

Project Title:**Rigorous higher-order DFT****Name:**

○ Bun Chan (Nagasaki University), Takahito Nakajima (RIKEN), Kimihiko Hirao (RIKEN), Subrata Tewary (RIKEN)

Laboratory at RIKEN:

Computational Molecular Science Research Team

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

Density functional theory (DFT) is the most prominent technique in computational chemistry. With computational resource support from RIKEN, we have in the past few years developed advanced DFT methods. This includes the formulation of improved “screened-exchange” procedures (e.g., HSEB and reHISB) that provide an optimal balance between accuracy and computational efficiency. The results have now been published.

We have showed that our new methods are accurate for thermochemical properties, but their performance for excited states is yet to be established. To expand the scope of our methods in this direction, it is instructive to better understand the most accurate DFT methods currently available for such a purpose. In this regard, long-range corrected (LC) methods provide an excellent platform for the study of relevant underlying physics.

While the good accuracy of our SX-DFT methods is achieved with rigorous physical principles, further improvement necessitates a pragmatic approach of fitting to accurate benchmark data. An obstacle to adopting this strategy is that it involves laborious optimization for a large data set. Thus, the compilation of small, yet high-quality, data set is vital for the improvement of our DFT methods in an efficient manner.

2. Specific usage status of the system and calculation method

The developmental work in this project employs customized Gaussian program on Hokusai, as well as

the standard quantum chemistry software package Q-Chem. A new LC-DFT method is being developed, using the so-called “corrected projection operator” mathematical technique, to unravel the correct physical behavior for excited states.

In the aspect of developing high-quality small data sets, we use the extensive MGCDB84 set with 5000 accurate data points of chemical properties as a baseline for comparison. Results for 200 DFT methods are used, leading to a total of a million unique data points. Subsets of MGCDB84 are then selected using standard statistical techniques.

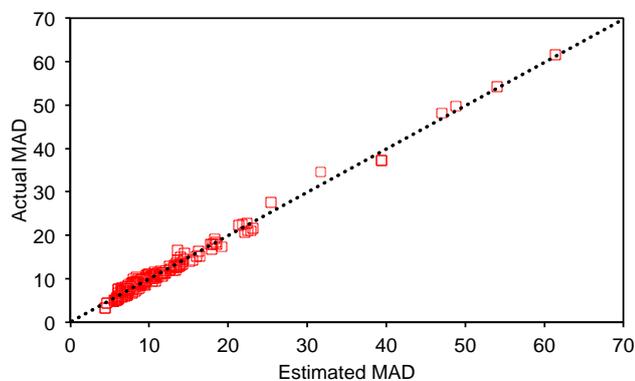
The large amount of computations required in these projects has consumed a significant portion of the assigned computer time on Hokusai. Overall, we have in FY2018 dedicated most of our resources toward the formulation of small data sets.

3. Result

Using the standard statistical techniques “lasso regularization” and “forward selection”, we find that a very small subset of MGCDB84 with just 30 of the 5000 data points correlates remarkably well with the complete data set, with a coefficient of determination (R^2) of 0.990. However, when we use this minimal subset to assess the accuracy of a given DFT method, we find that finer details such as how the method performs for different chemical properties (e.g., covalent versus non-covalent interactions) are lost.

We re-introduce additional data points within the MGCDB84 set to retain the finer details provided by the complete MGCDB84 set. Encouragingly, we find that a small collection that satisfies this criterion has just 60 data points; it still represents a

much more economical tool for the assessment of DFT methods when compared with the 5000-data-point MGCDB84. The overall good performance of this small data set is illustrated by an R^2 value of 0.995 (see the figure below).



With these promising results, we use the same statistical methodology to further devise small data sets that contain properties that are independent from those found in MGCDB84, which contains only main-group thermochemical data. Our preliminary results include some small data sets for transition-metal reaction energies, semi-conductor band gaps, and organic-dye excitation energies.

Regarding the application of DFT to the calculation of excitation energies, our preliminary findings show that, while LC-DFT indeed compares favorably to the highly accurate EOM-CCSD method for non-conjugated systems such as water and ethene, larger discrepancies are found for highly conjugated species, e.g., benzenoids. An examination of the fundamentals of electronic excitations suggest that the use of projected virtual orbitals may improve the accuracy.

4. Conclusion

Our previous studies in the development of accurate and computationally efficient DFT has already led to near-optimal methods that are based solely on the fundamental physics of DFT. Our current preferred strategy for further improving our methods involves optimization for a wide variety of chemical properties but this would have necessitated the computation of an enormous number of chemical data points using existing data sets.

To overcome this obstacle, we use statistical techniques to mine the most important data points from a large collection of main-group thermochemical quantities. Our work has led to the reduction of a highly robust 5000-data-point set to a minimal set of just 60 points without losing detail chemical information contained in the large set. With the emergence of this promising finding, we have started to build small, yet high-quality, data sets for a wider range of chemical properties.

We also aim to broaden the scope of our DFT methods to cover both ground and excited state properties. To this end, we have carried out preliminary studies into the fundamentals of applying LC-DFT to electronic excitations, because this type of DFT methods are generally considered most accurate for such a purpose. Our results suggest that using the advanced technique of “virtual orbital projection” may be critical to accurately describe these processes.

5. Schedule and prospect for the future

While the goal of our on-going study is to formulate a DFT method with optimal accuracy and efficiency, we will at this stage continue to build a solid foundation by devising high-quality and efficient benchmark data sets for a wider range of properties. This is vital for the optimization of new DFT methods. Notably, we will include additional transition metal properties in new data sets, as the chemistry of transition metals forms a critical part of modern chemical science.

We will also further examine electronic excitation properties, which are critical to applications in, e.g., solar energies. Specifically, we will implement algorithm for virtual orbital projection and examine its performance for describing different types of excited states for a wide range of substrates.

These will build on the collaboration between Dr Chan at Nagasaki University and Prof Nakajima and Prof Hirao at RIKEN AICS, with the provision of computational resources from RIKEN ACCC.

Fiscal Year 2018 List of Publications Resulting from the Use of the supercomputer

[Paper accepted by a journal]

1. Chan, B.; Kawashima, Y.; Hirao, K. The reHISS Three-Range Exchange Functional with an Optimal Variation of Hartree–Fock and Its Use in the reHISSB-D Density Functional Theory Method. *J. Comput. Chem.* **2019**, *40*, 29.
2. Chan, B. Formulation of Small Test Sets Using Large Test Sets for Efficient Assessment of Quantum Chemistry Methods. *J. Chem. Theory Comput.* **2018**, *14*, 4254.
3. Accurate Thermochemical and Kinetic Stabilities of C₈₄ Isomers. Waite, S. L.; Chan, B.; Karton, A.; Page, A. J. *J. Phys. Chem. A* **2018**, *122*, 4768.
4. Mechanism for Three-Component Ni-Catalyzed Carbonyl–Ene Reaction for CO₂ Transformation: What Practical Lessons Do We Learn from DFT Modelling? Chan, B.; Luo, Y.; Kimura, M. *Aust. J. Chem.* **2018**, *71*, 272.

[Oral presentation]

1. *3rd International Conference on Emerging Advanced Nanomaterials (ICEAN)*, 2018, Newcastle (invited oral presentation).

[Poster presentation]

1. *55th Kyushu Area Joint Meeting of the Chemistry-Related Societies*, 2018, Kitakyushu (poster presentation).
2. *21th Annual Meeting of Theoretical Chemistry Society*, 2018, Okazaki (poster presentation).