

Supporting Information:

Towards calculating electron impact mass spectra via automated transition state search - QCxMS2

Johannes Gorges and Stefan Grimme**

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

E-mail: grimme@thch.uni-bonn.de

S1 Statistical error measures

Statistical measures for a set of data points x_1, \dots, x_n with references r_1, \dots, r_n are:

- Average:

$$\bar{x} = \frac{1}{n} \sum_i^n x_i$$

- Mean deviation(MD):

$$MD = \frac{1}{n} \sum_i^n (x_i - r_i)$$

- Mean absolute deviation(MAD):

$$MAD = \frac{1}{n} \sum_i^n |x_i - r_i|$$

- relative mean absolute deviation (relMAD):

$$MAD = \frac{1}{n} \sum_i^n \frac{|x_i - r_i|}{r_i} \cdot 100\%$$

S2 Alternative Description via simplified RRKM equation

The usual way of describing reaction rates in mass spectrometry is based on the RRKM theory. In initial studies, we found the RRKM equations to be impractical to be used in an automated workflow, as the density of states has to be computed, which is computationally unfeasible due to the necessary highly accurate vibrational frequencies and very unstable for reactions with no clearly defined transition state. As a more robust choice, we employed the classical approximation of RRKM, in which the system consists only of harmonic oscillators.^{S1} The rate constant is given by

$$k(E) = \nu \cdot e^{-(n_{\text{vib}}-1)E_a/E}, \tag{1}$$

with ν being the internal reaction coordinate mode at the transition state, n_{vib} is the number of vibrational modes, E_a is the activation energy including the zero-point vibrational energy, and E the internal energy of the ion, given by the energy distribution $P(E)$.

However, for barrierless reactions with no clear transition state, ν is poorly defined and the equation leads to bad results for the rate in these cases. In the QCxMS2 workflow, just

the imaginary mode with the highest absolute value is used here. As these reactions often occur in the mass spectrometer, we decided to use the more robust Eyring equation (in the manuscript) instead. In Figure S1 spectra computed with ω B97X-3c//GFN2-xTB with the Eyring formalism and the RRKM formalism are shown for ethyl propyl ether.

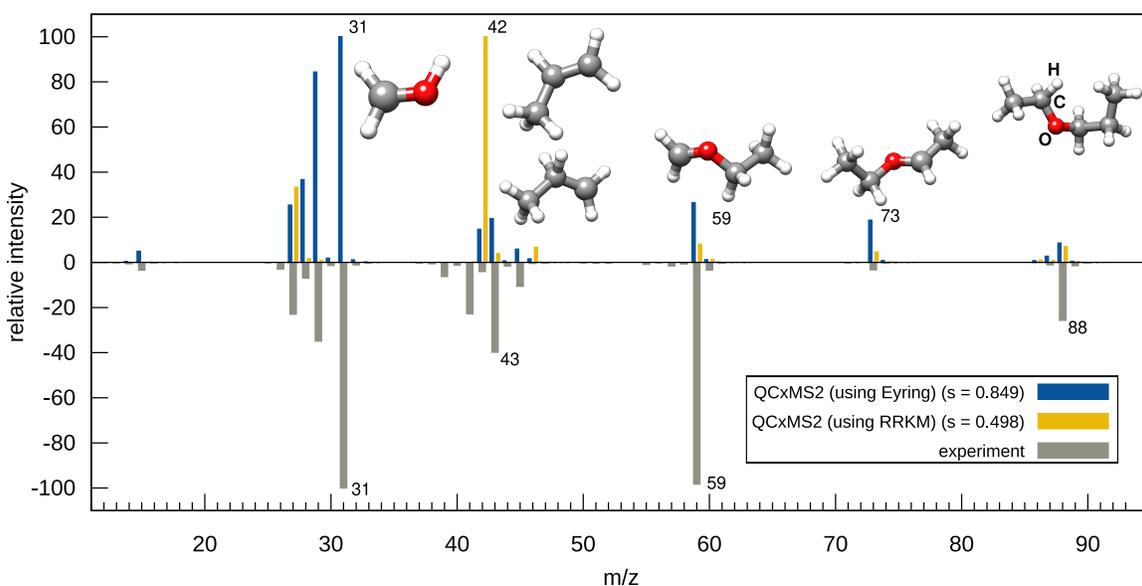


Figure S1: Mass spectrum of ethyl propyl ether computed with QCxMS2 at the ω B97X-3c//GFN2-xTB level with an intensity threshold for subsequent fragmentations of 5 % using the RRKM and Eyring equation for the calculation of the rate constants. The inverted experimental spectrum is shown for comparison and entropy similarity matching scores s are given.

Notably, the spectrum of ethyl propyl ether computed with RRKM theory is much worse than the one computed with the Eyring transition state theory, since the fragments with m/z values of 73, 43, 59 and its successor ion with m/z 31 are too low in intensity, whereas the peak at m/z 42 is too high. This is also reflected in the low matching score of 0.498 compared to 0.849. The respective rate constants are depicted in Figure S2 for both spectra. Values for $k(E)$ obtained with the Eyring and the RRKM approach are shown for the fragmentation reaction to the fragment of the signal at m/z 42. The internal energy, which is for this first fragmentation reaction just the impact excess energy (IEE) is also plotted. Here, a clear converged transition state was found with an imaginary mode and Eyring and RRKM give

very similar results across the complete sampled energy distribution. Note that absolute rate constants are not of interest here, and only relative rate constants are important for the correct intensities, and therefore the small deviations between both theories are not significant here. The rate constants for the formation of the other important fragments are shown in Figure b) (Eyring) and c) (RRKM), respectively. The fragments at m/z 59 and m/z 73 stem from C-C dissociations and m/z 43 stems from a C-O bond cleavage, which all have very late transition states and hence imaginary modes with a very low absolute value were found. Thus, the rate constants computed with RRKM are much lower compared to the Eyring approach leading to the worse agreement of the spectrum with the experiment. Therefore, using the Eyring equation is generally the more robust choice within the QCxMS2 workflow and the given computational constraints and predicts better spectra. Computing the rate constants via RRKM with equation 1 can still be used in QCxMS2 but is not recommended due to its instability.

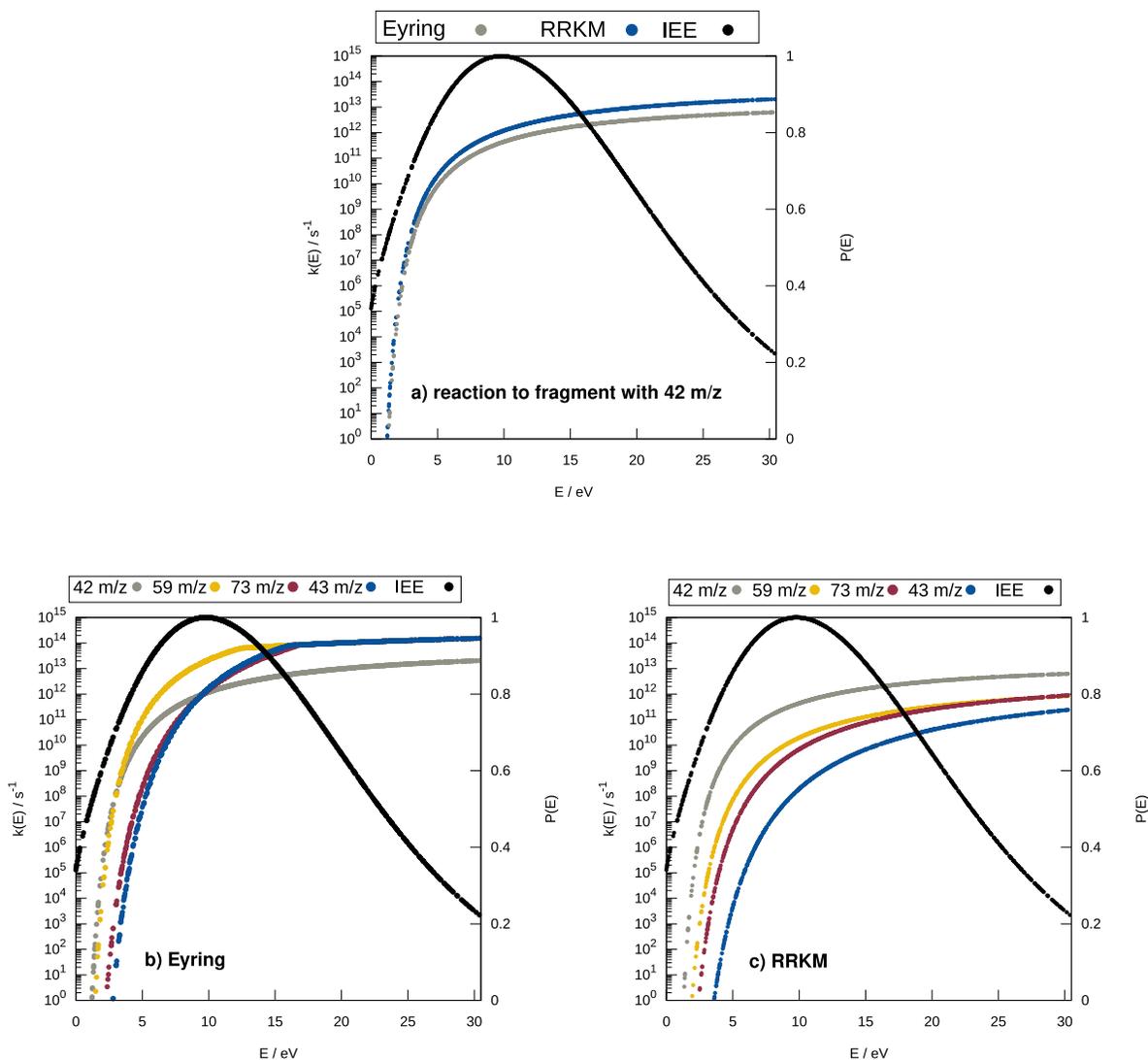


Figure S2: Rate constants $k(E)$ computed for the fragmentation reactions of ethyl propyl ether for the energy distribution sampled with 1000 points in the Monte Carlo simulation (default is 100000). The energy distribution $P(E)$ for the IEE is also plotted, to see the relevant range of internal energies E for the reaction. a): RRKM versus Eyring for reaction to 42 m/z, b): Eyring for important reactions, and c): RRKM for important reactions. The stepwise course with Eyring comes due to the fact that the thermal contributions for every energy are pre-computed for small energy windows to save computation time. 200 steps to sample the energy distribution are by default taken, the error of this is negligible. The $k(E)$ curves in Eyring for 59 m/z and 73 m/z 43 m/z strive toward the limit $k_B T/h$ for RRKM the limit is the imaginary frequency ν .

S3 Test set

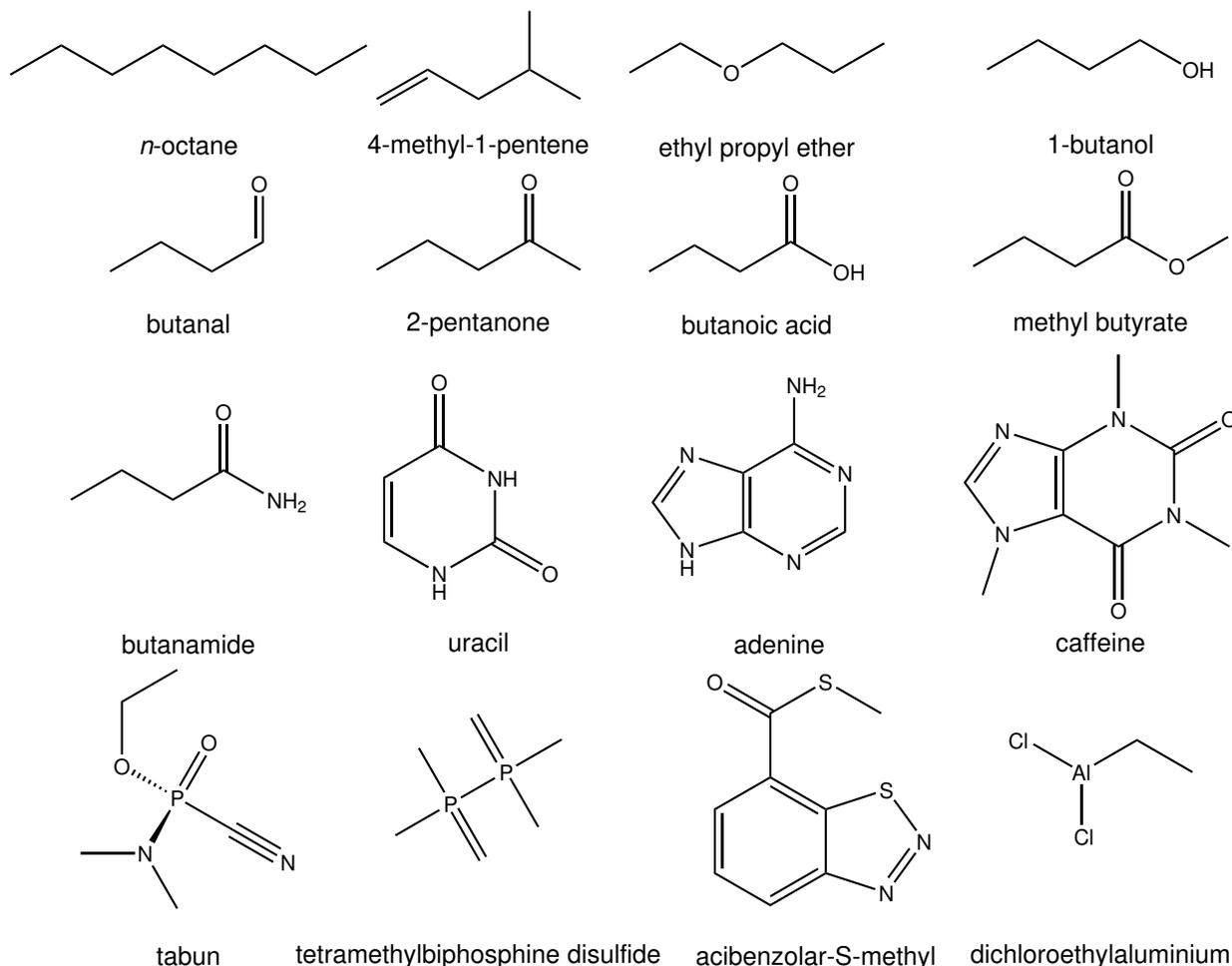


Figure S3: Lewis formulas for the test set used in this study.

S4 Variation of the impact excess energy (IEE)

The impact excess energy per atom (*ieeatm*) is a variable input parameter to fit the energy distribution to the experimental conditions. In Figure S4 spectra obtained with QCxMS2 at the mixed ω B97X-3c//GFN2-xTB level with 5 % intensity threshold are shown for butanal with the IEE set to 0.6 eV/atom, 0.8 eV/atom, 1.0 eV/atom, and 1.2 eV/atom, respectively. As expected, the intensity of the molecular peak decreases with increasing the IEE. Notably, the experimentally measured peak at 15 m/z is only predicted for the two higher energies.

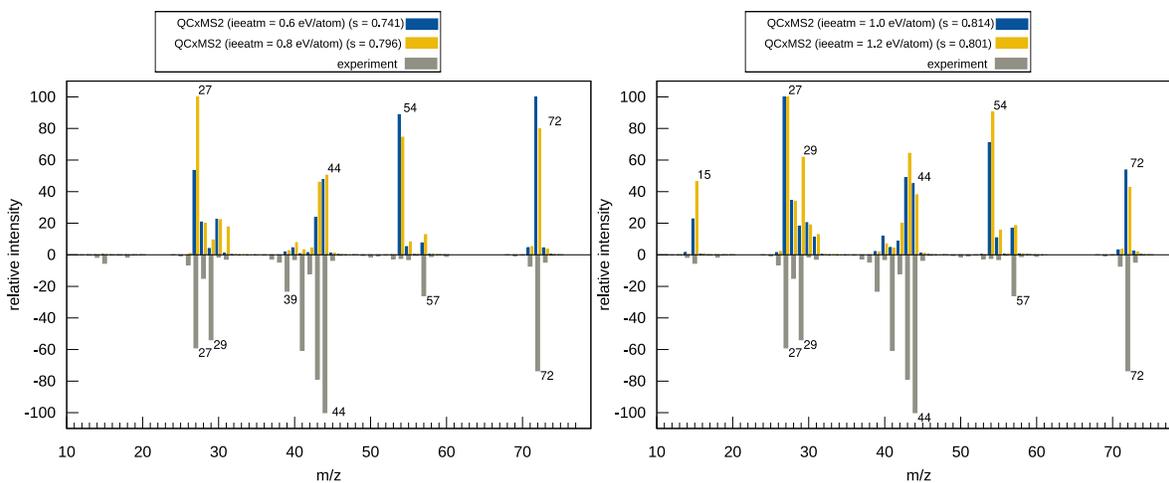


Figure S4: Spectra of butanal computed at the ω B97X-3c//GFN2-xTB level of theory with 5% percent intensity threshold for different *ieatm* values.

Here, the best score is obtained with 1.0 eV/atom. Matching scores for the entire test set obtained with QCxMS2 at the mixed ω B97X-3c//GFN2-xTB level with 5 % intensity threshold for subsequent calculations are given in Table S1 with *ieatm* set to 0.6, 0.8, and 1.0.

Table S1: Entropy similarity scores between experimental spectra and theoretical spectra computed with QCxMS2 at the ω B97X-3c//GFN2-xTB, level for all compounds of the test set with `ieeatm` set to 0.6 eV/atom, 0.8 eV/atom, 1.0 eV/atom, respectively. The intensity threshold for subsequent fragmentations was set to 5 %

compound	0.6	0.8	1.0
n-octane	0.734	0.717	0.729
4-methyl-1-pentene	0.667	0.718	0.686
ethyl propyl ether	0.874	0.849	0.832
1-butanol	0.762	0.751	0.777
butanal	0.741	0.796	0.814
2-pentanone	0.724	0.697	0.628
butanoic acid	0.682	0.692	0.616
methyl butyrate	0.643	0.665	0.646
butanamide	0.567	0.561	0.601
uracil	0.481	0.477	0.487
adenine	0.772	0.786	0.790
caffeine	0.571	0.592	0.592
tabun	0.441	0.442	0.489
tetramethylbiphosphine disulfide	0.780	0.757	0.760
acibenzolar-S-methyl	0.572	0.611	0.548
dichloroethylaluminium	0.670	0.709	0.665
average	0.668	0.676	0.666

Overall, the energy settings fitting best to the experiment are very system-dependent. In 4 cases, 0.6 eV/atom, in 6 cases 0.8 eV/atom, and in 6 cases 1.0 eV/atom shows the best matching. On average, the best matching scores with 0.676 were achieved for 0.8 eV/atom, which is therefore set as default for the best agreement with standard 70 eV electron ionization mass spectra.

S5 Internal energy scaling for H-dissociations

Figure S5 shows the spectrum of *n*-octane and butanoic acid computed at the ω B97X-3c// ω B97X-3c level. For *n*-octane, H-dissociation from the molecular peak is not observed in the experiment but predicted by QCxMS2. Although errors in the computed barriers due to the chosen QC method may be present—especially since these reactions are susceptible to

electron self-interaction error (SIE)—such errors should not be severe at the range-separated hybrid level. Instead, they likely reflect a systematic overestimation of the relative rate constants for H-dissociations. To partially account for this effect, we introduced a scaling factor for the internal energy of these reactions, addressing in part the inhomogeneous energy distribution that may occur briefly (on the order of several picoseconds)^{S2} after ionization. The scaling factor effectively scales down the temperature at which the rate constant of the reaction is computed. In tests on several molecules, it turned out that a factor of 0.5 works best for this. Figure S5 shows the spectra of *n*-octane and butanoic acid computed with a scaling factor of 0.5 and without scaling. For *n*-octane the changes in the spectrum by applying the scaling are very significant, especially for the peaks at m/z 113, 111, 109, and 107. This hydrogen dissociation cascade comes from the loss of H, and subsequent loss of H₂ respectively. This falsely predicted behavior is corrected by the scaling factor resulting in a much better matching score of 0.797 compared to 0.608. For butanoic acid, the changes are smaller than for *n*-octane. Here, the peak at M-1 is measured in low intensity of $\approx 2\%$ in the experiment and overestimated in the computed spectrum with scaling. In the computed spectrum with scaling it is slightly underestimated but still predicted and the matching score is virtually similar (0.753 versus 0.756). Matching scores computed at the three different theory levels for the entire test set with and without scaling are given for the entire test set in Table S2.

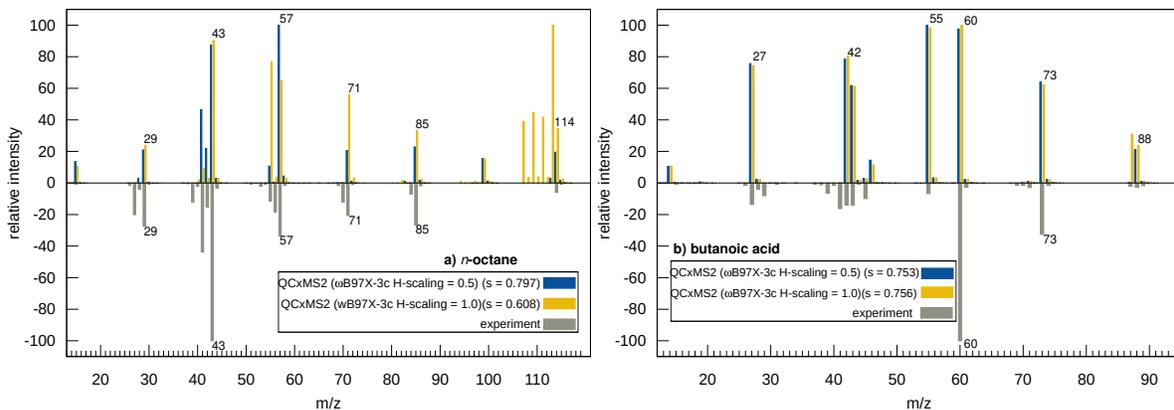


Figure S5: Mass spectrum of *n*-octane and butanoic acid computed with QCxMS2 at the ω B97X-3c level with scaling of the internal energy for H-dissociations and without scaling with an intensity threshold for subsequent fragmentations of 5 %. The inverted experimental spectrum is shown for comparison

Table S2: Entropy similarity scores between experimental spectra and theoretical spectra computed with QCxMS2 at the GFN2-xTB//GFN2-xTB, ω B97X-3c//GFN2-xTB, ω B97X-3c// ω B97X-3c levels for all compounds of the test set with scaling factor of 0.5 applied to H-dissociations and without scaling. The intensity threshold for subsequent fragmentations was set to 5 %

theory level scaling of IEE	GFN2-xTB		mixed		ω B97X-3c	
	1	0.5	1	0.5	1	0.5
<i>n</i> -octane	0.704	0.669	0.618	0.717	0.608	0.797
4-methyl-1-pentene	0.707	0.640	0.710	0.718	0.751	0.795
ethyl propyl ether	0.782	0.762	0.836	0.849	0.792	0.793
1-butanol	0.746	0.719	0.807	0.751	0.681	0.638
butanal	0.786	0.832	0.780	0.796	0.778	0.810
2-pentanone	0.742	0.734	0.714	0.697	0.824	0.824
butanoic acid	0.691	0.651	0.553	0.692	0.756	0.753
methyl butyrate	0.651	0.654	0.695	0.665	0.744	0.732
butanamide	0.403	0.450	0.396	0.561	0.549	0.605
uracil	0.644	0.617	0.524	0.477	0.488	0.658
adenine	0.472	0.754	0.608	0.786	0.546	0.629
caffeine	0.446	0.469	0.616	0.592	0.507	0.474
tabun	0.559	0.577	0.418	0.442	0.613	0.632
tetramethylbi- phosphine disulfide	0.689	0.695	0.740	0.757	0.752	0.752
acibenzolar-S-methyl	0.377	0.388	0.598	0.611	0.492	0.492
dichloroethylalumnium	0.790	0.680	0.588	0.709	0.709	0.677
average	0.637	0.643	0.638	0.676	0.662	0.691

For the GFN2-xTB//GFN2-xTB spectra in some cases the matching score is larger without scaling applied and on average no significant improvement is visible (0.637 versus 0.643). This is probably due to the limited accuracy of the barriers at this level of theory and possible loss of error compensation. However, for the ω B97X-3c//GFN2-xTB and ω B97X-3c// ω B97X-3c spectra, the matching scores are on average significantly improved from 0.638 to 0.676 and from 0.662 to 0.691, respectively. Therefore, all spectra shown in this work are computed with the default scaling factor of 0.5 applied.

S6 Investigation of the precision of the QCxMS2 workflow

Fragments from subsequent fragmentation reactions are obtained by applying the QCxMS2 workflow iteratively on generated fragments. To keep the total number of generated computational costs reasonable, only important fragments for the spectrum, i.e., with a significant intensity are further fragmented. A lower intensity threshold leads to more precision and - provided the QCxMS2 methodology works well on the molecule - to a more accurate spectrum. Entropy similarity matching scores to the experiment obtained with QCxMS2 at the ω B97X-3c//GFN2-xTB level computed spectra of 2-pentanone and caffeine are shown in Figure S6.

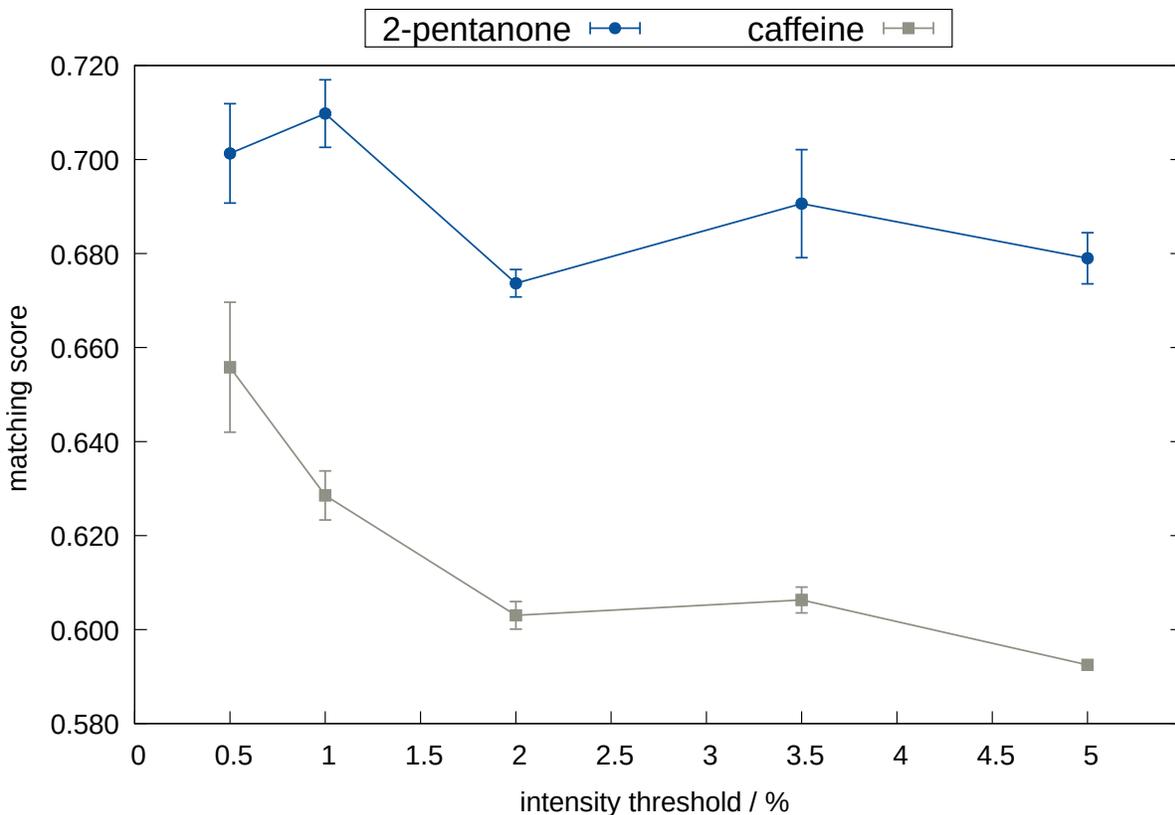


Figure S6: Entropy similarity matching scores with respect to the experimental values of QCxMS2 spectra computed at the ω B97X-3c//GFN2-xTB level at different intensity thresholds for subsequent fragmentation averaged over three independent runs. The error bars are the respective standard deviations.

As the QCxMS2 workflow is not completely deterministic due to the automated barrier calculation workflow which can lead to slightly deferring barriers, the computed intensity can vary between QCxMS2 calculations. Deviations due to numerical errors in the kinetic modeling of the Monte Carlo sampling also occur but are negligible at the default sample size of 10^5 points. Therefore, the matching scores shown in in Figure S6 are averaged over three independent runs and the respective standard deviations are also shown. Generally, the matching score increases with decreasing intensity threshold, which is a good sign for the QCxMS2 methodology. The decrease from 1 % to 0.5 % and from 3.5 % to 2 % for 2-pentanone is most likely due to inaccuracies in the employed level of theory. As expected the increase in accuracy with increasing precision is for caffeine more pronounced than for

2-pentanone, as the reaction network is larger and more peaks with lower intensity are measured for caffeine. However, the standard deviation obtained from three different runs also increases because more reaction barriers have to be calculated and the uncertainty increases with the size of the reaction network. This trend is also more pronounced for caffeine than for 2-pentanone. Also, the computational costs increase as shown in Figure S7. This is more pronounced for caffeine both for fragments and timings. At the 0.5 intensity threshold, the calculation produces 2218 fragments and isomer species and takes about 43 hours on 32 CPU cores. At 1% intensity threshold, only half of the computation time is needed.

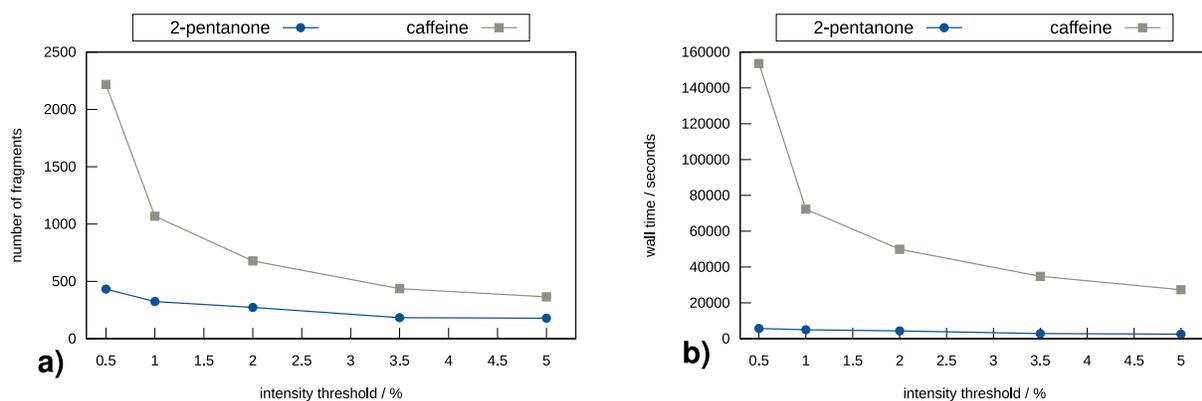


Figure S7: a): Number of computed fragments and isomers of 2-pentanone and caffeine with QCxMS2 at the ω B97X-3c//GFN2-xTB level at different intensity thresholds for subsequent fragmentation. b): Wall times on 32 Intel[®] Xeon[®] "Sapphire Rapids" v4 @ 2.10 GHz CPU.

Therefore, we set the intensity threshold by default to 1% as a compromise of precisions, reproducibility (reasonable standard deviation), and computational costs. Investigations on other parameters in the SI are performed with a 5% threshold since the trends are similar at much lower computation times.

S7 Variation of the rotor cutoff in the modified rigid

For the computation of the free energies of activation, the thermal contributions are computed via a SPH calculation at the GFN2-xTB level with the modified rigid-rotor harmonic-

oscillator (mRRHO) approximation.^{S3} Especially for "late" transition states, i.e., transition states, where the fragments are already dissociated from each other, low-lying frequencies may occur. To account for the high temperatures assumed, different values for the rotor cutoff which is by default 50 cm⁻¹ were tested. Table S3 shows entropy similarity scores of the spectra computed at the ω B97X-3c//GFN2-xTB level with 5 % intensity threshold with rotor cutoffs of 50 cm⁻¹, 100 cm⁻¹, 150 cm⁻¹, and 200 cm⁻¹.

Table S3: Entropy similarity scores of the spectra computed at the ω B97X-3c//GFN2-xTB level with 5 % intensity threshold with rotor cutoffs of 50 cm⁻¹, 100 cm⁻¹, 150 cm⁻¹, and 200 cm⁻¹ for the entire test set

rotor-cutoff	50	100	150	200
n-octane	0.767	0.724	0.717	0.707
4-methyl-1-pentene	0.694	0.688	0.718	0.684
ethyl propyl ether	0.825	0.820	0.849	0.822
1-butanol	0.763	0.779	0.751	0.764
butanal	0.795	0.762	0.796	0.799
2-pentanone	0.685	0.705	0.697	0.702
butanoic acid	0.608	0.652	0.692	0.743
methyl butyrate	0.667	0.649	0.665	0.652
butanamide	0.672	0.559	0.561	0.562
uracil	0.524	0.519	0.477	0.463
adenine	0.779	0.778	0.786	0.781
caffeine	0.648	0.585	0.592	0.577
tabun	0.460	0.460	0.442	0.436
tetramethylbiphosphine disulfide	0.714	0.765	0.757	0.760
acibenzolar-S-methyl	0.561	0.587	0.611	0.606
dichloroethylalumnium	0.625	0.647	0.709	0.686
average	0.674	0.668	0.676	0.671

Overall the differences are not that large. However, the cutoff of 150 cm⁻¹ showed the best average matching score of 0.676 for the entire set and was therefore chosen as default.

S8 Inclusion of excited states via TD-DFT

To investigate the effect of excited states on the reaction barriers and the resulting spectrum, time-dependent DFT calculations at the ω B97X-3c level were performed for each

species with ORCA for the spectra computed at the ω B97X-3c//GFN2-xTB level with 5 % intensity threshold. Therefore, the first ten electronically excited states were computed and the respective energies Boltzmann averaged at the given average fragment temperature T_{av} according to

$$p_i = \frac{e^{-\beta E_i}}{\sum_i e^{-\beta E_i}}, \quad (2)$$

where $\beta = \frac{1}{k_B T}$, and E_i is the electronic energy of the excited state. In Table S4 mean deviations (MDs), mean absolute deviations (MADs) between barriers computed at the ground state and barriers computed at the Boltzmann averaged energies are given for the QCxMS2 calculations for the complete test set, done with ω B97X-3c//GFN2-xTB at an intensity threshold for subsequent fragmentation of 5%. Only barriers that resulted in peaks with a relative intensity of at least 1 % are shown to exclude very high-lying, i.e., irrelevant barriers. Overall, most of the excited states are very in energy and hardly populated at the estimated temperatures and the differences for the barriers are very small. On average, the barriers are lowered by only 0.021 eV mol⁻¹ and the averaged MAD is 0.031 eV mol⁻¹ which corresponds to a relative deviation of only 1 %. Large differences in the barrier only occur, if the wrong ground state is selected which leads to negative relative energies of the supposedly excited states, which reduces the barrier. However, this results in barriers that are artificially too low, as TD-DFT is not suitable for such cases and these states are therefore not considered. This shows, that even the very simple approach chosen here to describe excited states is problematic for high-throughput calculations in the QCxMS2 workflow. As expected from the small differences in the barriers, the matching scores of the computed spectra, shown in Table S5, exhibit only small differences by inclusion of excited states. Large differences are only visible for *n*-octane and acibenzolar-S-methyl, for which the spectrum with the excited state calculation shows worse agreement. This is due to self-consistent field (SCF) convergence problems in the TD-DFT calculations for some barriers. As the excited state’s approach is generally less robust and does not bring any significant improvement, it was therefore not set as the default in QCxMS2.

Table S4: Number of barriers n_{barriers} , mean deviation (MD), mean absolute deviation (MAD), maximum deviation MAXD in eV/mol, and relative mean absolute deviation in % for computed barriers in a QCxMS2 calculation at the ω B97X-3c//GFN2-xTB for the test set

compound	n_{barriers}	MD	MAD	relMAD/%	MAXD
<i>n</i> -octane	21	0.000	0.036	1.649	0.127
4-methyl-1-pentene	20	-0.025	0.025	0.679	0.001
ethyl propyl ether	12	-0.013	0.014	0.454	0.003
1-butanol	13	-0.027	0.031	4.051	0.025
butanal	14	-0.051	0.051	1.564	-0.001
2-pentanone	18	-0.017	0.018	0.490	0.005
butanoic acid	17	-0.022	0.023	0.607	0.002
methyl butyrate	20	-0.014	0.015	0.779	0.002
butanamide	21	-0.008	0.027	0.184	0.043
uracil	8	-0.020	0.044	1.162	0.029
adenine	6	-0.019	0.037	0.612	0.054
caffeine	16	-0.020	0.037	0.921	0.062
tabun	22	-0.019	0.020	0.633	0.006
tetramethylbi- phosphine disulfide	14	-0.042	0.046	1.803	0.017
acibenzolar-S-methyl	17	-0.005	0.038	1.231	0.066
dichloroethylalumnium	6	-0.042	0.042	1.294	-0.006
average		-0.021	0.031	1.132	-

Table S5: Entropy similarity scores of the spectra computed at the ω B97X-3c//GFN2-xTB level with 5 % intensity threshold with and without inclusion of excited states for the entire test set

compound	default	inclusion of excited states
n-octane	0.717	0.663
4-methyl-1-pentene	0.718	0.720
ethyl propyl ether	0.849	0.853
1-butanol	0.751	0.786
butanal	0.796	0.795
2-pentanone	0.697	0.694
butanoic acid	0.692	0.694
methyl butyrate	0.665	0.681
butanamide	0.561	0.601
uracil	0.477	0.513
adenine	0.786	0.783
caffeine	0.592	0.594
tabun	0.442	0.436
tetramethylbiphosphine disulfide	0.757	0.742
acibenzolar-S-methyl	0.611	0.484
dichloroethylalumnium	0.709	0.678
average	0.676	0.670

S9 Kinetic energy release

The kinetic energy release (KER) upon fragmentation stems from two sources, namely the excess energy in the transition state and energy of the reverse reaction E_{ar} :

$$KER = KER_{\text{ex}} + KER_{\text{ar}}. \quad (3)$$

The first term is computed by:

$$KER_{\text{ex}} = \frac{E_{\text{ex}}}{0.44 \cdot s}, \quad (4)$$

with s being the number of vibrational degrees of freedom of the transition state, and empirically found parameters $a = 0.44$.^{S4} The second one is approximated by

$$KER_{\text{ar}} = 0.33 \cdot E_{\text{ar}}, \quad (5)$$

with $b = 0.33$ from experimental studies.^{S5}

S10 Effect of level of theory for calculation of thermal contributions

Thermal contributions are computed at the GFN2-xTB level of theory. This may introduce some errors. Was investigated for methylbutyrate, for which SPH calculations at the ω B97X-3c level were additionally computed using a development version of xTB 6.7.1 as driver for Turbomole version 7.7.1. Since SPH calculations were performed at the same level of theory as the geometry optimizations, the threshold for inverting imaginary frequencies was set to 20 cm^{-1} instead of the default of 100 cm^{-1} . Figure S8 shows the spectrum of methyl butyrate computed at the ω B97X-3c level of theory with GFN2-xTB SPH calculations and ω B97X-3c SPH calculations for the thermal contributions. A large difference is visible for the peak at m/z 74 which stems from a McLafferty-type rearrangement. In the GFN2-

xTB SPH spectrum, this peak is much too low in intensity compared to the experiment. This is corrected in the spectrum computed with thermal contributions from ω B97X-3c SPH calculations. Overall a better agreement is visible, which is also reflected in a slight improvement in the entropy similarity matching score (0.742 versus 0.762). However, as

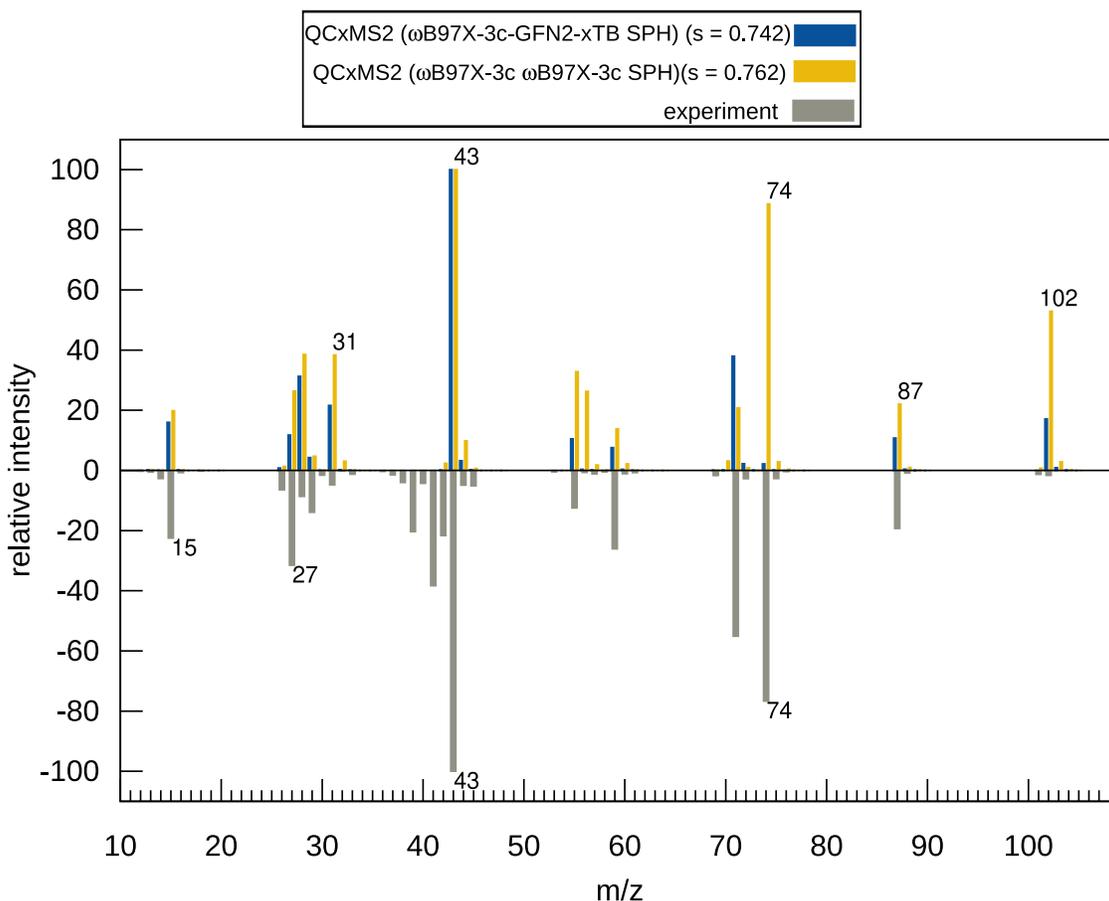


Figure S8: Spectrum of methyl butyrate computed at the ω B97X-3c level of theory with GFN2-xTB SPH calculations and ω B97X-3c SPH calculations for the thermal contributions.

performing numerical SPH calculations at the DFT level is very expensive, this was only tested for this small molecule.

S11 Additional computational details

Calculations, for which the SCF does not converge are restarted with Fermi Smearing at elevated electronic temperature. For GFN2-xTB, the electronic temperature was set to

5000 K, while for the range-separated hybrid ω B97X-3c, it was set to 15000 K.

S12 Results with cosine similarity matching score

The cosine similarity matching score is also often used. An often-used heuristic cutoff value of >0.7 ^{S6} is used for a positive match between theory and experiment and should be aimed for. Cosine similarity scores between experimental spectra and theoretical spectra computed with QCxMS at the GFN2-xTB level and QCxMS2 with the three method combinations for all 16 compounds of the test set are given in Table S6.

Table S6: Cosine similarity scores between experimental spectra and theoretical spectra computed with QCxMS2 at the GFN2-xTB//GFN2-xTB, "composite" ω B97X-3c//GFN2-xTB, and ω B97X-3c// ω B97X-3c levels for all compounds of the test set. Scores for spectra computed with QCxMS at the GFN2-xTB level are also given for comparison

level	GFN2-xTB	composite	ω B97X-3c	QCxMS
<i>n</i> -octane	0.450	0.506	0.869	0.742
4-methyl-1-pentene	0.708	0.517	0.858	0.757
ethyl propyl ether	0.502	0.813	0.622	0.404
1-butanol	0.579	0.630	0.535	0.487
butanal	0.870	0.670	0.740	0.657
2-pentanone	0.888	0.819	0.942	0.926
butanoic acid	0.418	0.545	0.715	0.142
methyl butyrate	0.468	0.646	0.765	0.563
butanamide	0.275	0.523	0.580	0.522
uracil	0.376	0.243	0.590	0.756
adenine	0.912	0.918	0.907	0.958
caffeine	0.383	0.714	0.626	0.537
tabun	0.634	0.376	0.575	0.545
tetramethylbi-phosphine disulfide	0.661	0.950	0.928	0.037
acibenzolar-S-methyl	0.214	0.491	0.370	0.014
dichloroethyl-aluminium	0.833	0.822	0.759	0.188
average	0.573	0.636	0.711	0.515

Of the GFN2-xTB spectra 5 spectra, of the composite theory level spectra, 6 spectra out of 16 compounds, of the ω B97X-3c spectra, 9, and from the QCxMS spectra 5 out of 16

compounds reach the target accuracy of over 0.7.

S13 Reversibility of isomer reactions

The Monte Carlo simulation for the calculation of the branching ratios is first performed for isomerization reactions, where the reverse reactions are considered to model an isomer equilibrium, and then for fragmentation reactions. This separation is justified because the isomerization reactions, which are mostly hydrogen rearrangements, are often orders of magnitude faster than the fragmentation reactions. To consider reversible isomer reactions, the survival yield is given by

$$I/I_0 = \exp(-k_{\text{for}}(T)t) + (1 - \exp(-k_{\text{for}}(T)t)) \cdot \frac{k_{\text{re}}(T)/k_{\text{for}}(T)}{k_{\text{for}}(T)/k_{\text{re}}(T) + k_{\text{re}}(T)/k_{\text{for}}(T)}, \quad (6)$$

where k_{for} and k_{re} are the rate constants of the forward and reverse reaction, respectively.

S14 Rigid ring system spectra

To assess, whether QCxMS2 has in general a problem with rigid ring systems we computed spectra for benzene and naphthalene as they are the simplest representatives of this class. QCxMS2 spectra computed at the ω B97X-3/GFN2-xTB level and QCxMS spectra computed at the GFN2-xTB level are shown in Figure S11. Notably, both for benzene and naphthalene many peaks are missing in the QCxMS2 spectrum. For the latter essentially no peaks apart from the molecular peak and the M-H peak are predicted. In contrast with QCxMS, many more peaks are predicted and a good agreement with the experiment is achieved.

To investigate if important fragments are missing, we computed the spectra with extended settings in the fragmentation generation and performed additionally 100 optimizations with randomly distorted atom positions. Figure S10 depicts the respective spectra for benzene and naphthalene.

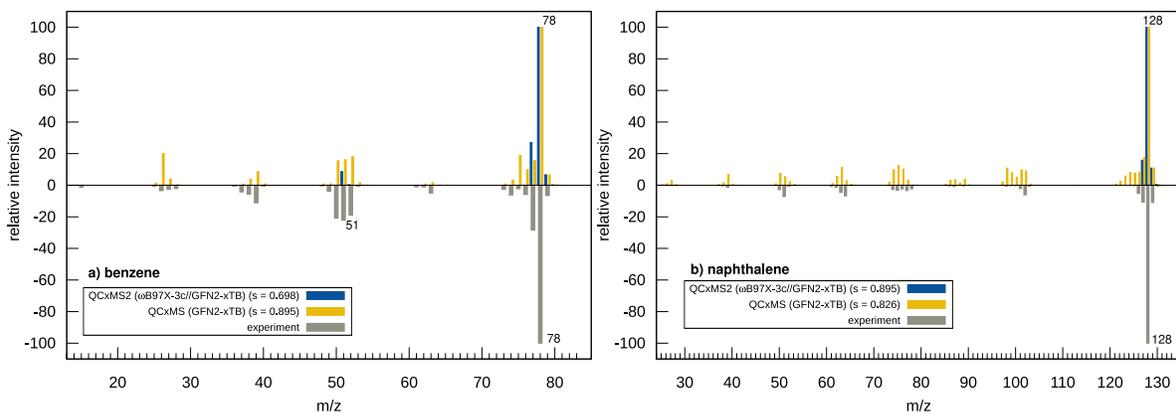


Figure S9: Mass spectra of a) benzene and b) naphthalene computed with QCxMS2 at the ω B97X-3/GFN2-xTB level and QCxMS at the GFN2-xTB level and inverted experimental spectrum for comparison with entropy similarity scores s .

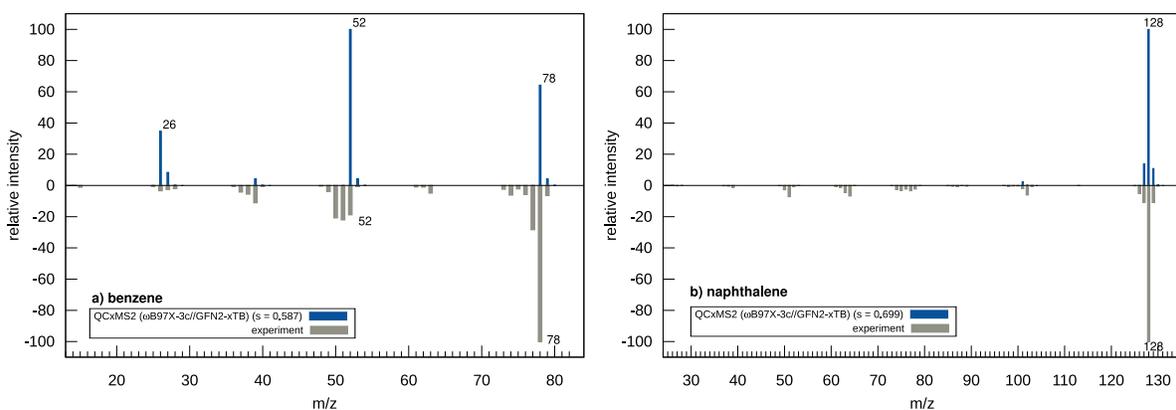


Figure S10: Mass spectra of a) benzene and b) naphthalene computed with QCxMS2 at the ω B97X-3/GFN2-xTB level at 5% intensity threshold and with 100 additional geometry optimizations with randomly distorted atom positions in the fragment generation step applied and inverted experimental spectrum for comparison with entropy similarity scores s .

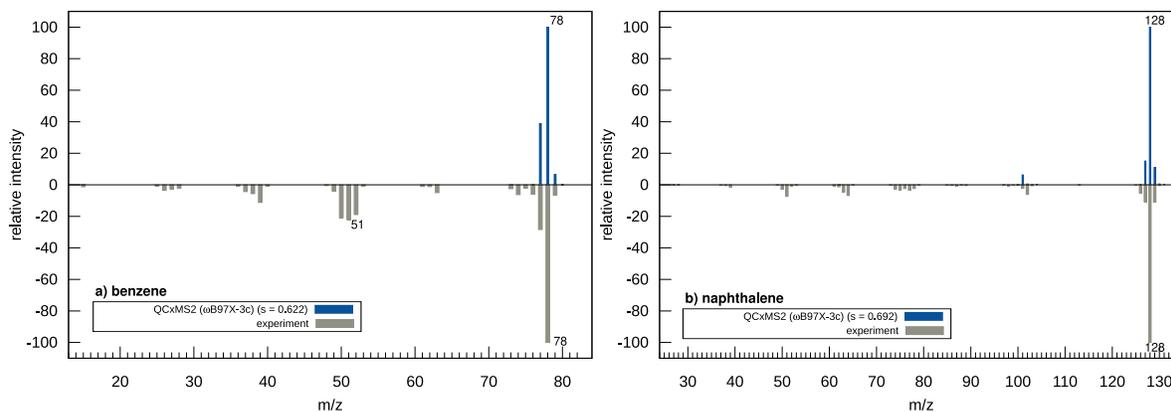


Figure S11: Mass spectra of a) benzene and b) naphthalene computed with QcXMS2 at the ω B97X-3 level and inverted experimental spectrum for comparison with entropy similarity scores s .

S15 Investigation of the errors in the spectrum of uracil

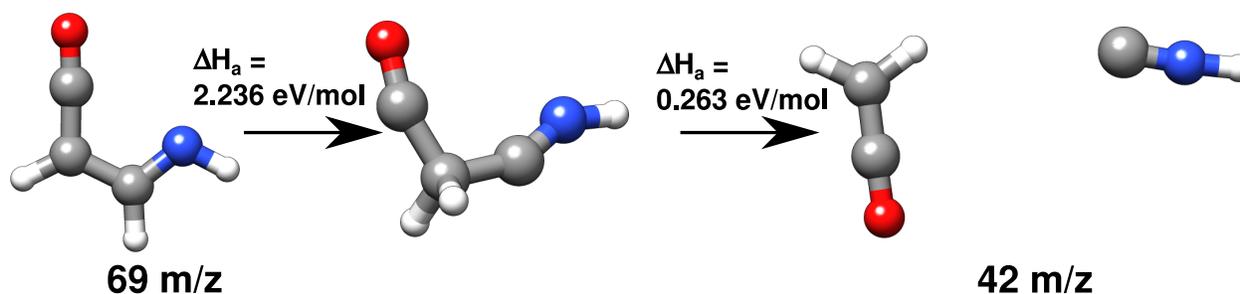


Figure S12: Proposed fragmentation cascade to fragment 42 of Uracil including computed enthalpy of activation ΔH_a and associated m/z values.

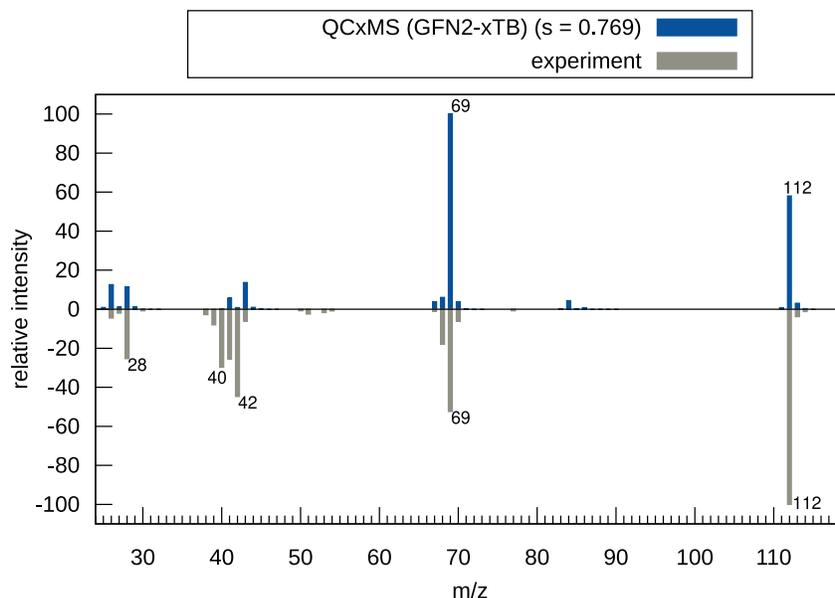


Figure S13: QCxMS computed spectrum for uracil and inverted experimental spectrum for comparison.

S16 Computed spectra of the test set

Spectra computed with QCxMS2 of the test set, which are not shown in the main paper are shown here including with the input geometry shown in the plot.

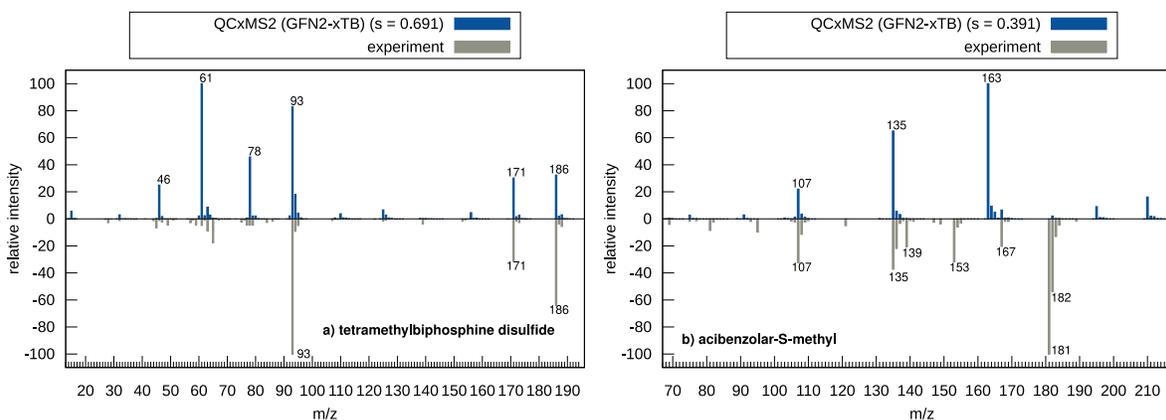


Figure S14: Mass spectra of a): tetramethylbiphosphine disulfide and b): acibenzolar-S-methyl computed with QCxMS2 at the GFN2-xTB level and inverted experimental spectrum.

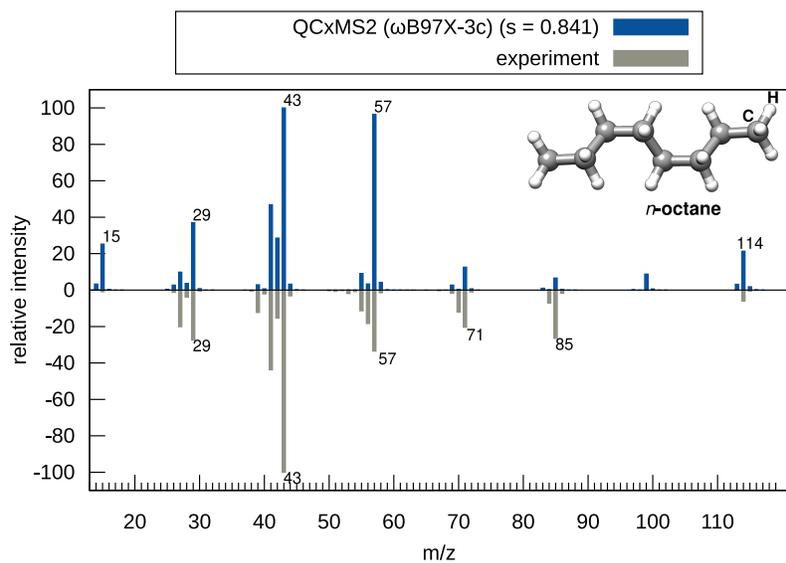


Figure S15: Mass spectrum of *n*-octane computed with QCxMS2 at the ω B97X-3c level and inverted experimental spectrum.

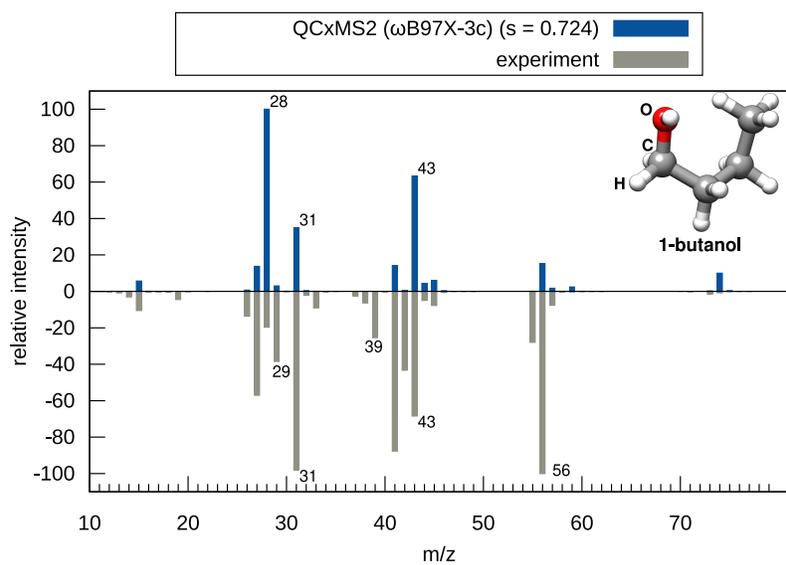


Figure S16: Mass spectrum of 1-butanol computed with QCxMS2 at the ω B97X-3c level and inverted experimental spectrum.

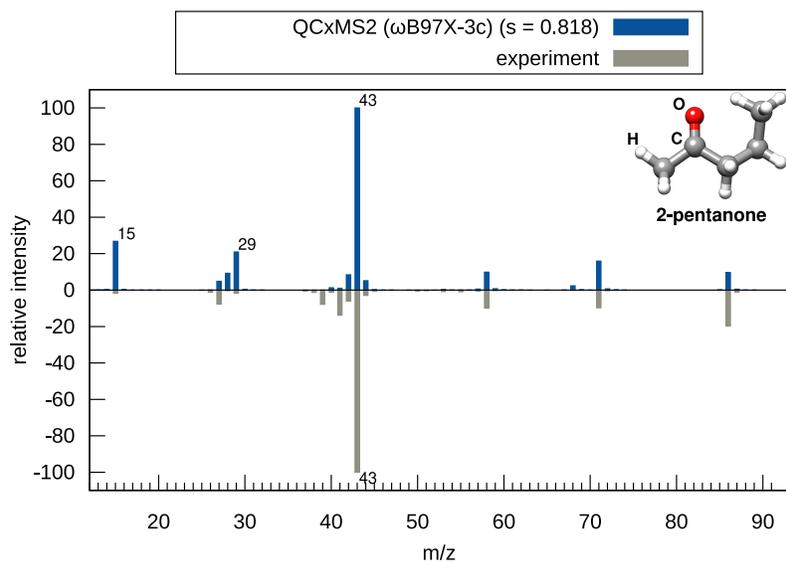


Figure S17: Mass spectrum of 2-pentanone computed with QCxMS2 at the ω B97X-3c level and inverted experimental spectrum.

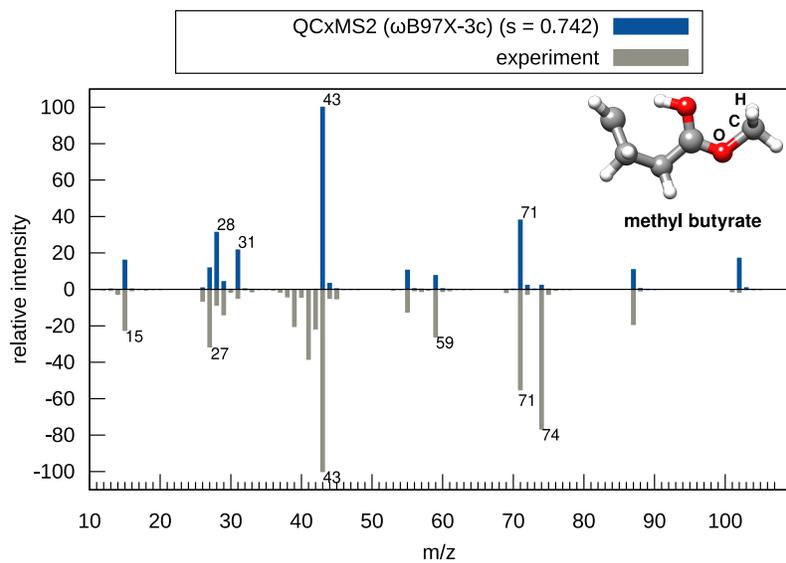


Figure S18: Mass spectrum of methyl butyrate computed with QCxMS2 at the ω B97X-3c level and inverted experimental spectrum.

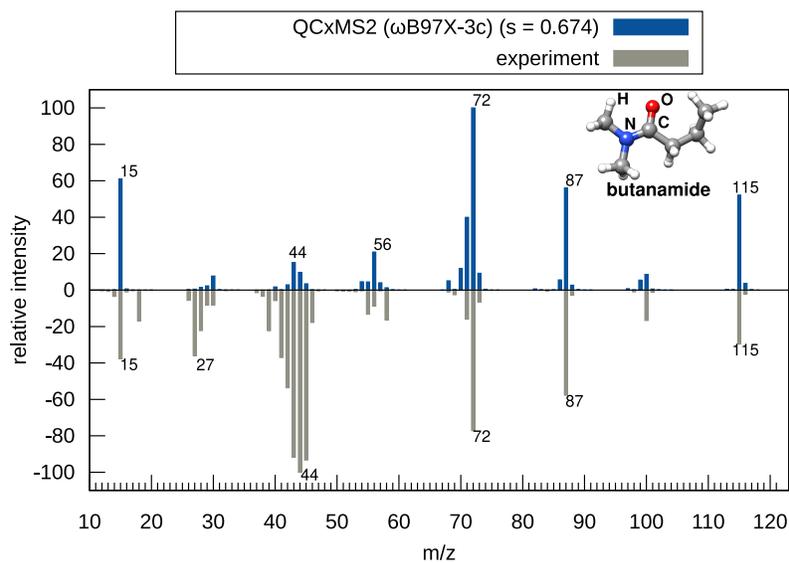


Figure S19: Mass spectrum of butanamide computed with QCxMS2 at the ω B97X-3c level and inverted experimental spectrum.

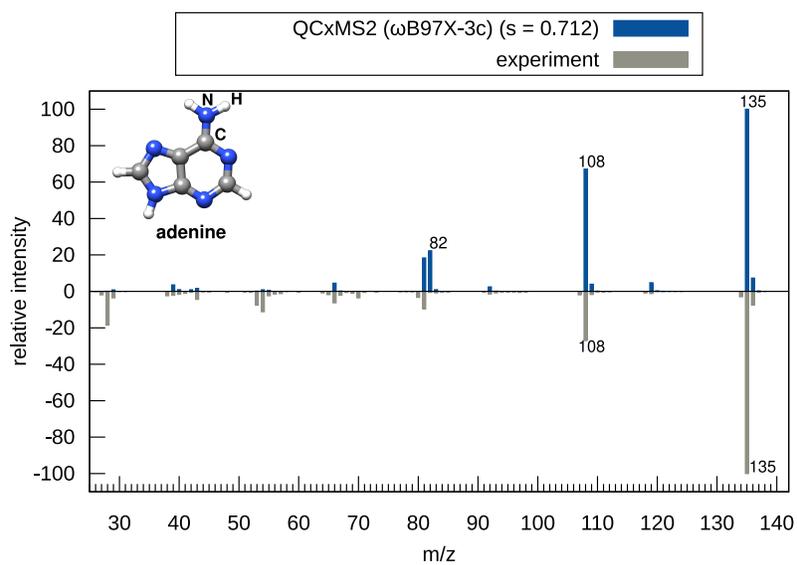


Figure S20: Mass spectrum of adenine computed with QCxMS2 at the ω B97X-3c level and inverted experimental spectrum.

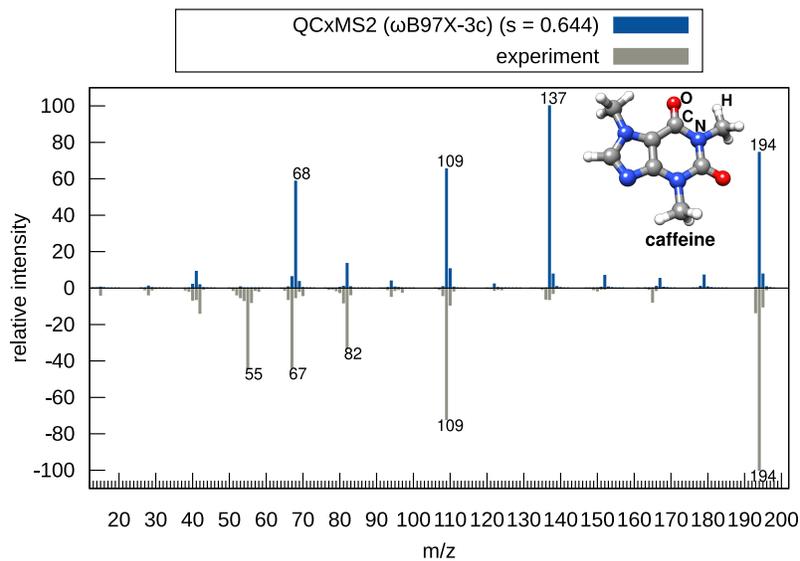


Figure S21: Mass spectrum of caffeine computed with QCxMS2 at the ω B97X-3c level and inverted experimental spectrum.

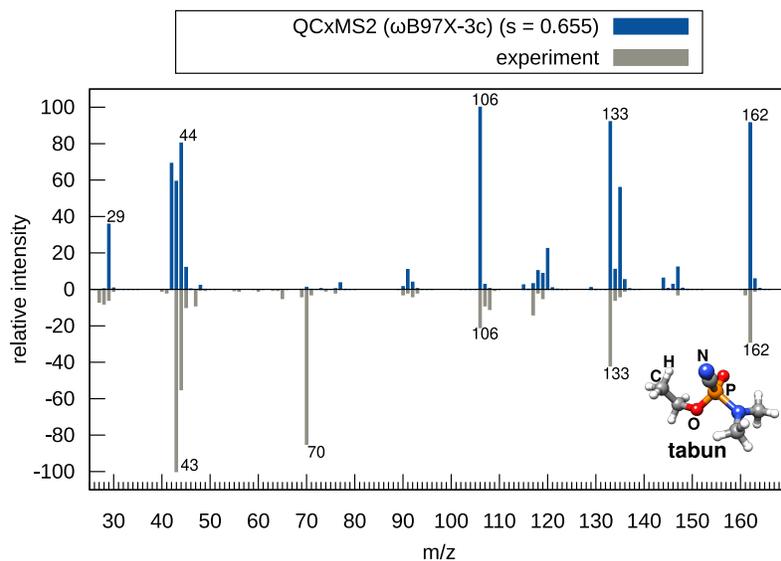


Figure S22: Mass spectrum of tabun computed with QCxMS2 at the ω B97X-3c level and inverted experimental spectrum.

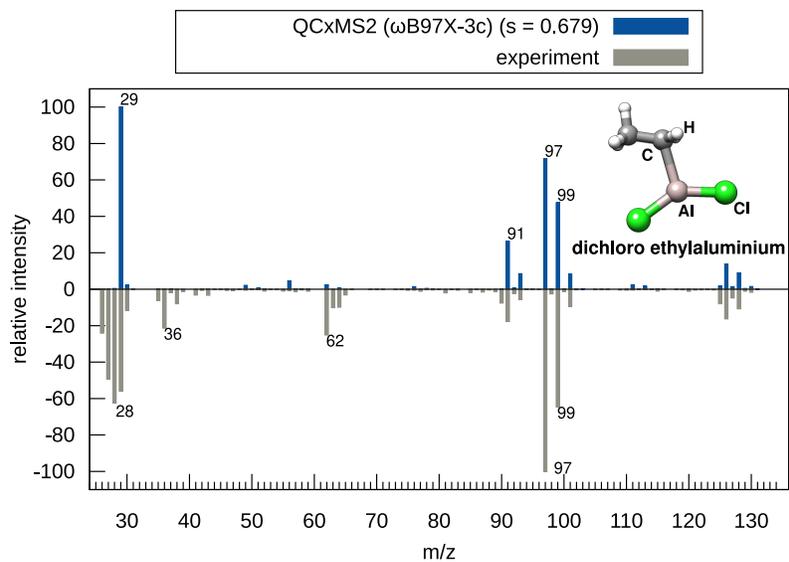


Figure S23: Mass spectrum of dichloroethylaluminum computed with QCxMS2 at the ω B97X-3c level and inverted experimental spectrum.

References

- (S1) Gross, J. H. *Mass spectrometry: a textbook*; Springer Science & Business Media, 2006.
- (S2) Dantus, M. *Acc. Chem. Res.* **2024**, 033003.
- (S3) Grimme, S. *Chem. Eur. J.* **2012**, *18*, 9955–9964.
- (S4) Haney, M. A.; Franklin, J. *J. Chem. Phys.* **1968**, *48*, 4093–4097.
- (S5) Kim, K.; Beynon, J.; Cooks, R. *J. Chem. Phys.* **1974**, *61*, 1305–1314.
- (S6) Wohlgemuth, G.; Mehta, S. S.; Mejia, R. F.; Neumann, S.; Pedrosa, D.; Pluskal, T.; Schymanski, E. L.; Willighagen, E. L.; Wilson, M.; Wishart, D. S.; Arita, M.; Dorrestein, P. C.; Bandeira, N.; Wang, M.; Schulze, T.; Salek, R. M.; Steinbeck, C.; Nainala, V. C.; Mistrik, R.; Nishioka, T.; Fiehn, O. *Nat. Biotechnol* **2016**, *34*, 1099–1101.