CATEGORY: LIFE & MATERIAL SCIENCE - LS04
POSTER P5191
CONTACT NAME
Yun Ding: yding8@lsu.edu

GPU TECHNOLOGY CONFERENCE

GeauxDock: an ultra–fast molecular docking package for computer-aided drug discovery

Yun Ding, Ye Fang, W. Feinstein, D.M. Koppelman, J. Moreno, J. Ramanujam, M.Brylinski, M. Jarrell
Louisiana State University, Louisiana Alliance for Simulation-Guided Materials Applications

LA-SiGMA Louisiana Alliance for Simulation-Guided Materials Applications

LSU

BACKGROUND

Computational modeling of binding drug to proteins has become an integral component of modern drug discovery pipelines. A typical application is structure-based virtual screening, which involves a large-scale modeling of pharmacological relevant associations between small molecules and their macromolecular targets (Fig. 1 and Fig. 2). The desire to improve state-of-the-art motivated us to develop an ultra-fast ligand docking approach that uses Monte Carlo as the sampling method and features computations on modern supercomputers. Combined with an effective scoring function, this new method will provide accurate predictions at a high performance/cost ratio, which is a critical factor for large-scale virtual screening applications.

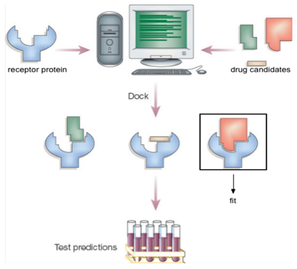

Figure 1. Computer-aided drug development holds a significant promise to speed up the discovery of novel pharmaceuticals at reduced costs.

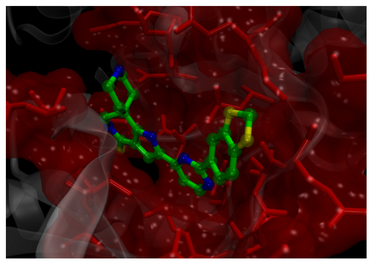

Figure 2. Docking simulations predict the native pose of the ligand by searching for the global minimum in the energy space.

SCORING FUNCTION

The native-like recognition capacity of the scoring function was optimized by maximizing the Z-score calculated across the training dataset. The Z-score is defined as:

$$Z-score=\frac{\overline{E_L}-\overline{E_H}}{\sqrt{\sigma_L^2+\sigma_H^2}}$$

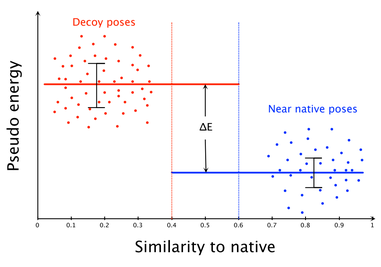

Figure 3. The native-like recognition capability of our force field is optimized by maximizing the Z-score.

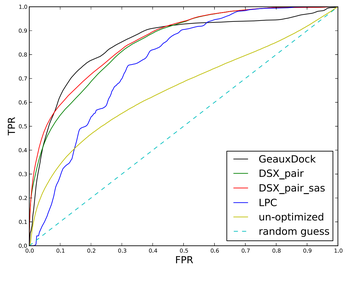

Figure 4. Receiver operating characteristic (ROC) analysis for the recognition of native-like conformations in docking ensembles by GeauxDock compared to other scoring functions. TPR – true positive rate, FPR – false positive rate.

CASE STUDY

One example of successful pose identification is shown to demonstrate the accuracy of our approach; GeauxDock was used to find the near-native pose of the inhibitor in Cathepsin K - P43235.

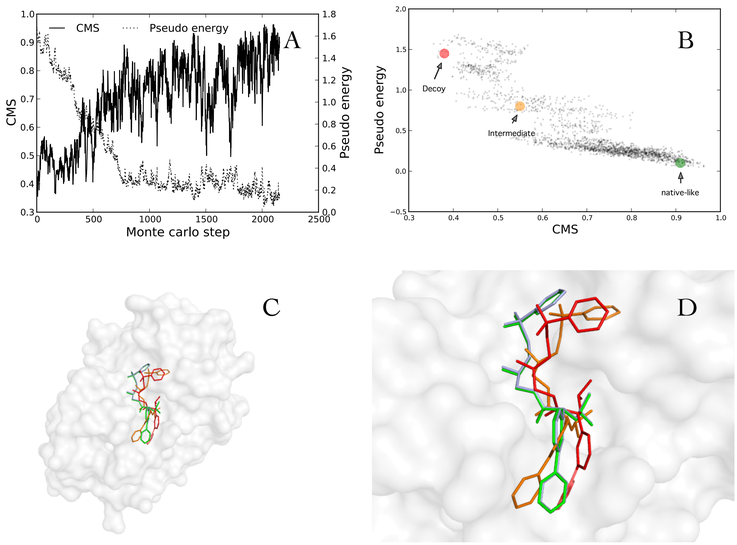

Figure 5. Docking simulation for Cathepsin K - P43235. From up left to down right. (A) Decreasing pseudo energy drives system to higher CMS regions during the Monte Carlo simulation. (B) Scatter plot of pseudo energy vs CMS. Three typical data points were selected from the simulation trajectory and their corresponding configurations (same color) are plotted in (C). (D) Close-ups. The crystallographic configuration is light blue.

BENCHMARKS

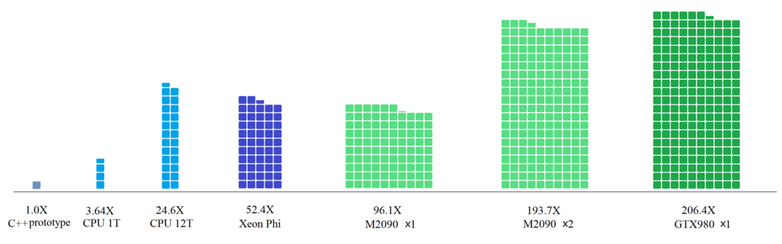

Figure 6. Performance comparison on different platforms. The benchmarks are carried on four platforms, a 6 core Xeon E5-2620 CPU, a Xeon Phi 3210A, two Tesla M2090, and a GeForce GTX 980. We use intel compiler 15.0.0 and CUDA 6.5 with optimization flag –O3. The results demonstrate a considerable speedup using heterogeneous accelerators. Our early tuning of a Maxwell GeForce GTX 980 GPU provides a two-fold speed up over a Fermi Tesla M2090 GPU.

IMPLEMENTATIONS

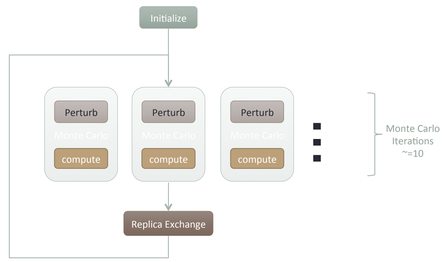

Figure 7. The program's flow chart. Many replicas are updated simultaneously using replica exchange Monte Carlo (REMC) algorithm.

The docking code is implemented with a highly efficient Replica Exchange Monte Carlo (REMC) sampling to take advantage of massively parallel capabilities provided by GPUs. In a nutshell, docking calculations exploits two levels of parallelism: coarse- and fine-grained. The former considers a simultaneous processing of many ligand configurations, which will be mapped to hardware thread blocks. The fine-grained level parallelizes the calculation of interactions for each individual configuration using multiple threads within each block.

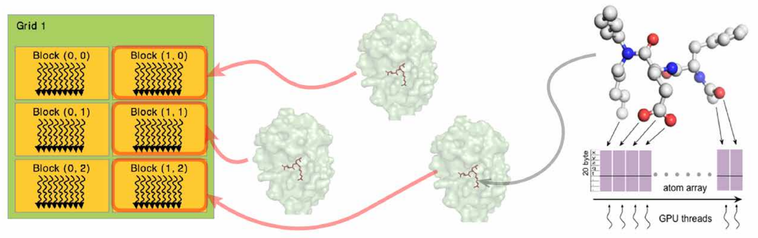

Figure 8. The mapping from computational resources to domain science objects**.** A protein-ligand replica is mapped to a GPU thread block; atom coefficients are stored in consecutive memory addresses, and mapped to GPU threads.

We implement and maintain fairly similar codes for three platforms: CPU-OpenMP, Phi-OpenMP, GPU-CUDA. The application front end loads data from input files, then apply strength reduction optimization and construct Struct of Array (SoA) data for GPU, or Array of Structure (AoS) data for CPU/Phi. A set of common wrappers for offload computing are developed to abstract the work flow on three different platforms: (1) data allocation, (2) copy-in, (3) kernel computation, and (4) copy-back, thus that CPU/Phi/GPU code can share the same high level infrastructure. While the low level kernel code for different architecture are forked from a highly optimized sequential code, and specifically tuned for the parallel performance.

CONCLUSIONS

GeauxDock provides sufficient accuracy for its application in large-scale virtual screening projects to support the reconstruction of protein-ligand interaction networks. GeauxDock adapts multiple commodity back-ends, among which, GPU provides considerable speedup. GeauxDock is freely available from our website: http://www.institute.loni.org/lasigma/package/dock/

This work is supported by the NSF EPSCoR LA-SiGMA project under award #EPS-1003897 and the Louisiana Board of Regents through the Board of Regents Support Fund [contract LEQSF(2012-15)-RD-A-05].