8-2017

Deciphering Delocalized Bonding in Excited States, Solvated Species and Novel 0-, 1-, 2-, and 3-Dimensional Chemical Systems

Alexander Boldyrev
Utah State University, a.i.boldyrev@usu.edu

Data Management Plan

1. Products of the Research

The results of the project include inputs and outputs for various quantum chemistry software packages, and technical details for performing the calculations. Input and output files will be in ACSII format. We estimate that about 200,000 files will be generated during this project. Further, the proposed research will result in new computer source code implementing new quantum chemistry methods, and underlying theoretical derivations. All generated computational inputs and outputs and other project data will be stored on the DigitalCommons@USU storage and be protected by hardware redundancy and a strict backup schedule (ranging from nightly backups to long term storage on a tape backup server). Utah State University, through the Merrill-Cazier Library, provides institutional repository services through the bepress Digital Commons platform. DigitalCommons@USU supports all file types and formats. Files are provided with persistent URLs, and if needed, Library staff can obtain DOIs for datasets. The system is able to produce license and copyright statement as needed, and creates standard citations. All files are backed up at multiple sites, including cloud storage. Preservation copies are stored in Amazon Web Services, with redundant storage across multiple facilities and are regularly verified for integrity of data using checksums. The files of different projects will be organized in accordance with the publication date. The output Gaussian files within the project will be organized according to their relative energies and summarized in a specific user-friendly output file. The input Gaussian files can be found according to the name of the output file. The data will be kept in this storage for a minimum of 5 years after completion of the grant. Primary research data are stored in Utah State University Cluster and in password protected for individual users and/or research group computers, but individuals designated by PI will have access to all generated data to fulfill any possible reporting, investigational, or compliance requirements. Computers are backed up on a regular schedule. All source code developed by the group within the program package Gaussian gets ultimately transferred to GitHub and will be available to user through this source.

2. Data Formats

The inputs and outputs of quantum chemistry programs are in universally accessible ASCII format. Computational protocols, theoretical derivations, as well as original and derived data will be published in reputable scientific publications, which will be made available per NSF policy. Program source code will be stored in zip-archive format. Most figures for our publications will be made via Molekel program and further modified using Adobe Photoshop program and stored in Tag Image File Format (TIFF). They will be archived together with the raw data and the manuscript files. They represent geometric structures of molecules and solids (usually a one or a few unit cells) with specification of atoms, geometric parameters (angles and distances) and also chemical bonding pictures showing lone pair, and two-center two-electron (2c-2e) bonds, three-center two-electron (3c-2e) bonds and in general nc-2e bonds over the geometric framework of a molecule or a crystal (over one or a few unit cells).

3. Access to Data and Data Sharing Practices and Policies

The primary means of providing access to research data will be through the publication of the results in peer-reviewed journals, as figures and tables within the journal publication as well as in the Supplementary Information or Supporting Information (SI) sections made available by journals for published data. The published results will be easily verifiable as the programs used for the research are available free of charge (see paragraph 2.). The PI has strong track records of publishing with his students, which will ensure the timely communication of results to the general scientific community. XYZ coordinates of molecular structures, visualized or not, will be published in article SI documents as well. Volume data are generally too voluminous for publication in a SI document and will be available in DigitalCommons@USU. The only case where sharing is limited is proprietary source code from commercial software. However, all theoretical derivations and technical aspects of a calculation using proprietary software in part or fully will be published together with the results from the project, such that the computations can be reproduced and verified.

4. Access to Data and Data Sharing Practices and Policies

Data will be available to reuse upon request and will be archived in DigitalCommons@USU. PI only request the user to acknowledge source of data.

5. Archiving of Data

Peer reviewed articles will be deposited in NSF Public Access Repository. Primary data will be archived in DigitalCommons@USU. DigitalCommons@USU's metadata is based on the Dublin Core metadata standard. Data files are accompanied by appropriate README and other documentation files to ensure future usability of data. Computer code will be archived in GitHub.

No human or animal subject experiments will be conducted during this project.