DOI 10.18699/MM-HPC-BBB-2018-13

HEDGE: Highly accurate GPU-powered protein-protein docking pipeline

T. Ermak*, A. Shehovtsov, P. Yakovlev
BIOCAD, Saint Petersburg, Russia
** e-mail: timofei.ermak@live.com*

Key words: protein-protein complexes prediction, docking, GPU, HPC, *in silico* drug design

Motivation and Aim: protein-protein interactions play key roles in living systems functioning: cell signaling, immune system reactions, microelements transport and many other processes are based on protein-protein complexes functions. Thus, protein-protein complexes prediction is very important task especially in terms of drug discovery. For example, *in silico* optimization stages of antibody-based drug development process requires to solve the problem hundreds of times. To perform *in silico* optimization and increase drug candidates' quality, the docking problem must be solved with high accuracy in short time ranges. But it is one of the hardest structural bioinformatics problems due to large solution space (possible molecules orientations), big sizes of protein systems and infinite space of molecules conformations.
Methods and Algorithms: the pipeline of algorithms in our tool called HEDGE can be described as follows: 1) scanning translational solution space using FFT correlation theorem; 2) calculation of Gibbs free energy change (ΔG), we use own highly optimized implementation of OPLS [1] force field. 3) minimization of a complex energy, Polak-Ribière-Polyak conjugate gradient method [2] is used to solve optimization problem.
Each step of the pipeline above is well-parallelizable, so, we utilize the full power of GPUs (graphics processing units), that allows to scan huge solution space and select best with solid metric of Gibbs free energy change. Moreover, different rotations of molecules can be processed independently, therefore, multi-GPU mode is supported to scale linearly and achieve maximal performance on multi-GPU supercomputers.
Results: HEDGE was tested on a subset of CAPRI [3] dataset showing 80 % of correct predictions for different types of proteins. Time required for prediction of one complex in rigid mode is about 7 minutes on Tesla V100 GPU, while other well-known tools (e. g. PIPER [4]) require about 90 minutes on 8 CPUs. Flexible mode requires much more calculations and takes about 1.5 hours on Tesla V100. Thus, our tool is one of the world's fastest in the field.
Conclusion: we developed highly accurate highly performant protein-protein docking tool called HEDGE, that successfully used in modern drug discovery pipelines.

References

1. Robertson M.J., Tirado-Rives J., Jorgensen W.L. (2015) Improved peptide and protein torsional energetics with the OPLS-AA force field. Journal of chemical theory and computation. 11(7):3499-3509.
2. Polak E., Ribiere G. (1969) Note sur la convergence de méthodes de directions conjuguées. Revue française d'informatique et de recherche opérationnelle. Série rouge. 3(16):35-43.
3. Janin J. (2002) Welcome to CAPRI: a critical assessment of predicted interactions. Proteins: Structure, Function, and Bioinformatics. 47(3):257-257.
4. Kozakov D. et al. (2006) PIPER: an FFT-based protein docking program with pairwise potentials. Proteins: Structure, Function, and Bioinformatics. 65(2):392-406.