Contrib. Astron. Obs. Skalnaté Pleso **55/2**, 81 – 87, (2025)
https://doi.org/10.31577/caosp.2025.55.2.81

MolD, EMol and ACol atomic and molecular databases for astrophysics: current stage and new directions of development

V.A. Srećković[1], **B.P. Marinković**[1], **Lj.M. Ignjatović**[1] **and V.Vujčić**[2]

[1] *University of Belgrade, Institute of Physics Belgrade, PO Box 57, (E-mail: vlada@ipb.ac.rs, ljuba@ipb.ac.rs)*

[2] *Astronomical Observatory, Volgina 7, 11060 Belgrade, Serbia (E-mail: veljko@aob.rs)*

Received: December 03, 2024; Accepted: December 16, 2024

Abstract. In this contribution, we provide the current state of the MolD, EMol and ACol databases, hosted within the Virtual Atomic and Molecular Data Center (VAMDC) and the Serbian Virtual Observatory (SerVO). MolD, EMol and ACol are atomic and molecular databases dedicated to modeling various laboratory plasmas, and stellar atmospheres. Moreover, in this contribution, we present plans for the further directions of development like machine learning, software updates, etc.

Key words: databases – atoms – molecules – collisions – atmospheres – modeling – ML

1. Introduction

Data volumes have grown rapidly from terabytes to tens or hundreds of petabytes, and will continue to grow at an accelerating rate in the age of huge sky surveys and large telescopes (Koekemoer et al., 2007; Tucker et al., 2006). Simultaneously, utilizing new technologies, vast amounts of data and information (generated by potent supercomputer simulations) are dispersed throughout the global network architecture and archived in networks (Ivezić et al., 2019). Thus, research methodologies, algorithms, and procedures, as well as the state of data-oriented science, become crucial. One of the main driving forces behind the astronomical Virtual Observatories is still the operational processing and scientific utilization of such massive data sets (Djorgovski & Williams, 2005).

Atomic and molecular (A&M) data and databases are becoming increasingly important for diagnostics, data interpretation, and the development of models and simulations of complex physical processes (Iacob, 2014; Srećković et al., 2014; Mihajlov et al., 2011; Vujčić et al., 2023; Dubernet et al., 2024). One of the key goals of the astronomical Virtual Observatories is still the scientific use

of such resources. In this contribution we provide an overview of the MolD, EMol and ACol databases, hosted within the Virtual Atomic and Molecular Data Center (VAMDC) and the Serbian Virtual Observatory (SerVO), their current state and future directions of development like e.g. involvement of machine learning (ML) models to produce and predict new A&M data (see papers of Michalenko et al. 2021; Harris & Nepomuceno 2024).

2. The Virtual Observatory and BG nodes

The SerVO (http://servo.aob.rs) hosts the Belgrade radiative and collisional nodes MolD, EMol and ACol of the Virtual Atomic and Molecular Data Centre (VAMDC) (Albert et al., 2020). The databases MolD (http://servo.aob.rs/mold), EMol (http://servo.aob.rs/emol/) and ACol (http://servo.aob.rs/acol) include astrophysically relevant data. Furthermore, the data and its analysis demonstrate their interdisciplinary character and applications, such as in physics, astrophysics, and chemistry. Fig.1 represents a snapshot of the SerVO web page. One can see on the left side are links to BG Nodes (MolD, ACol, BEAMdb/ EMol) as well as STARK-B (database for "Stark" broadening of isolated lines of atoms and ions in the impact approximation). In the middle are links to the photo plate archive as well as new information related to scientific meetings, etc.

Figure 1. The SerVO's homepage. On the left side are links to BG Nodes. In the middle are links to the photo plate archive as well as new information related to scientific meetings, etc.

The MolD, EMol and ACol Belgrade A&M databases have been linked with the VAMDC project (http://vamdc.eu) since their early phases of development (see e.g. Vujčič et al., 2015; Marinković et al., 2015). The VAMDC project

Albert et al. (2020) intends to bring together numerous current databases under a common standard, creating a centralized platform for accessing atomic and molecular data. Users can download A&M data in a consistent format using the XML Schema for Atoms, Molecules, and Solids (XSAMS) (Marinković et al., 2017; Jevremović et al., 2020; Albert et al., 2020). Technically, the data models were updated to effortlessly transition to the VAMDC's tree-structured serialization schema, XSAMS. All databases are stored on a MariaDB server (an open-source relational database management system based on MySQL) and backed up on a regular basis.

2.1. The MolD database

Information on a number of molecular species and their excited states can be found in the MolD database (see, for example, Srećković et al., 2017a; Vujčić et al., 2023). As a radiative database it includes associated averaged thermal photodissociation cross sections and the data, i.e., photodissociation cross sections, for the various ro-vibrational states of the diatomic molecular ions (see Fig. 2).

Beginning at the end of 2014, the MolD database underwent multiple rounds of development. The Astronomical Observatory Belgrade is hosting it. Astronomers regularly use the MolD, to study a variety of astro phenomena and model stellar atmospheres (Coppola et al., 2013; Mihajlov et al., 2011; Srećković et al., 2014). Research on plasma chemistry and experiments are among the other uses for the data (Srećković et al., 2021, 2022).

2.2. The EMol database

The EMol database curates cross sections for electron interactions with atomic particles (atoms, molecules and ions) as well as energy loss and threshold spectra. Cross sections may be differential, integrated (integral and momentum transfer) or total. Processes that are covered within the database are elastic scattering, excitations (electronic and vibrational) and ionization (partial or total). At the moment, within the database there are 16 neutral atomic species and two ionic ones, while there are 16 neutral molecules and five cations. These numbers grow slowly but steadily over the years.

Data maintained within EMol are published in scientific journals and have passed the referee procedure. They are obtained either experimentally or theoretically. Data sets coming from experiments are usually of limited range of experimental parameters, like impact energy, scattering angle or energy loss, while calculated data may be voluminous with dense grid. That is why some of the experimental data needs to be completed. There are already attempts to extrapolate experimental differential cross sections to inaccessible domains by using machine learning algorithms (Ivanović et al., 2020). One of the examples of successful application of such algorithms in the determination of cross

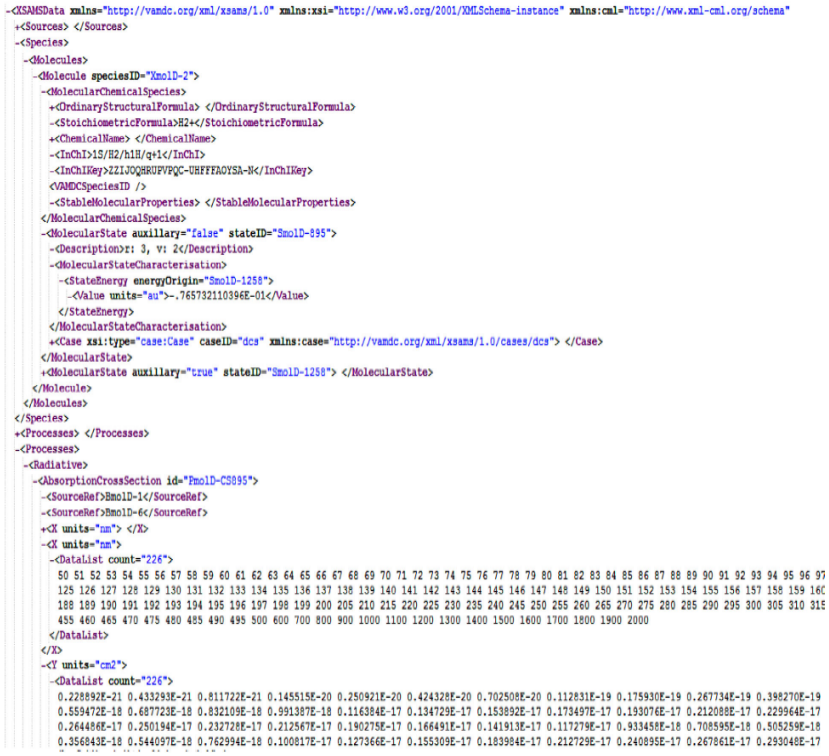

Figure 2. MolD sample output. Data displayed in XSAMS format.

sections from transport coefficients using deep neural networks is presented by Stokes et al. (2020).

2.3. The ACol database

The ACol database contains data on cross sections and rate coefficients for collisional processes of excitation/deexcitation and ionization/recombination in hydrogen, helium, and alkali metal plasmas (see e.g. Srećković et al., 2017b; Srećković et al., 2022; Vujčić et al., 2023, and references therein). This is the youngest BG VAMDC node. SerVO hosts the ACol database. The website executes the queries locally and transmits data in the VAMDC-specified XSAMS format via an AJAX-enabled web page. The dataset/database can be used for laboratory research and modeling of interstellar gas, weakly ionized layers in various atmospheres, and low temperature plasma.

3. Summary and further directions

In this paper we provided an overview of the MolD, EMol and ACol databases, hosted within the VAMDC and SerVO. MolD, EMol and ACol are atomic and molecular databases dedicated to modeling various laboratory plasmas, and stellar atmospheres.

To maintain and curate databases effectively, it's important to stay up-to-date with the rapidly evolving IT sector, in addition to following basic principles. In this regard, new standards and database future upgrades should be:

- Updating Node software, including Python and Django, on a regular basis.
- Adding new AJAX-enabled queries, visualizations, and extensions to the website user interface.
- Adding new calculated/measured datasets and radiative/collisional processes.
- Involvement of ML models to produce i.e. to fast predict new A&M data. We are currently in the process of preparation for training and testing datasets for development of advanced models.

While ML models can provide quick predictions, it is crucial to remember that they may not be very accurate. Consequently, there should be some skepticism regarding the results and datasets of machine learning models.

Acknowledgements. This work has been supported by the Institute of Physics Belgrade and Astronomical Observatory, through funds from the Ministry of Science, Technological Development and Innovations of the Republic of Serbia. The authors acknowledge the networking opportunities from COST Actions: CA20129 (MultIChem), CA21101 (COSY), CA22133 (PLANETS) supported by COST (European Cooperation in Science and Technology). VAS is grateful for the hospitality of West University Timisoara. BPM. acknowledges support of the Science Fund of the Republic of Serbia, Grant No. 6821, Atoms and (bio)molecules - dynamics and collisional processes on short time scale – ATMOLCOL.

References

Albert, D., Antony, B. K., Ba, Y. A., et al., A Decade with VAMDC: Results and Ambitions. 2020, *Atoms*, **8**, 76, DOI:10.3390/atoms8040076

Coppola, C. M., Galli, D., Palla, F., Longo, S., & Chluba, J., Non-thermal photons and H_2 formation in the early Universe. 2013, *MNRAS*, **434**, 114, DOI:10.1093/mnras/stt1007

Djorgovski, S. G. & Williams, R., Virtual Observatory: From Concept to Implementation. 2005, in Astronomical Society of the Pacific Conference Series, Vol. **345**, *From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science*, ed. N. Kassim, M. Perez, W. Junor, & P. Henning, 517

Dubernet, M. L., Boursier, C., Denis-Alpizar, O., et al., BASECOL2023 scientific content. 2024, *Astronomy and Astrophysics*, **683**, A40, DOI:10.1051/0004-6361/202348233

Harris, A. L. & Nepomuceno, J., A data-driven machine learning approach for electron-molecule ionization cross sections. 2024, *Journal of Physics B Atomic Molecular Physics*, **57**, 025201, DOI:10.1088/1361-6455/ad2185

Iacob, F., Spectral characterization of hydrogen-like atoms confined by oscillating systems. 2014, *Cent. Eur. J. Phys.*, **12**, 628, DOI:10.2478/s11534-014-0496-1

Ivanović, S., Uskoković, N., Marinković, B. P., & Mason, N. J., Determining Extrapolated Differential Cross Sections from Data Sets in BEAMDB Using Machine Learning Algorithms. 2020, in 30^{th} Summer School and International Symposium on the Physics of Ionized Gases, Vol. **99**, *Publications of the Astronomical Observatory of Belgrade*, ed. L. C. Popović, D. Borka, D. Ilić, & V. A. Srećković (Astronomical Observatory of Belgrade), 43–46, http://spig2020.ipb.ac.rs/Spig2020-Book-Onine.pdf

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al., LSST: From Science Drivers to Reference Design and Anticipated Data Products. 2019, *Astrophysical Journal*, **873**, 111, DOI:10.3847/1538-4357/ab042c

Jevremović, D., Srećković, V. A., Marinković, B. P., & Vujčić, V., Databases for collisional and radiative processes in small molecules needed for spectroscopy use in astrophysics. 2020, *Contributions of the Astronomical Observatory Skalnate Pleso*, **50**, 44, DOI:10.31577/caosp.2020.50.1.44

Koekemoer, A. M., Aussel, H., Calzetti, D., et al., The COSMOS Survey: Hubble Space Telescope Advanced Camera for Surveys Observations and Data Processing. 2007, *Astrophysical Journal, Supplement*, **172**, 196, DOI:10.1086/520086

Marinković, B. P., Jevremović, D., Srećković, V. A., et al., BEAMDB and MolD - databases for atomic and molecular collisional and radiative processes: Belgrade nodes of VAMDC. 2017, *European Physical Journal D*, **71**, 158, DOI:10.1140/epjd/e2017-70814-6

Marinković, B. P., Vujčić, V., Sushko, G., et al., Development of collisional data base for elementary processes of electron scattering by atoms and molecules. 2015, *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, **354**, 90, DOI:10.1016/j.nimb.2014.12.039, 26th International Conference on Atomic Collisions in Solids

Michalenko, J. J., Murzyn, C. M., Zollweg, J. D., et al., Machine Learning Predictions of Transition Probabilities in Atomic Spectra. 2021, *Atoms*, **9**, DOI:10.3390/atoms9010002

Mihajlov, A. A., Ignjatović, L. M., Srećković, V. A., & Dimitrijević, M. S., Chemi-ionization in Solar Photosphere: Influence on the Hydrogen Atom Excited States Population. 2011, *Astrophys. J., Suppl. Ser.*, **193**, 2, DOI:10.1088/0067-0049/193/1/2

Srećković, V., Ignjatović, L., Jevremović, D., Vujčić, V., & Dimitrijević, M., Radiative and Collisional Molecular Data and Virtual Laboratory Astrophysics. 2017a, *Atoms*, **5**, 31, DOI:10.3390/atoms5030031

Srećković, V. A., Jevremović, D., Vujčić, V., et al., Mol-D a database and a web service within the Serbian virtual observatory and the virtual atomic and molecular data centre. 2017b, *Proceedings of the International Astronomical Union*, **12**, 393, DOI:10.1017/S1743921316012643

Srećković, V. A., Mihajlov, A. A., Ignjatović, L. M., & Dimitrijević, M. S., Ion-atom radiative processes in the solar atmosphere: quiet Sun and sunspots. 2014, *Advances in Space Research*, **54**, 1264, DOI:10.1016/j.asr.2013.11.017

Srećković, V. A., Ignjatović, L. M., & Dimitrijević, M. S., Photodestruction of Diatomic Molecular Ions: Laboratory and Astrophysical Application. 2021, *Molecules*, **26**, DOI:10.3390/molecules26010151

Srećković, V. A., Ignjatović, L. M., Kolarski, A., et al., Data for Photodissociation of Some Small Molecular Ions Relevant for Astrochemistry and Laboratory Investigation. 2022, *Data*, **7**, DOI:10.3390/data7090129

Stokes, P. W., Cocks, D. G., Brunger, M. J., & White, R. D., Determining cross sections from transport coefficients using deep neural networks. 2020, *Plasma Sources Science and Technology*, **29**, 055009, DOI:10.1088/1361-6595/ab85b6

Tucker, D. L., Kent, S., Richmond, M. W., et al., The Sloan Digital Sky Survey monitor telescope pipeline. 2006, *Astronomische Nachrichten*, **327**, 821, DOI:10.1002/asna.200610655

Vujčić, V., Marinković, B. P., Srećković, V. A., et al., Current stage and future development of Belgrade collisional and radiative databases/datasets of importance for molecular dynamics. 2023, *Physical Chemistry Chemical Physics*, **25**, 26972, DOI:10.1039/D3CP03752E

Vujčič, V., Jevremović, D., Mihajlov, A. A., et al., MOL-D: A Collisional Database and Web Service within the Virtual Atomic and Molecular Data Center. 2015, *Journal of Astrophysics and Astronomy*, **36**, 693, DOI:10.1007/s12036-015-9344-y