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

Jeroen Koopman* and Stefan Grimme*

Mulliken Center for Theoretical Chemistry, Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany

Supporting Information

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

1. INTRODUCTION

Nowadays, structure elucidation of molecules or condensed phases is one of the key ingredients in everyday work in chemistry and related sciences. Over the past decades, several excellent experimental methods, namely NMR, IR, Raman, and UV−vis spectroscopy, have been developed enhancing the facility of solving molecular structures tremendously. To date, it has become computationally affordable to use quantum chemical (QC) methods to calculate the properties that are needed to predict such spectra.1 Another extremely important analytic tool for various areas in the organic and bioorganic chemistry is electron ionization mass spectrometry (EI-MS).2,3 Its daily application, e.g., in the field of forensic drug testing4 or pharmacokinetics,5 requires continuous investigation of many new substances and their structure−spectrum relationship. In practice, compound identification is often assisted using chemoinformatic approaches6−8 or database-driven programs.9,10 The ongoing development of computer-based neural networks aims to ease this procedure.11 However, these approaches lack the basic physics and chemistry of the EI-MS process and do not have the ability to determine the basic reaction mechanisms leading to the observed spectra. While it is possible to predict previously unknown structures of molecules by these methods7 the agreement with the experiments is often low.12,13

Unfortunately, the straightforward computation of the required properties to generate accurate EI-MS is not possible with standard theoretical techniques. To tackle this problem, statistical methods based upon the quasi-equilibrium theory (QET)14 or the Rice−Ramsperger−Kassel method (RRKM)15−17 have been developed. Downsides when using these methods arise for larger molecules, leading to computationally very demanding procedures, which are difficult to generalize. To overcome this problem, our group proposed to compute mass spectra using on-the-fly computed potential energy surfaces with Born−Oppenheimer ab initio molecular dynamics (BO-AIMD).18 Based on this idea, a widely applicable protocol for predicting mass spectra termed as the quantum chemistry electron ionization mass spectrometry (QCEIMS)19,20 method has been developed. It is, to our knowledge, the first attempt to use BO-AIMD to calculate EI-MS in a “close to the experiment” manner, without relying on any database or pretabulated fragmentation rules.21 QCEIMS yields standard 70 eV EI-MS for organic and inorganic molecules that agree reasonably well with corresponding experimental data and gives an unprecedented insight into the reaction mechanisms. Different QC methods for the calculation of MS have already been tested in the past for various molecules.19,21−24 It has been shown that at least Kohn−Sham density functional theory (KS-DFT) with small basis sets or alternatively semiempirical quantum mechanical (SQM) methods like DFTB325,26 or OM2/OM327,28 have to be used to gain an acceptable accuracy-to-cost ratio. With this in mind, we have implemented the GFN1-xTB and IPEA-xTB methods into the program, and this combination has outperformed other methods for predicting EI mass spectra.
with QCEIMS.29 Very recently, we have implemented its successor GFN2-xTB30 into the QCEIMS code. The improved physics of this method should increase the quality of the calculations, while the computational demands stay low.

2. METHODOLOGY

In an EI-MS experiment, a molecule is hit by a beam of high-kinetic-energy electrons. The impact of the accelerated electrons ejects a valence electron from the targeted molecule and, in positive ionization mode, creates a radical cation, as well as two out-going electrons with continuous energy, a so-called 1e−2e process. The deposited internal energy, if averaged over many molecules, is called the internal excess energy (IEE) and follows a complicated distribution for which we make reasonable assumptions.19 Its value determines which reaction channels are eventually possible. After the electron impact, the IEE is distributed in the vibrational modes of the molecule through internal conversion (IC) and converted to the nuclear kinetic energy of the corresponding atoms. The vibration can cause bond breakage or other chemical reactions, which mostly lead to fragmentation into smaller molecules. The potential energy surface (PES) of the ions, which can be calculated, determines the most energetically favorable reaction pathways.

For a more detailed description of these processes and a discussion of other important details that have to be considered in a theoretical EI-MS experiment, e.g., the question where the charge remains after the fragmentation process, the reader is referred to the original publication19 and the textbooks.2,3

2.1. QCEIMS. In the following section, we will briefly discuss the working principle of the program. For more details, the reader is referred to the original publication.19

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

1. Equilibration and conformer sampling: An initial guess of the neutral molecular starting structure will be equilibrated in the first MD run, and a predefined number of snapshots are randomly selected and saved to obtain starting coordinates. For complicated cases, a preceding detailed conformational analysis should be conducted and the entire QCEIMS procedure is then started separately for various conformers.

2. Assignment of IEE and IC: For each snapshot geometry, the molecular orbital spectrum is calculated by a single-point calculation after which a Mulliken population analysis31 is performed. With this information, the internal excess energy (IEE) and internal conversion (IC) time are estimated and assigned to all starting geometries. The IC time is calculated by the energy-gap law.32

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

The calculations done in the program are basically first-principles and fully theoretical, i.e., not based upon any experimental results. The EI-MS process is based upon a simplified theoretical model, and the PES is approximated by quantum chemical methods. Hence, the underlying QC method has a significant impact on the quality of the simulated spectrum. Furthermore, the number of production runs and the maximum simulation time can considerably alter the results. More subtle effects, like wrong assignments of ionization potentials to the fragments or the nature of the IEE distribution, can lead to false intensities or even signals from unphysical fragmentation (artifacts). A more detailed discussion of these influences can be found in the original publication.19

When computed spectra of this unbiased approach used by QCEIMS are compared to those from rule-based, chemoinformatic programs (see, e.g., refs 6−8, 11), it is to be kept in mind that this “black-box” method may give rise to results with lower accuracy. However, QCEIMS can predict EI spectra of unknown chemical compounds and is able to retrace the composition of the fragments created during the process from the MD trajectory. For this reason, the program allows the discussion of the computed spectra in a detailed way, from fragmentation patterns to recombinations and rearrangements.

Figure 1. Flowchart of the QCEIMS protocol.
occurring in the experiment. At no point, intermediate structures have to be guessed or altered.

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

In GFN2-xTB, the dispersion interactions are treated by means of a self-consistent variant of the D4 dispersion model instead of D3(Bj), used in GFN-xTB, and furthermore, the description of electrostatic interactions has been greatly improved. For a more detailed discussion of the differences between both methods, please refer to ref 30.

The originally proposed method is here on called GFN1-xTB for a better distinction between both schemes. The computed ionization potentials with GFN1-xTB are not sufficiently accurate. To remedy this, a special-purpose IPEA-xTB method has been developed. It is a reparametrization of GFN1-xTB and uses additional ($n + 1)s$ basis functions to better represent the electron affinities. Unfortunately, the spectra of some molecules containing transition metals were not described well with the IPEA-xTB method, which partially could be traced back to wrong charge assignments. The errors made on these systems can be fixed by calculating the ionization potentials at the hybrid DFT (PBE0/def2-SV(P)) level, for which the computational time is proportionally large in comparison to semiempirical methods. For GFN2-xTB, we did not find the need for a reparametrization because the IPs are being calculated qualitatively correctly with this method and are sufficiently accurate for the charge assignment of the fragments; thus, it is likely that use of this new method will achieve the proper results for these systems with less computational effort. To validate this statement and to test the performance of this method, we have tested combinations of GFN1- and GFN2-xTB using IPEA- and GFN2-xTB for IP calculations. To present the performance of the two methods more clearly, semiempirical OM3-D3

Figure 2. The benchmark set. Molecules (1)–(6) represent the organic group, molecules (7)–(9) represent the transition-metal group, and molecules (10)–(15) represent the main-group inorganic group.
calculations were performed for organic molecules. AM1 and PM3 calculations were omitted because of the bad performance of these methods with QCEIMS (see ref 19). DFT was used to cross-check molecules involving transition metal atoms. The combinations are noted as Method/IP-Method.

2.3. Technical Details. The calculations in this work were executed on Intel XEON E5-2660 2.00 GHz cores. Computations were performed using QCEIMS version 3.8. For OM3-D3 calculations, MNDQ2005 version 7.0 was used and DFT calculations were gained using the ORCA 4.0.1.2. suite of programs. Statistically converged results are obtained for 1000 production trajectories that were carried out for each molecule and method with a maximum MD simulation time of 10 ps. Each production run required about 10,000–20,000 QC calls. We did not alter any settings in QCEIMS, nor did we modify the tight-binding methods for this work. The results presented here therefore do not show the full capability of QCEIMS, which may be improved by choosing different simulation conditions in the program for different compound classes.

3. RESULTS AND DISCUSSION

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

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

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

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

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

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

For transition-metal-containing molecules (7–9), calculations of a total spectrum using different GFN1-xTB combinations take on average 200 h to complete, while the

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

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

The time per calculation with GFN1/IPEA-xTB for the main-group inorganic molecule group (10–15) averages for all four combinations of methods between 400 and 600 h, where use of GFN2-xTB for IP evaluation takes longest. This excludes the outlier dichloro-(ethyl)aluminum 13, for which the timing is in the range of the small organic molecules (e.g., 1-butanol (1)). For the molecule 15, timings with the GFN2-xTB combinations are lower than expected, but this is due to technical failures in the calculations, as discussed in the corresponding results Section 3.4.

3.3. Stability. The stability of the GFN1- and GFN2-xTB methods is evaluated by the number of successful production runs (see Figure 4). For both methods, the majority of runs complete properly, leading to an excellent average failure rate of less than 1%. For the GFN1/IPEA-xTB calculations, the largest failure rates are produced for hexafluorobenzene (2), bis(benzene)chromium (7), and octasulfur (15). For the GFN2-xTB method, failure rates are comparable, except for the transition metal involving molecule bis(benzene)-chromium (7), having a failure rate of 11.2%, and octasulfur (15) (>15%). The effect on the calculated spectrum is discussed in the corresponding results, Section 3.4.
In conclusion, low failure rates and acceptable timings indicate good applicability of the new GFN2-xTB method. This could not be expected because GFN2-xTB is inherently more involved mainly due to the additional multipole electrostatic treatment. Thus, GFN2/IPEA-xTB and GFN2/GFN2-xTB both can be used in QCEIMS as an alternative to the GFN1/IPEA-xTB method without significant restrictions in the computational demands or robustness.

Figure 5. Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the organic compounds (a) 1-butanol, (b) hexafluorobenzene, and (c) uracil. The structures of the precursor ion (denoted M⁺) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.
3.4. Comparison of Experimental and Theoretical Spectra. In the following section, we present the calculated QCEIMS spectra using GFN1/IPEA-xTB and GFN2/GFN2-xTB. The computed spectra are directly compared with their corresponding experimental EI-MS obtained from the NIST44 or SDBS45 databases, and the agreement between theory and experiment is determined by a composite matching score10,24 with a range of values between 0 (no match) and 1000 (perfect match).
match). Matching scores of organic molecules obtained with OM3-D3 are being listed for the purpose of validation of the quality of spectra gained using the GFN methods. The main differences between the results of the two GFN methods are being discussed, with a focus on the presence of important \( m/z \) signals as well as corresponding variation in signal intensity. We discuss determinative peak-series and point out interesting or important structures of the obtained fragments, which are shown explicitly as insets in the spectra. Major differences in signals or intensities between GFN2/IPEA- and GFN2/GFN2-xTB calculations were not observed or were of minor influence for the resulting spectra and are therefore not discussed further.

3.4.1. Organic Group. Small organic molecules have already been investigated in former studies\(^ {19,21,29} \) and are only briefly considered here.

For 1-butanol (Figure 5a), the main fragmentation pathways result from the loss of alkyl groups. These are reproduced well by both methods with a satisfying agreement of the simulation with the experiment. However, the survival rate of the precursor ion is too high in the simulations, meaning it does not decompose accordingly under given simulation conditions. This effect is due to the IEE distribution, which is of a Poisson-type variant and not obtained ab initio and thus can lead to a bad description of unusual electronic situations. This can partially be alleviated by applying higher IEE values and/or longer simulation times. These and other effects of various simulation conditions are discussed in the original publication.\(^ {19} \) Differences between the methods are found in the intensities of the signals. For GFN2-xTB, especially the peaks of \( \text{C}_4\text{H}_9^+ \) at \( m/z \) 39 and of \( \text{C}_6\text{H}_5^+ \) at \( m/z \) 56 are in better agreement with the experiment. Through the high survival rate of the precursor ion, the matching scores between the experiment and calculations are 225 for GFN1-xTB and 223 for GFN2-xTB. OM3-D3 calculations account better for the precursor ion signal, which leads to a matching score of 530.

Hexafluorobenzene (Figure 5b) represents an interesting case. The fragment pattern is dominated by the ring breakage products. In the experiment, the dominant path forms \( \text{C}_3\text{F}_3^+ \) (\( m/z \) 117) with the remainder being \( \text{CF}_2 \) (\( m/z \) 50) and a single fluorine atom (\( m/z \) 19). The loss of a single (neutral) fluorine atom from the precursor ion can be observed by the appearance of the signal at \( m/z \) 167. Furthermore, the breakage of the precursor into two units of \( \text{C}_2\text{F}_3^+ \) yields the signal at \( m/z \) 93. Both GFN methods provide similarly good results in this respect. However, the methods differ in the calculated intensities, as GFN2-xTB overestimates some of the signals, especially for the fragments \( \text{C}_6\text{F}_4^+ \) at \( m/z \) 148, \( \text{C}_4\text{F}_2^+ \) at \( m/z \) 110, and \( \text{C}_5\text{F}_3^+ \) at \( m/z \) 86. Matching scores are 734 for GFN1-xTB, 647 for GFN2-xTB, and 687 for OM3-D3.

An earlier QCEIMS work has been conducted for four different nucleobases.\(^ {22} \) These were calculated at the semi-empirical levels OM2-D3\(^ {25,26} \) and DFTB3-D3.\(^ {25,26} \) As an example molecule from this series, we present uracil (Figure 5c), computed with the two GFN methods. We find that the spectra produced with both GFN methods are in good agreement with the experiment, although some intensities of the calculated signals are either over- or underestimated when directly compared to the measured signals. Compared to the results obtained in the earlier work, the spectra calculated by the GFN methods are of better quality as the spectra are produced using DFTB3-D3 or OM2-D3. The fragmentation proceeds via the bond breakage of the precursor ion into units of \( \text{HCNO}^+ \) (\( m/z \) 43) and \( \text{HNC}_2\text{H}_4\text{O}^+ \) (\( m/z \) 69) and the subsequent dissociation into the fragments \( \text{HNCH}^+ \) and \( \text{CO}^+ \) at \( m/z \) 28. Between the calculated spectra of the two GFN methods, only minor differences in the intensities are observed. Matching scores are 780 for GFN1-xTB, 745 for GFN2-xTB, and 711 for OM3-D3.

The experimental spectrum of testosterone (Figure 6a) contains a large number of signals, which are overall reproduced well by the two GFN methods. Especially, the peak series of the lower and medium mass fragments (\( m/z \) 30–150) are replicated very accurately by both tight-binding methods when compared with the experiment. However, neither of the two computed spectra correctly reproduce the intense signals displayed in the experimental spectrum at \( m/z \) values 124, 203, and 246. Furthermore, the GFN-based calculations favor the formation of the fragment \( \text{C}_6\text{H}_4^+ \) at \( m/z \) 53 over the experimental found fragment \( \text{C}_6\text{H}_5^+ \) at \( m/z \) 55. Comparing the results produced by the two GFN methods with each other, the spectrum simulated with GFN1-xTB overestimates the intensities of the lower mass signals in the area of about \( m/z \) 30–90, while the spectrum calculated with the GFN2-xTB method is in overall better agreement with the experimental results and also accounts better for signals in the area between \( m/z \) 145 and 200. Matching scores are 532 for GFN1-xTB, 545 for GFN2-xTB, and 592 for OM3-D3.

The fragmentation scheme of sucrose (see Figure 6b) produces a considerable amount of highly intense signals in the experimental spectrum, notably in the area between \( m/z \) values 28 and 73. Large discrepancies between simulations and experiment can be observed in the number of hydrogen atoms bound to the fragments. Especially, the peaks from the experiment at \( m/z \) 28, belonging to \( \text{CO}^+ \), compared to the calculated signal at \( m/z \) 29 of \( \text{HCO}^+ \), as well as the experimental signal of \( \text{HC}_2\text{O}^+ \) at \( m/z \) 57 in contrast to the simulated peak of \( \text{H}_2\text{C}_2\text{O}^+ \) at \( m/z \) 59, are typical examples for this divergence. Furthermore, various signals that can be noticed in the experimental spectrum, e.g., \( m/z \) values 97, 221, and 293, and some of the less intense peaks in between, are not well recreated by the calculations using the GFN methods. A comparison of the simulated spectra of the two tight-binding methods with each other reveals no significant differences. Matching scores are 201 for GFN1-xTB, 217 for GFN2-xTB, and 169 for OM3-D3.

Leucylglycylglycine (see Figure 6c) is composed of one l-leucine and two glycine residues. In the EI-MS experiment, the main decomposition pathway leads to the fragments \( \text{H}_2\text{C}_3\text{N}_2\text{O}^+ \) (\( m/z \) 86), \( \text{H}_2\text{C}_2\text{O}^+ \) (\( m/z \) 59), \( \text{H}_2\text{C}_4^+ \) (\( m/z \) 57), and \( \text{HCNO}^+ \) (\( m/z \) 43). In a second step, \( \text{H}_2\text{C}_3\text{N}_2\text{O}^+ \) dissolves into \( \text{H}_2\text{C}_2\text{NO}^+ \) (\( m/z \) 56) and \( \text{H}_2\text{CN}^+ \) (\( m/z \) 30). The simulated spectra produced using either GFN1-xTB or GFN2-xTB account for the majority of these signals in accordance with the experiment. However, in the simulations, the survival rate of the precursor ion is too high so that less intense signals measured in the experiment do not appear in the calculated spectra. When comparing the GFN methods with each other, some of the intensities of various signals differ, e.g., peaks at \( m/z \) values 59, 86, and 159. The latter signal belongs to an intermediate product that gains stability through a ring formation of the peptide. The signal of \( \text{H}_2\text{O}^+ \) at \( m/z \) 18 in the experimental spectrum is probably measured due to the presence of water in the ionization chamber during the experiment and is therefore not produced by the simulations.
Matching scores are 192 for GFN1-xTB, 224 for GFN2-xTB, and 123 for OM3-D3.

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

For bis(benzene)chromium (Figure 8a), both GFN simulated spectra are in satisfying agreement with the experimental spectrum. A detailed comparison between the calculated and the experimental spectra reveals some divergences in the intensities, e.g., the simulated signal at m/z 77 and the experimental peak at m/z 78, resulting from a discrepancy between the simulated and the measured number of hydrogen atoms bound to fragmented benzene rings. Failure rates by both methods of the GFN-xTB family were high (see Section 3.3), especially for calculations with GFN2-xTB. This is most likely due to the wrong description of the Cr⁺ fragment, for which the corresponding signal at m/z 52 is almost completely missing in the calculated spectrum of this method. Instead, the protonated form HCr⁺ is preferred and completely missing in the calculated spectrum of this method. The IPEA-xTB method yields wrong IP values and therefore leads to false signals in the calculated EI-MS spectrum. This is visualized by the red circle in Figure 7, where the ionized iron cation is not being charged in the simulated spectrum (red, inverted) and thus the experimentally found signal of Fe⁺ is missing. Calculating the IPs with DFT or GFN2-xTB instead solves this problem. It is therefore recommended to use the GFN1/GFN2-xTB, GFN2/GFN2-xTB or the GFN1/DFT, GFN2/DFT combination for transition-metal-containing molecules. The results presented in this section for transition-metal-containing molecules were calculated with the GFN1/GFN2-xTB and GFN2/GFN2-xTB combinations.

For bis(benzene)chromium (Figure 8a), both GFN simulated spectra are in satisfying agreement with the experimental spectrum. A detailed comparison between the calculated and the experimental spectra reveals some divergences in the intensities, e.g., the simulated signal at m/z 77 and the experimental peak at m/z 78, resulting from a discrepancy between the simulated and the measured number of hydrogen atoms bound to fragmented benzene rings. Failure rates by both methods of the GFN-xTB family were high (see Section 3.3), especially for calculations with GFN2-xTB. This is most likely due to the wrong description of the Cr⁺ fragment, for which the corresponding signal at m/z 52 is almost completely missing in the calculated spectrum of this method. Instead, the protonated form HCr⁺ is preferred and simultaneously the simulation fails to account for signals with low mass-to-charge values between 20 and 40. Both calculations conducted with the GFN methods account for the signal at m/z 104, which is created by a fragment in which a H₂Cr molecule is bound to a C₄H₂ chain. Matching scores are 677 for GFN1-xTB and 688 for GFN2-xTB.

The Kaminsky catalyst can contain the group 4 metal components Ti, Zr, or Hf. We have decided to test the two GFN methods for EI-MS calculations of zirconocene dichloride (Figure 8b). The experimental spectrum reveals a variety of fragmentation pathways, from single chloride loss of the precursor ion (resulting in a signal at m/z 256) to the cyclopentadienyl dissociation (signals at m/z values 162 and 227). The EI-MS created using the GFN1-xTB method is in very good agreement with the experimental one. In contrast, the spectrum created using the GFN2-xTB method shows an overestimation of almost all intensities of the produced signals in comparison with those from the experiment. Furthermore, the new approach creates various artifacts that lead to a significant increase in the computational demands for the calculation of this system (see Section 3.2). However, the signals computed by GFN2-xTB around m/z 201 are in surprisingly good accordance with the experiment, which corresponds to the subsequent loss of C₂H₂ from the cyclopentadienyl fragment (m/z 227). These signals are not simulated well by GFN1-xTB. Matching scores are 671 for GFN1-xTB and 680 for GFN2-xTB.

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

3.4.3. Main-Group Inorganic Group. Bis(pinacolato)-diboron (Figure 9a) is an interesting compound related to Suzuki coupling reactions. In the EI-MS simulations, the fragmentation patterns can be retraced convincingly, as indicated by the good agreement between experimental and calculated spectra. The main fragmentation pathway is
characterized by the cleavage of CH$_3^+$ and C$_3$H$_6^+$ from the alkane groups of the molecule. These fragments themselves emerge in the lower mass area of the spectrum at m/z values 15 and 42 with various constellations in the number of hydrogen atoms bound to these fragments. This fragmentation is followed by various rearrangement reactions: In the first figure.

Figure 8. Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the transition metal compounds (a) bis(benzene)chromium, (b) zirconocene dichloride, and (c) nickel(II)bis(diphenyl-acetylacetonate). The structures of the precursor ion (denoted M$^+$) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.
step, one of the C₆H₁₂ (m/z 84) side chains splits off from the molecule, which leads to a rearrangement of the remaining fragment, creating a (O₂⁻)B-O-B(-O) chain (m/z 169). In the second rearrangement step, the remote O-B-O unit substitutes an oxygen atom at one of the side chains, which leads to the decomposition of the structure into two units of O-B-O-(CH₃)₂⁺, displayed by the signal at m/z 84. The detailed description of this process shows the outstanding capabilities.

Figure 9. Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the main-group inorganic compounds (a) bis(pinacolato)diboron, (b) chinalphos, and (c) triphenylstibine. The structures of the precursor ion (denoted M⁺) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their m/z values and discussed in the text.
of QCEIMS to analyze rearrangement procedures during EI-MS experiments, demonstrating its usefulness in elucidating structures and complex reaction mechanisms. Overall, we find the agreement between calculated spectra using the GFN methods and the experimentally measured spectrum to be very satisfying, as the two GFN methods account for the majority of

Figure 10. Comparison of the EI-MS computed by GFN1-xTB (left) and GFN2-xTB2 (right) to the experimental references (red, inverted) of the organic compounds (a) dichloro(ethyl)aluminum, (b) 2-(dimethyl(naphthalen-1-yl)silyl)phenyl)methanol, and (c) octasulfur. The structures of the precursor ion (denoted $M^+$) and selected signals/fragments have been superimposed on each spectrum. Important or interesting signals are highlighted by their $m/z$ values and discussed in the text.
signals. GFN2-xtb seems to perform better here since the agreement with the intensities measured in the experiment is slightly better than in the spectrum created with GFN1-xtb. Both methods overestimate the signals at m/z values 41 and 42, as well as the survival rate of the precursor ion. Matching scores are 245 for GFN1-xtb and 250 for GFN2-xtb.

The simulated MS of the insecticide chinalphos (Figure 9b) matches the experimental spectrum badly. The computed signal at m/z 145 is generated by the fragmentation of the precursor ion into a SP(OCC2H5)2+ fragment, which differs from the experimental signal at m/z 146 due to a hydrogen atom bonded less to the fragment. The signal at m/z 157, which is almost totally missing in the simulation, is generated due to the rearrangement of a CH3 fragment (which dissociates from the SP(OCC2H5)2+ group) between the benzene ring and one of the nitrogen atoms. This rearrangement is rarely reproduced in the simulation since the required migration of the methane group can progress into all spatial directions and is not often aligned at the ring system. In contrast to its predecessor, the simulation with the GFN2-xtb method displays a higher probability that in a first step the sulfur atom breaks from the precursor ion, which resolves in an intense fragment signal at m/z 266. This is followed by the decomposition of this fragment into the PS(OCC2H5)2+ fragment, generating the signal at m/z 121. Matching scores are 598 for GFN1-xtb and 510 for GFN2-xtb.

In the spectrum of triphenylstibine (Figure 9c), the dominant fragmentation pattern corresponds to the cleavage of the phenyl groups and their subsequent dissociation, which has already been reported in an earlier publication using GFN1-xtb for structure elucidation.29 We find that the simulated spectra using both GFN methods are in good agreement with the experimental data, but simulations using the GFN1-xtb method yield small artifacts throughout the spectrum. In contrast, the spectrum created using GFN2-xtb does not display any artifacts, and the method reduces the survival rate of the precursor ion to match the experiment almost perfectly. Furthermore, use of GFN2-xtb improves the intensities of the signals at m/z values 198 and 154 when compared to those from the experiment, where the latter signal describes the bond formation between two phenyl groups. The agreement between the simulated and the experimental signal of HSb+ at m/z 123 is better when using the GFN2-xtb method instead of GFN1-xtb. Matching scores are 642 for GFN1-xtb and 664 for GFN2-xtb.

The spectrum of dichloro(ethyl)aluminum (Figure 10a) indicates competing fragmentation reactions. The signals at m/z values 29 and 97 describe the loss of the ethane group, forming Cl2Al+ and C2H5+, while the signals at m/z values 36 and 91 are generated by the loss of the chlorine atoms. The simulated spectrum of the GFN2-xtb method nearly matches the experimental spectrum in better agreement than its predecessor GFN1-xtb. The substitution of the carbon—aluminum bond by a single chlorine is reproduced according to the experiment, generating the signal at m/z 64 in the GFN2-xtb-created spectrum. Matching scores are 514 for GFN1-xtb and 507 for GFN2-xtb.

In Hiyama cross-coupling reactions, so-called HOMSi reagents46 are used for carbon—carbon coupling reactions. As an example for this group of reagents, we have chosen to compare the simulated spectra created with GFN1- and GFN2-xtb to the experimental spectrum of (2-(dimethyl-(naphthalen-1-yl)silyl)phenyl)methanol (Figure 10b). The EI-MS spectrum measured in the experiment contains highly intense peaks in the larger mass region between m/z values 215 and 277 and less intense signals for small to medium mass regions with m/z values 20–200. The calculated spectra of the two GFN methods show a good agreement to the experiment in the lower and medium mass regions, but the main peaks at m/z values 215 and 259 and the survival rate of the precursor ion are displayed poorly with both methods compared to experiment. In the simulation, first, a rearrangement of the hydroxide molecule from the benzyl alcohol group (C7H7-OH) to the silicon atom takes place. The hydroxide molecule substitutes one of the ethane groups bound to the silicon atom, and the resulting structure creates a signal at m/z 277. From here on, the simulated fragmentation patterns do not seem to correspond to the bond breakage scheme of the experiment. While in simulation one methane group is cleaved off (forming the signal at m/z 261), in the experiment, a H2 molecule is additionally dissociated during this fragmentation process and the resulting fragment creates the signal at m/z 259. Subsequently, this fragment either decomposes into HOSi+, giving rise to the signal at m/z 45 and which has a dominating intensity in the simulated spectrum using the GFN1-xtb method, or, in contrast, forms the CH3Si+ fragment displayed at m/z 43, which is favored in the spectrum computed with the GFN2-xtb method. Both of these fragment signals are displayed in the experimental spectrum. Further differences between the simulated spectra of the GFN methods are visible in the overall signal intensity, as the lesser intense peaks created by the GFN2-xtb method produce a “cleaner” spectrum. However, some of the experimentally found signals are under-represented in the spectrum at this level. Matching scores are 315 for GFN1-xtb and 282 for GFN2-xtb.

The simulated spectra of octasulfur (Figure 10c) strongly depend on the method used. Calculations with the GFN2-xtb method result in a spectrum in which nearly all signals are missing. The majority of production runs show the immediate bond breakage of the sulfur ring into S2+ and S3+ fragments. Hereby, the intermediate products at m/z values 128 and 192 are strongly under-represented and the signals at m/z values 32, 160, and 224 do not appear. On the contrary, the spectrum simulated with GFN1-xtb is in very good agreement with the experimental results and reproduces all signals and their intensities satisfactorily, except for those signals corresponding to S4+ at m/z 224 and S5+ at m/z 32. Matching scores are 821 for GFN1-xtb and 213 for GFN2-xtb.

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

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

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

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

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

Differences between the GFN1-xTB and GFN2-xTB methods are mostly observed for the calculated intensities that sensitively depend on details of the computed PES. The new GFN2-xTB method produces fewer artifacts in the spectra but eventually misses important signals. Nevertheless, the results from both methods compare reasonably well to the experimental results. It is to be kept in mind that deviations like intensities, missing signals, and survival rates of the precursor ion depend on technical settings related to internal energy distribution and the ionization/heating procedure. They can be individually improved by changing the default settings, thereby significantly improving the simulated spectra.

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

The expansion of the QCEIMS program to involve collision-induced dissociation (CID) mass spectrometry techniques is currently being developed in our laboratory.
data and comparison to competitive fragmentation modeling. Int. J. Quantum Chem. 2018, 118, No. e25460.

(13) Hites, R. A.; Jobst, K. J. Response to “Letter to the Editor: Optimism for Nontarget Analysis in Environmental Chemistry”. Environ. Sci. Technol. 2019, 53, 5531−5533.

(14) Rosenstock, H.; Wallenstein, M.; Wahrhaftig, A.; Eyring, H. Absolute Rate Theory for Isolated Systems and the Mass Spectra of Polyatomic Molecules. Proc. Natl. Acad. Sci. USA 1952, 38, 667−678.

(15) Kassel, L. S. Studies in Homogeneous Gas Reactions. II. Introduction of Quantum Theory. J. Phys. Chem. A 1927, 32, 1065−1079.

(16) Marcus, R. A.; Rice, O. K. The Kinetics of the Recombination of Methyl Radicals and Iodine Atoms. J. Phys. Chem. A 1951, 55, 894−908.

(17) Marcus, R. A. Unimolecular Dissociations and Free Radical Recombination Reactions. J. Phys. Chem. A 1952, 20, 359−364.

(18) Marx, D.; Hutter, J. Ab Initio Molecular Dynamics; Cambridge University Press: Cambridge, 2009.

(19) Grimme, S. Towards First Principles Calculation of Electron Impact Mass Spectra of Molecules. Angew. Chem., Int. Ed. 2013, 52, 6306−6312.

(20) To obtain QCEIMS, please contact qceims@thch.uni-bonn.de. Further information can be found at the homepage of the Mulliken Center for Theoretical Chemistry: http://www.thch.uni-bonn.de/tc/.

(21) Bauer, C. A.; Grimme, S. How to Compute Electron Ionization Mass Spectra from First Principles. J. Phys. Chem. A 2016, 120, 3755−3766.

(22) Bauer, C. A.; Grimme, S. 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.

(23) Bauer, C. A.; Grimme, S. First principles calculation of electron ionization mass spectra for selected organic drug molecules. Org. Biomol. Chem. 2014, 12, 8737−8744.

(24) Bauer, C. A.; Grimme, S. Elucidation of Electron Ionization Induced Fragmentations of Adenine by Semiempirical and Density Functional Molecular Dynamics. J. Phys. Chem. A 2014, 118, 11479−11484.

(25) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. Phys. Rev. B 1998, 58, 7260−7268.

(26) Gaul, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent Charge Density-Functional Tight-Binding Method (SCC-DFTB). J. Chem. Theory Comput. 2011, 7, 931−948.

(27) Weber, W.; Thiel, W. Orthogonalization corrections for semiempirical methods. Theor. Chem. Acc. 2000, 103, 495−506.

(28) Dral, P. O.; Wu, X.; Spörkel, L.; Koslowski, A.; Weber, W.; Steiger, R.; Scholten, M.; Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Theory, Implementation, and Parameters. J. Chem. Theory Comput. 2016, 12, 1082−1096.

(29) Ågerfors, V.; Bauer, C. A.; Grimme, S. Quantum chemical calculation of electron ionization mass spectra for general organic and inorganic molecules. Chem. Sci. 2017, 8, 4879−4895.

(30) Bannwarth, C.; Ehler, S.; Grimme, S. GFN2-xTB: An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. J. Chem. Theory Comput. 2019, 15, 1652−1671.

(31) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. J. Chem. Phys. 1955, 23, 1833−1840.

(32) Turro, N. Modern Molecular Photochemistry; The Benjamin/Cummings Publishing Company Inc.: Menlo Park, California, 1978.

(33) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1−86). J. Chem. Theory Comput. 2017, 13, 1989−2009.