Usage Report for Fiscal Year 2022

Project Title:

Rigorous higher-order DFT

Name:
○Bun Chan (Nagasaki University), Takahito Nakajima (RIKEN)

Laboratory at RIKEN:
Computational Molecular Science Research Team

1. Background and purpose of the project, relationship of the project with other projects

In our previous investigations into the development of advanced DFT methods and related topics as part of this continuous project, we have devised several new methods on the basis of rigorous physical fundamentals. In more recent studies, we have used statistical and machine-learning approaches to assess the performance of DFT methods and unravel their strengths and weaknesses.

In our statistical studies, we have compiled new databases of accurate chemical data to reinforce the reliability of our analysis. We have in this FY furthered this endeavor and devised new sets of chemically independent data. As part of this effort, we have also developed new computational chemistry methods that are capable of producing data of the highest quality but with substantially better efficiency. These new methods have already, and will in the future, yield sizable databases of highly accurate data for the development of DFT.

2. Specific usage status of the system and calculation method

This project employs the Gaussian and Q-chem programs on Hokusai, as well as a wide range of standard quantum chemistry software packages such as Molpro, MRCC, and Orca. These tools enable us to access a diverse range of quantum chemistry methodologies, from highly accurate coupled-cluster methods at the one end, with which accurate reference data can be obtained, to a vast collection of DFT methods at the other end, with which insights into the fundamentals of a reliable DFT for a large variety of systems can be revealed.

In terms of assessment of DFT, we diversity our focus from small molecules to medium-sized to large macromolecules, in order to unravel potential deficiencies that may not be applicable to small systems. At this stage, our focus is on biologically relevant systems such as lignocellulose, which is the most abundant biomass on earth, and proteins that are the ubiquitous molecular machines in living organisms. We have also investigated redox chemical systems that are of relevance to bio-mimic energy harvesting systems.

Regarding the development of highly accurate methods for obtaining reliable reference quantities, we have previously optimized a major component of the high-level W*n* protocol, namely the basis set, to the fullest extent. Those studies have already led to advanced protocols including W*n*X and WG, as well as W*n*-P34 that is applicable to most of the periodic table in addition to light-main-group elements.

To further reduce the requirements on computational resources without compromising the accuracy would necessitate saving in other key components. In several studies, we have in this FY applied new algorithms of "local correlation" to significantly lower the scaling of computational requirement with respect to the size of the chemical system, which would guarantee the viable application to large systems.

Thus, our focus is on ensuring that the resulting protocol retain the accuracy, and to that end we have thoroughly tested the options used in several local correlation algorithms. We have used the W*n*X and the lower-cost and also widely used G4(MP2) protocols as the platform for our development.

3. Result

In our assessment of DFT methods for lignocellulose, we have screened a large set of over 50000 models using a multi-step, multi-level, approach. During this process, we have identified the MS1-D3 DFT method to be optimal for rapid determination of molecular structures of our model systems. In addition, we have found that the SCANh DFT method provides an accurate account for the relative energies of the model systems at a reasonable cost.

These methods have enabled us to narrow down the pool of models to just a few hundred, for which higher level DH-DFT and W*n*-type protocols can be applied to obtain energetics with the highest achievable level of accuracy. Using the high-quality results, we have further devised a machine-learning model for the initial screening of lignocellulose models at essentially zero cost. This model can in future studies be applied to highly realistic lignocellulose models to unravel their chemistry.

Our investigation into the appropriate DFT method for the calculation of biological macromolecular systems involve testing on not only proteins but also solvation of molecules, which is relevant to, e.g., interaction in hydrophilic and hydrophobic environment within organisms. Some of these are shown in Figure 1.

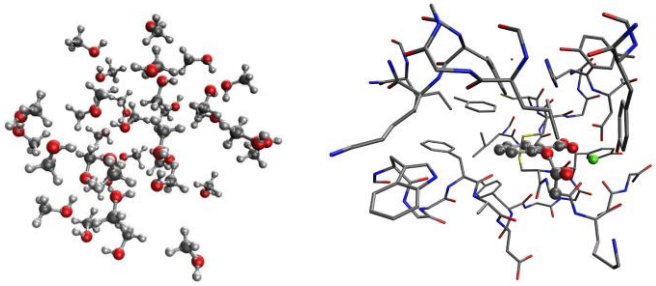

Figure 1. Example biologically related systems used for the assessment of DFT methods

They cover chemical systems and properties applicable to the ease of drug delivery and efficacy of drug actions. In total, our data set has more than 200 points of such systems. Using this data set, we have examined dozens of low-cost computational chemistry methods, including mostly DFT, as well even lower-cost methods that have become increasingly widely used in recent years. They include semi-empirical methods, tight-binding methods, and DFT-based 3c. Our assessment has led to us recommending the B97M-V method, which yields exceptionally accurate results (Figure 2).

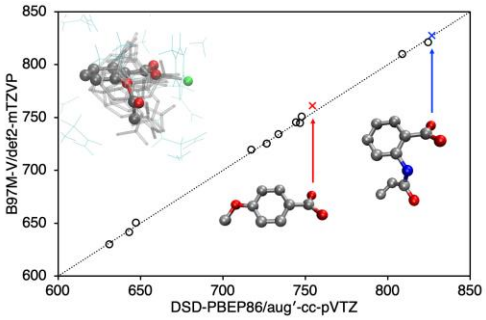

Figure 2. Correlation of low-cost B97M-V energies with accurate reference for drug–protein complexes.

In the area of development local-correlation-based high-accuracy methods, we have surveyed two main algorithms called DLPNO and LNO. For DLPNO, we have further investigated altering technical options including “grid size”, “PNO cut off” and “triples algorithm”. Our studies have suggested the “DefGrid3”, TightPNO, and the T1 algorithm to be the optimal. We have also examined the corresponding options for the LNO algorithm, and we have come to similar conclusions. Using these knowledges, we have developed the L-W1X and L-G4(MP2) methods. Both are capable of achieving the same level of accuracy as the methods that they are based on [W1X and G4(MP2)] but with at least an order of magnitude better efficiency.

4. Conclusion

Our continuous efforts in the assessment of DFT and high-level quantum chemistry methods have yielded valuable recommendations for the computation of large chemical systems. They have also created new methods that would enabled investigation of even larger species that have yet to be explored.

5. Schedule and prospect for the future

Our future investigations will involve further expansion of our data sets to cover a larger chemical space. This would ultimately lead to a fully unbiased data set, which would be a prerequisite for the development of reliable DFT methods.

Usage Report for Fiscal Year 2022

Fiscal Year 2022 List of Publications Resulting from the Use of the supercomputer

[Paper accepted by a journal]

1. Modeling the Conformational Preference of the Lignocellulose Interface and Its Interaction with Weak Acids. Chan, B.; Dawson, W.; Nakajima, T. *J. Phys. Chem. A* **2022**, *126*, 2119.
2. Searching for a Reliable Density Functional for Molecule–Environment Interactions, Found B97M-V/def2-mTZVP. Chan, B.; Dawson, W.; Nakajima, T. *J. Phys. Chem. A* **2022**, *126*, 2397.
3. Assessment of DLPNO-CCSD(T)-F12 and Its Use for the Formulation of the Low-Cost and Reliablel-W1X Composite Method. Chan, B.; Karton, A. *J. Comput. Chem.* **2022**, *43*, 1394.
4. High-Level Quantum Chemistry Reference Heats of Formation for a Large Set of C, H, N, and O Species in the NIST Chemistry Webbook and the Identification and Validation of Reliable Protocols for Their Rapid Computation. Chan, B. *J. Phys. Chem. A* **2022**, *126*, 4981.
5. Performance of Local G4(MP2) Composite Ab Initio Procedures for Fullerene Isomerization Energies. Karton, A.; Chan, B. *Comput. Theor. Chem.* **2022**, *1217*, 113874.
6. High-Level Quantum Chemistry Exploration of Reduction by Group-13 Hydrides: Insights into the Rational Design of Bio-Mimic CO_2 Reduction. Chan, B.; Masanari, K. *Electron. Struct.* **2022**, *4*, 044001.

[Oral presentation]

1. *Twelfth Triennial Congress of the World Association of Theoretical and Computational Chemists (WATOC)*, 2022, Vancouver (invited oral presentation).