Bioinformatics, 36(18), 2020, 4825–4826
doi: 10.1093/bioinformatics/btaa602
Advance Access Publication Date: 2 July 2020
Applications Note

Systems biology

RSSALib: a library for stochastic simulation of complex biochemical reactions

Vo Hong Thanh

Department of Computer Science, Aalto University, Finland

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on February 15, 2020; revised on April 24, 2020; editorial decision on June 17, 2020; accepted on June 20, 2020

Abstract

Motivation: Stochastic chemical kinetics is an essential mathematical framework for investigating the dynamics of biological processes, especially when stochasticity plays a vital role in their development. Simulation is often the only option for the analysis of many practical models due to their analytical intractability.

Results: We present in this article, the simulation library RSSALib, implementing our recently developed rejection-based stochastic simulation algorithm (RSSA) and a wide range of its improvements, to accelerate the simulation and analysis of biochemical reactions. RSSALib supports reactions with complex kinetics and time delays, necessary to model complexities of reaction mechanisms. Our library provides both an application program interface and a graphic user interface to ease the set-up and visualization of the simulation results.

Availability and implementation: RSSALib is freely available at: https://github.com/vo-hong-thanh/rssalib.

Contact: thanh.vo@aalto.fi

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Biochemical processes at the cellular level are intrinsically stochastic due to the discreteness of species and the randomness of reaction firings, leading to significant fluctuation in the cellular response. Stochastic chemical kinetics describes the stochastic dynamics of biochemical reactions through the chemical master equation. Stochastic simulation in many cases is the only approach to study the temporal dynamics of biological systems due to the high-dimensional state space (Marchetti *et al.*, 2017). Stochastic simulation of biochemical reactions poses many computational challenges not only due to the size of networks but also due to complex reaction mechanisms. Reactions with non-linear rate laws, such as enzymatic kinetics, often applied to model biochemical reactions to better match the experiments. In addition, reactions always take a certain time, called *delay*, from their initiation to finish. For many biological processes, such as the transcription and translation, where their completion time is slow, it is necessary to consider time delays to accurately describe the system dynamics. Finally, due to the stochastic nature of the approach, many simulation runs must be performed to obtain a reasonable statistical estimation of the expected behaviour of the system dynamics.

We present in this article, the simulation library RSSALib to offer the computational advantages of our recently developed rejection-based simulation algorithm (RSSA) (Thanh *et al.*, 2014). We implement RSSA and a wide range of improvements to cope with different aspects of biological processes. RSSALib supports reactions with complex kinetics, including Michaelis–Menten and Hill kinetics. It also allows reactions with time delays required to model complicated biological phenomena. The biological network can be described in our reaction format or imported directly from an SBML model. Our simulation library provides both an application program interface (API) for stand-alone applications and a graphic user interface (GUI) to ease the set up of simulation and visualization of results. We validate our implementation and demonstrate its applications on real biological models (see Supplementary Material for more details) to highlight its applicabilities and computational improvements in simulation performance of our computational tool.

2 RSSALib

The simulation library RSSALib provides a full implementation of all known RSSA formulations to offer their computational advantages in dealing with varying complexities of biological networks.

2.1 Theoretical background

RSSA is an exact simulation (see Thanh *et al.*, 2014, for a formal proof) that accelerates performance by reducing the average computations of reaction propensities. We consider a biological network consisting of N species S_i, $i = 1, \ldots, N$, interacting through M reactions R_j, $j = 1, \ldots, M$. The populations of species constitute the state $X(t)$. The probability that reaction R_μ fires in the infinitesimal time $[t, t + \tau + d\tau)$, given the state $X(t)$, is $p(\tau, \mu)d\tau$ in which the probability density function (pdf) $p(\tau, \mu) = a_\mu \exp\{-a_0\tau\}$ where a_j is called

reaction *propensity* and $a_0 = \sum_{j=1}^{M} a_j$. Instead of directly sampling pdf $p(\tau, \mu)$ using the propensity a_j, RSSA uses the bounds $[\underline{a}_j, \bar{a}_j]$ and the rejection-based technique. The selection of the next reaction firing in RSSA consists of two steps. First, a candidate reaction R_μ is selected with probability $\bar{a}_\mu/\bar{a}_0$ where $\bar{a}_0 = \sum_{j=1}^{M} \bar{a}_j$. The candidate reaction R_μ then enters a rejection-based test for validation with success probability $a_\mu/\bar{a}_\mu$. RSSA avoids computing propensity a_μ in this step as much as possible using the fact that if R_μ is accepted with probability $\underline{a}_\mu/\bar{a}_\mu$, then it is also accepted with probability $a_\mu/\bar{a}_\mu$. If R_μ is accepted, its firing time is then computed. The firing time τ of the accepted reaction R_μ is chosen following an *Erlang* distribution to ensure exactness of the selection specified by pdf $p(\tau, \mu)$.

2.2 Usage and implementation

Figure 1 shows the use of RSSALib's GUI to simulate and visualize the simulation result. The GUI allows one to load the model and perform simulation with a click-and-run. RSSALib can also be used as a developer API for building stand-alone applications. For this usage, we manually load the model, and call the runSim() method of the simulator to execute the simulation. In the following, we briefly describe the simulation algorithms implemented in RSSALib and their time complexities.

- RSSA and its extension for reactions with time delays (DelayedRSSA) realize the candidate reaction R_μ by linearly accumulating propensity upper bounds until it finds the reaction. We also provide the cache-friendly search, which reuses the previously computed sum of propensities in the last step. The time complexity of the search is $O(M)$.
- Partial-propensity RSSA (PRSSA) uses the factorization of the mass-action propensity to factorize the propensity bounds $\underline{a}_j/\bar{a}_j$ of reactions, which are then grouped by the common reactant species into the so-called partial propensity structure. The selection of candidate reaction in the PRSSA is performed in two consecutive steps in which the first search selects a group, and the second one locates the reaction in that group. The time complexity of the search in PRSSA is proportional to the number of species, i.e. $O(N)$.
- RSSA with tree-based search (RSSA-Binary) uses the tree-based search to reduce the time complexity for selecting the candidate reaction. First, a tree is built in which its leave store propensities $\bar{a}_j$ of reactions and internal nodes store the sum value of their children. The search for the next reaction will travel from the root to a leaf to discover the next reaction. The search depth is equal to the height of the tree, which is $O(\log(M))$, hence its time complexity.
- RSSA with composition–rejection search (RSSA-CR) implements the composition–rejection method to reduce the time complexity of the search for the candidate reaction to be independent with the number of reactions. Reactions are partitioned into L groups G_i, $i = 1, \ldots, L$ so that a reaction R_j is put into a group G_i if its propensity $\bar{a}_j$ satisfies $2^{q_i - 1} \leq \bar{a}_j \leq 2^{q_i}$. The selection of the candidate reaction is made in two steps. First, a group G_l is selected proportional to the sum of propensity bounds of reactions in the group. Then, the reaction R_μ in the group G_l is located by applying the acceptance–rejection with hat function 2^{q_l}. The selection of the candidate by the composition–rejection search depends only on the number of groups, i.e. $O(L)$.
- RSSA with table lookup search (RSSA-Lookup) reduces the time complexity of the search to be constant, i.e. $O(1)$; however, it requires to build the lookup tables which take linear time $O(M)$. The M probabilities $\bar{a}_j/\bar{a}_0$, for $j = 1 \ldots M$, are partitioned into an equi-probable mixture of M two-point distributions and store these values in two tables, called *cut-off table*, storing the probability of the first values of the two-point mixtures, and *alias table*, containing the alias to the second parts of the mixtures. For the selection of the next reaction, a random number r_1 is first used to lookup the position of the equi-probable mixture. It is then rescaled to select which part of the two-point.

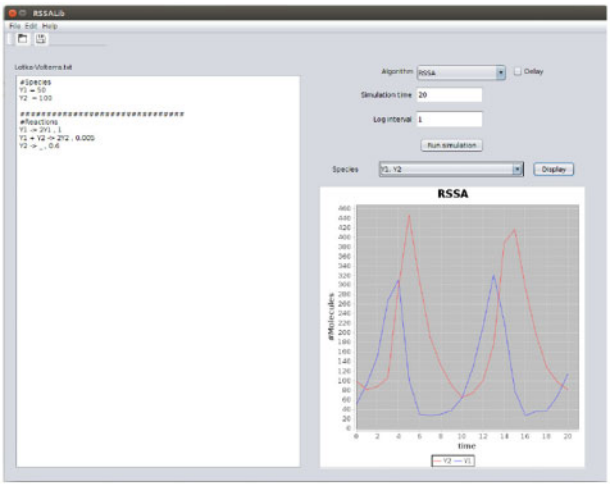

Fig. 1. GUI of RSSALib for setting up simulation and visualization

3 Conclusion

We presented the simulation library RSSALib, providing the implementation of the RSSA and a wide range of its improvements, to accelerate the simulation and analysis of biochemical reactions. Our computational tool enables investigating large, complex biological systems.

Funding

The research was supported by Academy of Finland [311639], 'Algorithmic Designs for Biomolecular Nanostructures (ALBION)'.

Conflict of Interest: none declared.

References

Marchetti,L. *et al.* (2017) *Simulation Algorithms for Computational Systems Biology*. Springer International Publishing, AG.

Thanh,V.H. *et al.* (2014) Efficient rejection-based simulation of biochemical reactions with stochastic noise and delays. *J. Chem. Phys.*, **141**, 134116.