

# Organic & Biomolecular Chemistry

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

# First Principles Calculation of Electron Ionization Mass Spectra for Selected Organic Drug Molecules<sup>†</sup>

Christoph Alexander Bauer,<sup>a</sup> and Stefan Grimme<sup>\*a</sup>

Received Xth XXXXXXXXXXXX 20XX, Accepted Xth XXXXXXXXXXXX 20XX

First published on the web Xth XXXXXXXXXXXX 200X

DOI: 10.1039/b000000x

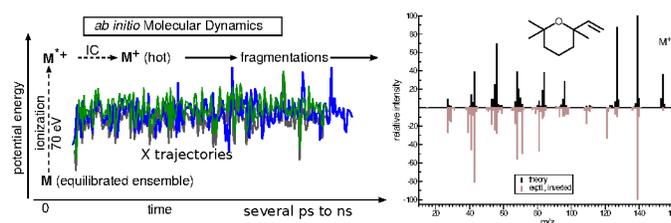
This study presents a showcase for the novel Quantum Chemistry Electron Ionization Mass Spectrometry (QCEIMS) method on five FDA-approved drugs. The method allows a first-principles electronic structure-based prediction of EI mass spectra in principle for any molecule. The systems in this case study 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.

## 1 Introduction

Modern quantum chemistry (QC) methods have made it possible to routinely compute and predict spectral properties of reasonably sized chemical compounds.<sup>1</sup> 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.<sup>2,3</sup> 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)<sup>4</sup> or Rice-Ramsperger-Kassel-Marcus (RRKM) theory.<sup>5,6</sup> However, even the most sophisticated attempts within these frameworks<sup>7</sup> have found no application on a regular basis.

The Quantum Chemistry Electron Ionization Mass Spectrometry (QCEIMS) method<sup>8</sup> 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)<sup>9</sup> 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 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.<sup>8</sup>



**Fig. 1** Schematic description of the QCEIMS procedure together with an illustrative result for a small organic molecule.

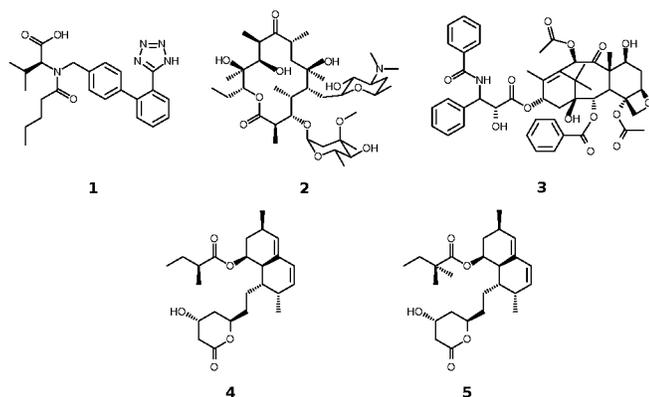
In the present study, we have taken our novel method to its current limits in order to assess whether it is workable for medicinally and pharmaceutically relevant organic compounds thereby becoming a new routine tool in theoretical chemistry. To this end five organic drugs (for structures,

<sup>†</sup> Electronic Supplementary Information (ESI) available: IUPAC names of compounds **1-5**. Optimized geometries of **1-5** and mean geometries of fragments **1a-1c** as .xyz files and a more detailed description of the simulation protocol as well as more computed results and computational wall times. See DOI: 10.1039/b000000x/

<sup>a</sup>Mulliken Center for Theoretical Chemistry, Institut für Physikalische und Theoretische Chemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Beringstr. 4, D-53115 Bonn, Germany. Fax: +49 (0)228/73-9064; Tel: +49 (0)228/73-2351; E-mail: grimme@thch.uni-bonn.de

see Figure 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<sub>24</sub>H<sub>29</sub>N<sub>5</sub>O<sub>3</sub>, nominal mass 435 u, **1**) is an angiotensin II receptor antagonist indicated e.g., against hypertension.<sup>10</sup> It is considered as the 1-*H*-tetrazol tautomer as found in the mass spectral database.<sup>11</sup> Erythromycin (C<sub>37</sub>H<sub>67</sub>NO<sub>13</sub>, nominal mass 733 u, **2**)<sup>12</sup> is a macrolide antibiotic<sup>13</sup> with a 14-membered macro-cycle and the two sugar moieties cladinose and desosamine (an amino sugar). Taxol (C<sub>47</sub>H<sub>51</sub>NO<sub>14</sub>, nominal mass 853 u, **3**) is an anticancer agent, which can be isolated from the pacific yew tree.<sup>14</sup> 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.<sup>15</sup> Lovastatin (C<sub>24</sub>H<sub>36</sub>O<sub>5</sub>, nominal mass 404 u, **4**) and simvastatin (C<sub>25</sub>H<sub>38</sub>O<sub>5</sub>, nominal mass 418 u, **5**) belong to a highly profitable class of cholesterol-lowering drugs, the statins.<sup>16</sup> Note that not all statins are as structurally similar to each other as **4** and **5**.



**Fig. 2** Formulas of the drug molecules chosen as examples for QCEIMS.

As can be seen from the structures in Figure 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.<sup>17</sup> 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 were picked more or less randomly from

intensive 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 Ref. 8 and hence not repeated here.

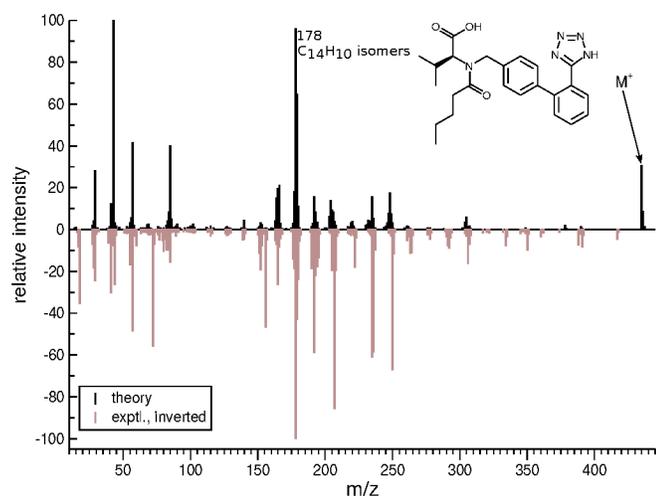
## 2 Results and discussion

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 decompose. 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 2.5.

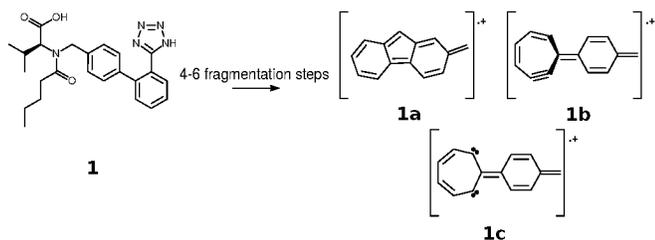
### 2.1 Valsartan

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 3 reveals surprisingly good results.



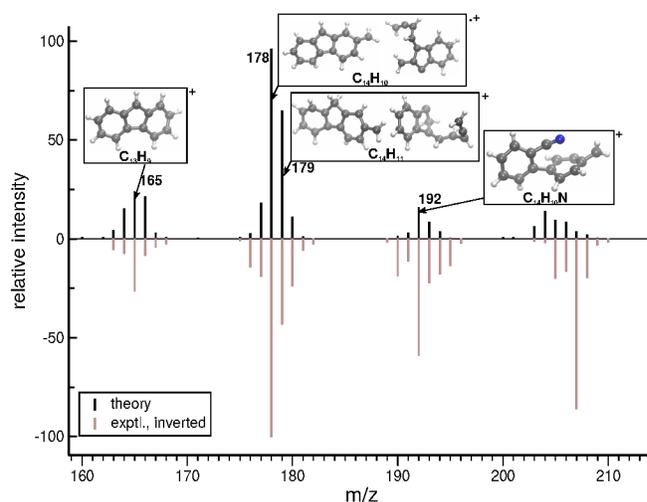
**Fig. 3** Calculated mass spectrum of valsartan in comparison with the experimental spectrum. The indicated  $m/z$  value is discussed in the text among others.

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 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.



**Fig. 4** Examples of calculated  $C_{14}H_{10}^+$  ( $m/z$  178) Valsartan fragment isomers.

Additionally, several other peaks of valsartan were assigned based on an analysis of the fragmentation trajectories. Figure 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 in Figure 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.



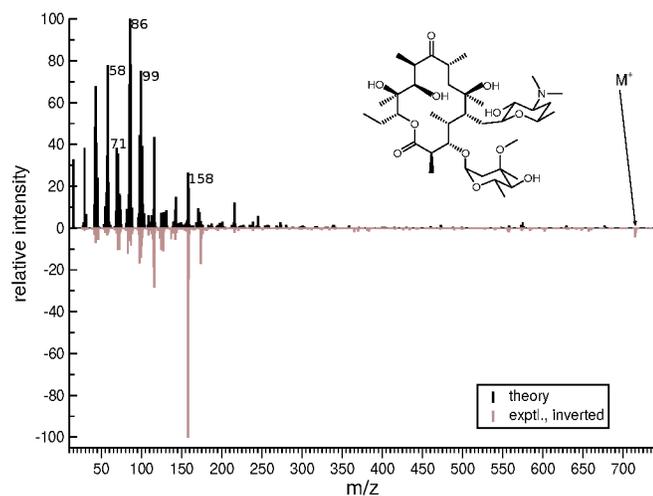
**Fig. 5** Comparison of experimental and computed spectrum of valsartan in the  $m/z$  range 160-210. Additionally, selected average fragment structures are displayed as given by the QCEIMS computations, and assigned to peaks in the mass spectrum of valsartan.

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 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.<sup>18</sup> Note that QCEIMS automatically derives from a very reasonable

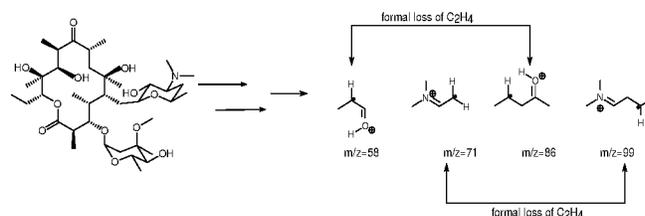
QC computation and the actual effective temperature the distribution of charge between fragments (assuming Boltzmann statistics) and based upon that decides which fragmentation cascade to follow. Undoubtedly, there are some deficiencies as well: false positives and negatives and wrong peak intensities are all clearly visible in Figure 3. Before jumping too hastily to conclusions we first examine the other simulated spectra in order to get a more comprehensive view.

## 2.2 Erythromycin

Figure 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 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 7, which resulted from different individual fragmentation runs.<sup>19</sup> The main peak of the experimental record,  $m/z$  158, most likely results from the desaminosyl unit ( $C_8H_{16}NO_2^+$ ) of erythromycin.



**Fig. 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.



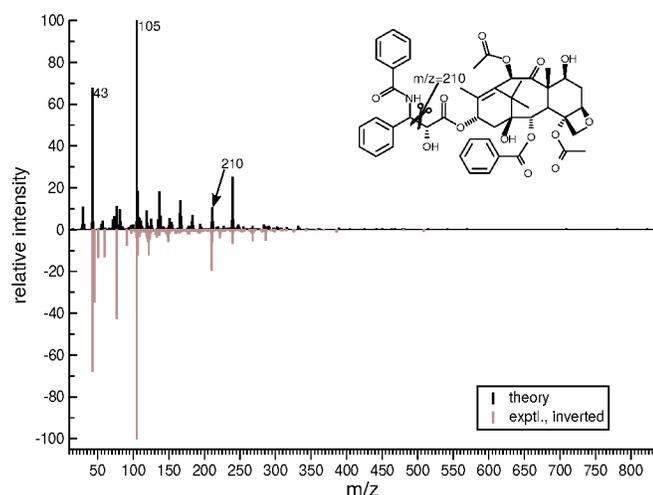
**Fig. 7** Calculated fragments of erythromycin yielding a series of peaks.

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 7. Note that we have not made any attempts for improving or fine-tuning of the calculations for the individual 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.

## 2.3 Taxol

Figure 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 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 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.

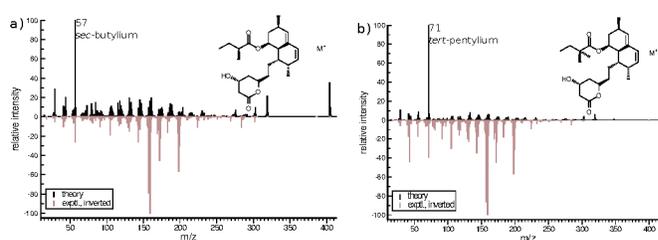


**Fig. 8** Calculated mass spectrum of taxol in comparison with the experimental spectrum. Indicated  $m/z$  values are discussed in the text. Note that the molecular ion ( $m/z$  853) is missing in both spectra.

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 2.5.

## 2.4 Statins

Figure 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.



**Fig. 9** Calculated mass spectra of a) lovastatin and b) simvastatin in comparison with the experimental spectra. Marked peaks are discussed in the text.

This is mainly related to an excess of *sec*-butylum ( $m/z$  57,  $C_4H_9^+$ ) and *tert*-pentylum ( $m/z$  71,  $C_5H_{11}^+$ ) signals for **4** and **5**, respectively. These groups are in  $\alpha$  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*-pentylum cations as depicted in Figure 9b. A similar problem of exaggerated alkyl loss has also been discussed in the original paper.<sup>8</sup> 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.

## 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 priori 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 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 be 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 make some changes compared to the original ansatz to distribute the IEE as discussed in the ESI. 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,<sup>20,21</sup> 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.

### 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.

### 4 Computational Details

The neutral ground state structures of the molecules **1-5** were optimized using dispersion-corrected DFT at the TPSS<sup>22</sup>-D3<sup>23-25</sup>/def2-TZVP<sup>26</sup> level as implemented in Turbomole 6.5.<sup>27</sup> The nature of the stationary point on the PES was confirmed to be a (local) minimum by calculating the harmonic vibrational frequencies.<sup>28</sup>

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 -  $\epsilon_{\text{HOMO}}$ , where the orbital energy  $\epsilon_{\text{HOMO}}$  was computed at the PBE12<sup>29,30</sup>/SVx<sup>31,32</sup>//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<sup>33</sup> with the D3 dispersion correction.<sup>23,24</sup> The initial temperature for each trajectory was set at 500 K, which is the default parameter (and sufficiently close to 250°C  $\hat{=}$  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 spec-

tra 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.<sup>34–36</sup> 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 the ESI. The statistical fragment charge assignment algorithm used the Boltzmann factor for ionization potentials (IPs),  $e^{-\frac{IP}{kT}}$ , 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.

For OM2 calculations, the MNDO program<sup>37</sup> was called and the DFT calculations were carried out by the ORCA suite of programs.<sup>38</sup>

Experimental spectra for comparison were downloaded from mass spectral databases available on-line.<sup>11,39,40</sup>

## Acknowledgements

This work has been supported by DFG grant no. 1927/10-1, "First Principles Calculation of Electron Impact Mass Spectra of Molecules". The authors thank Rebecca Sure and Dr. Stephan Ehrlich for helpful discussions.

## References

- 1 *Computational Spectroscopy*, ed. J. Grunenberg, Wiley-VCH, Weinheim, 2010.
- 2 J. Gross, *Mass Spectrometry - A Textbook*, Springer-Verlag, Heidelberg, 2nd edn, 2011.
- 3 F. W. McLafferty and F. Turecek, *Interpretation of Mass Spectra*, University Science Books, Mill Valley, CA, 3rd edn, 1993.
- 4 H. M. Rosenstock, M. B. Wallenstein, A. L. Wahrhaftig and H. Eyring, *Proc. Natl. Acad. Sci.*, 1952, **38**, 667–678.
- 5 R. A. Marcus and O. K. Rice, *J. Phys. Colloid Chem.*, 1951, **55**, 894–908.
- 6 R. A. Marcus, *J. Chem. Phys.*, 1952, **20**, 359–364.
- 7 L. Drahos and K. Vékely, *J. Mass Spec.*, 2001, **36**, 237–263.
- 8 S. Grimme, *Angew. Chem. Int. Ed.*, 2013, **52**, 6306–6312.
- 9 D. Marx and J. Hutter, *Ab Initio Molecular Dynamics*, Cambridge University Press, 2009.
- 10 P. B. Timmermans, P. C. Wong, A. T. Chiu, W. F. Herblin, P. Benfield, D. J. Carini, R. J. Lee, R. R. Wexler, J. A. Saye and R. D. Smith, *Pharmacol. Rev.*, 1993, **45**, 205–251.
- 11 SDBS Web: <http://sdb.srioddb.aist.go.jp> (National Institute of Advanced Industrial Science and Technology), accessed April 15, 2014.
- 12 P. F. Wiley, K. Gerzon, E. H. Flynn, M. V. Sigal, O. Weaver, U. C. Quarck, R. R. Chauvette and R. Monahan, *J. Am. Chem. Soc.*, 1957, **79**, 6062–6070.
- 13 S. Masamune, G. S. Bates and J. W. Corcoran, *Angew. Chem. Int. Ed.*, 1977, **16**, 585–607.
- 14 M. C. Wani, H. L. Taylor, M. E. Wall, P. Coggon and A. T. McPhail, *J. Am. Chem. Soc.*, 1971, **93**, 2325–2327.
- 15 P. B. Schiff, J. Fant and S. B. Horwitz, *Nature*, 1979, **277**, 665–667.
- 16 P. Gazerro, M. C. Proto, G. Gangemi, A. M. Malfitano, E. Ciaglia, S. Pisanti, A. Santoro, C. Laezza and M. Bifulco, *Pharmacol. Rev.*, 2012, **64**, 102–146.
- 17 A secondary motivation was the availability of experimental EI mass spectral data.
- 18 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 4 or the computational 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<sup>8</sup>, this may not be crucial.
- 19 It should be noted that the ions at  $m/z$  99 and  $m/z$  86 did not serve as immediate parent ions to the ions at  $m/z$  71 and  $m/z$  58, respectively. Thus, the loss of 28 atomic mass units between them can really only formally be attributed to the dissociation of ethylene.
- 20 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B*, 1998, **58**, 7260–7268.
- 21 M. Elstner, P. Hobza, T. Frauenheim, S. Suhai and E. Kaxiras, *J. Chem. Phys.*, 2001, **114**, 5149–5155.
- 22 J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.
- 23 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 24 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.
- 25 A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2005, **123**, 154101.
- 26 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
- 27 TURBOMOLE: R. Ahlrichs et al., Universität Karlsruhe 2009. See <http://www.turbomole.com>.
- 28 This in some cases yielded one imaginary vibrational 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.
- 29 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 30 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.
- 31 A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571–2577.
- 32 PBE12 is the same as PBE0, only with a Fock exchange mixing coefficient of 0.5. The SVx basis set is the same as SV(P), only without d-

polarization functions at the carbon atoms.

- 33 W. Weber and W. Thiel, *Theor. Chem. Acc.*, 2000, **103**, 495–506.
- 34 N. D. Mermin, *Phys. Rev.*, 1964, **134**, A112–A125.
- 35 R. W. Warren and B. I. Dunlap, *Chem. Phys. Lett.*, 1996, **262**, 384–392.
- 36 A. D. Rabuck and G. E. Scuseria, *J. Chem. Phys.*, 1999, **110**, 695–700.
- 37 *MNDO2005, Version 7.0*, W. Thiel, *MPI für Kohlenforschung, Mülheim, Germany*.
- 38 (a) 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.; (b) F. Neese, *WIREs Comp. Mol. Sci.*, 2012, **2**, 73–78.
- 39 S. E. Stein, in *Mass Spectra*, ed. P. J. Linstrom and W. G. Mallard, National Institute of Standards and Technology, Gaithersburg MD 20899, <http://wbbook.nist.gov>, (retrieved April 15, 2014).
- 40 SWGDRUG MS Library Version 2.0 (January 1, 2014), <http://www.swgdrug.org/ms.htm>, accessed in April 2014.