO, 28 u) as a second decomposition step after the rDA-reaction, giving $HC{-}C{=}NH^{\bullet+}$, m/z 41. This fragmentation path is recorded in 4 % of all QCEIMS production runs, which leads to a less intense computed signal compared to the experimental spectrum. Other peaks such as m/z 68 or m/z 40 are accounted for by loss of an H atom from the ions m/z 69 and m/z 41 displayed in Figure

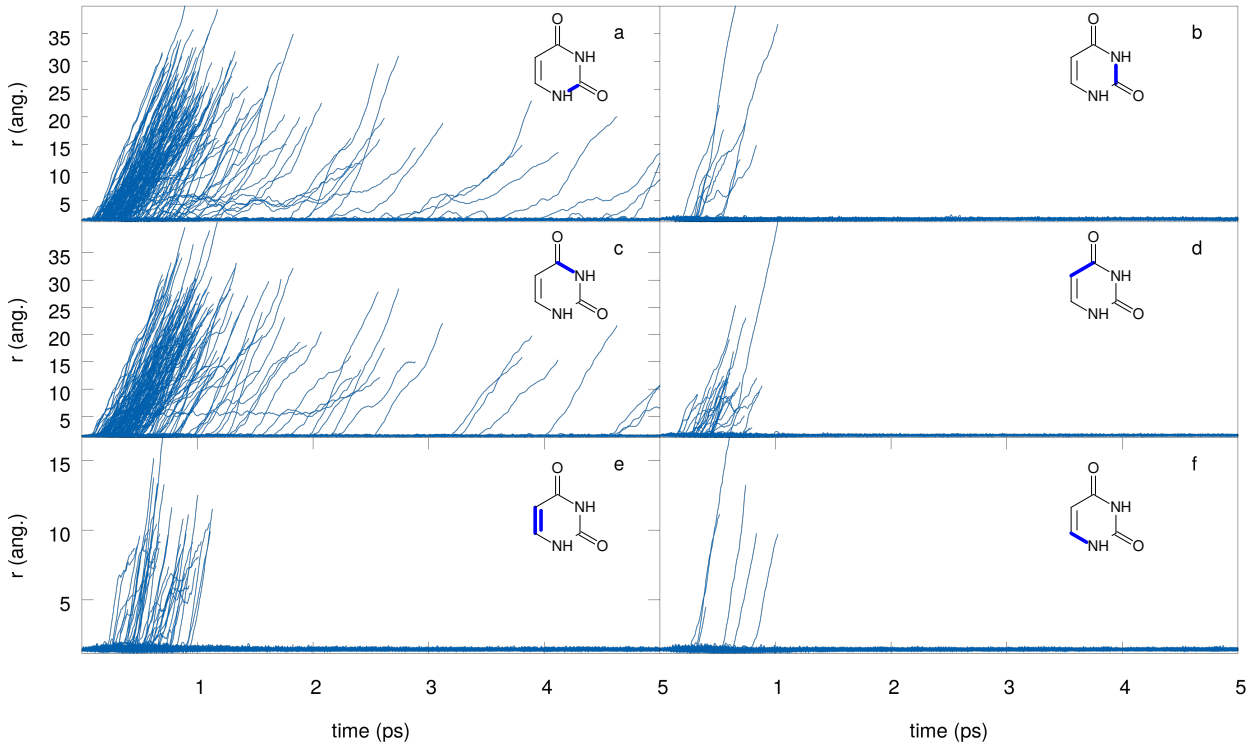

Figure 7.4.: Inter-atomic distances (in Å) of the indicated (**bold**) connections of $\mathbf{U}^{\bullet+}$ during 300 trajectories of the OM2-D3 production runs. Abrupt ending of a trajectory line indicates a recorded fragmentation event.

7.5. These are found among the results contained in the computed EI mass spectrum (see Figure 7.3). To summarize this subsection, our fragmentation scheme for $\mathbf{U}^{\bullet+}$ matches very well the models built on the experimental data[277]. We have been able to assign probabilities for specific fragmentation paths. While these do not match the experimental spectra perfectly, we submit that the qualitative agreement from an unbiased computation without defining any reaction coordinates beforehand is a very encouraging result. Note that no molecule-specific adjustments to our method have been made. We now proceed to **T**, which differs from **U** only in the substituent at the C5 position (methyl for **T**, hydrogen for **U**), and scrutinize what effect this difference has on the dissociation cascades.

Thymine

Figure 7.6 shows that the majority of intense signals of the EI mass spectrum of **T** is reproduced by our simulation. The base peak belongs to $\mathbf{M}^{\bullet+}$, m/z 126. The relative rate constants for the processes leading to the ions m/z 83, m/z 55, and m/z 28, respectively, are wrong, which is most probably caused by the insufficiency of the OM2-D3 PES. However, we analyze here the fragmentation paths of $\mathbf{T}^{\bullet+}$ on this PES to allow comparison with those of $\mathbf{U}^{\bullet+}$.

The relative number of simulation runs that do not show any fragmentation of the thymine radical cation is 18 %. This compares well to the spectrum reported by Rice[277], but not so

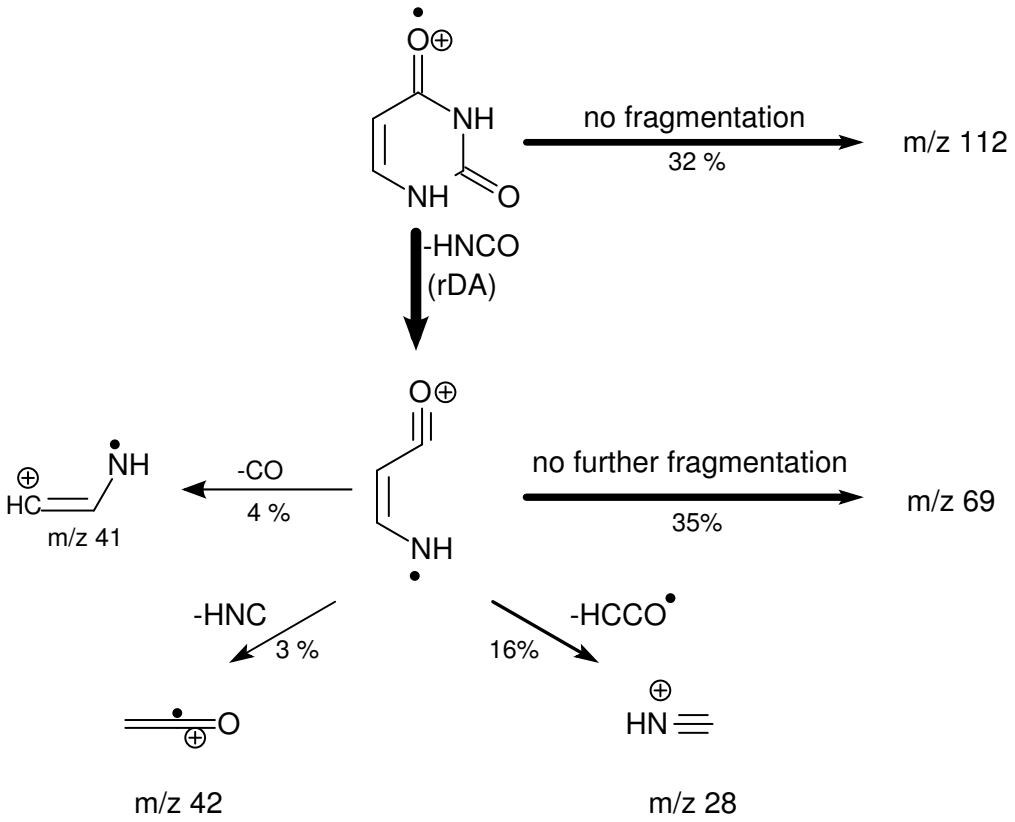

Figure 7.5.: Main fragmentation paths of $\mathbf{U}^{\bullet+}$ as given by our OM2-D3 simulations. The thicknesses of the arrows have been scaled in order to reflect the probabilities.

well to the one taken from the NIST database, where the molecular ion peak is the main one. The first step in the decay of $\mathbf{T}^{\bullet+}$ is the same as in $\mathbf{U}^{\bullet+}$. In a rDA-reaction, HNCO is lost and an ion m/z 83 is produced (see also Figure 7.7). This peak is very strong in the computed spectrum, which indicates that the OM2-D3 method provides quantitatively wrong barrier heights, facilitating the elimination of HNCO from $\mathbf{T}^{\bullet+}$ and hindering subsequent fragmentations of the ion yielding the m/z 83 signal. The next fragmentation step is very different from the one in uracil. $\mathbf{U}^{\bullet+}$ eliminates either HNC or HC=C=CO after the rDA reaction, whereas $\mathbf{T}^{\bullet+}$ eliminates CO giving rise to a signal at m/z 55 ($H_3C{-}C^{+}{=}CH{-}N^{\bullet}H$, 10 % of all runs). In order to cleave off H–C6–N1 or H–N1–C6, a hydrogen atom transfer has to take place, yielding $H_3CC^{\bullet}C{\equiv}O^{+}$, m/z 56, which happens in 2 % of all runs. The peak at m/z 28 corresponds to the HNCN ion upon a neutral loss of the $H_3CC^{\bullet}{=}C{=}O$ radical (27 % of all runs). This peak is again too large in the computed mass spectrum.

Some of the (m/z)-1 peaks in the mass spectrum of **T** originate by loss of a single H atom, and the most prominent example is the fragment m/z 54, displayed in Figure 7.7. We note at this point that the hydrogen positions in the fragment structures may be shifted. Furthermore, our calculations predict that H atom loss may happen at any point in a given fragmentation path. In this way, Figure 7.7 is somewhat incomplete, but for clarity's sake, only the most

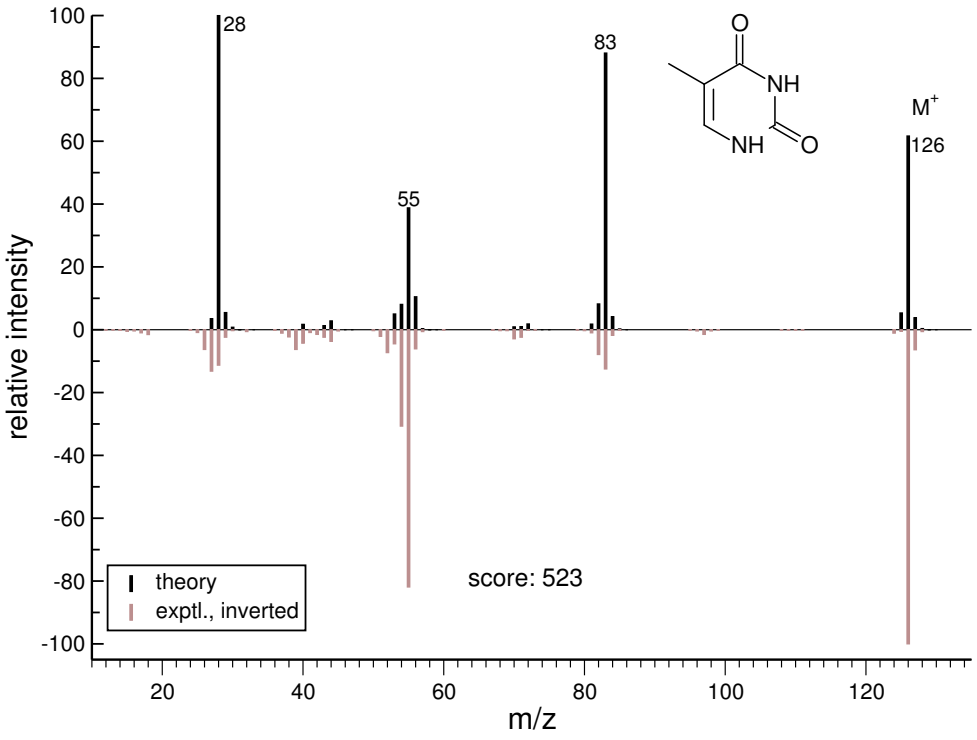

Figure 7.6.: Calculated (OM2-D3) vs. experimental EI mass spectrum of **T**. The peaks labeled by their m/z values are discussed in the text.

probable few sequences of reactions are displayed.

The similarities and differences between $\mathbf{U}^{\bullet+}$ and $\mathbf{T}^{\bullet+}$ are compiled in the histogram in Figure 7.8. The relative percentage of broken bonds reflects the relative rate constants for bond cleavage events. While the rDA reaction, by the excision of HNCO, is computed to happen in the same place (between the N1–C2 and N3–C4 bonds), the C4–C5 and C5–C6 cleavages are more pronounced in the simulations of $\mathbf{T}^{\bullet+}$. The loss of CO is more likely to happen in $\mathbf{T}^{\bullet+}$ than in $\mathbf{U}^{\bullet+}$. For the C2–N3 and C6–N1 bonds, there are similarly few bond splittings in $\mathbf{T}^{\bullet+}$ and $\mathbf{U}^{\bullet+}$. The overall picture compares well to experimental results[277] concerning the actual reaction cascades (albeit not in a quantitative way). The rDA reactions for $\mathbf{U}^{\bullet+}$ and $\mathbf{T}^{\bullet+}$ are captured, and further fragmentations, which are specific for $\mathbf{T}^{\bullet+}$, are found by analysis of the fragmentations.

Cytosine

By combining the calculated EI mass spectra of **C1**, **C2** and **C3** according to their Boltzmann factors, the spectral match score is significantly increased, as seen in Figure 7.9. This is a strong indication that our simulations can distinguish between the contributions of different tautomers to an EI mass spectrum. This is a very important result. In all three calculated single-tautomer spectra, $\mathbf{M}^{\bullet+}$, m/z 111, is the base peak. This hints at a similar stability

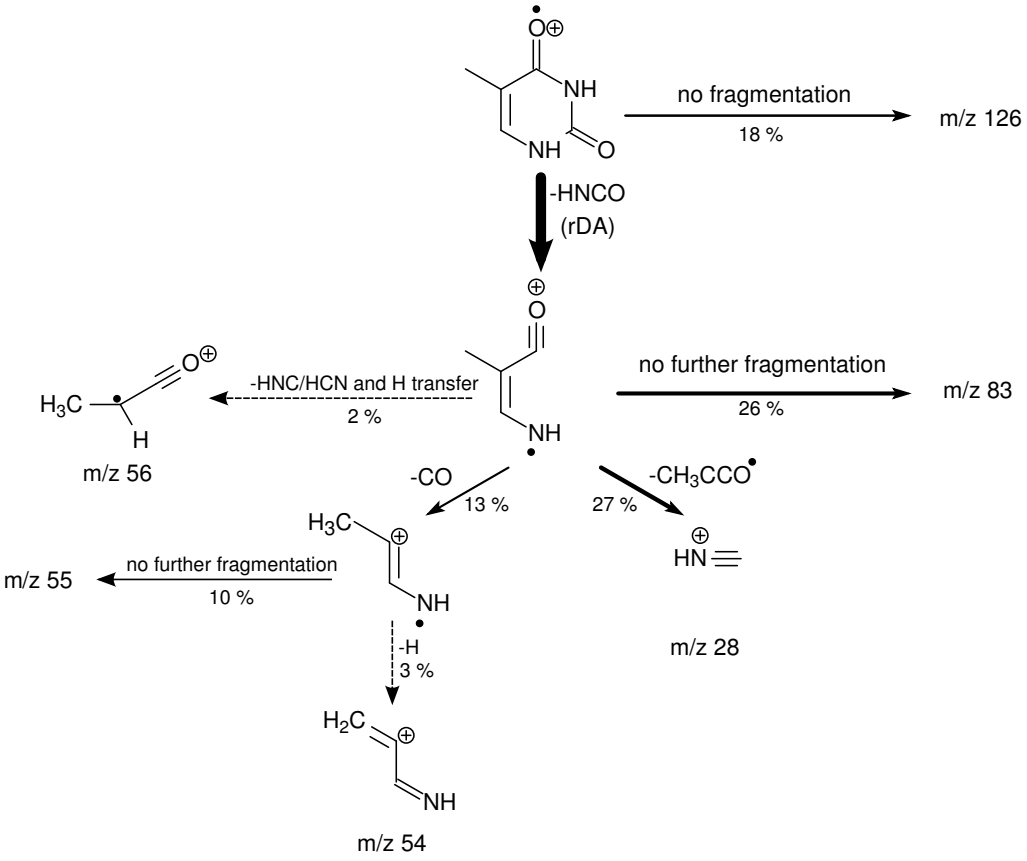

Figure 7.7.: Main fragmentation paths of $\mathbf{T}^{\bullet+}$ as given by our OM2-D3 dynamics simulations. The thicknesses of the arrows have been scaled to reflect the probabilities.

of $\mathbf{M}^{\bullet+}$ regardless of the tautomer. Large differences in calculated relative peak intensities are observed for m/z 68 and m/z 28. In order to elucidate the mechanistic reasons for these differences, we have again analyzed the fragmentation trajectories for the different tautomers of $\mathbf{C}^{\bullet+}$.

Figure 7.10 illustrates the tremendous effect of different protonation states on the fragmentation dynamics. The N1–C2 and N3–C4 bonds along the pyrimidine frame of $\mathbf{C}^{\bullet+}$ behave strikingly differently. In the case of the keto tautomer ($\mathbf{C1}^{\bullet+}$), the former breaks in almost all runs (see Figure 7.10 a) whereas in the enol tautomer it breaks much less often (Figure 7.10 c). The concerted splitting of N1–C2 and N3–C4 is the main part of the reaction coordinate of the rDA reaction, which is most likely in the imine tautomer ($\mathbf{C3}^{\bullet+}$), leading to the strong m/z 68 signal in its computed spectrum (Figure 7.9 c). This tautomer's gas phase population at 500 K is only 10 %, which explains the moderate relative intensity of the m/z 68 peak in both the tautomer averaged computational spectrum and the experimental spectrum.

By evaluating the bond lengths at the very end of each fragmentation trajectory one can measure how many of them were broken over the whole ensemble. Figure 7.11 quantifies the strong dependence on the initial tautomer state. This concerns especially the bonds where the

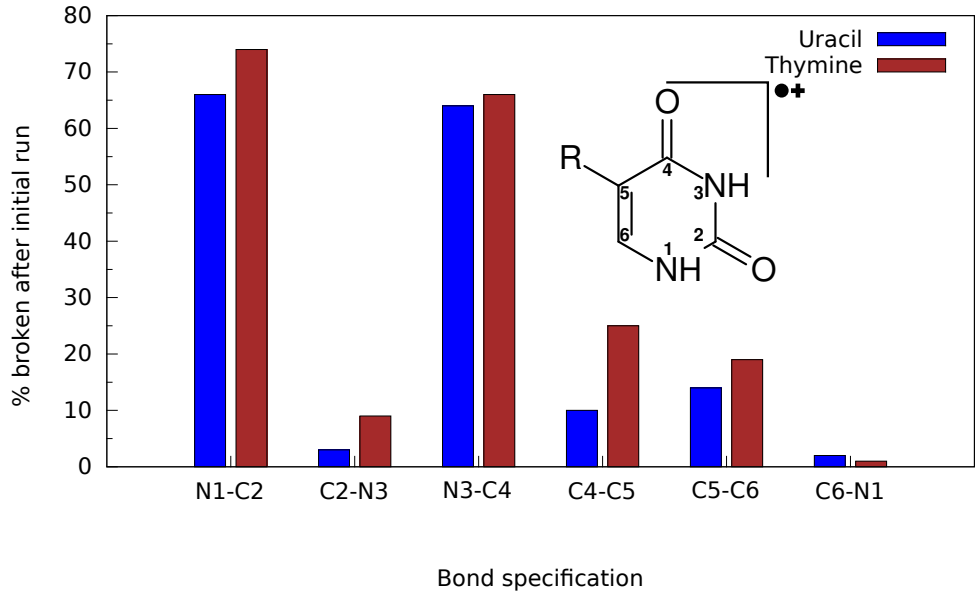

Figure 7.8.: Percentage of broken bonds (OM2-D3, interatomic distance ≥ 2 Å) along the ring perimeter of $\mathbf{U}^{\bullet+}$ and $\mathbf{T}^{\bullet+}$ in the MD runs. For $\mathbf{U}^{\bullet+}$, R=H, and for $\mathbf{T}^{\bullet+}$, R=CH_3.

prototropic shifts take place. The N1–C2 bond is weakened in the keto tautomer ($\mathbf{C1}^{\bullet+}$) and in the imino tautomer ($\mathbf{C3}^{\bullet+}$) whereas it breaks less often in the enol form ($\mathbf{C2}^{\bullet+}$). Similarly, the N3–C4 bond splits more easily in $\mathbf{C3}^{\bullet+}$. As already indicated above, this leads to a preferred rDA fragmentation path for $\mathbf{C3}^{\bullet+}$.

Interestingly, the N1–C2 bond breaks in approximately 90 % of all fragmentation runs of $\mathbf{C1}^{\bullet+}$. In contrast, Figure 7.9 a shows that the molecular ion $\mathbf{M}^{\bullet+}$ is the computed base peak for **C1**. This can be explained by the persistence of the open-chained molecular ion, which is counted after the maximum simulation time of 5 ps has been reached. See also Figure 7.10 a, where oscillations around 5 Å are seen up to 5 ps simulated time. The bonds C2–N3, C4–C5, C5–C6, and N1–C6 are not significantly affected by the initial tautomer state.

The principal fragmentation paths for the three tautomers investigated are displayed in Figure 7.12. The majority of runs yield the molecular ion $\mathbf{M}^{\bullet+}$ at m/z 111. For **C1**, 35 % of all runs end as an open-chained isomer of $\mathbf{M}^{\bullet+}$, whereas its cyclic counterpart survives in only 7 %. In the cases of $\mathbf{C2}^{\bullet+}$ and $\mathbf{C3}^{\bullet+}$, the cyclic $\mathbf{M}^{\bullet+}$ is produced in 31 % and 35 % of all runs, respectively. Figure 7.12 does not include the three high-energy reactions loss of $^{\bullet}NH_2$ (16 u), and the elimination of NH_3 and of $^{\bullet}OH$ (both 17 u). The number of counts for these processes are too small to allow a reliable quantification. The high-energy NH_2 (16 u, leading to an ion at m/z 95), and NH_3/OH (both nominally 17 u) expulsions leading to ions at m/z 94 can only be distinguished by high-resolution experiments due to differences in exact mass. The OH loss occurs from the enol tautomer $\mathbf{C2}^{\bullet+}$ and there are two pathways that lead to ions at m/z 83. One is the loss of carbon monoxide (CO, 28 u), the second one is the loss of one H atom and

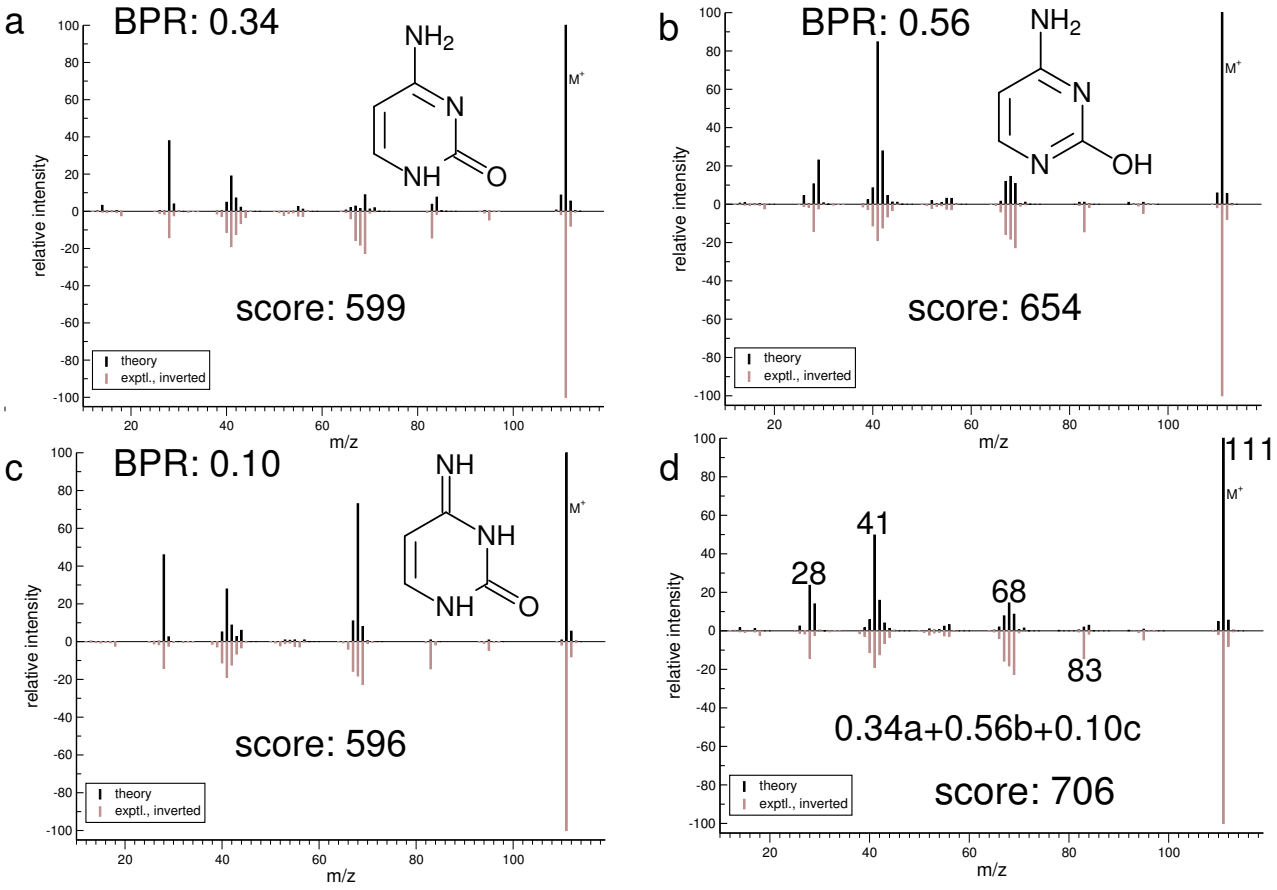

Figure 7.9.: Calculated (OM2-D3) vs. experimental mass spectra of **C1** (a), **C2** (b), and **C3** (c), respectively, and of the Boltzmann weighted sum (d). The peaks labeled by their m/z values are discussed in the text.

hydrogen cyanide (HCN, 27 u), again to be distinguished experimentally by high-resolution spectra. Interestingly, the pathway more likely to produce the m/z 83 computed signal is the latter one.

The loss of $^{\bullet}$NCO at positions N3–C2–O, occurs in 2 % of all simulations from $\mathbf{C1}^{\bullet+}$, but in virtually none from $\mathbf{C2}^{\bullet+}$ and $\mathbf{C3}^{\bullet+}$. Therefore, the ion m/z 69 originates mostly from fragmentations of $\mathbf{C1}^{\bullet+}$. The connectivity of this $C_3H_5N_2^+$ fragment ion is displayed in Figure 7.12. It shows the shift of a H-atom from the amino group to N1, which is mediated by the open-chained intermediate that is formed when the N1–C2 bond is split. This is a rather complex series of events that involves a rearrangement which is difficult to guess by chemical intuition and a good example for a mechanistic detail that warrants further investigation (visualization of a representative fragmentation trajectory is available at our website).

The rDA reaction, leading to the ion m/z 68 by expulsion of HNCO, is the second most probable fragmentation pathway for $\mathbf{C3}^{\bullet+}$, whereas in $\mathbf{C1}^{\bullet+}$ and $\mathbf{C2}^{\bullet+}$, it is a minor route. We observe the corresponding H-atom migration in the fragmentation dynamics of $\mathbf{C1}^{\bullet+}/\mathbf{C2}^{\bullet+}$, too. However, this is an energetically unfavorable and in most cases, competing pathways are

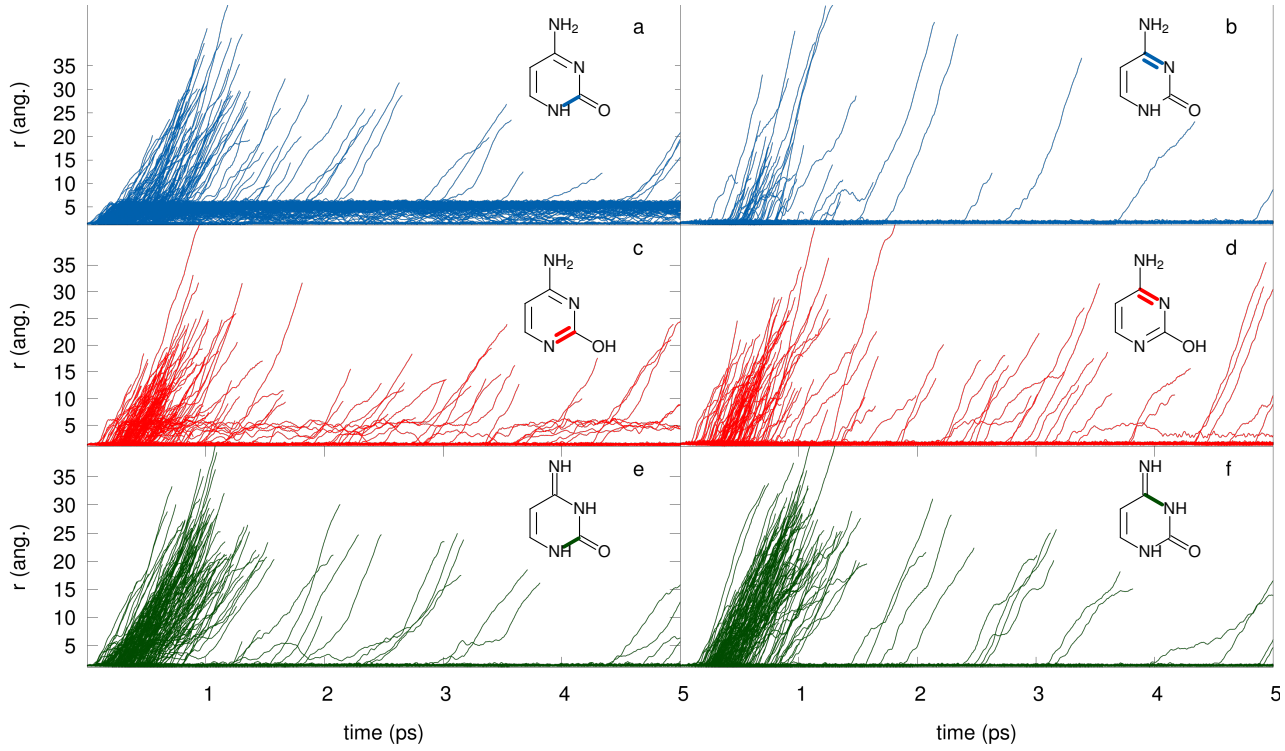

Figure 7.10.: Interatomic distances N1–C2 and N3–C4 (**bold bonds**) of $\mathbf{C1}^{\bullet+}$ (a,b), $\mathbf{C2}^{\bullet+}$ (c,d), and $\mathbf{C3}^{\bullet+}$(e,f) in Å during 325 trajectories of the OM2-D3 production runs. Abrupt ending of a line signifies a recorded fragmentation event.

entered.

One of the most elusive reaction products is $H_2C{-}C{=}NH^{\bullet+}$, m/z 41, which is formed through various pathways starting from all three tautomers. Its formation is most probable starting from $\mathbf{C2}^{\bullet+}$, where two neutral molecules, HOCN and HCN, are eliminated. Various isomers can occur by hydrogen shifts that differ only in the number of C–H/N–H bonds.

Lastly, the H_2CN^+ ion m/z 28 is formed predominantly from $\mathbf{C1}^{\bullet+}$ by simple N1–C6 and C5–C6 bond cleavages.

The overall fragmentation scheme with its main routes given in Figure 7.12 is more complicated than the EI mass spectrum suggests. The reactions deduced by Rice[277] are all depicted in this scheme. For example, the rDA reaction, the loss of the CNO radical, the subsequent loss of HCN (which may or may not be energetically available depending on the IEE), and many more, which are not found in the original studies, are covered by our approach. This scheme, even though the underlying QC simulation does not reproduce exactly the experimental EI mass spectrum, appears to be the best approximation yet to the true unimolecular reaction cascades of $\mathbf{C}^{\bullet+}$ after 70 eV electron ionization.

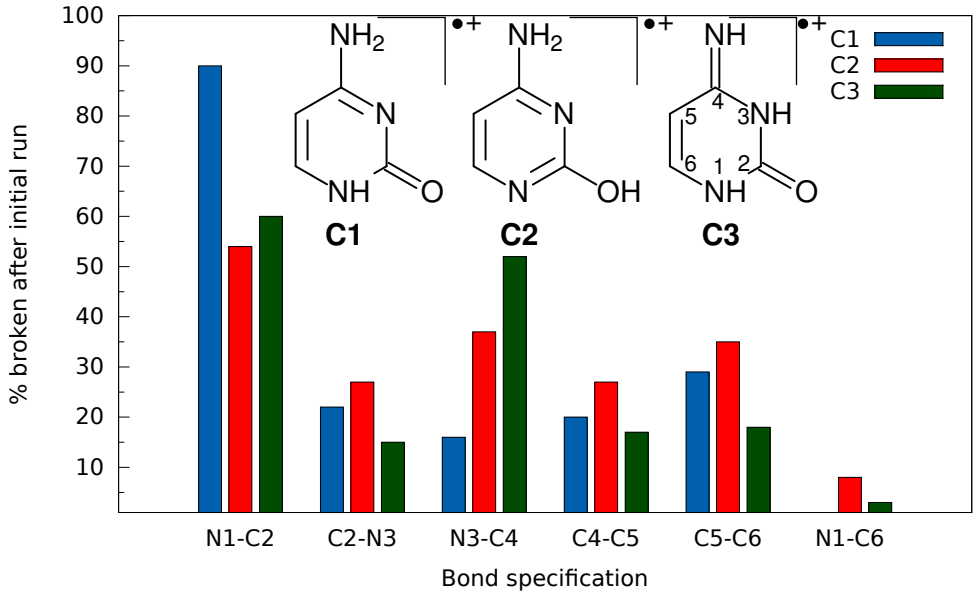

Figure 7.11.: Percentage of broken bonds ($\geq$ 2 Å) along the pyrimidine ring of $\mathbf{C}^{\bullet+}$ depending on the initial tautomer. 325 OM2-D3 trajectories were analyzed.

Guanine

The spectra of **G1**, **G2** (the keto-tautomers), **G3**, and **G4** (the enol-tautomers) are displayed in Figure 7.13 a–d. Although our simulation protocol based on the DFTB3-D3 PES cleaves too many bonds, the score of the calculated spectrum significantly improves when the individual spectra are combined with their Boltzmann weights. This points at a correct qualitative description of the different fragmentation channels of the tautomers, which we will address below. he largest difference between the $\mathbf{G1}^{\bullet+}/\mathbf{G2}^{\bullet+}$ and $\mathbf{G3}^{\bullet+}/\mathbf{G4}^{\bullet+}$ fragmentation paths is shown by the different computed relative intensities for m/z 43 (Figure 7.13 a,b/c,d). In order to further study the different dynamics associated with the tautomers, the bond distances along the purine ring were analyzed. Figure 7.14 reveals the bonds that are significantly affected thereby. The rupture of the C2-N3 and the N1-C6 bonds in $\mathbf{G1}^{\bullet+}$ and $\mathbf{G2}^{\bullet+}$ is more likely compared to the tautomers $\mathbf{G3}^{\bullet+}$ and $\mathbf{G4}^{\bullet+}$, which prefer the cleavage of the N1-C2 bond. The dissociation of the other bonds N3-C4, C4-C5, C5-C6, C5-N7, and C4-N9 depends less on the tautomer. The fragmentation paths of the 7*H*- and 9*H*- forms of $\mathbf{G}^{\bullet+}$ are rather similar. For $\mathbf{G2}^{\bullet+}$ and $\mathbf{G4}^{\bullet+}$, the N7–C8 bond breaks in 50–60 % of all runs, whereas the C8–N9 bond stays intact in the vast majority of all runs. In $\mathbf{G1}^{\bullet+}$ and $\mathbf{G3}^{\bullet+}$, the N7–C8 bond stays intact in $\approx$ 80 % of all runs, and the C8–N9 bond breaks in 20 %/35 % of all runs. We will omit further analysis of **G4** because its contribution to the EI mass spectrum is only 2 %. The main fragmentation paths of $\mathbf{G1}^{\bullet+}/\mathbf{G2}^{\bullet+}/\mathbf{G3}^{\bullet+}$ are displayed in Figure 7.15. In all three, $\mathbf{M}^{\bullet+}$, m/z 151, is the most abundant species within our 5 ps simulation time window. There are a few runs (between 1 and 2 % for each tautomer) where the elimination of the

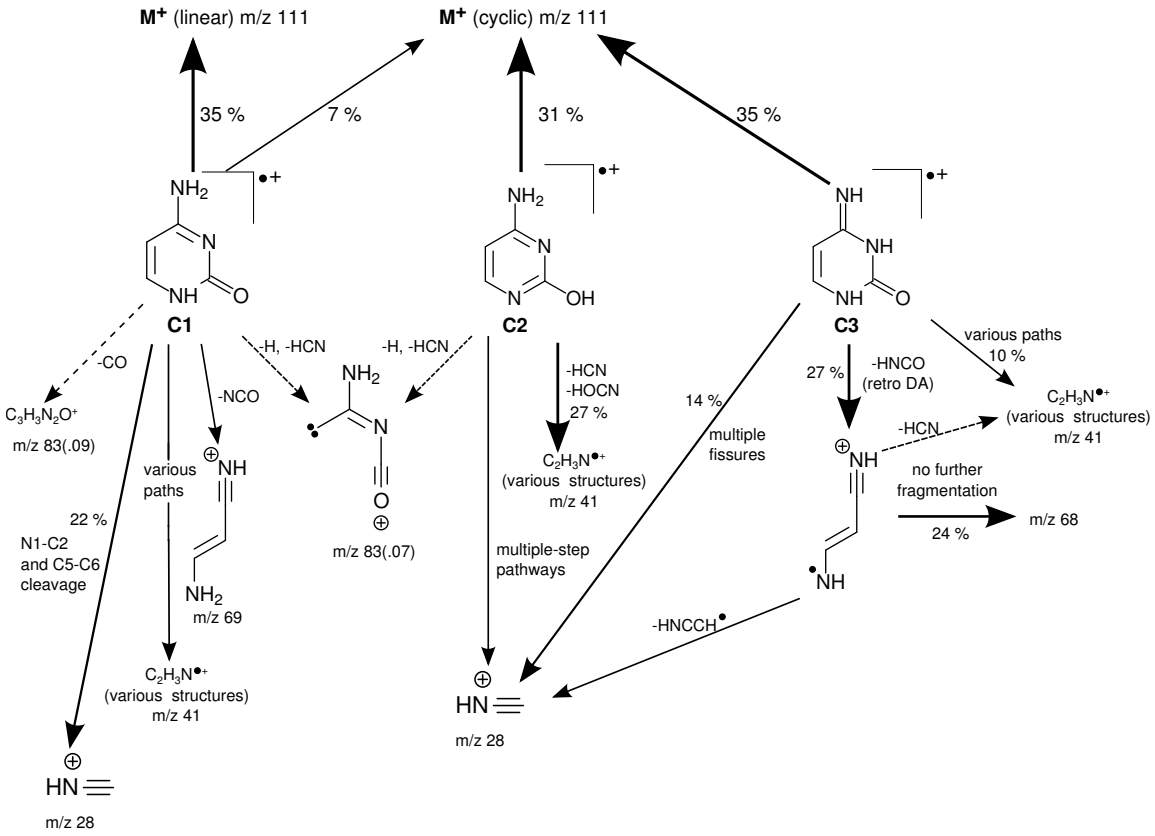

Figure 7.12.: Main fragmentation paths of $\mathbf{C1}^{\bullet+}/\mathbf{C2}^{\bullet+}/\mathbf{C3}^{\bullet+}$ as given by our OM2-D3 simulations. Each tautomer starts at 100 %, only paths with a probability $\geq 10\%$ are quantified. The thicknesses of the arrows have been scaled to reflect these probabilities.

NH_2 radical occurs, leading to ions of m/z 135. In a similarly low number of trajectories, ions m/z 134 form by expulsion of NH_3. This may happen after an H atom shift in $\mathbf{G1}^{\bullet+}$ or $\mathbf{G2}^{\bullet+}$, which we observe extremely rarely during our simulations. In the case of $\mathbf{G3}^{\bullet+}$, it is primarily the OH radical dissociation that lead to ions of m/z 134. There is a subtle difference between **G2** (the 7*H*-keto tautomer) and **G3** (the 9*H*-enol form) concerning the reaction pathway leading to $C_4H_3N_3O^{\bullet+}$ (m/z 109) radical cations. In the former, HN=C=NH is lost while in the latter, $H_2NC{\equiv}N$ is eliminated to give $C_4H_3N_3O^{\bullet+}$ in various structures, some open-chained, some with one partially intact ring system. These events are relatively rare throughout our simulations, and the computed relative intensity of the peak at m/z 109 does not match the experimental record well. Regarding the formation of ions m/z 81, we observe the largest difference between 7*H*- and 9*H*-isomers. Linear fragments $C_3H_3N_3{}^+$ (see Figure 7.15) arise by the sequential loss of HN1=C6=O and HC8–N9. The analysis of the origin of the m/z 54 signal is rather complex. There are various nearly isoenergetic pathways, all leading to radical cations $H_2C_2N_2{}^{\bullet+}$. In contrast, the pathways leading to $H_2N{-}C^+{=}NH$, m/z 43, are simple. The dissociation of this closed-shell ion proceeds along the coordinate of

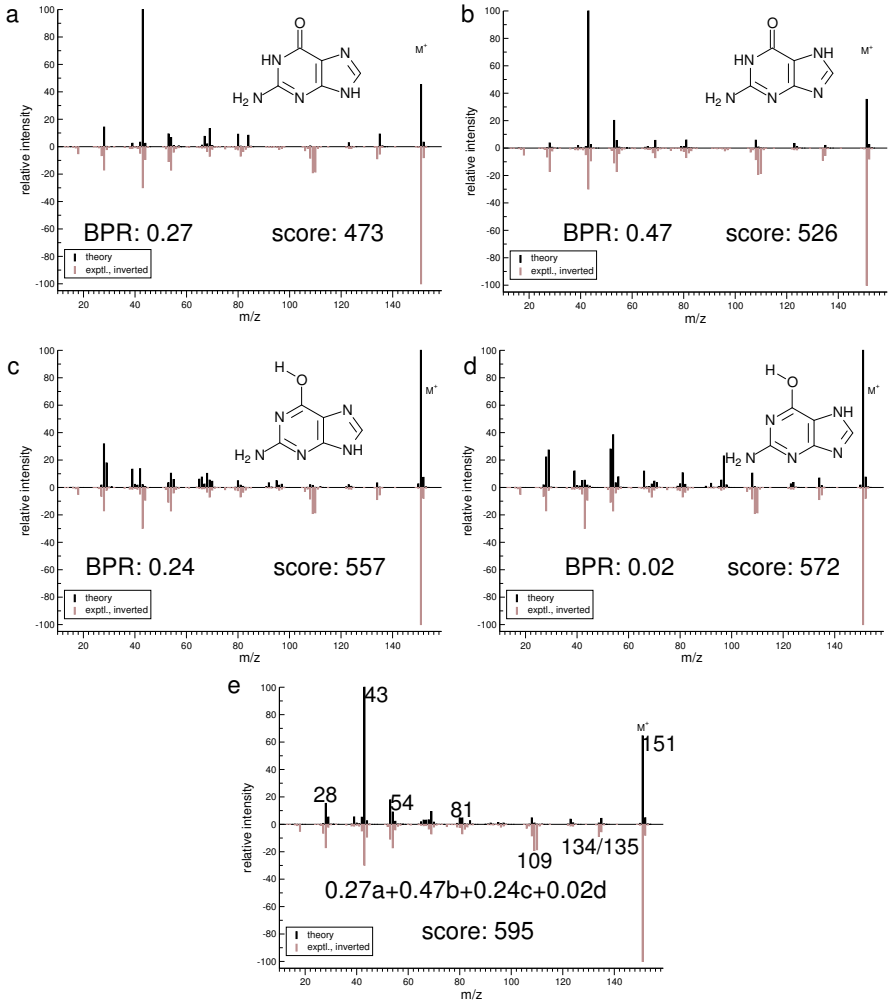

Figure 7.13.: Calculated (DFTB3-D3) vs. experimental mass spectra of **G1** (a), **G2** (b), **G3** (c), **G4** (d), and of the Boltzmann weighted sum (e). The peaks labeled by their m/z values are discussed in the text.

N1–C6 and C2–N3 bond cleavages in the keto tautomers $\mathbf{G1}^{\bullet+}$ and $\mathbf{G2}^{\bullet+}$ (and to a negligible extent, in $\mathbf{G3}^{\bullet+}$, after some unlikely H atom shifts). The aforementioned bonds are much weaker in $\mathbf{G1}^{\bullet+}$ and $\mathbf{G2}^{\bullet+}$ than in $\mathbf{G3}^{\bullet+}$. Therefore, nearly all contributions to the peak m/z 43 originate from $\mathbf{G1}^{\bullet+}/\mathbf{G2}^{\bullet+}$. Lastly, the $HC{\equiv}NH^{+}$ (m/z 28) ion is in most cases formed from HC8–N9H in $\mathbf{G1}^{\bullet+}$ and $\mathbf{G3}^{\bullet+}$, which are the 9*H*-keto and enol tautomers. In contrast, there are many rather complicated ways to form this ion from $\mathbf{G2}^{\bullet+}$ (and $\mathbf{G4}^{\bullet+}$). The larger number of $\mathbf{G3}^{\bullet+}$ runs ending with $HC{\equiv}NH^{+}$ can be explained by the lack of energetically attainable alternatives for the disintegration of this tautomer.

Rice and Dudek[257] discussed the effects of tautomerism on the fragmentation cascades of $\mathbf{G}^{\bullet+}$ in their experimental EI-MS study. They chose the well-established method of discussing the mass spectra of methyl derivatives of **G**. However, they did not discuss in detail that that $H_2N{-}C^{+}{=}NH$ (m/z 43) is formed by the C1–N6 and C2–N3 bond ruptures from $\mathbf{G2}^{\bullet+}$, which

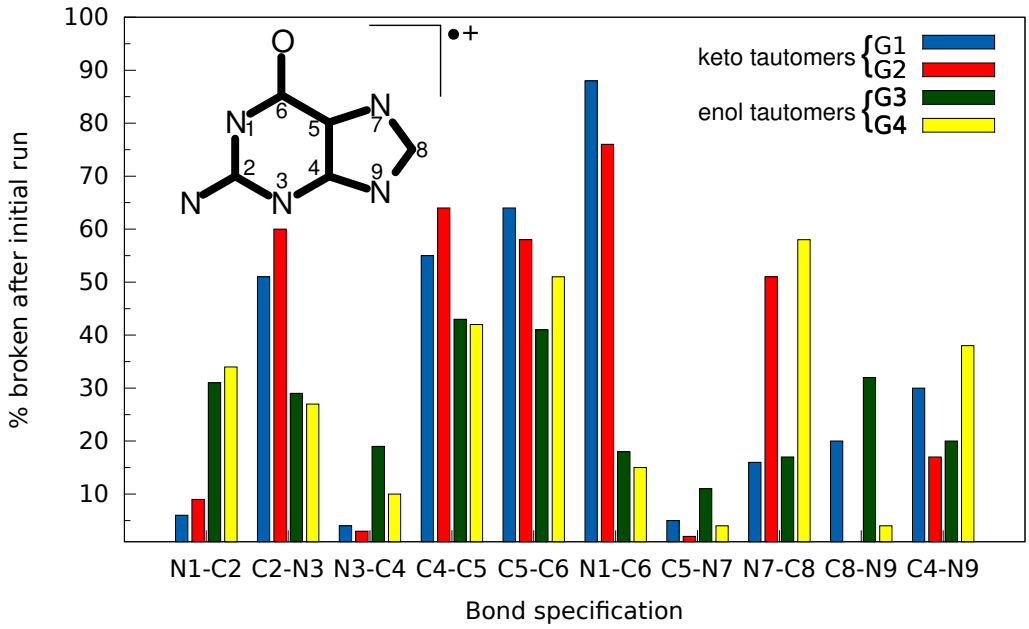

Figure 7.14.: Percentage of broken bonds ($\geq$ 2 Å) along the purine ring system of $\mathbf{G}^{\bullet+}$ during 400 DFTB3-D3 trajectories depending on the initial tautomer.

they correctly assigned as the prevailing tautomer in the vapor phase. From computational analysis by Improta et al.[200], some bond cleavages along the purine frame may be anticipated. By comparing the bond lengths in **G1** and $\mathbf{G1}^{\bullet+}$, they studied the weakening of the C2–N3 bond and the strengthening of the N3–C4 and C5–N7 bonds in this particular tautomer. This concurs nicely with our results obtained from molecular dynamics, see Figure 7.14. Main fragmentation processes can be indicated by those "standard" quantum chemical considerations. The advantage of our approach is the route to a more global picture (cf. Figure 7.15). In an unbiased way, we are able to provide an approximate, yet profoundly informative overview of the EI-induced fragmentation processes of $\mathbf{G}^{\bullet+}$, including tautomeric effects.

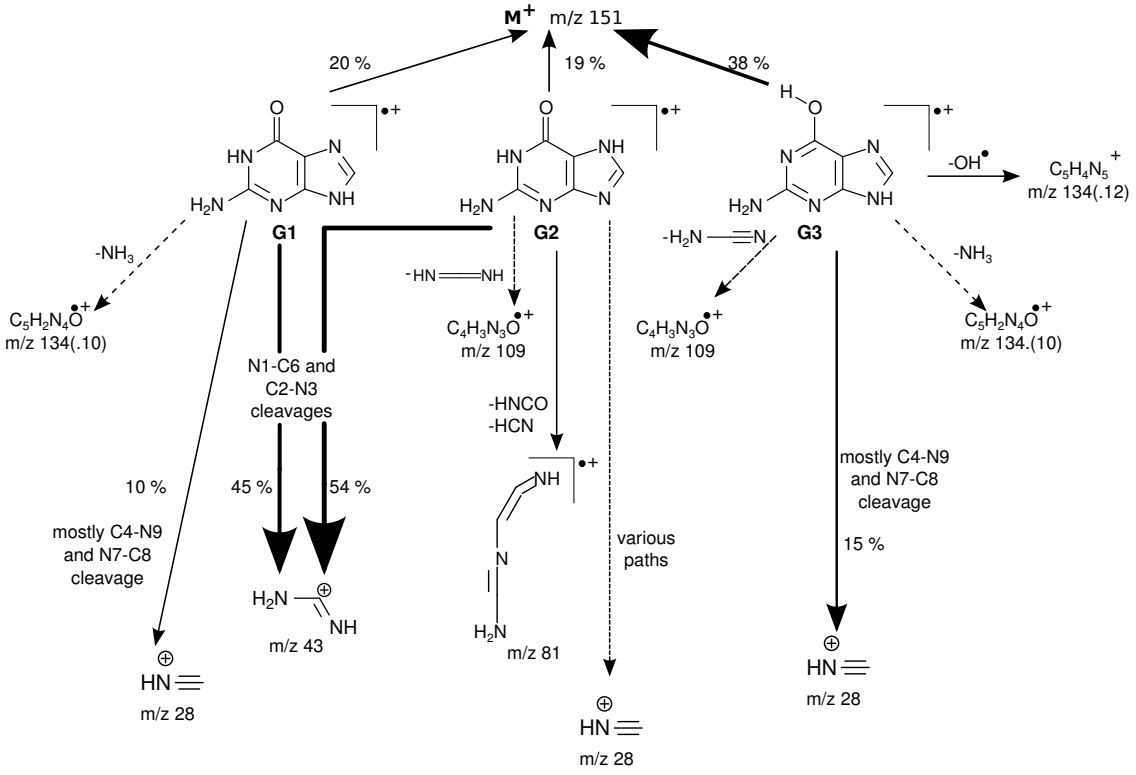

Figure 7.15.: Main fragmentation paths of $\mathbf{G1}^{\bullet+}/\mathbf{G2}^{\bullet+}/\mathbf{G3}^{\bullet+}$ as given by our DFTB3-D3 simulations. $\mathbf{G4}^{\bullet+}$ is omitted because of its low Boltzmann population. Only paths with a probability $\geq 10\%$ are quantified. The thicknesses of the arrows have been scaled to reflect the probabilities.

7.4. Conclusions and Outlook

We have presented a comprehensive theoretical study on the electron-ionization induced fragmentations of the four nucleobase molecules **U**, **T**, **C**, and **G** based on mixed quantum-classical MD. We have shown that QCEIMS produces theoretical EI mass spectra of these molecules that have a significant overlap with their experimental spectra. We have extracted the main fragmentation paths from our trajectories and shown that they concur with conclusions based on earlier studies.[257,267,277]. For **U**, rDA-reactions[306] (the loss of neutral HNCO, 43 u) are dominant in our simulations.

For **C** and **G**, there are several energetically similarly favorable tautomers in the gas phase equilibrium under typical EI-MS conditions ($T \approx 500$ K). Each isomer exhibits different fragmentation cascades upon EI. Specifically, in the imine tautomer of **C** HNCO elimination is much more likely compared to the much more populated keto and enol **C** tautomers. Similarly, in $\mathbf{G1}^{\bullet+}/\mathbf{G2}^{\bullet+}$ the N1–C6 bond is much more vulnerable compared to $\mathbf{G3}^{\bullet+}/\mathbf{G4}^{\bullet+}$. By combining computed spectra for the different tautomers of low energy of **C** and of **G** according to their Boltzmann population, the spectral overlap between the calculated and the experimental EI mass spectra is systematically improved. For many peaks with high relative intensities in the experiment, we reach nearly quantitative agreement between computed and experimental mass spectra.

The QCEIMS method with the current relatively simple (semiempirical) quantum chemical basis produces only semiquantitative agreement with experimental results for branching reactions and minor fragmentation channels. Nevertheless, we are confident that this represents important progress because the analysis of the fragmentation cascades by QCEIMS provides valuable mechanistic insight. The prevailing reaction trees gained by refining QCEIMS data may be subjected to further quantum chemical studies at a higher level of theory. Employing QCEIMS in combination with a semiempirical PES can indicate which parts of a potentially vast chemical reaction space are significant and warrant further investigation. Thereby, QCEIMS could become an efficient tool in order to pre-screen possible mechanisms pertaining to fragmentations subsequent to EI.

If the semiempirical PES cannot accurately enough describe the target system, at present a DFT-D3 PES may be used. Such significantly more expensive investigations are currently being conducted in our group. In the future, when high-level quantum chemical methods become more affordable, one may also use them more routinely as a basis for the MD computations, and we expect some increase in accuracy from such calculations in the mid-term future.

We are aiming at expanding QC(EI)MS to different ionization and fragmentation methods, especially electrospray ionization (ESI), where closed-shell protonated ions are produced and then possibly fragmented by collisions with Ar particles or the like. Additionally, we plan to include organometallic compounds, which are not be treated accurately by current semiempir-

ical QC methods. Finally, the implementation of a graphical user interface in order to increase user-friendliness of our program is a mid- to long-term goal.

8. Quantum Chemical Calculation of Electron Ionization Mass Spectra for General Organic and Inorganic Molecules

Vilhjálmur Ásgeirsson,* Christoph Alexander Bauer and Stefan Grimme†
Received 9th of February 2017, Published online 5th of May 2017

Reproduced (adapted) from
Vilhjálmur Ásgeirsson, Christoph Alexander Bauer and Stefan Grimme Chem. Sci. **2017**, *8*, 4879–4895
 DOI 10.1039/C7SC00601B

Own manuscript contribution

- Selection of compounds for EI mass spectra predictions
- Performance of EI mass spectra prediction calculations
- Discussion of the EI mass spectra predictions
- Writing the manuscript

*Faculty of Physical Sciences and Science Institute, University of Iceland, 107 Reykjavík, Iceland
†Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstraße 4, 53115 Bonn, Germany

Abstract We introduce a fully stand-alone version of the Quantum Chemistry Electron Ionization Mass Spectra (QCEIMS) program [S. Grimme, *Angew. Chem. Int. Ed.*, 2013, **52**, 6306] allowing efficient simulations for molecules composed of elements with atomic numbers up to Z=86. The recently developed extended tight-binding semi-empirical method GFN-xTB has been combined with QCEIMS, thereby eliminating dependencies on third-party electronic structure software. Furthermore, for reasonable calculations of ionization potentials, as required by the method, a second tight-binding variant, IPEA-xTB, is introduced here. This novel combination of methods allows the automatic, fast and reasonably accurate computation of electron ionization mass spectra for structurally different molecules across the periodic table. In order to validate and inspect the transferability of the method, we perform large-scale simulations for some representative organic, organometallic, and main-group inorganic systems. Theoretical spectra for 23 molecules are compared directly to experimental data taken from standard databases. For the first time, realistic quantum chemistry based EI-MS for organometallic systems like ferrocene or copper(II)acetylacetonate are presented. Compared to previously used semiempirical methods, GFN-xTB is faster, more robust, and yields overall higher quality spectra. The partially analysed theoretical reaction and fragmentation mechanisms are chemically reasonable and reveal in unprecedented detail the extreme complexity of high energy gas phase ion chemistry including complicated rearrangement reactions prior to dissociation.

8.1. Introduction

The application of *ab-initio* molecular dynamics (AIMD)[13] has become increasingly popular in recent years with the growing routinely available computational resources and the advent of efficient electronic structure methods and algorithms. In AIMD the chemical dynamics of a system is simulated directly by classically propagating the nuclear degrees of freedom, where the atomic forces along a potential energy surface (PES) are computed on the fly by a given quantum chemistry (QC) method. Recently, AIMD has been increasingly employed in relation to mass spectrometry to aid in the interpretation of, or even predict, experimental results *e.g.*, in electron ionization (EI) mass spectrometry,[10,307] collision induced dissociation (CID),[214,217,308] surface induced dissociation,[309,310,311] and dissociative electron attachment (DEA)[223,224,312]. AIMD simulations provide a promising alternative to the well-established statistical theories (*e.g.*, Eyring's Quasi-Equilibrium Theory,[14] and Rice-Ramsperger-Kassel-Marcus (RRKM)[22,23,24,25]). The problem of defining a set of decomposition channels *a priori* and locating the respective stationary points (minima and saddle points) on the potential energy surface (PES) is entirely avoided. As the number of viable decomposition channels grows rapidly with increasing molecular size, it can become very tedious and in some cases

even biased to use a statistical treatment for large molecular systems. However, if care is taken, such treatment can be very useful and yield valuable insights to mechanistic studies, by comparing the relative microcanonical rate constants for different unimolecular decomposition pathways. Therefore, RRKM has been widely applied in the context of mass spectrometry.[151]

AIMD simulations are able to explore automatically the energetically available regions of phase space and yield decomposition channels in an unbiased fashion. However, AIMD simulations are computationally expensive. Large-scale simulations beyond the picosecond time scale using density functional theory (DFT), or highly accurate wave function methods become computationally prohibitive. On these terms, fast and numerically robust semi-empirical schemes[43] provide a cost-efficient alternative.

Semi-empirical electronic structure methods are constructed by applying various approximations to Hartree-Fock (HF), yielding methods like the Parametric Models (PMx)[313,314] and the Orthogonalization-corrected Models (OMx)[228]. More recently such approximations have been applied to DFT, in particular to the exchange-correlation (XC) functional PBE[47], known as the DFTBx series[136,230]. These methods retain the fundamental limitation of the respective HF/DFT parent method, introduce further approximations to electronic integrals (*e.g.*, the neglect of three and four center integrals and the use of two center integral approximations) and employ minimal valence basis sets. Furthermore, the parametrization of a particular semi-empirical method often yields a poor description of molecular systems which differ from the training set and for properties that have not been included. The aforementioned approximations lead to an increase in computational efficiency by up to three orders of magnitude compared to HF/DFT. The price to pay is lowered accuracy and a poor description of certain chemical features. The considerable efforts devoted to develop corrections to these problems are summarized in Ref[43]. The high computational throughput of these methods render them valuable tools in large-scale quantum chemical calculations, *e.g.*, for biomolecular applications (>1000 atoms)[43,315] and long time-scale AIMD.

Regarding EI mass spectrometry, an original, exhaustive and widely applicable AIMD protocol has been devised and published, it is referred to as the **Q**uantum **C**hemistry **E**lectron **I**onization **M**ass **S**pectra (QCEIMS) program[10]. It is an automated, easy-to-use, dynamical procedure which combines AIMD with stochastic and statistical elements in order to predict reasonably accurate EI mass spectra (EI-MS) (see Fig. 8.1), without any preconceived notion of decomposition pathways. However, an almost non-empirical unbiased brute-force approach can not compete for fundamental reasons with the accuracy of database driven, knowledge based EI-MS predictors,[145,146,147,148] which should be kept in mind when judging the theoretical spectra.

The program is coupled to various third-party electronic structure software (*e.g.*, MOPAC,[316] ORCA[70,71], TURBOMOLE[63]), allowing the atomic forces required by the QCEIMS internal molecular dynamics procedure to be calculated with various semi-empirical methods (*e.g.*,

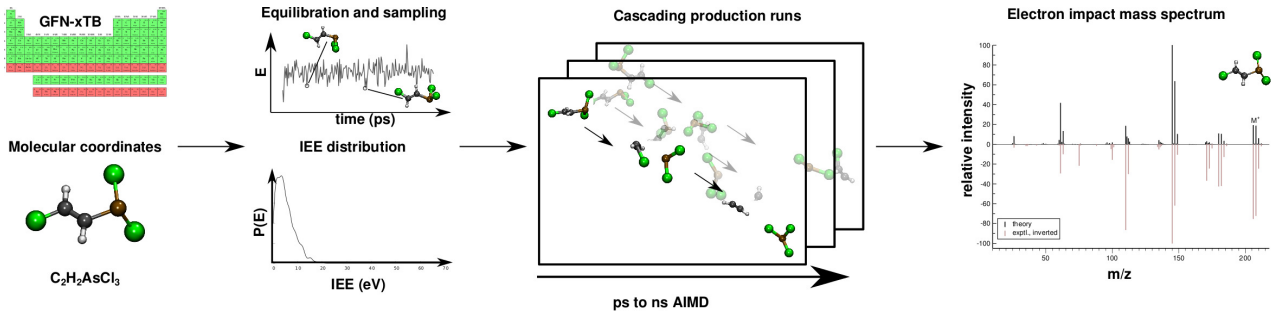

Figure 8.1.: Overview of the QCEIMS protocol. The number of production runs is chosen such that the simulation results are statistically converged.

DFTB3, OM2, PM3 and PM6) and standard DFT functionals. One of the main strengths of QCEIMS is that a simulated EI-MS result can be directly compared to the experiment. In addition the simulations provide valuable mechanistic insights into the dissociation dynamics, where bond ruptures and even complex molecular rearrangements prior to decomposition automatically occur during the simulations. All simulation trajectories are stored and can thus easily be inspected or post-processed. Therefore, the procedure is able to aid the user in mechanism-to-fragment-to-peak assignment. The decomposition pathways can also signify which channels are of high importance and can therefore be used in tandem with statistical theories, refined by higher level QC methods. In fact, DFT calculations have been used in the literature to study mass spectral fragmentation pathways, although the pathways were largely found because of prior experimental knowledge.[317,318,319,320,321]

A number of studies were conducted using the QCEIMS protocol on organic drug molecules[233] and the nucleobases[234,235] using different semi-empirical QC methods. In the most recent study, QCEIMS was extended to successfully predict the unimolecular decomposition pathways of four negatively charged nitrile compounds upon low energy electron attachment.[312]

In this study we have implemented two new semi-empirical methods, the GFN-xTB[41] and IPEA-xTB, in the QCEIMS program allowing spectral simulations for basically any reasonable molecule from the periodic table in a matter of minutes to a few hours of computation time, depending on the simulation conditions and number of available cores. The quantum chemical methods are tight-binding (TB) electronic structure schemes, where the former method was independently developed to accurately describe molecular geometries, atomic forces and non-covalent interactions of large molecules. The latter version of the same TB Hamiltonian was developed to accurately compute ionization potentials (IPs) and electron affinities (EAs), for QCEIMS as well as electrochemical applications (which are published separately). The GFN-xTB has been reported to be more computationally efficient, robust, and globally accurate than other similar semi-empirical methods, for the listed target properties.[41] Moreover, GFN-xTB has parameters available for elements with atomic numbers up to Z=86, making the approach applicable to a large range of molecular systems. Here we present the first, fully standalone

version of QCEIMS, where GFN-xTB is used for all energy and gradient computations. Note, that GFN-xTB which provides the PES for all reactions investigated in this work, was *not* modified specifically for the purpose studied here. Most IP evaluations, which are needed to compute the charge distribution on fragments, are conducted with the IPEA-xTB variant. This second parametrization is needed because at the TB level one can not simultaneously describe good PES and IP/EAs. Our new developments eliminate the necessity to employ third-party electronic structure software in QCEIMS and a fully stand-alone code is presented here for the first time. However, the option to use such software (*e.g.*, for DFT refinements) remains available. In the next section, a description is given for both of the TB variants and the QCEIMS protocol.

The purpose of this work is to assess the quality of simulated EI-MS produced by the combination of GFN-xTB/IPEA and QCEIMS along with its transferability to a diverse set of molecules. The basic QCEIMS scheme is not modified. Furthermore, the robustness and computational efficiency are investigated. For this purpose, we construct a molecular test set of 23 diverse molecules, composed of 24 different chemical elements. There are two criteria for the selection of the molecules. The first objective is to include as many elements as possible, in order to validate that the approach can predict accurate EI-MS for molecules composed of elements across the periodic table. The second objective is to compare the simulated spectra directly to the experimental spectra. Therefore, the molecules have to be well-known with well validated experimental EI-MS. All systems are obtained from the NIST[243] and SDBS[244] databases. Furthermore, the molecules should vary in structure, size, and chemical functionality. The chosen molecules are divided into three groups, organic, organometallic and main-group inorganic molecules. QCEIMS results for the later two groups are presented here for the first time.

The organic molecular group includes hexane (C_6H_{14}) (**1**), 1-flouro hexane ($C_6H_{13}F$) (**2**), 2-pentanone ($C_5H_{10}O$) (**3**), nitrobenzene ($C_6H_5NO_2$) (**4**), iodobenzene (C_6H_5I) (**5**) and testosterone ($C_{19}H_{28}O_2$) (**6**). The organometallic group includes ferrocene ($C_{10}H_{10}Fe$) (**7**), bis-benzene chromium ($C_{12}H_{12}Cr$) (**8**), copper(II)acetylacetonate ($C_{10}H_{14}O_4Cu$) (**9**), nickel(II)-bis(diphenyl-acetylacetonate) ($C_{30}H_{22}O_4Ni$) (**10**). The main group inorganic molecules include diborane (B_2C_6) (**11**), dichloro-ethylaluminium ($C_2H_5Cl_2Al$) (**12**), tetramethylsilane ($C_4H_{12}Si$) (**13**), dichloro-diphenylgermanium ($C_{12}H_{10}Cl_2Ge$) (**14**), tetramethylstannane ($C_4H_{12}Sn$) (**15**), tetraethyllead ($C_8H_{20}Pb$) (**16**), tetraethyl-diphosphane-disulfide (**17**), lewisite ($C_2H_2Cl_3As$) (**18**), triphenylstibine ($C_{18}H_{15}Sb$) (**19**), tris(para-tolyl)bismuthine ($C_{21}H_{21}Bi$) (**20**), octasulfur (S_8) (**21**), selenium hexamer (Se_6) (**22**) and diethyltelluride ($C_4H_{10}Te$) (**23**). In the inorganic molecular group, we have omitted molecules composed of elements for which EI-MS is not easily obtained *i.e.*, Gallium, Indium, Thallium and Polonium. Schematic representation of the molecules is given in Fig. 8.2, arranged in the order of the three given groups,

organic (**1**–**6**), organometallic (**7**–**10**) and main group inorganic molecules (**11**–**23**). Furthermore, the inorganic molecules are arranged by columns in analogy to groups 13–16 of the periodic table. We believe that the chosen molecular set will attest to the wide applicability and accuracy of the novel approach presented here. We aim to encourage the community to use QCEIMS in tandem to traditional experimental mass spectrometry. The `qceims` (3.62) program is available upon request* and should be suitable for any Linux distribution.

The paper is organized as follows: in Section 2, a general description of QCEIMS is given accompanied with a brief description of the underlying GFN-xTB and IPEA-xTB methods. Moreover, the computational expenses and robustness of the methods are discussed. In Section 3, we report the simulated EI-MS for the above molecular test sets and compare the results directly to the respective experimental data. We discuss each molecule individually and address interesting decomposition pathways with an emphasis on molecular rearrangements. In Section 4, concluding remarks are given.

8.2. Methodology

8.2.1. QCEIMS

The QCEIMS procedure is executed for each molecule, in three steps: (i) equilibration and sampling of (neutral) conformers, (ii) calculating the molecular orbital spectrum and (iii) production runs. A somewhat concise description of the three steps is given in the following. For a more involved discussion of QCEIMS, the reader is referred to Ref.[10]. The first and last steps of the procedure involve MD, where the neutral molecule or its positive ion, respectively, is propagated in time by numerically integrating Newton's equations of motion using the leap-frog algorithm. The time step is 0.5 fs. The atomic forces needed to integrate the equations of motion are calculated on the fly using GFN-xTB, which has been implemented in the QCEIMS program. The combination of QCEIMS and GFN-xTB is referred to as MS(GFN-xTB) if IPEA-xTB is used for the IP calculation (and MS(GFN-xTB/DFT) if DFT is used instead for the IP calculations) in the following discussion.

Equilibration and Sampling (i)

The neutral molecule of interest is equilibrated over a period of 12.5 ps in the canonical ensemble (NVT), with a constant temperature of 500 K. The equilibration is followed by a conformer sampling in the micro-canonical ensemble (NVE), where 1000 snapshots (geometry and nuclear velocities) are randomly selected and saved along a 25.0 ps trajectory. For consistency, the same simulation time (or trajectory length) is used for all molecules. The time is chosen such that a statistically uncorrelated sample of conformers is ensured, even for the

*To obtain `qceims`, please contact qceims@thch.uni-bonn.de

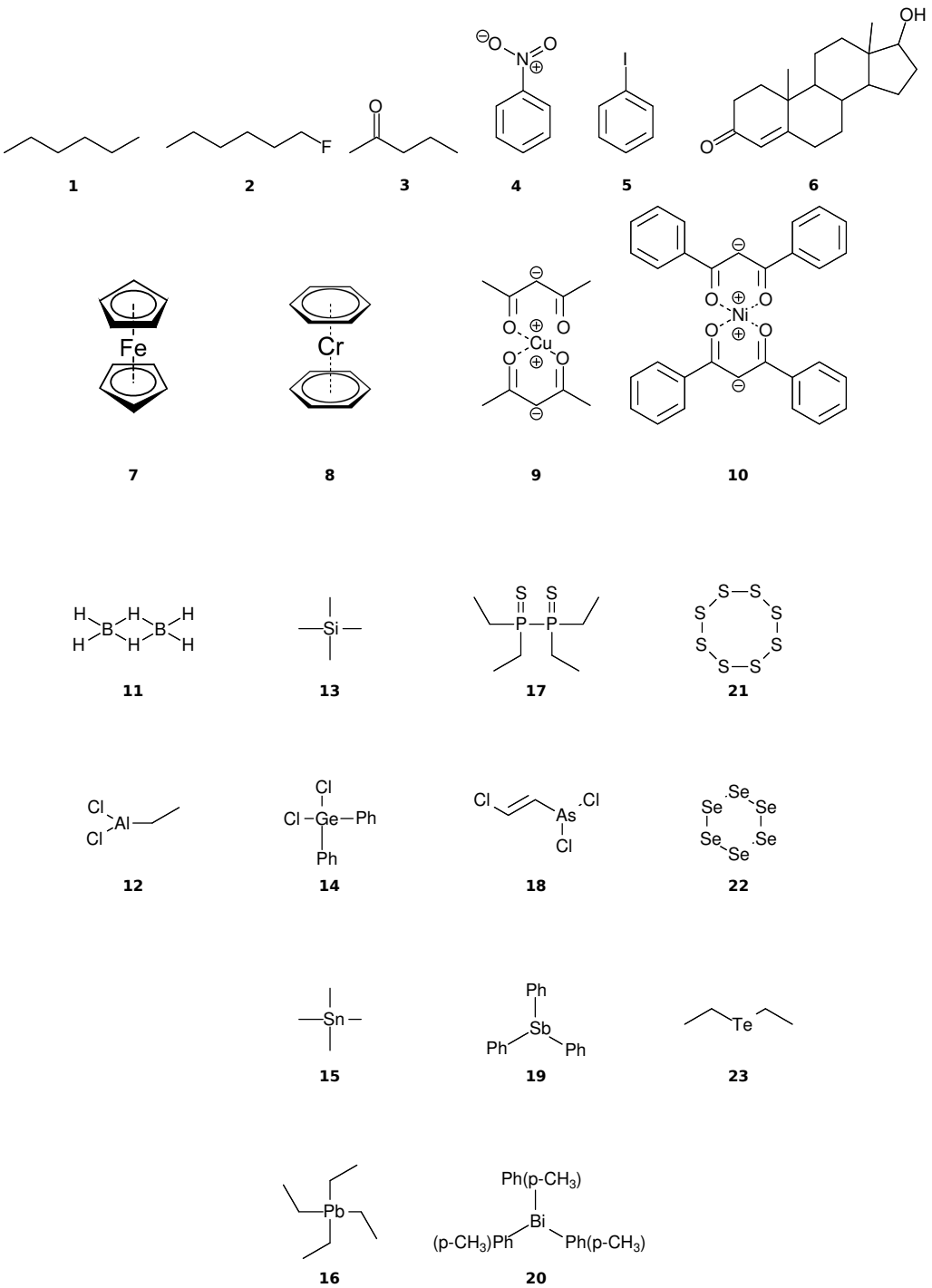

Figure 8.2.: Chart of the molecular test set (**1**–**23**). The first two rows represent organic and organometallic molecules, respectively. The last four rows represent main group inorganic molecules, arranged by columns in analogy to the periodic table (groups 13–16). Phenyl groups are denoted by Ph.

largest molecules in the test set *e.g.*, **6** and **10**. Very flexible systems, which are not considered here, will require longer ground state trajectories.

Molecular Orbital Spectrum (ii)

A single-point calculation with MS(GFN-xTB) is performed to determine the molecular orbital (MO) spectrum, followed by an MO resolved Mulliken population analysis.[322] This calculation is required, in order to estimate necessary ion state related quantities for the production run simulations, *i.e.*, the internal excess energy (IEE), internal conversion (IC) time, and MO-population derived nuclear velocity scaling factors.

The internal excess energy (IEE) represents the energy imparted on the molecule by the colliding electron and is distributed among the vibrational modes of the parent ion (also referred to as the molecular ion), by scaling the nuclear velocities (heating). The value of IEE for each production run is chosen in a stochastic manner, where it is assumed to be a Poisson-type variant,

$$P(E) = \frac{\exp(cE(1 + \ln(b/cE)) - b)}{\sqrt{aE + 1}} \tag{8.1}$$

$P(E)$ is the probability to have an IEE equal to E. The parameters a, b and c are given as $\approx$ 0.2 eV, 1.0 and $\frac{1}{aN_{\mathrm{el}}}$, respectively. The maximum value of IEE is equal to $E_{\mathrm{impact}} - \epsilon_{\mathrm{HOMO}}$, where E_{impact} is an input parameter and represents the kinetic energy of the free electron, before impact. It is set to 70 eV in analogy to standard EI experiments. The IEE distribution is set to have its mode at 0.6 eV per atom.

The internal conversion (IC) time is an interval over which the ion is heated. After the IC process, the IEE is entirely converted into nuclear kinetic energy. The IC time is calculated from the energy gap-law[271] and is dependent on the differences in MO energies. It is given by,

$$t_{\mathrm{IC}} = \sum_{j>i}^{M} \frac{k_{\mathrm{h}}}{N_{\mathrm{el}}} \exp(\alpha(\epsilon_i - \epsilon_j)) \tag{8.2}$$

where $\alpha = 0.5$ eV^{-1} and $k_{\mathrm{h}} = 2$ ps. M is the ordinal number of the HOMO and ϵ_i is the orbital energy of the i-th orbital. For molecules consisting of less than 35 atoms, MO-based velocity scaling factors are used. The scaling factor of a particular nucleus is proportional to the Mulliken population of that nucleus in the ionized MO. The idea is that ionization of localized MOs will yield localized structural distortions and therefore induce decomposition in the spatial vicinity of the MO. For larger molecules, the velocity scaling factors have been observed to yield some artifacts[233] (probably because the initial ionic states have more delocalised continuum character in larger systems) and are therefore set to unity.

Production runs (iii)

The randomly sampled conformers are instantaneously (valence) ionized and the coordinates and nuclear velocities are used as initial conditions for the propagation of the molecular parent ion in 1000 individual production runs. The production runs are performed in an embarrassingly parallel manner. The large number of runs, for each molecule, is to ensure that the resulting EI-MS are statistically converged with respect to the observed fragments. Furthermore, the maximum simulation time of an individual production run is chosen to be 10 ps (compared to a default value of 5 ps, used previously), to reduce the number of cases where the parent ion would otherwise not decompose, because of too short simulation time. The effect of this maximum simulation time is investigated in more detail for the two cases **2** and **13** as shown in appendix D. Note, that the overall simulation time in one run for a given parent molecular ion conformation can individually be much longer than the above maximum MD time of 10 ps because of the cascading technique used (see below).

In the beginning of each production run, the ion is heated by scaling the nuclear velocities, as described in the last subsection. The heating phase is usually conducted within the first 0.2-3.0 ps (IC time) of the simulation. The conceptual idea of the model is that after the EI of the molecule, an electronically excited ion will form, which relaxes to a vibrationally excited level of the electronic ground state (hot ion) through IC, followed by intramolecular vibrational redistribution (IVR), i.e., the excess energy imparted on the molecule from the colliding electron is transferred to the vibrational modes of the ion. Further propagation can then result in decomposition of the parent ion to favorable (radical) neutral and charged moieties.

If fragmentation occurs the algorithm will evaluate the vertical IP of each product by a Δ self-consistent field (SCF) or Δ self-consistent charge (SCC) calculation (see below), using IPEA-xTB, which is a differently parametrized TB variant of xTB implemented in the QCEIMS and solely employed for calculations of IPs. For molecules with a more difficult electronic structure *e.g.,* transition metal complexes, it can become vital to use more accurate QC methods for the IP evaluations in order to obtain more accurate peak intensities. Therefore, we use PBE0/def2-SV(P)[232] to evaluate the IPs for the organometallic molecules. All DFT calculations are performed using the ORCA 3.0.3 suite of programs. Moreover, a new feature is introduced in the QCEIMS protocol: when a fragment contains a 3d transition metal, the algorithm will begin by automatically finding the ground state multiplicity for both the ion and the neutral molecule, followed by the ΔSCF evaluation of the neutral and ion ground state. This new feature introduces few additional single point energy calculations but is found

to improve the IPs. The statistical weight of each product is then given by

$$C_i = \frac{\exp\left(\frac{-\Delta E_{\mathrm{SCF,i}}}{k_{\mathrm{B}} T_{\mathrm{Av}}}\right)}{\sum_j^M \exp\left(\frac{-\Delta E_{\mathrm{SCF},j}}{k_{\mathrm{B}} T_{\mathrm{Av}}}\right)} \tag{8.3}$$

where M is the number of fragments, T_{Av} is the average fragment temperature and k_{B} is the Boltzmann constant. The product of the statistical weight C_i and the total molecular charge yields the statistical charge of fragment i.

The fragment with the highest statistical charge is selected and propagated further in a so-called cascade, while the other fragments (with lower statistical charges) are counted and stored. In the cascade, the selected positively charged secondary fragment can decompose further. If this secondary fragment decomposes, the IPs of the newly formed tertiary fragments are calculated and the statistical charges are determined. Again, the tertiary fragment with the highest statistical charge is selected and a new cascade initiated. In each cascading run, the statistical weights are multiplied by the dominant statistical weight of the preceding run. In other words, the sum of the statistical charges of all order fragments, in a single production run, including all cascades, is equal to the total molecular charge (*i.e.*, 1). Therefore, the sum of the statistical charges for a specific fragment over the ensemble of production runs will yield the total relative intensity of the particular fragment, allowing the algorithm to predict EI-MS for an arbitrary molecule, as long as the PES and IP computations are reasonable. The natural isotope ratios are introduced in a post-simulation treatment. Furthermore, specific isotope labeling can easily be introduced by performing the simulations with altered nuclear masses. However, this option is not considered herein. The quality of the resulting, fully theoretical, basically first-principle, MS is determined by a number of factors, in particular the underlying QC method. It has been observed that the level of QC accuracy is reflected in the quality of the simulated spectrum.[10,234] Moreover, the number of production runs and the maximum simulation time accessible can affect the resulting spectra. More subtle effects which are harder to resolve are *e.g.*, the neglect of non-adiabatic effects (*i.e.*, where the charge is not assigned to the fragment with the lowest IP) and the nature of the IEE distribution[10]. Furthermore, in its current state, QCEIMS only allows for singly ionized species.

8.2.2. GFN-xTB and IPEA-xTB

The GFN-xTB method was developed and published very recently in our laboratories, and to familiarize the reader with the method, we give a brief but essential introduction to the features and ideas of GFN-xTB below, along with a short description of IPEA-xTB which is founded on the GFN-xTB method. For a more in-depth discussion the reader is referred to Ref.[41].

The GFN-xTB is a special-purpose semi-empirical approach analogous to the well-established DFTB3 method[230]. The GFN-xTB is motivated from the success of its predecessor sTDA-xTB[72,323,324], where an extended TB scheme is used to calculate, with good accuracy, electronic excitation spectra of large molecules. The new modified extended TB variant GFN-xTB, targets geometries, frequencies and non-covalent interactions (hence the namesake, "GFN"). It has been shown to yield more accurate results for the given target properties than other general semi-empirical methods, which usually attempt to capture both structural and energetic features (*e.g.,* thermochemistry) simultaneously.[41] The method is described as extended (denoted by "x" in xTB) because it employs partially polarized minimal basis sets, *i.e.*, with an additional s-function on H and d-functions for third row and higher elements. The use of an extended basis set largely alleviates problems in describing systems with polar bonding *e.g.*, hydrogen and hypervalent bonding situations for heavier elements. Furthermore, the GFN-xTB is found to be computationally faster than other comparable methods mainly due to quick and robust convergence of the electronic iterations.[41] Therefore, large-scale quantum chemical treatments of complex molecular systems can be performed routinely. The number of empirical method parameters is minimized and restricted to global and element-specific values, making it more transferable and easy to parametrize. There are only 19 global parameters and approximately 10 element specific parameters included in the GFN-xTB method. The parameters have been fitted to hybrid DFT data, where the target quantities are equilibrium and slightly distorted structures, harmonic vibrational frequencies, CM5 atomic charges[325] and non-covalent interactions energies and structures. Currently, parameters exist for elements up to Z=86, making the method applicable to a large range of chemical systems. The aforementioned properties: fast computations, robustness and wide applicability of GFN-xTB along with precise analytical nuclear gradients make the approach ideal to use in conjunction with QCEIMS. In the xTB approach, the total energy is expressed as a sum of four terms; the electronic energy (E_{el}), the repulsion energy (E_{rep}), the well-known D3(BJ)[33,34,35] dispersion energy (E_{disp}), and a classical correction for halogen-bonding interactions (E_{XB}). The electronic energy is computed by a SCC treatment, analogous to that of DFTB3. For a derivation and details of the GFN-xTB method, see Ref.[41] As discussed in the original publication, the use of a finite electronic temperature treatment[27,46,326] (see below) allows proper dissociation of covalent one- and two-electron chemical bonds which is of vital importance for QCEIMS.

The second TB variant, IPEA, is also of special-purpose and succeeds from the GFN-xTB. It is a straightforward re-parametrization to calculate reasonably accurate IPs and EAs up to a constant empirical shift. Moreover, the IPEA variant uses additional (n+1)s basis functions. It has been re-fitted to reference IP/EA values for parts of the original GFN-xTB training data set. The reference IP/EA values are computed by PW6B95/def2-TZVPD[327] with TURBOMOLE 7.1[62,63]. Typical errors for computed vertical IP/EA values by IPEA-

xTB compared to DFT are 0.2-0.4 eV. Both the GFN and IPEA parametrizations used by `qceims` are available from the authors and a more detailed discussion of the accuracy of IPEA-xTB for IP/EA will be given elsewhere in the context of electrochemistry applications. At this point, the special IPEA-xTB parameters are only available for parts of the periodic table, excluding the transition metals. For such species the standard GFN-xTB element parameters are used for the IP calculation step and we present MS(GFN-xTB) acquired spectra for the organometallic complexes in appendix D. As discussed above, in such cases MS(GFN-xTB/DFT) should currently be used and ongoing work in our group is devoted to cover all elements by IPEA-xTB.

The single point calculations involved in the first two steps of the QCEIMS protocol (equilibration, sampling and the MO spectrum calculation) and also the IP evaluations employ Fermi-smearing[27] at a default electronic temperature of 300 K. In the regular classical propagation of the nuclei, during the third step of QCEIMS (*i.e.*, production runs), the electronic temperature is chosen to be 5000 K (cf. the supporting information of Ref.[10]). Fermi smearing is found to facilitate SCC convergence and partially remedy electronically complicated situations which arise during the MD trajectories. It is essential to qualitatively correctly describe the dissociation of the parent ion and fragments to (radical) neutral and charged moieties, without resorting to impracticable multi-reference theory.

8.2.3. Performance

The production runs are executed in parallel on Intel(R) Xeon(R) E5-2660 2.00GHz cores, where each production run occupies only a single core. The total number of MS(GFN-xTB) single point energy/gradient calculations performed in the production runs of all included test set molecules surmounts to roughly 270 million. This sheer number emphasizes the need to use incredibly efficient electronic structure methods in conjunction with QCEIMS.

To further inspect the computational speed and robustness of MS(GFN-xTB) (and MS(GFN-xTB/DFT)), the average time per energy/gradient computation and the percentage of unsuccessful production runs is reported for each molecule in Fig. 8.3, with the exception of **10** (which is specifically addressed below). The average time per computation is found to be roughly 0.05 seconds, *i.e.*, 20 force evaluations per second, for both the main group inorganic and organic group molecules. The average time per energy/gradient computation for the organometallic molecules, is found to range from 0.15 to 0.30 seconds. The one order of magnitude increase in the computational time, from organic to organometallic molecules, is ascribed mainly to the overhead by the hybrid DFT IP evaluations. Moreover, multiplicity checks are employed for all fragments that include 3d transition metals, increasing the number of energy evaluations per IP. In the special case of **10**, we find the average time to be approximately 1.4 seconds per calculation resulting from the increased DFT overhead of the

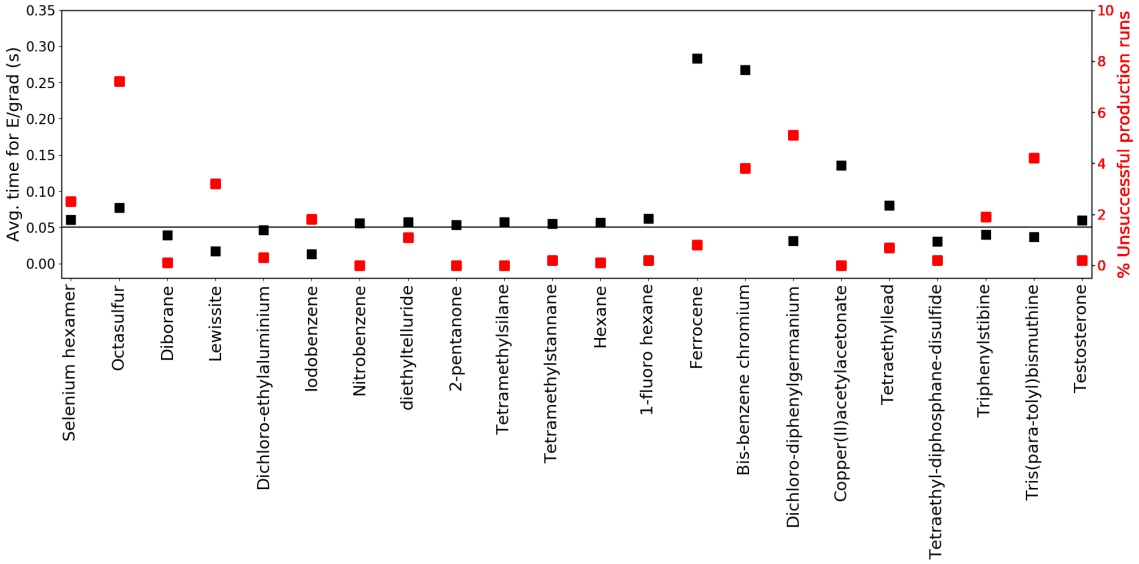

Figure 8.3.: The average time per single energy/gradient computation and robustness is depicted for all of the test-set molecules. The molecules are listed on the vertical axis, by ordering of increased molecular size (left to right). The average computational time is depicted on the left horizontal axis and denoted by black squares. The robustness (percentage of unsuccessful production runs out of 1000) is depicted on the right horizontal axis and denoted by red squares. **10** is an outlier and is omitted for clarity.

relatively large molecule.

It is hard to generalize about the wall-time required to simulate a EI-MS beforehand, since the computational time is heavily influenced by the input molecule itself, the number of production runs, cascades and fragments, and available computational resources. For this purpose we have listed the average and maximum number of energy/gradient computations required by the production runs, as well as, the computational times in appendix D. We find that the average number of energy/gradient computations per production run ranges from roughly 4000 (**12**) to 15000 (**19**) calculations. For the organic and main group inorganic molecules, the average wall-time per production run, can range from roughly 86 (**18**) to 857 (**6**) seconds. For the organometallic molecules the average wall-time per production run ranges from 1630 (**9**) to 18700 (**10**) seconds.

The robustness of MS(GFN-xTB) is found to be remarkably good. The majority of the molecules exhibit less than 2 % unsuccessful production runs. The number of unsuccessful runs for **10** is found to be exactly 4 % and only three molecules (**14**, **20** and **21**) have >4 % failure rate. For these three molecules, the number of unsuccessful runs is between 5 and 7 % which we consider as borderline for an unbiased sampling. In case of higher failure rates one would increasingly sample merely the electronically 'simple' part of the reaction space leading to biased results.

In order to predict relatively accurate theoretical EI-MS for an almost arbitrary molecule, the PES of the GFN-xTB has to parallel the 'true' PES for a wide range of nuclear arrangements. Therefore, we inspect a few simple exemplary reaction coordinates for decomposition pathways occurring in our simulations, using hybrid FT-DFT as reference, which are given in appendix D. Exemplary dissociation curves are additionally discussed in the original GFN-xTB publication[41]. Analysis of the data shows that potential energy curves for simple dissociation (using GFN-xTB) are relatively accurate despite the fact that the method was not primarily parametrized for energetic properties. We attribute this success (and that of the entire MS(GFN-xTB) scheme) to the inherent ability of TB methods to properly dissociate bonds in tandem with our specific fit to vibrational frequencies (yielding accurate force constants) and also to Fermi smearing.

8.3. Results and discussion

In this section, we present simulated EI-MS for all molecules of the test set (**1**–**23**) and compare the results directly to the respective experimental spectra. We address some fragment structures, investigate reaction pathways with an emphasis on molecular rearrangements and perform fragment-to-peak assignments for chosen signals, *e.g.*, determinative peak-series. In the analysis of the production run trajectories, we use the same script as previously reported[312] to identify fractional yields and distinguish between structural isomers that contribute to the same peak, or mass-to-charge ratio m/z. The comparison of the experimental and computed MS, for all molecules, is visualized in Figs. 8.4–8.9. The molecular ion and ionic fragments, which are discussed in the text, have been superimposed on the computed MS depicted in the figures. The visualized structures are taken as the average fragment structures over the last 50 MD steps in the production run trajectories and are also labeled by their m/z values.

8.3.1. Organic Molecules (1–6)

For hexane (**1**, Fig. 8.4a)), the simulated MS is found to be in very good agreement to the experimental spectrum, where the observed peak series m/z 57 ($C_4H_9^+$), m/z 43 ($C_3H_7^+$), and m/z 29 ($C_2H_5^+$) is reproduced by the simulations. Moreover, we find the parent ion to be slightly too stable, in the simulations. An inspection of the trajectories reveals that fragments m/z 58 and m/z 43 result from the formation of a tertiary cation, where a H atom migrates to a terminal position. Moreover, the simulations successfully predict the fragment m/z 29 to have the confirmed 'non-classical' ethyl cation structure.[328,329]

For 1-fluorohexane (**2**, Fig. 8.4b)), the experimental and computed MS are in good agreement. However, the parent ion does not decompose in a large number of production runs, meaning the survival rate of the **2** cation is too high under the given simulation conditions.

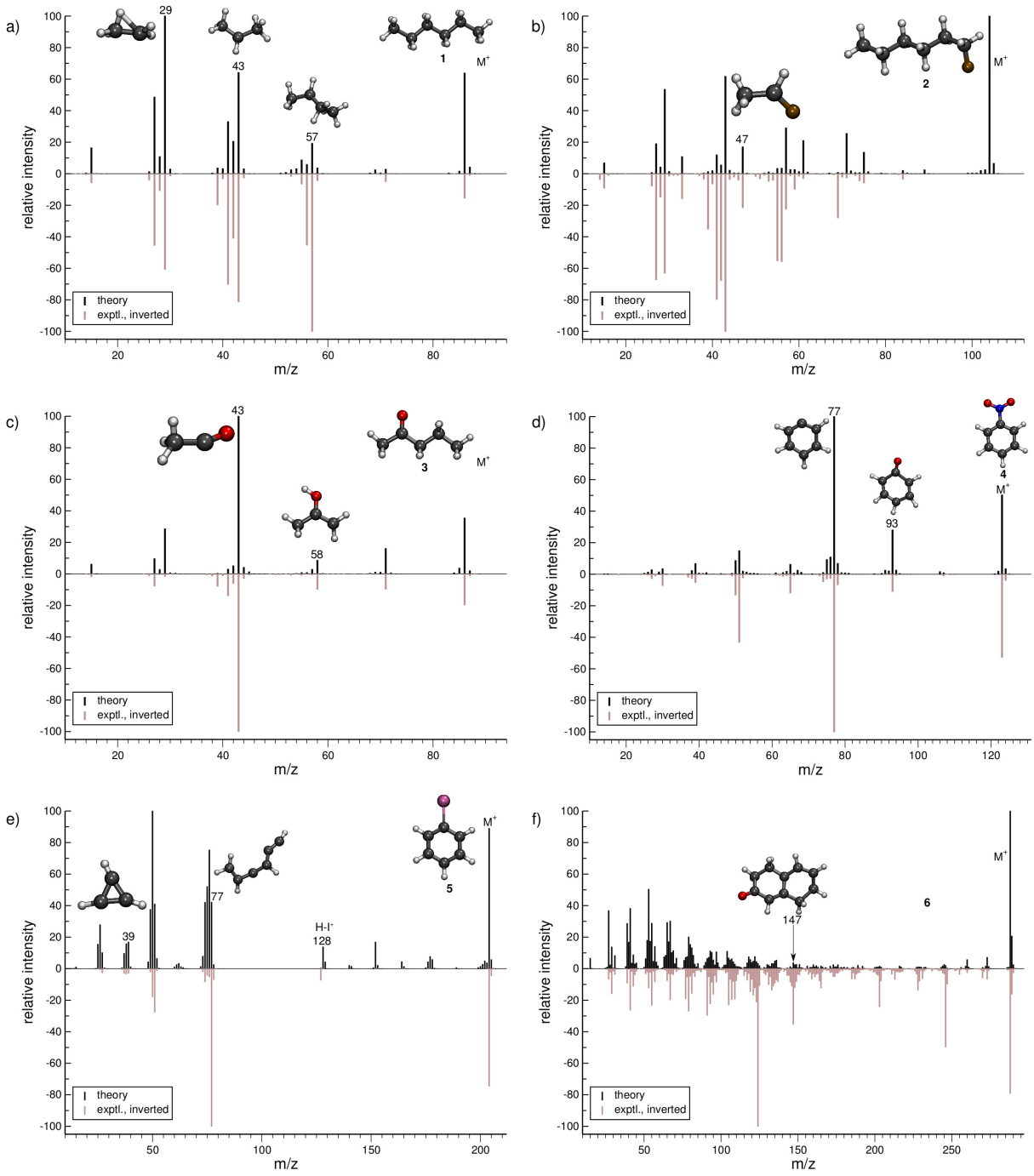

Figure 8.4.: Comparison of computed and experimental EI-MS for the organic set, including molecules **1**–**6** in a)-f), respectively. The structures of the parent ion (denoted by M^+) and selected ionic fragments have been superimposed on each computed spectrum. Moreover, the selected ions are marked by the respective m/z values and discussed in the text.

This problem can be partially alleviated by applying a higher IEE and/or longer simulation times. We would like to stress that the IEE distribution is not obtained specifically by *ab-initio* QC calculations, but rather assumed to be a Poisson type variant, for all molecules. Therefore, such effects are to be expected for certain molecules with unusual (1e-2e)-EI cross sections. The same determinative peak series, as observed in the MS of **1**, is observed for **2** and is again reproduced. Moreover, an additional signature peak is observed for **2**. It corresponds to the fragment m/z 47 ($C_2H_4F^+$) and involves a H atom migration, to form the more stable cation.

The next molecule is 2-pentanone (**3**, Fig. 8.4c)). This molecule is of a particular difficulty[10], since the parent ion can undergo the well-known McLafferty rearrangement[2,330]. This reaction is characterized by an H atom transfer to the carbonyl oxygen and a subsequent loss of a neutral olefin molecule, or propylene in our case. Indeed, the correct ion, m/z 58 ($C_3H_6O^{\bullet+}$), is reproduced in the simulations. Moreover, the base peak is correctly computed to be m/z 43 ($C_2H_3O^+$) and is found to result from α-cleavage. Overall, the experimental and computed MS of **3** are in excellent agreement.

The simulated MS of nitrobenzene (**4**, Fig. 8.4d)) contains all of the statistically significant peaks, found in the experimental spectrum. The relative intensity of the parent ion signal is reproduced quite well. The fragment m/z 93 ($C_6H_5O^+$) is an example of how complex molecular rearrangements are captured by the simulations. Here, the fragment can only form after loss of an NO molecule, subsequent to an oxygen atom migration. Moreover, the fragment m/z 77 ($C_6H_5^+$) is the cyclic phenyl cation and it forms the base peak in both the computed and experimental MS. We find the fractional yield of this phenyl cation to be 33 %. The agreement between the simulated and experimental spectra of **4** is good.

The phenyl cation is also one of the main products in the fragmentation of iodobenzene (**5**, Fig. 8.4e)). Interestingly, an analysis of the fragment m/z 77 reveals that the cyclic phenyl cation has a fractional yield of only 3.7 %, whereas various acyclic isomers of $C_6H_5^+$ are formed as well. This is in line with experimental studies that find a certain fraction of acyclic $C_6H_5^+$ in IR measurements, following the dissociation of halobenzenes.[331] Another peak that occurs in the MS of both **4** and **5** is m/z 39 ($C_3H_3^+$) which has the structure of the cyclopropenyl cation[332]. The comparison between the computed and experimental MS of **5** is reasonable, as there are a few artifacts observed in the computed MS. Additionally, certain peaks are found to be over pronounced *e.g.,* m/z 128 (H–I^+), in the computed spectrum.

The MS of testosterone (**6**, Fig. 8.4f)) contains a large number of peaks, and many, but not all of them are found in the computed MS. Most fragmentation pathways of **6** are underestimated by the simulations, even though the peak series in the lower-mass end of the spectrum is reproduced quite well. The stability of the parent ion appears to be estimated accurately. We have picked only one isomer contributing to the peak m/z 147 ($C_{10}H_{11}O^+$) to be displayed in Fig. 8.4f. As there are only a few production runs that yield this particular ion, it is not clear whether the displayed structure is the most abundant isomer. However, this structural

isomer results from the cleavage of two rings of the steroid scaffold, which appears to be a reasonable pathway.

In summary, the computed EI-MS for the organic group, **1**–**6**, are in general found to compare very well with the experimental spectra, using the new MS(GFN-xTB) approach. The quality of the spectra is comparable or slightly better than previous results acquired using other similar semi-empirical methods[233]. Furthermore, we show that the simulations are able to shed light on complex dissociation dynamics, where molecular rearrangements occur naturally in the simulation trajectories, *e.g.,* the McLafferty rearrangement for **3** and an oxygen atom transfer for **4**. Additional computed EI-MS of organic molecules have been included in appendix D.

8.3.2. Organometallic Molecules (7–10)

The treatment of organometallic molecules is challenging because of their generally (already in the neutral ground state) more complicated electronic structures.

The computed MS of ferrocene (**7**, Fig. 8.5a)) compares well to the experimental spectrum, which is a big success of the new approach. There are a number of peaks that are clearly over-pronounced in the computed spectrum, which can partially be attributed to the too low stability of the parent ion. The parent ion is not found to be the base peak in the computed spectrum. Instead, the base peak, m/z 121, is the $C_5H_5Fe^+$ fragment, which is formed by the loss of one cylopentadienyl ligand from the parent ion. Fe^+ (m/z 56) is observed, both experimentally and in the computed spectrum. The percentage of failed runs is only about 1 %, which is remarkable considering the electronic complexity of the ferrocene radical cation in particular.[333]

There are no major artifacts found in the computed MS of bis-benzene chromium (**8**, Fig. 8.5b). We consider the comparison between the experimental and computed spectrum to be good. We find the ion m/z 130 (formed by benzene loss from the parent ion), the benzene cation m/z 78, the chromium ion m/z 52 and a few less intense peaks. However, the latter cannot be considered representative, since the number of production runs corresponding to these peaks is smaller than the number of failed production runs, which is 3.8 %.

The computed spectrum of copper(II)acetylacetonate (**9**, Fig. 8.5c)) compares adequately to the experiment. Most of the peaks are reproduced in the simulations. However, there are a few artifacts in the simulated spectrum, but the extent of these is small. The fragment m/z 147 ($C_4H_4O_2Cu^+$) forms by the loss of an acetylacetonate ligand and a methyl radical. Moreover, the fragment m/z 105 ($C_2HOCu^{\bullet+}$) requires even more bond ruptures and results from the m/z 147 fragment *via* carbon monoxide and methyl radical loss. Comparable to **7**, the number of unsuccessful production runs is below 1 %.

The nickel(II)bis(diphenyl-acetylacetonate) (**10**, Fig. 8.5d)) molecule is the largest and in many ways the most challenging system, investigated in this study. This is reflected by the

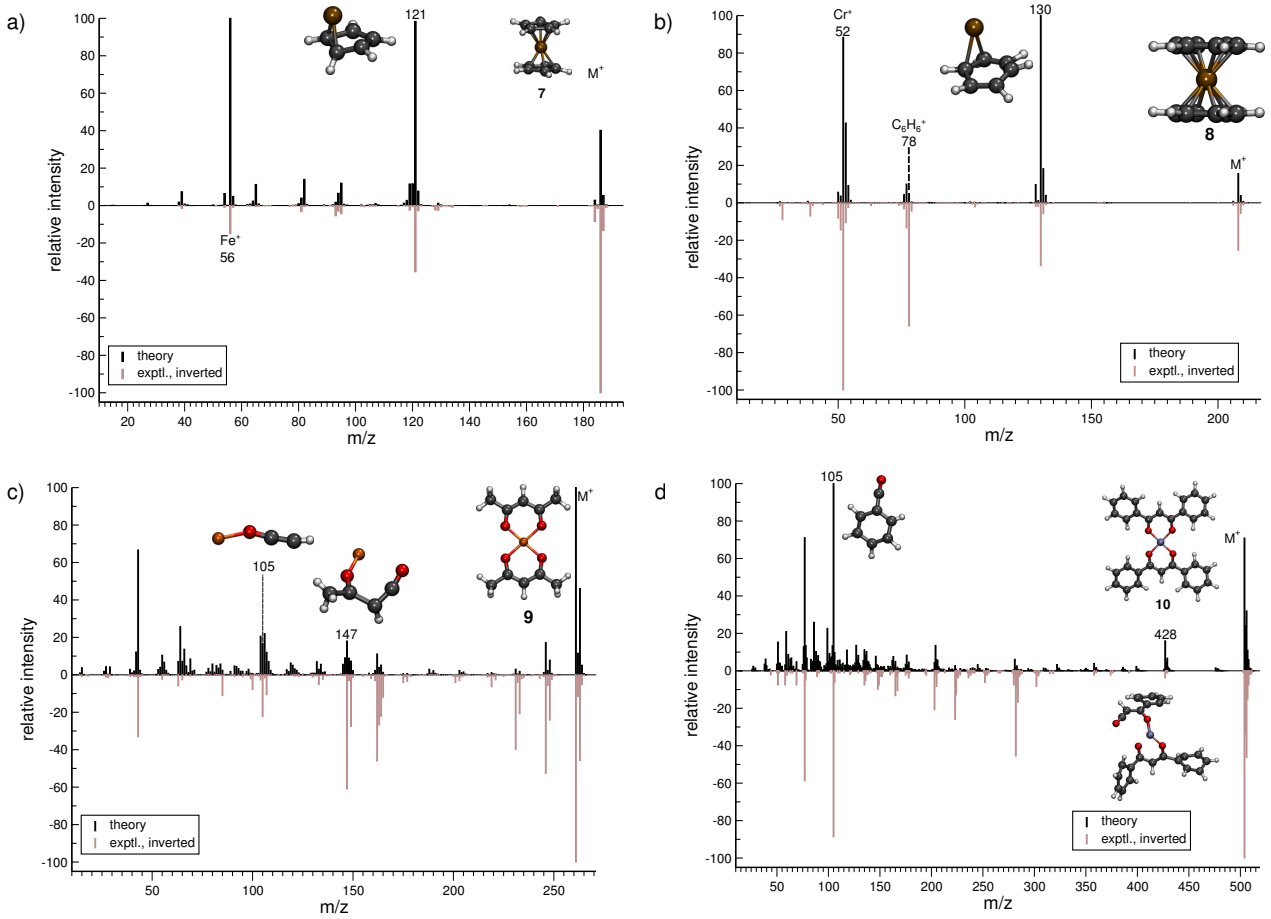

Figure 8.5.: Comparison of computed and experimental EI-MS for the organometallic group, including molecules **7**–**10** in a)-d), respectively. The structures of the parent ion (denoted by M^+) and selected ionic fragments have been superimposed on each computed spectrum. Moreover, the selected ions are marked by the respective m/z values and discussed in the text.

higher percentage failure rate of 4.0 %. Nevertheless, the computed MS compares moderately well to the experimental spectrum, where a number of peaks are successfully reproduced, *e.g.,* the fragment m/z 428 ($C_{24}H_{17}O_4Ni^+$), which forms after a phenyl radical loss and also the benzoyl cation at m/z 105 ($C_7H_5O^+$). We stress that quantum chemical calculations on Nickel complexes have remained a major challenge for DFT. Therefore, the computation of the EI-MS of **10** stands out among the results, even if the agreement between theory and experiment is not quantitative.

In summary, the comparison of experimental and computed MS for the organometallic molecules, **7**–**10**, provides us with the confidence that unimolecular decomposition pathways of cationic transition metal complexes can indeed be studied, in detail, with the novel MS(GFN-xTB/DFT) combination. This can also be seen from further computed MS shown in appendix D. The unprecedented success indicates the quality and robustness of the underlying GFN-xTB method, a really intriguing finding that could not be fully anticipated from its construction principle[41]. For organometallic complexes, we advocate at this point the use of hybrid DFT for the calculation of IPs, where reasonably accurate IP calculations play the central role in determining the computed peak intensities. As discussed in Section 2, even the relatively few DFT calculations will become the computational bottleneck of the whole procedure. Improving the preliminary parametrization of IPEA-xTB for organometallic compounds (which can be considered as a worst case scenario for the entire QCEIMS) might resolve this issue. It is in any case very encouraging to see the possibility of realistic theoretical EI-MS for organometallic compounds without any significant modifications or empirical adjustments of the procedure.

8.3.3. Inorganic Molecules (11–23)

Group 13 (11-12)

For diborane (**11**, Fig. 8.6a)), the comparison of the simulated and experimental MS is good. The fragmentation cascades consist of multiple hydrogen losses, both in the form of single H atoms and H_2 molecules. The base peak is correctly computed to be the $B_2H_4^{\bullet+}$ (m/z 26) fragment. We observe an interesting structure for this fragment, ascribed to the H atom mobility in the diborane cation where the H atoms can move freely between the boron centers. The fragment BH_3^+ (m/z 14) forms by a rupture of the boron-boron bond. However, the BH_3^+ fragment is found to be less abundant than fragments of m/z 13 and m/z 12 in the experimental MS, showing that H atom loss continues even after the B–B bond rupture. This is also reflected in the simulated spectrum.

In the case of dichloro-ethylaluminium (**12**, Fig. 8.6b)), the experimental and computed spectra are in a somewhat poorer agreement, where several peaks observed experimentally are missing in the computed spectra. The dominant reaction pathway is the dissociation of the ethyl moiety from the parent ion, resulting in the formation of $AlCl_2^+$ (m/z 97) and an

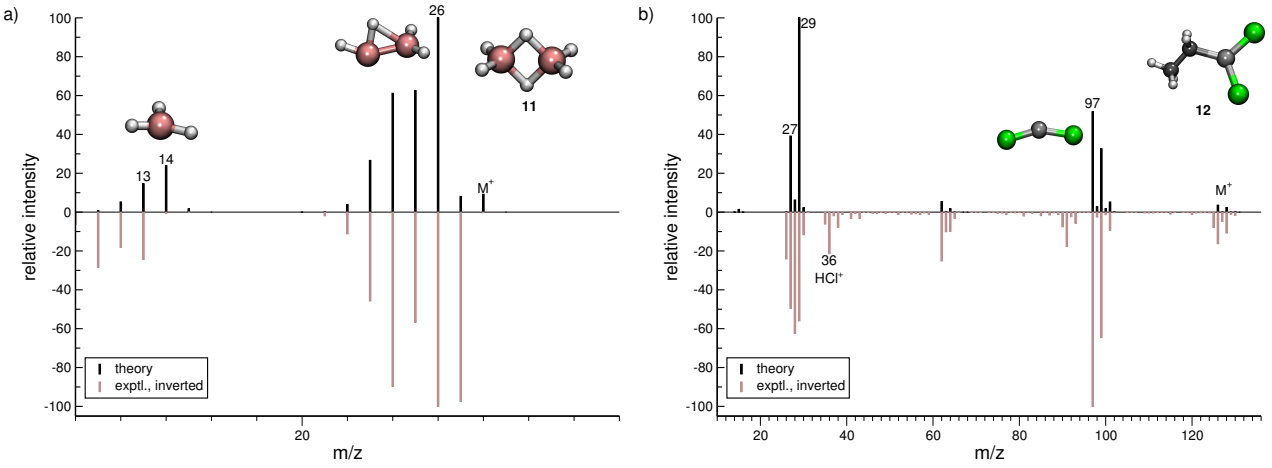

Figure 8.6.: Comparison of computed and experimental EI-MS for group 13 inorganic molecules, **11** and **12** in a) and b), respectively. The structures of the parent ion (denoted by M^+) and selected ionic fragments have been superimposed on each computed spectrum. Moreover, the selected ions are marked by the respective m/z values and discussed in the text.

ethyl cation (m/z 29). The IPs of these ions are comparable, which is why both ions are observed in the simulated spectrum. We note that $HCl^{\bullet+}$ is observed in the experimental MS of **12**. This fragment will be assigned a negligible statistical charge in the QCEIMS procedure because of the large IP of HCl. In the experiment the HCl^+ fragment may form from a reaction involving **12** and H_2O, prior to the ionization of **12**. For volatile compounds, such peaks can be rationalized too by our procedure.

Group 14 (13–16)

The computed MS of tetramethylsilane (**13**, Fig. 8.7a)) compares well to the experimental MS, although the computed survival rate of the parent ion is much too high. The base peak, assigned to the fragment m/z 73 ($C_3H_9Si^+$) is reproduced in the computed MS. This fragment is formed by the loss of a methyl group, from the parent ion. Interestingly, there is a very weak signal for the double methyl loss in the experiment, while the same signal is predicted to be strong in the computed spectrum. The fragment CH_3Si^+ (m/z 43) is correctly computed to have a relatively low abundance. In addition, in the MS of **13**, the importance of the statistical charge model becomes even more evident for the observed methyl ion (m/z 15) signal. The methyl cation acquires a non-negligible statistical charge because of the higher but still relatively similar IP of the methyl radical compared to the other reaction species.

For dichloro-diphenylgermanium (**14**, Fig. 8.7b)), the computed and experimental MS are in relatively good agreement. The peak series (m/z 264, 221, 186) reflects the loss of one chlorine atom, one phenyl group and one phenyl group as well as two chlorine atoms, respectively. The

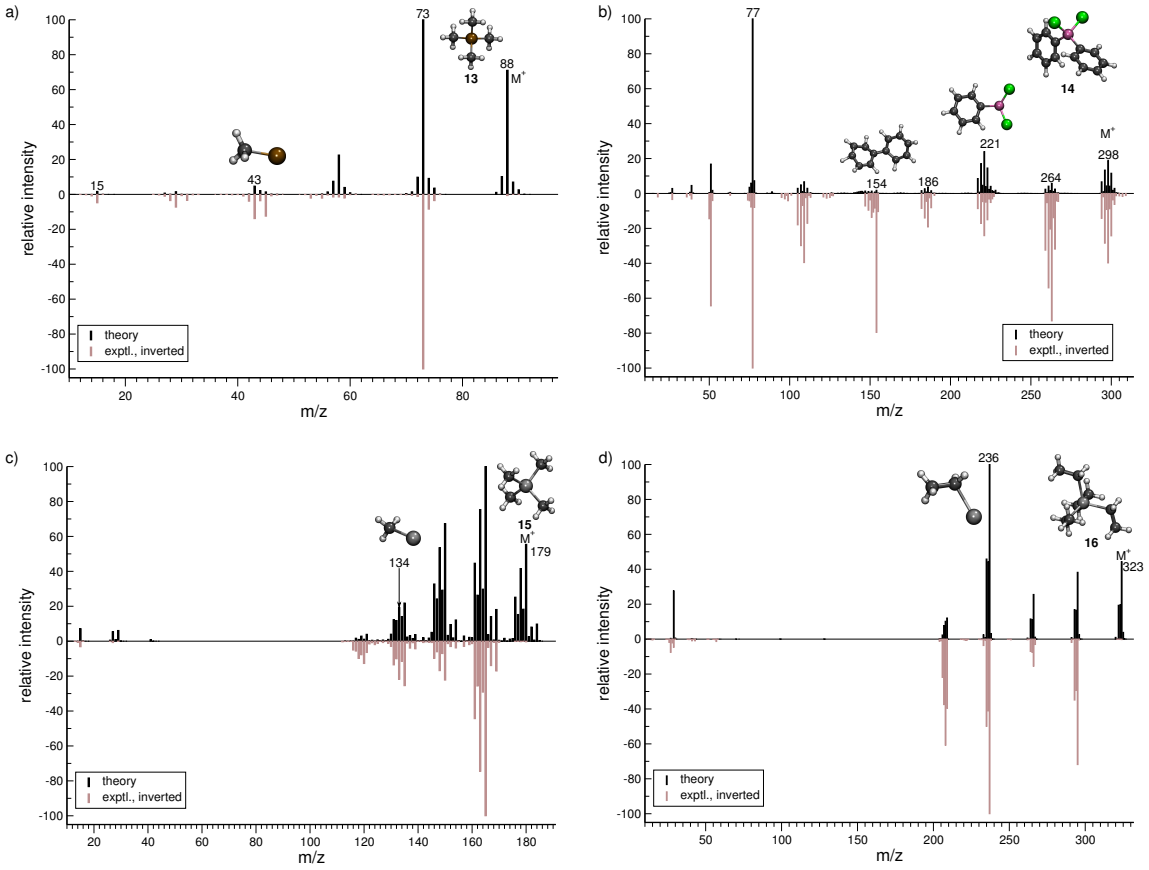

Figure 8.7.: Comparison of computed and experimental EI-MS for group 14 inorganic molecules **13**–**16** in a) to d), respectively. The structures of the parent ion (denoted by M^+) and selected ionic fragments have been superimposed on each computed spectrum. Moreover, the selected ions are marked by the respective m/z values and discussed in the text.

IP of the phenyl fragment (m/z 77) is relatively low, and consequently, it is computed to be the base peak. Here, the formation of the biphenyl cation, $C_{12}H_{10}^{\bullet+}$ (m/z 154) is observed. This interesting reaction pathway is underrepresented in the production runs with a fractional yield of only 0.7 %. It proceeds by a molecular rearrangement, followed by a C–C bond formation.

The computed MS of tetramethylstannane (**15**, Fig. 8.7c)) compares very well to the experimental MS. The decomposition of the parent ion is governed by a series of methyl losses, resulting in the formation of m/z 164 ($C_3H_9Sn^+$), 149 ($C_2H_6Sn^+$), 134 (CH_3Sn^+) and the naked Sn^+ (m/z 119) cation. This peak series is accurately captured by the simulations. The parent ion is observed to be statistically insignificant in the experiment, whereas we find it to be a relatively intense signal theoretically. This discrepancy is attributed to the high stability of the parent ion in the simulations which can be improved by adjusting the IEE simulation parameters as noted above.

For tetraethyllead (**16**, Fig. 8.7d)), the agreement between experiment and simulation is good. The MS reflects a series of ethyl losses, and the base peak is accurately predicted to be the $C_2H_5Pb^+$ (m/z 236) fragment. As for **15**, the peak corresponding to the parent ion is negligible in the experimental MS, while being a relatively intense signal computationally.

Group 15 (17–20)

Tetraethyl-diphosphane-disulfide (**17**, Fig. 8.8a)) exhibits an interesting isomerization reaction subsequent to an ethyl radical loss from the parent ion, yielding the $C_6H_{15}P_2S_2^+$ (m/z 213) fragment. As visualized in Fig. 8.8a, this fragment no longer has a P–P bond, instead the sulfur atom rearranges to form a bridging P–S–P bond. The base peak is correctly found to be the fragment $C_2H_{10}PS^+$ (m/z 121) and results from a rupture of the P–P bond, in which the phosphorous adopts a trigonal coordination. The overall agreement between the spectra is decent, where the molecular ion is predicted to be slightly too unstable compared to the experiment.

Lewisite (**18**, Fig. 8.8b)) is used as a chemical weapon and hence it is important to understand its MS and that of its derivatives, for analytical purposes.[334] We find the overall comparison between the experimental and simulated spectra to be good. One interesting reaction is the formation of the $AsCl_3^{\bullet+}$ (m/z 181) fragment. The pathway is observed in a number of trajectories, where it proceeds by a 1,3-chlorine atom shift in the parent ion. The most abundant fragment (base peak) is found to be $AsCl_2^+$ (m/z 145), in both the computed and experimental spectrum. The structure of the $C_2H_2Cl^+$ (m/z 61) fragment results from yet another case of H atom migration, where an H atom is transferred to the terminal carbon atom of the fragment.

For triphenylstibine (**19**, Fig. 8.8c)), we observe several artifacts in the computed MS, primarily of low abundance. The dominant peak series in the MS corresponds to the parent ion and subsequent dissociation of phenyl groups. The base peak of the computed spectrum

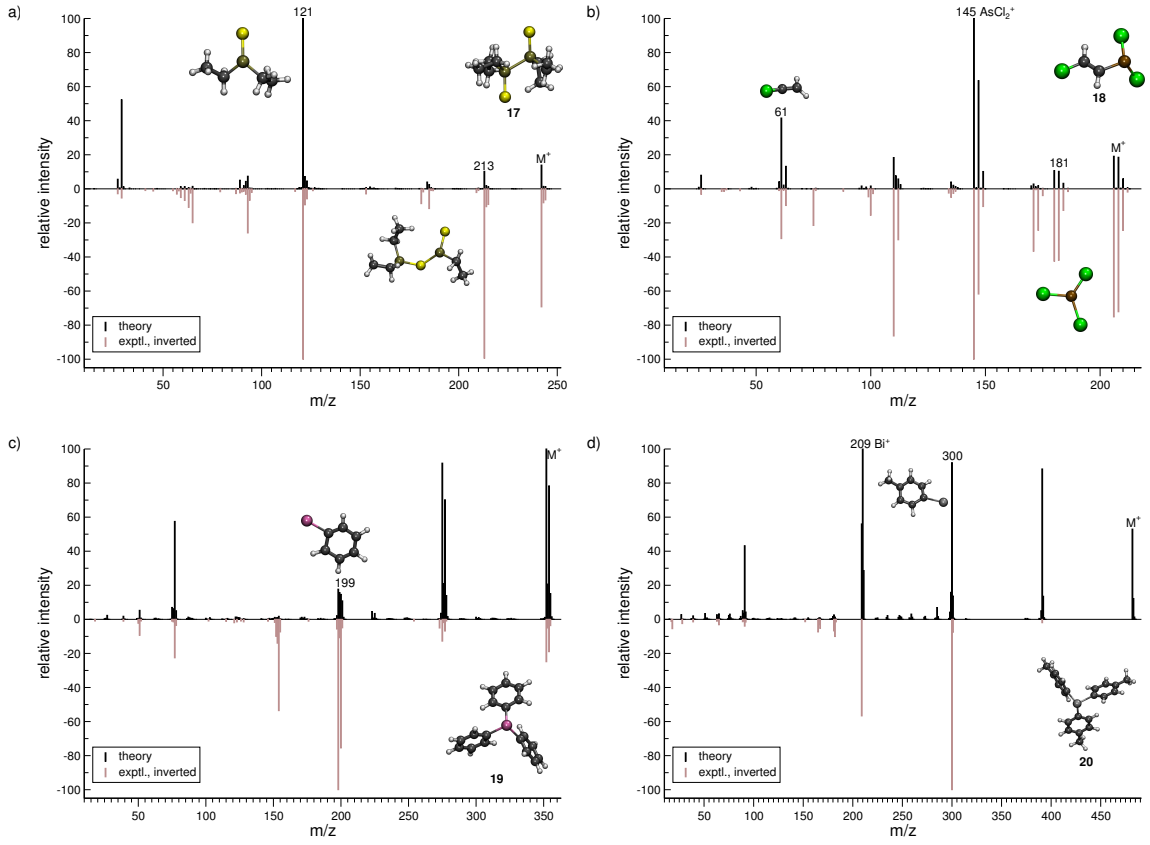

Figure 8.8.: Comparison of computed and experimental EI-MS for group 15 inorganic molecules, **17**–**20** in a) to d), respectively. The structures of the parent ion (denoted by M^+) and selected ionic fragments have been superimposed on each computed spectrum. Moreover, the selected ions are marked by the respective m/z values and discussed in the text.

is found to be the parent ion, while the experimental base peak corresponds to the fragment $C_6H_5Sb^{\bullet+}$ (m/z 199), *i.e.*, the survival rate of the parent ion is too high in the simulations. However, we find the m/z 199 peak to be statistically significant in the computed spectrum, even if it is not the main peak. The quality of the calculation in this case, may be considered as mediocre. However, it is still useful for molecular identification because the computed spectra accurately captures the characteristic peak series, of subsequent phenyl group losses.

For tris(para-tolyl)bismuthine (**20**, Fig. 8.8d)), the two spectra do not compare so well, with numerous low-intensity artifacts found in the computed spectrum. The parent ion is absent in the experimental spectrum, while being a relatively large signal theoretically. The characteristic peak series is represented by subsequent tolyl losses from the parent ion, and this peak series is reproduced in the computed spectrum. However, the first peak in the series has a much larger intensity in the computed than in the experimental spectrum. The second peak (m/z 300) forms by the loss of two para-tolyl groups, yielding the $C_6H_5Bi^{\bullet+}$ fragment. It is observed as the base peak in the experimental spectrum, while the last peak of the series is predicted to be the base peak in the computed spectrum. This peak corresponds to the naked Bi^+ cation, where the parent ion has lost all of the para-tolyl substituents. We honestly include (**20**) as an example of cases with relatively bad correspondence between theory and experiment. If this is rooted in an inaccurate parametrization of bismuth (in GFN-xTB) or related to some other problem, specific to very heavy elements, it will have to await further investigation.

Group 16 (21–23)

For the most prevalent allotrope of sulfur, cyclic S_8 (**21**, Fig. 8.9a)), the agreement between the experimental and computed spectra is found to be good. However, there are two peak-signals missing in the computed spectrum, the ion formed after a loss of a single S atom (m/z 226) and the S^+ ion (m/z 32), which most likely are complementary to one another. The remaining signals are captured by the simulations, where the fragment $S_6^{\bullet+}$ (m/z 192) adopts a cyclic structure and $S_4^{\bullet+}$ (m/z 128) is an open-chained structure. The base peak in the MS of (**22**) is correctly found to be m/z 64, corresponding to the $S_2^{\bullet+}$ fragment ion.

The quality of the simulation for Se_6 (**22**, Fig. 8.9b)) is comparable to that of **21**. The ion formed after a single Se atom loss (m/z 395) and the Se^+ ion (m/z 79) are underrepresented and missing in the simulations, respectively. The fragment $Se_2^{\bullet+}$ (m/z 158) is correctly found to be the base peak in the computed spectrum. Interestingly, the fragment m/z 316, $Se_4^{\bullet+}$, may assume a cyclic structure based on the visualization of the trajectories. This reaction could be investigated at a higher level of theory. Overall, the computed and experimental MS are in reasonable agreement.

The final molecule of this study is diethyltelluride (**23**, Fig. 8.9c). We observe a moderate agreement between the computed and experimental MS. Numerous peaks are missing in the

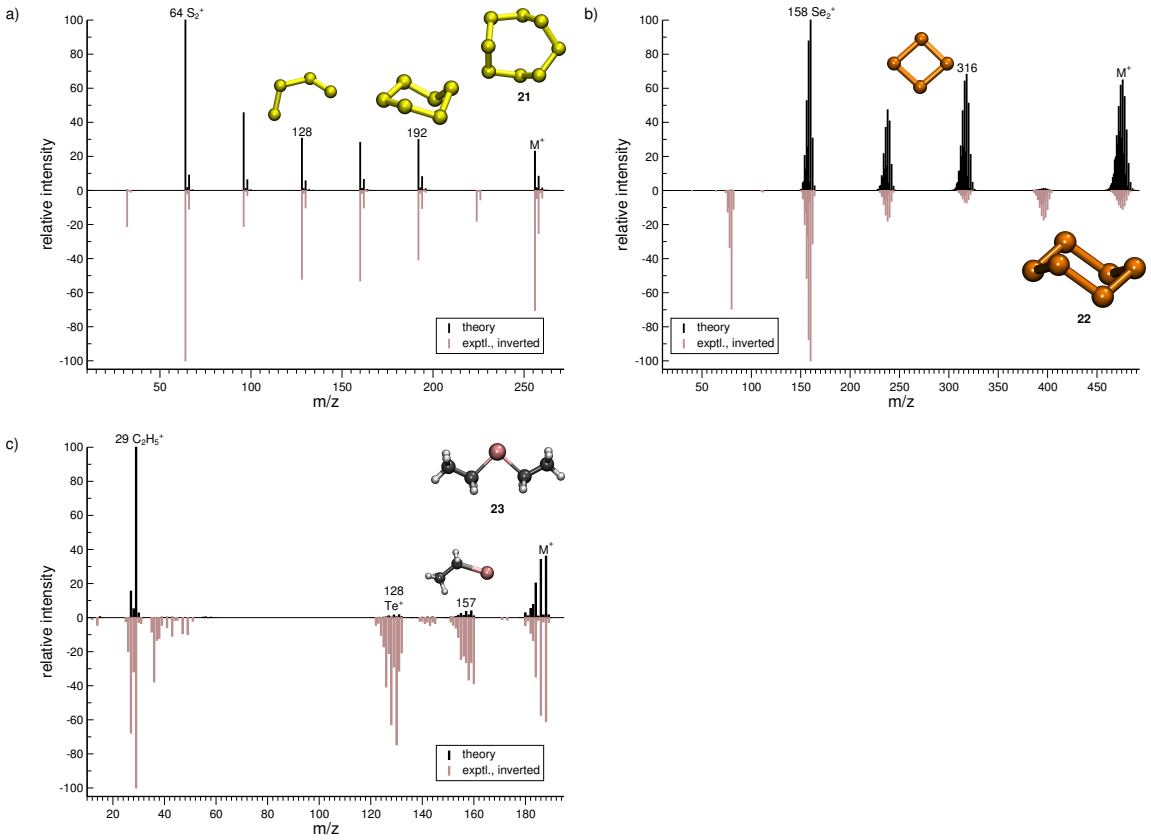

Figure 8.9.: Comparison of computed and experimental EI-MS for group 16 inorganic molecules, or molecules **21**–**23** in (a) to (c), respectively. The structures of the parent ion (denoted by M^+) and selected ionic fragments have been superimposed on each computed spectrum. Moreover, the selected ions are marked by the respective m/z values and discussed in the text.

computed MS *e.g.*, the fragments observed with m/z around 40 and 142. What is more, the intensities of few signals are drastically underrepresented. However, the intensity of the parent ion is correctly computed. Moreover, the simulations are able to reproduce the base peak m/z 29, which corresponds to the fragment $C_2H_5^+$, as well as the naked Te^+ (m/z 128).

For the main group inorganic molecules, **11**–**23**, the computed MS generally compare relatively well to the experimental spectra. The main peak series is usually fully reproduced by the simulations, even though the intensities can be somewhat inaccurate. We consider the applicability of the MS(GFN-xTB) combination to be evident from the computed EI-MS and think that the method can be convincingly applied to a large variety of molecular systems, comprising main group elements. Of course, the quality and furthermore the faults of the computed EI-MS will differ from one molecule to another, *e.g.*, the parent ion is predicted to be too stable for molecules **13**, **15**, **16**, **19** and **20**, several artifacts are found in the MS of **19** and **20**. The computed MS of **12**, **21**, **22**, **23** have a few missing peak-signals. It is important to note that all of the EI-MS are simulated using fixed conditions and nothing has been 'cherry-picked'. In general the quality of the spectra can be slightly improved by varying the simulation conditions for each case (mainly average IEE and simulation time).

8.4. Conclusions

We have implemented the recently developed, special-purpose, GFN-xTB and IPEA-xTB semi-empirical methods in QCEIMS, making `qceims` fully operational without inclusion of any third-party software. It is now applicable to molecules composed of elements with atomic numbers up to Z=86. The methods are devised to accurately compute *e.g.,* atomic forces and IPs, respectively, in a computationally efficient manner. The main method GFN-xTB which provides the PES for all occurring reactions was *not* modified for the present purpose. Because of their robustness and computational efficiency, GFN-xTB and IPEA-xTB are ideal to use in conjunction with QCEIMS. To evaluate the performance and transferability of MS(GFN-xTB), we have simulated EI-MS for 23 chemically diverse molecules. The molecules are divided into three groups comprising organic (**1**–**6**), organometallic (**7**–**10**) and main group inorganic molecules (**11**–**23**). Such extensive quantum chemistry calculations of EI-MS for molecules across the periodic table is unprecedented.

There were roughly 270 million single point energy and gradient calculations conducted for this study, using the GFN-xTB method (the number of IP evaluations is included in the count, but is negligible). We find GFN-xTB to be remarkably robust with typically less than 2 % unsuccessful production runs. Furthermore, because of its good convergence properties, GFN-xTB is extremely fast, where MS(GFN-xTB) performs, on average, around 20 energy/force evaluations per (real-time) second, irrespective of the given molecular size (up to 49 atoms as in **6**) and composition. It is evident that MS(GFN-xTB) is both robust and computationally

efficient, enabling exhaustive simulations of EI-MS for the first time.

For the organic molecules, the MS(GFN-xTB) computed MS compare generally well to the respective experimental spectra. The GFN-xTB computed spectra are of comparable quality to those published previously for similar compounds, using related semi-empirical methods in tandem to QCEIMS.[233] The visualization of the simulation trajectories reveals a number of interesting reaction pathways. As an example, for **1** and **2**, H atom migration to a terminal carbon atom is observed. For **3** the simulations are able to reproduce the McLafferty rearrangement and for **4**, an oxygen atom is transferred to a carbon atom, prior to NO loss. Interestingly, the peak-signal m/z 77 is observed in the MS of both **4** and **5**. For **4** this peak is solely ascribed to the cyclic phenyl cation (m/z 77). However, for **5** the peak results from various acyclic isomers of $C_6H_5^+$ and the cyclic phenyl cation. We conclude that one can convincingly simulate electron ionized fragmentation pathways of organic radical cations using MS(GFN-xTB).

The novel prediction of EI-MS of the organometallic molecules is achieved at the typical speed of semi-empirical QC calculations, for the first time. The quality of the computed MS of **8** is striking considering the complexity of the problem. The quality of the spectra of **7**, **9** and **10** is not as good, but can be considered satisfactory. For **7** the parent ion is found to be too unstable in the simulations, resulting in over pronounced fragment peak intensities. Furthermore, the spectra for the latter three molecules exhibit a few artifacts, or false-positive peak-signals. Nevertheless, the accuracy which MS(GFN-xTB) attains in the prediction of EI-MS for organometallic molecules is hard to achieve, even by simulations conducted using standard DFT methods. We hold that the quality of the MS is sufficient to enable investigations into the various fragmentation pathways of organometallic cations. Adjustments of the electronic parameters in the GFN-xTB Hamiltonian in particular for the transition metal complexes could further improve the quality of the theory. We stress that the fragment IPs play a pivotal role in the QCEIMS procedure, where they are used to determine the statistical charges and hence the peak intensities. Therefore, in the case of electronically complicated transition metal complexes, we advocate the use of hybrid DFT for the computations of IPs. On the downside this can drastically increase the overall simulation time, depending on the molecule under study.

The computed EI-MS of the inorganic main group molecules, (**11**–**23**), further attests to the transferability and accuracy of the MS(GFN-xTB) approach across the periodic table. Generally, the computed and experimental spectra compare relatively well. Therefore, the procedure allows for an unprecedented and unbiased insight into the fragmentation pathways of inorganic main group molecules. As an example, the simulations are able to capture many interesting reaction pathways *e.g.,* the formation of the biphenyl cation from **14**, the rearrangement and formation of a P–S–P bond subsequent to an ethyl radical loss from **17** and 1,3-chlorine atom shift of the cation of **18** required to form the $AsCl_3^{\bullet+}$ fragment. The worst

agreement between the computed and experimental spectra is observed for **12** and **23**. Also, for **23** some of the peak intensities are severely underestimated. We find that for alkylated and arylated compounds (**13**, **15**, **16**, **19**, **20**) the parent ion appears to be artificially too stable in the simulations, ascribed to an interplay of the simulation time and IEE.

Indeed, the typical errors in a computed MS are missing peaks, inaccurate intensities, artifacts and too high stability of the molecular parent ion. In most cases the MS can be improved by varying the simulation conditions, until an optimum spectrum is produced. More importantly, the implementation of GFN-xTB and IPEA-xTB (in QCEIMS) allows for further improvements to an individual MS, where the methods can be easily be re-parametrized to high-level reference data, for the system of interest. This direction and its impact on the quality of spectra will be explored in forthcoming work.

It has to be kept in mind that GFN-xTB and IPEA-xTB are semi-empirical methods and thus retain the fundamental deficiencies introduced by, *e.g.,* the parametrization, integral approximations and small basis sets. Therefore, it is to be expected that the MS(GFN-xTB) approach fails for some systems. Examples where the theoretical MS are of unacceptable quality, are given in appendix D. We also compare MS(GFN-xTB) to the semiempirical DFTB3-D3 and PM6-D2H[314,335] PES in three illustrative cases in appendix D, where GFN-xTB has given clearly superior results. We submit that the apparent accuracy of GFN-xTB for the majority of cases stems from error cancellation of systematically too deep potential wells leading to high barriers and the inherent TB self-interaction error which works in the opposite direction. Preliminary results comparing the GFN-xTB to high-level *ab initio* PES have been obtained and will be presented in a forthcoming report.

Nevertheless, the surprisingly high quality of the simulated EI-MS enables a fast overview of the unimolecular fragmentation space for a wide variety molecules. On these terms, one important aspect of MS(GFN-xTB) is the screening of possible reaction pathways, which are then later refined at a higher level of theory, thereby, avoiding prior knowledge (or assumption) of reaction channels. This may lead to the discovery of new reaction types and the elucidation of reaction mechanisms, especially concerning gas phase ion chemistry of transition metal complexes, for which DFT calculations are in high demand.[317,318,319,320,321]

The expansion of QCEIMS to simulations of electrospray ionization/collision induced dissociation (ESI/CID) mass spectrometry, techniques, where the initial conditions for an MD based theoretical QC treatment are more well-defined than for EI-MS, is underway in our laboratory.

Part IV.

Summary and Conclusions

In this thesis, I have reported on the progress concerning the computation of EI mass spectrometry using the QCEIMS method in conjunction with finite-temperature DFT and semi-empirical quantum chemical potential energy surfaces. The combination of statistical and BO-MD elements has proved to be versatile and robust for many classes of molecules. Figure 8.10 recalls the example given in Figure 1.4. It shows the experimental EI mass spectrum of limonene in comparison with the QCEIMS/GFN-xTB results, which is the state of the art. While there is certainly room for improvement, the parallel computation of this spectrum has only taken a computational wall time of one hour, which is a great achievement in my opinion. The aim set in Chapter 1 to advance the QCEIMS computational methodology and to thereby show that the calculation of EI mass spectra can become a routine task has therefore been accomplished.

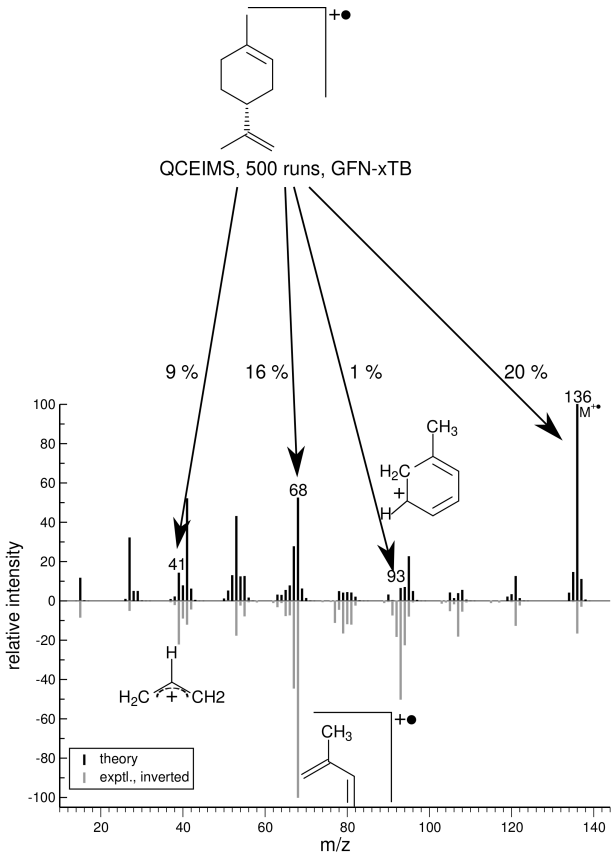

Figure 8.10.: Computed EI mass spectrum of limonene in comparison with the experiment (inverted, gray).[243] Selected fragment ions are displayed. The values in % indicate the number of QCEIMS production runs represented in the respective theoretical peaks.

Regarding the results presented in the previous Chapters of this thesis, I can draw the following conclusions:

1. The individual parts of the QCEIMS procedure are well-balanced and lead to a remarkable quality of theoretical EI mass spectra. The ground state sampling provides good initial conditions, along with the $P(E)$ estimate, which works surprisingly well. Most importantly, semi-empirical quantum chemistry methods can be employed to generate the PES on the fly, and the resulting computed EI mass spectra are still adequate. This means that the relative ordering of the competing fragmentation processes – thus, the relative values of $k_i(E)$ – are accurate enough at semi-empirical levels of theory for larger systems, while the calculations are still affordable at minutes to hours of computational wall-time per full spectrum when run in parallel.

2. The usage of finite-temperature DFT and semi-empirical methods is crucial for the robustness of the approach. As shown in Chapter 3, FT-DFT and FT-TB methods model static electron correlation effects which are often of great importance in the description of molecular dissociation processes. More importantly in the context of BO-MD, the Fermi Smearing procedure almost invariably leads to SCF convergence. Therefore, as reported in detail in Chapter 8, the number of failed BO-MD trajectories within QCEIMS is typically below 1 %. This low failure rate is encouraging as it avoids any additional methodological bias towards the description of electronically "well-behaved" processes. The successful computation of the EI mass spectra of the organometallic complexes in Chapter 8 is particularly encouraging since it has shown that the GFN-xTB method is even more flexible and robust than originally envisaged. The EI mass spectra of these compounds cannot be computed using standard DFT due their difficult electronic structures and SCF convergence issues.

3. QCEIMS in conjunction with semi-empirical PES is applicable to molecules as large as Taxol (113 atoms) in a reasonable computational wall-time of roughly one day for 1,000 parallel production runs. The quality of the predicted mass spectrum for such a molecule is astonishing given the thousands of possible fragmentations. This points to the advantage of the QCEIMS method over previous theories, where the reaction coordinates have to be known *a priori*.

4. The application of QCEIMS to the fragmentation of the nucleobases has revealed rich mechanistic details of their electron-ionization induced reactions. Several of these pathways are confirmed by comparison to isotope-labeled EI-MS experiments. This finding has the crucial implication that QCEIMS provides a purely *in silico* way of exploring the unimolecular reaction space. From the simulation data, reaction coordinates can be

extracted, which can in turn serve as the input for further quantum chemical calculations or chemical information frameworks.

5. Finally, the implementation of the GFN-xTB Hamiltonian has allowed for the fast prediction of mass spectra for compounds containing elements up to $Z = 86$. The GFN-xTB PES performs extremely well within the QCEIMS approach for organic, inorganic, and organometallic systems. This means that the electron-ionization induced fragmentations of compounds like ferrocene can be simulated using a BO-MD approach for the first time. The electronic structures of these systems are often so complicated that any BO-MD approach based on DFT (even FT-DFT) runs into SCF convergence problems. The incorporation and thorough evaluation of the GFN-xTB Hamiltonian into the `qceims` program, performed in collaboration with Viljhálmur Ásgeirsson, allows for the stand-alone usage of the program.

These findings are encouraging for the the ongoing endeavors to compute EI mass spectra in the context of structure identification of small molecules.[336] The following two major perspectives seem to be the most promising:

1. A large-scale benchmark evaluation of the GFN-xTB method for the possible generation of *in silico* libraries of EI mass spectra of organic molecules. If the quality of GFN-xTB computed mass spectra holds for a broad range of organic molecules, this could be of great usefulness for analytical chemists performing gas chromatography/EI-MS measurements. A thoroughly benchmarked GFN-xTB computed mass spectral library could serve in the future as a means of structure elucidation – as is the case today with large libraries of experimental EI mass spectra.

2. The QCEIMS approach of combining statistical elements with BO-MD can be extended to other mass spectrometry methods. There exist a variety of ionization methods which lead to a protonated molecule $[M+H]^+$ as the parent ion species. These methods are typically much "softer", *i.e.*, they cause much less fragmentation. The method of Collision Induced Decay (CID), where the $[M+H]^+$ species is accelerated to hit a neutral, inert gas (N_2 or Ar in most cases), leads to few, but characteristic and therefore analytically valuable fragmentations. In order to realize this perspective, a method of automatically predicting the $[M+H]^+$ protonation site has been devised and is submitted for publication. Additionally, a simulation protocol for the CID fragmentations has to be implemented and tested.

The results presented in this thesis have proved that the first-principles based calculation of EI mass spectra is no longer a fantasy but can be carried out on a moderately sized computer cluster for molecules composed of about 100 atoms. If the two perspectives enumerated

above become reality, then QCEIMS and related approaches will be of great assistance to experimental analytical chemists. Characteristic peaks or peak series ("fingerprints") that vary from isomer to isomer, may then be routinely rationalized by quantum chemical/BO-MD calculations. The future for computational mass spectrometry seems bright.

Part V.

Appendix

A. Supporting Information to Chapter 1

Appendix A contains:

- Additional Data on CAS-Selection
- Description of the used Structures

The structures themselves are contained in a file called `structures.zip`, which is available online as Supporting Information to

Christoph Alexander Bauer, Andreas Hansen and Stefan Grimme Chem.– Eur. J. **2017**, *23*, 6150–6164, DOI 10.1002/chem.201604682

Additional Data on CAS-Selection

Table A.1 lists the FT-DFT occupation numbers for Figure 6, calculated at the FT-BH-LYP/def2-TZVP level of theory at $T_{\mathrm{el}} = 15000$ K with the `ridft` program of the TURBOMOLE.6.6[63] suite of programs, using C_1 symmetry and the `m4` grid. It also lists the CASSCF/def2-TZVP occupation numbers, calculated with ORCA.

Table A.1.: Orbital occupations from FT-BH-LYP/def2-TZVP ($T_{\mathrm{el}} = 15000\,\mathrm{K}$) and CASSCF/def2-TZVP calculations. The gray cells mark the suggested choice of the active orbitals. CASSCF occupation numbers are NOONs in the active space, canonical orbital occupations (restricted Hartree-Fock wave function) otherwise. $w(|0\rangle)$ is the weight of the reference wave function.

	FT-BH-LYP occupations			CASSCF occupations		
index	C_4H_4 (D_{4h})	p-benzyne	retinal-90°	C_4H_4 (D_{4h})	p-benzyne	retinal-90°
LUMO+2	0.00	0.02	0.02	0.00	0.06	0.08
LUMO+1	0.00	0.06	0.03	0.00	0.07	0.09
LUMO	0.97	0.59	0.91	1.00	0.60	1.00
HOMO	0.97	1.57	1.16	1.00	1.39	1.00
HOMO-1	2.00	1.85	1.91	2.00	1.92	1.91
HOMO-2	2.00	1.90	1.96	2.00	1.93	1.91
N^{FOD}	2.03	1.36	2.56	–	–	–
$w(\|0\rangle)$	–	–	–	0.50	0.65	0.44

Description of the Structures

The structures that we have used for this study are supplied in a separate file, if they have not been obtained from other studies, which we have cited. Table A.2 describes the structures contained in `structures.zip`. The absolute energy of optimized structures is given at the level of theory at which the structure has been optimized in the second line of the `.xyz` files. If no file name for a structure is given in Table A.2, the citation lists from which study the original structure has been taken and – if applicable – at which level of theory it has been reoptimized.

Table A.2.: Details on the structures used for this study, see file `structures.zip`

molecule	file name	details
subsection: FOD as a Tool to Gauge Biradical Character		
Thiele's HC	`thiele.xyz`	TPSS-D3/def2-SV(P)
Tschitschibabin's HC	`tschitschibabin.xyz`	TPSS-D3/def2-SV(P)
DBHZ1	–	taken from the SI of Ref. 52[91] reoptimized at TPSS-D3/def2-SV(P)
DBHZ2	–	taken from the SI of Ref. 52[91] reoptimized at TPSS-D3/def2-SV(P)
TBP1	–	taken from the SI of Ref. 53[92] reoptimized at TPSS-D3/def2-SV(P)
1-7	–	taken from the SI of Ref. 44[83] reoptimized at TPSS-D3/def2-SV(P)
8	`8.xyz`	TPSS-D3/def2-SV(P)
9	`9.xyz`	TPSS-D3/def2-SV(P)
subsection: Selection of Active Spaces for Multiconfigurational Wave Function		
cyclobutadiene (D_{4h})	`cyclobutadiene-d4h.xyz`	TPSS-D3/def2-TZVP
para-benzyne	`p-benzyne.xyz`	TPSS-D3/def2-TZVP
retinal "90°"	–	contained as scanpoint no. 10 within `retinal-scan.trj` (see below)
anthracene	`anthracene.xyz`	TPSS-D3/def2-TZVP
pentacene	`pentacene.xyz`	TPSS-D3/def2-TZVP
heptacene	`heptacene.xyz`	TPSS-D3/def2-TZVP
$[Co^{II}(^2L_{NN}^{\bullet 1-})]$	`Co-bisphenylenediamine.xyz`	TPSS-D3/def2-TZVP
oxo-Mn salen	`oxo-Mn-salen.xyz`	TPSS-D3/def2-TZVP
$RuCl_4(Hind)(NO)^-$	`Ru-Cl4-HInd-NO.xyz`	TPSS-D3/def2-TZVP
subsection:FOD Analysis for Rotations around Double Bonds		
ethylene	`ethylene-scan.trj`	`ethylene-scan.trj` contains the structures of the 19 scan points.
TCNE	`TCNE-scan.trj`	`TCNE-scan.trj` contains the structures of the 19 scan points.
retinal	`retinal-scan.trj`	`retinal-scan.trj` contains the structures of the 19 scan points.
protonated merocyanine	–	taken from the SI of Ref. 89[127]
subsection:FOD as an Indication for Static Electron Correlation in Large Biochemical Systems		
CpdI minimal model	`CpdI-minimal_model.xyz`	TPSS-D3/def2-TZVP
1DZ9	`1dz9-prepped-maestro.pdb`	chain A only, saturated using maestro[135]
	`1dz9-prepped-maestro.xyz`	chain A only coordinates only
Cu active site of 5K49	`5k49-Cu-site.xyz`	coordinates from 5K49.pdb saturated using maestro[135]
5K49	`5k49-prepped-maestro.pdb`	saturated using maestro[135]
	`5k49-prepped-maestro.xyz`	coordinates only
2ASK	`2ask-prepped-maestro.pdb`	saturated using maestro[135]
	`2ask-prepped-maestro.xyz`	coordinates only

B. Supporting Information to Chapter 3

Appendix B contains:

- More details of the QCEIMS protocol as used for chapter 3
- Additional computational results

QCEIMS in more detail

For a comprehensive description of the inner workings of QCEIMS, the reader is referred to the SI of the original paper[10]. There, one of the authors (Grimme) has laid out the general QCEIMS architecture and discussed many technical details. What follows in this document is a more detailed overview of the procedure used for this study. The QCEIMS program was used in 3 steps:

1. **Initialization** - Production of a ground state trajectory to yield an ensemble from which to take snapshots (nuclear geometries and velocities) for the fragmentation runs. Optimized geometries were used as starting points. The velocity verlet algorithm[13] was used for propagation on the OM2-D3[33,228] potential energy surface (PES). For 1,000 trajectories, there were 25,000 steps of equilibration and 50,000 steps of production. The latter yielded the actual data from which the snapshots were obtained. With a timestep of 0.50 fs, this makes for 37.5 ps of ground state initialization. Geometry optimization at the given level of theory and initialization are actually the two most expensive steps in our procedure.

2. **Setup** - A random set of 1,000 equidistantly chosen nuclear geometry and velocity arrays was prepared and the ionization excess energy (IEE) was computed according to a Poisson distribution. Apart from $\varepsilon_{\mathrm{HOMO}}$ there was no information from the molecule itself. The distribution was calculated in such a way that the IEE/atom was equal to 0.6 eV.

3. **Production** - Each individual snapshot geometry was deprived of one electron, thereby rendering the systems radical cations. From this point, the propagation was started in the same way as for the closed-shell original system case, using either the OM2-D3

or the DFTB3-D3[136] PES. In the case of OM2-D3, unrestricted SCF calculations were carried out throughout every fragmentation run. The vibronic heating associated with internal conversion (IC) was simulated by scaling all nuclear velocities uniformly until the internal energy, predifined by the IEE value according to the Poisson distribution, was reached. This is a deviation from the protocol used in the original paper, where the velocities were scaled according to MO localizations (implications are discussed below). The heating time of the IC process was in the 1-2 ps time range. During the production runs, the 'electronic temperature' was set to a constant value of 21,000 K (OM2) or 5,000 K (DFTB3), yielding fractional occupation numbers (FON, 'Fermi Smearing')[27,44]. The FON approach had to be used in order to achieve SCF convergence[45]. Moreover, FON partially account for the mixing in of excited states and their multiconfigurational character. In the event of a fragmentation, a fragment assignment algorithm ordered the atoms according to their fragments. For each fragment, a mean geometry was computed from the trajectory. The ionization potentials (IPs) were then estimated at the OM2 level of theory for those mean fragment geometries. The charge was assigned statistically according to the Boltzmann distribution at the current (vibronic) temperature of the fragmentation event. The run was continued using the statistically most highly charged species, assigning the spin according to spin population analysis. This was done recursively up to a maximum of 7 times. The initial run maximum simulation time was 5 ps and decreased in each subsequent run. Given 4 or 5 runs during such a decomposition cascade, the typical total simulation time was between 5 and 10 ps.

At last, the fragments yielded by the production runs were counted with their statistical (Boltzmann-IP) weight and plotted vs. the experimental spectra (provided in either JCAMP-DX or peak table formats by the databases). The typical maximum count (100 % rel. int. signal in the computed spectra) was around 300. This is of course inferior to the experimental situation where there are millions of counts.

For OM2 calculations, the MNDO program[245] was called. DFTB+[136,246,337] was used for DFTB3 calculations.

The total computational times for spectra generation are found in Table B.1. These are only rough estimates of the timescale for QCEIMS calculations. The number of available cores on our computer cluster ranged from 100 to 500 and Intel CPUs were used as well as AMD CPUs. Note that up to this point, the 37.5 ps of ground state initialization is by far the most expensive step in our procedure with a wall time of about 1 day for systems **1-5**. Table B.1 also reveals that perhaps one of the strongest features of QCEIMS is the perfect parallelization - each production run is conducted totally independently from all the others. Thus, predictions for **1-5**, which are at the limit of experimental EI-MS, are accessible within

a few hours on a computer cluster and within a few days on a single workstation (assuming 8 modern CPU cores). It is also noteworthy that no part of the algorithm needs a lot of memory - provided semiempirical methods are used throughout. All in all, 8 to 16 GB RAM, which is standard nowadays, should be sufficient.

Table B.1.: Computational times (wall times) for QCEIMS production. Results marked (*) were produced on slow machines.

Compound	QC method	est. single core 1000 runs	average prod. run	total time
1	OM2-D3	51 d	50 min	4 h 0 min
1	DFTB3-D3	124 d	40 min	2 h 0 min
2	OM2-D3	266 d	176 min	13 h 0 min
2	DFTB3-D3	138 d	66 min	3 h 0 min
3	OM2-D3	308 d	546 min*	70 h 0 min*
3	DFTB3-D3	132 d	77 min	6 h 0 min
4	OM2-D3	56 d	53 min	8 h 45 min
4	DFTB3-D3	93 d	36 min	1 h 45 min
5	OM2-D3	69 d	54 min	2 h 45 min
5	DFTB3-D3	105 d	34 min	2 h 45 min

Additional computational results

Velocity scaling effect

The effect of localized heating by MO populations during a QCEIMS production run is dramatic for larger systems. As seen from the spectral comparison displayed in Figure B.1, a large number of artifacts results. This can be rationalized by inspecting the fragmentation runs. When heated locally, the Taxol frame breaks in certain positions very quickly, while other regions of the molecule stay 'cold'. If the charge is assigned to these 'cold parts', the runs are continued with these fragments, which then do not decompose further, leading to signals in the m/z 400-850 region. From the comparison with the experimental data one can see that this is wrong. While scaling the nuclear velocities uniformly leads to the reported computed spectrum of good quality, one cannot exclude other, non-beneficial effects of disregarding the localization of the ionization completely.

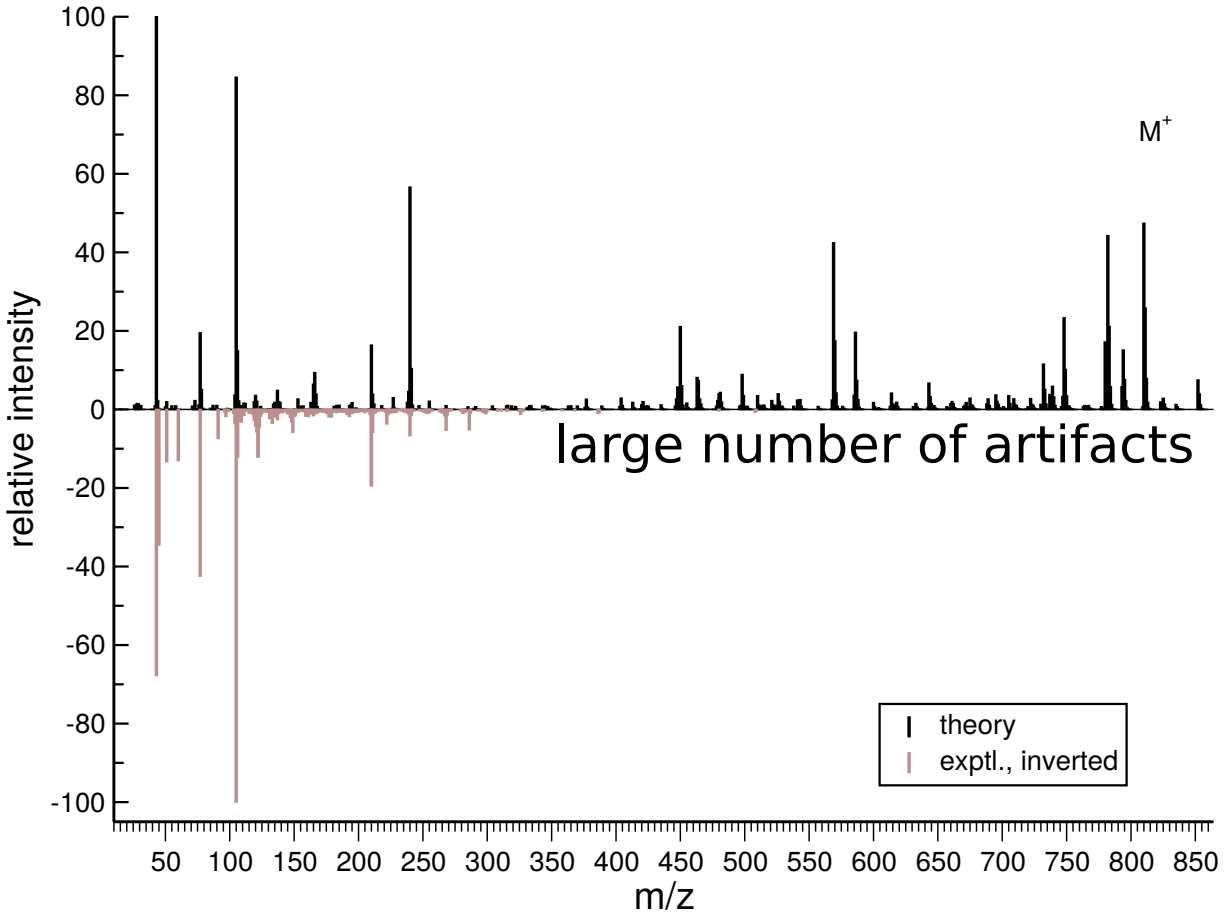

Figure B.1.: Computed (OM2-D3) vs. experimental spectrum of taxol (**3**) with local nuclear velocity scaling switched on. Note the artifacts in the heavier fragment region.

On the other hand, the computed spectra for other (smaller) compounds actually get better when the localization algorithm is switched on. However, in order to stay consistent, every reported spectrum (except where clearly noted) was produced with the localization routine deactivated.

IP estimate by OM2 - Comparison with DFT

Figure B.2 shows the difference for two computed spectra of **5**. For the first spectrum, fragment IPs were computed at the PBE0/SVx level of theory and for the second spectrum, OM2 was used for fragment IP calculation. The difference is marginal, with the maximum difference lower than 5 relative intensity units. One may thus safely assume that the semiempirical OM2 method is an adequate and efficient choice to gain a suitable IP estimate during a QCEIMS run.

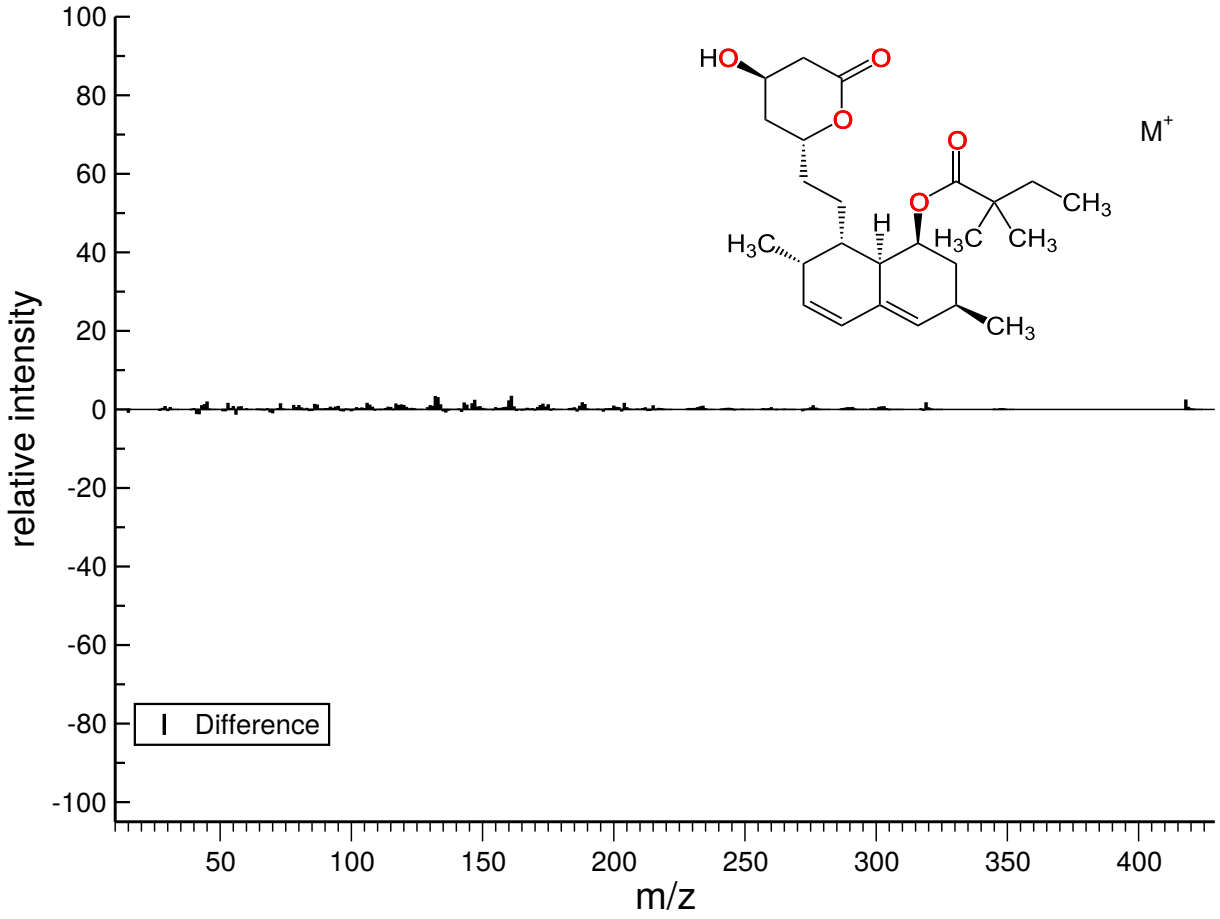

Figure B.2.: Difference between two computed spectra for Simvastatin: IP calculation with DFT (PBE0/SVx) - IP calculation with OM2.

DFTB3 spectra for 1-5

EI-MS spectra have been computed for compounds **1-5** using the DFTB3-D3 PES. As can be seen in Figures B.3-B.7, they are of similar or slightly worse quality to the OM2-D3 results. This shows that QCEIMS is in principle workable for many QC methods, producing results according to the PES generated by the method. In our cases, both DFTB3-D3 and OM2-D3 work well. The reason why OM2-D3 spectra are reported in the article is that OM2-D3 is better at describing intramolecular rearrangements such as the McLafferty reaction, see also original QCEIMS paper[10].

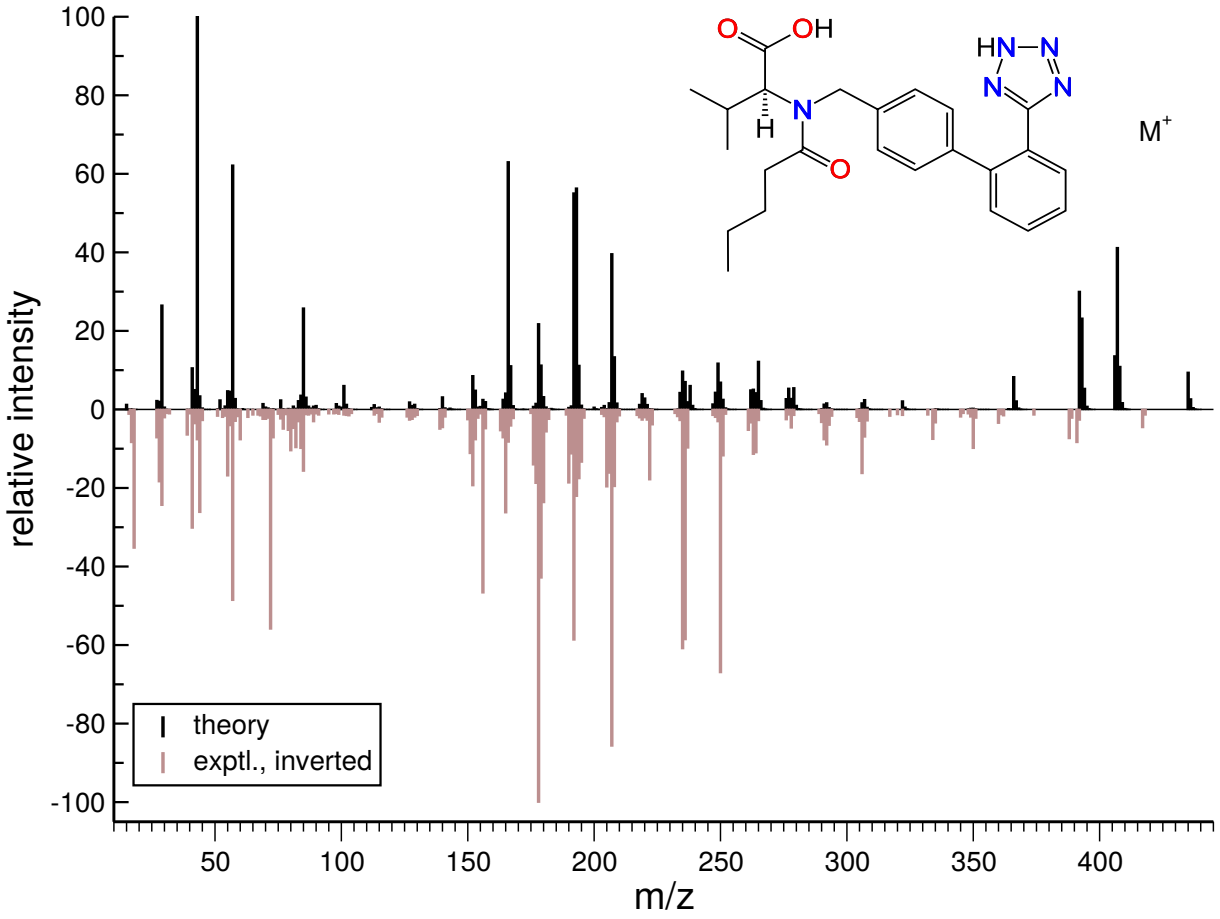

Figure B.3.: Calculated (DFTB3-D3) spectrum of **1** in comparison with the experimental spectrum.

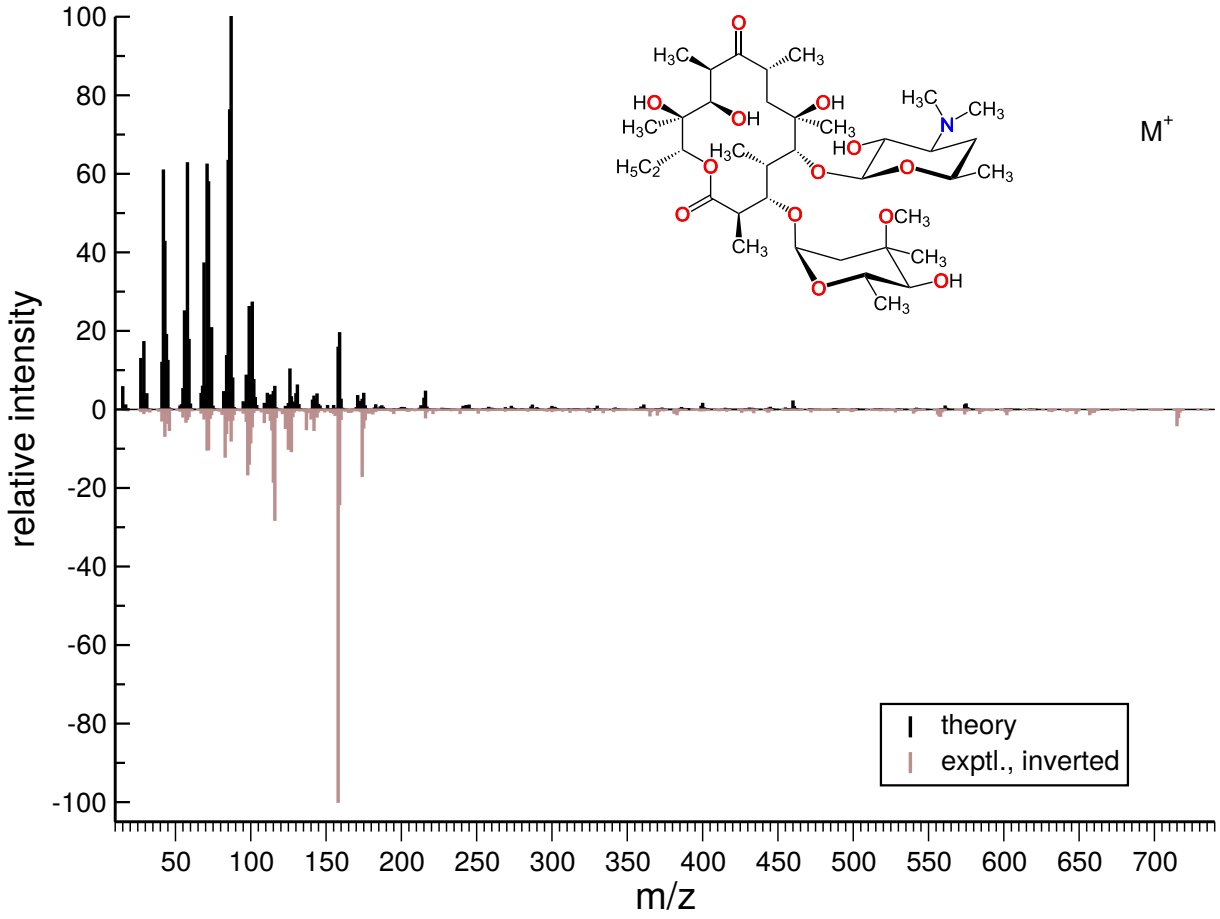

Figure B.4.: Calculated (DFTB3-D3) spectrum of **2** in comparison with the experimental spectrum. Note that the molecular ion (m/z=734) gives only a very weak signal in the experimental spectrum and none in the computed.

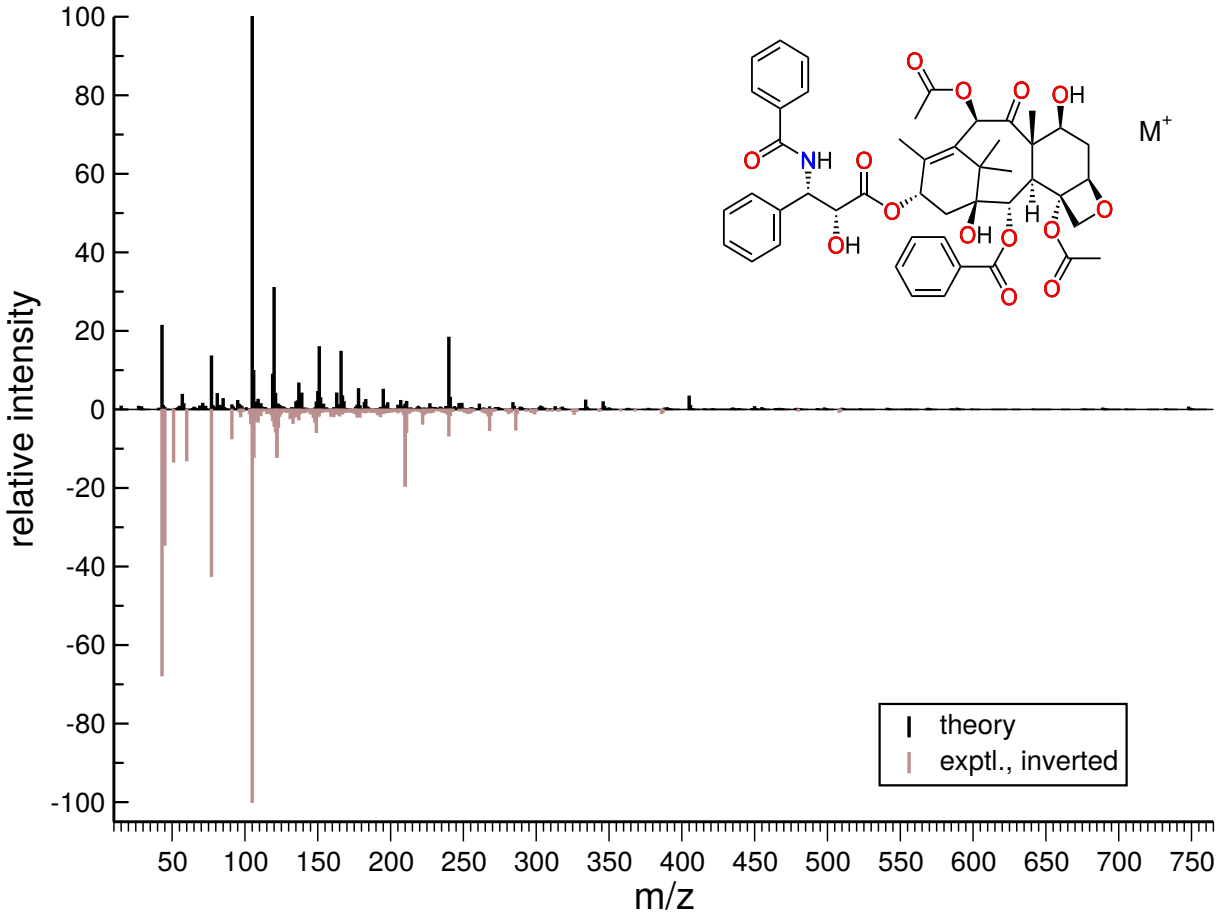

Figure B.5.: Calculated (DFTB3-D3) spectrum of **3** in comparison with the experimental spectrum. Note that the molecular ion (m/z=854) does not appear in either spectrum.

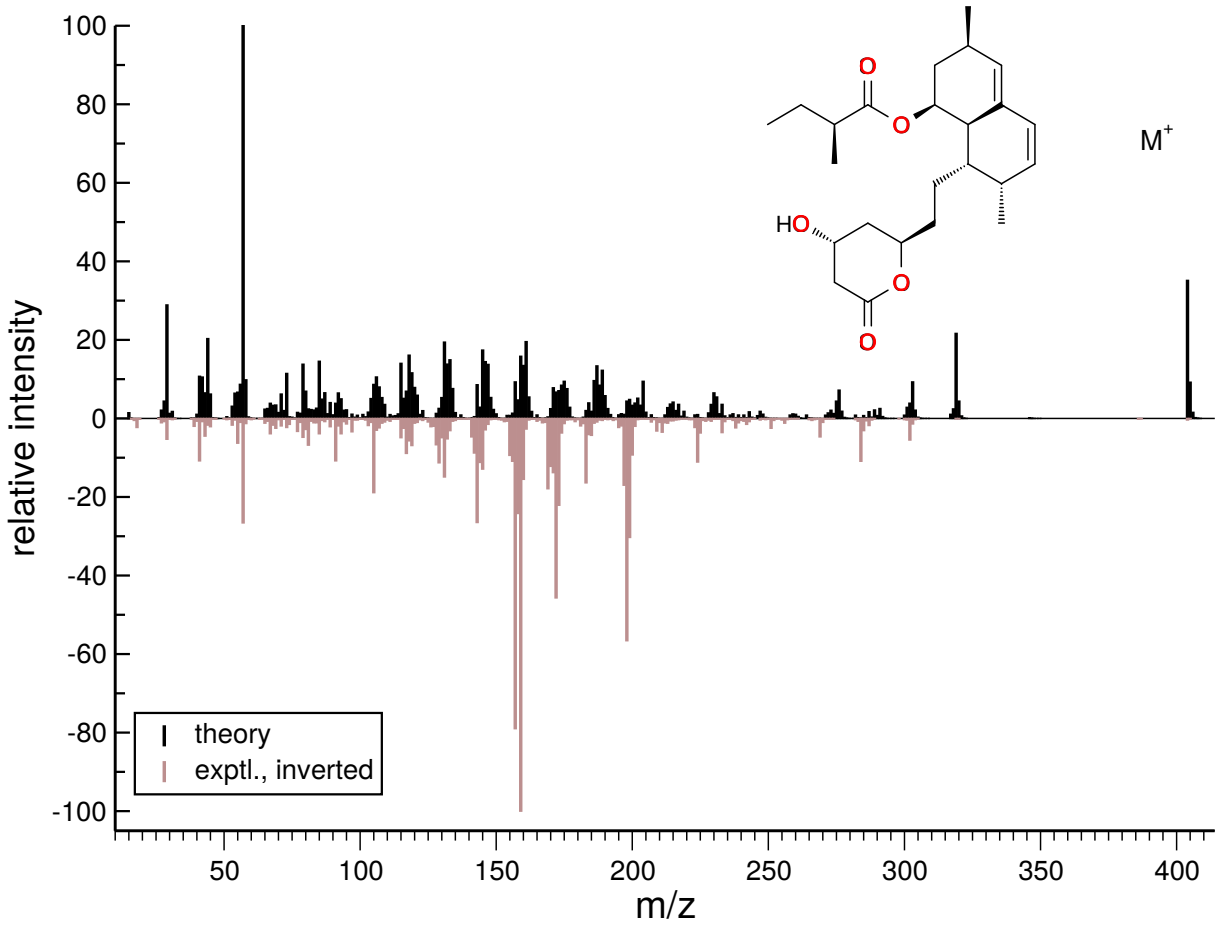

Figure B.6.: Calculated (DFTB3-D3) spectrum of **4** in comparison with the experimental spectrum.

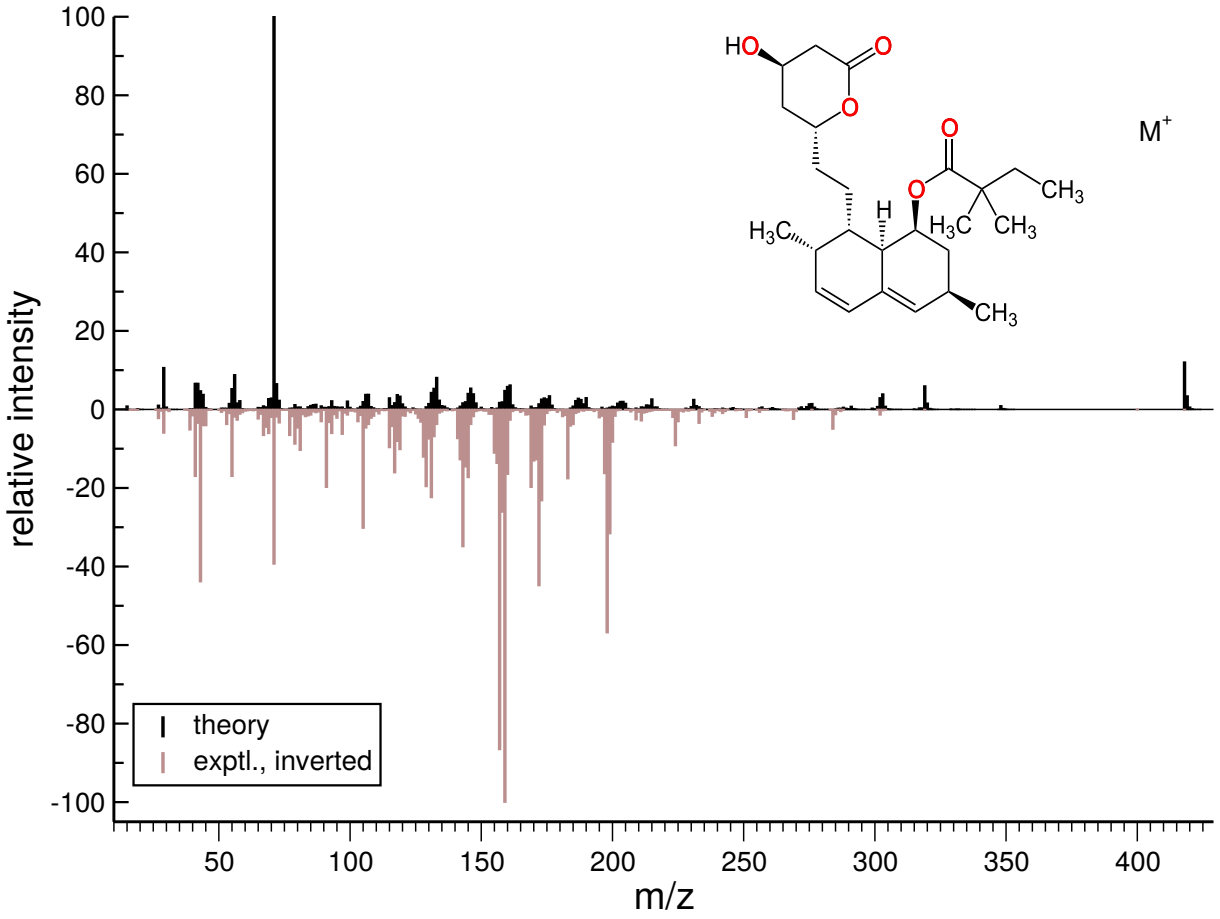

Figure B.7.: Calculated (DFTB3-D3) spectrum of **5** in comparison with the experimental spectrum.

C. Supporting Information to Chapter 4

Appendix C contains:

- Additional computed spectra
- MD analysis
- Example trajectory snapshots
- Mass spectral matching score

Additional Computed Spectra

In order to check for decomposition pathways occuring at a longer time scale, our calculations were re-run with a maximum simulation time of 100 ps on (i) the OM2-D3 (Figure C.1)and (ii) the DFTB3-D3 (Figure C.2) potential energy surfaces (PES). Additionally, the ionisation excess energy distribution was modified to allow for "milder" simulation conditions and identify possibilities for fragmentation reactions occuring late during our simulations.

The result of the test calculations with a longer time scale is that the quality of mass spectral prediction does not change significantly with increasing simulation times. The mass spectral match scores obtained from the computations with a maximum simulation time of 100 ps are almost the same as for the ones conducted over 5 ps, and the computed spectra look strikingly similar. The relative number of fragments detected by the QCEIMS algorithm at a simulated time of > 10 ps was less than 10 %.
DFT-D3 calculations over time scales of 100 ps are currently not computationally feasible, as the individual runs would consume several months of CPU time each.

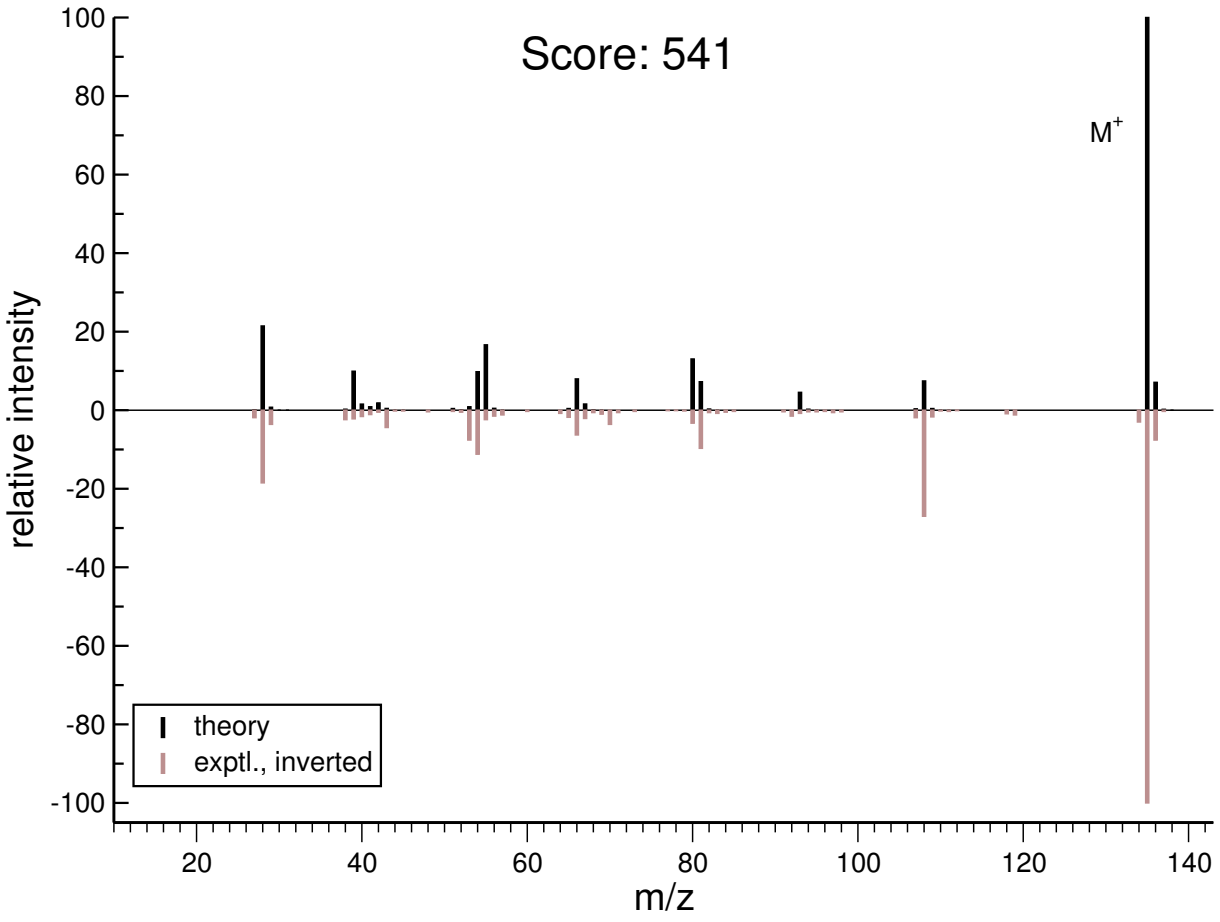

Figure C.1.: Calculated vs. experimental spectrum of **A**, OM2-D3 PES, with IEE/atom 0.3 eV, and a maximum simulation time of 100 ps.

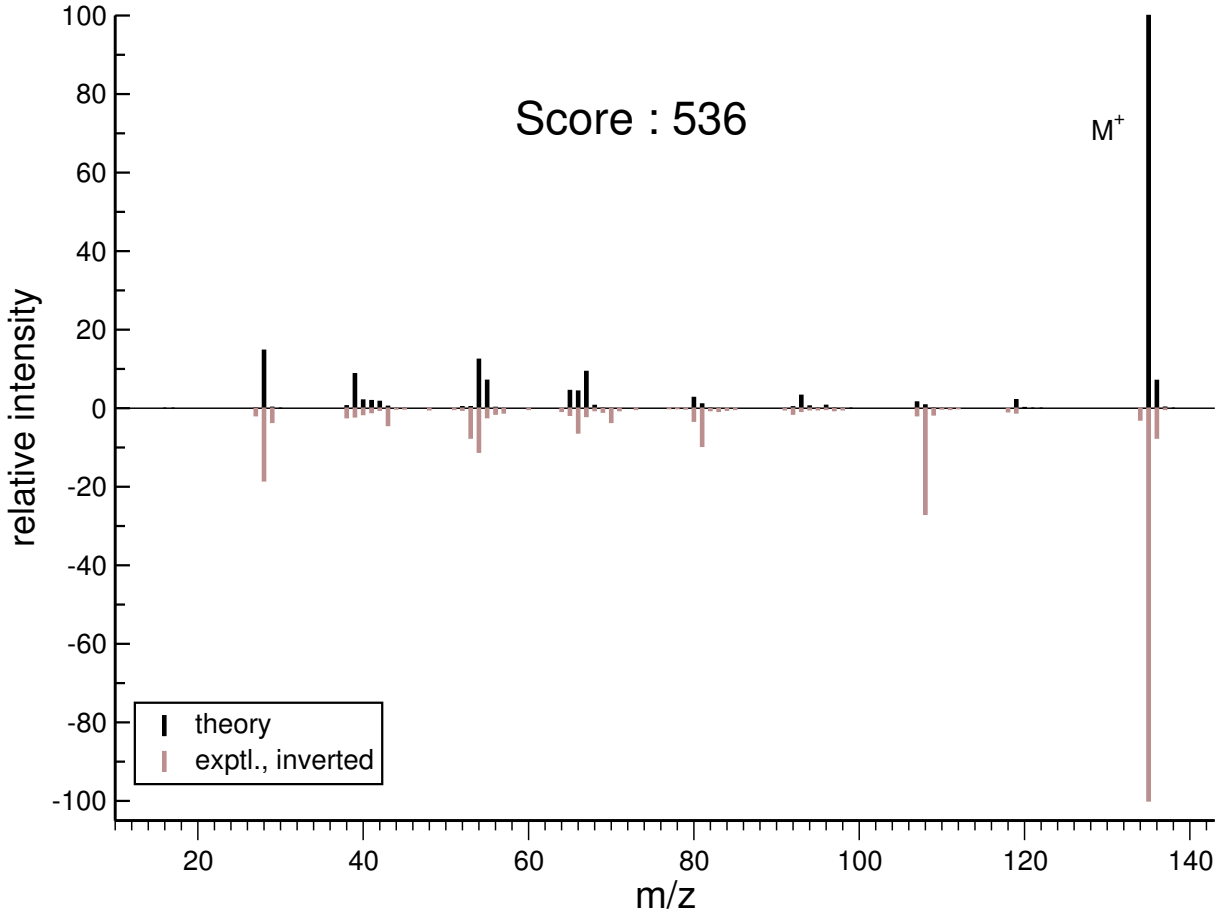

Figure C.2.: Calculated vs. experimental spectrum of **A**, DFTB3-D3 PES, with IEE/atom 0.3 eV, and a maximum simulation time of 100 ps.

MD Bond Length Analysis

Figures C.3, C.4, and C.5 show the analysis of interatomic distances during the initial fragmentation trajectories as given by the QCEIMS program output, ordered by quantum chemistry (QC) method. They hold information on two important matters:

1. The relative vulnerability of the bonds along the outer purine ring perimeter of **A** upon electron ionization. The more lines (trajectories) go to high distances in one specific subfigure, the more bond breaking events of the marked bonds are recorded by our algorithm. This information is then compared to predictions from non-dynamic calculations, see main text.

2. The time of the dissociative events. As one can see in all three figures, the vast majority of these occur within the first two picoseconds, regardless of the PES.

Comparison of Figures C.3, C.4, and C.5 reveals that there are many similarities between the three different PES generated by the respective QC methods. For example, Figures C.3 h, C.4 h, and C.5 h show that the C5–C6 bond is prone to dissociation within the first 2 ps of our simulations. However, the number of trajectories where this bond is broken decreases going from OM2-D3 to DFTB3-D3 to DFT-D3, as indicated by the lower number of lines going towards high distances in Figure C.5 h. This information is reflected by the diagram displaying the percentages of broken bonds along the purine ring perimeters in the main text.

Snapshots of an example trajectory

In order to highlight the complexity of the fragmentation network of **A**, we have taken snapshots from one exemplary fragmentation trajectory on the OM2[228]-D3 potential energy surface (PES), see Figure C.6. This Figure reveals that after 210 fs (Figure C.6 a) simulation time, the purine ring has broken in many places (C4–C5, N7–C8, and C4–N9). After 220 fs (Figure C.6 b), there is a proton transfer from a nitrogen (N9) to a carbon atom (C4), and the H–C8–N9 fragment dissociates. After 260 fs (Figure C.6 c), a second HCN (H–C4–N3) fragment is cleaved off. Finally, at 300 fs (Figure C.6 d), a third neutral loss of HCN (H–C2–N1) occurs, leaving the radical cation $C_2H_2N_2^+$, which is then counted as a contributor to the m/z=54 peak in the computed mass spectrum of **A**. The exemplary trajectory concurs with the data shown in the main letter, as the N5–C7 bond is not broken. Moreover, it is one of the 41 % of runs where the C5–C6 bond actually stays intact. This illustrative example indicates the intricacy of the fragmentation network of **A**. It also depicts and instance where the first HCN unit to leave in a temporal sense is not H–C2–N1.
Another important detail is the short simulated time scale on which the reactions happen. Apparently, it is not necessary to simulate for much more than 5 ps, given the fact that every

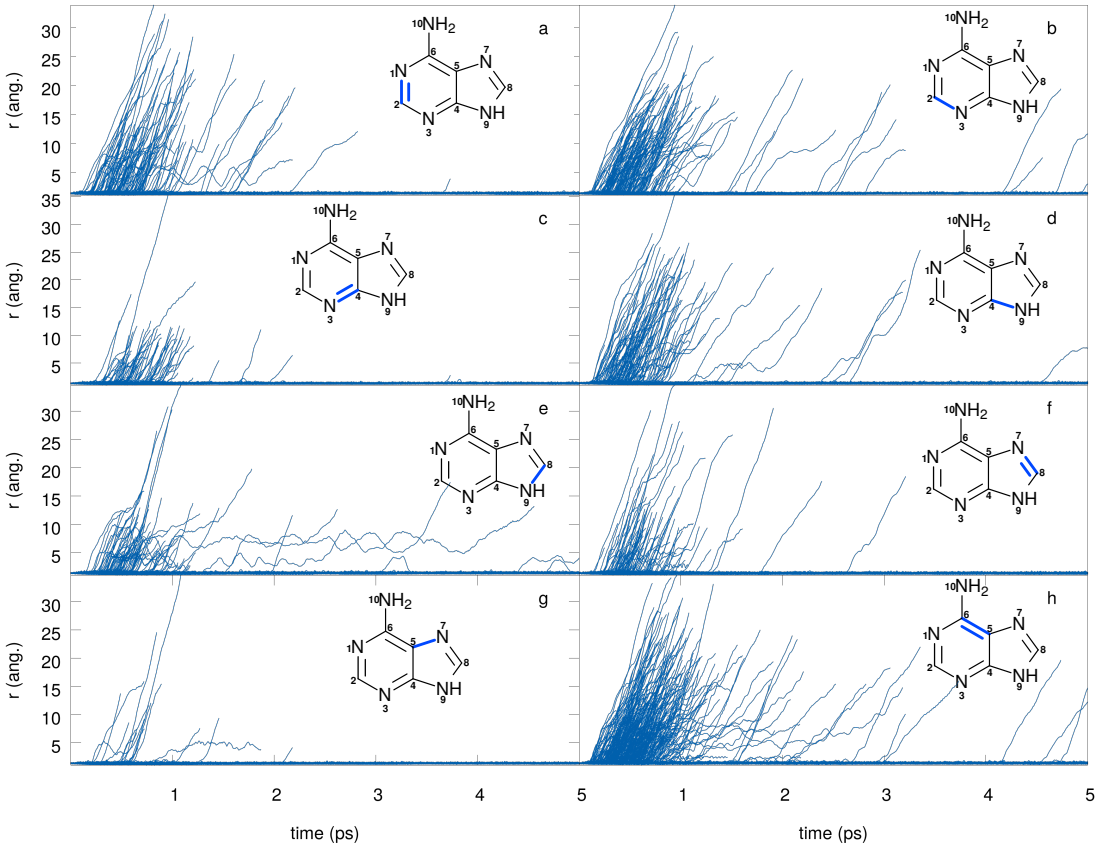

Figure C.3.: Bond lengths (in Å, indicated by the bold, blue lines in the inserts) along the outer ring perimeter of $\mathbf{A}^+$ during the initial fragmentation trajectories on the OM2-D3 PES. Abrupt ending of a line signifies a fragmentation event recorded by QCEIMS.

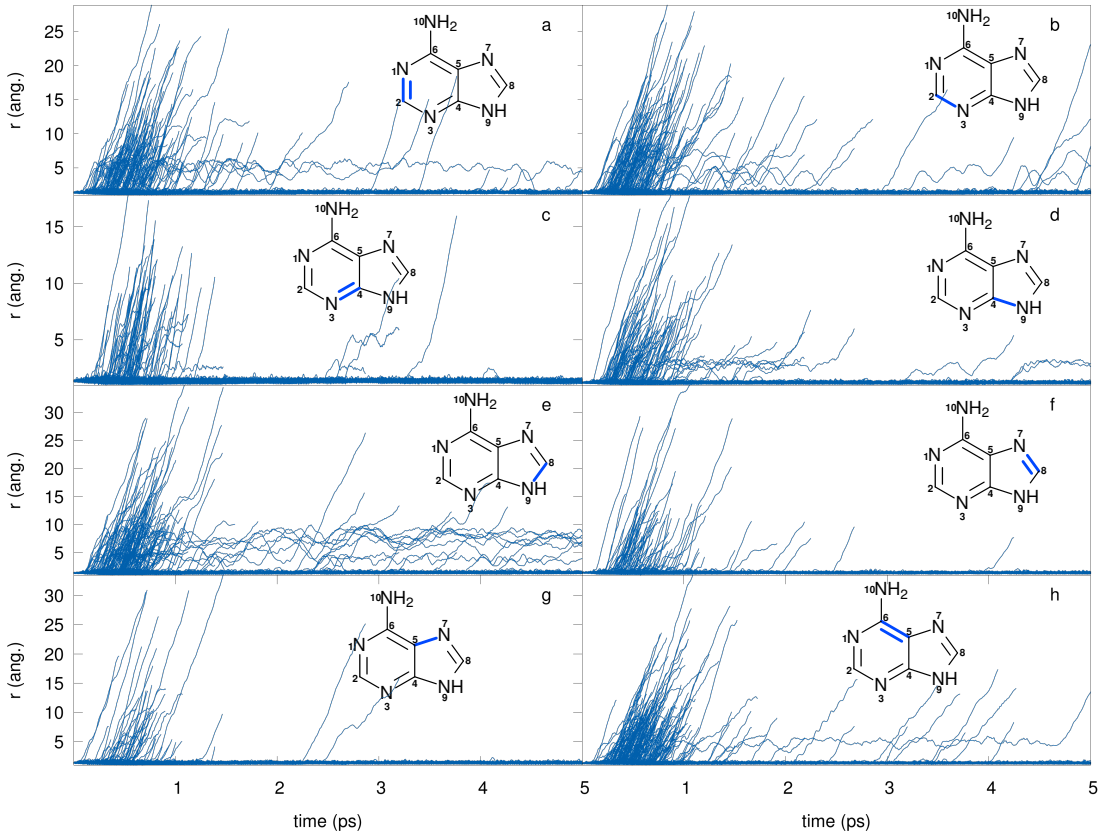

Figure C.4.: Bond lengths (in Å, indicated by the bold, blue lines in the inserts) along the outer ring perimeter of $\mathbf{A}^+$ during the initial fragmentation trajectories on the DFTB3-D3 PES. Abrupt ending of a line signifies a fragmentation event recorded by QCEIMS.

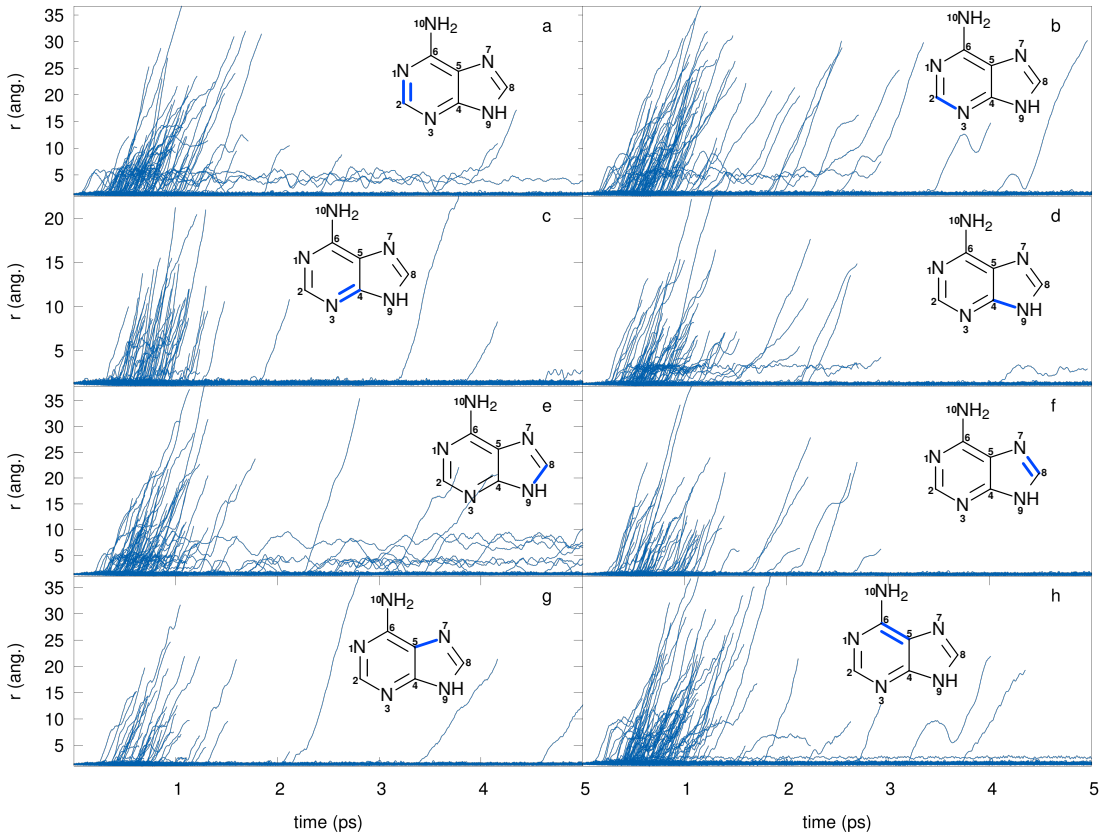

Figure C.5.: Bond lengths (in Å, indicated by the bold, blue lines in the inserts) along the outer ring perimeter of $\mathbf{A}^+$ during the initial fragmentation trajectories on the DFT-D3 PES. Abrupt ending of a line signifies a fragmentation event recorded by QCEIMS.

significant event in our calculations seems to occur before this threshold. This statement is also supported by the good quality of the theoretical mass spectra.

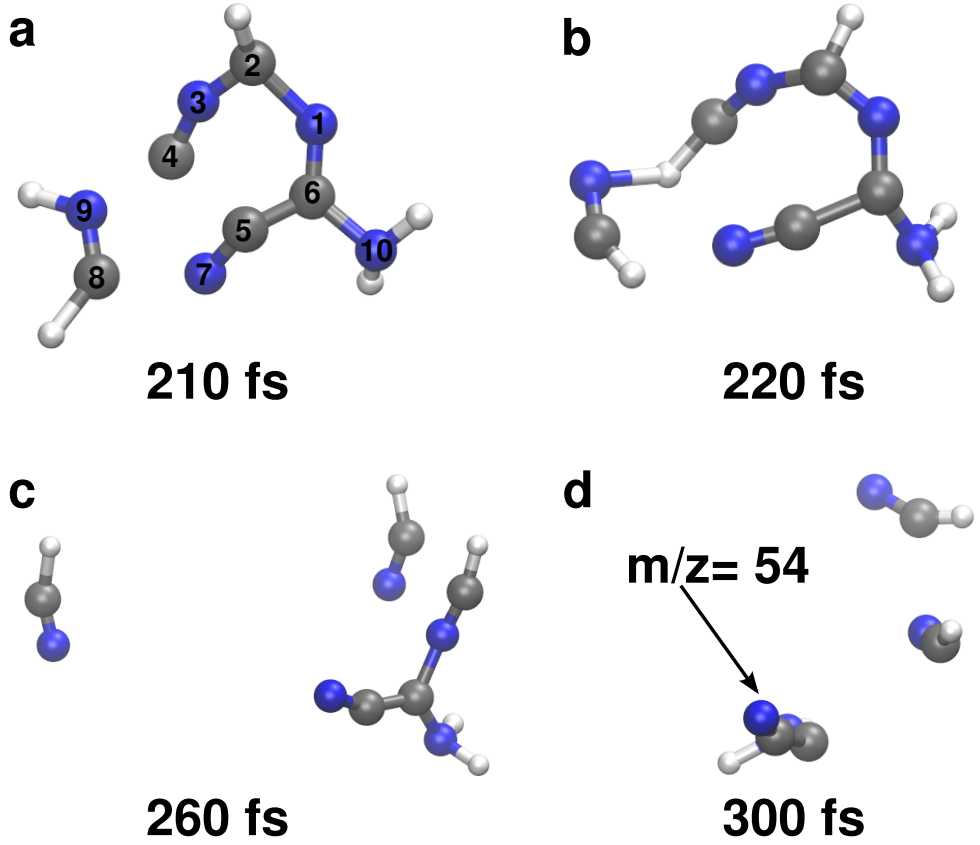

Figure C.6.: Snapshots with time index (simulated time) of an exemplary fragmentation trajectory of $\mathbf{A}^+$. Standard atom labels are displayed in the first frame (a) for orientation. The leftmost HCN unit in snapshot c is left out in snapshot d for clarity.

A sample of selected trajectories has been visualized as movies. They may be retrieved on the following website: http://www.thch.uni-bonn.de/tc/downloads/movies. In these movies, one can see various fragmentation runs on the OM2-D3 potential energy surfaces. The title of the video indicates which species carries the charge. In some of the videos, additional commentary is added to provide orientation.

Mass spectral matching score

In the mass spectrometry community, several kinds of quantitatively comparing of unknown spectra with spectral libraries (e.g. the NIST database) have been established.[9] We employ the "composite matching score" algorithm of the reference cited. Two mass spectra are compared. The first spectrum is the experimental one, the second the unknown (calculated) one. These

are then aligned. The spectral vectors carry weighted information and have the following elements $w_{\mathrm{exp/calc},i}$:

$$w_{\mathrm{exp/calc},i} = \left(\frac{m}{z}\right)^{m}_{\mathrm{exp/calc},i} \cdot I^{n}_{\mathrm{exp/calc},i}, \tag{C.1}$$

where I is the relative intensity. The argument for scaling the masses and the intensities by powers m, n is that higher m/z values often give diagnostic peaks, particularly for the molecular ion. The two weighted spectral vectors w_{exp} and w_{calc} are then normalized and the cos^2 of the angle between them is calculated. This gives a value between 0 and 1, where 0 means that the two spectra have absolutely nothing in common and 1 means that the experimental and the calculated spectra are the same. This value itself is called the dot-product score, F_{D}.
A second term is then introduced that sums over all peak pairs (that is peaks which occur in the experimental and in the calculated spectrum) and compares the elements of neighboring peaks. In this way, the topology of the spectra is accounted for. The second term looks like this:

$$F_{\mathrm{R}} = \frac{1}{N_{\mathrm{pairs}}} \sum_{i}^{\mathrm{pairs}} \left(\frac{w_{\mathrm{exp},i}}{w_{\mathrm{exp},i-1}} \cdot \frac{w_{\mathrm{calc},i-1}}{w_{\mathrm{calc},i}} \right). \tag{C.2}$$

The total composite match score that is used in our study is computed according to equation C.3:

$$\mathrm{score} = \frac{N_{\mathrm{calc}} F_{\mathrm{D}} + N_{\mathrm{pairs}} F_{\mathrm{R}}}{N_{\mathrm{calc}} + N_{\mathrm{pairs}}}, \tag{C.3}$$

where N_{calc} is the number of peaks in the calculated EI mass spectrum. Table C.1 gives the different mass and relative intensity scalings of the F_{D} and F_{R} terms.

Table C.1.: Scalings for the two different terms of the spectral matching score (taken from the reference cited, see above)

Term	mass power (m)	intensity power (n)
F_{D}	3	0.5
F_{R}	0	1

From Table C.1 one can see that the first term, F_{D}, manipulates the score in favor of heavy fragments and the molecular ion by weighting the masses to the third power, whereas the second term, F_{R}, introduces a similarity measure of the neighboring peaks topology (all masses count equally). Everything is still normalized, and the final number used in our article is the composite score multiplied by 1,000, which gives a number between 0 (orthogonal vectors, complete failure of spectral prediction) and 1,000 (perfect reproduction of the experimental mass spectrum, very unlikely).
It should be noted that the match scores only apply within a given system and/or for checking

against spectral libraries. This is mainly because the bigger the system is, the lower the match score will be just due to the fact that the chance for deviations (especially when the relative intensities of neighboring peaks are heavily involved) increases with the size of the system.

D. Supporting Information to Chapter 6

Appendix C contains:

- Reaction coordinates for decomposition pathways
- Additonal calculated spectra
- Computational statistics

Exemplary reaction coordinates for decomposition pathways

For the prediction of EI-MS by QCEIMS, the quality of the resulting spectra is reflected in the accuracy of the QC method used to compute the atomic forces. In other words, the PES of the QC method has to closely parallel the 'true' PES. Therefore, we compare potential energy curves obtained with GFN-xTB to its level of reference, hybrid DFT. Three simple exemplary reaction pathways involving single bond ruptures are examined: the loss of an ethyl residue from the hexane cation (see Fig. D.1), $AsCl_2^+$ from the lewisite cation (see Fig. D.2) and iodine from iodobenzene cation (see Fig. D.3). Snapshots along the reaction pathway have been superimposed on the figures. The three chosen pathways correspond to the formation of ions which were observed to have relatively intense peak-signals. Therefore, we consider the potential energy curves to be representative of the MD trajectories.

To compute the reaction pathways, we use a simple and intuitive approach, referred to here as a relaxed potential energy surface scan. Given an optimized reactant configuration (the equilibrium ion structure) and products (a neutral and a charged fragment). We perform a linear interpolation of the system with 30 system images placed between the reactant and product states. In the interpolation, only one degree of freedom is varied in an equidistant stepwise fashion, which corresponds to the dissociation process. Each image is then optimized with the dissociating bond distance constrained and all other degrees of freedom are allowed to relax. The optimization is performed using PBE0-D3(bj)/def2-TZVP with an electronic temperature of 10000 K. The energy of each optimized image (including the reactant and product configurations) is calculated with GFN-xTB (at 5000 K) and refined by PBE0-D3(bj)/def2-QZVP (at 10000 K). This methodology for computation of reaction pathways is known to fail for more complex reactions (*i.e.* reaction coordinates) than the ones presented here. This can

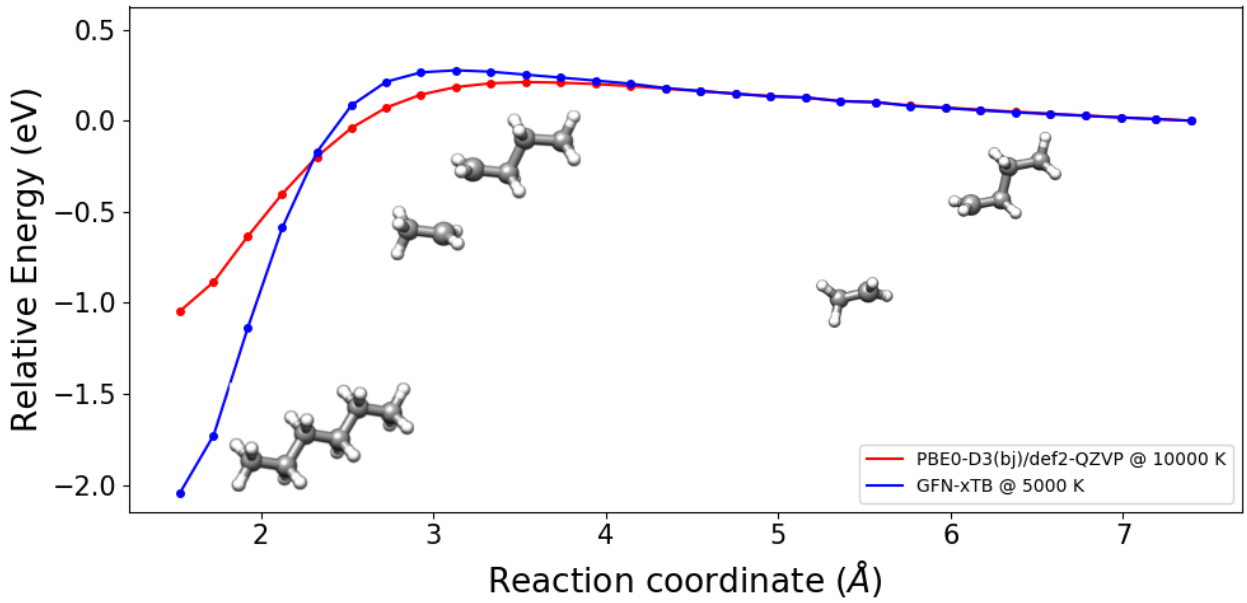

Figure D.1.: Potential energy curve for the loss of an ethyl residue from the hexane cation. The blue points were calculated using GFN-xTB (5000 K) and the red points by PBE0-D3(bj)/def2-QZVP (10000 K). For clarity, three snapshot along the reaction coordinate have been superimposed on the figure.

be seen by an introduction of discontinuity in the potential energy curve, where relaxation of the remaining degrees of freedom pulls the system away from the minimum energy path.

The agreement between the shape of the potential energy curves calculated with PBE0-D3(bj)/def2-QZVP and GFN-xTB is excellent, for all three reactions. However, we find the GFN-xTB to predict too strong binding, where the difference can range from roughly 0.25 (lewissite) to 1 eV (hexane).

The three cases shown here are only an initial assessment of GFN-xTB. It is nowhere near complete and a more extensive study is needed *e.g.,* by inclusion of a large number of representative 'real-world' systems and reactions with more complicated reaction coordinates and comparison to high-level *ab-initio* QC calculations and hybrid DFT.

Additional Spectra

All additional EI mass spectra are calculated with the same simulation parameters as in the main manuscript, except for the specific modifications, which are investigated in the first subsection.

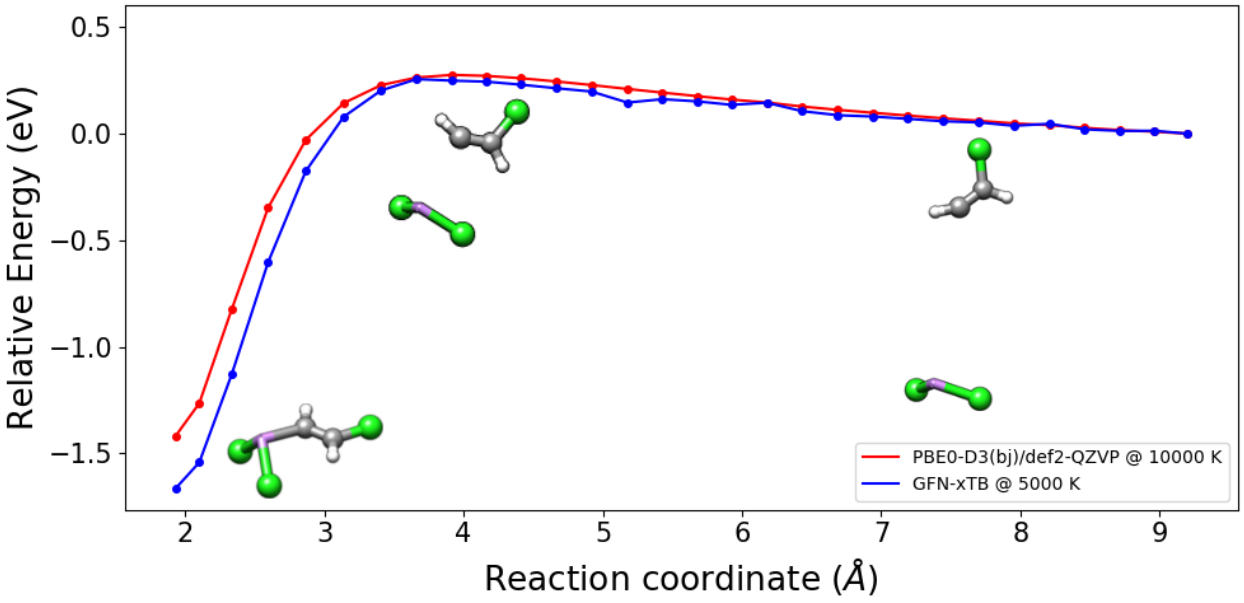

Figure D.2.: Potential energy curve for the loss of $AsCl_2^+$ from the lewissite cation. The blue points were calculated using GFN-xTB (5000 K) and the red points by PBE0-D3(bj)/def2-QZVP (10000 K). For clarity, three snapshot along the reaction coordinate have been superimposed on the figure.

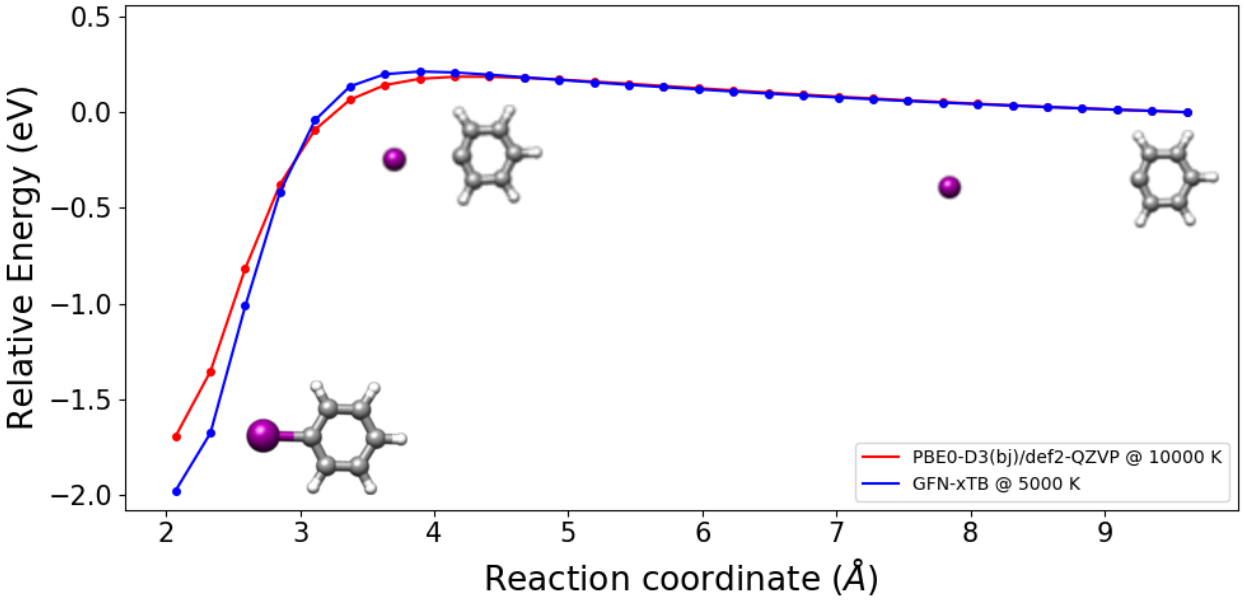

Figure D.3.: Potential energy curve for the loss of an iodine from the iodobenzen cation. The blue points were calculated using GFN-xTB (5000 K) and the red points by PBE0-D3(bj)/def2-QZVP (10000 K). For clarity, three snapshot along the reaction coordinate have been superimposed on the figure.

Effect of Simulation Time and IEE Distribution

We have investigated the effect of two important simulation parameters: (i) the maximum simulation time parameter, and (ii) the IEE distribution. The former determines one stop criterion in the QCEIMS production runs. It has been set to 10 ps for the results reported in the main manuscript. The latter determines the amount of internal energy deposed in each production run. It is set by default to have its maximum at 0.6 eV per atom. For details, see *Angew. Chem. Int. Ed.*, **2013**, *52*, 6306.

We have scanned these two simulation parameters in the following way : (i) the maximum simulation time is set to 5 ps, 10 ps, and 20 ps, respectively. (ii) The IEE distribution has been set to 0.6 eV per atom (the default value), and 0.3 eV per atom. The results are presented in Figures D.4, and D.5, respectively. This procedure was performed for the molecules 1-fluorohexane (**2**) and tetramethylsilane (**13**).

The results reveal that the simulation results are perhaps unexpectedly quite robust with respect to the choice of the two parameters. There are, of course, minor differences in the calculated EI-MS of the two compounds, but these are not visible in Figures D.4 and D.5, but are recorded in the respective output files. Since the purpose of QCEIMS is not to obtain a quantitatively accurate prediction of an EI-MS but rather to obtain a computed spectrum by which a compound may be identified and its unimolecular fragmentation pathways upon electron ionization explored, the finding that the variation of simulation parameters may not change the results significantly adds to our conclusion that QCEIMS is a stable and reliable program. The systematic exploration of much longer simulation times of 100 ps to a full nanosecond will be the subject of further research, which is beyond the scope of the present study.

Additional Calculated Spectra of Organic Molecules

We show additional calculated spectra of organic molecules below.

Spectra of Organometallic Molecules with ΔSCC (GFN-xTB) IP Evaluation

As seen in Figure D.12, the calculated spectrum shows a lot of artifacts that are due to the erroneous evaluation of the fragment ionization potentials, for which no specialized IPEA-xTB parameters exist. In contrast to the spectrum shown in the paper, the naked Fe^+ is not predicted correctly. For this reason, we recommend that the computation of ionization potential remain at the Δ SCF (PBE0/SV(P)) level of theory until the parametrization of IPEA-xTB will have been completed.

The comparisons of computed in experimental EI-MS (GFN-xTB) in Figure D.13 reveal

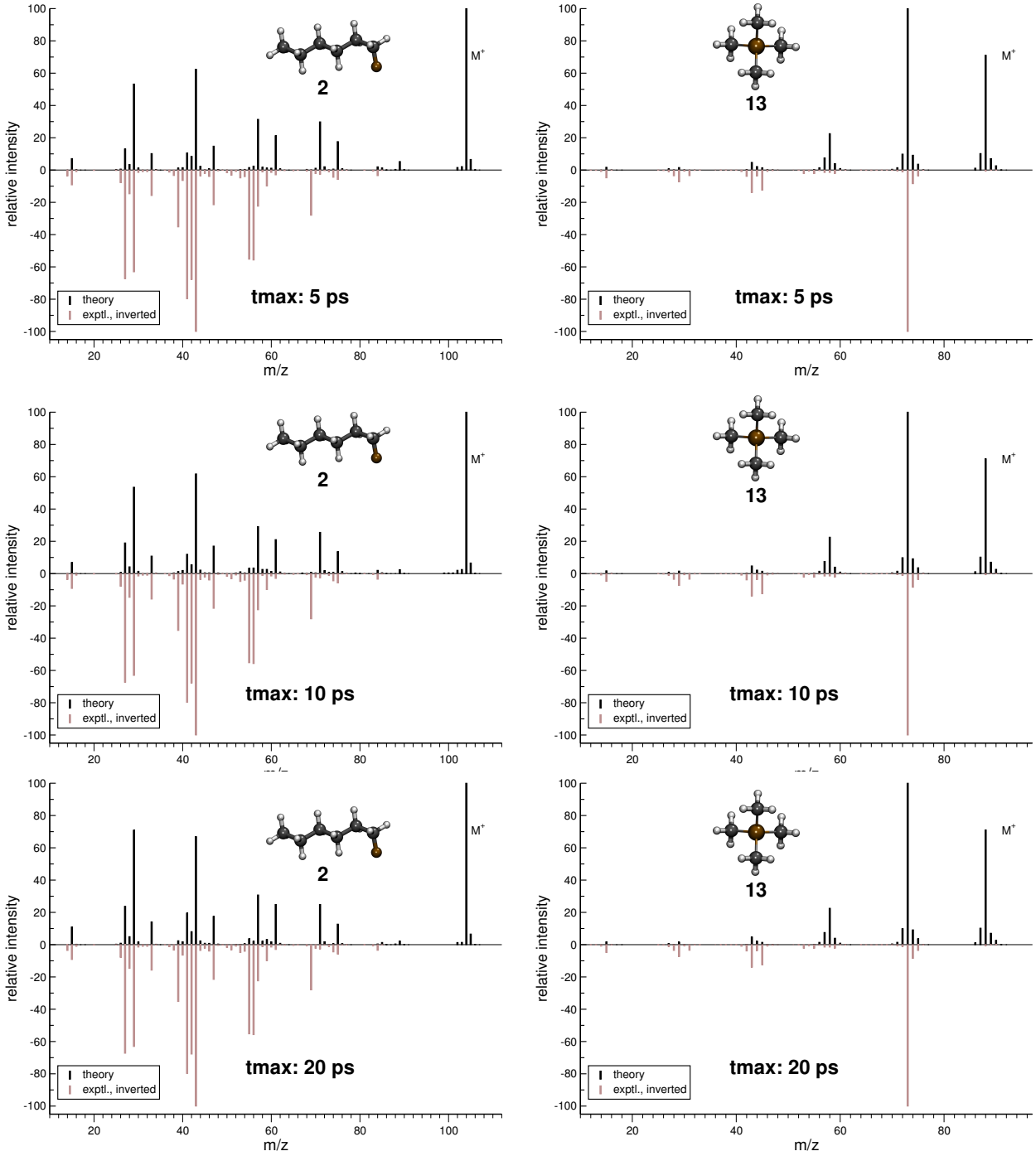

Figure D.4.: Comparison of computed and experimental EI-MS (GFN-xTB) for 1-F-hexane and tetramethylsilane depending on the maximum simulation time. Maximum of the IEE distribution at 0.6 eV per atom.

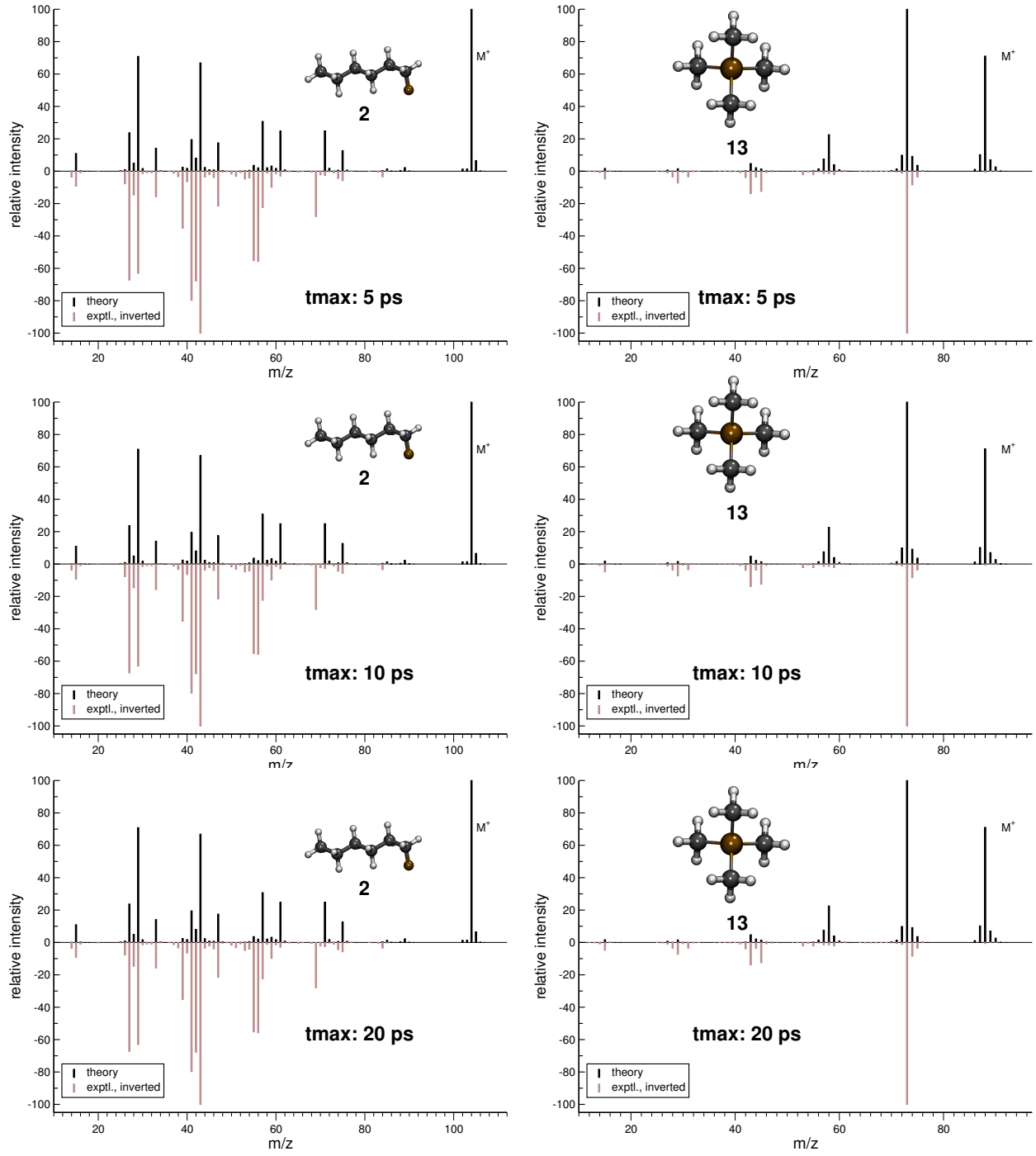

Figure D.5.: Comparison of computed and experimental EI-MS (GFN-xTB) for 1-F-hexane and tetramethylsilane depending on the maximum simulation time. Maximum of the IEE distribution at 0.3 eV per atom.

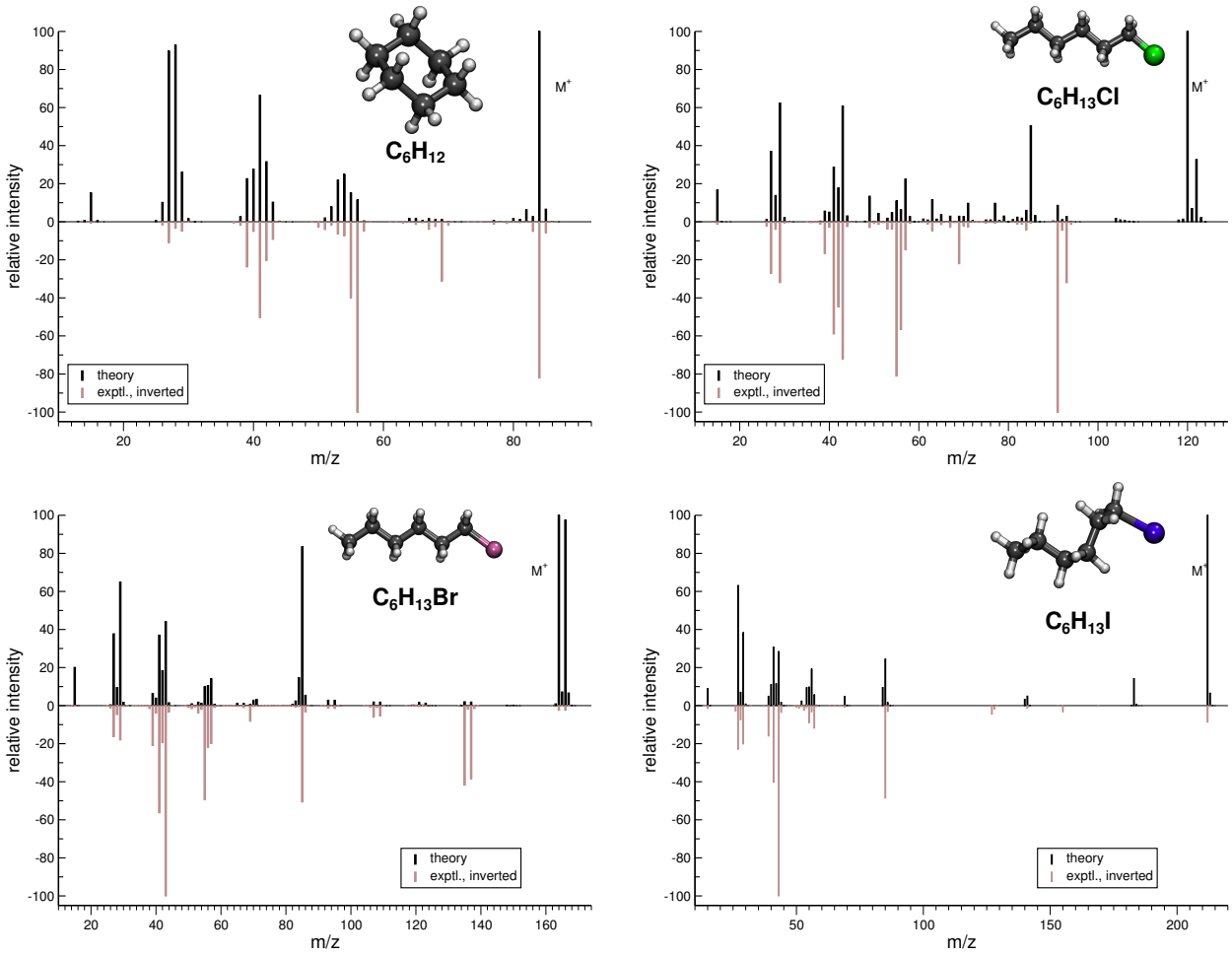

Figure D.6.: Comparison of computed and experimental EI-MS (GFN-xTB) for simple organic aliphatic and aliphatic halogenide molecules.

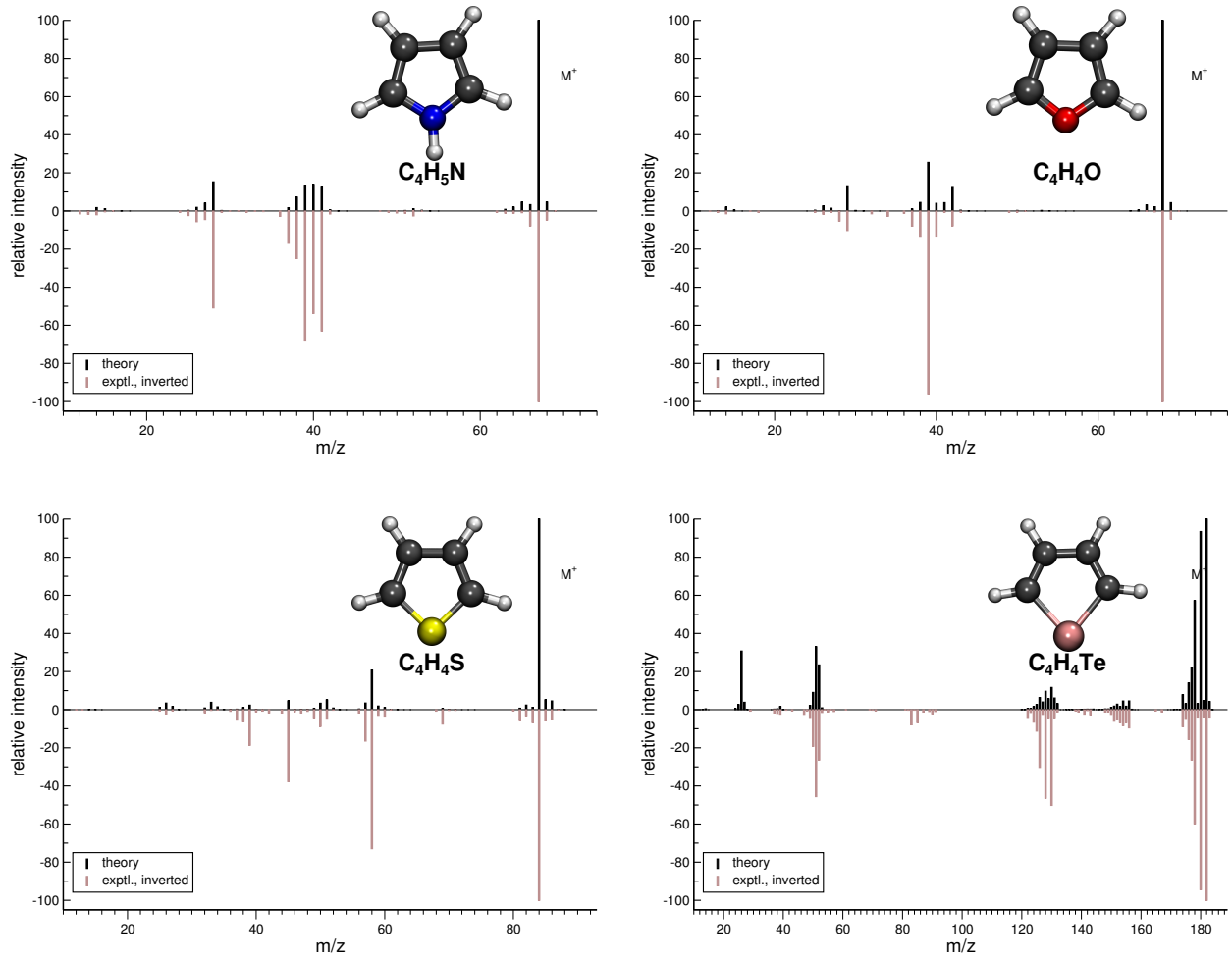

Figure D.7.: Comparison of computed and experimental EI-MS (GFN-xTB) for aromatic heterocyclic molecules.

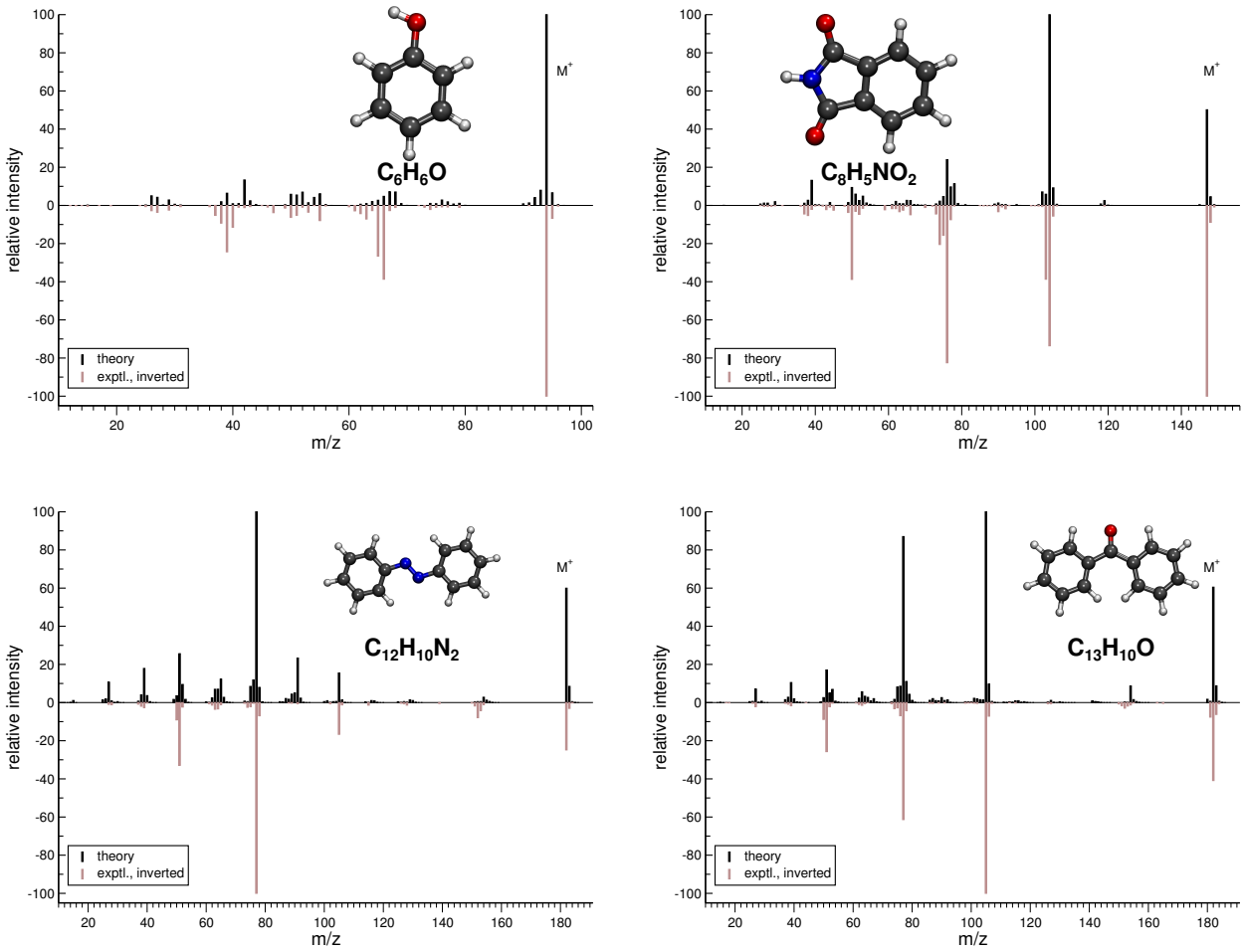

Figure D.8.: Comparison of computed and experimental EI-MS (GFN-xTB) for aromatic molecules.

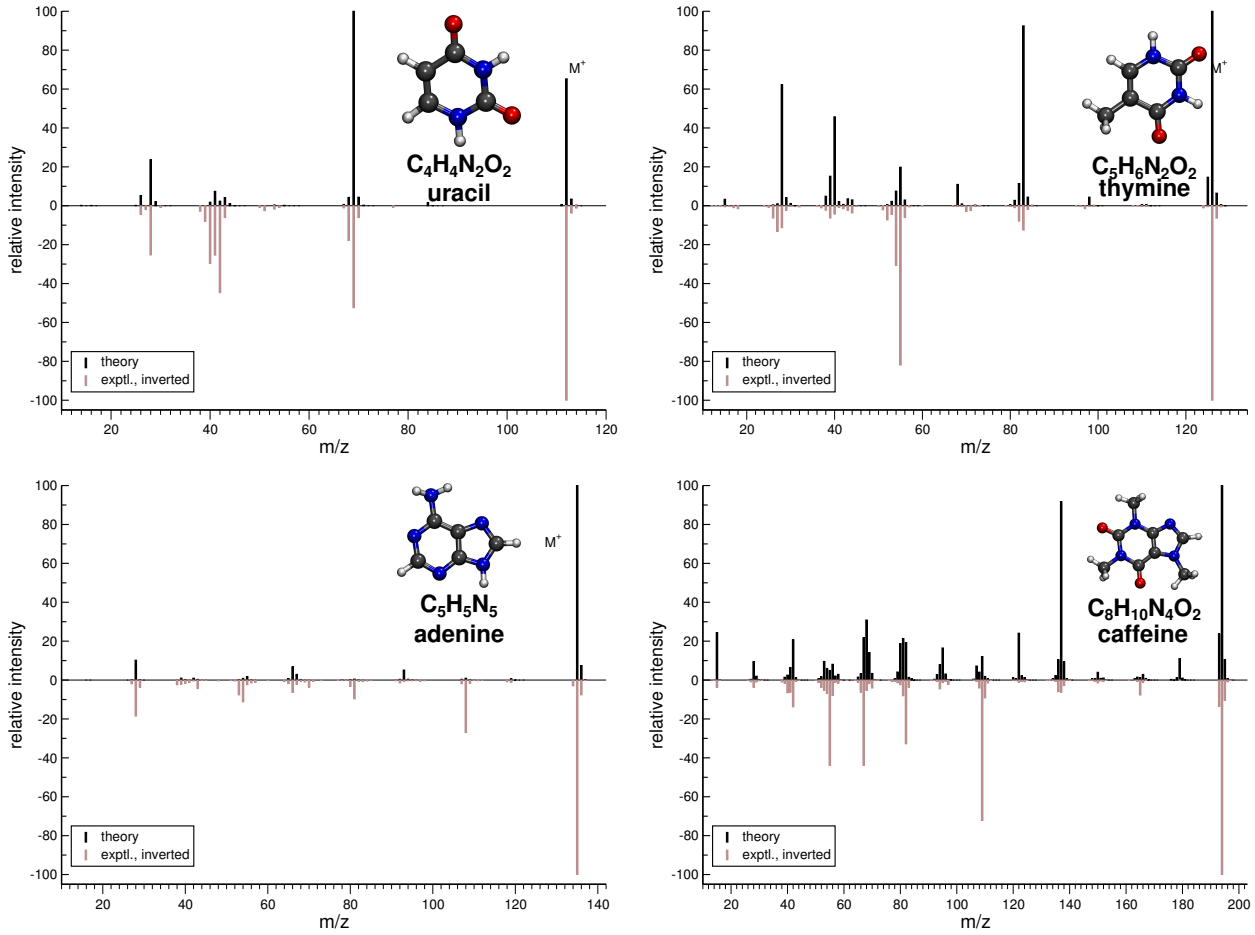

Figure D.9.: Comparison of computed and experimental EI-MS (GFN-xTB) for pyrimidine and purine derivative molecules.

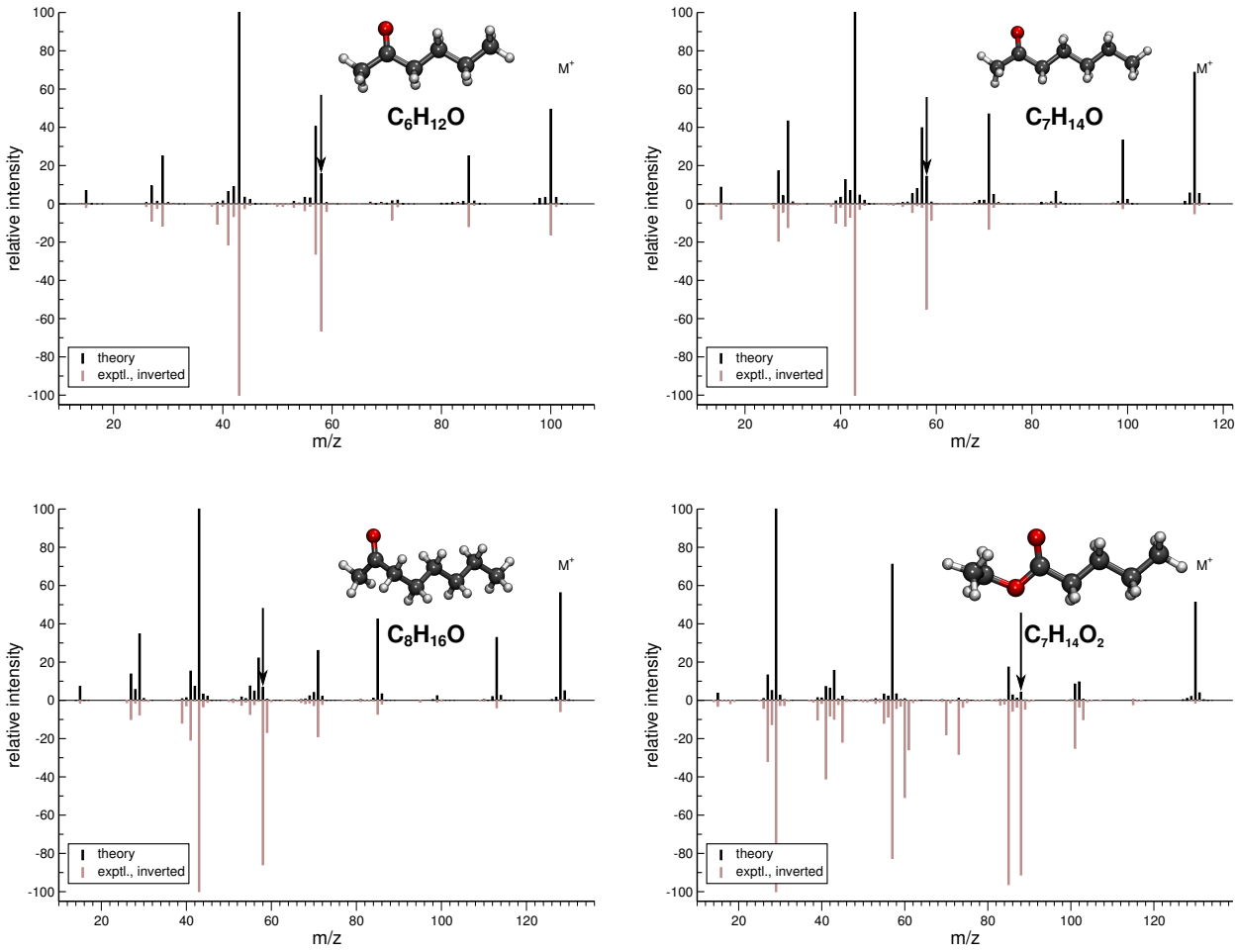

Figure D.10.: Comparison of computed and experimental EI-MS (GFN-xTB) for molecules that upon ionization undergo the McLafferty rearrangement. The corresponding peaks in the computed EI-MS are marked by the arrows.

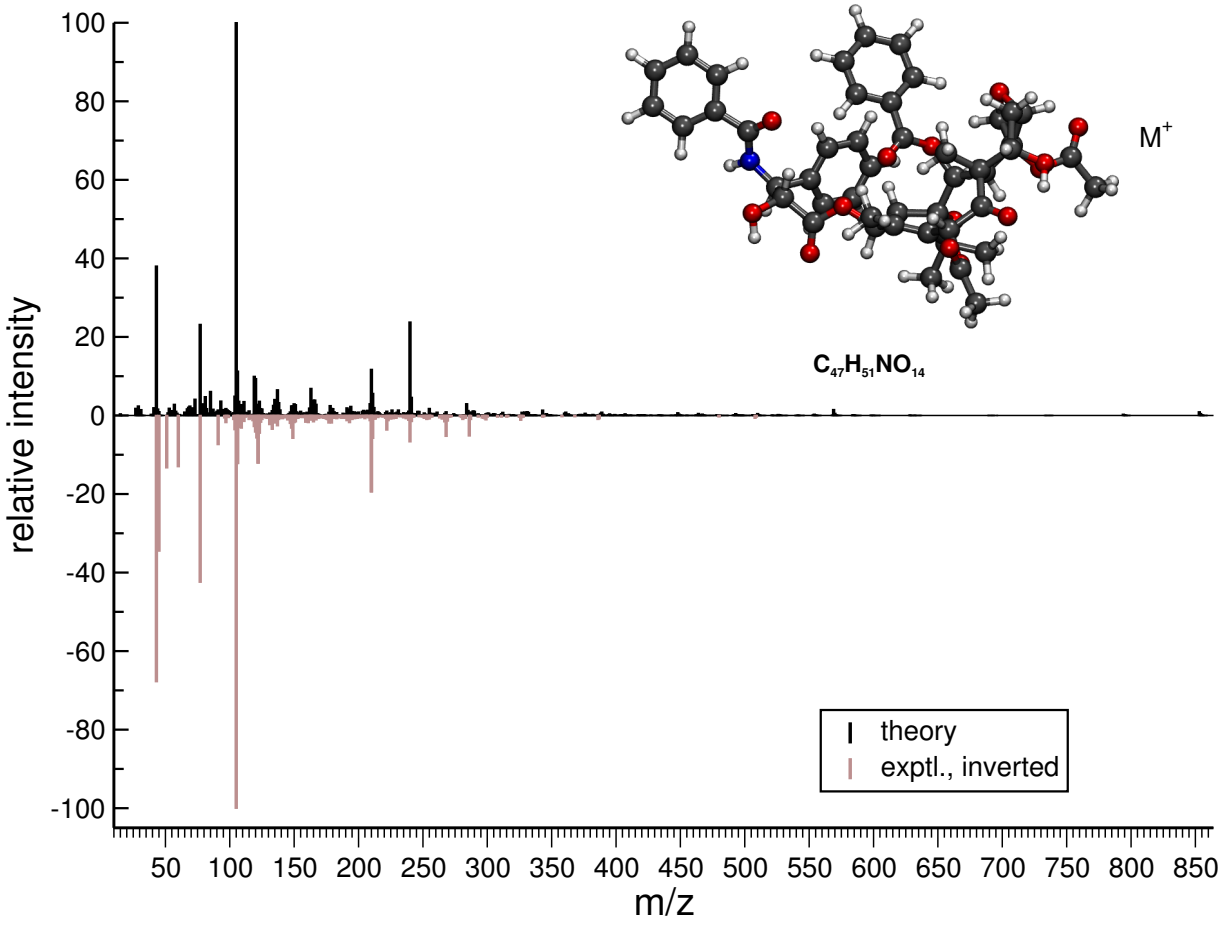

Figure D.11.: Comparison of computed and experimental EI-MS (GFN-xTB) for taxol.

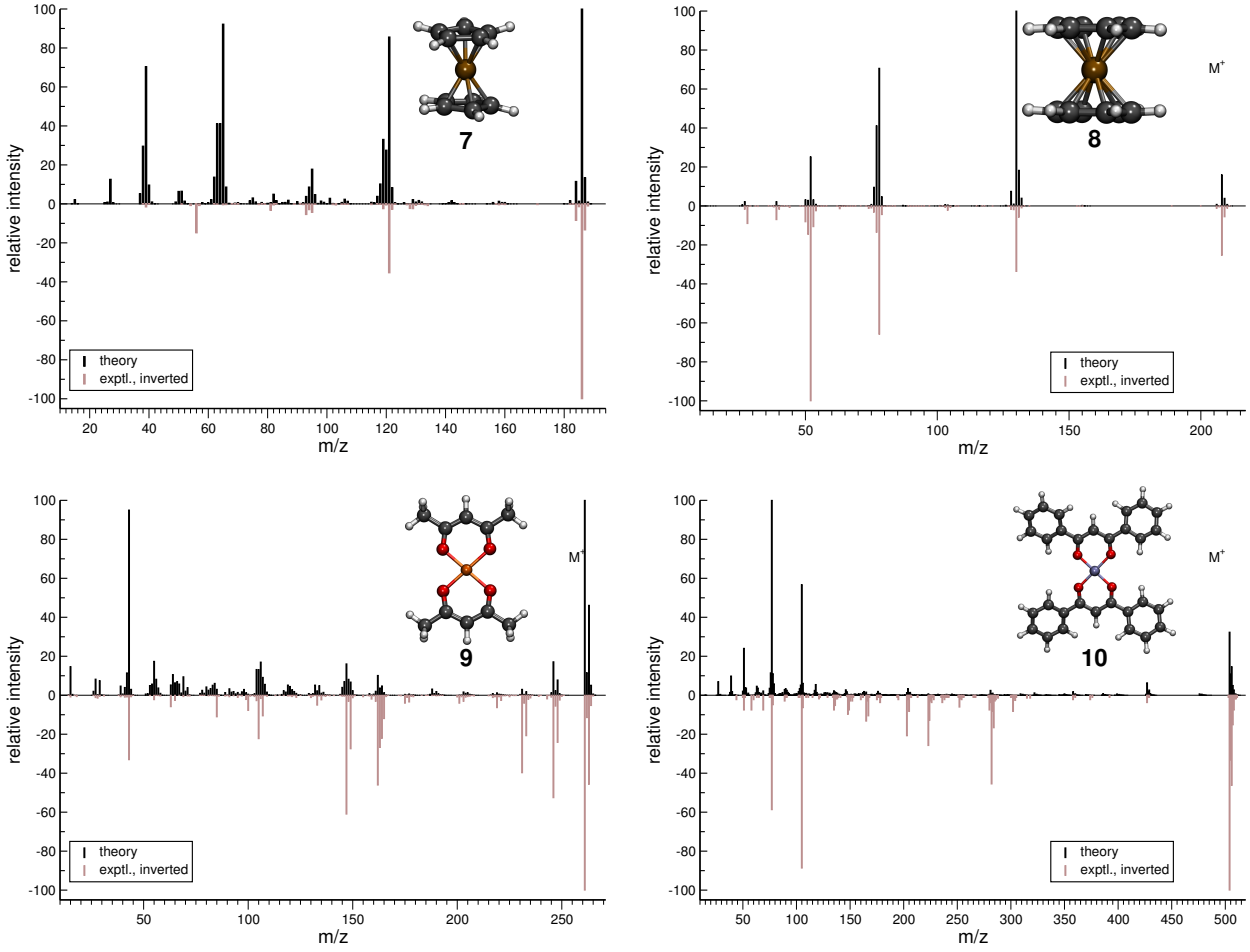

Figure D.12.: Comparison of computed and experimental EI-MS (GFN-xTB) for the organometallic group using fragment IPs calculated at the Δ SCC (GFN-xTB) level of theory.

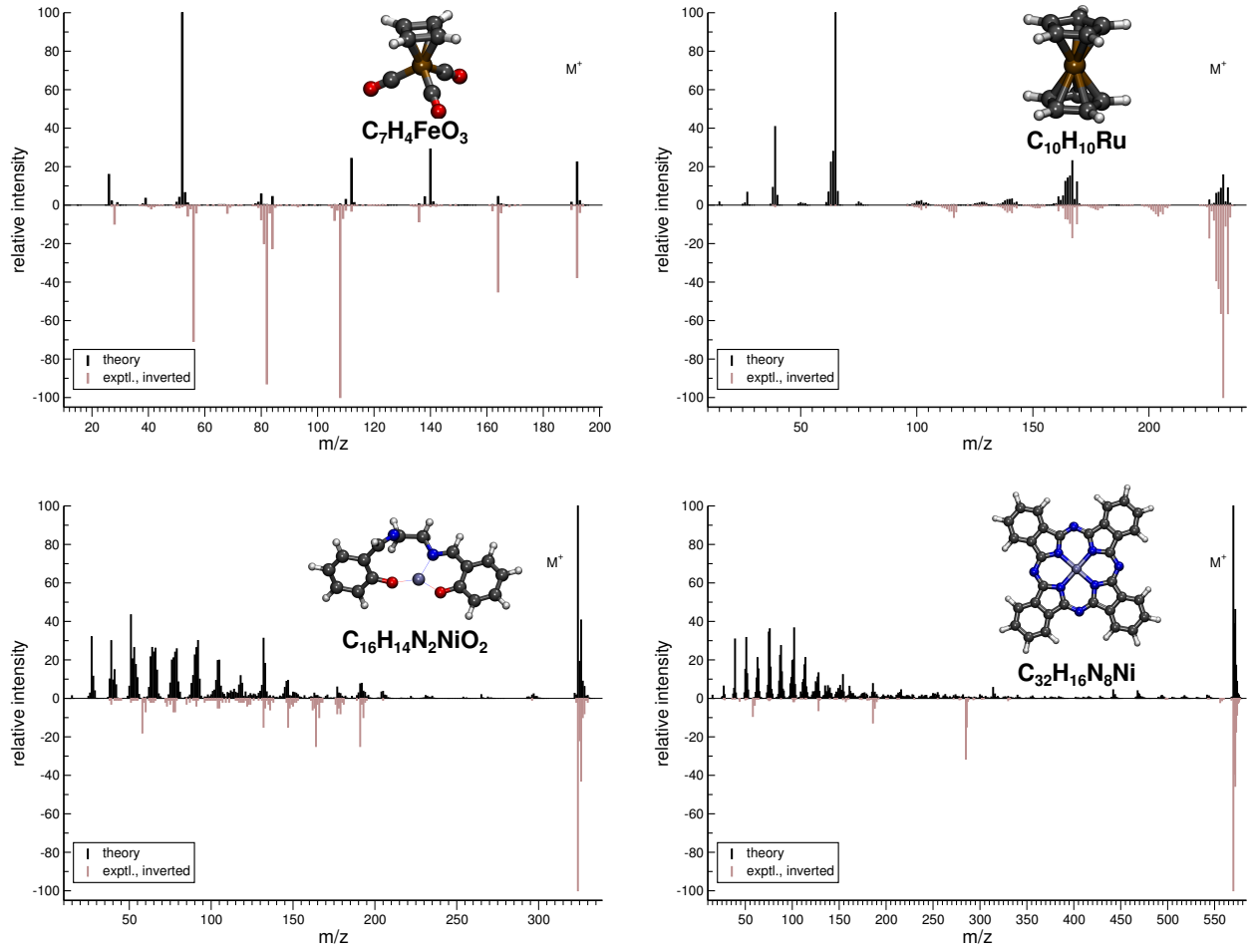

Figure D.13.: Comparison of computed and experimental EI-MS (GFN-xTB) for additional organometallic molecules using fragment IPs calculated at the Δ SCC (GFN-xTB) level of theory.

that (i) organometallics remain a challenging class of compounds for EI-MS prediction. Yet the problems that are encountered for the systems shown in Figure D.13 will be analysed and may thus provide the starting point for the continuous improvement of the QCEIMS method. This is also valid for the spectra presented in the next subsection.

Additional Calculated Spectra – Deficiencies of MS(GFN-xTB)

Here, we present a number of calculated EI-MS, which we do not consider of sufficient quality, and offer preliminary statements on how these failures may be explained. It should be said before all discussion below that GFN-xTB is a semiempirical, cost-efficient QC method, which cannot be expected to always yield a perfect description of the energetics of the unimolecular fragmentation reaction space.

The apparent failure of EI-MS prediction for several classes of biomolecules by MS(GFN-

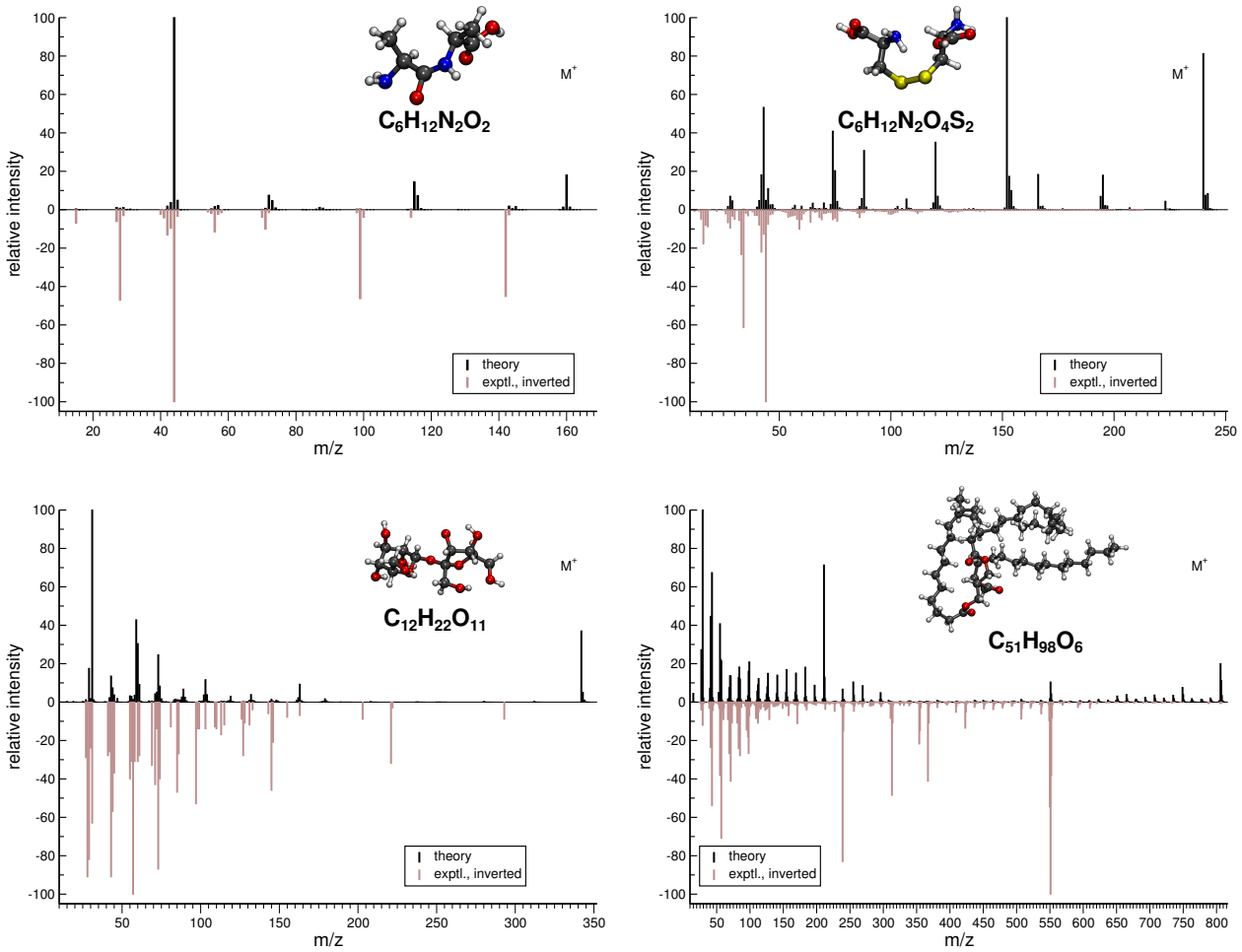

Figure D.14.: Comparison of computed and experimental EI-MS (GFN-xTB) for different classes of biomolecules.

xTB) (the dipeptides dialanine and cystine, sucrose and tripalmitin, shown in Figure D.14) could be considered a distressing finding. However, closer inspection of the simulation results reveals that some of the failures can be explained reasonably. The computed spectrum of dialanine, for instance, consists mostly of the base peak, which is an ion produced in a standard α cleavage channel. It is not unreasonable that the GFN-xTB PES should overrepresent this pathway by perhaps featuring a too low barrier for this reaction. Similar observations are made for sucrose and the triglyceride tripalmitin. For cystine, there are admittedly many artefacts, which, however, disappear when computing the IPs at the Δ SCF (PBE0/SV(P)) level, see Figure D.15. That spectrum has been calculated using 200 production runs. The final IP/EA xTB parametrization has not been performed for sulfur yet, and in sensitive cases, we recommend crosschecking the IP evaluation by switching on the Δ SCF (PBE0/SV(P)) level for that part of the simulation.

The computed EI-MS of saframycin A shown in Figure D.16 contains a lot of artefacts.

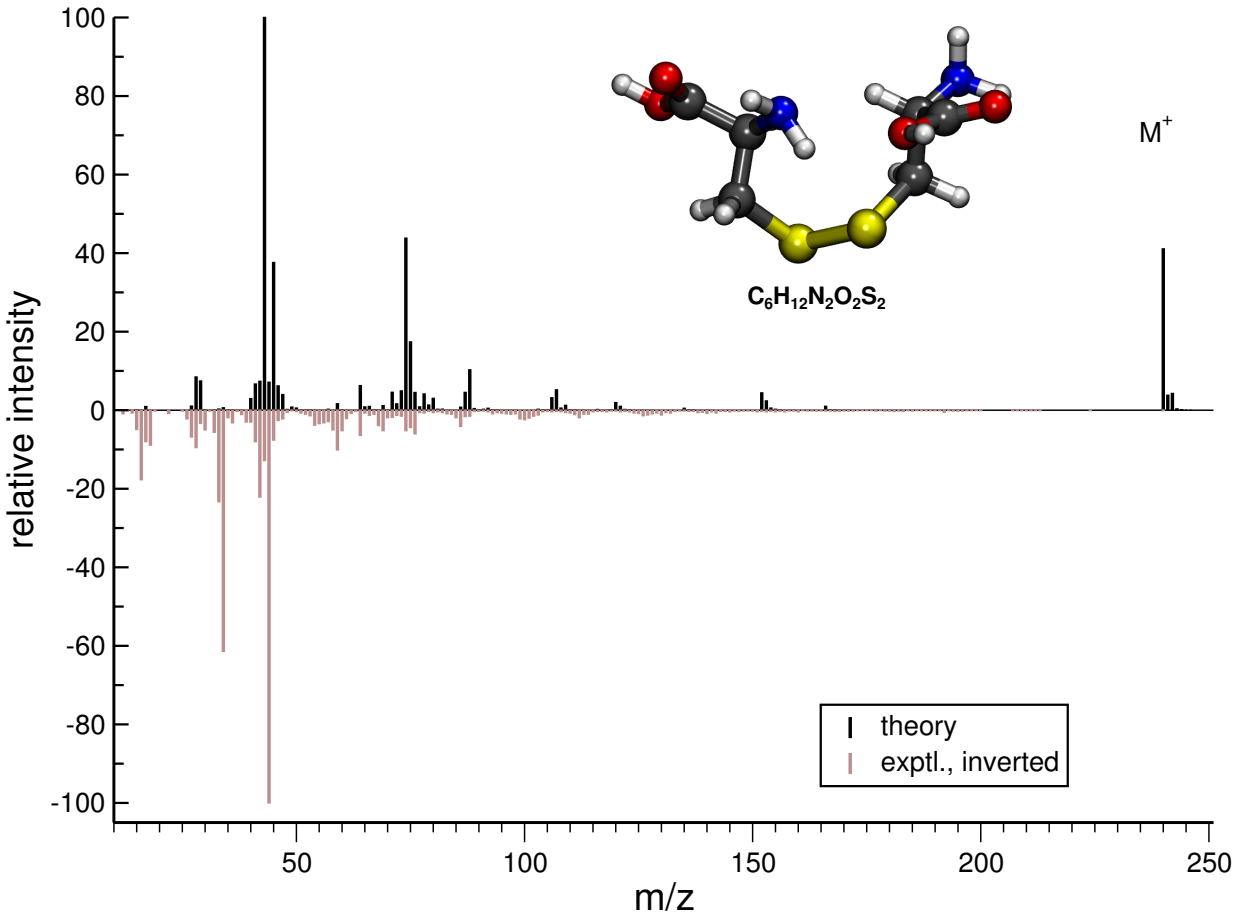

Figure D.15.: Comparison of computed and experimental EI-MS (GFN-xTB) for cystine with Δ SCF (PBE0/SV(P)) IP evaluation for the fragments.

Despite the ability of MS(GFN-xTB) to capture some of the main peaks, there are obviously some fragmentation pathways that are artificially overrepresented. Moreover, the internal energy distribution leads in this case to both heavy fragmentation in the production runs as well as survival of the molecular, which is not seen in the experiment. Therefore, the energy distribution may be unbalanced. It will be the topic of further research to investigate why our internal energy distribution model succeeds in many cases but fails in others. Similar observations are made for tecloftalam (Figure D.17).

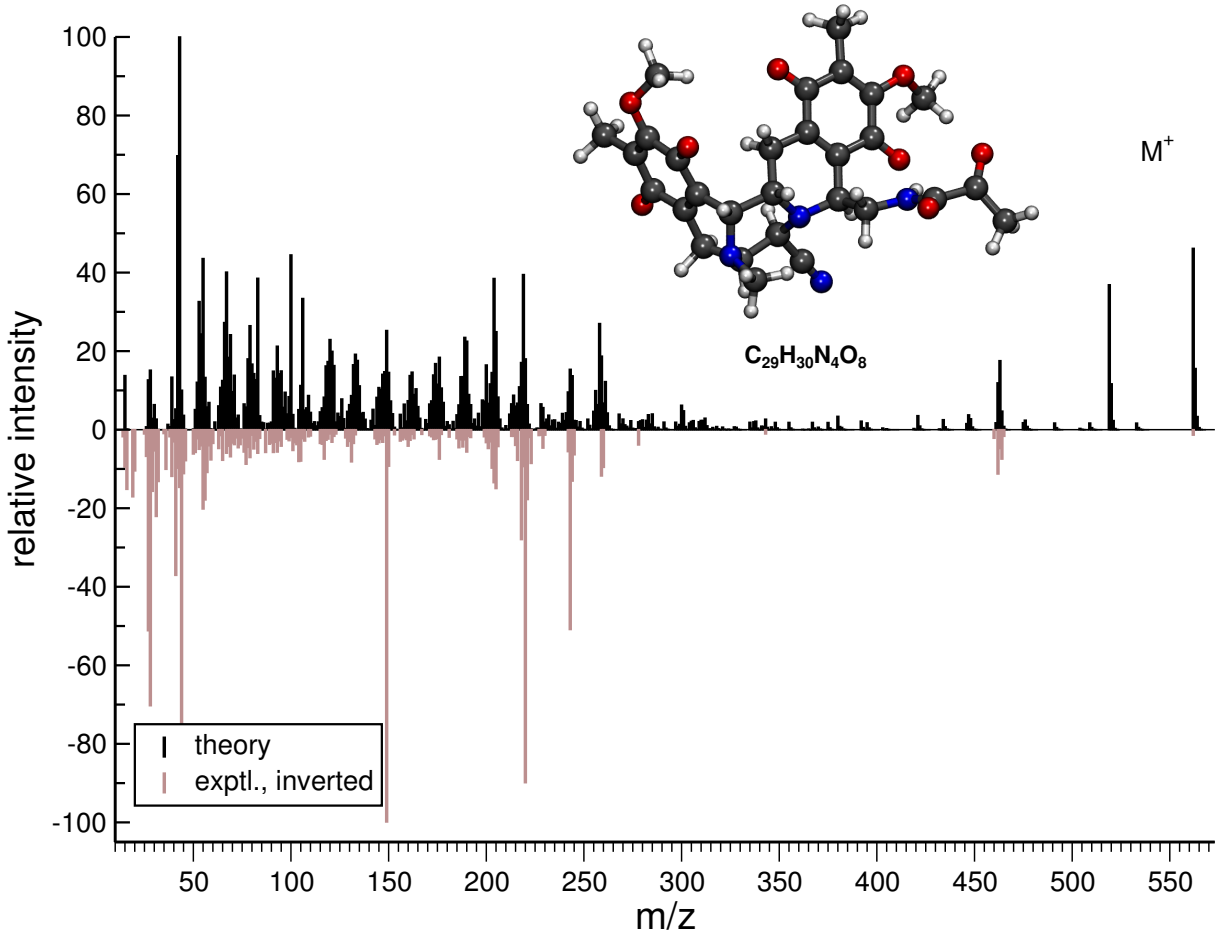

Figure D.16.: Comparison of computed and experimental EI-MS (GFN-xTB) for saframycin A.

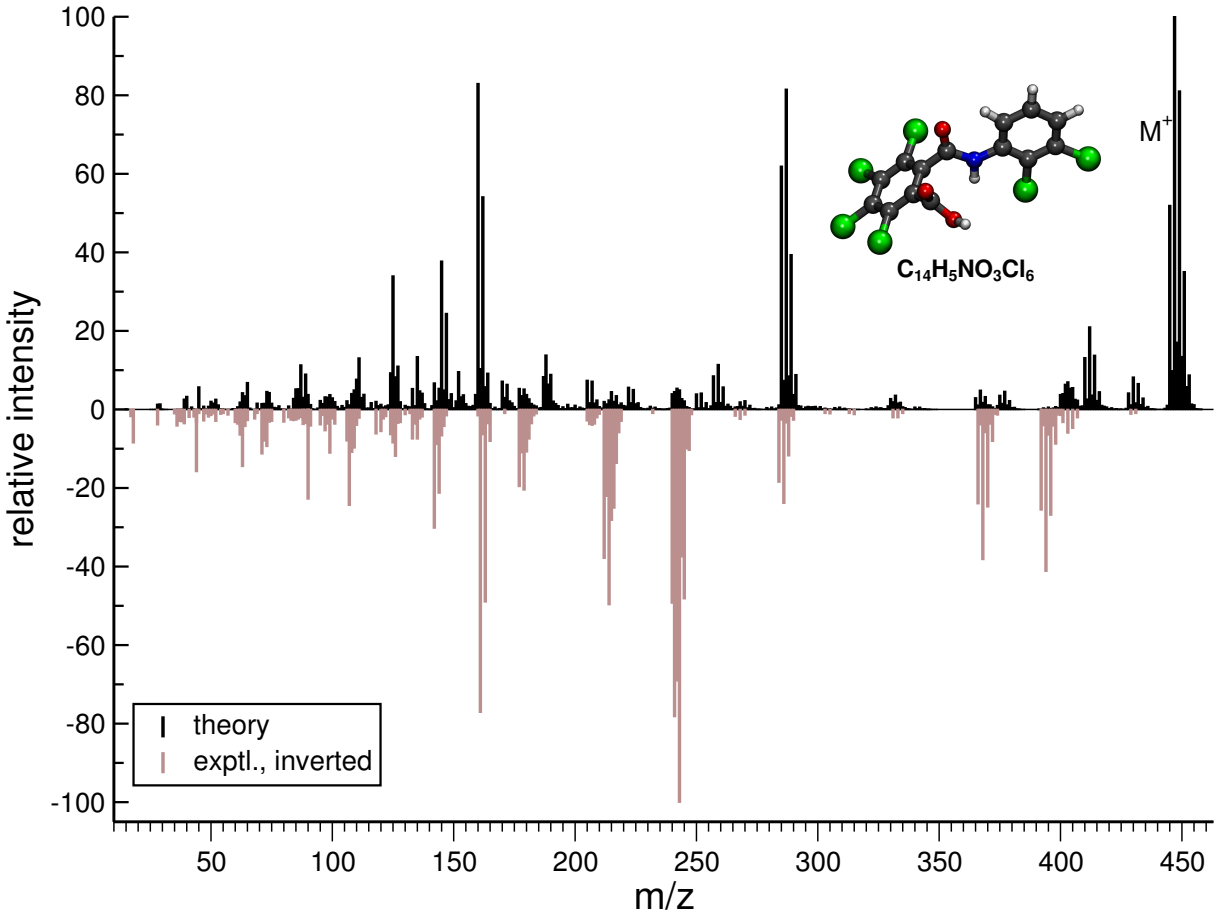

Figure D.17.: Comparison of computed and experimental EI-MS (GFN-xTB) for tecloftalam.

Additional Calculated Spectra – Comparison of Semi-empirical PES

In this subsection, we show three illustrative examples of the effect of the semi-empirical quantum chemical PES, which we hold is the largest error source for the computed spectra. The score that is given below is a modified dot-product score. It quantifies the overlap between the computed and experimental spectra. A score of 0 means no overlap, a score of 1,000 means identical spectra.

For the case of methyl sulfonamide, we compare between the DFTB3-D3 and GFN-xTB computed spectra, see Figure D.18. For DFTB3-D3, no molecular ion survives the simulation, and the base peak is not identified correctly, indicating that the dissociation energies of the S–N and S–C bonds are not in the right order. The intensity of the peaks in the GFN-xTB computed spectrum is of much higher quality, even if the base peak is misassigned (m/z 15 is the methyl cation, possibly a problem of the IP calculations, as addressed for cystine above). The higher PES quality of GFN-xTB for methyl sulfonamide leads to a much higher score for

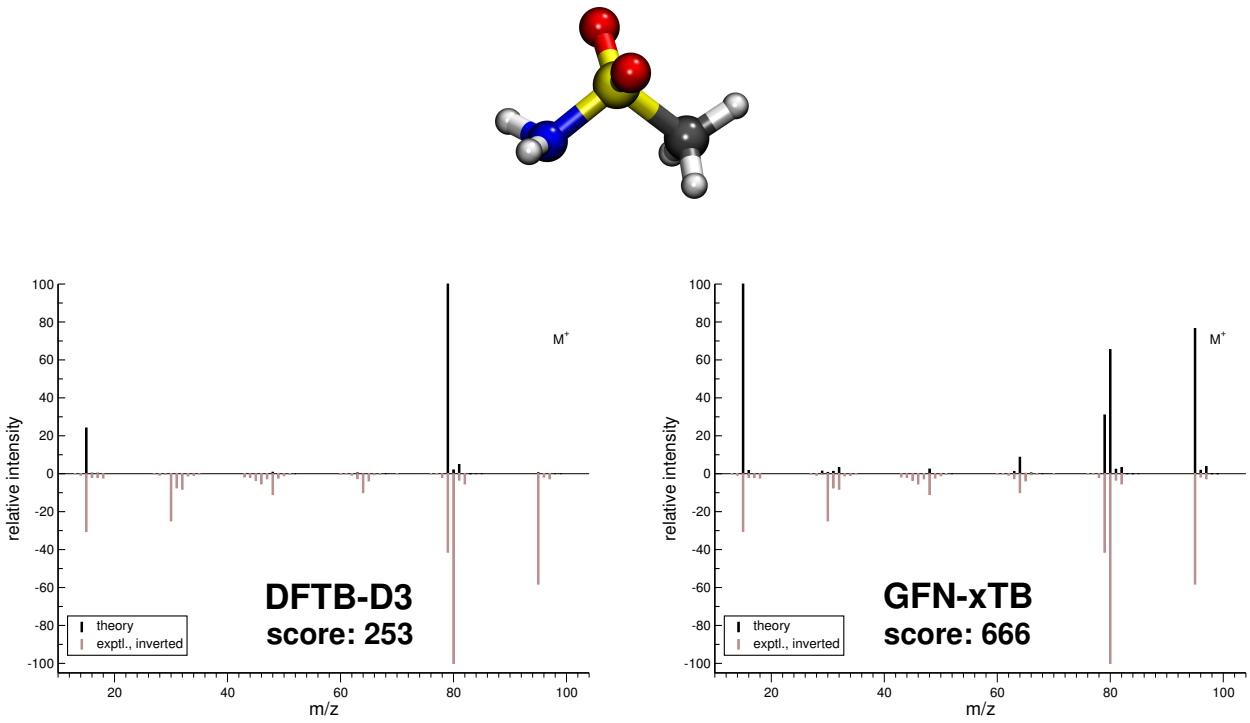

Figure D.18.: Comparison of computed and experimental EI-MS (DFTB3-D3/GFN-xTB) for methyl sulfonamide.

the comparison between computation and experiment.

Figure D.19 shows the comparison of DFTB3-D3 and GFN-xTB calculated mass spectra for 2-hexanone. This molecule undergoes a McLafferty rearrangement to yield the ion m/z 58, which is found in 4.6 % of all production runs at the GFN-xTB level of theory, whereas it is not found at all in the DFTB3-D3 production runs. The signal at m/z 58 in the DFTB3-D3 computed spectrum is only due to the isotope peak, which has been added post-simulation. This is another case where the PES quality is the main source of discrepancies between the simulation and the experiment. The McLafferty rearrangement pathway is accessible on the GFN-xTB PES. It appears to be inaccessible on the DFTB3-D3 PES, at least using our standard simulation conditions. Moreover, the base peak, m/z 43, is correctly predicted at the GFN-xTB level of theory whereas the base peak in the DFTB3-D3 computed spectrum is the ion m/z 57. This indicates that the GFN-xTB PES is of a higher quality for 2-hexanone compared to DFTB3-D3, which is also reflected in the spectral matching score difference.

Figure D.20 reveals how dramatically the quality of the computed spectrum may depend on pair-specific parameters of the GFN-xTB Hamiltonian. In the left spectrum, which is computed using the standard GFN-xTB parametrization (and has been part of the first submission of this manuscript), there are a lot of artifacts, especially the ion m/z57 (Fe-H$^+$). This artifact is completely removed in the new spectrum (which is now part of the main manuscript), simply by scaling down the Fe-H pair-specific parameter, which can be conveniently done via

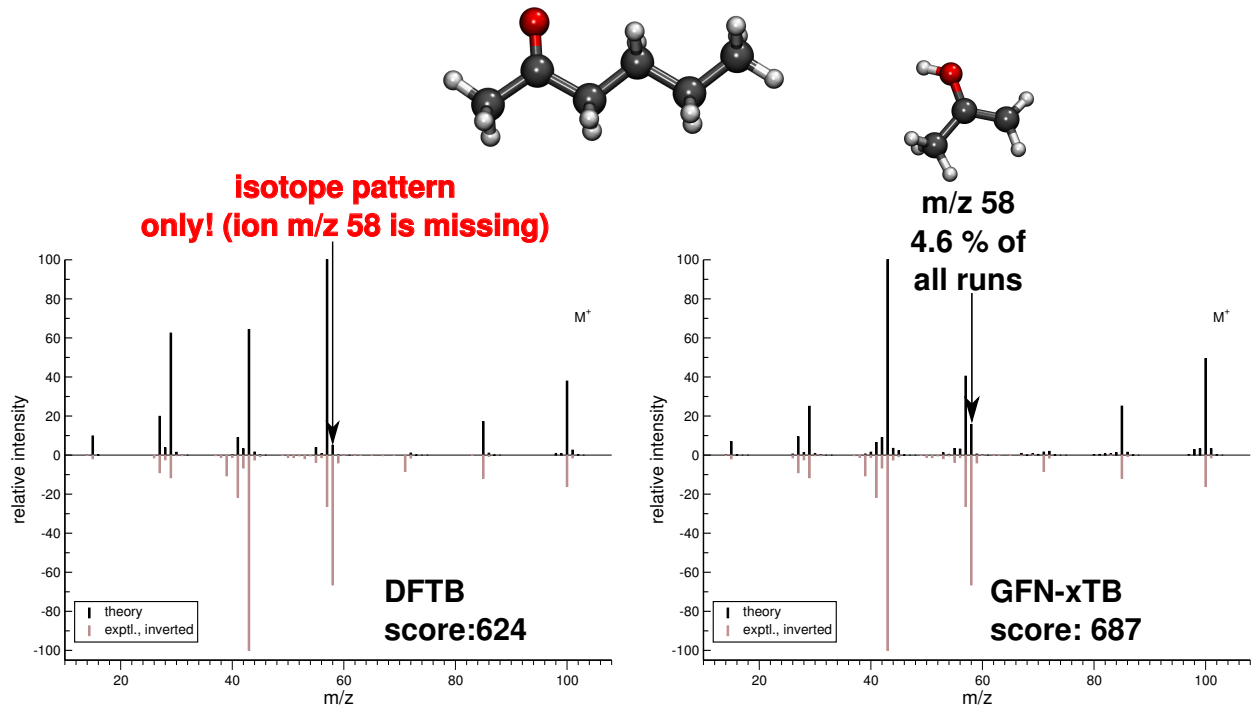

Figure D.19.: Comparison of computed and experimental EI-MS (DFTB3-D3/GFN-xTB) for 2-hexanone.

the parameter file of GFN-xTB read in by the program. The overall quality of the spectrum has thus greatly improved. Future research will be carried out in other cases where the standard parametrization of GFN-xTB, which has provided excellent results, see the spectra in the main part of the manuscript, apparently fails.

Lastly, we show a comparison between PM6-D2H and GFN-xTB calculated spectra of ferrocene in Figure D.21. The discussion here focuses on the ion m/z 105, FeC_4H^+, which is found in traces in the experimental spectrum. As displayed in Figure D.21, this ion has a chemically unreasonable structure, which is due to the short-range deficiencies of the PM6-D2H Hamiltonian (in essence, there is no Pauli repulsion), manifesting itself in the artificially short Fe–C bond length of 1.05 Å. The FeC_4H^+ ion also appears in the GFN-xTB calculated spectrum, although only as the results of one production run. Its structure is much more reasonable with a Fe–C distance of 1.97 Å. We therefore argue that GFN-xTB may produce artifacts, but they are to the best of our knowledge 'reasonable' artifacts, *e.g.*, due to a wrong ordering of reaction channels on the PES. We have not observed any completely unphysical structures of our simulated fragment ions.

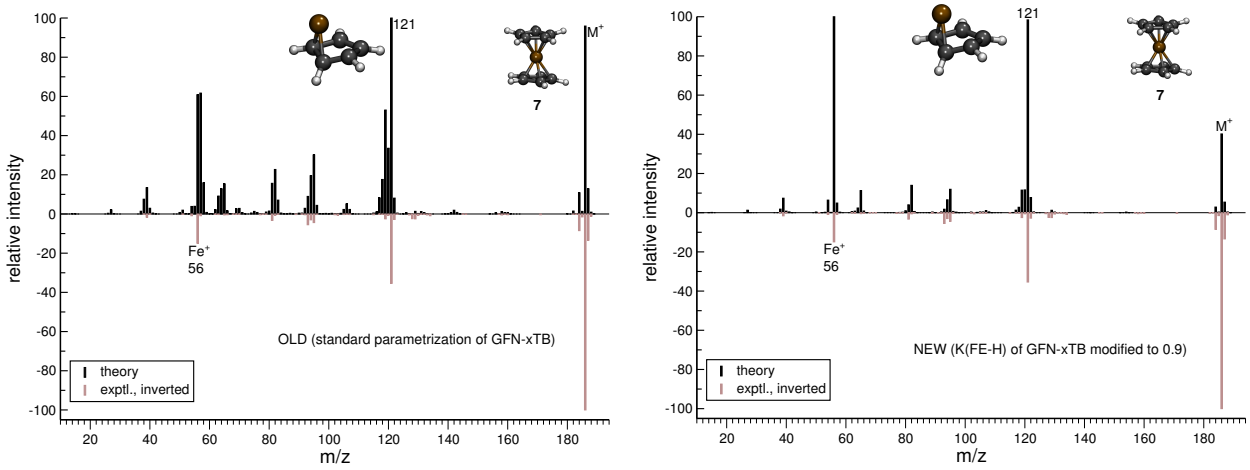

Figure D.20.: Comparison standard and slightly modified GFN-xTB PES computed and experimental EI-MS (GFN-xTB/IP: PBE0/SV(P)) for ferrocene.

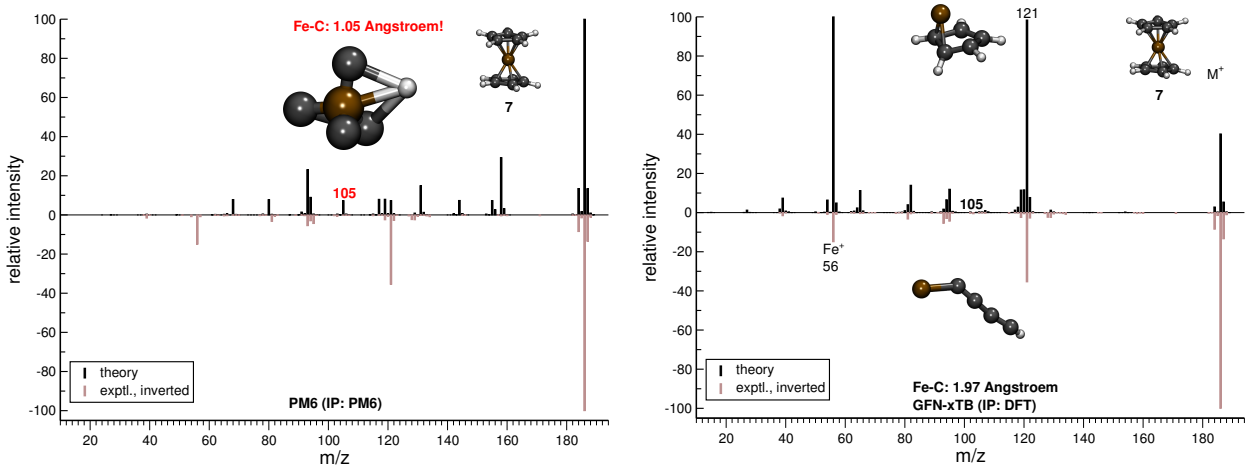

Figure D.21.: Comparison of computed and experimental EI-MS (PM6-D2H/GFN-xTB) for ferrocene.

Computational statistics

The average computational time required for a single point energy/gradient computation (QC call) and the number of unsuccessful production runs is depicted in Figure 3 (in the paper) for each molecule, with the exclusion of nickel(II)bis(diphenyl-acetylacetonate). This average computational time is calculated by the ratio of the total wall-time and the total number of QC calls, over all 1000 production runs. Moreover, for further transparency we report here (see Table D.1) the maximum and average number of QC calls per production run as well as the average and maximum computational time. The data show the large spread of computation times in the production runs depending on the fragmentation events and the stop criteria. The maximum computational time for a production run is often reached when the molecular ion survives while the maximum of QC calls is often related to production runs with many cascading trajectories, which are not necessarily more expensive due to the neutral losses being discounted. Table D.1 also reflects that the IP calculation by DFT (as was done for the organometallic systems **7-10** significantly increases the computational times.

Table D.1.: Average number of energy/gradient computations (QC calls) in a production run along with the standard deviation and the maximum number of QC calls, for the given molecules **1**–**23**. Furthermore, we show the average (along with standard deviation) and maximum computational time per production run, as well. To obtain an estimate of the total wall time, multiply the average t_{comp} by 1,000 and divide by the number of available cores (which has been, in our case 1,000).

Molecular index	**Avg. QC $_{calls}$**	**Max QC $_{calls}$**	**Avg. t_{comp} [s]**	**Max t_{comp} [s]**
1	8651.5 ± 6081.5	26319	489.8 ± 395.3	1592
2	10840.4 ± 6457.0	29979	670.45 ± 457.0	1633
3	7943.6 ± 5870.8	23277	424.48 ± 350.6	1294
4	10289.4 ± 5677.3	36008	574.6 ± 366.0	2071
5	7302.6 ± 5947.6	36008	99.3 ± 89.8	487
6	14312.5 ± 6403.7	44018	857.3 ± 595.6	3649
7	9403.9 ± 6423.2	36012	2664.4 ± 2176.4	24299
8	7700.9 ± 5336.5	28339	2061.9 ± 1010.7	5463
9	11976.1± 6718.1	36012	1627.7 ±939.5	7008
10	13622.0 ± 6467.3	33228	18697.2 ± 20969.7	135460
11	4536.6 ± 3760.5	22008	179.0 ± 147.7	893
12	3949.3 ± 2986.0	20002	183.5 ± 141.1	1031
13	12695.6 ± 5896.9	36008	734.0 ± 380.4	2030
14	14029.0 ± 5692.1	36008	442.3 ± 215.5	1090
15	9765.8 ± 5787.0	20596	538.6 ± 354.4	1278
16	10602.4 ± 6193.8	25066	850.6 ± 647.8	2393
17	11056.9 ± 5414.6	33222	340.1 ± 199.4	1067
18	5011.2 ± 5993.5	21535	85.9 ± 95.2	354
19	15218.0 ± 5323.3	29927	604.5 ± 272.2	1094
20	13378.5 ± 5770.2	42958	495.0 ± 248.0	1603
21	6196.1 ± 5938.1	22008	477.2 ± 481.8	1886
22	10457.0 ± 7647	22008	636.85 ± 482.6	1541
23	10931.9 ± 7872.8	20002	629.4 ± 490.4	1972

Bibliography

[1] Gross, J. *Mass Spectrometry - A Textbook*; Springer-Verlag: Weinheim, 2011.

[2] McLafferty, F. W.; Turecek, F. *Interpretation of Mass Spectra*; University Science Books, 1993.

[3] Sparkman, O. D.; Penton, Z.; Kitson, F. G. *Gas Chromatography and Mass Spectrometry*, 2nd ed.; Academic Press: Amsterdam, 2011.

[4] Gohlke, R. S.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 367–371.

[5] Hites, R. A. *Anal. Chem.* **2016**, *88*, 6955–6961.

[6] Tsivou, M.; Kioukia-Fougia, N.; Lyris, E.; Aggelis, Y.; Fragkaki, A.; Kiousi, X.; Simitsek, P.; Dimopoulou, H.; Leontiou, I.-P.; Stamou, M.; Spyridaki, M.-H.; Georgakopoulos, C. *Anal. Chim. Acta* **2006**, *555*, 1–13.

[7] Krasnopolsky, V. A.; Parshev, V. A. *Nature* **1981**, *292*, 610–613.

[8] Niemann, H. B.; et al., *Nature* **2005**, *438*, 779–784.

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

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

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

[12] Schymanski, E. L.; Gallampois, C. M. J.; Krauss, M.; Meringer, M.; Neumann, S.; Schulze, T.; Wolf, S.; Brack, W. *Anal. Chem.* **2012**, *84*, 3287–3295.

[13] Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*; Cambridge University Press: Cambridge, 2009.

[14] Rosenstock, H. M.; Wallenstein, M. B.; Wahrhaftig, A. L.; Eyring, H. *Proc. Natl. Acad. Sci.* **1952**, *38*, 667–678.

[15] Harrison, A. G.; Finney, C. D.; Sherk, J. A. *Org. Mass Spectrom.* **1971**, *5*, 1313–1320.

[16] Meisels, G. G.; Chen, C. T.; Giessner, B. G.; Emmel, R. H. *J. Chem. Phys.* **1972**, *56*, 793–800.

[17] McCarthy, I.; Weigold, E. *Phys. Rep.* **1976**, *27*, 275 – 371.

[18] McCarthy, I. E.; Weigold, E. *Rep. Prog. Phys.* **1991**, *54*, 789.

[19] Ying, J. F.; Zhu, H.; Mathers, C. P.; Gover, B. N.; Banjavčić, M. P.; Zheng, Y.; Brion, C. E.; Leung, K. T. *J. Chem. Phys.* **1993**, *98*, 4512–4519.

[20] Shojaei, S. H. R.; Morini, F.; Deleuze, M. S. *J. Phys. Chem. A* **2013**, *117*, 1918–1929.

[21] Shojaei, S. H. R.; Vandenbussche, J.; Deleuze, M. S.; Bultinck, P. *J. Phys. Chem. A* **2013**, *117*, 8388–8398.

[22] Rice, O. K.; Ramsperger, H. C. *J. Am. Chem. Soc.* **1927**, *49*, 1617–1629.

[23] Kassel, L. S. *J. Phys. Chem.* **1928**, *32*, 225–242.

[24] Marcus, R. A.; Rice, O. K. *J. Phys. Colloid Chem.* **1951**, *55*, 894–908.

[25] Marcus, R. A. *J. Chem. Phys.* **1952**, *20*, 359–364.

[26] Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data 17, suppl. 1*; 1988; p 861pp.

[27] Mermin, N. D. *Phys. Rev.* **1964**, *134*, A112–A125.

[28] Szabo, A.; Ostlund, N. *Modern Quantum Chemistry*, 1st ed.; McMillan Publishing Company: New York, 1982.

[29] Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69–89.

[30] Hall, G. G. *Proc. Roy. Soc. A* **1951**, *205*, 541–552.

[31] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864–B871.

[32] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133–A1138.

[33] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

[34] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.

[35] Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

[36] Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, NY, 1989.

[37] Perdew, J. P.; Schmidt, K. *AIP Conference Proceedings* **2001**, *577*, 1–20.

[38] Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.

[39] Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670–6688.

[40] Sun, J.; Ruzsinszky, A.; Perdew, J. P. *Phys. Rev. Lett.* **2015**, *115*, 036402.

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

[42] Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657–2664.

[43] Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. *Chem. Rev.* **2016**, *116*, 5301–5337.

[44] Warren, R. W.; Dunlap, B. I. *Chem. Phys. Lett.* **1996**, *262*, 384–392.

[45] Rabuck, A. D.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *110*, 695–700.

[46] Grimme, S.; Hansen, A. *Angew. Chem. Int. Ed.* **2015**, *54*, 12308–12313.

[47] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865, erratum: *Phys. Rev. Lett.* **1997** *78*, 1396.

[48] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

[49] Cremer, D. *WIREs Comp. Mol. Sci.* **2011**, *1*, 509–530.

[50] Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. *Chem. Rev.* **2016**, *116*, 5105–5154.

[51] Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. *Chem. Rev.* **2012**, *112*, 4–74.

[52] Alavi, A.; Kohanoff, J.; Parrinello, M.; Frenkel, D. *Phys. Rev. Lett.* **1994**, *73*, 2599–2602.

[53] Silvestrelli, P. L.; Alavi, A.; Parrinello, M.; Frenkel, D. *Phys. Rev. Lett.* **1996**, *77*, 3149–3152.

[54] Abe, M. *Chem. Rev.* **2013**, *113*, 7011–7088.

[55] Pulay, P.; Hamilton, T. P. *J. Chem. Phys.* **1988**, *88*, 4926–4933.

[56] Keller, S.; Boguslawski, K.; Janowski, T.; Reiher, M.; Pulay, P. *J. Chem. Phys.* **2015**, *142*, 244104.

[57] Stein, C. J.; Reiher, M. *J. Chem. Theory Comput.* **2016**, *12*, 1760–1771.

[58] Knecht, S.; Hedegård, E. D.; Keller, S.; Kovyrshin, S.; Ma, Y.; Muolo, A.; Stein, C. J.; Reiher, M. *Chimia* **2016**, *70*, 244–251.

[59] Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. *Nature* **1996**, *381*, 272.

[60] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

[61] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

[62] Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. *WIREs Comput. Mol. Sci.* **2014**, *4*, 91–100.

[63] TURBOMOLE: R. Ahlrichs et al., Universität Karlsruhe 2009. See http://www.turbomole.com.

[64] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. *J. Chem. Phys.* **2013**, *139*, 134101.

[65] Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. *J. Chem. Phys.* **2016**, *144*, 024109.

[66] Roos, B. O. *Advances in Chemical Physics*; John Wiley & Sons, Inc.: New York, NY, 1987; Vol. 69; pp 399–445.

[67] Angeli, C.; Cimiraglia, R.; Evangelisti, S.; Leininger, T.; Malrieu, J.-P. *J. Chem. Phys.* **2001**, *114*, 10252–10264.

[68] Angeli, C.; Cimiraglia, R.; Malrieu, J.-P. *Chem. Phys. Lett.* **2001**, *350*, 297 – 305.

[69] Angeli, C.; Cimiraglia, R.; Malrieu, J.-P. *J. Chem. Phys.* **2002**, *117*, 9138–9153.

[70] Neese, F. *WIREs Comp. Mol. Sci.* **2012**, *2*, 73–78.

[71] F. Neese, ORCA - an Ab Initio, Density Functional and Semiempirical Program Package, Ver. 2.9 (Rev 0), Max Planck Institute for Bioinorganic Chemistry, Germany, 2011.

[72] Grimme, S.; Bannwarth, C. *J. Chem. Phys.* **2016**, *145*, 054103.

[73] Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[74] Lu, T.; Chen, F. *J. Comput. Chem.* **2012**, *33*, 580–592.

[75] Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

[76] Sun, Z.; Ye, Q.; Chi, C.; Wu, J. *Chem. Soc. Rev.* **2012**, *41*, 7857–7889.

[77] Sun, Z.; Zeng, Z.; Wu, J. *Acc. Chem. Res.* **2014**, *47*, 2582–2591.

[78] Chikamatsu, M.; Mikami, T.; Chisaka, J.; Yoshida, Y.; Azumi, R.; Yase, K.; Shimizu, A.; Kubo, T.; Morita, Y.; Nakasuji, K. *Appl. Phys. Lett.* **2007**, *91*, 043506.

[79] Koike, H.; Chikamatsu, M.; Azumi, R.; Tsutsumi, J.; Ogawa, K.; Yamane, W.; Nishiuchi, T.; Kubo, T.; Hasegawa, T.; Kanai, K. *Adv. Func. Mater.* **2016**, *26*, 277–283.

[80] Morita, Y.; Nishida, S.; Murata, T.; Moriguchi, M.; Ueda, A.; Satoh, M.; Arifuku, K.; Sato, K.; Takui, T. *Nat. Mater.* **2011**, *10*, 947–951.

[81] Thompson, N. J.; Wilson, M. W. B.; Congreve, D. N.; Brown, P. R.; Scherer, J. M.; Bischof, T. S.; Wu, M.; Geva, N.; Welbornd, M.; Voorhis, T. V.; Bulović, V.; Bawendi, M. G.; Baldo, M. A. *Nat. Mater.* **2014**, *13*, 1039–1043.

[82] Kamada, K.; Ohta, K.; Kubo, T.; Shimizu, A.; Morita, Y.; Nakasuji, K.; Kishi, R.; Ohta, S.; Furukawa, S.-i.; Takahashi, H.; Nakano, M. *Angew. Chem. Int. Ed.* **2007**, *46*, 3544–3546.

[83] Kamada, K.; Ohta, K.; Shimizu, A.; Kubo, T.; Kishi, R.; Takahashi, H.; Botek, E.; Champagne, B.; Nakano, M. *J. Phys. Chem. Lett.* **2010**, *1*, 937–940.

[84] Doehnert, D.; Koutecky, J. *J. Am. Chem. Soc.* **1980**, *102*, 1789–1796.

[85] Nakano, M.; Kishi, R.; Nitta, T.; Nakasuji, T. K. K.; Kamada, K.; Ohta, K.; Champagne, B.; Botek, E.; Yamaguchi, K. *J. Phys. Chem. A* **2005**, *109*, 885–891.

[86] Thiele, J.; H., B. *Chem. Ber.* **1904**, *37*, 1463.

[87] Chichibabin, A. E. *Chem. Ber.* **1907**, *40*, 1810.

[88] Montgomery, L. K.; Huffman, J. C.; Jurczak, E. A.; Grendze, M. P. *J. Am. Chem. Soc.* **1986**, *108*, 6004–6011.

[89] Clar, E. *The Aromatic Sextet*; Wiley: London, 1972.

[90] Solà, M. *Front. Chem.* **2013**, *1*, 22.

[91] Sun, Z. et al. *J. Am. Chem. Soc.* **2013**, *135*, 18229–18236.

[92] Liu, J.; Ravat, P.; Wagner, M.; Baumgarten, M.; Feng, X.; Müllen, K. *Angew. Chem. Int. Ed.* **2015**, *54*, 12442–12446.

[93] Nakano, M.; Fukui, H.; Minami, T.; Yoneda, K.; Shigeta, Y.; Kishi, R.; Champagne, B.; Botek, E.; Kubo, T.; Ohta, K.; Kamada, K. *Theor. Chem. Acc.* **2011**, *130*, 711–724, erratum: 130, 725 (2011).

[94] Das, A.; Müller, T.; Plasser, F.; Lischka, H. *J. Phys. Chem. A* **2016**, *120*, 1625–1636.

[95] Li, Y. et al. *J. Am. Chem. Soc.* **2012**, *134*, 14913–14922.

[96] Ishida, M.; Shin, J.-Y.; Lim, J. M.; Lee, B. S.; Yoon, M.-C.; Koide, T.; Sessler, J. L.; Osuka, A.; Kim, D. *J. Am. Chem. Soc.* **2011**, *133*, 15533–15544.

[97] Baerends, E. J.; Gritsenko, O. V. *J. Phys. Chem. A* **1997**, *101*, 5383–5403.

[98] Cremer, D. *Mol. Phys.* **2001**, *99*, 1899–1940.

[99] Veryazov, V.; Malmqvist, P. A. k.; Roos, B. O. *Int. J. Quant. Chem.* **2011**, *111*, 3329–3338.

[100] Kollmar, H.; Staemmler, V. *J. Am. Chem. Soc.* **1977**, *99*, 3583–3587.

[101] Wierschke, S. G.; Nash, J. J.; Squires, R. R. *J. Am. Chem. Soc.* **1993**, *115*, 11958–11967.

[102] Li, H.; Yu, S.-Y.; Huang, M.-B.; Wang, Z.-X. *Chem. Phys. Lett.* **2007**, *450*, 12 – 18.

[103] Wang, E. B.; Parish, C. A.; Lischka, H. *J. Chem. Phys.* **2008**, *129*, 044306.

[104] Aiga, F. *J. Phys. Chem. A* **2012**, *116*, 663–669.

[105] Bill, E.; Bothe, E.; Chaudhuri, P.; Chlopek, K.; Herebian, D.; Kokatam, S.; Ray, K.; Weyhermüller, T.; Neese, F.; Wieghardt, K. *Chem. Eur. J.* **2005**, *11*, 204–224.

[106] Zhang, W.; Loebach, J. L.; Wilson, S. R.; Jacobsen, E. N. *J. Am. Chem. Soc.* **1990**, *112*, 2801–2803.

[107] Irie, R.; Noda, K.; Ito, Y.; Matsumoto, N.; Katsuki, T. *Tetrahedron Lett.* **1990**, *31*, 7345 – 7348.

[108] Ivanic, J.; Collins, J. R.; Burt, S. K. *J. Phys. Chem. A* **2004**, *108*, 2314–2323.

[109] Sears, J. S.; Sherrill, C. D. *J. Chem. Phys.* **2006**, *124*, 144314.

[110] Ma, D.; Li Manni, G.; Gagliardi, L. *J. Chem. Phys.* **2011**, *135*, 044128.

[111] Wouters, S.; Bogaerts, T.; Van Der Voort, P.; Van Speybroeck, V.; Van Neck, D. *J. Chem. Phys.* **2014**, *140*, 241103.

[112] Olivares-Amaya, R.; Hu, W.; Nakatani, N.; Sharma, S.; Yang, J.; Chan, G. K.-L. *J. Chem. Phys.* **2015**, *142*, 034102.

[113] Ignarro, L. J., Ed. *Nitric Oxide: Biology and Pathobiology*; Academic Press: San Diego, CA, 2000.

[114] Freitag, L.; Knecht, S.; Keller, S. F.; Delcey, M. G.; Aquilante, F.; Bondo Pedersen, T.; Lindh, R.; Reiher, M.; Gonzalez, L. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14383–14392.

[115] Bonačić-Koutecký, V.; Koutecký, J.; Michl, J. *Angew. Chem. Int. Ed. Engl.* **1987**, *26*, 170–189.

[116] Olivucci, M.; Ragazos, I. N.; Bernardi, F.; Robb, M. A. *J. Am. Chem. Soc.* **1993**, *115*, 3710–3721.

[117] Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158.

[118] Terakita, A. *Genome Biol.* **2005**, *6*, 1–9.

[119] Garavelli, M.; Celani, P.; Bernardi, F.; Robb, M. A.; ; Olivucci, M. *J. Am. Chem. Soc.* **1997**, *119*, 6891–6901.

[120] González-Luque, R.; Garavelli, M.; Bernardi, F.; Merchán, M.; Robb, M. A.; Olivucci, M. *Proc. Natl. Acad. Sci.* **2000**, *97*, 9379–9384.

[121] Send, R.; Sundholm, D. *J. Phys. Chem. A* **2007**, *111*, 8766–8773.

[122] Tuna, D.; Lefrancois, D.; Łukasz Wolanski,; Gozem, S.; Schapiro, I.; Andruniów, T.; Dreuw, A.; Olivucci, M. *J. Chem. Theory Comput.* **2015**, *11*, 5758–5781.

[123] Manathunga, M.; Yang, X.; Luk, H. L.; Gozem, S.; Frutos, L. M.; Valentini, A.; Ferrè, N.; Olivucci, M. *J. Chem. Theory Comput.* **2016**, *12*, 839–850.

[124] Gozem, S.; Schapiro, I.; Ferré, N.; Olivucci, M. *Science* **2012**, *337*, 1225–1228.

[125] Gozem, S.; Huntress, M.; Schapiro, I.; Lindh, R.; Granovsky, A. A.; Angeli, C.; Olivucci, M. *J. Chem. Theory Comput.* **2012**, *8*, 4069–4080.

[126] Saha, S.; Stoddart, J. F. *Chem. Soc. Rev.* **2007**, *36*, 77–92.

[127] Markworth, P. B.; Adamson, B. D.; Coughlan, N. J. A.; Goerigk, L.; Bieske, E. J. *Phys. Chem. Chem. Phys.* **2015**, *17*, 25676–25688.

[128] Hajjar, L.; Hicks, R. G.; Zeng, T. *J. Phys. Chem. A* **2016**, *120*, 7569–7576.

[129] Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. *Chem. Rev.* **2005**, *105*, 2253–2278.

[130] Meunier, B.; de Visser, S. P.; Shaik, S. *Chem. Rev.* **2004**, *104*, 3947–3980.

[131] Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279–2328.

[132] Jung, C. *Biochim. Biophys. Acta* **2011**, *1814*, 46 – 57.

[133] Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science* **2000**, *287*, 1615–1622.

[134] Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221–234.

[135] Schrödinger Release 2016-2: Schrödinger Suite 2016-2 Protein Preparation Wizard; Epik version 3.7, Schrödinger, LLC, New York, NY, 2016; Impact version 7.2, Schrödinger, LLC, New York, NY, 2016; Prime version 4.5, Schrödinger, LLC, New York, NY, 2016.

[136] Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.

[137] Radoń, M.; Broclawik, E. *J. Chem. Theory Comput.* **2007**, *3*, 728–734.

[138] Chen, H.; Lai, W.; Shaik, S. *J. Phys. Chem. B* **2011**, *115*, 1727–1742.

[139] Im, W.; Feig, M.; III, C. L. B. *Biophys. J.* **2003**, *85*, 2900 – 2918.

[140] Mann, S. I.; Heinisch, T.; Weitz, A. C.; Hendrich, M. P.; Ward, T. R.; Borovik, A. S. *J. Am. Chem. Soc.* **2016**, *138*, 9073–9076.

[141] Yu, F.; Cangelosi, V. M.; Zastrow, M. L.; Tegoni, M.; Plegaria, J. S.; Tebo, A. G.; Mocny, C. S.; Ruckthong, L.; Qayyum, H.; Pecoraro, V. L. *Chem. Rev.* **2014**, *114*, 3495–3578.

[142] Silvian, L.; Jin, P.; Carmillo, P.; Boriack-Sjodin, P. A.; Pelletier, C.; Rushe, M.; Gong, B.; Sah, D.; Pepinsky, B.; Rossomando, A. *Biochemistry* **2006**, *45*, 6801–6812.

[143] Chai, J.-D. *J. Chem. Phys.* **2012**, *136*, 154104.

[144] Becke, A. D. In *Top. Curr. Chem.*; Johnson, E. R., Ed.; Springer International Publishing: Cham, 2014; pp 175–186, doi: 10.1007/128_2014_581.

[145] Rasche, F.; Svatoš, A.; Maddula, R. K.; Böttcher, C.; Böcker, S. *Anal. Chem.* **2011**, *83*, 1243–1251.

[146] Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. *Nucleic Acids Res.* **2014**,

[147] Allen, F.; Greiner, R.; Wishart, D. *Metabolomics* **2014**, *11*, 98–110.

[148] Mistrík, R. Mass Frontier^{TM} 3.0; http://www.highchem.com.

[149] Baer, T.; Hase, W. L. *Unimolecular Reaction Dynamics, Theory and Experiments*; Oxford University Press: New York, NY, 1996.

[150] Murray, K. K.; Boyd, R. K.; Eberlin, M. N.; Langley, G. J.; Li, L.; Naito, Y. *Pure Appl. Chem.* **2013**, *85*, 1515–1609.

[151] Baer, T.; Mayer, P. M. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 103–115.

[152] Beyer, T.; Swinehart, D. F. *Commun. ACM* **1973**, *16*, 379.

[153] Yao, L.; Mebel, A. M.; Lu, H. F.; Neusser, H. J.; Lin, S. H. *J. Phys. Chem. A* **2007**, *111*, 6722–6729.

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

[155] Ervin, K. M. *Int. J. Mass Spectrom.* **2000**, *195-196*, 271 – 284.

[156] Lorquet, J. *Int. J. Mass Spectrom.* **2000**, *200*, 43 – 56, Volume 200: The State of the Field as We Move Into a New Millenium.

[157] Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. *J. Phys. Chem.* **1996**, *100*, 12771–12800.

[158] Armentrout, P. B.; Ervin, K. M.; Rodgers, M. T. *J. Phys. Chem. A* **2008**, *112*, 10071–10085.

[159] Armentrout, P. B.; Heaton, A. L.; Ye, S. J. *J. Phys. Chem. A* **2011**, *115*, 11144–11155.

[160] Lifshitz, C. *Acc. Chem. Res.* **1994**, *27*, 138–144.

[161] Lifshitz, C.; Gotkis, Y.; Ioffe, A.; Laskin, J.; Shaik, S. *Int. J. Mass Spectrom. Ion Processes* **1993**, *125*, R7 – R11.

[162] Dewar, M. J. S.; Landman, D. *J. Am. Chem. Soc.* **1977**, *99*, 2446–2453.

[163] Moon, J. H.; Choe, J. C.; ; Kim, M. S. *J. Phys. Chem. A* **2000**, *104*, 458–463.

[164] Choe, J. C. *J. Phys. Chem. A* **2006**, *110*, 7655–7662.

[165] Kim, Y. H.; Choe, J. C.; ; Kim, M. S. *J. Phys. Chem. A* **2001**, *105*, 5751–5758.

[166] Choe, J. C. *J. Phys. Chem. A* **2008**, *112*, 6190–6197.

[167] Choe, J. C. *Int. J. Mass Spectrom.* **2008**, *278*, 50 – 58.

[168] Vékey, K. *J. Mass Spectrom.* **1996**, *31*, 445–463.

[169] Knyazev, V. D.; Stein, S. E. *J. Phys. Chem. A* **2010**, *114*, 6384–6393.

[170] Chiavarino, B.; Crestoni, M. E.; Dopfer, O.; Maitre, P.; Fornarini, S. *Angew. Chem. Int. Ed.* **2012**, *51*, 4947–4949.

[171] Halbert, S.; Bouchoux, G. *J. Phys. Chem. A* **2012**, *116*, 1307–1315.

[172] Muntean, F.; Armentrout, P. B. *J. Phys. Chem. B* **2002**, *106*, 8117–8124.

[173] Muntean, F.; ; Armentrout, P. B. *J. Phys. Chem. A* **2003**, *107*, 7413–7422.

[174] Kim, S. Y.; Choe, J. C. *Int. J. Mass Spectrom.* **2010**, *295*, 65 – 71.

[175] Letzel, M.; Barth, D.; Kuck, D.; Grützmacher, H.-F.; Kim, S. Y.; Yim, M. K.; Choe, J. C. *Int. J. Mass Spectrom.* **2013**, *336*, 1 – 16.

[176] Wolken, J. K.; Yao, C.; Tureček, F.; Polce, M. J.; Wesdemiotis, C. *Int. J. Mass Spectrom.* **2007**, *267*, 30–42, Sharon G. Lias Honour Issue.

[177] Lewis, W. K.; Applegate, B. E.; Sztáray, J.; Sztáray, B.; Baer, T.; Bemish, R. J.; Miller, R. E. *J. Am. Chem. Soc.* **2004**, *126*, 11283–11292.

[178] Vallejo-Narváez, W. E.; Villota, P. V. B.; Espinoza, E. A. S. *J. Phys. Chem. A* **2012**, *116*, 12136–12147.

[179] Tureček, F.; Julian, R. R. *Chem. Rev.* **2013**, *113*, 6691–6733.

[180] Lai, C.-K.; Mu, X.; Hao, Q.; Hopkinson, A. C.; Chu, I. K. *Phys. Chem. Chem. Phys.* **2014**, *16*, 24235–24243.

[181] Sztáray, B.; Bodi, A.; Baer, T. *J. Mass Spectrom.* **2010**, *45*, 1233–1245.

[182] Sztáray, B.; Baer, T. *J. Am. Chem. Soc.* **2000**, *122*, 9219–9226.

[183] Rennie, E. E.; Cooper, L.; Shpinkova, L. G.; Holland, D. M. P.; Shaw, D. A.; Guest, M. F.; Mayer, P. M. *J. Phys. Chem. A* **2009**, *113*, 5823–5831.

[184] Bodi, A.; Stevens, W. R.; Baer, T. *J. Phys. Chem. A* **2011**, *115*, 726–734.

[185] Bodi, A.; Johnson, M.; Gerber, T.; Gengeliczki, Z.; Sztáray, B.; Baer, T. *Rev. Sci. Instrum.* **2009**, *80*, 034101.

[186] West, B.; Joblin, C.; Blanchet, V.; Bodi, A.; Sztáray, B.; Mayer, P. M. *J. Phys. Chem. A* **2012**, *116*, 10999–11007.

[187] West, B.; Joblin, C.; Blanchet, V.; Bodi, A.; Sztáray, B.; Mayer, P. M. *J. Phys. Chem. A* **2014**, *118*, 1807–1816.

[188] West, B.; Sit, A.; Mohamed, S.; Joblin, C.; Blanchet, V.; Bodi, A.; Mayer, P. M. *J. Phys. Chem. A* **2014**, *118*, 9870–9878.

[189] Chabot, M.; Béroff, K.; Gratier, P.; Jallat, A.; Wakelam, V. *Astrophys. J.* **2013**, *771*, 90.

[190] Tsyshevsky, R.; Garifzianova, G.; Shamov, A.; Khrapkovskii, G. *Int. J. Mass Spectrom.* **2014**, *369*, 36 – 43.

[191] Solano, E. A.; Mayer, P. M. *J. Chem. Phys.* **2015**, *143*, 104305.

[192] Bouwman, J.; de Haas, A. J.; Oomens, J. *Chem. Commun.* **2016**, *52*, 2636–2638.

[193] Mayer, I.; Gömöry, A. *Int. J. Quant. Chem.* **1993**, *27*, 599–605.

[194] Mayer, I.; Gömöry, A. *J. Mol. Struct. Theochem* **1994**, *311*, 331–341.

[195] Mayer, I.; Gömöry, A. *Chem. Phys. Lett.* **2001**, *344*, 553 – 564.

[196] Gömöry, A.; Somogyi, A.; Tamás, J.; Stájer, G.; Bernáth, G.; Komáromi, I. *Int. J. Mass Spectrom. Ion Processes* **1991**, *107*, 225 – 246.

[197] Vékey, K.; Somogyi, A.; Tarnás, J.; Pócsfalvi, G. *Org. Mass Spectrom.* **1992**, *27*, 869–875.

[198] McCormack, A. L.; Somogyi, A.; Dongre, A. R.; Wysocki, V. H. *Anal. Chem.* **1993**, *65*, 2859–2872.

[199] Zayed, M.; Hawash, M.; Fahmey, M. *Spectrochim. Act. Mol. Biomol. Spectrosc.* **2006**, *64*, 363 – 371.

[200] Improta, R.; Scalmani, G.; Barone, V. *Int. J. Mass Spectrom.* **2000**, *201*, 321–336.

[201] Arani, L. S.; Mignon, P.; Abdoul-Carime, H.; Farizon, B.; Farizon, M.; Chermette, H. *Phys. Chem. Chem. Phys.* **2012**, *14*, 9855–9870.

[202] Sadr-Arani, L.; Mignon, P.; Chermette, H.; Abdoul-Carime, H.; Farizon, B.; Farizon, M. *Phys. Chem. Chem. Phys.* **2015**, *17*, 11813–11826.

[203] Minaev, B. F.; Shafranyosh, M. I.; Svida, Y. Y.; Sukhoviya, M. I.; Shafranyosh, I. I.; Baryshnikov, G. V.; Minaeva, V. A. *J. Chem. Phys.* **2014**, *140*, 175101.

[204] Dawley, M. M.; Tanzer, K.; Cantrell, W. A.; Plattner, P.; Brinkmann, N. R.; Scheier, P.; Denifl, S.; Ptasinska, S. *Phys. Chem. Chem. Phys.* **2014**, *16*, 25039–25053.

[205] Cheng, P.; Li, Y.; Li, S.; Zhang, M.; Zhou, Z. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4667–4677.

[206] Karton, A.; Goerigk, L. *J. Comput. Chem.* **2015**, *36*, 622–632.

[207] Tully, J. C. *Faraday Discuss.* **1998**, *110*, 407–419.

[208] Sun, L.; Hase, W. L. *Reviews in Computational Chemistry*; John Wiley & Sons, Inc., 2003; pp 79–146.

[209] Liu, J.; Song, K.; Hase, W. L.; Anderson, S. L. *J. Chem. Phys.* **2003**, *119*, 3040–3050.

[210] Hu, X.; Hase, W. L.; Pirraglia, T. *J. Comput. Chem.* **1991**, *12*, 1014–1024.

[211] Chen, W.; Hase, W. L.; Schlegel, H. *Chem. Phys. Lett.* **1994**, *228*, 436 – 442.

[212] González-Vázquez, J.; Martínez-Núñez, E.; Vázquez, S.; Santamaría, J.; Bañares, L. *Chem. Phys. Lett.* **2004**, *396*, 442 – 447.

[213] Martínez-Núñez, E.; Fernández-Ramos, A.; Vázquez, S. A.; Marques, J. M. C.; Xue, M.; Hase, W. L. *J. Chem. Phys.* **2005**, *123*, 154311.

[214] Spezia, R.; Salpin, J.-Y.; Gaigeot, M.-P.; Hase, W. L.; Song, K. *J. Phys. Chem. A* **2009**, *113*, 13853–13862.

[215] Martín-Sómer, A.; Yáñez, M.; Gaigeot, M.-P.; Spezia, R. *J. Phys. Chem. A* **2014**, *118*, 10882–10893.

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

[217] Spezia, R.; Lee, S. B.; Cho, A.; Song, K. *Int. J. Mass Spectrom.* **2015**, *392*, 125 – 138.

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

[219] Spezia, R.; Martens, J.; Oomens, J.; Song, K. *Int. J. Mass Spectrom.* **2015**, *388*, 40 – 52.

[220] Boudaïffa, B.; Cloutier, P.; Hunting, D.; Huels, M. A.; Sanche, L. *Science* **2000**, *287*, 1658–1660.

[221] Sanche, L. *Eur. Phys. J. D* **2005**, *35*, 367–390.

[222] Flosadóttir, H.; Denifl, S.; Zappa, F.; Wendt, N.; Mauracher, A.; Bacher, A.; Jónsson, H.; Märk, T.; Scheier, P.; Ingólfsson, O. *Angew. Chem. Int. Ed.* **2007**, *46*, 8057–8059.

[223] Flosadottir, H. D.; Jonsson, H.; Sigurdsson, S. T.; Ingolfsson, O. *Phys. Chem. Chem. Phys.* **2011**, *13*, 15283–15290.

[224] Omarsson, B.; Bjarnason, E. H.; Haughey, S. A.; Field, T. A.; Abramov, A.; Klupfel, P.; Jonsson, H.; Ingolfsson, O. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4754–4766.

[225] Townsend, D.; Lahankar, S. A.; Lee, S. K.; Chambreau, S. D.; Suits, A. G.; Zhang, X.; Rheinecker, J.; Harding, L. B.; Bowman, J. M. *Science* **2004**, *306*, 1158–1161.

[226] Morton, T. H. *Org. Mass Spectrom.* **1992**, *27*, 353–368.

[227] Chronister, E. L.; Morton, T. H. *J. Am. Chem. Soc.* **1990**, *112*, 133–139.

[228] Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495–506.

[229] Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149–5155.

[230] Gaus, M.; Goez, A.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.

[231] Kruse, H.; Grimme, S. *J. Chem. Phys.* **2012**, *136*, 154101.

[232] Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.

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

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

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

[236] Sethi, S.; Gupta, S. P.; Jenkins, E. E.; Whitehead, C. W.; Townsend, L. B.; McCloskey, J. A. *J. Am. Chem. Soc.* **1982**, *104*, 3349–3353.

[237] Stevenson, D. P. *Discuss. Faraday Soc.* **1951**, *10*, 35–45.

[238] Bursey, M. M.; Wolfe, E. S. *Org. Mass Spectrom.* **1968**, *1*, 543–549.

[239] Griller, D.; Lossing, F. P. *J. Am. Chem. Soc.* **1981**, *103*, 1586–1587.

[240] Andrews, L.; Dyke, J. M.; Jonathan, N.; Keddar, N.; Morris, A. *J. Am. Chem. Soc.* **1984**, *106*, 299–303.

[241] Dyke, J. M. *J. Chem. Soc., Faraday Trans. 2* **1987**, *83*, 69–87.

[242] Ruscic, B.; Berkowitz, J. *J. Chem. Phys.* **1992**, *97*, 1818–1823.

[243] Stein, S. E. In *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; Linstrom, P. J., Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg MD 20899, http://webbook.nist.gov, (retrieved Dec 15, 2014).

[244] SDBS Web: http://sdbs.riodb.ais.go.jp (National Institute of Advanced Industrial Science and Technology), accessed December 15, 2014.

[245] MNDO2005, Version 7.0, W. Thiel, MPI für Kohlenforschung, Mülheim, Germany.

[246] Aradi, B.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.

[247] Grunenberg, J., Ed. *Computational Spectroscopy*; Wiley-VCH: Weinheim, 2010.

[248] Drahos, L.; Vékey, K. *J. Mass Spec.* **2001**, *36*, 237–263.

[249] Timmermans, P. B.; Wong, P. C.; Chiu, A. T.; Herblin, W. F.; Benfield, P.; Carini, D. J.; Lee, R. J.; Wexler, R. R.; Saye, J. A.; Smith, R. D. *Pharmacol. Rev.* **1993**, *45*, 205–251.

[250] Wiley, P. F.; Gerzon, K.; Flynn, E. H.; Sigal, M. V.; Weaver, O.; Quarck, U. C.; Chauvette, R. R.; Monahan, R. *J. Am. Chem. Soc.* **1957**, *79*, 6062–6070.

[251] Masamune, S.; Bates, G. S.; Corcoran, J. W. *Angew. Chem. Int. Ed.* **1977**, *16*, 585–607.

[252] Wani, M. C.; Taylor, H. L.; Wall, M. E.; Coggon, P.; McPhail, A. T. *J. Am. Chem. Soc.* **1971**, *93*, 2325–2327.

[253] Schiff, P. B.; Fant, J.; Horwitz, S. B. *Nature* **1979**, *277*, 665–667.

[254] Gazzerro, P.; Proto, M. C.; Gangemi, G.; Malfitano, A. M.; Ciaglia, E.; Pisanti, S.; Santoro, A.; Laezza, C.; Bifulco, M. *Pharmacol. Rev.* **2012**, *64*, 102–146.

[255] Folkard, M.; Prise, K.; Vojnovic, B.; Davies, S.; Roper, M.; Michael, B. *Int. J. Radiat. Biol.* **1993**, *64*, 651–658.

[256] Huels, M. A.; Boudaïffa, B.; Cloutier, P.; Hunting, D.; Sanche, L. *J. Am. Chem. Soc.* **2003**, *125*, 4467–4477.

[257] Rice, J. M.; Dudek, G. O. *J. Am. Chem. Soc.* **1967**, *89*, 2719–2725.

[258] Occolowitz, J. L. *Chem. Commun. (London)* **1968**, 1226–1227.

[259] Leonard, N. J.; Henderson, T. R. *J. Am. Chem. Soc.* **1975**, *97*, 4990–4999.

[260] Barrio, M. D. C. G.; Scopes, D. I. C.; Holtwick, J. B.; Leonard, N. J. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 3986–3988.

[261] Nelson, C. C.; McCloskey, J. A. *J. Am. Chem. Soc.* **1992**, *114*, 3661–3668.

[262] Nolting, D.; Weinkauf, R.; Hertel, I. V.; Schultz, T. *ChemPhysChem* **2007**, *8*, 751–755.

[263] Alvarado, F.; Bari, S.; Hoekstra, R.; Schlathölter, T. *J. Chem. Phys.* **2007**, *127*, 034301.

[264] Brédy, R.; Bernard, J.; Chen, L.; Montagne, G.; Li, B.; Martin, S. *J. Chem. Phys.* **2009**, *130*, 114305.

[265] Tabet, J.; Eden, S.; Feil, S.; Abdoul-Carime, H.; Farizon, B.; Farizon, M.; Ouaskit, S.; Märk, T. *Int. J. Mass Spectrom.* **2010**, *292*, 53–63.

[266] Cheong, N. R.; Nam, S. H.; Park, H. S.; Ryu, S.; Song, J. K.; Park, S. M.; Perot, M.; Lucas, B.; Barat, M.; Fayeton, J. A.; Jouvet, C. *Phys. Chem. Chem. Phys.* **2011**, *13*, 291–295.

[267] Jochims, H.-W.; Schwell, M.; Baumgärtel, H.; Leach, S. *Chem. Phys.* **2005**, *314*, 263–282.

[268] Tureček, F.; Chen, X. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1713–1726.

[269] Piacenza, M.; Grimme, S. *J. Comput. Chem.* **2004**, *25*, 83–99.

[270] Verlet, L. *Phys. Rev.* **1967**, *159*, 98–103.

[271] Turro, N. J. *Modern Molecular Photochemistry*; The Benjamin/Cummings Publishing Company Inc.: Menlo Park, CA, 1978.

[272] Bursey, J. T.; Bursey, M. M.; Kingston, D. G. I. *Chem. Rev.* **1973**, *73*, 191–234.

[273] Steenken, S. *Chem. Rev.* **1989**, *89*, 503–520.

[274] Wallace, S. S. *Free Rad. Bio. Med.* **2002**, *33*, 1–14.

[275] Kanvah, S.; Joseph, J.; Schuster, G. B.; Barnett, R. N.; Cleveland, C. L.; Landman, U. *Acc. Chem. Res.* **2010**, *43*, 280–287.

[276] Ptasińska, S.; Denifl, S.; Gohlke, S.; Scheier, P.; Illenberger, E.; Märk, T. D. *Angew. Chem. Int. Ed.* **2006**, *45*, 1893–1896.

[277] Rice, J. M.; Dudek, G. O.; Barber, M. *J. Am. Chem. Soc.* **1965**, *87*, 4569–4576.

[278] Coupier, B.; Farizon, B.; Farizon, M.; Gaillard, M.; Gobet, F.; de Castro Faria, N.; Jalbert, G.; Ouaskit, S.; Carré, M.; Gstir, B.; Hanel, G.; Denifl, S.; Feketeova, L.; Scheier, P.; Märk, T. *Eur. Phys. J. D.* **2002**, *20*, 459–468.

[279] Imhoff, M.; Deng, Z.; Huels, M. A. *Int. J. Mass Spectrom.* **2005**, *245*, 68–77.

[280] Barc, B.; Ryszka, M.; Spurrell, J.; Dampc, M.; Limão Vieira, P.; Parajuli, R.; Mason, N. J.; Eden, S. *J. Chem. Phys.* **2013**, *139*, –.

[281] Zhou, C.; Matsika, S.; Kotur, M.; Weinacht, T. C. *J. Phys. Chem. A* **2012**, *116*, 9217–9227.

[282] Arani, L. S.; Mignon, P.; Abdoul-Carime, H.; Farizon, B.; Farizon, M.; Chermette, H. *Chem. Phys. Lett.* **2013**, *583*, 165–169.

[283] Sadr-Arani, L.; Mignon, P.; Chermette, H.; Douki, T. *Chem. Phys. Lett.* **2014**, –.

[284] Kotur, M.; Weinacht, T. C.; Zhou, C.; Kistler, K. A.; Matsika, S. *J. Chem. Phys.* **2011**, *134*, 184309.

[285] Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. *Nat. Chem.* **2014**, *6*, 1044–1048.

[286] Hase, W. L.; Buckowski, D. G. *J. Comput. Chem.* **1982**, *3*, 335–343.

[287] Yang, L.; Sun, R.; Hase, W. L. *J. Chem. Theory Comput.* **2011**, *7*, 3478–3483.

[288] Nir, E.; Müller, M.; Grace, L.; de Vries, M. *Chem. Phys. Lett.* **2002**, *355*, 59–64.

[289] Choi, M. Y.; Miller, R. E. *J. Am. Chem. Soc.* **2006**, *128*, 7320–7328.

[290] Feyer, V.; Plekan, O.; Richter, R.; Coreno, M.; Vall-llosera, G.; Prince, K. C.; Trofimov, A. B.; Zaytseva, I. L.; Moskovskaya, T. E.; Gromov, E. V.; Schirmer, J. *J. Phys. Chem. A* **2009**, *113*, 5736–5742.

[291] Plekan, O.; Feyer, V.; Richter, R.; Coreno, M.; Vall-llosera, G.; Prince, K. C.; Trofimov, A. B.; Zaytseva, I. L.; Moskovskaya, T. E.; Gromov, E. V.; Schirmer, J. *J. Phys. Chem. A* **2009**, *113*, 9376–9385.

[292] Zhou, J.; Kostko, O.; Nicolas, C.; Tang, X.; Belau, L.; de Vries, M. S.; Ahmed, M. *J. Phys. Chem. A* **2009**, *113*, 4829–4832.

[293] Trygubenko, S. A.; Bogdan, T. V.; Rueda, M.; Orozco, M.; Luque, F. J.; Sponer, J.; Slavicek, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4192–4203.

[294] Fogarasi, G. *J. Phys. Chem. A* **2002**, *106*, 1381–1390.

[295] Bravaya, K. B.; Kostko, O.; Dolgikh, S.; Landau, A.; Ahmed, M.; Krylov, A. I. *J. Phys. Chem. A* **2010**, *114*, 12305–12317.

[296] Marian, C. M. *J. Phys. Chem. A* **2007**, *111*, 1545–1553.

[297] Mons, M.; Piuzzi, F.; Dimicoli, I.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. A* **2006**, *110*, 10921–10924.

[298] Dreyfus, M.; Bensaude, O.; Dodin, G.; Dubois, J. E. *J. Am. Chem. Soc.* **1976**, *98*, 6338–6349.

[299] Colarusso, P.; Zhang, K.; Guo, B.; Bernath, P. F. *Chem. Phys. Lett.* **1997**, *269*, 39–48.

[300] Choi, M. Y.; Miller, R. E. *J. Phys. Chem. A* **2007**, *111*, 2475–2479.

[301] Plekan, O.; Feyer, V.; Richter, R.; Coreno, M.; de Simone, M.; Prince, K.; Trofimov, A.; Gromov, E.; Zaytseva, I.; Schirmer, J. *Chem. Phys.* **2008**, *347*, 360–375.

[302] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243–252.

[303] Neese, F.; Valeev, E. F. *J. Chem. Theory Comput.* **2011**, *7*, 33–43.

[304] Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[305] Diels, O.; Alder, K. *Chem. Ber.* **1929**, *62*, 2337–2372.

[306] Kwart, H.; King, K. *Chem. Rev.* **1968**, *68*, 415–447.

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

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

[309] Meroueh, S. O.; Wang, Y.; Hase, W. L. *J. Phys. Chem. A* **2002**, *106*, 9983–9992.

[310] Wang, Y.; Hase, W. L.; Song, K. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 1402–1412.

[311] Park, K.; Deb, B.; Song, K.; Hase, W. L. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 939–948.

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

[313] Stewart, J. J. P. *J. Comput. Chem.* **1991**, *12*, 320–341.

[314] Stewart, J. J. *J. Mol. Mod.* **2013**, *19*, 1–32.

[315] Cui, Q. *J. Chem. Phys.* **2016**, *145*, 140901.

[316] MOPAC2012: Steward, James JP, Stewart Computational Chemistry: Colorado Springs, CO, 2012.

[317] Wee, S.; O'Hair, R. A.; McFadyen, W. *Int. J. Mass Spectrom.* **2004**, *234*, 101 – 122, Special Issue in Honour of Alan G. Marshall.

[318] Koszinowski, K.; Müller, C.; Brand, H.; Fleckenstein, J. E. *Organometallics* **2012**, *31*, 7165–7170.

[319] Zavras, A.; Khairallah, G. N.; Krstić, M.; Girod, M.; Daly, S.; Antoine, R.; Maître, P.; Mulder, R. J.; Alexander, S.-A.; Banačić-Koutecký, V.; Dugourd, P.; O'Hair, R. A. J. *Nat. Commun.* **2016**, *7*, 11746.

[320] Geng, C.; Li, J.; Weiske, T.; Schlangen, M.; Shaik, S.; Schwarz, H. *J. Am. Chem. Soc.* **2017**, *139*, 1684–1689.

[321] Sun, X.; Zhou, S.; Schlangen, M.; Schwarz, H. *Chem. Eur. J.* **2017**, *23*, 1498–1501.

[322] Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833–1840.

[323] Grimme, S. *J. Chem. Phys.* **2013**, *138*, 244104.

[324] Bannwarth, C.; Grimme, S. *Comput. Theor. Chem.* **2014**, *1040*, 45–53.

[325] Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2012**, *8*, 527–541.

[326] Bauer, C. A.; Hansen, A.; Grimme, S. *Chem. Eur. J.* **2017**, doi:10.1002/chem.201604682.

[327] Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656–5667.

[328] Klopper, W.; Kutzelnigg, W. *J. Phys. Chem.* **1990**, *94*, 5625–5630.

[329] Andrei, H.-S.; Solcà, N.; Dopfer, O. *Angew. Chem. Int. Ed.* **2008**, *47*, 395–397.

[330] McLafferty, F. W. *Anal. Chem.* **1959**, *31*, 82–87.

[331] Patzer, A.; Chakraborty, S.; Solcà, N.; Dopfer, O. *Angew. Chem. Int. Ed.* **2010**, *49*, 10145–10148.

[332] Breslow, R.; Groves, J. T.; Ryan, G. *J. Am. Chem. Soc.* **1967**, *89*, 5048–5048.

[333] Phung, Q. M.; Vancoillie, S.; Pierloot, K. *J. Chem. Theory Comput.* **2014**, *10*, 3681–3688.

[334] Szostek, B.; Aldstadt, J. H. *J. Chromatogr. A* **1998**, *807*, 253 – 263.

[335] Korth, M.; Pitoňák, M.; Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 344–352.

[336] Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.; Mehta, S. S.; Wohlgemuth, G.; Barupal, D. K.; Showalter, M. R.; Arita, M.; Fiehn, O. *Mass Spectrom. Rev.* **2017**, DOI:10.1002/mas.21535.

[337] T. Frauenheim, DFTB+ (Density Functional based Tight Binding) 2008, see http://www.dftb.org/.

Acknowledgments

I thank my supervisor Prof. Dr. Stefan Grimme for being an adviser who has always given me freedom to research whatever I pleased, within and without the bounds of computational mass spectrometry. He has set a shining example of how science ought to be conducted. I will always remember the trust that he has put in me and my co-workers during my time as a PhD student at the University of Bonn.

I thank Prof. Dr. Barbara Kirchner for being my second referee and for our collaboration concerning the practical course in computational chemistry for undergraduate students.

Moreover, I thank Priv. Doz. Dr. Marianne Engeser for being a member of the commission and for our fruitful collaboration, which led to a joint experimental and theoretical study published in the Journal of Mass Spectrometry.

Many thanks also go to the (former) members of the Grimme group, especially Vilhjálmur Ásgeirsson (now of the University of Iceland) and Dr. Andreas Hansen, with whom I have co-written scientific publications.

At last, I thank my parents. They have supported me financially and otherwise through nearly 25 years of my life and thus enabled my success at university.

Curriculum Vitae

Christoph Bauer
Personal Data

Scientific Conferences (Selection)

09/2016	*Symposium on Theoretical Chemistry*, Bochum, Germany Poster: "Application of the Fractional Occupation Number Weighted Density Analysis"
09/2016	*Future of Chemical Physics*, Oxford, UK Poster: "On the Importance of Statistical Charge Assignment for the Computation of Electron Ionization Mass Spectra"
03/2016	*Annual meeting of the German Society for Mass Spectrometry*, Hamburg, Germany Contributed Talk: "Electron Ionization Induced Fragmentations of Organometallic Compounds from First Principles"
03/2015	*Annual meeting of the German Society for Mass Spectrometry*, Wuppertal, Germany Contributed Talk: "First Principles Calculation of Electron Ionization Mass Spectra"

Scientific Publications (Selection)

- C. A. Bauer, A. Hansen, S. Grimme, *The Fractional Occupation Number Weighted Density as a Versatile Analysis Tool for Molecules with a Complicated Electronic Structure*, Chem. Eur. J. **2017**, *23*, 6150–6164
- C. A. Bauer, S. Grimme, *How to Compute Electron Ionization Mass Spectra from First Principles*, J. Phys. Chem. A **2016**, *120*, 3755–3766
- C. A. Bauer, 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. **2015**, *21*, 125–140
- C. A. Bauer, S. Grimme, *Elucidation of Electron Ionization Induced Fragmentations of Adenine by Semiempirical and Density Functional Molecular Dynamics*, J. Phys. Chem. A **2014**, *118*, 11479–11484
- C. A. Bauer, S. Grimme, *First principles calculation of electron ionization mass spectra for selected organic drug molecules*, Org. Biomol. Chem. **2014**, *12*, 8737–8744

Scholarships

- Scholarships for academic performance of the University of Vienna, awarded separately for the consecutive academic years from 2008/09-2011/12
- *Pro Scientia* scholarship, funding period from 03/2011-03/2014

IT Skills

- OS: Windows, Linux (daily use)
- Fortran (maintenance and modifications of the qceims mass spectral prediction code)
  bash scripting, LaTeX, Gnuplot, MS/Open Office

Volunteer Work and Interests

- Member of the Collegium musicum choir committee, Bonn, Germany (2014-2017, co-organisation of concerts and rehearsals, writing of funding applications)
- Member of the Roman Catholic parish council of Kottingbrunn, Austria (2008-2014, co-organisation of parish community events)

Other Qualifications

- Military service completed (10/2007-03/2008, Austrian Army)
- Driving Licence "B" (since 08/2007)

Bonn, July 25, 2017