DI
GA
TN
ML
BI

RADICAL-CYBERTOOLS: A BUILDING-BLOCKS APPROACH TO HARNESSING THE POWER OF MANY

Allocation: NSF PRAC/3,450 Knh
PI: Shantenu Jha[1]

[1]Rutgers, the State University of New Jersey

EXECUTIVE SUMMARY

RADICAL-Cybertools (RCT) is a set of software tools that provides extensible building-block capabilities to enable the scalable utilization of production-distributed cyberinfrastructure. We use RCT to address the computational challenges in running ensemble-based methods for free-energy protocols (HTBAC) and enhanced sampling techniques (ExTASY). We are also using RCT to provide deep-learning workflows for classification of high-resolution satellite imagery.

RESEARCH CHALLENGE

RADICAL-Cybertools has been used to develop three toolkits, which have been utilized to support scalable science applications:

- HTBAC (High-Throughput Binding Affinity Calculator)—HTBAC aims to facilitate the running of drug-screening campaigns of molecular simulations for computing binding affinities.
- ExTASY (Extensible Toolkit for Advanced Sampling and analYsis)—ExTASY uses enhanced sampling techniques to explore the same conformational space using computationally efficient approaches to study more complex proteins at longer timescales.
- ICEBERG (Imagery Cyberinfrastructure and Extensible Building Blocks to Enhance Research in the Geosciences)—ICEBERG aims to facilitate pan-Arctic and pan-Antarctic science projects requiring manipulation and interpretation of high-resolution satellite imagery.

METHODS & CODES

We have developed the HTBAC toolkit to address the need for scalable and accurate protein–drug binding affinity results in acute precision medicine. In a computational drug campaign that requires the screening of thousands of compounds, accuracy comes at the cost of capturing subtle chemical properties of individual drug candidates. Accurate free-energy calculations require adaptive workflows that intervene during runtime to analyze intermediately generated simulations to optimally place the next set of simulations. In the second citation under Publications and Data Sets (below), we demonstrate that adaptive simulation methodologies can improve binding affinity accuracy given a fixed amount of compute time to make drug campaigns more computationally effective. Our code is open source and provided at https://github.com/radical-cybertools/htbac.

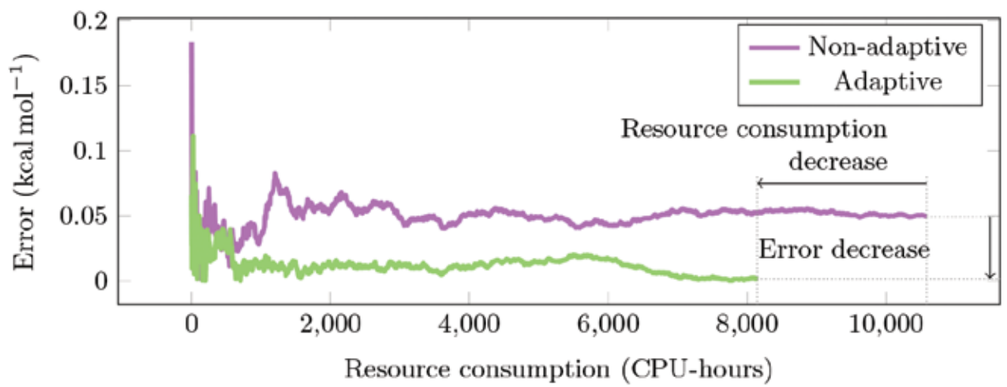

Figure 1: Adaptive simulation methodologies reduce both resource consumption and uncertainty errors in free energy, leading to a reduction in time-to-solution and improvement in accuracy. The "optimal" simulation parameters are difficult to determine *a priori*, thus requiring runtime determination and adaptation of the workflow based on intermediately generated data.

The ExTASY project aims to address the fact that the adaptive sampling of proteins is more complex to set up than plain molecular dynamics since it requires periodic iterations of molecular dynamics and analysis steps. By using a workflow system, this complex setup is simplified. The fourth citation under Publications and Data Sets (below) provides results demonstrating scalability and early adaptive execution of ExTASY on Blue Waters. ExTASY is a Python library supporting domain scientists at Rice University. Our code is open source and provided at https://github.com/radical-collaboration/extasy-grlsd.

The ICEBERG project aims to address the following challenge: methods of using pixel-based classification models to extract feature representations for low- and medium-resolution images provide limited use of classification for high-resolution imagery. ICEBERG will provide deep-learning workflows for classification of high-resolution satellite imagery, and its integration with other data streams (e.g., Digital Elevation Models). The code of this work may be found at https://github.com/iceberg-project.

RESULTS & IMPACT

The first and second citations under Publications and Data Sets (below) provide results demonstrating scalability and early adaptive execution of HTBAC on Blue Waters. In addition, we were recently awarded the 2018 IEEE SCALE Challenge for using HTBAC on Blue Waters to explore simulations that calculate the strength of drug binding in order to reduce costs and improve computational efficiency for drug design and to underpin development in personalized medicine. We were able to extend the previous version of ExTASY to use GPUs and also increased its flexibility with regard to what kind of adaptive sampling algorithms can be used. So far, studies on Blue Waters have helped to pinpoint the most critical parameters to ExTASY performance and to improve sampling techniques. In the coming months, full-scale simulations of Chignolin, using ExTASY, will be completed in attempts to show that these methods can reproduce the work of [1].

WHY BLUE WATERS

Blue Waters is essential for our computational campaign in order to reach the scales needed to make a clinical impact. HTBAC and ExTASY require a scalability to at least 2,000 nodes per use-case. The full use-case of ICEBERG requires scalability up to 1,000 GPU nodes to classify images using deep neural networks.

PUBLICATIONS & DATA SETS

Dakka, J., et al., High-throughput binding affinity calculations at extreme scales. *Computational Approaches for Cancer Workshop at Supercomputing17* (2017).

Dakka, J., et al., Concurrent and Adaptive Extreme Scale Binding Free Energy Calculations. https://arxiv.org/pdf/1801.01174.pdf (2018).

Dakka, J., et al., Enabling Trade-off Between Accuracy and Computational Cost: Adaptive Algorithms to Reduce Time to Clinical Insight. SCALE Award Winner, *IEEE/ACM CCGrid'18* (2018).

Balasubramanian, V., et al., ExTASY: Scalable and flexible coupling of MD simulations and advanced sampling techniques. *Proceedings of the 2016 IEEE 12th International Conference on e-Science*, (March 3, 2017), pp. 361–370.