RESEARCH ARTICLE

pSSAlib: The partial-propensity stochastic chemical network simulator

Oleksandr Ostrenko[1,2], Pietro Incardona[3,4,5], Rajesh Ramaswamy[6]*, Lutz Brusch[1,2], Ivo F. Sbalzarini[2,3,4,5]*

1 Center for Information Services and High Performance Computing, TU Dresden, Dresden, Germany, **2** Center for Advancing Electronics Dresden, TU Dresden, Dresden, Germany, **3** Chair of Scientific Computing for Systems Biology, Faculty of Computer Science, TU Dresden, Dresden, Germany, **4** MOSAIC Group, Center for Systems Biology Dresden, Dresden, Germany, **5** Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany, **6** Max Planck Institute for the Physics of Complex Systems, Dresden, Germany

* raj.r11235@gmail.com (RR); ivos@mpi-cbg.de (IFS)

Abstract

Chemical reaction networks are ubiquitous in biology, and their dynamics is fundamentally stochastic. Here, we present the software library pSSAlib, which provides a complete and concise implementation of the most efficient partial-propensity methods for simulating exact stochastic chemical kinetics. pSSAlib can import models encoded in Systems Biology Markup Language, supports time delays in chemical reactions, and stochastic spatiotemporal reaction-diffusion systems. It also provides tools for statistical analysis of simulation results and supports multiple output formats. It has previously been used for studies of biochemical reaction pathways and to benchmark other stochastic simulation methods. Here, we describe pSSAlib in detail and apply it to a new model of the endocytic pathway in eukaryotic cells, leading to the discovery of a stochastic counterpart of the cut-out switch motif underlying early-to-late endosome conversion. pSSAlib is provided as a stand-alone command-line tool and as a developer API. We also provide a plug-in for the SBMLToolbox. The open-source code and pre-packaged installers are freely available from http://mosaic.mpi-cbg.de.

This is a *PLOS Computational Biology* Software paper.

Introduction

Modeling chemical reaction networks is key to unraveling the mechanisms underlying cell physiology and pathology [1–3]. Given the complexity of biochemical reaction networks and the large number of interacting species, computational studies are often necessary to guide experimental assays and to identify candidates for wet-lab perturbation. Numerical computer simulations provide full control over all parameters of the model, enabling "what if" studies that would be hard to address otherwise.

Such computational studies, however, are often complicated by the fact that molecules move and react stochastically. Biochemical reaction networks in sub-cellular compartments are frequently characterized by low copy numbers for some of the involved molecules, which leads to stochastic kinetics also of macroscopic quantities that can be qualitatively different from the deterministic predictions of standard reaction-rate equations [4–7]. In addition, biochemical reactions often incur a time delay (e.g., chaperoning through the nuclear pore complex, multistage binding, and mRNA translation), causing additional fluctuations in the molecular concentrations. Lastly, the spatial distribution of chemical species may play an important role within the same compartment, as is for example the case for diffusion in the endoplasmic reticulum [8]. A constructive role of fluctuations may also lie at the core of biological information processing, e.g., in transcription and translation [9, 10].

Accounting for the exact stochastic kinetics of a reaction network is typically done using exact stochastic simulation algorithms (SSAs). They sample species populations from the exact solution of the chemical master equation (CME) [11], preserving all information about the molecular fluctuations in the system, and their dynamics. To accommodate time delays, some of these algorithms have been extended accordingly [12]. Finally, a classical way of accommodating spatial heterogeneity is to subdivide the system into a number of well-mixed sub-spaces and to perform the reaction sampling in each of them according to the original scheme, augmented by jump reactions between sub-spaces in order to model diffusion. This classical algorithm is known as the nextsubvolume method (NSM) [13]. Together, these approaches provide a toolbox to account for stochastic effects in chemical reaction networks with elementary rates.

The use of this toolbox for simulating large reaction networks with sufficient statistics is mainly limited by the computational cost of the algorithms. Therefore, numerous improvements to the computational efficiency of exact SSAs have been presented in the past, notably the next reaction method (NRM) by Gibson and Bruck [14] and the SSA with Composition-Rejection Sampling (SSA-CR) [15]. While all other known variants of exact SSA have a computational cost that scales linearly with the number of reactions in the network [16], NRM reduced the runtime to scale with the logarithm of the number of reactions and SSA-CR further reduced it to constant-time. These improvements, however, only hold for weakly coupled networks [16]. For strongly coupled networks, the scaling of the computational cost remains linear with the number of reactions, which is unfavorable when simulating strongly coupled biochemical networks that contain many more reactions than species. A new class of SSAs based on factored-out reaction propensities has therefore been proposed [16] to improve the cost also for strongly coupled networks (see Partial-propensity methods in Sec. Design and implementation for the concept underlying this new class of methods). These so-called partial-propensity SSAs have a computational cost that scales at most linearly with the number of species, but not with the number of reactions, even for strongly coupled reaction networks. Two such algorithms are the partial-propensity direct method(PDM) [16] and the sorting partial-propensity direct method (SPDM) [16]. For weakly coupled networks, the computational cost is further reduced to constant-time using the partial-propensity SSA with Composition-Rejection Sampling (PSSA-CR) [17]. Partial-propensity methods [18] have also been extended to reactions with time delays (dPDM, the delay partial-propensity direct method) [19], and to spatiotemporal reaction-diffusion systems (PSRD, the partial-propensity stochastic reaction-diffusion method) [20]. Due to the state-of-the-art computational performance of partial-propensity methods, they are frequently used to benchmark newer SSAs [21–24]. In addition, particular instances of partial-propensity methods have been included in third-party software libraries, such as *ngss* [21] and *ssapredict* [24]. However, no complete and concise software

implementation of all partial-propensity algorithms is available with support for delayed chemical reactions and spatiotemporal dynamics.

Here, we present the partial-propensity stochastic simulation algorithms library (pSSAlib), a fast, flexible, and user-friendly implementation of all known partial-propensity SSAs. This open-source library is implemented in C++, providing both an Application Programming Interface (API) for incorporation into other software projects and a stand-alone simulator application for direct use. The software imports biochemical model specifications from Systems Biology Markup Language (SBML) files and provides tools for statistical analysis and visualization of the simulation results. Moreover, it provides a plug-in for the SBMLToolbox [25], enabling the user to change the model using a graphical user interface, and it supports the use of parallel computing resources to parallelize over simulation trajectories using the Message Passing Interface (MPI) [26]. pSSAlib adds partial-propensity methods to the family of already available SSA simulation software packages, including Cain [27], StochKit2 [28], libRoadRunner [29], and URDME [30]. This further extends the choice for users to pick the appropriate and most efficient SSA formulation for a problem at hand.

We present the software implementation of pSSAlib, validate pSSAlib, and benchmark its computational efficiency in standard test cases, comparing with the efficient NRM implementations provided by Cain [27]. Then, we use the software in order to gain insight into a new stochastic model of early-to-late endosome conversion in the endocytic pathway [31]. Besides demonstrating practical use of pSSAlib, the results show that early-to-late endosome conversion is robust to intrinsic fluctuations arising from low copy numbers of signaling molecules in the endocytic pathway. Crucially, the stochastic effects modify the non-monotonic signal-response characteristics of the deterministic counterpart of the cut-out switch and may explain the observed variability of Rab5 density at the onset of early-to-late endosome conversion in the endocytic pathway [32].

Design and implementation

Partial-propensity methods

As any Monte Carlo (MC) method, SSAs requires sampling a significant number of trajectories from the CME in order to acquire sufficient statistics of the system's behavior [11]. At each time step, sampling is performed over the reaction propensities, which scales linearly with the number of reactions in the network. Partial-propensity methods reduce this computational cost by factorizing the reaction propensities. The propensity of a reaction is defined as the product of its specific probability rate and its reaction degeneracy, which counts the number of distinct ways in which reactant molecules can interact [33]. All molecules are assumed to be homogeneously distributed within the reactor. The partial propensity of a reaction is defined as the *propensity per molecule* of one of the reactants [16]. This requires that the reaction propensity has a structure that can be factorized. This is obviously the case for elementary chemical reactions, i.e., reactions involving at most two reactants, but also for other reactions where the propensity has a product structure.

The partial propensities are stored in a special data structure, called the *partial-propensity structure*. Instead of sampling over reaction propensities, partial-propensity SSAs sample over partial propensities, which can be interpreted as sampling *reaction partners* instead of reactions. Hence, instead of asking "which reaction happens next?", partial-propensity SSAs ask: "which molecule is going to participate in the next reaction?" and then, given that: "which partner is it going to react with?" This means that instead of choosing from the set of reactions, the methods choose from the set of species, which explains why their computational cost scales with species number rather than reaction number.

Overview of pSSAlib

We give an overview of the partial-propensity stochastic simulation algorithms library (pSSAlib), a complete and portable implementation of all partial-propensity SSAs, complemented with tools for running simulations and analyzing the results.

Features. pSSAlib has been designed to facilitate the process of simulating stochastic chemical reaction networks using partial-propensity methods. pSSAlib is written in C++ and is designed to be highly efficient, also supporting parallel execution on distributed-memory computer clusters. It provides a simple and unified interface to SSAs for simulating networks of elementary chemical reactions, including delayed reactions (both consuming and non-consuming) and spatiotemporal reaction-diffusion systems in box-shaped domains. In addition to elementary reactions, pSSAlib can also simulate non-elementary reactions with at most two reactant species of which one may have a stoichiometry larger than one (i.e., reactions of the type $A + NB \to \ldots$, $N = 1, 2, 3, \ldots$).

The following SSAs are currently implemented in the library:

- Gillespie's direct method (DM) [11] as a reference;

- partial-propensity direct method (PDM) [16];

- sorting partial-propensity direct method (SPDM) [16];

- partial-propensity SSA with Composition-Rejection Sampling (PSSA-CR) [17];

- delay partial-propensity direct method (dPDM) [19]

- partial-propensity stochastic reaction-diffusion method (PSRD) [20].

Apart from the classes provided by the C++ Standard Template Library (STL), pSSAlib requires other external open-source libraries: random number generators from the GNU Scientific Library (GSL) [34], Boost header-only libraries [35], and the SBML library [36] for importing model descriptions. The supported model specifications are based on SBML Level 2 Version 4 specifications. A feature comparison between pSSAlib and other popular SSA simulation packages is shown in Table 1. pSSAlib does not support the definition of Events, and spatial reaction-diffusion simulations are limited to uniform Cartesian grids using the standard NSM [13], whereas URDME supports arbitrary geometries using tetrahedral meshes. Since the diffusion part in pSSAlib is standard, we expect speedups to mainly come from the reaction part, where the use of partial-propensity methods is particularly beneficial for strongly-coupled reaction networks and for reaction-dominated dynamics.

User interface. pSSAlib functionality can be accessed by either (i) direct calls from C++ code using the library's API or (ii) using the pSSAlib Command Line Interface (CLI). The input can be either an SBML model from a file or defined dynamically and passed to the library using the API. Parameters, including diffusion constants, reaction rates, and time delays are defined in the SBML model as annotations. In the case of a model file, this can either be done by manual editing or through the SBMLToolbox [25] plug-in of pSSAlib.

Integration with the SBMLToolbox. In order to facilitate manual editing of reaction rates, reaction delays, and diffusion constants, pSSAlib includes a plug-in for the SBMLToolbox [25], see Fig 1a. For ease of installation, we provide a pre-packaged distribution of the SBMLToolbox for download on the pSSAlib website. Alternatively, the plug-in can also be added to any existing installation of the SBMLToolbox by following the instructions on our website. The plug-in is activated via the respective menu item and, when active, indicates whether the model components (reactions and species) contain valid annotations that can be processed by pSSAlib. The plug-in also provides a graphical user interface for editing the

Table 1. Feature comparison of pSSAlib with other exact SSA simulation packages. (SBML: supports SBML model input, pSSA: provides partial-propensity SSAs, Diffusion: provides spatiotemporal stochastic reaction-diffusion, Delays: provides chemical reactions with time delays, Events: provides event triggers).

Software package	SBML	pSSA	Diffusion	Delays	Events
pSSAlib 2.0.0	☑	☑	Cartesian NSM	☑	☐
libRoadRunner 1.3.0 [29]	☑	☐	☐	☐	☑
StochKit 2.0.13 [28]	☑	☐	☐	☐	☑
Cain 1.10 [27]	☑	☐	☐	☐	☑
URDME 1.3 [30]	☐	☐	Tetrahedral NSM	☐	☐

https://doi.org/10.1371/journal.pcbi.1005865.t001

respective annotations. Reaction annotations contain the reaction rates and time delays with their respective values, see Fig 1b. The plug-in uses a color code to highlight SBML nodes that can be edited, marking a node without annotation with a yellow background, red background for an invalid annotation, and green background for a valid node, see Fig 1.

Validation

We use two standard reaction networks (see Validation test cases in Sec. Supporting information) to validate the correctness of the partial-propensity methods implemented in pSSAlib. For each case, different numbers of MC samples are acquired, starting from time $t = 0$ until steady state. The simulated steady-state probability distribution functions (PDFs) of species populations are shown in Fig 2a and 2b along with the corresponding analytical PDFs. To quantify similarity between empirical and analytical PDFs, we use the Kullback-Leibler divergence. Fig 2c and 2d show that the simulated PDFs converge to the analytical ones. The
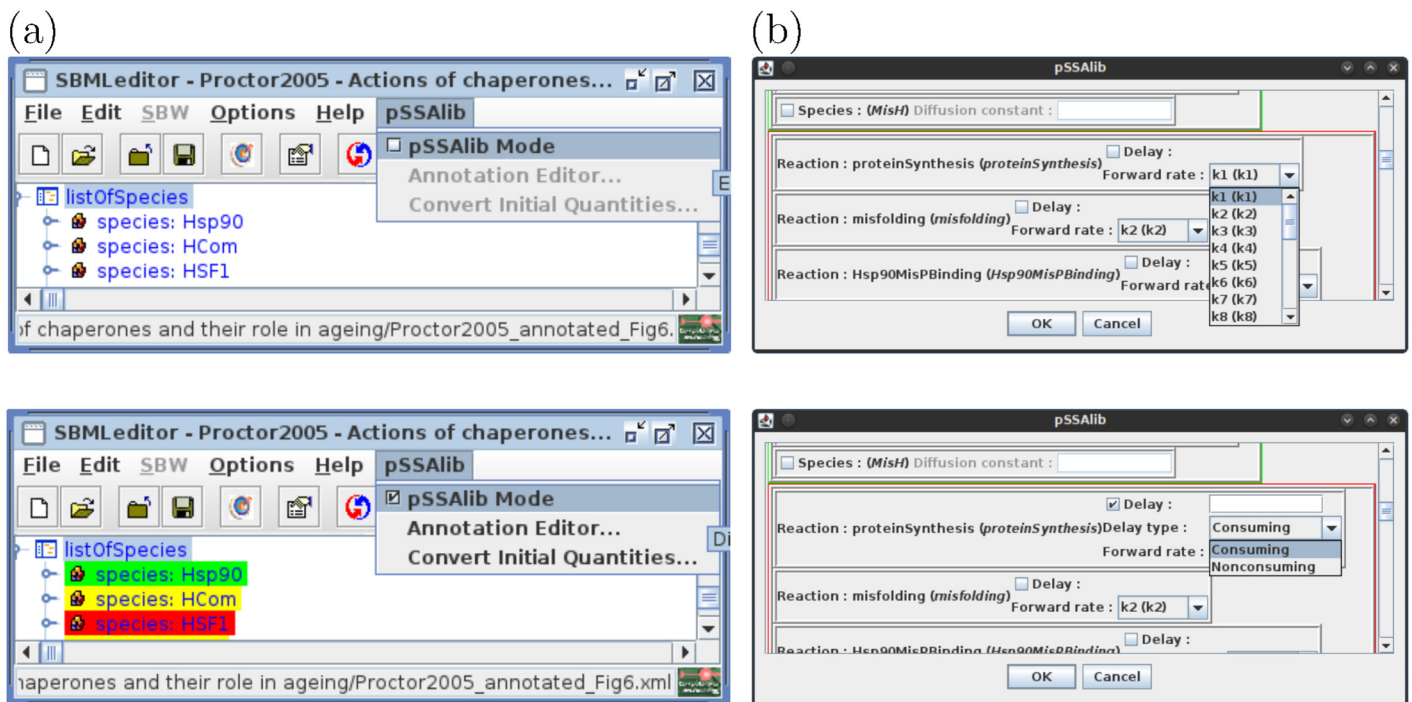
Fig 1. a. Screenshots of the new pSSAlib plug-in for the SBMLeditor, offering a graphical user interface for model editing using the SBMLToolbox [25]. **b.** Screenshots of the annotation editor for simulations using pSSAlib.

https://doi.org/10.1371/journal.pcbi.1005865.g001
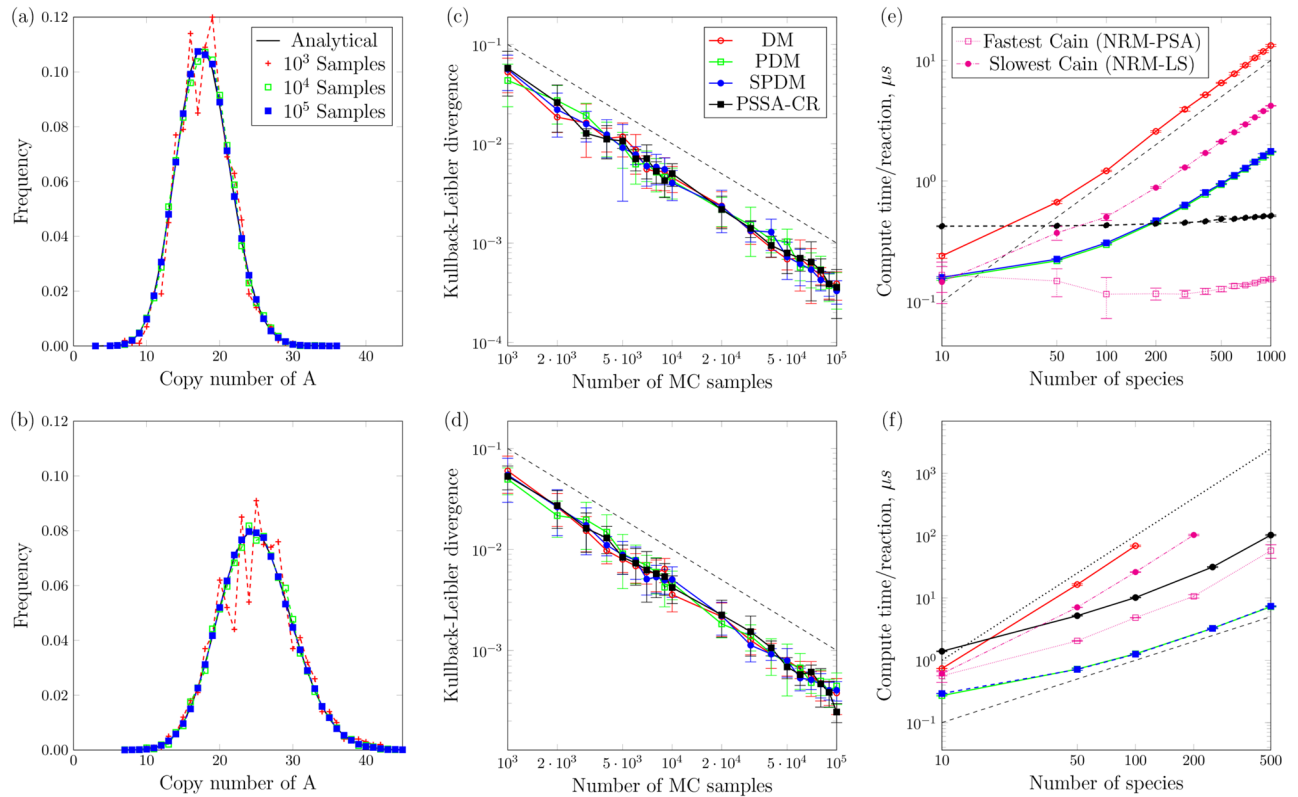
Fig 2. a-d. Results for the homo- (top row) and heteroreaction (bottom row) test cases (see Validation test cases in Sec. Supporting information): **a,b**. Simulated PDFs computed from different numbers of MC samples (symbols) using DM and plotted in comparison with the analytical PDF (solid line); **c,d**. Kullback-Leibler divergence between the analytical PDFs and the simulated ones using DM, PDM, SPDM, and PSSA-CR, computed for different numbers of samples in log—log scale. Each point represents 10 repetitions with mean and standard deviation indicated. As a guide, the black dashed line represents a slope of S^{-1}, where S is the number of MC samples. **e,f**. Comparison of computational costs for various partial-propensity SSAs with respect to DM and the fastest (on this problem NRM-PSA) and slowest (on this problem NRM-LS) NRM variants implemented in Cain [27]. **e**. Results for the weakly coupled Cyclic Linear Chain model (see Benchmark test cases in Sec. Supporting information). **f**. Results for the strongly coupled Colloidal Aggregation model. In both panels, the black dashed line represents a slope of N^1 (i.e., linear scaling), and the black dotted line in panel *(f)* has a slope of N^2, where N is the number of species. Error bars in *c-f* represent standard deviations.

https://doi.org/10.1371/journal.pcbi.1005865.g002

weak (i.e., in distribution) order of convergence of −1 (dashed black line) is expected for a MC method with a strong order of convergence of −1/2.

Benchmarking

We benchmark the computational performance of pSSAlib on weakly coupled and strongly coupled reaction networks and compare with the highly optimized nrm implementations provided by Cain [27]. We use the cyclic linear chain model and the colloidal aggregation model as representative examples of the two classes of reaction networks (see Benchmark test cases in Sec. Supporting information). Fig 2e and 2f show the average computer runtime per simulated reaction as a function of the number of species for the pSSAlib implementations of PDM, SPDM, PSSA-CR, and DM, compared with the fastest (on this problem: next reaction method partition size adaptive (NRM-PSA)) and slowest (on this problem: next reaction method linear search (NRM-LS)) NRM implementations offered by the Cain software package [27]. The results confirm that the runtime of PDM and SPDM is linear in the number of species. As a result, they are orders of magnitude faster than traditional SSAs such as DM, whose runtime is
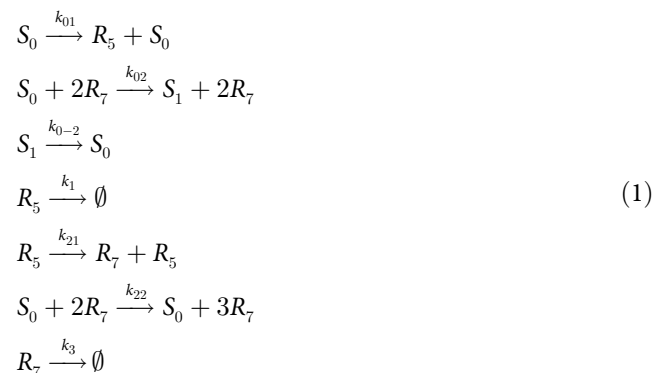
proportional to the number of reactions. In addition, for weakly coupled networks, PSSA-CR shows constant-time computational cost. For strongly coupled networks (Fig 2f), PDM and SPDM outperform all other methods. For weakly coupled networks (Fig 2e), NRM-PSA of Cain is the fastest, while PSSA-CR shows the expected constant-time scaling.

Results

New insights for a biochemical switch

Eukaryotic cells sort signaling molecules and nutrients through a number of distinct intracellular membrane compartments called endosomes. The decision to switch endosome types depends on the abundance of competing Rab GTPases, especially membrane-associated GTP-bound Rab5 (denoted R_5) and Rab7 (denoted R_7), see Fig 3a where the inactive GDP-bound forms are superscripted with a minus. Experimental evidence [31] suggests that a change in endosome type, characterized by a sharp decrease in R_5 copy number, is triggered at peak abundance of R_5 as the copy number of R_5's effectors increases (see Fig 3b). Theoretical investigation has revealed that such a behavior is characteristic of a cut-out switch [32, 37].

Previously, this cut-out switch has been modeled at the macroscopic level by ordinary differential equations (ODEs) using empirical species interactions [32]. Here, we propose a novel model for this switch using a reaction network that is composed of chemical reactions that obey mass-action kinetics (see Eq (1)). The corresponding species interactions are shown in Fig 3a. Here, S_0 and S_1 are the active and inactive forms of R_5's effector proteins, respectively. Likewise, S_2 and S_3 denote the active and inactive forms of R_7's effector proteins. The dynamics of S_2 and S_3, however, are considered identical to those of S_0 and S_1 in the present model and, hence, S_2 is replaced by S_0, and S_3 by S_1 in our model (see Eq (1)). By construction, the sum of copy numbers of S_0 and S_1 is a conserved quantity, which we denote S_{01}. As an early endosome accumulates signaling molecules and effector proteins (S_0, S_1) during multiple rounds of endosome fusion and fission [38], we interpret different fixed values of S_{01} as different degrees of endosome progression from nascent early endosomes at low S_{01} to large early endosomes at high S_{01}. The switch should be triggered by increasing the value of S_{01}. As shown below, our mass-action-based model reproduces the experimental observations of Refs. [31, 32]. Importantly, this novel set of 7 biochemical reactions (1) allows us to use pSSAlib to study the exact stochastic system behavior for low copy numbers of molecules. Low copy numbers are expected for the small (100–500 nm diameter [39]) endosomes.

$$
\begin{aligned}
S_0 &\xrightarrow{k_{01}} R_5 + S_0 \\
S_0 + 2R_7 &\xrightarrow{k_{02}} S_1 + 2R_7 \\
S_1 &\xrightarrow{k_{0-2}} S_0 \\
R_5 &\xrightarrow{k_1} \emptyset \\
R_5 &\xrightarrow{k_{21}} R_7 + R_5 \\
S_0 + 2R_7 &\xrightarrow{k_{22}} S_0 + 3R_7 \\
R_7 &\xrightarrow{k_3} \emptyset
\end{aligned}
\tag{1}
$$

We perform stochastic simulations of this model using PDM and SPDM as implemented in pSSAlib. Parameter values are given in the caption of Fig 3, and model analysis is repeated for a range of fixed parameter values S_{01}. The simulations reveal a cut-off switch pattern in the copy numbers of R_5 and R_7. Specifically, as the abundance of effector proteins crosses a
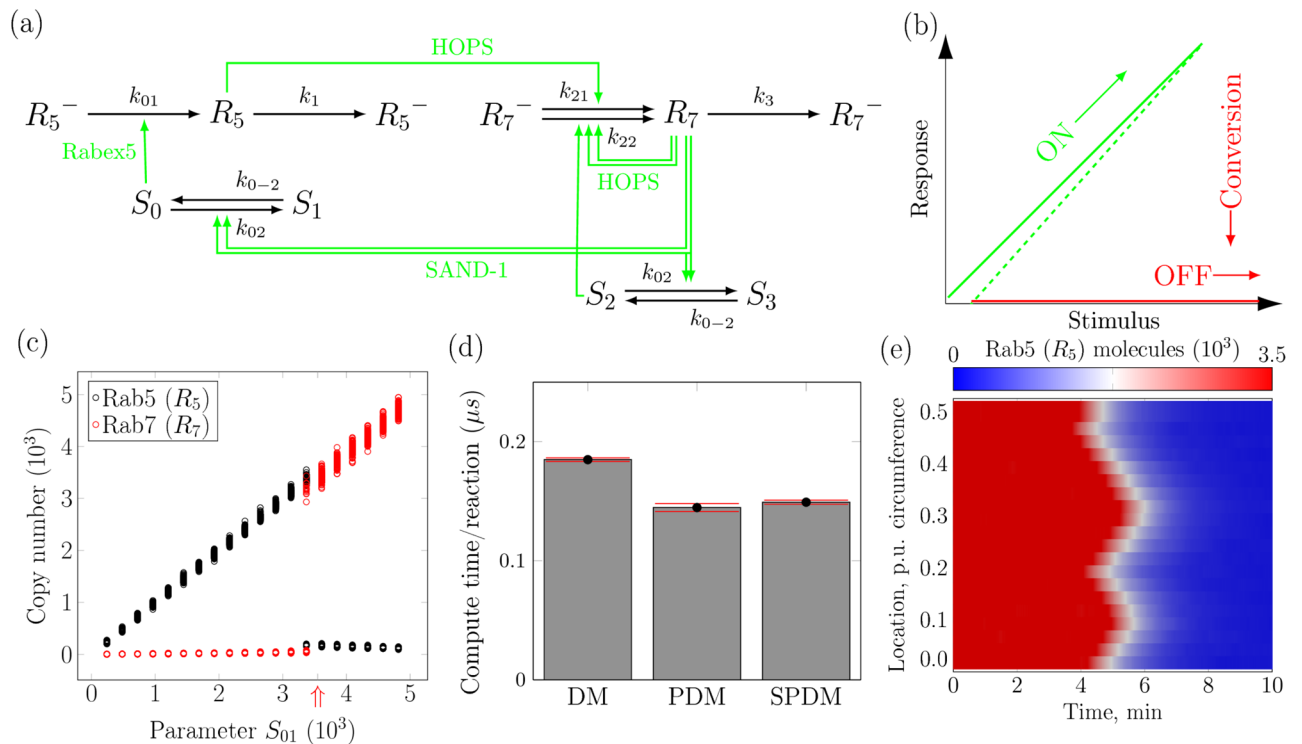
Fig 3. Biochemical network and self-organization underlying early-to-late endosome conversion in the endocytic pathway of eukaryotic cells. a. Novel reaction network diagram extending previous models of this system [32, 37]. See main text for species description; inactive GDP-bound forms are superscripted with a minus. **b**. Cut-out switch motif represented by its idealized stimulus-response diagram showing the underlying bifurcation structure. **c**. Simulated bifurcation diagram with respect to the total number of S_0 and S_1 molecules (parameter S_{01}). Symbols represent final states of stochastic simulations for 100 individual samples at each parameter value, each run for 100 min of simulated time using PDM with initial conditions corresponding to the early endosome state (R_5 = 1.0 mol m^{-3}, R_7 = 0, $S_0 = S_{01}$ and $S_1 = 0$). The S_{01} value indicated by the red arrow is used in panels d,e. Other parameter values are: k_{01} = 1 min^{-1}, k_{02} = 100 m^6 mol^{-2} min^{-1}, k_{0-2} = 10 min^{-1}, k_1 = 1 min^{-1}, k_{21} = 0.1 min^{-1}, k_{22} = 100 m^6 mol^{-2} min^{-1}, k_3 = 10 min^{-1}. **d**. Computer runtime per simulated reaction for DM, PDM, and SPDM with initial condition $R_5 = R_7 = 1$ mol m^{-3}. **e**. Kymograph of local Rab5 molecule number within 20 subvolumes with diffusive R_5 and R_7 coupling, mimicking sectors on the endosomal surface. Stochastic dynamics results in spontaneous local triggering of the switch, with fronts robustly propagating the switch across space. Compartment volume was 4×10^{-21} m^3 in all cases and both diffusion constants for R_5 and R_7 equal 10^{-21} m^2 min^{-1}.

threshold, we observe a sharp decrease in R_5 abundance accompanied by a sharp increase in R_7 abundance (see Fig 3c). This observation confirms that the present model, based solely on a chemical reaction network that obeys mass-action kinetics, reconstitutes the cut-out switch that triggers transition from early to late endosomes. In addition, we observe that even for such a small chemical reaction network, both PDM and SPDM are approximately 20% faster than DM (see Fig 3d).

Until now, we have considered only temporal dynamics of the cut-out switch model by neglecting diffusion of R_5 and R_7. Next, we perform spatiotemporal simulations using the PSRD implementation of pSSAlib. Specifically, we consider a one-dimensional membrane section of the endosome between its two poles and account for finite diffusivity of R_5 and R_7. Starting from uniform spatial profiles and abundance corresponding to early endosomes, we perform stochastic simulations using PSRD. In these simulations, the abundance of effector proteins is set close to the cut-out switch threshold observed in the well-mixed simulations. Fig 3e shows the kymograph of R_5 copy number obtained from the simulation. We observe that the switch in R_5 abundance is triggered at multiple locations on the membrane. These

locations are not deterministic, but determined by stochastic dynamics. Once the switch is triggered locally, the change in R_5 abundance propagates to the rest of the membrane until the stochastic activation waves fuse (see Fig 3e). This behavior is observed over many parameter combinations and for different initial conditions (data not shown).

Through these simulations, we show for the first time that the cut-out switch model is robust against fluctuations, which are intrinsic to biological cells. In addition, we show that spatial coupling renders the concentration of Rab GTPases spatially homogeneous at steady state, despite a local trigger, thereby ensuring unique identity of individual endosomes.

Discussion and conclusion

We have presented pSSAlib, the first complete and concise implementation of all exact partial-propensity SSAs. Partial-propensity SSAs reduce the computational cost of exact stochastic simulations of chemical reaction networks from scaling with the number of reactions to scaling at most with the smaller number of species. This is particularly beneficial for strongly coupled networks, where the number of reactions can approach the square of the number of species. In weakly coupled networks, the scaling of the computational cost of partial-propensity SSAs is identical to that of previous SSAs and is independent of system size.

pSSAlib implements and provides partial-propensity methods for strongly and weakly coupled networks, for networks involving time delays, and for spatiotemporal stochastic reaction-diffusion simulations in uniform Cartesian grids. It is implemented in C++ and can be used both stand-alone and as part of other software projects. pSSAlib is based on widely-used open-source libraries, including Boost [35], the GNU Scientific Library [34], the Message Passing Interface (MPI) [26] for parallel processing, and the SBML library [36] for portable model definition and exchange. The pSSAlib library is complemented by stand-alone software to simulate SBML models and to analyze the results. Additionally, we provide a plug-in to the SBMLToolbox for user-friendly model editing and parameter annotation (see Overview of pSSAlib in Sec. Design and implementation).

pSSAlib has previously been applied to a large biochemical reaction pathway with ATP turnover, revealing fluctuations in metabolite copy numbers on the order of 20% [23]. In the Heat-shock-response tutorial in Sec. Supporting information, we apply pSSAlib to a previously developed model of the heat shock response pathway in single cells [40]. In this manuscript, we have shown the application of pSSAlib to a novel stochastic model of endosome maturation.

We have described the architecture, functionality, and use of pSSAlib, and we have compared its functionality and computational performance with popular implementations of classical, full-propensity SSAs. Validation cases for which the exact solution of the chemical master equation is known have been used to verify correctness of the software implementation. The performance benchmarks have shown that the user-friendly implementation in pSSAlib has the beneficial scaling of partial-propensity SSAs and outperforms traditional SSAs for large or strongly coupled networks. Finally, we have applied pSSAlib and its analysis workflow to the open biological question of robust intracellular compartmentalization despite low copy numbers of the controlling molecules. To this end, we developed a new model of the endocytic pathway in eukaryotic cells and uncovered the stochastic counterpart of the cut-out switch motif underlying early-to-late endosome conversion.

Even though pSSAlib offers all partial-propensity SSAs, the choice of the most efficient method depends on problem type and size. As shown in our benchmarks, other implementations, such as the NRM variants offered by Cain [27], may be more efficient in some cases. In addition, for very small reaction networks, possibly none of the partial-propensity SSAs

amortize the overhead incurred by the partial-propensity data structures and the memory look-up therein. However, pSSAlib complements the family of SSA simulation packages and provides an efficient way for existing simulation packages, like Morpheus [41] and Copasi [42], to make partial-propensity methods available to their users.

Availability

The pSSAlib source code and pre-packaged installers are freely available from mosaic.mpi-cbg. de as open source under the GNU LGPL v3 license. The web site also contains an online manual, user tutorial, and a complete developer's reference to the API.

Supporting information

Heat-shock-response tutorial

We demonstrate the application of pSSAlib to a model of age-related impairment of the heat-shock response, showing that efficient simulations enable large sample sizes and that the correspondingly accurate population averages reproduce published results [40].

Chaperones play an important role in cell physiology by assisting in proper folding of nascent proteins, refolding of misfolded ones, catalyzing protein degradation in the lysosomes, and preventing protein aggregation. However, the amount of damaged protein in a cell increases with cell age, suggesting an impairment of the heat-shock response.

An SBML model [43] of this system is publicly available from the BioModels Database [44]. After downloading this model, we added the reaction annotations using the pSSAlib plug-in for the SBMLToolbox and saved the file under a new name. The model was then ready to be processed by pSSAlib. We first reproduced the results from Fig. 2 of Ref. [40], covering tens of thousands of chaperone-protein interactions in an unstressed cell.

We let the simulation engine sample 100 trajectories using SPDM as the simulation algorithm. We chose to collect simulation output every 10 seconds of simulated time. The respective CLI call to the simulator read:

```
<path-to-pSSA-CLI>/simulator -m spdm -i <path-to-model-file> -n 100
--tend 10000 --dt 10 -o <output-path>
```

**Listing 1**. Simulator usage for 100 samples using SPDM.

We then used the pSSAlib analyzer to statistically analyze the simulation data and to plot the results. We selected the species to be included in the analysis using the −s command-line option followed by the respective species names as a comma-separated list. The two calls to the analyzer below serve to plot a single trajectory for the selected species by specifying both −n 1 and −r trajectories.

```
<path-to-pSSA-CLI>/analyzer -i <path-to-simulator-output>
    -s NatP, MisP, AggP, Hsp90 -n 1 -r trajectories -o <output-path>
--gnuplot
<path-to-pSSA-CLI>/analyzer -i <path-to-simulator-output>
    -s ATP, ADP, ROS -n 1 -r trajectories -o <output-path> --gnuplot
```

Listing 2. Analyzer usage to plot individual trajectories, as shown in Fig 4.

The −−gnuplot option causes the analyzer to write gnuplot scripts for visualizing the results. These were directly visualized by the open-source plotting program gnuplot, including legends with species names as given in the model specification; see Fig 4 for an example.

The results are shown in Fig 4a and 4b and confirm Figs. 2a and 2b of Ref. [40]. Similarly, a model extension that incorporates Hsp90 misfolding in a stressed cell (see Fig 4c and 4d) reproduced Fig. 6 of Ref. [40]. Both stochastic results were thereby confirmed using a different method than in the original work.
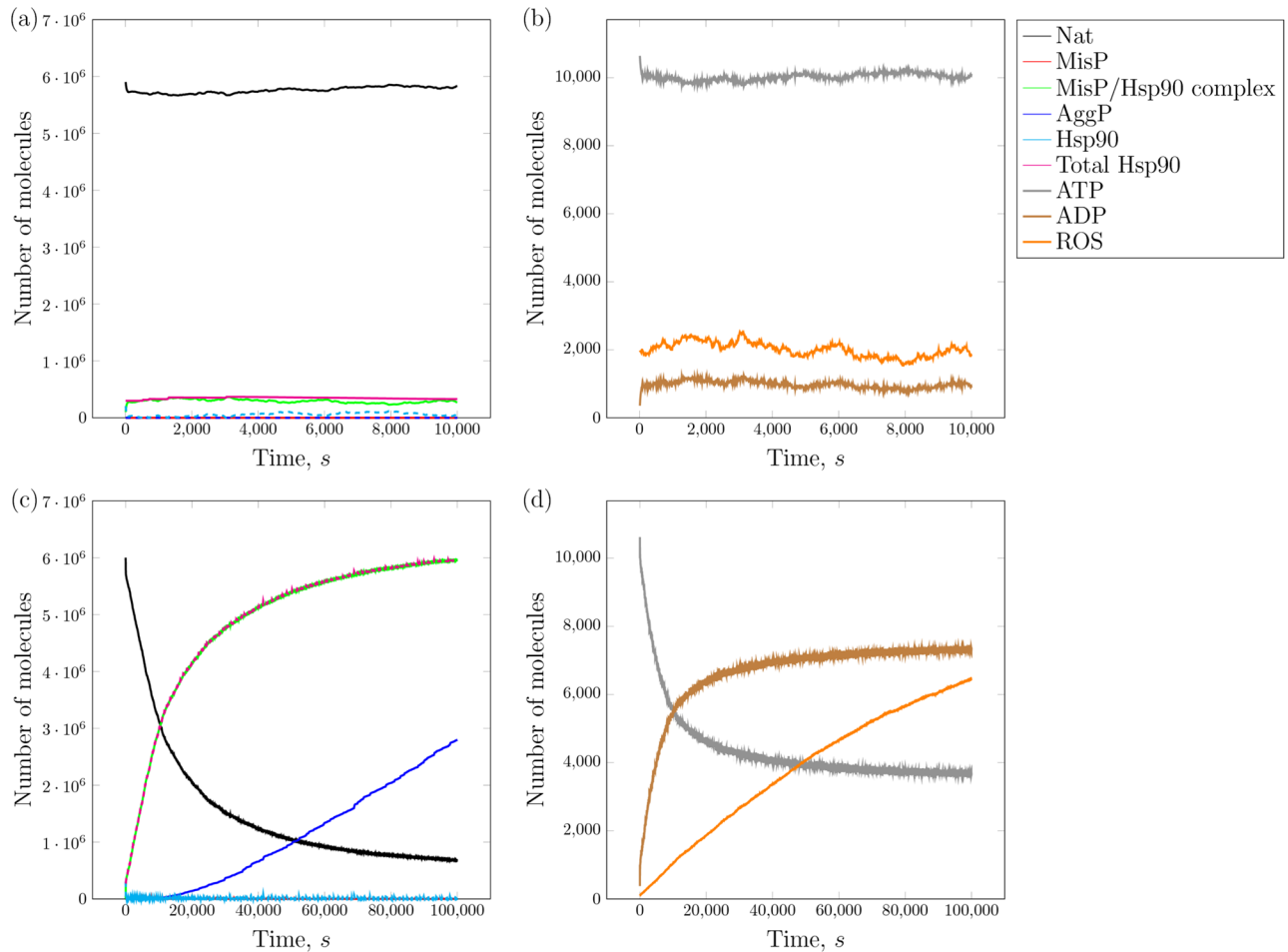
Fig 4. a,b. Heat-shock response of an unstressed cell. Visualization of the pSSAlib simulation results showing time courses for a single cell, confirming Figs. 2a and 2b from Ref. [40]. The legend of panel b applies to all panels. **c,d.** Heat-shock response of a stressed cell. Visualization of the pSSAlib simulation results showing time courses for a single cell, confirming Figs. 6a and 6b from Ref. [40]. The results shown here were obtained using SPDM as implemented in pSSAlib.

https://doi.org/10.1371/journal.pcbi.1005865.g004

Validation test cases

Homoreaction model. The following two reactions define the homoreaction test case [45]:

$$\mathrm{A} + \mathrm{A} \xrightarrow{k_1} \emptyset, \quad \emptyset \xrightarrow{k_2} \mathrm{A}. \tag{2}$$

Here, k_1 and k_2 are macroscopic rates. In addition, we consider that these reactions happen within a finite-sized compartment of volume V. Using simulations, we track the copy number $n(t)$ of molecules A. The parameter values are: $k_1/V = 0.016$ s^{-1}, $k_2V = 10$ s^{-1}. The initial copy number $n(t = 0)$ is set to 25. The analytical steady-state population distribution is:

$$\phi(n) = \frac{1}{\sqrt{2}I_1(2\sqrt{2K})} \frac{(\sqrt{K})^n}{n!} I_{n-1}\left(2\sqrt{K}\right), \tag{3}$$

where $\phi(n)$ is the discrete PDF for finding $n \geq 0$ molecules of species A at steady state, $K = \frac{k_2 V^2}{k_1}$, and $I_m(\cdot)$ is the modified Bessel function of the first kind.

Heteroreaction model. The heteroreaction test case [45] is defined by the following two reactions:

$$A + B \xrightarrow{k_1} B, \quad \emptyset \xrightarrow{k_2} A. \tag{4}$$

Here again k's are macroscopic rates, and we consider these reactions happening within a compartment of volume V. Using simulations, we track the copy number $n_a(t)$ of species A. For this reaction network, the copy number n_b of species B remains unchanged from its initial value. The parameter values are $k_1/V = 0.04 \text{ s}^{-1}$, $k_2 V = 1 \text{ s}^{-1}$. The initial copy numbers are $n_a(t = 0) = 25$ and $n_b(t) \equiv n_b(t = 0) = 1$. The analytical steady-state population distribution for species A is:

$$\phi(n_a) = \frac{1}{n_a!} \left(\frac{K}{n_b} \right)^{n_a} \exp\left(-\frac{K}{n_b} \right), \tag{5}$$

where $\phi(n_a)$ is the discrete PDF for finding n_a molecules A at steady state, n_b is the constant copy number of species B, and $K = \frac{k_2 V^2}{k_1}$.
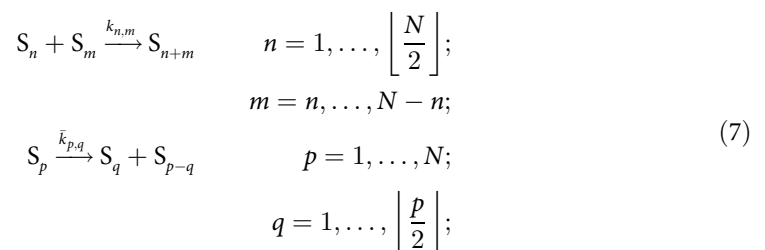
Benchmark test cases

Cyclic linear chain reaction network. We benchmark partial-propensity SSAs on a weakly coupled model in order to assess their limitations in cases where other SSA formulations might be more efficient. We choose the cyclic linear chain model from [16] since it is the most weakly coupled reaction network possible. For M reactions, it involves the minimum number of species $N = M + 1$, and the maximum out-degree of the dependency graph is constant at the smallest possible value of 2, since every reaction at most influences the population of its only reactant and of the only reactant of the subsequent reaction. The reactions of the cyclic linear chain model are given by:

$$\begin{aligned} S_i &\xrightarrow{k_i} S_{i+1} \quad i = 1, \ldots, N - 1. \\ S_N &\xrightarrow{k_N} S_1. \end{aligned} \tag{6}$$
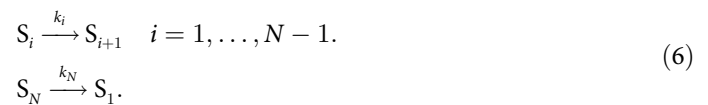
where S_i is chemical species i, k_i are macroscopic reaction rates, and N is the total number of species in the model. For the simulations we set $k_i = 1$ and $S_i(t = 0) = 1$ for all $i = 1, \ldots, N$.

Colloidal aggregation model. We use the colloidal aggregation model from [16] as an example of a strongly coupled reaction network that can be used to model, e.g., aggregation of solvated proteins, nanobeads, or viruses. For N chemical species it consists of $M = \lfloor \frac{N^2}{2} \rfloor$ reactions. The maximum out-degree of the dependency graph is $3N - 7$ and hence scales with system size, which is the definition of strong coupling. The reactions of the colloidal aggregation model are given by:

$$\begin{aligned} S_n + S_m &\xrightarrow{k_{n,m}} S_{n+m} \qquad n = 1, \ldots, \left\lfloor \frac{N}{2} \right\rfloor; \\ & \qquad\qquad\qquad m = n, \ldots, N - n; \\ S_p &\xrightarrow{\bar{k}_{p,q}} S_q + S_{p-q} \qquad p = 1, \ldots, N; \\ & \qquad\qquad\qquad q = 1, \ldots, \left\lfloor \frac{p}{2} \right\rfloor; \end{aligned} \tag{7}$$

where S$_i$ is chemical species i, $k_{n,m}$ and $\bar{k}_{p,q}$ are macroscopic reaction rates, and N is the total number of species in the model. For the simulations we set all rate constants to 1 and S$_i(t = 0) = 1$ for all $i = 1,\ldots, N$.

Acknowledgments

Author Contributions

Conceptualization: Ivo F. Sbalzarini.

Data curation: Oleksandr Ostrenko, Pietro Incardona.

Formal analysis: Rajesh Ramaswamy, Ivo F. Sbalzarini.

Funding acquisition: Lutz Brusch, Ivo F. Sbalzarini.

Investigation: Oleksandr Ostrenko, Pietro Incardona, Rajesh Ramaswamy, Ivo F. Sbalzarini.

Methodology: Rajesh Ramaswamy.

Project administration: Ivo F. Sbalzarini.

Resources: Pietro Incardona.

Software: Oleksandr Ostrenko, Pietro Incardona.

Supervision: Rajesh Ramaswamy, Lutz Brusch, Ivo F. Sbalzarini.

Validation: Oleksandr Ostrenko, Pietro Incardona, Rajesh Ramaswamy, Ivo F. Sbalzarini.

Visualization: Oleksandr Ostrenko, Ivo F. Sbalzarini.

Writing – original draft: Oleksandr Ostrenko, Rajesh Ramaswamy, Lutz Brusch, Ivo F. Sbalzarini.

Writing – review & editing: Lutz Brusch, Ivo F. Sbalzarini.

References

1. Tyson JJ, Chen KC, Novak B. Sniffers, buzzers, toggles and blinkers: Dynamics of regulatory and signaling pathways in the cell. Curr Opin Cell Biol. 2003; 15(2):221–231. https://doi.org/10.1016/S0955-0674(03)00017-6 PMID: 12648679

2. Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, et al. A community-driven global reconstruction of human metabolism. Nat Biotech. 2013; 31(5):419–425. https://doi.org/10.1038/nbt.2488

3. Sbalzarini IF. Modeling and simulation of biological systems from image data. Bioessays. 2013; 35 (5):482–490. https://doi.org/10.1002/bies.201200051 PMID: 23533152

4. Grima R. An effective rate equation approach to reaction kinetics in small volumes: Theory and application to biochemical reactions in nonequilibrium steady-state conditions. J Chem Phys. 2010; 133 (3):035101. https://doi.org/10.1063/1.3454685 PMID: 20649359

5. Ramaswamy R, Sbalzarini IF, González-Segredo N. Noise-Induced Modulation of the Relaxation Kinetics around a Non-Equilibrium Steady State of Non-Linear Chemical Reaction Networks. PLoS ONE. 2011; 6(1):e16045. https://doi.org/10.1371/journal.pone.0016045 PMID: 21297975

6. Ramaswamy R, Sbalzarini IF. Intrinsic noise alters the frequency spectrum of mesoscopic oscillatory chemical reaction systems. Sci Rep. 2011; 1:154. https://doi.org/10.1038/srep00154 PMID: 22545192

7. Ramaswamy R, González-Segredo N, Sbalzarini IF, Grima R. Discreteness-induced concentration inversion in mesoscopic chemical systems. Nat Commun. 2012; 3:779. https://doi.org/10.1038/ncomms1775 PMID: 22491327

8. Sbalzarini IF, Mezzacasa A, Helenius A, Koumoutsakos P. Effects of organelle shape on fluorescence recovery after photobleaching. Biophysical Journal. 2005; 89(3):1482–1492. https://doi.org/10.1529/biophysj.104.057885 PMID: 15951382

9. Paulsson J. Summing up the noise in gene networks. Nature. 2004; 427(6973):415–418. https://doi.org/10.1038/nature02257 PMID: 14749823

10. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. Nature. 2010; 467(7312):167–173. https://doi.org/10.1038/nature09326 PMID: 20829787

11. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry. 1977; 81(25):2340–2361. https://doi.org/10.1021/j100540a008

12. Cai X. Exact stochastic simulation of coupled chemical reactions with delays. J Chem Phys. 2007; 126 (12):124108. https://doi.org/10.1063/1.2710253 PMID: 17411109

13. Elf J, Ehrenberg M. Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. Systems Biology. 2004; 1(2):230–236. https://doi.org/10.1049/sb:20045021 PMID: 17051695

14. Gibson MA, Bruck J. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. J. Phys. Chem. A. 2000; 104:1876–1889. https://doi.org/10.1021/jp993732q

15. Slepoy A, Thompson AP, Plimpton SJ. A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks. J. Chem. Phys. 2008; 128:205101. https://doi.org/10.1063/1.2919546 PMID: 18513044

16. Ramaswamy R, Gonzalez-Segredo N, Sbalzarini IF. A new class of highly efficient exact stochastic simulation algorithms for chemical reaction networks. J Chem Phys. 2009; 130(24):244104. https://doi.org/10.1063/1.3154624 PMID: 19566139

17. Ramaswamy R, Sbalzarini IF. A partial-propensity variant of the composition-rejection stochastic simulation algorithm for chemical reaction networks. The Journal of Chemical Physics. 2010; 132 (4):044102. https://doi.org/10.1063/1.3297948 PMID: 20113014

18. Ramaswamy R, Sbalzarini IF. Fast Exact Stochastic Simulation Algorithms Using Partial Propensities. In: Proc. ICNAAM, Numerical Analysis and Applied Mathematics, International Conference. AIP; 2010. p. 1338–1341.

19. Ramaswamy R, Sbalzarini IF. A partial-propensity formulation of the stochastic simulation algorithm for chemical reaction networks with delays. J Chem Phys. 2011; 134(1):014106. https://doi.org/10.1063/1.3521496 PMID: 21218996

20. Ramaswamy R, Sbalzarini IF. Exact on-lattice stochastic reaction-diffusion simulations using partial-propensity methods. J Chem Phys. 2011; 135:244103. https://doi.org/10.1063/1.3666988 PMID: 22225140

21. Sanassy D, Fellerman H, Krasnogor N, Konur S, Mierla LM, Gheorghe M, et al. Modelling and stochastic simulation of synthetic biological boolean gates. In: Proceeding 16th IEEE International Conference on High Performance Computing and Communication; 2014.

22. Thanh VH, Zunino R, Priami C. On the rejection-based algorithm for simulation and analysis of large-scale reaction networks. J Chem Phys. 2015; 142:244106. https://doi.org/10.1063/1.4922923 PMID: 26133409

23. Ehlert K, Loewe L. Lazy Updating of hubs can enable more realistic models by speeding up stochastic simulations. The Journal of Chemical Physics. 2014; 141(20):204109. https://doi.org/10.1063/1.4901114 PMID: 25429935

24. Sanassy D, Widera P, Krasnogor N. Meta-Stochastic Simulation of Biochemical Models for Systems and Synthetic Biology. ACS Synthetic Biology. 2015; 4(1):39–47. https://doi.org/10.1021/sb5001406 PMID: 25152014

25. Keating SM, Bornstein BJ, Finney A, Hucka M. SBMLToolbox: an SBML toolbox for MATLAB users. Bioinformatics (Oxford, England). 2006; 22(10):1275–1277. https://doi.org/10.1093/bioinformatics/btl111

26. The Message Passing Interface Forum. MPI: A Message-Passing Interface Standard, Version 2.1. 2008.

27. Mauch S, Stalzer M. Efficient formulations for exact stochastic simulation of chemical systems. IEEE/ACM Trans. Comput. Biol. Bioinform. 2011; 8(1):27–35. https://doi.org/10.1109/TCBB.2009.47 PMID: 21071794

28. Sanft KR, Wu S, Roh M, Fu J, Lim RK, Petzold LR. StochKit2: software for discrete stochastic simulation of biochemical systems with events. Bioinformatics. 2011; 27(17):2457–2458. https://doi.org/10.1093/bioinformatics/btr401 PMID: 21727139

29. Somogyi ET, Bouteiller J-M, Glazier JA, König M, Medley JK, Swat MH, Sauro HM. libRoadRunner: a high performance SBML simulation and analysis library. Bioinformatics. 2015; 31(20):3315–3321. https://doi.org/10.1093/bioinformatics/btv363 PMID: 26085503

30. Drawert B, Engblom S, Hellander A. URDME: a modular framework for stochastic simulation of reaction-transport processes in complex geometries. BMC Systems Biology. 2012; 6:76. https://doi.org/10.1186/1752-0509-6-76 PMID: 22727185

31. Rink J, Ghigo E, Kalaidzidis Y, Zerial M. Rab conversion as a mechanism of progression from early to late endosomes. Cell. 2005; 122(5):735–749. https://doi.org/10.1016/j.cell.2005.06.043 PMID: 16143105

32. del Conte-Zerial P, Brusch L, Rink J, Collinet C, Kalaidzidis Y, Zerial M, et al. Membrane identity and GTPase cascades regulated by toggle and cut-out switches. Mol Syst Biol. 2008; 4:206. https://doi.org/10.1038/msb.2008.45 PMID: 18628746

33. Gillespie DT. A rigorous derivation of the chemical master equation. Physica A. 1992; 188(1-3):404–425. https://doi.org/10.1016/0378-4371(92)90283-V

34. Galassi M. GNU Scientific Library: reference manual for GSL version 1.12. Network Theory; 2009. Available from: http://www.worldcat.org/isbn/9780954612078.

35. Siek J, Lee LQ, Lumsdaine A. The Boost Library; 2011. http://www.boost.org/.

36. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics. 2003; 19(4):524–31. https://doi.org/10.1093/bioinformatics/btg015 PMID: 12611808

37. Poteryaev D, Datta S, Ackema K, Zerial M, Spang A. Identification of the Switch in Early-to-Late Endosome Transition. Cell. 2010; 141(3):497–508. https://doi.org/10.1016/j.cell.2010.03.011 PMID: 20434987

38. Foret L, Brusch L, Jülicher F. Theory of Cargo and Membrane Trafficking. In: Bradshaw RA, Stahl PD, editors. Encyclopedia of Cell Biology. vol. 4. Waltham: Academic Press; 2015. p. 56–62. Available from: http://www.sciencedirect.com/science/article/pii/B978012394447440009X.

39. Helmuth JA, Burckhardt CJ, Greber UF, Sbalzarini IF. Shape reconstruction of subcellular structures from live cell fluorescence microscopy images. J Struct Biol. 2009; 167:1–10. https://doi.org/10.1016/j.jsb.2009.03.017 PMID: 19358891

40. Proctor CJ, Söti C, Boys RJ, Gillespie CS, Shanley DP, Wilkinson DJ, et al. Modelling the actions of chaperones and their role in ageing. Mechanisms of Ageing and Development. 2005; 126(1):119–131. https://doi.org/10.1016/j.mad.2004.09.031 PMID: 15610770

41. Starruß J, de Back W, Brusch L, Deutsch A. Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology. Bioinformatiks. 2014; 30(9):1331–1332. https://doi.org/10.1093/bioinformatics/btt772

42. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, et al. COPASI: a COmplex PAthway SImulator. Bioinformatics. 2006; 22(24):3067. https://doi.org/10.1093/bioinformatics/btl485 PMID: 17032683

43. Proctor CJ. BIOMD0000000091—Proctor2005—Actions of chaperones and their role in ageing; 2005. http://www.ebi.ac.uk/biomodels-main/BIOMD0000000091.

44. Li C, Donizelli M, Rodriguez N, Dharuri H, Endler L, Chelliah V, et al. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. BMC systems biology. 2010; 4:92. https://doi.org/10.1186/1752-0509-4-92 PMID: 20587024

45. Erban R, Chapman SJ. Stochastic modelling of reaction—diffusion processes: algorithms for bimolecular reactions. Physical Biology. 2009; 6(4):046001. https://doi.org/10.1088/1478-3975/6/4/046001 PMID: 19700812