université PARIS-SACLAY

Thèse de doctorat

Theoretical modelling of gas phase collision induced dissociation of biomolecules

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'université d'Evry val d'Essonne

École doctorale n°571 Science chimiques : molécules, matériaux, instrumentation et biosystèmes, 2MIB
Spécialité de doctorat: chimie

Thèse présentée et soutenue à Evry, le 21 Septembre 2018, par

Veronica Macaluso

Composition du Jury :

Jean-Yves Salpin Directeur de recherche CNRS, LAMBE	Président
Saulo Vasquez Professeur des Universités, Universidad de Santiago de Compostela	Rapporteur
Valerie Gabelica Directrice de recherche INSERM, INSERM U1212, Université de Bordeaux	Rapporteure
Anne Boutin Directrice de recherche CNRS, Ecole Normale Supérieure	Examinatrice
Debora Scuderi Maitre de Conférence, Université Paris XI	Examinatrice
Aude Simon Chargée de recherche CNRS, LCPQ, Tolouse	Examinatrice
Riccardo Spezia Chargé de recherche CNRS, LAMBE	Directeur de thèse

Contents

Acknowledgements

There are certainly many people I would like to say thank you, which for different reasons have contributed to make great these three years of PhD. I first want to thank my PhD advisor, Dr. Riccardo Spezia, for the precious time, support and help he gave me, especially during these last months, in which he has been correcting my thesis with patience and as fast as I wrote it. I thank him for all the scientific discussions, to have introduced me to other researchers and to have helped me in any situation. I say many thanks to the people in Evry (included people that have left), in particular to Dr. Sana Bougueroua, Flavio Sirio Brigiano, Dr. Daria Galimberti, Fabrizio Creazzo, Simone Pezzotti, Louis Potier, Dr. Ana Martin-Somer, Dr. Estefania Rossich-Molina, Anastasiia Halushkina, Dr. Yannick Jeanvoine, Prof. Marie-Pierre Gaigeot and Dr. Alvaro Cimas. I want to specifically thank Dr. Ana Martin-Somer and Dr. Estefania Rossich-Molina for their precious friendship, working together and supporting each other. I thank Dr. Ana-Martin Somer and Dr. Yannick Jeanvoine to have been developing together the software that analyzes the trajectories and Dr. Ariel Perez-Mellor for the useful discussions. I thank a lot also Dr. Daria Galimberti, not being in my research group but however, she gave me a lot of useful advise. I want to thank, in particular, her and Flavio Sirio Brigiano for the company during the "long trip" from Paris to Evry and for enjoying time together after work.

I thank a lot Dr. Debora Scuderi for my PhD annual revision, for the precious collaboration and to be always really kind in all the discussions. I thank Prof. W. L. Hase, with whom I collaborated during these three years, for his advise, help and for hosting me at the Texas Tech University in Lubock (USA). Thank you also to Prof. Emilio Martinez-Nuñez, that hosted me in the Universidad de Santiago de Compostela for one week and helped me in my work.

I would like to thank the Dr. Valerie Gabelica and Prof. Saulo Vazquez to have accepted to be reviewers of the thesis and and Dr. Anne Boutin, Dr. Debora Scuderi, Dr. Aude Simon and Dr. Jean-Yves Salpin for accepting being part of the jury.

I want finally to express my gratitude to my family for their continuous help and support.

Publications and scientific communications

Publications

The work presented in the thesis gave place to the following scientific publications:

1. V. Macaluso, Z. Homayoon, R. Spezia and W. L. Hase. Threshold for Shattering Fragmentation in Collision-Induced Dissociation of the Doubly Protonated Tripeptide TIK$(H^+)_2$. Phys.Chem.Chem.Phys. 20, 19744 (2018). (Chapter 5);

2. Z. Homayoon*, V. Macaluso*, A. Martin-Somer, M. C. N. B. Muniz, I. Borges Jr., W. L. Hase and R. Spezia. Chemical dynamics simulations of CID of peptide ions: comparisons between TIK$(H^+)_2$ and TLK$(H^+)_2$ fragmentation dynamics, and with thermal simulations. Phys.Chem.Chem.Phys. 20, 3614-3629 (2018). *ZH and VM equally contributed. (Chapter 5);

3. Z. Homayoon, S. Pratihar, E. Dratz, R. Snider, R. Spezia, G. L. Barnes, V. Macaluso, A. Martin-Somer and W. L. Hase. Thermal Dissociation of the TIK$(H^+)_2$ Tripeptide. Mechanisms, Kinetic Parameters, and Comparison with CID. J. Phys. Chem. A 120, 8211-8227 (2016). (Chapter 5);

4. R. Spezia, A. Martin-Somer, V. Macaluso, Z. Homayoon, S. Pratihar and W. L. Hase. Unimolecular dissociation of peptides: statistical vs. non-statistical fragmentation mechanisms and time scales. Faraday Discuss. 195, 599-618 (2016) (Chapter 4).

Other manuscripts, that are not yet published but which are in advanced preparation, are:

1. V. Macaluso, D. Scuderi, M. E. Crestoni, S. Fornarini, E. Dalloz, E. Martinez-Nuñez and R. Spezia. Fragmentation of L-Cysteine sulfate anion. In preparation. (Chapter 6);

2. V. Macaluso, A. Martin-Somer, Z. Homayoon, V. Macaluso, G. Barnes, S. Pratihar, K. Song, W. L. Hase and R. Spezia. Role of Chemical Dynamics Simulations in Mass Spectrometry Studies of Collision Induced Dissociation and Collisions of Biological Ions with Surfaces. In preparation. (Chapter 4).

Scientific communications

The above work led to the following scientific communications:

1. Effect of Different Activation Methods on CID: Experiments and Chemical Dynamics Simulations on the L-Cysteine Sulfate Anion. Oral communication at the *American Society of Mass Spectrometry* (ASMS) in San Diego (CA, USA) (June 2018);

2. Effect of Different Activation Methods on CID: Experiments and Chemical Dynamics Simulations on the L-Cysteine Sulfate Anion. Oral communication at the *Journees Theorie, Modelisation et Simulation* (JTMS) in Paris (June 2018);

3. V. Macaluso, D. Scuderi, M. E. Crestoni, E. Martinez-Nuñez, W. L. Hase and R. Spezia. Understanding the Gas-phase Fragmentation Mechanism of Deprotonated L-Cysteine-sulfate via Direct Dynamics. Poster presentation at WATOC in Munich (August 2017);

4. V. Macaluso, A. Martin-Somer, Z. Homeyoon, W. L. Hase and R. Spezia. Statistical *vs* Non-Statistical effects in Unimolecular Fragmentation of Peptides by Direct Molecular Dynamics. The effect of the activation method. Poster presentation at CCP2017 IUPAP conference in Paris (July 2017);

5. V. Macaluso, A. Martin-Somer, Z. Homeyoon, W. L. Hase and R. Spezia. Statistical *vs* Non-Statistical effects in Unimolecular Fragmentation of Peptides by Direct Molecular Dynamics. The effect of the activation method. Poster presentation at CHITEL conference in Paris (July 2017);

6. V. Macaluso, D. Scuderi, M. E. Crestoni, W. L. Hase and R. Spezia. Understanding the MS dissociation mechanism the L-Cysteine-sulfate via molecular dynamics. Correlation between the activation method in MD and simulations. Poster presentation at the journee des doctorant de l'Ecole 2MIB de l'Universite Paris Saclay (October 2016). Poster prize;

7. V. Macaluso, A. Martin-Somer, Z. Homayoon, S. Pratihar, W. L. Hase and R. Spezia. Gas-Phase Collision Induced Dissociation of peptide ions by Molecular Dynamics Simulations. Poster presentation at the CECAM workshop in Paris (April 2016);

8. V. Macaluso, A. Martin-Somer, Z. Homayoon, S. Pratihar, W. L. Hase and R. Spezia. Gas-Phase Collision Induced Dissociation of the TLK$(H^+)_2$ peptide ion by Molecular Dynamics Simulations. Poster presentation at the ASMS conference in San Antonio, Texas, USA (June 2016).

Chapter 1

Introduction

Mass spectrometry (MS) is a widely used analytical technique that can identify unknown chemical species within a sample and quantify known compounds [1]. It can also elucidate structures and chemical properties, like the reactivity, of chemical systems [2]. Important applications of MS belong to the fields of proteomics, metabolomics, drug discovery, pollution control, forensic science, gas-phase ion chemistry, etc. [3, 4, 5, 6].

In mass spectrometry ions are analysed in the gas-phase as a function of their different mass over charge (m/z) ratio and mass spectra with the ions relative abundances as a function of the m/z values are obtained. Originally the main limit of MS was the non-applicability of the technique to non-volatile species, as peptides, oligosaccharides, etc. and that it was possible to analyse only compounds with a limited m/z ratio [2]. MS became a really powerful technique when it has been possible to couple the instrument with *electrospray ionization* (ESI) and *matrix-assisted laser desorption ionization* (MALDI) ion sources, which ones are capable to bring non-volatile ions into the gas phase and can ionize large molecules forming multi-charged ions, so that the m/z ratio is decreased [2]. ESI and MALDI appeared on the MS scene almost simultaneously in the 1988 and revolutionized biological MS, so that are still the most used ions sources for macromolecules [1].

The ions that are produced through the ion source can be then analysed as a function of m/z through the MS analyser and detected. Another possibility is to fragment the ions and to analyse and detect the consequent fragmentation products (MS/MS). Collision-induced dissociation (CID) [2], for example, is used in MS to fragment the ions through the impact with an inert gas in the collision cell. In this way, the collisional gas-ion translational energy that is

converted into internal energy of the ion can induce its dissociation, which can be influenced by many factors. Statistical theories of dissociation (see section 3.2), based on the assumption that a complete internal energy redistribution takes place before the fragmentation, can only be applied if certain experimental conditions are fulfilled. The characteristics of collision cell, in particular, are really important in determining the experimental results [2, 7]. CID spectra are measured using triple quadrupole (the first quadrupole selects an ion that is fragmented in the second one and the third quadrupole analyses the fragments), Paul ion trap, ion cyclotron resonance (ICR) etc. [2]. The quadrupole is made up of four rods of circular hyperbolic section and, when working as a collision cell, the ions are accelerated trough, colliding with the inert gas present. Other kinds of collision cells are the ion traps: they accelerate the ions in a loop, so that the ions are subjected to a "trapped" motion along with they collide with the gas and eventually fragment [2]. The collision set-up used for the experiment is really important and should be always considered to interpret CID spectra. In fact, the collisional energy in a quadrupole is generally "higher" than in an ion trap and one collision can be enough to allow fragmentation. It can also cause a local activation and the ion can fragment before any internal vibrational relaxation (IVR). On the other hand, in an ion trap the ions collide almost continuously with the buffer gas. These multiple and low-energy (lower than in a quadrupole) collisions can thermalize the ion and the dissociation reaction intermediates, leading generally to statistical fragmentations [7].

In the present thesis, we focus on the fragmentation kinetics and mechanisms of CID and on the role of chemical dynamics in the interpretation of experimental results.

Reactivity and kinetics of chemical reactions are generally studied theoretically through a quantum chemistry (QC) approach based on the depiction of a reaction potential energy surface (PES) [8, 9]. The minima and transition states structures that are found can also be used to determine microcanonical rate constants [10]. Beside when studying flexible and large systems the number of stationary points, constituting the PES, increases considerably and this approach becomes inconvenient, a major disadvantage is represented by the necessity of knowing all the reaction products. An alternative is to use chemical dynamics simulations, which possibility was pioneered by W. L. Hase and co-workers in studying CID of different chemical systems [11, 12, 13, 14, 15]. Chemical dynamics allows to discover unknown dissociation products and mechanisms, which discovery component is really important because it allows to obtain theoretical mass spectra independently from the experiments [10]. There is in fact the necessity of a better understanding of fragmentation mechanisms in order to identify and clarify missing or misidentified data in the current search databases for mass spectra.

Studies have shown that 40-70 % of high signal/noise tandem mass spectrometry (MS/MS) spectra cannot be matched to predicted protein spectra by widely used search engines [16]. Peptides, for example, often fragment differently from what it is expected, due to scrambling phenomena [17], formation of cyclic structures [18], dissociation in more than one point along the backbone [19, 20], or side chain fragmentation along backbone fragmentation [21], this last relevant in the case of post-translational modifications.

To show how simulations can be used to help experiments in recognizing peaks in the mass spectra of different biological systems, indpendently by a mass spectra search engine, the group of R. Spezia has been studying CID of systems belonging to different fields, like nucleobases [22], carbohydrates [23], peptides [24] and amino acids [25]. For them we analysed mechanisms, reaction products and in some cases we obtained theoretical mass spectra. In order to simulate better the experiments we activate the fragmentation in simulations in two different ways (or in one way depending by the the kind of fragmentation expected or obtained in the experiments). In the so-called collisional simulations we simulate explicitly the collision between the gas and the ion and this allows to recover information, like the energy transfer, not measurable by the experiments. Unfortunately, this activation mode is limited in time-scale that is explored: in the experiments there is always time for a complete internal vibrational relaxation before the fragmentation, but we can simulate only the first 10-20 ps after the collision, observing principally the less statistical pathways. However, we can investigate longer time scales using internal energy activation simulations (in which we redistribute statistically the ion internal vibrational energy at the beginning of the trajectories), from which, generally, it is possible to obtain canonical rate constants by observing the reactant population decay as a function of the time. The rate constants obtained in such a way include the intrinsic anharmonicity of the system [26], unlike results coming from electronic structure theory calculations. Thus, another important advantage of simulations over a static PES approach is the possibility of observing both statistical and non-statistical (or dynamical) fragmentations. In particular, the last one has been shown to be also an important component to explain the fragmentation of ions in the gas phase [27, 28, 29, 30, 31]. One example of non-statistical behaviour is represented by shattering fragmentation [32, 33], which represents one of the two limits fragmentation dynamics for CID: the ion fragments immediately, i.e. "shatters", after the collision.

We thus used both collisional and internal energy (or thermal) activation in order to obtain both the statistical (long time-scale) and dynamical (short time-scale) contributions to the fragmentation and in order to interpret the results obtained using the two limit activation modes, i.e. the quadrupole and the ion trap collision cells.

The manuscript is thus organized as follows:

- Chapter 2: description of the principles of CID experiments, a general overview of CID studies on biomolecules and of theoretical models for CID;
- Chapter 3: basis of the theoretical methods used for this thesis, i.e. for the Rice, Ramsperger, Kassel and Marcus theory(i), for chemical dynamics simulations (ii) and for QC methods (iii).
- Chapter 4: description of the methodology. It is shown how chemical dynamics can be used to study CID and the information that can be recovered. We first show how to model the ion-gas interaction energy and the consequent information about the energy transfer that we collected; then we report our understanding of shattering obtained by the analysis of the trajectories; finally we compare statistical and non statistical fragmentations in simulations using the di-proline anion as a model.
- Chapter 5: analysis of the fragmentation of the doubly charged tripeptides ions $TIK(H^+)_2$ and $TLK(H^+)_2$ as obtained by chemical dynamics simulations.
- Chapter 6: study of the L-Cysteine sulfate anion by means of experiments and a coupling between a PES approach and simulations.

Chapter 2

Principles of collision induced dissociation

Collision induced dissociation (CID) or collision activated dissociation (CAD) [3] is a well known mass spectrometry fragmentation technique whose aim is to study the fragmentation of ions through the impact with an inert buffer gas. In this way the collisional translational energy of the ion is converted into internal energy which can induce its dissociation.

Mass spectrometry is mainly used as an analytical tool, i.e. for the identification, sometimes also for the quantification, of one or more species through simply mesuring their mass over charge (m/z) ratio or by detection of their fragmentation patterns. Important applications belong to the biological interest, like proteomics, metabolomics, drug discovery, pollution control, forensic science, etc. [3, 4]. Another important application is the gas-phase ion chemistry, which mainly studies the reactivity and kinetics of uni- and bi-molecular reactions [5].

In the present chapter an overview of the experimental apparatus is first given in section 2.1. Then the actual state of art in CID studies on amino acids on peptides is summarized in section 2.2, while a general overview about theoretical methods that can be used to study collision-induced dissociation reactions is given in section 2.3.

2.1 Experimental overview

Mass spectrometry is a technique by which it is possible to investigate the nature and the chemistry of ions in the gas phase. For the study of amino acids and peptides, which are normally dissolved in solution, it is first necessary to bring them in the gas phase. This is done in the *ion-source*. The second step consists in separating the ion(s) as a function of their different m/z. They are further detected and the resulting signal will be elaborated to give a *mass spectrum* with a signal for each m/z having an area proportional to the relative abundance. An example of mass spectrum is shown in figure 2.1, while figure 2.2 shows a schematic mass spectrometry instrument.

Collision-induced dissociation in mass spectrometry uses an intermediate step between the

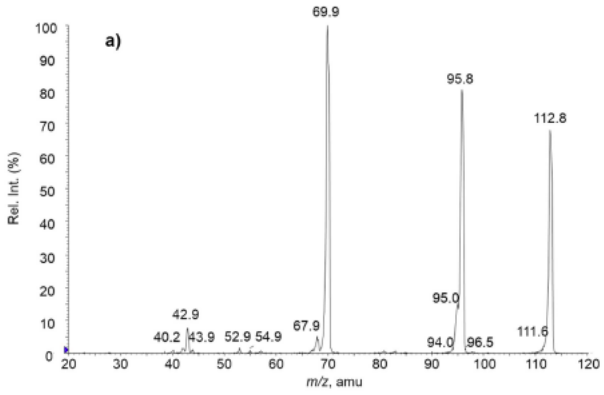


Figure 2.1: Experimental MS/MS spectrum of protonated uracil (taken from reference [22]).

gas-phase ions generation and m/z analysis, that is the dissociation of the ions by collisions with an inert gas in the so-called *collision cell*. This step produces new fragmented ions:

$$m_p^+ \rightarrow m_f^+ + m_n \tag{2.1}$$

where m_p^+ is named *precursor ion* and m_f^+ and m_n are the ion and neutral fragments respectively. A further step can precede the fragmentation, that is the ion selection: another mass spectrometer (used as a filter) is used to select, by m/z separation, the ion that will be further fragmented in the collision cell. Since in this case two mass spectrometers are used, the experiment (MS) is referred to as Tandem Mass Spectrometry (MS/MS or MS^2). In figure 2.3 is then shown, schematically, a triple-quadrupole instrument that can be used to perform MS/MS. The first and third quadrupole (Q1 and Q3) (see section 2.1.3) are mass spectrometers or mass

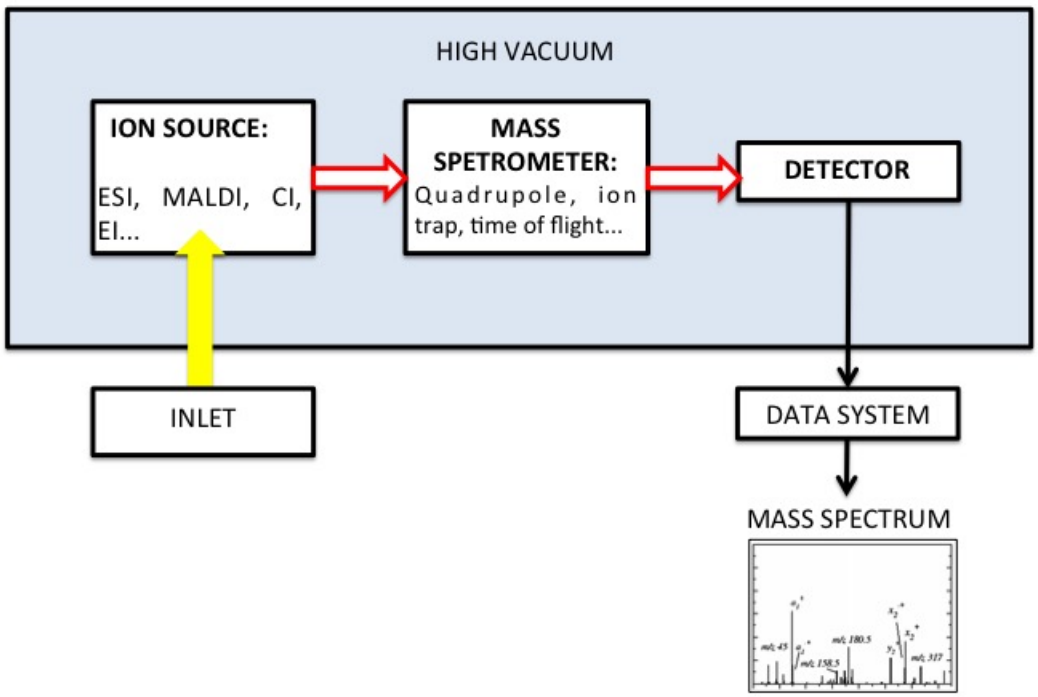


Figure 2.2: Schematic picture of a MS apparatus. The sample is injected into the high vacuum zone in which it is brought to the gas phase through the ion source. The ions are then analysed as function of their different m/z and detected. The signal is elaborated from a data system and finally converted in a mass spectrum.

analysers, while the second one is a collision cell (q2) (see section 2.1.2).

Instruments with several analysers in series, as in figure 2.3 can be scanned in many ways [2]. For example, in the so-called "product ion scan" Q1 selects an ion with a specific m/z ratio and it collides inside q2, fragmenting. Then Q3 will analyse the fragments produced in the q2. In this way one can obtain information about the fragmentation of a specific ion. Another possibility is to perform the "precursor ion scan", that is used to understand which precursor produces a certain fragment. In this case Q3 select the fragment of interest while the Q1 is analysing the different m/z [2].

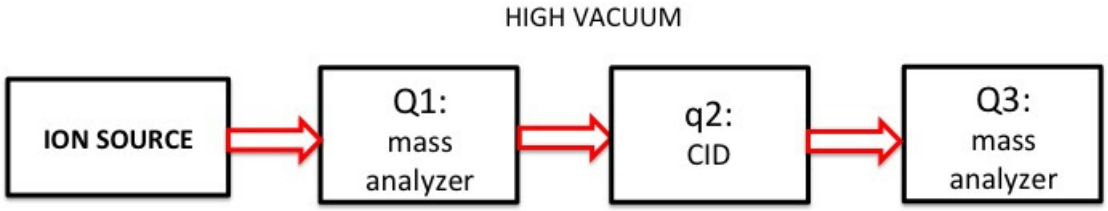


Figure 2.3: Representation of a triple quadrupole mass spectrometer instrument. Q1 and Q3 are mass analysers or mass spectrometers, while q2 is a collision cell.

2.1.1 Ion source

The most used ion sources for MS of biomolecules are *laser desorption* (LD, or MALDI for matrix assisted laser desorption ionization) and *electrospray ionization* (ESI) [34]. They are both considered "soft" techniques for the capacity of letting molecules intact during the ionization process.

In LD the molecules are crystallized and a laser is used to make them to absorb energy. Gas-phase molecules are produced and, successively, reacting with a proton donor species, they produce finally the gas-phase ions. This technique is often matrix assisted (MALDI), causing generally high chemical background, which can be a problem in the low m/z range (m/z <500) for fragments detection, [2]. For this reason for certain small size biomolecules ESI is more indicated.

The electrospray ionization works forcing the solution containing the analyte through a capillary to which a voltage is applied producing an electric field. Charged droplets emerge from the spray (from which the name of the technique), forming the so-called *Taylor cone*. From the droplets is then possible to form bare ions in accord with three different mechanisms [7]:

- The *ion evaporation* model (IEM) [7] generally holds for ions with low molecular weight. The mechanism is based on the ejection of small solvated ions from the droplet surface due to the droplets electric field [35];

- The *charge residue* model (CRM) is more used for globular species. According to this model, due to the heat and the bath gas encountered by the droplets, they start to evaporate until the charge repulsion causes a Coulomb explosion. This produces "naked ions", i.e. ions not solvated by molecules of the solvent;

- Finally, for disordered polymers, the *chain ejection* model (CEM) was recently proposed by Konermann and coworkers [36].

The analyte molecules M can now emerge as ions MH^{n+} (or $((M\text{-}nH)^{-}$ in the negative mode). Larger molecules can be produced often with a high charge. Afterward, the ions are ready to enter in the high vacuum system of the mass instrument [37].

The ions can be now selected or directly activated by collisions, slow heating, electron transfer, etc.. Given the purpose of the thesis, only CID will be considered in the following.

2.1.2 Ion activation in collision-induced dissociation

In CID an ion is activated through one or more collisions with an inert gas in the vacuum. The collisional translational energy is thus converted into internal energy of the ion, i.e. in vibrational and rotational energy.

Statistical theories of dissociation (see section 3.2) are based on the assumption that a complete internal energy redistribution between the normal modes takes place before the fragmentation. However, this is not what always happens. Many factors, like e.g. the number of collisions, the collisional energy, the dimension of the ion or its structure, as well as its flexibility, etc. can influence the way in which the internal energy is distributed between the normal modes. Moreover, all those factors can also influence the percentage of translational energy converted in vibrational and/or rotational energy of the ion. The number of collisions and the collisional energy of each shock depends on the instrument that is used.

The quadruple combines together an electrostatic field and a radio-frequencies oscillating electric field and works as an ion transmitter: the ions are accelerated "linearly" toward the quadrupole plates in order to collide with the gas and to further reach the detector.

If quadrupole plates are used to form an *ion trap*, the instrument takes the name of *Paul trap*. As in all the traps, the ions are accelerated in a "trapped" motion in the collision cell.

In the limit of low pressure, the activation mode of a quadruple can be thought as a linear single collision activation. Note that this is a limit an ideal case; in reality it is very rare to have less than ten collisions. The number of collisions depends on the length of the quadrupole and the gas pressure and the mean free path can be estimated by a simple kinetic gas theory

relationship, as it follows:

$$L = \frac{k_B T}{\sqrt{2} p \sigma} \tag{2.2}$$

where k_B is the Boltzmann constant, T the temperature, p the pressure and σ the collision cross-section. σ equals to πd^2, where d is the sum of the radii of the stationary and colliding molecules [7].

Anyway, the collisional energy in a quadrupole is generally "high" and one collision can be enough to allow fragmentation. It can also cause a local activation: the ion can fragment before any internal vibrational relaxation (IVR). On the other hand, in an ion trap the ions collide almost continuously with the buffer gas. These multiple and low energy (lower than in a quadrupole) collisions can thermalize the ion and the dissociation reaction intermediates, leading generally to statistical fragmentations. In fact, in this way, a complete IVR is assumed to be done for each intermediate of the reaction. The result is a less energetic and more controlled fragmentation than what can be achieved using a quadruple. When modeling CID the disadvantage of ion trap experiments is that in simulations it is not possible to set the activation energy accordingly to experiments since in the latter the energy is not well defined.

2.1.3 Ions analysis

The ion activation and consequent dissociation result in the formation of many fragments that can be characterized by their m/z ratio. Depending on the mass analyser that is used the principles with which the ions are analysed can be different. In the following the analysis principles of some of the most common mass analysers will be outlined.

Time of flight mass analyser

Time of flight (`TOF`) mass analyser separates the ions as a function of their different velocities. The ions present in the cell are accelerated into the instrument by a potential V, so that their kinetic energy, E_k, is equal to:

$$E_k = \frac{1}{2} m v^2 = zV \tag{2.3}$$

Thus the ions with a different m/z will have a different velocity:

$$v = \sqrt{\frac{2zV}{m}} \tag{2.4}$$

What it is actually calculated is the time that each group of ions needs to cover the distance L of the TOF tube in which they are accelerated, i.e.:

$$t = \frac{L}{v} = \frac{L\sqrt{\frac{m}{z}}}{\sqrt{2V}} \tag{2.5}$$

However, in modern days the ions do not travel anymore in the tube in such a simple way but modifications as a reflector and/or orthogonal accelerations can be encountered during the path. Continuous ionization sources, as ESI and LD, are generally coupled with a TOF analyser or other similar instrument variations.

Quadrupole

A quadrupole is formed by four electrodes with hyperbolic or cylindric sections perfectly parallel to each other [2]. An electrostatic field U exists between the sections and an oscillating electric field V is applied using radio frequencies. The ions are separated using the trajectories stability as a function of different combinations of U and V. The potential that is applied has then the general form:

$$\Phi_0 = U - V\cos\omega t \tag{2.6}$$

where ω is the angular frequency of the RF field.

The ions accelerated along the $z - axis$ enter the space between the quadrupole rods with constant velocity along this axis. However, they are submitted to accelerations along x and y, due to the forces induced by the electric fields, that are:

$$F_x = m\ddot{x} = -ze\frac{\partial\Phi}{\partial x} \tag{2.7}$$

and

$$F_y = m\ddot{y} = -ze\frac{\partial\Phi}{\partial y} \tag{2.8}$$

where Φ is function of Φ_0:

$$\Phi = \frac{\Phi_0\,(x^2 - y^2)}{r_0^2} = \frac{(x^2 - y^2)\,(U - V\cos\omega t)}{r_0^2} \tag{2.9}$$

By differentiation it is possible to obtain the motion equations of the ions that are subjected to the potential Φ, that are:

$$\ddot{x} + \frac{2ze}{mr_0^2}(U - V\cos\omega t)x = 0 \tag{2.10}$$

$$\ddot{y} - \frac{2ze}{mr_0^2}(U - V\cos\omega t)y = 0 \tag{2.11}$$

$$m\ddot{z} = 0 \tag{2.12}$$

They can be re-written obtaining the Mathieu equations:

$$\ddot{x} + (a + 2q\cos 2\tau)x = 0 \tag{2.13}$$

$$\ddot{y} - (a + 2q\cos 2\tau)y = 0 \tag{2.14}$$

where:

$$a = \frac{8zU}{m\omega^2 r_0^2} \tag{2.15}$$

$$q = \frac{4zV}{m\omega^2 r_0^2} \tag{2.16}$$

$$\tau = \omega t/2 \tag{2.17}$$

x and y values can be found along the time as a function of U and V. In this way two kinds of trajectories will be solution of the equations 2.12: stable trajectories that oscillate around the $z - axis$ with amplitude less than r_0 and exponential increasing amplitude trajectories, that are unstable. a and q can be used to draw stability areas for the ions trajectories, as it is shown in figure 2.4. The red shaded region in panel A represents the stability area for an ion of mass m, i.e. the area with values of U and V such that x and y do not reach the value r_0. In panel B it is possible to see that three ions with m/z m_1, m_2 and m_3 ($m_1 < m_2 < m_3$) have stability regions in common. The black line corresponds to varying linearly a with q, indeed U with V, in order to end up in a selective $a - q$ space for each of the three ions (that are the three shaded zones above the line).

The ions that have stable trajectories pass through the quadrupole along the $z - axis$ and reach the detector.

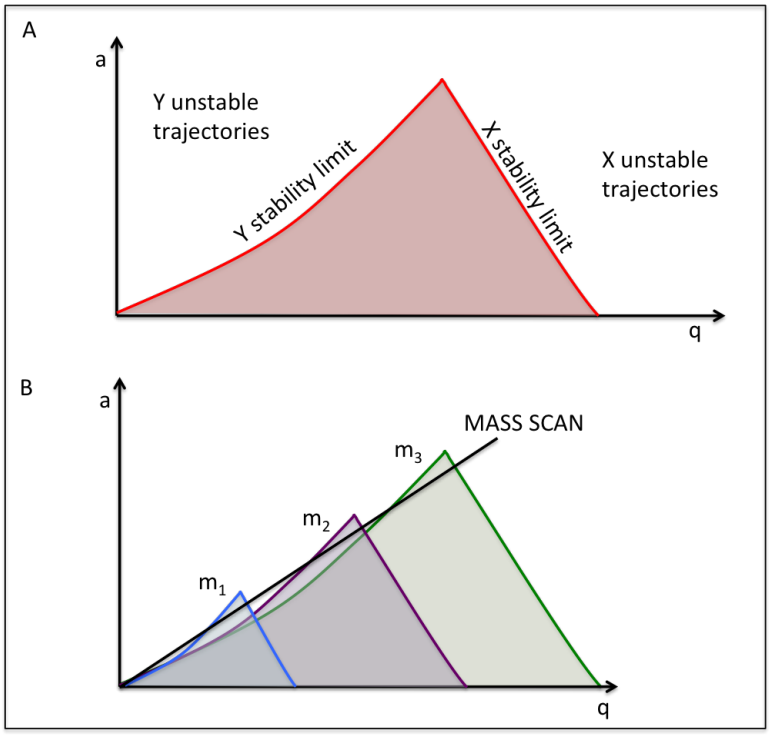


Figure 2.4: Panel A: stability areas as a function of a and q, i.e. values of U and V are such that x and y for the ion of mass m does not reach r_0. Panel B: U (or a) can be changed linearly with V (or q) in order to observe ions with different masses ($m_1 < m_2 < m_3$) successively.

Ion traps

The ion traps use a combination of magnetic and electric fields to trap the ions. With a similar principle by which a quadrupole lets the ions pass with "linear" stable trajectories and reach the detector, ion traps collect the ions with "in-circle" trajectories to send them "packed" to the detector. Examples of ion traps are the Paul trap (electrostatic and oscillating electric field), that is a quadrupole ion trap and the Penning trap (magnetic and oscillating electric field) [2, 7].

The Paul trap is one of the most used ion traps and it is often referred to as, simply, *ion trap*. It can also be called tridimensional trap or 3D ion trap, linear trap or 2D trap and cylindrical trap depending on the trap configuration.

The ion cyclotron resonance (ICR) is a Penning trap and it uses a combination of a fixed magnetic field, B, and an electric field. The magnetic field is used to induce a *cyclotron motion*.

An electric field is then used to increase the cyclotron radii and make the ions orbits coherent. The ions are thus analysed as a function of their different cyclotron frequency:

$$\nu_c = \frac{zB}{2\pi m} \tag{2.18}$$

(This is the equation for ions subject only to a magnetic field).

As suggested by the equation 2.18, ions with the same m/z ration will rotate in phase at their cyclotron frequency and they will induce a current to the detection plates when rotating closer to them. An image current at the different cyclotron frequencies will be registered and thus converted to a mass spectrum through a Fourier transform (FT-ICR). An ICR is thus at the same time a mass analyser and a detector.

2.1.4 Detector

After the ions are analysed they have to be counted. The detection generally is achieved through the impact of the ions (that will arrive in differnt m/z packets) on a surface efficient for the emission of electrons [2]. The current generated from the electrons will be further amplified, registered and converted to give a mass spectrum.

As we explained previously, in some cases, as for the ICR, the analysis and detection can be done at the same time.

2.2 CID of amino acids and peptides

One of the main applications of mass spectrometry is in the field of proteomics [3, 19, 38, 39]. Generally the aim is the identification of proteins in a sample and their characterization, as in the case of post-translational modifications. Identification of proteins is made through one of two approaches, that are the *top-down* and *bottom-up* methods [3, 40]. Both of them involve MS/MS experiments. In top-down proteomics the whole selected protein is fragmented in order to obtain the amino acids sequence. In bottom-up approach, instead, the protein is first digested into peptides. They are then selected and fragmented. In both the approaches the ions generated through fragmentation are mostly identified by comparing them with an existent database [41]. Being able to rely on a peptides fragmentation database is indeed crucial in the case of bottom-up proteomics. However, in some cases the produced mass spectra

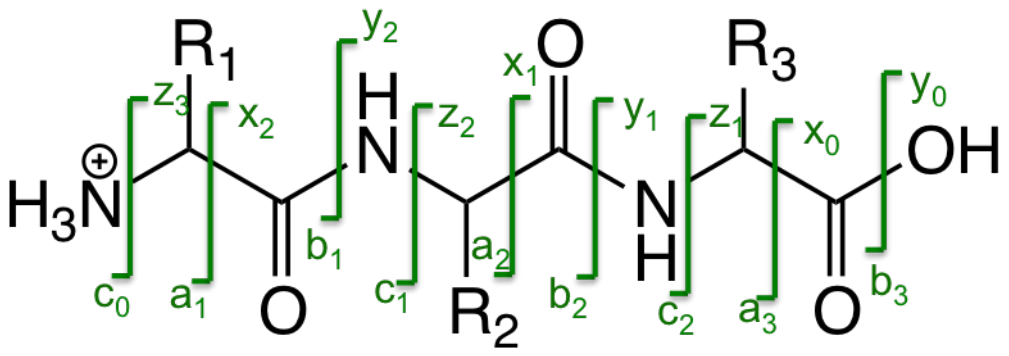


Figure 2.5: Backbone fragmentation sites in singly charged tripeptide with generic R_1, R_2 and R_3 amino acids using the fragmentation nomenclature of Roepstorff and Fohlman [42].

may not match exactly the ones present in the database. This could be due to synthesis or isolation errors, post-translational modifications, missing reference data, etc.. For these cases, to obtain sequence information it is necessary a deeper understanding of the fragmentation.

Dissociation of peptides in mass spectrometry is generally achieved through CID of the protonated peptide ions. The possible backbone fragmentation sites for a protonated peptide are schematized in figure 2.5. a^+, b^+, c^+ fragments contain the N-terminus amino acid residue, while x^+, y^+, z^+ the C-terminus one [42]. The number in subscript indicates the the number of amino acids in the fragment. Whereas all backbone fragments are really important to obtain the sequence-information, low energy collision conditions enhance primarly the amide-bond cleavage leading to b^+ and y^+ ions [43].

For bottom-up proteomics it is fundamental to obtain b_n^+ or y_n^+ ions in an ordered series in order obtain the amino acid sequence. In fact an amino acid mass can be calculated by the mass difference of the ions b_{n+1}^+ and b_n^+. However, some ions in a series can be missing. An ion absence in the mass spectrum may be due, for example, to a successive neutral loss. b_n^+ ions can form particular and different structures that can enhance different consequent fragmentations [19]. b_1 ions, for example, are generally unstable for the consecutive losses of H_2O and CO. By the way, b_1 of nucleophilic amino acids can be detected: a stable cycle can be formed due to the attack of this nucleophile site on the carbonyl function [18]. Larger b ions are instead easier detected and many different structures have been studied and suggested both experimentally and theoretically [19, 44, 45]. Whereas CID of peptides produce generally b or y ions, there are cases in which these fragments are not predominant [19].

The proton mobile model [19, 46, 47, 48] has been developed to explain the different-peptides fragmentation behaviour in the gas phase, but it can also be used to understand fragmentation of amino acids and other organic molecules. It is based on the assumption that ions produced in the gas phase with a soft ionization technique, as ESI or LD, localize the proton in the most basic site. In protonated peptides the most basic sites are the N-terminus and the basic amino acids in the side chains. The "mobile" proton [19] will indeed move to other sites and initiate fragmentations. Exceptions for this model are represented from *charge-remote* fragmentations [3, 19], where the cleavage occurs far from the excess proton.

Differently from CID of peptides, CID of amino acids are not used for proteins sequencing. These experiments are performed, for example, for quantitative analysis, characterize their post-translational modification structures or to confirm their synthesis in the gas phase.

2.3 Theoretical methods for CID

Reactivity and kinetics of chemical reactions are generally studied theoretically through a quantum chemistry (QC) approach based on the localization of the stationary points [8, 9, 49, 50], i.e. minima and transition states (TS), along the reaction coordinate. The result is the minimum energy path, or the minimum potential energy surface (PES), connecting the reagent(s) with the product(s). An example is given in figure 2.6. In order to study the dissociation reaction of a CID experiment the reagent would be the so called *precursor ion*, i.e. the dissociating ion, while the fragments would be then the products. Scanning the PES does not allow only to probe mechanisms for a specific fragmentation pathway but it can also be used to determine the energetics of the different protonation sites [19].

If the studied system is small and dynamical effects have not a big role in the reactivity a static approach is a good method. In fact, it may be possible to depict a complete PES for the system reactivity. Furthermore, the energy barriers obtained can be used to predict kinetic rate constants through existent statistical theories [29, 49]. In particular, in mass spectrometry the focus reaction is the dissociation of the ion that is unimolecular (the gas does not influence the dissociation rate constant) and the most used theory for such kind of reactions is the RRKM (Rice, Ramsperger, Kassel and Marcus) theory, see section 3.2.

One of the most used computational strategies to obtain minima structures of peptides, for example, is based on dynamics simulations using classical force fields (FF) [19]. In fact,

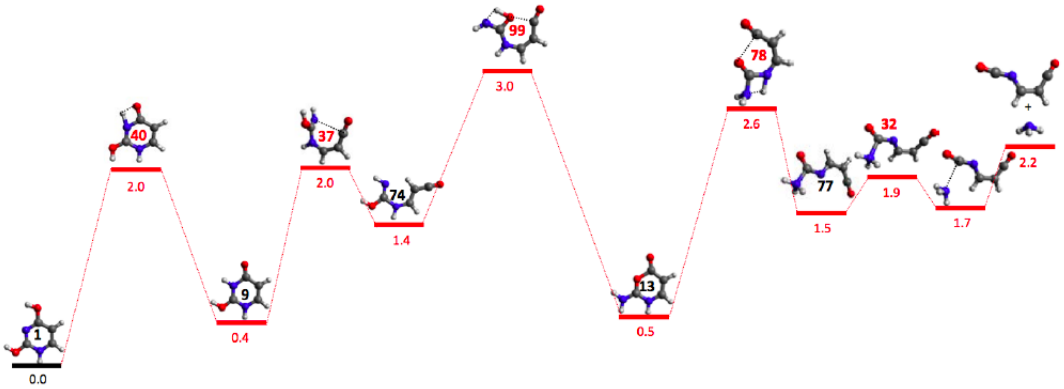


Figure 2.6: Potential energy surface for a dissociation pathway example of protonated uracil [27]. The transition states are labelled in black and the minima in red.

increasing the system size and flexibility it becomes impossible to find all the possible minima structures by means of a PES approach. However, FFs need an accurate parametrization and cannot be used to sample the fragmentations because the bonds are fixed.

In order to study the fragmentation of larger systems semi-automatic methods to find TS were proposed, like the automatic *Transition State Search using Classical Dynamics Simulations* (TSSCDS) [51]. This algorithm is coupled with a code that performss internal energy activation simulations (see section 3.3.3) to sample the reaction phase-space, where it can further search for candidate TSs structures. The latter are further optimized and if they satisfy certain criteria they are saved as TSs. From the new structures it is possible to find minima using an algorithm that follows the intrinsic reaction coordinate (IRC). The new minima can be used then as new initial structures for further simulations [52]. In the end, transitions states and minima can be hopefully connected to give a PES.

Recently, the TSSCDS algorithm was modified in order to search for possible transition states in the phase-space explored by external trajectories. This last version was developed and used in order to study the unimolecular fragmentation of the L-Cysteine-sulfate anion (see chapter 6). In fact, for this study, chemical dynamics were employed using two different activation methods: explicit collision and internal energy activation. One important pathway, responsible for an abundant fragment in the experiments, was obtained with difficulty and mainly from collisional simulations. For this reason an external coupling with the simulations

was adopted. Unfortunately, it was not possible to find all the stationary points for the desired reactions using the coupling with the automated tool, but one important pathway was completely characterized. Also, it was possible to obtain new minima for the L-Cysteine-sulfate anion, one of which resulted to be the global potential energy minimum.

The problem encountered using the TSSCDS code in finding the desired pathway PES can be assigned to the *non-saddle point* transition state character of the last step of the reaction. Moreover, the system is very flexible, so that the automatic tool would spend a lot of computational effort only to find various configurational isomers and pathways.

A solution to study the fragmentation of flexible and/or large systems is to use direct chemical dynamics, which can calculate *on-the-fly* the force by using a QC method [10]. One of the principal advantage of direct dynamics, upon the static PES approach, is that it is not necessary to know which are the reactions products, that are not obvious in the case of the fragmentation of large systems. Moreover, reactions can be dynamically driven and a potential energy profile cannot give a satisfactory description of the phenomenon. In this case the reaction cannot be described by standard kinetics statistical theories [28, 29].

In the literature it is possible to find different cases of *non-statistical* fragmentations behaviours, as for example shattering (see section 4.2). This mechanism was recently reported for the dissociation of two doubly charged tripeptides ions, diprotonated threonine-isoleucine-lysine and threonine-leucine-lysine (see section 5.3), or in both experiments and simulations of the protonated urea [30, 31], protonated uracil [22, 27] and for the Ca^{+}/ formamide complex [29]. Non-statistical *roaming* mechanism (see chapter 6) was also observed in both experiments and simulations, whose studied were recently reviewed [53, 54].

Chapter 3

Theoretical methods

3.1 Introduction

In CID the ion can get a certain excess energy through one or more collisions with the gas. Then the ion can dissociate with an unimolecular mechanism, following first order kinetics. The rate constant of such a reaction is described by the RRKM (Rice, Ramsperger, Kassel and Marcus) theory. The computation of a potential energy surface (PES) has been for long time the standard theoretical method to study CID reactions [7]. However, for large molecules the identification of all the pathways and all the stationary points becomes less feasible. Moreover, especially for high energy conditions, some pathways are unpredictable and cannot be explained using statistical theories.

Chemical dynamics simulations are then illuminating, giving the possibility to discover unexpected fragmentation mechanisms and (or) products. Simulations allow in fact to explore simultaneously the statistical (based on RRKM theory) and dynamical behaviour of collision induced fragmentations.

Going then to apply theoretical methods on the collision induced dissociation processes, chemical dynamics is indispensable to investigate the *dynamics* of the collision and its effect on the consequent ion fragmentation.

First in this chapter, it will be reported a summary about the development of the RRKM theory. Then the "effective" methodologies used in this thesis will be presented, and notably: (i) the final formulation of RRKM theory; (ii) "our" Chemical Dynamics methodology, with

particular importance given to the initial conditions and the semi-empirical Hamiltonians.

3.2 Theory of unimolecular reaction rate constant

An unimolecular reaction happens when a molecule, that is excited by any kind of external stimulus, will have enough energy to react by itself. More precisely, the evolution in time of such a system will depend only on the concentration of one species.

If we consider the reaction A $\longrightarrow$ P, where A is the energized molecule and P the product(s) of the decomposition/isomerization of A, we have an unimolecular reaction, for which:

$$-\frac{d[A]}{dt} = \textit{reaction velocity} = k_{uni}[A] \tag{3.1}$$

where $[A]$ is the concentration of the species A, t the time and k_{uni} is the rate constant.

By integration, it gives the time evolution of the concentration of A as a function of time:

$$[A] = [A_0]e^{-k_{uni}t} \tag{3.2}$$

where $[A_0]$ is the initial concentration of A.

The first and simplest theory to calculate the unimolecular reactions rate constants was proposed already at the beginning of 20's, then further developed by Rice/Rampsperger [55] and Kassel [56] (RRK) theory in the middle 20's. In the 30's, at a time when more was known about potential energy surfaces, also transition state theory (TST) was formulated [57, 58]. In 1951-1952 [59, 60] Marcus blended together RRK and TST theories leading to the actual RRKM formulation.

3.2.1 Collision theory and Arrhenius equation

Between 1916 and 1918 [61] it was proposed the collision theory for bimolecular reactions, A+B $\longrightarrow$ P, for which the fraction N of molecules having enough energy to react is:

$$N = Ze^{-\frac{E_0}{RT}} \tag{3.3}$$

This is the number of molecules that, colliding between each other, acquired enough energy (E_0), so that they will go into a chemical change. E_0 is the activation energy, $e^{-\frac{E_0}{RT}}$ is the probability for a molecule to have energy E_0, while Z the collision frequency, equal to:

$$Z = N_A N_B \sigma \sqrt{\frac{8 k_B T}{\pi \mu_{AB}}} \tag{3.4}$$

where N_A and N_B is the number of molecules A and B respectively, σ is the cross-section, k_B is the Boltzmann constant and μ_{AM} the reduced mass of the two reactants.

The collision theory bimolecular rate constant is then given by dividing equation 3.3 by the concentration of the species A and B (and by the steric factor). This theory gave the first theoretical explanation from the molecular point of view to the empirical Arrhenius [62] equation:

$$k(T) = A e^{-\frac{E_0}{RT}} \tag{3.5}$$

where $k(T)$ is the canonical rate constant, E_0 is the activation energy, R is the Regnault gas constant, T the temperature and A the pre-exponential factor. In particular, the pre-factor A is a function of the collision frequency and the collision orientation.

When this equation fits the under-exam data, it can be used to evaluate the reaction activation energy, or thermodynamic properties as reaction enthalpy and entropy.

3.2.2 Lindemann theory

The first explanation of unimolecular reactions on the basis of the collision theory, was given by Lindemann [5] in 1922, then re-elaborated by Hinshelwood [63] in 1927. It is the simplest theory for these kinds of reactions and the mechanisms to which it refers is known as Lindemann-Hinshelwood mechanism. It consists of a two-step reaction: a second-order bimolecular step, in which the reactant molecule is excited by collisions in a thermal system, followed by the rate determining unimolecular one.

$$\mathrm{A} + \mathrm{M} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} A^* + \mathrm{M} \tag{3.6}$$

$$A^* \xrightarrow{k_2} \mathrm{P} \tag{3.7}$$

A^* are excited species and their lifetime will be really short, thus the steady-state approximation can be considered:

$$\frac{d[A^*]}{dt} = 0 = k_1[M][A] - k_{-1}[M][A^*] - k_2[A^*] \tag{3.8}$$

It follows that:

$$[A^*] = \frac{k_1[A][M]}{k_{-1}[M] + k_2} \tag{3.9}$$

Using equation 3.9, it is then possible to write the overall reaction rate constant for the reaction described by the reactions 3.6 and 3.7:

$$\frac{d[P]}{dt} = k_2[A^*] = \frac{k_1 k_2[A][M]}{k_{-1}[M] + k_2} = k_{uni}[A] \tag{3.10}$$

This constant is phenomenologically unimolecular:

$$k_{uni} = \frac{k_1 k_2[M]}{k_{-1}[M] + k_2} \tag{3.11}$$

As we can see, there is no dependence or information about the energy of the reactants, neither about their nature (for example which is the number of normal modes of A). The important feature of this unimolecular rate constant is that it has a different behaviour at high and low pressure of M. The reaction rate will, in fact, pass from a second order (in $[A]$ and in $[M]$) to a first order (only in $[A]$) kinetics respectively, as it follows:

$$\text{For } k_{-1}[M] << k_2 : \frac{d[P]}{dt} = k_1[M][A] \tag{3.12}$$

$$\lim_{[M]\to\infty} \frac{d[P]}{dt} = \frac{k_1}{k_{-1}} k_2[A] = k_\infty[A] \tag{3.13}$$

k_1 and k_{-1} do not depend on the energy and can be calculated directly from kinetic collision theory. If the probability of obtaining species A with energy higher than E_0 is $e^{-\frac{E_0}{RT}}$, then we have:

$$k_1 = Ze^{-\frac{E_0}{RT}} \tag{3.14}$$

and

$$k_{-1} = Z \tag{3.15}$$

k_2 can be calculated from equation 3.13, using equations 3.14 and 3.15.

Such a change from a second (equation 3.12) to a first order (equation 3.13) kinetics has been observed experimentally and it was this evidence that brought Lindemann to propose its mechanism against the radiation theory of Perrin [64, 65].
We can derive k_∞ from equation 3.13, and replacing it in 3.11 we obtain:

$$k_{uni} = \frac{k_\infty[M]}{[M] + \frac{k_\infty}{k_1}} \tag{3.16}$$

that with some adjustment becomes:

$$\log \frac{k_{uni}}{k_\infty} = \log(k_1[M]) \tag{3.17}$$

Plotting the experimental observable $\frac{k_{uni}}{k_\infty}$ *vs* $\log[M]$ it is possible to obtain the famous fall-off plots. An example of these plots is given in figure 3.1, which shows a change from second order (the straight-line at low pressures of M) to unimolecular (independent of the pressure). k_1 can be calculated considering equation 3.12. Similarly, k_∞ can be estimated by extrapola-

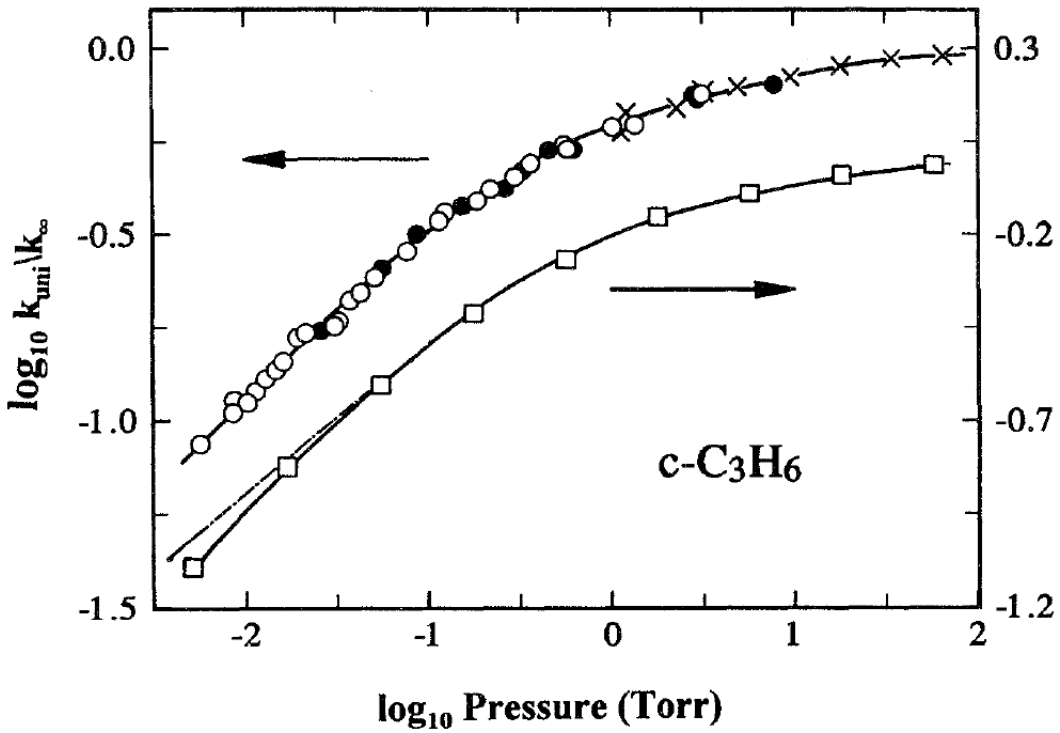


Figure 3.1: Pressure dependence of the unimolecular rate constant for
C-C_3H_6 $\longrightarrow$ H_2=CH-CH_3. The open and closed circles are data from Prichard et al. (1953) while the x's are data of Chambers and Kistiakowsky (1934). The solid lines are the experimental results, the open squares are calculated by Slater (1953) assuming 13 active oscillators and the dashed curve is a Kassel or RRK calculation with 13 oscillators by Prichard *et all* (1953). The figure is taken from reference [5].

tion at high pressures using equation 3.13. Accordly to these Lindeman theory-based calculations, the fall-off region must appear at higher pressure than what observed experimentally.

This was the first failure of Lindemann theory. In addition, plotting $\frac{1}{k_{uni}}$ *vs* $\frac{1}{[M]}$, using calculated values, there is a deviation from linearity.

Today we know that Lindemann theory fails for two main points: (i) there is no reactant internal energy dependence for the activation and neither degrees of freedom are considered; (ii) it is not considered (TST was not still formulated) that a specific internal mode, or more, should be activated to allow the unimolecular reaction step to take place. In Hinshelwood theory it is introduced a dependence on the energy and the ion internal-degrees of freedom in the rate constant of the activation process k_1.

3.2.3 Hinshelwood theory

In unimolecular reactions many more molecules react than the simple collision theory allows for. The Lindemann theory suggested that there might be enough collisions to account for if the energy could be distributed among a considerable number of degrees of freedom. In fact, A was modelled as constituted by s simple equivalent harmonic oscillators with frequency ν, that were used to calculate the probability of the molecule to be activated.

So, we can consider the number of ways of distributing n quanta in s oscillators (so the number of degenerate states for n quanta), that is:

$$g_n = \frac{(n+s-1)!}{n!(s-1)!} \tag{3.18}$$

To calculate the activation rate constant two assumptions are considered: (i) consequently at a collision the probability of deactivation of A^* is unity, as in equation 3.15 (strong collision hypothesis); (ii) molecules A and A^* are in equilibrium. For the (i) it results that the activation rate constant k_1 for the state n is equal to the collision frequency Z of A^*, as in equation 3.14. For (ii) the probability of forming a particular state n is given by the Boltzmann distribution. It follows that the activation rate constant, k_1^n, is equal to:

$$k_1^n = Z\frac{g_n e^{\frac{-nh\upsilon}{kT}}}{q} \tag{3.19}$$

with q the partition function.
The overall activation reaction constant k_1 will be:

$$k_1 = \sum_1^m k_1^n = \sum_1^m Z\frac{g_n e^{\frac{-nh\upsilon}{kT}}}{q} \tag{3.20}$$

where m is the number of molecules in the state n.

Since in thermal experiments the energies are usually large, meaning $E_0 >> kT$, in Hinshelwood theory it is assumed that the energy levels are continuous:

$$dk_1 = Z\frac{N(E)e^{-\frac{E}{kT}}}{q}dE \tag{3.21}$$

where $N(E)$ is the density of states, $N(E)dE$ consequently is the number of states with energy between E and $E + dE$ and dk_1 is the collisional reaction constant to activate the molecule into the said range. The classical overall reaction constant for the activation process will be obtained integrating equation 3.21 between E_0 and ∞, resulting in:

$$k_1 = \frac{Z}{(s-1)!}\left(\frac{E_0}{kT}\right)^{s-1}\exp\left(\frac{-E_0}{kT}\right) \tag{3.22}$$

The formulation of k_1, as in equation 3.22, is still used in following RRK theory [55, 56], but it is used differently. In fact, in Hinshelwood theory it is not specified the number of degrees of freedom that must be considered: they are chosen by test as the ones giving the best accord with the experiments. In the following theories all the degrees of freedom of the molecule A are evaluated in k_1. A consequent internal energy flow is then made responsible of the energy confinement of the energy in the reactive part of the molecule.

3.2.4 RRK theory

The RRK theory was developed by Rice and Rampsperger [55] and independently by Kassel [56]. First of all they stated that E_0, the activation energy, shall be defined as the minimum energy that a molecule needs in order to dissociate. Only internal energy must be included. Moreover, it was recognized that, beyond k_1, also the rate of the unimolecular step should depend on the energy of the energized molecule. In particular, only the molecules with a minimum of energy ε_t deposed in a certain normal mode will be able to react. This energy can be obtained via collisions and distributed with a consequent internal energy flow between the modes, that is a valid assumption when $E >> s$. This means that the oscillators are coupled, therefore they are anharmonic. Consequently, they calculated the number of ways of distributing $n - n_0$ quanta over the s oscillators, with n_0 the number of quanta in the critical normal mode, that is:

$$g_{n-n_0} = \frac{(n-n_0+s-1)!}{(n-n_0)!(s-1)!} \tag{3.23}$$

that divided by 3.18, the way of distributing n quanta in s oscillators, gives:

$$P = \frac{(n - n_0 + s - 1)!n!}{(n - n_0)!(n + s - 1)!} \tag{3.24}$$

that is the probability for the molecule of having n quanta of energies distributed in such a way that n_0 of them are in the critical normal mode.

If the spacing is very small, so n and n_0 large, we can apply the Stirling approximation, that is $h! = \frac{h^h}{e^h}$, and if $(n - n_0) >> s$, then:

$$P = \left(\frac{n - n_0}{n}\right)^{s-1} \tag{3.25}$$

Multiplying equation 3.25 by the critical vibrational frequency, it is possible to obtain the RRK formulation for the unimolecular rate constant.

$$k_2(E) = \nu^{cr} \left(\frac{n - n_0}{n}\right)^{s-1} \tag{3.26}$$

or

$$k_2(E) = \nu^{cr} \left(\frac{E - E_0}{E}\right)^{s-1} \tag{3.27}$$

k_2 in equation 3.26 is the Kassel quantum formulation of the RRK rate constant [56]. In fact, the internal degrees of freedom are treated by means of discrete vibrational energy levels. Instead, equation 3.27 gives the classical version of k_2, which was given by Rice and Rampsperger [55]. By the way, since the equations were obtained for the classical limit $(n - n_0) >> s$, it is common to refer to both the formulations as the classical (microcanonical) rate constant. Note that, differently from RRKM theory, RRK refers to the Lindemann-Hinshelwood mechanism.

k_1 formulation is given by Hinshelwood theory, with the expedient that all the degrees of freedom of A must be included in. Next step from the previous theory is recognizing that the number of ways in which the energy can be distributed into the molecule in order to dissociate should be constrained by confinement of a certain energy in the critical normal mode.

We see that the higher is the excess energy, the higher probability there will be for the dissociating normal mode to reach an energy equal (or higher) to ε_t and indeed the rate constant will have also an higher value. Instead, the higher is the number of oscillators, the smaller will be the rate constant.

Unfortunately the theory gave still big errors in the evaluation of the rate constant calculation. One of the reasons is the approximation $(n - n_0) >> s$ that holds only in few cases. Furthermore, no zero point energy, neither rotations are considered. Moreover, all the oscillators have all the same frequency ν. Finally, they did not take into account the energy flow

process between the activation and the fragmentation process.

3.2.5 RRKM theory

The formulation of RRKM theory, by Marcus and Rice [59] and then by Marcus ([60]), was given much later than the previous ones. In the 30's Wigner and Hirschfelder [66, 67] developed the classical transition state theory, TST. In accord with the formation of an activated complex, the Lindeman-Hinshelwood mechanism was thus reconsidered:

$$\mathrm{A} + \mathrm{M} \underset{k_{-1}}{\overset{k_1(E)}{\rightleftharpoons}} A^*(E,E+dE) + \mathrm{M} \tag{3.28}$$

$$A^*\left(E, E+dE\right) \xrightarrow{k_2(E)} A^\dagger \tag{3.29}$$

$$A^\dagger \xrightarrow{k^\dagger} \mathrm{P} \tag{3.30}$$

where the first activation step is still given by the Lindemann mechanism of equation 3.6, while a new intermediate step is now added, consisting in the IVR. After the IVR, the activated complex, $A^\dagger$, is formed and finally the dissociation step can take place.

The overall reaction rate constant is:

$$\frac{dP}{dt} = k^\dagger[A^\dagger] \tag{3.31}$$

The steady-state approximation is applied before on $[A^\dagger]$, leading to:

$$k^\dagger[A^\dagger] = k_2[A^*] \tag{3.32}$$

and then on $[A^*]$:

$$k_2[A^*] = \frac{k_1 k_2(E)[A][M]}{k_{-1}[M] + k_2(E)} \tag{3.33}$$

For the equations 3.31, 3.32 and 3.33 it is possible to write:

$$\frac{d[P(E)]}{dt} = \frac{k_1 k_2(E)[A][M]}{k_{-1}[M] + k_2(E)} = k_2[A^*] = k_{uni}(E)[A] \tag{3.34}$$

All the k's can be first evaluated for a small energy range, so that $k_{uni}(T)$ can be obtained by integration over the energy between E_0 and E:

$$k_{uni}(T) = \int_{E_0}^{\infty} k_{uni}(E)dE \tag{3.35}$$

$$dk_{uni} = \frac{dk_1 k_2(E)[M]}{k_{-1}[M] + k_2(E)} = \frac{k_2(E)d(k_1/k_{-1})}{1 + \frac{k_2(E)}{k_{-1}[M]}} \tag{3.36}$$

k_2 in equation 3.36 can be estimated considering that at the equilibrium the ratio between $[A^{\dagger}]$ and $[A^*]$ is given by $2k_2/k^{\dagger}$ (with the reverse rate constant of equations 3.29 and 3.30 assumed to be identical). The quantity $2k_2/k^{\dagger}$ can be evaluated calculating the number of ways of distributing energy between the active molecule and the activated complex. The low limit is the fixed energy: the sum of activation and zero point energies. Moreover, in the case in which rotation is thought to be important in the formation of the activated complex, its energy contributions are considered. $k^{\dagger}$ is calculated assuming to be equal to the frequency of the dissociating normal mode.

The term $d(k_1/k_{-1})$ in equation 3.36 is instead the chance for the active molecule of having an energy between E and $E + dE$ with k_{-1} calculated again from kinetic theory using the strong collision assumption (equation 3.15).

$k_2(E)$, the microcanonical unimolecular rate constant, is the most important feature of RRKM theory. While at the beginning comparison with experiments were performed using the derived RRKM expression for the canonical rate constant, described in the next session, later advances in energy selected experiments allowed to test directly the microcanonical rate constant $k_2(E)$. This allowed to remove the contribution to the calculated rate constants of the activation and de-activation steps, which determination is still uncertain. The rest of this section will be, indeed, focused on the derivation of $k_2(E)$. Many formulations of the RRKM theory exist, often differing only in the way the rotational contribution and the transition state are treated. Here it will be followed the derivation of the theory given by Baer and Hase [5], that uses the variational transition state theory [66] in which the reaction is seen as a reaction flux in the phase space. In this derivation the transition state is "rigid", indeed located at the saddle point, and the rotation contribution to the reaction rate constant is not considered.

Derivation of microcanonical RRKM theory

A molecule, having m degrees of freedom, can be described by m momenta p and m positions q in a phase-space of $2m$ dimensions. If the molecule is part of the microcanonical system

($E = \text{const}$), it will be then described by a $2m-1$ hyper-surface of the phase space.

A unimolecular reaction can be, thus, described as a reaction flux in the phase space, where the non-turning reactants-products dividing surface is called critical surface. For "rigid" transition-states the latter is located at the saddle-point, whose coordinates, $q^\dagger$ and $p^\dagger$, are perpendicular to all the others.

To continue with a statistical mechanics treatment, statistical mechanical quantities have to be introduced, as the sum and density of states. The sum of states $N(E)$ in the range of energy between 0 and E, is given by:

$$N(E) = 1/h^n * PSV = 1/h^n \int_0^E \int dp_1...dp_n dq_1...dq_n \tag{3.37}$$

with $n = 3N-6$ and N the number of molecules; p and q are momenta and positions respectively and PSV is the phase space volume. Since $\Delta p \Delta q \geq \hbar/2$, it follows that h^n is the smallest unit of a N-molecules PSV and $N(E)$ represents the smallest representation of the PSV and gives the sum of states in the range of energy between 0 and E. Note that the quantization is obtained dividing by h^n, otherwise the quantity in equation 3.37 would be classical. Thus we can define the density of states, $\rho(E) = dN(E)/dE$, i.e. the number of states per energy unit. From equation 3.37, it can be written as:

$$\rho(E) = 1/h^n * PSV = 1/h^n \int_E \int dp_1...dp_n dq_1...dq_n \tag{3.38}$$

Indeed, to use such statistical quantities, it must be assumed that the population density is uniform over all the phase-space. In this way, the fraction of molecules at the critical surface can be given by the fraction of phase-space at the dividing-surface, i.e. at any time:

$$\frac{dN(q^\dagger, p^\dagger)}{N} = \frac{dq^\dagger dp^\dagger \int_{E-\varepsilon_t-E_0} ... \int ... \int dq_1^\dagger ... dq_{n-1}^\dagger dp_1^\dagger ... dp_{n-1}^\dagger}{\int_{E_0} ... \int dq_1 ... dq_n dp_1 ... dp_n} \tag{3.39}$$

where $\frac{dN(q^\dagger,p^\dagger)}{N}$ is the fraction of molecules described by the coordinates into the range $q^\dagger, q^\dagger + dq^\dagger$ and $p^\dagger, p^\dagger + dp^\dagger$, E_0 is the activation energy and ε_t is the translation energy at the saddle point. The energy is referred to the zero-point energy.

The reaction rate, so the reaction flux, can then be calculated as the time derivative of the molecules at the critical surface:

$$Flux = \frac{dN(q^\dagger, p^\dagger)}{dt} \tag{3.40}$$

Replacing equation 3.39 in equation 3.40, and considering that $dq^\dagger/dt = \text{velocity} = p^\dagger \mu$, we obtain:

$$\frac{dN(q^\dagger, p^\dagger)}{dt} = \frac{N\frac{p^\dagger dp^\dagger}{\mu} \int_{E-\varepsilon_t-E_0} ... \int ... \int dq_1^\dagger ... dq_{n-1}^\dagger dp_1^\dagger ... dp_{n-1}^\dagger}{\int_{E_0} ... \int dq_1 ... dq_n dp_1 ... dp_n} \tag{3.41}$$

with μ the reduced mass of the two separating fragments.

Writing the translational energy as $\varepsilon_t = \frac{p^{\dagger 2}}{2\mu}$, we notice that the term $\frac{p^\dagger}{\mu}$ is equal to $\frac{d\varepsilon_t}{dp^\dagger}$, and so re-arranging we have that $\frac{p^\dagger dp^\dagger}{\mu} = d\varepsilon_t$, that can be replaced in equation 3.41, giving:

$$\frac{dN(q^\dagger, p^\dagger)}{dt} = \frac{N d\varepsilon_t \int_{E-\varepsilon_t-E_0} ... \int ... \int dq_1^\dagger ... dq_{n-1}^\dagger dp_1^\dagger ... dp_{n-1}^\dagger}{\int_{E_0} ... \int dq_1 ... dq_n dp_1 ... dp_n} \tag{3.42}$$

Equation 3.42 is indeed the reaction rate, expressed as the number of molecules N multiplied by the rate constant $k(E, \varepsilon_t)$. Recalling equation 3.38, we can see that the denominator of equation 3.42 is the density of state, $\rho(E)$, multiplied by h^n. We can see that the numerator is still a density of state, but obtained integrating over one less dimension. Therefore we can write:

$$k(E, \varepsilon_t) = \frac{\rho^\dagger(E - E_0 - \varepsilon_t)}{h\rho(E)} \tag{3.43}$$

This represents the rate constant that takes into account all the possible ways of partitioning the "non-fixed" energy between the vibrational degrees of freedom of the activated complex, given the translational energy of the separating fragments. To obtain, instead, the overall reaction rate constant, it is necessary to integrate over all the translational energies available, that is:

$$k(E) = \frac{\int_0^{E-E_0} \rho^\dagger(E - E_0 - \varepsilon_t) d\varepsilon_t}{h\rho(E)} = \frac{N^\dagger(E - E_0)}{h\rho(E)} \tag{3.44}$$

where $N^\dagger(E - E_0)$ is the sum of states of the activated complex from 0 to $E - E_0$. This is the standard RRKM expression, for the unimolecular rate constant $k(E)$, neglecting the rotational term. Thus, when a microcanonical system is formed, if the IVR is very fast compared to the reaction rate, i.e. a statistical distribution of the energy is maintained until the decomposition, then a single exponential decay of the population versus the time is observed, with $k_2(E)$, now called $k(E)$ being the observed rate constant.

Also, the RRKM theory is derived considering separability between the critical coordinates and all the others. This, however, is a good assumption only at relatively low energies and in absence of strongly anharmonic normal modes.

The equation 3.44 is generally used in a similar form:

$$k(E) = \sigma \frac{N^\dagger(E - E_0)}{h\rho(E)} \tag{3.45}$$

where σ is the reaction symmetry factor, or reaction degeneracy. The meaning of σ can be understood through a simple example. Consider the system $X - Y_3$, having three symmetric

bonds $X-Y$. Any of the three identical bonds $X-Y$ can be broken. The sum of states is then calculated for the transition state relative to one of the three possible dissociative bonds. However, it will have to be multiplied by 3 (the number of equivalent bonds) in order to calculate the overall reaction rate constant (there will be more possibilities to form the product $X-Y_2+Y$).

When the rate constant $k(E)$ is calculated using partition functions and the rotational contribution is also included, the role of σ will be replaced by the ratio between the transition state and reagent moment of inertia I.

3.2.6 Canonical rate constant

While we have just formulated the microcanonical rate constant, $k(E)$, the canonical rate constant, $k(T)$, is the most used in chemical kinetics studies. The most common formulation for $k(T)$ is given by transition state theory [57, 58]. However, it has been really useful to directly obtaining $k(T)$ from $k(E)$ in order to test the theory when it was not possible to perform energy-selected experiments. $k(T)$ can thus be derived from $k(E)$ by integration over the distribution of internal energy at the temperature T [5]:

$$k(T) = \int_{E_0}^{\infty} k(E)P(E,T)dE \tag{3.46}$$

where:

$$P(E,T) = \frac{\rho(E)e^{-\frac{E}{k_BT}}}{\int_{E_0}^{\infty}\rho(E)e^{-\frac{E}{k_BT}}dE} = \frac{\rho(E)e^{-\frac{E}{k_BT}}}{Q(t)} \tag{3.47}$$

Using equation 3.47, $k(T)$ becomes:

$$k(T) = \int_{E_0}^{\infty} \frac{N^{\dagger}(E-E_0)\rho(E)}{h\rho(E)Q(T)}e^{-\frac{E}{k_BT}}dE = \frac{e^{-\frac{E_0}{k_BT}}}{hQ(T)}\int_{E_0}^{\infty} N^{\dagger}(E)e^{-\frac{E}{k_BT}}dE \tag{3.48}$$

The last integral is the Laplace transform of the sum of states ($L[N^{\dagger}]$) and can be converted by the integration theorem of Laplace transforms to:

$$L[N^{\dagger}] = L\left[\int_{E}^{E_0}\rho^{\dagger}(E)dE\right] = \frac{L\left[\rho^{\dagger}(E)\right]}{\beta} = Q^{\dagger}(T)k_BT \tag{3.49}$$

Thus we obtain:

$$k(T) = \frac{k_BTQ^{\dagger}(T)}{hQ(T)}e^{-\frac{E_0}{k_BT}} \tag{3.50}$$

which is the transition state theory rate constant obtained in the high-pressure limit. In this way it is demonstrated that the RRKM theory and canonical TST rate constants are identical when the first is averaged over the Boltzmann energy distribution, i.e. we have demonstrated that RRKM theory is exactly the microcanonical version of the TST theory [5].

3.2.7 Evaluation of the density and sum of states

Evaluation of the microcanonical rate constant $k(E)$ involves the calculation of density and sum of states. The direct count of states is the most accurate procedure for it. Unfortunately it has been feasible for years only for small molecules. For bigger systems (amino acids size) the Whitten-Rabinovitch (WR) [68] semi-classical method has been the standard choice for many years until the development of a more efficient direct count algorithm (BS) by Beyer and Swinehart [69].

Howsoever, none of these methods are really computationally convenient for big systems (the computation time of BS algorithm increases exponentially with the molecular size) and new methods have been developed, as for example the improved WR algorithm for proteins [70]. All the methods that will be further presented use the harmonic approximation in the calculation of the vibrational frequencies. Anharmonic corrections can be added afterward.

Direct count: Beyer-Swinehart (BS) algorithm

The direct count algorithm [69] is the most accurate, simple, but time consuming method to calculate sums and densities of harmonic vibrational states. It can be applied only to small molecules.

Let's consider a system constituted by s harmonic oscillators. Each harmonic oscillator will have equally spaced states at energies $E_i = \left(n + \frac{1}{2}\right) h\omega_i$, where $n = 0, 1, 2...$ and ω_i is the vibrational frequency in cm^{-1}. Using the zero point energy oscillator as reference energy, the same equation can be re-written as:

$$E_i = n_i \omega_i \tag{3.51}$$

For a given total internal energy, E, and a given oscillator energy, E_i, the contribution to the density of states (the number of ways of distributing the energy E between the oscillators) is given by the convolution of the product $\rho(E)\rho(E - E_i)$. To obtain all the density of states

it must be integrated over all the range of E and over all the oscillator energies E_i:

$$\int_0^E \rho(E)\rho(E - E_i)dE \tag{3.52}$$

that is the convolution integral method. The internal vibrational energy is then divided into bins of 1 cm^{-1}. Thus, the algorithm to generate the density of states of s harmonic oscillators is given by:

```
 ρ(i) =[1,0,0,0,...]    !(initialization of vector \rho(i))
Do j=1, s   !(s=number of oscillators)
   Do i=ω(j), M !(M=maximum energy)
      ρ(i)=ρ(i)+(ρ(i)-ω(j))
   End do
End do
```

Note that in this notation energy and ω have the same units, so that, e.g., $\rho(\omega(1)) = \rho(E(\omega(1)) = n\omega(i)$ with $n = 1$.

When $j = 1$ the first oscillator is filled: $\rho(i) = 1$ when $\rho(i - \omega(j))$ is equal to zero (first loop on i) or equal to a multiple of $\omega(j)$. In the end of the first j loop the ρ vector will have some 0 and some 1. In the second j loop the $\rho(i)$ value will be increased by 1 when $\rho(i-\omega(j))$ is equal to 0 or if $\rho(i - w(j))$ is equal to a multiple of $\omega(i)$ or $\omega(i - 1)$. In the end the values of $\rho(i)$ will correspond to the degeneracy of the energy level.

The sum of states, $N(E)$, can be then obtained by direct numerical integration of the density of states. Otherwise (if the frequencies of the transition state are known or if they can be approximated) it can be also calculated using the same BS scheme:

```
N(i)=[1,1,1...1]    !(initialization of the vector N(i))
Do j=1, s
   Do i=ω(j),M
      N(i)=N(i)+N(i-ω(j))
   End do
End do
```

Whitten-Rabinovitch algorithm

The Whitten-Rabinovitch (WR) algorithm is based on the classical sum of states for s harmonic oscillators. The energy for an harmonic oscillator is:

$$E = \frac{p^2}{2\mu} + \frac{1}{2}kq^2 \tag{3.53}$$

i.e. the kinetic energy plus the harmonic potential energy. μ is the reduced mass and k is the force constant of the oscillator. The phase space volume PSV to describe such a system is the integral over the positions and the momenta:

$$PSV = \int\int dqdp \tag{3.54}$$

where q and p are subject to the conditions of equation 3.53, that can be written also as:

$$1 = \frac{p^2}{2\mu E} + \frac{q^2}{2E/k} \tag{3.55}$$

that is the equation for an ellipse of semi a-axis $= \sqrt{2\mu E}$ and semi b-axis$= \sqrt{\frac{2E}{k}}$. The area of an ellipse is πab, so the phase space volume for a classical harmonic oscillator is:

$$PSV = 2\pi E\sqrt{\frac{\mu}{k}} = \frac{E}{\omega} \tag{3.56}$$

The energy for a system consisting of s harmonic oscillators is $\sum_i E_i$ and the PSV can be written as:

$$PSV = \frac{E^s}{s!\prod \omega_i} \tag{3.57}$$

The sum of states is then given by:

$$N(E) = \frac{E^s}{s!\prod h\omega_i} \tag{3.58}$$

while the density of states, that is the derivative of $N(E)$ in $d(E)$, is:

$$\rho(E) = \frac{E^{s-1}}{(s-1)!\prod h\omega_i} \tag{3.59}$$

In their method, Whitten and Rabinovitch use the modified form by Rabinovitch and Diesen for the sum of states:

$$N(E) = \frac{(E + aE_z)^s}{s!\prod h\omega_i} \tag{3.60}$$

where E is the internal vibrational frequency and E_z is the zero point energy. Next step for the WR method consists in obtaining a from:

$$\omega = \frac{1-a}{\beta} \tag{3.61}$$

with

$$\beta = \frac{s-1<\omega^2>}{s<\omega>^2} \tag{3.62}$$

and with ω equal to:

$$\omega = (5.00\varepsilon + 2.73\varepsilon^{0.5} + 3.51)^{-1} \ , \ 0.1 < \varepsilon < 1.0 \tag{3.63}$$

$$\omega = exp(-2.4191\varepsilon^{0.25}) \ , \ 1.0 \leq \varepsilon \tag{3.64}$$

where $\varepsilon = E/E_z$.

The density of state will be then:

$$\rho(E) = \frac{(E+aE_z)^{s-1}}{(s-1)!\prod h\omega_i}\left[1-\beta\left(\frac{d\omega}{d\varepsilon}\right)\right] \tag{3.65}$$

with

$$\frac{d\omega}{d\varepsilon} = -\frac{1(5.0+1.365\varepsilon)}{(5.0\varepsilon + 2.73\varepsilon^{0.5} + 3.51)^2} \tag{3.66}$$

WR-modified algorithm

The WR method does not fit well with results given using the BS algorithm in low internal energy range. This represents a problem for studying proteins that in the field of mass spectrometry are known to dissociate also at values of energy close to the ZPE (corresponding to $\varepsilon = E/E_0 < 0.1$). Modified WR algorithms were formulated by Laskin [71] and Sun [70] for peptides and proteins. The results of these two methods, the WR and BS algorithms are compared and shown in figures C.1 and C.2 in *Appendix C* for the example tripeptide system $TIK(H^+)_2$.

The version of the WR-method given by Julia Laskin consists in a simple modification of the ω equation at low values of ε:

$$\omega = (1.783\varepsilon^{1.4135} + 6.192\varepsilon^{0.6209} + 3.265)^{-1} \ , \ \varepsilon < 1.0 \tag{3.67}$$

The Sun version [70] of the WR method uses again a different analytical form for ω. Moreover, they parametrized a different ω_i for each of the 20 amino acids. The rate constant

of an n-amino acid protein is then calculated averaging ω using the different ω_i. The analytical form of ω used for the parametrization is:

$$\omega = (c_3\varepsilon^{1.5} + c_2\varepsilon + c_1\varepsilon^{0.5} + c_0)^{-1} \ , \ 0.0005 < \varepsilon < 2.0 \tag{3.68}$$

3.3 Chemical Dynamics

Molecular dynamics (MD) propagates the classical Newton equations of motion for the nuclei. Using the nuclear position vector $\mathbf{r}$ instead of the Cartesian coordinates $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, we can summarize those equations with:

$$\mathbf{F} = m\ddot{\mathbf{r}} \tag{3.69}$$

where $\mathbf{F}$ is the force, $\ddot{\mathbf{r}}$ the second time derivative of the nuclear coordinates, and m the mass.

The kinetic energy T for a molecular system composed by N atoms is defined by:

$$T = \frac{1}{2}\sum_{i=1}^{N} m_i|\dot{\mathbf{r}}_i|^2 \tag{3.70}$$

For a conservative system then it is possible to define the potential energy V, such that the force acting to each particle can be written as:

$$\mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i} \tag{3.71}$$

Generally, the kinetic energy is function of the momenta and the potential energy of the coordinates only. The total energy of the system, is then defined as the sum of the kinetic and potential energy, the so-called Hamiltonian function:

$$H = T + V \tag{3.72}$$

The Newton equations of motion can be written in the Hamiltonian (or also Lagrangian) form, that is easier to be integrated. Integration of the Newton equations in fact permits, knowing initial coordinates and velocities, to obtain coordinates and velocities at any time t.

Thus, using a numerical method, the time is discretized and given nuclei positions, velocities and the forces (from the electronic potential energy) at time t, positions and velocities will be returned at the time $t + \Delta t$. This is shown in figure 3.2.

As it will be shown in section 3.3.1, since the nuclei are much slower than the electrons,

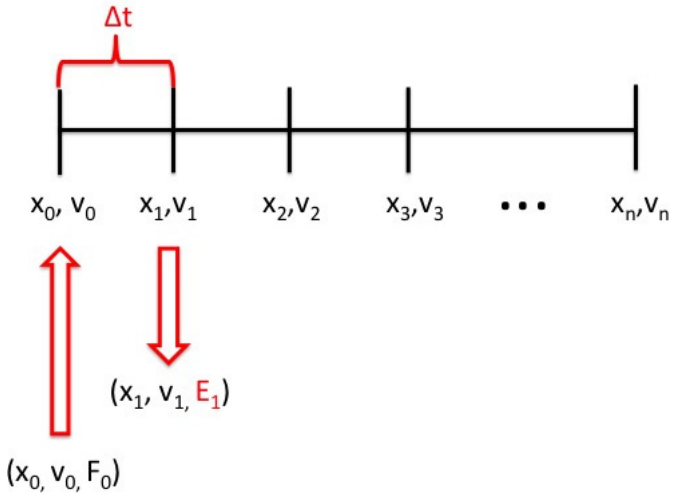


Figure 3.2: The time is discretized in time steps Δt. Coordinates and velocities are taken from initial conditions and the forces are calculated at the time t_0. Using a numerical algorithm is possible to consequently obtain coordinates, velocities and the energy at the time $t + \Delta t$.

it is possible to approximate that their motion is independent, i.e. nuclei and electrons velocities can be decoupled. The nuclei are considered "fixed" and the time-independent electronic Schrödinger equation is resolved for the nuclear positions. The electric field generated by the electrons will induce a force on the nuclei, which will advance by a small Δt along the potential energy surface of the electrons.

Traditionally, the electronic potential energy surface is represented by an analytical function, that can be either a standard model [72] or fitted using electronic structure theories. Another possibility is to perform direct dynamics simulations, i.e. to calculate the electronic energy and its gradient *on-the-fly* using an electronic structure code.

Molecular dynamics studies equilibrium properties and for processes at equilibrium it can be assumed that a long enough trajectory can describe the proper ensemble of states to describe the phenomenon. This assumption is well known as *ergodic hypothesis*.

However, we are focused on studying reactions and chemical dynamics must be performed on an ensemble of trajectories. The difference, in fact, between molecular and chemical dynamics is that the second simulates quasi-equilibrium and/or non-reversible processes (like unimolecular fragmentation in the gas phase) for which it is not possible to explore all the phase-space of interest with only one trajectory.

Thus, the key ingredients to perform chemical dynamics are: (i) the application of the Born-Oppenheimer approximation; (ii) the choice of a numerical integration method for the classical motion equations; (iii) the definition of the initial conditions for each trajectory; (iiii) the choice of how to calculate energies and forces for each point of the trajectory.

3.3.1 Born-Oppenheimer approximation

The aim of the Born-Oppenheimer approximation is to separate the motion of the electrons from the one of the nuclei. This will allow to derive classical molecular dynamics.

In order to derive those equations for a chemical, thus quantum, system it is necessary to start from the time-dependent Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Phi\left(\{\mathbf{r}_i\},\{\mathbf{R}_I\};t\right)=\mathbf{H}\Psi_{tot}\left(\{\mathbf{r}_i\},\{\mathbf{R}_I\};t\right) \tag{3.73}$$

where the vectors $\mathbf{r}$ and $\mathbf{R}$ identify the electrons and nuclei coordinates.

The total Hamiltonian for a chemical system can be written as it follows:

$$\mathbf{H}_{tot}=\mathbf{T}_n+\mathbf{T}_e+\mathbf{V}_{nn}+\mathbf{V}_{ee}+\mathbf{V}_{ne} \tag{3.74}$$

where $\mathbf{T}_n$ and $\mathbf{T}_e$ are the kinetic energy operators for the nuclei and electrons respectively; $\mathbf{V}_{nn}$, $\mathbf{V}_{ee}$ and $\mathbf{V}_{ne}$ are the nucleus-nucleus, nucleus-electron, and electron-electron potential energy operators, respectively. Expressing all these terms in atomic units, equation 3.74 becomes:

$$\begin{aligned}\mathbf{H}_{tot}=&-\frac{1}{2}\sum_{I}^{M}\frac{1}{M_I}\nabla_I^2(\mathbf{R}_I)-\frac{1}{2}\sum_{i}^{N}\nabla_i^2(\mathbf{r}_i)+\\&\sum_{I<J}^{M}\frac{Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|}+\sum_{i<j}^{N}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}-\sum_{I<J}^{M}\sum_{i<j}^{N}\frac{Z_I}{|\mathbf{r}_i-\mathbf{R}_I|}\end{aligned} \tag{3.75}$$

where N is the number of electrons, M the number of nuclei and Z the nuclear charge.

The total Hamiltonian, 3.74, can be further separated in two parts:

$$\mathbf{H}_{tot}=\mathbf{H}_N+\mathbf{H}_{el} \tag{3.76}$$

i.e. in the nuclear and electronic Hamiltonian, where

$$\mathbf{H}_N=\mathbf{T}_n+\mathbf{V}_{nn} \tag{3.77}$$

and

$$\mathbf{H}_{el} = \mathbf{T}_{el} + \mathbf{V}_{ee} + \mathbf{V}_{ne} \tag{3.78}$$

Since the nuclei are much slower than the electrons, it is possible to approximate that the kinetic energy of the nuclei is zero, so that the electrons move in the field of fixed nuclei. The Hamiltonian for such a system is:

$$\mathbf{H}_{el}^{FN} = \mathbf{H}_{el} + \mathbf{V}_{nn} \tag{3.79}$$

The time-independent Schrödinger equation for $\mathbf{H}_{el}$ is:

$$\mathbf{H}_{el}^{FN}\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right)\varphi_k\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right) = E_{el,k}\left(\{\mathbf{R}\}\right)\varphi_k(\{\mathbf{r}_i\};\{\mathbf{R}_I\}) \tag{3.80}$$

having as solution the stationary electronic wave-function φ_k with eigenvalue $E_{el,k}$.

There is an adiabatic electronic wave-function φ_k for each fixed nuclear position $\mathbf{R}_I$, such that:

$$\int \varphi_k^*\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right)\varphi_l\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right)d\mathbf{r} = \delta_{k,l} \tag{3.81}$$

where $\delta_{k,l} = 1$ if $k = l$ and $\delta_{k,l} = 0$ if $k \neq l$, i.e. the eigenfunctions φ_k are orthonormal.

It is thus possible to write the total wave-function of the time-independent Schrödinger equation as an expansion over these electrostatic wave-function and the time-dependent nuclear ones:

$$\Phi\left(\{\mathbf{r}_i\},\{\mathbf{R}_I\};t\right) = \sum_{l=0}^{\infty}\varphi_l\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right)\chi_l\left(\{\mathbf{R}_I\};t\right) \tag{3.82}$$

that couples the adiabatic electronic states with the nuclear wave functions.

Inserting the wave-function of equation 3.82 in equation 3.73 and writing the total Hamiltonian as a sum of the stationary Hamiltonian H_{el}^{FN} and the kinetic energy T_n, we have:

$$\left(\mathbf{H}_{el}^{FN} + \mathbf{T}_n\right)\left(\sum_{l=0}^{\infty}\varphi_l\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right)\chi_l\left(\{\mathbf{R}_I\};t\right)\right) = i\hbar\frac{\partial}{\partial t}\left(\sum_{l=0}^{\infty}\varphi_l\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right)\chi_l\left(\{\mathbf{R}_I\};t\right)\right) \tag{3.83}$$

When the operator $\mathbf{T}_n\,(=\nabla_n^2)$ operates on the product $(\chi_k\varphi_k)$ the result is:

$$\begin{aligned}\nabla_I^2\left(\chi_k\varphi_k\right) &= \nabla_I\nabla_I\left(\chi_k\varphi_k\right)\\ &= \varphi_k\nabla_I^2\chi_k + \chi_k\nabla_I^2\varphi_k + 2\nabla_I\varphi_k\nabla_I\chi_k\end{aligned} \tag{3.84}$$

Inserting now equation 3.82 in equation 3.73, multiplying $\varphi_k^*(\{\mathbf{r}_i\};\{\mathbf{R}_I\})$ by the left, and integrating over all the electronic coordinates $\mathbf{r}$, we obtain:

$$\left(-\sum_I\frac{1}{2M_I}\nabla_I^2 + E_{el,k}(\{\mathbf{R}_I\})\right)\chi_k + \sum_l C_{k,l}\chi_l = i\hbar\frac{\partial}{\partial t}\chi_k \tag{3.85}$$

where $C_{k,l}$ is the non-adiabatic coupling operator, which is a non-diagonal term containing both φ_k and φ_l. Neglecting terms (adiabatic approximations), it is possible to write the total wave-function of equation 3.82 as:

$$\Phi\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right) \approx \varphi_l\left(\{\mathbf{r}_i\};\{\mathbf{R}_I\}\right)\chi_l(\{\mathbf{R}_I\};t) \tag{3.86}$$

i.e. into a product of the electronic and nuclear wave-function. Now, neglecting all $C_{k,l}$, we have:

$$\left(-\sum_I \frac{1}{2M_I}\nabla_I^2 + E_{el,k}\left(\{\mathbf{R}_I\}\right)\right)\chi_k = i\hbar\frac{\partial}{\partial t}\chi_k \tag{3.87}$$

that represents the Born-Oppenheimer approximation.

This equation is the starting point to derive classical dynamics, i.e. to transform the quantum motion of the nuclei to a classical one. To do so, we have first to write the nuclei wave-function as:

$$\chi_k\left(\{\mathbf{R}_I\};t\right) = A_k\left(\{\mathbf{R}_I\};t\right)\exp\frac{iS_k\left(\{\mathbf{R}_I\};t\right)}{\hbar} \tag{3.88}$$

Replacing equation 3.88 in equation 3.87 and taking $\hbar \longrightarrow 0$, i.e. in the classical limit, it is possible to derive [73]:

$$M_I\ddot{\mathbf{R}}_I(t) = -\nabla_I V_k^{BO}(\{\mathbf{R}_I\}(t)) \tag{3.89}$$

for each electronic state k and:

$$V_k = min_{\varphi_0}\{\langle\varphi_0|H_{el}^{FN}|\varphi_0\rangle\} \tag{3.90}$$

Thus, with the Born-Oppenheimer approximation, we can use the classical Newton equations for the nuclei, where the force is obtained by the gradient (changed of sign) of the electronic potential energy surface. This one is simultaneously calculated resolving the electronic time-independent Schrödinger equation.

3.3.2 Integration of the motion equations

We have just derived that it is possible, using the Born-Oppenheimer approximation, to use the Newton classical equations of motion for the nuclei.

Integration of the Newton equations allows to know positions and momenta of a system along the time. However, the motion of a chemical system is complex and it is not possible to resolve these equations analytically and it is necessary to use numerical methods. First, the

simulation time is discretized in small integration steps. The smaller the integration time, the higher will be the accuracy (a good algorithm should allow to use longer time steps Δt and at the same time conserve the total energy).

One of the most used numerical method to integrate the classical equations of motion in Molecular Dynamics is the Velocity Verlet [74], (a variation of the previous Verlet algorithm [75]) that was used for many of the simulations reported here. The Velocity Verlet equations for the motion are thus:

$$\mathbf{r}\left(t+\Delta t\right)=\mathbf{r}(t)+\mathbf{v}\Delta t+\frac{1}{2}\frac{\mathbf{F}(t)}{m}\Delta t^{2} \tag{3.91}$$

$$\mathbf{v}\left(t+\Delta t\right)=\mathbf{v}(t)+\frac{1}{2}\Delta t\left[\frac{\mathbf{F}(t)}{m}+\frac{\mathbf{F}(t+\Delta t)}{m}\right] \tag{3.92}$$

where $\mathbf{r}$, $\mathbf{v}$ are the position and velocity vectors, while $\mathbf{F}$ is the force.

Summarizing, in order to implement the Velocity Verlet algorithm, we have to:

1. Give initial conditions for positions and velocities.
2. Use $\mathbf{r}(t)$ to calculate the force $\mathbf{F}(t)$
3. Use $\mathbf{r}(t)$, $\mathbf{v}(t)$ and $\mathbf{F}(t)$ to calculate $\mathbf{r}\left(t+\Delta t\right)$.
4. Use $\mathbf{r}\left(t+\Delta t\right)$ to calculate $\mathbf{F}\left(t+\Delta t\right)$.
5. Use $\mathbf{v}(t)$, $\mathbf{F}(t)$ and $\mathbf{F}\left(t+\Delta t\right)$ to calculate $\mathbf{v}\left(t+\Delta t\right)$
6. Re-iterate from step 2.

This algorithm is numerically stable, simple to implement and shows excellent energy conservation. Another integration method that was used is a sixth order symplectic algorithm [76].

3.3.3 Initial conditions

To propagate the trajectories one has first to give the initial conditions, i.e. nuclei positions and momenta at $t=0$. This set is selected in order to obtain a statistical ensemble of trajectories. For example, a quantum or classical microcanonical [77] or canonical sampling for

positions and momenta may be chosen for the ion.

In order to simulate CID two chemical dynamics activation methods were employed: explicit collision and internal energy activation, which differ in the initial conditions. We will now describe the two separately.

Explicit collision activation

In section 2.1.2 of chapter 2 it was introduced the idea that an ion can be activated through a "linear" acceleration toward the collision cell containing an inert gas. In these kinds of instruments, as for example in a triple quadrupole, the energy transmitted with only one collision can be "high" enough to allow fragmentation. In this case, the dynamical effect of the collision is important, for example the dissociation can take place close to the collision site. The *high* energy collision can in fact induce a fast dissociation, i.e. the ion can dissociate before an internal energy relaxation (IVR), see *panel A* of figure 3.3. These are all simplifications made in order to model the experiment, that it is indeed idealized through a *single* and *high* collision between the ion and an inert gas.

In order to make the model "real" two aspects must be considered: (i) the configurational space must sampled for both the ion and the gas and (ii) random conditions for the collision must be chosen, see figure 3.4.

1. First, an initial equilibrium geometry at a minimum electronic potential energy surface must be given for both the ion and the collision gas.

2. For the ion a classical canonical sampling for the normal modes can be chosen at the temperature of 300 K. At this temperature the classical and quantum energy distributions are similar and the ion vibrational energy will be higher than its zero point energy (ZPE), that is a quantum property. Since initial conditions are chosen in order to reproduce the ZPE, it is often referred to this sampling as quasi-classical initial conditions [76]. However, classical Newton equations are used to describe the nuclei motion and thus there will not be any guarantee that the ZPE will be preserved. The harmonic vibrational Hamiltonian of a chemical system consisting of s normal modes is given by:

$$H = \sum_{i=1}^{s} \frac{P_i^2 + \lambda_i Q_i^2}{2} \tag{3.93}$$

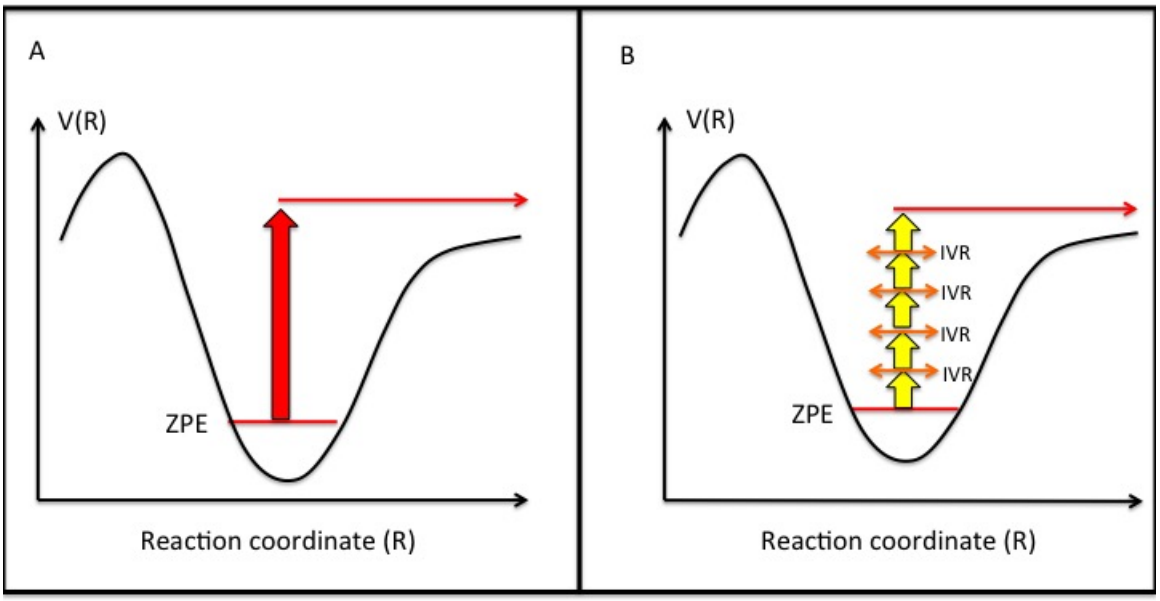


Figure 3.3: Panel A: fast excitation process. The molecular system dissociates before any internal energy redistribution; Panel B: slow excitation process: the internal excess energy is statistically distributed before the dissociation.

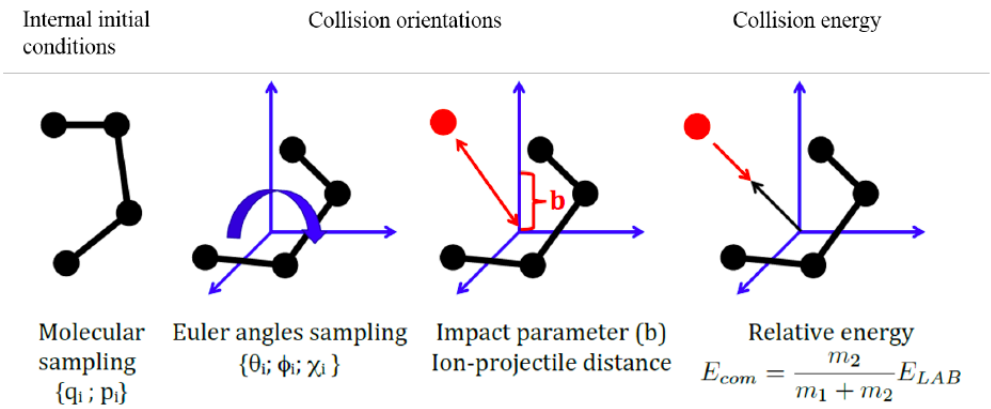


Figure 3.4: Generic set-up for collisional simulations.

where Q_i are the normal mode coordinates that are equal to:

$$Q_i = \sum_l a_{il} q_l \tag{3.94}$$

and where q_l are the mass-weighted displacement (from the equilibrium geometry) Cartesian coordinates and a_{il} are the matrix elements that define the transformation between Q_i and q_l.
Thus, the energy is calculated for each Q_i, so that we can calculate the probability of each normal mode quantum number n_i to be populated at the temperature T_{VIB}:

$$P(n_i) = exp\left(-\frac{n_i h\nu_i}{k_B T_{VIB}}\right)\left[1 - \exp\frac{-h\nu_i}{k_B T_{VIB}}\right] \tag{3.95}$$

where h is the Plank constant, ν_i is the vibrational frequency of the normal mode i, k_B is the Boltzmann constant and T_{VIB} is the vibrational temperature, for example 300 K. For each trajectory a single normal mode quantum number, and thus energy, is chosen in accord to this probability. The same normal mode energy can be given by different combinations of the normal mode coordinates, Q_i, and its time derivative, $\dot{Q}_i$:

$$Q_i = A_i \cos(2\pi R_i) \tag{3.96}$$

$$\dot{Q}_i = -\tilde{\omega}_i A_i \sin(2\pi R_i) \tag{3.97}$$

where A_i is the normal mode i amplitude and $\tilde{\omega}_i = 2\pi\nu_i$, while R_i is a random number by which Q_i and $\dot{Q}_i$ are chosen. When this procedure is done for all the normal modes these coordinates can be linearly transformed in atomic positions and momenta. In particular, the transformation is made in order to displace the atomic positions around the given potential energy equilibrium geometry.

3. An angular momentum will be then applied to the ion in accord to the rotational temperature chosen. This energy can be given in accord to a Boltzmann distribution at 300 K and assuming a *symmetric-top* geometry for the ion, or giving $\frac{k_B T_{ROT}}{2}$ to each axis of inertia.

4. The gas used to enhance the ion dissociation can be an atom, like Ar, or diatomic, like N_2. For the Ar it is simple because it has no vibrational or rotational degrees of freedom. For N_2, having only one vibrational normal mode, its quantum number can be sampled with a Boltzmann distribution (it will be almost always $n = 0$ and sometimes $n = 1$). Rotational energy can be given selecting the rotational quantum number or giving $\frac{k_B T_{ROT}}{2}$ to the two axes of inertia.

Once the reactants initial conditions are given, it is necessary to set randomly also the "collision process". This is done as follows.

1. The center of mass translational collisional energy and the ion-gas initial distance are set and they are a constant for all the trajectories. In particular the collision energy, given in the center-of-mass framework, can be related to the collision energy in the laboratory framework used in triple quadrupole mass spectrometer (see section 2.1 in chapter 2):

$$E_{CM} = \frac{m_{gas}}{m_{ion} + m_{gas}} E_{LAB} \tag{3.98}$$

   where E_{CM} and E_{LAB} are the energies in the center-of-mass and in the laboratory framework, respectively, and m_{ion} and m_{gas} are the masses of the ion and the gas, respectively.

2. The ion is rotated about its Euler angles (using random numbers) in order to obtain random directions of the collisions.

3. The impact parameter b is the gas collision aiming distance (calculated from the ion center of mass), see figure 3.5. It can be a fixed number or randomly selected imposing the maximum value b_{max}:

$$b = b_{max}\sqrt{R} \tag{3.99}$$

R is a random number between 0 and 1 and the square root of such a number provides a higher probability of higher b values.

The initial distance between the ion and the gas is set so that the initial potential energy

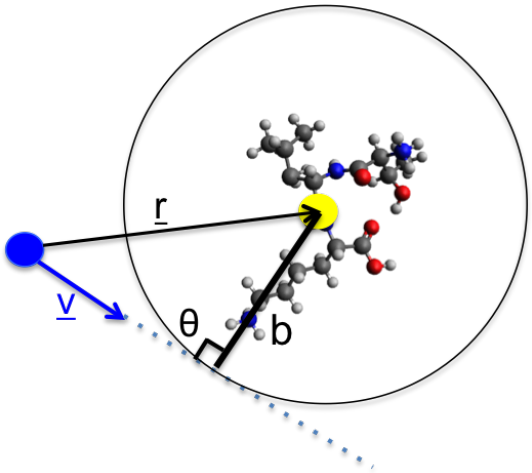


Figure 3.5: Impact parameter for a collision between the gas and a generic ion. The impact parameter b is the distance between the center of mass of the ion and the velocity **v** of the gas.

repulsion is *zero*. The maximum impact parameter can be chosen with a geometric or energetic criterion, generally leading to a similar value of b_{max}. A more rigorous way is to analyse, via a preliminary set of trajectories, the energy transfer as a function of b. b_{max} is then chosen as the value of b at which the transferred energy is less than 5% of the collision energy.

Internal energy activation

As already mentioned in section 2.1.2 of chapter 2, the collision activation in a generic ion-trap is different from the activation mode of a quadrupole filter. This regards two aspects of the two instruments: in the quadrupole the ions are only accelerated through the quadrupole plates in order to reach the detector, while in an ion-trap the ions are collected in a loop-motion and only after they will reach the detector. The second aspect to be considered regards the amount of energy transfer: in the quadrupole the energy of one collision can be enough to allow fragmentation, while multiple collisions are generally needed in an ion-trap.

Thus, one collision in a trap cannot allow immediate fragmentation, but the energy is stored in the ion, and ideally, statistically distributed between the normal modes. The internal energy of the ions can grow up slowly and each intermediate of a reaction coordinate is thermalized by this "external bath" (see *panel B* of figure 3.3). The evolution of a thermalized system does not depend on the activation way. Such a system can be thus idealized as a system having energy statistically distributed between its degrees of freedom.

Indeed, in internal energy activation simulations the ion is excited by increasing its internal energy and the collision is not simulated. This can be done using a classical microcanonical sampling scheme [77] for the vibrational energy. The harmonic oscillators energies must be chosen randomly and in order to have a total vibrational energy E_{VIB}. Normal modes coordinates and consequently atomic positions and momenta are obtained as discussed previously.

If the ion is large enough $s - 1$ can be approximated to s, the number of normal modes. If also the reactant energy E is large, so that $E_0/E << 1$, the classical microcanonical and canonical rate constant, $k(E)$ and $k(T)$ are equivalent. This will be better shown in section 4. Microcanonical simulations can then be used to obtain $k(T)$ choosing $E_{VIB} = sk_BT$.

3.4 Quantum Chemistry methods

Given the Born-Oppenheimer approximation (see section 3.3.1), it is possible to separate the nuclear and electronic motions. In section 3.3.2 the algorithms used to apply the Newton equation on the nuclei were described. In order to apply those algorithms on the nuclei it is necessary to define the initial conditions (section 3.3.3) and to provide the forces responsible for the nuclear motion itself. The force is calculated by the gradient of the nuclear potential energy and the electronic energy. To perform our direct Chemical Dynamics simulations they

are generally calculated using semi-empirical Hamiltonians. They are based on the Hartree Fock (HF) method, but provide more efficiency by the introduction in the HF equations of some approximations and empirical parameters. The Hartree Fock method and the successive semi-empirical Hamiltonians [23, 24, 78, 79] will be presented in this section.

More accurate quantum chemistry methods are used to calculate the electronic energies and frequencies like density functional theory (DFT) or post-HF methods.

3.4.1 Hartree Fock

Hartree Fock is one of the oldest quantum chemistry method used to find an approximated electronic wave-function representing a system stationary electronic state and the associated energy.

Often it is referred to as the self-consisted field (SCF) theory because the two goals are achieved through an iterative search of the best wave-function coefficients and the calculation of the corresponding energy.

The method is based on the Born-Oppenheimer and on other two main approximations, the mean-field and the pseudo-single particle. The mean-field approximation regards the interaction term between a mono-electronic function and all the others, which can be approximated by a mean electronic field. This is the approximation that allows to write E_{el}, the total electronic energy, as a sum of mono-electronic terms and to resolve the electronic problem. The pseudo single-particle approximation is a consequence of the previous approximation. If the electronic energy is written as a sum, the wave-function can be written as a product of mono-electronic functions. However, an electronic function must be anti-symmetric for the exchange of two electrons coordinates (because electrons are indistinguishable).

The first iterative process step consists in building an initial anti-symmetric electronic wave-function guess, that can be obtained building a Slater determinant: a set of orthonormal mono-electronic wave-functions is chosen, which indices are along the columns and the electronic coordinates along the rows.

$$\Phi = \frac{1}{\sqrt{N}} \begin{bmatrix} \phi_1(1) & \phi_2(1) & \phi_3(1) & \dots & \phi_N(1) \\ \phi_1(2) & \phi_2(2) & \phi_3(2) & \dots & \phi_N(2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_1(N) & \phi_2(N) & \phi_3(N) & \dots & \phi_N(N) \end{bmatrix}$$

Each mono-electronic wave-function is a product of an orbital spatial function and a spin function, α or β. The ϕ_i are now be called spin-orbitals. In particular, if the objective is to calculate the energy of a molecular system, each spin-orbital is called molecular-orbital.
The variational principle states that any approximated wave-function has an energy higher or equal to the exact one. It is then possible to optimize Φ by minimizing the energy.

To simplify the Hartree Fock equation that will follow, it is better to write Φ differently, i.e. as the product of the anti-symmetrization operator $\mathbf{A}$ and the determinant diagonal product Π:

$$\Phi = \mathbf{A}\Pi \tag{3.101}$$

with:

$$\Pi = [\phi_1(r_1)\phi_2(r_2)\phi_3(r_3)\dots\phi_N(r_N)]\,, \tag{3.102}$$

$$\mathbf{A} = \frac{1}{\sqrt{N}}\sum_{p=0}^{N-1}(-\mathbf{1})^p\mathbf{P} = \frac{1}{\sqrt{N}}\left[\mathbf{1} - \sum_{i,j}\mathbf{P}_{i,j} + \sum_{i,j,k}\mathbf{P}_{i,j,k} - \dots\right] \tag{3.103}$$

where $\mathbf{P}$ is the *permutation* operator and $\mathbf{1}$ the identity operator. $\mathbf{P}_{i,j}$ is the permutation over two electronic coordinates, so $\mathbf{P}_{i,j,k}$ is the permutation over three and so on.

The electronic Hamiltonian for a molecular system can be written as:

$$\mathbf{H}_{el} = \sum_{i}^{N}\mathbf{h}_i + \sum_{j>i}^{N}\mathbf{g}_{i,j} + \mathbf{V}_{nn} \tag{3.104}$$

where $\mathbf{h}_i$ is the one-electron Hamiltonian, i.e. the sum of one-electron kinetic energy and electron-nuclei attractive interaction:

$$\mathbf{h}_i = -\frac{1}{2}\nabla_i^2 - \sum_{m}^{M}\frac{Z_m}{\mid \mathbf{R}_m - \mathbf{r}_i \mid} \tag{3.105}$$

where m is the index over the nuclei, Z_m the m-atomic number, M the number of nuclei, $\mathbf{R}_m$ the m-nucleus coordinate and $\mathbf{r}_i$ the electron one.

$\mathbf{g}_{i,j}$ is then the two-electron operator, i.e. that gives electron-electron repulsion:

$$\mathbf{g}_{i,j} = \frac{1}{\mid \mathbf{r}_i - \mathbf{r}_j \mid} \tag{3.106}$$

while V_{nn} is the nuclear-nuclear repulsion operator, independent of the electronic coordinates, and indeed leads to a constant along all the iterative process.

Φ is an approximated and normalized wave-function, so that the energy can be calculated as:

$$E_{el} = \langle\Phi|\mathbf{H}_{el}|\Phi\rangle \tag{3.107}$$

Replacing Φ with $\mathbf{A}\Pi$ and considering the two following properties of $\mathbf{A}$:

$$\mathbf{AH} = \mathbf{HA} \tag{3.108}$$

$$\mathbf{AA} = \sqrt{N!}\mathbf{A} \tag{3.109}$$

the equation 3.107 can be rewritten as:

$$E_{el} = \sum_P (-1)^P \langle \Pi | \mathbf{H} | \mathbf{P}\Pi \rangle \tag{3.110}$$

Looking at equation 3.104, it is possible to consider separately the electron's coordinates dependent terms $\mathbf{h}_i$ and $\mathbf{g}_{i,j}$. The one-electron operator, $\mathbf{h}_i$ gives a non-zero contribution to the energy only when $\mathbf{P}$ is equal to the identity operator $\mathbf{1}$. Indicating with E_1 the energy given by the one-electron operator acting on the electron coordinate 1, h_1:

$$\begin{aligned} E_1 = \langle \Pi | \mathbf{h}_1 | \Pi \rangle &= \langle \phi_1(1) | \mathbf{h}_1 | \phi_1(1) \rangle \langle \phi_2(2) | \phi_2(2) \rangle \langle \phi_3(3) | \phi_3(3) \rangle \dots \langle \phi_N(N) | \phi_N(N) \rangle \\ &= \langle \phi_1(1) | \mathbf{h}_1 | \phi_1(1) \rangle \end{aligned} \tag{3.111}$$

Similarly to what was done to calculate equation 3.111, it is possible to see that the two-electron operator $\mathbf{g}_{i,j}$, instead, gives a non-zero contribution to the total electronic energy when the operator $\mathbf{P}$ is equal to the identity or to the permutation of two-electron operators. Considering the two-electron operators $\mathbf{g}_{1,2}$ acting on the electron coordinates 1 and 2, it follows that:

$$E_{1,2} = \langle \phi_1(1)\phi_2(2) | g_{1,2} | \phi_1(1)\phi_2(2) \rangle + \langle \phi_1(1)\phi_2(2) | g_{1,2} | \phi_2(1)\phi_1(2) \rangle = J_{1,2} + K_{1,2} \tag{3.112}$$

where $J_{1,2}$ is called Coulomb integral and it is a two-charge distribution classical repulsion, while $K_{1,2}$ is called exchange integral and it is a pure quantum quantity.

The total electronic energy is now:

$$E_{el} = \sum_{i=1}^{N} h_i + \sum_{j>i}^{N} (J_{i,j} - K_{i,j}) + V_{nn} \tag{3.113}$$

Note that using this notation, h_i and V_{nn} are integrals and not operators, as well as $J_{i,j}$ and $K_{i,j}$.

The purpose is then to find the stationary wave-function that minimize the energy. However, the search must be constrained in order to maintain the set of mono-electronic wave-functions orthonormal. This is done using the method of the Lagrangian multipliers that

consists in finding the coefficients that make the Lagrangian function L stationary with their little variation:

$$\delta L = \delta E_{el} - \sum_{i,j}^{N} \lambda_{i,j} \delta \left(\langle \phi_i | \phi_j \rangle - \delta_{i,j} \right) = 0 \tag{3.114}$$

Note that in 3.114, as henceforth, a different notation is used to specify the electrons coordinates, so that, for example, $\langle \phi_i(1)\phi_j(2)|\phi_i(2)\phi_j(1)\rangle = \langle \phi_i\phi_j|\phi_j\phi_i\rangle$.

It is possible to derive :

$$\begin{aligned} \delta E_{el} &= \sum_{i}^{N} \left(\langle \delta\phi_i | \phi_i \rangle + \langle \phi_i | \delta\phi_i \rangle \right) + \sum_{j>i}^{N} \left(\langle \delta\phi_i | \mathbf{J}_i - \mathbf{K}_j | \phi_i \rangle + \langle \phi_i | \mathbf{J}_i - \mathbf{K}_j | \delta\phi_i \rangle \right) \\ &= \sum_{i}^{N} \left(\langle \delta\phi_i | \mathbf{F}_i | \phi_i \rangle + \langle \phi_i | \mathbf{F}_i | \delta\phi_i \rangle \right) \end{aligned} \tag{3.115}$$

from which it is possible to define the Fock operator $\mathbf{F}$, as:

$$\mathbf{F}_i = \mathbf{h}_i + \sum_{j}^{N} \left(\mathbf{J}_j - \mathbf{K}_j \right) \tag{3.116}$$

where $\mathbf{J}_j$ and $\mathbf{K}_j$ are the Coulomb and Exchange operators:

$$\mathbf{J}_j = \left\langle \phi_j(j) | \mathbf{g}_{i,j} | \phi_j(j) \right\rangle \tag{3.117}$$

$$\mathbf{K}_j = \left\langle \phi_j(j) | \mathbf{g}_{i,j} | \phi_i(j) \right\rangle \tag{3.118}$$

If we compare the Hamiltonian with the Fock operator we have:

$$\mathbf{H}_{el,i} = \mathbf{F}_i - \frac{1}{2} \sum_{j}^{N} \left(\mathbf{J}_j - \mathbf{K}_j \right) + \mathbf{V}_{nn} \tag{3.119}$$

Replacing equation 3.115 in equation 3.114:

$$\delta L = \sum_{i}^{N} \left(\langle \delta\phi_i | \mathbf{F}_i | \phi_i \rangle + \langle \phi_i | \mathbf{F}_i | \delta\phi_i \rangle \right) - \sum_{j>i}^{N} \lambda_{i,j} \left(\langle \delta\phi_i | \phi_j \rangle + \langle \phi_i | \delta\phi_j \rangle \right) = 0 \tag{3.120}$$

For the following properties:

$$\langle \phi | \delta\phi \rangle = \langle \delta\phi | \phi \rangle^* \tag{3.121}$$

$$\langle \phi | \mathbf{F} | \delta\phi \rangle = \langle \delta\phi | \mathbf{F} | \phi \rangle^* \tag{3.122}$$

and with a rearrangement equation 3.120 becomes:

$$\sum_{i}^{N} \langle \delta\phi_i | \mathbf{F} | \phi_i \rangle - \sum_{j>i}^{N} \lambda_{i,j} \langle \delta\phi_i | \phi_j \rangle = - \sum_{i} \langle \delta\phi_i | \mathbf{F} | \phi_i \rangle^* + \sum_{j>i}^{N} \lambda_{i,j} \langle \delta\phi_i | \phi_j \rangle^* \tag{3.123}$$

We can also write the same equation differently:

$$\sum_{i}^{N} \langle \delta\phi_i | \mathbf{F} | \phi_i \rangle - \sum_{i,j}^{N} \lambda_{i,j} \langle \delta\phi_i | \phi_j \rangle + \sum_{i} \langle \delta\phi_i | \mathbf{F} | \phi_i \rangle^* - \sum_{i,j}^{N} \lambda_{i,j} \langle \delta\phi_i | \phi_j \rangle^* = 0 \tag{3.124}$$

From the 3.123 it is clear that to have the 3.124, so to make δL equal to 0, it is sufficient the following condition:

$$\sum_{i}^{N} \langle \delta\phi_i | \mathbf{F} | \phi_i \rangle - \sum_{j>i}^{N} \lambda_{i,j} \langle \delta\phi_i | \phi_j \rangle = 0 \tag{3.125}$$

From 3.125 it is possible to write N^2 Hartree Fock equations:

$$\mathbf{F}_i \phi_i = \sum_{j}^{N} \lambda_{i,j} \phi_j \tag{3.126}$$

By choosing the matrix of the coefficients of Φ as an unitary transformation that makes the matrix of the Lagrangian multipliers diagonal, where $\lambda_{i,i} = \epsilon_i$, the set of these new wave-functions with energy ϵ_i are called canonical orbitals. The Hartree Fock equations become:

$$\mathbf{F}_i \phi_i^{'} = \epsilon_i \phi_j^{'} \tag{3.127}$$

The canonical orbitals, $\phi_i^{'}$, are the wave-functions that make stationary and minimum the electronic energy, which should be still calculated using equation 3.113. In analogy with equation 3.119, the total electronic energy can be written as:

$$E_{el} = \sum_{i}^{N} \epsilon_i - \frac{1}{2} \sum_{i,j}^{N} (J_{i,j} - K_{i,j}) + V_{nn} \tag{3.128}$$

The Coulomb and Exchange integrals make the energy minimization an iterative process. In fact, they are in turn functions of the wave-function that must be determined. It is particularly useful to see how the iterative process works in the case of Φ being the total electronic molecular wave-function and ϕ_i being the electronic molecular orbitals, indeed wave-functions obtained as a linear combination of a number Mb of atomic wave-functions basis, so-called atomic orbitals (AO). Each atomic-orbital is taken as a pre-computed set of functions. We will now on refer to the ϕ_i as molecular orbitals (MO), that are:

$$\phi_i = \sum_{\alpha}^{Mb} c_{\alpha,i} \chi_\alpha \tag{3.129}$$

where χ_α is an atomic orbital.

Using equation 3.127 and replacing ϕ_i with equation 3.129 the Hartree Fock equations can be written as:

$$\mathbf{F}_i \sum_{\alpha}^{Mb} c_{\alpha,i}\chi_\alpha = \epsilon_i \sum_{\alpha}^{Mb} c_{\alpha,i}\chi_\alpha \tag{3.130}$$

Multiplying by the left equation 3.130 and by further integration we obtain the Roothaan-Hall equations for the molecular orbital ϕ_i that can be expressed in a matrix notation:

$$\mathbf{FC} = \mathbf{SC}\epsilon \tag{3.131}$$

Where $\mathbf{F}$ is the Fock matrix with elements $F_{\alpha,\beta} = \langle\chi_\alpha|\mathbf{F}|\chi_\beta\rangle$, $\mathbf{S}$ is the overlap matrix, i.e. the matrix formed by all the elements $S_{\alpha,\beta} = \langle\chi_\alpha|\chi_\beta\rangle$, $\mathbf{C}$ is the matrix of the coefficients c_α, and ϵ the vector of the eigenvalues ϵ_i.

Given an initial coefficients guess, i.e. an N^2 coefficients matrix, it will be possible to find an unitary transformation that makes the Fock matrix diagonal, with the diagonal elements being the eigenvalues, ϵ_i. This unitary transformation is a new coefficients matrix. If this is used to calculate again the two-electron integrals then the Fock matrix is not any more diagonal and it must be found a new unitary transformation until convergence, indeed until the coefficients matrix used to calculate the two-electron integrals is the one used to diagonalize the Fock matrix.

A useful notation for equation 3.131 can be obtained by writing the two-electron integrals as a product of a density matrix and a tensor containing two-electron integrals:

$$\begin{aligned}
\langle\chi_\alpha|\mathbf{F}|\chi_\beta\rangle &= \langle\chi_\alpha|\mathbf{h}|\chi_\beta\rangle + \sum_{j}^{occ.MO} \left(\langle\chi_\alpha\phi_j|\mathbf{g}|\chi_\beta\phi_j\rangle - \langle\chi_\alpha\phi_j|\mathbf{g}|\phi_j\chi_\beta\rangle\right) \\
&= \langle\chi_\alpha|\mathbf{h}|\chi_\beta\rangle + \sum_{j}^{occ.MO}\sum_{\gamma,\delta}^{Mb} c_{\gamma,j}c_{\delta,j}\left(\langle\chi_\alpha\chi_\gamma|\mathbf{g}|\chi_\beta\chi_\delta\rangle - \langle\chi_\alpha\chi_\gamma|\mathbf{g}|\chi_\delta\chi_\beta\rangle\right) \\
&= \langle\chi_\alpha|\mathbf{h}|\chi_\beta\rangle + \sum_{\gamma,\delta}^{Mb} D_{\gamma,\delta}\left(\langle\chi_\alpha\chi_\gamma|\mathbf{g}|\chi_\beta\chi_\delta\rangle - \langle\chi_\alpha\chi_\gamma|\mathbf{g}|\chi_\delta\chi_\beta\rangle\right)
\end{aligned} \tag{3.132}$$

or:

$$F_{\alpha,\beta} = h_{\alpha,\beta} + \sum_{\gamma,\delta} G_{\alpha,\beta\gamma,\delta}D_{\gamma\delta} \tag{3.133}$$

that can be written in a matrix form as:

$$\mathbf{F} = \mathbf{h} + \mathbf{G}\cdot\mathbf{D} \tag{3.134}$$

An Hartree Fock (HF) calculation scales formally as Mb^4, i.e. as the number of two-electron integrals. In the next-session the semi-empirical Hamiltonians methods will be described that, based on the HF formalism, reduce the computational cost by reducing the number of these two-electron integrals.

3.4.2 Semi-empirical Hamiltonian methods

The number and type of Mb atomic basis, that form the so-called *basis set*, provide the computational cost of any SCF iteration. This is the first point considered by semi-empirical Hamiltonian methods to reduce the computational cost of the SCF. A basis set should give better results as the number of basis functions increases. However, semi-empirical methods reduce their number taking advantage of the empirical parameters that are introduced in the modified HF equations to reduce the calculation cost and use a minimum set of basis, the so-called *minimal basis set*. Furthermore, only basis for the valence electrons take part to the integrals, while the *core* electrons are treated by introducing an analytical function or reducing the nuclear charge:

$$\mathbf{h} = -\frac{1}{2}\nabla^2 - \sum_{a}^{N_{nuclei}} \frac{Z_a^{red}}{|\mathbf{R}_a - \mathbf{r}|} = -\frac{1}{2}\nabla^2 - \sum_{a}^{N_{nuclei}} \mathbf{V}_a \tag{3.135}$$

where Z_a^{red} is the nuclear charge reduced by the number of *core* electrons.

Most of the semi-empirical methods adopt the minimal basis set, using only one s- and three p-orbitals (only one s-orbital for H is used) for atoms of the first two rows of the periodic table.

The central approximation of all the semi-empirical Hamiltonians models is the *Zero Differential Overlap* (ZDO), that neglects all the basis functions products (not their integrals) that locate the same electron on different atoms:

$$\langle \mu_A(1) | \nu_B(1) \rangle = 0 \tag{3.136}$$

where in semi-empirical theory the χ_s functions generally used by *ab initio* methods are replaced by μ, ν, γ and σ. $\mu_A(1)$ indicates the electron with coordinate 1 located on the atomic basis μ of the atom A and $\nu_B(1)$ the electron with coordinate 1 located on the atomic basis ν of the atom B. With this notation a Fock matrix element becomes:

$$F_{\mu,\nu} = h_{\mu,\nu} + \sum_{\lambda,\sigma}^{Mb} D_{\lambda,\sigma} \left(\langle \mu\nu | \lambda\sigma \rangle - \langle \mu\nu | \sigma\lambda \rangle \right) \tag{3.137}$$

where $\langle \mu\nu|\lambda\sigma\rangle$ is an abbreviation for the two-electron integrals.

From Eq. 3.136, it follows that the overlap matrix, $\mathbf{S}$, is an *unitary* matrix; the one-electron three-center integrals and the two-electron three- and four-center integrals are zero, that are given respectively by:

$$S_{\mu,\nu} = \langle \mu|\nu\rangle = \delta_{\mu,\nu}\delta_{A,B} \tag{3.138}$$

$$\langle \mu_A|\mathbf{V}_C|\nu_B\rangle = 0 \tag{3.139}$$

$$\langle \mu_A\nu_B|\lambda_C\sigma_D\rangle = \delta_{AC}\delta_{BD}\,\langle \mu_A\nu_B|\lambda_A\sigma_B\rangle \tag{3.140}$$

This reduces the computational cost but also the accuracy, and in order to compensate for the errors introduced in this way, the remaining integrals are parametrized using empirical data or high level quantum chemistry data. The ZDO is also known as NDDO, that is *Neglect of Diatomic Differential Overlap Approximation*. This is also the name of the family of semi-empirical methods that obey to it. Other two important families of semi-empirical methods are the INDO (*Intermediate Neglect of Differential Overlap Approximation*) and CNDO (*Complete Neglect of Differential Overlap Approximation*) that differ each other in the way the two-electron integrals are treated. In particular, while in the NDDO approximation the one- and two-center two-electron integrals are calculated on-the-fly, in the other two they are made into parameters.

All the cited approximations use Z^{red} in equation 3.135 in order to recover for the *core* electrons neglect. The semi-empirical methods of a same family are then differentiated depending on the way in which the parametrization is done. For example the so-called MINDO models are modified-INDO methods that differ in the data set used in the parametrization. Between the most famous or used methods we find modified-NNDO models [80], namely MNDO, as AM1 [81] (*Austin Model number one*), RM1 [82] (*Recife Model number one*) and different PM-N variations (*Modified Neglect of Diatomic Overlap, Parametric Method number N*). The main advantage of MNDO methods over the MINDO ones is that the parametrization was done to reproduce molecular properties and not atomic ones. However, the first released MNDO model could not reproduce the Hydrogen bond and in all the consequent models the *core-core* repulsion was modified by adding Gaussian stabilizing functions.
The resulting Hamiltonian equations are the same for AM1, RM1 and PM3 [83], while other further modifications were carried out by the successive PM6-D [84] and PM7 [85]. In particular, the RM1 and PM-N models differ from the older MNDO ones [80, 81] by the fact that the parameters are optimized simultaneously and it was adopted a larger reference data set.

Moreover, RM1, PM6-D and PM7 include for many elements a set of five d-functions inside

the *basis set*, allowing the study of heavier elements. Also a dispersion (and hydrogen bonds) contribution *post*-SCF correction was added to PM6, resulting in PM6-D, while in the case of PM7 the parametrization has been achieved after having incorporated those corrections into the method. This last correction should allow PM7 to give more correct heats of formation than the previous method PM6-D [86]. The most recent parametrization model consist then in PM6-D3, where the dispersion term is replaced by the DFT-D3 [87] correction.

Independently from the corrections terms included, the accuracy of each method is also strictly dependent on the system and it should be tested, since different data sets were used to obtain the different parametrizations.

Chapter 4

Chemical Dynamics simulations for Collision-Induced Dissociation of biomolecules

Collision induced dissociation (CID) is widely used in experiments for fundamental and analytical studies of the unimolecular dissociation of molecular ions [2]. The standard theoretical approaches for CID were discussed in chapter 2, as well as the experimental methodologies and applications in Mass Spectrometry.

By using direct dynamics, and employing two different activation models, two fragmentation limits of CID may be investigated (see section 3.3.3): (1) explicit collision simulations, in the tens of picosecond time scale, provide the products observed mainly through non-statistical fragmentations, i.e. products formed on time scales much faster than full IVR; and (2) simulations with internal energy activation of the peptide ion may be used to establish fragmentations with thermal random activation, providing also the Arrhenius parameters for the observed pathways. Thermal activation pathways and time scales may be thus compared with those observed for collision dynamics, while activation energies may be compared with electronic structure theory calculations. This will allow to obtain information on the statistical or non-statistical behaviour of the observed unimolecular dissociation.

Direct dynamics simulations have been extensively used to study the CID of organic and biological ions [27, 29, 30, 31, 88, 89, 90, 14, 91, 25, 92, 93, 24]. In this chapter they are compared and summarized in the context of describing the role of chemical dynamics simulations

for CID experiments.

In particular, we first need to analyse three aspects relative to the unimolecular dissociation induced by collision that are: (i) the amount of energy that each collision transfers to the ion; (ii) the vibrational-rotational partitioning of the energy that is transferred upon the collision; (iii) the localization of the energy in a particular internal mode after the collision and the way this energy can further flow through the other internal modes.

The first two aspects are discussed in the following section, 4.1. The third one, a limit case of energy localization, the so-called shattering mechanism, will be presented in section 4.2. The way in which the energy is distributed influences the consequent fragmentation dynamics. Thus, an example study is given in section 4.3 and it is used to discuss the complementarity in CID of the two dissociation mechanisms (statistical or non-statistical) and their relationship with the activation mode in simulations. Particular importance will be given to the information that may be recovered when the fragmentation is the consequence of a statistical internal energy distribution.

4.1 Energy transfer to vibration and rotation in collision simulations

In order to simulate CID, it is crucial to learn about the energy transfer to the ion due to the collision with the inert gas. This can be done by simulating explicitly one single collison with fixed collision energy, as we will see in section 4.3. In particular, using equation 3.98 it is possible to relate the collision energy in the gas-ion center-of-mass framework, used in the simulations, with the collision energy used in a quadrupole mass spectrometer [94]. Thus, simulations and experiments can be compared directly.

One approach to simulate an explicit collision consists on modelling the ion, that must be able to dissociate, with a quantum chemistry method (QM), while the interaction with the gas and the gas itself using an analytical function (MM). A general analytical expression for this function is:

$$V = Ae^{-Br} + \frac{C}{r^n} \tag{4.1}$$

where A, B and C are positive (pure repulsive potential) and can be obtained by fitting interaction energies calculated *ab initio*, while n can be fixed or also fitted.

When studying large molecules (more than 10 atoms), like peptides, a building block approach can be employed:

- The interaction of the gas with, for example, CH_4, NH_3 and other "blocks" is calculated using high-level electronic calculations (e.g. QCISD(T)).
- Parameters A, B and C for the equation 4.1 are obtained for each building block by fitting the high-level calculations total energies.
- Each building block will be a "portion" of a larger molecule/ion, which interaction potential with a second species (the gas), will be represented by the formula:

$$V = \sum_i \left(A_i e^{-B_i r} + \frac{C_i}{r^n} \right) \tag{4.2}$$

  where A_i, B_i, and C_i are obtained for each building block.

Generally, the interaction is set for each hybridization or environments: e.g. for a carbon, C, a different value for C sp^2 or C sp^3 is used, while for an hydrogen, H, the values are function of the different groups to which it can be attached etc. Employing this approach, and fixing $n = 9$, Meroueh and Hase [12] developed an interaction potential between building blocks of poly-glycines. The same functional form that they reported was then used to model the interaction of N_2 with the building blocks of poly-glycines. The parameters obtained were then used to study CID of protonated urea [31].

The possibility of modelling interactions via simple Lennard-Jones potentials, which parameters are atomic, was investigated [12], but the energy transfer was found to be significantly different from what obtained by accurate parametrizations.

For the $TIK(H^+)_2$ ion, namely threonine-isoleucine-lysine, we have used a full semi-empirical Hamiltonian (RM1) (see section 4.2 of chapter 5). Semi-empirical methods can provide correct interactions for N_2 while they showed to be inefficient for treating rare gas atoms. The choice of a full semi-empirical Hamiltonian or an analytical expression for the two colliding partners interaction depends essentially on the availability of the parameters. Results of RM1 and analytical potential (equation 4.2) for the simple case of N_2 interacting with CH_3OH shows that both provide reasonable short-range (which is very important for CID) interaction potentials in comparison with accurate QCISD(T)/6-31^{++}G** calculations (see figure 4.1).

The energy transfer, over an ensemble of trajectories, can assume a spectrum of values that go from zero to the nominal collision energy. Analysing the data in table 4.1, we should notice that the values, % E_{tr}, observed do not exceed the 50% of the collision energy and they are often largely smaller than this. Furthermore, these values decrease as the collision energy increases, reflecting the fact that the increase of collision energy corresponds to the increase of the projectile velocity and, thus, the time the ion and the projectile interaction decreases. The distribution of energy transfer is generally quite spread, even always resulting in a relatively small value when reporting only the average. If then the energy transfer is decomposed between the reactive and non-reactive trajectories, the first involves always higher energies, as expected. In figure 4.2 it is shown the distribution of energy transfer for collision simulations of N_2 with the $TIK(H^+)_2$ tripeptide ion, that is decomposed between the contribution of non-reactive and reactive trajectories.

The results obtained for the energy transfer in last years simulations for different systems are summarized in table 4.1.

The unimolecular RRKM rate constant is function of both rotational and the vibrational energy [59, 60, 55, 56]. However, rotational energy is often disregarded from the rate constant calculation. In fact its contribution to the rate constant is important for high values of rotational energy only or in the case in which reactant and transition state are quite different.

By the way it is not clear how the collision activation process affects the way in which the energy is partitioned between vibrational and rotational energy. In this sense, simulations can recover this information. For example, they pointed out that rotational activation is higher for planar shape ions and smaller for globular ones [31, 11]. Results of simulation on the percentage of rotational activation are reported in table 4.1. We can immediately notice that rotational activation is more important for small ions and its percentage decreases with the increasing of the collisional energy. This was clearly shown in the collision simulation study of protonated uracil with Ar [27].

In table 4.1 it is also reported the percentage of shattering fragmentation. It is, in fact, possible to define two classes of fragmentation mechanisms for the ion as a function of the time spanned from the collision to the unimolecular fragmentation, that are shattering (see section 4.2) and non-shattering mechanisms. They are thus related to the localization or random distribution of the ion internal energy after the collision.

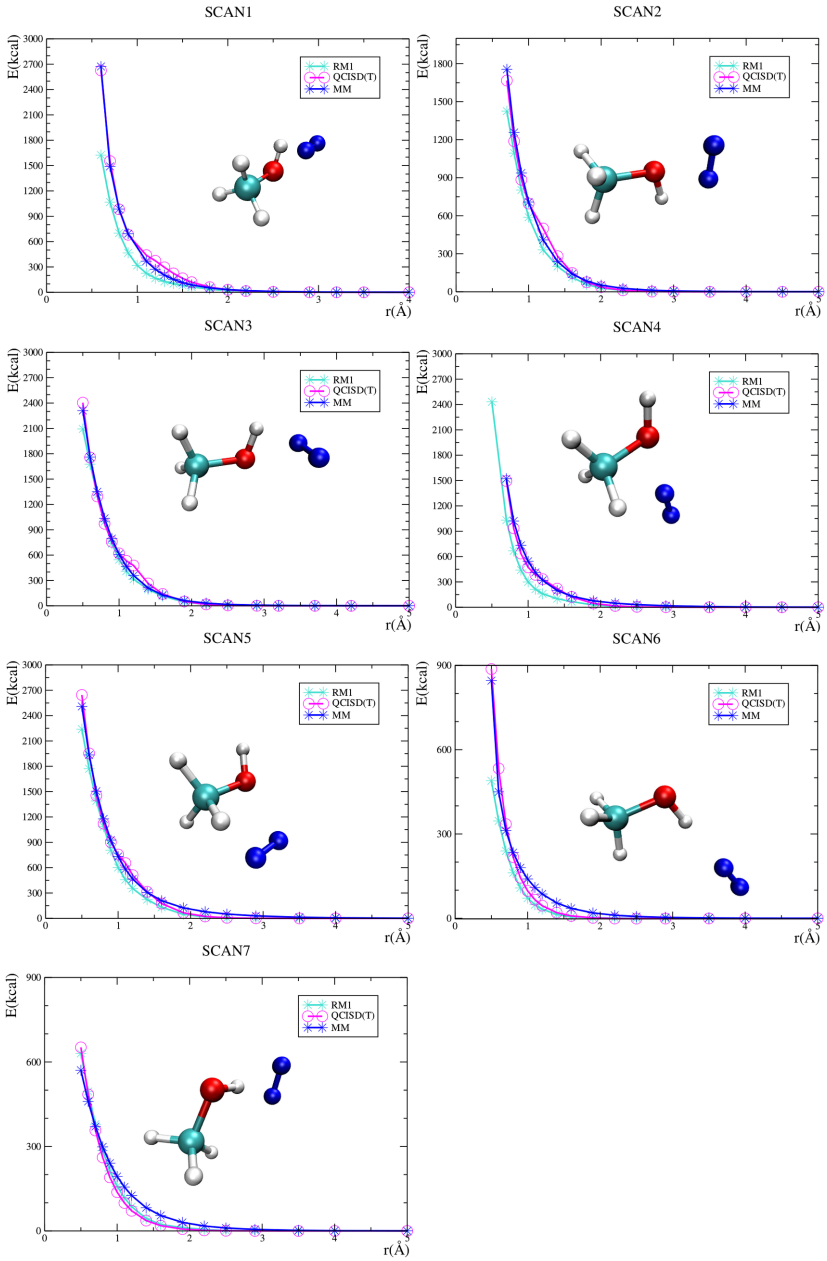


Figure 4.1: N_2-CH_3OH interaction energy curves (different orientations are shown) as obtained from RM1, QCSID(T) and analytical potential (MM) of equation 4.1.

Table 4.1: Summary of results obtained for explicit collision dynamics of different systems as a function of the collision energy E_{col}.

Collision System	E^a_{col}	$\%E^b_{tr}$	$\%E^c_{rot}$	$\%Shattering^d$	Reference
$UreaH^+$ + Ar	4.40	33	22	1	[30]
$UreaH^+$ + Ar	5.66	32	20	33	[30]
$UreaH^+$ + Ar	6.29	35	21	41	[30]
$UreaH^+$ + N_2	6.29	19	13	14	[31]
$CaUreaH^+$ + Ar	9.11	2	13	–	[88]
$CaUreaH^+$ + Ar	13.00	20	8	–	[88]
$CaForm^{2+}$ + Ar	7.80	14	11	–	[29]
$CaForm^{2+}$ + Ar	9.97	10	8	–	[29]
$CaForm^{2+}$ + Ar	12.14	9	6	–	[29]
$SrForm^{2+}$ + Ar	7.80	9	6	–	[29]
$SrForm^{2+}$ + Ar	9.97	8	5	–	[29]
$H^+Gly_2NH_2$ + Ar	15.18	8	3	–	[93]
$H^+Gly_3NH_2$ + Ar	15.18	12	0.7	–	[93]
$H^+Gly_5NH_2$ + Ar	15.18	51	3	51	[24]
$H^+Gly_5NH_2$ + Ar	15.18	14	0.6	50	[24]
$H^+Gly_8NH_2$ + Ar	15.18	22	0.7	55	[24]
Pro_2^- + Ar	13.01	20	–	–	[79]
$TIK(H^+)_2$ + N_2	10.8	23	–	50.8	[95]
$TIK(H^+)_2$ + N_2	12.9	18	–	57.8	[95]
$TIK(H^+)_2$ + N_2	30	3.5	–	91.6	[95]
$TLK(H^+)_2$ + N_2	12.9	11	1	42	[95]
$TestosteroneH^+$ + Ar	30	–	–	22*	[96]

[a] Collision energy in eV and in the center-of-mass framework

[b] Percentage of total energy transferred to the ion (average values calculated over the number of non-reactive trajectories).

[c] Percentage of the energy transferred to the ion that is rotational energy (average values calculated over the number of non-reactive trajectories).

[d] Percentage of fragmentation that is shattering (average values calculated over the number of reactive trajectories).

* % Shattering for the most important fragments.

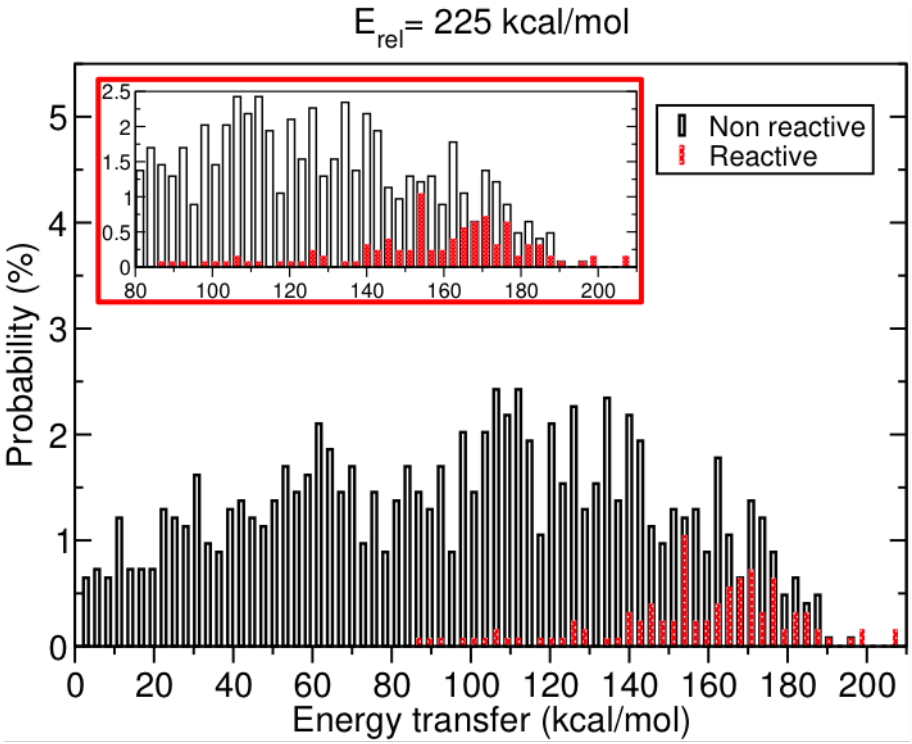


Figure 4.2: Distribution of the energy transfer as obtained in explicit collision dynamics for N_2 + TIK$(H^+)_2$ at the collision energy, E_{rel}, of 225 kcal/mol in the center-of-mass framework. The contribution of energy transfer of reactive trajectory is in red and of non-reactive trajectories in black.

4.2 A collision dynamical effect: Shattering fragmentation

A shattering mechanism consists on a bond cleavage within one or less vibrational period after the collision. In this case the collision gives the energy to a specific bond that consequently breaks almost immediately, i.e. before the IVR takes place. This mechanism was initially observed in surface induced dissociation simulations (SID) [32, 33], where the ion can "shatter", i.e. breaking in many "pieces" just after colliding with the surface.

This mechanism is a typical example of non-RRKM dynamics, due to the collisional activation. Even if, generally, the ions break where the gas hits, this happens not because the phase space of the ion is partitioned (like in intrinsic non-RRKM behavior), but because a bond received energy enough to break and consequently the ion fragments really fast.
A second mechanism can be defined, called non-shattering, that involves an energy flow through the internal modes before the fragmentation. In particular, if the energy flow consists on a full-IVR, a statistical RRKM behaviour of the fragmentation kinetics should be expected. If this is not the case, we expect an intrinsic non-RRKM behavior [97].

In table 4.1 we report the percentage of shattering obtained from collision simulations of different systems. As expected, the percentage of shattering increases with the collision energy. The fraction of shattering is particularly high for the penta- and octa-glycines, where it occurs especially at the side chains, that are more exposed to the collision. It has also to be noticed that larger molecules will need more time for a complete or semi-complete IVR. However, the simulation time is limited ($\approx$ 10-20 ps), and for this reason the fraction of shattering over the number of reactive trajectories will be intrinsically higher for larger ions. The limitation due to the simulation time will, anyway, always overestimate the fast processes.

4.3 Statistical *vs* Non-Statistical fragmentation mechanisms and time scales

As it was discussed in section 3.3.3, it is possible to model the activation obtained in quadrupole-CID experiments with a single collision in simulations. In the same way, it is possible to model CID ion-traps experiment using internal energy activation simulations. Another possibility is

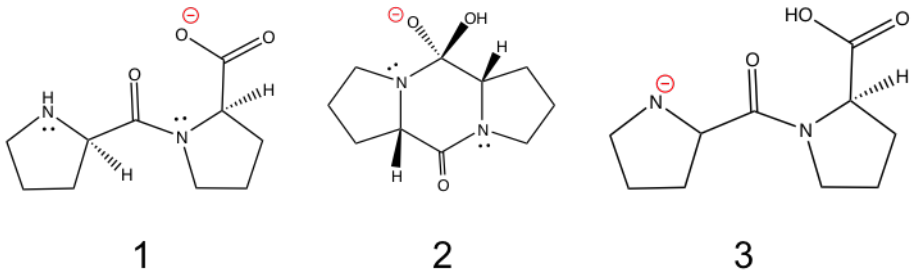


Figure 4.3: Initial structures used in the simulations of $[Pro_2\text{-}H]^-$: 1 "carboxylate", 2 "3-cycles" and 3 "N-terminus anion".

to model a whole generic CID experiment using both the single collision and internal energy activation.

In fact, one limiting aspect for simulations is the time-scale, especially for collision activation. The statistical fragmentation, i.e. the long time-scale processes, can then be recovered using internal energy activation. On the other hand, the dynamical effects, missing using only internal activation, are recovered by simulating explicitly the collision.

This possibility was recently investigated studying the unimolecular dissociation of a relatively simple dipeptide, the di-proline anion.

4.3.1 Internal energy activation simulations

Simulations are performed with the VENUS [98] software coupled with MOPAC [99] for the semi-empirical Hamiltonians. Internal energy activation was implemented using the classical microcanonical sampling, that was discussed in section 3.3.3.

As starting structures for the simulations three isomers of the di-proline anion $[Pro_2\text{-}H]^-$ were used and are reported in figure 4.3, where the carboxylate is the most stable one. The PM3 [83] semi-empirical Hamiltonian (see section 3.4.2) was used for their optimization and for the simulations, while the Velocity Verlet [74] algorithm was used to integrate the nuclear motion equations with a time step of 0.2 fs. Trajectories were propagated up to 10 ps.

An ensemble of 200 trajectories was used for each internal energy and for each initial isomer, and they were propagated up to 10 ps. In table 4.2 it is shown the correspondence

Table 4.2: Correspondence between internal vibrational energy and temperature for $[Pro_2\text{-}H]^-$ thermal activation simulations.

Vibrational energy (kcal/mol)	Temperature (K)
208.64	1250
250.36	1500
292.09	1750
333.82	2000
375.54	2250
417.27	2500
459.00	2750
500.72	3000

between temperature and total vibrational energy. In fact, it is possible to demonstrate that the classical RRKM rate constant (equation 3.27 of section 3.2) and the classical transition state theory (TST) rate constant:

$$k(T) = \nu \exp\left(-\frac{E_0}{k_B T}\right) \tag{4.3}$$

are equal for large energies ($E >> E_0$) and large number of normal modes ($s \approx s - 1$). Then for classical harmonic oscillators, the energy and the temperature, T, are related by:

$$E = s k_B T \tag{4.4}$$

where k_B is the Boltzmann constant.

Thus, using equation 4.4 it is possible to relate the energy of the microcanonical ensemble to the temperature and, consequently, to set the microcanonical rate constant $k(E)$ equal to canonical rate constant $k(T)$. For this reason, we often refer to internal energy activation simulations also as to thermal simulations, especially when they are used to calculate thermal rate constants.

Rate constants

For each set of simulations it is possible to follow the decay of the reactants population as a function of time and consequently to obtain the canonical rate constant, $k(T)$. In fact, if the decay is statistical, the population decay will be described by a single exponential:

$$N(t) = N(0)e^{-k(T)t} \tag{4.5}$$

where $N(t)$ is the number of reactant molecules as a function of the time, and $N(0)$ at the time 0. *k(T)* is the rate constant and t the time.

The decay of the initial population is found to be a single exponential if using carboxylate as the starting structure, as it is shown in figure 4.4. Note that the resulting rate constant reflects the anharmonicity of the potential energy surface (PM3) and of its vibrations [26].

Since, using the carboxylate as the starting structure the population decay is exponential, it is possible to plot (see figure 4.5) the $k(T)$ as a function of the temperature and obtain Arrhenius parameters using the Arrhenius equation:

$$k(T) = A \exp\left(-\frac{E_a}{k_B T}\right) \tag{4.6}$$

where A is the Arrhenius pre-exponential factor and E_a is the activation energy, that corresponds to the total reactivity.

The activation energy obtained from the fit of the Arrhenius plot of figure 4.5 is then 36.3 $\pm$ 3.3 kcal/mol, that is in the same order of magnitude of typical peptides activation energies.

If we plot the decay of the initial population using the other two isomers as the initial structure we do not always obtain single exponential decay. For example, as we can see in figure 4.6 for the N-terminus simulations there is an initial fast decay that can be fit by a single exponential, followed by a plateau region, corresponding to isomerization mainly to the carboxylate isomers, and eventually further reactivity.

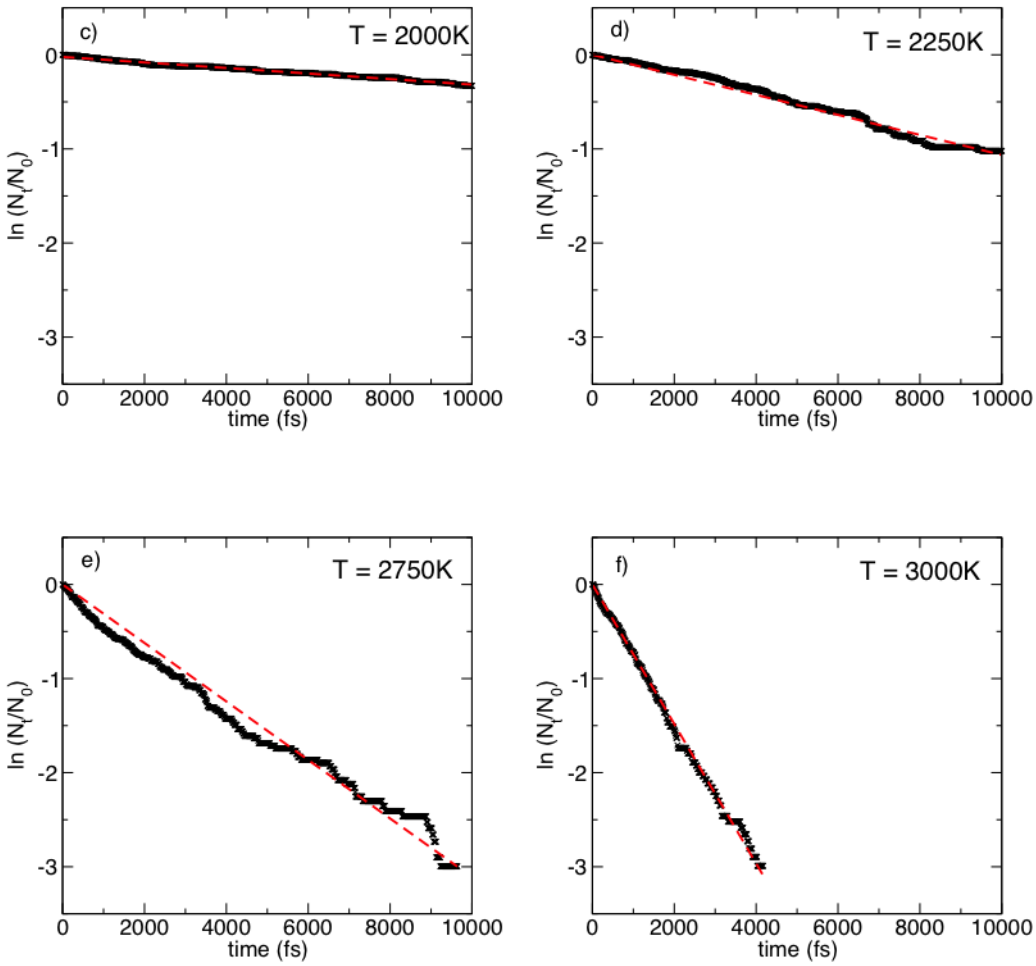


Figure 4.4: Population decay (in logarithmic scale) of $[Pro_2\text{-}H]^-$ using the carboxylate as starting structure for thermal simulations at different temperatures.

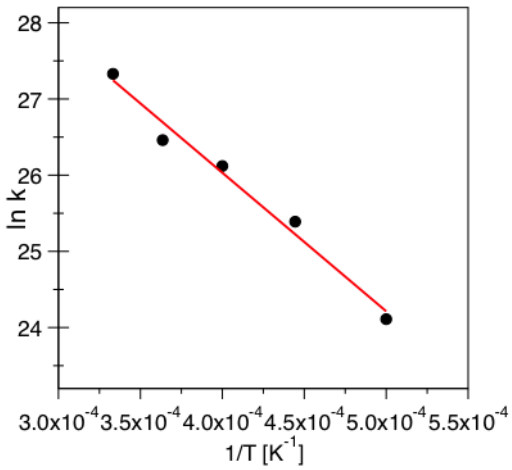


Figure 4.5: Arrhenius plot obtained from temperature activation simulations using carboxylate (structure 1 in figure 4.3) as the initial structure. Black dots are the values obtained from carboxylate simulations, the red straight line corresponds to the fit using 4.6.

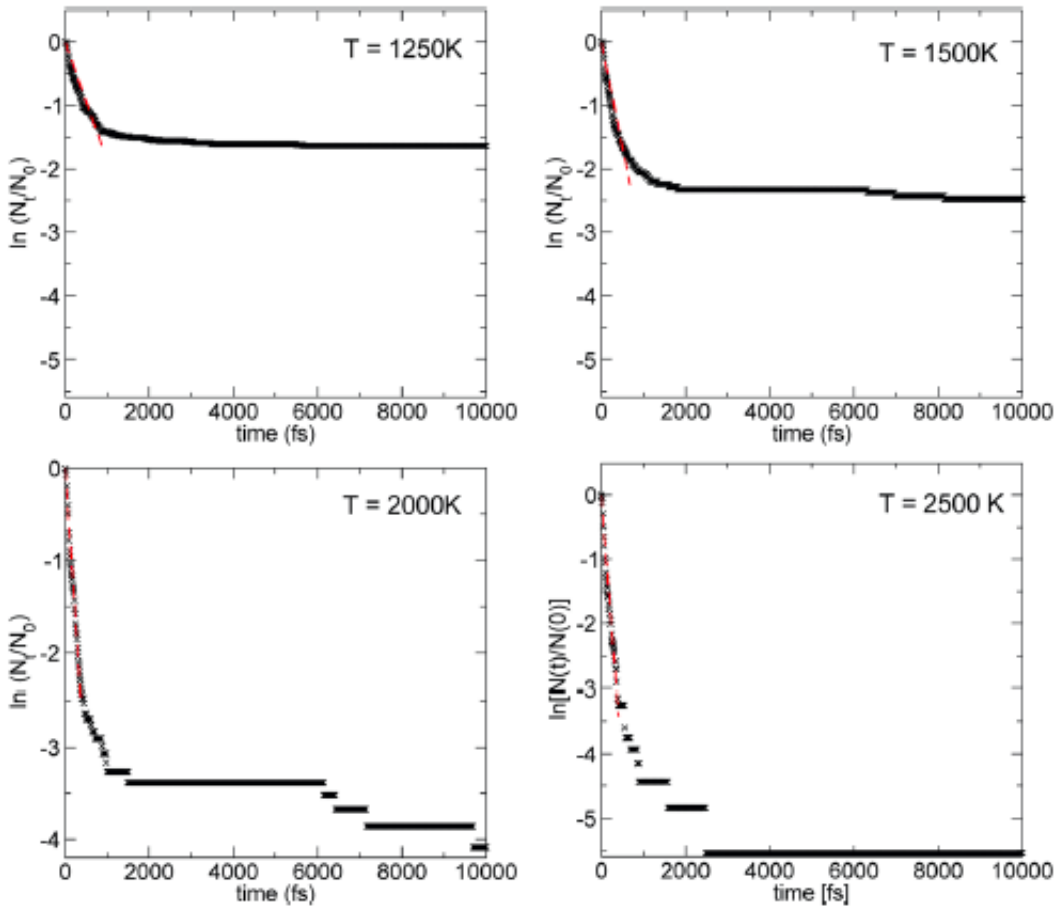


Figure 4.6: Population decay obtained using the N-terminus anion at 1250, 1500, 2000 and 2500 K. In red we show a linear fitting corresponding to the fast decay. The corresponding k values are: $1.87 \times 10^{12} \pm 6.22 \times 10^{10}$ s^{-1} (1250 K), $3.20 \times 10^{12} \pm 1.19 \times 10^{11}$ s^{-1} (1500 K), $6.61 \times 10^{12} \pm 1.52 \times 10^{11}$ s^{-1} (2000 K) and $8.33 \times 10^{12} \pm 2.54 \times 10^{11}$ s^{-1} (2500 K).

4.3.2 Explicit collision activation simulations

In these simulations the $[Pro_2\text{-}H]^-$ ion collides with an Ar atom, where the three isomers in figure 4.3 were used as initial structure. A QM+MM approach was used, where the internal energy of the ion is represented by the PM3 semi-empirical Hamiltonian, while the analytical function of equation 4.2 is used for the $[Pro_2\text{-}H]^-$-Ar interaction.

The initial conditions were given as described in section 3.3.3. In particular, a collision energy of 300 kcal/mol was considered and the impact parameter was sampled between 0 and 5.0 Å. The b_{max} value of 5.0 Å was chosen by evaluating the average energy transfer as a function of the impact parameter b. This is shown in figure 4.7. For $b = 5$ Å there is less than 10% of energy transfer, and higher values were not considered in the simulations because the reactivity would be too low. The ion and the Ar atom were placed at the initial relative distance of 10 Å in the center-of-mass framework. The distance is chosen such that the gas-ion interaction is negligible. The simulations were stopped when the relative distance reached 400 Å allowing a simulation time in the range $5 - 40$ ps. To integrate the nuclei motion equation the *Velocity Verlet* [74] algorithm was used. The number of trajectories simulated was chosen in order to have about 1000 reactive trajectories for each set of simulations.

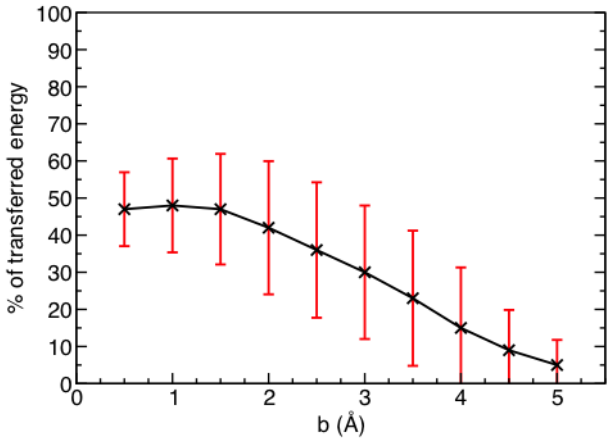


Figure 4.7: Percentage of energy transfer as a function of the impact parameter, b, for Ar + $[Pro_2\text{-}H]^-$ simulations at a 300 kcal/mol of collision energy.

Collisional time scales

By grouping the trajectories as a function of the energy transfer it is possible to compare the collisional simulations results with the thermal ones. The trajectories, so grouped, were analysed in order to plot, in figure 4.8 the decay of the initial population *vs* the time. As expected, with larger energy transfer the decay rate of the initial population increases. But more interesting, these decays have a highly non-exponential character, especially at small time. Then, after the initial highly non-exponential decay (due to the non-random collisional excitation of the ion and its consequent non-statistical fragmentation), there is a plateau region with no reactivity. In this time it is reasonable to think that IVR may be occurring so that the ions will dissociate statistically at times longer than the simulation time-length.

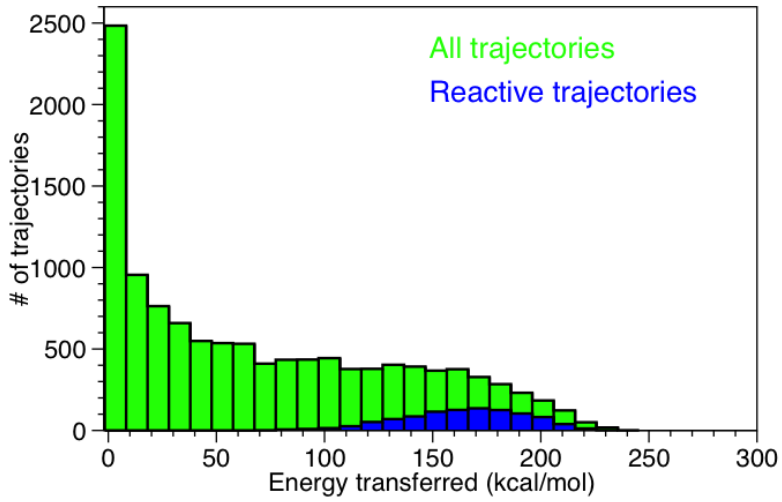


Figure 4.8: Distributions of the energy transfer after the collision ion-Ar at the collision energy of 300 kcal/mol for all the trajectories (in green) and for reactive ones (in blue).

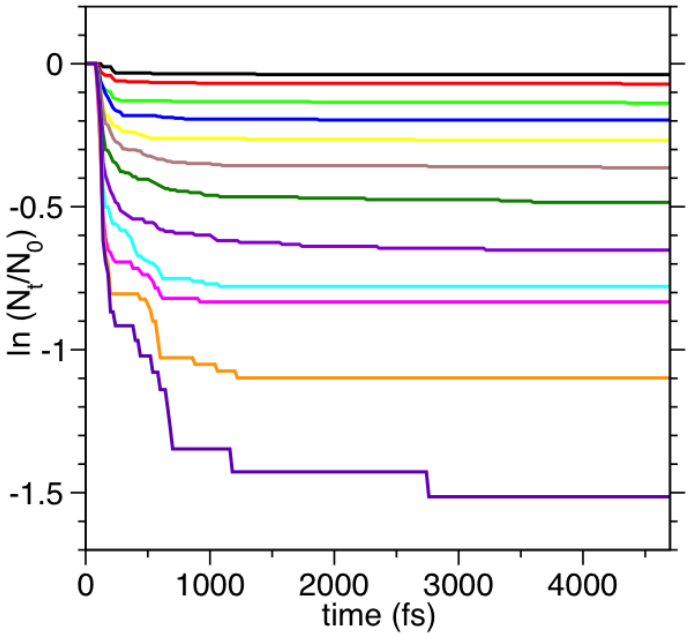


Figure 4.9: Decay of the initial population for the collisional simulations for different energy transfer values. From top (black) to down (purple) we show energies in the $110-230$ kcal/mol range (a bin of 10 kcal/mol is used).

4.3.3 Collisional *vs* thermal simulations results

We can compare the results obtained from collisional and thermal simulations starting from the fragments that were obtained.

Different reaction products can be observed as a function of the activation method and the temperature (for thermal simulations). Mass distribution results for the collisional simulations and one prototypical temperature are shown in figure 4.10. The temperature of 2250 K was chosen because of the high reactivity percentage: 66%, 77% and 100% for the carboxylate, 3-cycles and N-terminus anion simulations, respectively. These values reflect the energy ordering: carboxylate is the most stable structure, followed by 3-cycles (9.4 kcal/mol higher) and N-terminus anion (34.3 kcal/mol higher).

Note that in figure 4.10 we show the mass distribution of the different products (we report both charged and neutral products, this information is wider than what is obtained in the experimental mass spectra where only charged ions are present), the structures of the most important ions obtained are shown in figure 4.11 with the corresponding typical nomenclature used in mass spectrometry of peptides [42].

From figure 4.10, it can be shown that some of the products can be obtained from both collisional and thermal simulations, for example CO (mass 28.01 *amu*), CO_2 (mass 44.01 *amu*), a_1^- (mass 70.11 *amu*), b_1^- (mass 98.12 *amu*) or the ion with mass 167.23 that correspond to the loss of CO_2. Some other products, instead, like ethylene (mass 28.05 *amu*) are present only in the thermal activation simulations, while x_1^- (mass 142.13 *amu*) is present mainly in the collisional simulations. In particular this product is mostly obtained when the initial structure is the 3-cycle or the N-terminus anion. The most relevant peaks in experiment are also obtained through simulations and they are b_2^-, y_1^- and m/z 165. We observed additional fragments, which probably further react in the experiments. Note that in figure 4.10 we plot the mass distribution and not a theoretical MS spectrum.

It is also interesting to note that when using the 3-cycles structure as the initial isomer (see panel *b* of figure 4.10) we obtain both H_2O and OH^- loss in the case of CID, while for thermal dissociation we obtain only H_2O. This reaction pathway is particularly relevant for peptide dissociation: in the case of H_2O loss this corresponds to the formation of the b_2^- ion typical of peptide reactivity. To form this ion, the OH^- group has to leave with not much translation energy along the breaking C-O bond such that it is possible to take a proton from the adjacent C-H group and form H_2O and b_2^-. The mechanism is shown in figure 4.12.

In thermal dissociation the second step, crucial for formation of H_2O and the b_2^- ion, is possible, since the energy is distributed over all the modes and thus it is not concentrated on the C-O stretching. In this way, the leaving group has time to take the adjacent H atom. In collision simulations, the C-O bond mainly breaks when the energy is concentrated on the bond just after the collision with Ar. This is an example of shattering fragmentation (see previous section 4.2). Under this regime, the OH^- group has a relatively high translational energy that comes from a high energy along the C-O stretching, such that it may have insufficient time to interact with the adjacent H atom and form H_2O. Once the OH^- group is too far from the C-H bond the proton transfer is no more possible and the next step providing H_2O and b_2^- is forbidden.

We can conclude that in this case the final product distribution is only slightly affected by the different activation modes. However, it is possible to find some remarkable difference in a

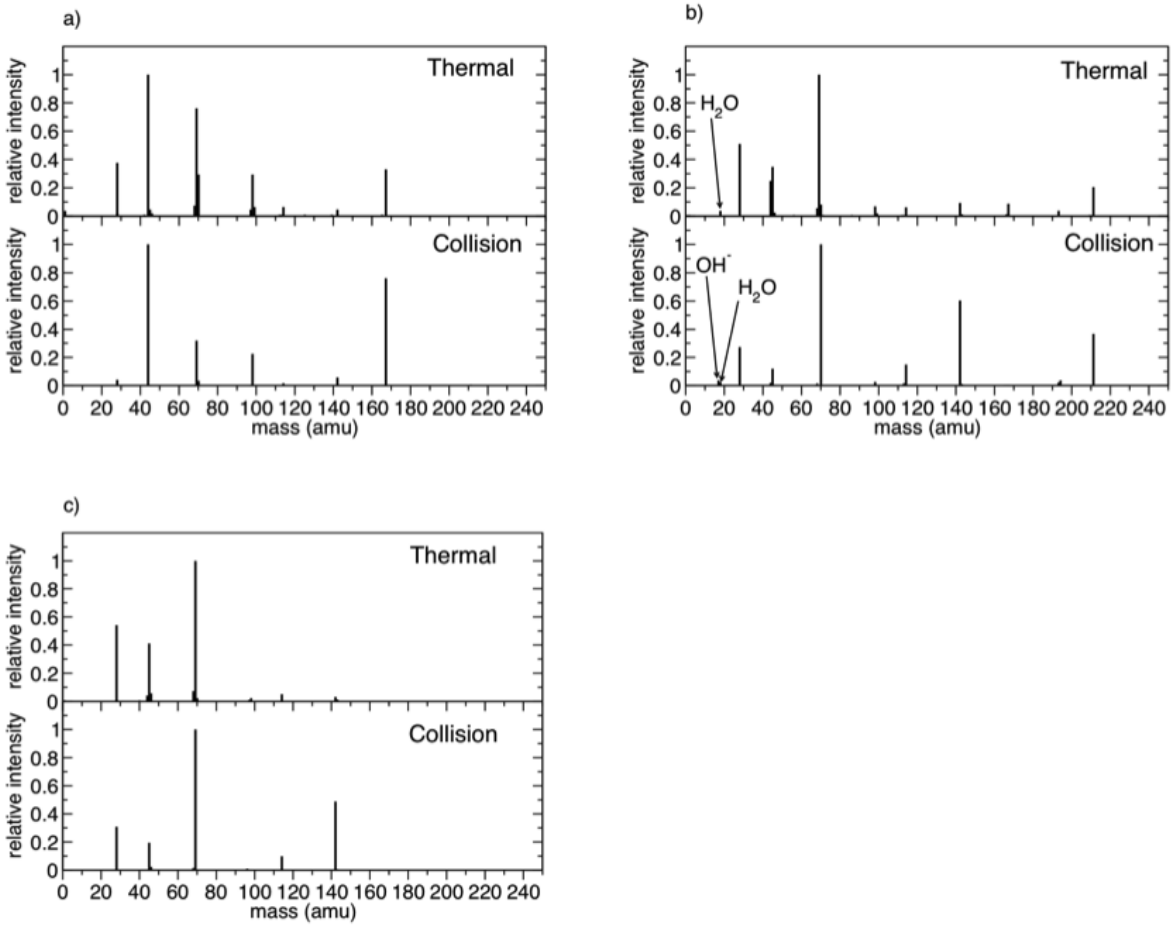


Figure 4.10: Mass distribution of the reaction products observed for collisional and thermal simulations (T= 2250 K), using the carboxylate (panel a), the 3-cycles (panel b) and the N-terminus anion (panel c) as initial structures.

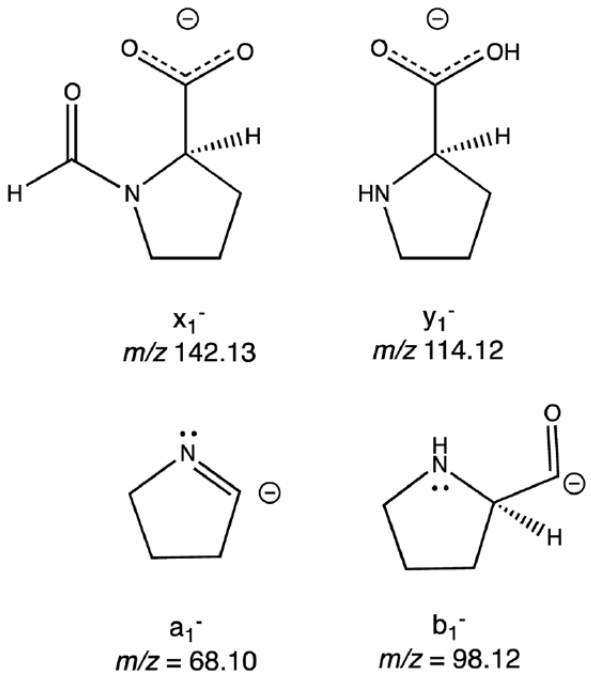


Figure 4.11: Structures of typical ion products, with associated nomenclature generally used in peptide gas-phase fragmentation. Products with trivial structures (e.g. H_2O, CO) are omitted.

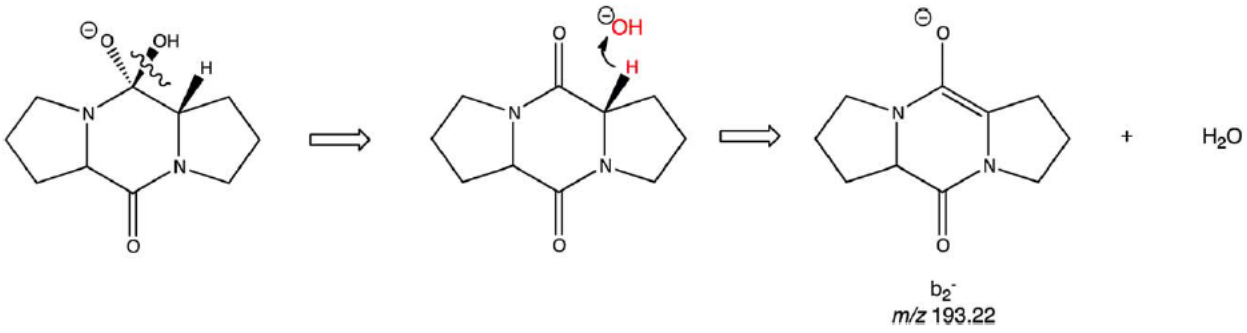


Figure 4.12: Mechanism leading to H_2O loss and formation of the b_2^- ion (m/z 193.22) starting from the 3-cycles isomer.

fragmentation mechanism when crucial dynamical effects are involved, as for the case of the shattering fragmentation responsible of the OH^- product in collision simulations.

4.4 Summary and conclusions

In order simulate a CID experiment and rationalize the fragmentation pathways it is necessary to evaluate different aspects, which have been analysed:

1. The interaction energy of the ion and the gas in collisional simulations can be modelled using analytical functions. However, a full QM approach can be used for relatively big systems, like for $TIK(H^+)_2$ and $TLK(H^+)_2$ tripeptide ions colliding with N_2, this avoiding the parametrization for the ion-gas interaction.

2. Semi-empirical Hamiltonians show to well reproduce the interaction between a molecule and N_2. Anyway, they cannot be used when treating rare gases.

3. The way in which the energy is partitioned between vibration and rotation upon collision is not evident experimentally. Chemical dynamics can recover this information.

4. Chemical dynamics, in particular if the activation is collisional, can show dynamical effects, like fragmentation.

5. Internal energy activation simulations may be used to calculate thermal data, like rate constants and activation energies either of the global fragmentation or of single pathways. These data contain the intrinsic anharmonicity of the vibrations, unlike results coming from electronic structure theory calculations.

6. Collisional and internal energy simulations can be used together to rationalize a generic CID experiment. One can be used for short time-scale and more dynamical fragmentation; the other for the longer time-scale and more statistical one.

7. CID in a quadrupole, in ideal conditions, may be modelled through explicit collision dynamics.

8. CID in an ion-trap may be modelled performing thermal simulations.

9. Mass spectra distributions resulting from simulations can provide information about the different fragments, both charged and neutral, and the fragmentation mechanisms. Note that the masses of the neutrals are not directly observable in experiments and the structures of ions can be obtained only if coupling with infra-red multiple photon dissociation (IRMPD) spectra [7, 100].

Chapter 5

On the fragmentation of the TIK$(H^+)_2$ and TLK$(H^+)_2$ tripeptides

5.1 Introduction

In this chapter we report a chemical dynamics study on two relatively "large" biological systems, i.e. the threonine-isoleucine-lysine TIK$(H^+)_2$ and threonine-leucine-lysine TLK$(H^+)_2$ tripeptides ions.

It was previously outlined (see chapter 2) the importance of collision induced dissociation (CID) of peptides in proteomics [2, 19], and more in particular in the bottom-up approach [19]. Despite the large number of studies on peptides fragmentation, all the reaction pathways and associated mechanisms are not definitively understood and classified [44, 47, 101, 102, 103, 104].

I and L have identical masses but different side-chain structures, thus knowing their fragmentation dynamics will be informative to protein identification when these side-chain losses occur. In particular there is limited experimental information about their side-chain fragmentation, but there is some evidence for a difference in side-chain fragmentation for I and L in the literature [43]. To interpret biological MS/MS experiments for peptides containing I and L it is important to establish their side-chain fragmentation mechanisms. Thus, the likelihood of a particular fragment being lost from I or L needs to be better understood. These fragmentation probabilities for I and L may depend on the neighboring amino acids, so this positional context also needs to be explored. A complete positional characterization will require fragmentation

studies of X-I-Y and X-L-Y (where X/Y is an amino acid), which means there are a total of 800 possible tripeptides that contain isoleucine or leucine in the center. To reduce the number, we first choose as X threonine, which has well-documented side-chain losses [43, 105]. For Y we chose lysine because it can be easily protonated. It is at the N-terminus of many tryptic peptides commonly used for protein identification, and it has a long side-chain that provides for interesting dynamics. The study of a doubly protonated tripeptide guarantees that only those with a lysine or arginine are likely.

In section 4.3 we discussed about the possibility of modelling a generic CID experiment using both collision and internal energy (or thermal) activation simulations. The first activation way can give information about the short-time scale and dynamical fragmentation, while the second can recover longer-time scale and statistical dissociation pathways.

Thus, we first performed internal energy activation simulations for the TIK$(H^+)_2$ ion in order to study its thermal fragmentation (section 5.2). Successively, we performed collision activation dynamics on both the ions (section 5.3). A comparison is made between thermal and collisional simulations results and with a mass spectrometry database. For collisional trajectories a limit fragmentation mechanism, namely shattering (see section 4.2 of the previous chapter), was found to be important.

An important issue for CID of a peptide ion is the threshold for the shattering mechanism. Thus, we investigated the shattering threshold for both backbone and side-chain fragmentations of the doubly protonated tripeptide threonine-isoleucine-lysine ion, TIK$(H^+)_2$ (section 5.4).

5.2 Model Simulations of the Thermal Dissociation of the TIK$(H^+)_2$ Tripeptide: Mechanisms and Kinetic Parameters

We now report chemical dynamics simulations performed for the thermal dissociation of the doubly protonated tripeptide threonine-isoleucine-lysine, TIK$(H^+)_2$. The primary structure of TIK$(H^+)_2$ is shown in figure 5.1 and it represents the fragmentation nomenclature given by Roepstorff and Fohlman [42]. From the simulations we can obtain the fragmentation pathways, their rate constants versus temperature, and their Arrhenius parameters.

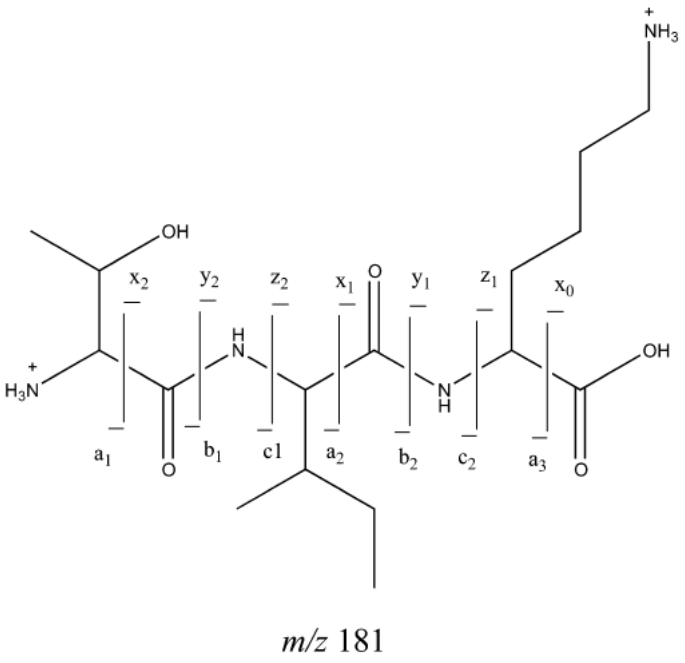


Figure 5.1: Primary structure of $TIK(H^+)_2$ and relative fragmentation nomenclature of Roepstorff and Fohlman [42].

5.2.1 Methodology

The methodology employed to simulate internal energy activation simulations is reported in section 3.3.3 of chapter 3. The RM1 semi-empirical Hamiltonian [82] was used for the simulations. Semi-empirical Hamiltonians are discussed in section 3.4.2 of chapter 3. Three low-energy conformers of $TIK(H^+)_2$ were found performing optimizations both with RM1 and a higher level theory method (DFT) and they gave similar results. The three conformers are shown in figure 5.2.

The lowest energy conformer of figure 5.2 was chosen as initial structure for the trajectories. However, to investigate if the fragmentation dynamics would change exciting a higher-energy conformer, an additional simulation of 100 trajectories was performed at 2000 K for excitation of the highest-energy conformer in figure 5.2. Simulations were performed using the package consisting of VENUS chemical dynamics computer program [98, 106] interfaced with MOPAC electronic structure code [107].

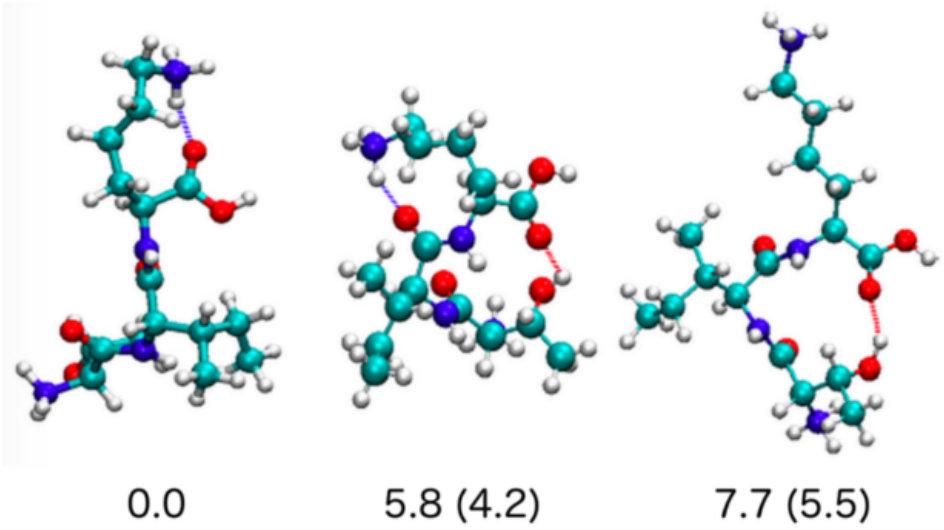


Figure 5.2: Three lowest-energy conformers found for TIK$(H^+)_2$ using RM1. The relative energies are given in kcal/mol, with DFT B3LYP$/6-31G^*$ values in parentheses. The blue, white, cyan, and red colors identify the nitrogen, hydrogen, carbon, and oxygen atoms, respectively, while blue and red dashed lines represent hydrogen bonds.

As seen in section 4.3.1 of the previous chapter, for large excess energies (E $>$ E_0) and large number of normal modes ($s \approx s-1$), the microcanonical rate constant $k(E)$ is equal to the canonical rate constant $k(T)$. In order to obtain $k(T)$, the energy, E, of the internal energy activation simulations is related to the temperature T through the equation 4.4.

We performed 600 trajectories with RM1 at four different temperatures (or energies), i.e. at 1250, 1500, 2000, and 2500 K, corresponding to 423, 508, 677 and 846 kcal/mol internal energies. The trajectories were stopped after the simulation time of 90, 30, 10, and 1.5 ps respectively. The nuclei equations of motion were numerically integrated with a sixth-order symplectic algorithm [108, 109] using a time step of 0.3, 0.1, 0.05 and 0.01 fs respectively for trajectories at 1250, 1500, 2000 and 2500 K. Additional 100 trajectories were performed at 2000 K using as initial structure the highest-energy conformer in figure 5.2.

5.2.2 Results

Only results for the RM1 TIK$(H^+)_2$ thermal simulations will be reported extensively. A more extensive study about the explicit collision activation simulations for TIK$(H^+)_2$ + N_2 and TLK$(H^+)_2$ + N_2 will be reported in the next session 5.3.

Dissociation Pathways and Their Probabilities.

In figures 5.4, 5.5 and 5.6 the eight principal (over 41) primary dissociation pathways are reported for the thermal excitation between 1250 and 2500 K and about the potential-energy minimum in figure 5.2. For 1250, 1500, and 2000 K the dominant primary dissociation is pathway 1. As the temperature is increased more pathways become important, and at 2500 K pathway 3 dominates. At 2000 and 2500 K there are many additional pathways. The number of primary dissociation pathways are 3, 14, 32, and 61 for temperatures of 1250, 1500, 2000, and 2500 K, respectively.

The dissociation mechanisms for additional pathways, which had two or more trajectory events, are described in figures A.1, A.2, A.3 and A.4 in *Appendix A*. They are pathways 9-20. The probabilities for these pathways are summarized in table 5.1.

The increase in the number of products ions as the temperature is increased can be observed in figure 5.3, where we report theoretical mass spectra, obtained as m/z values abundance as a function of m/z.

Experimentally, only a few ions are generally obtained with a relevant abundance. We thus analysed the significant number of m/z values obtained from the simulations for each temperature. This total number of m/z values is 7, 29, 71, and 98 for the respective temperatures of 1250, 1500, 2000, and 2500 K. This large number of ions decreases significantly if we only consider the number of m/z values that contribute 2% or more of the primary dissociation.

Note that the number of pathways do not necessarily correspond to the number of m/z values since different pathways can lead to the same ion(s) (in particular here where the precursor ion is doubly charged).

The relative importance of side-chain fragmentation versus backbone fragmentation is of particular interest. Of the pathways in figures 5.4, 5.5 and 5.6, pathways 5, 7 and 8 involve side-chain fragmentation. Among the pathways 9-20 (shown in figures A.1, A.2, A.3 and A.4 in *Appendix A*), pathways 9-15 and 18 are side-chain fragmentations. From the lower to the higher temperature, the percentages of the fragmentations that are side-chain are 0.0, 2.1,

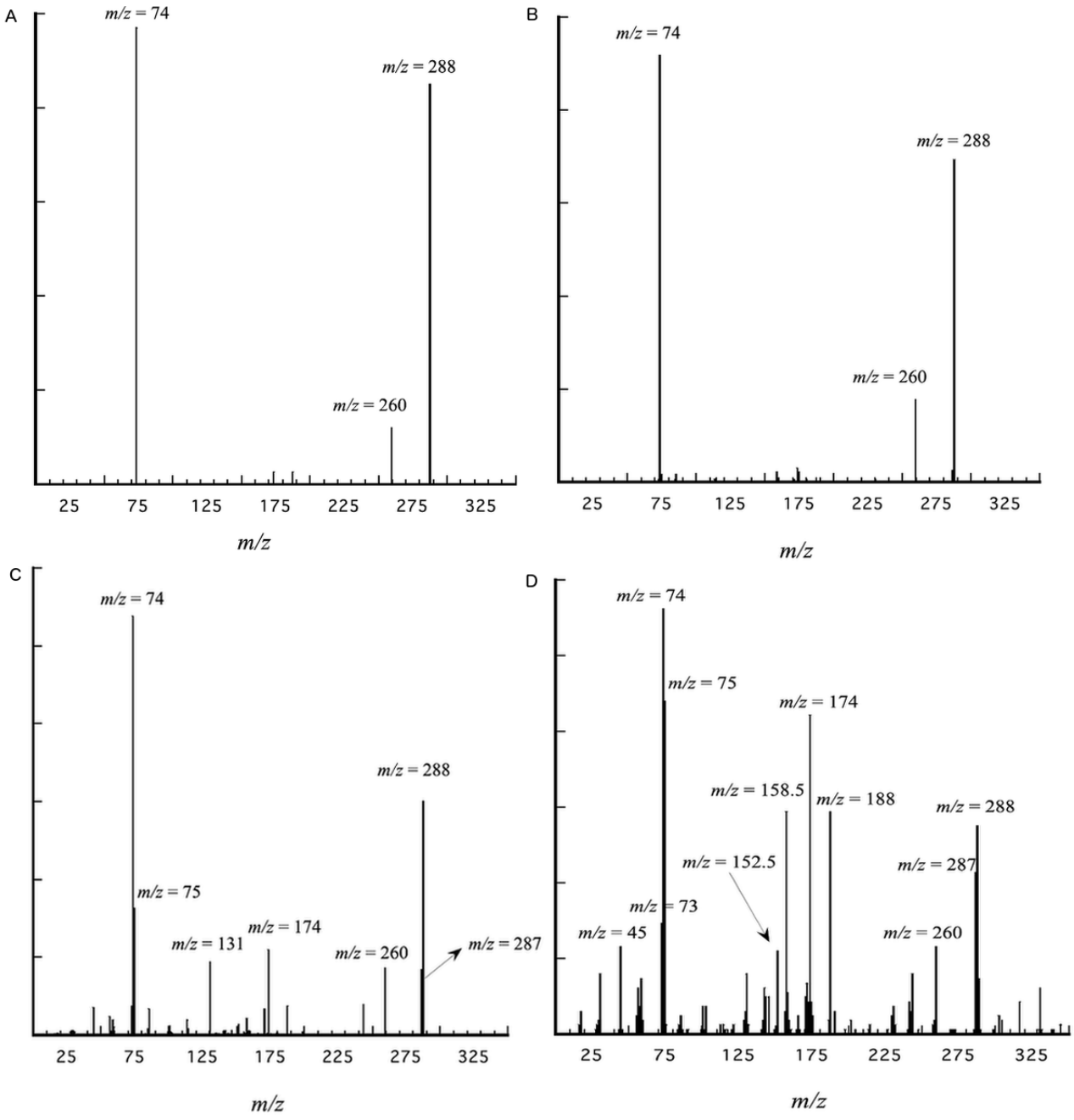


Figure 5.3: Mass spectra as obtained from RM1 trajectories at 1250 K (panel A), 1500 K (panel B), 2000 K (panel C) and 2500 K (panel D).

Table 5.1: Probabilities of the TIK$(H^+)_2$ dissociation pathways as a function of the simulations temperature as obtained from the RM1 direct dynamics.

Pathway	$1250K$	$1500K$	$2000K$	$2500K$
1	55.5(85.5)	49.7(73.2)	31.0(35.4)	11.4(12.9)
2	7.9(12.1)	12.7(18.7)	8.4(9.6)	3.9(4.5)
3	1.6(2.4)	0.5(0.8)	11.0(12.6)	14.6(16.5)
4	0.0	1.1(1.6)	13.3(15.2)	9.6(10.9)
5	0.0	0.7(1.1)	6.0(6.7)	7.1(8.1)
6	0.0	0.2(0.3)	1.9(2.2)	3.2(3.6)
7	0.0	0.0	1.4(1.7)	2.1(2.4)
8	0.0	0.4(0.5)	3.4(3.9)	1.4(1.6)
Other paths	0.0	2.6(3.8)	11.2(12.7)	34.9(39.5)

The probabilities are given as percentages. The percentage of the dissociating trajectories by the pathway is given in parentheses.

15.6, and 36.3, respectively. side-chain fragmentation becomes significantly more important as the temperature (internal energy) of TIK$(H^+)_2$ is increased.

The above results are for exciting about the minimum of the lowest-energy TIK$(H^+)_2$ conformer in figure 5.2. Similar fragmentation results were obtained using the 100 trajectories at 2000 K using as initial structure the highest energy TIK$(H^+)_2$ conformer in figure 5.2. In particular, when exciting this higher-energy conformer, we observe an insensitivity of the dissociation probabilities to which conformer is excited. This result is not unexpected, since the conformer structures are not that dissimilar, and given the high temperature excitations unimolecular isomerizations between the conformers should be rapid and much faster than the dissociations.

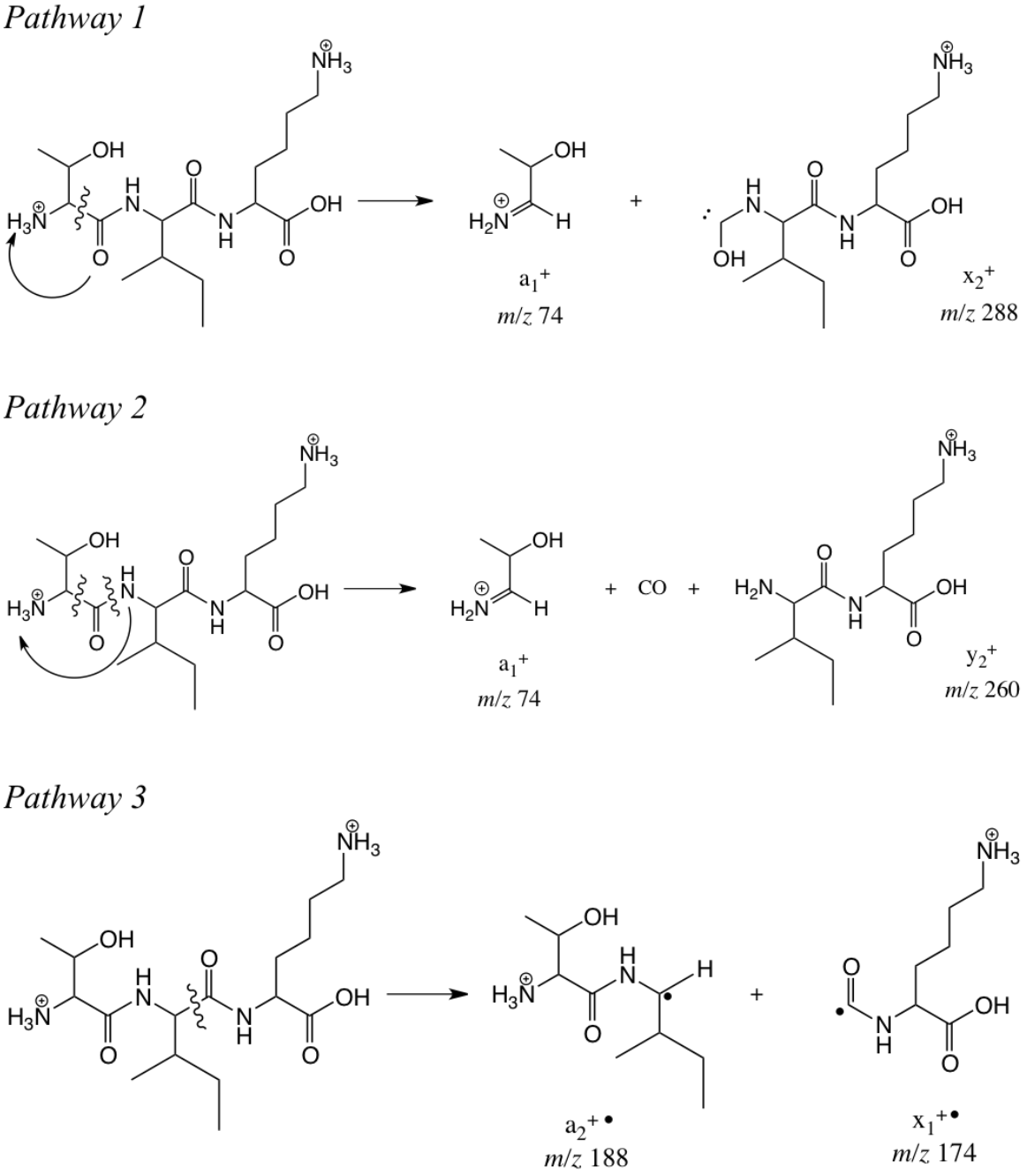
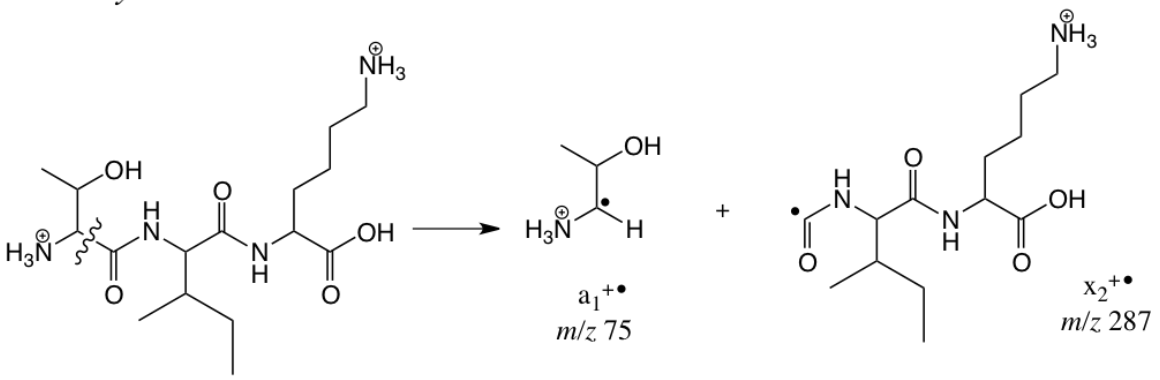
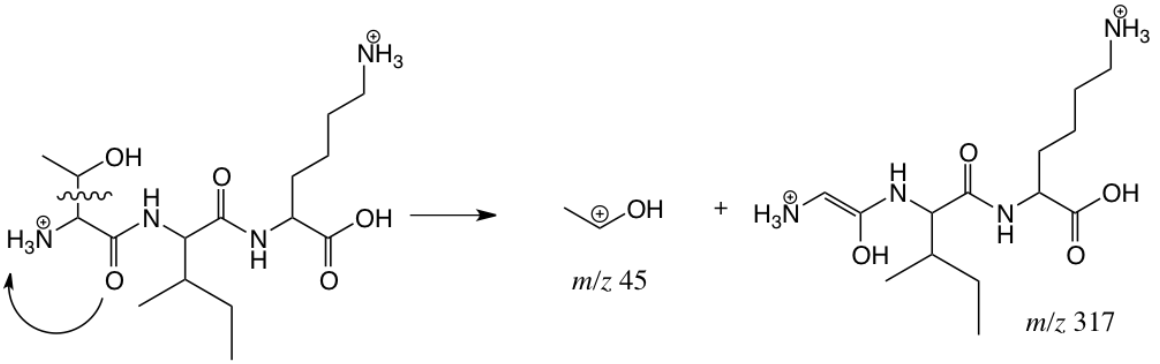


Figure 5.4: Mechanisms for dissociation pathways 1-3 of TIK$(H^+)_2$.

Pathway 4



Pathway 5



Pathway 6

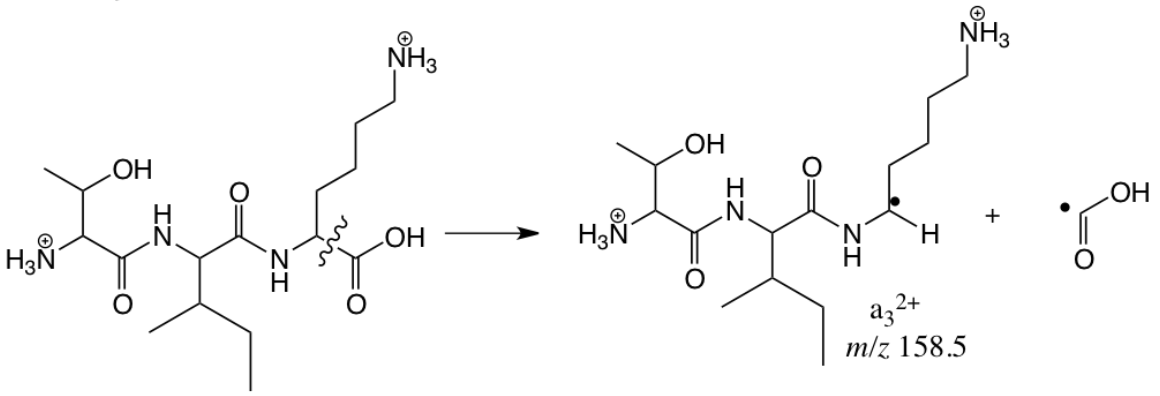


Figure 5.5: Mechanisms for dissociation pathways 4-6 of TIK$(H^+)_2$.

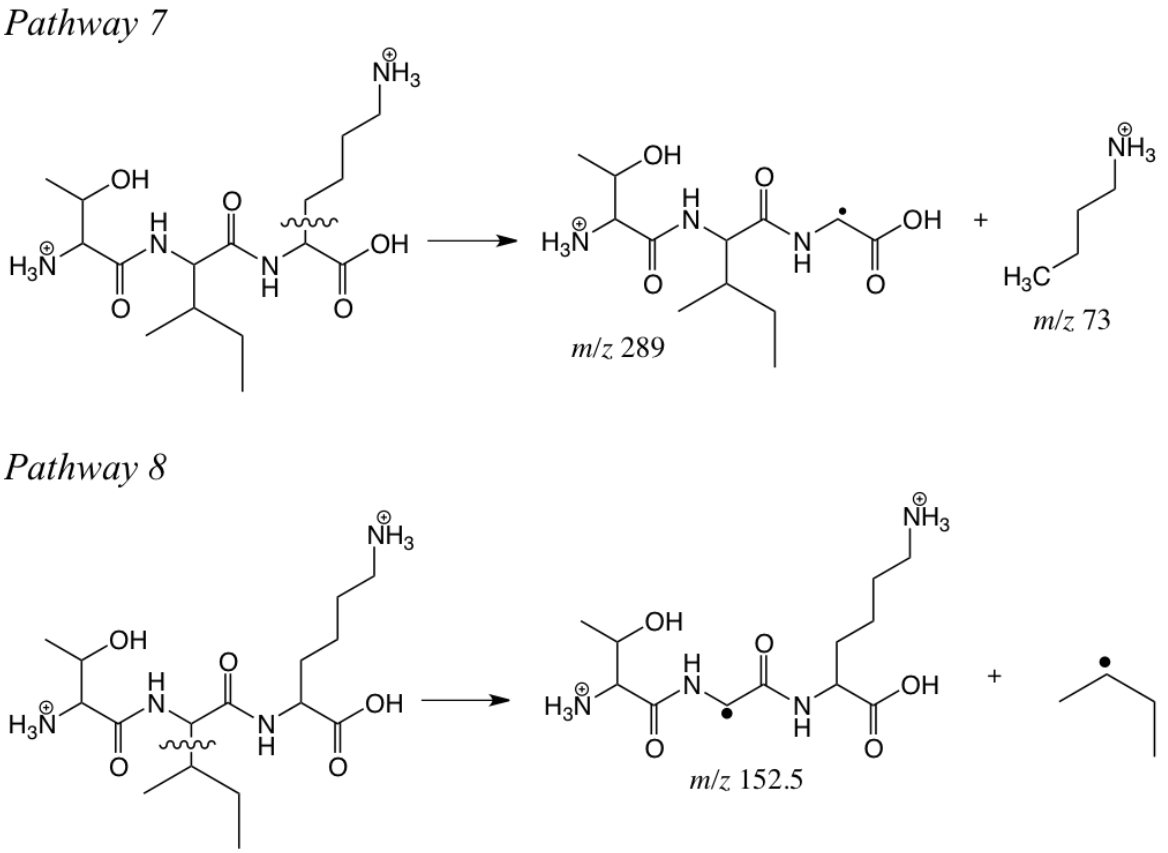


Figure 5.6: Mechanisms for dissociation pathways 7-8 TIK$(H^+)_2$.

Overall Thermal Unimolecular Rate Constant.

If the unimolecular dissociation of TIK$(H^+)_2$ is ergodic, as assumed by RRKM theory, the relative number of TIK$(H^+)_2$ molecules *versus* time, N(t)/N(0), will decay exponentially yielding the unimolecular rate constant. This exponential dissociation dynamics was found for each of the temperatures, for excitation about the lowest-energy potential-energy minimum. An illustrative plot of $\ln[N(t)/N(0)]$ *vs* time is shown in figure 5.7 for the simulation at 1500 K.

The same methodology applied on the thermal simulations $[Pro_2\text{-}H]^-$ (see section 4.3 of chapter 4) is now applied here: the unimolecular rate constant are reported as the slope of this curve. The respective rate constants for the 1250, 1500, 2000, and 2500 K simulations are $(7.50 \pm 0.06) \times 10^9$, $(3.18 \pm 0.03) \times 10^{10}$, $(2.35 \pm 0.02) \times 10^{11}$, and $(8.74 \pm 0.79) \times 10^{11}$ s^{-1}.

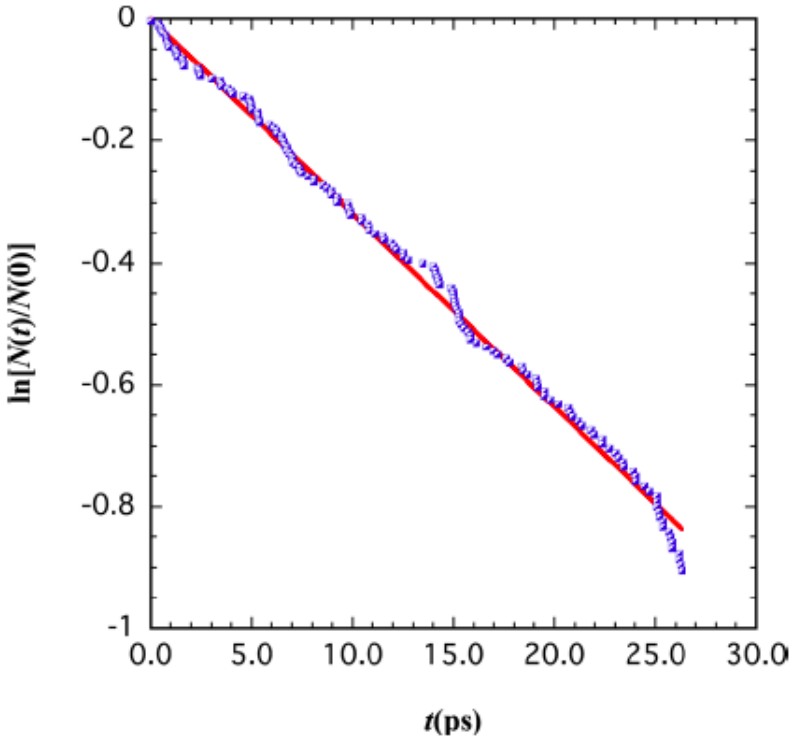


Figure 5.7: Plot of $ln[N(t)/N(0)]$ *vs* t for the 1500 K thermal simulation. $N(t)$ is the number of TIK$(H^+)_2$ non-reactive molecules at time t. The R value for the fit is -0.9867.

A plot of the logarithm of these rate constants *va* temperature is given in figure 5.8. The resulting Arrhenius parameters are are found from the intercept and the slope of the curve and they are: A $= (1.22 \pm 0.07) \times 10^{14} s^{-1}$ and $E_a = 24.3 \pm 0.2$ kcal/mol. From the plot of $\ln[N(t)/N(0)]$ *vs* time for excitation of the highest energy conformer in figure 5.2 at 2000 K we found a rate constant of $2.36 \pm 0.08 \times 10^{11} s^{-1}$ and the same as that above was found for excitation about the lowest-energy conformer ($2.35 \pm 0.02 \times 10^{11}$ s^{-1}).

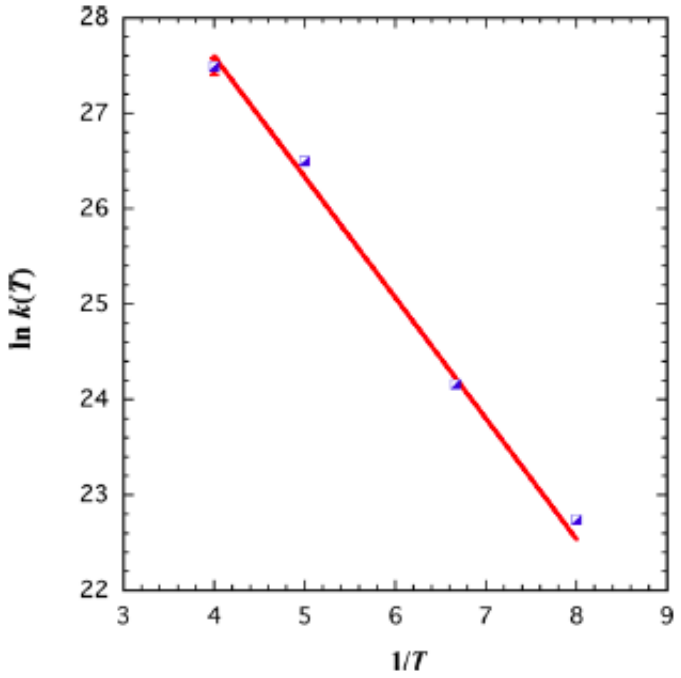


Figure 5.8: Natural logarithm of the overall rate constant in s^{-1} for TIK$(H^+)_2$ dissociation plotted *vs* $1/T$ (1×10^{-4} K^{-1}). The Arrhenius parameters are found from the intercept and the slope of the curve and they are respectively: the pre-exponential factor A $= 1.22 \pm 0.07 \times 10^{14}$ s^{-1} and $E_a = 24.3 \pm 0.2$ kcal/mol. The R value for the fit is -0.9961.

Table 5.2: Arrhenius parameters for TIK$(H^+)_2$ dissociation.

Pathway	A (s^{-1})	E_a (kcal/mol)
Overall	$1.22 \pm 0.07 \times 10^{14}$	24.2 ± 0.2
1	$2.80 \pm 0.32 \times 10^{12}$	14.7 ± 0.5
2	$2.15 \pm 0.33 \times 10^{12}$	18.6 ± 0.5
3	$2.03 \pm 0.75 \times 10^{14}$	36.5 ± 1.2
4	$4.73 \pm 1.17 \times 10^{14}$	40.1 ± 1.0

Thermal Unimolecular Rate Constants and Arrhenius Parameters for the Individual Pathways.

The temperature dependent rate constant $k(T)$ (as discussed in section 4.3.1 of previous chapter) is the sum of the rate constants for the individual dissociation pathways, i.e. $k = \sum_i k_i$. Thus, the individual $k_i(T)$ equals $k(T)$ multiplied by fraction of reactive trajectories dissociating through the i^{th} pathway p_i, that is:

$$k_i(T) = p_i k(T) \tag{5.1}$$

Overall, the rate constants for each pathway are well represented by the Arrhenius expression; however, there are some deviations from linearity.

Arrhenius parameters for pathways 1 to 4 are given in table 5.2.

A-factors and activation energies are smaller for pathways 1 and 2, which is consistent with the combined H atom transfer and bond rupture for these paths. For paths 3 and 4 there is only a single C-C bond rupture.

Comparison of Arrhenius and Quantum Chemistry Dissociation Barriers.

To compare with the activation energies determined from the Arrhenius plots, RM1 calculations were performed to determine dissociation barriers for pathways 1 and 2. The classical Arrhenius activation energy E_a for unimolecular dissociation equals E_0, the difference between the classical potential energies of the dissociation TS and the unimolecular reactant, plus the difference between the average thermal energies of the TS and reactant [110].

Pathways 1 and 2 are concerted reactions, consisting of simultaneous proton transfer and

C-C bond cleavage, that form singlet products. The potential-energy barrier for the pathway 1, found performing RM1 calculations, corresponds to 15.0 kcal/mol, which value agrees with the Arrhenius activation energy of 14.7 ± 0.5 kcal/mol of table 5.2. An agreement is found also or pathway 2, for which the RM1 TS energy is 183.1 kcal/mol and the Arrhenius activation energy is 185.9 ± 5.0 kcal/mol (table 5.2).

However, dissociation activation energies for pathways 3 and 4 are only in a weak qualitative agreement with the value found performing thermal dissociations.

5.2.3 Summary and conclusions

Classical direct chemical dynamics simulations were performed using the semi-empirical RM1 method RM1 for the potential energy surface of $TIK(H^+)_2$.

We performed ensembles of randomly excited trajectories at 1250, 1500, 2000, and 2500 K, corresponding to classical energies in the range between 425 and 850 kcal/mol.

1. For each temperature $TIK(H^+)_2$ dissociates exponentially *versus* time, in accord with the RRKM and TS theories; that is, $TIK(H^+)_2$ is an intrinsic RRKM molecule [110]. The unimolecular rate constants is obtained for each temperature for the total fragmentation and includes the complete anharmonicity of the RM1 [82] potential-energy surface. Fragmentation may involve either the backbone or side chains, and both occur by concerted reactions, with simultaneous proton transfer and bond rupture, and also homolytic C-C bond ruptures without proton transfer. Side-chain fragmentation becomes more important as the temperature is increased; that is, backbone fragmentation decreases from 100 to 63.7% as T is increased from 1250 to 2500 K.

2. The number of primary fragmentation pathways obtained from simulations are 3, 14, 32, and 6 for the temperatures of 1250, 1500, 2000, and 2500 K, respectively.

3. The rate constant k(T) is a sum of the rate constants $k_i(T)$ for the individual dissociation pathways, which is equal to k(T) multiplied by the probability of dissociation by pathway i.

4. The E_a values determined from the simulation Arrhenius plots are in very good agreement with the reaction barriers for the RM1 method used for the simulations. This and additional studies [26, 52] illustrate that chemical dynamics simulations may be used to study thermochemical kinetics and obtain Arrhenius parameters.

These thermal simulations provide "benchmark" results for comparison with extensive simulations of TIK$(H^+)_2$ fragmentation by CID. As mentioned above, in thermal simulations the excitation energy is randomly and uniformly distributed within the ion, and the simulations provide information regarding its dissociation dynamics when it is excited randomly/statistically. In CID of peptide ions, the activating collision excites the ion non-randomly, and the ions' initial dissociation dynamics may reflect this non-random excitation. Only after longer times, when IVR is complete, can one be assured that the ion's dissociation is statistical. In comparing thermal and collisional excitations, the localized, non-random excitation by the collisions, as compared to the random thermal excitation, may be important.

5.3 Collisional simulations of CID of peptide ions: comparisons between TIK$(H^+)_2$ and TLK$(H^+)_2$ fragmentation dynamics, and with thermal simulations

As identified by related surface-induced dissociation (SID) studies [33, 111, 112], there are two limiting fragmentation atomistic dynamics for the ion excited by collision [33]. For one, there is translation-to-vibration energy transfer during the collision, followed by efficient and statistical IVR within the ion so that its fragmentation dynamics are accurately described by RRKM theory. For the other limit, the ion fragments, i.e., "shatters", upon collision with the bath gas. The orientation of the collision directs the translation to vibration energy transfer so that a fragmentation transition state (TS) is accessed during the collision. CID shattering fragmentation was observed in chemical dynamics simulations and experiments [27, 29, 30, 113, 15]. Direct chemical dynamics simulations [114] may be used to study both thermal, random, and collisional, non-random, excitations. We now report and discuss extensive collisional simulations done for TIK$(H^+)_2$ and TLK$(H^+)_2$. There are two principal reasons for this study. One is to compare the side chain fragmentations for I and L components of the ions. The other is to compare the fragmentation dynamics for TIK$(H^+)_2$ following non-random, collisional

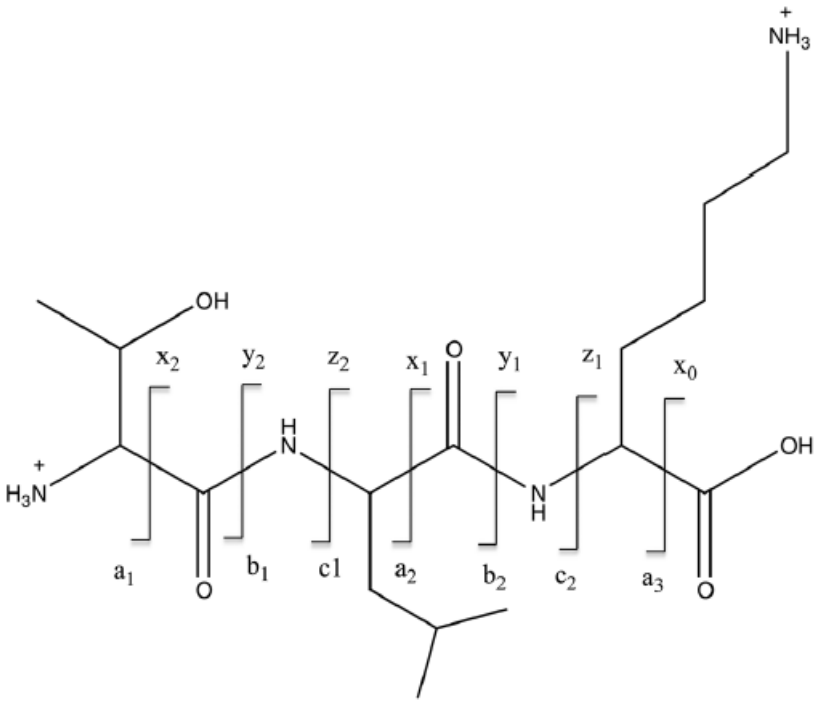


Figure 5.9: Primary structure of $TLK(H^+)_2$ and relative fragmentation nomenclature of Roepstorff and Fohlman [42].

(non-random) and thermal (random) excitations. The latter will identify the importance of non-statistical and non-RRKM unimolecular dynamics for $TIK(H^+)_2$ following collisional excitation, and presumably also for collisional excitation of $TLK(H^+)_2$ and other peptide ions.

The primary structure for $TIK(H^+)_2$ is given in figure 5.1 and for $TLK(H^+)_2$ in figure 5.9.

This section will be, thus, focused on four different aspects: (i) the comparison of $TIK(H^+)_2$ and $TLK(H^+)_2$ fragmentation dynamics using explicit collision simulations; (ii) how to distinguish between isoleucine and leucine side chains as determined by direct chemical dynamics; (iii) the different fragmentation behaviour of the unimolecular fragmentation of $TIK(H^+)_2$ as a function of the activation mode (collisions or thermal activation); (iv) the comparison of simulations results with an online experimental database.

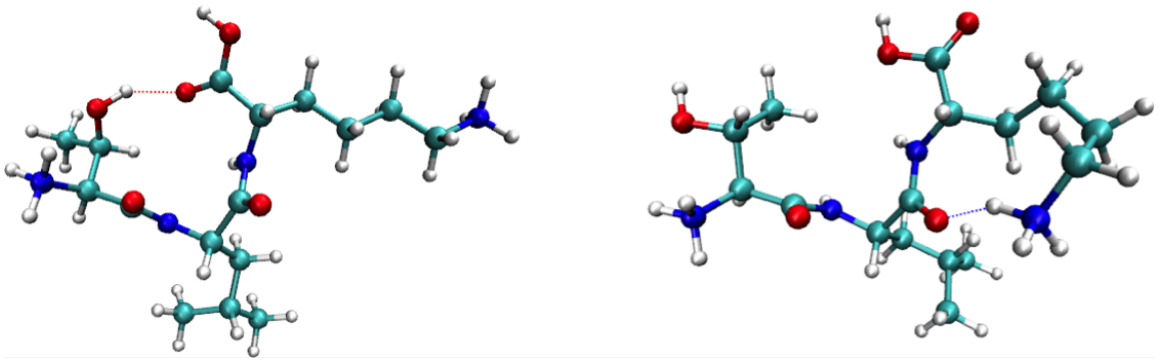


Figure 5.10: The two low-energy conformers of TLK$(H^+)_2$ determined by RM1 optimizations. The classical potential energy for the conformer on the left is 0.6 kcal/mol lower in energy than the one on the right.

5.3.1 Methodology

The way to perform collisional activation simulations was illustrated in section 3.3.3 of chapter 3. The RM1 semi-empirical Hamiltonian [82] was used for the previous study (section 5.2) of the thermal and collisional fragmentation of TIK$(H^+)_2$ and the same method is used for the simulations reported here for both the peptide ions.

In order to find low energy conformers structures for TLK$(H^+)_2$, we replaced I with L in the optimized TIK$(H^+)_2$ structures (see figure 5.2) and we optimized them again using RM1. The two lowest energy conformers found for TLK$(H^+)_2$ have similar RM1 energies and are shown in figure 5.10. For the previous TIK$(H^+)_2$ simulations (section 5.2) the lowest-energy conformer and the next lowest-energy conformer gave statistically the same fragmentation dynamics. For the current TLK$(H^+)_2$ collisional simulations, the two conformers in figure 5.10 also give statistically the same fragmentation dynamics. A major reason that the two TLK$(H^+)_2$ conformers give similar dynamics is that the small difference in their potential energy minima is overwhelmed by their zero-point energies (ZPEs) which is included in the trajectory simulations (see below). For the conformers on the left and right in figure 5.10, the harmonic ZPEs are 351.5 and 350.6 kcal/mol respectively.

The N_2 + TIK$(H^+)_2$ simulations were performed using as initial structure for TIK$(H^+)_2$ its global lowest energy conformer (see figure 5.2). 1000 trajectories were simulated at 10.8 and 30 eV (250 and 690 kcal/mol) of relative collision energies (E_{rel}). They can be compared

with the simulations performed for the previous work (section 5.2) at 13.0 eV (300 kcal/mol). The impact parameter was randomly chosen between 0 and b_{max}, with the b_{max} value set to 8.5 and 9 Å.

The N_2 + TLK$(H^+)_2$ simulations were performed for E_{rel} of 10.8, 13.0, and 26.0 eV. At E_{rel} = 13.0 eV, 2000 and 1000 trajectories were calculated for the lowest and higher energy TLK$(H^+)_2$ conformers in figure 5.10, respectively. 1000 trajectories were simulated at 10.8 and 26.0 eV simulations using TLK$(H^+)_2$ lowest energy conformer. For each set of simulations, i.e. for each E_{rel}, the impact parameter was chosen randomly between 0 and 9 Å. The trajectories were stopped when the $N2-$ion separation was 80, 250, and 250 Å for simulations at 10.8, 13.0, and 26.0/30.0 eV of E_{rel}. A connectivity matrix was used to identify reactive trajectories and the fragments. See *Appendix B* for details about the analysis code that was developed and used. Once the fragments are identified, the charge is then assigned to each fragment from a Mulliken population analysis based on RM1 calculations of the whole system. The analysis return partial charges on each atom. For each fragment the charge is then determined summing the charges of its atoms.

The maximum value of the impact parameter (b_{max}) was assigned as it was done for the collisional simulations of the di-proline anion [78] (see section 4.3 of chapter 4).

5.3.2 Results

In this section we will first report and discuss TIK$(H^+)_2$ and TLK$(H^+)_2$ collisional simulations, with particular emphasis on the effect of the collision energy and similarities and differences in the fragmentation pathways of the two ions.

Primary dissociation pathways

As shown in section 5.2, 41 primary dissociation pathways were found for the fragmentation of TIK$(H^+)_2$ and they were labelled as 1, 2, 3,.... We now use the same labels for the TLK$(H^+)_2$ pathways, except that they are labelled as 1′, 2′, 3′,.... Pathways 5 and 5′ observed in collisional simulations differ from the TIK$(H^+)_2$ thermal ones by the charge distribution of the products. In the thermal simulations the products are neutral radical threonine plus a doubly-

charge ion (m/z 158.5), while in the collisional simulations most of them are composed of two singly charged ions, protonated threonine (m/z 45) and the ion with m/z 317. Pathways 5 and 5′, corresponding to ions collisional simulations, are shown in figure 5.11. A slight difference is found between the two ions fragmentation for these pathways. In fact, for TIK$(H^+)_2$ the C-C bond breaking is always coupled with a proton transfer from the N-terminus group to the vicinal C=O (pathway 5). On the other hand, for TLK$(H^+)_2$ there are two other possibilities, represented by pathways 5'-2 and 5'-3: (i) the C-C bond breaks without any proton transfer; (ii) the bond cleavage is coupled with a proton transfer from the N-terminus group to the central C=O group respectively. This can be due to the TIK$(H^+)_2$ and TLK$(H^+)_2$ different conformations used in the simulations, which involve two different hydrogen bonds.

Primary dissociation pathways that are unique for the fragmentation of TLK$(H^+)_2$ are then shown in figure 5.12.

Probabilities of the different fragmentation pathways for TIK$(H^+)_2$ + N_2 and TLK$(H^+)_2$ + N_2 collision simulations can be found in tables A.1, A.2, A.3, A.4, A.5 and A.6 in *Appendix A*. They report the probabilities of the dissociation pathways with respect to the total number of dissociating trajectories, the probability for the dissociation to be shattering, and whether the fragmentation is backbone or side chain. In particular a trajectory has been identified as shattering if the dissociation was observed within 40 fs by the collision time. In particular, once modification of the connectivity matrix occurs providing two separated species, a fragmentation event is defined if the reactant is not reformed (see *Appendix B* for details about the analysis code). The fragmentation time is then obtained by calculating the time lapse between fragmentation and collision. The time of 30 fs was considered as the up boundary limit for shattering fragmentation in order to use the same criterion for all the bonds (C-H, N-H, C-C ...).

At the collision energy of 10.8 eV, 5.7 and 5.3% of the TIK$(H^+)_2$ and TLK$(H^+)_2$ ions trajectories fragmented, respectively; at 13.0 eV of E_{rel} the 8 and 9%, respectively. The number of primary pathways and final primary products for the two are also slightly different. Minor differences are seen for the more probable pathways, generally falling in the uncertainties. The highest energy simulations were performed at 30.0 and 26.0 eV of E_{rel} for the two TIK$(H^+)_2$ and TLK$(H^+)_2$ ions, respectively. At the higher collision energy we found 45 fragmentation pathways for TIK$(H^+)_2$ and 25 for TLK$(H^+)_2$. Furthermore, more than 90% of all reactive trajectories are shattering (95% for TIK$(H^+)_2$ and 92% for TLK$(H^+)_2$). Except for pathways 21 and 22 (and 21′ and 22′), that correspond to formation of H^+ and N_2H^+ due to H^+ losses from different sites and that are the by far the most abundant fragmentations, there are no

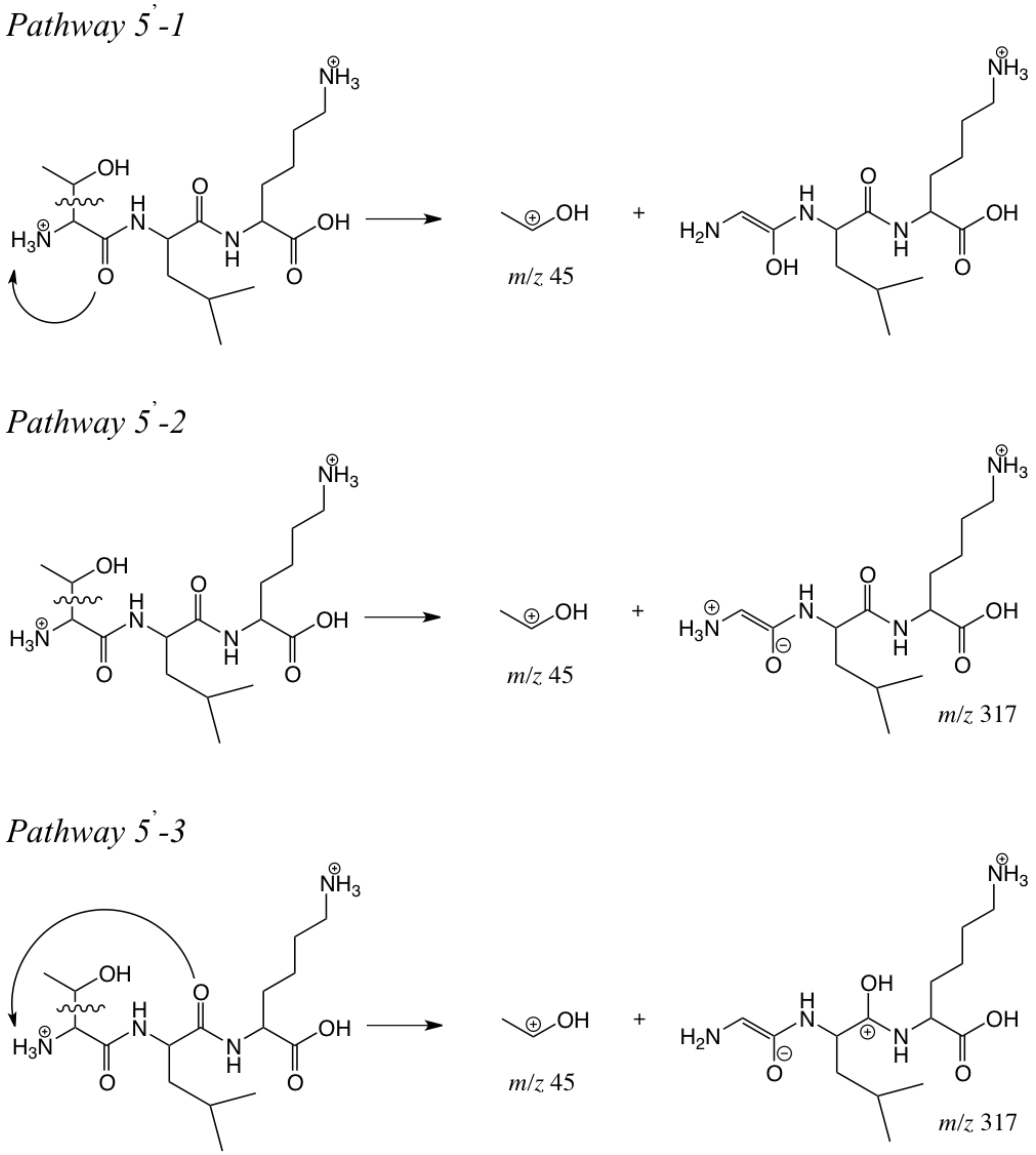


Figure 5.11: Fragmentation mechanisms corresponding to pathway 5 for collisional simulations of TLK$(H^+)_2$ primary dissociation.

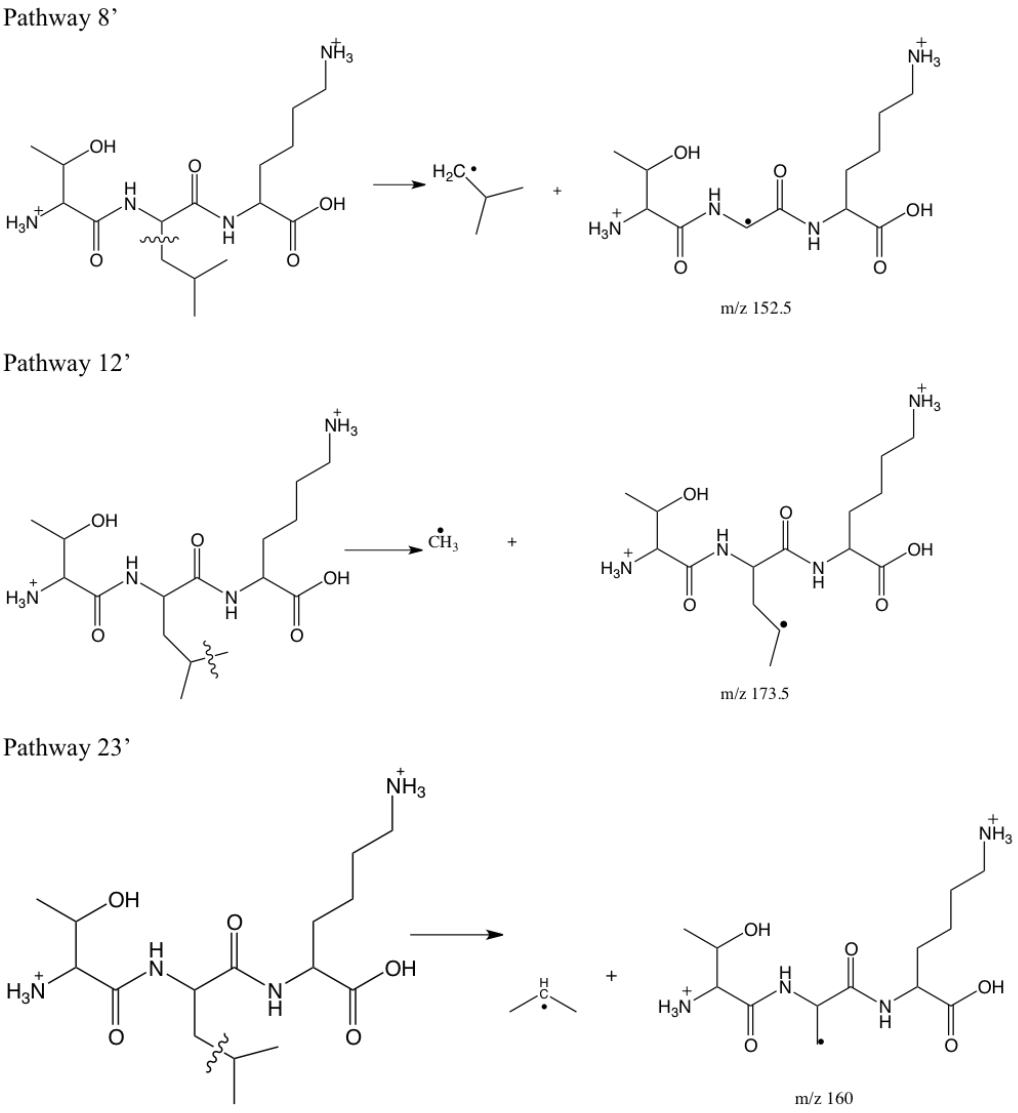


Figure 5.12: TLK$(H^+)_2$ primary dissociation pathways, which are different from the TIK$(H^+)_2$ ones.

Table 5.3: Percentages of backbone and side-chain fragmentation[a].

Collision energy (eV)	$TIK(H^+)_2$		$TLK(H^+)_2$	
10.8	54 ± 7;	28 ± 6	62 ± 7;	23 ± 6
13.0	51 ± 3;	33 ± 3	44 ± 3;	32 ± 3
30.20,26.0[b]	16 ± 3;	23 ± 3	14 ± 3;	13 ± 2

[a]The first percentage is for backbone and the second for side chain. Trajectories that fragmented forming H^+ and N_2H^+ are not categorized as backbone or side chain. The standard deviation of the percentage is also given. [b]The first collision energy is for $TIK(H^+)_2$ and the second for $TLK(H^+)_2$.

dominant fragmentation pathway, which seems reasonable considering the percentage of shattering at these high energies.

Table 5.3 shows the percentages of the fragmentation that is backbone and side chain. At the lower collision energies of 10.8 and 13.0 eV, backbone fragmentation is more important than side-chain for both the ions, while at 26 and 30 eV the two fragmentations have similar probabilities.

Table 5.4 shows the percentages of the fragmentation that is shattering. We can see that shattering is more important for $TIK(H^+)_2$ than for $TLK(H^+)_2$. However, for both ions shattering backbone fragmentation becomes more important as the collision energy increases, where it approaches 100% at the highest collision energies. In general, shattering is very important for the two ions side-chain fragmentations. This is expected, in fact the side chains are more exposed to the collision with N_2, which can enhance immediate fragmentation due to an energy localization.

In figure 5.13 an example is given of non-shattering fragmentation corresponding to pathway 1. The time-evolution of four characteristic inter-nuclear distances is shown in the figure. While the collision takes place within about 100 fs (and N_2 rapidly departs), the first dissociation is at about 2.4 ps, typical of a non-shattering mechanism. For this case, we observe first, at about 2.4 ps, a proton transfer from the N-terminus to the C=O group (distance R2 in figure 5.13) and then the cleavage of a C-C bond (distance R4) at about 2.8 ps.

In figure 5.14 an example of shattering trajectory is shown for pathway 3 of $TLK(H^+)_2$ at $E_{rel} = 13$ eV. The collision takes place in the first 200 fs, depicted by the black line indicating the inter-nuclear distance $R1$ of a N-atom of N_2 colliding with a C-atom of the ion. The R2 distance shows fast (shattering) fragmentation of the C-C bond, occurring within the first 40

Table 5.4: Percentage of shattering fragmentation[a].

Collision energy (eV)	Backbone		Side chain	
	$TIK(H^+)_2$	$TLK(H^+)_2$	$TIK(H^+)_2$	$TLK(H^+)_2$
10.8	37 ± 9	21 ± 7	53 ± 12	83 ± 12
13.0	41 ± 4	17 ± 4	88 ± 4	74 ± 5
30.20, 26.0[b]	78 ± 8	42 ± 10	95 ± 3	96 ± 4

[a]The percentage of the total fragmentation that is shattering. Trajectories that fragmented to the H^+ and N_2H are not categorized as backbone or side chain. The standard deviation of the percentage is also given. [b]The first collision energy is for $TIK(H^+)_2$ and the second for $TLK(H^+)_2$.

fs after the collision. Interestingly, after other about 2.6 ps from the shattering event, the R3 C-N inter-nuclear distance increases, showing further fragmentation. Thus for this last C-N bond dissociations occurs *via* intramolecular vibrational redistribution.

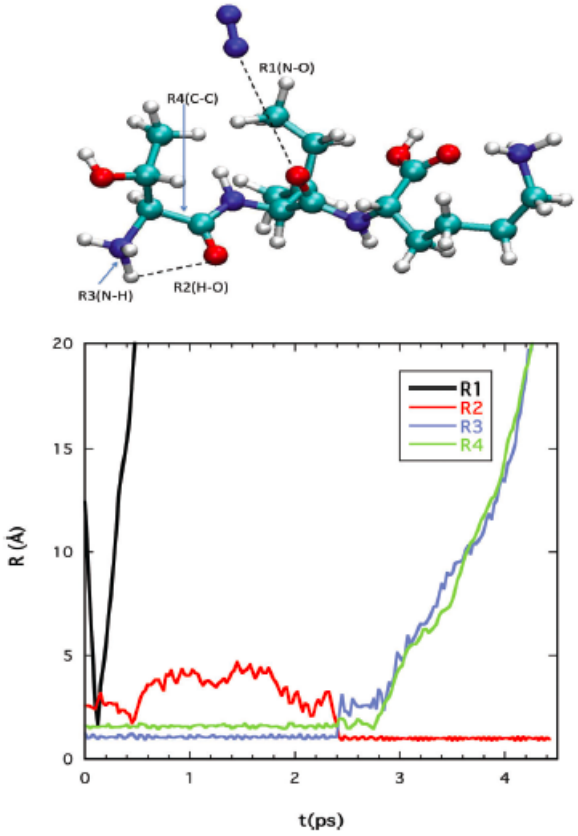


Figure 5.13: Variation of internuclear distances of a representative TIK$(H^+)_2$ CID trajectory following pathway 1 at E_{rel} = 13.0 eV. The geometry with distances defined is reported above the graph.

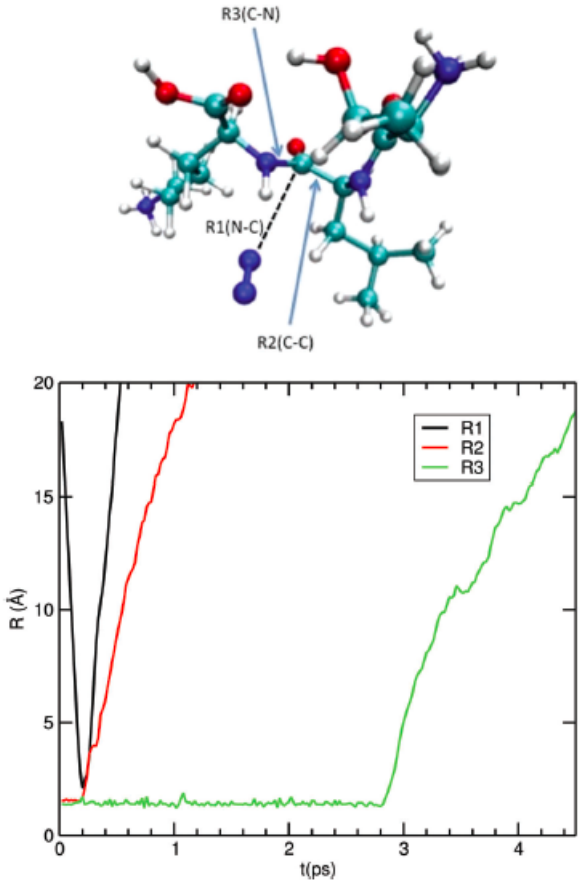


Figure 5.14: Variation of internuclear distances of a representative TLK$(H^+)_2$ CID trajectory following pathway 3′ at E_{rel} = 13.0 eV. The geometry with distances defined is reported above the graph.

Secondary dissociation pathways and comparison with a database

We report theoretical mass spectra that result by counting the occurrence of final products m/z values.

For the lower collision energies of 10.8 and 13.0 eV the relative theoretical mass spectra of TIK$(H^+)_2$ and TLK$(H^+)_2$ ions are shown in figure 5.15. For the high collision energy simulations, the number of fragments increases dramatically and the mass spectra are less informative, so they are not reported.

Looking at the mass spectra of figure 5.15, the total number of different ions formed is higher than what generally observed in a CID spectrum of a peptide. However, the number of fragments having relevant contribution to the final theoretical mass spectrum is much lower. In particular, at the lowest collision energy there are about 10 significant peaks. The primary peaks shown by the mass spectra of the two ions are the same. The most relevant are: m/z a_1^+ (from pathways 1 and 2), x_2^+ (from pathway 1), y_2^+ (from pathway 2), m/z 45 and 317 (from pathway 5, figure 5.11), m/z 158.5 (from pathway 6), and m/z 180.5 (corresponding to H^+ loss). The small differences are within statistical uncertainties. However, it should be noticed that most of the differences in the fragmentation of the two ions regard the low intensity peaks. The main effect of increasing the collision energy is to increase the intensity of less important peaks (considered as a statistical noise in mass spectrometry in analytical characterizations) and, consequently, decreasing the relative intensity of the characteristic peaks. The mass spectra show a limited number of relevant peaks as generally obtained in the CID of peptides [43].

Unfortunately, there are no CID experiments reported for the TIK$(H^+)_2$ and TLK$(H^+)_2$ ions, but is possible to make a qualitative comparison with information present in the literature. "Fragment ion calculators" can be considered, as done by popular codes like MASCOT [41]. In particular, we used one on-line [115] from which we found agreement for the backbone fragmentation. Unluckily, the approach used from this automatic code neglects side-chain fragments (it reports only possible $a/b/c/x/y/z$ fragmentation).

Another comparison can be done using the NIST database [116] for the fragmentations of separated amino acids. For threonine, peaks m/z 74 and 75 show good correspondence between the simulations and the NIST spectrum. For lysine, fragments m/z 31 and 73 are present in the NIST spectrum and are relatively abundant in the simulations. Finally, for side-chain loss from leucine and isoleucine corresponding to neutrals, there is no corresponding information available in the NIST database.

Anyway, our aim consists mainly in providing information for plausible fragmentation

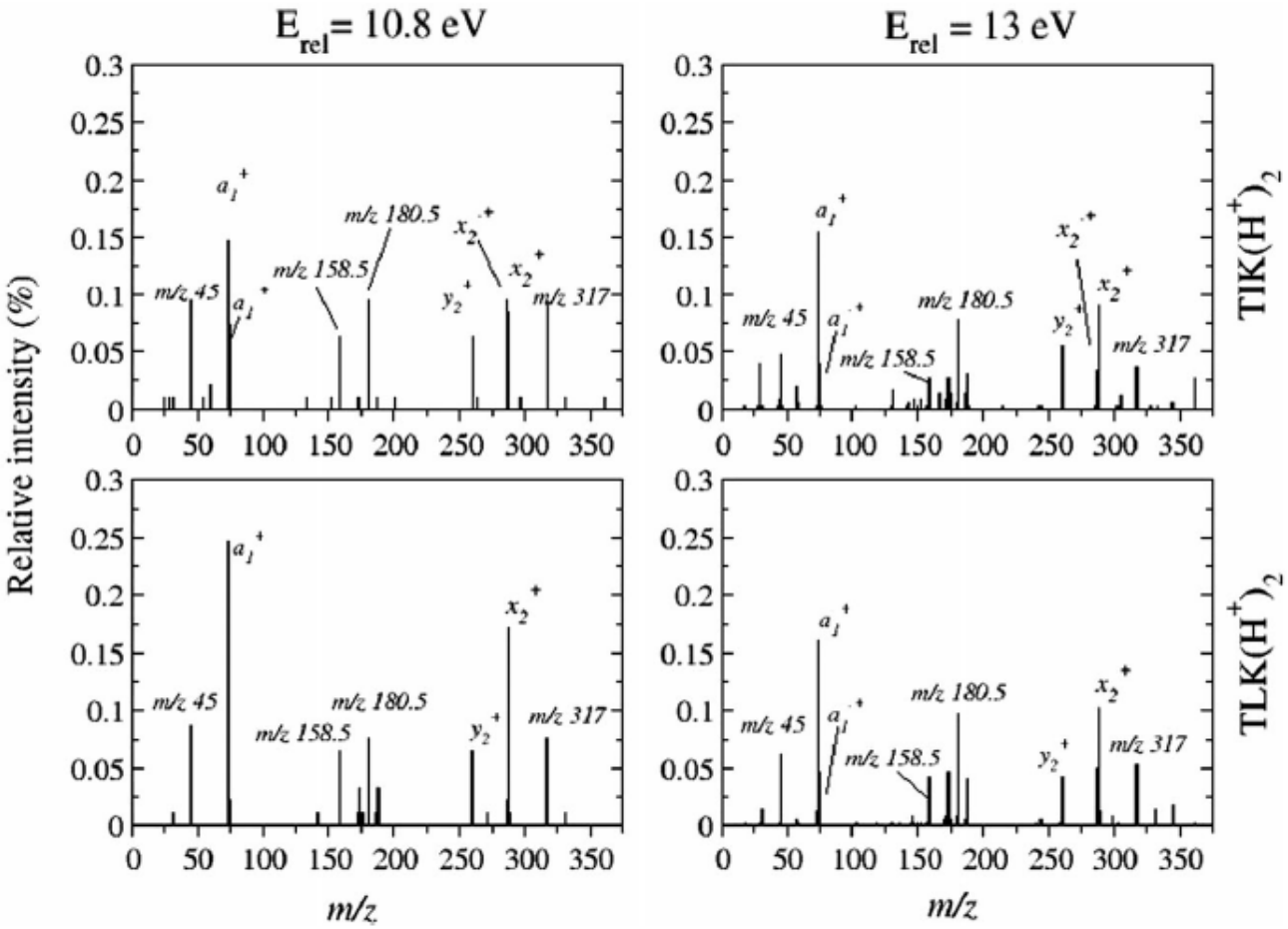


Figure 5.15: m/z mass spectra of TIK$(H^+)_2$ (upper) and TLK$(H^+)_2$ (lower) trajectories for CID at E_{rel} of 10.8 and 13.0 eV.

Table 5.5: Energy transfer properties.

	$TIK(H^+)_2$			$TLK(H^+)_2$		
E_{rel}	10.8	13.0	30.0	10.8	13.0	26.0
$\langle E \rangle^a_{react}$ (kcal/mol)	199 ± 32	236 ± 43	407 ± 131	168 ± 36	196 ± 42	327 ± 111
$\langle E \rangle^b_{non-react}$ (kcal/mol)	59 ± 61	54 ± 62	24 ± 59	34 ± 51	33 ± 56	27 ± 56
$E^0_{coll}{}^c$ (kcal/mol)	113.0	85.3	93.9	84.3	80.8	84.6
$P(E > E^0_{coll})^d$	18%	20%	9%	18%	17%	13%

$^a E_{react}$ is the average transferred energy for fragmenting trajectories. $^b E_{non-react}$ is the average transferred energy for non-reactive trajectories. $^c E^0_{coll}$ is the minimum energy transfer for fragmenting trajectories. $^d P(E>E^0_{coll})$ is the percentage of non-reactive trajectories with transferred energy larger than E^0_{coll}.

patterns with relative mechanisms, and understanding the differences between two limiting thermal and collisional activation modes.

Energy transfer

Collisional simulations provide useful information not only in terms of fragmentation products and mechanisms but also in the energy transferred to the ions due to the collision. The distributions of energy transfer to $TIK(H^+)_2$ and $TLK(H^+)_2$ for both the ions are reported in figure 5.16 for both the reactive trajectories (i.e. that fragmented in the simulation time) and the non-reactive ones.

The average energy transfer properties are summarized in table 5.5. We note that, as expected, the energy transfer to reactive trajectories is larger than that for non-reactive ones, and it is always more than 50% of the collision energy.

For each collision energy and ion, there is a minimum energy transfer for which fragmentation occurred. The minimum energy transfer for $TIK(H^+)_2$ simulations from the lowest to the higher collision energy is 113.0, 85.3 and 93.9 kcal/mol, respectively. For the $TLK(H^+)_2$ it is 84.3, 80.8, and 84.6 kcal/mol, respectively. These energies may be viewed as dynamical fragmentation thresholds, E^0_{coll}, which depend on the simulation time scale. If the trajectories were integrated for longer periods of time, fragmentations at lower energies are expected,

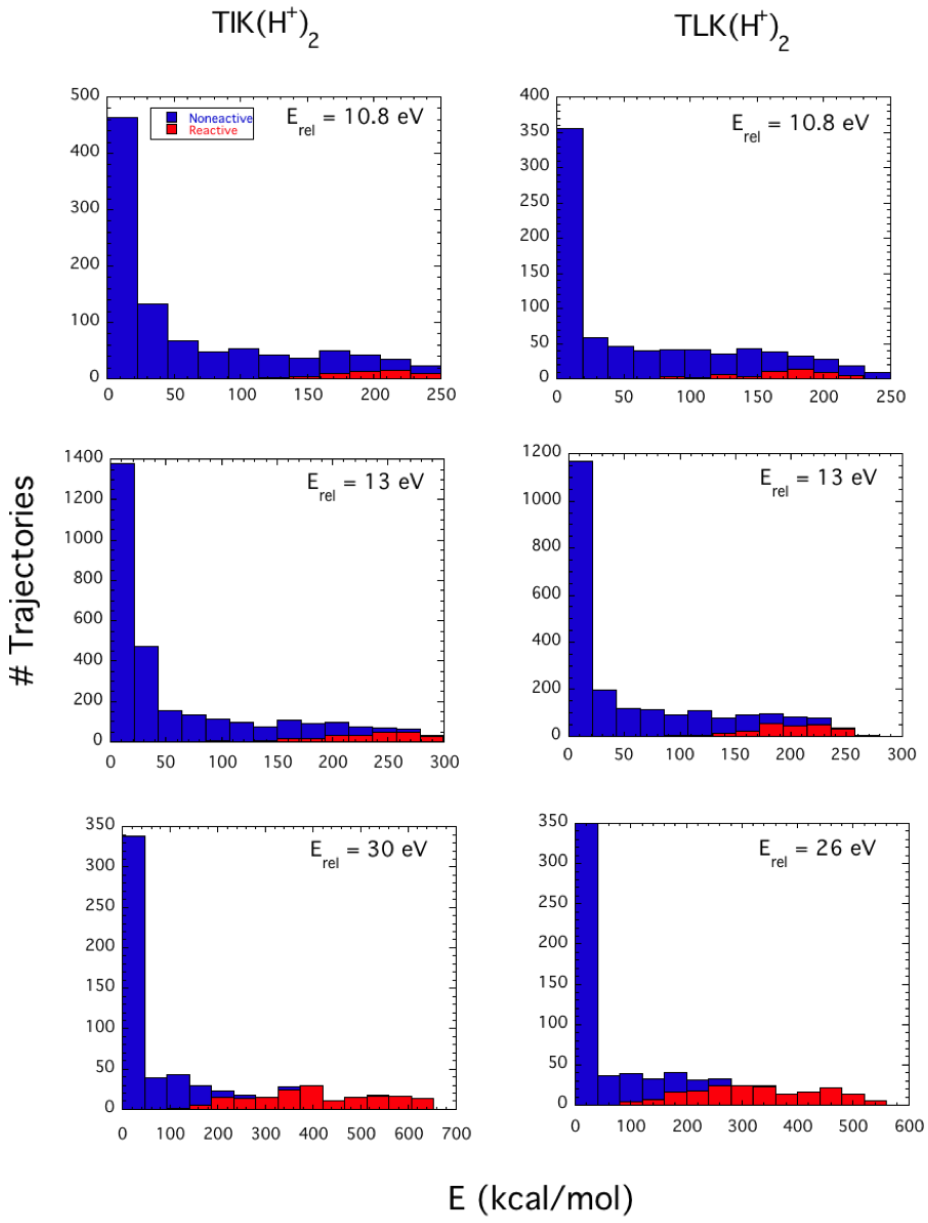


Figure 5.16: Energy transfer distributions for $TIK(H^+)_2$ and $TLK(H^+)_2$ fragmenting and non-fragmenting trajectories at different collision energies.

resulting in lower E^0_{coll}. For the current simulations the fraction of non-reactive trajectories for which the energy transfer is greater than E^0_{coll} is given in table 5.5 for both ions and each collision energy. The non-reactive trajectories may have sufficient time to complete IVR [117] and would then dissociate in accord with RRKM theory [118]. In future work, one could then use this theory to calculate the unimolecular rate constant k of the ions, and resulting lifetime $t = 1/k$, versus energy. The fraction of ions which have lifetimes shorter than the mass spectrometry experimental time scale could then be determined. A challenging aspect would be to include an accurate anharmonic correction to the unimolecular rate constant, since anharmonicity may be quite important for the ion's RRKM rate constant.

Impact parameter distribution

We analysed the distribution of shattering as a function of the impact parameter (that was chosen randomly between 0 and b_{max} in all the simulations) for the TIK$(H^+)_2$ simulation at E_{rel} = 13.0 eV. This distribution, and the ones for backbone shattering and side-chain shattering trajectories, as a function of b, are shown in figure 5.17.

We can see that the probability of shattering versus b is nearly proportional to b up to b = 3 Å. The probability of shattering decreases for large b. The maximum value of b at which side-chain fragmentation is observed is 8 Å, while 6.5 Å for backbone shattering. This shows that it is important to use an high value of b_{max} in order to consider all the possible fragmentations. However, a too high value of b_{max} chosen for the simulations would increase the computational cost and most of the trajectories would be, anyway, non-reactive.

The average impact parameters for total, backbone and side-chain shattering are found to be 3.2, 2.5, and 3.5 Å, respectively.

Differentiating CID of TIK$(H^+)_2$ and TLK$(H^+)_2$

TIK$(H^+)_2$ and TLK$(H^+)_2$ have the same mass, so that it is important to understand how they can be differentiated in their CID mass spectra. More in general, it is important to understand how to differentiate isoleucine (I) and leucine (L) in peptides or proteins.

In figure 5.18 the possible pathways involving I and L side-chain fragmentations are shown, consisting in the whole I/L loss, or in the cleavage of one C-C bond in the I/L side chain.

Pathways in panels A and D form ions in commons between TIK$(H^+)_2$ and TLK$(H^+)_2$

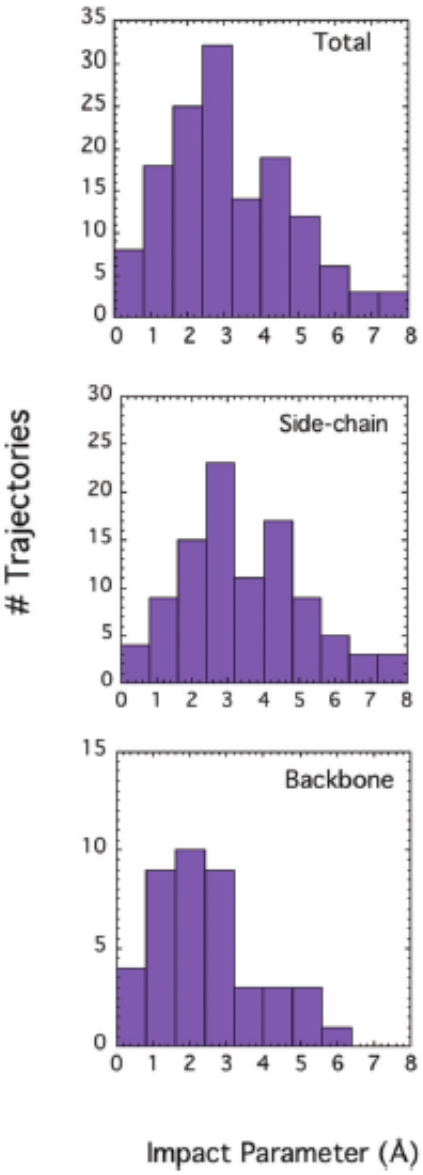


Figure 5.17: Impact parameter distributions leading to TIK$(H^+)_2$ fragmentation at E_{rel} = 13.0 eV for total, backbone and side-chain shattering trajectories.

(m/z 152.5 and 173.5), while pathways B and C form unique ions with m/z 159.5 for L and m/z 166.5 for I. The occurrence of these ions for TIK$(H^+)_2$ and TLK$(H^+)_2$ is then reported in figure 5.19 at different collision energies. Unfortunately, their probabilities are low, the highest

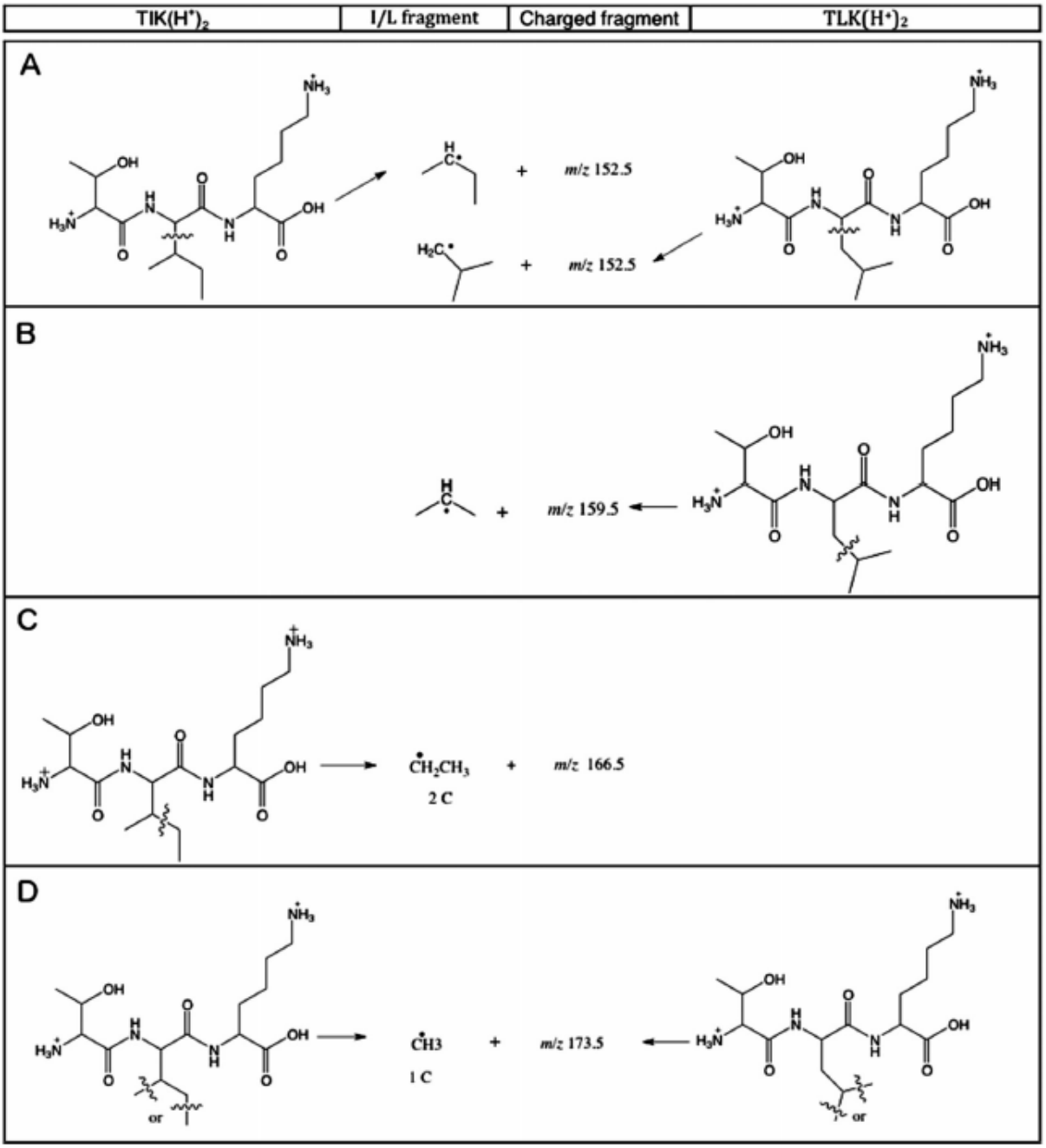


Figure 5.18: Fragmentation pathways that can differentiate between I and L.

being only 4%.

The common ion m/z 152.5 is considerably more probable for TIK$(H^+)_2$ than for TLK$(H^+)_2$.

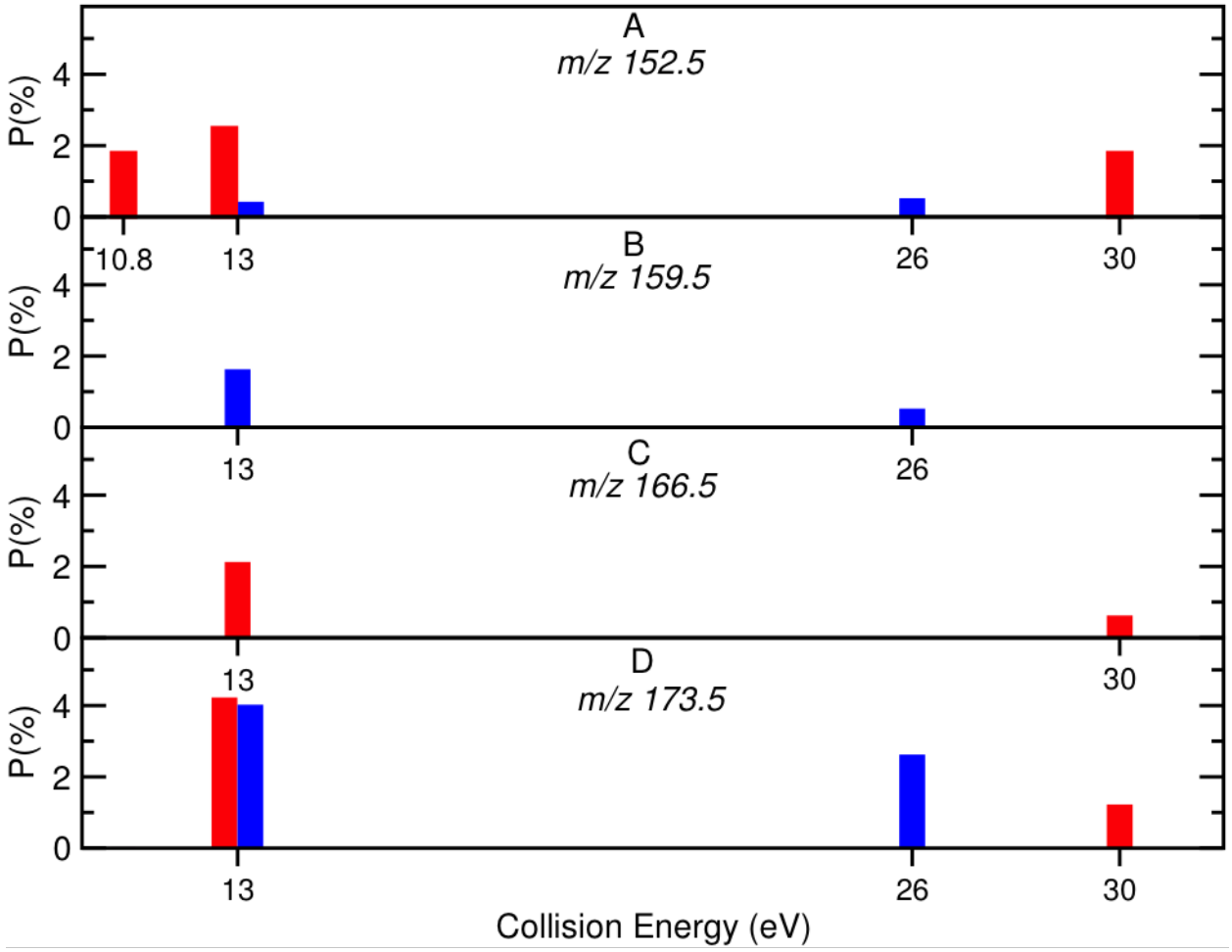


Figure 5.19: Percentages for forming product ions, with respect to total fragmentation, for pathways A-D in fgure 5.18. Percentages for forming the TIK$(H^+)_2$ and TLK$(H^+)_2$ fragmentation ions are given by the red and blue bars, respectively. Fragmentation percentages are given for 10.8, 13.0, 26.0, and 30.0 eV.

The m/z 173.5 ion has almost the same abundance in both systems and, of the four graphs in figure 5.19, has the largest probability of about 4% for E_{rel} = 13.0 eV. The two ions which are characteristic of the two isomers, m/z 159.5 for TLK$(H^+)_2$ and 166.5 for TIK$(H^+)_2$, are obtained with low probability, the highest values at E_{rel} = 13.0 eV. An interesting consideration

can be made inspecting the theoretical mass spectra of the two ions in figure 5.15. The m/z 159.5-170 region is empty, thus making m/z 166.5 (which is characteristic of I) a peak which might be used to distinguish I from L. On the contrary, near the m/z 159.5 peak (which is characteristic of L) there is another peak at m/z 159 and also an intense peak at 158.5, making m/z 159.5 difficult to be detected. Even using high-resolution mass spectrometry (HRMS) analysers, which should actually be able to resolve peaks separated by 0.5 m/z units, the detection would be still problematic. In fact, HRMS instruments have generally a low-resolution mass filter. In other words, it is well known that detection problems in low-resolution can become a transmission limitation in HRMS (e.g. "trap overfilling").

However, in principle it is possible for experimental work to focus on this region of the spectrum. Unluckily, the intensity of m/z 166.5 is very low, such that experimental spectra with very low noise will be necessary.

Comparison of thermal and collisional fragmentation of TIK$(H^+)_2$ and TLK$(H^+)_2$

For the thermal simulations on TIK$(H^+)_2$ and TLK$(H^+)_2$, reported in the previous section 5.2, it was found a statistical unimolecular behaviour. On the other hand, when performing collision activation the energy is not deposited randomly amongst the vibrational degrees of freedom of the ions and the fragmentation can occur before a complete IVR. However, as it was discussed previously (see section 4.3 in chapter 4), both statistical and non-statistical unimolecular mechanisms are responsible for the peaks in a CID experimental mass spectrum. It is for this reason that it is important to compare fragmentation pathways and product ions as obtained from different excitation methods.

We should note that in thermal excitation all fragmentations will be by definition non-shattering, while in collisional excitation both shattering and non-shattering fragmentation may occur.

We can start comparing the theoretical mass spectra resulting by applying one or the other activation mode, i.e. the spectra from thermal simulations in figure 5.3 and the spectra from the collision ones in figure 5.15. In the low temperature thermal spectra there are few peaks, while in the collisional simulations there are always many peaks. However, at higher temperature the number of peaks increases, so that it is possible to see as many peaks as from collision simulation. Note that the highest temperature simulation at 2500 K corresponds to an internal energy of 850 kcal/mol, and for the collisional simulation with E_{rel} = 13 eV we have an average ion energy upon collision of about 780 kcal/mol, resulting from the ZPE and average energy transfer.

The thermal and collisional fragmentation dynamics may be compared for the most important peaks in the spectra, i.e. m/z 45, 74, 158.5, 174, 180.5, 260, 288 and 317. It is instructive to underline the role of shattering fragmentation in this comparison. The probabilities for forming these ions are summarized in tables 5.6 and 5.7 as a function of collision energy and temperature for the collisional and thermal simulations, respectively.

The two most abundant ions in the thermal simulations, m/z 74 and 288, correspond to the products formed through non-shattering mechanisms in the collisional simulations. They correspond to a_1^+ and x_2^+ ions formed via a proton transfer/backbone breaking mechanism (see pathways 1 and 2 in figure 5.4 of previous chapter). Upon increasing the temperature their relative abundances decrease, mainly because other fragmentation channels become important.

The ions m/z 174 and 260 are also obtained in the collisional simulations, mainly through non-shattering mechanisms, and they are also observed in the thermal simulations. The ion

Table 5.6: Shattering *vs* non-shattering % are reported for collisional trajectories. We show the TIK$(H^+)_2$ results and in parenthesis the corresponding TLK$(H^+)_2$ ones.

	E_{rel} = 10.8 eV		E_{rel} = 13 eV	
m/z	shatt.	non-shatt.	shatt.	non-shatt.
45	44 (62.5)	55.6 (37.5)	92.0 (100)	8.0 (0.0)
74	0.0 (8.7)	100.0 (91.3)	1.7 (3.1)	98.3 (96.9)
174	0.0 (33.3)	100.0 (66.7)	0.0 (11.8)	100.0 (88.2)
158.5	100.0 (80.0)	0.0 (20.0)	0.0 (25.0)	100.0 (75.0)
180.5	100.0 (100.0)	0.0 (0.0)	100.0 (100.0)	0.0 (0.0)
260	0.0 (0.0)	100.0 (100.0)	0.0 (5.9)	100.0 (94.1)
288	0.0 (6.2)	100.0 (93.8)	2.50 (5.0)	97.5 (95.0)
317	100.0 (71.4)	0.0 (28.6)	100.0 (100.0)	0.0 (0.0)

Table 5.7: For thermal simulations we report the % of different ions over the reactive trajectories.

m/z	1250 K	1500 K	2000 K	2500 K
45	-	-	1.7	2.3
74	49.0	46.0	27.0	11.0
174	1.2	1.6	5.4	8.4
158.5	-	-	-	5.8
180.5	-	-	-	-
260	6.0	9.0	4.3	2.3
288	42.0	35.0	15.0	5.5
317	-	-	-	-

m/z 260 shows good intensity at low temperatures, becoming much less abundant with increasing temperature, while m/z 174 is abundant only at high temperatures. Both m/z 174 and 260 correspond to backbone fragmentation. Backbone fragmentations are often obtained through the non-shattering mechanism for two main reasons: (i) the backbone is protected by side chains, such that direct exposure to the projectile (which favors shattering) is reduced; and (ii) non-shattering is associated with proton transfer which trigers backbone fragmentation, as in the mobile proton model [48, 119, 120, 121].

On the other hand, ions m/z 45, 180.5 and 317 are obtained mainly via shattering mechanisms in the collisional simulations, while in the thermal simulations m/z 180.5 and 317 are not observed and m/z 45 is only observed at high temperatures. The formation of m/z 45 and 317 corresponds to the loss of threonine (see figure 5.11), while ion m/z 180.5 corresponds to the loss of H^+.

The ion m/z 158.5 merits individual discussion. It is obtained by shattering at low collisional energies, while from non-shattering only upon increasing the collision energy. It is also obtained from thermal simulations (by definition via non-shattering mechanism) at high temperatures, for which the internal energy of the ions is comparable to high collision energies. However, there is an important difference in the nature of these ions between the collisional and thermal simulations. In the collisional simulations it corresponds to the $a_3^{2+}\cdot$ radical ion, i.e. loss of the terminal COOH group (pathway 6 in figure 5.5). In the thermal simulations it corresponds to the formation of not only the a_3^{2+} ion, but also of neutral threonine thus forming a doubly charged ion (see pathway 5 in figure 5.5). In fact, when breaking the C-C bond connecting threonine to the backbone, there are two different behaviors for the thermal and collisional simulations. In the thermal simulations, the loss of neutral threonine side-chain and the formation of the doubly charged ion m/z 158.5 are found, while in the collisional simulations the two singly charged ions m/z 45 and 317 are obtained (see figure 5.11).

In general, non-shattering products are similar in the thermal and collisional simulations, in particular when energies are comparable as for the highest temperature in the thermal simulations, while the collisional simulations often give unique shattering products. Some products are obtained only through shattering and, thus, they will only be seen with a sufficiently high collision energy and localized non-random excitation, which is possible in some triple-quadrupole experimental set-ups [2]. In particular, side-chain fragmentation and fragmentations involving terminal groups (like the C-terminus COOH or H^+ loss) are peculiar for the collisional simulations and are mainly obtained via shattering mechanisms.

5.3.3 Conclusions

In the simulations reported here the two ions TIK$(H^+)_2$ and TLK$(H^+)_2$, were activated by collisions with N_2 at different collision energies. The two systems only differ in the central residue, isoleucine (I) or leucine (L). Collisional simulations were also compared with previous simulations presented in section 5.2, in which TIK$(H^+)_2$ fragmentation was studied by means of thermal activation. We can conclude that:

1. The TIK$(H^+)_2$ and TLK$(H^+)_2$ fragmentations are very similar, with few differences which are within statistical uncertainties and concern low abundant peaks.

2. Shattering fragmentation is important for collision simulations, and it mainly corresponds to side-chain fragmentation;

3. A detailed study of differences in TIK$(H^+)_2$ and TLK$(H^+)_2$ side-chain fragmentations, which could distinguish I from L, found that a peak at m/z 166.5 would be characteristic of I. This peak originates from a shattering mechanism. Unfortunately, the abundance is quite low and thus, to be used experimentally as a detection peak, the experimental noise should be drastically reduced;

4. Collisional energy transfer was analysed for the trajectories and the minimum energy transfer leading to fragmentation, identified as E^0_{coll}, was determined for each ion and collision energy. This is a dynamical fragmentation threshold, which will be smaller if the trajectories were integrated for a longer period of time. For the current simulation 10-20% of the non-reactive trajectories had collision energy transfers higher than E^0_{coll}.

5. The thermal and collisional simulations show similar product distributions when high temperatures (i.e. high internal energies) are considered. However, some peaks and mechanisms are different between the two activation modes. In fact in collisional dynamics there is both statistical and non-statistical dissociation, while in thermal activation only statistical products are observed.

In summary, this is the first study which considers in detail, by direct dynamics unimolecular, the fragmentation of a relatively large doubly charged peptide (three residues with about 60 atoms). Two activation modes, random/thermal and non-random/collisional are compared,

which are two limiting conditions for CID. The possibility of treating large systems with direct dynamics, for both thermal and collisional activation, paves the way to a more quantitative theoretical mass spectrometry, in particular because they can investigate different energy ranges and distinguish differences between random and non-random excitation.

Finally, for the study of such large systems, it has been crucial the development and use of a semi-automatic analysis code, which is reported in *Appendix B*.

5.4 Threshold for Shattering Fragmentation in Collision-Induced Dissociation of the Doubly Protonated Tripeptide $\mathbf{TIK(H^+)_2}$

In previous section, we have shown that a non-statistical fragmentation, shattering, is important in the collisional simulations of the $TIK(H^+)_2$ and $TLK(H^+)_2$ doubly protonated tripeptide ions (see section 4.2 of chapter 4 for a more extensive description of shattering fragmentation). In order to model and interpret experimental CID spectra of peptides, it is important to know the collision energy threshold for shattering. For example, is shattering important at low collision energies at or near thresholds for peptide ion fragmentation ? In particular, shattering is unique (by definition) for collision simulations, where translation to vibration energy transfer during the collision can directly access the fragmentation transition state (TS).

For the simulations presented here, shattering thresholds are investigated for both backbone and side-chain fragmentations of the doubly protonated tripeptide threonine-isoleucine-lysine ion, $TIK(H^+)_2$. We have thus investigated in deeper details the threshold energy needed to activate shattering fragmentation on this system. We have previously studied the same system via statistical fragmentation (see section 5.2), such that it will be possible to compare shattering *vs* RRKM energy threshold. Atomistic details of the fragmentation mechanisms are discussed in the previous sections 5.2 and 5.3 and here we refer to them, while, by decreasing the collision energy, here we focus on how decreasing the collision energy the amount of shattering evolves, thus making possible to identify an energy threshold for this particular way of fragmentation.

The same methodology described previously for $TIK(H^+)_2$ and $TLK(H^+)_2 + N_2$ collisional direct dynamics (see section 5.3) was used here. To establish an impact parameter b for the

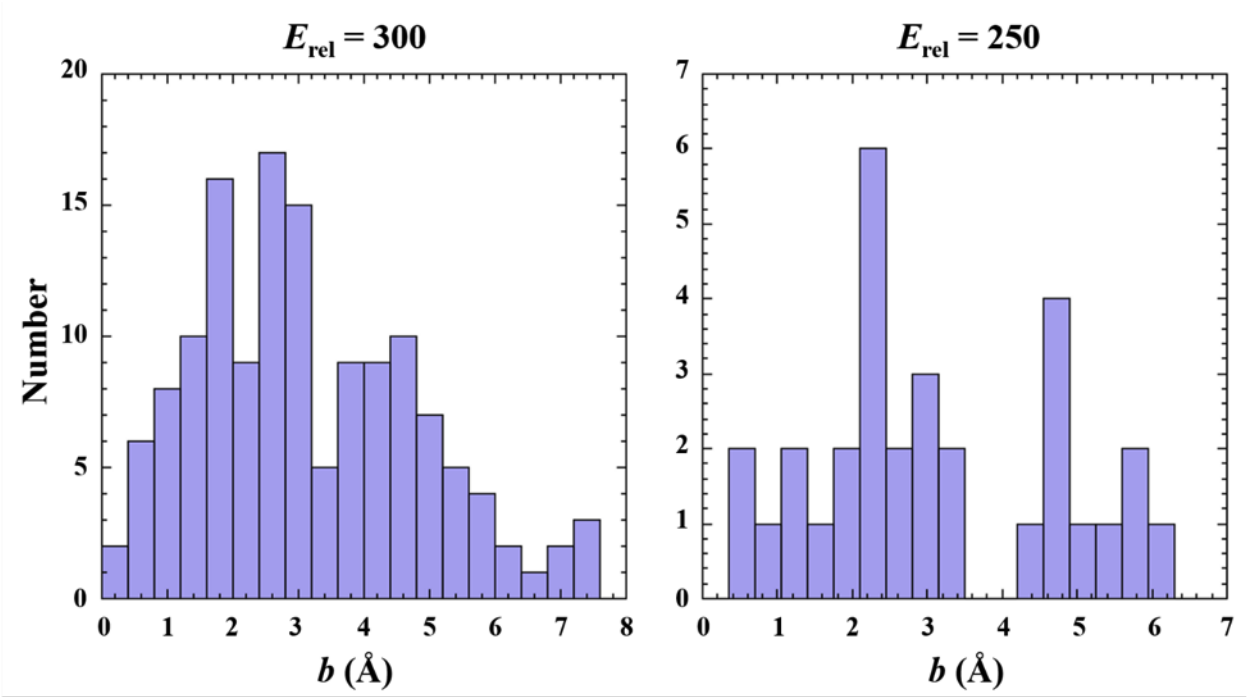


Figure 5.20: Number of shattering trajectories versus impact parameter for N_2 + TIK$(H^+)_2$ collisions at E_{rel} of 300 and 250 kcal/mol. The impact parameter is chosen randomly between 0 and 8.5 Å.

current simulations, we used the distributions of impact parameters leading to shattering computed in the previous TIK$(H^+)_2$ CID simulations at collision energies of 250 and 300 kcal/mol. In figure 5.20 we can thus see that the probability of shattering versus b is approximately proportional to the probability of a collision versus b for a value for b up to about 2.5 Å. Following this result, a fixed value for b of 2.5 Å was chosen for the simulations reported here in order to enhance the probability of shattering.

The N_2 + TIK$(H^+)_2$ simulations were performed for collision energies E_{rel} of 100, 150, 175, 200, and 225 kcal/mol. These energies were chosen in order to slowly decrease from the original 250 kcal/mol value studied in the previous collision simulations study (section 5.3) up to a lower value for which no (or few) shattering (and more in general reactive) trajectories are found. The respective number of trajectories calculated for these energies were 10338, 7491, 3259, 1928, and 1237. The total number of trajectory for each value of E_{rel} was chosen in

Table 5.8: Results from N_2 + TIK$(H^+)_2$ collision simulations at different collision energies (E_{rel})

E_{rel}	100	150	175	200	225
Number of trajectories	10338	7491	3259	1928	1237
% Fragmentation	0.22 ± 0.05	1.3 ± 0.1	3.0 ± 0.3	5.7 ± 0.5	9.5 ± 0.8
% Shattering[a]	17 ± 8	29 ± 5	31 ± 5	29 ± 4 39 ± 5	
% Shattering[b]	0.04 ± 0.02	0.39 ± 0.07	0.95 ± 0.17	1.7 ± 0.3	3.7 ± 0.5

[a]Percentage of shattering over the number of fragmentation trajectories. [b]Percentage of shattering over the total number of trajectories.

order to have at least about 100 reactive trajectories, limiting the maximum number of total trajectories for each set to about 10000 (this last condition was applied only to the trajectories with the lowest E_{rel} value). Each trajectory was integrated for 2.5 ps.

Four different types of fragmentation mechanisms were observed in the current simulations; i.e. non-shattering fragmentation, backbone and side-chain shattering, and shattering forming an H^+ or NH_2^+ ion. In table 5.8 we report three percentages which characterize the reactivity as a function of collision energy (E_{rel}): (i) the percentage of the trajectories which fragmented, (ii) the percentage of shattering trajectories over the number of fragmentation trajectories, and (iii) the percentage of shattering trajectories over the total number of trajectories. There is a substantial decrease in the percentage of the trajectories which fragmented, with decrease in E_{rel}, from 9.5% to 0.22%. Amongst the trajectories which fragmented, the percentage which fragmented by shattering decreased by approximately a factor of 2 with decrease in E_{rel} from 225 to 100 kcal/mol, i.e. from 9.3% to 17.4%. Since the number of shattering trajectories does not change with increase in integration time, but the number of fragmenting trajectories does, these percentages decrease with integration time. The percentage which does not depend on integration time is the percentage of shattering trajectories with respect to the total number of trajectories, which is given in table 5.8 and plotted in figure 5.21. This percentage decreases from 3.7% to 0.04% respectively from 225 to 100 kcal/mol.

The percentages of the shattering fragmentations which are side chain, backbone, or formation of an H^+ or N_2H^+ ions are given by the bar graph in figure 5.22. For each E_{rel}, the sum of the percentages is 100%. In the case of backbone shattering, they come mainly from pathways 4 and 6 (see figure 5.4, 5.5 and 5.5 of section 5.2) which correspond to loss

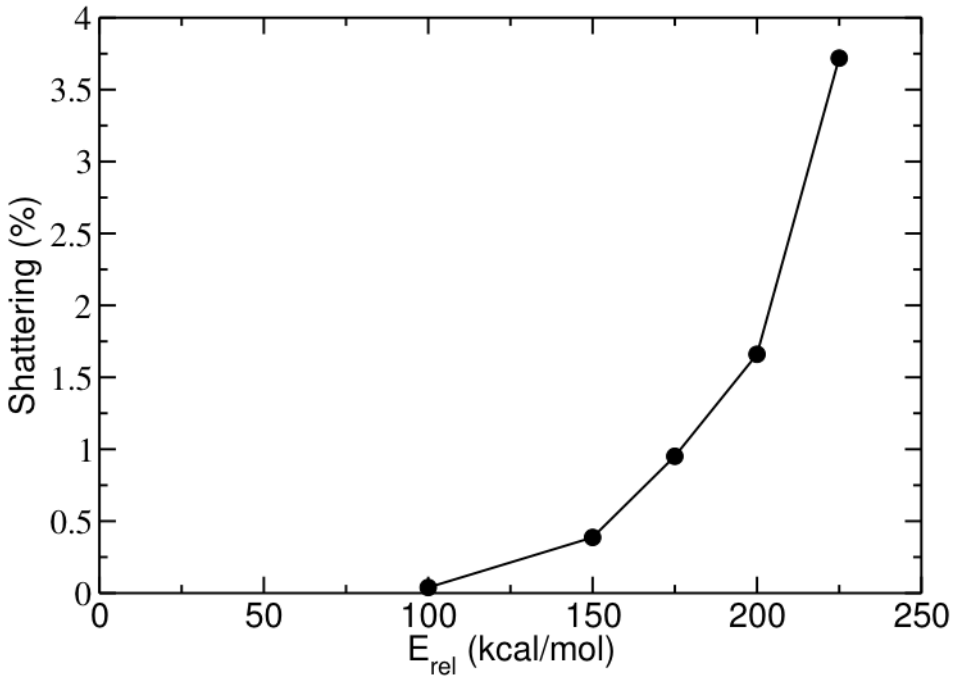


Figure 5.21: Percentage of trajectories which are shattering, with respect to the total number of trajectories, versus N_2 + $TIH(H^+)_2$ relative translational energy E_{rel}. The impact parameter is 2.5 Å.

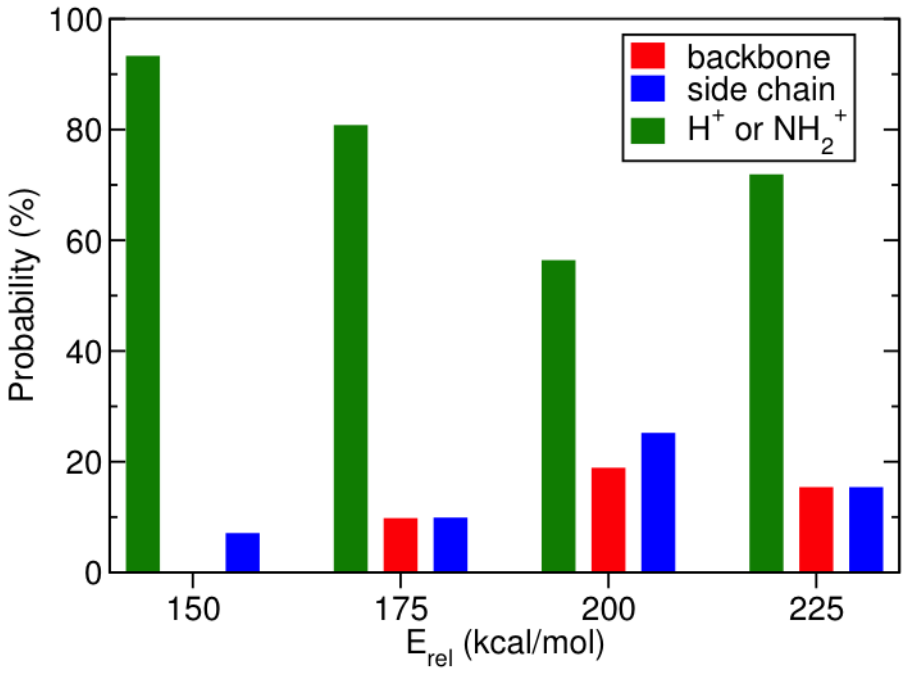


Figure 5.22: Bar graph of the percentages of the shattering fragmentation which are side-chain (blue), backbone (red), and formation of an H^+ or N_2H^+ (green). Total percentage for each energy is 100%. The impact parameter is 2.5 Å.

of N-terminus and C-terminus groups, forming ions a_1^+ and a_3^{2+}, respectively. The side-chain shattering fragmentations correspond mainly to the pathway 5 (loss of threonine side chain), pathway 8 (loss of isoleucine side chain) and pathway 9 (loss of terminal part, $CH_2NH_3^+$, of lysine side chain). At 175, 200, and 225 kcal/mol, the percentages for side-chain and backbone shattering are similar. However, at 150 kcal/mol only side-chain shattering is observed and its percentage is much smaller than that for H^+ or N_2H^+ formation. The threshold for backbone shattering fragmentation is higher than that for side-chain shattering fragmentation or shattering fragmentation forming an H^+ or N_2H^+ ion. While also the backbone shattering fragmentations concern parts of the molecule most exposed to the projectile, the side chains (and also the leaving H^+) are slightly more exposed, being probably at the origin of a lower shattering energy threshold.

As shown in figure 5.21, the percentage of all the trajectories which are shattering at a specific E_{rel} decreases with decrease in E_{rel}, becoming 0.04% for E_{rel} = 100 kcal/mol. A

collision energy threshold for shattering will be thus less than 100 kcal/mol. This energy value corresponds to the collision energy, i.e. the relative energy set for the collisional system. From the trajectories, it is possible to obtain the amount of transferred energy and decompose it as a function of shattering and non-shattering trajectories for different values of E_{rel}. It results in a distribution of values which is reported in figure 5.23. The energy distributions, in case of non-shattering trajectories, will be a function of simulation time: in fact, in the long-time limit they will provide statistical fragmentation. For this reason, the minimum values of internal energy in what here we identified as "non-shattering" do not correspond to what obtained from a previous statistical study in which the threshold was obtained from an Arrhenius plot: present simulations cannot, by definition, reach the statistical limit. On the other hand, shattering trajectories are defined as the one who react in less than 40 fs, and thus they are largely inside our simulation time. It is thus possible to identify the minimum transferred energy value for which shattering occurs. We should remark that the internal energy is always less than the collision energy (E_{rel}), and it corresponds to the actual activation provided to the ion after the collision. From the energy distributions it is possible to identify the average energy transfer which are listed in table 5.9 as a function of E_{rel} as obtained in all of the fragmentations and then only for the shattering fragmentations. For shattering we report in the same table also the minimum energy transfer which provides an estimation of the internal energy threshold for shattering. The threshold of 55 kcal/mol is substantially higher than the lowest activation energy of 14.7 kcal/mol, found from direct dynamics simulations for the thermal dissociation of $TIK(H^+)_2$ (see section 5.2). Thus, the shattering threshold is significantly higher than that for statistical redistribution of the ion's vibration energy and RRKM unimolecular kinetics.

Concluding, this is the first time the details of shattering fragmentation is reported from a relatively large system. An approach on how obtain shattering threshold is shown and applied to a tripeptide, providing an energy value which is largely different with respect to RRKM value. This would stimulate the searching of shattering and RRKM energy threshold by comparing collisional simulations at different relative energies with statistical trajectory decompositions done at different temperatures. The method does not need the location of transition states and thus it can be easily applied to complex molecules for which geometrical identification of saddle points (and more in general of reactive potential energy surfaces) is generally very problematic.

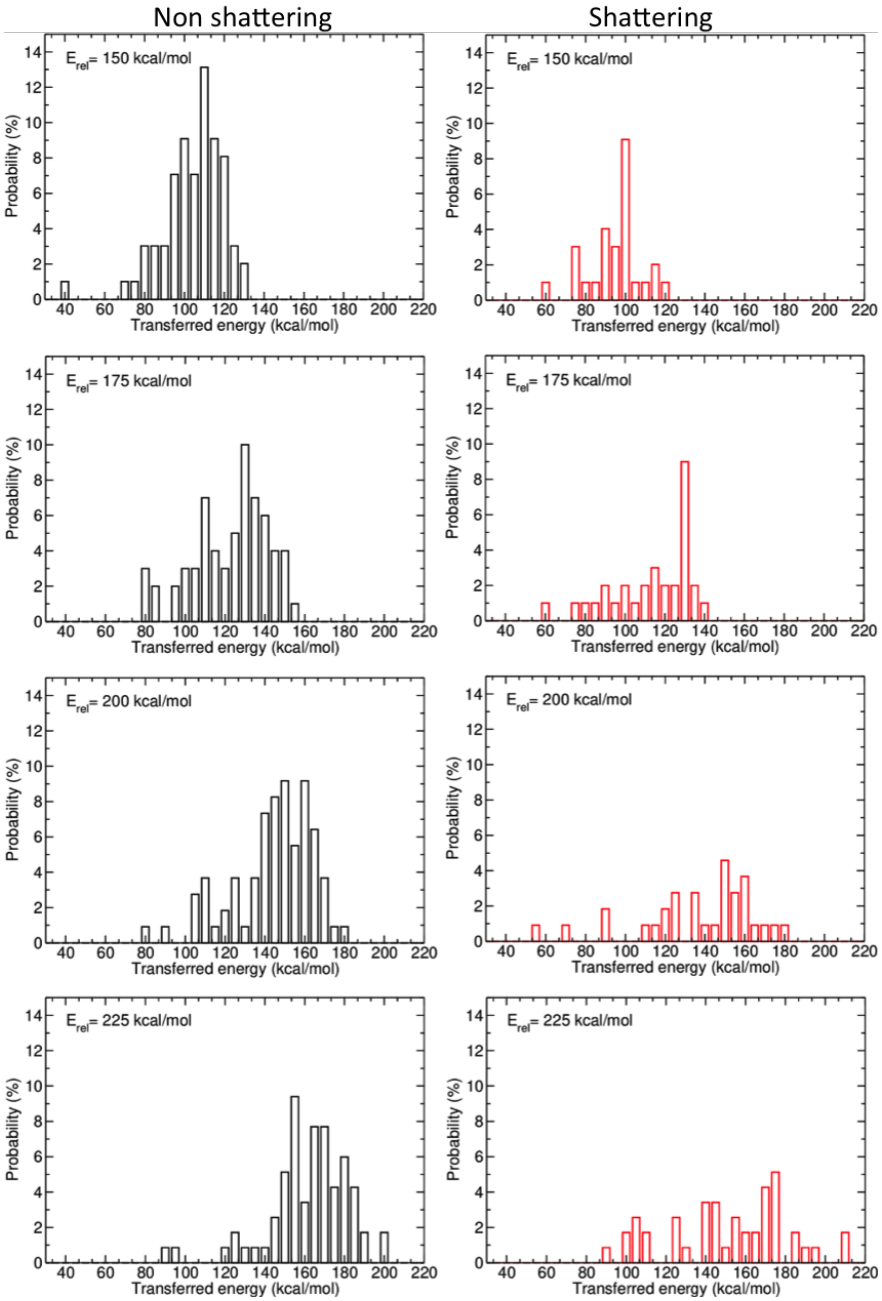


Figure 5.23: Probability of collisional energy transfer to TIK$(H^+)_2$ internal energy for non-shattering and shattering trajectories, with respect to the total number of shattering and non-shattering trajectories. The total probability of each graph is 100%, with results for different relative translational energies E_{rel}. The impact parameter is 2.5 Å.

Table 5.9: Energy transfer results for reactive trajectories as a function of the collision energy.

E_{rel}	150	175	200	225
Avarage energy transfer for reactive trajectories	100 ± 15	117 ± 20	139 ± 23	156 ± 25
Average energy transfer when shattering	93 ± 14	111 ± 20	134 ± 29	149 ± 30
Minimum shattering energy transfer	58	55	55	88

[a]Energies in kcal/mol

Chapter 6

Fragmentation of L-Cysteine sulfate anion

6.1 Introduction

L-Cysteine sulfenic acid (CysSOH) is a well known post-translational protein modification (PTM) [122]. It is involved as intermediate in oxidation processes that induce the reduction of L-Cystein-thiols (CysSH) forming disulfure bonds and consequent proteins folding [122], regenerating the proteins activity. Thus, CysSOH formation, along the proteins folding process, plays different roles in different cellular processes, in particular, to prevent the oxidative stress [123, 124, 122], to mediate redox signaling [122, 125], etc..

Even if sulfenic acids have been identified in several biochemical works, they are transient and not isolable in small molecules [6]. In a recent work [6], however, it was possible to form the reduced form L-Cysteine sulfenate ($CysSO^-$) by means of low energy collision-induced dissociation (CID). Although the structure of this species, with mass over charge (m/z) 136, was unambiguously characterized [6], it was not possible to unravel the fragmentation mechanism leading to it. Thus, a combined study on the CID dissociation of L-Cysteine-sulfate anion ($CysSSO_3^-$) was performed by combining chemical dynamics simulations, quantum chemistry calculations and experiments. Experiments were performed by the groups of D. Scuderi at the Laboratoire de Chimie-Physique at Universitè Paris-Sud (Orsay) and Prof. M. E. Crestoni at the University of Roma, La Sapienza (Italy).

6.2 Experimental methods

Here we summarize the different experiments performed by the above-mentioned groups.

The fragmentation of $CysSSO_3^-$ induced by CID has been studied by means of three different mass spectrometers: a commercial hybrid triple quadrupole/linear ion-trap (QTRAP), a modified Paul ion trap (Bruker, ESQUIRE 3000) [126] and a hybrid Fourier transform-ion cyclotron resonance (FT-ICR, Bruker, Apex Qe) mass spectrometer [127]. The three experimental sets up are equipped with an ESI (electrospray ionization) source. Mass spectrometry experiments of energy-resolved CID are carried out in the hybrid triple quadrupole (in Rome) which is endowed with three quadrupoles, Q1, q2 and Q3. The ions of interest are mass-selected using Q1. The first and the third analysers (Q1 and Q3) are separated by a collision cell into which an inert gas (N_2) is admitted to collide with the selected sample ions. CID experiments are performed in the quadrupole cell q2 at a variable energy (E_{lab}= 5-50 eV). The nominal pressure of the N_2 gas was typically set at about 10^{-5} mbar. The ionic products were monitored by scanning the Q3 sector which may act as either a quadrupole analyser or a linear trap.

Collision-induced dissociation has also been performed in Orsay in two modified ion traps: a Paul ion trap and a hybrid FT-ICR mass spectrometer. In the Paul trap, after ESI of the solution of the sample, ions are guided through a capillary and two octopoles into the trap, where they are spatially confined for extended period of time from 100 ms to 1 s. The CID is then performed at a pressure of He estimated to be about 10^{-3} mbar.

In the hybrid FT-ICR mass spectrometer, after ESI of the solution of the sample, ions are guided through a capillary and an ion funnel into a first hexapole, where they are collected and then pulse-extracted into a *radio-frequencies* (RF) quadrupole ion guide, where mass selection can be performed by resonant RF ejection of other present species. Selected ions are then collected in a second hexapole cell, where CID can be performed using a flow of high-purity Ar, at the pressure of about 10^{-6} mbar, and applying a collisional voltage from 1 up to 40 V. The typical collection time for the current experiment in the hexapole collision cell varies from 500 ms to 1 s in these experiments. Ions are then pulse-extracted towards the ICR cell, where the mass spectrum is the Fourier transform of a time-domain transient averaged four times. Before entering the IRC cell, the ions are decelerated by two ring electrodes of which the one closest to the trapping electrodes is segmented into two half-electrodes, allowing for a more efficient trapping.

6.3 Theoretical methods

The L-Cysteine sulfate anion ($CysSSO_3^-$) is first optimized using three semi-empirical Hamiltonians: PM6-D [86], PM7 [85] and RM1 [82]. These semi-empirical methods were then used in the subsequent chemical dynamics simulations. Both internal energy and explicit collision activation are employed, which are explained in section 3.3.3 of chapter 3.

Four collision energies (in the experimental range) were chosen for the collision simulations, i.e. 1, 2, 3 and 5 eV. For all the simulations the initial distance between the anion and N_2 was set to 10 Å. The impact parameter b was sampled between 0 and $b_{max} = 4.5$ Å. b_{max} was chosen as done for the collision simulations on the di-proline anion (see section 4.3 of chapter 4) and on the TIK_2H^+ and TLK_2H^+ tripeptides ions (see section 5.3 of chapter 5).

The nuclear motion equations were then integrated using a sixth-order symplectic algorithm [108, 109], with an integration step size of 0.2 fs, which assures energy conservation. The trajectories with $E_{rel} = 1$ and 2 eV were stopped when the reactants reached a relative distance of 200 Å after the collision, while the trajectories at 3 and 5 eV were stopped at the distance of 250 Å. These conditions correspond to total simulation time up to 15 ps. We performed between 1500 and 2000 trajectories for each E_{rel} and for each of the semi-empirical Hamiltonians.

When performing internal energy activation only the $CysSSO_3^-$ is explicitly considered. The sampling of initial conditions is the same as discussed in section 3.3.3 of chapter 3. Simulations were performed for two internal energies: 87 and 166.7 kcal/mol. It should be reminded that these energies are added to the electronic energy minimum and not to the zero potential vibrational energy (ZPE) level. Trajectories at 87 and 166.7 kcal/mol were simulated respectively for 30 and 14 ps, in consideration of the decreasing anion lifetime with the increasing of its internal energy. Between 1500 and 2000 trajectories were propagated for each internal energy and for each semi-empirical Hamiltonian.
Calculations were performed by using VENUS96 [98] coupled with MOPAC version 5.022mn software for the semi-empirical Hamiltonian calculations.

The geometries of the products found as result of chemical dynamics simulations were first optimized using the correspondent semi-empirical Hamiltonian and than optimized again at $B3LYP/6-311^{++}G^{**}$ level of theory. Vibrational analysis were carried out to verify that the structure were minima. The energies were than corrected at CCSD(T)/aug-cc-pVDZ level of theory.

To determine the potential energy surface (PES), in terms of minima (intermediates) and

transition states (TS) connecting the reactants with the products, we used the reactive chemical dynamics simulations. In particular, trajectories were analysed using a modified version of the Transition State Search Chemical Dynamics Simulations [51, 52] (TSSCDS) method, recently developed by Martinez-Nuñez, or manually. When a pathway was identified (at semi-empirical Hamiltonian level) minima and TS were further optimized, as described previously for the simulations products. Note that the TSSCDS method was also able to find other local minima of the reactant, which were used as further initial structures in reactive dynamics to investigate the role of the initial conformation on final results.

B3LYP and CCSD(T) calculations were performed with the Gaussian09 [128] software.

6.4 Experimental results

We now summarize the experimental results performed in Rome and in Orsay. The $CysSSO_3^-$ anion (m/z 200) was produced in the gas phase via electrospray ionization (ESI) (see section 2.1 of chapter 2). It was fragmented in three different experimental devices: (i) an ESQUIRE, which is a Paul trap; (ii) a hexapole (of a FT-ICR instrument) and (iii) a quadrupole collision cell (of a triple quadruple).
As reported in a previous work [6], the fragmentation of m/z 200 in the Paul-trap provides only one product, m/z 136, which corresponds to $CysSO^-$. Also by changing the voltage of the trap, m/z 136 was the only product obtained. The structure of this ion was characterized in detail in a previous work [6] and it is schematically shown in figure 6.1.

A different picture is obtained by performing CID in the other two instruments. CID

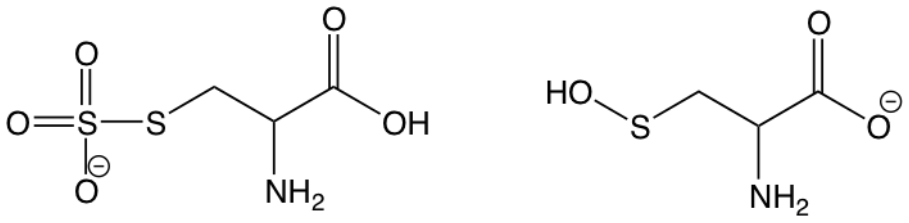


Figure 6.1: Structure formula of the reactant m/z 200 (left) and product m/z 136 (right).

of $CysSSO_3^-$ performed in the hexapole of the hybrid FT-ICR mass spectrometer produces $CysSO^-$ starting from a collisional voltage of 6 V. The product at m/z 136 is the main frag-

ment employing a collisional voltage up to 11 V. Two other products can be observed at m/z 113 and 120 respectively, moving at higher collision voltages. Results of hexapole fragmentation are shown in figure 6.2, which illustrates a plot with the relative abundance of the parent ion (at m/z 200) and daughter ions (at m/z 120 and 113) as a function of collision voltage. To calculate the ion relative abundances, the intensity of each fragment was calculated by the sum of all ions. In this way the relative abundances are normalized. These results show that in the hexapole other products can be obtained, in particular increasing the collision voltage, but that m/z 136 is still the most characterizing fragmentation product of L-cysteine sulfate anion.

While in the ion trap the ion get multiple collisions before fragmenting, in the hexapole the pressure (about 10^{-6} mbar) is much lower than in the Paul trap (estimated to be about 10^{-3} mbar). This can be at the origin of the different reactivity: in the ion trap the reaction can proceed via thermalization (due to multiple collisions) with the bath gas much more than in the hexapole.

Finally, a very different picture is obtained by using the quadrupole, as can be observed

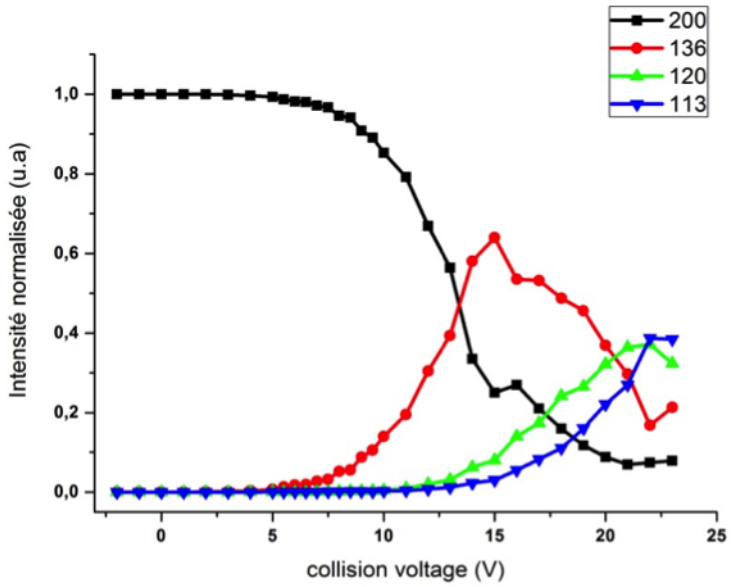


Figure 6.2: Fragmentation products of m/z 200 in the hexapole of a FT-ICR instrument at different collision voltage values.

in figure 6.3, in which the appearance of two different products is shown as a function of col-

lision energy, E_{rel} (E_{CM} in the figure, where it is expressed in the center of mass framework). Normalized relative abundances are shown also for this plot. Experiences have been recorded at different pressure values of N_2 collision gas (experiences have been recorded at different pressure values of N_2 collision gas to exclude the multicollisional events in the collisional cell. We can notice that at very low E_{rel} values, the product at m/z 136 is mainly obtained, but the product at m/z 81 becomes dominant already for $E_{rel} = 2$ eV. This ion was never observed in the Paul trap or hexapole. Furthermore, we have other peaks, some in common with the hexapole, that are m/z 120 and 113. Others are obtained only in the triple quadrupole, i.e. m/z 81, already mentioned, but also m/z 74 and 33 (this last being observed only at high energies). Results show that fragmentation obtained in a quadrupole collision cell are very different from what it is observed in both Paul trap and hexapole.

The origin of the differences observed should reside in the way the energy is transferred

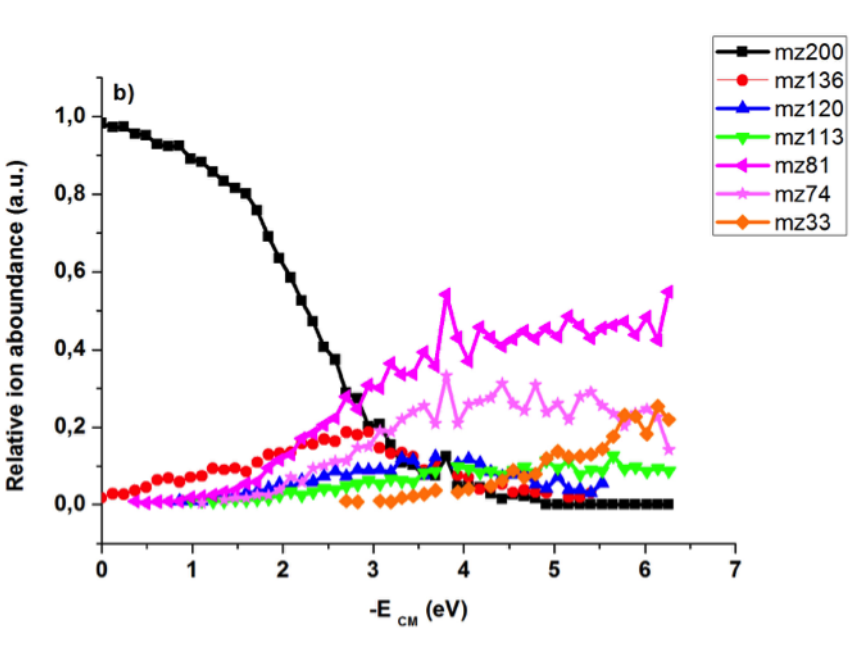


Figure 6.3: Fragmentation products of m/z 200 in the triple quadrupole as a function of the collision energy (E_{CM}).

to the ion and, more generally, the way the ion is activated and behaves in the instrument. Globally, we should note that a quadrupole collision cell gives a richer fragmentation pattern than the other ones, suggesting that linear motion, close to single collision limit, may have an important role. We have thus used chemical dynamics, using both explicit collision and

internal energy activation and coupled with quantum chemistry calculations, to elucidate the reaction mechanisms and to understand what is at the origin of the observed behavior.

6.5 Chemical dynamics simulations

Chemical dynamics simulations were performed by using two ways of activating the ion initially: by explicit collision with N_2 and by distributing internal energy into the ion (so without considering the gas). In explicit collision simulations we set the collision energy, which is partially transferred to the internal modes of the ion due to the collision. The effect of transferring the energy in such a way is that the energy is generally localized in certain parts (normal mode(s)) of the ion after the collision. Consequently the energy can be distributed between the normal modes or cause fragmentation before any or a full internal vibrational relaxation (IVR). On the other hand, internal energy activation corresponds to an initial microcanonical distribution of a given amount of energy through the different vibrational modes of the molecular ion. This last corresponds to a statistical unimolecular dissociation limit of the initial structure given that the exit channel is sufficiently close to the reactant. In other words, the statistical conditions are set for the initial structure and if an isomerization occurs the isomer(s) eventually formed does not have time to re-randomize for a further statistical reaction.

Fragments m/z 136, 120, 113, 81, 74 and 33 were all obtained using the simulations. Some differences are obtained as a function of the method, as it is possible to see in the tables A.7, A.8, A.9, A.10, A.11 and A.12 in *Appendix A*. In particular, RM1 tends to overestimate m/z 120 product, which corresponds to the cysteine anion. PM6−D on the other hand, overstimates the product m/z 113, while PM7 overstimates both m/z 120 and 113. The structures of the fragments ions obtained from simulations and further optimized at $B3LYP/6-311^{++}G^{**}$ level of theory are reported in figure 6.4.

Based on the products obtained from chemical dynamics sampling of the reactivity, it was possible to identify the different reactions and calculate the energetics. Results are reported in table 6.1.

We should notice that m/z 113 and 136 have similar formation energy, but the most observed one in Paul trap (m/z 136) is not the most stable thermodynamically. Furthermore, m/z 81, which is the most abundant one at high collision energies in the triple quadrupole,

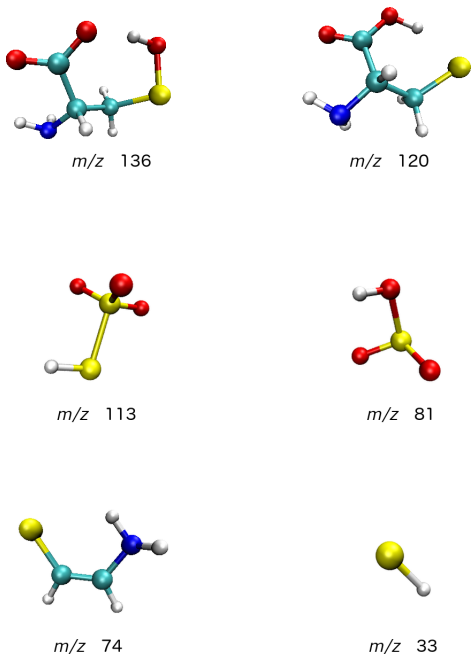


Figure 6.4: Structures of the fragment ions obtained as optimized at B3LYP/$6-311^{++}G^{**}$ level of theory.

is higher in energy. Finally, we should note that some large differences between B3LYP and CCSD(T) results are found for some reactions (formation of m/z 113 in particular). Thus, in the following we will discuss potential energy surface results calculated from CCSD(T) electronic energies on B3LYP structures and vibrational frequencies.

Simulations provide not only the products but also the mechanisms connecting the precursor ion with the different products. Thus, we investigated the mechanisms in three different ways: (i) we directly watched the corresponding reactive trajectories; (ii) we used the reactive trajectories as phase space sampling from which minima and TSs, connecting the precursor with products, were localized, mainly thank to the TSSCDS approach; (iii) using some key intermediates obtained as new initial structures to understand better some regions of the phase space.

Table 6.1: Fragmentation reactions corresponding to the formation of the different product ions. Reaction energies (ΔE = $E_{products}$ - $E_{reactant}$ in kcal/mol) are reported at B3LYP/6 − 311^{++}G** and CCSD(T)/aug-cc-pVDZ levels of theory with ZPE correction (calculated in both cases at B3LYP level).

Reaction	m/z product	B3LYP	CCSD(T)
$CysSSO_3^- \rightarrow C_3NO_3SH_6^- + SO_2$	136	27.15	28.59
$CysSSO_3^- \rightarrow$ L-$Cys^- + SO_3$	120	52.26	54.97
$CysSSO_3^- \rightarrow HS_2O_3^- + C_2NH_5 + CO_2$	113	9.38	28.42
$CysSSO_3^- \rightarrow HSO_3^- + C_3NO_2SH_5$	81	40.39	41.66
$CysSSO_3^- \rightarrow C_2NSH_4^- + SO_3H_2 + CO_2$	74	29.49	38.49
$CysSSO_3^- \rightarrow HS^- + C_2NH_5 + SO_3 + CO_2$	33	92.18	83.29

6.6 Fragmentation mechanisms

We investigated the mechanisms responsible of the formation of the different products. This aspect will be important to fully understand their appearance. In fact, two aspects should be addressed based on the experimental and simulation results: (i) why there are differences in product distribution as a function of the way the fragmentation is done experimentally? (While not large differences are obtained from different activation modes in simulations); (ii) why the most stable product (m/z 113) is not the most abundant and why a high energy product is the most observed one at high collision energies in the triple quadrupole instrument?

As remarked previously, we now use the simulations to find the mechanisms and to help in determining the reaction pathways and in particular the relevant transition states.

6.6.1 Formation of m/z 113

The ion with m/z 113 is largely observed in all simulations. The main mechanism responsible of this product is reported in figure 6.5. As shown, the fragmentation is initiated by a proton transfer from -COOH to -NH_2 group, followed by a concerted step in which three fragments are formed: CO_2, NH_2CHCH_2 and the ion m/z 113. As it results from simulations, the ion m/z 113 has a deprotonated S group, but the other isomer, deprotonated on the O is 1.21 kcal/mol more stable (as obtained at CCSD(T)/aug-cc-pVDZ//B3LYP/6 − 311$^{++}G^{**}$ level of

theory, see figure 6.6). The oxygen-deprotonated species is not observed in trajectories probably because the system has no time to isomerize. In fact, as we will detail from a PES analysis, the fragmentation mechanism is such that the sulfur-deprotonated species is obtained first. Using the reactive trajectories, in fact, it was possible, by employing the TSSCDS approach, to locate the minima and TSs connecting the reactant with the products. The resulting potential energy surface is reported in figure 6.6.

The analysis of the PES shows that the rate determining step is the loss of neutrals from

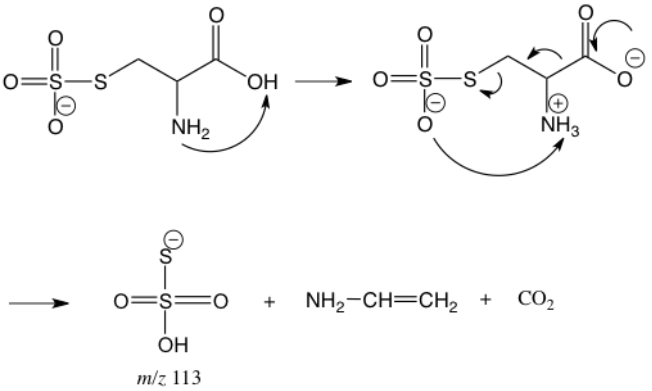


Figure 6.5: Mechanism responsible of the formation of ion m/z 113 as obtained in chemical dynamics simulations.

the reactant tautomer where the -COO group is deprotonated (structure M_A1 in figure 6.6), while the barrier to obtain this tautomer from the initial one (M_0) is much smaller. Note also that the TSSCDS analysis on the simulations results have found different isomers of the initial structure, one of them being slightly more stable than the initial structure (M_0). We have thus run trajectories also with this new conformer as initial structure and results are comparable with the previous ones.

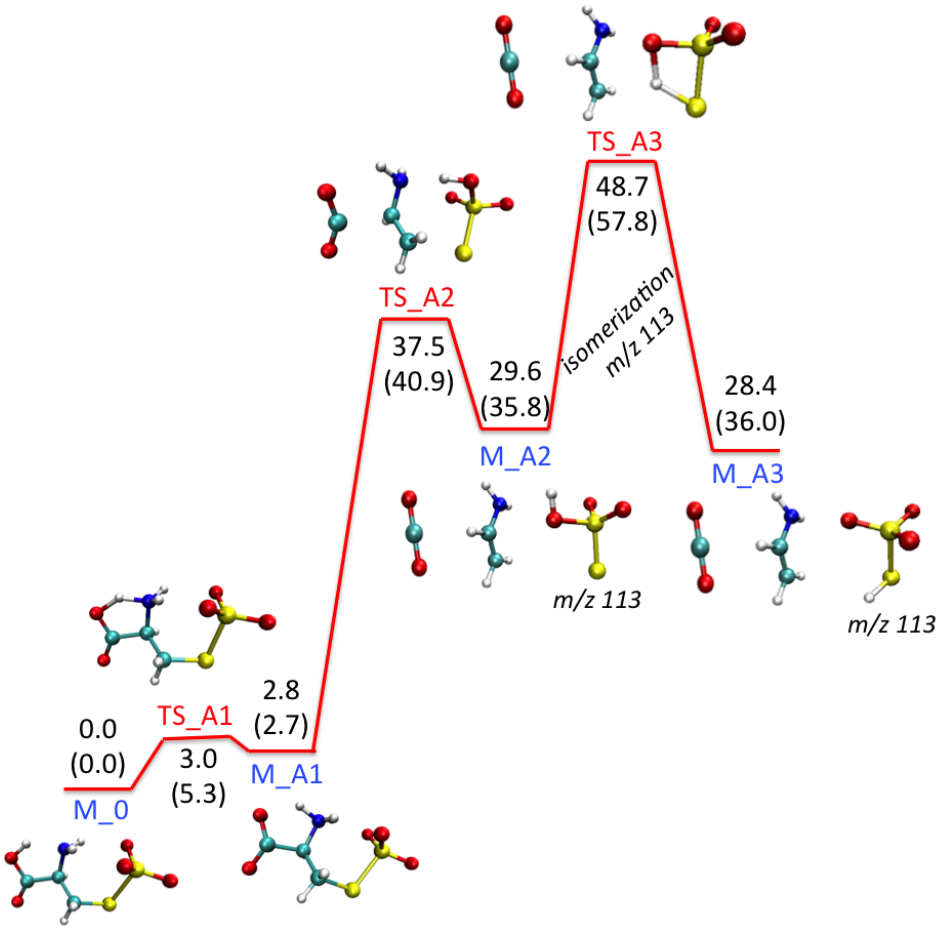


Figure 6.6: Potential energy surface corresponding to the formation of m/z 113. Results are shown at CCSD(T)/aug-cc-pVDZ//B3LYP/$6-311^{++}G^{**}$ level of theory with ZPE correction. Relative energies only considering electronic energies are shown in parenthesis. Energies are in kcal/mol. Transition state structures are labelled in red, minimum structures in blue.

6.6.2 Formation of m/z 136, 81 and 74

Ions m/z 74, 81, 136 are also obtained in simulations. By inspecting the trajectories it turns out that these three fragments are connected, in particular m/z 81 and 136, while for m/z 74 there is a second reaction pathway which is more probable. The fragmentation mechanism obtained is schematized in figure 6.7. The reaction proceeds as follows: first there is a proton

transfer from -COOH to -SO_3^- group. If a particular conformer of the anion is formed, it is then possible to observe a concerted mechanism consisting in a nucleophilic attack of the oxygen atom of the carboxylate group to the S atom of Cysteine with the simultaneous cleavage of the S-S bond. Thus, by a further simple separation of the two incipient fragments it is possible to form m/z 81 and a neutral 5-member ring. Alternatively, the two fragments can remain trapped and roam around each other. The two fragments can now react again, since the neutral molecule can pick an -OH group from the HSO_3^- group leaving neutral SO_2 and forming m/z 136. Another possibility is instead that the HSO_3^- group picks a proton from the neutral ring, forming the neutral H_2SO_3. The ring is now open and negatively charged: CO_2 leaves leading to fragment m/z 74. The potential energy surface for the formation of m/z 81 is shown in figure 6.8, together with the energies of m/z 136 and 74 products. In fact, it is well known that roaming mechanisms do not proceed via tight transition state [54].

The formation of m/z 136 is also illustrated in figure 6.9 using the snapshots of a repre-

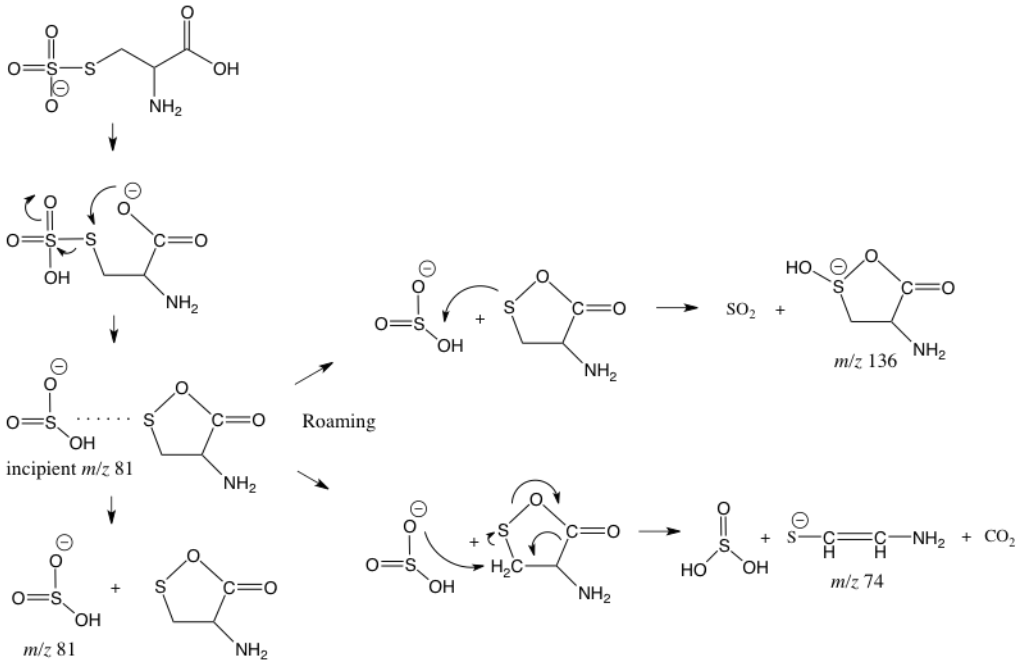


Figure 6.7: Mechanisms responsible of the formation of ions m/z 81, 136 and 74.

sentative trajectory and the time evolution of some representative distances.

While m/z 74 can be formed via a roaming mechanism, the mechanism mostly observe in simulations is much simpler ad it is shown in figure 6.10. The fragmentation starts with a proton transfer from -COOH to -SO_3^-. Subsequently a second proton is transferred from the

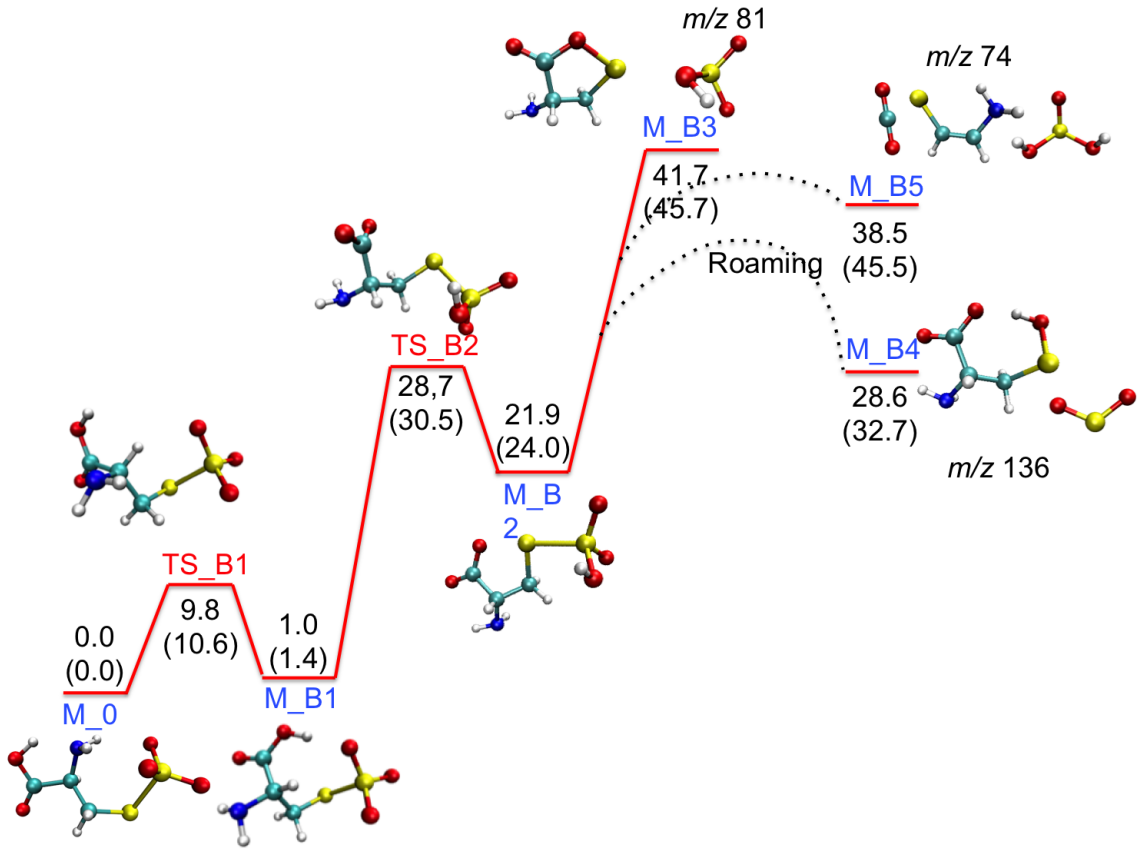


Figure 6.8: Potential energy surface corresponding to the formation of m/z 81. Results at CCSD(T)/aug-cc-pVDZ electronic level of theory with B3LYP/$6-311^{++}G^{**}$ ZPE correction are shown. Only electronic energy in parenthesis. Energies are in kcal/mol. Transition state structures are labelled in red, minimum structures in blue.

-CH_2- to the SO_3H^- group, that forming H_2SO_3 leaves. The exit of the neutral loss CO_2 leads finally to the formation of m/z 74. This second mechanism is similar to the other one but now the proton is taken from a linear species.

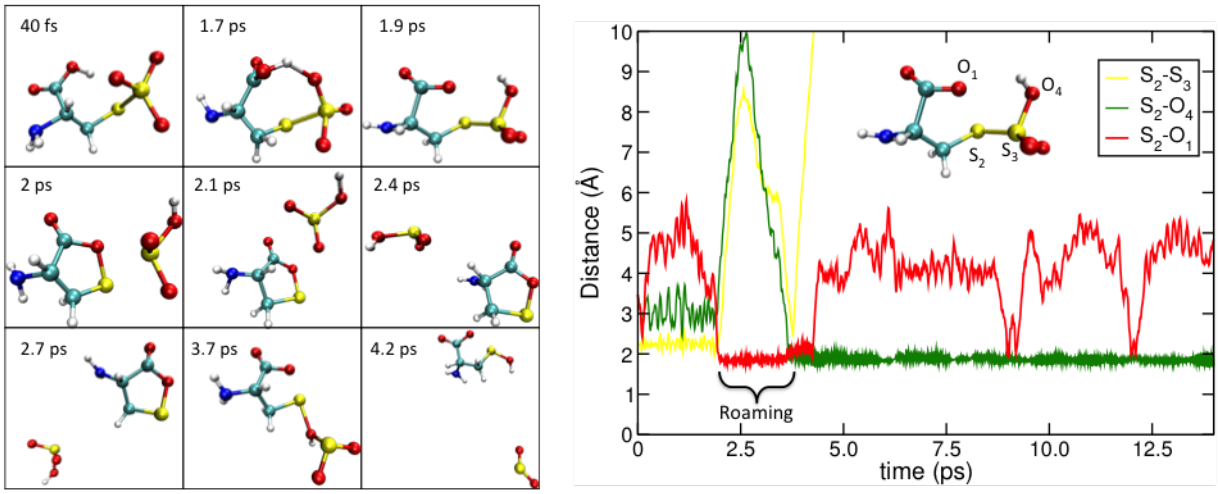


Figure 6.9: A PM6-D trajectory example for the formation of m/z 136 via roaming mechanism. In the panel on the left we show some representative snapshots while in the right panel the time evolution of three interatomic distances which characterize the fragmentation highlighting the roaming time.

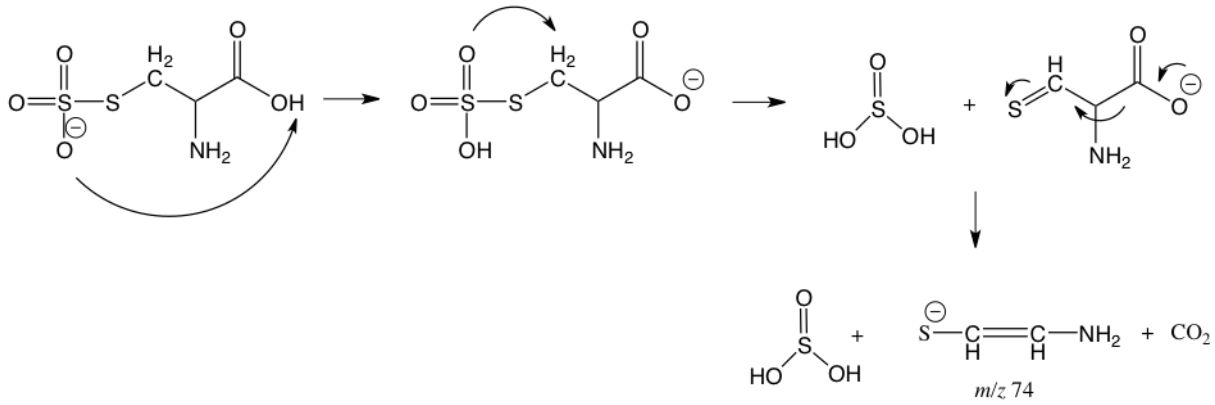


Figure 6.10: Mechanism for the formation of ion m/z 74 mostly observed in chemical dynamics simulations.

6.6.3 Other fragmentation mechanisms

In figure 6.11 the fragmentation mechanisms, as obtained from simulations, leading to the fragment ions m/z 120 and 33 are depicted.

The fragmentation mechanism leading to ion m/z 120 consists in the simple cleavage of

the S-S bond, resulting in the formation of the anion itself and the neutral SO_3.

In the case of the formation of the ion m/z 33 many different mechanisms are observed. In fact, it has been possible to form this fragment only at high collision energies (5 eV) or at high internal energies (166.7 kcal/mol), respectively in collision and thermal simulations. Note that in experiments it is observed only at high collision energies and only in the triple quadrupole.

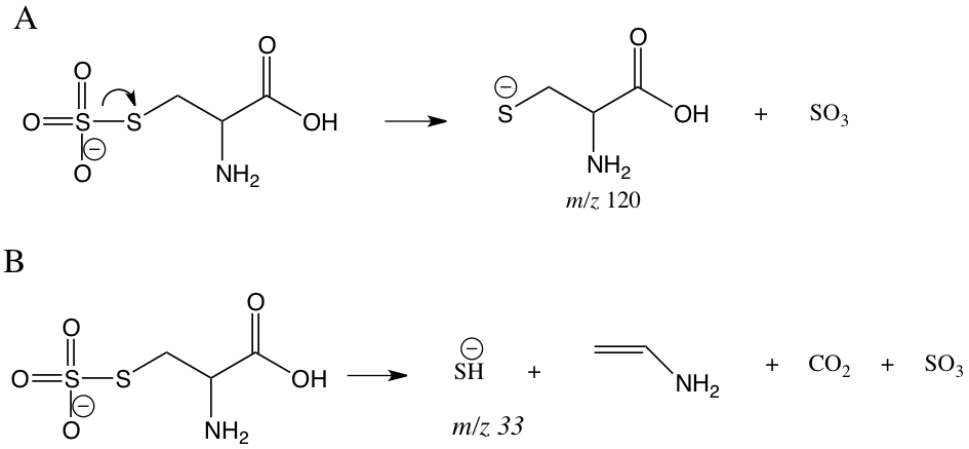


Figure 6.11: Fragmentation pathways for the formation of ions m/z 120 (A) and of m/z 33 (B).

6.7 Discussion

We can now rationalize what obtained in the experiments on the basis of the simulations and PES results, at CCSD(T)/aug-cc-pVDZ//B3LYP-6311^{++}G** level of theory. We start analysing the results of the two most different instruments here used: the Paul trap and the triple quadrupole. To do it we take in account the products that are most abundant in these two experiments (m/z 81 and m/z 136 in the quadrupole and in the Paul trap respectively) and the low energy product m/z 113. We can summarize the results obtained for these ions in figure 6.12.

In the quadrupole, differently from a Paul trap, the reaction is not thermalized continuously from multiple collisions and the fragmentation may not be under statistical conditions.

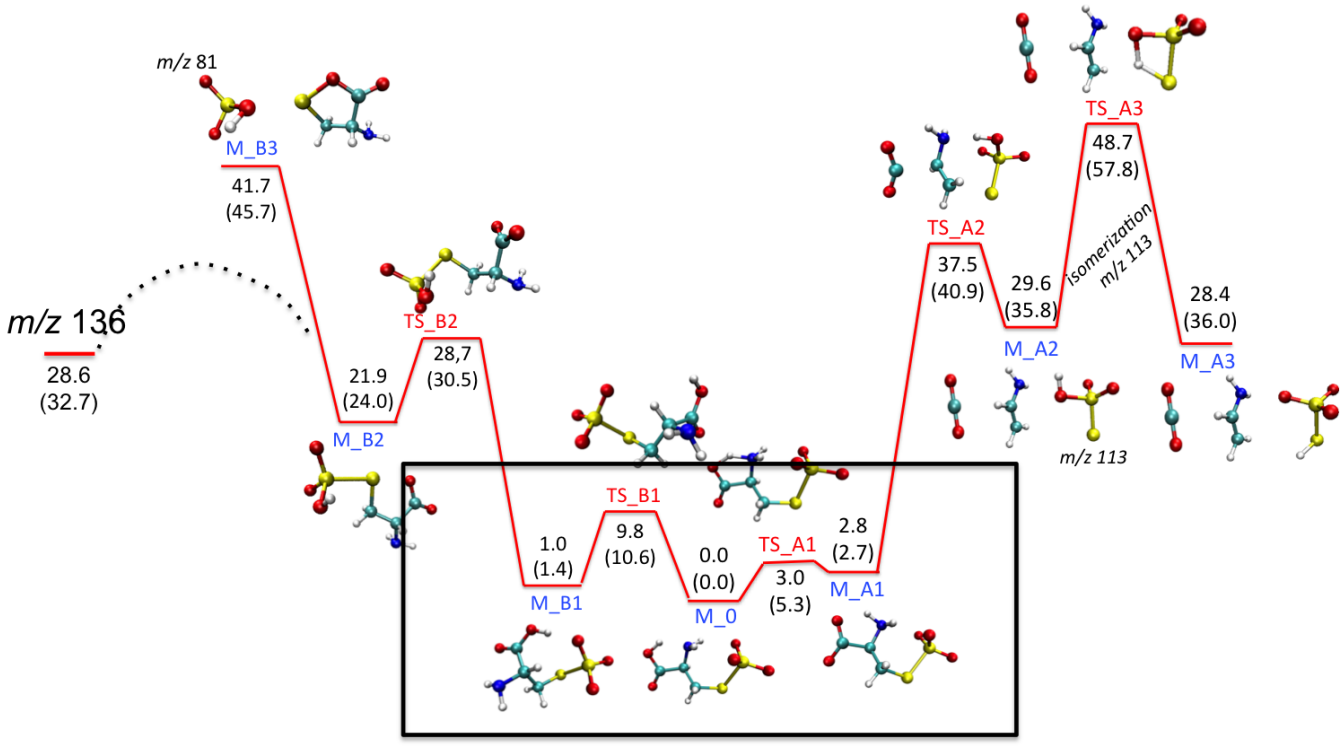


Figure 6.12: Potential energy surface corresponding to the formation of m/z 81, m/z 136 and m/z 113. Results at CCSD(T)/aug-cc-pVDZ electronic level of theory with B3LYP/6 − $311^{++}G^{**}$ ZPE correction are shown. Only electronic energy in parenthesis. Energies are in kcal/mol. Transition state structures are labelled in red, minimum structures in blue. In the black square there are the lowest energy minima conformers.

Furthermore, and especially increasing the collision energy, the energy can be localized in certain normal mode(s) of the ion, as we discussed before, and this may lead to a fast dissociation, i.e. without IVR. When the ion is brought to the gas phase all the minima (see structures in the black square in figure 6.12) can be populated (with different probabilities): the reaction can then follow the path on the left (m/z 81 and m/z 136), requiring as first step a proton transfer, or the path in the right (m/z 113), requiring the cleavage of two bonds. The reaction will proceed with more probability toward the formation of m/z 81 and m/z 136 because the energy of TS_B2 is lower than of TS_A2 (see figure 6.12). Moreover a fast process will favor a dynamically easy process, such as the proton transfer leading to M_B2. Moreover, a two-bond cleavage will require a more complex localization of the energy (in two bonds indeed) and it

is less probable. From the minimum M_B2 it would still possible for the reaction to come back in the square region of the lower energy minima, but there is not equilibrium due to the absence of thermalization. From M_B2 the reaction will end with the formation of m/z 81 and not m/z 136 because of the following: especially at higher energy collisions the two incipient fragments (m/z and the neutral) will go far away using the internal S-S vibrational energy conversion to translational energy of the two new fragments forming m/z 81. Furthermore, m/z 81 is formed by M_B2 through a dissociative transition state (TS). This kind of TS has not saddle point and it is called *loose* TS, while a *tight* TS is characterized by a saddle point on a potential energy surface. An important consequence of the fact that m/z 81 is formed through a loose transition state (while m/z 136 and 113 are not) is that respective free energy barrier (due to the entropic contribution) is energy dependent and, in particular, it will decrease as the internal energy (so also the collision energy) is increased. Subsequently, at higher energies we can expect that the formation of m/z 81 will be strongly favoured.

On the other hand, if the energy is not enough (low energy range) the incipient fragments formed by M_B2 can interact, roam around and rearrange forming m/z 136.

The activation mode of an hexapole can be compared to the Paul trap one. However, it was used at lower pressure and this can be at the origin of a difference in the fragmentation due to a minor thermalization capacity. The effect is that more products can be formed (with energy barriers not too high). However, the energy of one collision is never high as much as in the quadrupolar collision cell and even increasing the voltage it is not possible to strongly localize the energy in a specific normal mode of the ion. As a consequence m/z 81 is never formed confirming its role as the dynamical product of the L-Cysteine sulfate fragmentation.

6.8 Conclusions

We can summarize the results obtained as follows:

1. In the Paul trap only one product is obtained, that is m/z 136.

2. In the quadrupole, at the lowes energies m/z 136 is the only product obtained, while at higher collision energies many other fragments appear, m/z 81 being the most abundant one.

3. In the hexapole, at low voltage only m/z 136 is obtained, and other fragments, m/z 120

and 113 appear.

4. From simulations we obtained all the experimental products, for which we could propose the correspondent structures and energies.

5. We could characterize in detail, with a relatively accurate level of theory, fragmentation mechanisms for the formation of the most abundant products in the experiments (m/z 81 and m/z 136) and the low energy product (m/z 113).

6. We rationalized two instruments different behaviour in terms of statistical/kinetic control (Paul trap) and non-statistical/dynamical control (triple quadrupole).

7. The activation mode in the hexapole (of the FT-ICR instrument) is intermediate between the Paul trap and quadrupole: not only m/z 136 is formed, but the collisions are too low in energy to give m/z 81.

8. The PM6-D semi-empirical Hamiltonian recovered better than the other methods the most abundant fragments in the experiments (see tables A.7, A.8, A.9, A.10, A.11 and A.12 in *Appendix A*) but it overestimates m/z 113. Product m/z 120 was obtained only using PM7 and RM1.

9. Collisional activation results (from PM6-D) are qualitatively in agreement (almost all the products are obtained but with different abundance) with triple quadrupole experiments.

10. Results underlined that internal energy activation simulations cannot be compared with Paul trap experiments for the fragmentation of $CysSSO_3^-$: in the experiments there is a continuous thermalization while in the simulations only the reactant is thermalized and not the following intermediates.

11. We performed additional simulations starting from minimum M_B2 of figure 6.12 (see A.13 in *Appendix A*): PM6-D simulations still form preferentially m/z 113, while DFT simulations formed m/z 136 and m/z 81. This shows that the disagreement between experiments and simulations can be largely attributed to the semi-empirical Hamiltonians.

Chapter 7

General conclusions

In last years, the role of chemical dynamics simulations in the understanding of CID of different biological systems has been investigated. In particular, it was shown how simulations can help in the identification and clarification of peaks in experimental mass spectra of relatively simple systems, like organic molecules or simple peptides [22, 79, 27, 25, 91], i.e. how they can be used as an alternative to a standard comparison with mass spectra data-bases [41, 115]. In the present thesis we have then investigated the possibility of studying collision-induced dissociation (CID) of more complex amino acids and peptides by means of chemical dynamics simulations.

It was first necessary to underline the coupling of collisional and thermal activation in simulations: the first are limited in the time-scale and recover the more dynamical processes, while the second can be used to study more statistical and longer time-scale processes. A first comparison between results coming out from collision and thermal simulations was done studying the fragmentation of the di-proline ion: few differences were found in terms of products. However, some of them can be formed through different mechanisms depending on the activation mode.

We thus reported a first detailed study on the fragmentation of relatively large systems (59 atoms), two tripeptide ions, $TIK(H^+)_2$ and $TLK(H^+)_2$. This study has been very useful in order to develop and test a software that can be used to analyse faster and more easily the simulation results. We first performed thermal simulations on $TIK(H^+)_2$ ion in order to find the statistical pathways, rate constants and activation energies for the most probable ones. In particular, the statistical fragmentation threshold was found to be 14.7 kcal/mol, which value

corresponds to the energy barrier of the most probable pathway obtained using thermal simulations. We further performed collisional simulations on both the TIK$(H^+)_2$ and TLK$(H^+)_2$ ions, which showed product distributions similar to the ones obtained with thermal simulations when high temperatures (i.e. high internal energies) were considered. As expected, some peaks and mechanisms are different between the two activation modes. We analysed in more details the TIK$(H^+)_2$ and TLK$(H^+)_2$ side-chain fragmentations and we found that a peak would be characteristic of I, so it could be used to distinguish I from L in a mass spectrum. However, this peak was found to be really low in intensity so that, in order to be detected, the signal/noise should be drastically reduced in the experiments. We also investigated the limit case of non-statistical fragmentation, shattering, by performing lower energy collisional simulations for N_2 + TIK$(H^+)_2$. We identified the energy of 55 kcal/mol as the minimum transferred energy value for which shattering occurs. This threshold of 55 kcal/mol is substantially higher than the lowest activation energy of 14.7 kcal/mol, found from simulations for the thermal dissociation of TIK$(H^+)_2$. This study showed that in future, possibly coupling chemical dynamics with experiments, collisional simulations can be used to recover information about the side-chain fragmentation, while thermal simulations about backbone fragments and, in general, more statistical products. Moreover, information extrapolated by performing simulations, like the shattering or statistical fragmentation energy thresholds, may be used in order to simplify mass spectra by guiding the experimental conditions.

Finally, we presented a combined experimental and theoretical study on the CID dissociation of L-Cysteine-sulfate anion. Three different instruments were used and results showed to be dependent by their instrinsic characteristics. By performing chemical dynamics simulations, coupled with a guided PES study (the trajectories were used to obtain guess structures), we were able to find mechanisms, structures for many of the products and to understand the different instruments behaviour. The unique product obtained with the Paul trap instrument is fragment m/z 136. We expected to find this fragment with more facility using thermal activation than collisional, but we obtained similar results. We understood that thermal simulations cannot simulate efficiently a complex multi-step reaction involving a continuous energy redistribution because in the simulations only the reactant is thermalized and not the following intermediates. In the case of the L-Cysteine sulfate the semi-empirical Hamiltonians that were used (PM6-D, PM7 and RM1) showed to be really useful for an initial sampling. Despite we could not explain the fragmentation only by means of simulations, mechanisms for two of the most important products (m/z 136 and 81) could not been obtained only by using a static approach. In particular, we were able to detect a roaming mechanism for the fragmentation of

the L-Cysteine sulfate anion, mechanism that was confirmed from DFT thermal simulations that we have recently performed. Roaming mechanism has been recently detected in different experiments through the coupling with simulations [53], but it is has not found yet a roaming transition state [54], confirming the important role of chemical dynamics in the understanding of chemical reactions.

In future studies, by implementing the quantum thermal bath [129], path integral [130] or a combination of the two [131], will be likely possible to solve the problem of the incomplete thermalization of simulations and simultaneously the classical zero point energy (ZPE) missing in the products (it is possible to obtain products with less than the ZPE energy due to the classical treatment of the nuclear motion).

Concluding, when studying CID of large peptides, it was underlined the necessity of developing an efficient and automatic way to calculate the final charges on the fragments products obtained from simulations. This would represent the final step to produce, fully automatically, theoretical mass spectra for peptides, once the simulations have been performed. More computational resources and a complete automatic analysis of the trajectory can surely open the possibility of computing theoretical protein mass spectra. In this context, it is already been possible an user-friendly application of the analysis code. In fact, Dr. A. Carrà, a visiting experimentalist post-doctoral researcher (from the University of Minneaopolis (USA)) was able in two months to carry out mass distribution spectra which were compared with experimental results giving an excellent agreement. After this short period this study was also presented at the Americal Society of Mass Spectrometry conference in San Diego (USA, June 2018).

Appendix A

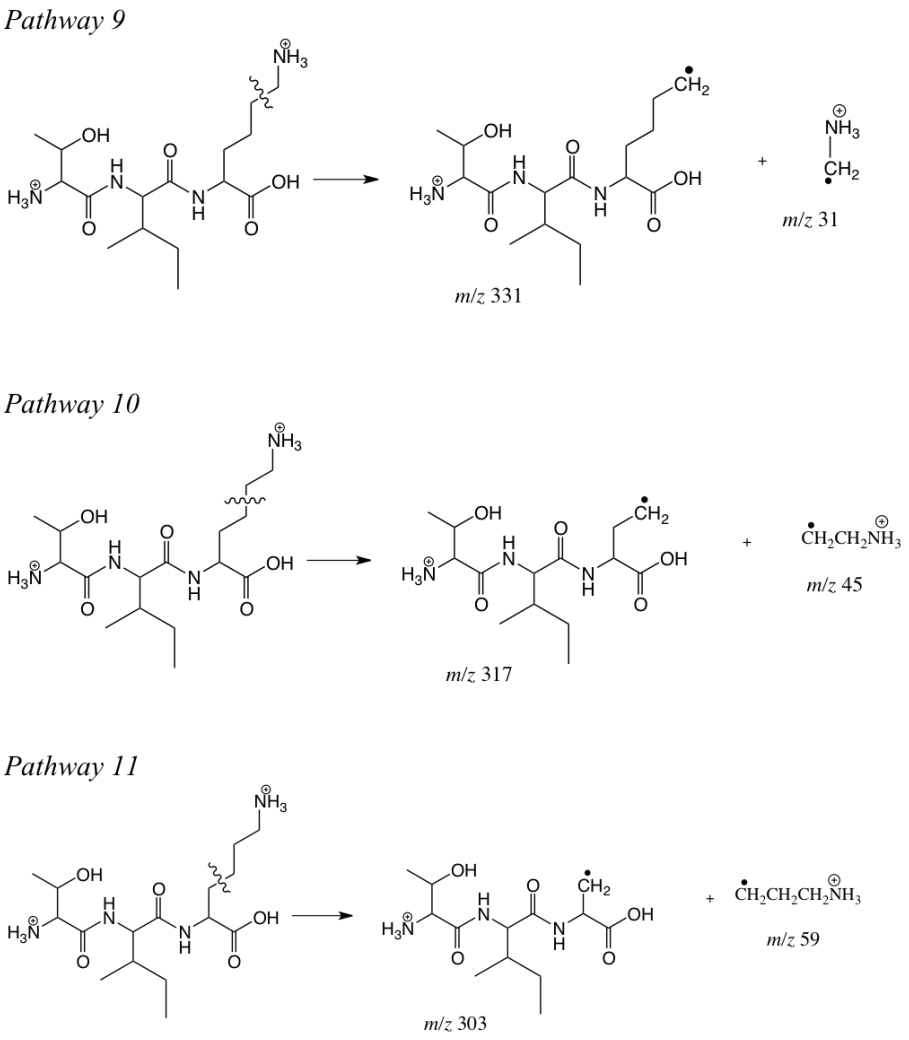
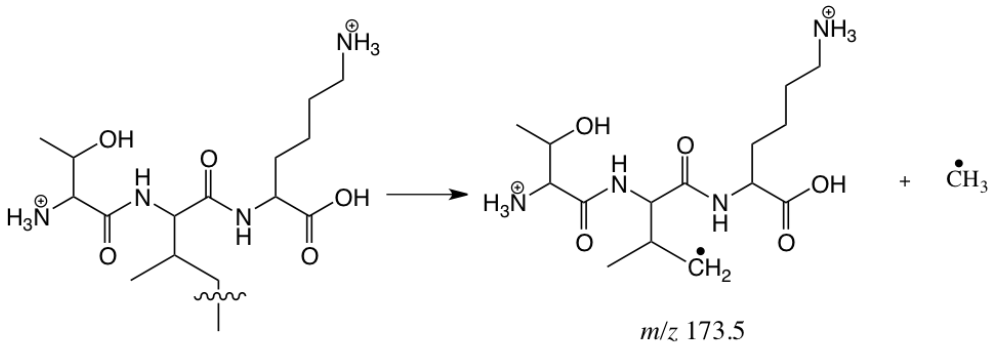


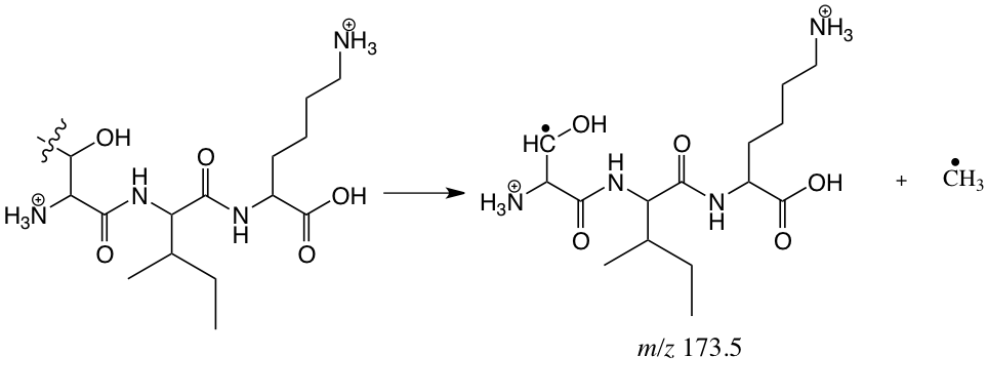
Figure A.1: Mechanisms for dissociation pathways $9 - 11$ of $\mathrm{TIK(H^+)_2}$

.

Pathway 12



Pathway 13



Pathway 14

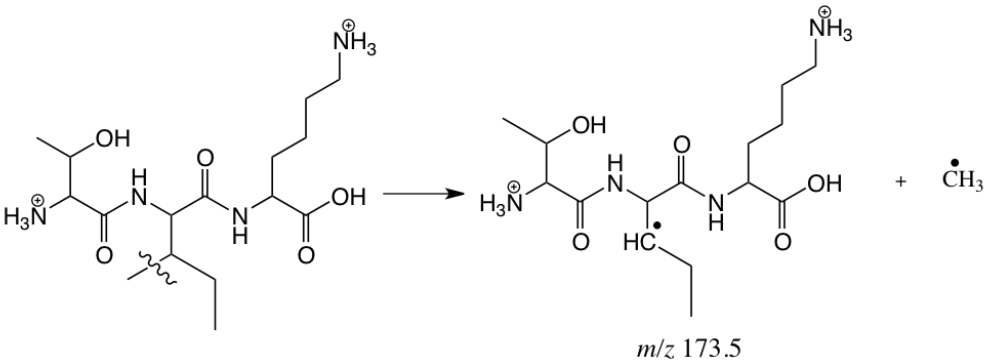
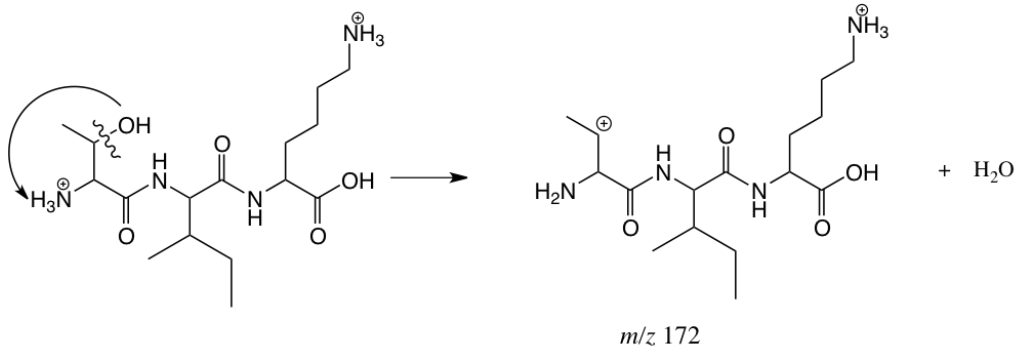


Figure A.2: Mechanisms for dissociation pathways $12 - 14$ of TIK$(H^+)_2$

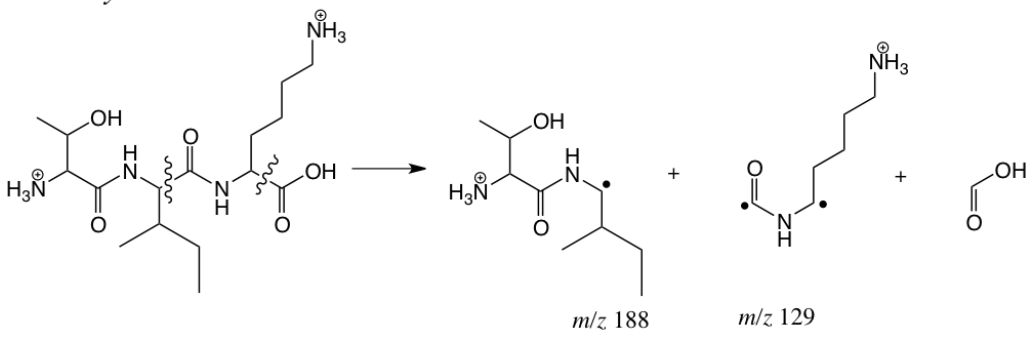
.

Pathway 15



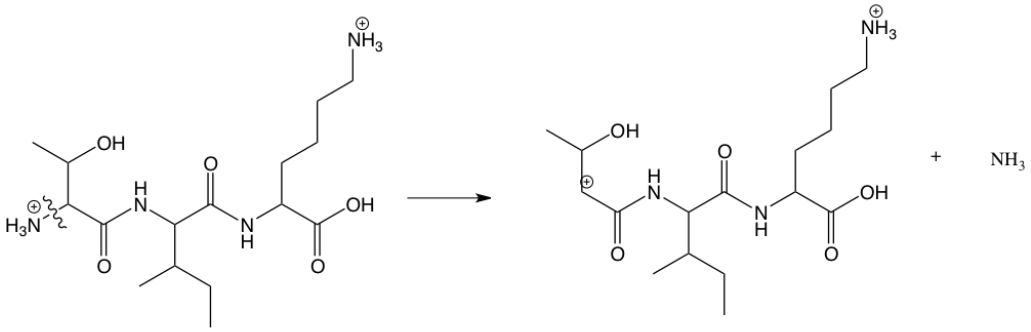
m/z 172

Pathway 16



m/z 188 *m/z* 129

Pathway 17

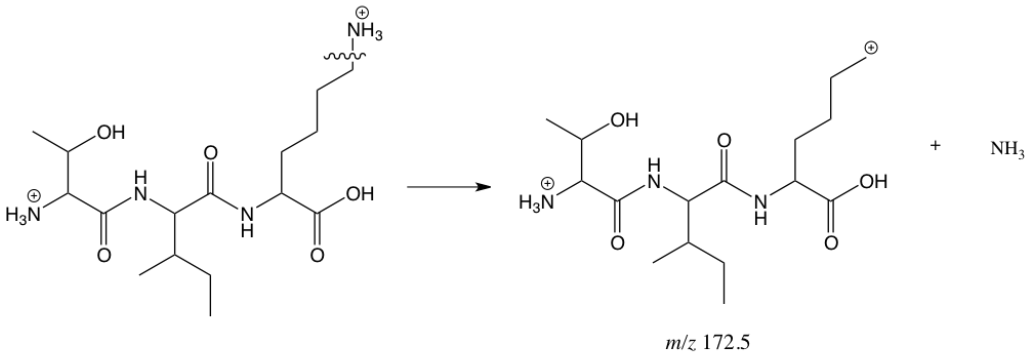


m/z 172.5

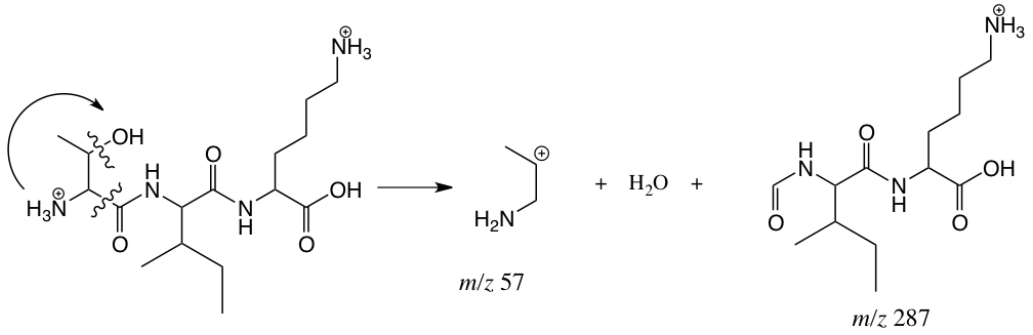
Figure A.3: Mechanisms for dissociation pathways $15 - 17$ of TIK$(H^+)_2$

.

Pathway 18



Pathway 19



Pathway 20

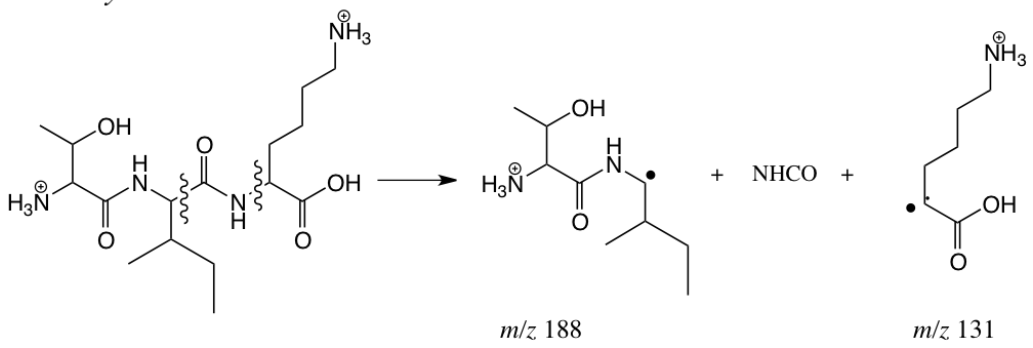


Figure A.4: Mechanisms for dissociation pathways $18 - 20$ of $\mathrm{TIK(H^+)_2}$

.

Table A.1: Probabilities of TIK$(H^+)_2$ dissociation pathways for CID at $E_{rel} = 10.8$ eV

Pathway	Dissociation site	Probability[a]
1	Backbone	14.0 ± 4.6 (0.0)
2	Backbone	10.5 ± 4.1 (0.0)
3	Backbone	1.8 ± 1.8 (0.0)
4	Backbone	14.0 ± 4.6 (50.0 ± 17.7)
5	Side chain	15.8 ± 4.8 (44.4 ± 16.6)
6	Backbone	15.8 ± 4.8 (44.4 ± 16.6)
7	Side chain	-
8	Side chain	1.8 ± 1.8 (100.0)
9	Side chain	1.8 ± 1.8 (100.0)
10	Side chain	-
11	Side chain	-
12-14	Side chain	1.8 ± 1.8 (0.0)
15	Side chain	-
16	Backbone	-
17	Backbone	-
18	Side chain	1.8 ± 1.8 (0.0)
19	Backbone	-
20	Backbone	-
21	H^+	5.2 ± 2.9 (100.0)
22	N_2H^+	12.3 ± 4.3 (100.0)
23	Side chain	-
24	Side chain	-
Other[b]		8.7 ± 3.7 (60.0 ± 21.9)

[a]Probability of this pathway with respect to the dissociating trajectories. The number in parenthesis is the probability that the trajectories for this pathway are shattering fragmentation.
[b] Probability of other fragmentation pathways with respect to the dissociating trajectories.

Table A.2: Probabilities of TLK$(H^+)_2$ dissociation pathways for CID at $E_{rel} = 10.8$ eV

Pathway	Dissociation site	Probability[a]
1'	Backbone	30.2 ± 6.3 (6.2 ± 6.0)
2'	Backbone	11.3 ± 4.3 (0.0)
3'	Backbone	5.7 ± 3.2 (3.3 ± 10.3)
4'	Backbone	3.8 ± 2.6 (50.0 ± 35.2)
5'	Side chain	15.1 ± 4.9 (62.5 ± 17.11)
6'	Backbone	9.4 ± 4.0 (80.0 ± 17.9)
7'	Side chain	1.9 ± 1.9 (100.0)
8'	Side chain	-
9'	Side chain	1.9 ± 1.9 (100.0)
10'	Side chain	-
11'	Side chain	-
12-14'	Side chain	1.9 ± 1.9 (100.0)
15'	Side chain	-
16'	Backbone	-
17'	Backbone	-
18'	Side chain	3.8 ± 2.6 (100.0)
19'	Backbone	-
20'	Backbone	-
21'	H^+	7.5 ± 3.6 (100.0)
22'	N_2H^+	5.7 ± 3.2 (100.0)
23'	Side chain	-
24'	Side chain	-
Other[b]		1.9 ± 1.9 (0.0)

[a]Probability of this pathway with respect to the dissociating trajectories. The number in parenthesis is the probability that the trajectories for this pathway are shattering fragmentation.
[b] Probability of other fragmentation pathways with respect to the dissociating trajectories.

Table A.3: Probabilities of TIK$(H^+)_2$ dissociation pathways for CID at E_{rel} = 13 eV

Pathway	Dissociation site	Probability[a]
1	Backbone	16.4 ± 2.4 (2.6 ± 2.5)
2	Backbone	7.6 ± 1.7 (0.0)
3	Backbone	3.8 ± 1.2 (88.9 ± 10.5)
4	Backbone	5.9 ± 1.5 (100.0)
5	Side chain	10.1 ± 1.9 (91.7 ± 5.6)
6	Backbone	2.1 ± 0.9 (100.0)
7	Side chain	-
8	Side chain	2.9 ± 1.1 (85.7 ± 13.2)
9	Side chain	-
10	Side chain	-
11	Side chain	0.4 ± 0.4 (100.0)
12-14	Side chain	5.0 ± 1.4 (100.0)
15	Side chain	2.1 ± 0.9 (100.0)
16	Backbone	-
17	Backbone	-
18	Side chain	1.8 ± 1.8 (0.0)
19	Backbone	-
20	Backbone	-
21	H^+	6.3 ± 1.6 (100.0)
22	N_2H^+	10.0 ± 1.9 (100.0)
23	Side chain	2.1 ± 0.9 (100.0)
24	Side chain	3.8 ± 1.2 (44.4 ± 16.5)
Other[b]		16.0 ± 2.4 (10.0 ± 4.9)

[a]Probability of this pathway with respect to the dissociating trajectories. The number in parenthesis is the probability that the trajectories for this pathway are shattering fragmentation.
[b] Probability of other fragmentation pathways with respect to the dissociating trajectories.

Table A.4: Probabilities of TLK$(H^+)_2$ dissociation pathways for CID at $E_{rel} = 13$ eV

Pathway	Dissociation site	Probability[a]
1'	Backbone	15.9 ± 2.3 (5.0 ± 3.4)
2'	Backbone	5.9 ± 1.48 (0.0)
3'	Backbone	6.3 ± 1.5 (56.2 ± 12.4)
4'	Backbone	6.7 ± 1.6 (41.2 ± 12.09)
5'	Side chain	9.5 ± 1.8 (100.0)
6'	Backbone	4.8 ± 1.3 (75.0 ± 12.4)
7'	Side chain	2.4 ± 1 (66.7 ± 19.2)
8'	Side chain	0.4 ± 0.4 (100.0)
9'	Side chain	2.0 ± 0.9 (60.0 ± 21.8)
10'	Side chain	0.4 ± 0.4 (100.0)
11'	Side chain	0.4 ± 0.4 (100.0)
$12 - 14$'	Side chain	7.9 ± 1.7 (80.0 ± 9.0)
15'	Side chain	1.2 ± 0.7 (66.7 ± 27.1)
16'	Backbone	0.4 ± 0.4 (0.0)
17'	Backbone	1.2 ± 0.7 (33.7 ± 27.1)
18'	Side chain	3.1 ± 1.1 (50.0 ± 17.9)
19'	Backbone	-
20'	Backbone	-
21'	H^+	14.3 ± 2.2 (100.0)
22'	N_2H^+	10.3 ± 1.9 (0.0)
23'	Side chain	1.6 ± 0.8 (0.0)
24'	Side chain	0.8 ± 0.6 (0.0)
Other[b]		4.4 ± 1.3 (54.5 ± 14.9)

[a]Probability of this pathway with respect to the dissociating trajectories. The number in parenthesis is the probability that the trajectories for this pathway are shattering fragmentation.
[b] Probability of other fragmentation pathways with respect to the dissociating trajectories.

Table A.5: Probabilities of TIK$(H^+)_2$ dissociation pathways for CID at E_{rel} = 30 eV

Pathway	Dissociation site	Probability[a]
1	Backbone	0.6 ±0.6 (0.0)
2	Backbone	0.6 ±0.6 (0.0)
3	Backbone	0.6 ±0.6 (0.0)
4	Backbone	0.6 ±0.6 (0.0)
5	Side chain	$1.2 \pm 0 - 8$ (50.0 ± 35.3)
6	Backbone	$1.2 \pm 0 - 8$ (50.0 ± 35.3)
7	Side chain	-
8	Side chain	1.8 ± 1.8 (100.0)
9	Side chain	0.6 ± 0.6 (100.0)
10	Side chain	-
11	Side chain	-
12-14	Side chain	2.4 ± 1.2 (100.0)
15	Side chain	-
16	Backbone	-
17	Backbone	0.6 ±0.6 (100.0)
18	Side chain	1.2 ± 0.8 (100.0)
19	Backbone	-
20	Backbone	-
21	H^+	8.2 ± 2.1 (100.0)
22	N_2H^+	10.0 ± 1.9 (100.0)
23	Side chain	0.6±0.6 (100.0)
24	Side chain	0.6 ±0.6 (0.0)
Other[b]		26.0 ± 3.4 (84.4 ± 5.4)

[a]Probability of this pathway with respect to the dissociating trajectories. The number in parenthesis is the probability that the trajectories for this pathway are shattering fragmentation.
[b] Probability of other fragmentation pathways with respect to the dissociating trajectories.

Table A.6: Probabilities of TLK$(H^+)_2$ dissociation pathways for CID at E_{rel} = 26 eV

Pathway	Dissociation site	Probability[a]
1'	Backbone	0.5 ±0.5 (0.0)
2'	Backbone	-
3'	Backbone	1.1 ±0.8 (50.0 ± 34.6)
4'	Backbone	0.5 ±0.5 (0.0)
5'	Side chain	2.1 ±1.0 (100.0)
6'	Backbone	3.2 ±1.3 (66.7 ± 19.2)
7'	Side chain	1.1 ±0.8 (100.0)
8'	Side chain	1.6 ±0.9 (100.0)
9'	Side chain	0.5 ±0.5 (100.0)
10'	Side chain	-
11'	Side chain	0.5 ±0.5 (100.0)
$12 - 14$'	Side chain	3.7 ±1.4 (100.0)
15'	Side chain	0.5 ±0.5 (100.0)
16'	Backbone	-
17'	Backbone	2.1 ±1.0 (100.0)
18'	Side chain	0.5 ±0.5 (100.0)
19'	Backbone	-
20'	Backbone	-
21'	H^+	57.1 ± 3.6 (100.0)
22'	N_2H^+	15.3 ± 2.6 (0.0)
23'	Side chain	1.1 ± 0.8 (0.0)
24'	Side chain	-
Other[b]		8.5 ± 2.0 (100.0)

[a]Probability of this pathway with respect to the dissociating trajectories. The number in parenthesis is the probability that the trajectories for this pathway are shattering fragmentation.
[b] Probability of other fragmentation pathways with respect to the dissociating trajectories.

Table A.7: Probability of forming the different products for the fragmentation of the L-Cysteine sulfate anion as obtained by thermal simulations at the internal energy of 87 kcal/mol using M_B1 as initial structure.

Method	PM6-D	PM7	RM1
136	0.0	0.0	0.0
120	0.0	18.4 ± 4.1	100.0
113	88.8 ±1.9	44.9 ±5.3	0.0
81	0.3 ±0.3	0.0	0.0
74	2.7 ±0.9	0.0	0.0
72	0.0	0.0	0.0
33	0.0	0.0	0.0
React.[a]	15.7 ±0.8	3.1 ±0.3	0.5 ± 1.9

Results are reported as a function of the internal energy and the semi-empirical method. The products are identified by their masses. [a] Percentage of total reactivity.

Table A.8: Probability of forming the different products for the fragmentation of the L-Cysteine sulfate anion as obtained by thermal simulations at the internal energy of 166.7 kcal/mol using M_B1 as initial structure.

Method	PM6-D	PM7	RM1
136	0.2 ±0.1	0.1 ±0.08	0.0
120	0.0	46.8 ± 1.3	54.8 ±3.1
113	83.4 ±1.1	35.0 ±1.3	16.0 ± 2.3
81	3.1 ±0.5	0.6 ± 0.2	0.0
74	3.8 ±0.6	0.4 ± 0.2	8.8 ± 1.8
72	0.5 ± 0.2	0.0	0.0
33	0.4 ± 0.2	7.3 ± 0.7	0.0
React.[a]	94.0 ±0.7	79.4 ±1.0	13.9 ± 0.8

Results are reported as a function of the internal energy and the semi-empirical method. The products are identified by their masses. [a] Percentage of total reactivity.

Table A.9: Probability of forming the different products for the fragmentation of the L-Cysteine sulfate anion as obtained by collisional simulations at the collision energy of 1 eV in the center of mass framework using M_B1 as initial structure.

m/z	PM6-D	PM7	RM1
136	0.0	0.0	0.0
120	0.0	39.5 ± 7.9	71.4 ±9.9
113	66.2 ±5.4	23.7 ±6.9	0.0
81	18.2 ±4.4	0.0	0.0
74	3.9 ±2.2	0.0	0.0
72	0.0	0.0	0.0
33	0.0	0.0	0.0
React.[a]	2.7 ±0.3	2.0 ±0.3	1.5 ± 0.3

Results are reported as a function of the internal energy and the semi-empirical method. The products are identified by their masses. [a] Percentage of total reactivity.

Table A.10: Probability of forming the different products for the fragmentation of the L-Cysteine sulfate anion as obtained by collisional simulations at the collision energy of 2 eV in the center of mass framework using M_B1 as initial structure.

m/z	PM6-D	PM7	RM1
136	1.6 ± 1.6	0.0	0.0
120	0.0	48.9 ± 7.4	64.9 ±7.8
113	64.5 ±6.1	20.0 ±6.0	2.7 ± 2.7
81	17.7 ±4.9	0.0	5.4 ± 3.7
74	0.0	0.0	0.0
72	0.0	0.0	0.0
33	0.0	0.0	0.0
React.[a]	4.3 ±0.5	2.4 ±0.3	9.3 ± 0.8

Results are reported as a function of the internal energy and the semi-empirical method. The products are identified by their masses. [a] Percentage of total reactivity.

Table A.11: Probability of forming the different products for the fragmentation of the L-Cysteine sulfate anion as obtained by collisional simulations at the collision energy of 3 eV in the center of mass framework using M_B1 as initial structure.

m/z	PM6-D	PM7	RM1
136	2.6 ± 1.5	0.0	0.0
120	0.0	61.6 ± 4.3	61.4 ±8.2
113	57.3 ±4.6	22.8 ±3.7	22.8 ± 7.1
81	15.7 ±3.4	0.3 ± 0.5	0.0
74	0.8 ± 0.8	0.3 ± 0.5	0.0
72	0.0	0.0	0.0
33	0.0	0.0	0.0
React.[a]	9.3 ±0.8	5.5 ±0.5	2.7 ± 0.4

Results are reported as a function of the internal energy and the semi-empirical method. The products are identified by their masses. [a] Percentage of total reactivity.

Table A.12: Probability of forming the different products for the fragmentation of the L-Cysteine sulfate anion as obtained by collisional simulations at the collision energy of 5 eV in the center of mass framework using M_B1 as initial structure.

m/z	PM6-D	PM7	RM1
136	2.9 ± 1.0	0.0	0.0
120	1.8 ± 0.8	74.7 ± 2.2	86.8 ±3.3
113	51.3 ±3.0	14.0 ±1.8	0.0
81	0.7 ±0.5	0.3 ± 0.3	0.0
74	0.8 ± 0.8	0.3 ± 0.5	0.0
72	0.0	0.0	0.0
33	1.8 ± 0.8	3.8 ± 1.0	0.0
React.[a]	16.8 ±0.9	18.2 ±0.8	6.1 ± 0.6

Results are reported as a function of the internal energy and the semi-empirical method. The products are identified by their masses. [a] Percentage of total reactivity.

Table A.13: Probability of forming the different products for the fragmentation of the L-Cysteine sulfate anion as obtained by PM6-D thermal simulations at the internal energy of 87 kcal/mol using M_B2 as initial structure.

136	0.0
120	0.0
113	39.3 ± 3.1
81	6.9 ± 1.6
74	6.5 ± 1.6
72	0.4 ± 0.4
33	0.4 ± 0.4

Results are reported as a function of the internal energy and the semi-empirical method. The products are identified by their masses. [a] Percentage of total reactivity.

Appendix B

Analysis code

In figure B.1 it is represented schematically the code that we recently developed to automatically analyse the simulation results. The blue and green rectangles represent different subroutines, while white rectangles inputs and output files. Magenta ellipses represent non-automated procedures.

The analysis code is based on graph theory: the different ion structures are identified by different matrices (adjacency matrices, see next section) that can thus be compared.

1. For collisional simulations, it is possible to study first the energy transfer distribution for all the trajectories.

2. Another possibility, in order to obtain a different distribution for non-reactive and reactive trajectories is to use first the trajectory classification code, which, by comparing the matrices, differentiate between non-reactive, isomerization and reactive trajectories (in order to study energetic properties non-reactive and isomerization trajectories are generally analysed together).

3. Using the partition energy code, is then possible to calculate separately the rotational, vibrational and translational energetic contribution (using the rigid body approximation) for non-reactive trajectories.

4. Isomers can be classified obtaining isomeric distribution. Note that this procedure is not fully automatized. In the case, for example, of SO_3H, we will obtain three different isomers, i.e. one for each O-H bond possible. In order to improve this analysis graph isomorphism [132] should be implemented in future.

5. For collisional trajectories, it is possible to use the fragmentation-time routine in order to calculate the collision time, the fragmentation-time and thus to differentiate between

shattering and non-shattering trajectories. For both collisional and thermal trajectories the fragmentation time code can also be used to obtain a fragmentation time distribution.

6. Once the reactive trajectories are identified, it is possible to identify the fragments along time and obtain the fragmentation mechanism. However, this procedure has not still been automated. On the other hand, the fragments are identified at the end of the trajectories and it is possible to determine the trajectories final pathways. The mass of each fragment is also calculated in this step.

7. The fragments formed in all the trajectories can be counted (fragments are identified by their masses) and it is possible to obtain the mass distribution.

8. If the charge is calculated for each fragment it is then possible to obtain the theoretical mass spectrum, i.e. the m/z distribution. This last passage is not trivial. It is necessary to find and test a cheap (computationally) electronic density localization method. Moreover, the same fragment, depending on the specific path, the activation method (thermal or CID) or the kind of fragmentation (statistical or non-statistical) can have a different electronic density. The consequence is that the calculation must be performed on each fragment. Another problem is represented by "close" fragments, i.e. fragments that at the end of the trajectory are not far enough from each other, which makes more difficult the electronic density localization.

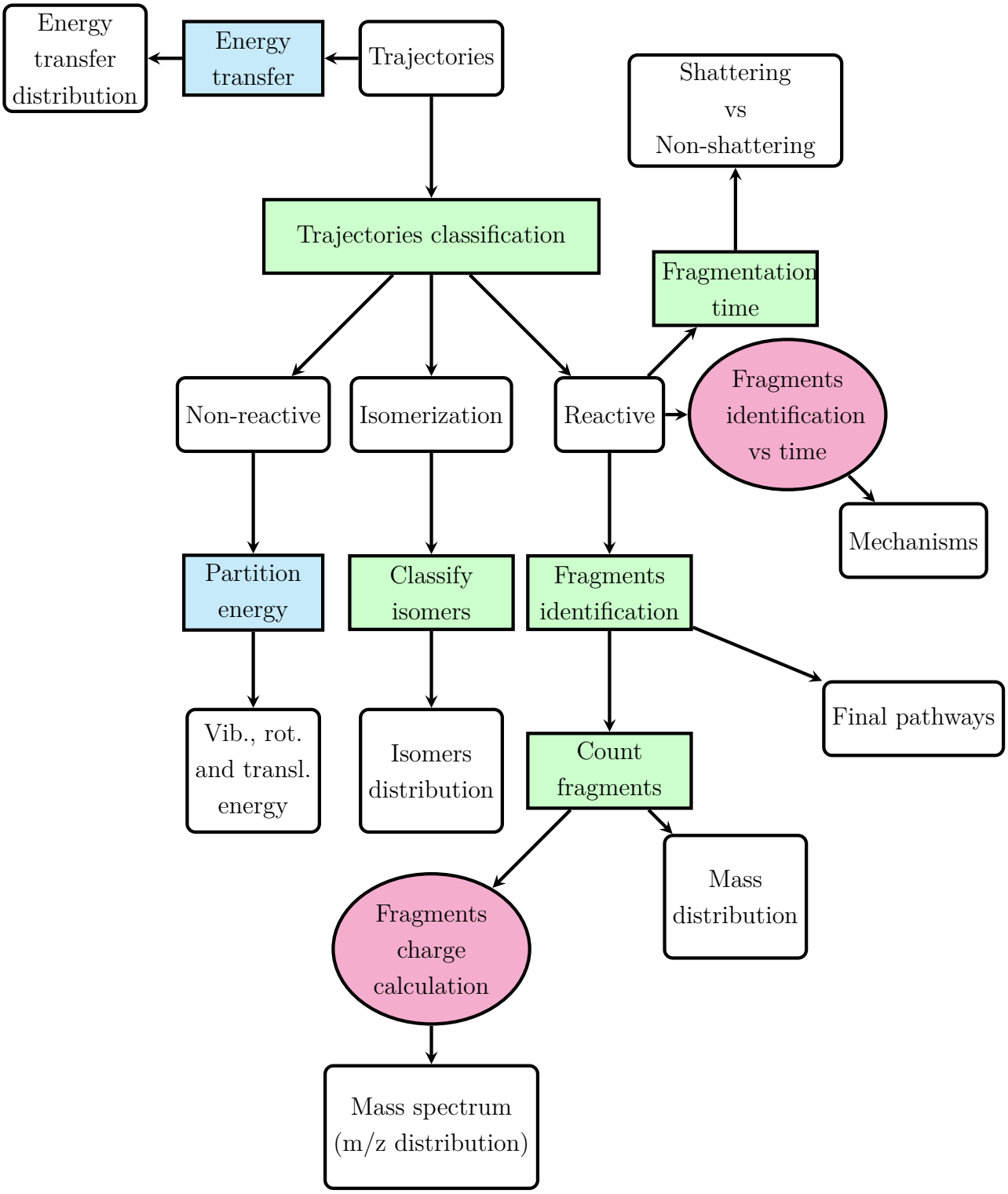


Figure B.1: Analysis code scheme

Adjacency matrix and trajectories classification

The analysis code is based on the comparison of the trajectories adjacency matrices. If the distance between two atoms A and B is less than the A-B bond criterion value (chosen from the sum of covalent radii augmented by a factor in order to consent bond elongation), it will be assigned the value of 1 to the AB and BA elements of the matrix, otherwise it will be assigned 0. The adjacency matrix can thus be built as it follows:

```
Allocate( A(1:nat,1:nat) )  !(A= adjacency matrix,
                                !nat=ion number of atoms)
      !Adjency matrix initialization
      Do i=1,nat
         Do j=1,nat
            A(i,j) = 0
         End do
      End do

      Do i=1,nat
         Do j=1,nat
            dx=(x(i)-x(j))^2
            dy=(y(i)-y(j))^2
            dz=(z(i)-z(j))^2
            dist=sqrt(dx+dy+dz)
            crit=(r_at(i)+r_at(j))*1.264
            if(i.ne.j.AND.dist.le.crit) then
               A(i,j)=1
            end if
         End do
      End do
```

The ion adjacency matrix is compared with the same matrix of the final "nsteps_check" steps of the trajectory. If at the end of the trajectory the two matrices are the same the trajectory is classified as non-reactive. In the case in which the two matrices are different but (using the subroutine "frag_assignemet") only one fragment is detected the trajectory is classified as "isomer". If more than two fragments are found a second criterion is used: the

ion must remain broken for a minimum number of steps ("nsteps_check"). An extract of the procedure is the following:

```
!****isomer or non-reactive trajectory**
if (ANY (A_0 .ne. A_f)) then
    call frag_assignement(N_ion, A_f, FRAG_f, nbfrag_f)

!****isomer trajectory*****************
    if (( nbfrag_f==1 ) .OR. ((bond_broken.eqv..true.)&
    & .AND. (cc.lt.nsteps_check))) then
       fp = 21
       open(unit=fp, action='write', position='append',&
       & file='isomers.dat')

...

!****reactive trajectory***************
    elseif ((bond_broken.eqv..true.) .AND. (cc==nsteps_check))&
    & then
       fp = 22
       open(unit=fp, action='write', position='append', &
       & file='react.dat')

...

    end if
End if
```

In the end, we obtained lists with non-reactive trajectories, isomers and reactive trajectories, where the last one contains also the formula and mass of each fragment.

Fragmentation time: shattering *vs* non-shattering

This code can be used directly only on the reactive trajectories (of collisional simulations), which have been classified using the previous code (see previous section). In order to know if the trajectory is shattering or not, we have first to calculate the collision time:

```
      Read(2,*) nat
      Do while ((IO==0) .AND. (collision.eqv..false.))
       If (j.ge.1) then
          Read(2,*)
       end if
       Read (2,*) !step
       Do i=1, nat
         Read(2,*) atom, x(i), y(i), z(i)
       End do
       j=j+1 !I am counting the steps
       if (j.ge.2) then
          Dist_proj_ion_check=Dist_proj_ion
       end if

       Call Calc_Dist_proj_ion(N_proj, nat , x,y,z, &
       & Dist_proj_ion)
       If (j .ge. 2) then

           if (Dist_proj_ion .gt. Dist_proj_ion_check) then
               collision=.true.
               coll_step=j-1
           end if
       end if
      End do
```

More complicate is the calculation of the fragmentation time: we have to use the "frag_assignement" subroutine to understand if the ion has fragmented. The trajectory must be analysed until the ion remains broken for at least "nsteps_check" steps. Afterword, we can discriminate between shattering and non-shattering trajectories:

```
d_step=check_step-coll_step !check_step is the fragmentation time
      if (d_step.le.threshold_shatt) then
         shatt=.true.
      else
         shatt=.false.
      end if
```

Appendix C

Shattering in the L-Cysteine-sulfate fragmentation

We analysed shattering for the collision simulations reported in chapter 6 about the fragmentation of L-Cysteine-sulfate anion at the collision energies of 23 and 46 eV. Results are summarized in table C.1.

Table C.1: Probability of Shattering (%)

Method	E_{rel} = 23 eV	E_{rel} = 46 eV
PM6-D	0.0	0.0
RM1-D	0.0	18.9
PM7	2.6	0.0

Calculation of RRKM rate constants for proteins

In order to calculate the RRKM rate constant for proteins (a tripeptide for the moment) I implemented two Whitten-Rabinovitch (WR) [68] modified algorithms [70, 71] in the code to calculate rate constants developed by W. L. Hase. Here it is shown an extract of the code to calculate JSUM (the sum of the states) using WR or the two modified versions (the code is in fortran77):

```
!Option for semi-classical thecnique
!     JSUM_OP_WR=0 STANDART SEMICLASSICAL TECHNIQUE(W-R METHOD)
!     JSUM_OP_WR=-1 W-R METHOD MODIFIED BY JULIA LASKIN
```

```
!      JSUM_OP_WR=-2 W-R METHOD MODIFIED BY SUN

       If(JDEN.EQ.0).OR.(JSUM.EQ.0))THEN
         If((JDEN_OP_W-R.EQ.-2).OR.(JSUM_OP_W-R.EQ.-2))THEN
            READ(5,*) N_AA
            if(N_AA.GT.N_AA_MAX) THEN
               Stop
            End if
            Read(5,57)(M_AA(I),I=1,N_AA)
            !M_AA are the peptide amino acids
            Call AA_C(N_AA_MAX, N_AA, M_AA, C0, C1, C2, C3)
         End if
       End if

       If(JSUM_OP_W-R.EQ.0)THEN
         WW=1.0/(5.0*EPM+2.73*EPM**0.5+3.51)
       Else if(JSUM_OP_W-R.EQ.-1)THEN
         WW=1.0/(1.783*EPM**1.4135+6.192*EPM**0.6209+3.265)
       Else if(JSUM_OP_W-R.EQ.-2)THEN
         WW=1.0/(C3*EPM**1.5+C2*EPM+C1*EPM**0.5+C0)
       End if
       Continue

!Subroutine to calculate the W-R parameters as done
!by M. Sun et all.
SUBROUTINE AA_C(N_AA_MAX, N_AA, M_AA, C0, C1, C2, C3)
!This subroutine associates to each amino acid
!M_AA of a peptide four parameters C0, C1, C2, C3
!and calculate the main values C0, C1, C2, C3
!
       IMPLICIT DOUBLE PRECISION(A-H,O-Z)
       CHARACTER*1 M_AA(N_AA_MAX), M_AAP(20)
       DIMENSION C0_P(20), C1_P(20), C2_P(20), C3_P(20)
```

```
      C0=0.0
      C1=0.0
      C2=0.0
      C3=0.0

      Do J=1, 20
        Read(10,*) M_AAP(J), C3_P(J), C2_P(J), C1_P(J), C0_P(J)
      End do

      Do I=1, N_AA
         Do J=1, 20
            If (M_AA(I).eq.M_AAP(J))THEN
               C0=C0+C0_P(J)
               C1=C1+C1_P(J)
               C2=C2+C2_P(J)
               C3=C3+C3_P(J)
            End if
         End do
      End do

      C0=C0/N_AA
      C1=C1/N_AA
      C2=C2/N_AA
      C3=C3/N_AA

      Return
      End
```

In order to test the two algorithms and the implementation in the code we performed a preliminary study to calculate the RRKM rate constant for pathway 2 of TIK$(H^+)_2$, which results are shown in figure C.1 and C.2.

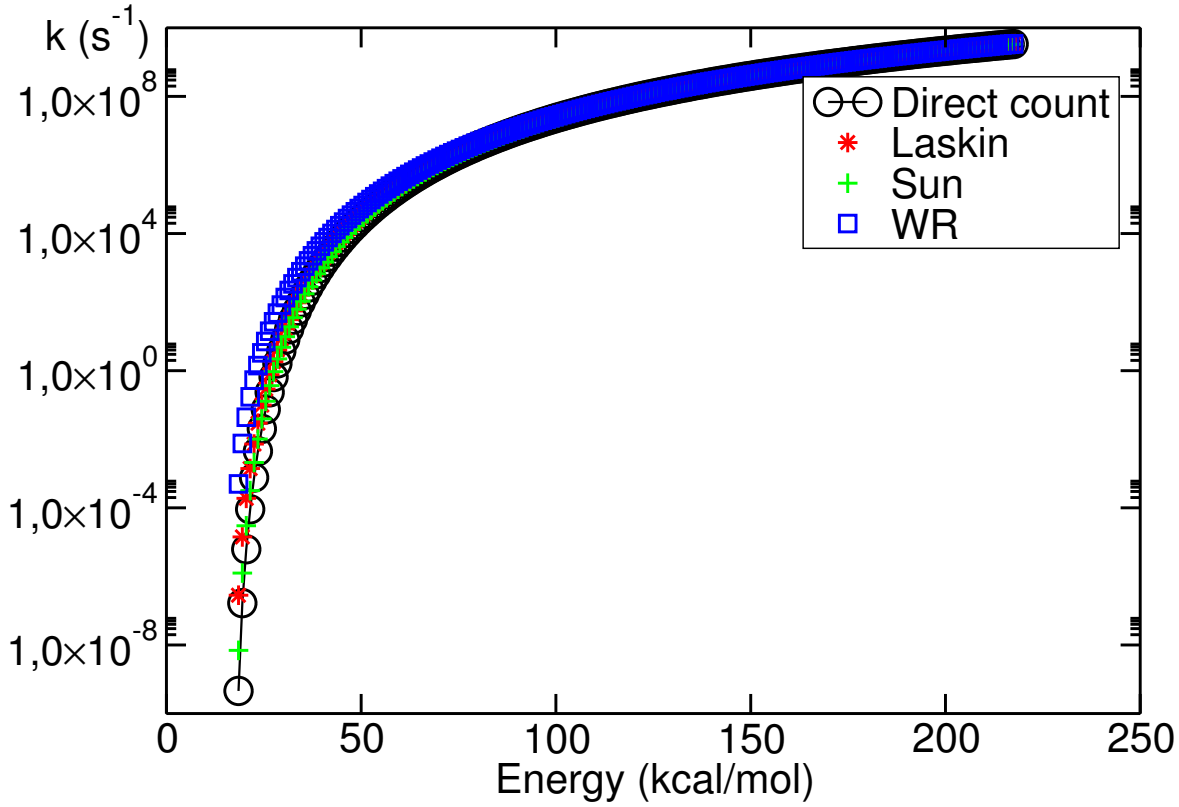


Figure C.1: Rate constant *versus* the internal energy as obtained using the Beyer-Swinehart direct count, the original WR algorithm and the two WR improved algorithm modified by Sun et al. [70] and by Laskin [71].

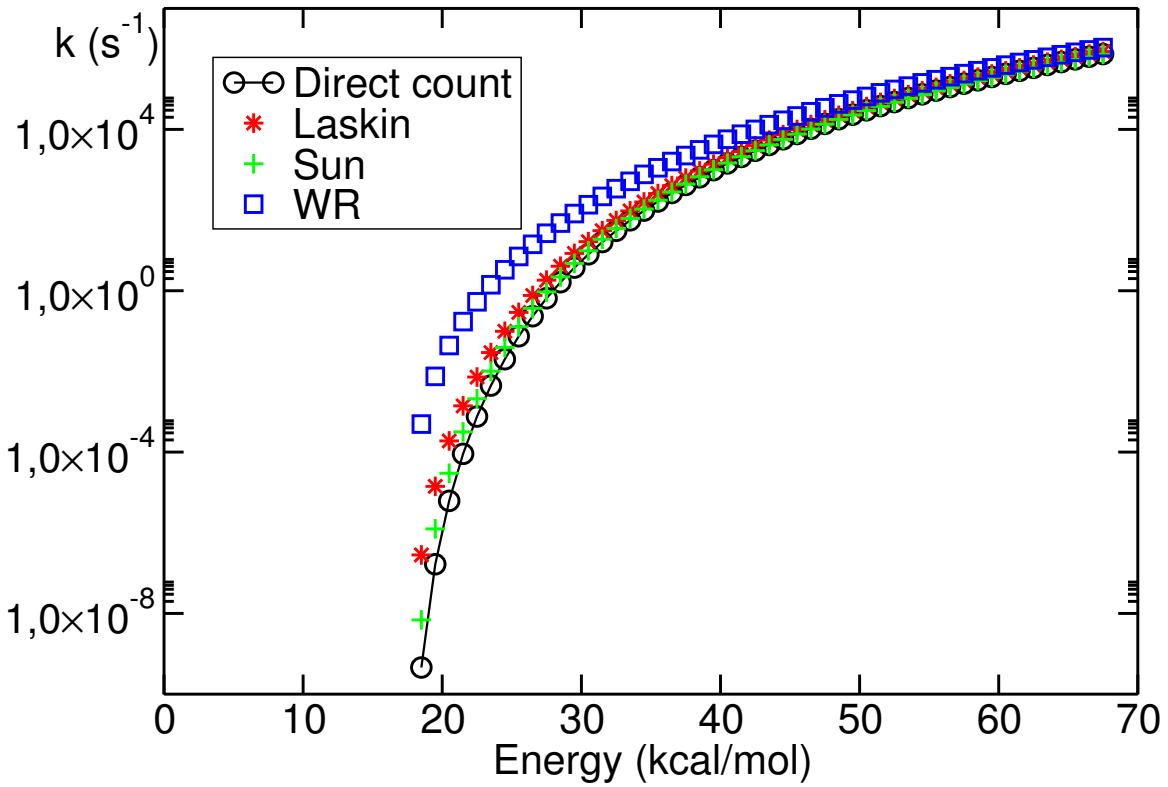


Figure C.2: Zoom at low energies of the rate constant *versus* the internal energy as obtained using the Beyer-Swinehart direct count, the WR algorithm and the two WR improved algorithm modified by Sun et al. [70] and by Laskin [71].

List of Figures

64.260

Bibliography

[1] J. Griffiths. A brief history of mass spectrometry. *Anal. Chem.*, 80:5678–5683, 2008.

[2] E. de Hoffmann and V. Stroobant. *Mass Spectrometry. Principles and Applications.* . Wiley, Chichester, West Sussex, England, 2007.

[3] J. S. McIndoe W. Henderson. *Mass Spectrometry of Inorganic and Organometallic Compounds: Tools - Techniques - Tips.* . John Wiley & Sons, Ltd., 2005.

[4] M. Akesson J. Smedsgaard S.G. Villas-Boas, S. Mas and J. Nielsen. Mass spectrometry in metabolome analysis. *Mass Spectrom. Rev.*, 24:613–646, 2005.

[5] T. Baer and W. L. Hase. Unimolecular reaction dynamics. theory and experiments. *J. Chem. Educ.*, 75:1098, 1998.

[6] D. Scuderi, E. Bodo, B. Chiavarino, S. Fornarini, and M. E. Crestoni. Amino-acids oxidation: a combined study of cysteine oxo-forms by IRMPD spectroscopy and simulation. *Chem. Eur. J.*, 22:17239–17250, 2016.

[7] R. Spezia K. Song. *Theoretical Mass Spectrometry. Tracing Ions with Classical Trajectories.* . De Gruyter, 2018.

[8] M. Yáñez J-Y. Salpin J. Tortajada. C. Trujillo, O. Mó. Gas-phase reactions between Thi-Urea and Ca^{2+}: new evidence for the formation of $[(NH_3)]^{2+}$ and other doubly charged species. *ChemPhysChem*, 8:1330–1337, 2007.

[9] B. Paizs, G. Lendvay, K. Vekey, and S Suhai. Formation of b_2^+ ions from protonated peptides: an ab initio study. *Rapid Commun. Mass Spectrom.*, 13:525–533, 1999.

[10] M. Paranjothy, Z. Homayoon X. Ma, G. L. Barnes, and W. L. Hase. Direct chemical dynamics simulations. *J. Am. Chem. Soc.*, 139:3570–3590, 2017.

[11] P. de Sainte Claire, G. H. Peslherbe, and W. L. Hase. Energy transfer dynamics in the collision-induced dissociation of Al_6 and Al_{13} clusters. *J. Phys. Chem.*, 99:8147–8161, 1995.

[12] O. Meroueh and W. L. Hase. Collisional activation of small peptides. *J. Phys. Chem. A*, 103:3981–3990, 1999.

[13] O. Meroueh and W. L. Hase. Energy transfer pathways in the collisional activation of peptides. *Int. J. Mass Spectrom.*, 201:233–244, 1999.

[14] O. Meroueh, Y. Wang, and W. L. Hase. Direct dynamics simulations of collision- and surface-induced dissociation of n-protonated glycine. shattering fragmentation. *J. Phys. Chem. A*, 106:9983–9992, 2002.

[15] E. Martinez-Nunez, A. Fernandez-Ramos, S.A. Vazquez, J. M. C. Marques, M. Xue, and W. L. Hase. Quasiclassical dynamics simulation of the collision-induced dissociation of $Cr(CO)_6^+$ with Xe. *J. Chem. Phys.*, 123:154311, 2005.

[16] A. M. Mendoza L. D. Aveline-Wolf K. R. Jonscher K. G. Pierce W. M. Old H. T. Cheung H. T. Cheung S. Russel J. L. Wattawa; et al. K. A. Resing, K. Meyer-Arendt. Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics. *Anal. Chem.*, 76:3556–53568, 2004.

[17] A. G. Harrison, A. B. Young, C. Bleiholder, S. Suhai, and B. Paizs. Scrambling of sequence information in collision-induced dissociation of peptides. *J. Am. Chem. Soc.*, 128:10364–10365, 2006.

[18] A. G. Harrison. To b or not to b: The ongoing saga of peptide b ions. *Mass Spectrom. Rev.*, 28:640–654, 2009.

[19] B. Paizs and S. Suhai. Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.*, 24:508–548, 2005.

[20] P. B. Armentrout and A. L. Heaton. Thermodynamics and mechanisms of protonated diglycine decomposition: A computational study. *J. Am. Soc. Mass Spectrom.*, 23:621–631, 2012.

[21] O. NØrregaard Jensen. Modification-specific proteomics : characterization of post-translational modifications by mass spectrometry. *Curr. Op. Chem. Biol.*, 8:33–41, 2004.

[22] E. Rossich Molina, D. Ortiz, J.-Y. Salpin, and R. Spezia. Elucidating collision induced dissociation products and reaction mechanisms of protonated uracil by coupling chemical dynamics simulations with tandem mass spectrometry experiments. *J. Mass Spectrom.*, 50:1340–1351, 2015.

[23] E. Rossich Molina, A. Eizaguirre, V. Haldys, D. Urban, G. Doisneau, Y. Bourdreux, J.-M.Beau, J.-Y. Salpin, and R.Spezia. Characterization of protonated model disaccharides from tandem mass spectrometry and chemical dynamics simulations. *ChemPhysChem*, page DOI: 10.1002/cphc.201700202., 2017.

[24] R. Spezia, S. B. Lee, A. Cho, and K. Song. Collision induced dissociation mechanisms of protonated penta- and octa-glycine as revealed by chemical dynamics simulations. *Int. J. Mass Spectrom.*, 392:125, 2015.

[25] D. Ortiz, P. Martin-Gago, A. Riera, K. Song, J.-Y. Salpin, and R. Spezia. Gas-phase collision induced dissociation mechanisms of peptides. theoretical and experimental study of n-formylalanylamide fragmentation. *Int. J. Mass Spectrom.*, 335:33–44, 2013.

[26] S. Kolakkandy, A. K. Paul, S. Pratihar, S. C. Kohale, G. L. Barnes, H. Wang, and W. L. Hase. Energy and temperature dependent dissociation of the Na^+(benzene)1,2 clusters. importance of anharmonicity. *J. Chem. Phys.*, 142:044306, 2015.

[27] E. Rossich Molina, J.-Y.Salpin, R.Spezia, and E. Martinez-Nunez. On the gas phase fragmentation of protonated uracil: a statistical perspective. *Phys. Chem. Chem. Phys.*, 18:14980–14990, 2016.

[28] A. Martin-Somer, R. Spezia, and M. Yanez. Gas-phase reactivity of $[Ca(formamide)]^{2+}$ complex: an example of different dynamical behaviours. *Phil. Trans. A*, A 375:20160196, 2017.

[29] A. Martin-Somer, M.-P. Gaigeot, M. Yanez, and R.Spezia. RRKM study and DFT assessment on gas-phase fragmentation of Formamide-M_2^+ (M=Ca, Sr). *Phys. Chem. Chem. Phys.*, 16:14813–14825, 2014.

[30] M.-P. Gaigeot W.L. Hase R. Spezia, J.-Y. Salpin and K. Song. Protonated urea collision-induced dissociation. comparison of experiments and chemical dynamics simulations. *J. Phys. Chem. A*, 113:13853–13862, 2009.

[31] Y. Jeanvoine, M.-P. Gaigeot, W.L. Hase, K. Song, and R. Spezia. Collision induced dissociation of protonated urea with a diatomic projectile : Effects on energy transfer and reactivity via chemical dynamics simulations. *Int. J. Mass Spectrom.*, 308:289–298, 2011.

[32] K. J. Song Y. Wang, W. L. Hase. Direct dynamics study of n-protonated diglycine surface-induced dissociation. influence of collision energy. *J. Am. Soc. Mass Spectrom.*, 14:1402–1412, 2003.

[33] O. Meroueh and W. L. Hase. Effect of surface stiffness on the efficiency of surface-induced dissociation. *Phys. Chem. Chem. Phys.*, 3:2306–2314, 2001.

[34] J. B. Fenn, M .Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246:64–71, 1989.

[35] J. V. Iribarne and B. A. Thomson. On the evaporation of small ions from charged droplets. *J. Chem. Phys.*, 64:2287–2294, 1976.

[36] Lars Konermann, Elias Ahadi, Antony D. Rodriguez, and Siavash Vahidi. Unraveling the mechanism of electrospray ionization. *Anal. Chem.*, 85:2–9, 2013.

[37] P. Kebarle and U. H. Verkerk. Electrospray: from ions in solution to ions in the gas phase, what we know now. *Mass Spectrom. Rev.*, 28:898–917, 2009.

[38] R. Aebersold and M. Mann. Mass spectrometry-based proteomics. *Nature*, 422:198–207, 2003.

[39] X. Han, A. Aslanian, and J. R. Yates III. Mass spectrometry for proteomics. *Curr. Op. Chem. Biol.*, 12:483–490, 2008.

[40] B. Bogdanov and R. D. Smith. Proteomics by fticr mass spectrometry: top down and bottom up. *Mass Spectrom. Rev.*, 24:168–200, 2005.

[41] D. M. Creasy D. N. Perkins, D. J. Pappin and J. S. Cottrell. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20:3551–3567, 1999.

[42] P. Roepstorff and J. Fohlman. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.*, 11:601, 1984.

[43] I. A. Papayannopoulos. The interpretation of collision-induced dissociation tandem mass spectra of peptides. *Mass Spectrom. Rev.*, 14:49–73, 1995.

[44] Brittany R. Perkins, Julia Chamot-Rooke, Sung Hwan Yoon, Ashley C. Gucinski, Arpad Somogyi, and Vicki H. Wysocki. Evidence of diketopiperazine and oxazolone structures for ha b_2^+ ion. *J. Am. Chem. Soc.*, 131:17528–17529, 2009.

[45] J. Oomens, S. Young, S. Molesworth, and M. van Stipdnok. Spectroscopic evidence for an oxazolone structure of the b_2 fragment ion from protonated tri-alanine. *J. Am. Soc. Mass Spectrom.*, 20:334–339, 2009.

[46] A. R. Dongré, J. Jones, A. Somogyi, and V. H. Wysocki. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *J. Am. Chem. Soc.*, 118:8365–8374, 1996.

[47] R. Boyd and A. Somogyi. The mobile proton hypothesis in fragmentation of protonated peptides: A perspective. *J. Am. Soc. Mass Spectrom.*, 21:1275–1278, 2010.

[48] N. C. Polfer, J. Oomens, S. Suhai, and B. Paizs. Infrared spectroscopy and theoretical studies on gas-phase protonated leu-enkephalin and its fragments: Direct experimental evidence for the mobile proton. *J. Am. Chem. Soc.*, 129:5887–5897, 2007.

[49] G. Lendvay S. Suhai IP. Csonka, B. Paizs. Proton mobility in protonated peptides: A joint molecular orbial and RRKM study. *Rapid Commun. Mass Spectrom.*, 14:417–431, 2000.

[50] G. Lendvay S. Suhai IP. Csonka, B. Paizs. Proton mobility and main fragmentation pathways of protonated lysilclycine. *Rapid Commun. Mass Spectrom.*, 15:1457–1472, 2001.

[51] E. Martinez-Nunez. An automated method to find transition states using chemical dynamics simulations. *J. Comput. Chem.*, 36:222–234, 2015.

[52] E. Martinez-Nunez. An automated transition state search using classical trajectories initialized at multiple minima. *Phys. Chem. Chem. Phys.*, 17:14912–14921, 2015.

[53] J. M. Bowman and P. L. Houston. Theories and simulations of roaming. *Chem. Soc. Rev.*, 46:7615,7624, 2017.

[54] Z. C. Kramer B. K. Carpenter-G. S. Ezra S. C. Farantos F. A. L. Mauguiere, P. Collins and S. Wigging. Roaming: A phase-space perspective. *Anny. Rev, Phys, Chem.*, 23:1,26, 2017.

[55] H. C. Ramsperger O. K. Rice. Theories of unimolecular gas reactions at low pressures.

[56] L. S. Kassel. Studies in homogeneous gas reactions.

[57] H. Eyring. The activated complex in chemical reactions. *J. Chem. Phys.*, 3:107–115, 1935.

[58] M. G. Evans and M. Polanyi. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc.*, 31:875–894, 1935.

[59] R. A. Marcus and O. K. Rice. The kinetics of the recombination of methyl radical and iodine atoms. *J. Phys. Colloid Chem.*, 55:894, 1951.

[60] R. A. Marcus. Unimolecular dissociations and free radical recombination reactions. *J. Chem. Phys.*, 20:359, 1952.

[61] F. A. L. Mauguière, P. Collins, Z. C. Kramer, B. K. Carpenter, G. S. Ezra, S. C. Farantos, and S. Wigging.

[62] S. Arrhenius. On the reaction velocity of the inversion of cane sugar by acids. *Zeitschrift für Physikalische Chemie*, 4:226, 1889.

[63] C. N. Hinshelwood. Quasi-Unimolecular Reactions. The Decomposition of Diethyl Ether in the Gaseous State. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 114:84–97, 1927.

[64] G. N. LeWis and J. E. Mayer. A disproof of the radiation theory of chemical activation. *Chemistry*, 13:623–625, 1927.

[65] J. Perrin. Matière et lumière: essai de synthèse de la mécanique chimique. *Ann. Phys.*, 11:1–5, 1919.

[66] E. Wigner. The transition state method. *Trans. Faraday Soc.*, 34:29–41, 1938.

[67] J. O. E. Hirschfelder and E. Wigner. Some quantummechanical considerations in the theory of reactions involving an activation energy. *J. Chem. Phys.*, 7:616, 1939.

[68] G. Z. Whitten and B. S. Rabinovitch. Accurate and facile approximation for vibrational energy-level sums. *J. Chem. Phys.*, 38:2466–2473, 1963.

[69] T. Beyer and D. F. Swinehart. Algorithm 448: Number of multiply-restricted partitions. *Commun. ACM*, 16:1973, 379.

[70] M. Sun, J. H. Moon, and M. S. Kim. Improved Whitten-Rabinovitch approximation for the Rice-Ramsperger-Kassel-Marcus calculation of unimolecular reaction rate constants for proteins. *J. Phys. Chem.*, 111:2747–2751, 2007.

[71] J. Laskin. Private conversation.

[72] D. L. Bunker. Monte Carlo calculation of triatomic dissociation rates. i. N_2O. *J. Chem. Phys.*, 37:393–403, 1962.

[73] D. Marx and J. Hutter. *Ab Initio Molecular Dynamics*. . Cambridge University Press, Cambridge, 2009.

[74] L. Verlet. Computer "experiments" on classical fluids. iv. transport properties and time-correlation functions of the Lennard-Jones liquid near its triple point. *Phys. Rev. A*, 7:1690–1700, 1973.

[75] L. Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159:98–103, 1967.

[76] W. L. Hase G. H. Peslherbe, H. Wang. Monte Carlo sampling for classical trajectory simulations. *Adv. Chem. Phys.*, 105:171–201, 1999.

[77] W. L. Hase and D. G. Buckowski. Monte Carlo sampling of a microcanonical ensemble of classical harmonic oscillators. *Chem. Phys. Lett.*, 74:284–287, 1980.

[78] R. Spezia, A. Martin-Somer, V. Macaluso, Z. Homayoon, S. Pratihar, and W.L. Hase. Unimolecular dissociation of peptides: Statistical vs non-statistical fragmentation mechanisms and time scales. *Faraday Discuss.*, 195:599–618, 2016.

[79] A. Martin-Somer, J. Martens, J. Grzetic, W. L. Hase, J. Oomens, and R. Spezia. Unimolecular fragmentation of deprotonated diproline $[Pro_2\text{-}H]^-$ studied by chemical dynamics simulations and IRMPD spectroscopy. *J. Phys. Chem. A*, 122:2612–2625, 2018.

[80] Michael J. S. Dewar and Walter Thiel. Ground states of molecules. 38. the mndo method. approximations and parameters. *J. Am. Chem. Soc.*, 99(15):4899–4907, 1977.

[81] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. Development and use of quantum mechanical molecular models. 76. am1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, 107:3902, 1985.

[82] G. B. Rocha, R. Oliveira Freire, A. Mayall Simas*, and J. J. P. Stewart. Rm1: a reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.*, 27:1101–1111, 2006.

[83] J. J. P. Stewart. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.*, 10:209–220, 1989.

[84] J. J. P. Stewart. Optimization of parameters for semiempirical methods. v. modification of nddo approximations and application to 70 elements. *J. Mol. Model.*, 13:1173–1213, 2007.

[85] J. J. P. Stewart. Optimization of parameters for semiempirical methods vi: more modifications to the nddo approximations and re-optimization of parameters. *J. Mol. Model.*, 19:1–32, 2013.

[86] Jan Rezac, Jindrich Fanfrlik, Dennis Salahub, and Pavel Hobza. Semiempirical quantum chemical pm6 method augmented by dispersion and h-bonding correction terms reliably describes various types of noncovalent complexes. *J. Chem. Theory Comput.*, 5(7):1749–1760, 2009.

[87] S. Ehrlich S. Grimme, J. Antony and H. Kriegh. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.*, 132:154104, 2010.

[88] R. Spezia, A. Cimas, M.-P. Gaigeot, J.-Y. Salpin, K. Song, and W.L. Hase. Collision induced dissociation of doubly-charged ions: Coulomb explosion vs neutral loss in $[Ca(urea)]_2^+$ gas phase unimolecular reactivity via chemical dynamics simulations. *Phys. Chem. Chem. Phys.*, 14:11724–11736, 2012.

[89] J. Liu, B. W. Uselman, J. M. Boyle, and Scott L. Anderson. The effects of collision energy, vibrational mode, and vibrational angular momentum on energy transfer and dissociation in NO_2^+- rare gas collisions: An experimental and trajectory study. *J. Chem. Phys.*, 125:133115, 2006.

[90] J. Liu, K. Song, W. L. Hase, and S. L. Anderson. Direct dynamics study of energy transfer and collision-induced dissociation: Effects of impact energy, geometry, and reactant vibrational mode in H_2CO^+-Ne collisions. *J. Chem. Phys.*, 119:3040–3050, 2003.

[91] D. Ortiz, J.-Y. Salpin, K. Song, and R. Spezia. Galactose 6-sulfate collision induced dissociation using qm+mm chemical dynamics simulations and ESI-MS/MS experiments. *Int. J. Mass Spectrom.*, 358:25–35, 2014.

[92] H. Bednarski, K. Sohlberg, M. Domanski, J. Weszka, G. Adamus, M. Kowalczuk, and V. Cozan. A combined theoretical and experimental study of mechanisms of fragmentation active for phb oligomers in negative-ion mode multistage mass spectrometry. *Int. J. Mass Spectrom.*, 304:15–24, 2011.

[93] R. Spezia, J. Martens, J. Oomens, and K. Song. Collision-induced dissociation pathways of protonated Gly_2NH_2 and Gly_3NH_2 in the short time-scale limit by chemical dynamics and ion spectroscopy. *Int. J. Mass Spectrom.*, 388:40–52, 2015.

[94] D. J. Douglas. Applications of collision dynamics in quadrupol mass spectrometry. *J. Am. Soc. Mass Spectrom.*, 9:101–113, 1998.

[95] Z. Homayoon, V. Macaluso, A. Martin-Somer, M. C. N. B. Muniz, I. Borges Jr., W. L. Hase, and R. Spezia. Chemical dynamics simulations of CID of peptide ions: comparisons between $TIK(H^+)_2$ and $TLK(H^+)_2$ fragmentation dynamics, and with thermal simulations. *Phys. Chem. Chem. Phys.*, 5:20, 2018.

[96] G. Lee, E. Park, H. Chung, Y. Jeanvoine, K. Song, and R. Spezia. Gas phase fragmentation mechanisms of protonated testosterone as revealed by chemical dynamics simulations. *Int. J. Mass Spectrom.*, 407:40–50, 2016.

[97] D. L. Bunker and W. L. Hase. On non-RRKM unimolecular kinetics: Molecules in general and CH_3NC in particular. *J. Chem. Phys.*, 59:4621–4632, 1978.

[98] W. L. Hase, R. J. Duchovic, X. Hu, A. Komornicki, K. F. Lim, D.-H. Lu, G. H. Peslherbe, K. N. Swamy, S. R. V. Linde, A. Varandas, H. Wang, and R. J. Wolf. Venus. a general chemical dynamics computer program. *QCPE Bull.*, 16:671, 1996.

[99] J. J. P. Stewart, L. J. Fiedler, J. Zheng, I. Rossi, W. P. Hu, G. C. Lynch, Y. P. Liu, P. Zhang, Y. Chuang, J. Pu, J. Li, P. L. Fast, C. J. Cramer, J. Gao, and D. G. Truhlar. MOPAC-version 5.022mn, HTTP://OpenMOPAC.net.

[100] M. Thevis, S. Beuck, S. Höppner, A. Thomas, J. Held, M.Schäfer, J. Oomens, and W.Schänzer. Structure elucidation of the diagnostic product ion at m/z 97 derived from androst-4-en-one-based steroids by ESI-CID and IRMPD spectroscopy. *J. Am. Soc. Mass Spectrom.*, 23:537–546, 2012.

[101] B. J. Bythell, P. Maitre, and B. Paizs. Cyclization and rearrangement reactions of fragment ions of protonated peptides. *J. Am. Chem. Soc.*, 132:14766–14779, 2010.

[102] J. Grzetic and J. Oomens. Spectroscopic identification of cyclic imide b2-ions from peptides containing gln and asn residues. *J. Am. Soc. Mass Spectrom.*, 24:1228–1241, 2013.

[103] X. Chen, M. Tirado, J. D. Steill, J. Oomens, and N. C. Polfer. Cyclic peptide as reference system for b ion structural analysis in the gas phase. *J. Mass Spectrom.*, 46:1011–1015, 2011.

[104] X. Chen, L. Yu, J. D. Steill, J. Oomens, and N. C. Polfer. Effect of peptide fragment size on the propensity of cyclization in colliison-induced-dissociation: Oligoglycine $b_{(2)}$-$b_{(8)}$. *J. Am. Chem. Soc.*, 131:18272–18282, 2009.

[105] M. M. Savitski, M. L. Nielsen, and R. A. Zubarev. Side-chain losses in electron capture dissociation to improve peptide identification. *Anal. Chem.*, 79:2296–2302, 2007.

[106] X. Hu, W. L. Hase, and T. Pirraglia. Vectorization of the general Monte Carlo classical trajectory program venus. *J. Comput. Chem.*, 12:1014–1024, 1991.

[107] J. J. P. Stewart. MOPAC2012, 2012.

[108] A. Seiter C. Schlier. Integration of classical trajectories: a case study. *J. Phys. Chem. A*, 102:9399–9404, 1998.

[109] A. Seiter C. Schlier. High-order symplectic integration: an assessment. *Comput. Phys. Commun.*, 130:176–189, 2000.

[110] J. I. Steinfeld, J. S. Francisco, and W. L. Hase. *Chemical kinetics and dynamics.* . Prentice Hall, Upper Saddle River, N.J., 1999.

[111] J. Laskin and J. H. Futrell. Surface-induced dissociation of peptide ions: Kinetics and dynamic. *J. Am. Soc. Mass Spectrom.*, 14:1340–1347, 2003.

[112] S. Pratihar, G. L. Barnes, J. Laskin, and W. L. Hase. Perspective: Dynamics of protonated peptide ion collisions with organic surfaces. consonance of simulation and experiment. *J. Phys. Chem. Lett.*, 7:3142, 2016.

[113] S. A. Vàsquez E. Martinez-Nunez and J.M.C. Marques. Quasiclassical trajectory study of the collision-induced dissociation of CH_3SH^+ + Ar. *J. Chem. Phys.*, 121:2571–2577, 2004.

[114] M. T. Rodgers and P. B. Armentrout. Absolute binding energies of Lithium ions to short chain alcohols, $C_nH_{2n+2}O$, n = 1-4, determined by threshold collision-induced dissociation. *J. Phys. Chem. A*, 101:2614–2625, 1997.

[115] NIST Atomic Spectra Database(ver. 5.3), [Online]. Available: `http://physics.nist.gov/asd`.

[116] A. Kramida, Yu. Ralchenko, J. Reader, and and NIST ASD Team. NIST Atomic Spectra Database (ver. 5.3), [Online]. Available: `http://physics.nist.gov/asd` [2017, July 23]. National Institute of Standards and Technology, Gaithersburg, MD., 2015.

[117] W. L. Hase. Unimolecular and intramolecular dynamics. relationship to potential energy surface properties. *J. Phys. Chem.*, 90:365–374, 1986.

[118] T. Baer and W. L. Hase. *Unimolecular reaction dynamics: theory and experiments.* . Oxford University Press, New York, 1996.

[119] I. Komaromi, A. Somogyi, and V. H. Wysocki. Proton migration and its effect on the MS fragmentation of N-acetyl OMe proline: MS/MS experiments and ab initio and density functional calculations. *Int. J. Mass Spectrom.*, 241:315–322, 2004.

[120] B. Paizs and S Suhai. Theoretical study of the main fragmentation pathways for protonated glycylglycine. *Rapid Commun. Mass Spectrom.*, 15:651, 2001.

[121] K. A. Cox, S. J. Gaskell, M. Morris, and A. Whiting. Role of the site protonation in the low-energy decomposition of gas-phase peptide ions. *J. Am. Soc. Mass. Spectrom.*, 7:522–531, 1996.

[122] D. R. Rehder and C. R. Borges. Cysteine sulfenic acid as an intermediate in disulfide bond formation and nonenzymatic protein folding. *Biochemistry*, 49:7748–7755, 2010.

[123] L. Turell H. Botti B. A. Freeman S. Caraballal, B. Alvarez and R. Radi. Sulfenic acid in human serum albumin. *Amino Acids*, 32:543, 551, 2003.

[124] M.C. Kirk S. Barnes B.A. Freeman S. Caraballal, R. Radi and B. Alvarez. Sulfenic acid formation in human serum albumin by hydrogen peroxide and peroxynitrite. *Biochemistry*, 32:543, 551, 2003.

[125] T. C. Mallett J. Luba E.J. Crane-E. J. V. Charrier A Claiborne, J. I. Yeh and D. Parsonage. Protein-sulfenic acids: Diverse roles for an unlikely player in enzyme catalysis and redox regulation. *Biochemistry*, 38:155407, 15416, 1999.

[126] T. B. McMahon J. M. Ortega D. Scuderi J. Lemaire L. Mac Aleese, A. Simon and P. Maitre. Mid-ir spectroscopy of protonated leucine methyl esther performed with an ft-icr or an paul type ion-trap. *Int. J. Mass Spectrom.*, 249:14–20, 2006.

[127] T. B. McMahon J. M. Ortega D. Scuderi J. Lemaire L. Mac Aleese, A. Simon and P. Maitre. Gas-phase structure of a π-allyl-palladium complex: Efficient infrared spectroscopy in a 7 t Fourier Transform Mass Spectrometer. *J. Phys. Chem. A*, 111:13415–13424, 2007.

[128] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo,

R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09, Revision D.01, 2009.

[129] Fabien Brieuc, Yael Bronstein, Hichem Dammak, Philippe Depondt, Fabio Finocchi, and Marc Hayoun. Zero-point energy leakage in quantum thermal bath molecular dynamics simulations. *J. Chem. Theory Comput.*, 12:5688–5697, 2016.

[130] R. P. Feynman and A. R. Hibbs. *Quantum mechanics and path integrals*. McGraw-Hill, New York, 1965.

[131] Fabien Brieuc, Hichem Dammak, and Marc Hayoun. Quantum thermal bath for path integral molecular dynamics simulation. *J. Chem. Theory Comput.*, 12:1351–1359, 2016.

[132] Brendan D. McKay and Adolfo Piperno. Practical graph isomorphism, {II}. *Journal of Symbolic Computation*, 60(0):94 – 112, 2014.

Titre : Modélisation théorique de la dissociation induite par collision en phase gazeuse de biomolécules

Mots clés : Dynamique moléculaire, réactivité chimique, fragmentation, acides aminés, peptides, spectrométrie de masse

Résumé : Dans la présente thèse, nous rapportons l'étude de la dissociation induite par collision (CID) de biomolécules. La CID est une technique de spectrométrie de masse (MS) bien connue, dont le but est la dissociation d'ions par l'impact avec un gaz inerte. L'énergie de translation collisionnelle est convertie en énergie interne de l'ion qui peut ainsi se dissocier. La CID est donc une technique largement utilisée en MS qui permet d'identifier, ou de quantifier, une ou plusieurs espèces par la détection des fragments générés. La réactivité et la cinétique des réactions chimiques sont généralement étudiées théoriquement par la recherche des points stationnaires sur la coordonnée de réaction. Il est ainsi possible d'identifier le chemin d'énergie minimum de réaction ou la surface d'énergie potentielle (PES). Une autre possibilité est d'effectuer des simulations de dynamique chimique, qui permettent d'explorer la réactivité d'une espèce sans connaitre les produits, ce qui est un point crucial pour des molécules plus grosses. En plus, pour interpréter la MS il est important d'avoir une compréhension fondamentale de la dynamique de la fragmentation de l'ion, et des informations importantes peuvent être récupérées avec des simulations. Dans le présent travail, nous avons étudié et développé des modèles physiques pour étudier des biomolécules complexes et flexibles, comme les acides aminés et les peptides. Une fois que l'ion est excité par une seule collision, le transfert d'énergie peut être suivi d'une redistribution statistique interne de l'énergie vibrationnelle (IVR) de l'ion et des produits statistiques sont typiquement obtenus. D'autre part, la collision peut causer une localisation de l'énergie et une excitation rapide, donnant des produits différents de ceux observés après une IVR. En particulier, une situation limite est celle où l'ion se fragmente juste après la collision avec le gaz. Afin de récupérer ces fragmentations moins statistiques, il est important de modéliser la collision, ce qui peut être fait par une dynamique chimique de collision explicite. Cependant, ce type de simulation est limité dans le temps (~ 10-15 ps). La dynamique chimique par activation statistique interne (ou thermique) peut être utilisée pour obtenir une échelle de temps plus longue et une réactivité statistique. De plus, en observent le déclin de la population par rapport au temps, il est possible d'obtenir les constantes de vitesse globales et individuelles. Les deux modes d'activation ont été appliqués pour étudier la réactivité de l'anion di-proline, les deux tripeptides doublement chargés TIK(H+)2 et TLK(H+)2 et l'anion L-cystéine-sulfate. Pour l'étude de ce dernier système en particulier, nous avons utilisé les résultats de nos simulations pour interpréter des expériences faites avec différents montages expérimentaux.

Title : Theoretical modelling of gas phase collision induced dissociation of biomolecules

Keywords : Molecular dynamics, chemical reactivity, fragmentation, amino acids, peptides, mass spectrometry

Abstract : In the present thesis, we focus on the study of the collision induced dissociation (CID) of biomolecules. CID is a well known mass spectrometry (MS) fragmentation technique which aim is the dissociation of ions through the impact with an inert buffer gas. The collisional translational energy is converted in internal energy of the ion that can thus dissociate. CID is thus widely used in MS in order to identify or quantify one or more species through detection of the mass over charge ratio of the fragments products. Once the ion is excited by single-collision, the translation-to-vibration energy transfer can be followed by a statistical internal vibrational energy redistribution (IVR) of the ion and typical statistical products are obtained. On the other hand, the collision can cause localization of the energy and a fast excitation, giving different products than those observed after an IVR. In particular a limit situation is when the ion fragments right after the collision with the gas. In order to recover these (less or fully) non-statistical fragmentations it is important to

Reactivity and kinetics of chemical reactions are generally studied theoretically through the research of the stationary points along the reaction coordinate. It is thus, possible, to identify the reaction minimum energy path or potential energy surface (PES). Another possibility is to perform chemical dynamics simulations, which allow to explore the reactivity of one specie without the knowledge of the products, that is a crucial point for larger molecules. Moreover, to interpret MS it is important to have a fundamental understanding of the ion's fragmentation dynamics and important information can be recovered with simulations. In the present work, we have studied and developed physical models to address the study of complex and flexible biomolecules, like amino-acids and peptides.

model the collision, which can be done performing explicit collision chemical dynamics. However, this activation way in simulations is limited in the time-scale (~ 10-15 ps). Statistical internal energy (or thermal) activation chemical dynamics can be used to obtain longer time scale and statistical reactivity. Moreover, observing the population decay versus the time it is possible to obtain the global and single pathways rate constants. Both activation modes have been applied to study the reactivity of the di-proline anion, the two doubly charged tri-peptides TIK(H+)2 and TLK(H+)2 and the L-Cysteine sulphate anion. In particular for the study of this last system we used our understanding of simulations to interpret experiments done with different set-ups.