In Silico Tandem Mass Spectrometer: an Analytical and Fundamental Tool

Andrea Carrà, Riccardo Spezia

In Silico Tandem Mass Spectrometer: an Analytical and Fundamental Tool

Andrea Carrà[a] and Riccardo Spezia*[b]

Authors like to dedicate this article to the memory of Prof. *William L. Hase*

In this article, we summarize some aspects of the recently developed computational approach to model and predict collision induced dissociation mass spectra. In particular, we describe how initial conditions can be set to model fragmentation conditions and then obtain different results from the analysis of an ensemble of reactive trajectories. This approach was studied and validated on different systems, from small organic molecules to large biomolecules. Recently an iterative procedure was proposed to increase the fragmentation sampling, reducing computational time and providing a more comprehensive fragmentation pattern. All these fundamental developments are at the basis of the application of this approach to analytical problems. One important and possible outcome will be in creating an *in silico* data base which will be an useful complement to the experimental ones. This is discussed within other possible future outlooks.

1. Introduction

Tandem mass spectrometry is a physical chemistry method which is nowadays used in several analytical applications. The basic principles are that once a given ionic species is isolated in the gas phase it is activated and fragments are produced (generally this occurs in an ion trap, or through quadrupoles). The activation can be done via the collision with an inert gas (N_2, He, Ne, Ar, Xe), and, in this case, we talk about collision induced dissociation (CID).[1] Alternatively, the ions can be activated by collision with an inert surface (e.g. diamond), leading to surface induced dissociation (SID).[2] Photons can also activate the ions:[3] when in UV/Vis region this results in the so-called UV photon dissociation (UVPD), while if in the IR we have IR(multiple) photon dissociation (IRPD and IRMPD) or the blackbody infrared radiative dissociation (BIRD). The capture of an electron is also a way of activating the ion, as in the electron transfer and electron capture dissociations (ETD and ECD).[4] There exist, of course, other activation methods like ion-molecule reaction, electron impact ionization, negative electron transfer dissociation or electron detachment dissociation.[5]

The resulting fragmentation products are characteristic of the activated ion and of the activation mode. They can be used, for example, as a fingerprint of the molecule, or to study fundamental properties of the chemical bonds, or to reconstruct the action IR or UV/Vis spectrum.

Theoretical chemistry and more in general computer-based approaches are very powerful tools to help in understanding fragmentation spectra.[6] The aim of such *in silico* methods can be manifold: propose the structures of reactants and products, identify the reaction pathways, help in kinetic modelling associated with some of these techniques etc.

One key issue in MS techniques, is that experimentally only indirect information on the structure of the species is available. While the mass is carefully determined, such that the object under analysis is much more clearly identified with respect to in solution approaches, only the charged species are analyzed. Furthermore, it is very difficult to isolate the intermediates of a reaction mechanism. Theoretical and computational chemistry is a largely used tool to help experimentalists to better understand molecular properties of molecules under study.

Often, quantum chemistry is used to identify reactants, products and transition states via geometry optimizations, using the highest level of theory which is computationally doable given the size of the system.[7] This is done after the products are known and it becomes very complicated when the system size is growing. The reason is not only that the theory level should decrease as systems size increases, but more importantly the conformational sampling size increases enormously such that "manual" identification of structures and reaction pathways become almost impossible (and what suggested is often questionable because of the uncertainty in the correctness of the sampling). Recently, Blockhuys and co-workers proposed a method to predict fragments based on bond order obtained from electronic structure calculations of the reactants.[8]

One approach is to use computer-based approaches to identify in an automatic fashion minima and transition states along a reaction pathway.[9] In this way, a kinetic picture can also be built.[10] In the context of MS/MS this was applied to uracil[11] and, partially, to L-sulfated cysteine.[12] The drawback is that the procedure fails if the potential energy surface is too flat

[a] *Dr. A. Carrà*
Agilent Technologies Italia
Via Piero Gobetti 2/C
20063 Cernusco SN, Milano (Italy)
[b] *Dr. R. Spezia*
Laboratoire de Chimie Théorique
Sorbonne Université, UMR 7616 CNRS
4, Place Jussieu
75005 Paris (France)
E-mail: riccardo.spezia@sorbonne-universite.fr

(which occurs in certain regions in particular for flexible molecules).

Another approach, pioneered by Hase and co-workers,[13] is based on chemical dynamics simulations. Ions are activated due to explicit collision with an inert gas and resulting trajectories can lead to products. In this way, products are obtained automatically without the need of knowing them in advance. Furthermore, also non-statistical mechanisms, which can be important in both CID and SID, are obtained. This approach was successfully used in MS/MS of different systems, paving the way to a fully *in silico* fragmentation determination.[14] A similar approach based on multiple trajectories was proposed in 2013 by Grimme[15] and developed by the group, called quantum chemistry ionization mass spectra (QCEIMS).[16] It is designed to model electron ionization (EI) spectra by means of multiple trajectories. The theoretical spectra are obtained by counting the abundance of the different fragmentation products. The most important difference comes from the activation way, as it is obvious since QCEIMS was particularly designed for EI. Simulations were run using different electronic structure methods and were successfully applied to fragmentation of different organic molecules and nucleobases.[17]

In the following we will discuss methods and issues related to the use of chemical dynamics in CID.

2. Methodologies

An ion in a mass spectrometer is activated by collisions with an inert gas followed (eventually) by unimolecular fragmentation. This corresponds to ro-vibrational activation and many aspects depend on the experimental apparatus. Simulations were not developed to a specific one but to grasp the general process. In particular, they are aimed in modeling limit (somehow abstract) conditions, and notably: (i) single collision with an inert gas atom or molecule; (ii) multiple low-energy collisions. These two situations were modeled differently. In the following we report the basic approaches used. They are important to correctly identify the physical-chemistry framework on which the simulations are located.

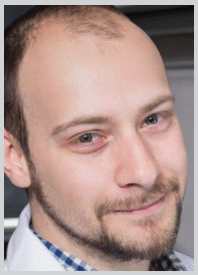
All the studies presented in sections 2.1 and 2.2 were performed with VENUS software[18] developed by the Hase group which was coupled with different electronic structure theory codes, while the analysis was done with home-made codes.

2.1. Single Collision

The single collision limit is modeled by simulating explicitly the fragmenting molecular ion and the neutral gas. This last is generally a rare gas atom (Ar, Xe, Ne,…) but it can be also N_2. The total system is treated by an interaction potential:

$$V = V_{ion} + V_{ion/neutral} + V_{neutral} \quad (1)$$

composed by the intermolecular potential of the ion molecule, V_{ion}, the internal potential of the projectile, $V_{neutral}$ (this is zero for atoms) and a potential describing the interaction between the ion and the gas, $V_{ion/neutral}$. For V_{ion} electronic structure calculations are generally used to allow the system to react. First studies focusing on energy transfer used molecular mechanics potential,[19] while later semi-empirical Hamiltonians, density functional theory or MP2 methods were used.[14] To study relatively large ions, semi-empirical Hamiltonians are necessary and recent studies suggest that PM3, RM1, PM6-D or PM7 show the best performances.[12,14c–e,20] For $V_{ion/neutral}$ it is possible to use analytical functions, parametrized on accurate quantum chemistry calculations,[14b,e,19,20b,21] while semi-empirical Hamiltonians are well performing for N_2.[22]



Andrea Carrà obtained his Ph.D. in physical chemistry with Prof. A. Vertova at University of Milan in 2014. He got intrigued by mass spectrometry since he was a student, with a focus in omics research and method development. During the years at university he collaborated as scientist with both academic and industrial research teams (IRF M.Negri and SAES Getters). In 2016, he moved for a post-doctoral position in the group Prof S. Balbo, working on Adductomics. During his postdoc Andrea has found an emerging interest for the development of computational model aimed at predicting MS spectra. Thanks to the collaboration with Dr. R. Spezia at Sorbone Univeristy in Paris, they figured out a way to map in silico the DNA adducts reactivity, finding a promising correlation between experimental and computational data. Upon this experience Andrea got a permanent position within Agilent Technologies (2018).

Riccardo Spezia studied chemistry at Università di Roma "La Sapienza" where he also obtained his PhD in theoretical physical chemistry with Prof. A. Di Nola in 2004. He moved for post-doctoral position in the group of Prof. J.T. Hynes at ENS Paris, and then to Université d'Evry where he was enrolled as CNRS researcher. In 2018, he moved to Sorbonne Université where he was promoted CNRS research director. His research activity is in the field of molecular dynamics and recently he focused in particular on reaction dynamics. He developed the concept of theoretical mass spectrometry and extended this approach also to study complex ion-molecule reactions relevant to astrochemistry.

Initial conditions (positions and velocities) are obtained by semi-classical sampling:[23] the ion minimum energy geometry is determined (in principle it needs to be just a local minimum) and normal modes calculated. Then, the population of each mode is determined either by setting a temperature (normal mode Boltzmann sampling) or a total internal vibrational energy. Normally, 300 K (or related vibrational energy) is considered, assuming that in MS/MS experiments no particular temperature control is done. If this is not the case, a different initial temperature can be set. Rotational motion is considered classical and added on top of vibrational one. The same is done for N_2 when it is used as neutral collision partner.

Then, the ion and the neutral are placed at a given distance (far enough to have $V_{ion/neutral} \sim 0$) and random relative orientations are sampled: Euler angles to consider the possible side of approach and impact parameter (which measures the distance from the center-of-mass). This shows that to correctly model an explicit collision, many trajectories are needed (in the order of hundreds or thousands depending on the size of the ion and the complexity of fragmentation).

Finally, the relative energy is set as an input parameter. This corresponds to the collision energy in the center-of-mass framework, which is a typical parameter of mass spectrometer instruments like quadrupole time-of-flight (QqTof) or triple-quadrupole (QqQ). The collision energy can be varied as in experiments and relative abundance of products followed as a function of collision energy. An example is reported in Figure 1, which corresponds to the fragmentation of protonated urea (extracted from Ref. [14a]).

Once the initial conditions are set, chemical dynamics simulations are run using the on-the-fly potential and numerical integration of Newton's equations of motion. The simulation times are usually in the order of tens of picosecond for each trajectory.

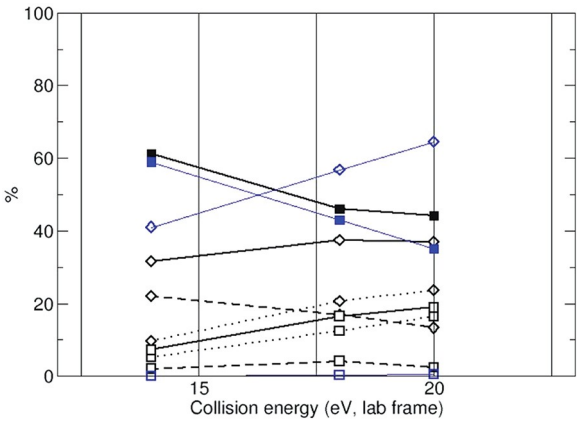


Figure 1. Results of final products of $UreaH^+ + Ar$ collisional simulations: reactants (■), $NH_3 + CONH_2^+$ (□) and $NH_4^+ + CONH$ (△) products. Full lines are total, while dotted lines are for shattering and dashed for non-shattering trajectories. In blue the same results as obtained from statistical kinetic analysis. Reprinted with permission from Ref. [14a]. Copyright 2009 American Chemical Society.

2.2. Multiple Collisions

When ions are activated by multiple (and generally low-energy) collisions, it is typically assumed that after each collision the transferred energy is redistributed to the whole ion and then another collision occurs and the process is repeated a huge number of times. This is modeled in simulations by providing to the ion a given internal energy which is randomly distributed on the vibrational (and eventually rotational) degrees of freedom. Only the ion under investigation is explicitly considered, so at odds with single collision method described previously the neutral gas is not modeled explicitly.

The fragmenting ion is thus treated using electronic structure theory as before. Even if collisions are not explicitly simulated, it is necessary to perform an ensemble of trajectories also in this case: in fact, the initial excess energy can be distributed in different ways in the internal degrees of freedom. Usually, micro-canonical normal mode sampling is performed to internally activate the ion.[23] The value of the internal energy cannot be set directly from an experimental parameter, but it can be estimated in different ways. One possibility is to use explicit collisions to evaluate the amount of transferred energy and then use the average value or a range of obtained values. Another possibility, is to run simulations as a function of the internal energy and analyze the evolution of results (fragmentation percentage, product ions, mechanisms…) as a function of this parameter. In some way, this is analogous to what is done experimentally by modifying the activation voltage. At odds with collisional simulations, which can be compared with QqQ or QqTof experiments where the collision energy is set, the excess energy cannot be directly related to the collision voltage of an ion trap, for example. In simulations, one can control precisely the excess energy given to the ion. The analysis of theoretical results and their evolution as a function of the internal energy could become a fully in silico approach to study properties of ions even disconnected from experiments.

2.3. Data Analysis

Once the simulations are done, an accurate analysis of the results is needed to obtain an in silico mass spectrum. First, trajectories provide geometries of atoms as a function of time: a key point is that one must analyze them to have the first important information, namely if the ion fragmented. When only few trajectories were performed, like in early studies,[13a,d,14a,21] this analysis was often done just looking the generated animations, while today with the increasing number of trajectories this is no longer possible. Another important information is the chemical species obtained. In simple systems, one can identify all the possible products before the simulations and then monitor their formation (in terms of abundance, time etc). Again, when moving to large systems this is no more possible. At this end, automatic tools were developed based on graph theory.[20b,24] From the atom-atom distance, with a distance cut-off criterion which must be carefully checked, it is possible to know which atoms are connected and, using

standard graph-theory algorithms, to identify the fragments and their atomic composition.

Once the product molecules are identified in the whole ensemble of trajectories, one needs to localize the charge. For simple systems (in particular singly charged with few products), this can be done very easily by simple chemical intuition. On the other hand, when systems grow in size and charge, the approach used so far is to couple with some charge projection method, like Mulliken charges or Natural Bond Orbitals (NBO) analysis.[25] These methods provide the partial charge on each atom and by summing them it is possible to obtain the charge on each fragment. In principle, non-integer charges can result on the fragments: this can be due to an intrinsic electronic structure problem or as an indication that the two fragments are still strongly interacting.

By counting the abundance of charged products at the end of simulations one can obtain a theoretical mass spectrum. An example is reported in Figure 2. However, due to limitation in time-length of the actual simulations, it is possible that at the end some ion-molecule complexes still survive or some activated ions do not have enough simulation time to fragment. They can evolve in a simple fragmentation but it is also possible that they rearrange providing different products. This is why the obtained spectra are time-dependent spectra.

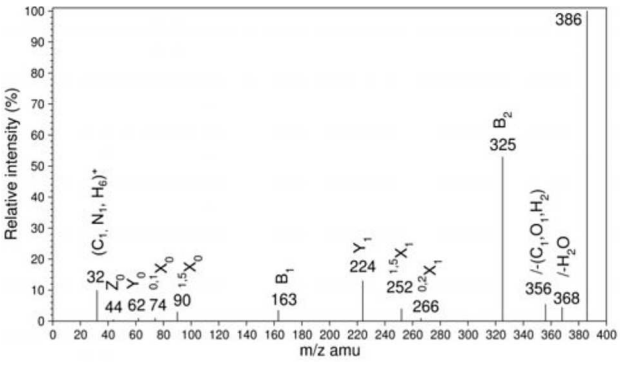


Figure 2. Theoretical mass spectrum of protonated b-amminoethylcellobiose as obtained from CID simulations. Reproduced from Ref. [14b]. Copyright 2017 Wiley.

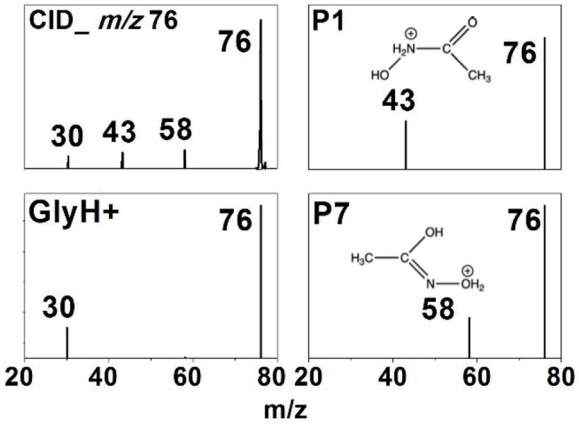


Figure 3. CID spectra of protonated ion m/z 76 as obtained from IR-induced ion-molecule reaction. Experimental (top left) and theoretical (different isomers as noted). Reproduced from Ref. [25]. Copyright 2020 Wiley.

3. Fragmentation Fingerprint

Using the methodology described previously, it is possible to obtain the fragmentation products of a molecular system just using as input a molecular geometry. Often, different isomers are possible for a given chemical formula, and the fragmentation products can be obtained for each of them. This is potentially useful to determine the fingerprint of a given molecular structure and compare with experiments. In most cases, the isomers differ from the protonation site and this will be discussed in detail in section 3.1. Recently, different isomers of protonated glycine were studied by fragmentation simulations, and used to suggest that multiple isomers are formed as result of laser-induced ion-product reaction.[26] In Figure 3 we show the experimental CID spectrum of ion *m/z* 76 which is the result of $NH_3OH^+ + CH_3COOH$ reaction catalyzed by IR laser, with theoretical spectra of three isomers. Notably, peaks *m/z* 43 and 58, which were not typical of protonated glycine, can come from different structures and simulations show such different CID fingerprints.

3.1. Mobile Proton

It is well-established that gas phase fragmentation of peptides follows the so-called mobile proton model.[27] In fact, the gas phase ions produced by electrospray ionization are protonated in the most basic site(s). Collisional activation can provide fragmentation directly from the most stable tautomer in the gas phase, but also induce proton transfer(s) populating tautomers which are higher in energy. These intermediates are often crucial to further provide final fragmentation products. Simulations have confirmed this picture, and found that also other molecules follow a similar mechanism.[14c,20b,28] However, simulating the full process (proton transfer plus subsequent fragmentation) needs a huge amount of computing time. Shortcuts can be used by running different sets of simulations with the different tautomers as initial structure. In small systems, where the number of tautomers is relatively small, this can be done directly.[29] However, when the possibilities grow due to a large number of possible protonation sites, a more general approach is needed. Recently, we have proposed an iterative scheme to account for different tautomers in fragmentation of protonated methyl-guanine.[24d] This approach is totally general and it is based on automatic recognition of isomers using graph theory. Once an isomer (which is mostly a tautomer) is obtained, it is used as new input if the structure is new and this iteratively up to any new structure is recognized. A schematic picture of the iterative procedure is shown in Figure 4.

In this way, we can gain a lot of computing time: in fact, the new tautomer starts from equilibrated initial conditions, thus

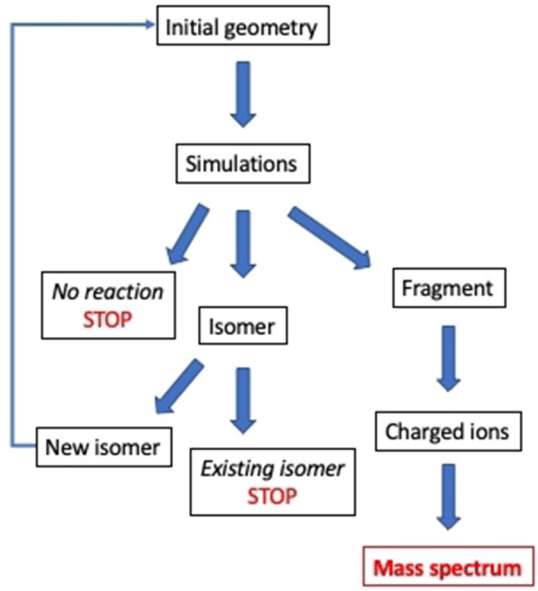


Figure 4. Schematic representation of the iterative simulation approach to account for proton mobile model in fragmentation simulation. Here we show the general procedure which holds for any kind of isomerization.

mimicking the equilibration step done by the buffer gas in real instruments.

This approach, so far used to sample tautomerization space, is not limited in principle to proton transfer, but it can be extended to all possible isomers which are populated within the energy considered in the activation energy range applied to the initial structure.

4. Database and "Omics" Sciences

In the era of genome-wide association studies and development of -omics techniques, new technologies are needed to allow, for example, the investigation of molecular effects arising from the environmental exposure in all its complexity.

In this scenario, mass spectrometry has played a key role in the last decade, defining several biomarkers for the early diagnosis of complex diseases.[30] Recently, computational mass spectrometry has emerged as an additional tool, useful for the identification of unknown compounds, independently on any prior assumption about their fragmentations.[22a,31] Even though computational approaches, e.g. Mascot, can drive mass spectrometry data analysis across super dense set of data,[32] sample preparation is still playing a key role at isolating the class of molecules, which are aimed to be investigated.

The exploration of certain molecular classes allows the systematic study of their expression under specific boundary conditions, related to the experimental design. The most typical example is the identification of proteins within proteomic associated studies.[33] Nevertheless, proteins are not the only class of molecules deserving the attention of fundamental research. For example, the investigation of peptides, lipids, carbohydrates and nucleic acids supports glycomics,[34] lipidomics[35] and genomics.[36] Today, the evaluation of each sample batch is based upon the possibility of loading the experimental file on a dedicated database and get feedbacks about the direct comparison of both theoretical and experimental MS data e.g. METLIN.[37] This approach is extremely efficient for compounds which have a known, or at least a predictable, fragmentation mechanism. This is the case, for example, for peptides,[6a] nucleic acids[38] or aldehydes.[39] At the state of the art, the biggest limitation for data analysis is the lack of fragmentation information. These data are necessary for the unambiguous identification of analytes, which can be candidate biomarkers.

In future perspective, the growing number of high throughput methodologies will expand the classes of candidate compounds characterized by biological interest. On our present days, we are already affording this open challenge. Most of the unknown compounds can only be identified by evaluating manually their MS spectra,[40] but some speculations are hardly supported by further experimental evidences, because of the lack of isotopically labeled internal standards.

For the time being, in silico mass spectrometry is relatively in its infancy, in particular concerning applications to large systems. Further developing and systematic application to different classes of compounds will be important to keep sharing and rationalize MS data arising from un-harmonized experimental conditions.[41] In fact, MS fragmentations differ from instrument to instrument according to collision energy, geometry, lens voltages etc. To achieve a global coverage of MS data, it will be necessary to introduce an algorithm aimed at defining unique fragmentation models, based on an unambiguous chemical approach. In silico mass spectrometry can be the instrument of choice to support future progresses of mass spectrometry, especially in case of unknown compounds. At this end, simulations must be computationally fast and reliable at the same time. The use of chemical dynamics (as described in previous sections) with semi-empirical Hamiltonians seem to be the best choice at the time, while other methods, like tight-binding density functional theory (DFTB)[42] will merit a deeper investigation. First direct dynamics simulations using this approach are encouraging.[43]

The iterative approach described in section 3.1 for the mobile proton case, is surely a practical way of enhancing the fragmentation sampling without the need of conceiving all the possible isomers. In fact, its use is not limited to proton transfer but it can be used for any isomerization process. Furthermore, also secondary fragmentation can be considered in an analogous way, by using, for example, the main first fragmentation products as input for new runs and determination of new fragments. The ultimate goal of iterative simulations approach will be the full coverage of the chemical pathway behind each experimental fragmentation pattern.

By means of high-resolution mass spectrometers, nowadays analysts can easily get accurate mass spectra, wherein each fragmentation signal has a chemical formula directly associated. Differently from today, the future database search will directly use chemical formula instead of *m/z* data. In this way, all the

unknown analytes will be partially, or even completely, identified using a pull of diagnostic fragments, instead of computing for the number of theoretical signals covered by experiential result-set.

Furthermore, the intensive use of computational models along with the development of dedicated facility and network project, will allow to merge together results from any sort of MS facility. Hopefully, disputes arising from un-harmonized results will be minimized and eventually computational mass spectrometry will bring research teams deeply into the systematic investigation of molecular biology.

These different aspects suggest that in silico MS/MS can be used to build a new database which would be complementary to experimental ones and can help in better analysis of a large class of compounds. At this end, after the first pioneering phase of theoretical fragmentation studies, a more systematic approach to cover the different classes of fragments is clearly needed, together with more standard and user-friendly computational tools.

5. Fundamental aspects

Chemical dynamics simulations can provide also fundamental aspects of unimolecular reactivity in MS/MS experiments. As already discussed, the fragmentation patterns and branching ratio are time-dependent quantities, so their absolute values are related to the simulation time. With this respect, the intensities from *in silico* mass spectra should not be directly compared with experiments.

On the other hand, since simulations provide time dependent properties, it is possible to obtain kinetic information. In particular, by following in time the abundance of the precursor ion, it is possible to characterize its unimolecular kinetics. Simulations have shown that, when using the internal energy activation approach, the precursor ion shows an exponential decay, in line with Rice-Ramsperger-Kassel-Marcus (RRKM) theory.[44] This decay can be fitted with an exponential function obtaining the uni-molecular rate constant. In case of multiple reaction products, the rate constant of each pathway (k_i) can be obtained from the overall one (k) by the simple expression:

$$k_i = P_i\, k \tag{2}$$

where P_i is the probability of obtaining a given pathway *i*. Of course, this is statistically meaningful only for abundant pathways (i.e. when P_i is not small and its statistical uncertainty not too large).

Furthermore, by measuring the unimolecular rate constants at different internal energies it is possible to obtain activation energies via Arrhenius-like or RRK fits (RRK is the classical version of the general RRKM theory). Arrhenius-like expressions are grounded in particular for systems with a large number of degrees of freedom and when the activation energy is much higher than the barrier. In this case, a classical correspondence between energy and temperature can be done and it is possible to describe the temperature behavior via a typical Arrhenius plot. Also, pre-exponential factors can be obtained.[20a,c,24a]

Similarly, RRK fits can be done if the system does not hold the previously mentioned conditions, providing the threshold energies.

Recently, a simple model of unimolecular reaction was simulated including nuclear quantum effects with relatively non-expensive algorithms.[45] This paves the way of a more quantitative calculations of unimolecular reaction barriers.

Using this approach, it was possible to estimate the statistical fragmentation threshold of different peptides. The comparison between these simulations and direct collision ones could show differences between the two fragmentation processes.[20d,46] In particular, while unimolecular decay is exponential in internal energy activation simulations, it is not when doing single collision ones.[24a] In these simulations a direct fragmentation mechanism, the so-called shattering,[47] can play an important role: this is characterized by a sudden bond breaking when the ion and the inert gas collide (typically in less than a vibrational period of the breaking bond). Different simulations have shown how this mechanism can be important and possibly at the basis of the formation of high energy products which cannot be justified simply by the study of the potential energy surface.[13b–d,14a,48] Furthermore, it was possible to estimate the shattering threshold energy (i.e. the minimum energy needed for shattering formation) in peptides and compare with statistical threshold, showing that this last is clearly smaller.[20d,46] A direct consequence is that in threshold-CID experiments the effect of shattering fragmentation should be minimal (while it can be important when increasing the collision energy).

Finally, these fundamental aspects related to kinetics and activation energy have a direct relation with applications. In fact, they can be used to make a direct link between internal energy used in simulations and experimental conditions, for example through the dissociation of the thermometer ions.[43b] Recent simulations have shown that they can be successfully studied with different methods and different kinetic fits,[49] thus paving the way of using them to a more accurate calibration of energy used in simulations.

6. Conclusions and Outlooks

This minireview has summarized the main advances made recently to use *in silico* technology in order to model and predict collision induced dissociation spectra of complex molecules. Chemical dynamics simulations are at the basis of the approach and two limit activation modes are discussed and used recently to study different classes of molecules. We should recall that the same approach can be used in SID.[50] However, SID is less used in analytical chemistry laboratories and from a computational point of view the ion-surface interaction is more complex and should be parametrized for each new system. A similar approach, also based on multiple trajectories but with a different activation mode, was developed and applied by

Grimme and co-workers to model *in silico* electronic ionization spectra.[15–17]

The internal energy activation simulations are very general and can provide a reference for future building of an *in silico*, fragmentation database. A systematic study of different compounds should be done to this end and this will also need a technological development to provide user friendly tools which can be used by not skilled operators. Different options are under considerations, like developing a web-server database where submitting the system under interest and/or build a library which can be accessed by the users.

Combining together the different developments done so far (electronic structure method, activation modeling, graph-theory based analysis, iterative procedure etc.) a full *in silico* mass spectrometer is available. While it should be made more user-friendly from a technical point of view, the different studies have shown that it is scientifically solid and it can move on to a production stage. Further developments are likely in some fundamental aspects related to including nuclear quantum effects but also in using faster methods (like e.g. reactive force fields) in order to study bigger systems and/or increase the statistical sampling. One possibility would be to use machine-learning (ML) based approaches. ML is used in mass spectrometry to obtain fragmentations from data-bases of existing spectra or to predict MS/MS of specific classes of molecules from common rules (typically peptides).[51] However, recently many progresses were done in using ML to build reactive potentials,[52] with applications also to excited state reactivity.[53] It can be clearly a powerful method to build reactive force fields paving the way to study, for example, full proteins.

Acknowledgements

The authors thank the American Society of Mass Spectrometry for a post-doctoral travel award.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: mass spectrometry · computer simulations · gas-phase reactivity · omics sciences · fragmentation data base

[1] a) R. B. Cody, B. S. Freiser, *Int. J. Mass Spectrom. Ion Phys.* **1982**, *41*, 199–204; b) J. Laskin, J. H. Futrell, *Mass Spectrom. Rev.* **2005**, *24*, 135–167.
[2] a) M. A. Mabud, M. J. Dekrey, R. G. Cooks, *Int. J. Mass Spectrom. Ion Processes* **1985**, *67*, 285–294; b) M. E. Bier, J. W. Amy, R. G. Cooks, J. E. P. Syka, P. Ceja, G. Stafford, *Int. J. Mass Spectrom. Ion Processes* **1987**, *77*, 31–47; c) V. H. Wysocki, J. M. Ding, J. L. Jones, J. H. Callahan, F. L. King, *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 190–208.
[3] a) N. C. Polfer, *Chem. Soc. Rev.* **2011**, *40*, 2211–2221; b) D. Scuderi, V. Lepere, G. Piani, A. Bouchet, A. Zehnacker-Rentien, *J. Phys. Chem. Lett.* **2014**, *5*, 56–61; c) J. P. Simons, *Mol. Phys.* **2009**, *107*, 2435–2458; d) T. Rizzo, J. Stearns, O. Boyarkin, *Int. Rev. Phys. Chem.* **2009**, *28*, 481–515; e) P. D. Schnier, W. D. Price, R. A. Jockusch, E. R. Williams, *J. Am. Chem. Soc.* **1996**, *118*, 7178–7189.
[4] a) R. A. Zubarev, N. L. Kelleher, F. W. McLafferty, *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266; b) F. W. McLafferty, D. M. Horn, K. Breuker, Y. Ge, M. A. Lewis, B. Cerda, R. A. Zubarev, B. K. Carpenter, *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 245–249; c) J. J. Coon, M. J. Schoreder, J. Shabanowitz, D. F. Hung, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9528–9533.
[5] a) J. S. Brodbelt, *Mass Spectrom. Rev.* **1997**, *16*, 91–110; b) T. D. Märk, G. H. Dunn, *Electron Impact Ionization*, Springer Science & Business Media, **2013**; c) J. Griffiths, *Anal. Chem.* **2008**, *80*, 5678–5683; d) B. A. Budnik, K. F. Haselmann, R. A. Zubarev, *Chem. Phys. Lett.* **2001**, *342*, 299–302.
[6] a) Y. Hoppilliard, G. Ohanessian, S. Bourcier, *J. Phys. Chem. A* **2004**, *108*, 9687; b) B. Paizs, S. Suhai, *Mass Spectrom. Rev.* **2005**, *24*, 508–548; c) B. Paizs, S. Suhai, *Rapid Commun. Mass Spectrom.* **2001**, *15*, 651–663; d) P. B. Armentrout, A. L. Heaton, *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 621–631.
[7] a) A. Eizaguirre, O. Mo, M. Yanez, J.-Y. Salpin, *Phys. Chem. Chem. Phys.* **2011**, *13*, 18409–18417; b) A. Eizaguirre, O. Mo, M. Yanez, J.-Y. Salpin, J. Tortajada, *Org. Biomol. Chem.* **2012**, *10*, 7552–7561; c) C. Trujillo, A. Lamsabhi, O. Mo, M. Yanez, J.-Y. Salpin, *Int. J. Mass Spectrom.* **2011**, *306*, 27–36.
[8] a) J. Cautereels, M. Claeys, D. Geldof, F. Blockhuys, *J. Mass Spectrom.* **2016**, *51*, 602–614; b) J. Cautereels, F. Blockhuys, *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 11227–1235.
[9] a) E. Martínez-Nuñez, *Phys. Chem. Chem. Phys.* **2015**, *17*, 14912–14912; b) E. Martínez-Nuñez, *J. Comput. Chem.* **2015**, *36*, 222–234.
[10] A. Rodríguez, R. Rodríguez-Fernandez, S. A. Vazquez, G. L. Barnes, J. J. P. Stewart, E. Martínez-Nuñez, *J. Comput. Chem.* **2018**, *39*, 1922–1930.
[11] E. Rossich Molina, J. Y. Salpin, R. Spezia, E. Martínez-Nuñez, *Phys. Chem. Chem. Phys.* **2016**, *18*, 14980–14990.
[12] V. Macaluso, D. Scuderi, M. E. Crestoni, S. Fornarini, D. Corinti, E. Dalloz, E. Martinez-Nuñez, W. L. Hase, R. Spezia, *J. Phys. Chem. A* **2019**, *123*, 3685–3696.
[13] a) P. de Sainte Claire, W. L. Hase, *J. Phys. Chem.* **1996**, *100*, 8190–8196; b) S. O. Merouch, Y. Wang, W. L. Hase, *J. Phys. Chem. A* **2002**, *106*, 9983–9992; c) J. Liu, K. Song, W. L. Hase, S. L. Anderson, *J. Chem. Phys.* **2003**, *119*, 3040–3050; d) E. Martínez-Núñez, A. Fernández-Ramos, S. A. Vázquez, J. M. C. Marques, M. Xue, W. L. Hase, *J. Chem. Phys.* **2005**, *123*, 154311.
[14] a) R. Spezia, J.-Y. Salpin, M.-P. Gaigeot, W. L. Hase, K. Song, *J. Phys. Chem. A.* **2009**, *113*, 13853–13862; b) R. Rodriguez-Fernandez, S. A. Vazquez, E. Martinez-Nuñez, *Phys. Chem. Chem. Phys.* **2013**, *15*, 7628–7637; c) R. Spezia, J. Martens, J. Oomens, K. Song, *Int. J. Mass Spectrom.* **2015**, *388*, 40–52; d) E. Rossich Molina, A. Eizaguirre, V. Haldys, D. Urban, G. Doisneau, Y. Bourdreux, J.-M. Beau, J.-Y. Salpin, R. Spezia, *ChemPhysChem* **2017**, *18*, 2812–2823; e) D. Ortiz, J.-Y. Salpin, K. Song, R. Spezia, *Int. J. Mass Spectrom.* **2014**, *358*, 25–35; f) G. Lee, E. Park, H. Chung, Y. Jeanvoine, K. Song, R. Spezia, *Int. J. Mass Spectrom.* **2016**, *407*, 40–50.
[15] S. Grimme, *Angew. Chem. Int. Ed.* **2013**, *52*, 6302–6312; *Angew. Chem.* **2013**, *125*, 6422–6425.
[16] a) C. A. Bauer, S. Grimme, *Org. Biomol. Chem.* **2014**, *12*, 8737–8744; b) C. A. Bauer, S. Grimme, *J. Phys. Chem. A* **2016**, *120*, 3755–3766; c) V. Asgeirsson, C. A. Bauer, S. Grimme, *Chem. Sci.* **2017**, *8*, 4879–4895.
[17] a) C. A. Bauer, S. Grimme, *J. Phys. Chem. A* **2014**, *118*, 11479–11484; b) C. A. Bauer, S. Grimme, *Eur. J. Mass Spectrom.* **2015**, *21*, 125–140; c) V. Asgeirsson, C. A. Bauer, S. Grimme, *Phys. Chem. Chem. Phys.* **2016**, *18*, 31017–31026; d) M. Engeser, C. Mundt, C. A. Bauer, S. Grimme, *Int. J. Mass Spectrom.* **2017**, *52*, 452–458; e) J. Koopman, S. Grimme, *ACS Omega* **2019**, *4*, 15120–15133.
[18] X. Hu, W. L. Hase, T. Pirraglia, *J. Comput. Chem.* **1991**, *12*, 1014–1024.
[19] O. Merouch, W. L. Hase, *J. Phys. Chem. A* **1999**, *103*, 3981–3990.
[20] a) Z. Homayoon, S. Pratihar, E. Dratz, R. Snider, R. Spezia, G. L. Barnes, V. Macaluso, A. Martin-Somer, W. L. Hase, *J. Phys. Chem. A* **2016**, *120*, 8211–8227; b) A. Martin-Somer, J. Martens, J. Grzetic, W. L. Hase, J. Oomens, R. Spezia, *J. Phys. Chem. A* **2018**, *122*, 2612–2625; c) A. Malik, Y.-F. Lin, S. Pratihar, L. A. Angel, W. L. Hase, *J. Phys. Chem. A* **2019**, *123*, 6868–6885; d) A. Malik, L. A. Angel, R. Spezia, W. L. Hase, *Phys. Chem. Chem. Phys.* **2020**, *22*, 14551–14559.
[21] a) Y. Jeanvoine, M.-P. Gaigeot, W. L. Hase, K. Song, R. Spezia, *Int. J. Mass Spectrom.* **2011**, *308*, 289–298; b) R. Spezia, A. Cimas, M.-P. Gaigeot, J.-Y. Salpin, K. Song, W. L. Hase, *Phys. Chem. Chem. Phys.* **2012**, *14*, 11724–11736.
[22] a) A. Martin-Somer, V. Macaluso, G. L. Barnes, L. Yang, S. Pratihar, K. Song, W. L. Hase, R. Spezia, *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 2–24; b) Z. Homayoon, V. Macaluso, A. Martin-Somer, M. C. N. Barbosa Muniz,

I. Borges Jr., W. L. Hase, R. Spezia, *Phys. Chem. Chem. Phys.* **2018**, *20*, 3614–3629.
[23] G. H. Peslherbe, H. Wang, W. L. Hase, *Adv. Chem. Phys.* **1999**, *105*, 171–201.
[24] a) R. Spezia, A. Martin-Somer, V. Macaluso, Z. Homayoon, S. Pratihar, W. L. Hase, *Faraday Discuss.* **2016**, *195*, 599–618; b) Y. Jeanvoine, A. Largo, W. L. Hase, R. Spezia, *J. Phys. Chem. A* **2018**, *122*, 869–877; c) S. Bougueroua, R. Spezia, S. Pezzotti, S. Vial, F. Quessette, D. Barth, M.-P. Gaigeot, *J. Chem. Phys.* **2018**, *149*, 184102; d) A. Carrà, V. Macaluso, P. W. Villalta, R. Spezia, S. Balbo, *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2771–2784.
[25] a) R. Mulliken, *J. Chem. Phys.* **1955**, *23*, 1833–1840; b) E. Carpenter, F. Weinhold, *J. Mol. Struct.*, **1988**, *139*, 41–62.
[26] D. Scuderi, A. F. Perez-Mellor, J. Lemaire, S. Indrajith, J.-X. Bardaud, A. Largo, Y. Jeanvoine, R. Spezia, *ChemPhysChem* **2020**, *21*, 503–509.
[27] A. R. Dongré, J. L. Jones, A. Somogyi, V. H. Wysocki, *J. Am. Chem. Soc.* **1996**, *118*, 8365–8374.
[28] D. Ortiz, P. Martin-Gago, A. Riera, K. Song, J.-Y. Salpin, R. Spezia, *Int. J. Mass Spectrom.* **2013**, *335*, 33–44.
[29] E. Rossich Molina, D. Ortiz, J.-Y. Salpin, R. Spezia, *J. Mass Spectrom.* **2015**, *50*, 1340–1351.
[30] X.-W. Zhang, Q.-H. Li, Z.-D. Xu, J.-J. Dou, *RSC Adv.* **2020**, *10*, 3092–3104.
[31] K. Song, R. Spezia, *Theoretical Mass Spectrometry*, De Gruyter, Berlin, **2018**.
[32] M. Bantscheff, S. Lemeer, M. M. Savitski, B. Kuster, *Anal. Bioanal. Chem.* **2012**, *404*, 939–967.
[33] M. Bantscheff, M. Schirla, G. Sweetman, J. Rick, B. Kuster, *Anal. Bioanal. Chem.* **2007**, *389*, 1017–1031.
[34] a) J. E. Turnbull, R. Sasisekharan, *OMICS* **2010**, *14*, 385–387; b) J. Zaia, *OMICS* **2010**, *14*, 401–418.
[35] S. J. Blanksby, T. W. Mitchell, *Ann. Rev. Anal. Chem.* **2010**, *3*, 433–465.
[36] J. S. Andersen, M. Mann, *FEBS Lett.* **2000**, *480*, 25–31.
[37] C. Guijas, J. R. Montenegro-Burke, X. Domingo-Almenara, A. Palermo, B. Warth, G. Hermann, G. Koellensperger, T. Huan, W. Uritboonthai, A. E. Aisporna, D. W. Wolan, M. E. Spilker, H. P. Benton, G. Siuzdak, *Anal. Chem.* **2018**, *90*, 3156–3164.
[38] S. Balbo, S. S. Hecht, P. Upadhyaya, P. W. Villalta, *Anal. Chem.* **2014**, *86*, 1744–1752.
[39] R. Dator, A. Carrà, L. Maertens, V. Guidolin, P. W. Villalta, S. Balbo, *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 608–618.
[40] Y. Chen, S. W. Kwon, S. C. Kim, Y. Zhao, *J. Proteome Res.* **2005**, *4*, 998–1005.
[41] A. L. Rockwood, M. S. Lowenthal, C. Bystrom, *Clin. Lab. Med.* **2018**, *38*, 499–513.
[42] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, *Phys. Rev. B* **1998**, *58*, 7260.
[43] a) R. Nieman, R. Spezia, B. Jayee, T. Minton, W. L. Hase, H. Guo, *J. Chem. Phys.* **2020**, *153*, 184702; b) A. Malik, R. Spezia, W. L. Hase, *J. Am. Soc. Mass Spectrom* **2021**, *32*, 169–179.
[44] T. Baer, W. L. Hase, *Unimolecular Reaction Dynamics. Theory an Experiments*, Oxford University Press: New York, **1996**.
[45] R. Spezia, H. Dammak, *J. Phys. Chem. A* **2019**, *123*, 8542–8551.
[46] V. Macaluso, Z. Homayoon, R. Spezia, W. L. Hase, *Phys. Chem. Chem. Phys.* **2018**, *20*, 19744–19749.
[47] a) D. G. Schultz, L. Hanley, *J. Chem. Phys.* **1998**, *109*, 10976; b) T. Raz, R. D. Levine, *J. Chem. Phys.* **1996**, *105*, 8097.
[48] a) E. Martínez-Nuñez, S. A. Vazquez, J. M. C. Marques, *J. Chem. Phys.* **2004**, *121*, 2571; b) E. Martínez-Nuñez, S. A. Vaźquez, E. J. Aoiz, J. F. Castillo, *J. Phys. Chem. A* **2006**, *110*, 1225–1231; c) A. Martin-Somer, M. Yañez, M.-P. Gaigeot, R. Spezia, *J. Phys. Chem. A* **2014**, *118*, 10882–10893.
[49] a) F. Derwa, E. de Pauw, P. Natalis, *Org. Mass Spectrom.* **1991**, *26*, 117–118; b) R. Rahrt, T. Auth, M. Demireva, P. B. Armentrout, K. Koszinowski, *Anal. Chem.* **2019**, *91*, 11703–11711.
[50] a) S. Pratihar, G. L. Barnes, W. L. Hase, *Chem. Soc. Rev.* **2016**, *45*, 3595–3608; b) S. Pratihar, G. L. Barnes, J. Laskin, W. L. Hase, *J. Phys. Chem. Lett.* **2016**, *7*, 3142–3150; c) D. G. Schultz, S. B. Weinhaus, L. Hanley, P. de Sainte Claire, W. L. Hase, *J. Chem. Phys.* **1997**, *106*, 10337–10348; d) S. B. M. Bosio, W. L. Hase, *Int. J. Mass Spectrom. Ion Processes* **1998**, *174*, 1–9; e) O. Meroueh, W. L. Hase, *Phys. Chem. Chem. Phys.* **2001**, *3*, 2306–2314; f) K. Song, O. Meroueh, W. L. Hase, *J. Chem. Phys.* **2003**, *118*, 2893–2902.
[51] a) K. A. Neilson, N. A. Ali, S. Muralidharan, M. Mirzaei, M. Mariani, G. Assadourian, A. Lee, S. C. van Sluyter, P. A. Haynes, *Proteomics* **2011**, *11*, 535–553; b) D. N. Perkins, D. J. Pappin, D. M. Creasy, J. S. Cottrell, *Electrophoresis* **1999**, *20*, 3551–3567; c) U. W. Liebal, A. N. T. Phan, M. Sudhakar, K. Raman, L. M. Blank, *Metabolites* **2020**, *10*, 243; d) Y.-M. Lin, C.-T. Chen, J.-M. Chang, *BMC Genomics* **2019**, *20*, 906; e) C. Ji, R. J. Arnold, K. J. Sokoloski, R. W. Hardy, H. Tang, P. Radivojac, *Proteomics* **2013**, *13*, 756–765; f) S. Li, R. J. Arnold, H. Tang, P. Radivojac, *Anal. Chem.* **2011**, 83, 790–796; g) S. Degroeve, L. Martens, *Bioinformatics* **2013**, *29*, 3199–3203; h) S. Degroeve, D. Maddelein, L. Martens, *Nucleic Acids Res.* **2015**, *43*, W326-W330.
[52] a) H. Nakata, S. Bai, *J. Comput. Chem.* **2019**, *40*, 2000–2012; b) K. Y. S. Chang, S. Fias, R. Ramakrishnan, A. O. von Lilienfeld, *J. Chem. Phys.* **2016**, *144*, 174110; c) A. O. von Lilienfeld, *Angew. Chem. Int. Ed.* **2018**, *57*, 4164–4169; *Angew. Chem.* **2018**, *130*, 4235–4240.
[53] a) J. Westermayr, P. Marquetand, *Mach. Learn.: Sci. Technol.* **2020**,*1*, 043001; b) J. Westermayr, M. Gastegger, P. Marquetand, *J. Phys. Chem. Lett.* **2020**, *11*, 3828–3834; c) J. Westermayr, M. Gastegger, M. Menger, S. Mai, L. Gonzalez, P. Marquetand, *Chem. Sci.* **2019**, *10*, 8100–8107; d) J. Westermayr, F. A. Faber, A. S. Christensen, O. A. von Lilienfeld, P. Marquetand, *Mach. Learn.: Sci. Technol.* **2020**, *1*, 025009; e) J. Westermayr, P. Marquetand, *Chem. Rev.* In press, DOI: 10.1021/acs.chemrev.0c00749.

Manuscript received: December 1, 2020
Version of record online: ■■, ■■■