Title

Beyond the Ground State: Predicting Electron Ionization Mass Spectra Using Excited-State Molecular Dynamics

Journal

Journal of Chemical Information and Modeling, 62(18)

ISSN

1549-9596

Authors

Wang, Shunyang

Kind, Tobias

Bremer, Parker Ladd

et al.

Publication Date

2022-09-26

DOI

10.1021/acs.jcim.2c00597

Peer reviewed

HHS Public Access

Author manuscript

J Chem Inf Model. Author manuscript; available in PMC 2024 October 21.

Published in final edited form as:

J Chem Inf Model. 2022 September 26; 62(18): 4403–4410. doi:10.1021/acs.jcim.2c00597.

Beyond the Ground State: Predicting Electron Ionization Mass Spectra Using Excited-State Molecular Dynamics

Shunyang Wang,
West Coast Metabolomics Center, UC Davis Genome Center, University of California, Davis, California 95616, United States

Department of Chemistry, University of California, Davis, California 95616, United States

Tobias Kind,
West Coast Metabolomics Center, UC Davis Genome Center, University of California, Davis, California 95616, United States

Parker Ladd Bremer,
West Coast Metabolomics Center, UC Davis Genome Center, University of California, Davis, California 95616, United States

Department of Chemistry, University of California, Davis, California 95616, United States

Dean J. Tantillo,
Department of Chemistry, University of California, Davis, California 95616, United States

Oliver Fiehn
West Coast Metabolomics Center, UC Davis Genome Center, University of California, Davis, California 95616, United States

Abstract

Here, we provide an algorithm that introduces excited states into the molecular dynamics prediction of the 70 eV electron ionization mass spectra. To decide the contributions of different electronic states, the ionization cross section associated with relevant molecular orbitals was calculated by the binary–encounter–Bethe (BEB) model. We used a fast orthogonalization model/single and double state configuration interaction (OM2/CISD) method to implement excited

Corresponding Author Oliver Fiehn – West Coast Metabolomics Center, UC Davis Genome Center, University of California, Davis, California 95616, United States; ofiehn@ucdavis.edu.

Author Contributions

S.W. and T.K. developed the method and designed the experiments. S.W., T.K., P.B., D.J.T., and O.F. provided guidance and wrote the manuscript. All authors read and approved the final manuscript.

The authors declare no competing financial interest.

The code of Modified QCEIMS and processing scripts is available at https://github.com/Shunyang2018/EXMD. Input structures, MD trajectories, output files are available at https://zenodo.org/record/5796857. In silico Mixed MS spectra have been uploaded to MassBank of North America: https://massbank.us/spectra/browse?query=tags.text%3D%3D%22QCEIMS%22&text=&size=10

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c00597.

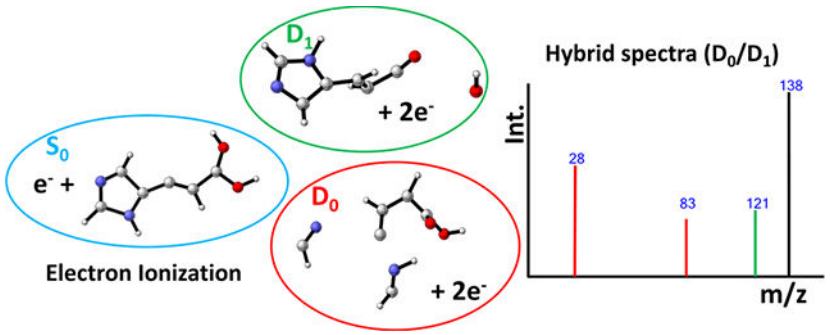
Molecular orbital contours; head-to-tail mass spectra examples; additional details for parameter settings (PDF)

Input molecular structures; similarity scores and failure rate of the molecular dynamic simulations (XLSX)

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c00597

states calculations and combined this with the GFN1-xTB semiempirical model. Demonstrated by predicting the mass spectrum of urocanic acid, we showed better accuracies to experimental spectra using excited-state molecular dynamics than calculations that only used the ground-state occupation. For several histidine pathway intermediates, we found that excited-state corrections yielded an average of 73% more true positive ions compared to the OM2 method when matching to experimental spectra and 16% more true positive ions compared to the GFN method. Importantly, the exited state models also correctly predict several fragmentation reactions that were missing from both ground-state methods. Overall, for 48 calculated molecules, we found the best average mass spectral similarity scores for the mixed excited-state method compared to the ground-state methods using either cosine, weighted dot score, or entropy similarity calculations. Therefore, we recommend adding excited-state calculations for predicting the electron ionization mass spectra of small molecules in metabolomics.

Graphical Abstract



INTRODUCTION

Using quantum chemical methods along with statistical methods to predict electron ionization (EI) mass spectra (MS)[1,2] has been explored for many types of molecules, including organic molecules, inorganic molecules,[3] and heavy-metal-containing molecules.[4] The QCEIMS software based on quantum chemistry[1,3–6] can provide reasonable results and detailed reaction pathways. In a recent publication, the QCxMS software was introduced, combining electron ionization and collision-induced dissociation modeling into a single software package.[7] However, a shortcoming of these methods is the ground-state potential energy surface (PES) may have inaccurate results even with density functional theory (DFT) methods,[6,8] which can cause missing fragment ions. For example, Wang et al. showed that only 50% of the observed ions were generally captured in 681 trimethylsilylated molecules, which compromised the accuracy of QCxMS simulations. To improve predictions of relative energies of structures on PESs, alternative theoretical methods (e.g., post-Hartree–Fock methods[9]) can be utilized. In addition, the inclusion of the excited states might improve predictions by accounting for fragmentation reactions that are not accessible on the ground-state PES.[10] The focus of the current study was to examine the impact of including excited states in the semiempirical molecular dynamics method for the prediction of 70 eV EI mass spectra.

The EI ionization process can be described as a (1e, 2e) gas phase reaction

$$M + e^- \rightarrow (AB)^{+\cdot} + 2e^-$$

In this process, the analyte molecule M is impacted by e^-, which is scattered and causes the loss of another e^-, resulting in the reaction complex $(AB)^{+\bullet}$, which can undergo further reactions.[11] Ionization cross sections play an important role in providing information about the EI process and the binary–encounter–Bethe (BEB) model[12] provides an ab initio means of calculating ionization cross sections without any fitting parameters. The 70 eV electron can traverse the molecule in a few femtoseconds, a much shorter time than the bond vibration period. Therefore, the transition from the $v = 0$ vibrational state of the ground electronic state to an excited state generally obeys the Frank–Condon principle and can be modeled as "vertical ionization":[13] one electron is removed from the neutral molecule with the molecular structure unchanged.

The impact energy can be divided into ionization energy and impact excess energy. According to the vertical ionization model, upon ionization, the impact excess energy is saved in highly excited vibrational modes, and this is the driving force of future fragmentation. With the increase in molecular size, energy must be distributed to more degrees of freedom (DOF), and thus more energy is needed for redistribution.[14] QCxMS assumes that the excited ion state goes through an internal conversion to the vibrationally hot ion ground state.[1] This model introduces the impact of excess energy (IEE) on the nuclear DOF in a continuous time by increasing the velocity of each atom. QCxMS has been tested successfully in many cases,[3–9] and it has been shown that the IEE distribution model[1] has a small effect on the simulated spectra.

We hypothesized that simulations including excited states could reveal reactions not encountered in ground-state simulations. Because we focused on finding more reactions, nonadiabatic coupling between excited states[15] was not considered in this project. Instead, in the excited-state molecular dynamics, the molecule starts at different states but jumps back to the ground state once the fragmentation is detected. In addition, one must determine which methods are the most appropriate to describe relevant excited state PESs in terms of both accuracy and computing resource feasibility. We found that the orthogonal-corrected semiempirical quantum chemical methods OMx[16] can be used for ground-state mass spectral predictions.[8] To include the dynamic and static electron-correlation effects for excited states, configuration interaction (CI) and multireference (MR) methods are needed.[17] A graphical unitary-group approach (GUGA)[18] can be combined with OMx models. In this way, we can capture the excited-state correlation effects with state-based semiempirical treatments.[19,20] In this paper, we provide a prototype of a mass spectra prediction model and discuss additional method improvements for large-scale MS predictions. This model combines ground-state and excited-state molecular dynamics based on the BEB model.

METHODS

Ionization Cross Section by the BEB Model.

The BEB model is simplified from the binary–encounter–dipole model[12] and has wide applications, including mass spectrometer normalization, plasma modeling, and material radiation effect calculation.[21] The electron impact ionization cross section for molecular orbital i (MO_i) is calculated by

$$\sigma_i = \frac{S_i}{t_i + u_i + 1}\left[\frac{\ln t_i}{2}\left(1 - \frac{1}{t_i^2}\right) + 1 - \frac{1}{t_i} - \frac{\ln t_i}{t_i + 1}\right]$$

$$u = \frac{U}{B}, t = \frac{T}{B}, S = \frac{4\pi a_0^2 N R^2}{B^2}$$

$$a_0 = 0.592 \text{ Å}, R = 13.61 \text{ eV}$$

where T is the energy of the impact electrons, B is the electron-binding energy of MO_i, and U is the kinetic energy of MO_i. The occupation number N of MO_i is 2 for ground-state molecules. The GAMESS[22] package is used to calculate the molecular orbital properties under the Hartree–Fock method with a 6–31G basis set. The GAUSSIAN program[23] is used to optimize the structure and calculate the molecular orbital contours. The Avogadro v1.2 software[24] was used to visualize the molecular orbitals. A python script package (https://github.com/Shunyang2018/EXMD) was developed to calculate the electron impact ionization cross section and the relative ratio of each electronic state after ionization.

Modified QCEIMS Algorithm.

The QCEIMS v4.0[1,3–6] code was used with several modifications to perform excited-state molecular dynamics. Default settings were applied for the ground-state calculations at the GFN2-xTB level.[25] For the excited state, uniform velocity scaling was enforced during the internal conversion step. The MNDO99 program[26] was used for the semiempirical OM2[16,27] level gradient calculations for excited-state molecular dynamics (MD).[18,20,28] The active space is decided by the following rule set. Each π bond provides a pair of occupied/unoccupied orbitals, and each oxygen or nitrogen atom provides a lone pair orbital. Because the radical cation system is an open shell, only one reference occupation is used. Restricted Open-shell Hartree–Fock (ROHF) is used, while single and double excitations are allowed for the reference configurations for simplicity. Different parameter settings of the excited-state molecular dynamics are tested in Figures S8–S10. Once fragmentation is detected, the ionization potential (IP) of each fragment is calculated and the partial charge is assigned by the Boltzmann distribution. Another MD simulation is performed on the fragment with the largest partial charge for secondary fragmentations. This MD simulation is for the ion ground state, where the fractional orbital occupations[29] in unrestricted OM2/SCF calculations are used. Ions larger than 15 Da are counted and used to generate

in silico spectra of the ground state and excited state separately. Then, the excited-state (D_1, D_2···) spectra are used as corrections to the ground-state (D_0) spectrum per their relative ratios obtained from the BEB model, assuming that the D_0 state is ionized from the highest occupied molecular orbital (HOMO) with an ionization cross section σ_0, D_1 is from HOMO – 1 with σ_1 and so on. Theoretically, D_2 and higher excited states can contribute to fragmentation reactions; however, we found that D_1 state calculations already predicted most of the experimentally observed reactions. Importantly, we did not aim to theoretically or comprehensively compare all different methods but to give practical applications of molecules that are typically encountered in metabolomics research. The OM2/CISD method had an average failure rate of 0.64 across all trajectories of all 48 molecules due to self-consistent field convergence problems (Supporting S8). For simplicity, only the two lowest electronic states (D_0, D_1) are taken as reference states and nonadiabatic crossing is neglected. Because excited-state calculations are used as corrections, this approach can be extended to other higher excited states in the future.

RESULTS

Urocanic Acid as a Demonstration Case.

Urocanic acid is an intermediate of histidine catabolism.[30] We chose urocanic acid as an example because it contains both nitrogen and oxygen elements, an imidazole aromatic system, and a carboxylic acid functional group and features typical of many organic molecules of biological interest. The MOs of urocanic acid are shown in Table S1 and visualizations of the MO contours are shown in Figure S2. The HOMO, HOMO – 1, HOMO – 3, and HOMO – 5 orbitals are π orbitals, and the HOMO – 2, HOMO – 4, HOMO – 6, HOMO – 7, HOMO – 8, and HOMO – 9 are n orbitals associated with lone pairs on oxygen and nitrogen. Consequently, an active space of 11 electrons and 10 orbitals (11, 10) should be sufficient for modeling the first excited state of the urocanic acid radical cation. The MO ionization cross section according to electron kinetic energy of the four highest occupied MOs is shown in Figure 1. Vertical ionization from the HOMO generates the D_0 state (remove one electron from HOMO), while that from HOMO – 1 generates the D_1 state. According to the ratio of σ at 70 eV, the ground state is significantly more likely than other states, thus the spectrum from ground-state MD contributes most to the final corrected spectrum. The apex of the ionization cross section curve is slightly lower than 70 eV and shifts to 70 eV with lower energy MOs, which is consistent with the region of the highest ionization efficiency. That is the basic reason why 70 eV is the classic experimental energy for electron ionization in gas chromatography–mass spectrometry.

Different quantum chemistry methods, including GFN1-xTB, GFN2-xTB, OM2, and PBE0[31]-D3[32]/SV(P)[33] (density functional theory) were tested on urocanic acid in the ground electronic state (Figure 2). OM2 (Figure 2c) is one of the fastest methods available in the QCEIMS program, but it is only parameterized for five elements: C, H, O, N, and F. The default GFN1-xTB method and its advanced GFN2-xTB version (Figure 2a) yielded only minute differences in predicted mass spectra when applied to the chemical urocanic acid. More GFN2-xTB calculations can be found in Supporting File S13. Interestingly, neither the PBE0/SV(P) method (Figure 2b) nor GFN1-xTB or GFN2-xTB correctly predicted the

experimental *m*/*z* 45 fragment ion. The PBE0/SV(P) method requires larger computing resources. This disadvantage is amplified if hundreds of trajectories need to be calculated. In comparison to the GFN1-xTB method, the prediction of relative intensities of other fragment ions (*m*/*z* 138, 93, 39) did not improve with PBE0/SV(P) method; more false positive ions (*m*/*z* 98, 44) were captured and ions missing in simulations with the semiempirical methods were still not found. The comparison between these three different methods showed that optimizing the potential energy surface did not solve all of the problems in mass spectral predictions.

The spectrum calculated from the first excited-state MD and the mixed spectrum after correction of the first excited state are compared in Figure 2c,d. When moving from the ground state to the excited state, the molecular ion intensity decreased dramatically, and more fragment ions and higher intensities of these ions were found in the low-mass range. This change can be explained by a higher reactivity of the excited state. We used the weighted dot-product score[8] to evaluate the similarity between in silico spectra and the experimental reference spectra given in the NIST17 mass spectral library. Although the dot-product score slightly decreased from 907 to 894 with the excited-state correction, the details of the in silico spectrum improved. For example, the group of fragment ions around *m*/*z* 28, 39, and 67 were better captured, giving higher confidence when comparing the spectra. The *m*/*z* 39 fragment ion was identified as C_2HN by Mass Spectrum Interpreter[34] software, a product that was missing in all ground-state simulations. In the first excited-state simulation, the *m*/*z* 39 ion was found in 30 out of 400 trajectories, arising from C_2HN^+ and $C_3H_3^+$. Overall, the mixed spectra predicted 62 instead of only 44 fragment ions from the D_0 spectrum. The number of true positives, i.e., predicted ions that were found in experimental spectra, increased from 31 to 37. Some problems remained unsolved even when including the D_1 excited state, however. For example, the excited-state MD overestimated the intensity of the low-mass ions, especially *m*/*z* 28 and 40. This problem might be solved by including more electronic states.

When we compared the experimental spectrum of urocanic acid with all in silico generated spectra by all molecular dynamic methods, we found that three significant fragment ions were missed or underestimated, especially evident when removing isotope ions (Figure S12). Here, *m*/*z* 68 was missing from all simulations, whereas *m*/*z* 120 and *m*/*z* 94 were represented much found at much lower intensity in the predicted spectra compared to experimental spectra (Figure 2). All three ions shared a similar pattern: while the exact fragment ion product is missing or has a very low abundance, there was an abundant ion within 1 Da of its expected location (*m*/*z* ± 1). This observation implied that hydrogen rearrangements were not described correctly in the simulations. For example, the *m*/*z* 94 fragment ion results from a loss of CO_2, which is a product of a three-membered ring rearrangement reaction of hydrogen that is transferred to a double bond. We found that the other two fragment ions *m*/*z* 120 and 68 were also generated by hydrogen rearrangements, as discussed later in more detail.

Rearrangement Reaction in the Unimolecular Dissociation.

To validate the overall performance of excited-state corrections and to investigate the rearrangement reactions, we selected several molecules from the histidine biosynthetic pathway.[35] Seven of the 10 major pathway intermediates had 70 eV EI mass spectra included in the NIST17 database. Table 1 shows the parameters used in the calculation. On average, the ground-state simulation contributed around 60% to the final mixed spectra, while the first excited state contributed around 40%. If we added the second excited state with a similar weight as the D_1 state, the contribution of the ground state would be decreased to around 40%. The mass spectral prediction with the OM2 method had an average running time of 1.55 h per molecule, while GFN1-xTB required 7.2 h per molecule, both on 16 CPU threads. The OM2/CISD simulation took about twice as long as the OM2 ground-state calculation. MD on the first excited state required a similar amount of simulation time as the ground-state simulation. GFN1-xTB did not allow for computations of excited-state MD trajectories. The OM2 method had its own disadvantage because it was only parameterized for limited elements, C, H, N, O, and F, while GFN1-xTB included 86 elements.

As shown in Figure 3, most ion fragments derived from alanine were correctly predicted. The in silico spectra predicted the loss of two or three hydrogen fragments ($[M - 1]^+$, $[M - 2]^{+\bullet}$ ions) from the molecular ion m/z 89. Such hydrogen losses were not verified by the experimental spectrum and originated from the OM2 method parametrization.[8] The prediction of the histamine mass spectrum benefited from the additional ions predicted by including the D_1 excited state. However, the base ion intensity for m/z 82 was underestimated in both the D_0 and D_1 simulations. As in the urocanic acid example described above, this problem is a result of hydrogen rearrangement reactions.

Head-to-tail plots for the other molecules from Table 1 can be found in Figures S4–S7. For four of the seven molecules examined, the base ions were missing (m/z 94 of urocanic acid, m/z 82 of histidine and histamine, m/z 96 of 1-methylhistidine). The rule-based Mass Frontier software (Thermo Fisher Scientific Inc.) predicted that these ions were generated by hydrogen rearrangement reactions or a mobile proton mechanism.[36] The MD trajectories in our calculations also supported this model; however, we observed very few occurrences of these specific hydrogen rearrangement reactions. As an example, we plotted the energy vs time for two trajectories at ground-state simulations of the fragmentation of the urocanic acid radical cation (Figure 4). Trajectory #13 generated m/z 93 ($C_5N_2H_5^+$), with a neutral loss of $CO_2H^\bullet$, while trajectory #318 generated m/z 94 ($C_5H_2H_6^{+\bullet}$), with a neutral loss of carbon dioxide. The difference was caused by a hydrogen rearrangement between two fragments in the latter trajectory. Although there was mobile hydrogen in trajectory #13, it was only intramolecular rearrangement of the $C_5N_2H_5^+$ fragment. To better visualize the electronic energy change, we plotted the moving average trend line per 100 fs. This trend line showed that energies stayed unchanged once the fragments were generated in trajectory #13. In trajectory #318, the energy decreased after an energy barrier was surmounted. Because unimolecular reactions did not include collision and energy exchange, the electronic energy of the system did not change for most homolytic bond cleavages. In other words, energy curves stayed at the same level after passing through transition states. However, rearrangement fragmentations face different conditions because forming

new bonds during a rearrangement reaction decreased the system's total electronic energy level. Therefore, rearrangement reactions were usually exothermic. After the transition state was passed, products were found at energetically lower states, and this extra energy was converted to kinetic energy (translational energy of fragments). This was the so-called kinetic energy release (KER) process.[37] In the QCxMS simulation, all molecular dynamics steps were under the constant total energy (NVE ensemble), except the heating process. These two different energy change types (Figure 4) were consistent with our observations that the electronic energy converted to translational energy of fragments and caused them to depart while the total energy was conserved. This observation proved that the model captured the hydrogen rearrangement reactions as an exit channel on the ground-/excited-state potential energy surface (PES). Yet, relative intensities of fragment ions (selectivity of the reactions) were still underestimated.

One possible reason is the simulation time scale (on the level of a picosecond) is limited when compared to the total time that mass spectrometers use (on the level of a microsecond[38]). The reaction time scale of unimolecular rearrangement reaction in a mass spectrometer is 10^{-11}–10^{-6} s, while a simple bond dissociation fragmentation proceeds much faster at 10^{-12} s.[39] The typical ion flight time in a quadrupole or time-of-flight mass spectrometer is around 50 μs.[40] Our parameter settings simulated molecular dynamics for up to 5,000 femtoseconds, which reproduced most reactions well but failed to account for the frequency of rearrangement reactions. However, even for simple fragmentation reactions, 5 ps might not be sufficient time as we found 42/400 trajectories that did not yield any fragmentations within the maximum simulation time. For the PBE0/SV(P) method, we found even 115 trajectories that did not yield any fragmentation, increasing the prediction of the relative abundance of the unfragmented molecular ion and thereby lowering the weighted dot similarity score by 230 units, a decrease of 15% in prediction scores. Another explanation why rearrangement reactions were missed is that such reactions might need a specific starting conformation to overcome positional barriers.[41] More comprehensive conformer sampling might overcome this problem. Yet, when we tested conformer-rotamer sampling and Wigner distribution sampling, we did not improve results for the lacking rearrangement reactions. Hence, for correctly predicting hydrogen rearrangements, the limitation of simulation time might remain the main obstacle.

Method Test on Small Molecules.

The examples discussed before do not provide a theoretical validation of the value of adding an excited-state method to the ground-state method. Unfortunately, more accurate ab initio techniques are prohibited by the size of metabolites tested here. Instead, to compare QCxMS-based ground-state simulations to the new excited-state method, we randomly selected 48 molecules from our previously published study.[8] We used both weighted dot-score similarity and the Jaccard index that calculated the number of true positive predicted ions divided by the number of all ions observed in the combination of the in silico and experimental spectra. For 48 calculated molecules, we found an improvement from an average dot product similarity score of 681 in the OM2 method to 724 for the GFN1-xTB method and 726 using the mixed method, adding the excited states to supplemented ground-state simulations (Supporting S8). We have recently introduced the concept of

entropy similarity scores that outperformed dot-product similarities and improved False Discovery Rates.[42] Comparing spectral entropy similarities indeed showed a significant improvement for the mixed method: the two ground-state methods resulted in an average entropy similarity score of 600, whereas the mixed method gave an average score of 680 and a narrower Kernel density distribution (Figure 5). The OM2 method showed a much lower average Jaccard index of 0.34 than the default GFN1-xTB or the mixed method. While the average Jaccard index of these two methods was not statistically significantly different, the mixed method clearly showed less variance and a higher maximum density of the Jaccard indices of the modeled in silico spectra in comparison to the GFN1-xTB method. On average, the mixed method yielded 16 and 73% more true positive ions in pairwise comparisons than GFN1-xTB and OM2 ground-state methods, respectively. We found that the mixed method avoided extremely poor simulations that were observed for the OM2 methods with a dot score of <200 with a true positive rate of <0.4 (Supporting File S13).

CONCLUSIONS

We here show for the first time that molecular dynamics can be utilized with excited-state calculations using the BEB model to scale contributions based on ionization cross sections. We provided and tested this excited-state correction method for quantum chemistry molecular dynamics prediction of standard 70 eV mass spectra and showed that it improved the existing GFNn-xTB method by generating about 16% more correctly predicted fragment ions. For example, the mixed method presented here added hydrogen shift reactions that were missed by the classic methods. When comparing this mixed method with other tools such as DFT and semiempirical methods like OM2, we found clear improvements in accuracy that came with only 20% increased computational times. Although our OM2/CISD excited-state simulations discovered more fragmentation reactions than the standard model, improvement in spectra similarities to experimental spectra were limited because the GFNn-xTB method generally already yielded a high number of excellent predictions, as shown in detail for the molecule urocanic acid. However, predicting rearrangement reactions is the bottleneck of QCxMS. Because much longer simulation times might be prohibitively expensive with respect to computational costs, we propose that machine learning methods are needed to recognize rearrangement reactions. We recommend this mixed model to generate in silico electron ionization mass spectral libraries for small molecules.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank Stefan Grimme (University of Bonn, Germany) for providing the source code of QCEIMS. They thank C. William McCurdy (University of California, Davis) for inspiring suggestions and enjoyable discussions about electron ionization and excited-state molecular dynamics.

Funding

Funding for the “West Coast Metabolomics Center for Compound Identification” was provided by the National Institutes of Health under the award number NIH U2C ES030158.

ABBREVIATIONS

BEB	binary–encounter–Bethe model
QCEIMS	quantum chemistry electron ionization mass spectrometry
PES	potential energy surface
DFT	density functional theory
IEE	impact excess energy
OMx	orthogonal-corrected semiempirical quantum chemical methods
CI	configuration interaction
MR	multireference methods
MO	molecular orbital
MD	molecular dynamics
IP	ionization potential

REFERENCES

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

(2). Borges RM; Colby SM; Das S; Edison AS; Fiehn O; Kind T; Lee J; Merrill AT; Merz KM; Metz TO; Nunez JR; Tantillo DJ; Wang L-P; Wang S; Renslow RS Quantum Chemistry Calculations for Metabolomics. Chem. Rev 2021, 121, 5633–5670. [PubMed: 33979149]

(3). Ásgeirsson V; Bauer CA; Grimme S Quantum chemical calculation of electron ionization mass spectra for general organic and inorganic molecules. Chem. Sci 2017, 8, 4879–4895. [PubMed: 28959412]

(4). Koopman J; Grimme S Calculation of Electron Ionization Mass Spectra with Semiempirical GFNn-xTB Methods. ACS Omega 2019, 4, 15120–15133. [PubMed: 31552357]

(5). Koopman J; Bauer C; Ásgeirsson V QCEIMS 4.0 Installation Guide & Manual, 2020.Bauer CA; Grimme S How to Compute Electron Ionization Mass Spectra from First Principles. J. Phys. Chem. A 2016, 120, 3755–3766. [PubMed: 27139033] Bauer CA; Grimme S First principles calculation of electron ionization mass spectra for selected organic drug molecules. Org. Biomol. Chem. 2014, 12, 8737–8744. [PubMed: 25260171] Bauer CA; 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.

(6). Bauer CA; 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. [PubMed: 25392949]

(7). Koopman J; Grimme S From QCEIMS to QCxMS: A Tool to Routinely Calculate CID Mass Spectra Using Molecular Dynamics. J. Am. Soc. Mass Spectrom. 2021, 32, 1735–1751. [PubMed: 34080847] Koopman J; Grimme S Calculation of mass spectra with the QCxMS method for negatively and multiply charged molecules. chemrxiv 2022, 1–16, DOI: 10.26434/chemrxiv-2022-w5260.

(8). Wang S; Kind T; Tantillo DJ; Fiehn O Predicting in silico electron ionization mass spectra using quantum chemistry. J. Cheminf 2020, 12, No. 63.

(9). Townsend J; Kirkland JK; Vogiatzis KD Chapter 3 - Post-Hartree-Fock Methods: Configuration Interaction, Many-body Perturbation Theory, Coupled-cluster Theory. In Mathematical Physics in Theoretical Chemistry; Blinder SM; House JE, Eds.; Elsevier, 2019; pp 63–117.

(10). Lorquet JC Basic questions in mass spectrometry. Org. Mass Spectrom 1981, 16, 469–482.Lorquet JC Landmarks in the theory of mass spectra. Int. J. Mass Spectrom. 2000, 200, 43–56.

(11). Mark TD Fundamental aspects of electron impact ionization. Int. J. Mass Spectrom. Ion Phys 1982, 45, 125–145.

(12). Kim Y-K; Rudd ME Binary-encounter-dipole model for electron-impact ionization. Phys. Rev. A 1994, 50, 3954–3967. [PubMed: 9911367]

(13). Gross JH Principles of Ionization and Ion Dissociation. In Mass Spectrometry: A Textbook; Gross JH, Ed.; Springer: Berlin Heidelberg, 2011; pp 21–66.Franck J; Dymond EG Elementary processes of photochemical reactions. Trans. Faraday Soc. 1926, 21, 536–542.Condon E A Theory of Intensity Distribution in Band Systems. Phys. Rev. 1926, 28, 1182–1201.

(14). Bente PF; McLafferty FW; McAdoo DJ; Lifshitz C Internal energy of product ions formed in mass spectral reactions. Degrees of freedom effect. J. Phys. Chem. A 1975, 79, 713–721.

(15). Nelson TR; White AJ; Bjorgaard JA; Sifain AE; Zhang Y; Nebgen B; Fernandez-Alberti S; Mozyrsky D; Roitberg AE; Tretiak S Non-adiabatic Excited-State Molecular Dynamics: Theory and Applications for Modeling Photophysics in Extended Molecular Materials. Chem. Rev. 2020, 120, 2215–2287. [PubMed: 32040312]

(16). Dral PO; 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. [PubMed: 26771204]

(17). Gerber RB; Shemesh D; Varner ME; Kalinowski J; Hirshberg B Ab initio and semi-empirical Molecular Dynamics simulations of chemical reactions in isolated molecules and in clusters. Phys. Chem. Chem. Phys. 2014, 16, 9760–9775. [PubMed: 24569494]

(18). Koslowski A; Beck ME; Thiel W Implementation of a general multireference configuration interaction procedure with analytic gradients in a semiempirical context using the graphical unitary group approach. J. Comput. Chem. 2003, 24, 714–726. [PubMed: 12666163]

(19). Tuna D; Lu Y; Koslowski A; Thiel W Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Benchmarks of Electronically Excited States. J. Chem. Theory Comput. 2016, 12, 4400–4422. [PubMed: 27380455] Silva-Junior MR; Thiel W Benchmark of Electronically Excited States for Semiempirical Methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. J. Chem. Theory Comput 2010, 6, 1546–1564. [PubMed: 26615690] Foresman JB; Head-Gordon M; Pople JA; Frisch MJ Toward a systematic molecular orbital theory for excited states. J. Phys. Chem. B 1992, 96, 135–149.

(20). Liu J; Thiel W An efficient implementation of semiempirical quantum-chemical orthogonalization-corrected methods for excited-state dynamics. J. Chem. Phys. 2018, 148, No. 154103.

(21). Możejko P; Sanche L Cross section calculations for electron scattering from DNA and RNA bases. Radiat. Environ. Biophys. 2003, 42, 201–211. [PubMed: 14523567]

(22). Schmidt MW; Baldridge KK; Boatz JA; Elbert ST; Gordon MS; Jensen JH; Koseki S; Matsunaga N; Nguyen KA; Su S; Windus TL; Dupuis M; Montgomery JA General atomic and molecular electronic structure system. J. Comput. Chem. 1993, 14, 1347–1363.

(23). Gaussian 16 Rev. C.01; Wallingford, CT, 2016.

(24). Hanwell MD; Curtis DE; Lonie DC; Vandermeersch T; Zurek E; Hutchison GR Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J. Cheminf 2012, 4, No. 17.

(25). Bannwarth C; Ehlert 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. [PubMed: 30741547]

(26). Dewar MJS; Thiel W Ground states of molecules. 38. The MNDO method. Approximations and parameters. J. Am. Chem. Soc. 1977, 99, 4899–4907.

(27). Thiel W Semiempirical quantum–chemical methods. WIREs Comput. Mol. Sci. 2014, 4, 145–157.

(28). Tuna D; Spörkel L; Barbatti M; Thiel W Nonadiabatic dynamics simulations of photoexcited urocanic acid. Chem. Phys. 2018, 515, 521–534.

(29). Bauer CA; Hansen A; Grimme S 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. [PubMed: 27906486]

(30). Koltai T; Reshkin SJ; Harguindey S Chapter 15 - Pharmacological Interventions Part III. In An Innovative Approach to Understanding and Treating Cancer: Targeting pH, Koltai T; Reshkin SJ; Harguindey S, Eds.; Academic Press, 2020; pp 335–359.

(31). Perdew JP; Ernzerhof M; Burke K Rationale for mixing exact exchange with density functional approximations. J. Chem. Phys. 1996, 105, 9982–9985.

(32). Grimme S; Antony J; Ehrlich S; Krieg H A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. 2010, 132, No. 154104.

(33). Schäfer A; Horn H; Ahlrichs R Fully optimized contracted Gaussian basis sets for atoms Li to Kr. J. Chem. Phys. 1992, 97, 2571–2577.

(34). Mass Spectrum Interpreter2019 https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:interpreter.

(35). Alifano P; Fani R; Liò P; Lazcano A; Bazzicalupo M; Carlomagno MS; Bruni CB Histidine biosynthetic pathway and genes: structure, regulation, and evolution. Microbiol. Rev 1996, 60, 44–69. [PubMed: 8852895]

(36). Cautereels J; Blockhuys F Quantum Chemical Mass Spectrometry: Verification and Extension of the Mobile Proton Model for Histidine. J. Am. Soc. Mass Spectrom. 2017, 28, 1227–1235. [PubMed: 28349436] Tsugawa H; Kind T; Nakabayashi R; Yukihira D; Tanaka W; Cajka T; Saito K; Fiehn O; Arita M Hydrogen rearrangement rules: computational MS/MS fragmentation and structure elucidation using MS-FINDER software. Anal. Chem. 2016, 88, 7946–7958. [PubMed: 27419259] Demarque DP; Crotti AEM; Vessecchi R; Lopes JLC; Lopes NP Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products. Nat. Prod. Rep. 2016, 33, 432–455. [PubMed: 26673733]

(37). Holmes JL; Terlouw JK The scope of metastable peak observations. Org. Mass Spectrom. 1980, 15, 383–396.Williams DH A transition-state probe. Acc. Chem. Res. 1977, 10, 280–286.

(38). Wollnik H Time-of-flight mass analyzers. Mass Spectrom. Rev. 1993, 12, 89–114.

(39). Holmes JL Assigning structures to ions in the gas phase. Org. Mass Spectrom. 1985, 20, 169–183.

(40). Hesse M Mass Spectrometry. A Textbook. By Jürgen H. Gross. Angew. Chem., Int. Ed. 2004, 43, 4552.

(41). Semialjac M; Schröder D; Schwarz H Car–Parrinello Molecular Dynamics Study of the Rearrangement of the Valeramide Radical Cation. Chem. - Eur. J 2003, 9, 4396–4404. [PubMed: 14502626]

(42). Li Y; Kind T; Folz J; Vaniya A; Mehta SS; Fiehn O Spectral entropy outperforms MS/MS dot product similarity for small-molecule compound identification. Nat. Methods 2021, 18, 1524–1531. [PubMed: 34857935]

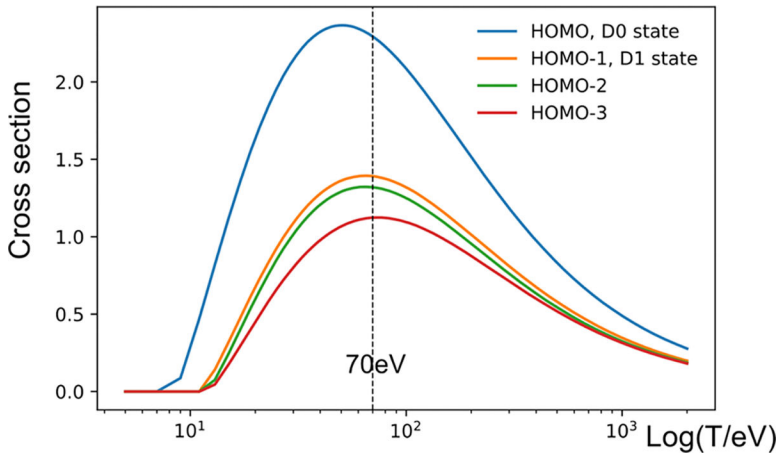


Figure 1.

Ionization cross section of four highest molecular orbitals of urocanic acid. Dashed line denotes the 70 eV kinetic energy used in EI. Blue line denotes the highest occupied molecular orbital, which has the largest ionization cross section.

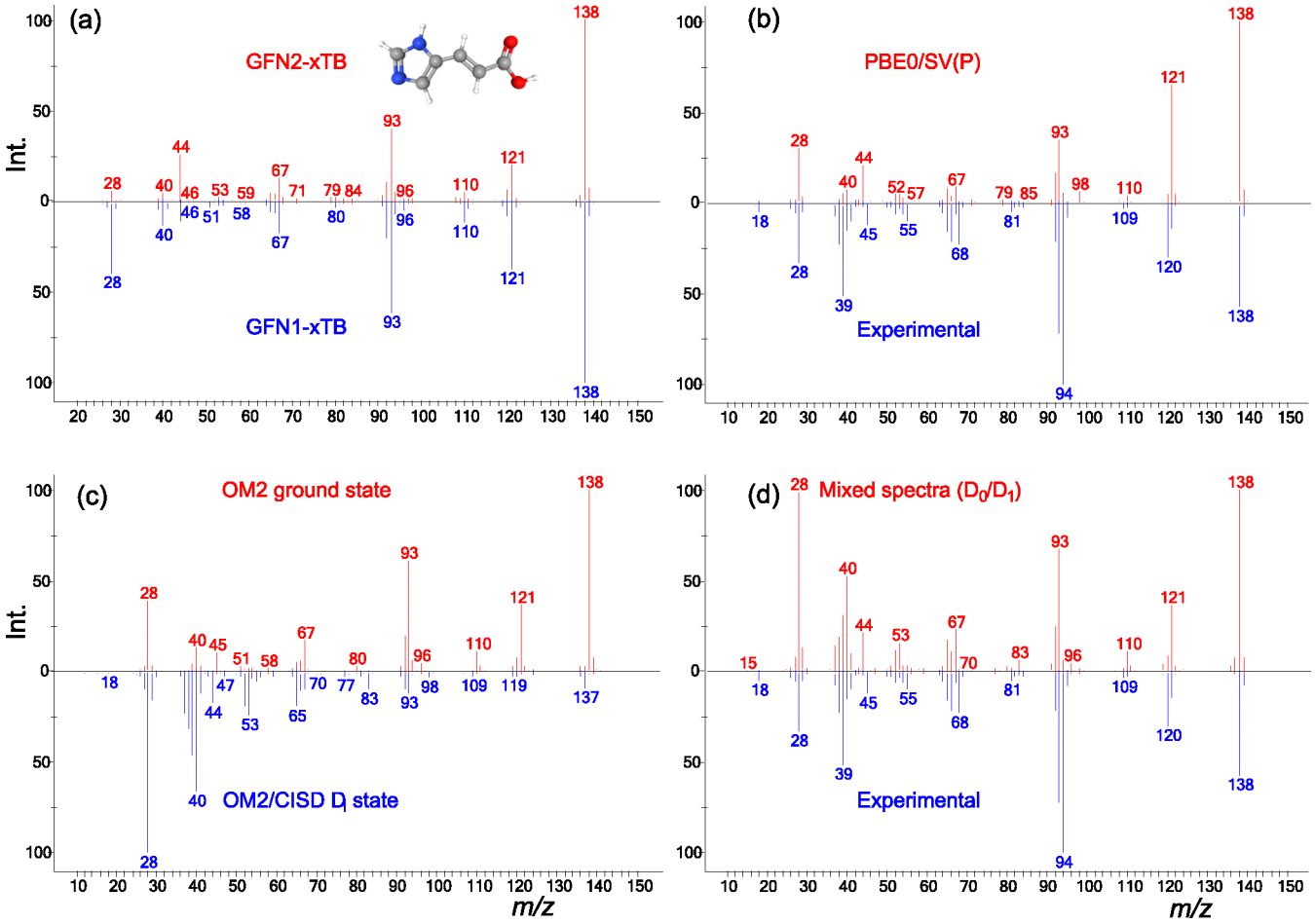


Figure 2.

In silico spectra of urocanic acid using different simulation methods. (a) Semiempirical level GFN2-xTB vs GFN1-xTB. (b) DFT level PBE0/SV(P) vs the experimental spectrum from the NIST17 library. (c) Semiempirical level OM2 at ground state vs the semiempirical configuration interaction level OM2/CISD (first excited state). (d) Mixed spectrum of D0 (OM2) and D1 (OM2/CISD) simulations vs the NIST17 experimental spectrum.

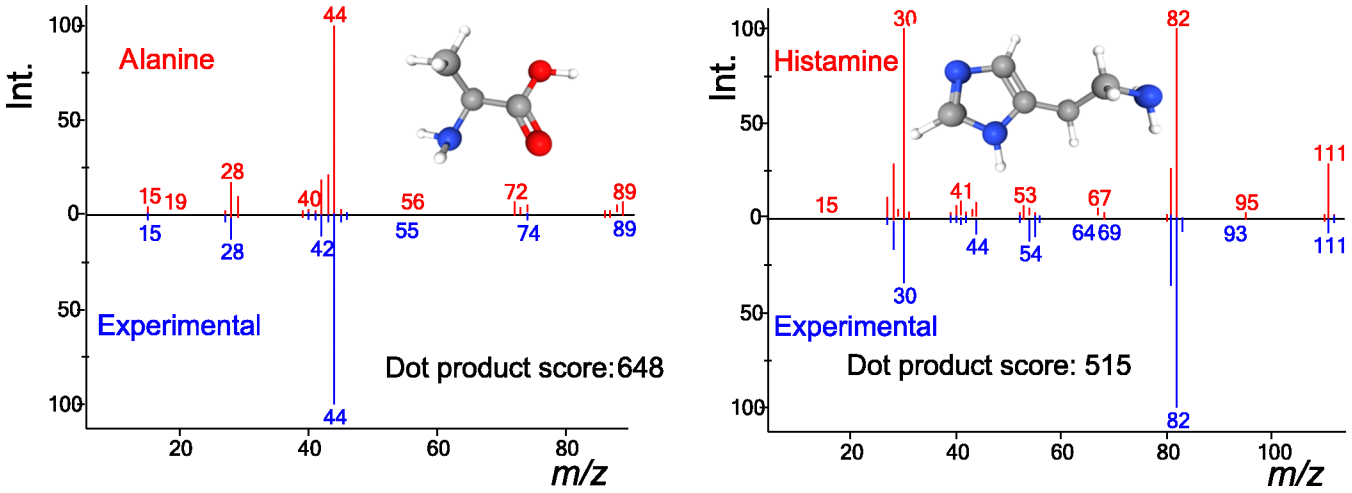


Figure 3.

Head-to-tail spectra of alanine and histamine. Top: mixed spectra obtained by GFN1-xTB/OM2/CISD. Bottom: experimental spectra from the NIST17 library.

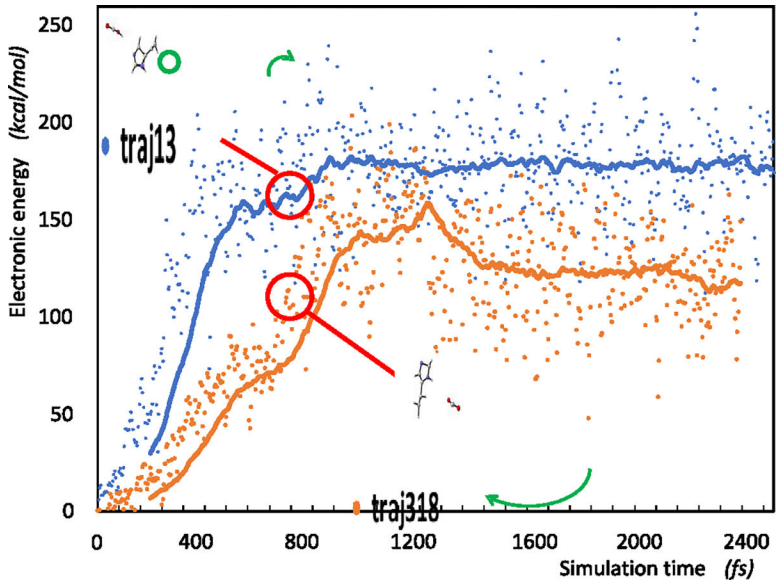


Figure 4.

Electronic energy change of the whole system according to the simulation time of urocanic acid: (1) blue, trajectory 13: $C_5N_2H_5^+$ and neutral $CO_2H^\bullet$; (2) orange, trajectory 318: $C_5H_2H_6^{+\bullet}$ and neutral CO_2; (3) all trend lines are moving average per 100 fs; and (4) inset: the structures at around 780 fs.

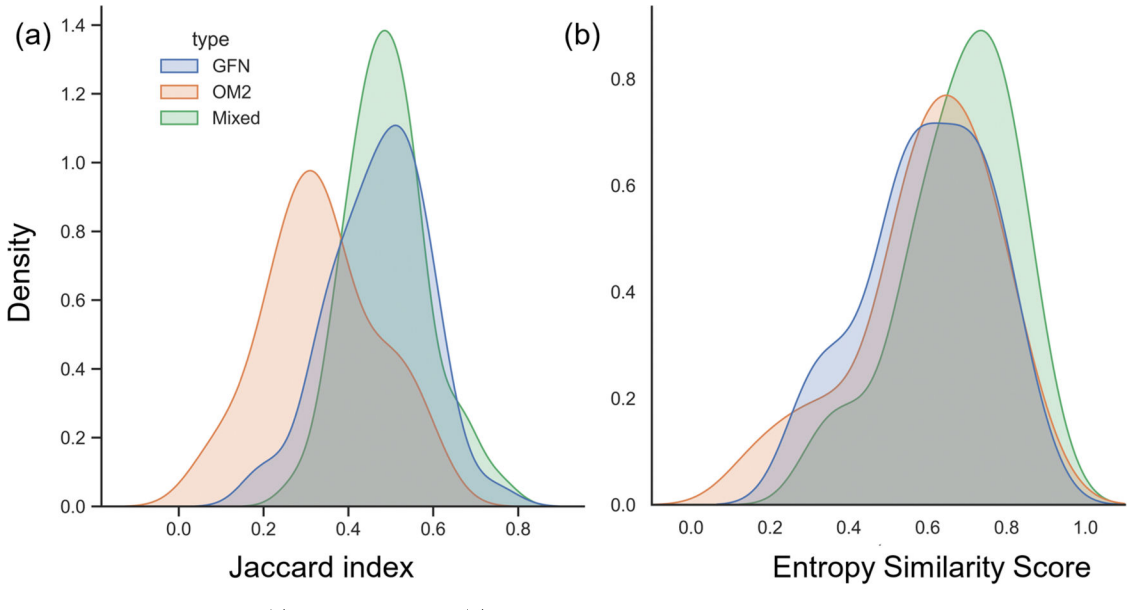


Figure 5.

Kernel density estimate plot of (a) Jaccard index and (b) entropy similarity score of 48 small molecules. Blue: ground-state spectra predicted by GFN1-xTB. Orange: ground-state spectra simulated by the OM2 method. Green: mixed spectra of D_0 and D_1 using the GFN1-xTB/OM2/CISD method.

Table 1.

Parameters Used for Simulating Ground State/Excited State Ratios[a]

name	fraction D_0	fraction D_1	unoccupied MO	occupied MO	active electron
alanine	0.58	0.42	2	4	7
glutamic acid	0.58	0.42	2	7	13
histidine	0.60	0.40	3	6	11
histamine	0.60	0.40	2	5	9
carnosine	0.59	0.41	4	11	21
1-methyl-histidine	0.59	0.41	3	8	15
urocanic acid	0.62	0.38	4	6	11

[a]The fraction of D_0 and D_1 was decided by the ionization cross section. The sum of unoccupied and occupied molecular orbital is the size of active space. The number of active electrons is calculated as the product of electrons in the molecular orbitals −1.