

**Project Title:****Designing a Bioremediator: Mechanistic Models Guide Cellular and Molecular Specialization****Name:**

O William Dawson (1), Bun Chan(1)

**Laboratory at RIKEN:**

(1) RIKEN Center for Computational Science, Computational Molecular Science Research Team

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

Density Functional Theory (DFT) is a widely-used method in to predict properties of molecules and materials. Traditionally, its use has been limited to small systems (a few hundred atoms) due to high computational requirements. However, recent advancements in algorithms and computing power have enabled DFT to handle much larger systems (several thousand atoms).

A current challenge in DFT calculations of large systems is that most advanced algorithms, which avoid the eigensolver step, can't provide information about orbital energies. This affects the ability of DFT to estimate ionization energies and excitations, which are crucial for understanding processes important for bioremediation. Our study aims to address this issue by employing the Quantum Mechanics - Complexity Reduction (QM-CR) framework, a solution developed in previous years of our project.

2. Specific usage status of the system and calculation method

BigDFT is a program designed for large-scale Density Functional Theory calculations, featuring a unique wavelet formalism that makes it suitable for computing large systems. Initially, we developed our QM-CR method specifically for BigDFT, making it the reference point in our study.

NTChem is another program for large-scale DFT calculations. It differs from BigDFT in that it uses Gaussian orbitals, making it capable of conducting more accurate hybrid functional calculations on

large systems and simulating core electrons.

PyBigDFT is a Python interface for both BigDFT and NTChem. It simplifies the management of the complex workflows needed for simulating large systems. Importantly, PyBigDFT is the package where the QM-CR method is implemented.

3. Result

We first analyzed the transferability of our QM-CR method to the NTChem program, using Molnupiravir surrounded by a cluster of water molecules as a case study. We compared QM-CR's measures of fragment quality and fragment interaction strength when using different types of density functionals and basis sets (Figure 1).

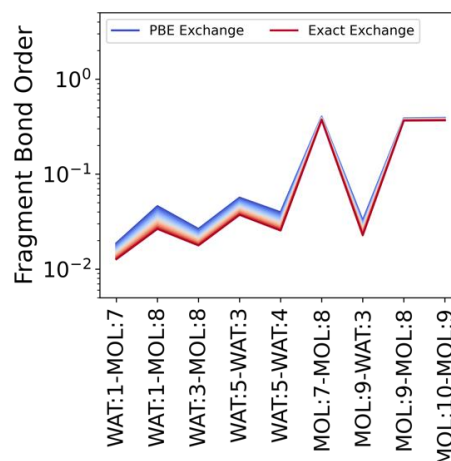


Figure 1: Fragment interaction strength, determined by the QM-CR approach using different fractions of exact-exchange. The ability to analyze systems using hybrid functionals (i.e. exact exchange) was a benefit of incorporating QM-CR into NTChem.

Our findings indicate that trends in fragment quality and interaction strength were well preserved across calculation methods. However, the use of hybrid and meta-GGA functionals lead to a more compact

description, characterized by higher quality fragments and weaker interactions. In contrast, using higher cardinality basis sets or diffuse functions produced the reverse effect. These observations confirm that, with appropriate calibration, the QM-CR method can be effectively adapted for use in NTChem. Additionally, by leveraging NTChem's capacity to simulate core electrons, we could extend our QM-CR metrics analysis beyond BigDFT's capabilities. This extended analysis revealed that core orbitals exhibit high locality, yet still interact with neighboring atoms.

Encouraged by the locality revealed by our QM-CR analysis, we proposed two new algorithms for computing the orbital energies of a system.

- 1) **Locality in Space:** In this approach, the QM-CR method partitions the system into distinct fragments within automatically generated embedding environments. Then we diagonalize the small submatrix specific to each fragment and its environment. Finally, the sub-density matrix of the target fragment is used to project out the eigenvalues related to the environment.
- 2) **Locality in Energy:** This method involves calculating the pivoted Cholesky decomposition of the density matrix corresponding to the desired energy window. Utilizing the resulting Cholesky vectors, we then condense the Hamiltonian into a more manageable, smaller matrix. The final step is the diagonalization of this reduced matrix to get the target orbital energies.

An example of these algorithms is shown in Figure 2. Our goal was to compute the core orbital energies of a protein in a large water cluster. We started by using a diagonalization free approach to compute the density matrix associated with the core orbitals of the system (i.e. local in energy). We then applied the QM-CR scheme to verify the suitability of the protein as a fragment and embed it in a minimal

environment. This analysis revealed the significant influence of hydrogen-bonding water molecules on core orbital energies. By diagonalizing the relevant submatrix and projecting out non-protein eigenvalues using the density matrix, we achieved an average error in reproducing orbital energies of less than 0.01 eV. Similar accuracy was observed with the Local in Energy algorithm.

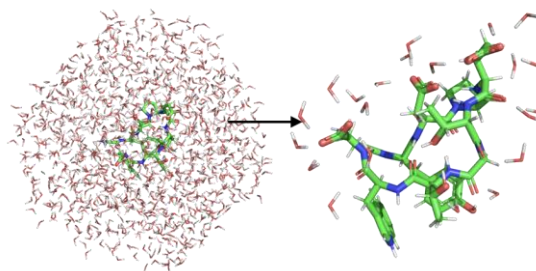


Figure 2: Using the QM-CR approach, we can extract a minimal embedding environment that describes the spectral quantities of a target moiety.

#### 4. Conclusion

We applied our QM-CR method to data generated by the NTChem code, observing a consistent correlation between NTChem and BigDFT results, with minor adjustments needed for basis set and functional choice. Encouraged by these findings, we developed two new algorithms for calculating orbital energies of large systems. These algorithms capitalize on the inherent locality of systems and reveal a deeper understanding of environmental impacts on orbital energies. In the realm of bioremediation, this work offers the potential to unravel the mechanisms by which enzymes oxidize substrates.

#### 5. Schedule and prospect for the future

In the following year, we plan to exploit the new computing system to study even larger systems. We will begin by benchmarking our codes on the new machine, using the new capabilities in PyBigDFT for remote job submission. We then plan to perform detailed tuning of the codes maximize performance on the new system.

#### 6. If no job was executed, specify the reason.

N/A

**Fiscal Year 2023 List of Publications Resulting from the Use of the supercomputer**

**[Paper accepted by a journal]**

• Dawson, William, Eisuke Kawashima, Laura E. Ratcliff, Muneaki Kamiya, Luigi Genovese, and Takahito Nakajima. "Complexity reduction in density functional theory: Locality in space and energy." *The Journal of Chemical Physics* 158, no. 16 (2023).

**[Conference Proceedings]**

**[Oral presentation]**

- William Dawson and Takahito Nakajima. "Complexity Reduction in Density Functional Theory: Locality in Space and Energy" Numerical Methods in Quantum Chemistry 2023. June 2023. Tromsø, Norway.
- William Dawson, Louis Beal, Luigi Genovese, and Takahito Nakajima. "Productive Large Scale Quantum Mechanical Calculations" 15<sup>th</sup> JLESC Workshop. March 2023. Bordeaux, France.

**[Poster presentation]**

- William Dawson, Luigi Genovese, and Takahito Nakajima. "Complexity Reduction in Density Functional Theory: Locality in Space and Energy" CCP2023. August 2023. Kobe, Japan.
- William Dawson, Luigi Genovese, and Takahito Nakajima. "Productive Large Scale DFT Calculations on the Fugaku Supercomputer" The 17<sup>th</sup> Annual Meeting of Japan Society for Molecular Science. September 2023. Osaka, Japan.

**[Others (Book, Press release, etc.)]**