Integrating molecular databases in a proteome-wide simulation environment.

Bence Keömley-Horváth[1], Simone Rizzeto[2], Alex Szendrei[1], István Reguly[1], and Attila Csikász-Nagy[1]

[1]Pázmány Péter Catholic University, Faculty of Information Technology and Bionics, Budapest
[2]Viral Immunology Systems Program, Kirby Institute School of Medical Sciences, UNSW, Sydney

ABSTRACT

In the last few decades due to new high-throughput techniques, such as mass spectrometry, we gained a large amount of data about Protein-Protein Interactions (PPI) and Domain-Domain Interactions (DDI). This opened the opportunity for computational methods to predict protein complexes. Clusterization methods produce only qualitative information about the complexes even though they are formed dynamically in various parts of the cells. The number of complexes formed is limited by proteins which have many possible binding partners while present in low abundance.
Here we propose a new version of SiComPre[1], a simulation based protein complex prediction method, that is based on Gillespie's multiparticle algorithm.[2] This combines the classic Gillespie's algorithm with diffusion by splitting the simulation space into subvolumes (SVs) which are small enough that we can assume that the reactant are able to find each others. In given times the species diffuse between the SVs[2].
The previous SiComPre program has some weaknesses. One of these is that the model of the cell is only two dimensional. In the new version, we want to implement a three dimensional cell model, thus the diffusion will be much less limited. The membranes were represented as barriers, and through these only limited interactions could happen. We will handle it as special thinner double layer of sub-volumes, which have inner and outer sides, that is a much more realistic representation of the lipid bilayer of membranes.
A proteom-wide diffusion coefficient database is not available for any organisms, so we developed a method that predicts these constants from protein masses with the help of Stokes-Einstein equation.
We have implemented the core Gillespie algorithm that works on multiple sub-volumes and we already parallelized it on CPU. We are working on the implementation of the diffusion and we have to develop an applicable representation of the complexes. After that we will move it to GPU clusters for faster parallel computing.

References

1. Rizzetto S, Priami C, Csikász-Nagy A. *Qualitative and Quantitative Protein Complex Prediction Through Proteome-Wide Simulations.* PLoS Comput Biol 11(10): e1004424. doi:10.1371/journal.pcbi.1004424, 2015
2. Stundzia, A.B. and Lumsden, C.J. *Stochastic simulation of coupled reaction-diffusion processes.* Journal of Computational Physics (1996), 127, 196-207.
3. M. E. Young P. A. Carroad R. L. Bell *Estimation of diffusion coefficients of proteins* Biotechnology and Bioengineering, Vol. XXII, Pp. 947-955 (1980)