

IT.ENG.2

EVALUATING A DATA LEVEL PARALLELISM APPROACH IN DRUG DISCOVERY RESEARCH

U. Senanayake, A. Sivanathan, R. Ragel

*Department of Computer Engineering,
Faculty of Engineering, University of Peradeniya*

Bioinformatics has progressed into a widely explored research area due to the significant innovation it places on conventional methods. Structural Bioinformatics is a main branch where molecular interactions are studied at the structural level. This paper addresses a problem in molecular docking that has taken the experimental level molecular interaction to in-silico interactions. The need for this is closely associated with the exponential growth in the identified number of proteins and ligands.

Protein ligand docking can be considered as a lock and key problem where we have to find out the correct orientation of the ligand that can bind to the protein receptor. It is considered as an integral segment in drug designing which is known as lead identification. Repeated application of docking operation known as Virtual Screening (VS) is used for this purpose and *Autodock Vina* can be considered as standard software that is utilized by professionals. This paper explores an approach that utilizes data level parallelism and evaluates its scalability and performance on a clustered environment. The rationale behind this evaluation is to understand the feasibility of using this mechanism in large scale drug discovery research. As such, it should be scalable on top of a cluster to effectively materialize the full potential of the cluster.

The parallelizable component and serial component of *Autodock Vina* is experimentally derived using Amdahl's law and the throughput improvement is further evaluated according to the Gustafson's principle. By deriving the parallelizable fraction, one can understand the lower bound of the throughput enhancement that can be introduced by parallelizing *Autodock Vina*. It is also implied that the molecular docking problem cannot be modelled using Gustafson's principle. This is an important segment in academic research where it provides a solid theoretical interface between the implementation and theoretically derived implications. A complete scalability evaluation was carried out by observing the aforementioned approach in a clustered environment and measuring the individual and collective elapsed times with respect to different conditions. The results were then compared with the implications derived by Amdahl's law and Gustafson's principle to understand the applicability of the relevant models.

Hence it is rational to conclude that the DLP approach in drug discovery research can be used in large scale virtual screening processes carried out around the world therein making an important contribution to drug discovery research.