Data Management Plan

Action Number: 101040356

Action Acronym: ATTOP
Action title: ATTOsecond Photochemistry: controlling chemical reactions with electrons

Date: 04/05/2023

DMP version: 1.0

1. Data Summary

ATTOP is a theoretical chemistry project that will explore the synergy between the two fields of attoscience and photochemistry. The central goal of ATTOP is to tackle the challenge of designing and suggesting, using modelling approaches, molecular systems, electronic wavepackets and attosecond experiments that manipulate the outcome of chemical reactions for diverse applications. To that end, simulations will be run with a number of quantum chemistry and quantum dynamics codes. Among them are Quantics, OpenMolcas, BAGEL and SHARC.

OpenMolcas, BAGEL and SHARC are opensource codes, all available on the GitHub platform, at https://github.com/Molcas/OpenMolcas, https://github.com/nubakery/bagel and https://github.com/sharc-md/sharc respectively. Quantics code is available from Mendeley Data: https://data.mendeley.com/datasets/x9dcpc2r5c/1. The simulations will generate text and binary files, the content of which will describe quantitatively the electronic and nuclear behaviour upon photo-excitation, and the size of which is expected to reach up to several tens of To.

We may re-use the xyz geometries optimised in previous theoretical studies. Similarly, the xyz of key geometries that we will find in the course of this project may be useful for other researchers in the photochemistry community.

One of the objectives of ATTOP is to develop a computational protocol to simulate quantum dynamics simulations of large systems (≈ 30 atoms and more), with a reduced number of coordinates chosen objectively, while controlling the impact of the approximations made. This is critical to be able to apply atto-photochemistry to a wide range of reactions. To achieve this, we will use machine learning algorithms in order to identify the most relevant nuclear coordinates for the studied process. We think it is important to emphasize that, while to design the computational workflow, we will generate data for specific test reactions, these will not be useful for other chemical reactions since one will need to create new data for each specific reaction of interest. Once developed, we expect the workflow itself however to be extremely useful to the quantum dynamics community.

Towards the end of ATTOP, we will develop descriptors and analysis tools for electronic wavepackets, with the idea to use them to assess the expected suitability of a given electronic wavepacket to induce a desired chemical reaction. These developments will be made in the OpenMolcas code and are expected to be useful to the atto-chemistry community.

2. FAIR data

2.1. Making data findable, including provisions for metadata

We will create a GitHub for ATTOP, with a mirror on the GitLab repository of Nantes Université, gathering for instance, the workflow developed, all important xyz data (starting geometries and velocities of dynamics simulations to ensure reproducibility of the simulations), the preprint of our publications, any poster or presentation pdfs, notebooks etc. Care will be taken to provide useful and rich metadata (example: attochemistry, charge migration, charge-directed reactivity).

The developed workflow will be assigned a persistent identifier with Zenodo.

We will also add a tab for the ATTOP project on Morgane Vacher's website (https://morganevacher.wordpress.com/), centralizing all the links to the archived data.

2.2. Making data accessible

The ATTOP GitHub will be publicly accessible, and so the data shared via it (eg. xyz data, workflow developed, publications).

Preprints of publications will also be deposited on ArXiv (an openaccess repository) and HAL platforms and be under CC BY 4.0, as required by French National Research Agency regulations.

The workflow developed will be released under GPLv3 license.

Any other code development in already existing packages will be constrained by the specific licences of the packages themselves. OpenMolcas code is released under the LGPLv2.1 license and so the descriptors and analysis tools developed in that code will also be released under the LGPLv2.1 license. Similarly, SHARC and Quantics codes are released under the GPLv3 and LGPLv3 licenses respectively, so any development in those codes will be released under the same licenses.

2.3. Making data interoperable

Standard xyz format is used for atomic coordinates.

Quantics is already interfaced with several electronic structure packages: Gaussian, Molpro, OpenMolcas. We will create another interface with the BAGEL code. SHARC is also already interfaced with many electronic structure packages: Gaussian, Molpro, OpenMolcas, BAGEL, Columbus, Orca. The two dynamics codes are thus compatible with many quantum chemistry codes.

The workflow will be developed using standard python libraries and thus easily interoperable. Moreover, it will take as input standard xyz molden format files, as output from widely used mixed quantum-classical dynamics code (OpenMolcas and SHARC).

2.4. Increase data re-use

Documentation on how to use the developed workflow will be provided on ATTOP GitHub and Zenodo. This will include notebooks, test sets and documentation which will be made freely available in the public domain to permit the widest re-use possible.

The same is true for the new descriptors and analysis tools for electronic wavepackets through the publicly available OpenMolcas package. In addition, OpenMolcas being widely used in the quantum chemistry community, it will facilitate the re-use of these developments by other researchers. The developments of these new descriptors will be accompanied by additional documentation and tests, as required by any development part of the OpenMolcas project.

3. Allocation of resources & Data security

In the ATTOP project, part of the budget is provided for open-access publication fees from Journals.

The mirroring of the GitHub repository in the GitLab of the coordinator's University provides a backup of the data stored there. The additional upload of the workflow on Zenodo provides secure storage and long-term preservation of this developed code.

We have a local dedicated research storage space managed by the scientific interest group (GIS) Glicid. As the sub-projects end, the respective data will be sshed locally to our storage space from the computing centers. We also plan to buy a storage server (several tens of To) to be installed locally in the CEISAM laboratory, in order to back up the data stored in the Glicid storage space.

Aymeric Blondel is the person responsible for data management in the ATTOP project.