OPEN ACCESS
Journal of Bioinformatics and Computational Biology
Vol. 20, No. 3 (2022) 2250018 (21 pages)
© The Author(s)
DOI: 10.1142/S0219720022500184

World Scientific
www.worldscientific.com

TemporalGSSA: A numerically robust R-wrapper to facilitate computation of a metabolite-specific and simulation time-dependent trajectory from stochastic simulation algorithm (SSA)-generated datasets

Siddhartha Kundu

Department of Biochemistry
All India Institute of Medical Sciences
Ansari Nagar, New Delhi 110029, India
siddhartha_kundu@yahoo.co.in; siddhartha_kundu@aiims.edu

Received 12 April 2022
Accepted 2 July 2022
Published 8 August 2022

Whilst data on biochemical networks has increased several-fold, our comprehension of the underlying molecular biology is incomplete and inadequate. Simulation studies permit data collation from disparate time points and the imputed trajectories can provide valuable insights into the molecular biology of complex biochemical systems. Although, stochastic simulations are accurate, each run is an independent event and the data that is generated cannot be directly compared even with identical simulation times. This lack of robustness will preclude a biologically meaningful result for the metabolite(s) of concern and is a significant limitation of this approach. "TemporalGSSA" or temporal Gillespie Stochastic Simulation Algorithm is an R-wrapper which will collate and partition SSA-generated datasets with identical simulation times (trials) into finite sets of linear models (technical replicates). Each such model (time step of a single run, absolute number of molecules for a metabolite) computes several coefficients (slope, intercept, etc.). These coefficients are averaged (mean slope, mean intercept) across all trials of a technical replicate and along with an imputed time step (mean, median, random) is incorporated into a linear regression equation. The solution to this equation is the number of molecules of a metabolite which is used to compute the molar concentration of the metabolite per technical replicate. The summarized (mean, standard deviation) data of this vector of technical replicates is the outcome or numerical estimate of the molar concentration of a metabolite and is dependent on the duration of the simulation. If the SSA-generated dataset comprises runs with differing simulation times, "TemporalGSSA" can compute the time-dependent trajectory of a metabolite provided the trials-per technical replicate constraint is complied with. The algorithms deployed by "TemporalGSSA" are rigorous, have a sound theoretical basis and have contributed meaningfully to our comprehension of the mechanism(s) that drive complex biochemical systems. "TemporalGSSA", is robust, freely accessible and easy to use with several readily testable examples.

Keywords: Biochemical network; linear model and linear regression equations; metabolite-specific and simulation time-dependent trajectory; R-wrapper; stochastic simulation algorithm (SSA)-generated dataset.

1. Introduction

Simulation studies of metabolites or distinct molecular species whence part of biochemical networks, can provide insights into the biochemical processes that they partake in. These may be coarse grained or atomistic and are inferential. Molecular trajectories that are deterministic are computed as solutions of first-order ordinary differential equations (ODEs),

$$y = f(t), \tag{1}$$

$$\dot{y} = \frac{df}{dt}. \tag{2}$$

These equations are formulated as the time differential of various reactants at steady state,

$$\dot{y} = \frac{df}{dt} = 0, \tag{3}$$

$$y = \int dt = \Delta t + C. \tag{4}$$

However, most biological systems comprise of several interacting metabolites and are therefore better modeled as systems of higher-order coupled partial differential equations (PDEs),

$$y = f(\phi_1, \phi_2, \ldots, \phi_x), \tag{5}$$

$$\partial^n y = a_0 \cdot \frac{\partial^n f}{\partial \phi_1^n} + a_1 \cdot \frac{\partial^n f}{\partial \phi_1^{n-1} \cdot \cdot \partial \phi_x} \cdots + a_{n-1} \cdot \frac{\partial^n f}{\partial \phi_x^n}. \tag{6}$$

It is clear that even when numerically feasible, ODE- and PDE-models of a biochemical process are dependent on available empirical data. Additionally, and even when not limiting, the data generated are from *in vitro* studies and is therefore, not physiological. The rigor and robustness of the deterministically computed trajectories notwithstanding, the aforementioned lacunae ensures that the data are of limited biological relevance. This limitation is further compounded by the lack of a suitable model to account for emergent phenomena when using ODEs and PDEs. A modification, is the stochastic differential equation (SDE), which draws a pseudorandom number from a normal distribution and appends this to the SDE. Stochastic simulations, on the other hand, are accurate, avoid bias and more likely to mimic the intracellular environment.[1,2] Emergent behavior, is, also, by definition encouraged and may provide insights into unforeseen and therefore, potentially impactful scenarios. However, computing and then comparing time-dependent trajectories for every metabolite of a biochemical network is not trivial and mandates extensive pre- and *post-hoc* processing.

The chemical master equation (CME), is an analytically intractable numerical expression of a well-stirred mixture of metabolites/reactants/molecular species ($j = 1, 2, \ldots, M$) that undergoes explicit interactions or reactions between distinct combinations of metabolites ($i = 1, 2, \ldots, R$). The CME can be modeled as a homogenous Poisson's process where the change of state of a metabolite is dependent on an average number of times a reaction fires over a length of time. This is used to compute the reaction-specific propensity from a Poisson's distribution $\mathcal{P}(\cdot)$,

$$m_j(t+\delta t) = m_j(t) + \sum_{i=1}^{i=R} \mathcal{P}(\phi_i)\cdot s_{ij} \,|\, s_{ij} \in (s_{ij}) = \boldsymbol{S} \subset \mathbb{Z}^{M\times R}, \tag{7}$$

$$\phi_i = r_i \cdot [m_{ji}]. \tag{7.1}$$

Here,

$$\begin{aligned} \phi &:= \text{Rate of an arbitrary reaction} \\ r &:= \text{Rate constant for an arbitrary reaction} \\ [m_j] &:= \text{Molar concentration of an arbitrary metabolite } (\mathrm{mol\,L^{-1}}). \end{aligned}$$

The change in the number of molecules can be used to compute the probable states of the network. The stochastic simulation algorithms (SSAs) utilize two pseudorandom numbers drawn from a uniform distribution to compute the length of the time interval between two reaction-simulating events the choice for a specific reaction.[1,2] A major limitation of the SSAs is the increased computational complexity for a large biochemical network. Here, the generation of a statistically relevant data will necessitate an increase in the duration of simulation times. An alternate strategy to numerically approximate the CME is to reformulate the change in the number of metabolites as a continuous variable,

$$\mathcal{P}(\phi) \approx \mathcal{N}(\phi) = \mathcal{N}(\mu(\phi), \sigma(\phi)\cdot\gamma) \,|\, \gamma \in \mathbb{R} \cap \mathcal{N}(0,1). \tag{8}$$

The chemical Langevin equation (CLE) for an arbitrary metabolite (m_j) is as follows:

$$m_j(t+\delta t) = m_j(t) + \sum_{i=1}^{i=R} \mathcal{N}(\phi_i)\cdot s_{ij} \,|\, s_{ij} \in (s_{ij}) = \boldsymbol{S} \subset \mathbb{Z}^{M\times R}, \tag{9}$$

$$= m_j(t) + \mu(\phi_i)\cdot s_{ij} + \sigma(\phi_i)\cdot s_{ij}\gamma_i, \tag{9.1}$$

$$= m_j(t) + \frac{1}{R}\cdot \sum_{i=1}^{i=R}(\phi_i)\cdot s_{ij} + \sum_{i=1}^{i=R}\sqrt{Q}, \tag{9.2}$$

$$= m_j(t) + \frac{1}{R}\cdot\sum_{i=1}^{i=R}(\phi_i)\cdot s_{ij} + \sqrt{Q}. \tag{9.3}$$

Here,

$$Q = \frac{\sum_{i=1}^{i=R}(\phi_i - \mu(\phi_i))^2\cdot s_{ij}\cdot\gamma_i}{R}, \tag{9.4}$$

$$\gamma_i \in \mathbb{R}\cap\mathcal{N}(0,1) = \frac{e^{\frac{-i^2}{2}}}{\sqrt{2\pi}}\,|\, i = 1,2,\ldots,R;\;\, i^2 \neq -1; i^2 \in \mathbb{N}. \tag{9.5}$$

The major limitation of this form of the CLE is that if the numbers of molecules becomes negative, computing the square root of the stochastic term becomes impossible.[3] The related chemical Fokker–Planck equation (CFPE), i.e. the partial derivative form of the probability function after truncating higher-order terms ($n > 2$) in the Taylor's series expansion of the CME, is[3,4]

$$\frac{\partial}{\partial t}P(m_j,t) = \left(\sum_{i=1}^{i=R}\frac{\partial}{\partial t}(m_{ji})\cdot\Delta m_{ji} + \sum_{i=1}^{i=R}\frac{1}{2}\cdot\frac{\partial^2}{\partial t^2}\cdot(m_{ji})\cdot{\Delta m_{ji}}^2\right)\cdot P(m_j,t). \tag{10}$$

For every metabolite of a generic biochemical network ($\mathbf{m}$) the CFPE is as follows:

$$\forall\, m,\quad \frac{\partial}{\partial t}P(\mathbf{m},t) = \left(\sum_{i=1}^{i=R}\frac{\partial}{\partial t}(\mathbf{m})\cdot\tilde{\mathbf{m}} + \sum_{i=1}^{i=R}\frac{1}{2}\cdot\frac{\partial^2}{\partial t^2}\cdot(\mathbf{m})\cdot\tilde{\mathbf{m}}^2\right)\cdot P(\mathbf{m},t), \tag{11}$$

where, $\quad \mathbf{m} = [m_{1i}m_{2i}\ldots m_{j=M,i}]^{\boldsymbol{T}} \in \mathbb{N}^M; \quad \tilde{\mathbf{m}} = [\Delta m_{1i}\Delta m_{2i}\ldots\Delta m_{j=M,i}]^{\boldsymbol{T}} \in \mathbb{R}_+^M$

$$\frac{\partial}{\partial t}P(\mathbf{m},t) = \left(\sum_{i=1}^{i=R}\sum_{j=1}^{j=M}\frac{\partial}{\partial t}(m_{ji})\cdot\Delta m_{ji} + \sum_{i=1}^{i=R}\sum_{j=1}^{j=M}\frac{1}{2}\cdot\frac{\partial^2}{\partial t^2}\cdot(m_{ji})\cdot{\Delta m_{ji}}^2\right)\cdot P(\mathbf{m},t). \tag{11.1}$$

A spatial form of the CLE (SCLE) has been investigated as a trade-off between the computationally intensive and accurate SSAs *vis-a-vis* the processivity of the CLE or CFPE.[3,4]

The SSAs (exact, τ-leap, Poisson, continuous), for well-stirred mixtures, are widely implemented computationally at both the package level and as part of multi-scale software environments. These software include those that implement newer algorithm variants of the SSAs (BioNetGen, PySB, Stochkit2, GillespieSSA) and/or novel solvers (COPASI, SUNDIALS).[5–12] Alternatively, multi-scale integrated and cloud-based computing environments for 2D (V-cell)-, 2D3D (StochSS, E-cell)- and 3D (Smoldyn, PyURDME, Readdy, STEPS, mesoRD) allow users to mix models, determine parameters and perform microscale simulations.[13–19] Despite these developments, the data generated by stochastic runs of a biochemical network are specific to that run even with identical simulation times. This means that unlike a deterministic model, the numeric values of the metabolites cannot be mapped to a specific simulation time and are therefore incomparable. This greatly limits the utility of stochastic simulations in making inferences into the genesis and mechanism(s) of physiological and pathological processes.

"TemporalGSSA", is an R-wrapper which will collate and partition SSA-generated datasets with identical simulation times into well-defined bins (trials, technical replicates, outcomes). Additional algorithms will utilize a combination of inferential (linear models, regression) and descriptive-statistical measures to transform this pre-processed raw data into a numerical estimate of the molar concentration of a metabolite which is dependent on the duration of the simulation. This will allow metabolites to be compared directly despite the stochastic nature of the raw data (Fig. 1). Case studies will be discussed in detail to highlight the importance of deriving comparable time-dependent trajectories from independent stochastic runs for non-enzyme- and enzyme-based biochemical networks.

This paper introduces generic definitions utilized to formulate the CME of a biochemical network. These are needed to fully comprehend, utilize, develop, and implement "TemporalGSSA". This is then followed by a detailed description of the functions and usage of "TemporalGSSA" along with the necessary mathematical formalism. Case studies (published models, feasibility study) that highlight the utility of the algorithms and the biological relevance of "TemporalGSSA" are discussed in context of complex biochemical networks (non-enzyme, enzyme). This paper concludes with a summary of the salient features, limitations and possible future studies that will utilize "TemporalGSSA".

2. Definitions and General Concepts

A biochemical network comprises of multiple interactions (non-enzyme) or reactions-(enzyme) of reactants. A generic enzyme-catalyzed reaction under the assumption of Michaelis–Menten (MM)-kinetics is as follows:

$$[E] + [Su] \underset{r_{-1}}{\overset{r_1}{\leftrightarrow}} [ESu] \overset{r_2}{\rightarrow} [E] + [Pr] \quad \text{Reaction, (1)}$$

$$\mathrm{Km} = \frac{r_1 + r_2}{r_{-1}}. \tag{12}$$

Here,

$$\begin{aligned}
E &:= \text{Generic enzyme,}\\
Su &:= \text{Substrate,}\\
ESu &:= \text{Enzyme substrate complex,}\\
Pr &:= \text{Product,}\\
r_x &:= \text{Rate constant of a reaction,}\\
Km &:= \text{Michaelis} - \text{Menten constant.}
\end{aligned}$$

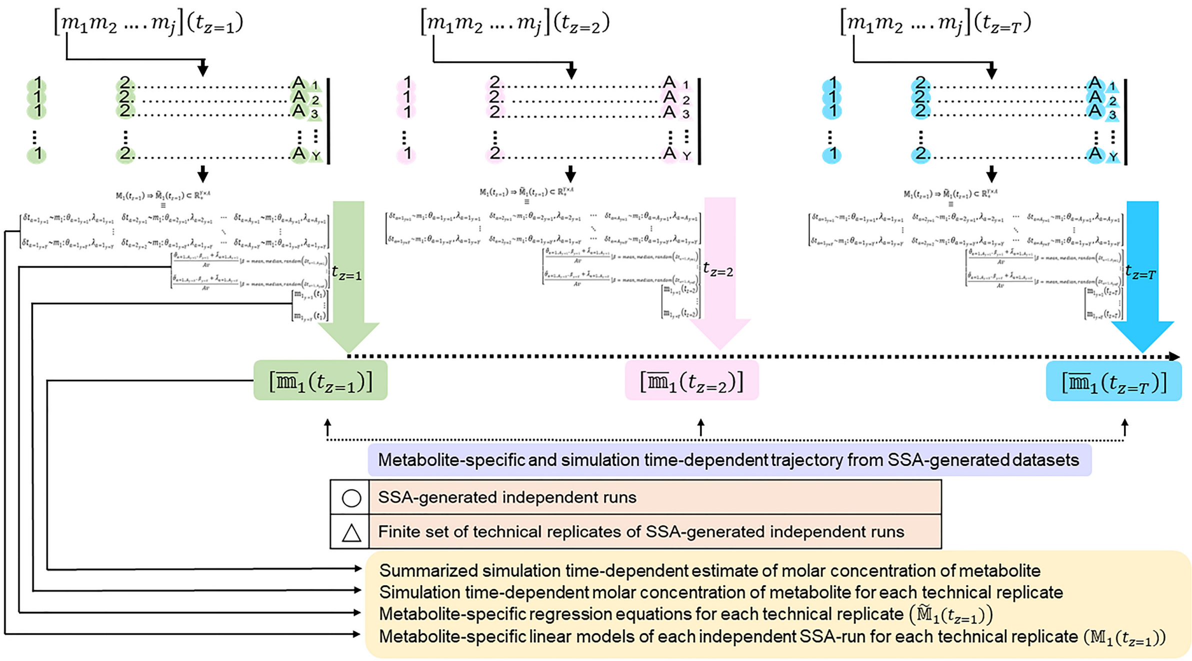
$[m_1 m_2 \dots m_j](t_{z=1})$
$[m_1 m_2 \dots m_j](t_{z=2})$
$[m_1 m_2 \dots m_j](t_{z=T})$
1
2
A
Y
$t_{z=1}$
$t_{z=2}$
$t_{z=T}$
$[\overline{\mathbb{m}}_1(t_{z=1})]$
$[\overline{\mathbb{m}}_1(t_{z=2})]$
$[\overline{\mathbb{m}}_1(t_{z=T})]$
Metabolite-specific and simulation time-dependent trajectory from SSA-generated datasets
○ SSA-generated independent runs
△ Finite set of technical replicates of SSA-generated independent runs
Summarized simulation time-dependent estimate of molar concentration of metabolite
Simulation time-dependent molar concentration of metabolite for each technical replicate
Metabolite-specific regression equations for each technical replicate $(\widetilde{\mathbb{M}}_1(t_{z=1}))$
Metabolite-specific linear models of each independent SSA-run for each technical replicate $(\mathbb{M}_1(t_{z=1}))$

Fig. 1. Schematic representation of the computational pipeline and algorithms deployed by "TemporalGSSA". "TemporalGSSA", is an R-wrapper which collates and partitions SSA-generated datasets with identical simulation times into well -defined bins (trials, technical replicates, outcomes). A linear model of each trial relates the time step of a simulation run with changes in the numbers of molecules for every metabolite that partakes in the biochemical network. The averaged coefficients from these linear models (mean slope, mean intercept) across all trials that comprise a technical replicate along with the imputed time step are incorporated into a regression equation. The solution to this equation is the number of molecules of a metabolite and is used to compute the molar concentration of the metabolite per technical replicate. The metabolite-specific vector of all the technical replicates is then summarized (arithmetic mean, variance) and is an estimate of the molar concentration of a metabolite which is dependent on the duration of simulation. If the raw dataset was generated with varying simulation times, "TemporalGSSA", can compute the time-dependent trajectory for every metabolite provided the constraints are complied with. **Abbreviations**: **A**, Number of trials in each technical replicate and corresponds to the number of independent runs from an SSA-generated dataset; *Av*, Avogadro's number of molecules per gram; **Y**, Number of technical replicates used to obtain a statistically sound estimate of the molar concentration of a metabolite which is dependent on the duration of simulation; **LM**, linear model, and **SSA**, stochastic simulation algorithm.

Partitioning this into partial reactions,

$$[E] + [Su] \underset{r_{-1}}{\overset{r_1}{\leftrightarrow}} [ESu], \quad \text{Reaction, } (2) \text{ and } (3)$$

$$[ESu] \overset{r_2}{\rightarrow} [E] + [Pr]. \quad \text{Reaction, } (4)$$

Clearly, modeling an enzyme-based biochemical network in not trivial and must take into account the number of partial reactions and feedback- and feedforward-mechanisms. In contrast, a non-enzyme- based reaction (interaction) is one that follows simple association/disassociation kinetics, is easily represented and can be derived as follows:

$$\lim_{\text{Rate of reaction } (4) \to 0} r_2 \to 0, \tag{13}$$

$$Km = \frac{r_1}{r_{-1}} = Kd \in \mathbb{R}_+. \tag{14}$$

Rewriting these as simpler forward and reverse interactions,

$$a \cdot [R] + b \cdot [S] \underset{r_{-1}}{\overset{r_1}{\leftrightarrow}} [R]^a [S]^b. \quad \text{Reaction, } (5)$$

$$Kd = \frac{r_1}{r_{-1}}.$$

Here,

$$\begin{aligned} R, S &:= \text{Generic reactants,} \\ a, b &:= \text{Stochiometry of reactants,} \\ RS &:= \text{Enzyme substrate complex,} \\ r_x &:= \text{Rate constant of the interaction,} \\ Kd &:= \text{Disassociation constant.} \end{aligned}$$

This interaction-specific disassociation constant ($Kd \in \mathbb{R}_+$) can be used to infer the dominant direction of every such interaction of a biochemical network.

$$\text{Forward } Kd > 1.0, \tag{15}$$

$$\text{Reverse } Kd \in (0, 1.0), \tag{15.1}$$

$$\text{Equivalent } Kd \simeq 1.0. \tag{15.2}$$

2.1. *The CME and its representation in terms of the modeled metabolites*

For an indexed set ($\boldsymbol{R}$) of R-interactions, $i = 1, 2, \ldots, R$, and M-metabolites, $j = 1, 2, \ldots, M$, then the following definitions hold for the collection of well-stirred molecules at a particular state ($\mathcal{K}$) which is as follows:

$$\{\mathbf{m}(\mathcal{K}(z)) = \mathbf{m}_{\mathcal{K}_z} \mid z = 1, 2, \ldots, T\}, \quad \text{Def. } (1)$$

where,

$$\mathbf{m} = (m_1 m_2, \ldots, m_{j=M})^{\boldsymbol{T}} \in \mathbb{N}^M.$$

The change in molecular numbers and thence the change of state is governed by the stoichiometry of the reacting molecules (L_j),

Let,

$$\mathbf{m}_{K_z} \xrightarrow{\phi_{\mathcal{K}_z\mathcal{K}_k} \equiv \phi_i} \mathbf{m}_{\mathcal{K}_k} \,|\, k = z \pm z'; \quad z' \in \mathbb{N} \quad \text{Def. (2)}$$

$$\mathbf{m}_{\mathcal{K}_k} \xrightarrow{\phi_{\mathcal{K}_k\mathcal{K}_z} \equiv \phi_k} \mathbf{m}_{\mathcal{K}_z} \,|\, k = z \pm z'; \quad z' \in \mathbb{N} \quad \text{Def. (3)}$$

Here,

$$m_j(t + \delta t) = m_j(t) + [\phi_i \cdot \boldsymbol{s}_{ij} + \phi_k \cdot \boldsymbol{s}_{kj}] \,|\, \boldsymbol{s}_{ij}, \boldsymbol{s}_{kj} \in \boldsymbol{\mathcal{S}} \subset \mathbb{Z}^{R\times M}, \tag{16}$$

where,

$$\phi_i = \frac{\text{Rate of the } i^{\text{th}} - \text{interaction}(s^{-1})}{\text{Molar concentration of reactants}([m_j]^{L_j \geq 0} \ldots)}, \tag{16.1}$$

$$\phi_k = \frac{\text{Rate of the } k^{\text{th}} - \text{interaction}(s^{-1})}{\text{Molar concentration of reactants}([m_j]^{L_j \geq 0} \ldots)}. \tag{16.2}$$

Here,

$$\omega = \sum_{j=1}^{j=M} L_j \,|\, L \in Z_+ \cap 0 \tag{16.3}$$

$$\begin{aligned}
&1M(\cdot) \equiv [\cdot] = 6.023 \times 10^{23} \text{molecules of } 1g(\cdot)\\
&\qquad \phi_i \gg 1\\
&\phi_i := \text{Interaction} - \text{specific rate constant},\\
&\omega := \text{Order of a generic interaction},\\
&L := \text{Stoichiometry of a reactant or metabolite}.
\end{aligned}$$

Rewriting this for the defined molecular entity (m_j),

$$\phi_i = \frac{1}{\frac{[m_j]^{L_j \geq 0}}{\text{Rate of the } i^{\text{th}} - \text{interaction}}}, \tag{17}$$

$$\forall\, j, \frac{1}{\frac{\bar{\mathbf{m}}}{\text{Rate of the } i^{\text{th}} - \text{interaction}}} = \phi_i(m_{\mathcal{K}_z \to \mathcal{K}_k}) = \tilde{\phi}_i, \tag{18}$$

where,

$$\bar{\mathbf{m}} = ([m_1]^{L_1 \geq 0}[m_2]^{L_2 \geq 0} \ldots [m_{j=M}]^{L_{j=M} \geq 0}). \tag{18.1}$$

The conditional probabilities for each change of state can be written as follows:

$$P(\mathbf{m}_{\mathcal{K}_z} \,|\, \mathbf{m}_{\mathcal{K}_k}) = P(\mathbf{m}_{\mathcal{K}_k}(t + \delta t) - \mathbf{m}_{\mathcal{K}_z}(t)) = \boldsymbol{\Delta}\mathbf{m}_{kz} \cdot t^{-1} \cdot dt = \phi_k \cdot dt, \tag{19}$$

$$P(\mathbf{m}_{\mathcal{K}_k} \,|\, \mathbf{m}_{\mathcal{K}_z}) = P(\mathbf{m}_{\mathcal{K}_z}(t + \delta t) - \mathbf{m}_{\mathcal{K}_k}(t)) = \boldsymbol{\Delta}\mathbf{m}_{zk} \cdot t^{-1} \cdot dt = \phi_i \cdot dt. \tag{20}$$

The conditional probability of the set of metabolites for two distinct states $(\mathbf{m}_{\mathcal{K}_z}, \mathbf{m}_{\mathcal{K}_k})$ is as follows:

$$P(\mathbf{m}_{\mathcal{K}_z} \,|\, \mathbf{m}_{\mathcal{K}_k}) = \frac{P(\mathbf{m}_{\mathcal{K}_k} \,|\, \mathbf{m}_{\mathcal{K}_z}) \cdot P(\mathbf{m}_{\mathcal{K}_z})}{P(\mathbf{m}_{\mathcal{K}_k})}, \tag{21}$$

$$P(\mathbf{m}_{\mathcal{K}_z} \,|\, \mathbf{m}_{\mathcal{K}_k}) \cdot P(\mathbf{m}_{\mathcal{K}_k}) = P(\mathbf{m}_{\mathcal{K}_k} \,|\, \mathbf{m}_{\mathcal{K}_z}) \cdot P(\mathbf{m}_{\mathcal{K}_z}), \tag{21.1}$$

$$(1 - \phi_k \cdot dt) \cdot P(\mathbf{m}_{\mathcal{K}_k}) = (1 - \phi_i \cdot dt) \cdot P(\mathbf{m}_{\mathcal{K}_z}), \tag{21.2}$$

$$P(\mathbf{m}_{\mathcal{K}_k}) - P(\mathbf{m}_{\mathcal{K}_k}) \cdot (\phi_k \cdot dt) = P(\mathbf{m}_{\mathcal{K}_z}) - P(\mathbf{m}_{\mathcal{K}_z}) \cdot (\phi_i \cdot dt). \tag{21.3}$$

Rearranging,

$$\begin{aligned} P(\mathbf{m}_{\mathcal{K}_k}) - P(\mathbf{m}_{\mathcal{K}_z}) &= -[P(\mathbf{m}_{\mathcal{K}_z}) \cdot (\phi_i \cdot dt)] + P(\mathbf{m}_{\mathcal{K}_k}) \cdot (\phi_k \cdot dt), \\ \frac{d}{dt}(P(\mathbf{m}_{\mathcal{K}_z}, t)) &= -[P(\mathbf{m}_{\mathcal{K}_z}) \cdot (\phi_i \cdot dt)] + P(\mathbf{m}_{\mathcal{K}_k}) \cdot (\phi_k \cdot dt), \qquad (21.4) \\ &= P(\mathbf{m}_{\mathcal{K}_k}) \cdot (\phi_k \cdot dt) - P(\mathbf{m}_{\mathcal{K}_z}) \cdot (\phi_i \cdot dt), \end{aligned}$$

$$\frac{d}{dt} P(\mathbf{m}_{\mathcal{K}_z}, t) = [P(\mathbf{m}_{\mathcal{K}_k}) \cdot \phi_k \cdot dt - P(\mathbf{m}_{\mathcal{K}_z}) \cdot \phi_i \cdot dt], \quad k = z \pm z^{'}; z^{'} \in \mathbb{N}. \tag{21.5}$$

For an arbitrary number of states this is as follows:

$$\frac{\partial}{\partial t} P(\mathbf{m}_{\mathcal{K}_z}, t) = \sum_{i=1}^{i=R} [P_{\mathcal{K}_k}(\mathbf{m} - \mathbf{s}_i, t) \cdot c_i(\mathbf{m} - \mathbf{s}_i) - P_{\mathcal{K}_z}(\mathbf{m}, t) \cdot c_i(\mathbf{m})], \tag{22}$$

where,

$$\mathbf{s} \in \mathbb{Z}^M, \tag{22.1}$$

$$c_i = \tilde{\phi}_i \cdot h(\mathbf{m} - \mathbf{s}_i), \tag{22.2}$$

$$h = \sum_{j=1}^{j=M} \binom{\mathbf{m} - \ \mathbf{s}_i}{m_j + s_{ij}}. \tag{22.3}$$

$$\begin{aligned} c_i &:= \text{Interaction} - \text{specific propensity factor}, \\ \phi_i &:= \text{Interaction} - \text{specific rate constant}, \\ \boldsymbol{S}_i &:= \text{Interaction} - \text{specific stoichiometry vector for all reactants}, \\ h &:= \text{set of unique combinations of reactants}, \end{aligned}$$

2.2. *Collating, formatting, annotating SSA-generated datasets*

The CME can be numerically approximated by the SSAs.[1,2] "TemporalGSSA", is an R-wrapper and can be used with any SSA-generated dataset. However, this needs to be pre-processed in accordance with the guidelines of "TemporalGSSA". The parameter of choice is the time step and is characteristic of each run. SSA-runs with identical simulation wall times are collated into trials $(a = 1, 2, \ldots, A)$ for a finite set of technical replicates $(y = 1, 2, \ldots, Y)$. In vector notation,

$$\{\boldsymbol{s}_y(a) = \boldsymbol{s}_{y_a} = [s_{y_1} s_{y_2} \dots s_{y_{a=A}}]|\}, \tag{23}$$

where,

$$t_z(s_{y_1}) = t_z(s_{y_2}) \dots = t_z(s_{y_{a=A}}); \quad t_z \in \mathbb{N} \,|\, z = 1, 2, \dots, T. \tag{23.1}$$

The SSA-generated data for a specific simulation time (t_z) are then the rectangular matrices of data points (time steps, numbers of molecules of all metabolites),

$$\boldsymbol{S}_y = (s_{y_{ab}}) \in \mathbb{R}_+^{A \times b}. \tag{24}$$

Here,

$$b = M + 1, \tag{24.1}$$

$$\boldsymbol{s}_{y_{b=1}} = [\alpha_1 \alpha_2 \dots \alpha_{a=A}]^{\boldsymbol{T}} \in \mathbb{R}_+^A \cap (0, 1), \tag{24.2}$$

$$\boldsymbol{s}_{y_{b>1}} = [m_{j_1} m_{j_2} \dots m_{j_{a=A}}]^{\boldsymbol{T}} \in \mathbb{N}^A. \tag{24.3}$$

Here, the first column represents the time steps for all trials. Each subsequent column corresponds to a specific metabolite and represents their altered and absolute numbers across all trials. This is in accordance with the rate constant of the interaction(s) that they partake in and stoichiometry. Expanding this for an arbitrary technical replicate $(\mathcal{S}_y \,|\, y \in Y)$,

$$\left[\begin{array}{lcccc} & \text{Col 1 (Time step)} & \text{Col 2 (metabolite)} & \cdots & \text{Col } b \text{ (metabolite)} \\ \text{Run or trial}(s_{y_{a=1b}}) & ts1 & m_{1_1} & & m_{j_a} \\ & \vdots & & \ddots & \vdots \\ \text{Run or trial}(s_{y_{a=Ab}}) & tsA & m_{1_A} & \cdots & m_{j_A} \end{array}\right]$$

The finite set of all technical replicates $(y = 1, 2, \dots, Y)$ is as follows:

$$\mathfrak{S} = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \dots, \boldsymbol{S}_{y=Y}\} \,|\, \boldsymbol{S}_y = (s_{y_{ab}}) \in \mathbb{R}_+^{a \times b}. \tag{25}$$

Since these are generated with identical simulation times $(t_z \in \mathbb{N} \,|\, z = 1, 2, \dots, T)$ we can rewrite this as follows:

$$\mathfrak{S}(t_z) = \mathcal{S}_1(t_z)\mathcal{S}_2(t_z) \dots \mathcal{S}_{y=Y}(t_z). \tag{26}$$

These data can be used to establish a statistical profile (mean, standard deviation, variance) for an arbitrary metabolite at a given simulation time.

2.3. *Availability and implementation of* "*TemporalGSSA*"

"TemporalGSSA" is available from CRAN (https://cran.r-project.org/package = TemporalGSSA). The package can be automatically updated and installed directly using the available functions update.packages('TemporalGSSA') and install-packages('TemporalGSSA'). "TemporalGSSA", is built in RStudio (1.4.1717) and tested in R-4.1.x. "TemporalGSSA" comprises of two functions (*calculate_TemporalGSSA*, *check_TemporalGSSA*). The package depends on the functions imported rom the R-package ("stats"). The data that will be used as input for "TemporalGSSA", will be collated from SSA-generated datasets, formatted and binned. The package includes the necessary documentation on all the functions, examples and tests of functionality.

3. Results and Discussion

The algorithm(s) used to construct, implement, and utilize "TemporalGSSA" are outlined in detail and are accompanied with the necessary mathematical formalism (Fig. 1).

3.1. *Checking the user-defined data that "TemporalGSSA" will process*

The *check_TemporalGSSA()*function of "TemporalGSSA" will check the data that are input by the user before processing the same. The maximum value, i.e. Y-number of y-indexed technical replicates is decided by the user and is optional. However, since the purpose of "TemporalGSSA" is to provide a numerical estimate of the molar concentration of a metabolite which is dependent on the duration of the simulation, the following constraint is recommended:

$$Y \geq 3. \tag{27}$$

There are, however, two mandatory constraints without which "TemporalGSSA" will return an error. This includes the minimum number of trials for each technical replicate and the choice of statistic to compute the representative time step (Fig. 1). The constraint of a lower bound to the number of trials is imposed such that the data computed by the regression models are statistically significant,

$$A \geq 30. \tag{28}$$

The trials-per-technical replicate is an important test that combines these constraints which the SSA-generated datasets must comply with get a statistically significant estimate of the number of molecules of each metabolite.

$$\xi = \frac{A}{Y} \geq 30. \tag{29}$$

The metric of choice for computing the imputed time step ($\mathcal{C}$) from the set of trials and for a technical replicate is as follows:

$$\text{Choice}(\mathcal{C}) = 1 \vee 2 \vee 3 \vee 4, \tag{30}$$

where,

$$\begin{gathered}\text{Mean} \equiv \mathcal{C} = 1, \quad \text{Def. (4)}\\ \text{Median} \equiv \mathcal{C} = 2, \quad \text{Def. (5)}\\ \text{Random selection} \equiv \mathcal{C} = 3, \quad \text{Def. (6)}\\ \text{All} \equiv \mathcal{C} = 4. \quad \text{Def. (7)}\end{gathered}$$

A numeric flag which indicates suitability of data ("0") *vis-à-vis* with one that mandates revisions ("1") is returned after this step. The data generated by the linear models in subsequent steps are not trivial. A logical argument, which is optional, will be used to indicate whether the console output for the models with coefficients for the intercept, slope, miscellaneous statistical descriptors, etc., will be curtailed (*TRUE*; default) or verbose (*FALSE*). The console data for the latter can be redirected to a local text file.

3.2. *Computing the representative time step of a technical replicate*

The major task of "TemporalGSSA" is to facilitate the computation of the simulation time-dependent trajectory for every participating metabolite of a biochemical network. This is accomplished with the *calculate_TemporalGSSA()* function. Each run of an SSA-generated dataset is independent and is characterized by a time step which is chosen randomly from a uniform distribution. This implies that a comparative assessment of the number of molecules of a metabolite is possible by imputing a single time step across all trials of a technical replicate (β_y) and inferring the numbers of molecules of each metabolite at that time step. A number of choices are available to the user. These include the mean, median, arbitrary or a combination of all these.

$$\beta_y = \begin{cases} \frac{1}{A} \cdot \|\boldsymbol{s}_{a=1..Ab=1}\|_1, \\ \text{median}(\boldsymbol{s}_{a=1..Ab=1}), \\ \text{rand}(\boldsymbol{s}_{a=1..Ab=1}). \end{cases} \tag{31}$$

3.3. *Formulating the regression equation of a metabolite for each technical replicate*

The raw data matrices ($\mathfrak{S}$) of the SSA-generated dataset(s) will be converted into metabolite-specific matrices ($\mathbb{M}_j$). These will be processed and be used to compute an estimate of the molar concentration of a metabolite which is dependent on the duration of the simulation. The first step is accomplished by creating a set of linear models between the time steps of each trial, and for each metabolite,

$$s_{y_{a=1..Ab=1}}(t_z) \sim s_{y_{a=1...Ab=2...M}}(t_z). \tag{32}$$

Rewriting this data for an arbitrary metabolite (m_j) across all technical replicates and will give the following matrix (Fig. 1):

$$\mathbb{M}_j(t_z) = (\mathcal{L}m_{j_{ya}}(t_z)) \in \mathbb{R}_+^{Y \times A} \,|\, \mathcal{L}m_{j_{ya}}(t_z) = \delta t_a(t_z) \sim m_{j_{ya}}(t_z). \tag{33}$$

Here,

$$\delta t_a(t_z) = s_{y_{ab=1}}(t_z), \tag{33.1}$$

$$m_{j_{ya}}(t_z) = s_{y_{ab>1}}(t_z), \tag{33.2}$$

$$a = 1, 2, \ldots, A; \quad y = 1, 2, \ldots, Y; \quad z = 1, 2, \ldots, T.$$

"TemporalGSSA" iteratively invokes the R-package "stats" to make these calculations. The detailed list of coefficients such as the mean ($\theta_{m_{j_{ya}}}$) and intercept

($\lambda_{m_{j_{ya}}}$) along with several others is sent to a data file and is used to populate each entry of the matrix (Fig. 1),

$$\tilde{\mathbb{M}}_j(t_z) = ([\theta_{m_{j_{ya}}}(t_z), \lambda_{m_{j_{ya}}}(t_z)])_{ya} \in \mathbb{R}_+^{Y\times A}, \tag{34}$$

where,

$$\theta_{m_{j_{ya}}}(t_z) = \theta_{\mathcal{L}m_{j_{ya}}(t_z)}, \tag{34.1}$$

$$\lambda_{m_{j_{ya}}}(t_z) = \lambda_{\mathcal{L}m_{j_{ya}}(t_z)}. \tag{34.2}$$

Since each row corresponds to a technical replicate with A-trials, the mean-slope ($\bar{\theta}_{m_{j_y}}$) and mean-intercept ($\bar{\lambda}_{m_{j_y}}$) can be calculated for a technical replicate:

$$\bar{\theta}_{m_{j_y}} = \bar{\theta}_{m_j(y)} = \frac{1}{A}\cdot \|\boldsymbol{\theta}_{m_{j_{ya=1\ldots A}}} = \boldsymbol{\theta}^A_{m_{j_y}}\|_1, \tag{35}$$

where,

$$\boldsymbol{\theta}^A_{m_{j_y}} = [\theta_{m_{j_{ya=1}}}, \theta_{m_{j_{ya=2}}}, \ldots, \theta_{m_{j_{ya=A}}}]. \tag{35.1}$$

$$\bar{\lambda}_{m_{j_y}} = \bar{\lambda}_{m_j(y)} = \frac{1}{A}\cdot \|\boldsymbol{\lambda}_{m_{j_{ya=1\ldots A}}} = \boldsymbol{\lambda}^A_{m_{j_y}}\|_1, \tag{36}$$

where,

$$\boldsymbol{\lambda}^A_{m_{j_y}} = [\lambda_{m_{j_{ya=1}}}, \lambda_{m_{j_{ya=2}}}, \ldots, \lambda_{m_{j_{ya=A}}}]. \tag{36.1}$$

These data, along with imputed time step (β_y) are incorporated into a linear regression model and will be used to compute the molar concentration of a metabolite for a technical replicate

$$m_{j_y} = \frac{((\bar{\theta}_{m_{j_y}})\cdot(\beta_y)) + \bar{\lambda}_{m_{j_y}}}{Av}, \tag{37}$$

where

$$Av = 6.023 \times 10^{23}\,\text{molecules}\,g^{-1}. \tag{37.1}$$

3.4. *Computing the simulation time-dependent trajectory for a metabolite of a biochemical network*

For the finite set of replicates this is the vector (Fig. 1),

$$\mathbb{mm}_j^Y = [m_{j_1}, m_{j_2}, \ldots, m_{j_{y=Y}}]^T. \tag{38}$$

Since, this representation is for a specific simulation time, we can rewrite this as follows:

$$\mathbb{mm}_j^Y(t) = [m_{j_1}(t), m_{j_2}(t), \ldots, m_{j_{y=Y}}(t)]^T \in \mathcal{M} \subset \mathbb{R}_+^Y. \tag{39}$$

These data, for an arbitrary metabolite, can be summarized using the arithmetic mean (μ) and variance (σ^2) and is the outcome (numerical estimate of a molar concentration of a metabolite which is dependent on the duration of the simulation).

$$\mu(\mathbb{m}_j^Y(t)) = \frac{1}{Y} \cdot \|\mathbb{m}_j^Y(t)\|_1 = \frac{1}{Y} \cdot \sum_{y=1}^{y=Y} m_{j_y}(t), \tag{40}$$

$$\sigma^2(\mathbb{m}_j^Y(t)) = \sum_{y=1}^{y=Y} (m_{j_y}(t) - \mu(\mathbb{m}_j^Y(t)))^2. \tag{41}$$

These equations can be combined for every metabolite of a biochemical network.

$$\mu(\mathcal{M}) \pm \sigma^2(\mathcal{M}) = (\mu \pm \sigma^2)(\mathcal{M}), \tag{42}$$

$$= [[(\mu \pm \sigma^2)(\mathbb{m}_{\boldsymbol{j}}^{\boldsymbol{Y}}(t))]_{j=1} \cdots [(\mu \pm \sigma^2)(\mathbb{m}_{\boldsymbol{j}}^{\boldsymbol{Y}}(t))]_{j=M}]^T, \tag{42.1}$$

$$= \bar{\mathbb{m}}^M(t). \tag{42.2}$$

The time-dependent trajectories for M-metabolites of a biochemical network for varying simulation times is the sequence (Fig. 1),

$$\bar{\mathbb{m}}^M(t=1), \bar{\mathbb{m}}^M(t=2), \ldots, \bar{\mathbb{m}}^M(t=T) = \bar{\mathbb{m}}_1^M, \bar{\mathbb{m}}_2^M \ldots \bar{\mathbb{m}}_{t=T}^M = (\bar{\mathbb{m}}_t^M)_{t\in[1,T]}. \tag{43}$$

3.5. *Biological relevance of "TemporalGSSA"*

Biochemical networks comprise of non-enzyme mediated interactions (Kd, Ka), enzyme-based reactions (Km) or both. The algorithms deployed by "TemporalGSSA" have been utilized to glean valuable insights into biochemical systems that favor the association and disassociation of molecular complexes from simulation studies conducted previously.[20,21] These studies were able to offer mechanistic explanations into seemingly contradictory empirical data. Additionally, simulation data from a multi-compartment (intracellular, membrane, extracellular) mathematical model of inflammation suggested that the effects of external noxious stimuli could be integrated at the membrane of an advancing neutrophil and thence transduced into the cytosol and nucleus. In contrast, biochemical networks which are enzyme-based have to account for partial reactions, substrate and product interconversions, inhibitory kinetics, etc. Details of these are presented as case studies to highlight the biological relevance and feasibility of this approach to modelling complex biochemical systems.

Case 1. *Low-affinity peptide-driven (LAPD)-model of MHC1-mediated export of high-affinity peptides*

Low affinity peptides are continually processed by the assembly and disassembly of the peptide loading complex (PLC). The PLC is an Endoplasmic Reticulum (ER) membrane associated multiprotein complex (Tapasin, major histocompatibility complex (MHC)-I, Erp57) that functions to funnel cytosolic peptides from the ubiquitin proteasome into the lumen of the ER.[22–25] However, only a minimal fraction (≈ 0.00001%) of this pool, referred to as the high-affinity peptides, are exported to the plasma membrane and presented to cytotoxic T-lymphocytes. Despite the large volume of empirical data, the mechanism(s) by which this occurs inside the cell is not clear. Plausible explanations included the role of stoichiometry of the PLC and peptide editing in association with Tapasin.[26–28] However, the outcome of these models, i.e. export of high-affinity peptides to the plasma membrane, under the kinetics of PLC-assembly/disassembly and peptide editing had not been elaborated upon.

These problems were addressed by the LAPD-model, a mathematical representation of a simple biochemical network of the MHC1-mediated exported nigh-affinity peptides to the plasma membrane.[20] Here, the model (number of interactions = 30, number of reactants = 18) is well stirred.[20] Whilst, the first part of that study was carried out to parameterize the network in terms of the peptide-editing retrograde pathway, the second was to determine the time-dependent trajectory of the metabolites under varying simulation wall times (60–900 s).[20] The study, which used the algorithms deployed by "TemporalGSSA" was able to provide clear insights into the molecular biology of high-affinity peptide export to the plasma membrane of a nucleated cell.[20] In particular, the simulation data revealed that low-affinity peptides are not passive players and may function as key regulators, along with Tapasin of this process.[20,26,27] The analysis of the data also suggested that several-fold excess of low-affinity peptides could initiate and maintain the PLC in a primed state.[20] Thus, MHC1-mediated export and thence cytotoxic T-lymphocyte triggered damage can be efficient and rapid.[20,29,30] These observations have important consequences for the cell in the presence of an acute (viral infection) and/or chronic (dysplastic transformation) molecular insult to the nucleated cell.[29,30]

Case 2. *Insights into the mechanism(s) of ROS/RNS-mediated phagocyte chemotaxis* Inflammation is a physiological reaction to noxious stimuli-mediated cell injury. Here, circulating neutrophils exit the circulation, exhibit transcellular migration and are able to effectively neutralize several well characterized chemical mediators of inflammation.[21,31] However, the precise manner in which this occurs has eluded a comprehensive mechanistic explanation. More specifically, the manner in which transient and random membrane protrusions of a migrating neutrophil are converted into a single dominant lamellipodium was not clear.[32] Further confounding the available empirical data are the roles of the reactive oxygen species (ROS) and reactive nitrogen species (RNS).[33,34] A simple multiscale non-enzymatic model that included the extracellular matrix (ECM), lipid rafts and the cytoskeleton of a migrating neutrophil was presented and simulated.[21] Lipid rafts, are ordered microdomains (8–200 nm; saturated fatty acids, sphingomyelin, gangliosides, cholesterol) present in the plasma membrane.[35] These favor the formation of multiprotein complexes and serve as efficient platforms for signal transduction.[33–35] The molecular components of this model included a fully saturable isoform of an Superoxide Dismutase present in the ECM (EcSOD), superoxide anions and hydrogen peroxide along with the monomeric small GTPase (Rac, Rho, Rab)-proteins.[21,36,37] The cellular event incorporated into the model was the propensity of the lipid rafts to stochastically transit between monomeric and higher-order complexes (dimer, tetramer, octamer).[21]

The time-dependent trajectory of the lipid rafts of an advancing neutrophil as they assemble/disassemble suggested that they are able to integrate and thence transduce the effect of extracellular ROS/RNS into the actin cytoskeleton.[21,33,34] These complexes, whence formed, and after several futile cycles will convert the stochastic event of transient protrusion formation into a single dominant and stable lamellipodium.[21] In fact, calculations from the simulation data suggested a concentration in excess of 0.08 aM for the octamer form of lipid raft-multiprotein complex was critical for dominant and stable lamellipodium formation.[21] Interestingly, the data also attributed success of formation of a dominant lamellipodium of an advancing neutrophil to EcSOD being saturated.[21] Since, the half-life of EcSOD is approximately 85 h, a noxious stimulus of a significant degree will completely saturate EcSOD, whilst actively interfering with its interaction with the ECM.[21,38–40] Additionally, the study also suggested that the ratio of superoxide anions and hydrogen peroxide in the ECM was critical in bringing about these changes.[21]

Case 3. *Feasibility study of a non-haem iron(II)- and 2-oxoglutarate-dependent dioxygenase based biochemical network*

As discussed previously, enzyme-catalyzed reactions and the networks they partake in cannot be modeled as discrete states. However, these can be reduced to simpler interactions under perturbing conditions which can then be modeled and simulated by the SSAs. Consider the case of a biochemical network of non-haem iron(II)- and 2-oxoglutarate-dependent-dioxygenases (i2OGdd), a superfamily of enzymes that is well represented in all kingdoms of life. These require iron(II), 2-oxoglutarate as a co-substrate and will catalyze a generic hydroxylation reaction, although they can concomitantly catalyze one or more specialized reactions.[40–42] A generic i2OGdd-catalyzed reaction is as follows:

$$[i2OGdd] + [O_2] + [2OG] + [Su] \overset{r_1}{\underset{r_{-1}}{\leftrightarrow}} [i2OGdd - 2OG - Su], \quad \text{Reactions (1a), (2a)}$$

$$[i2OGdd - 2OG - Su] \xrightarrow{r_2} [i2OGdd] + [SA] + [CO_2] + Pr, \quad \text{Reactions (3a)}$$

$$Km(i2OGdd) = \frac{r_1 + r_2}{r_{-1}}. \tag{44}$$

Here,

$SA :=$ Succinic acid,

$CO_2 :=$ Carbon dioxide,

$Su :=$ Substrate,

$Pr :=$ Specialized product (hydroxylated, sulfated, etc.),

$Km :=$ Michaelis $-$ Menten constant.

The biochemical network of these enzymes exhibits multiple degrees of complexity. These can be in terms of the taxonomic distribution, protein structure (superfamily; protein sequence identity $\approx 1 - 15\%$), isoenzyme distribution and localization, co-factor and co-substrate requirement, and a flexible active site geometry which results in numerous substrates, reactions, and products.[40–42]

A network of i2OGdd-members acting in concert to execute a systems-level function (iron homeostasis) while concomitantly performing enzyme-specific catalysis has been described.[43,44] Since i2OGdd-members possess different kinetic parameters $(Km_{Fe^{2+}}Km_{2OG}Km_{O_2})$, an initial strategy is to choose a set of data points which is best represented and has maximal variance. On the basis of available kinetic data for ferrous iron $(Km_{Fe^{2+}})$, i2OGdd-members can be categorized as follows:[43]

$$\boldsymbol{H}_{\text{i2OGdd}} \subset \boldsymbol{i2OGdd} \,|\, Km_{Fe^{2+}} \leq 0.02\,\text{mM}, \tag{45}$$

$$\boldsymbol{L}_{\text{i2OGdd}} \subset \boldsymbol{i2OGdd} \,|\, Km_{Fe^{2+}} > 0.02\,\text{mM}. \tag{46}$$

Rather than modeling the generic state, a response-network i.e. the behavior of the i2OGdd-members in response to a stressor agent can be investigated. In that study, iron deficiency and the compensatory response of the network of i2OGGdd-members to mitigate the effects of this stressor were studied.[43,44] A regression model of i2OGdd-members with lower affinities for ferrous iron $(\boldsymbol{L}_{\textbf{i2OGdd}})$ suggested that this subset may exhibit an increased sensitivity to falling iron levels. Functionally, this means that despite the continued presence of 2OG, this subset of enzymes may exhibit compromised catalytic activity.

$$\lim_{\text{Rate of reaction }(3a)\to 0} r_2(\boldsymbol{L}_{\textbf{i2OGdd}}) \to 0. \tag{47}$$

In contrast, and by definition, the subset of enzymes with higher affinities $(\boldsymbol{H}_{i2OGdd})$ will be able to retain iron at their active sites. However, this tight binding, too, will not allow i2OGdd-members to function optimally as catalysts.

$$\lim_{\text{Rate of reaction }(3a)\to 0} r_2(\boldsymbol{H}_{\textbf{i2OGdd}}) \to 0. \tag{48}$$

Clearly, in the presence of progressive iron deficiency state the catalytic activity of every i2OGdd-member of the modeled network will be compromised.

$$\lim_{\text{Rate of reaction }(3a)\to 0} r_2(\boldsymbol{H}_{\textbf{i2OGdd}} \cup \boldsymbol{L}_{\textbf{i2OGdd}}) \to 0. \tag{49}$$

Rewriting the value of MM-constant for all i2OGdd-members,

$$Km(\boldsymbol{H}_{\textbf{i2OGdd}} \cup \boldsymbol{L}_{\textbf{i2OGdd}}) = \frac{r_1}{r_{-1}} = Kd(\boldsymbol{H}_{\textbf{i2OGdd}} \cup \boldsymbol{L}_{\textbf{i2OGdd}}) \in \mathbb{R}_+. \tag{50}$$

This set of interactions is then easily modeled as a well stirred mixture of interacting reactants and whose trajectories can be computed from an SSA-generated dataset.

4. Conclusions

Stochastic simulations are accurate, unbiased, and support emergent phenomena. However, a major limitation of this approach in establishing biological relevance is the absence of robustness between runs even with identical simulation times. "TemporalGSSA" is an R-wrapper and addresses this problem by collating and partitioning SSA-generated datasets with identical simulation times into well-defined bins (trial, technical replicates, outcomes). The algorithms deployed by "TemporalGSSA" utilize a combination of linear and regression models to relate the time step (a parameter specific to an independent stochastic run) with the absolute numbers of molecules of each metabolite. The computed and thence summarized data are numerical estimates of the molar concentration of a metabolite which is dependent on the duration of the simulation. If the datasets are generated by varying the simulation times, "TemporalGSSA", will construct a time-dependent trajectory that provides the trials-per-technical replicate constraint is complied with. "TemporalGSSA", can thus, render stochastically generated metabolite data from independent runs comparable and thereby provide insights into mechanism(s) that govern complex immunological and biochemical systems.

Acknowledgment

This work is funded by an early career intramural grant (Code No. A-766) awarded to SK by the All India Institute of Medical Sciences (AIIMS, New Delhi, India). SK designed the study, formulated and developed the algorithms, conducted the analysis, collated and analyzed the data, and wrote all the codes and the paper.

Availability and Implementation

"TemporalGSSA", is available at Comprehensive R archive network (CRAN) with the URL (https://cran.r-project.org/package=TemporalGSSA). "TemporalGSSA", can also be updated or installed directly from the GUI (R-4.1.x) as "update.packages ('TemporalGSSA')" and "install.packages('TemporalGSSA')" from any of the CRAN mirrors.

References

1. Gillespie DT, Stochastic simulation of chemical kinetics, *Annu Rev Phys Chem* **58**:35–55, 2007, doi: 10.1146/annurev.physchem.58.032806.104637.
2. Gillespie DT, Roh M, Petzold LR, Refining the weighted stochastic simulation algorithm, *J Chem Phys* **130**:174103, 2009, doi: 10.1063/1.3116791.
3. Schnoerr D, Sanguinetti G, Grima R, The complex chemical Langevin equation, *J Chem Phys* **141**(2):024103, 2014, doi: 10.1063/1.4885345, PubMed PMID: 25027995.
4. Ghosh A, Leier A, Marquez-Lago TT, The spatial chemical Langevin equation and reaction diffusion master equations: Moments and qualitative solutions, *Theor Biol Med Model* **12**:5, 2015, doi: 10.1186/s12976-015-0001-6.

5. Lopez CF, Muhlich JL, Bachman JA, Sorger PK, Programming biological models in Python using PySB, *Mol Syst Biol* **9**:646, 2013, doi: 10.1038/msb.2013.1.
6. Hepburn I, Chen W, Wils S, De Schutter E, STEPS: Efficient simulation of stochastic reaction-diffusion models in realistic morphologies, *BMC Syst Biol* **6**:36, 2012, doi: 10.1186/1752-0509-6-36.
7. Sanft KR, Wu S, Roh M, Fu J, Lim RK, Petzold LR, StochKit2: Software for discrete stochastic simulation of biochemical systems with events, *Bioinformatics* **27**(17):2457–2458, 2011, doi: 10.1093/bioinformatics/btr401.
8. Andrews SS, Addy NJ, Brent R, Arkin AP, Detailed simulations of cell biology with Smoldyn 2.1., *PLoS Comput Biol* **6**(3):e1000705, 2010, doi: 10.1371/journal.pcbi.1000705.
9. Faeder JR, Blinov ML, Hlavacek WS, Rule-based modeling of biochemical systems with BioNetGen, *Meth Mol Biol* **500**:113–167, 2009, doi: 10.1007/978-1-59745-525-1_5.
10. Kerr RA *et al.*, Fast Monte Carlo simulation methods for biological reaction-diffusion systems in solution and on surfaces, *SIAM J Sci Comput* **30**(6):3126, 2008, doi: 10.1137/070692017.
11. Harris LA, Clancy P, A "partitioned leaping"" approach for multiscale modeling of chemical reaction dynamics, *J Chem Phys* **125**(14):144107, 2006, doi: 10.1063/1.2354085.
12. Hoops S *et al.*, COPASI–a COmplex PAthway SImulator, *Bioinformatics* **22**(24):3067–3074, 2006, doi: 10.1093/bioinformatics/btl485.
13. Drawert B *et al.*, Stochastic simulation service: Bridging the gap between the computational expert and the biologist, *PLoS Comput Biol* **12**(12):e1005220, 2016, doi: 10.1371/journal.pcbi.1005220.
14. Hattne J, Fange D, Elf J, Stochastic reaction-diffusion simulation with MesoRD, *Bioinformatics* **21**(12):2923–2924, 2005, doi: 10.1093/bioinformatics/bti431, PubMed PMID: 15817692.
15. Tomita M *et al.*, E-CELL: Software environment for whole-cell simulation, *Bioinformatics* **15**(1):72–84, 1999, doi: 10.1093/bioinformatics/15.1.72.
16. Schaff J, Fink CC, Slepchenko B, Carson JH, Loew LM, A general computational framework for modeling cellular structure and function, *Biophys J* **73**(3):1135–1146, 1997, doi: 10.1016/S0006-3495(97)78146-3.
17. Drawert B, Trogdon M, Toor S, Petzold L, Hellander A, MOLNs: A cloud platform for interactive, reproducible, and scalable spatial stochastic computational experiments in systems biology using PyURDME, *SIAM J Sci Comput* **38**(3):C179–C202, 2016, doi: 10.1137/15M1014784.
18. Somogyi ET *et al.*, libRoadRunner: A high performance SBML simulation and analysis library, *Bioinformatics* **31**(20):3315–3321, 2015, doi: 10.1093/bioinformatics/btv363.
19. Schoneberg J, Noe F, ReaDDy–a software for particle-based reaction-diffusion dynamics in crowded cellular environments, *PLoS One* **8**(9):e74261, 2013, doi: 10.1371/journal.pone.0074261.
20. Kundu S, Mathematical modeling and stochastic simulations suggest that low-affinity peptides can bisect MHC1-mediated export of high-affinity peptides into "early"- and "late"-phases, *Heliyon* **7**(7):e07466, 2021, doi: 10.1016/j.heliyon.2021.e07466.
21. Kundu S, Stochastic modelling suggests that an elevated superoxide anion — hydrogen peroxide ratio can drive extravascular phagocyte transmigration by lamellipodium formation, *J Theor Biol* **407**:143–154, 2016, doi: 10.1016/j.jtbi.2016.07.002.
22. Thomas C, Tampe R, Structure of the TAPBPR-MHC I complex defines the mechanism of peptide loading and editing, *Science* **358**:1060–1064, 2017, doi: 10.1126/science.aao6001.
23. Thomas C, Tampe R, Proofreading of peptide-MHC complexes through dynamic multivalent interactions, *Front Immunol* **8**:65, 2017, doi: 10.3389/fimmu.2017.00065.

24. Dong G, Wearsch PA, Peaper DR, Cresswell P, Reinisch KM, Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer, *Immunity* **30**:21–32, 2009, doi: 10.1016/j.immuni.2008.10.018.
25. Wang Z, Zhang L, Qiao A, Watson K, Zhang J, Fan GH, Activation of CXCR4 triggers ubiquitination and down-regulation of major histocompatibility complex class I (MHC-I) on epithelioid carcinoma HeLa cells, *J Biol Chem* **283**:3951–3959, 2008, doi: 10.1074/jbc.M706848200.
26. Momburg F, Tan P, Tapasin-the keystone of the loading complex optimizing peptide binding by MHC class I molecules in the endoplasmic reticulum, *Mol Immunol* **39**:217–233, 2002, doi: 10.1016/s0161-5890(02)00103-7.
27. Praveen PV *et al.*, Tapasin edits peptides on MHC class I molecules by accelerating peptide exchange, *Eur J Immunol* **40**:214–224, 2010, doi: 10.1002/eji.200939342.
28. Schneeweiss C *et al.*, The mechanism of action of tapasin in the peptide exchange on MHC class I molecules determined from kinetics simulation studies, *Mol Immunol* **46**: 2054–2063, 2009, doi: 10.1016/j.molimm.2009.02.032.
29. Hewitt EW, The MHC class I antigen presentation pathway: strategies for viral immune evasion, *Immunology* **110**:163–169, 2003, doi: 10.1046/j.1365-2567.2003.01738.x.
30. Dissemond J, Kothen T, Mors J, Weimann TK, Lindeke A, Goos M, Wagner SN, Downregulation of tapasin expression in progressive human malignant melanoma, *Arch Dermatol Res* **295**:43–49, 2003, doi: 10.1007/s00403-003-0393-8.
31. Hattori H, Subramanian KK, Sakai J, Luo HR, Reactive oxygen species as signaling molecules in neutrophil chemotaxis, *Commun Integr Biol* **3**(3):278–281, 2010, doi: 10.4161/cib.3.3.11559.
32. Andrew N, Insall RH, Chemotaxis in shallow gradients is mediated independently of PtdIns 3-kinase by biased choices between random protrusions, *Nat Cell Biol* **9**(2):193–200, 2007, doi: 10.1038/ncb1536.
33. Forman HJ, Fukuto JM, Torres M, Redox signaling: Thiol chemistry defines which reactive oxygen and nitrogen species can act as second messengers, *Am J Physiol Cell Physiol* **287**(2):C246–C256, 2004, doi: 10.1152/ajpcell.00516.2003.
34. Mikkelsen RB, Wardman P, Biological chemistry of reactive oxygen and nitrogen and radiation-induced signal transduction mechanisms, *Oncogene* **22**(37):5734–5754, 2003, doi: 10.1038/sj.onc.1206663.
35. Pike LJ, Lipid rafts: Bringing order to chaos, *J Lipid Res* **44**(4):655–667, 2003, doi: 10.1194/jlr.R200021-JLR200.
36. Grigoriev I, Borisy G, Vorobjev I, Regulation of microtubule dynamics in 3T3 fibroblasts by Rho family GTPases, *Cell Motil Cytoskeleton* **63**(1):29–40, 2006, doi: 10.1002/cm.20107.
37. Marklund SL, Human copper-containing superoxide dismutase of high molecular weight, *Proc Natl Acad Sci USA* **79**(24):7634–7638, 1982, doi: 10.1073/pnas.79.24.7634.
38. Gao F *et al.*, Extracellular superoxide dismutase inhibits inflammation by preventing oxidative fragmentation of hyaluronan, *J Biol Chem* **283**(10):6058–6066, 2008, doi: 10.1074/jbc.M709273200.
39. Karlsson K, Sandstrom J, Edlund A, Marklund SL, Turnover of extracellular-superoxide dismutase in tissues, *Lab Invest* **70**(5):705–710, 1994.
40. Stralin P, Marklund SL, Effects of oxidative stress on expression of extracellular superoxide dismutase, CuZn-superoxide dismutase and Mn-superoxide dismutase in human dermal fibroblasts, *Biochem J* **298**(Pt 2):347–352, 1994, doi: 10.1042/bj2980347.
41. Kundu S, Co-operative intermolecular kinetics of 2-oxoglutarate dependent dioxygenases may be essential for system-level regulation of plant cell physiology, *Front Plant Sci* **6**:489, 2015, doi: 10.3389/fpls.2015.00489.

42. Herr CQ, Hausinger RP, Amazing diversity in biochemical roles of Fe(II)/2-oxoglutarate oxygenases, *Trends Biochem Sci* **43**(7):517–532, 2018, doi: 10.1016/j.tibs.2018.04.002.
43. McDonough MA, Loenarz C, Chowdhury R, Clifton IJ, Schofield CJ, Structural studies on human 2-oxoglutarate dependent oxygenases, *Curr Opin Struct Biol* **20**(6):659–672, 2010, doi: 10.1016/j.sbi.2010.08.006.
44. Kundu S, Fe(2)OG: An integrated HMM profile-based web server to predict and analyze putative non-haem iron(II)- and 2-oxoglutarate-dependent dioxygenase function in protein sequences, *BMC Res Notes* **14**(1):80, 2021, doi: 10.1186/s13104-021-05477-z.

Siddhartha Kundu is an Assistant Professor of Biochemistry at the prestigious All India Institute of Medical Sciences (New Delhi, India). He completed his post-graduate training in Biochemistry after graduating from medical school (M.B.B.S., M.D.). Thereafter, he worked at several research and training institutions, an effort which culminated in a Ph.D. in Computational Biology and Bioinformatics. His training in medicine, molecular biology and research mathematics has resulted in a considerable body of work over the last seven years that is interdisciplinary and involves a mathematical treatment of biochemical processes. He has extensive experience in coding with PERL and R and is able to complement his research work by developing web servers and software. These are freely accessible and available from repositories such as CRAN. He is deeply interested in comprehending and contributing to the mathematics that underlies and characterizes biochemical networks. He has modeled complex biochemical and cellular processes at various levels (atomistic, molecular, compartmental). These theoretical studies have provided valuable insights into immune-system-relevant chemotaxis, peptide transport and catalysis. He also has a deep interest in comprehending the mechanistic behavior of biologically relevant enzymes. He has developed and rigorously established a mathematical expression which integrates the output of Hidden Markov Models (HMM) into an Artificial Neural Network (ANN).