Modelling spatial heterogeneity and macromolecular crowding with membrane systems

Ettore Mosca[1], Paolo Cazzaniga[2], Dario Pescini[2], Giancarlo Mauri[2], and Luciano Milanesi[1]

[1] Institute for Biomedical Technologies, National Research Council, Via Fratelli Cervi, 20090 Segrate (MI), Italy
{ettore.mosca,luciano.milanesi}@itb.cnr.it
[2] Department of Informatics, Systems and Communications, University of Milan-Bicocca, Viale Sarca 336, 20126 Milan, Italy
{cazzaniga,pescini,mauri}@disco.unimib.it

Abstract. In biological processes, intrinsic noise, spatial heterogeneity and molecular crowding deeply affect the system dynamics. The classic stochastic methods lack of the necessary features needed for the description of these phenomena. Membrane systems are a suitable framework to embed these characteristics; in particular, the variants of τ-DPP and Sτ-DPP allow the modelling and stochastic simulations of multi-volume biochemical systems, in which diffusion and size of volumes and chemicals are taken into account improving the description of these biological systems. In this paper we show, by means of two models of reaction-diffusion and crowded systems, the correctness and accuracy of our simulation methods.

1 Introduction

Membrane systems [29] (also called P systems) have been recently exploited for the modelling of biological systems and for the investigation of their dynamical properties (we refer the reader to [32] for an updated bibliography). There are several important features that makes P systems suitable for the modelling of biological systems: a membrane structure is used to describe a compartmentalized environment, in which membranes are organized according to a specified hierarchy. Inside different membranes, different sets of objects (e.g. molecular species) can be defined along with sets of multiset rewriting rules (e.g. chemical reactions) to describe the evolution of the system.

On the other hand, there are some characteristics of the basic definition of P systems which are not adequate for the description of the biological reality. Among others, the maximal parallelism of the rules application, which consists in the application of all rules by consuming all objects present in the system, or the non deterministic selection of concurrent rules, are not suitable to the description of the stochasticity usually present in biological systems where some molecular

species occur in small quantities. In order to achieve a better description of biological systems, the variant of dynamical probabilistic P systems (DPPs) [31] has been introduced. In DPPs, the maximal parallelism has been mitigated by assigning probabilities to the rules, and these values vary according to the system state. By exploiting these values, it is possible to provide a description of the system's dynamics, that is, DPPs allow to reproduce the stochastic variations of the elements (i.e. chemical species) occurring in the system. However, this description is only *qualitative*, in the sense that an effective (physical) time streamline cannot be directly associated to the evolution steps of the system.

τ-DPP has been introduced to overcome the limitations of DPPs, providing a *quantitative* description of a system dynamics, by extending the single-volume algorithm of tau-leaping [8].

In order to correctly describe the behaviour of a system, τ-DPP runs in parallel inside each volume. A modified version of the tau-leaping procedure presented in [8] is exploited to compute the length of the step τ. In this novel version of the simulation algorithm, the least value for the time increment, among those computed inside each volume, is used to sample the number of reactions to execute (as in the original tau-leaping algorithm). Thanks to this "common" time increment, shared by all volumes, the simulation is synchronized at each step, allowing the correct passage of the molecules involved in communication rules.

A novel variant of τ-DPP, presented in [11], has been introduced to consider the size of volumes and objects involved in a system, in order to better describe systems where the "space" play an important role in the dynamics, such as crowded systems. This variant of our multi-volume stochastic simulation algorithm, called Sτ-DPP, is based on the same modelling framework of τ-DPP, and it exploits the same strategy for the simulation of the system's behaviour.

τ-DPP and Sτ-DPP algorithms can be used in the modelling and simulation of reaction–diffusion (RD) systems and crowded environments. RD systems are mathematical models used to describe those chemical systems for which the spatial distribution of chemicals influence the overall dynamics. The standard methods used to describe such systems are based on partial differential equations; however, when the intrinsic fluctuations of the chemical system play a major role in the dynamic, as in the case of many systems of interest for biology, a stochastic approach is more suitable. The intracellular environment is considered crowded since it is characterised by the presence of high concentrations of soluble and insoluble macromolecules; therefore, the classic approaches which consider molecules as points (without specifying their size), are no longer adequate. Under crowded conditions, the rate of some cellular processes can be increased or decreased, according to the "free space" in the system. By using Sτ-DPP, it is possible to consider the size of reaction volumes and chemicals, achieving a correct description of a crowded system by computing the reactions probability according to the free space occurring in a volume.

In this paper we show the simulation of the heat equation by means of τ-DPP, proving its correctness in the simulation of diffusive processes. To this aim, we

compare the results of the stochastic simulations with the exact solution of the heat equation. Afterwards, we present the results of the simulation of a crowded environment, showing how macromolecules affect the reactions probability and the overall system dynamics.

The paper is organised as follows. In the next section we give a description of spatial heterogeneity and macromolecular crowding in living cells, and we briefly present the classic computational approaches used in the description of these phenomena; in Section 3, we describe the stochastic simulation algorithms we have developed for the modelling and description of multi-volume systems. In order to validate our approaches for the simulation of RD systems and crowded environments, we present in Section 4 the results obtained from the simulation of the heat equation, and in Section 5 the simulation of a biological system with macro molecules that induce crowding effects. We conclude with some remarks about the results of the presented systems and with some possible future extensions.

2 Spatial heterogeneity and macromolecular crowding in living cells

Living cells are very far from the homogeneous and diluted compartment that is often used for their modelling. These requirements can be considered satisfied in many cases without taking them explicitly into account; however, there are several processes in which the effects of spatial heterogeneity (due to diffusive processes) and crowding (caused by the presence of macromolecules) must be considered in order to capture the correct system dynamics.

2.1 Reaction-diffusion systems

RD systems are mathematical models used to describe those chemical systems for which the spatial distribution of chemicals influences the overall dynamics. The standard approach exploits a continuous time and space domain description of the systems, such as the partial differential equations, where the mass transport, the chemical kinetics and the conservation laws, together with the boundary conditions, are embedded within the same set of equations that can be solved analytically or numerically.

When the intrinsic fluctuations of the chemical system play a major role in the dynamics, as is the case for many systems of interest for Biology, a master equation approach is more suitable[4, 20]. The chemical master equation formulation adopts a mechanistic perspective on the chemical system describing it as a sequence of collision events among molecules. Each of these scattering events can lead either to a new compound (reactive collision) or to an elastic scattering (diffusive collision) which does not alter the chemical species distributions but only the particles speed and direction. Which of the two collisions pathways will be followed by each scattering event is determined by the energy involved in the process: if this energy exceed the Arrhenius threshold (activation energy)

then the two molecules will react to form the new compound. According to this scheme it is possible to separate the RD process into a free flight phase followed by the interaction phase. The resulting dynamics is the superposition of a brownian motion (random walk) with an interaction/reaction process. The existence of an activation energy imposes that the diffusive events are the most probable ones, if the environment in which the reactions take place is homogeneous, this picture corresponds to a well stirred reactor and the dynamics can be tracked by means of a stochastic simulation algorithm such as the Gillespie's one [20].

A natural extension of the master equation approach to heterogeneous (space) systems consists of dividing the original volume V into smaller sub-volumes V_v, each one with a characteristic length h ($V_v = h^d$, being $d \in \{1, 2, 3\}$ the spatial dimensions), such that each of these sub-volumes can be considered homogeneous. Moreover, it is possible to define a mean jump frequency $\widetilde{D}_{i,v}$ [4] for each chemical species S_i in the sub-volume V_v, in order to connect the microscopic description of the master equation with the macroscopic Fick's diffusion coefficient D_i

$$\widetilde{D}_{i,v} = \frac{2d}{h_v^2} D_i, \tag{1}$$

and to verify if the well stirred condition still holds. The latter requirements is equivalent to impose in each sub-volume that the diffusion time $\tau_D \cong \frac{h^2}{2dD}$ is much smaller then the reaction waiting time τ_R [5], condition that should be also granted for the molecules diffusing across the sub-volumes. This observation allows to define the expression for the stochastic kinetic constant associated to the "diffusive" reactions

$$c_D = \frac{D}{h^2} \tag{2}$$

that mimic the molecules movement from one sub-volume to another.

2.2 Macromolecular crowding

The intracellular environment is characterised by the presence of high concentrations of soluble and insoluble macromolecules [16, 27, 38]. This medium is termed "crowded", "confined" or "volume-occupied", rather than "concentrated", because single molecular species may occur at low concentrations, but all species taken together occupy a considerable fraction of the total volume [28].

The term "macromolecular crowding" refers to the non-specific influence of steric repulsions (i.e., a consequence of the mutual impenetrability of molecules due to the Pauli exclusion principle) on molecular processes that occur in highly volume-occupied media [33].

Due to macromolecular crowding, biochemical, biophysical, and physiological processes in living cells may be quite different from those under idealized conditions [37], and order-of-magnitude effects of crowding have been demonstrated by both experimental and theoretical works on a broad range of processes [33]. All these effects are related to variations occurring in macromolecular thermodynamics activities [37] and diffusion [15].

To understand whether a crowded medium will increase or decrease the rate of a process, it is important to take into account the changes induced by the process itself on the available volume inside the system. Among others, binding of macromolecules to one another, folding of proteins and nucleic-acid chains into more compact shapes, the formation of aggregates, are all processes stimulated in crowding conditions due to the induced net increase of the available volume [15].

The other main effect of macromolecular crowding is related to anomalous diffusion. In crowded media, the mean squared displacement, $\langle r^2 \rangle$, of a solute particle in three dimensions is related to the diffusion coefficient D, but it is no longer linearly proportional to the time t:

$$\langle r^2 \rangle = 6Dt^{\alpha} \tag{3}$$

If $\alpha < 1$, the diffusion is called *anomalous subdiffusion*; on the other hand, if $\alpha > 1$ the diffusion is called *anomalous superdiffusion*; if $\alpha = 1$ the diffusion is normal. Crowding can reduce the rate of diffusion (according to the size of the diffusing molecule and to the degree of volume occupancy) and can lead to anomalous diffusion [3]. Large reductions in solute diffusion are probably indicators of interactions between the solute and cellular components, such as membranes [13]. Therefore, the rates of diffusion-controlled biochemical processes – mainly affected by the diffusion of the reactants – will be reduced in crowded media. The decrease in the diffusion rates due to crowding may also lead to complex phenomena like fractal kinetics (anomalous reaction orders and time-dependent reaction rate coefficients [24]) and spatial segregation of molecules [6]. In the latter case, as a consequence of the increased probability of recollision, crowding will determine the increase of the reaction rate of processes characterised by a low reaction probability [23].

2.3 Classic computational approaches

Computational approaches aimed at studying spatially heterogeneous systems and molecular crowding have to deal with the tabulation of spatial position of particles as a function of time. Several computational frameworks can be used to analyse such kind of systems [36]; we report hereafter the classic and most used methods.

Molecular dynamics (MD) simulations provide detailed trajectories, but they are computationally too expensive for simulating systems formed by a large number of atoms or with time scales above μs. MD has only been used in problems involving time-scales of ns and space-scales of tens of nm.

Brownian dynamics (BD) is a particle-based stochastic approach used to describe the time and space motion of molecules. Time and space are continuous, and noise is modelled by means of the Langevin equation. Crowded media can be explicitly described since it is possible to represent crowder molecules. However, as the number of particle collisions increases, the BD simulations demands a very

high computational cost. Examples of methodologies based on BD are Green's function reaction dynamics algorithm [39], Smoldyn [2] and MCell [35].

Partial differential equations (PDEs) are a continuous and deterministic approach. PDEs represent the classical method used to model RD systems. Each equation relates the time variation of a species concentration to its space variation and to the other species concentration. Crowding effects can be implicitly represented acting on diffusion coefficients (e.g., by lowering their values) and kinetic constants (e.g., by increasing their values). PDEs are usually solved using numerical methods (only in a few case the analytical solution is available); moreover, as the time-step and the sub-volume size (the space domain is usually divided in a number of elements) are reduced, the solutions becomes more accurate while the computational effort increases.

Cellular automata (CA) consist of a grid of cells (in any number of dimensions), each cell has a finite set of states, and it evolves according to the neighbours state. CA can be used to simulate RD systems at both microscopic and mesoscopic scales, depending on the number of molecules associated with each cell of the lattice. Crowding can be explicitly represented by considering crowder molecules or fixed barriers. For instance, people of the CyberCell project modelled a virtual cell membrane using discrete automata [7].

Lastly, spatial approaches based on Gillespie's method extend the stochastic simulation algorithm [21] in order to represent an RD system as a set of a well-stirred chemical reactors that communicate particles. The next sub-volume method [14], spatial τ-leaping [34] and the method described in [5] are algorithms that follow this approach; MesoRD [22] and SmartCell [1] are popular simulators. As molecules are considered point particles, these tools cannot be utilised to describe molecular crowding.

3 Multi-volume stochastic simulation algorithms based on P systems

The standard algorithms for the simulation of biochemical systems (see, for instance, the stochastic simulation algorithm (SSA) [18] and the next reaction method [17]) have been developed for the description of the exact behaviour of systems enclosed in a single volume. Recently, novel approaches have been introduced to simulate spatial heterogeneity, as in the next sub-volume method [14] or in the binomial τ-leap spatial stochastic simulation algorithm [25]. By using these methods, the volume of a system is divided into a number of separated sub-volumes, whose size is small enough to satisfy the requirements of the SSA, so that the probabilities of the reactions and the diffusive events occurring inside each sub-volume can be properly described.

A limitation of these stochastic methods consists in the fact that the size of chemicals and of the volumes in which reactions take place is not considered during the simulation of the system dynamics. To overcome the issues related to spatiality and size of chemicals, we have recently introduced two multi-volume stochastic algorithms called τ-DPP [9] and Sτ-DPP [11], which will be presented

in the next subsections. These methods combine a variant of P systems called *dynamical probabilistic P systems* [31] with the simulation method of tau-leaping [8]: the first one is suitable for the description of chemical, biological and ecological systems, and can be easily applied to the modelling and simulation of reaction-diffusion systems; the second one can be used for the simulation of crowded systems, since the size of chemicals and volumes is taken into account during the description of the system behaviour.

There exist different simulation methods based on P systems: among others we recall here the multicompartmental Gillespie's algorithm [30]. This algorithm is used to simulate systems composed by many volumes (also called compartments), exploiting the Gillespie's procedure. In particular, the direct method is used for the computation of the time τ and the index of the next reaction to execute within each compartment. This information is stored in a list which is updated at each iteration, modifying the values related to the compartments affected by the executed reaction. This strategy is very similar to that of the next subvolume method [14], with the difference that the multicompartmental Gillespie's algorithm does not use a particular data structure such as a heap or an indexed queue to efficiently handle compartments information.

In the description of biochemical systems modelled by means of this method, the so called *boundary rules* are exploited. Such kind of reactions are used to capture the features of the communication and the transformation of molecules. In particular, the authors consider special cases of boundary rules which involve molecules occurring in different compartments, though it is not very clear how Gillespie's theory for single volume systems can be used to describe the propensity functions of reactions that are simultaneously active in two compartments, nor in which compartment this information is used to compute the value of τ.

τ-DPP

τ-DPP [11] is a computational method which can be used to describe and perform stochastic simulations of complex biological or chemical systems. The "complexity" of the systems that can be managed by means of τ-DPP, resides both in the number of the (chemical) reactions and of the species involved, and in the topological structure of the system, that can be composed by many volumes. For instance, cellular pathways involving several spatial compartments (as the extracellular ambient, the cytoplasm, the nucleus, etc.), or multicellular systems like bacterial colonies, or multi-patched ecological systems as metapopulations, are all examples of complex systems that could be investigated with τ-DPP.

The correct behaviour of the whole system is achieved by letting all volumes evolve in parallel, and by using the following strategy for the choice of time increments. At each iteration of τ-DPP, we consider the current state of each volume (determined by the current number of molecules), and we calculate a time increment independently in each volume (according to the standard tau-leaping algorithm [8]). Then, the smallest time increment is selected and used to evaluate the next-step evolution of the entire system. Since all volumes *locally* evolve according to the same time increment, τ-DPP is able to correctly work

out the *global* dynamics of the system. Moreover, by adopting this procedure, the simulated evolutions of all volumes get naturally *synchronized* at the end of each iterative step. The synchronization is also necessary – and exploited together with a parallel update of all volumes – to manage the communication of molecules among volumes (i.e., diffusive events), whenever prescribed by specific (communication) rules.

The system is defined by means of a set of N volumes organised according to the hierarchy specified by the membrane structure. The state of the whole system is characterised by all multisets M_v occurring inside each volume V_v $(1 \leq v \leq N)$.

Inside the volumes, the sets of rules $R_1, \ldots, R_N$ are defined along with the sets of stochastic constants $C_1, \ldots, C_N$.

Each volume V_v can contain two different kinds of rules, termed *internal* and *communication* rules. An internal rule describes the modification, or evolution, of the objects inside the single volume where it is applied, while a communication rule sends the objects from the volume where it is applied to an adjacent volume (possibly modifying the form of these objects during the communication step).

More precisely, internal rules have the general form $\alpha_1 S_1 + \alpha_2 S_2 + \cdots + \alpha_m S_m \rightarrow \beta_1 S_1 + \beta_2 S_2 + \cdots + \beta_m S_m$, where $S_1, \ldots, S_m$ belong to the set of distinct object types $\mathcal{S}$, and $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m \in \mathbb{N}$. For instance, $S_1, \ldots, S_m$ can correspond to molecular species, and, in this case, $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m$ represent stoichiometric coefficients. The objects appearing in the left-hand side of the rule are called *reagents*, while the objects on the right-hand side are called *products*. Note that, usually, we will consider the case where (at most) three objects appear in the reagents group. The rational behind this is that we require biochemical reactions to be (at most) of the third-order, since the simultaneous collision and chemical interaction of more than three molecules at a time, has a probability to occur close to zero in real biochemical systems. Moreover, the interaction among more than three molecules can be modelled by using a set of successive reactions with lower order. In what follows, we will refer to rules or reactions without distinction.

When dealing with communication rules inside a volume, besides defining the sets of reagents and products, it is necessary to specify the target volume where the products of this rule will be sent[3]. Formally, a communication rule has the form[4] $\alpha_1 S_1 + \alpha_2 S_2 + \cdots + \alpha_m S_m \rightarrow (\beta_1 S_1 + \beta_2 S_2 + \cdots + \beta_m S_m, tar)$, where $S_1, \ldots, S_m \in \mathcal{S}$ are distinct object types, $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m \in \mathbb{N}$, and *tar* represents the volume where the products of the reaction diffuse.

A complete an extensive description of the τ-DPP algorithm and some applications can be found in [9, 10].

---

[3] This definition can be easily extended in order to assign a different target volume to each object appearing in the set of products.

[4] The condition that at most three objects appear as reagents is usually required also for communication rules.

Sτ-DPP

Sτ-DPP [11] is obtained combining the structure definition of tissue P systems [26] with the simulation strategy used in τ-DPP [9]. Here, nodes are arranged in a tissue–like fashion, but each of them can have a complex internal hierarchy, organised in a tree–like structure. Moreover, in this new variant we consider sizes for both membranes and objects, and the rules defined inside each membrane will be enabled only in the case there is sufficient free space in the membrane where the rule is applied, for instance, to "create" new objects or to receive objects from other volumes. The size considered here can be used in the modelling and simulation of biochemical systems where diffusive processes play an important role, and it is necessary to avoid the unlimited accumulation of objects in a region of finite size.

In order to correctly describe the hierarchy of complex nodes of the system we first need a graph representing the topology of the membranes. In particular, this graph can have undirected edges to indicate that two membranes are placed at the same hierarchical level (as in the standard definition of tP systems [26]). On the other hand, directed edges of the graph are used to denote that the "source" membrane contains the "target" membrane.

A second directed graph is needed to represent the communication channels among membranes. Clearly, the arrows of the edges indicate the direction of the (permitted) flow of objects between different compartments. Note that, this communication graph can contain edges that are not indicated in the graph which describes the membrane structure. The meaning of these particular edges is to represent communication channels that connect non adjacent membranes. Using these arcs, it is then possible to create privileged pathways of communication between membranes.

The features of Sτ-DPP can be exploited to represent (among other real life systems) reaction–diffusion systems [12]. In this case, the membrane structure can be used to represent a reaction volume as the composition of a number of finite size sub-volumes, and the communication graph will describe the diffusion directions through the system.

With respect to the definition of τ-DPP, when using this simulation strategy we need to specify additional information about the system under investigation. First, we need to provide the sizes of the volumes composing the system and of the molecular species occurring within them. Second, we have to compute the initial free space of each volume.

Given the internal state M_v of a membrane V_v together with size $s_{S_1}, \ldots, s_{S_m}$ of the species, and the size $s_{V_1}, \ldots, s_{V_N}$ of the volumes, it is possible to define the *occupied volume* in V_v as:

$$O(V_v) = \sum_{i=1}^{m} (m_i \cdot s_{S_i}) + \sum_{V_l \in a_{\mathcal{T}}(V_v)} s_{V_l}$$

where m_i represents the amount of the species S_i and $a_{\mathcal{T}}(V_v)$ is the set of volumes contained within the volume V_v. Hence, it is possible to define the value of the

free space in V_v as:

$$F(V_v) = s_{V_v} - O(V_v)$$

Internal and communication rules are defined as in τ-DPP, but, at each rule execution (internal or communication), apart from updating the molecular amounts of the species involved in the reaction, also the free space value has to be updated. The update operation adds to the free space value the "volume" of the objects consumed or sent by the rule and subtracts the "volume" of the objects produced by the rule or received from other membranes. In particular, after the execution of an internal rule, the free space in V_v is updated as $F(V_v) = F(V_v) - \sum_{i=1}^{m}(\beta_i - \alpha_i) \cdot s_{S_i}$. On the contrary, when a communication rule is applied, we need to update the free space of V_v (i.e., the membrane where the reaction is applied) as $F(V_v) = F(V_v) + \sum_{i=1}^{m} \alpha_i \cdot s_{S_i}$ and the free space of each target volume V_{tgt_k} indicated by the rule: $F(V_{tgt_k}) = F(V_{tgt_k}) - \sum_{i=1}^{m} \beta_{i,k} s_{S_i}$.

At each iteration of the algorithm, in order to obtain a correct description of the system's dynamics, we need to check if a rule r (internal or communication) is applicable. A complete description of Sτ-DPP can be found in [11]; some additional details about the implementation of the algorithms will be provided in the next section.

Handling diffusive events and crowding in τ-DPP and Sτ-DPP algorithms

The iterations of the simulation algorithms described in the previous sections are composed of three main stages: (i) computation of the reactions probability; (ii) calculation of the τ value; (iii) selection of the reactions to execute and check of the system state consistency. During these stages we have to keep into account some details in order to correctly describe diffusive events and the effect due to crowded media, as described in what follows.

In the first stage, given the system state $\mathbf{x}$, the probability $a(\mathbf{x})$ of a rule application (i.e. the propensity function) is generally computed as follows: $a(\mathbf{x}) = c \cdot h$, where c is the stochastic constant associated to the rule and h is a combinatorial function depending on the left-hand side of the rule [18]. This definition is used in τ-DPP to compute the propensity functions of each volume of the system. Note that, this operation is performed independently in each volume; hence, the propensity function of the reactions occurring inside a volume V_v depend only on its current state (defined as M_v).

For what concerns the Sτ-DPP algorithm, the propensity functions of the internal reactions are computed by also considering the value of the free space of the current volume. So doing, we can correctly simulate crowded systems: we suppose that while first order reactions (e.g. $a \rightarrow b$) are not affected by the value of the free space, in the case of reactions of higher orders, the lack of free space enhances the reaction probability. Therefore, the propensity functions of second and third order reactions are computed as follows:

$$a(\mathbf{x}) = \frac{c \cdot h}{F(V)}. \tag{4}$$

On the other hand, communication rules representing diffusive events are not influenced by the free space left in the volume; therefore, their propensity functions are computed as in the standard procedure.

During the second phase of the algorithms, inside each volume of the system, a candidate length τ of a step is obtained by using the following equation:

$$\tau = \min_{i \in \mathcal{S}} \left\{ \frac{max\{\varepsilon m_i / g_i, 1\}}{|\mu_i(\mathbf{x})|}, \frac{max\{\varepsilon m_i / g_i, 1\}^2}{\sigma_i^2(\mathbf{x})} \right\},$$

where g_i is a value depending on the highest order of reaction in which a species $i \in \mathcal{S}$ is involved and ε is an error control parameter, while $\mu_i(\mathbf{x})$ and $\sigma_i^2(\mathbf{x})$ are calculated as described below (according to the definition presented in [19]):

$$\mu_i(\mathbf{x}) = \sum_{j \in R_{ncr} \cap R_{int}} (v_{i,j}\ a_j(\mathbf{x})) + \sum_{j \in R_{ncr} \cap R_{comm}} (-lhs_{i,j}\ a_j(\mathbf{x}))\,, \quad \forall i \in \mathcal{S},$$

$$\sigma_i^2(\mathbf{x}) = \sum_{j \in R_{ncr} \cap R_{int}} \left(v_{i,j}^2\ a_j(\mathbf{x})\right) + \sum_{j \in R_{ncr} \cap R_{comm}} \left(-lhs_{i,j}^2\ a_j(\mathbf{x})\right), \quad \forall i \in \mathcal{S},$$

where j belongs to the set R of reactions of the considered volume, the restriction on the set of noncritical reactions R_{ncr} is present, due to the conditions of the modified non-negative Poisson tau-leaping [8], while R_{int} represents the set of internal rules and R_{comm} the set of communication rules. During the computation of μ_i and σ_i^2, we consider the variation of the species i due to the reaction j (specified by the value $v_{i,j}$), for what concerns internal rules. On the other hand, we only consider the variation of the species i described by the left-hand side of a communication rule j ($lhs_{i,j}$), since the current volume is affected only by these variations.

There exists different approaches in which communication rules (i.e. diffusive processes) are considered as special events and their probabilities are computed by using a deterministic formulation, and during the calculation of μ_i and σ_i^2 besides the changes of the species due to the rule, also the contribution of the neighbourhood volumes is taken into consideration (see, for instance, [34]).

In the last part of the algorithms, the set of reactions to be executed inside each volume is selected, and before the system update, the applicability of this set has to be verified in order to obtain a consistent system state.

Both τ-DPP and Sτ-DPP select the number of occurrences of each reaction j (inside each volume) by sampling a random number from a Poisson distribution having mean and variance equal to $a_j(\mathbf{x})\tau$. Afterwards, the applicability of the set of selected reactions is verified: both algorithms check if the system state resulting from the execution of the reactions contains negative values for the amount of some molecular species. Moreover, in the case of Sτ-DPP, the set of reactions selected in a volume is applicable only if there is enough free space after the rules application [11].

As stated above, a rule can be executed only if the free space of the volume, after the rule application, is greater or equal to zero. The rule applicability is com-

puted differently for internal and communication rules. Given an internal rule occurring inside volume V_v, we need to check if, after the rule execution, $F(V_v) \geq 0$. For what concerns a communication rule r, we also need to check all the free space of all the target volumes indicated by the rule: $\forall\ tgt_l$ of r, $F(V_{tgt_l}) \geq 0$, where the values β_j are the stoichiometric coefficients of the molecular species associated with V_{tgt_l}.

Note that, using a strategy based on the tau-leaping algorithm to describe the behaviour of the system, at each iteration step a certain number of rules is applied in parallel. Hence, the applicability of the entire set of selected rules has to be verified. This operation is realised by computing the free values of each volume V_v considering the contribution of all the selected rules; if the values of $F(V_v)$ is greater or equal to zero (for each volume), then the execution is allowed.

If any of these requirements is not satisfied, then the value of τ is reduced by half and a new set of reactions is selected (this strategy has been proposed in [8]). On the other hand, the iteration of the τ-DPP and Sτ-DPP algorithms proceeds by updating the system state and the simulation time; and the simulation finishes if a termination criterion is reached.

4 Validation of the diffusion implemented with τ-DPP

In this section we present some results obtained from simulations performed by using τ-DPP in order to verify if diffusion is correctly handled by our simulation algorithm. Berstein [5] showed that it is possible to simulate mesoscopic RD systems using the Gillespie's algorithm comparing the simulations with the solution of diffusion equations. We adopted the same strategy in order to show that τ-DPP can be used to reproduce diffusion introducing a reasonably small error.

4.1 A popular diffusion equation: the heat equation

The heat equation is a partial differential equation which describes the heat distribution in a region during time, and it is a special case of diffusion equation where the diffusion coefficient D is constant in time and space:

$$\frac{\partial u(\overrightarrow{x}, t)}{\partial t} = D\Delta u(\overrightarrow{x}, t) \tag{5}$$

where $u(\overrightarrow{x}, t)$ is the density of the diffusive material in $\overrightarrow{x}$ at time t, and $\Delta = \frac{\partial^2}{\partial x} + \frac{\partial^2}{\partial y} + \frac{\partial^2}{\partial z}$ is the Laplacian operator.

To test the accuracy of τ-DPP in reproducing diffusion, we studied the unidimensional diffusion of the molecule S in the region $\Omega \subset \mathbb{R}$:

$$\frac{\partial [S]}{\partial t}(x, t) = D\frac{\partial^2}{\partial x}[S](x, t),\ \forall x \in \Omega \tag{6}$$

where $[S](x, t)$ indicates the concentration of molecule S in position x at time t. In particular, we considered the region $\Omega = [0, 1]$ and the Neumann boundary conditions:

$$\frac{\partial[S]}{\partial x}(0,t) = \frac{\partial[S]}{\partial x}(1,t) = 0, \tag{7}$$

indicating that the flux from outside into Ω is null. Considering $D = 1$, an exact solution in the region Ω satisfying Eq. 7 is:

$$[\mathrm{S}](x,t) = S^{\Omega}[1 + \tfrac{1}{2}\mathrm{e}^{-\pi^2\gamma^2 t}\cos(\gamma\pi x)] \tag{8}$$

where S^{Ω} is the total number of S molecules inside the system and γ is a non negative integer. In all the cases that we will discuss in the next section we have considered $\gamma = 3$ and $S^{\Omega} = 500$.

4.2 Comparison between τ-DPP and the heat equation

In order to compare the simulations performed by using τ-DPP with the continuous exact solution of the heat equation (Eq. 8) at a given time instant, we dived the region Ω into N smaller adjacent regions V_v such that $\Omega = \bigoplus_v V_v$.

A series of issues have to be handled to realize a meaningful comparison. First, since τ-DPP algorithm is stochastic, it is crucial to consider a high number of simulations in order to obtain a significant comparison with the heat equation. We accomplished this task averaging the results of a sufficiently high number G of simulations. Note that, in general, this average is not representative of the final state of a system, like in the case of multistable systems, in which averaging may lead to fictitious states. However, when the average solution converges to the system state – as in the case we are considering here – the deviations from the exact solution can be considered as a type of sampling error and the average solution is a good representative of the system state.

Second, since τ-DPP simulator works with molecules, rather than molecule concentration, we must calculate the initial distribution of S molecules within V_v, i.e., $S_1, \ldots, S_N$, from the solution of Eq. 8 at time $t = 0$, in order to use this distribution as input for τ-DPP. Moreover, we must calculate τ-DPP predicted concentrations, $[S^*]_1, \ldots, [S^*]_N$, from the distribution of S molecules over the sub-volumes at a particular time point. These conversions have been defined according to the following relation between concentration $[S]$ and molecule number S:

$$[S_v] = \frac{S_v}{s_{V_v}},\ \forall v \tag{9}$$

Note that the solution of Eq. 8 must be calculated using the appropriate vector $\boldsymbol{x} = (x_1, \ldots, x_v, \ldots, x_N)$, whose members are located at the middle of each volume s_{V_v}:

$$x_v = \frac{s_{V_v}}{2} + (v-1)s_{V_v},\ \forall x_v \in \boldsymbol{x} \tag{10}$$

Third, since the time increments τ are randomly generated, it is very unlikely that the simulator will output a numerical solution exactly at a specific time

point t. Therefore, the molecule distribution computed by τ-DPP at a particular time point t has been calculated as a linear interpolation of the two numerical solutions at $t_1 < t$ and $t_2 > t$, where t_1 and t_2 are, respectively, the points computed by the τ-DPP that immediately precede and follow t. An example of comparison between the heat equation and the τ-DPP simulations is shown in Fig. 1, where it is possible to observe the high closeness between the simulations and exact solution.

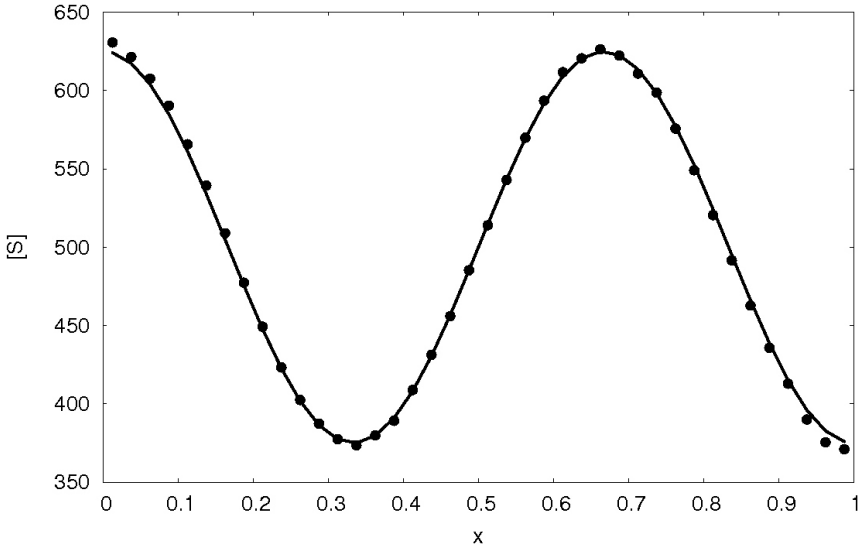


Fig. 1. The heat equation exact solution (line) and τ-DPP average results (dots) at $t = 0.0078034$, $S^{\Omega} = 500$, $G = 10000$, $\gamma = 3$, $s_{V_v} = 0.025$.

Quantitatively, the quality of the τ-DPP simulation $e = 1, \ldots, G$ has been assessed considering, as in [5], the error due to the difference between the exact solutions $[S_{v,e}]$ and the concentrations computed using the numerical results of τ-DPP $[S^*_{v,e}]$:

$$\epsilon_v = \frac{1}{G} \sum_{e=1}^{G} (1 - \frac{[S^*_{v,e}]}{[S_v]}) \tag{11}$$

in the volume v, considering a pool of G simulations ran with the same settings. Note that $\epsilon_v \to 0$ as $[S^*_{v,e}] \to [S_{v,e}]$ (obviously) and in the case in which the distribution of $[S^*_{v,e}]$ is symmetric with respect to the exact value $[S_v]$. The errors ϵ_v have been averaged considering all the elements of Ω:

$$\bar{\epsilon} = \frac{1}{N} \sum_{v=1}^{N} \mid \epsilon_v \mid \tag{12}$$

We studied the relation of $\bar{\epsilon}$ with the number of simulations $10 \leq G \leq 10000$ and the number of volumes $10 \leq N \leq 40$ used to discretise the spatial region Ω

$$s_{V_v} = \frac{\Omega}{N}. \tag{13}$$

As the number of simulations increases the sampling error decreases (as shown in Fig. 2). In particular, we observed a decrease of one order of magnitude passing from 10 to 10^4 simulations in all cases with exception of $s_{V_v} = 0.1$ ($N = 10$), where the decrease has been lower. Note that as $G \to 0$ the lowest error is associated to settings with a higher s_{V_v}, while as $G \to \infty$ the lowest error is associated to lower s_{V_v}. This can be attributed to the noise: as $s_{V_v} \to 0$, the volumes will contain a lower number of molecules (high noise); hence, a higher number of simulations is required to eliminate noise. This phenomenon has been particularly evident in the study presented here due to the relatively low quantity of molecules used, $S^{\Omega} = 500$, with respect to the number of volumes for the discretisation of Ω, $10 \leq N \leq 40$: we passed from $S_v \in [10, 10^2]$ when $s_{V_v} = 0.1$ ($N = 10$) to $S_v \in [1, 10]$ when $s_{V_v} = 0.025$, ($N = 40$).

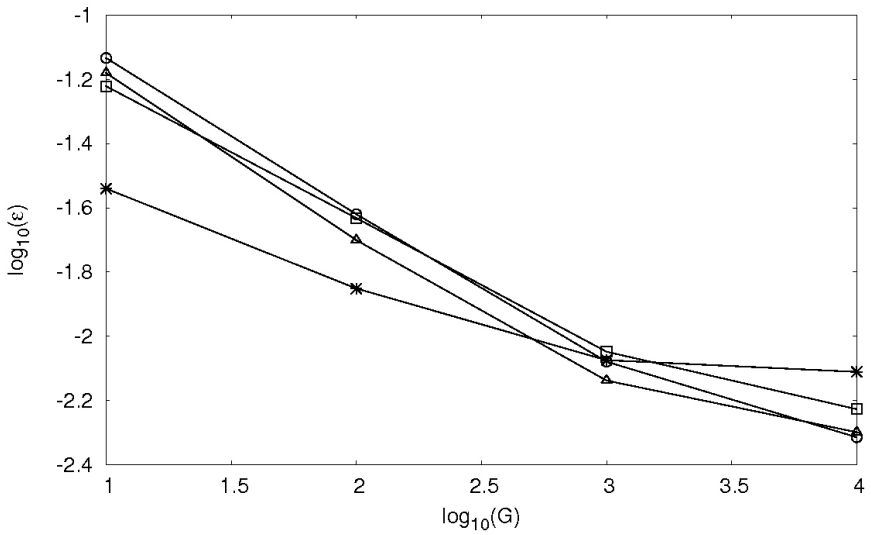


Fig. 2. Relation between the error ϵ and the number of simulations (G). $s_{V_v} = 0.1$ (*), $s_{V_v} = 0.05$ ($\square$), $s_{V_v} = 0.03\bar{3}$ ($\triangle$), $s_{V_v} = 0.025$ ($\circ$), $\gamma = 3$, $t = 0.0078034$.

Another source of error is associated with the spatial discretisation, i.e. with the number of membranes in which Ω is divided into. In order to reduce the contribution of the sampling error it is crucial to study the behaviour of the spatial discretisation error with a high G. This error decreases as $s_{V_v} \to 0$ (Fig. 3).

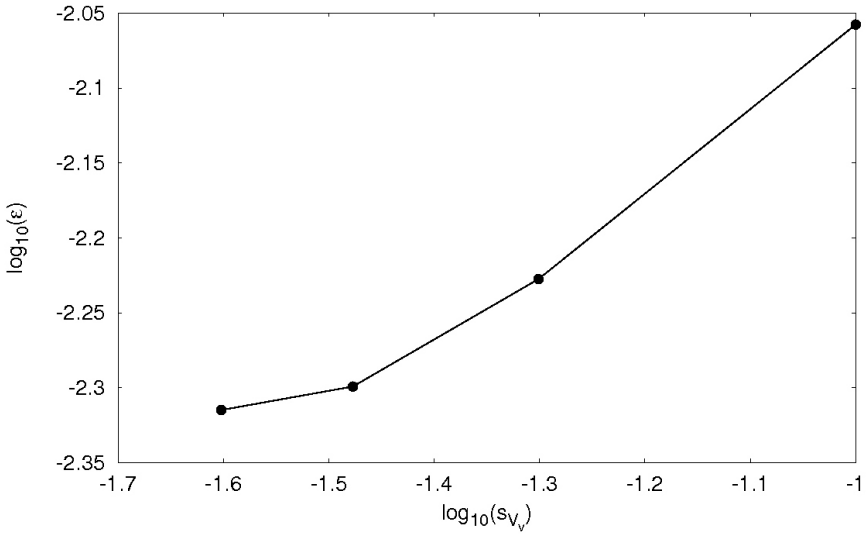


Fig. 3. Relation between the spatial discretisation error and the volumes size at time $t = 0.0078034$.

5 Macromolecular crowding with Sτ-DPP

Sτ-DPP allows to model objects of arbitrary size that react and diffuse through a spatial region composed by the union of a set of membranes of arbitrary size; therefore, Sτ-DPP can be used to simulate crowded RD systems. In a previous work, we have shown that Sτ-DPP are able to reproduce systems in which crowding determines a slower motion of molecules [11].

In this section we show the effects of macromolecular crowding which increases a biochemical reaction rate due to the increase of reaction collision probabilities, that is in turn determined by the reduction of the free space induced by crowding. In particular, the reduction of the space in a volume determines the increase of the propensity functions of second and third order reactions, as described in Section 3.

We accomplished this task by modelling a crowded RD system composed by four species, A, B, C, Z, within the spatial region $\Omega \subset \mathbb{R}^2$, represented using a set of 81 membranes organised as a 9x9 bidimensional lattice, and two biochemical reactions:

$$r_1 : \mathrm{A} + \mathrm{B} \xrightarrow{c_1} \mathrm{C} \tag{14}$$

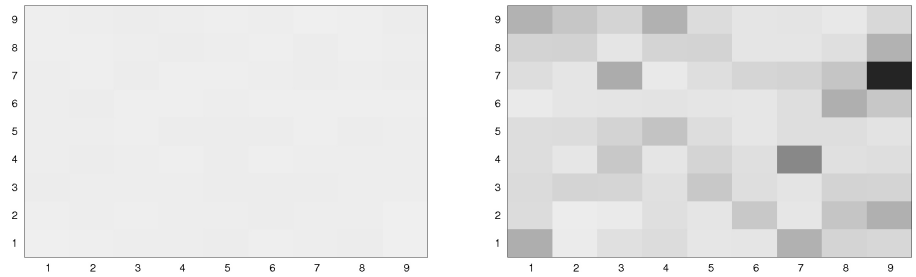
$$r_2 : \mathrm{C} \xrightarrow{c_2} \mathrm{A} + \mathrm{B} \tag{15}$$

where $c_1 = 0.0001$ and $c_2 = 0.001$ are the stochastic constants. Stochastic constants for diffusion rules have been set three orders of magnitude higher than

c_1, c_2, that is $c_{D_A} = c_{D_B} = c_{D_C} = 0.16$, in order to ensure $\tau_D \ll \tau_R$. The size of species Z is three orders of magnitudes bigger than that of the other species ($s_D = 0.1$, $s_A = s_B = 0.0001$ and $s_C = 0.0002$) and it has been used to represent macromolecular crowding. The simulation has been done considering 100 molecules of A and 100 molecules of B in each membrane while molecule Z have been randomly distributed among the membranes; the number of Z has been chosen in order to occupy 0 (diluted media) or approximately $\frac{1}{3}$ (crowded media) of the total volume. Note that, while molecules A, B, C can diffuse to each first next neighbour, molecules Z do not diffuse.

For each membrane we determined the average value assumed by the propensity function associated with the reaction defined in Eq. 14 (i.e. production of C), during the simulation in the time interval $[0, 10]$.

In the diluted condition the reaction propensity assumes approximately the same value within all membranes, with mean value $\mu = 0.9237$ and standard deviation $\sigma = 0.001842$ (Fig. 4(a)). The distribution of the propensity functions values has been affected by the addiction of crowding molecules: in this condition we obtained mean $\mu = 1.451$ and standard deviation $\sigma = 0.498$ (Fig. 4(b)). The average reaction propensity values in the crowded media are up to 6 folds higher then in the diluted case, and their homogeneity over the system is affected by the presence of highly reactive volumes in which the production of molecules C is faster, as shown in Fig. 5.



(a) (b)

Fig. 4. Average propensity function values of reaction r_1 within the spatial domain in (a) diluted ($\mu = 0.9237, \sigma = 0.001842$) and (b) crowded conditions ($\mu = 1.451, \sigma = 0.498$).

6 Conclusions

In this work we examined the application of two membrane systems variants, τ-DPP and Sτ-DPP, in order to model and simulate spatial heterogeneity and molecular crowding, two important characteristics of living cells.

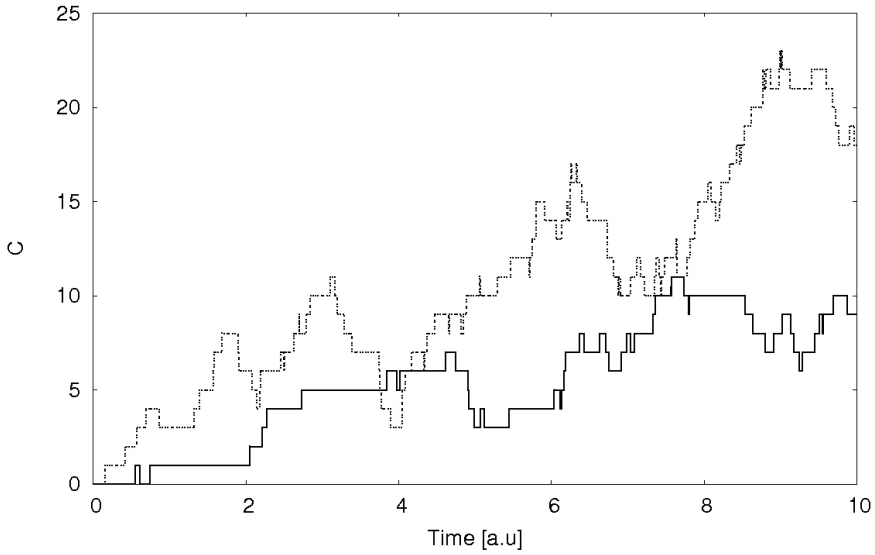


Fig. 5. Number of C molecules in the crowded volume 78 (row 9 and column 7 of Fig. 4) in diluted (solid line) and crowded (dashed line) conditions.

To capture spatial heterogeneity at the molecular level, a spatial domain must be defined and the time evolution of possibly reacting molecules must be tracked in different locations of the domain. We showed that τ-DPP can reproduce the diffusion of molecules within a spatial region divided in a set of sub-volumes (membranes) that can move (communicate) objects, according to a defined topology.

We tested diffusion using the same strategy followed by Bernstein [5], i.e., we compared the τ-DPP simulations with an analytical solution of a PDE for the heat equation (a diffusion equation). Note that for most applications of real interest biological processes analytical solutions are hardly available.

The error that we reported is slightly greater than the one found in [5]. This is due to the fact that the τ-leaping algorithm – which stands at the basis of τ-DPP – generates an approximate dynamics with respect to the exact solution of the chemical master equation, whereas the Gillespie's algorithm used in [5] is exact. We think that this loss of accuracy (that can be a priori controlled) is well balanced by the increase in performance that enables simulations of more complex systems. The quantitative characterisation of this discrepancy is currently under investigations and it will be published in a future work.

Molecular crowding can be explicitly modelled by the Sτ-DPP variant, an extension of τ-DPP, in which - among other features - sizes are associated to objects and membranes. We have previously shown that Sτ-DPP captures the delay in the communication of objects between volumes due to crowded conditions [11]. In this contribution we showed that Sτ-DPP can also be used to

reproduce the reaction rate increase observed in crowded media due to the increase of recollision probability determined by the reduction of the available free space in a volume. This effect has been captured by modifying the propensity function calculation of second and third order reactions, and it has been shown with an example concerning a crowded RD system in a bidimensional region.

Considering the results provided in this work, Sτ-DPP can be successfully used to model and simulate crowded RD systems. It is worthy to note that, the current version of Sτ-DPP has all the features to capture the major effects that a crowded medium determines over RD system dynamics, and we plan to do an extensive study of this topic in future. Moreover, an interesting direction of investigation, enabled by the possibility of arbitrarily defining the volumes size and the topology of their communication, consists in the use of Sτ-DPP to study RD systems dynamics in structured regions (as the cytoplasm of living cells).

7 Acknowledgements

This work has been supported by the NET2DRUG, EGEE-III, BBMRI, EDGE European projects and by the MIUR FIRB LITBIO (RBLA0332RH), ITALBIONET (RBPR05ZK2Z), BIOPOPGEN (RBIN064YAT), CNR-BIOINFORMATICS initiatives and FAR-09 "Modelli di calcolo naturale e applicazioni" Italian initiatives.

References

1. Ander, M., Beltrao, P., Ventura, B.D., Ferkinghoff-Borg, J., Foglierini, M., Kaplan, A., Lemerle, C., Toms-Oliveira, I., Serrano, L.: Smartcell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. Systems Biology 1(1), 129–138 (2004)
2. Andrews, S.S., Addy, N.J., Brent, R., Arkin, A.P.: Detailed simulations of cell biology with smoldyn 2.1. PLoS Computational Biology 6(3), e1000705 (2010)
3. Banks, D.S., Fradin, C.: Anomalous diffusion of proteins due to molecular crowding. Biophysical Journal 89(5), 2960–2971 (2005)
4. Baras, F., Mansour, M.: Reaction-diffusion master equation: A comparison with microscopic simulations. Physical Review E 54(6), 6139–6148 (1996)
5. Bernstein, D.: Simulating mesoscopic reaction-diffusion systems using the gillespie algorithm. Physical Review E 71(4 Pt 1), 041103 (2005)
6. Berry, H.: Monte carlo simulations of enzyme reactions in two dimensions: fractal kinetics and spatial segregation. Biophysical Journal 83(4), 1891–1901 (2002)
7. Broderick, G., Ru'aini, M., Chan, E., Ellison, M.J.: A life-like virtual cell membrane using discrete automata. In Silico Biology 5(2), 163–178 (2005)
8. Cao, Y., Gillespie, D.T., Petzold, L.R.: Efficient step size selection for the tau-leaping simulation method. Journal of Chemical Physics 124, 044109 (2006)
9. Cazzaniga, P., Pescini, D., Besozzi, D., Mauri, G.: Tau leaping stochastic simulation method in p systems. In: Hoogeboom, H. J.; Paun, G.; Rozenberg G.; Salomaa.A. (eds.) Proc. of the 7th International Workshop on Membrane Computing. LNCS vol. 4361, pp. 298–313 (2006)

10. Cazzaniga, P.: Stochastic algorithms for biochemical processes. Ph.D. thesis, Università degli Studi di Milano-Bicocca (2010)
11. Cazzaniga, P., Mauri, G., Milanesi, L., Mosca, E., Pescini, D.: A novel variant of tissue p systems for the modelling of biochemical systems. In: Paun, G., Perez-Jimenez, M., Riscos-Nunez, A., Rozenberg, G., Salomaa, A. (eds.) Proc. of the 10th International Workshop on Membrane Computing. LNCS vol. 5957, pp. 210–226 (2009)
12. De Wit, A.: Spatial patterns and spatiotemporal dynamics in chemical systems. Advances in Chemical Physics 109, 435–513 (1999)
13. Dix, J.A., Verkman, A.S.: Crowding effects on diffusion in solutions and cells. Annual Review of Biophysics 37, 247–263 (2008)
14. Elf, J., Ehrenberg, M.: Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. IEE Proceedings Systems Biology 1(2), 230–236 (2004)
15. Ellis, R.J., Minton, A.P.: Cell biology: join the crowd. Nature 425(6953), 27–28 (2003)
16. Fulton, A.B.: How crowded is the cytoplasm? Cell 30(2), 345–347 (1982)
17. Gibson, M., Bruck, J.: Efficient exact stochastic simulation of chemical systems with many species and many channels. Journal of Physical Chemistry A 104(9), 1876–1889 (2000)
18. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry 81(25), 2340–2361 (1977)
19. Gillespie, D.T., Petzold, L.R.: Improved leap-size selection for accelerated stochastic simulation. Journal of Chemical Physics 119, 8229–8234 (2003)
20. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics 22(4), 403–434 (1976)
21. Gillespie, D.T.: Stochastic simulation of chemical kinetics. Annual Review of Physical Chemistry 58, 35–55 (2007)
22. Hattne, J., Fange, D., Elf, J.: Stochastic reaction-diffusion simulation with mesoRD. Bioinformatics 21(12), 2923–2924 (2005)
23. Kim, J.S., Yethiraj, A.: Effect of macromolecular crowding on reaction rates: a computational and theoretical study. Biophysical Journal 96(4), 1333–1340 (2009)
24. Kopelman, R.: Fractal reaction kinetics. Science 241(4873), 1620–1626 (1988)
25. Marquez-Lago, T.T., Burrage, K.: Binomial tau-leap spatial stochastic simulation algorithm for applications in chemical kinetics. Journal of Chemical Physics 127(10), 104101 (2007)
26. Martín-Vide, C., Pazos, J., Păun, G., Rodríguez-Patón, A.: A new class of symbolic abstract neural nets: Tissue p systems. In: COCOON '02: Proc. of the 8th Annual International Conference on Computing and Combinatorics. pp. 290–299. Springer-Verlag, London, UK (2002)
27. Minton, A.P.: The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. Molecular and Cellular Biochemistry 55(2), 119–140 (1983)
28. Minton, A.P.: The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. Journal of Biological Chemistry 276(14), 10577–10580 (2001)
29. Păun, G.: Membrane Computing. An Introduction. Springer-Verlag, Berlin (2002)
30. Pérez-Jiménez, M.J., Romero-Campero, F.J.: Transactions on Computational Systems Biology VI, chap. P Systems, a New Computational Modelling Tool for Systems Biology, LNCS vol. 4220, pp. 176–197. Springer Berlin / Heidelberg (2006)

31. Pescini, D., Besozzi, D., Mauri, G., Zandron, C.: Dynamical probabilistic P systems. International Journal of Foundations of Computer Science 17(1), 183–204 (2006)
32. The P systems web page, http://ppage.psystems.eu
33. Rivas, G., Ferrone, F., Herzfeld, J.: Life in a crowded world. EMBO Reports 5(1), 23–27 (2004)
34. Rossinelli, D., Bayati, B., Koumoutsakos, P.: Accelerated stochastic and hybrid methods for spatial simulations of reaction-diffusion systems. Chemical Physics Letters 451(1-3), 136–140 (2008)
35. Stiles, J.R., Bartol, T.M.: Computational Neuroscience: Realistic Modeling for Experimentalists, chap. Monte Carlo Methods for Simulating Realistic Synaptic Microphysiology Using MCell, pp. 87–127. CRC Press (2001)
36. Takahashi, K., Arjunan, S.N.V., Tomita, M.: Space in systems biology of signaling pathways–towards intracellular molecular crowding in silico. FEBS Letters 579(8), 1783–1788 (2005)
37. Zimmerman, S.B., Minton, A.P.: Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annual Review of Biophysics and Biomolecular Structure 22, 27–65 (1993)
38. Zimmerman, S.B., Trach, S.O.: Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of escherichia coli. Journal of Molecular Biology 222(3), 599–620 (1991)
39. van Zon, J.S., ten Wolde, P.R.: Green's-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space. Journal of Chemical Physics 123(23), 234910 (2005)