Article

Mass Spectra Resulting from Collision Processes

Felix Duensing * [ID] and Paul Scheier [ID]

Institut für Ionenphysik und Angewandte Physik, Universität Innsbruck, Technikerstr. 25,
6020 Innsbruck, Austria; paul.scheier@uibk.ac.at
* Correspondence: felix.duensing@uibk.ac.at

Abstract: A new database and viewer for mass spectra resulting from collision processes is presented that follows the standards of the *Virtual Atomic and Molecular Data Centre* (VAMDC). A focus was placed on machine read and write access, as well as ease of use. In a browser-based viewer, mass spectra and all parameters related to a given measurement can be shown. The program additionally enables a direct comparison between two mass spectra, either by plotting them on top of each other or their difference to identify subtle variations in the data.

1. Introduction

In recent years, open data has become more important, and journals push authors to make their data available with the FAIR Guiding Principles (findability, accessibility, interoperability, and reuse of digital assets [1]). A few years ago, a consortium of scientists started an open data infrastructure project called the *Virtual Atomic and Molecular Data Centre* (VAMDC) [2]. The idea was to provide a standard for requesting data, a standard format of returned data, and an infrastructure which can but does not have to be used. Scientists can create a node, i.e., a database which provides data of certain atomic or molecular radiative and collisional processes. These nodes are available online, either directly or via the VAMDC portal where all available nodes are registered. Progress of the VAMDC and its nodes was published recently [3]. In 2013, two VAMDC nodes were created at our institute, providing data on radiation damage to biomolecules [4] and relative cross sections for dissociative electron attachment measurements [5]. The entries in this database are ion efficiency curves where the ion yield of a specific product ion is measured as function of the electron energy. So far, mass filters (quadrupole or sector field instrument) have been utilized and for each ion of interest, an electron energy scan has been recorded. By replacing these mass filters with time of flight instruments, ion efficiency curves can be measured for all ions in one electron energy scan. The resulting data file is a three-dimensional array, i.e., ion yield as a function of mass per charge and electron energy and does not fit into the existing database.

Besides free electron attachment, several other topics in the field of atomic and molecular physics are covered in our institute utilizing mass analyzers, predominantly time-of-flight mass spectrometers that have been constantly modified [6–8]. Three of these instruments are equipped with a helium droplet source and designed to investigate electronic excitation of cold ions [9,10] as well as cluster chemistry and physics at low temperatures [11–13]. Mass spectra obtained from such helium droplet experiments critically depend on several parameters that often are scanned to obtain optimum conditions for specific experiments. This requires the comparison of mass spectra taken at different conditions and sometimes recorded a long time ago.

In the case of our oldest instrument [14], we have recorded more than 6000 mass spectra since 2011, including simple high-resolution mass spectra as well as two-dimensional

spectra where mass spectra are recorded while a certain parameter (e.g., collision energy) is scanned. Particularly large files are produced from action spectroscopy measurements [9,10,15] which are typically recorded for several hours up to a few days. The resulting *HDF5* files easily exceed several GB of data. Another setup is investigation of ion-surface-collisions by secondary ion mass spectrometry (SIMS) [16].

In order to keep track of this amount of data and to be able to find a particular measurement, even years after it was recorded, it is important to have a tool at hand that supports the user with this task. Keywords and parameters help to narrow down the number of potential hits but still, several mass spectra might have to be plotted to find the right one.

Mass spectrometers are not only used in many areas of research but also in commercial operations where mass spectra databases have already been established for common applications like electrospray ionization liquid gas chromatography for bio-molecules or electron impact ionization of gases [17–19]. While functionality and features of these databases are excellent they address different applications. There, investigations were done on a single substance, i.e., showing the fragmentation pattern during ionization or in a collision induced dissociation. Most of these mass spectra show a rather limited number of ions and the data are often presented as simple bar graphs. In contrast, mass spectra from cluster experiments in general and resulting from doped helium droplet experiments [11,12] in particular often contain an enormous number of ions, resulting from the fact that besides pristine He_n^+, helium-solvated dopant cluster ions $He_nA_m^+$ with n of up to a few hundred are also present [13].

There is a trend of journals, particularly high-impact ones, and also funding agencies requiring to provide open access to all data and evaluations leading to a publication, including raw data. With this work, we want to present a new database which takes the special conditions of cluster production and classification into account and is compatible with the standards defined by the VAMDC consortium. An important goal was to provide simple access to visual presentation of data without the need to install additional software. Furthermore, this viewer enables quick access to mass spectra that were recorded in the past and it enables an efficient comparison of two mass spectra when looking for optimum settings of the experiment.

2. Implementation

2.1. Software

The software is free and open-source. It is designed to be used on a common Linux server. Incoming requests are managed by the webserver software nginx which serves static content and is a reverse proxy to gunicorn, a Python web server gateway interface (WSGI) managing HTTP requests. Gunicorn serves a Python application by using the web framework django which follows the model–template–views (MTV) architectural pattern. Data is stored in a relational MariaDB database. The system is connected to an external storage system to create backups of database dumps on a regular basis. The django application is based on the VAMDC NodeSoftware (https://github.com/VAMDC/NodeSoftware (accessed on 29 April 2022). One table in the relational databases represents one class of the application.

2.2. Adding Datasets

A dataset consists of:

- a mass spectrum with one column of mass-to-charge values in atomic mass units and one column of ion yields in arbitrary units; Both columns must have equal length;
- one or more authors responsible for the data;
- the used experiment, which should be described in a peer-reviewed article;
- a set of parameters explaining the experimental conditions.

Registered users can upload datasets by using a form. Mass spectral data can be uploaded either as a csv file or as json formatted data.

2.3. Accessing Data Sets

Data gets served by a .XSAMS file (Section 2.3.1) or by a graphical user interface (GUI, Section 2.3.2).

2.3.1. XML Schema for Atomic Molecular and Solid Data (XSAMS)

To fulfill the VAMDC requirements, all datasets can be downloaded as XSAMS files. These files can contain one or more datasets as well as all associated information like the used experiment, experimental parameters, and authors. These files can also be obtained directly through the VAMDC portal wherein a query, e.g., for a molecule, returns all datasets of all VAMDC-related databases, and thus also this database. A benefit is that the XSAMS file structure is already common in some communities.

2.3.2. Graphical User Interface (GUI)

Datasets can be viewed in a GUI accessible by common web browsers. A table-structured view shows all entries with the possibility to search and filter for certain properties, like the measured sample or the used experiment. In a first step, one can select one (or more) datasets to be viewed in the mass spectra viewer (Figure 1). The datasets can be displayed alone or on top of each other for comparison. Common plotting features are available, for example the possibility to zoom in both axes or toggling between a linear or logarithmic y-scale for the ion yield. In order to overcome differences in the detection efficiencies of the mass analyzers, the ion yields of the individual datasets can be individually scaled (Figure 2). When exactly two sets are displayed, the difference of both sets can be shown which becomes useful in terms of background subtraction. As the mass-to-charge is stored as discrete values but two datasets can have different mass axis values, an interpolation of the mass-to-charge axes is performed. The lower half of the screen shows the mass spectra-related metadata about authors, experiments, and parameters.
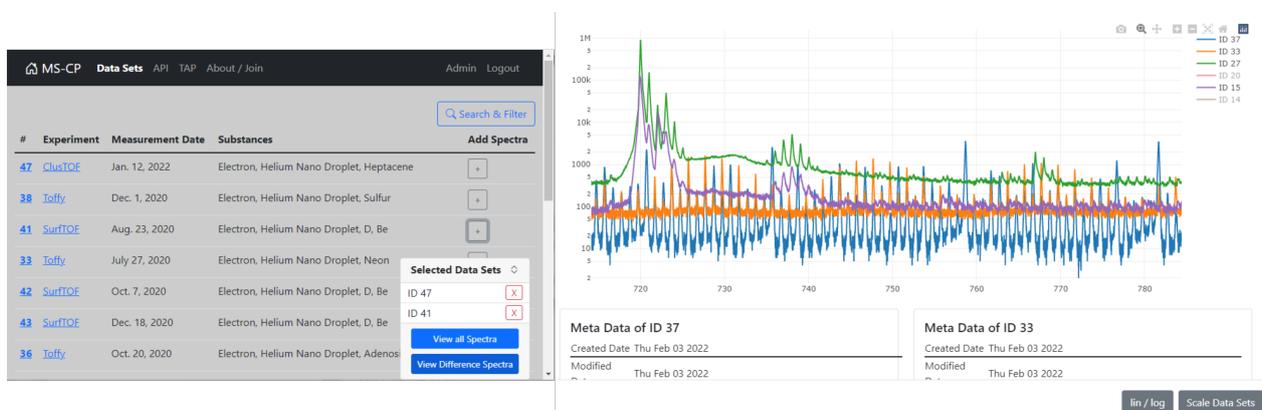
Figure 1. Screenshots of the GUI. Left: List of available datasets with the option to search and filter the entries. Selected datasets can be viewed on top of each other or as a differential mass spectrum (see Figure 2). Right: data representation. The upper part contains the visual representation of the selected datasets, the lower part shows metadata of the selected datasets. At the bottom, two buttons allow for the toggling between a linear and a logarithmic ion yield axis as well as the scaling of the individual mass spectra.

2.4. REST-API Endpoint

A representational state transfer (REST) at an application programming interface (API) is available for adding and serving data. Independent of the language used by an authenticated user, this endpoint can be used to programmatically add json-formatted data. This is especially sensible if many datasets (for example an already existing database) are to be migrated to this database.

This endpoint can serve json formatted data as well, also to non-authenticated users.

Figure 2. Detailed view of the data representation. Left: comparison of two mass spectra plotted on top of each other with a linear y-axis. Right: difference between the same two mass spectra, with a scaling factor of 0.95 applied to the first mass spectrum.

2.5. Access

- The database including the list of measurements, the data viewer, and additional information can be reached at https://ideadb.uibk.ac.at/mscp (accessed on 29 April 2022). In the "Data Sets" section one can add measurements to the list of interest and start the visualization.
- The API is accessible at https://ideadb.uibk.ac.at/mscp/api/ (accessed on 29 April 2022).
- The code repository including is hosted on https://github.com/nano-bio/NodeSoftware/tree/cluster-db/nodes/mscp (accessed on 29 April 2022).

3. Outlook

In the first step, a high degree of complexity was dispensed with in favour of ease of use. In the next step, however, more possibilities for a more detailed description of data are to be offered. Finding and filtering of datasets will also be improved. Mass spectra are not the only results of collision processes, so that other data, such as the dependence of the ion yield of a specific ion on a scan parameter, such as the electron energy in the case of cross sections or the laser wavelength in case of ion spectroscopy, can also be provided in the future.

4. Conclusions

Out of the need to publish our mass spectra according to the open data guidelines, an open database was created. Depending on the user's selection, the output can either be a .xsams or .json file, or an interactive representation in a GUI. With .xsams a growing standard for atomic and molecular data was chosen, whereas .json can be accessed from a REST-API for machine–machine communication. The interactive GUI data viewer provides fast insight into data sets in a user-friendly manner.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]
2. Dubernet, M.L.; Antony, B.K.; Ba, Y.A.; Babikov, Y.L.; Bartschat, K.; Boudon, V.; Braams, B.J.; Chung, H.K.; Daniel, F.; Delahaye, F.; et al. The virtual atomic and molecular data centre (VAMDC) consortium. *J. Phys. B At. Mol. Opt. Phys.* **2016**, *49*, 074003. [CrossRef]
3. Albert, D.; Antony, B.; Ba, Y.A.; Babikov, Y.L.; Bollard, P.; Boudon, V.; Delahaye, F.; Del Zanna, G.; Dimitrijević, M.S.; Drouin, B.J.; et al. A Decade with VAMDC: Results and Ambitions. *Atoms* **2020**, *8*, 76. [CrossRef]
4. Denifl, S.; Garcia, G.; Huber, B.A.; Marinković, B.P.; Mason, N.; Postler, J.; Rabus, H.; Rixon, G.; Solov'yov, A.V.; Suraud, E.; et al. Radiation damage of biomolecules (RADAM) database development: current status. *J. Phys. Conf. Ser.* **2013**, *438*, 012016. [CrossRef]
5. Ribar, A.; Fink, K.; Probst, M.; Huber, S.E.; Feketeová, L.; Denifl, S. Isomer Selectivity in Low-Energy Electron Attachment to Nitroimidazoles. *Chem. Eur. J.* **2017**, *23*, 12892–12899. [CrossRef] [PubMed]
6. Tiefenthaler, L.; Ameixa, J.; Martini, P.; Albertini, S.; Ballauf, L.; Zankl, M.; Goulart, M.; Laimer, F.; von Haeften, K.; Zappa, F.; et al. An intense source for cold cluster ions of a specific composition. *Rev. Sci. Instrum.* **2020**, *91*, 033315. [CrossRef] [PubMed]
7. Albertini, S.; Bergmeister, S.; Laimer, F.; Martini, P.; Gruber, E.; Zappa, F.; Ončák, M.; Scheier, P.; Echt, O. SF6+: Stabilizing Transient Ions in Helium Nanodroplets. *J. Phys. Chem. Lett.* **2021**, *12*, 4112–4117. [CrossRef]
8. Martini, P.; Albertini, S.; Laimer, F.; Meyer, M.; Gatchell, M.; Echt, O.; Zappa, F.; Scheier, P. Splashing of Large Helium Nanodroplets upon Surface Collisions. *Phys. Rev. Lett.* **2021**, *127*, 263401. [CrossRef] [PubMed]
9. Kuhn, M.; Renzler, M.; Postler, J.; Ralser, S.; Spieler, S.; Simpson, M.; Linnartz, H.; Tielens, A.G.G.M.; Cami, J.; Mauracher, A.; et al. Atomically resolved phase transition of fullerene cations solvated in helium droplets. *Nat. Commun.* **2016**, *7*, 13550. [CrossRef] [PubMed]
10. Meyer, M.; Martini, P.; Schiller, A.; Zappa, F.; Krasnokutski, S.A.; Scheier, P. Electronic Spectroscopy of Anthracene Cations and Protonated Anthracene in the Search for Carriers of Diffuse Interstellar Bands. *Astrophys. J.* **2021**, *913*, 136. [CrossRef]
11. Mauracher, A.; Echt, O.; Ellis, A.; Yang, S.; Bohme, D.; Postler, J.; Kaiser, A.; Denifl, S.; Scheier, P. Cold physics and chemistry: Collisions, ionization and reactions inside helium nanodroplets close to zero K. *Phys. Rep.* **2018**, *751*, 1–90. [CrossRef]
12. Albertini, S.; Gruber, E.; Zappa, F.; Krasnokutski, S.; Laimer, F.; Scheier, P. Chemistry and physics of dopants embedded in helium droplets. *Mass Spectrom. Rev.* **2021**, *Early View*. [CrossRef]
13. González-Lezana, T.; Echt, O.; Gatchell, M.; Bartolomei, M.; Campos-Martínez, J.; Scheier, P. Solvation of ions in helium. *Int. Rev. Phys. Chem.* **2020**, *39*, 465–516. [CrossRef]
14. Schöbel, H.; Bartl, P.; Leidlmair, C.; Denifl, S.; Echt, O.; Märk, T.D.; Scheier, P. High-resolution mass spectrometric study of pure helium droplets, and droplets doped with krypton. *Eur. Phys. J. D* **2011**, *63*, 209–214. [CrossRef]
15. Gatchell, M.; Martini, P.; Laimer, F.; Goulart, M.; Calvo, F.; Scheier, P. Spectroscopy of corannulene cations in helium nanodroplets. *Faraday Discuss.* **2019**, *217*, 276–289. [CrossRef] [PubMed]
16. Ballauf, L.; Duensing, F.; Hechenberger, F.; Scheier, P. A high sensitivity, high resolution tandem mass spectrometer to research low-energy, reactive ion–surface interactions. *Rev. Sci. Instrum.* **2020**, *91*, 065101. [CrossRef] [PubMed]
17. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837. [CrossRef]
18. Schulze, T.; Neumann, S.; Rainer, J.; Schymanski, E. MassBank Europe—Progress Report 2021. 2021. Available online: https://zenodo.org/record/5504121#.Yo8VpP5BxPY (accessed on 29 April 2022).
19. MassBank of North America. Available online: https://mona.fiehnlab.ucdavis.edu (accessed on 29 April 2022).