WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Brief Survey on Biological Simulation Tools

Nayana. G. Bhat[1], Omkar. R. Deshpande[2], K. J. Sumedh Hebbar[3], Nitinkumar. S. M[4], Akshaykumar. J. G[5]

[1-]Assistant professor, Dept of CSE, Jyothy Institute of Technology
[2,3,4,5-]Student, Dept of CSE, Jyothy Institute of Technology

ABSTRACT—

The Lac Operon in the E. coli is the main component that is used to simulate the reactions happening in the cell. The stochastic simulation of the reaction taking place in the lac operon requires high computational power and thus it takes a large amount of time to produce the desired output. There are multiple tools that can be used to simulate biological reactions such as BioPSiS [1], Smoldyn [9], Dizzy [2] etc. These tools can be used in high performance computing environment to get the desired output via parallelization techniques. This paper explores various such tools that are used for the purpose of simulation. The main aim of this survey is to reduce the simulation time of the reaction without decreasing its accuracy.

Keywords—Lac Operon, E. coli, stochastic, parallelization

INTRODUCTION

Overview of Bioinformatics

The Bioinformatics is the study of biological data that is obtained by applying various tools to the biological reactions taking place in the environment. Methods and software tools for understanding biological data, particularly when the data sets are large and complex, are developed in the multidisciplinary field of bioinformatics. In today's biology and medicine, data management relies heavily on bioinformatics.

The data obtained from the simulation of these reactions are used for analysis and research purpose. Furthermore, these analytics of biological data can be used to manufacture precise medicines. A bioinformatician's primary tools are computer software and the internet. Computer programs like BLAST and Ensembl, which aid in the simulation of biological reactions, are included in the bioinformatics toolbox.

There are various applications in the field of bioinformatics. The primary focus of bioinformatics and its application is the extraction of useful facts and figures from a natural-world data set. Bioinformatics is used in a wide range of areas, including drug development, image analysis, 3D cell modeling, and 3D image processing. The most significant application of bioinformatics can be seen in medicine, where we heavily rely on its data to develop therapeutics for infectious and harmful diseases.

E. coli

Escherichia coli, more commonly referred to as E. coli, is a Gram-negative, pole-shaped, anaerobic, coliform bacterium of the Escherichia variety that is frequently found in the lower digestive tract of warm-blooded animals. It is generally harmless and quite good for the organism's health. However, there are variants of these bacteria that can cause severe food poisoning and are extremely harmful.

One of the bacteria that can harm humans is known as E. coli. If infected, these kinds of bacteria can result in diarrhea, kidney failure, and even death. E. coli can infect humans in a number of ways, including by coming into contact with infected animals and drinking or swimming in infected water. E. coli's survival depends heavily on water as a medium. E. coli bacteria can grow and live in a variety of water types at varying temperatures. Basically, high temperatures are more likely to contain E. coli than low temperatures [16].

E. coli is known to cause pneumonia and urinary tract infections, the latter of which can result in serious health complications, in addition to food poisoning. Shiga, a potent toxin produced by certain strains of E. coli bacteria, is one of the primary causes of illness. The Shiga toxin has the potential to harm the lining of the intestine, which can eventually result in a wide range of other diseases. As a result, the specific strains of E. coli that produce this toxin are referred to as Shiga Toxin-producing E. coli (STEC).

Lac Operon

The lac operon, first depicted by Jacob and Monod (1961), codes for the cell apparatus to ship and process lactose. In its environment, E. coli encounters numerous different sugars. For the metabolism of these sugars, like lactose and glucose, distinct enzymes are required.

The lac operon contains three enzymes involved in lactose metabolism: lacZ, lacY, and lacA. LacZ contains the gene for the enzyme galactosidase, which breaks down lactose into its two sugars: galate and glucose.

To get lactose into the cell, the first lactose permease, a PMF-mediated transport protein called LacY, must be expressed and embedded in the cell membrane. The production of the enzyme b-galactosidase (LacZ) is next. Before the disaccharide can be catabolized by the cell, this breaks it down into its two simple sugars, glucose and galactose.

Escherichia coli and a few other enteric bacteria require the lac operon for lactose transport and metabolism. Numerous factors, including glucose and lactose availability, regulate the lac operon. One of the most prominent examples of prokaryotic gene regulation is the lac operon's gene regulation, which was the first complex genetic regulatory mechanism to be elucidated [15].

Gene regulation proteins have the ability to turn on and off gene transcription. One illustration of this dual control is the lac operon in E. coli. The lac operon's initiation of transcription, or whether it is switched "ON" or "OFF," is controlled by glucose and lactose levels.

Three structural genes, a promoter, a terminator, a regulator, and an operator make up the lac operon. The three underlying genes which code for polypeptides of around 1000, 260, and 275 amino acids are : The lacZ, lacY, and lacA.

- The intracellular enzyme lacZ, which cleaves the disaccharide lactose into glucose and galactose, is encoded by lacZ.
- LacY is a membrane-bound transport protein that transports lactose into the cell and is encoded by lacY.
- LacA is an enzyme that transfers an acetyl group from acetyl-CoA to β-galactoside. This enzyme is called lacA.

Another gene, lacI, or simply "I," is located close to the lac operon. It encodes the protein lac repressor, which is necessary for controlling lac operons. The lac repressor gene is expressed "constitutively," which indicates that it is active at all times (though at a low level) [15].

LITERATURE SURVEY

Computational, modeling, and visualization tools for multidimensional simulations have become essential for the creation of new models in cellular electrophysiology that have significant computational requirements. Computational cellular electrophysiology has seen the presentation of numerous tools [1].

The simulation can be carried out using a variety of methods with the help of numerous available tools. Over the past 30 years, a number of stochastic methods for simulating biochemical reactions have been developed [2]. The tools are helpful in simulating complex biological reactions and understanding them.

The tools make use of Ordinary Differential Equations (ODE) such as the *Chemical Master Equation* and *Reaction Diffusion Master Equation (RDME)* to carry out the simulations.

Chemical Master Equation

The chemical master equation, which is a special case of the forward Chapman Kolmogorov equation [17], is perhaps the stochastic model for biochemical systems that is the most widely accepted one [18].

According to the Chemical Master Equation (CME), the probability that a simulation will be in a particular state is determined by how the number of each modeled species changes over time [3]. An example for the CME is shown in equation 1. Based on Gillespie's Stochastic Simulation Algorithm (SSA) [4], several methods and extensions for sampling the CME have been developed, including the τ-leaping method [5], composite rejection method, and optimized direct method [6], among others.

For a system to be eligible for the CME, a number of conditions must be met. The first is the "well-stirred" assumption, which assumes that, similar to an ODE system, the reactions happen much more slowly than diffusion. Another is that each reaction is a distinct event that can be referred to as a Markovian process [3]. A spatially-resolved approach is required to investigate scenarios that relax the first of these conditions.

Reaction Diffusion Master Equation

A less well-known technique for modeling chemical reactions in slow diffusion conditions is the RDME [7]. When diffusion and reaction occur on timescales that are comparable enough to allow for species concentration gradients in the system, the Reaction-Diffusion Master Equation (RDME) is required to describe the time evolution of spatially resolved states.

By dividing the system volume into discrete sub volumes with molecules diffusing between the sub volumes and reacting only with other molecules in the local sub volume, the RDME extends the CME's master equation formalism to account for spatial degrees of freedom.

Elf and Ehrenberg invented the next-sub volume method for precisely sampling the RDME [8]. The list of sub volumes is sorted by the time of their subsequent diffusion or reaction event using this strategy, which makes use of a priority queue resembling a next reaction. The standard Gillespie direct method is used to identify the specific reaction or diffusion event that occurred in the sub volume after a sub volume has been chosen for an event.

Tools and Methods

Numerous tools have been developed to facilitate the simulation of a complex biological system. A portion of those strategies likewise utilize the GPU power and parallelization procedures to accomplish the ideal result.

BioPSiS(Biological Process Simulation System) [1] is one of the tools that is used to simulate and visualize biological reactions. It is a computational framework addresses two distinct requirements: the requirement for an easy-to-use tool for evaluating the results and the requirement for real-time monitoring of the simulation processing.

In BioPSiS, the mathematical calculations can be done in a distributed or simple processing unit. An interactive tool is provided by the subsystem's functionality as demonstrated here. The evaluation of computational efficiency and the analysis of the simulation results are both made easier by this intuitive visualization subsystem.

Smoldyn is a widely used algorithm for stochastic simulation of chemical reactions with single molecule detail and spatial resolution [9]. An innovative alternative is proposed that takes advantage of the parallelism of GPUs.

A diffusion-influenced systems-specific modified Smoluchowski model is used in the Smoldyn BD method. Smoldyn employs a few distinctions from the Smoluchowski model: Smoldyn approximates time using discrete time steps of fixed length to make the simulation's computational aspects easier to understand.

The Smoldyn algorithms makes use of the GPUs to perform simulations parallelly. The NVIDIA GPU Architecture [10] is used in the algorithm. With the introduction of the Compute Unified Device Architecture (CUDA) [11], NVIDIA was the first company to specifically address GPU computing.

Dynamic load balancing for workstations with GPUs of varying performance and memory capacity is supported by a novel spatial decomposition-based multi-GPU parallel implementation of the MPD-RDME method [12]. For peer-to-peer GPU memory transfers and evaluating our algorithms' performance on cutting-edge GPU devices, the method makes use of CUDA's high-performance features.

Another method for effectively simulating a genetic switch system in a eukaryotic cell with 37 species, 75 reactions and millions of metabolite particles is the implementation of the hybrid CME–ODE algorithm [13], which is now compatible with LM/pyLM [14].

The setup conditions for hybrid simulations of much more computationally intensive, spatially resolved whole cell RDME studies can be influenced by the outcomes of these effective hybrid CME–ODE simulations. The multiple-GPU computation and optimized propensity calculation that were developed for RDME simulations can be used in hybrid simulations without the user having to do anything extra.

There exist three additional techniques for recreating the biological systems. Those devices are known as the Dizzy, Systems Biological Toolbox and Copasi. The Gillespie's Stochastic Simulation Algorithm (SSA) [4] is used to test the tools.

Dizzy is a simulation software package that is written in Java. The Dizzy package has four ways to simulate the reactions: two ways of SSA [4] (the direct method and the next reaction method) and two types of Gillespie's τ-leap algorithm [5]. Dizzy is capable of displaying the models graphically. The output can be plotted, listed as tables or saves as .csv files. The error messages that appear during the simulation are not clear enough. The output plot does not have labels and also does not provide zooming options for the users. The parameter estimation and sensitivity analysis options are not available.

Systems Biology Toolbox comes under the MATLAB [20] software package. It is mostly used for stochastic simulations rather than the deterministic simulations. The output can be plotted or stored in the MATLAB structured arrays. The error messages that appear during the simulation are not very clear. The parameter estimation and the sensitivity analysis options are provided by the Systems Biology Toolbox.

The Copasi is a simulation software that is used through a Graphical User Interface (GUI) and hence it is easier for the users. Whenever the user sets up the reaction, Copasi automatically determines the reactants from the reaction. The output obtained can be plotted as a graph or saved in an

ASCII file. The error messages are not very informative to the user. The Copasi provides an output assistant that helps in plotting. Copasi also provides additional tools such as the parameter estimation, sensitivity analysis and the stochiometric analysis.

OBJECTIVES

This study's primary objectives are

1) The tools have to be identified and designing small models for sample data.
2) To design precise model by making use of available whole-cell model software and tools.
3) To represent functions and characteristic of Lac-operon by modelling and simulating.
4) To speed up the model simulations using high-performance heterogeneous parallelized computing technology.
5) To validate the new model with experimental data.

PROPOSED METHODOLOGY

Analyze the chemical reactions in the cell

Through decades of biochemical research, the majority of cellular reactions' kinetic parameters have been measured in closely related organisms. In order to construct the mathematical model, these parameters need to be thoroughly examined.

Build the mathematical model for the reactions

Utilizing the kinetic parameter analysis, the numerical model for the reactions is constructed utilizing the equations known as the Chemical Master Equations (CME) and the Reaction Diffusion Master Equations (RDME).

Simulation Execution, Output and Analysis

To precisely test the measurements of a CME or RDME model, one requires to create numerous autonomous trajectories. To work with this cycle on huge process groups, every autonomous trajectory is figured out utilizing equal projects. The yield trajectories from each program can be melded into a solitary HDF5 (Hierarchical Data Format Version 5) file format. Tools like MATLAB, Lattice Microbes and so on can be utilized for reproduction.

Speeding up the simulation

The numerical model of the substance responses is given as contribution to the simulation algorithms for testing. Normally, approaches for precisely inspecting the CME are variations of stochastic simulation algorithm (SSA). Every response in the framework is treated as a singular response. These algorithms can be worked on such that the simulation time is decreased.

Visualization

There exist a few perception programs that can be utilized to see the sub-atomic unique simulations. These projects have progressed data structures and algorithms that can envision cellular models and quicken cellular simulation trajectories. Tools, for example, VMD can be utilized to picture the responses

CONCLUSIONS

The biological systems are extremely complicated to understand and require computational systems that provide assistance in the simulation of these systems. At first, the mathematical model for the reactions is composed utilizing the Chemical Master Equation (CME) and the Reaction Diffusion Master Equation (RDME). Then the equations are given to fitting apparatuses or tools with the end goal that the simulation can occur.

It is fundamental for these systems to utilize profoundly strong GPU framework to do the simulation method. It is important to lessen the simulation season of these reaction so it can help the exact medication producing process. To make the simulation quicker, the algorithms should be grown with the end goal that they utilize high computational power as well as different parallelization techniques to simulate the biological reactions.

Thus, this paper shows various tools that are already being used by scientists to simulate biological processes to understand the mechanisms lying under it. The paper also points out the differences between various such tools. The survey also focuses on the equations such as Chemical Master Equation (CME) and the Reaction Diffusion Master Equation (RDME).

References

Petsios SK, Fotiadis DI. *A computational framework for the analysis of biological models.* Annu Int Conf IEEE Eng Med Biol Soc. 2007;2007:1101-4. doi: 10.1109/IEMBS.2007.4352488. PMID: 18002154.

Manninen T, Mäkiraatikka E, Ylipää A, Pettinen A, Leinonen K, Linne ML. *Discrete stochastic simulation of cell signaling: comparison of computational tools.* Conf Proc IEEE Eng Med Biol Soc. 2006;2006:2013-6. doi: 10.1109/IEMBS.2006.260023. PMID: 17945691.

Peterson, Joseph & Hallock, Michael & Cole, John & Luthey Schulten, Zaida. (2013). *A Problem Solving Environment for Stochastic Biological Simulations*. 10.13140/2.1.3207.7440.

D. Gillespie, "Exact Stochastic Simulation of Coupled Chemical Reactions," J. Phys. Chem., vol. 81, no. 25, pp. 2340–2361, 1977.

M. Rathinam, L. Petzold, Y. Cao, and D. Gillespie, "*Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method*," J. Chem. Phys., vol. 19, no. 24, pp. 12 784–12 794, 2003.

A. Slepoy, A. Thompson, and S. Plimpton, "*A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks,"* J. Chem. Phys., vol. 128, no. 20, p. 205101, 2008.

Roberts, Elijah & Stone, John & Luthey Schulten, Zaida. (2013). *Lattice Microbes: High-Performance Stochastic Simulation Method for the Reaction-Diffusion Master Equation. Journal of computational chemistry*. 34. 10.1002/jcc.23130.

Hattne J, Fange D, Elf J. Bioinformatics. 2005; 21:2923–4. [PubMed: 15817692]

L. Dematte, "*Smoldyn on Graphics Processing Units: Massively Parallel Brownian Dynamics Simulations,*" in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 9, no. 3, pp. 655-667, May-June 2012, doi: 10.1109/TCBB.2011.106.

E. Lindholm, J. Nickolls, S. Oberman, and J. Montrym, "*Nvidia Tesla: A Unified Graphics and Computing Architecture,*" IEEE Micro, vol. 28, no. 2, pp. 39-55, Mar./Apr. 2008.

J. Nickolls, I. Buck, and M. Garland, "*Scalable Parallel Programming with Cuda,*" Queue—GPU Computing , vol. 6, no. 2, pp. 40-53, 2008.

Hallock, Michael & Stone, John & Roberts, Elijah & Fry, Corey & Luthey Schulten, Zaida. (2014). *Simulation of reaction diffusion processes over biologically relevant size and time scales using multi-GPU workstations*. Parallel Computing. 40. 86-99. 10.1016/j.parco.2014.03.009.

Bianchi DM, Peterson JR, Earnest TM, Hallock MJ, Luthey-Schulten Z. *Hybrid CME-ODE method for efficient simulation of the galactose switch in yeast*. IET Syst Biol. 2018 Aug;12(4):170-176. doi: 10.1049/iet-syb.2017.0070. PMID: 33451183; PMCID: PMC8687183.

Roberts, Elijah & Stone, John & Luthey Schulten, Zaida. (2013). *Lattice Microbes: High-Performance Stochastic Simulation Method for the Reaction-Diffusion Master Equation*. Journal of computational chemistry. 34. 10.1002/jcc.23130.

K. Kosev, P. Melo-Pinto and O. Roeva, "*Generalized net model of the lac operon in bacterium E. coli*," 2012 6th IEEE International Conference Intelligent Systems, 2012, pp. 237-241, doi: 10.1109/IS.2012.6335224.

S. Z. Aziz, M. F. Jamlos and M. A. Jamlos, "*Escherichia coli detection in different types of water*," 2014 IEEE Symposium on Wireless Technology and Applications (ISWTA), 2014, pp. 125-129, doi: 10.1109/ISWTA.2014.6981170.

NG Van Kampen, "*Stochastic processes in physics and chemistry*", North Holland, 2007.

E. Yeung, J. L. Beck and R. M. Murray, "Modeling environmental disturbances with the chemical master equation," 52nd IEEE Conference on Decision and Control, 2013, pp. 1384-1391, doi: 10.1109/CDC.2013.6760076.

T. Manninen, E. Makiraatikka, A. Ylipaa, A. Pettinen, K. Leinonen and M. -L. Linne, "*Discrete stochastic simulation of cell signaling: comparison of computational tools*," 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, 2006, pp. 2013-2016, doi: 10.1109/IEMBS.2006.260023.

L. Yu, "*Matlab Programming Environment Based on Web*," 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), 2018, pp. 509-512, doi: 10.1109/ITOEC.2018.8740716.