Highly Scalable Parallel implementation of Drug Discovery Applications

Amit Upadhyay
Joint Institute for Computational Sciences,
University of Tennessee-Oak Ridge National Laboratory,
Oak Ridge, TN, USA
aupadhy1@vols.utk.edu

Bhanu Rekepalli
Joint Institute for Computational Sciences,
University of Tennessee-Oak Ridge National Laboratory,
Oak Ridge, TN, USA
brekapal@vols.utk.edu

ABSTRACT

The average cost to bring a new drug in the market is $1.8 billion which is increasing rapidly. In addition there has been a steady decline in the approval of new drugs every year. Majority of the drug candidates fail in the clinical trial phases which account for 63% of the cost. Thus it is important to accurately identify drug candidates in the initial stages of the drug discovery pipeline. Virtual screening of large compound libraries is a common technique that is widely used in pharmaceutical industries for hit identification as well as lead optimization. Virtual screening involves docking of compounds into the binding site of receptor molecules of biological interest in order to determine the best pose as well as the binding affinity of the compounds. However, its use is challenged by the increasing database size of the available protein structures as well as the compound libraries. The Protein Data Bank currently contains 3D structures for more than 99,000 proteins and with the initiatives such as Structural Genomics Initiative, this number is expected to increase rapidly. ZINC, a database of commercially available compounds for virtual screening, presently contains over 35 million compounds. Docking of such large number of molecules necessitates availability of tools that can make use of high performance parallel computing. DOCK6 is one of the most popular tools for docking. An MPI implementation is available but its efficiency steadily declines with increasing number of cores. We have earlier developed a Highly Scalable Parallel Wrapper (HSP-Wrap) that was successfully used to scale BLAST, HMMer and MUSCLE to tens of thousands of cores. HSP-Wrap makes use of shared memory to efficiently manage I/O in addition to efficient load balancing. In this study we describe a parallel implementation of DOCK6, using HSP-Wrap. We carried out weak and strong scaling studies for docking of multiple copies of the ligand nevirapine (ranging from 5760 to 1 million) with the receptor Human Immunodeficiency Virus-1 (HIV-1) reverse transcriptase and found that HSP-Wrap DOCK6 shows efficient scaling. We were able to dock millions of compounds from the ZINC database in few hours using thousands of processors on XSEDE HPC resources when compared to months on a typical university based clusters. HSPDOCK thus provides a powerful means to carry out large scale virtual screening for finding drug candidates thus reducing the costs and time to market, thus getting a step closer to era of personalized medicine.

Keywords

Drug discovery, virtual screening, Dock6, docking, high performance computing, parallel, MPI