

A simple statistical approach for fragmentation studies of doubly ionized cytosine, thymine and uracil bases

Jean-Philippe Champeaux, P Çarçabal, Martine Sence, Patrick Moretto-Capelle, Pierre Cafarelli

To cite this version:

Jean-Philippe Champeaux, P Çarçabal, Martine Sence, Patrick Moretto-Capelle, Pierre Cafarelli. A simple statistical approach for fragmentation studies of doubly ionized cytosine, thymine and uracil bases. Journal of Physics B: Atomic, Molecular and Optical Physics, 2011, 44 (4), pp.45205. $10.1088/0953-4075/44/4/045205$. hal-00596621

HAL Id: hal-00596621 <https://hal.science/hal-00596621>

Submitted on 28 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A simple "statistical" approach for fragmentation studies of doubly ionized cytosine, thymine and uracil bases

J.-P. Champeaux1,2 , P. Çarçabal 3 , M.Sence 1,2 , P Moretto-Capelle 1,2 , P. Cafarelli 1,2

¹Université de Toulouse; UPS; Laboratoire Collisions Agrégats Réactivité, IRSAMC; F-31062 Toulouse, France

²CNRS; UMR 5589; F-31062 Toulouse, France

³Institut des Sciences Moléculaires d'Orsay - CNRS, Université Paris Sud, F-91405 Orsay, Cedex, France.

Email: jean-philippe.champeaux@irsamc.ups-tlse.fr

Abstract

A simple statistical model describing the dissociation of molecular dications into correlated fragment pairs has been developed. This model is based on a combinatory approach in which all possible fragments are enumerated and is refined by taking into account the initial structure of the parent molecule, considering the number of chemical bonds to be broken to give rise to the fragments. We show how this model can be used as a tool to help interpreting experimental results of coincidence experiments. It shows that dissociation of doubly ionized molecules upon 100 keV proton irradiation is dominated by statistical processes but it also enables an easy identification of the dissociation products originating from non-statistical processes requiring further investigation, possibly conveying information on the radiation-molecule interaction itself.

1. Introduction

Over the last decade, ionization and fragmentation of molecules of biological interest such as DNA and RNA bases [1-4] , or more complex systems such as nucleoside or clusters of biomolecules [5, 6] , submitted to high energy radiation (X or UV photons, ions, electrons), received a renewed interest. One of the motivations for such work is related to the advent of the wide-spread radio-therapy techniques, which prompted the need for a better understanding of the interaction between ionizing radiation and systems of biological relevance at the microscopic (molecular or atomic) scale. These studies focus on the initial physical stage (in the femto- or pico-second range) of the radiationmolecule interaction, and aim at identifying and characterizing its outcome, *i.e.* product ions and electrons that may lead, directly or indirectly, to tumor cell death.

The description of the ionization and fragmentation processes is a complex task. It has been investigated by Classical Trajectory Monte Carlo method [7, 8] and by state-of-the-art quantum mechanical calculations from the description of the initial physical process of electron(s) removal (i.e ionization) from valence molecular orbitals [9] to the molecular relaxation leading to dissociation [10] . Another approach, based on the analysis of fragments size distributions, has also been adopted, bringing evidence for strong similarities between fragmentation of a variety of finite size systems like atomic nuclei, molecular clusters and fullerenes [11-14] . For instance, percolation theory has been applied on hydrogen clusters [15] and to the analysis of C_{60} fragmentation in ion– C_{60} collisions [11, 16] , indicating a possible phase transition upon dissociation. In the case of the dissociation of biomolecules studied by mass spectrometric methods, rather complex mass spectra are observed and their interpretation can be challenging. Heinonen *et al.* [17] proposed a method based on a combinatorial approach in order to help the interpretation of molecular fragmentation.

We present here a simple statistical model based on fragments enumeration combined with bond breaking calculation. The mass spectra simulated with this model can be compared to experimental data and used as a tool to identify, among the many observed dissociation products of complex molecules, those resulting from specific (non-statistical) dissociation pathways. After describing the model, we show how it can be used to help the interpretation of our recent experimental results on the dissociation of cytosine, thymine and uracil (Figure 1), doubly ionized upon collisions with 100 keV protons.

2. Method

2.1. Experimental context

The model we have developed is a tool for the interpretation of experiments of multi-coincidence detection of fragments originating from gas phase bio-molecules doubly ionized by collisions with 100 keV protons. The experimental setup has been described in detail elsewhere [18] . In these experiments, coincidence detection relies on very high detection efficiency and very low detection rate. The density of the proton bunches impinging on the gaseous target molecules is kept very low, so that collision products are detected much less often than once per proton pulse cycle (typically, once every 100 pulses). Consequently, when one singly positive charged product (fragment or parent ion) is detected, it is considered to arise from a singly charged parent molecule. As a consequence, when two singly positive charged fragments are detected during one proton cycle, they are considered as a correlated fragment pair A^{+}/B^{+} originating from the same doubly ionized molecule, and not as two uncorrelated fragments issued from two singly charged molecules (low detection rate) or as two correlated fragments emitted from a triply ionized parent with a third ion product that eluded detection (high detection efficiency). Although we attempted to detect negatively charged products by operating the mass spectrometer in inverse mode, only positively charged products have been detected. Then, the correlated fragment pairs come from the following general reaction path:

$$
M + p^{+} \rightarrow M^{2+} + p^{+} + 2e^{-} \rightarrow A^{+} + B^{+} + {C}^{0+} + p^{+} + 2e^{-}
$$

The stars indicate that the fragments are not necessarily in their electronic ground state, and ${C}$ refers to the neutral undetected fragments.

This correlation treatment leads to experimental insights into the dissociation kinematics and in some cases, into the dissociation dynamics [18-20] .

2.2. Statistical model

The idea underlying our statistical approach is to consider that all possible pathways leading to any correlated A^{+}/B^{+} fragment pair can be explored by the system with an equal probability. This implies

that all possible initial states can be populated with the same probability, independently of the amount of energy deposited in the molecule upon the collision process. We also assume that all rearrangements of the molecule or the fragments during dissociation are possible and equally probable. These assumptions imply that parent molecules have enough internal energy after double ionization to overcome all possible potential barriers in the dissociation and rearrangement pathways. Experiments performed at different collision energy (in the range 25-100 keV) support this assumptions as they revealed similar fragmentation mass spectra, independently of the incident protons energy. Therefore, specific dissociation schemes and associated reaction rates, that are often considered in lower energy processes such as photodissociation [21], are not taken into account here. In this approximation, every atom from the parent molecule can be combined to form all possible correlated A^{+}/B^{+} fragment pairs without any initial geometrical or charge density consideration. The only constrains imposed in the model are the initial number of atoms in the parent molecule, the singly charged state of the fragments and the valence of the atoms involved in the fragments. In this context, all dissociation pathways leading to two correlated mono-cation A_i^+ and B_j^+ , from a doubly charged parent molecule M^{2+} , can be generalized into the two following cases:

- initial charge separation: $M^{2+} \rightarrow X^+ + \underline{X}^+ \rightarrow \{X A_i\}^0 + A_i^+ + \{\underline{X} B_j\}^0 + B_j^+$ (a)
- initial neutral(s) emission: $M^{2+} \to X^0 + \underline{X}^{2+} \to {X}^0 + {\underline{X}^2}^{4} \to {X}^0 + {\underline{X}^2}^{4} + A_i^+ + B_j^+$ (b)

where $\{\alpha\}^0$ represents one or several neutral species with a total mass of α , and X and X stand for all species that can be possibly composed of a combination of atoms of $M (M=X+X)$. For example in uracil $C_4N_2O_2H_4$ (M=112 amu), a possible X/X pair is 1/111 with 4 possible distinct H/(U-H) pairs for the first step of the dissociation following (a) or (b).

The valence of each atom is used to verify that X^+ and X^+ for case (a), and X^{2+} for case (b), are chemically possible. For example, in case (a), $X=H_3O^+$ is acceptable while $X=H_4^+$ is not. Similarly, $X^0 = C_4N_2O_2H_3$ and $\underline{X}^{2+} = H^{2+}$ will not be treated in case (b).

For each A_i^+ , we can compute and enumerate all the possible ways to produce A_i^+ / B_j^+ correlations following initial charge separation (a), or initial neutral emission (b), from the M^{2+} parent molecule, by cycling over all possible X/ \underline{X} pairs and, for each X/ \underline{X} , all possible B_j^+ (according to $M_{Ai} + M_{Bj} \leq M$ and atoms valence). Finally, by counting the number of pathways for each A_i^+ / B_j^+ pair we obtain, after normalization to unity, the probability distribution to produce a given A_i^+ cation correlated with all possible B_j^+ cations with respect to the initial equiprobability criteria postulated above. In the case of uracil with $A^{\dagger} = H^{\dagger}$, such enumeration yields a total number of dissociation pathways of nearly 1.4 million, the major one being the $H^+/28^+$ with more than 60000 ways giving rise to these correlated fragments, all redundancies being removed.

2.3. Structural extension of the statistical model

Since the statistical model assumes that all fragmentation and rearrangement pathways are equally probable, it is clear that it will largely overestimate some fragment pairs that are highly unlikely to occur for energetic, structural or physical reasons. For example, initial state separation (case (a)) is more probable than initial neutral evaporation (case (b)), simply because of coulombic explosion. Also, in spite of the very high degree of rearrangement required, the formation of fragments like *e.g.* H_3^+ or C_3^+ from uracil will have the same probability in this model as the formation of CH or OCNH fragments, made of concomitant atoms.

To circumvent the overestimation of fragmentation pathways involving unreasonably high degrees of rearrangement, one can consider locality criteria in the initial structure of the parent molecule by taking into account the number of chemical bonds that have to be broken to produce the fragments. To achieve this, a Monte Carlo algorithm, named *CM(n)*, has been developed. This code identifies all fragments obtained by breaking a given maximum number of chemical bonds in the parent molecule. For each event of the Monte Carlo procedure, the number of bonds to be broken is chosen randomly from 1 to n. The major difficulty is then to identify the connected parts constituting the fragments wich is achieved by a recursive procedure based on graph theory. Typically $10⁴$ events where generated to achieve convergence for cytosine, thymine and uracil, but the maximum number of events can be fixed by user. The number of bonds that are allowed to be broken also introduces, indirectly, the notion of excess energy available in the molecule after ionization. For instance, allowing a maximum number of 4 bond breakage (*CM(4)*) implicitly requires on average twice as much internal energy than allowing only 2 bond breakage (*CM(2)*). This algorithm can be coupled to quantum chemistry codes to obtain the optimized geometries and energies of the fragments but this aspect is not developed here. Analog algorithms, using more accurate bond energy estimates, have been devoted to mass spectrometry analysis like the FiD code [17] .

Using fragments found by *CM* to restrict the A_i and B_j possibilities will drastically decrease the number of explored pathways following cases (a) or (b). The drawback of the geometrical extension is that it allows fragments made of initially bound atoms only, and rearrangement is forbidden. It has the advantage of removing highly improbable fragments but, since some rearrangement can physically occur before or during dissociation, this structural constrain may eliminate some physical pathways.

3. Results and discussion

3.1. Energetic considerations

In proton-molecule collision experiments, two major aspects of the interaction are not controlled: the impact parameter (distance between the incident proton and the target molecule) and the molecular orientation. Consequently, the amount of energy effectively transferred by the proton to the molecule, and where this energy is deposited, are not controlled. The only experimental information on the amount of deposited energy can be obtained from the degree of ionization of the target molecules. In the case of DNA/RNA bases, 90% of all events are due to single ionization of molecules (1 cation detected) and the remaining 10% are nearly entirely due to the double ionization (2 correlated cations detected). Higher degrees of ionization exist although they remain marginal (<< 1%).

To evaluate the order of magnitude of the available excess energy after ionization, we have calculated the vertical single, double and triple ionization potentials at the optimized geometry of the neutral parent in its electronic ground state, for the two lowest spin multiplicity of the ions, at the B3LYP/6- 311G** level of theory with the Gaussian 98 software [22] . Results are given in Table 1.

Table 1: Vertical ionization potentials (in eV) for single (IP1), double (IP2) and triple (IP3) ionization of uracil, thymine and cytosine, calculated at the B3LYP/6-311G** level of theory. The numbers in parenthesis indicate the spin multiplicity of the ion. Italic values in the IP1 (2) column are

The minimum amount of energy required to produce dications lies around 25 eV, but the upper limit of the amount of energy deposited in the interaction is not easy to estimate due to the possible excitation of vibrational or electronic states in the ionized molecule. For 100 keV protons, the interaction is dominated by electronic processes and, using the Lindhart-Winter model, we have previously shown for adenine [2] that the number of events with an amount of deposited energy in the molecule greater than 100 eV falls drastically down to zero. Similar results have been obtained for all the bases treated here and the intensity of the high energy part of the spectrum follows an exponential law of the type I≈exp(-E/30), where E is expressed in eV. Nevertheless, since we detect some triple and higher ionization events, it means that the amount of excess energy in the doubly ionized molecules may reach several tens of eV. This relatively high excess of energy suggests that the systems may explore many pathways, including those involving potential energy barriers, as supposed by the initial assumptions of the statistical model.

3.2. Experimental fragmentation of dications

The 2D correlated time of flight spectrum obtained in 100keV proton collision with uracil is given in Figure 2. Characteristic masses are also indicated. Only one channel corresponding to two-body fragmentation and leading to the formation of cations of mass 43 amu (HNCO⁺) and 69 amu $(C_3NOH_3^+)$, is observed. The 2D TOF spectrum is dominated by all the other so-called 'correlation islands' corresponding to dissociation channels where two product ions and one or more neutrals are emitted.

Figure 2: 2D Correlated time of flight spectrum obtained from dissociation of Uracil dication. The grayed area highlights the 'slice' of the H^+ / B_j^+ correlations events.

For cytosine, thymine and uracil, the most intense fragmentation channels following double ionization is the emission of a proton correlated with another cation. Table 2 summarizes the measured branching ratio of the overall proton emission upon double ionization of these molecules, with respect to the total number of recorded events.

Table2: Measured branching ratios (%) of H⁺ emission from doubly ionized molecules. The number after the molecule name indicates the number of hydrogen atoms in the molecule.

Molecule	Branching Ratio
cytosine (5)	20.4 ± 0.2
thymine (6)	24.4 ± 0.2
uracil (4)	27.7 ± 0.2

One can extract partial integrated spectra by 'slicing' the 2D spectrum along one selected mass to obtain the mass spectrum of the fragments correlated to the selected one. Such partial integrated spectra of the fragments correlated to the H⁺ fragment $(A_i^+ = H^+)$ are given in Figure 3 for cytosine, thymine and uracil.

Figure 3: Partial integrated mass spectra of fragments correlated to H^+ emission following dissociation of cytosine, thymine and uracil dication. The black triangles indicate the mass of the parent molecule.

These spectra exhibit very similar patterns. We can notice that the most intense channels correspond to B⁺ fragments of mass between 20 and 50 amu and the signal intensity drastically decreases for heavier masses with a cutoff around 70 amu.

3.3. Calculated mass spectra for the proton emission channels from dications (A+=H+)

The simulated mass spectra of the fragments correlated to proton emission $(H⁺/B_j⁺)$ obtained by applying our statistical model and its structural extension are compared to the experimental spectra in Figure 4.

Figure 4: Comparison between calculated spectra associated to proton emission with experimental partial integrated spectra (black line) for cytosine, thymine and uracil. The pink spectra have been obtained with the statistical model alone, red triangles (resp, blue circles) are fragments distribution obtained by the structural extension with a maximum of 3 (resp. 4) bond breakage (Stat $+CM(3)$, resp. Stat+*CM(4)*). All spectra are normalized to unity. The black triangles indicate the mass of the parent molecule.

Qualitatively, the experimental spectra are relatively well reproduced by the statistical model. All the observed correlated fragments are found with a satisfactory intensity agreement especially for fragments of low masses, below 30 amu. At first sight, this suggests that most dissociation mechanisms involved in 100 keV proton induced dissociations are driven by statistics.

The cutoff in the H^+ correlated spectra observed around 70 amu is also well reproduced. It can be understood since the number of channels leading to an heavy fragment correlated to H^+ (*e.g.* (U- H^{\dagger}/H^{\dagger}) is necessarily lower than the number of possible ways to form H^{\dagger} correlated to intermediate mass fragments. The same conclusion is drawn for the H^{\dagger}/H^{\dagger} correlation, the number of ways to produce such correlation being very low.

In all cases, the statistical curve underestimates the number of events compared to the experiment for the low masses. This is expected since the statistical model overestimates the number of possible fragments by authorizing high degrees of molecular rearrangement, the global intensity of the distribution is then reduced by the normalization. This is slightly corrected if we impose locality criteria using $Stat+CM(3)$ (red triangle in figure 4) that reduces the number of possible fragments. It is clear that for each group of mass peaks, the structural extension eliminates fragments at both ends of the groups. The main reasons for this is that the low mass end must be dominated by fragments resulting from pathways with more bond cleavages than limited by *CM* and the high mass end must be mostly due to fragments obtained after recombination, that are excluded by *CM*. Evidently, using this same locality criteria but increasing the number of possible bound breaks to 4 (Stat+*CM(4)*) in the molecule improves the agreement with the experimental results, as shown in the case of uracil where masses omitted by Stat+*CM(3)* appear, like fragments of mass 35 amu as well as 65-67 amu in uracil.

Statistical dissociation can be considered as random dissociation of 'hot' molecules where the energy is distributed over highly excited vibrational states of the system and not much is to be learned about the molecule-radiation interaction from such processes. It is then clear that attention must be paid to features of the experimental spectra that are *not* reproduced by the statistical model or by Stat+*CM*, as these features must relate to *specific* mechanisms, and possibly to the quantum nature of the interaction. In Figure 4, one can easily identify three regions where both the statistical model and its structural extension fail to reproduce the intensity of the observed spectra. These correspond to correlated fragment pairs at $H^+/12^+$, $H^+/[35-40]^+$ and $H^+/[50-55]^+$. The underestimation of the carbon peak $(H^+/12^+)$ by the model is unexpected and cannot be interpreted on the basis of the models presented here, a contamination of the experiment due to the heating the oven can not be excluded. In the recent study of 100 keV proton dissociations of Halouracils [20] , it has been shown that fragments of mass 38-40 amu, correlated to the emission of the halogen atom (replacing the H atom in position 5 of Uracil) exhibit relative intensity trends upon halogen atom substitution, suggesting such specific dissociation pathways. The discrepancy between the experimental intensity and the number of pathways leading to these products, as predicted by the statistical approach presented here, may reveal quantum mechanisms increasing their probability.

Other correlated fragment pairs are predicted to exhibit significant intensities by the model but are not observed experimentally. This is the case for the $H^+/[55-60]^+$ fragment pairs for all molecules studied here. This suggests that these fragments are likely to further dissociate, leading to secondary products in the experimental spectra, or are not formed at all.

Finally, the mass peaks for the $H^+/16^+$ and $H^+/17^+$ fragment pairs of uracil²⁺ also show a striking disagreement with the intensities predicted by the models. These products (most likely, O^+ and OH^+) could be due to specific, highly efficient dissociation pathways from the uracil dication but we cannot exclude that they could also arise from residual water in the background vacuum of the chamber. Another example that the statistical model can be used to point out peculiars feature in experimental spectra.

3.4. Fragmentation of dications: emission of fragment with mass 43 (A+=43+)

The mass 43 amu fragment is another major component of the observed 2D correlated spectra [24] . Figure 5 shows the comparison between the experimental partial integrated mass spectrum associated to mass 43 and the (Stat+*CM(3)*) calculation for uracil.

Figure 5 : Comparison between the experimental partial integrated mass spectrum (black curve), the global stat distribution (pink curve), the Stat+*CM(3)* distribution (red triangles) and Stat+*CM(4)* (blue circles) of fragments correlated to the fragment cation of mass 43 in Uracil.

The agreement between the experimental spectrum and the spectrum obtained from our model is qualitatively satisfactory in the mass range 10-60 amu except for fragments of mass 28 and 40 amu. Also the model completely underestimates the $43^{+}/69^{+}$ channel, which is the most intense correlation island of the 2D correlated mass spectrum (figure 2).

This indicates that these dissociation channels $(28^+/43^+, 40^+/43^+, 43^+/69^+)$ must be driven by molecular mechanisms with very high cross sections that considerably increase their probability, with respect to their statistical weight. In the case of $43^{+}/69^{+}$, it can be understood since this dissociation channel corresponds to two-body dissociation that may occur on a very fast time scale, by simply breaking 2 chemical bonds, in a non-statistical regime.

4. Conclusion

A simple statistical approach based on the enumeration of the dissociation pathways, assuming equiprobability of all pathways has been developed to help interpreting the emission of correlated fragment pairs A^{+}/B^{+} observed in the dissociation of the pyrimidine DNA and RNA bases cytosine, thymine and uracil, after double ionization induced by 100 keV proton collisions.

The method gives a good agreement for low mass correlations and suggests that statistical dissociations are dominant for such high energy collision induced dissociations, where the amount of deposited energy is not controlled. Evidently, this approach cannot account properly for some particular fragmentation channels which may imply quantum effects that counter-balance their low statistical weight. For instance, in the case of the DNA/RNA bases investigated here, the fragment pair for which further investigation could be informative are : $H^+/12^+$, $H^+/35-40$ ⁺ and $H^+/50-60$ ⁺, as well as $28^{+}/43^{+}$ and $[38^{+}-42^{+}/43^{+}]$ for uracil.

The statistical approach can then be seen as a powerful tool to identify the products of non-statistical mechanisms that have to be investigated further, possibly conveying information on the radiationmolecule interaction itself. This can prove very helpful to guide and narrow down sophisticated and computationally expensive studies using *ab-initio* quantum chemistry or TDDFT approaches [10, 25] .

References:

- [1] Schlathölter, T. et al. 2006 Physica Scripta **73** C113-C7
- [2] Moretto-Capelle, P. et al. 2007 Journal of Chemical Physics **127** 234311
- [3] Moretto-Capelle, P. and Le Padellec, A. 2006 Physical Review A **74** 062705
- [4] Alvarado, Fresia et al. 2007 The Journal of Chemical Physics **127** 034301
- [5] Schlathölter,T. et al. 2006 ChemPhysChem **7** 2339-45
- [6] Manil, B. et al. 2006 Photonic, Electronic and Atomic Collisions 599
- [7] Lekadir, H. et al. 2009 Physical Review A **79** 062710
- [8] Abbas, I. et al. 2008 Physics in Medicine and Biology **53** N41-N51
- [9] Dal Cappello, C. et al. 2008 Physical Review A **78** 042702
- [10] Gaigeot, M. P. et al. 2010 Mutation Research-Reviews in Mutation Research **704** 45-53
- [11] Lebrun, T. et al. 1994 Physical Review Letters **72** 3965
- [12] Bonasera, A. and Schulte, J., in Proceedings *Similarities and differences between atomic nuclei and clusters. Toward a unified development of cluster science*, 1998), Vol. 416, p. 463.
- [13] Campi, X. 1986 J. Phys. A **19** L917
- [14] Bauer, W. 1988 Phys. Rev. C **38** 1297
- [15] Gobet, F. et al. 2001 Phys. Rev. A **63** 033202
- [16] Rentenier, A. et al. 2005 J. Phys. B **38** 789
- [17] Heinonen, M. et al. 2008 Rapid Communications in Mass Spectrometry **22** 3043
- [18] Cafarelli, P. et al. 2008 Radiation Damage in Biomolecular Systems **1080** 71
- [19] Eland, J.H.D., Wort, F.S. and Royds, R.N. 1986 J. Electron Spectr. and Related Phenomena **41** 297
- [20] Champeaux, J. P. et al. 2010 Physical Chemistry Chemical Physics **12** 5454
- [21] Wang, Q. Q. et al. 2008 Journal of Chemical Physics **129** 204302
- [22] Frisch, a M. J. et al. Gaussian, Inc., Pittsburgh PA, 2001
- [23] Hush, N. S. and Cheung, A. S. 1975 Chemical Physics Letters **34** 11
- [24] Le Padellec, A. et al. 2008 Radiation Damage in Biomolecular Systems **101** 12007
- [25] Adoui, L. et al. 2009 Journal of Physics B-Atomic Molecular and Optical Physics **42** 075101