Poster J-10

Dedicated Computer System for Genome Wide Simulation of Signal Transduction Networks

Authors:
Tetsuya Maeshiro (*University of Tsukuba*)
Katsunori Shimohara (*Doshisha University*)

Short Abstract: We have developed a computer system dedicated to simulate and predict signal transduction networks in biological cells. The simulation speed is at least ten thousand times faster than software based simulations executed on conventional computers. We present a simulation result of a simple cell model, and the potentiality of our system.

Long Abstract:
Dedicated Computer System for Genome Wide Simulation of Signal Transduction Networks

Introduction

Starpack (Signal Transduction Advanced Research PACKage) is a dedicated computer system to simulate and predict biological signal transduction networks. Its execution speed is at least ten thousand times faster than conventional simulation systems. Computer simulation of biological phenomena is useful for system level understanding of biological systems described at molecular level, and drug screening, among others. Simulation of the whole network is important because interactions of individual components are ultimately responsible for an organism's form and functions. Several systems have been proposed for simulation, for instance E-Cell, but they lack the simulation speed required to treat real biological organisms, which has a large number of elements and interactions among elements. For instance, a human cell has about 30,000 types of genes and these genes produce substances that participate in a signal transduction network composed of around hundred thousand types of elements. Conventional simulation systems are software based, and even running on cluster systems consisting of thousands of processors, it is not possible to execute simulations in a practical time.

Simulation Performance

The system architecture of Starpack is completely different from ordinary computers. Designed specifically to simulate models described as large networks, Starpack is a scalable system with several thousands of dedicated processors (CPUs). Each processor stores the quantity of a substance, and interactions among substances are simulated by modifying the values that correspond to substances' quantities. Currently, Starpack has 4,096 processors.

Starpack is faster by orders of magnitude than software simulations. Conventional simulation systems model biochemical reactions as differential equations and solve the equations by numerical methods, which require enormous computational time. On the other hand, Starpack is based on direct models of biochemical reactions, analogous to the first principle

modeling of physical phenomena. This results in high speed, further accelerated by hardware encoding. Moreover, Starpack is also orders magnitudes faster than real biochemical reactions, thus suitable to perform `virtual experiments'. The speedup is significant for simulations of slow growing organisms such as plants.

The current architecture of Starpack is optimized to simulate two types of biological phenomena, both related to the prediction of signal transduction network structures: (i) Detailed biochemical reactions, modeled from reaction kinematics; (ii) Gene networks. The preliminary simulation of simplified erythrocite model presented here corresponds to the first category. Simulations of biochemical reactions focus mainly on variations of substances that participate in reactions. On the other hand, simulations of gene networks, the second category, hide biochemical details and focus on how genes are functionally interconnected. For instance, some genes are periodically activated and suppressed following circadian rythms, and one is interested in which genes trigger or silence other genes.

The two categories request different types of computational speed. The first category, deteiled biochemical simulations, requires speed to simulate reactions. However, the second category, gene network simulations, requires the speed to explore large number of network configurations. While only the genes are treated in the simulation of gene networks, simulation of biochemical reactions requires additional substances that participate in reactions.

As a preliminary test of simulation of biochemical reactions, a simple model of human erythrocyte was used, where 142 substances participate in 45 reactions. Although this network is smaller than real biological reaction networks, it enables the comparison of speed and precision with software based simulations. Same model was simulated on a software based simulation system (E-Cell). An advantage of Starpack is its ability to simulate with the precision of real reactions, which is practically impossible with software simulation because of unrealistic computation time that is required. Starpack was one million times faster than the software simulation for the same simulation precision. E-Cell was executed on a single PC with high performance configuration, but even a cluster system with thousand PCs cannot fill the gap of six orders of magnitude, further aggravated by the problem of interprocess communication latency.

For the simultation of the second category, the simulation of gene networks, the strategy is (i) execute low precision but fast simulations to explore parameter spaces; then (ii) execute high precision simulation to improve the accuracy. Primary input data for the simulation of gene networks is the microarray experiment data. Microarray experiments provide analysis of the whole organism, suitable for the objective of Starpack, which is the system level simulation of organisms. Microarray experiments generate time course variation of quantity of genes, denoted as expression patterns. Then activation relationships, describing which gene triggers which genes, should be estimated. Each activation pattern gives a different network structure. A widely used method to analyze microarray is the clustering, which groups genes based on similarity of expression patterns. However, clustering algorithms indicate only the binary relationships among genes. Due to its high speed, Starpack explores possible activation combinations among genes, which corresponds to generating network structures. Again, conventional simulation systems lack the performance for this type of simulation.

Conclusions

The preliminary simulation of biochemical reactions indicates clear advantage of Starpack over conventional systems, as Starpack is one million times faster than the software simulation for the same simulation precision. Such improvement by orders of magnitude considerably reduces the turn-around-time of simulation experiments, and opens possibilities for new research methodologies, as numerous parameter sets can be tested simultaneously. This advantage becomes larger for real networks found in biological organisms, which have more substances and more complex signal transductions.