TN

MI

CI

IMPROVING NWCHEM SCALABILITY USING THE DATASPACES FRAMEWORK

Allocation: Innovation and Exploration/500 Knh
PI: Gregory Bauer[1]
Collaborators: Victor Anisimov[1], Manish Parashar[2], Melissa Romanus[2]

[1]National Center for Supercomputing Applications
[2]Rutgers University

EXECUTIVE SUMMARY

The purpose of this project is to speed up coupled cluster CCSD(T) computation in NWChem and then incorporate this highly reliable level of theory into parameter optimization procedures for the AMBER classical force field for DNA, with the potential in future of extending this approach to optimization of other popular force fields. The development approach relies on revising the most memory-demanding part of the CCSD algorithm with the possibility of offloading the largest memory arrays to a DataSpaces server. The purpose of using CCSD(T) computation in the parameter optimization procedure is to employ the wealth of available experimental data, which is typically underutilized in common parameter-refining techniques. Moving the parameter optimization engine to a supercomputing platform, and automating the entire process of data extraction and parameter refining, makes the process reproducible, extendable, and portable.

RESEARCH CHALLENGE

Molecular Dynamics (MD) simulations based on classical force fields are the major tool to breech the gap between experiment and theory in materials science, engineering, and biomedical research. However, the predictive ability of MD simulations depends heavily on the quality of the underlying parameters. The limited pool of experimental data and the extremely labor-intensive nature of parameter optimization represent the major limiting factors in improving the quality of classical force field parameters. Theoretically, the use of highly reliable electronic structure computations at the CCSD(T) level of theory could help mitigate the lack of experimental data in parameter optimization. However, the high computational cost of the CCSD(T) method precludes its routine use even on modern supercomputing platforms. Therefore, faster and more computationally efficient CCSD(T) implementations are greatly needed. Additionally, the labor-intensive nature of classical force field determination, and the manually driven optimization, makes it difficult to reproduce the work, to revise and improve the protocol, and to introduce a quality control into the optimization process. Because parameter optimization is a computationally demanding effort, the efficient use of supercomputing resources becomes an additional challenge.

METHODS & CODES

The work on performance optimization of the CCSD(T) method in NWChem, which is maintained at Pacific Northwest National Laboratory, relies on offloading large memory arrays from compute nodes to a dedicated data storage employing the DataSpaces data management framework developed by the team headed by Manish Parashar at Rutgers University. While this work is in progress, the previously optimized version of the CCSD(T) method [1] in NWChem [2] has been used to generate a database of experimental data on intermolecular interactions from X-ray crystallographic data of molecular crystals of DNA bases. An automated procedure was developed to extract unique intermolecular pairs from the crystal data, to relax hydrogen atoms, to compute CCSD(T) interaction energies, and to arrange the data in a structured form. A scalable tool to optimize Lennard-Jones parameters in the AMBER (Assisted Model Building with Energy Refinement) force field to fit the parameters to intermolecular interaction energies for experimental geometry of monomers has been developed. It has been tested to run on 1,000 nodes using 32,000 processing units on Blue Waters. The optimization tool generates 200 alternative parameter sets of comparable quality determined by the fitting criterion. The number of alternative parameter sets is easily adjustable in this method. The optimized parameter sets proceed to the final validation step running in parallel to perform MD simulations, to compare the geometry of the system to that from the X-ray data, and to reproduce experimental heat of sublimation, unit cell volume, and interatomic distances.

RESULTS & IMPACT

This project introduces a new approach to parameter optimization for classical force fields that combines a high-level electronic structure calculation method to extract additional previously inaccessible information from experimental data, to reengineer the optimization procedure, and to tailor it to maximally utilize high-performance computational resources. The developed procedure shortens the time needed for parameter optimization by roughly a factor of 10. It resolves the issue of accessibility to only a few highly capable teams, transforms the force field optimization from an empirical to a well-structured discipline, educates, and makes the results easy to reproduce by the community.

WHY BLUE WATERS

Blue Waters, with its fast interconnect and large memory per core, is unique in its ability to conduct CCSD(T) computations of molecular systems encountering a thousand basis functions, which is vital for the success of the developed parameter optimization procedure. Since the parameter optimization procedure is extremely resource demanding, the availability of large numbers of nodes is essential for the exhaustive exploration of parameter space.